

DATASET BRIEF

An extensive proteome map of tomato (*Solanum lycopersicum*) fruit pericarp

Jiixin Xu^{1,2}, Laura Pascual¹, Rémy Aurand^{1,3}, Jean-Paul Bouchet¹, Benoît Valot⁴, Michel Zivy⁴, Mathilde Causse¹ and Mireille Faurobert¹

¹INRA, UR1052, Unité de Génétique et Amélioration des Fruits et Légumes, Montfavet, France

²College of Horticulture, Northwest A&F University, Yang Ling, P. R. China

³INRA, UR1115 Plantes et Systèmes de Culture Horticoles, Avignon, France

⁴INRA/Université Paris-Sud/CNRS, Plateforme d'Analyse Protéomique de Paris Sud-Ouest, UMR 0320/UMR 8120 de Génétique Végétale, Gif-sur-Yvette, France

Tomato (*Solanum lycopersicum*) is the model species for studying fleshy fruit development. An extensive proteome map of the fruit pericarp is described in light of the high-quality genome sequence. The proteomes of fruit pericarp from 12 tomato genotypes at two developmental stages (cell expansion and orange-red) were analyzed. The 2DE reference map included 506 spots identified by nano-LC/MS and the International Tomato Annotation Group Database searching. A total of 425 spots corresponded to unique proteins. Thirty-four spots resulted from the transcription of genes belonging to multigene families involving two to six genes. A total of 47 spots corresponded to a mixture of different proteins. The whole protein set was classified according to Gene Ontology annotation. The quantitative protein variation was analyzed in relation to genotype and developmental stage. This tomato fruit proteome dataset is currently the largest available and constitutes a valuable tool for comparative genetic studies of tomato genome expression at the protein level. All MS data have been deposited in the ProteomeXchange with identifier PXD000105.

Received: December 20, 2012

Revised: May 16, 2013

Accepted: July 20, 2013

Keywords:

Fruit pericarp proteome / Genome sequence / Liquid chromatography–mass spectrometry / Plant proteomics / Tomato / Two-dimensional gel electrophoresis



Additional supporting information may be found in the online version of this article at the publisher's web-site

Protein expression integrates posttranscriptional and PT modifications that modulate the quantity, localization, and efficiency of the final product within the cell. The plasticity of a phenotype is driven by genetic variations in the levels of proteins and metabolites [1]. Recently, protein metabolism and, especially, protein stability were suggested to play a major role in plant growth, yield, and heterosis [2]. Knowledge about the fruit proteome is thus a challenging area of research, as reviewed by Palma et al. [3].

Apart from being one of the most important vegetables consumed worldwide, tomato (*Solanum lycopersicum*) is also the model species for studying fleshy fruit development [4]. This self-pollinating species exhibits a large range of phenotypes but a limited molecular diversity. Studies on the tomato fruit proteome have been recently reviewed by Faurobert et al. [5]. Most of the studies were carried out on one or two genotypes and identified a relatively small number of proteins, failing to provide an insight into the extent of natural genetic variation [6–12]. Improvement of protein identification technologies and the availability of the high-quality tomato genome sequence [13] have increased the efficiency of protein identification. We herein provide an extensive characterization of the tomato fruit proteome and complete the study of Faurobert et al. [12] in light of the genome

Correspondence: Dr. Mathilde Causse, INRA, UR1052, Unité de Génétique et Amélioration des Fruits et Légumes, CS 60094–84143 Montfavet Cedex, France

E-mail: Mathilde.Causse@avignon.inra.fr

Fax: +33-04-32-72-27-02

sequence. The proteome map presented here results from two distinct experiments. It consists of 506 protein spots identified from 11 692 peptide sequences. Such resources may help to improve genome sequence annotation.

