Short title: Control of isometric gigantism in tomato

Authors for contact: Agustin Zsögön agustin.zsogon@ufv.br
Lázaro Peres lazaro.peres@usp.br

Article title: The *ORGAN SIZE* (*ORG*) locus contributes to isometric gigantism in domesticated tomato

Mateus Henrique Vicente, Kyle MacLeod, Cassia Regina Fernandes Figueiredo, Antonio Vargas de Oliveira Figueira, Fady Mohareb, Zoltán Kevei, Andrew J. Thompson, Agustin Zsögön*, Lázaro Eustáquio Pereira Peres*.

Laboratory of Plant Developmental Genetics. Departamento de Ciências Biológicas, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, CP 09, 13418-900, Piracicaba, SP, Brazil (M.H.R., C.R.F.F., L.E.P.P.); Cranfield Soil and AgriFood Institute, Cranfield University, Cranfield, MK43 0AL, UK (K.M., F.M., Z.K., A.J.T.); Laboratory of Plant Breeding, Centro de Energia Nuclear na Agricultura (CENA), USP, Av. Centenário, 303, CEP 13400-970, Piracicaba, SP, Brazil (A.V.O.F.); Departamento de Biologia Vegetal, Universidade Federal de Viçosa, CEP 36570-900, Viçosa, MG, Brazil (A.Z.).

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instruction for Authors (www.plantphysiol.org) is: Lázaro E. P. Peres (lazaro.peres@usp.br)

One sentence summary: A locus that controls isometric size increase in vegetative and reproductive organs of tomato through changes in cell division

M.H.V performed the experiments using A.V.O.F. methods and analysed the data; C.R.F.F. provided technical assistance to M.H.V.; K.M. performed experiments using F.M. methods and analysed the data with assistance from Z.K.; A.Z. and A.J.T. analysed the data and wrote the article with contributions of all the authors; L.E.P.P

conceived, designed and supervised the research, and agrees to serve as corresponding author.

*For correspondence: agustin.zsogon@ufv.br; lazaro.peres@usp.br

Abstract

Gigantism is a key component of the domestication syndrome, a suite of traits that differentiates crops from their wild relatives. Allometric gigantism is strongly marked in horticultural crops, causing disproportionate increases in the size of edible parts such as stems, leaves or fruits. Tomato (*Solanum lycopersicum*) has attracted attention as a model for fruit gigantism, and many genes have been described controlling this trait. However, the genetic basis of a corresponding increase in size of vegetative organs contributing to isometric gigantism, has remained relatively unexplored. Here, we identified a 0.4 Mbp region on chromosome 7 in introgression lines (ILs) from the wild species *Solanum pennellii* in two different tomato genetic backgrounds (cv. M82 and cv. Micro-Tom) that controls vegetative and reproductive organ size in tomato. The locus, named *ORGAN SIZE* (*ORG*), was fine-mapped using genotype-by-sequencing. A survey of literature revealed that *ORG* overlaps with previously mapped QTLs controlling tomato fruit weight during domestication. Alleles from the wild species led to reduced cell number in different organs, which was partially compensated by greater cell expansion in leaves but not in fruits. The result was a proportional reduction in leaf, flower and fruit size in the ILs harbouring the wild alleles. These findings suggest that selection for large fruit during domestication also tends to select for increases in leaf size by influencing cell division. Since leaf size is relevant for both source-sink balance and crop adaptation to different environments, the discovery of *ORG* could allow fine-tuning of these parameters.

Introduction

The domestication syndrome is the suite of phenotypic changes that occurred through artificial selection to transform wild species into crops (Evans 1996). Some of the most commonly found traits in crops are increased apical dominance, determinate growth and loss of natural seed dispersal (Meyer et al. 2012; Denham et al. 2020). An

increase in the size of certain organs, or gigantism, is also widespread, particularly in horticultural crops (Schwanitz 1957). Gigantism can be isometric, *i.e.* a proportional increase in all body parts, but most generally occurs through allometric alterations in the relative size of certain plant structures (Niklas 2004). A prime example is the species *Brassica oleracea*, where multiple cultivated strains were produced through artificial selection on the differential growth of edible organs such as stems (kohlrabi), buds (cabbage, Brussels sprouts), leaves (kale) and flowers (broccoli, cauliflower) (Prakash et al. 2011). Although increased organ size can be explained by alterations in cell division and expansion (Krizek 2009), it also requires developmental alterations to transform larger organs into stronger photosynthetic sources or sinks (Gifford et al 1984). Given that photosynthesis as a biochemical process has not been improved by crop domestication or breeding (Orr et al. 2017; Batista-Silva et al. 2020), most of the genetic gains in productivity have occurred indirectly through changes in plant development (Greenland et al. 1997; Zsögön and Peres 2018).

In tomato (*Solanum lycopersicum* L.), gigantism is evidenced in the phenomenal increase in fruit size when compared to its wild progenitor *S. pimpinellifolium* (Tanksley 2004). The genetic basis of fruit size control has attracted considerable attention (reviewed in Azzi *et al.*, 2015). Increased fruit size in tomato involves mutations in multiple loci, some of which have been characterized at the molecular level, for instance *fruit weight 2.2* (*fw2.2*), *fw3.2*, *fw11.3*, *fasciated* (*fas*), *locule number* (*lc*) and *EXCESSIVE NUMBER OF FLORAL ORGANS* (*ENO*). All of them are involved in fundamental processes of plant developmental such as cell division, expansion and differentiation. The *FW2.2* gene is a negative regulator of cell division responsible for up to 30% of the increase in fruit size when comparing lines harbouring small- and big-fruit alleles (Frary et al. 2000). *FW3.2* and *FW11.3* were identified as a P450 enzyme of the CYP78A subfamily (*SlKLUH*) and a *Cell Size Regulator* (*CSR*), controlling cell division and expansion, respectively (Chakrabarti et al. 2013; Mu et al. 2017). Unlike *fw2.2*, *fw3.2* and *fw11.3*, which mostly affect fruit size, *fas* and *lc* also control fruit shape. The big-fruit *fas* and *lc* alleles increase the number of carpels, altering cell diferentiation through the CLAVATA3-WUSCHEL module (Schoof et al. 2000). The increase in the number of carpels often results in larger and wider fruits with many locules and pronounced ribbing (Lippman and Tanksley 2001;

van der Knaap and Tanksley 2003). The *lc* mutant phenotype is caused by two single-nucleotide polymorphisms (SNPs) downstream of the coding region of the *WUSCHEL* (*WUS*) gene (Muños et al., 2011). The *fas* mutation is a partial loss of expression caused by a chromosome inversion with a break point in the vicinity of the *CLAVATA3* (*CLV3*) gene (Xu et al 2015), a negative regulator of *WUS* (Schoof et al. 2000). Lastly, *ENO* is an AP2/ERF transcription factor that interacts synergistically with *lc* and *fas*, causing a substantial increase of the *WUS* expression domain, which results in enlarged floral meristems (Fernández-Lozano et al., 2015; Yuste-Lisbona et al., 2020). Thus, the *ENO* domestication allele (a promoter deletion that knocks down its expression) also affects stem cell fate, giving rise to multilocular fruits that derive from the larger floral meristem.

Compared with the genetic regulation of fruit growth, relatively little is known about the control of vegetative organ size. In many crops, including tomato (Supp Fig. S1) but also peppers (Jarret et al., 2019), sunflower (Warburton et al., 2017), soybeans (Kofsky et al., 2018) and common beans (Herron et al., 2020), domestication entailed the selection of plants with bigger shoots and leaves. In tomato, the proportional increase in the size of vegetative parts is likely to be a component of isometric gigantism during domestication. Herein, we hypothesized that if vegetative gigantism is under genetic control, the wild species' alleles leading to reduced organ size could be found through wide crosses between cultivated tomato and its wild relative species. We selected *S. pennellii* as a wild parental, due to its annotated genome sequence (Bolger et al. 2014) and its rich repertoire of genomic tools, such as fully sequenced introgression lines (Alseekh et al. 2013; Chitwood et al. 2014). We crossed it to the cultivated tomato cv. Micro-Tom (MT) and after successive backcrosses and phenotypic selection, we isolated an introgression line with reduced vegetative and reproductive organs compared to the recurrent parental MT. We mapped this introgression to chromosome *7* and named the locus *ORGAN SIZE* (*ORG*). We show that *ORG* leads to reduced organ size through changes in cell division, and that it segregates as a monogenic, semi-dominant Mendelian locus. Our fine mapping results show that the *ORG* candidate genes overlap a previously described domestication sweep (Lin et al. 2014). We speculate on the impact of this locus in the tomato domestication syndrome and discuss its potential exploitation for crop breeding.

Results

Natural genetic variation for leaf size in tomato

Compared to domesticated tomato cultivars, most wild relatives of the tomato have small leaves (Supplemental Figure S1). Thus, we decided to look for a genetic determinant of leaf size in the wild species. We crossed *S. pennellii* to the cultivated tomato cv. Micro-Tom (MT). Upon self-fertilization of the F_1 population, we selected F_2 plants with small leaves, from which we collected pollen to backcross (BC) to MT. After six rounds of backcrossing to the recurrent parental (MT), self-fertilization (BC_6F_2), phenotypic screening, and further self-fertilization (BC_6F_n), we produced an introgression line (IL) with reduced leaf size in the MT background, which we called *ORGAN SIZE* (*ORG*) (Figure 1). *ORG* plants show a very conspicuous phenotype for leaf size: the difference in leaf size between MT and *ORG* was consistent across all leaves and developmental stages (Figure 1). Monogenic segregation of *ORG* was verified on a segregating population of MT and *ORG*. We determined leaf size in F_1 hybrids between MT and *ORG*, and the intermediate phenotype suggested that *ORG* behaves as a semi-dominant gene (Supplemental Figure S2).

