Rapid identification of causal mutations in tomato EMS populations via mapping-by-sequencing

Virginie Garcia[θ1] , Cécile Bres[θ1]*, Daniel Just[θ1], Lucie Fernandez[θ1], Fabienne Wong Jun Tai[1], Jean-Philippe Mauxion[1], Marie-Christine Le Paslier[2], Aurélie Bérard[2], Dominique Brunel[2], Koh Aoki[3], Saleh Alseekh[4], Alisdair R. Fernie[4], Paul D. Fraser[5], Christophe Rothan[1]

[1]Institut National de la Recherche Agronomique and Université de Bordeaux Unité Mixte de Recherche 1332 Biologie du Fruit et Pathologie, F-33140 Villenave d'Ornon, France. [2]Institut National de la Recherche Agronomique US1279 Etude du Polymorphisme des Génomes Végétaux, CEA-Institut de Génomique-CNG, F-91057 Evry, France. [3]Graduate School of Life and Environmental Sciences Osaka Prefecture University 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka, 599-8531 Japan. [4]Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, 14476, Potsdam-Golm, Germany. [5]School of Biological Sciences, Royal Holloway University of London, Egham, Surrey TW20 0EX, United Kingdom

*Corresponding author; email: cecile.bres@bordeaux.inra.fr; telephone: +33-557122545; fax: +33-557122541

Other authors;

virginie.garcia@bordeaux.inra.fr; daniel.just@bordeaux.inra.fr; lucie.fernandez@bordeaux.inra.fr; fabienne.wong@bordeaux.inra.fr, jpmauxio@bordeaux.inra.fr; le.paslier@cng.fr; berad@cng.fr; Dominique.Brunel@versailles.inra.fr; kaoki@plant.osakafu-u.ac.jp; alseekh@mpimp-golm.mpg.de; fernie@mpimp-golm.mpg.de; P.Fraser@rhul.ac.uk; christophe.rothan@bordeaux.inra.fr

[θ] These authors contributed equally to the establishment of the protocol described herein

Key words: tomato EMS mutant, causal mutation, forward genetics, mapping-by-sequencing

ABSTRACT

Tomato is the model species of choice for fleshy fruit development and for the *Solanaceae* family. EMS mutants of tomato have recently proved their utility for discovering new functions in plants leading to improved breeding stock for superior tomato varieties. However, until recently, the identification of causal mutations underlying remarkable phenotypes has been a very lengthy task that many labs could not afford due to spatial and technical limitations. Here, we describe a simple protocol for mapping-by-sequencing causal mutations in tomato, in which the phenotypes of interest are first isolated by screening a mutant collection generated in the miniature cultivar Micro-Tom. A recombinant F_2 population is then generated by crossing the mutant with a wild type (non-mutagenized) genotype and the F_2 segregants displaying the same phenotype are pooled. Finally, whole genome sequencing and analysis of allele distributions in the pools allows the identification of the causal mutation. The whole process takes 6 to 12 months, from the isolation of the tomato mutant to the identification of the causal mutation. This strategy overcomes many previous limitations, is of simple use and can be applied in most labs with limited facilities for plant culture and genotyping.

INTRODUCTION

Tomato is the model species of choice both for fleshy fruits and *Solanaceae*, in general. Considerable genetic tools and resources have been developed in tomato, increasing its use as a translatable model among the plant-based scientific community. In recent years, a high quality tomato genome sequence has been obtained for the inbred tomato cultivar 'Heinz 1706' which serves as a reference genome. The 900 megabases (Mb) tomato genome includes ~35 000 predicted protein coding genes supported by RNA sequencing data. More recently, whole genome sequences of an increasing number of cultivated tomato genotypes and wild tomato relatives have been made available. To date, more than 300 tomato cultivars and most tomato wild species have been sequenced to variable depths ([1-4]). These include deep sequencing of the Micro-Tom ([5]) and M82 ([3]) cultivars supporting the considerable genetic resources generated, such as collections of *S. pennellii* Introgression Lines (Ils) in M82 ([6]) and/or of fast-neutron and EMS (ethyl methanesulfonate) mutant collections for the M82 ([7]) and Micro-Tom ([8-13]) cultivars.

Tomato displays some specific features. For example, 727 gene groups found in tomato genome are limited to fleshy fruit species such as tomato, grape and potato and are not found in dry fruit species such as *Arabidopsis* ([1]). Conversely, most of the genes, pathways and other physiological processes are shared between tomato and other flowering plant species making Tomato can be an excellent model, for example for studying cuticle composition and properties ([14-19]) and carotenoid biosynthesis and regulation ([20-25]) because of the thick cuticle covering the fruit and of the large accumulation of carotenoids in the flower and in the ripening fruit. The successful use of tomato as a working model extends to other fields of plant biology, such as plant metabolism ([26-31]), plant architecture ([32-37]), leaf shape ([38-40]) and resistance to pathogens ([41]) and abiotic stress ([42,3]). In the post-genome era, the main bottleneck preventing the use of tomato as a more widespread model is the current paucity of tools available in tomato for linking genes to plant phenotypes. Two major approaches known as the reverse genetics and the forward genetics approaches are being used for linking gene to phenotype and assign a function to the genes in this species.

In the reverse genetics approach, a candidate gene possibly involved in the trait-of-interest is first identified and its function and role *in planta* is then analyzed. With a high quality tomato genome sequence available, the function of a large number of tomato genes can be predicted by *in silico* annotation and further inferred by the information on when and where they are expressed. In classical forward genetics approaches, tomato genetic resources (wild species, mutant collections, heirloom varieties) are first screened for phenotypes-of-interest and the underlying genes are subsequently identified by map-based (positional) cloning. In this process, the chromosomal region responsible for the trait variation is identified by looking for linkage between the trait studied, which can be controlled by a mendelian mutation or Quantitative Trait Loci (QTLs), and markers with known physical location in the genome. Map-based cloning of the underlying genes usually involves crossing the genotype carrying the trait studied with a genotype that does not and for which markers allowing the discrimination between both genotypes are available. Though the toolbox necessary to perform positional cloning in tomato has considerably improved in the last years, identifying the allelic variant responsible for the trait of interest can be a lengthy task. In addition, outcrossing the genotype studied with a distant phenotype may produce in the segregating population used to map the allelic variant a large phenotypic diversity which, depending on the trait studied, may prevent its detection.

In forward genetics screening, the large phenotypic diversity generated by mutagenesis is explored to identify mutants displaying the phenotypes of interest. Using the map-based cloning strategy to identify the causal mutations has the same limitations as those described above. The current advances in deep sequencing technologies provide one alternative strategy, which is to identify the allelic variant underlying the phenotype of interest by whole genome sequencing. The so-called mapping-by-sequencing strategy is very well adapted to tomato thanks to the genetic resources and tools already developed in this crop species. It considerably shortens the time necessary to identify a causal mutation, can be performed by most labs with limited expertise in tomato and in genotyping and, when the miniature tomato cultivar described in the protocol is used, requires a limited space for plant culture. The protocol described should therefore contribute much to the use of tomato as a plant model.

Using Artificially-induced genetic diversity

Artificially-induced genetic variability has been used for decades to generate new valuable traits and alleles, resulting in the creation of commercial cultivated varieties in many crop species, including tomato ([43]). Several physical or chemical mutagenic agents can be used to generate genetic and phenotypic diversity in crop plants, the most commonly used being fast-neutron or gamma-ray bombardments or EMS. EMS mutagenesis creates point mutations evenly distributed with high density over the whole genome thus generating large genotypic and phenotypic diversity in relatively small populations. Technological advances in the last 20 years triggered a renewed interest for the generation of EMS mutant populations in crop species, mainly because of the development of the TILLING (Targeting Induced Local Lesions IN Genomes) strategy, whereby allelic series for a given gene can be found by screening EMS mutant collection ([44]). Many crop plants for which functional analysis of target genes was difficult until then became therefore amenable to studies involving reverse genetics approaches. At the same time, because of the large phenotypic diversity created by EMS mutagenesis, which usually is much beyond the natural genetic diversity found in cultivated species, many of the mutant collections were simultaneously screened for phenotypic alterations. In tomato, EMS mutant collections have been generated in several determinate and indeterminate cultivars ([7,11,13,45-47]), including the miniature cultivar Micro-Tom well suited for genomic studies ([8,9]). Databases describing tomato mutant phenotype have been made publicly available for the processing tomato M82 cultivar ([7]) and for the Micro-Tom cultivar ([11]). Mutants for specific traits in uniform genetic background can be ordered from these collections. In addition, single-gene mutations in various backgrounds affecting a wide range of plant and fruit characters are available at the Tomato Genetics Resource Center.

Development of the protocol

A forward genetics approach in tomato (*Solanum lycopersicum*), specifically focusing on the forward genetic screening of tomato EMS mutants was the approach used in the present studies. We recently showed how two successive rounds of EMS mutagenesis (**Fig. 1**) can produce very high mutation frequencies in tomato, up to 1 mutation/125 Kb as demonstrated by TILLING experiments ([13,46,48]). The use of the

miniature cultivar Micro-Tom for mutagenesis further allows the screening of hundreds or thousands of M_2 plants for major agronomic traits such as plant and flower truss architecture, yield and fruit characteristics (shape, weight, colour, ripening). Screening can be done in a limited space available in many plant labs. We successfully applied this strategy to the study of fruit cuticle mutants and showed that systematic visual screening of a fraction of the collection for fruit colour (pink fruits) and more subtle changes (glossy fruits) can readily lead to the identification of tens of mutants for cuticle traits ([18,19]). The genes underlying two cuticle mutants were further identified by mapping the mutant traits ([18,19]).

Mapping the mutant trait with the help of markers is usually the first step after mutant isolation in forward genetics screening of tomato EMS mutants ([16,18,19,24,33,37,49]). The causal mutation is then identified through several approaches including literature mining for candidate genes and gene sequencing ([16,18,19,24,37,49]), RNA seq analysis ([37]) or whole genome sequencing ([24]). In this classical scenario, the genetic mapping of the mutation represents a major bottleneck since there is the necessity to outcross the mutant, usually with related wild species or at least with cultivated varieties for which genetic markers are available. Outcrossing tomato varieties well adapted to mutagenesis produces large phenotypic variability in the segregant population used for identifying the causal mutation. When taken together with the space required for growing the segregant population (fields, greenhouses etc.), this fact has prevented many in the plant science community from using tomato mutants.

In the recent years, taking advantage of the current possibilities offered by deep sequencing technologies, several forward genetics screening strategies based on whole genome sequencing have been designed in the model plant species *Arabidopsis* and rice (recently reviewed in ([50]). Among these are the SHOREmap ([51]), NGM ([52]) and MutMap ([53-56]) methods, which use very similar approaches and tools. These methods, termed here mapping-by-sequencing, are based on phenotypic screening of EMS mutant collections and subsequent bulked-segregant analysis of F_2 progeny derived from a cross between a mutant-of-interest and a wild-type (non-mutagenized) parental line. Whole genome sequencing of bulked F2 plants displaying the same phenotype and alignment to a reference genome allows the identification of single nucleotide polymorphism (SNP) variants. The SNP variants used as DNA markers are derived from the outcross of the mutant with a distantly

related genotype ([51,52]) or from the backcross of the mutant with the parental wild-type line ([53,55]). In that later case, the DNA polymorphism observed is only due to mutagenesis. The analysis of SNP frequencies is then used to delineate the chromosomal region harboring the causal mutation ([50-53]).

Following the recent publication of a high quality tomato genome sequence from Heinz 1706 ([1]), our early attempts to identify causal mutations in Micro-Tom by mapping-by-sequencing were unsuccessful. We used the published automated pipelines NGM ([52]) and MutMap ([53]), which require less than 100 F2 progeny for bulking. However, a very high background noise prevented the identification of the regions harboring the causal mutations. Several reasons likely explain this negative result. Analysis of the tomato reference genome ([1]) emphasized the importance of gene duplications in tomato. Therefore, to improve mapping of the reads to the reference sequence, whole-genome sequencing of the bulks is now done by producing Paired-End (PE) Illumina reads. In addition, genome sequencing of Micro-Tom demonstrated the presence of as many as 1.23 million SNPs and 0.19 million indels between Micro-Tom and Heinz 1706 reference genome ([5]). The results were recently confirmed by the sequencing of several Micro-Tom wild-type and mutant lines. In addition, copy number variation, deletions and insertions of chromosomal segments were further detected in both the mutant and wild-type lines ([57]). To take into account that natural polymorphism, one way is to construct a Micro-Tom reference sequence by replacing nucleotides in the Heinz 1706 reference sequence with those polymorphic in the mutagenized cultivar, as was done in rice ([53,56]). Considering the lack of *de novo* Micro-Tom genome sequence assembly, the high natural polymorphism between Micro-Tom and Heinz 1706, and the variability of Micro-Tom WT lines ([5,57,58]), we preferred to use the Heinz 1706 sequence as the reference tomato sequence. Read mapping and variant calling could be performed using standard bioinformatics tools (BWA[59] and SAMtools[60]). However, before filtering using allelic frequencies, we included in the procedure a step designed for identifying and removing natural polymorphism between Micro-Tom and Heinz 1706.