The first experiment describes the proteome diversity in eight genotypes, which represent a large range of tomato genetic and phenotypic diversity identified in a collection of 360 accessions [14]. It describes the proteomes of the fruit pericarps of four *S. lycopersicum* lines (Levovil, Stupicke Polni Rane, LA0147, and Ferum) and four *S. lycopersicum* var *cerasiforme* lines (Cervil, Criollo, Plovdiv 24A, and LA1420). Cervil produces small fruits (less than 10 g). Levovil, Ferum, and LA0147 genotypes have large fruits (more than 100 g). Stupicke Polni Rane, Criollo, Plovdiv 24A, and LA1420 have intermediate fruit size. The second experiment focused on fruit proteome in relation with sensory quality. Three parental lines, contrasted for fruit quality (Cervil, Levovil, and VilB, a large fruited tomato), and three lines with intermediate fruit size carrying introgressed chromosome fragments from Cervil in Levovil (L4 and L9) and VilB backgrounds (B9) were compared [15]. Plants were grown from February to June 2010 under greenhouse conditions (16/20°C) in Avignon (South of France) for the first experiment and in Bellegarde (40 km west from Avignon) for the second one. Plants were grown under natural light and identical ferti-irrigation conditions. Given the large effect of developmental stage on fruit proteome, which has been previously demonstrated [12], fruits were collected at two stages of development in both experiments, cell expansion (25, 20, and 14 days after anthesis for large-, intermediate-, and small-fruited lines, respectively) and orange-red stage, according to the fruit color. Three to four biological pools of 5 to 20 fruits per genotype and per developmental stage were harvested. Fruit pericarp was isolated and immediately frozen, then ground in liquid nitrogen and stored at –80°C. Proteins were extracted using a phenol extraction method developed by Faurobert et al. [16] and separated by 2DE according to Page et al. [7]. After Coomassie colloidal staining, image analysis was performed with Samespot version 4.1 software (Nonlinear Dynamics) and the normalized spot volumes were assessed. Statistical analyses were performed using the R software (R Development Core Team 2005). Spots volumes were compared through two-way ANOVA (testing the differences between genotypes, stages, and their interaction). A PCA was carried out on spot abundance in the first experiment.

Within each experiment, the varying spots ($p < 0.05$) were picked for identification by nano-LC-MS/MS method as detailed in Supporting Information text 1. FASTA sequences of the identified proteins were used to re-annotate the proteins using the Blast2GO package [17]. Sequences were compared against the NCBI-NR database of nonredundant protein sequences using BLASTX with the default setting.

In the first experiment, we compared the pericarp proteomes of eight genotypes at cell expansion and orange-red stages, and 1230 spots were detected (Table 1). The ANOVA analysis revealed 424 spots that significantly changed in pro-

Table 1. Summary of the protein spots studied and identified

	Number
Total number of protein spots detected	2307
Number of spots identified by MS	833
Nonredundantly identified spots across experiments	506
Number of identified spectra for the 506 spots	16 598
Number of unique peptides for the 506 spots	11 692
Number of spots resulting from protein mix	47
Number of spots encoded by a unique gene locus	333
Number of spots corresponding to multigene family	34
Number of loci contributing to one spot	1–6
Number of protein functions displaying multiple spots	63

tein abundance, 333 according to the developmental stage, 321 according to the genotype, and 215 according to a stage and genotype interaction. In the second experiment, 1077 spots were detected comparing six genotypes at the two stages, among which 419 showed variation in abundance, 351 varying according to the genotype, 251 according to the stage, and 176 varied according to an interaction between genotype and stage (Supporting Information Table 1).

The large variation in the tomato pericarp proteome during fruit development agrees with the results of Faurobert et al. [12] and other studies on fruits reviewed by Palma et al. [3]. In both experiments, developmental stage was the main discriminating factor, but the variation across the genotypes was also very important at both stages, as illustrated in Fig. 1 for the first experiment. The first PCA plan accounted for 30% of the variation at the cell expansion stage and 27% at the orange-red stage (Fig. 1). The three biological replicates of each genotype appeared well clustered. The genotypes showed the same pattern of clustering for the two developmental stages. Cervil was distinct from all other genotypes. The *S. lycopersicum* genotypes were more tightly clustered than the *S. lycopersicum* var *cerasiforme* genotypes. The possibility to discriminate genotypes according to their proteome profile has already been demonstrated in roots under salt stress condition by Manaa et al. [8].

The 424 and 419 variable spots (with significant ANOVA) of each experiment were submitted to MS. For both experiments, a total of 16 598 spectra were identified by nano LC-MS/MS (corresponding to 11 692 unique peptides). Ten spots could not be identified. This very low level of failure was due to the high sensitivity of MS and the high quality of the genome annotation. On average, the detected peptides covered 54% of protein sequences. Supporting Information Table 1 lists the spots and their main features.