The smaller leaf size in ORG is caused by reduced cell division

Change in organ size is due to either altered cell proliferation or expansion, or a combination of both (Krizek 2009). We analysed *ORG* leaves and found enlarged epidermal and mesophyll cells compared to MT (Supplemental Figure S3). This suggests that the smaller leaves of *ORG* are caused by reduced cell proliferation as evidenced by cell number and density of *ORG* compared to MT (Supplemental Figure S3). The greater palisade parenchyma cell size promoted an increase in leaf thickness in *ORG*. We next performed a time course analysis of reproductive growth starting eight days before anthesis and until 16 days after anthesis and verified a decrease in the size of styles, ovaries and fruits in *ORG* (Figure 2). As in the case of leaves, the reduction was caused by lower cell numbers, which we verified as a reduced number of cell layers in the pericarp. The ovary cells of *ORG* were also smaller than MT cells at anthesis and post-anthesis. Other floral organs, namely, petals and sepals, were also reduced in *ORG*

flowers compared to MT (Supplemental Figure S4). The reduced size of floral organs may have strong consequences on fruit development, given their impact on ovary size (Supplemental Figure S4e-h).

Fruit weight and yield are reduced in ORG

The size and shape of the ovary before anthesis is strongly correlated with the final size and shape of the fruit (Grandillo et al., 1999; Azzi et al., 2015). Thus, we next analysed the potential impact of *ORG* on fruit development. Fruit set was reduced in heterostylic *ORG* flowers, so we hand-pollinated emasculated MT and *ORG* flowers in a reciprocal cross. Several ovaries per plant were pollinated, but after fruit set confirmation (five days after pollination), we performed selective fruit removal to allow only five fruits to set on each plant. The presence of *ORG* ovaries had a substantial impact on the final fruit size regardless of pollen origin (Figure 3). Fruit weight was 31-37% lower in *ORG* than in MT ($P < 0.0001$, Supplemental Table S2). *ORG* fruits have higher total soluble solids content (°Brix) compared to MT (Supplemental Figure S5). We further observed that *ORG* had a similar frequency of locule number per fruit and reduced seed number (Supplemental Figure S5). Reciprocal crosses indicated that the reduction in seed number is determined by *ORG* ovaries rather than pollen (Supplemental Figure 5c).

We next addressed the possibility that reduced fruit size could be the consequence of altered photosynthetic source-sink relationships due to reduced leaf area. We thus manipulated the plants to maintain the availability of sources (leaves) constant and altered the source:sink ratio by changing the number of sinks (fruits). Three treatments were performed: either three, six or nine fruits were allowed to set on each plant. To ensure that additional sinks did not interfere in the results, we also pruned all the plants to remove side shoots. The results are summarized on Figure 3c-e. *ORG* plants produced consistently smaller fruits than MT in all treatments (Figure 3). The increase in fruit number, from three to six, promoted a reduction in fruit weight only in MT plants, suggesting that leaf area was a limiting factor to the final fruit weight in MT, since the leaf area was similar in both experimental conditions (Figure 3). On the other hand, when the number of fruits was increased from six to nine, there was a reduction in the final fruit weight for both genotypes. These results suggest that the

smaller leaf size of *ORG* could also account for its reduced fruit size, but only under full fruit load. Therefore, the primary cause of the reduced fruit size in *ORG* is likely a direct effect of this organ development since the pre-anthesis (Fig. 2c). In addition, the presence of the ORG introgression reduced the yield in all treatments.

Expression patterns are altered in genes related to cell division and expansion in ORG

The results described so far suggest that the transcriptional activity of genes involved in the control of cell division and expansion could be altered in *ORG*. To assess this, we extracted mRNA from ovaries/fruits at -8, -4, 0, 4 and 8 days pre/post anthesis, and fruit pericarps at 12 and 16 days to analyse the transcriptional profile of a set of genes related to the control of cell division: *CYCLIN B2;1* (Solyc02g082820), *FW2.2* (Solyc02g090730), *FW3.2/SlKLUH* (Solyc03g114940) and *EXPANSIN PRECURSOR 5* (Solyc02g088100).

In ovary/fruit tissues, we verified that the mRNA levels of the cell-division genes *CYCB2;1* and *FW3.2* showed greatest expression in both genotypes at 4 days pre-anthesis (Figure 4). *CYCB2;1* was higher in MT than *ORG* especially in pre-anthesis and anthesis stages (at -4, -8, and 0 days), while *FW3.2* was higher in anthesis and post-anthesis stages (at 0, 4 and 12 days). On the other hand, *FW2.2*, another cell-division gene, but a negative regulator, was highly expressed at 4 and 8 days post-anthesis in both genotypes. Quantitative variation in *FW2.2* expression was observed pre- and post-anthesis between genotypes (at -4 and 8 days, respectively), whereas *ORG* ovaries showed significant increased levels of this transcript compared than MT (Figure 4). After 4 days post-anthesis, the expression of the cell-expansion gene *EXPA5*, a member of the α-expansin gene family, increased in in both genotypes (Figure 4). However, ovaries of *ORG* plants displayed a significant decrease in the expression of this gene at anthesis compared to MT. Similar behavior was observed 16 DPA.

The ORG locus is located on chromosome 7

We next conducted a genotyping by sequencing (GBS) analysis to determine the size and location of the *S. pennellii* introgression in *ORG*. The results show a discrete region in the terminal end of the long arm of chromosome *7* encompassing ~11 Mb

(Figure 5). No further segments of *S. pennellii* genome were found on other chromosomes. Based on the SL2.50 tomato genome annotation, the introgression region contains 1169 genes. A closer look at the introgressed region revealed a small double recombination, from position 64,826,717 to 65,444,176, encompassing 78 tomato genes which score as *S. lycopersicum* (Figure 5b).

Fine-mapping of OS using introgression lines

To reduce the list of candidate genes for *ORG*, we next analysed two other introgression lines (ILs) of *S. pennellii* in the MT background previously generated in our laboratory: *Brilliant corolla* (*Bco)* and *Regeneration 7H* (*Rg7H),* both of which partially overlap either end of the *ORG* introgression (Figure 6). We used the span of the introgressions in *Bco* and *Rg7H* and the extent of their overlap with *ORG* (Supplemental Figure 6 for *Bco* and Pinto *et al.*, 2017 for *Rg7H*) to narrow down the candidate region for *ORG*. Given that neither of these ILs show the reduced organ phenotype of *ORG*, the resulting candidate region is located between positions 65,444,176 and 66,373,175 (Figure 6a).

We took advantage of the existing collection of ILs from *S. pennellii* in tomato cv. M82 as a tool to further refine the above chromosome location (Zamir and Eshed 1994; 1995). The introgressions were precisely delimited by sequencing by Chitwood et al. (2014), who also characterized terminal and lateral leaflet size in the ILs. Their results revealed the existence of a QTL for reduced leaflet size on both IL7-2 and IL7-3 (Figure 7a-b). We also cultivated ILs harbouring *S. pennellii* genomic segments on chromosome 7 (IL7-1; IL7-2, IL7-3; IL7-4 and IL7-5) and determined their leaf and ovary size. We found a reduction in the ovaries of both IL7-2 and IL7-3, compared to M82, but under our growth conditions only IL7-2 showed consistently smaller leaves than the parental line (Supplemental Figure S7). We found a discrepancy between the Chitwood et al. dataset and ours for leaf size on IL7-1, but the consistently smaller pistils in IL7-2 and IL7-3 helped us delimit the right border of the candidate region to 65,865,655, narrowing the interval to 421,479 bp (Figure 7c).

Genomic analysis of ORG and identification of candidate genes

The resequenced dataset of tomato and wild relative accessions (Aflitos et al., 2014) was used to identify the polymorphisms of *S. pennellii* when aligned with *S. lycopersicum* (SL2.50) in the *ORG* region. We found 58 CDS within the *ORG* region in the *S. pennellii* genome and 65 CDS within *S. lycopersicum*, with considerable synteny (Figure 7d). Within the *ORG* region, an alignment of the *S. pennellii* genome sequence (Spenn-ch07:76,477,056-76,940,423) with *S. lycopersicum* (SL2.50ch07:65,444,176-65,865,655) showed that the two genomes were structurally similar (Supplemental Figure S8). We therefore investigated the similarities and differences in the coding sequences (CDS) between the two genomes with BLAST (Supplemental Table S3). We found a total of 6,009 polymorphisms, 5,093 of which were single-nucleotide polymorphisms (SNPs) and 916 were insertions-deletions (InDels). Additionally, there were 304 moderate effect missense variants affecting 58 genes (Supplemental Table S4) and 18 high effect polymorphisms (*e.g.* frameshift variants, stop gained) (Supplemental Table S5). There was one *S. pennellii* CDS without a corresponding match in *S. lycopersicum*, *i.e.* a new gene within the *ORG* region, namely Sopen07g031050 (hypothetical protein). Additionally, there were six presence-absence variants (PAVs) in *S. lycopersicum* without a corresponding match in *S. pennellii* (Supplemental Table S6), *i.e.* six genes lost in the *ORG* region, namely, a Yippee family protein (Solyc07g062900), a nucleolar GTP-binding protein 2 (Solyc07g063280), a Tir 2C resistance protein (Solyc07g063360) and three CDS annotated as 'unknown protein'. The genes Sopen07g031090 and Sopen07g031100, both being putative Yippee family zinc-binding proteins, produced multiple significant matches with Solyc07g062880, Solyc07g062890 and Solyc07g062910. Additionally, Sopen07g031530 (beta glucosidase 46) and Sopen07g031540 (hypothetical protein) produced only partial matches with Solyc07g063370 (beta glucosidase) and Solyc07g063380 (unknown protein), respectively; indicating that the gene pairs share conserved regions but are otherwise dissimilar (Figure 7d).