The simple protocol described allowed identification of the causal mutations for several tomato mutants affected in ascorbate regulation, fruit size, cuticle formation and, provided here as an example, the carotenoid pathway gene *PSY1* (*Phytoene synthase 1*) that controls fruit colour. One additional advantage of this approach over

fine-mapping strategies classically used is that it does not require outcrossing the mutant. It thereby avoids the large diversity in segregating F_2 population associated with such crosses and considerably increases the precision of the phenotyping.

Applications of the method

Tomato EMS mutant collections have already demonstrated their utility in the TILLING identification of allelic series for genes involved in virus resistance, hormonal signaling, central and secondary metabolism and plant development ([34,35,45,46,48,61,62]). Recently, forward genetics screening of tomato mutant collections also proved extremely useful for discovering new functions, not yet known in plants, such as cutin synthase involvement in cutin polymerization in the cuticle ([16,19]) and the role of the phytyl ester synthase (PES1/PYP1) in carotenoid sequestration in the plastoglobules ([24]). Tomato mutants can also help shed new light on old questions in plant science such as the mechanisms of heterosis in the control of fruit yield ([33]). A remarkable paper recently published by ([37]) further showed how screening EMS mutants for suppression of bushy and determinate growth habits of tomato allowed the isolation of new alleles which, when combined, modified the architecture from the plant and considerably increased fruit yield.

Protocols described in this manuscript for linking phenotype to genotype in Micro-Tom tomato by using EMS mutagenesis and mapping-by-sequencing can be applied to other tomato cultivars with minor modifications (**Box 3**). Among these are the M82 cultivar, for which both mutant resources ([7]) and high quality sequence ([3]) are available, but also several other processing tomato cultivars for which mutant resources but not genomic sequences are available ([45,47,62]). In addition, EMS mutagenesis is increasingly used to generate genetic diversity in fresh market tomato genotypes. One of the reasons is that these varieties are considerably different from processing tomatoes with respect to plant architecture (indeterminate growth), cultural practices (greenhouse-types) and fruit morphology and composition. The advantage of applying the forward genetics screening strategy directly to the elite lines is to reduce the complexity and time which would be necessary to introgress the allele of interest from distantly related genotypes. Although the procedure is described for the identification of a recessive causal mutation, it can also be used for dominant mutations. Specific applications to the identification of a dominant mutation

and the use of different tomato genetic backgrounds are further documented in **Box 3**.

Advantages and limitations

The protocol described uses the miniature tomato cultivar Micro-Tom, which presents several advantages for EMS mutagenesis and forward genetics screening: (*i*) it can be grown at high density in greenhouse or growth chamber *i.e.* less than 4 square meters are necessary for the culture of a segregating population of 600 plants; (*ii*) short cycle of three months allows four generations per year; (*iii*) large phenotyped mutant collections are publicly available ([11]); (*iv*) deep whole genome sequencing of Micro-Tom has been done recently ([5]). The Micro-Tom/Heinz 1706 variants are listed ([5]) and their type and effect on protein function of EMS mutations have been analyzed ([61,57,63]). As a consequence, the mapping-by-sequencing approach using Micro-Tom considerably reduces the cost for identifying the causal mutation. For a crop species like tomato, plant culture can be a major limitation. Cultivating a population of 600 greenhouse-type tomatoes may require a whole greenhouse of up to 350 m^2 during 6 months. The corresponding cost can be prohibitive when working on various mutants.

One of the limitation of using Micro-Tom is that this cultivar is a miniature determinate mutant carrying a mutation in the brassinosteroid biosynthesis pathway ([64]). This may affect the plant architecture and several traits related to hormone regulation in the vegetative tissues. Note that such features are not specific to Micro-Tom and that all cultivated genotypes may carry deleterious mutations in several genes ([65,63]). If a strong interaction between the mutated allele and the genetic background is likely, it is advisable to cross the mutant with another tomato cultivar (*e.g.* an indeterminate tomato variety) and study the phenotype-of-interest and allele effect in the F_1 and F_2 generations. It is also important to keep in mind that most phenotypic changes induced by EMS are due to loss-of-function mutations. Therefore, screening EMS mutant populations for identifying alleles responsible for gain-of-function mutations will probably remain unsuccessful. For example, extreme phenotypes such as the elongated fruit phenotype found in the *sun* mutant, which is due to a gene duplication event mediated by a retrotransposon ([66]) will likely never occur in tomato EMS mutant populations.

Overview of the procedure

There are a number of studies and reviews already dealing with EMS mutagenesis and TILLING in tomato and in other plant species ([11,44,45,46,48,61,62]); hence, we will succinctly describe the steps necessary to obtain a highly EMS mutagenized tomato collection and mostly restrict the protocols to points that are essential for the completion of the procedure for mapping-by-sequencing a mutation, starting with the isolated mutant carrying the trait-of-interest to the successful identification of the causal mutation (see Overview of the Experimental Design in **Fig. 1** and **Box 1**).

Briefly, following the generation of the EMS mutant population and the selection of the mutant plant carrying the trait-of-interest, the genetic inheritance of the mutation is analysed. A homozygous mutant is selected and crossed with the wild-type Micro-Tom line used for mutagenesis to produce a BC_1F_1 plant which is selfed. The phenotype of the BC_1F_2 population plants is then analysed. Two bulks displaying respectively the mutant-like phenotype and the wild-type like phenotype are constituted and sequenced (paired-end Illumina sequencing) to a read depth of ~20 to 40X. Sequences are first trimmed on base quality criteria and subjected to bioinformatics analysis using the BWA alignment tool ([59]) for mapping the reads onto the tomato reference genome (Heinz 1706) and the SAMtools ([60]) for variant calling. The resulting variant files contain all variants (SNPs + indels) between the Micro-Tom mutant and the Heinz 1706 reference genome, including EMS-induced mutations and natural polymorphism. If not done previously, the sequencing of the wild type Micro-Tom parental line allows the inventory of all natural Micro-Tom/Heinz polymorphisms. These are then subtracted to leave the EMS-induced polymorphisms mapped onto the reference genome. Analysis of allelic frequencies from SNP variants in the two bulks next allows the identification of a chromosomal region with higher than average SNP frequency in the mutant-like bulk (>0.5 and close to 1 in the vicinity of the causal mutation) and lower than average SNP frequency in the WT-like bulk (<0.5 and close to 0.33 in the vicinity of the causal mutation). Once the chromosomal region carrying the causal mutation has been precisely identified, the determination of variant types (including the indels) is done and false positives are excluded. If necessary, in genomic regions with high linkage disequilibrium (LD) or with high mutation density, a recombinant analysis is performed to identify unequivocally the causal mutation.

Towards this end, the EMS-induced SNPs are used as markers to detect recombination events in the BC_1F_2 segregating population previously generated.
The protocol described does not require any specific expertise since it is simply based on mutation frequency analyses in mutant bulks and makes uses of tools available on the web and of simple scripts provided. It will thus likely contribute much to the use of tomato as a model plant species.

Experimental design

Backcrossing the mutant. We advise to use the mutant as male to control that the backcross was successful. Despite the precautions taken to insure that cross-pollination is effective, the possibility of a self-pollination is not totally excluded. If the mutant is used as a female and that self-pollination occurred, all the BC_1F_1 plants would display the mutant trait when the mutation controlling the phenotype-of-interest is homozygous recessive in the mutant parent. The same result would be obtained when a homozygous dominant mutant is backcrossed with Micro-Tom line. Distinguishing both cases and determining if the backcross was successful would therefore require additional experiments.

We prefer to use the cone of anthers for pollination because pollen can be difficult to collect in the high hygrometry conditions often prevalent in greenhouses. The cone of anthers further protects from undesirable pollen contamination. Alternatively, when hygrometry is low, pollen can be easily collected from the mutant and used to hand pollinate the Micro-Tom line. In that case, a paper bag covering the flower should be used to avoid any spurious pollination. However, due to the small size of plant and flower trusses from the Micro-Tom cultivar, this may be tricky and lead to flower abortion.

Size of the segregant population and constitution of the bulks. In our design (step 3), only half of the BC_1F_2 individuals presenting the mutant-like phenotype are included in the mutant-like bulk. We take this precaution of analyzing a large segregating BC_1F_2 population for two reasons: (*i*), some problems may arise during the culture (low germination rate, fungal attacks…). Therefore, performing a clear cut phenotyping of the mutant trait can be difficult, *e.g.* in case of incomplete penetrance of the mutation; (*ii*), the number of pooled individuals chosen represents a good compromise between the minimum sequencing depth required (20-40X tomato

genome coverage) and the large number of individuals necessary to lower the incidence of phenotyping errors on subsequent mutation frequency analysis. Considerations on the incidence of misclassification of the phenotype on the detection of the causal mutation have been detailed for the MutMap method in rice ([53]). In addition, a large number of individuals in the BC_1F_2 population will considerably reduce the time necessary for subsequent recombinant analysis, if necessary, because the DNA can readily be extracted from the individual plants displaying the mutant phenotype.

Recombinant analysis. Successful identification of the causal mutation through mapping-by-sequencing depends on the density of EMS mutations and on the extent of linkage disequilibrium (LD) in the region of interest. For tightly linked EMS mutations, similar allelic frequency patterns are expected, limiting the direct identification of the causal mutation in a single step through mapping-by-sequencing. In this case, further confirmations using recombinant analysis are required to identify the mutation at the origin of the phenotype of interest. For this purpose, fine linkage mapping using BC_1F_2 population is proposed as an alternative to identify recombination events without the need to increase the read depth (DP).

The first step for recombinant analysis includes the production of a segregating population phenotyped for the trait of interest. The same BC_1F_2 segregating population previously used for the mapping-by-sequencing procedure can be used if the population was maintained or if plant material for DNA extraction was already collected from each F_2 individual. Otherwise a newly sown and phenotyped BC_1F_2 population or BC_1F_3 population produced from the selfed BC_1F_2 individuals (if corresponding seeds were collected) needs to be developed. The production of a new population will require at least 4 months for a complete analysis.

The second step is plant material harvesting and DNA extraction from each BC_1F_2. DNA extraction in 96-well plates is recommended for large-scale analysis. The DNA extraction can be performed using the procedure described in the protocol for DNA extraction from bulks (steps 4 to 10). Mutant and WT Micro-Tom lines should be included in the analysis as controls of the homozygous mutant allele and wild type allele, respectively.

In the third step, the EMS mutations detected by the mapping-by-sequencing procedure are used as markers for genotyping the BC_1F_2 individuals. This can be

done by using the KASP assay, which is a cost effective, easy to implement and efficient method to genotype SNPs on a large scale ([67]). Specific primer design can be done using batchprimer3 software ([68]; http://probes.pw.usda.gov/batchprimer) and KASP genotyping procedures (LGC Genomics, Queens Road, Teddington, Middlesex, TW11 0LY, UK http://www.lgcgroup.com/our-science/genomics-solutions/genotyping/). The recombinants are detected by a change in the genotypic state (homozygous WT, homozygous mutant or heterozygous) at the different EMS mutation loci (**Fig. 2C**). The causal mutation corresponds to the only one locus homozygous for the mutant allele in all the BC_1F_2 individuals displaying the mutant phenotype. Number of BC_1F_2 individuals analyzed needs to be increased if no recombination was detected between two linked candidate mutations. The size of the population required for this analysis should be adapted according to the physical distance between the linked EMS mutations and their chromosomal localization. Information on low- or high-recombination regions in tomato is available from the sequenced reference genome ([1]).