The eye comparison between the two sets of gel images was facilitated by the presence of Levovil and Cervil in both experiments. According to gel comparison and MS identification, 337 spots were common to both experiments. Their position on 2D gels is illustrated in Supporting Information

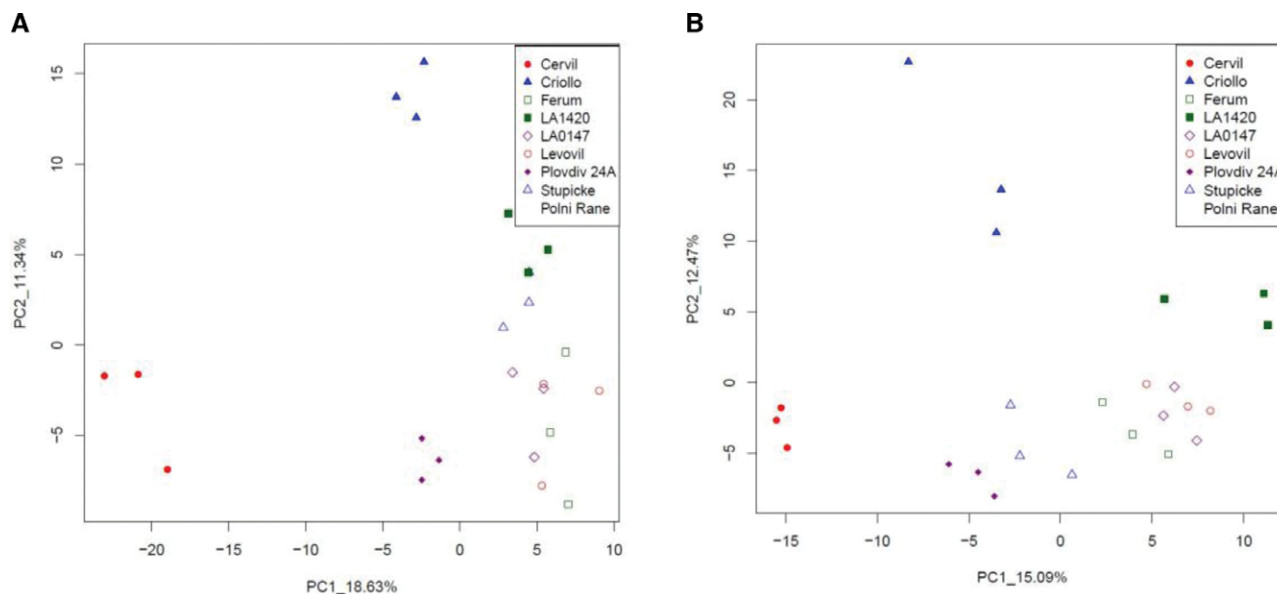


Figure 1. PCA based on the volume of 424 variable spots detected at cell expansion (A) and orange-red (B) fruit stages, in the first experiment. The eight genotypes included four large-fruited lines (LA0147, Levovil, Ferum, and Stupicke Polni Rane) and four cherry tomato lines (Cervil, Criollo, Plovdiv 24A, and LA1420). Each dot corresponds to a biological replicate. The percentages of total variation accounted for by each component are indicated along the axes.

Fig. 1. Taking into account the specific and common spots from each experiment, we identified a total of 506 different spots (Table 1).

Mass spectrometry of the 506 analyzed spots allowed identification of 425 spots corresponding to 425 unique proteins. Regarding the remaining 81 spots, 47 were identified as containing a mix of proteins with diverse functions (Supporting Information Table 1). The remaining 34 spots corresponded to the translation of several members of a multigenic family (Supporting Information Table 2). The number of loci that contributed to one protein spot varied from two, in 20 cases, to a maximum of six loci in the case of spot 74. This particular spot was identified as an HSP of high molecular weight (around 70 kDa). The multigenic nature of HSPs is well known [18]. In eight cases, two or more identified genes were located on the same chromosome in a very narrow region (available on the Solanaceae Genomic Network site) and corresponded to duplicated genes in tandem. For example, spot 36 corresponded to 1-aminocyclopropane-1-carboxylate-oxidase resulting from the transcription of two neighbor loci, *Solyc07g049530.2.1* and *Solyc07g049550.2.1*. Alternatively, these two loci could correspond to an incorrect annotation since according to the relative abundance index (Protein Abundance Index (PAI), [19]), the first locus seemed to contribute more to the final protein abundance. On the other hand, several loci equally contributed to protein abundance for four spots (e.g. spot 30 corresponded to an Actin protein encoded by two different genes, one located on chromosome 3, the other on chromosome 11). Apart from the spots resulting from the transcription of several genes, 63 genes were detected in several spots (from two

to five). In most cases (56 spots), the spots were close together on the 2D gel. For instance, two closely located spots (spots 125 and 142) corresponded to the metacaspase 7 protein (*Solyc09g098150.2.1*, Supporting Information Fig. 1). In seven cases, both closely and distantly located spots corresponded to the same protein. For example, the acid beta-fructofuranosidase (*Solyc03g083910.2.1*) was represented by the closely located spots 51 and 87 and by spots 237, 239, and 284, located further away. The presence of multiple spots may be due to PTMs, splice variants, protein degradation, or allelic variation, and has already been reported for acid beta-fructofuranosidase by Faurobert et al. [12]. For instance, five closely located spots (115, 205, 214, 256, and 323) corresponded to enolase protein (*Solyc09g009020.2.1*). Each spot was picked from one specific genotype and they showed significant quantitative differences according to the genotypes.