Discussion

The genetic basis of fruit gigantism has been extensively explored in tomato and a number of major genes controlling that trait have been identified (Nesbitt and Tanksley 2001; Causse et al. 2004; Muños et al. 2011; Chakrabarti et al. 2013; Mu et al. 2017). However, the genetic mechanisms behind isometric gigantism between

vegetative and reproductive organs are unknown. Are they driven pleiotropically by genes for fruit gigantism that operate on the meristem simultaneously controlling vegetative and reproductive development, or are they the product of indirect selection on independent loci necessitated by the altered source-sink relationships between vegetative or reproductive organs? As a starting point to address this question, we set out to discover genetic determinants for changes in the size of vegetative organs in the tomato. We thus identified *ORGAN SIZE* (*ORG*), an introgression with reduced leaf size but which also showed smaller reproductive organs, namely flowers and fruits.

Instead of the conventional approach of QTL mapping, which sometimes is followed by fine-mapping and gene cloning, we revisited the alternative, forward genetics strategy, of wide cross followed by controlled introgression (Rick, 1969). We crossed *S. pennellii* to the tomato cv. Micro-Tom (MT) and conducted multiple rounds of crosses and backcrosses to the recurrent domesticated parental, selecting plants with smaller leaves in each generation. Our results, which identified the *ORG* locus, tie up previous, independent studies of the genetic control of leaf (Holtan and Hake, 2003; Chitwood *et al.*, 2014) and fruit (Grandillo et al. 1999; van der Knaap and Tanksley 2003; Causse et al. 2004; Barrantes et al. 2016) size in tomato using QTL analysis. Hence, a survey of previous studies that identified putative QTLs for increased fruit weight during tomato domestication and breeding revel a chromosomal region overlapping *ORG* (Supplemental Figure 9). However, none of these studies reported alterations in vegetative development associated to fruit weight QTLs. This indicates that controlled introgression guided by phenotypic selection is a powerful tool that, unlike QTL mapping, allows the detection of genes (or closely linked genes) that control more than one trait simultaneously. Either QTL mapping, or its more up-to-date variant, genome-wide sequencing analysis (GWAS), are useful to detect multiple genes spread out in the genome controlling one trait, but on the other hand, are prone to miss pleiotropic or tightly linked genes controlling multiple traits, because generally only one phenotype is analysed at a time (Korte and Farlow, 2013).

Genotyping-by-sequencing showed *ORG* to harbour 1169 genes in approximately 11 Mb of *S. pennellii* genome. This represents 1.15% of the tomato genome, which is a good fit with the theoretically expected proportion of donor genome after six rounds of back-crossing (Stam and Zeven 1981). Although the segregation data

indicate that *ORG* behaves as a Mendelian, semi-dominant gene, we cannot at this stage exclude the possibility that the IL harbours two or more genes controlling similar traits on chromosome 7. However, we showed that the common denominator for the reduced size of vegetative and reproductive organs in *ORG* is a reduction in the number of cells, possibly through alteration of cell division rate, as suggested by our gene expression analyses for *CYCB2;1*, *FW2.2* and *FW3.2*. This trait could be under pleiotropic control of a single gene. In fact, our analysis of the genes contained in the candidate region shows variation between *S. pennellii* and *S. lycopersicum* for genes predicted to be involved in the control of cell division, as well as regulatory genes that could control the size of organs (Supplemental Tables S4 and S5). An interval containing 19 putative domestication genes was also identified on chromosome 7 by Lin *et al.* (2014) by analyzing the genome sequence of 360 tomato accessions. All 19 genes are contained within the list of 58 candidates for the *ORG* region. This paves the way for the future identification and validation of, potentially, a single gene with a unique underlying variant (*e.g.* SNP, InDel, PAV) controlling organ size.

Increased organ size, or gigantism, is a recurrent domestication trait observed in many crops. Selection for increased size of edible parts led to allometric increases in reproductive organs. However, domesticated plants also tend to present gigantism in vegetative parts, *e.g.* larger leaves and thicker stems in *Phaseolus vulgaris* (Donald and Hamblin, 1983), larger leaves in eggplant (Page *et al.*, 2019) and soybean (Kofsky *et al.*, 2018). The tomato shows striking increases in fruit size (Tanksley 2004), but also leaf area, and stem thickness compared to its wild relatives (Milla and Matesanz, 2017). This isometric size change could lead to a better balance between photosynthetic sources and fruit sinks. When we altered the relative strength of the sinks by allowing only three, six or nine fruits to develop in either MT or *ORG* plants, we found an inverse correlation between fruit number and size in MT but not in *ORG*. In addition, the reduction in fruit size of MT has no penalty in its final yield. These results suggest two things. First, that the reduced size of *ORG* fruits is an intrinsic trait, possibly a developmental result of smaller ovaries, and not an indirect consequence of reduced leaf area (photosynthetic source). The second is that leaf area is not always directly limiting fruit (sink) size and/or yield. In agreement with this, both experimental and modelling work have shown that defoliation does not have a negative effect on crop yield,

352 implying that source strength is not limiting (provided water and nutrient availability
353 are sufficient and that photosynthesis is not light limited) (Heuvelink *et al.*, 2005). An
354 extreme situation is found in garden peas (*Pisum sativum*), where leaf area reduction has
355 been a breeding goal to reduce interplant competition and increase yield (Cousin, 1997).
356 Mutants of the 'leafless' and 'semi-leafless' type show 40% lower leaf area with up to
357 20% higher yield and better standing ability, which in turn facilitates mechanical
358 harvesting (Checa *et al.*, 2020). The increased popularity and growing market niche for
359 'gourmet' cherry tomatoes opens up the perspective of breeding varieties with smaller
360 leaves to improve agronomic management (*e.g.* reduced fertilizer, water use) (Sarlikioti
361 et al., 2011).
362
363 **Conclusions**
364 Based on the analysis of natural genetic variation, we have described a potential
365 genetic determinant for increased leaf size in cultivated tomato. Our results could unveil
366 a novel link in the genetic control of isometric fruit and leaf gigantism in tomato.
367 Further research to determine the molecular identity of the gene(s) underlying the *ORG*
368 phenotype is underway. This knowledge would be a valuable addition in the repertoire
369 of gene targets that can be manipulated with ideotype breeding (Donald, 1968; Zsögön
370 et al. 2017) or *de novo* domestication platforms (Gasparini et al., 2021).
371
372 **Materials and methods**
373
374 *Plant material*

375 The wild relatives of tomato used in this work were *S. pennellii* (LA0716), *S. chilense*
376 (LA1969), *S. peruvianum* (LA1537), *S. neorickii* (LA1322), *S. chmieslewskii* (LA1028),
377 *S. habrochaites* f. *glabratum* (PI134417), *S. habrochaites* f. *hirsutum* (LA1777), *S.*
378 *galapagense* (LA1401), *S. pimpinellifolium* (CNPH384), and *S. lycopersicum* var.
379 *cerasiforme* (LA1320). Domesticated tomatoes of the cultivars Micro-Tom (MT)
380 (LA3911), M82 (LA3475), Moneymaker (LA2706) and Santa Clara (Brazilian local
381 cultivar) were also used. The *S. pennellii* chromosome 7 introgression lines (ILs)
382 harboring alleles of *ORGAN SIZE* (*ORG*), *BRILLIANT COROLLA* (*Bco*) (Chetelat
383 1998) and *Rg7H* (Pinto et al. 2017) were obtained through repeated backcrossing

384 between cultivated MT as a pollen receptor and *S. pennellii*, as described in Carvalho et
385 al. (2011). Seeds of the tomato wild relatives were obtained from the UC Davis/C.M.
386 Rick Tomato Genetics Resource Center, maintained by the Department of Plant
387 Sciences, University of California, Davis, CA 95616. Seeds of MT were kindly donated
388 by Prof. Avram Levy (Weizmann Institute of Science, Israel) in 1998 and kept as a true-
389 to-type cultivar through self-pollination.

390 *Growth conditions*

391 Plants were grown in a greenhouse at the Laboratory of Plant Developmental
392 Genetics, ESALQ-USP, (543 m a.s.l., 22º 42' 36" S; 47º 37' 50" W), Piracicaba, SP,
393 Brazil. Automatic irrigation took place four times a day. Growth conditions were: mean
394 temperature of 28°C, 11.5 h/13 h (winter/summer) photoperiod, 250–350 µmol photons
395 m^{-2} s^{-1} PAR irradiance, attained by a reflecting mesh (Aluminet, Polysack Indústrias
396 Ltda, Leme, Brazil). Seeds were germinated in 350 mL pots with a 1:1 mixture of
397 commercial potting mix Basaplant® (Base Agro, Artur Nogueira, SP, Brazil) and
398 expanded vermiculite supplemented with 1 g L^{-1} 10:10:10 NPK and 4 g L^{-1} dolomite
399 limestone ($MgCO_3$ + $CaCO_3$). Upon the appearance of the first true leaf, seedlings were
400 transplanted to pots containing the soil mix described above, except for NPK
401 supplementation, which was increased to 8 g L^{-1}. In addition, MT and *OS* plants
402 received a supplementary fertilization of 0.5g of NPK formulation 10:10:10 after
403 flowering. Cultivated and wild tomato plants were supplemented with 2g of NPK
404 formulation 10:10:10 per plant.
405
406 *Phenotypic characterization*

407 We scanned all leaves of the MT and *ORG* plants 40 days after germination
408 (dag) and determined the leaf area using ImageJ software (http://rsbweb.nih.gov/ij/).
409 For the characterization of floral whorls, we evaluated: length of petals and
410 sepals; corolla area; and ovary weight, height and diameter. To measure ovary length
411 and height we used a magnifying glass (Leica S8AP0, Wetzlar, Germany), coupled to a
412 camera (Leica DFC295 Wetzlar, Germany). To determine ovary weight we determined
413 the weight of 1.5 mL Eppendorf microtubes with 1 mL of distilled water, before and
414 after collection of 10 ovaries. Ovary weight was then determined as the difference
415 between initial and final tube weight. We also evaluated the leaf area and ovary weight

of M82 plants and introgression lines (ILs) from chromosome 7, using the same methodology as for MT and *ORG*.