BOX 1 |

MATERIALS

PLANT MATERIAL

To avoid confusion we use a strict nomenclature for plants:

Micro-Tom line: non-mutagenized Micro-Tom

WT-like bulk: pooled BC_1F_2 individuals that exhibit the WT phenotype

Mutant-like bulk: pooled BC_1F_2 individuals that exhibit the mutant phenotype

REAGENTS

- Liquid nitrogen for freezing samples.

! CAUTION Liquid nitrogen is a low temperature refrigerant and should be handled with protective glasses and gloves. In addition, liquid nitrogen is not used in closed rooms for danger of suffocation.

- Sorbitol (SIGMA, cat. no. S6021)
- NaCl (Euromedex, cat. no. 1112)
- Tris base (***Sigma***-Aldrich, cat. no. T1503)
- EDTA Ethylenediaminetetraacetic acid (Eurobio, cat. no. GAUEDT0066)

- CTAB cetyl-trimethyl-ammonium bromide (SIGMA, cat. no. H6269)
- Sarcosyl (***Sigma***-Aldrich, cat. no. 61743)
- Sodium bisulfite (***Sigma***-Aldrich, cat. no. 243973)
- Chloroform GPR Rectapur (VWR International, cat. no. 22706.292)

! CAUTION Chloroform is toxic and should be handled under a fume hood.

- Isoamyl alcohol (***Sigma***-Aldrich, cat. no. I9392)
- 2-propanol (VWR International, cat. no. 20842.312)

! CAUTION 2-propanol is inflammable and should be handled under a fume hood.

- RNAse A (Macherey-Nagel cat. no. 740505)
- Electran agarose DNA grade (VWR International, cat. no. 438794L)
- 6X loading dye (Promega, cat. no. G190A)
- Gelgreen nucleic acid gel stain (Interchim cat. no. CJ2730)
- 2-Log DNA Ladder 0.1–10.0 kb (Ozyme, cat. no. N3200L)
- Quant-iT™ dsDNA Assay Kit, broad range/high sensitivity (Invitrogen, cat. no. Q-33130/ Q-33120)
- Tube for Covaris focused-ultrasonicators: used for this study, microTUBE 6 x 16mm Round Bottom glass tube, AFA fiber, and cap system, 100ul (Covaris, cat. no. 520045)
- Illumina TruSeq DNA PCR-Free LT Sample Preparation Kit (Illumina, cat. no. FC-121-3001/3002).
- High Sensitivity DNA Kit/Reagents (Agilent Technologies, cat. no. 5067-4626/5067-4627)

EQUIPMENT

- 15 ml tube 120/17 (Sarstedt cat. no. 62.554.502)
- Pipets 2-10-20-200-1000 µl (Gilson, Inc)
- Tips for pipet, fine quality.
- Waterbath at 65°C
- Centrifuge machine with swinging bucket rotor for 15 ml tubes (Sigma Laborzentrifugen cat. no. 2-16K)
- Gel Doc EZ Imager (Bio-Rad laboratories)
- NanoDrop ND-1000 UV-VIS spectrophotometer version 3.2.1 (NanoDrop Technologies)

- Fluorescence Microplate Reader Gemini (Molecular Devices)
- A Covaris focused-ultrasonicator model E210 (Covaris)
- 2100 Bioanalyzer (Agilent Technologies, cat. no. G2940CA)
- A compatible Illumina DNA sequencing instrument and associated equipment. HiSeq2000 with cBot were used for this study (Illumina, cat.no SY-401-1001/ SY-301-2002)
- Computer operating systems: Windows XP, GNU Linux or Mac OS X
- BWA v0.7.12 (http://bio-bwa.sourceforge.net/): software package for mapping low-divergent sequences against a large reference genome, such as the human genome[90]
- SAMtools v1.2 (http://samtools.sourceforge.net/): provides various utilities for manipulating alignments in the SAM format, including sorting, merging, indexing and generating alignments in a per-position format[91].

REAGENT SETUP

- **CTAB extraction buffer** 145 mM sorbitol, 850 mM NaCl, 125 mM Tris-HCl [pH 8.0], 25 mM EDTA, 0.8% [w/v] Sarcosyl [n-lauroyl-sarcosine], 0.8% [w/v] CTAB.
  ! CAUTION the solution should be prepared fresh.
- **Chloroform:isoamyl alcohol (24:1)** made by mixing isoamyl alcohol in chloroform. Store at 4°C several months.
  ! CAUTION Reagents are extremely toxic and should be handled under the fume hood.
- **50X TAE buffer** 2M Tris Acetate 500mM EDTA. Store at room temperature several months.
- **1X TAE buffer** solution made by diluting (1/50) 50X TAE buffer stock in H_2O. Store at room temperature several months.

PROCEDURE

Production of segregant BC_1F_2 population ● TIMING ~6 months

1| **Backcross.** Cross the Micro-Tom line (female) with the EMS mutant (male) displaying the phenotype-of-interest. For crossing, dissect unopened flower (stage 16-18[69]) from female plant to remove immature anthers and prevent self-pollination without damaging the pistil. Cross-pollinate the female flower using flower at anthesis (freshly opened) from the male (mutant). The cone of anthers removed from the male flower is deposited over the style-stigma of the female. It will protect internal floral organs during pollination and fertilization. BC_1F_1 seeds are then collected on red ripe fruits. . **? TROUBLESHOOTING**

2| **Phenotyping of BC_1F_1 individuals and self-pollination.** In order to determine whether the mutation is recessive or dominant, analyze 8 to 10 BC_1F_1 individuals by segregation analysis. Analyze EMS mutant and Micro-Tom line in the same culture, as controls, to accurately phenotype the mutant trait. Segregation ratios are indicative of the genetic inheritance and of the state of the mutation in the parental EMS mutant:

- Case 1: for a recessive mutation at homozygous state in the parental mutant, 0% of mutant-like phenotypes would be observed in the BC_1F_1 individuals
- Case 2: for a dominant mutation at homozygous state in the parental mutant, 100% of mutant-like phenotypes would be observed in the BC_1F_1individuals
- Case 3: for a dominant mutation at heterozygous state in the parental mutant, 50% of mutant-like phenotypes would be observed in the BC_1F_1individuals

As the following procedure is focused on the analysis of recessive mutations, collect only BC_1F_2 seeds from self-pollinated individuals corresponding to case 1.

3| **BC_1F_2 segregant population and bulk constitution.** Produce and phenotype segregant population of 500 BC_1F_2 individuals. Analyze EMS mutant and Micro-Tom parental lines in the same culture, as controls, to accurately phenotype the mutant character. Around 125 individuals (25%) are expected to present the mutant-like phenotype in agreement with Mendel's laws involving a single locus. Pool 60 (+/-20) individuals presenting clear mutant-like and WT-like phenotypes to constitute the mutant-like bulk and the WT-like bulk, respectively.

. **? TROUBLESHOOTING**

Isolation and quantification of genomic DNA from the two bulks of BC_1F_2 plants

● **TIMING ~4 h**

4| Harvest with a punch of 5 mm diameter 5 leaf disks from each BC_1F_2 plant (~600 mg Fresh Weight) included in the mutant-like bulk and the WT-like bulk, and from Micro-Tom line.

5| Pool it in 15 ml tube. Freeze immediately in liquid nitrogen. Transfer into a cold mortar, add liquid nitrogen and grind to a fine powder with a pestle. Transfer again in a 15 ml tube.

6| Extract genomic DNA by a cetyl-trimethyl-ammonium bromide (CTAB) method. Add 4 ml of CTAB extraction buffer with 2.2 mg of sodium metabisulfite and RNase A at 120µg/ml per sample. Incubate at 65°C for 90 min in a waterbath.

7| Add 3.6 ml of chloroform:isoamyl alcohol (24:1) to each sample, mix by inversion and centrifuge for 10 min at 4500g. Subsequently, pipet and transfer the supernatant to a fresh tube. **<CRITICAL STEP>** be careful to pipet only the supernatant.

8| Treat with 4 ml of cold 2-propanol. Dry DNA 15 min at 45°C under vacuum, suspend in 200 µl of distilled water and samples. **<PAUSE POINT>** Genomic DNA samples can be stored at -20 or -80°C for months to years.

9| To control the genomic DNA quality and integrity, add 5µl of the DNA solution to 1µl of 6X loading dye and load onto 1% (w/v) agarose TAE gel. Also load 5µl of 2-log DNA ladder into a lateral well.

10| Quantify using a Nanodrop spectrophotometer, or equivalent.

Libraries construction and Next-Generation Sequencing ● TIMING ~4 weeks

11| Prior to libraries construction, perform a quality control (QC) for all DNA samples *e.g.* by checking samples for concentration by fluorometric measurement with Quant-iT™ dsDNA assay kit and sample quality (absorbance, electrophoretic profil on agarose).

12| Prepare Illumina paired-end shotgun indexed libraries: by using the TruSeq DNA PCR-Free LT Sample Preparation Kit following the manufacturer recommendations. In order to shoot any trouble during the library construction, we

recommend to implement QC on Agilent HS DNA chip at each step *i.e* fragmentation, sizing, ligation and enrichment if performed, despite the increase of the total cost.

13| Conduct Illumina sequencing using an Illumina sequencer in PE (Paired-end) mode, with read length and lane number adapted to the expected depth: Process the raw data fasta Q files for data quality control: (traceability, cleaning (N, redundancy) and trimming on base quality criteria (Phred value=30).

Bioinformatic analysis ● TIMING ~1 to 2 weeks

Following Illumina HiSeq2000 sequencing, two sequence read files (R1 and R2 corresponding to the paired-ends) are available for each bulk. R1 file is sample_R1.fastq.gz. R2 file is sample_R2.fastq.gz. Sample refers to the sample name.

All scripts are provided as supplementary information (1 to 4).

14| Index the tomato reference genome *Solanum lycopersicum* v2.50. Genome index file allows simplifying and accelerating further genome mapping.

- bwa index S_lycopersicum_chromosomes.2.50.fasta

15| Map the reads on the tomato reference genome *Solanum lycopersicum* v2.50 (ftp://ftp.solgenomics.net/tomato_genome/wgs/assembly/build_2.50/S_lycopersicum_chromosomes.2.50.fa.gz).

- bwa aln S_lycopersicum_chromosomes.2.50.fasta sample_R1.fastq.gz -f sample_R1.sai
- bwa aln S_lycopersicum_chromosomes.2.50.fasta sample_R2.fastq.gz -f sample_R2.sai
- bwa sampe -a [INSERT_SIZE] S_lycopersicum_chromosomes.2.50.fasta sample_R1.sai sample_R2.sai sample_R1.fastq.gz sample_R2.fastq.gz | samtools view -Sb - | samtools sort – sample

16| Create a BAM index file (optional). Index file is usually needed when visualizing a BAM file in a genome browser/viewer (like IGV Integrative Genome Viewer[70,71] http://www.broadinstitute.org/igv/home). It simplifies and accelerates search operation.

- samtools index sample.bam

17| Perform the variant calling using SAMtools. Variant calling is used to identify the polymorphisms (SNPs and indels) between the Micro-Tom genome and the reference genome *Solanum lycopersicum* v2.50.

- samtools mpileup -B –t DP -u -f S_lycopersicum_chromosomes.2.50.fasta sample.bam | bcftools call –O v –v -c - > sample.vcf

As the tomato reference genome (Heinz 1706 cultivar) used to map the reads is different from that of the Micro-Tom cultivar, the variants identified would include both Heinz 1706 /Micro-Tom natural polymorphisms and EMS mutations. In this context, additional sequencing to a minimum depth of 20X of the Micro-Tom line can be required to take into account and further remove all natural polymorphisms between Micro-Tom and Heinz 1706. Thus, steps 15 to 17 are to be applied to the Micro-Tom line, mutant-like bulk and WT-like bulk. The Micro-Tom line output file (.vcf) would include all variants (SNPs + indels) corresponding to natural polymorphisms between Micro-Tom and Heinz 1706 (reference genome used to map the reads). The two .vcf output files obtained for the mutant-like and WT-like bulks would include variants (SNPs + indels) corresponding to natural polymorphisms between Micro-Tom and Heinz 1706 and also to EMS mutations. The output VCF file includes various quality parameters relevant to sequencing and mapping that can be used to subsequently filter the variants.

Note: alignment and variant calling using bwa and SAMtools can also be performed using Galaxy Web based platform ([72-74]).