Many proteins have pleiotropic functions within the cell and it is therefore complicated to draw major conclusions from basic protein functional classification. However, we propose here a functional classification of the proteins according to the GO terms obtained after re-annotating gene sequences with Blast2Go based on biological process term (level 3). Proteins were assigned to 16 categories (Fig. 2). Primary metabolic process represented the largest class (27%). Several enzymes were involved in sugar synthesis, such as fructokinase (spots 15, 218, and 363), glucose-6-phosphate 1-dehydrogenase (spot 329), fructose-bisphosphate aldolase (spot 241), UTP-glucose-1-phosphate uridylyltransferase (spots 344 and 403), or in cysteine synthesis (cysteine synthase, spots 26, 179, and 402). Malate dehydrogenase (spots 4 and 177) is involved in the tricarboxylic

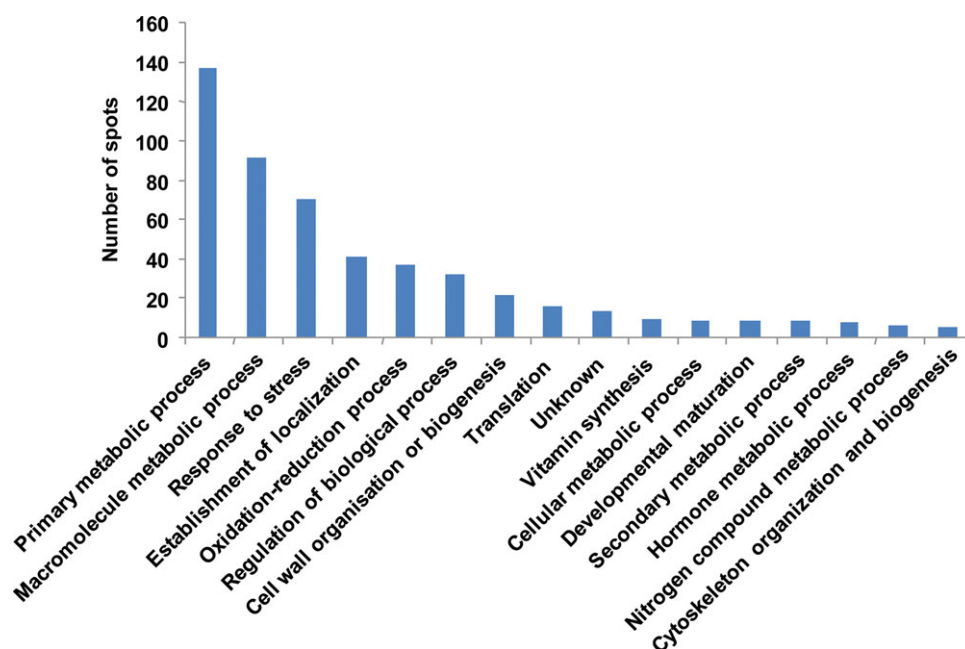


Figure 2. Distribution of the 506 tomato protein spots identified by MS into biological classes according to gene ontology classification.

acid cycle, which provides the energy necessary for ripening and was also identified in the primary metabolism subgroup. Proteins involved in macromolecular metabolic process accounted for 18% of all identified proteins. Chaperonins (spots 152, 181, 201, 249, 330, 382, 390, and 414) and proteasome subunits (spots 114, 182, 336, 375, 387, RA311, and RA 394) were identified in this subgroup. Proteins related to stress responses represented the third largest subgroup (14%). This group was clearly dominated by HSPs (23 spots), but also included some enzymes such as GST (spots 2, 80, 112, 204, 293, and 374) that has a role in detoxification processes. HSPs are modulated by a wide range of environmental stresses, but also during fruit development [12, 20, 21]. Other smaller functional groups, which represented a relatively low number of proteins, were identified. Finally, 17 spots were classified into unknown category.

In conclusion, we provide the first comprehensive proteome reference map of the tomato fruit pericarp at two developmental stages from 12 genotypes representing a wide phenotypic and genotypic diversity. We identified 506 spots corresponding to 333 proteins expressed in fruit and described the physiological function of these proteins. These data provide experimental evidence for tomato fruit proteins that had only been predicted by genome annotation and constitute valuable tools for comparative studies of protein expression. Furthermore, they can be useful for further investigation of genetic variation. We are now characterizing the fruit proteome of hybrids from the genotypes used in this study to assess the inheritance pattern of the variable proteins.