MT and ORG productivity traits

We hand-pollinated MT and *ORG* plants with pollen from MT and *ORG* plants, because the *ORG* genotype displayed low fruit set. Various ovaries were pollinated, but after fruit set confirmation (five days after pollination), we performed selective fruit removal to allow only five fruits to set on each plant.

Productive performance of plants was assessed 90 days after germination. The following parameters were determined: mean weight per fruit; total soluble solids content in fruits (Brix); locule number and number of seeds per fruit; and weight of 10 seeds. Total soluble solids content of fruits was assessed using a digital refractometer (PR-101, Atago, Tokyo, Japan).

Source-sink ratio in MT and ORG plants

To determine whether leaf area of *ORG* plants is a limiting factor for fruit development (since leaves and fruits are the major sources and sinks of photoassimilates, respectively), we manipulated plants creating three categories based on different source-to-sink ratios. Thus, we kept the same amount of source tissue (leaves) in all plants of each genotype and altered the sink strength by changing fruit number (either three, six or nine per plant, to produce high, medium or low source-to-sink ratios, respectively). We removed side branches to prevent them from acting as alternative sinks. The following parameters were then determined: total fruit weight per plant (yield); average fruit weight and whole-plant leaf area.

Mapping and PCR amplification of DNA markers

We designed molecular markers to discover polymorphisms between tomato and *S. pennellii* in the region comprising the IL-7-2 and part of the IL 7-4 (Chitwood et al., 2014). The sequences and types of molecular makers are shown on Supplemental Table S1. Two further genotypes harbouring genome segments of *S. pennellii* for chromosome 7, *Brilliant corolla* (*Bco*) and *Regeneration 7h* (*Rg7H*), both in cv. MT, were characterized molecularly and phenotypically. Cross-referencing information from these

448 genotypes and the ILs in the M82 background we constructed a map with the putative
449 location of the *ORG* locus.

450 Genomic DNA extraction from young leaves was performed as described by
451 Fulton et al. (1995). PCR was performed using the following program: a denaturation
452 step at 95°C for 2 min, 35 cycles of 30 s at 95°C, 60s at 56°C, 90 s at 72°C, and a final
453 cycle at 72°C for 7 min. When required, restriction enzyme analysis (Supplemental
454 Table S1) was performed following the manufacturer's recommendations (NEB,
455 Bethesda, USA). The final PCR products were analyzed via 1.5% (m/v) agarose gel
456 electrophoresis, stained with SYBR Gold (Invitrogen).

457

458 *Histological and microscopic analyses*

459 Samples of MT and *ORG* ovaries/fruits at -8, -4, 0, 4 and 8 days, and fruit
460 pericarps at 12 and 16 days (anthesis=0), were collected and fixed in Karnovsky
461 solution (Karnovsky 1965), and vacuum-infiltrated for 15 min. The times referred to as
462 -8 and -4 days correspond to 8 and 4 days before anthesis, respectively. We based these
463 on the length of the closed flower buds (Faria 2014).

464 Samples were next dehydrated in an increasing ethanol series (10–100%), and
465 infiltrated into synthetic resin, using a HistoResin embedding kit (Leica, www.leica-
466 microsystems.com), according to the manufacturer's instructions. The tissues were
467 sliced using a rotary microtome (Leica RM 2045, Wetzlar, Germany), stained with
468 toluidine blue 0.05% (Sakai 1973), and photographed in a microscope (Leica DMLB,
469 Heidelberg, Germany), coupled to a Leica DFC310 camera (Wetzlar, Germany).
470 Histological analysis of ovaries was performed in the central region of the outer
471 pericarp of the fruits, and the area and number of cells were determined using ImageJ
472 software (http://rsbweb.nih.gov/ij/). This histological analysis also was performed in the
473 mature leaves of these genotypes adopting the procedures described above. The area
474 and number of cells in the adaxial leaf epidermis of the MT and *ORG* genotypes was
475 also evaluated using the leaf dental resin imprinting technique (Weyers and Johansen
476 1985).

477

478 *Quantitative real-time reverse transcription PCR*

Total RNA was extracted from ovaries/fruits at -8, -4, 0, 4 and 8 days, and fruit pericarps at 12 and 16 days (anthesis = 0), using Trizol reagent (Invitrogen), as indicated by the manufacturer, and treated with RQ1 RNAse-Free DNAse (Promega). Fruit pericarps were carefully collected from the central region of the outer pericarp of the fruits, at 12 and 16 days. After DNase treatment, a single-strand cDNA was synthesized from total RNA (1µg) by reverse-transcription, using RevertAid RT Reverse Transcription Kit (Thermo Fisher Scientific).

Gene expression analyses were performed on a Rotor-Gene Q real-time PCR cycler (Qiagen), using Kapa Sybr Fast qPCR Master Mix (Kapa Biosystems) and specific primers for *CYCB2;1* (Solyc02g082820), *FW2.2* (Solyc02g090730), *FW3.2* (Solyc03g114940) and *EXP5* (Solyc02g088100) genes. The reactions were amplified for 2 min at 95 °C, followed by 40 cycles of 95 °C for 15 s and 60 °C for 30 s. The threshold cycle (C_T) was determined. Melting curve analysis was performed with each primer set to confirm the presence of only a single peak before the gene expression analyses. Two technical replicates were analyzed for each of three or four biological samples. The relative transcript accumulation was normalized to an *ACTIN* (Solyc04g011500) gene. The fold changes for each gene were calculated using the equation $2^{-\Delta\Delta CT}$ (Livak and Schmittgen 2001). Primer sequences used to qRT-PCR are shown in Supplemental Table S1.

In silico *analysis of probable* ORG *region*

The genomes of *S. lycopersicum* cv. Heinz 1706, SL2.50 (https://solgenomics.net/) and *S. pennellii* LA716 (Bolger et al., 2014b) were aligned and plotted with Mummer v4.0.0 (Marcais et al., 2018). Variants of *S. pennellii* LA716 versus SL2.50 within the *OS* region were obtained through the Wageningen resequencing project (Aflitos et al., 2014). The coding sequences of the genes within the region were obtained from Solanaceae Genomics Network (https://solgenomics.net/) and similarities between Heinz 1706 and LA716 were tested with BLAST v2.10.0 (Camacho et al., 2009). The Circos plot was created with Circos v0.69.9 (Krzywinski et al., 2009) on Windows 10. The synteny plot was created with the genoPlotR package (Guy et al., 2011) within R (Team, 2017).

Genotyping by sequencing (GBS)

DNA was extracted from young leaf samples (~10 mm length) that were freeze-dried (CoolSafe™ 55-9; Scanvac, Lynge, Denmark) overnight. Leaf samples were powdered in a Star-Beater (VWR, Lutterworth, UK) at 30 Hz for 30s in 2 mL microcentrifuge tubes containing two 5 mm acid-rinsed soda-glass balls. DNA was extracted from ~50 mg samples with an E.Z.N.A® Plant DNA Kit (VWR, Lutterworth, UK). DNA fragment size was assessed on a 1% agarose gel in Tris/Borate/EDTA to confirm that all samples had the majority of DNA fragments >10 kilobases.

The GBS library was prepared using the restriction enzyme *Msl*I and sequenced on an Illumina NextSeq 500 V2 by LGC Genomics (Berlin, Germany). The 150 base-pair paired-end reads were aligned to the *Solanum lycopersicum* Heinz 1706 reference genome (SL2.50) with BWA v0.7.15 (Li and Durbin, 2009). The SAM files were processed with Samtools Fixmate v1.3.1 (Li et al., 2009). InDels were realigned with GATK's IndelRealigner v3.8-0 (Mckenna et al., 2010; Depristo et al., 2011) before variant calling with Samtools Mpileup v1.3.1 and Bcftools Call v1.3 (Li, 2011).

The raw VCF files of the GBS sample, a 40× resequenced Micro-Tom (Cranfield University, unpublished data) and the resequencing of *S. pennellii* LA716 (Aflitos et al., 2014), were combined into an index using Tersect (Kurowski and Mohareb, 2020). Tersect was used to determine which variants were shared between the *ORG* IL and *S. pennellii* LA716, excluding the variants shared with Micro-Tom. The variants output from Tersect were then filtered as follows: all variants with a quality score less than 20, a mapping quality score below 40 and a raw read depth either below 10 and above 200 were removed. In addition, heterozygous variants were removed. The variant density of the filtered variants over a 10 kb window (5 kb sliding) were plotted across all 12 chromosomes with ggplot2 (Wickham, 2016) within R.

Statistical analysis

Statistical analysis was performed using SAS software (SAS Institute Inc., Cary, NC, USA). The variables data were submitted to analysis of variance (ANOVA) and the means compared by the Student's t- or Tukey's test. When the data did not meet the assumptions of ANOVA, we performed to non-parametric analysis, using Wilcoxon rank sum or Dunn's test to compare the means.

Supplemental Data

Supplemental Figure S1. Leaf size increases during tomato domestication and improvement.

Supplemental Figure S2. Heterozygous *ORG* plants (*ORG*/+) show an intermediate leaf area compared to MT and *ORG* plants.

Supplemental Figure S3. Smaller leaf size in *ORG* is caused by reduced cell division.

Supplemental Figure S4. *ORG* reduces organ size in all floral whorls

Supplemental Figure 5. Fruit traits are altered in *ORG* plants.