18| Perform the variant analysis. This step includes the allelic frequency computation, variant type determination and exclusion of false positives

The *compare_WT_mutant_samtools_vcf_v5.py* script computes the allelic frequency and compares the .vcf files of the mutant and WT-like bulks to produce two output files for variants between the both bulks. Only one variant type at a time would be considered at this step (SNPs or indels)

- python compare_WT_mutant_samtools_vcf_v5.py mutant.vcf WT.vcf [snp] common_file.csv

We recommend focusing the analysis on SNP variants which represent 98% of the EMS mutations ([57]). Most of indels identified through commonly used bioinformatic

pipelines correspond to false positive variants due to mapping error in low read coverage regions or to complex and repetitive genomic regions. We suggest analyzing indels only after the chromosomal region carrying the causal mutation has been identified, especially when no causal EMS mutation has been detected through the SNP variant analysis. In this last case, only indels identified in the genomic region carrying the mutation (previously defined through SNP variant analysis) should be analyzed. **? TROUBLESHOOTING**

19| Annotate the variant positions according to the tomato reference genome annotation *ITAG2.4_gene_*models using the script *annotate_csv_with_solyc.py* and the GFF file downloadable here: ftp://ftp.solgenomics.net/genomes/Solanum_lycopersicum/annotation/ITAG2.4_release/ITAG2.4_gene_models.gff3. This step allows identifying the variants present in the annotated genes available in the .gff3 file.

- python annotate_csv_with_solyc.py common_file.csv ITAG2.4_gene_models.gff3 common_file_annotated.csv

20| Implement the .csv files with the Micro-Tom line .vcf file using the script *snp_present_in_other.py*. This step allows discriminating natural polymorphisms between Micro-Tom and Heinz 1706 from specific EMS mutations.

- python snp_present_in_other.py common_file_annotated.csv [FILE_LIST] common_file_enriched.csv

FILE_LIST: must be a text file with one line per genotype line, each line should have 2 columns (separated table) *i.e.* column 1 for the genotype line name and column 2 for the path of the corresponding .vcf file.

Example: microtom /path/to/microtom.vcf

21| Split the .CSV files by chromosome to facilitate file manipulation using the script *split_csv_by_chr.py* (optional).

- python split_csv_by_chr.py common_file_enriched.csv output/dir

22| EMS candidate mutation identification. For recessive mutation, only variant file needs to be studied (allelic frequency tends to 1 for mutant-like bulk and tends to 0.33 for WT-like bulk). For dominant mutation, see **Box 3**.

- filter each .csv file on Depth_Mut column with "Depth_Mut is superior to 10 and Depth_Mut is inferior to 100" to remove false positives SNPs due to mapping error.
- filter "Freq_allele_alt_Mut" and "Freq_allele_alt_WT" columns with the expected frequencies ie 1.00 for the Freq_allele_alt_Mut and 0.33 for the Freq_allele_alt_WT, in case of a recessive mutation. For a dominant mutation, see **Box 3.**

?TROUBLESHOOTING

BOX 2 |

BOX 3 |

? TROUBLESHOOTING

Troubleshooting advice for several steps can be found in the following section.

Step	Problem	Possible reason	Possible solution
Step 1	**Unsuccessful pollination/fertilization during BC_1F_1 production**	Key flower stage and environmental conditions are not adequate	Successful conditions need to be adjusted
		EMS mutations alter reproductive organs (pollen quality from the mutant needs to be checked)	If the mutant presents a male sterility, step 1 can be performed using the mutant as female.
Step 3:	**No clear phenotype segregation in the BC_1F_2 population**	Absence of the mutant phenotype due to unsuccessful pollination with the mutant pollen at step 1 and selfing of the Micro-Tom line	Repeat step 1
		Deviation of the phenotypic segregation ratio from the expected Mendelian ratio 1:3 due to the involvement of various EMS mutations in the phenotype of interest	If segregation ratios correspond to the involvement of 2 unlinked recessive loci (6% of mutant-like phenotypes expected) mapping-by-sequencing procedure can be developed using a larger population in step 3 (1000 BC_1F_2) and bulking 30 individuals per bulks. Size of the populations would proportionally increase with number of loci involved, which can be a limit depending on greenhouse or field space available
		Absence or deviation of the phenotypic segregation ratio due to (i) incomplete or reduced penetrance or (ii) variation in expressivity resulting in a range of	Both the penetrance and expressivity are in some cases explained by the action of gene modifiers and/or due to environmental interactions. For these particular cases, accurate phenotyping of the BC_1F_2 population would be

		phenotypes	limited and appears incompatible with this mapping-by-sequencing approach
Step 15	**Accurate identification of indels**	False positive indels are suspected to be caused by mapping error, in regions of low read coverage as well as in highly repetitive regions	Analysis of .vcf output files for indels variant type all along the genome would represent a time consuming task to discriminate false positives from true variants. Once a reasonable confidence interval region encompassing the causal mutation has been identified using SNPs variant type, then analysis and confirmation of the indels included in this associated region is conceivable using IGV
Step 19	**No mutation detected**	Filtering out allelic frequencies is too stringent	Because of possible errors in the constitution of the bulks and in the phenotyping process, less stringent thresholds for the analysis of sequencing data should be tested as well as assaying different combinations of frequency filters for both bulks (see example in **Table 3**).
		The causal mutation is located in a region not represented in the genome of reference Heinz 1706.	Map sequence reads against other available tomato genomes ([1,3]).
	Many candidate mutations detected	Many mutations may present the same pattern of allelic frequency either due to close vicinity between the adjacent EMS mutations or to the high LD in the chromosomal region carrying the causal mutation *e.g.* regions next to the centromere ([1]).	Recombinant analysis is required to fine map the causal mutation (see **Box 2**). For causal mutations close to but not within the low recombinant regions, the recombination could be sufficient to break the LD and identify the causal mutation. However for causal mutation located in non- or low recombinant regions, recombinant analysis would require large segregant population. This therefore limits the benefit of this approach. We propose in this case to use SnpEff tool ([87]http://snpeff.sourceforge.net/SnpEff.html), which is compatible with the tomato genome, to predict the effects of variants on expressed genes and focus only on the mutations that possibly alter protein functions.
BOX 3	**No causal mutation identified after recombinant analysis of the associated region**	The causal mutation does not probably correspond to a SNPs variant type	In this case proceed to indels analysis (step 18).

ANTICIPATED RESULTS

We have used the protocol with a mutant line called *yellow* previously identified in a Micro-Tom EMS mutant collection generated by two rounds of mutagenesis with 1% EMS as described in **Fig. 1**. Several mutations affecting various aspects of fruit development (fruit growth, fruit ripening, cuticle composition, vitamin C content) have been readily identified in this manner in our group. Many other causal mutations underlying a large range of phenotypic variations present in publicly available tomato EMS mutant collections ([7-11]) are prone to be identified using the same procedure.

Our mapping by sequencing approach was used prior any information on the nature of the mutation. The *yellow* fruit mutant was first checked for recessive genetic inheritance of the mutation that confers the *yellow* fruit phenotype, then crossed to a wild-type Micro-Tom line to produce BC_1F_1 plant. A segregating population of 600 BC_1F_2 plants was then screened, from which a bulk of 150 plants displaying the yellow (mutant-like) fruit phenotype and a bulk of 150 plants displaying the red (WT-like) fruit phenotype were sequenced (Step 1 to 9) to a read depth of ~43X (mutant-like bulk) and ~26X (WT-like bulk) (**Table 1**). Sequences were analysed following the bioinformatics procedure (step 14 to 22; **Fig. 2a**). Since the *de novo* assembly of the Micro-Tom genome has not been performed yet, the reference genome used was Heinz 1706 (*Solanum lycopersicum* genome v2.5) and variability inherent to Micro-Tom cultivar was accounted for by considering the Heinz 1706/Micro-Tom natural polymorphisms. A total of 4130 DNA polymorphisms due to mutagenesis were detected (**Table 2**), a value very close to those obtained for other mutants from the Micro-Tom EMS mutant collection generated according to **Box 1**. Distribution of the allelic frequencies over the whole genome was not even, as anticipated. In chromosome 3, allelic frequency was much higher than average in the mutant-like bulk (**Fig. 2b**). In particular, 4 mutations showed frequencies>0.98 in the mutant-like bulk and <0.40 in the WT-like bulk (**Table 3**). Two of these mutations affected genes predicted to encode proteins and were located at ~80 kb one from each other (**Table 4**). Genotyping a subset of the recombinant BC_1F_2 population (150 plants with mutant-like phenotype) easily allowed us to unequivocally attribute the mutant yellow phenotype to a mutation in the phytoene synthase *PSY1* gene (**Table 4; Fig. 2c and 3a**). Carotenoid profiling of the WT (Micro-Tom line) and *yellow* fruit mutant confirmed that the mutation in the *PSY1* gene was responsible for the block of carotenoid accumulation (**Fig. 3b and Supplementary information 5**), consistent with previous data obtained in the *yellow flesh* mutant lines carrying knockout alleles of *PSY1* ([62]Gady et al. 2012). Other approaches of reverse genetics described in **Box 2** can be used to confirm the link between the mutation and the phenotype-of-interest ([22]). This includes the identification of allelic series by TILLING using the same mutant collection ([46,13]). Several large modifications in primary metabolites independent of the ripening stage of the *yellow* mutant fruit were evidenced using the method described by [75], in the heatmap of fruit primary metabolite composition in mutants displaying a WT-like (red) or a mutant-like (yellow) phenotype (**Fig. 3c**). Thus, our

mapping-by-sequencing analysis of a yellow-coloured tomato fruit mutant enabled the identification of a point mutation in the carotenoid pathway gene phytoene synthase *PSY1.* This mutation leads to alterations of both carotenoid and primary metabolite accumulations in the fruit, therefore providing tools to further explore the interplay between secondary and primary metabolism in the tomato fruit.

Table 1

	Mutant-like bulk	**WT-like bulk**
Scored phenotype	yellow ripe fruit	red ripe fruit
Tomato reference genome	Heinz 1706 (v2.50)	Heinz 1706 (v2.50)
Nb of BC_1F_2 plants per bulk	150	150
Nb of paired reads	403 552 936	244 334 344
% mapped reads	94.52%	95.11%
Insert size	378	382
Sequence length	101	101
Read depth	43X	26X
Nb of SNPs	2 273 267	2 232 386

Table 2

Chromosome	**Total variants**[a]	**EMS mutations**[b]
1	75 143	572
2	189 025	352
3	107 225	303
4	101 282	416
5	499 401	371
6	39 699	262
7	46 148	318
8	24 090	333
9	36 718	315
10	18 669	305
11	209 195	253
12	58 814	333
all	1 405 409	4133

a: total variants (natural polymorphisms and EMS mutations) obtained in the mutant and WT-like bulks (10<read depth<100)
b: EMS mutations obtained after removing natural polymorphisms between Micro-Tom and Heinz 1706 (10<read depth<100)

Table 3

Mutant-like bulk	**WT-like bulk**	**No of EMS mutations**

AF>0.9	0.2<AF<0.4	30
AF>0.95	0.2<AF<0.4	11
AF>0.98	0.2<AF<0.4	4

AF is the allelic frequency

Table 4

EMS Position	Mutation	Read depth	Gene position	Mutation type	Tomato gene id	Gene annotation
4327086	G to A	26	exonic	nonsense (W to stop)	Solyc03g031860	PSY1
4408241	C to T	53	exonic	Missense (E to K)	Solyc03g031940	Acyl-CoA synthetase /AMP-acid ligase II
39268042	C to T	29	intergenic	silent	-	-
39467288	G to A	32	intergenic	silent	-	-

ACKNOWLEDGMENTS

This work was supported by CEA-IG/CNG for conducting QC of DNA and Illumina sequencing. Thanks to A Boland, MT Bihoreau and their staff. We are grateful to the Genotoul bioinformatics platform Toulouse Midi-Pyrenees and Sigenae group for providing help and/or computing and/or storage resources thanks to Galaxy instance (http://sigenae-workbench.toulouse.inra.fr). Thanks to S Maman. This project was funded by INRA AIP Bioressources and ERANET “TomQML”. FW was supported by a grant from ANR Bioadapt “Adaptom”.

AUTHOR CONTRIBUTIONS

CB, DJ, LF, VG and CR developed the original protocol. DB, JPM and AB performed the sequencing experiments. FW performed computational analyses. CB, LF, DJ, FW, MCLP, KA, SA, ARF, PDF and CR contributed sections to the manuscript. CB, LF, DJ and CR collated and standardized the text. All authors read and approved the final version.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

BOX 1 | MUTAGENESIS AND PHENOTYPING ● TIMING ~6-12 months

Forward genetics screens aims at identifying the causal mutation in a gene underlying a remarkable phenotypic alteration in a collection of mutants. EMS mutagenesis can produce a very large range of visible phenotypic alterations at whole plant level ([7,11]), the number of which will depend on the intensity of the mutagenesis. EMS mutants carrying the trait of interest can be obtained by mining various tomato mutant phenotypic databases publicly available for various tomato cultivars. Alternatively, a new mutant collection can be created. Generation of new mutant collection can be useful, especially if ones want to screen for traits not described in available public databases (e.g. the resistance to pathogens or the root development) or wants to use a specific tomato cultivar or a wild tomato species.

Glossary

Micro-Tom: Micro-Tom is a miniature tomato cultivar ([76]) highly suitable for large-scale experiments because it can be grown at high density in greenhouse or growth chambers (*e.g.* 150 plants/m^2 for fruit phenotyping and mapping-by-sequencing). It carries three recessive mutations affecting brassinosteroid synthesis in vegetative tissues (*dwarf*), determinate growth of the plant (*sp*) and unknown affecting internode length ([77]).

EMS: ethyl methanesulfonate is a chemical mutagenic agent that produces random mutations in DNA by base-base substitution. Most mutations are point mutations, particularly transition mutations due to guanine alkylation (G/C to A/T). In tomato transversion mutations (~35%) and small insertion/deletions (indels 1.3%) can also be found ([57]).

Mutagenesis: Untreated seeds are called M_0 seeds. Treated seeds that carry heterozygous mutations are called M_1 seeds. Following self-pollination, the seeds on the M_1 plants, which segregate for the mutations, are called M_2 seeds.

TILLING: Targeting Induced Local Lesions IN Genomes is a reverse genetics approach aiming at identifying allelic series for target genes by screening EMS mutant collection. Either one single M_2 plant or a whole family (pool of 12 M_2 plants) are usually selected for DNA extraction. Mutation detection can be done using a range of technologies including NGS sequencing. This approach can be very useful for the functional validation of causal mutations in candidate genes previously identified through mapping-by-sequencing (**Box 2**).

Phenotyping: a range of plant, root, leaf, or fruit phenotypic descriptors can be used for phenotyping the mutant collection. Phenotypic descriptions can then be stored in searchable databases. Phenotypic databases for publicly available tomato mutants are TGRC (http://tgrc.ucdavis.edu/); SGN (http://solgenomics.net/); TOMATOMA (http://tomatoma.nbrp.jp/); Genes That Make Tomato (http://zamir.sgn.cornell.edu/mutants/).

Main steps (see Fig. 1)

- Soak the M_0 tomato seeds in water for 4 h, incubate them overnight (15 h) with 1% EMS (Ethyl MethaneSulfonate SIGMA cat. no. M0880) under gentle stirring and rinse them 3 times for 3 h with water. **<CRITICAL STEP>** EMS is mutagen and carcinogenic for mammals. The use of gloves and hood is mandatory. All material in contact with EMS, including the water used for seed rinsing, must be treated with 1M NaOH (Euromedex cat. no. 2020A) solution during 15 h. **! CAUTION** EMS and NaOH are toxic and should be handled under a fume hood.
- Sow the M_1 seeds in moist vermiculite. After germination, transplant the M_1 plantlets to a substrate (compost plus perlite) suitable for sub-irrigation and grow the plants till they carry red ripe fruits. For Micro-Tom, conditions and management of plant culture are described in ([78]). **<CRITICAL STEP>** Efficiency of EMS mutagenesis can be assessed in the M_1 plants by the frequency of plants carrying white sectors. If less than 0.5% of the M_1 plants display partial albinos or variegated phenotype, mutagenesis was likely poorly efficient. **<CRITICAL STEP>** Because EMS mutagenesis efficiency will depend on the physiological state of the seeds used, pilot experiments can be done previously by treating seeds with various concentrations of EMS (0.4 to 1.2%).
- Collect the M_2 seeds from M_1 plants. To increase mutation frequency, steps [1 – 2] can be repeated using M_2 mutagenized seeds for EMS mutagenesis. This will typically increase the mutation frequency (up to 6000 mutations per plant in Micro-Tom) but will strongly increase the number of non-germinating seeds (up to 50%) and the number of sterile M_1 plants (up to 60%).
- Sow M_2 seeds for phenotyping M_2 plants (12 plants / family) or store them until use. **<CRITICAL STEP>** To avoid contamination by pathogens, seeds are treated with sodium hypochlorite (2.5% of active chloride during 15 min), rinsed with

water (30 min) and air-dried under a hood for one day. Seeds can be stored at 4°C in sealed boxes with desiccant (Silica Gel) for up to 10 years.

- If M_2 mutant plants are also to be used for TILLING, collect leaf materials from 12 four-weeks old plants (two 5 mm diameter discs per plant) and pool them for DNA extraction. Store at -80°C until use.
- Analysis of the phenotypic alterations of M_2 plants is highly dependent on the focus of the mutant collection and the tomato cultivar. For a thorough description of the visual phenotype of the plants, phenotyping must be done at least at seven developmental stages: germination, seedling, plantlet, flowering plant, immature green fruit, ripening fruit, harvest.
- Store M_3 seeds until use.

End of Box 1

BOX 2 | Functional validation TIMING ~6 to 12 months

Once a causal mutation is identified, involvement of the locus/gene underlying the mutant phenotype needs to be confirmed. During the last few years, there have been impressive advances in tomato functional genomics and many tools have been developed or are under development ([79-85,45,46,48]). Two approaches of reverse genetics can be developed: (i) gene complementation in order to restore the WT phenotype in the mutant background or (ii) gene knockdown or knockout to copy the mutant phenotype in a WT background. Different strategies for functional genes validation in tomato are listed below:

Methods	Description	Expected results	Advantages	Limits
TILLING[46]	To identify allelic variants for a candidate gene in mutagenized populations	Gene knockout or knockdown to copy the mutant phenotype	Large scale screening New mutation/allele identification	Wide variability in gene knockdown efficiency for the different allelic variants Needs to have a large population of mutagenized plant available
Gene-overexpression[79]	Transgenic plants that overexpress a target gene	Complement the mutant phenotype	Efficient tomato transformation	Availability of specific promoters to direct transgene expression in a spatial/time manner
Gene suppression[79]	Transgenic plants that silence a target gene by RNAi or amiRNA strategies	Gene knockdown to copy the mutant phenotype	Efficient tomato transformation Specificity of gene silencing	Gene expression not totally abolished and limited availability of specific promoters to direct transgene expression in a spatial/time manner
VIGS (Virus-Induced Gene Silencing)[83-85]	Transient system to silence target gene expression through a plant RNAi-mediated antiviral defense mechanism	Transient knockdown to copy the mutant phenotype	Rapid and effective	Critical to choose the right developmental stage Mosaic phenotype because of non-homogeneous viral infection
CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats)[86]	RNA-Guided Endonuclease technology for targeted gene mutagenesis	Gene knockout to copy the mutant phenotype or complement the mutant phenotype	Easy design and high efficiency Possibility to remove the T-DNA Possible introduction of specific nucleotide modifications at the target gene	Possible off-targets

End of Box 2

BOX 3 | Mapping-by-sequencing SETUP for analysis of dominant mutation and for other tomato genotypes

Dominant and incomplete dominant mutations

For EMS mutations, most of dominant phenotypes will result from loss of function mutations in haploinsufficient genes. In this case, the mutant allele is dominant over the WT allele as haploinsufficient genes require both alleles to be functional and therefore express the WT phenotype. For dominant mutation, the heterozygote displays the mutant phenotype. However it is also possible that the heterozygote display an intermediate, less severe phenotype in the case of incomplete dominant mutations. The following steps of the procedure need to be used in case of dominant or incomplete dominant mutations.

Procedure step2:

For case 2 and 3 select one BC_1F_1 individuals displaying the mutant-like phenotype to collect BC_1F_2 seeds.

Procedure step3:

For a dominant mutation. The mutant-like bulk displaying the mutant phenotype is composed in a 1:2 ratio of individuals that are homozygous for the mutant allele and of individuals that are heterozygous thus resulting in an allelic frequency (AF) of the mutant allele of 0.77. The WT-like bulk is expected to be free of the mutant allele. In view to facilitate the analysis of the sequencing data, 10% individuals presenting the mutant-like phenotype are introduced in the WT like bulk. For this purpose, combine 50 individuals exhibiting the WT like phenotype with 6 individuals exhibiting the mutant–like phenotype in order to constitute the pool termed WT-like bulk. In this case, allelic frequency of the mutant allele in the WT-like bulk should be comprised between $0.05<AF<0.1$.

For an incomplete dominant mutation. To constitute the mutant-like bulk select the individuals displaying the most severe phenotype that would be mainly represented by individuals homozygous for the mutant allele. Allelic frequency should be close to 1 in the mutant-like bulk. As for dominant mutation, combine 50 individuals exhibiting the WT-like phenotype with 10 individuals exhibiting the mutant-like phenotype to constitute the WT-like bulk. In this case, frequency of the mutant allele in the WT-like bulk should be comprised between $0.05<AF<0.1$.

Procedure step 22:

For an incomplete dominant mutation. Filter “Freq_allele_alt_Mut” and “Freq_allele_alt_WT” columns with the expected frequencies i.e. 0.7<Freq_allele_alt_Mut<0.8 and 5<Freq_allele_alt_WT<10.

For a dominant mutation. Filter “Freq_allele_alt_Mut” and “Freq_allele_alt_WT” columns with the expected frequencies i.e. 0.80<Freq_allele_alt_Mut<1 and 5<Freq_allele_alt_WT<10.

Other tomato genotypes

EMS tomato mutant collections have been produced in other tomato cultivars than Micro-Tom, such as M82. The mapping-by-sequencing procedure described can also be used to identify key mutations in these mutant collections. All steps are similar to the ones described for Micro-Tom. If the tomato line is already sequenced, sequencing of the mutant-like and WT-like bulks are sufficient for the identification of natural polymorphisms between the line considered and Heinz 1706. In this case, begin at step 4 for the non-mutagenized line. **?TROUBLESHOOTING**

End of Box 3

FIGURE LEGENDS

Fig 1 | An overview of the experimental design of forward genetics screening and detection of causal mutation by mapping-by-sequencing in tomato. Most such experiments start with the generation of a highly mutagenized EMS mutant collection. Alternatively, publicly available mutant collections can be screened *in silico* for mutants displaying the phenotype-of-interest. **(a) EMS mutagenesis and phenotyping.** Untreated seeds (M_0) are treated with EMS to give M_1 (EMS-treated) seeds. M_1 plants are grown, producing M_2 seeds which are either stored, treated again with EMS to increase mutation frequencies (2^{nd} round of mutagenesis) or sown to give M_2 plants. At this step, DNA can be collected for TILLING. M_2 plants are further screened using various phenotypic descriptors and the phenome data are stored in a database to allow *in silico* mining of the mutant collection for phenotype-of-interest. At each generation, seeds are collected and stored. **(b) Detection of the causal mutation by mapping-by-sequencing.** The experimental design is shown for a recessive mutation, the most commonly found in EMS mutants. Once a homozygous mutant carrying a recessive mutation responsible for the phenotype-of-interest (*e.g. yellow* for yellow-colored fruit) has been selected, the mutant is backcrossed (BC_1) with the WT genotype used for generating the EMS mutant collection. The BC_1F_1 plant displays a WT-like phenotype (red fruit) because the *yellow* mutation is recessive. A BC_1F_2 population (usually 500 plants) obtained by selfing the BC_1F_1 plants is screened for mutant-like phenotype and WT-like phenotype. Two bulks of pooled plants are set-up: one displaying the mutant-like phenotype (*yellow* fruit) and one displaying the WT-like phenotype (red fruit). Each bulk is sequenced to a depth of 20 to 40X the tomato genome, trimmed sequences are mapped onto the tomato reference genome and EMS mutation variants are filtered. Analysis of the allelic variant frequencies in the two bulks (usually ~60 plants) leads to the identification of the causal mutation which displays very high frequency in the mutant-like bulk (ideally 100% of variant allele) and lower than average frequency in the WT-like bulk (ideally 33% of variant allele).