Supplemental Figure 6. GBS defines the span of the introgression in the *Brilliant corolla (Bco)* introgression line.

Supplemental Figure 7. Characterization of *S. pennellii* introgression lines (IL) in chromosome *7*.

Supplemental Figure 8. Alignment plot of the *S. pennellii* and *S. lycopersicum* genomes within the *ORG* region.

Supplemental Figure 9. Colocalization of ORG and previously mapped fruit size QTLs.

Supplemental Table S1. Oligonucleotide sequences used for genotyping and quantitative PCR analyses in this work.

Supplemental Table S2. Fruit weight of MT and ORG plants.

Supplemental Table S3. Similarities and discrepancies between coding sequences of S. pennellii v. S. lycopersicum candidate genes.

Supplemental Table S4. Polymorphisms with a moderate effect on gene function for S. pennellii v S. lycopersicum within the ORG region.

Supplemental Table S5. Polymorphisms with a high effect on gene function for S. pennellii v S. lycopersicum within the ORG region.

Supplemental Table S6. Coding sequences of S. lycopersicum not producing a match on the S. pennellii genome assembly.

Acknowledgements

AZ was partly funded by a grant (RED□00053□16) from the Foundation for Research Assistance of the Minas Gerais State (FAPEMIG, Brazil). AT, FM, ZK were partially funded by BBSRC BB/S007970/1; AT and AZ were partially funded by the

Royal Society's Newton Mobility Grant NMG\R2\170027 and the Global Challenges Research Fund (GCRF, UK Research and Innovation). LEPP was supported by grant 306518/2018-0 from CNPq (Brazil) and 2018/050003-1 from FAPESP (Brazil). MHV gratefully acknowledges a PhD scholarship from FAPESP (16/05566-0). We thank Diego S. Reartes for valuable technical assistance.

Literature cited

Aflitos S, Schijlen E, De Jong H, De Ridder D, Smit S, Finkers R, Wang J, Zhang G, Li N, Mao L, et al (2014) Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing. Plant Journal 80: 136–148

Alseekh S, Ofner I, Pleban T, Tripodi P, Di Dato F, Cammareri M, Mohammad A, Grandillo S, Fernie AR, Zamir D (2013) Resolution by recombination: Breaking up Solanum pennellii introgressions. Trends in Plant Science 18: 536–538

Azzi L, Deluche C, Gévaudant F, Frangne N, Delmas F, Hernould M, Chevalier C (2015) Fruit growth-related genes in tomato. Journal of Experimental Botany 66: 1075–1086

Batista-Silva W, da Fonseca-Pereira P, Martins AO, Zsögön A, Nunes-Nesi A, Araújo WL (2020) Engineering Improved Photosynthesis in the Era of Synthetic Biology. Plant Communications 1: 100032

Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, Quesneville H, Alseekh S, Sørensen I, Lichtenstein G, et al (2014a) The genome of the stress-tolerant wild tomato species Solanum pennellii. Nature Genetics 46: 1034–8

Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, Quesneville H, Alseekh S, Sørensen I, Lichtenstein G, et al (2014b) The genome of the stress-tolerant wild tomato species Solanum pennellii. Nature Genetics 46: 1034–1039

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: Architecture and applications. BMC Bioinformatics 10: 1–9

Causse M, Duffe P, Gomez MC, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-Chamley M, Rothan C (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. Journal of Experimental Botany 55: 1671–1685

Chakrabarti M, Zhang N, Sauvage C, Muños S, Blanca J, Cañizares J, Diez MJ, Schneider R, Mazourek M, McClead J, et al (2013) A cytochrome P450 regulates a domestication trait in cultivated tomato. Proceedings of the National Academy of Sciences 110: 17125–17130

Checa OE, Rodriguez M, Wu X, Blair MW (2020) Introgression of the Afila Gene into Climbing Garden Pea (Pisum sativum L.). Agronomy 10: 1537

Chitwood DH, Ranjan A, Kumar R, Ichihashi Y, Zumstein K, Headland LR, Ostria-Gallardo E, Aguilar-Martínez JA, Bush S, Carriedo L, et al (2014) Resolving distinct genetic

612 regulators of tomato leaf shape within a heteroblastic and ontogenetic context. Plant
613 Cell 26: 3616–29

614 Cousin R (1997) Peas (Pisum sativum L.). Field Crops Research 53: 111–130

615 Denham T, Barton H, Castillo C, Crowther A, Dotte-Sarout E, Florin SA, Pritchard J, Barron A,
616 Zhang Y, Fuller DQ (2020) The domestication syndrome in vegetatively propagated
617 field crops. Annals of Botany 125: 581–597

618 Depristo MA, Banks E, Poplin R, Garimella K V., Maguire JR, Hartl C, Philippakis AA, Del
619 Angel G, Rivas MA, Hanna M, et al (2011) A framework for variation discovery and
620 genotyping using next-generation DNA sequencing data. Nature Genetics 43: 491–501

621 Donald CM (1968) The breeding of crop ideotypes. Euphytica 17: 385–403

622 Donald CM, Hamblin J (1983) The Convergent Evolution of Annual Seed Crops in Agriculture.
623 Advances in Agronomy 36: 97–143

624 Evans LT (1996) Crop Evolution, Adaptation and Yield. Cambridge University Press,
625 Cambridge, UK

626 Fernández-Lozano A, Yuste-Lisbona FJ, Pérez-Martín F, Pineda B, Moreno V, Lozano R,
627 Angosto T (2015) Mutation at the tomato EXCESSIVE NUMBER OF FLORAL
628 ORGANS (ENO) locus impairs floral meristem development, thus promoting an
629 increased number of floral organs and fruit size. Plant Science 232: 41–48

630 Frary A, Nesbitt TC, Frary A, Grandillo S, Knaap E van der, Cong B, Liu J, Meller J, Elber R,
631 Alpert KB, et al (2000) fw2.2: A Quantitative Trait Locus Key to the Evolution of
632 Tomato Fruit Size. Science 289: 85–88

633 Gasparini K, Moreira J dos R, Peres LEP, Zsögön A (2021) De novo domestication of wild
634 species to create crops with increased resilience and nutritional value. Current Opinion
635 in Plant Biology 60: 102006

636 Gifford RM, Thorne JH, Hitz WD, Giaquinta RT (1984) Crop productivity and photoassimilate
637 partitioning. Science 225: 801–8

638 Grandillo S, Ku HM, Tanksley SD (1999) Identifying the loci responsible for natural variation
639 in fruit size and shape in tomato. TAG Theoretical and Applied Genetics 99: 978–987

640 Greenland DJ, Gregory PJ, Nye PH, Evans LT (1997) Adapting and improving crops: the
641 endless task. Philosophical Transactions of the Royal Society of London Series B:
642 Biological Sciences 352: 901–906

643 Guy L, Kultima JR, Andersson SGE, Quackenbush J (2011) GenoPlotR: comparative gene and
644 genome visualization in R. Bioinformatics 27: 2334–2335

645 Herron SA, Rubin MJ, Ciotir C, Crews TE, Van Tassel DL, Miller AJ (2020) Comparative
646 Analysis of Early Life Stage Traits in Annual and Perennial Phaseolus Crops and Their
647 Wild Relatives. Front Plant Sci. doi: 10.3389/fpls.2020.00034

648 Heuvelink E, Bakker MJ, Elings A, Kaarsemaker R, Marcelis LFM (2005) Effect of leaf area on
649 tomato yield. Acta horticulturae

650 Holtan HEE, Hake S (2003) Quantitative Trait Locus Analysis of Leaf Dissection in Tomato
651 Using Lycopersicon pennellii Segmental Introgression Lines. Genetics 165: 1541–1550

652 Jarret RL, Barboza GE, Batista FR da C, Berke T, Chou Y-Y, Hulse-Kemp A, Ochoa-Alejo N,
653 Tripodi P, Veres A, Garcia CC, et al (2019) Capsicum—An Abbreviated Compendium.
654 Journal of the American Society for Horticultural Science 144: 3–22

655 Karnovsky MJ (1965) A formaldehyde-glutaraldehyde fixative of high osmolality for use in
656 electron microscopy. Journal of Cell Biology 27: 137A-137A

657 van der Knaap E, Tanksley SD (2003) The making of a bell pepper-shaped tomato fruit:
658 identification of loci controlling fruit morphology in Yellow Stuffer tomato. Theor Appl
659 Genet 107: 139–147

660 Kofsky J, Zhang H, Song B-H (2018) The Untapped Genetic Reservoir: The Past, Current, and
661 Future Applications of the Wild Soybean (Glycine soja). Front Plant Sci. doi:
662 10.3389/fpls.2018.00949

663 Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a
664 review. Plant Methods 9: 29

665 Krizek BA (2009) Making bigger plants: key regulators of final organ size. Current Opinion in
666 Plant Biology 12: 17–22

667 Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA
668 (2009) Circos: An information aesthetic for comparative genomics. Genome Research
669 19: 1639–1645

670 Kurowski TJ, Mohareb F (2020) Tersect: A set theoretical utility for exploring sequence variant
671 data. Bioinformatics 36: 934–935

672 Li H (2011) A statistical framework for SNP calling , mutation discovery , association mapping
673 and population genetical parameter estimation from sequencing data. 27: 2987–2993

674 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler
675 transform. Bioinformatics 25: 1754–1760

676 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
677 Data GP, et al (2009) The Sequence Alignment/Map format and SAMtools.
678 Bioinformatics 25: 2078–2079

679 Lin T, Zhu G, Zhang J, Xu X, Yu Q, Zheng Z, Zhang Z, Lun Y, Li S, Wang X, et al (2014)
680 Genomic analyses provide insights into the history of tomato breeding. Nature Genetics
681 46: 1220–1226

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

685 Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time
686 quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods (San Diego, Calif) 25:
687 402–8