Fig 2 | Mapping-by-sequencing of Micro-Tom EMS mutants. (**a**) **Two-step bioinformatic pipeline for analysis of whole genome sequencing data.** First, raw reads are mapped to the SL2.50 Heinz 1706 tomato genome sequence using BWA v0.7.12 aligner. Variant calling is performed using SAMtools v1.2 and output VCF

files of total variants are generated for the mutant and WT-like bulks as well as for the Micro-Tom line. As the reference genome from Heinz 1706 used to map the reads is different from that of the Micro-Tom line, the variants identified include Heinz 1706 /Micro-Tom natural polymorphisms and EMS mutations. At this step, an additional Micro-Tom sequencing is required as a control to further remove natural polymorphisms. In-house python script for the analysis of variants includes the (i) comparison of variants obtained in each bulk to generate SNPs files, (ii) computation of variant allelic frequencies for each bulk, (iii) indication about the presence of the variants in the Micro-Tom line, and (iv) indication of variants located in tomato genes according to the Heinz 1706 genome annotation. Natural polymorphisms between Micro-Tom and Heinz can be excluded by analyzing the variants present in the Micro-Tom line (control). Only EMS mutations are further considered to specifically identify the causal mutation using two filtering parameters: read depth (10<DP<100) to exclude false positive variants, and the allelic frequency expected for a recessive mutation (allelic frequency tends to 1 for mutant-like bulk and tends to 0.33 for WT-like bulk). All the workflow is available in supplementary information 1-4. (**b**) **Identification of the chromosome associated with the yellow-fruit phenotype.** Pattern of the mutation allelic frequencies obtained in the mutant and WT-like bulks are represented along tomato chromosomes by yellow and red lines, respectively. Chromosome 3 exhibits a dramatic increase in the mutation allelic frequency for the mutant-like bulk compared to the WT-like bulk, and therefore likely corresponds to the chromosome carrying the causal mutation. For an optimal representation, sliding windows of 20 positions are used that average the allelic frequency for 20 successive EMS mutations **(c) Fine mapping of the causal mutation using BC_1F_2 population.** Recombinant analysis of BC_1F_2 individuals displaying the yellow fruit phenotype allowed us to define more precisely the chromosomal region associated with the mutant phenotype. Recombination events between linked mutations are used to discriminate the causal mutation from the adjacent ones. Only mutant alleles are represented for the mutation at position 4 327 086 (*yellow* mutant lines) while WT alleles (red lines) are identified for the adjacent mutations. Recombinant scoring of 150 BC_1F_2 individual plants allows the accurate identification of the causal mutation at position 4 327 086 on chromosome 3.

Fig 3 | Mutation in the phytoene synthase gene *PSY1* affects fruit metabolism. **(a)** *PSY1* gene (above) consists in 6 exons. Trans-isoprenyl diphosphate synthase domain is indicated in grey. Mutation in *PSY1* in the coding region (below) results in a premature stop codon at position 152. **(b)** Scheme showing block tomato fruit carotenoid pathway resulting from *yellow* mutation. Red keys indicate the degree of reduction in amounts, green represents accumulation in red ripe (Breaker+7) fruit. **(c)** Metabolite profiling by GC-MS further revealed differences in primary metabolite content in the WT (Micro-Tom line) and *psy1* mutant lines. Heat map showing the fruit metabolic profiling of *yellow* mutant (*psy1*) and WT, harvested during fruit ripening at Breaker+5 stage (yellow Br+5) and at red ripe Breaker+7 stage (WT-Br+7, WT-Br+7, yellow-Br+7), respectively. Red and blue rectangles depict increases and decreases with respect to average of all lines. Hierarchical clustering of samples and metabolites is shown in the dendrogram.

Table 1 | Illumina sequencing of BC_1F_2 bulked individuals displaying a mutant yellow fruit color or a WT red fruit color. Sequence of the reference genome is the tomato whole genome shotgun chromosomes from version 2.50 of the Wageningen University and Research center (WUR) assembly, available on the SGN website (http://solgenomics.net).

Table 2 | Number of SNPs in the mutant and the WT-like bulks for the yellow fruit mutant. Only variants with a read depth between 10<DP<100 were considered to remove false positive variants due to erratic read mapping. Number of total variants (natural polymorphisms between Heinz 1706 and Micro-Tom and EMS mutations) and only EMS mutations per chromosome are reported in the second and third column, respectively.

Table 3 | Identification of the putative causal mutations associated with the yellow fruit phenotype based on allelic frequency analysis in the mutant and WT-like bulks. In case of a recessive mutation (ie most of EMS mutation), all BC_1F_2 individuals that exhibit the mutant phenotype are homozygous for the causal mutation (frequency =1 in the mutant-like bulk). On the contrary, in the WT-like bulk, the EMS mutation segregates as a mendelian locus (frequency ~0.33 in the WT-like bulk) The table reports the number of putative causal mutations corresponding to

three different allelic frequency filters applied to detect the recessive causal mutation. The number of candidate causal mutations decrease with stringency of allelic frequency cut-offs. Four mutations on chromosome 3 appear as likely causal mutation candidates. AF= Allelic Frequency.

Table 4 |Annotation of the high scoring causal mutations identified on chromosome 3 for the yellow fruit mutant. Two silent EMS mutations (mutations at positions 39268042 and 39467288 on chromosome 3) are located in intergenic regions and are supposed to have no effect on protein structure. Two other ones located in coding regions cause non-synonymous changes. The mutation 4408241 corresponds to a missense mutation and the mutation 4327086 corresponds to a nonsense mutation involving a premature codon STOP that probably unable the protein function. In addition, this mutation is located in the *PSY1* gene, coding for an enzyme of the carotenoid biosynthesis pathway. The mutation 4408241 is likely the causal mutation at the origin of the yellow-fruit phenotype.

References

1| Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635-641 (2012).

2| Aflitos, S. *et al.* Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing. *Plant J*. **80**, 136-148 (2014).

3| Bolger, A. *et al*. The genome of the stress-tolerant wild tomato species Solanum pennellii. *Nat. Genet.* **46**, 1034-1038 (2014).

4| Lin, T. *et al*. Genomic analyses provide insights into the history of tomato breeding Nat. Genet. **46**, 1220-1226 (2014).

5| Kobayashi, M. *et al.* Genome-wide analysis of intraspecific DNA polymorphism in 'Micro-Tom', a model cultivar of tomato (Solanum lycopersicum). *Plant Cell. Physiol.* **55**, 445-454 (2014).

6| Eshed, Y. & Zamir, D. An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* **141**, 1147-1162 (1995).

7| Menda, N., Semel, Y., Peled, D., Eshed, Y. & Zamir, D. In silico screening of a saturated mutation library of tomato. *Plant J.* **38**, 861-872 (2004).

8| Meissner, R. *et al.* A new model system for Tomato Genetics. *Plant J.* **12**, 1465-1472 (1997).

9| Meissner, R., Chague, V. Zhu, Q. Emmanuel, E. Elkind, Y. & Levy, AA. (2000) A high throughput system for transposon tagging and promoter trapping in tomato. *Plant J.* **38**, 861-872 (2000)

10| Emmanuel, E. & Levy, AA. Tomato mutants as tools for functional genomics. *Curr. Opin. Plant. Biol.* **5**, 112-117 (2002).

11| Saito, T. *et al*. TOMATOMA: a novel tomato mutant database distributing Micro-Tom mutant collections. *Plant Cell Physiol.* **52**, 283-296 (2011)

12| Carvalho, R.F. *et al.* Convergence of developmental mutants into a single tomato model system: 'Micro-Tom' as an effective toolkit for plant development research. Plant Methods. **7**, 18 (2011)

13| Monteiro, CC. *et al.* Biochemical and histological characterization of tomato mutants. *An. Acad. Bras. Cienc.* **84**, 573-585 (2012).

14| Just, D. *et al.* Micro-Tom mutants for functional analysis of target genes and discovery of new alleles in tomato. *Plant Biotechnol.* **30**, 225-231 (2013).

15| Isaacson, T. *et al.* Cutin deficiency in the tomato fruit cuticle consistently affects resistance to microbial infection and biomechanical properties, but not transpirational water loss. *Plant J.* **60**, 363-377 (2009).

16| Nadakuduti, S.S. *et al.* Pleiotropic phenotypes of the sticky peel mutant provide new insight into the role of CUTIN DEFICIENT2 in epidermal cell function in tomato. *Plant Physiol.* **159**, 945-960 (2012).

17| Yeats, T.H. *et al.* The identification of cutin synthase: formation of the plant polyester cutin. *Nat. Chem. Biol.* **8**, 609-611 (2012).

18| Kimbara, J. *et al.* Inhibition of CUTIN DEFICIENT 2 Causes Defects in Cuticle Function and Structure and Metabolite Changes in Tomato Fruit. *Plant Cell. Physiol*. **54**, 1535-1548 (2013)

19| Shi, J.X. *et al.* The tomato SlSHINE3 transcription factor regulates fruit cuticle formation and epidermal patterning. *New Phytol.* **197**, 468-480 (2013).

20| Petit, J. *et al.* Analyses of tomato fruit brightness mutants uncover strong cutin-deficient mutants among which a new allele of GDSL lipase. *Plant Physiol.* **164**, 888-906 (2014)

21| Ronen, G., Carmel-Goren, L., Zamir, D. & Hirschberg, J. An alternative pathway to beta -carotene formation in plant chromoplasts discovered by map-based cloning of beta and old-gold color mutations in tomato. *Proc. Natl. Acad. Sci. USA* **97**, 11102-11107 (2000).

22| Isaacson, T., Ronen, G., Zamir, D., & Hirschberg, J. Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of beta-carotene and xanthophylls in plants. *Plant Cell* **14**, 333-342 (2002).

23| Fraser, P.D. *et al.* Manipulation of phytoene levels in tomato fruit: effects on isoprenoids, plastids, and intermediary metabolism. *Plant Cell* **19**, 3194-3211 (2007).

24| Galpaz, N., Wang, Q., Menda, N., Zamir, D. & Hirschberg, J. Abscisic acid deficiency in the tomato mutant high-pigment 3 leading to increased plastid number and higher fruit lycopene content. *Plant J.* **53**, 717-730 (2008).

25| Ariizumi, T. *et al.* Identification of the carotenoid modifying gene PALE YELLOW PETAL 1 as an essential factor in xanthophyll esterification and yellow flower pigmentation in tomato (*Solanum lycopersicum*). *Plant J.* **79**, 453-465 (2014).

26| Neuman, H., Galpaz, N., Cunningham, F.X., Zamir, D. & Hirschberg, J. The tomato mutation nxd1 reveals a gene necessary for neoxanthin biosynthesis and demonstrates that violaxanthin is a sufficient precursor for abscisic acid biosynthesis. *Plant J*. **78**, 80-93 (2014).

27| Fridman, E., Carrari, F., Liu, Y.S., Fernie, A.R. & Zamir, D. Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* **305**, 1786-1789 (2004).

28| Schauer, N. *et al.* Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* **24**, 447-454 (2006).

29| Mounet, F. *et al.* Gene and metabolite regulatory network analysis of early developing fruit tissues highlights new candidate genes for the control of tomato fruit composition and development. *Plant Physiol.* **149**, 1505-1528 (2009).

30| Itkin, M. *et al.* Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science* **341**, 175-179 (2013).

31| Sauvage, C. *et al.* Genome-Wide Association in Tomato Reveals 44 Candidate Loci for Fruit Metabolic Traits. *Plant Physiol*. **165**, 1120-1132 (2014).

32| Colombié, S. *et al.* Modelling central metabolic fluxes by constraint-based optimization reveals metabolic reprogramming of developing *Solanum lycopersicum* (tomato) fruit. *Plant J.* **81**, 24-39 (2015).

33| Schmitz, G. *et al.* The tomato Blind gene encodes a MYB transcription factor that controls the formation of lateral meristems. *Proc. Natl. Acad. Sci. USA* **99**, 1064–1069 (2002).

34| Krieger, U., Lippman, Z.B. & Zamir, D. The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nat. Genet.* **42**, 459-463 (2010).

35| Busch, B.L. *et al.* Shoot branching and leaf dissection in tomato are regulated by homologous gene modules. *Plant Cell* **23**, 3595-3609 (2011).

36| Martín-Trillo, M. *et al.* Role of tomato BRANCHED1-like genes in the control of shoot branching. *Plant J.* **67**, 701-714 (2011).

37| MacAlister, C.A. *et al.* Synchronization of the flowering transition by the tomato TERMINATING FLOWER gene. *Nat. Genet.* **44**, 1393-1398 (2012).

38| Park, S.J. *et al.* Optimization of crop productivity in tomato using induced mutations in the florigen pathway. *Nat. Genet.* **46**, 1337-1342 (2014).

39| Burko, Y. *et al.* A role for APETALA1/fruitfull transcription factors in tomato leaf development. *Plant Cell* **25**, 2070-2083 (2013).

40| Ichihashi, Y. *et al.* Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. *Proc. Natl. Acad. Sci. USA* **111**, 2616-2621 (2014).

41| Ori, N. *et al.* Regulation of LANCEOLATE by miR319 is required for compound-leaf development in tomato. *Nat. Genet.* **39**, 787-791 (2007).

42| Li, F. *et al.* MicroRNA regulation of plant innate immune receptors. *Proc. Natl. Acad. Sci. USA* **109**, 1790-1795 (2012).

43| Semel, Y. *et al.* Overdominant quantitative trait loci for yield and fitness in tomato. *Proc. Natl. Acad. Sci. USA* **103**, 12981-12986 (2006).

44| Kharkwal, M.C. & Shu, Q.Y. The role of induced mutations in world food security. In induced plant mutations in the genomics era. QY Shu (ed) Food and Agricultural Organization of the United nations, Rome, 33-38 (2009).

45| Henikoff, S. & Comai, L. Single-nucleotide mutations for plant functional genomics. *Annu. Rev. Plant Biol.* **54**, 375-401 (2003).

46| Minoia, S. *et al.* A new mutant genetic resource for tomato crop improvement by TILLING technology. *BMC Res. Notes* **3**, 69 (2010).

47| Okabe, Y. *et al.* Tomato TILLING technology: development of a reverse genetics tool for the efficient isolation of mutants from Micro-Tom mutant libraries. *Plant Cell Physiol.* **52**, 1994-2005 (2011).

48| Sreelakshmi, Y. *et al.* NEATTILL: A simplified procedure for nucleic acid extraction from arrayed tissue for TILLING and other high-throughput reverse genetic applications. *Plant Methods* **6**, 3 (2010).

49| Baldet, P. *et al.* TILLING identification of ascorbate biosynthesis tomato mutants for investigating vitamin C in tomato. *Plant Biotechnol*. **30**, 309-314 (2013).

50| Yifhar, T. *et al.* Failure of the tomato trans-acting short interfering RNA program to regulate AUXIN RESPONSE FACTOR3 and ARF4 underlies the wiry leaf syndrome. *Plant Cell* **24**, 3575-3589 (2012).

51| Schneeberger, K. Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nat. Rev. Genet.* **15**, 662-676 (2014).

52| Schneeberger, K. *et al.* SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat. Methods* **6**, 550-551 (2009).

53| Austin, R.S. *et al*. Next-generation mapping of Arabidopsis genes. *Plant J.* **67**, 715-725 (2011)

54| Abe, A. *et al.* Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* **30**, 174-178 (2012)

55| Fekih, R., *et al*. MutMap+: genetic mapping and mutant identification without crossing in rice. *PLoS One* **8,** e68529 (2013)

56| Takagi, H. *et al*. MutMap-Gap: whole-genome resequencing of mutant F2 progeny bulk combined with de novo assembly of gap regions identifies the rice blast resistance gene Pii. *New Phytol.* **200**, 276-283 (2013)

57| Takagi, H. *et al.* MutMap accelerates breeding of a salt-tolerant rice cultivar. *Nat. Biotechnol.* 33, 445-449 (2015)

58| Shirasawa, K., Hirakawa, H., Nunome, T., Tabata, S. & Isobe, S. Genome-wide survey of artificial mutations induced by ethyl methanesulfonate and gamma rays in tomato. *Plant Biotechnol. J.* [Epub ahead of print] (2015)

59| Shirasawa, K. *et al.* SNP Discovery and Linkage Map Construction in Cultivated Tomato. *DNA Res* **17**, 381-391 (2010).

60| Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform *Bioinformatics* **25**, 1754-1760 (2009).

61| Li, H., *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

62| Piron, F. *et al.* An induced mutation in tomato eIF4E leads to immunity to two potyviruses. *PLoS One* **5**, e11313 (2010)

63| Gady, A.L. *et al.* Induced point mutations in the phytoene synthase 1 gene cause differences in carotenoid content during tomato fruit ripening. *Mol Breed* **29**, 801-812 (2012)

64| Hirakawa, H. *et al.* Genome-Wide SNP Genotyping to Infer the Effects on Gene Functions in Tomato. *DNA Research* **20**, 221-233 (2013).

65| Martí, E., Gisbert, C., Bishop, G.J., Dixon, M.S., García-Martínez, J.L. Genetic and physiological characterization of tomato cv. Micro-Tom. *J. Exp. Bot.* **57,** 2037-2047 (2006).

66| Causse, M. *et al.* Whole genome resequencing in tomato reveals variation associated with introgression and breeding events. *BMC Genomics* **14**, 791 (2013).

67| Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., & van der Knaap, E. A. Retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **319**, 5869 (2008)

68| Smith, S.M. & Maughan, P.J. SNP genotyping using KASPar assays. *Methods Mol. Biol.* **1245**, 243-256 (2015).

69| You, F.M. *et al.* BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* **9**, 253 (2008).

70| Brukhin, V., Hernould, M., Gonzalez, N., Chevalier, C. & Mouras A. Flower development schedule in tomato *Lycopersicon esculentum* cv. sweet cherry. *Sex Plant Reprod* **15**, 311-320 (2003).

71| Robinson, J.T. *et al*. Integrative Genomics Viewer. *Nat Biotechnol* **29**, 24–26 (2011).

72| Thorvaldsdóttir, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192 (2013).

73| Goecks. *et al.* Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*. **11**, R86 (2010).

74| Blankenberg, D. "Galaxy: a web-based genome analysis tool for experimentalists". *Current Protocols in Molecular Biology*. Chapter 19: Unit 19.10.1-21 (2010).

75| Giardine, B. *et al.* "Galaxy: a platform for interactive large-scale genome analysis." *Genome Res.* **15**, 1451-1455 (2005).

76| Lisec, J., Schauer, N., Kopka, J., Willmitzer, L. & Fernie, A.R. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* **1**, 387-396 (2006).

77| Scott, J.W. & Harbaugh, B.K. Micro-Tom. A miniature dwarf tomato. *Florida Agricultural Experimental Station Circular* S-**370**, 1–6 (1989).

78| Martí, E., Gisbert, C., Bishop, G.J., Dixon, M.S. & García-Martínez, J.L. Genetic and physiological characterization of tomato cv. Micro-Tom. *J. Exp. Bot*. **57**, 2037-2047 (2006).

79| Rothan, C. *et al*. Culture of the tomato Micro-Tom cultivar in Greenhouse. In *Plant Signal Transduction: methods and protocols*, second edition. JR Botella Mesa and MA Botella Mesa eds. Springer publisher. *Methods in Molecular Biology* (accepted) (2015).

80| Fernandez, A.I. *et al.* Flexible tools for gene expression and silencing in tomato. *Plant Physiol.* **151**, 1729-1740 (2009).

81| Ferreira e Silva, G.F. *et al.* microRNA156-targeted SPL/SBP box transcription factors regulate tomato ovary and fruit development. *Plant J.* **78**, 604-618 (2014).

82| Brooks, C., Nekrasov, V., Lippman, Z.B. & Van Eck, J. Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiol.* **166**, 1292-1297 (2014).

83| Ron, M. et al. Hairy root transformation using Agrobacterium rhizogenes as a tool for exploring cell type-specific gene expression and function using tomato as a model. *Plant Physiol.* **166**, 455-469 (2014).

84| Orzaez, D. *et al.* A visual reporter system for virus-induced gene silencing in tomato fruit based on anthocyanin accumulation. *Plant Physiol.* **150**, 1122-1134 (2009).

85| Quadrana, L. *et al.* Coupling virus-induced gene silencing to exogenous green fluorescence protein expression provides a highly efficient system for functional genomics in Arabidopsis and across all stages of tomato fruit development. *Plant Physiol.* **156**, 1278-1291 (2011).

86| Senthil-Kumar, M. & Mysore, K.S. Tobacco rattle virus-based virus-induced gene silencing in Nicotiana benthamiana. *Nat. Protoc.* **9**, 1549-1562 (2014).

87| Brooks, C., Nekrasov, V., Lippman Z.B., Van Eck, J. Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiol*. **166**, 1292-7(2014).

88| Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w^{1118}; iso-2; iso-3. *Fly* **6**, 80-92 (2012).

MUTAGENESIS AND PHENOTYPING

6 – 12 months

6 months

EMS 1 %
1st round

M_0 seeds → M_1 seeds → M_1 plants → M_2 seeds

2nd round

M_2 seed bulk from 1st round of EMS mutagenesis

M_2 family culture

DNA extraction and storage
(for TILLING)

M_2 family phenotyping

Phenotypic database

MUTATION SEGREGATION AND NGS MAPPING

6 months

6 months

Selection of mutant-of-interest

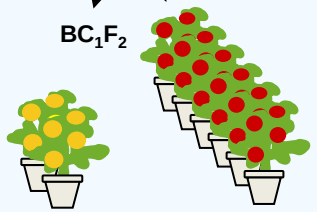
e.g. Yellow fruit colour
Recessive mutation