688 Marcais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Aleksey Z (2018) MUMmer4:
689 A fast and versatile genome alignment system. PLoS Comput Biol 14(1): 14: 1–14

690 Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,
691 Altshuler D, Gabriel S, Daly M, et al (2010) The Genome Analysis Toolkit: A
692 MapReduce framework for analyzing next-generation DNA sequencing data. Genome
693 Research 20: 1297–1303

694 Meyer RS, DuVal AE, Jensen HR (2012) Patterns and processes in crop domestication: an
695 historical review and quantitative analysis of 203 global food crops. New Phytologist
696 196: 29–48

697 Milla R, Matesanz S (2017) Growing larger with domestication: a matter of physiology,
698 morphology or allocation? Plant Biology 19: 475–483

699 Mu Q, Huang Z, Chakrabarti M, Illa-Berenguer E, Liu X, Wang Y, Ramos A, van der Knaap E
700 (2017) Fruit weight is controlled by Cell Size Regulator encoding a novel protein that is
701 expressed in maturing tomato fruits. PLOS Genetics 13: e1006930–e1006930

702 Muños S, Ranc N, Botton E, Bérard A, Rolland S, Duffé P, Carretero Y, Le Paslier M-C,
703 Delalande C, Bouzayen M, et al (2011) Increase in tomato locule number is controlled
704 by two single-nucleotide polymorphisms located near WUSCHEL. Plant physiology
705 156: 2244–54

706 Nesbitt TC, Tanksley SD (2001) fw2.2 Directly Affects the Size of Developing Tomato Fruit,
707 with Secondary Effects on Fruit Number and Photosynthate Distribution. Plant
708 Physiology 127: 575–583

709 Niklas KJ (2004) Plant allometry: is there a grand unifying theory? Biological reviews of the
710 Cambridge Philosophical Society 79: 871–89

711 Orr DJ, Pereira AM, da Fonseca Pereira P, Pereira-Lima ÍA, Zsögön A, Araújo WL (2017)
712 Engineering photosynthesis: progress and perspectives. F1000 Research 6: 1891–1891

713 Page AML, Daunay M-C, Aubriot X, Chapman MA (2019) Domestication of Eggplants: A
714 Phenotypic and Genomic Insight. *In* MA Chapman, ed, The Eggplant Genome. Springer
715 International Publishing, Cham, pp 193–212

716 Pinto M de S, Abeyratne CR, Benedito VA, Peres LEP (2017) Genetic and physiological
717 characterization of three natural allelic variations affecting the organogenic capacity in
718 tomato (Solanum lycopersicum cv. Micro-Tom). Plant Cell, Tissue and Organ Culture
719 (PCTOC) 129: 89–103

720 Prakash S, Wu X-M, Bhat SR (2011) History, Evolution, and Domestication of Brassica Crops.
721 Plant Breeding Reviews. John Wiley & Sons, Ltd, pp 19–84

722 Rick CM (1969) Controlled Introgression of Chromosomes of Solanum pennellii into
723 Lycopersicon esculentum: Segregation and Recombination. Genetics 62: 753–768

724 Sarlikioti V, De Visser PHB, Buck-Sorlin GH, Marcelis LFM (2011) How plant architecture
725 affects light absorption and photosynthesis in tomato: towards an ideotype for plant
726 architecture using a functional-structural plant model. Annals of Botany 108: 1065–73

Schoof H, Lenhard M, Haecker A, Mayer KF, Jürgens G, Laux T (2000) The stem cell population of Arabidopsis shoot meristems in maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. Cell 100: 635–644

Schwanitz F (1957) Von der Wildpflanze zur Kulturform. *In* F Schwanitz, ed, Die Entstehung der Kulturpflanzen. Springer, Berlin, Heidelberg, pp 1–52

Stam P, Zeven AC (1981) The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. Euphytica 30: 227–238

Tanksley SD (2004) The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. Plant Cell 16: S181–S189

Team RC (2017) R: A language and environment for statistical computing.

Warburton ML, Rauf S, Marek L, Hussain M, Ogunola O, Gonzalez J de JS (2017) The Use of Crop Wild Relatives in Maize and Sunflower Breeding. Crop Science 57: 1227–1240

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis, 1st ed. doi: 10.1007/978-0-387-98141-3

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

Yuste-Lisbona FJ, Fernández-Lozano A, Pineda B, Bretones S, Ortíz-Atienza A, García-Sogo B, Müller NA, Angosto T, Capel J, Moreno V, et al (2020) ENO regulates tomato fruit size through the floral meristem development network. PNAS 117: 8187–8195

Zsögön A, Cermak T, Voytas D, Peres LEP (2017) Genome editing as a tool to achieve the crop ideotype and de novo domestication of wild relatives: Case study in tomato. Plant Science 256: 120–130

Zsögön A, Peres LEP (2018) Molecular control of plant shoot architecture. Plant Cell 30: tpc.118.tt1218-tpc.118.tt1218

Figure legends

Figure 1. A tomato introgression line (IL) from *S. pennellii* with reduced vegetative organs (*ORG*) size. (a) Crossing scheme to create an introgression line with smaller leaves in the tomato cv Micro-Tom (MT) background **(b)** Representative population of MT (left) and *ORG* (right) plants, 25 days after germination (dag). **(c)** Side and top view of MT (top) and *ORG* (bottom) plants. **(d)** Leaf series of MT (top) and *ORG* (bottom) genotypes from cotyledons (C1) to fifth leaf (L5). Scale bar=5 cm. **(e)** Leaf area of the leaf series of MT (gray bar) and *ORG* (white bar) plants, 40 dag. Data are mean ± s.e.m. (n=14 leaves). Statistical significance was tested by Student's *t*-test (***$p<0.001$).

Figure 2. *ORG* affects cell number and size during fruit development. (a) Developing ovary/fruit at -12, -8, -4, 0, 4, 8, 12 and 16 days (anthesis = 0). MT (top) and *ORG* (botton). Scale bar=5mm. **(b)** Longitudinal sections of MT (top) and *ORG* (bottom) pericarp at -12, -8, -4, 0, 4, 8, 12 and 16 days (anthesis = 0). Scale bar = 150µm. **(c)** Time course of the number of

cell layers in the longitudinal sections of MT (gray bar) and *ORG* (white bar) ovary/fruit pericarp. Insert in top of this figure represents how the counting of the cells was performed and red lines delimited cell perimeter (n=30). **(d)** Time course of cell area in the cell layers of MT (gray bar) and *ORG* (white bar) (n=30). Data are mean ± s.e.m. Statistical significance was tested by Student *t*-test (*$p<0.05$, ***$p<0.001$, *ns* indicates non-significant differences).

Figure 3. Fruit growth and source-sink relationships are altered in *ORG*. **(a)** Representative MT (♀, left) and *ORG* (♀, right) ripe fruits pollinated with MT (♂, left) and *ORG* (♂, right) pollen. Scale bar=1 cm. **(b)** Mean (red) and median (black) values of fruit weight of MT (gray box) and *ORG* (white box) ripe fruits pollinated with MT (n=10) and *ORG* (n=14) pollen. **(c)** Frequency of locule number per fruit in MT and *ORG* fruits (n=125). **(d)** Seeds per fruit of MT and *ORG* pollinated with MT (n=11) and *ORG* (n=15) pollen. Data are mean±s.e.m. Statistical significance was tested by Student's *t*-test (***$p<0.001$). **(e-g)** Average values of fruits weight (e), leaf area (f) and yield (g) from MT (gray bar) and *ORG* (white bar) plants pruned to three, six and nine fruits (n=6 plants per treatment). Data are mean±s.e.m. Different capital and lowercase letters on the symbols indicate significant differences by Tukey's test ($p<0.001$) between the treatments in MT and *ORG* genotypes, respectively.