X

Backcross

BC_1F_1

BC_1F_2



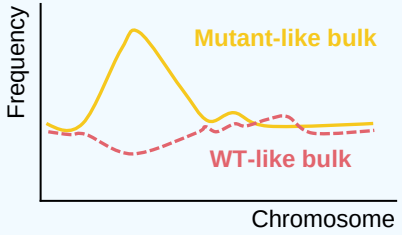
Mutant-like bulk

WT-like bulk

Mutation segregation
Phenotyping - Bulk set-up

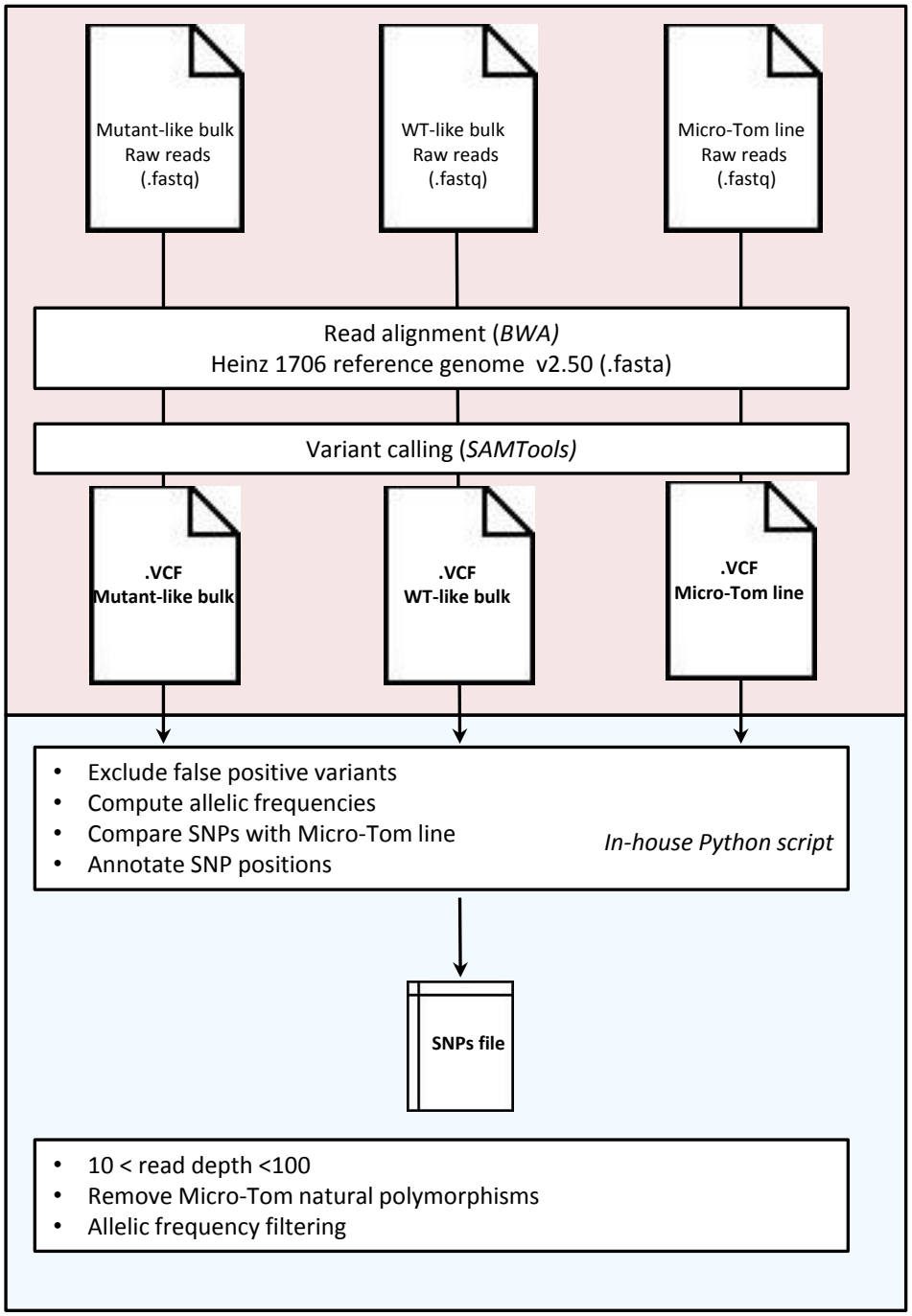
Sequencing
20 – 40 X
each bulk

Sequence mapping
Variant calling
Frequency analysis

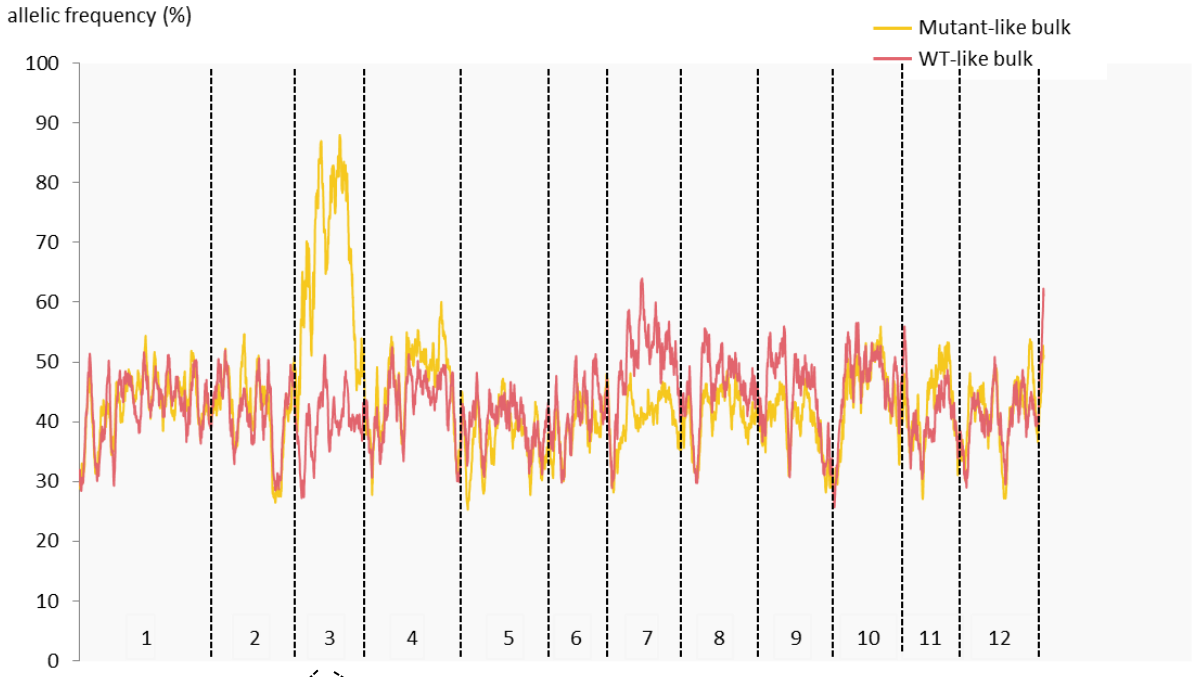


Detection of causal mutation

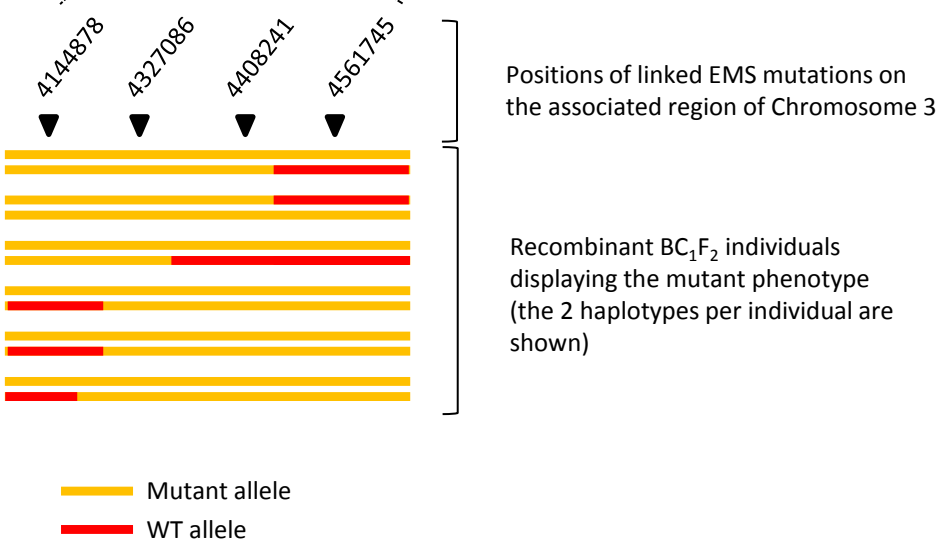
a


Mutant-like bulk
Raw reads
(.fastq)
WT-like bulk
Raw reads
(.fastq)
Micro-Tom line
Raw reads
(.fastq)
Read alignment (*BWA)*
Heinz 1706 reference genome v2.50 (.fasta)
Variant calling (*SAMTools)*
.VCF
Mutant-like bulk
.VCF
WT-like bulk
.VCF
Micro-Tom line
• Exclude false positive variants
• Compute allelic frequencies
• Compare SNPs with Micro-Tom line
• Annotate SNP positions
In-house Python script
SNPs file
• 10 < read depth <100
• Remove Micro-Tom natural polymorphisms
• Allelic frequency filtering

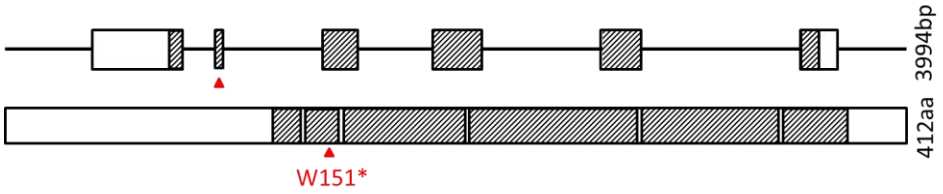
b


allelic frequency (%)
Mutant-like bulk
WT-like bulk
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12

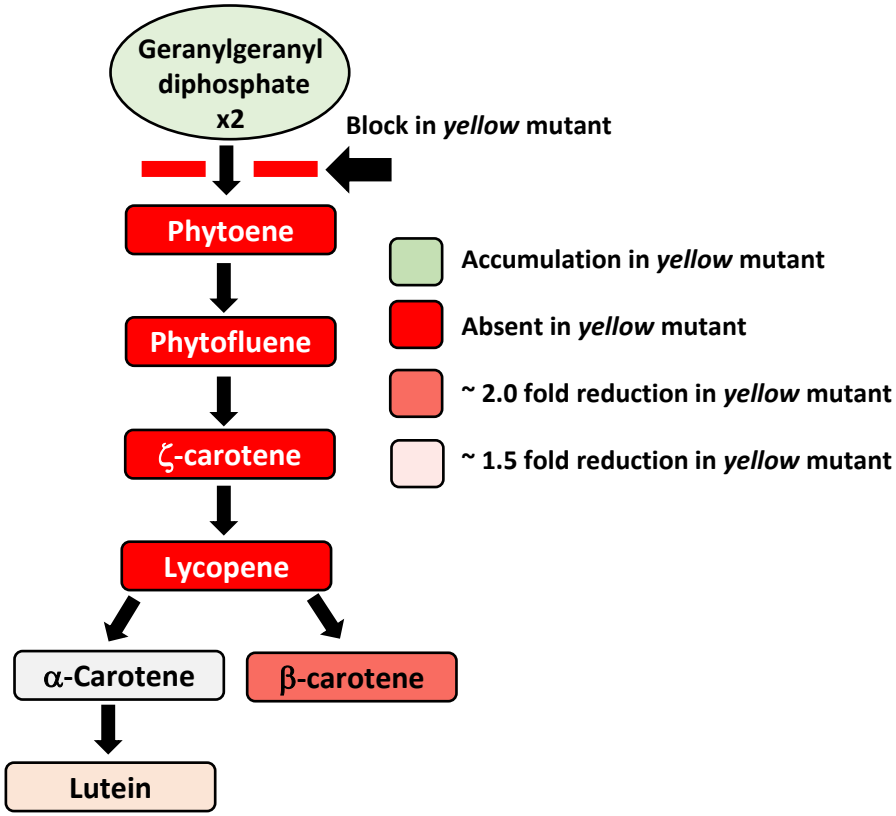
c


4144878
4327086
4408241
4561745
Positions of linked EMS mutations on the associated region of Chromosome 3
Recombinant BC1F2 individuals displaying the mutant phenotype (the 2 haplotypes per individual are shown)
Mutant allele
WT allele

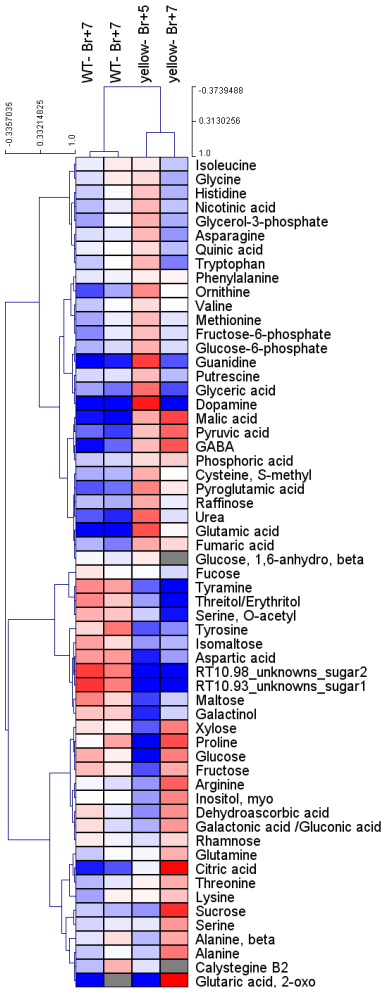
a


3994bp
412aa
W151*

b


Geranylgeranyl diphosphate x2
Block in yellow mutant
Phytoene
Phytofluene
ζ-carotene
Lycopene
α-Carotene
β-carotene
Lutein
Accumulation in yellow mutant
Absent in yellow mutant
~ 2.0 fold reduction in yellow mutant
~ 1.5 fold reduction in yellow mutant

c


WT- Br+7
WT- Br+7
yellow- Br+5
yellow- Br+7
Isoleucine
Glycine
Histidine
Nicotinic acid
Glycerol-3-phosphate
Asparagine
Quinic acid
Tryptophan
Phenylalanine
Ornithine
Valine
Methionine
Fructose-6-phosphate
Glucose-6-phosphate
Guanidine
Putrescine
Glyceric acid
Dopamine
Malic acid
Pyruvic acid
GABA
Phosphoric acid
Cysteine, S-methyl
Pyroglutamic acid
Raffinose
Urea
Glutamic acid
Fumaric acid
Glucose, 1,6-anhydro, beta
Fucose
Tyramine
Threitol/Erythritol
Serine, O-acetyl
Tyrosine
Isomaltose
Aspartic acid
RT10.98_unknowns_sugar2
RT10.93_unknowns_sugar1
Maltose
Galactinol
Xylose
Proline
Glucose
Fructose
Arginine
Inositol, myo
Dehydroascorbic acid
Galactonic acid /Gluconic acid
Rhamnose
Glutamine
Citric acid
Threonine
Lysine
Sucrose
Serine
Alanine, beta
Alanine
Calystegine B2
Glutaric acid, 2-oxo