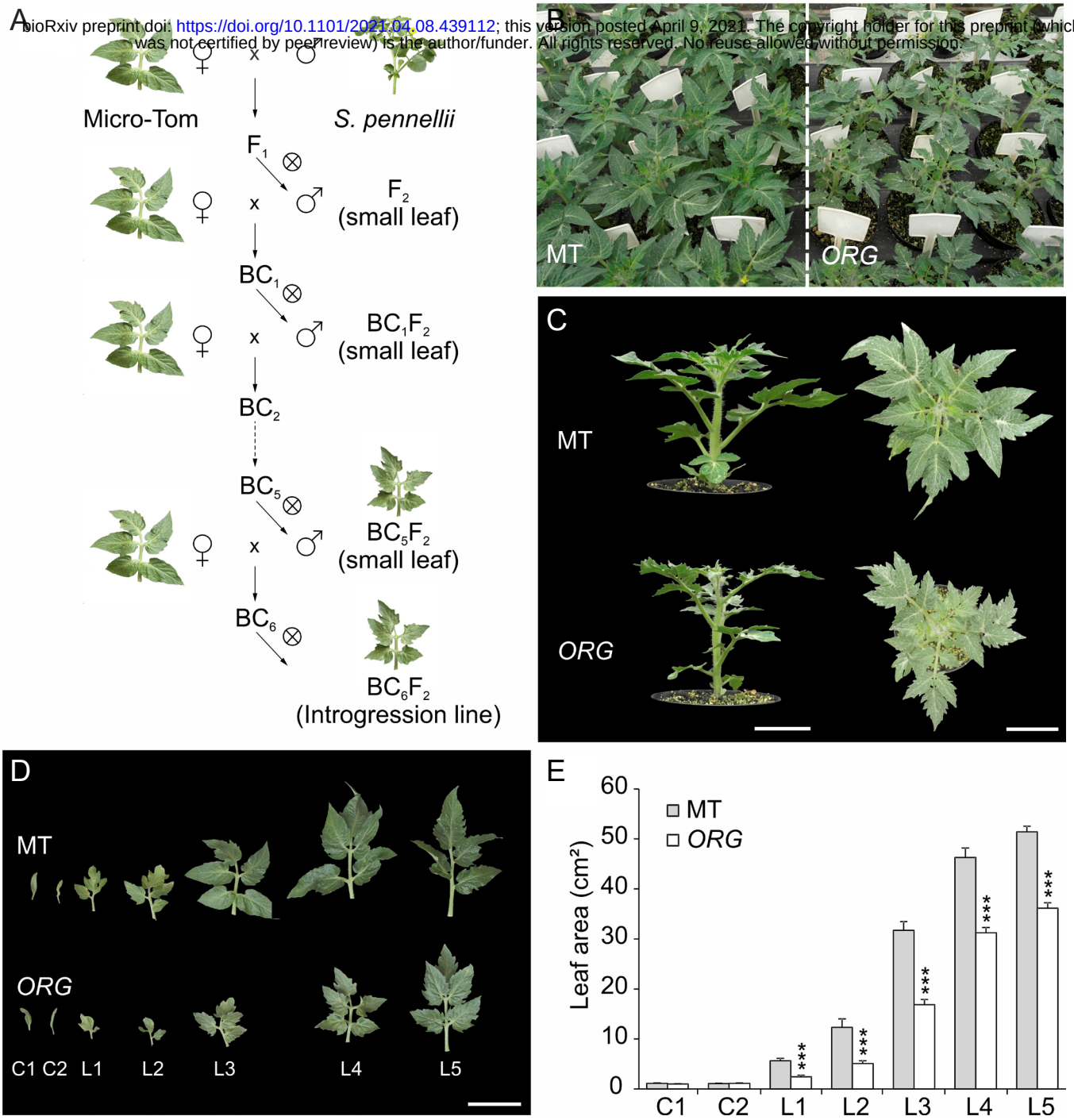
Figure 4. Altered patterns of gene expression in *ORG*. Time course of transcript levels of cell division- and expansion-related genes in ovaries/fruits of MT (gray bar) and *ORG* (white bar) genotypes. Relative (to actin control) transcript levels of *CYCB2;1* **(a)**, *FW2.2* **(b)**, *FW3.2* **(c)** and *EXP5* **(d)** in ovaries/fruit at -8, -4, 0, 4, 8 days and fruit pericarp at 12 and 16 days (anthesis = 0). Data are mean±s.e.m (n=3 biological replicates indicated with black dots). Statistical significance was tested by Student's *t*-test (*$p<0.05$, **$p<0.01$, ***$p<0.001$).

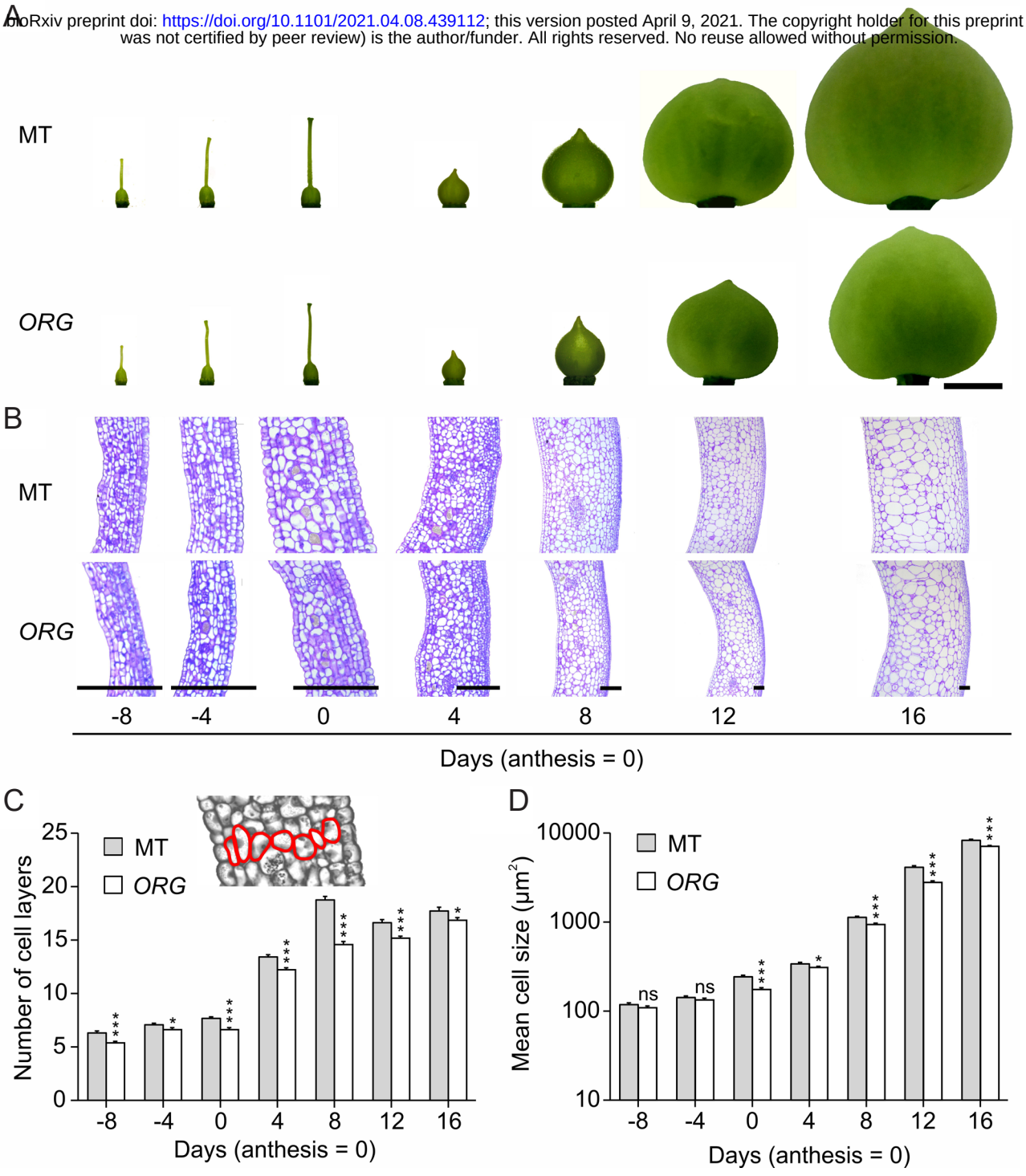
Figure 5. GBS defines the span of the introgression in the *ORG* introgression line (IL). (a) Genome-wide density of unique variants shared between *ORG* and *S. pennellii* LA716 in the genetic background of tomato cv Micro-Tom. **(b)** Close up view of the introgression on chromosome 7.

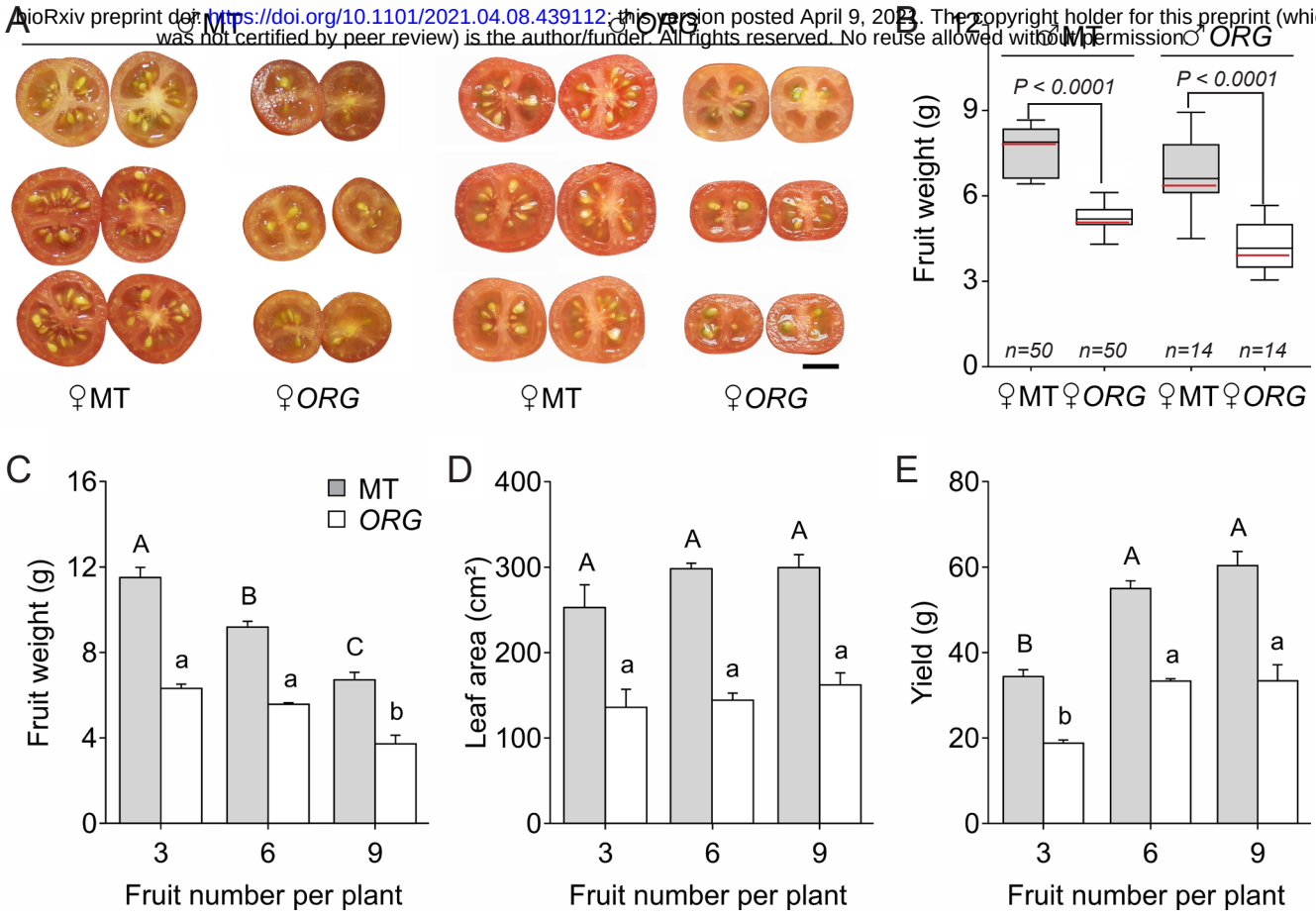
Figure 6. Mapping refines the candidate region for *ORG*. (a) Two introgression lines (ILs) in the tomato cv Micro-Tom (MT) background that contain different segments from *S. pennellii* on chromosome 7 (*Bco* and *Rg7H*) were mapped to refine the candidate region harboring the *ORG* locus (red segment). **(b)** Representative leaf of MT, *ORG*, *Rg7H* and *Bco* genotypes. Scale bar = 5 cm. **(c-d)** Leaf area (c) and ovary weight (d) of MT, *ORG*, *Rg7H* and *Bco* (n=10). Statistical significance was tested by Tukey's test ($p<0.05$). Different letters indicate significant difference between genotypes.

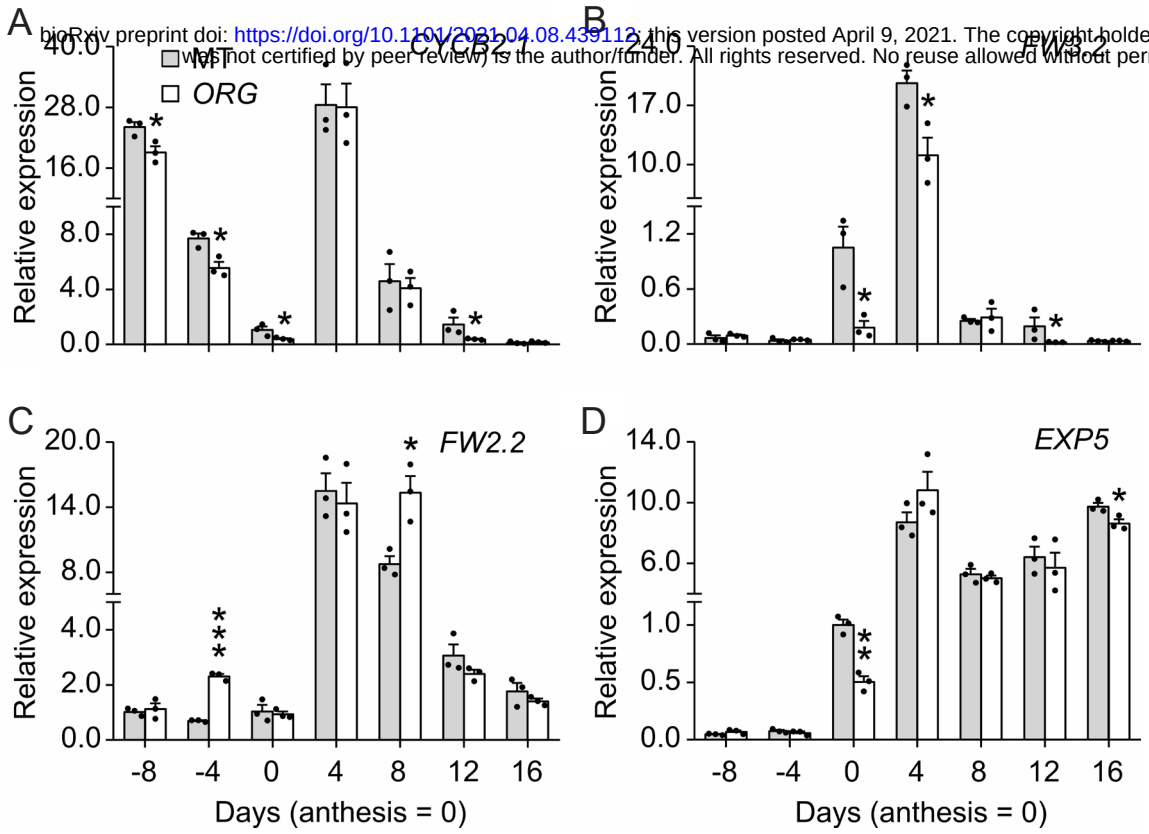
Figure 7. Analysis of the genomic region containing *ORG*. Terminal **(a)** and lateral **(b)** leaflet area of M82 and chromosome *7* introgression lines (ILs) from *S. pennellii*. Statistical significance was tested by ANOVA followed by Tukey's HSD test. Redrawn from Chitwood et al. (2014). **(c)** Chromosomal position of *S. pennellii* genomes segments in tomato cv. M82 background in chromosome *7*. The location of the *ORG* candidate region is shown in red. **(d)** Synteny plot of the coding sequences (CDS) within the *ORG* region between *S. lycopersicum* and *S. pennellii* genomes. The similarity between the CDS of *S. lycopersicum* (SL2.50) and *S. pennellii* (Spenn) were tested with BLAST+ and variant effect prediction was obtained from the resequenced dataset (Aflitos *et al.* 2014). Key: Dark green, CDS that match with a high level of similarity, but *S. pennellii* alleles contain single nucleotide polymorphisms (SNPs). Light green, *S. pennellii* alleles contain insertions and deletions (InDels). Red, *S. pennellii* alleles contain variants predicted to cause loss of function. Blue, complex relationship between *S. lycopersicum* and *S. pennellii* alleles, *i.e.* multiple matches between different genes. Grey, partial matches

819 between *S. lycopersicum* and *S. pennellii* alleles, *i.e.* CDS with conserved regions but otherwise
820 dissimilar. Black, genes present in *S. lycopersicum* or *S. pennellii* only.

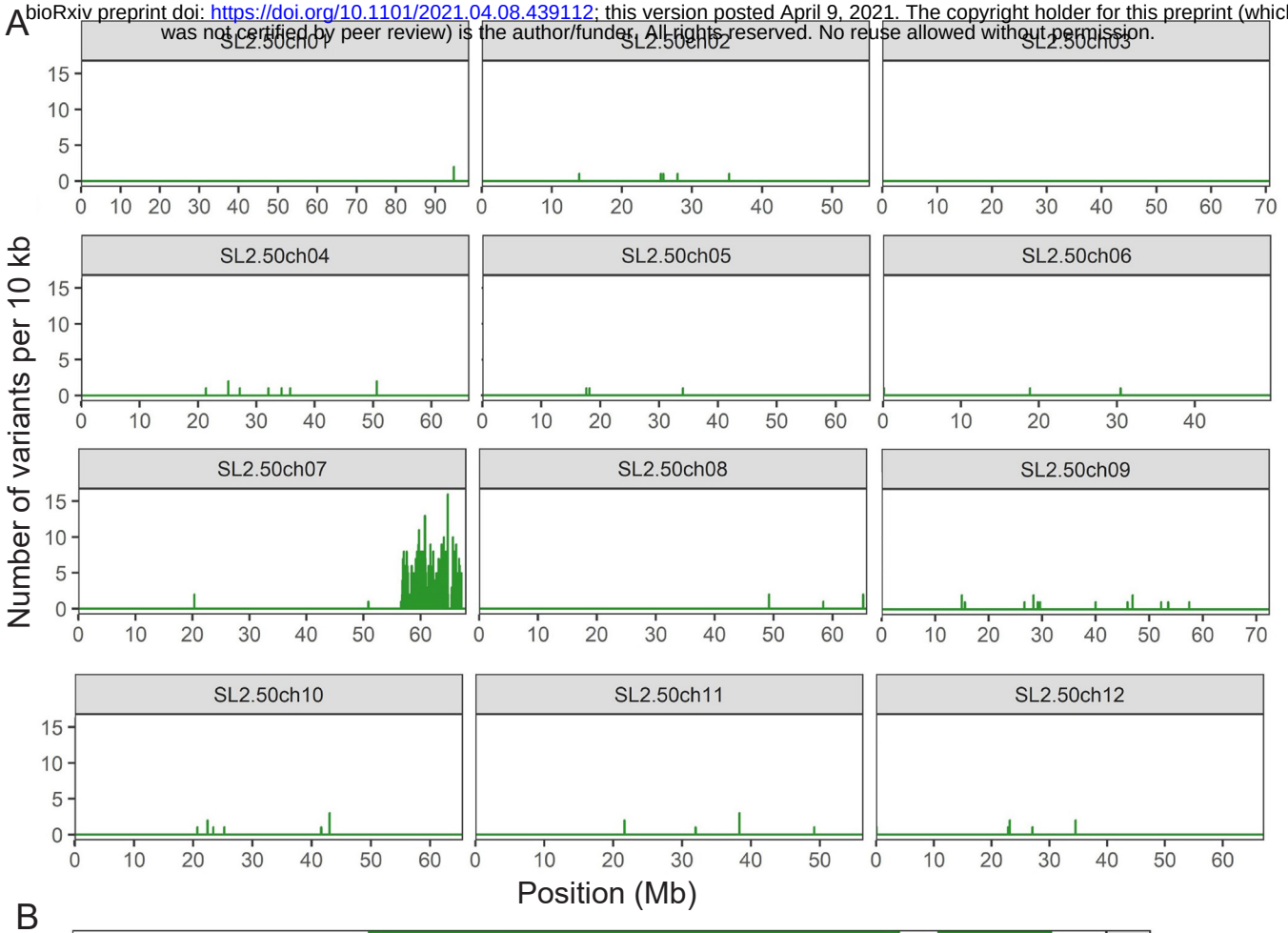








A



B