The MS proteomics data in this paper have been deposited in the ProteomeXchange Consortium

(<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository [22]; dataset identifier PXD000105.

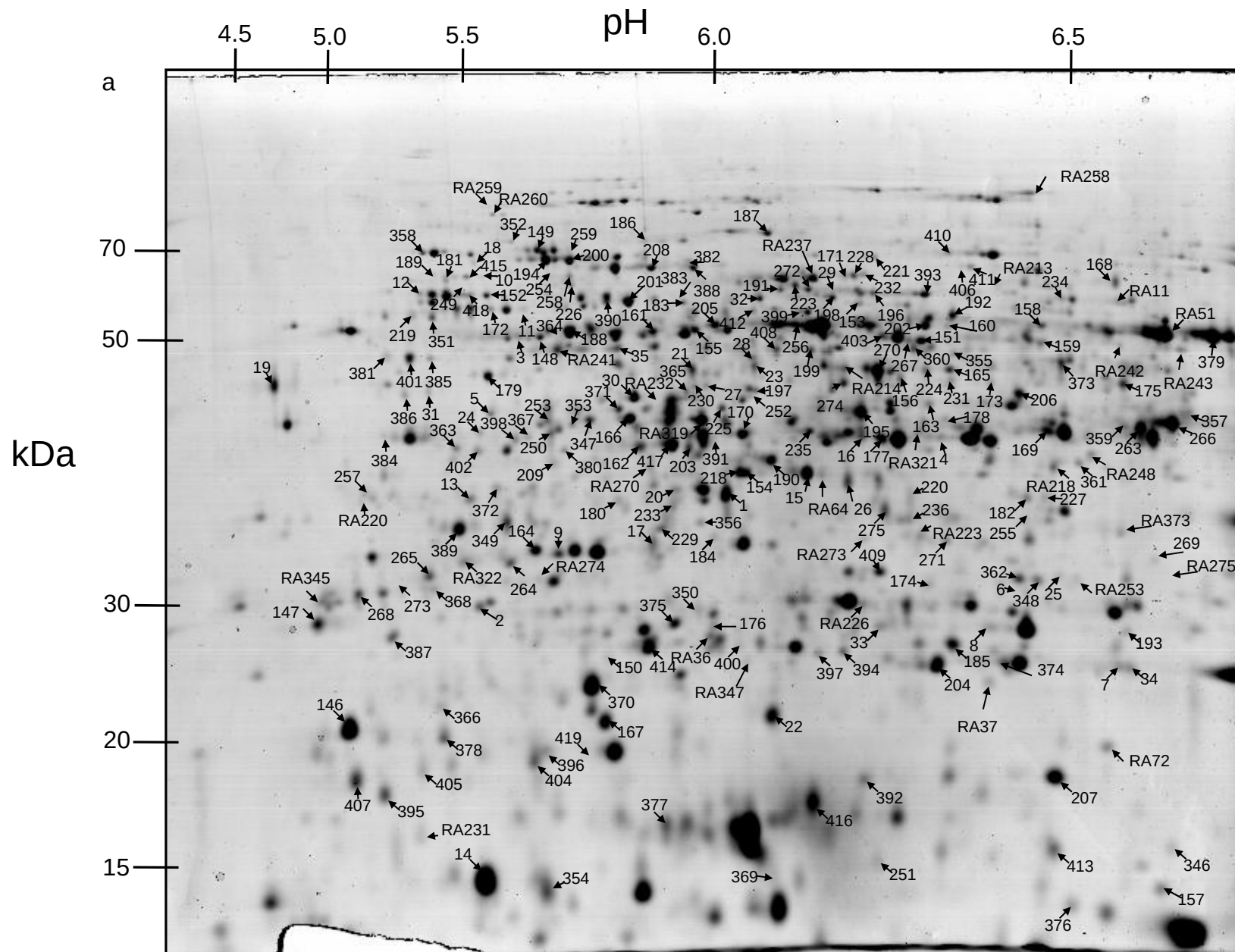
We thank Caroline Callot and Karine Leyre for technical help. We also thank Esther Pelpoir for her help in fruit sampling, Yolande Carretero for taking care of the plants in the greenhouse, and Rebecca Stevens for English revision. Finally, we thank CTIFL (Centre Technique Interprofessionnel des Fruits et Légumes) from Balandran for plant cultivation. This work was supported by ANR Genomic project MAGICTomSNP project and EUSOL European project PL016214–2.

The authors have declared no conflict of interest.

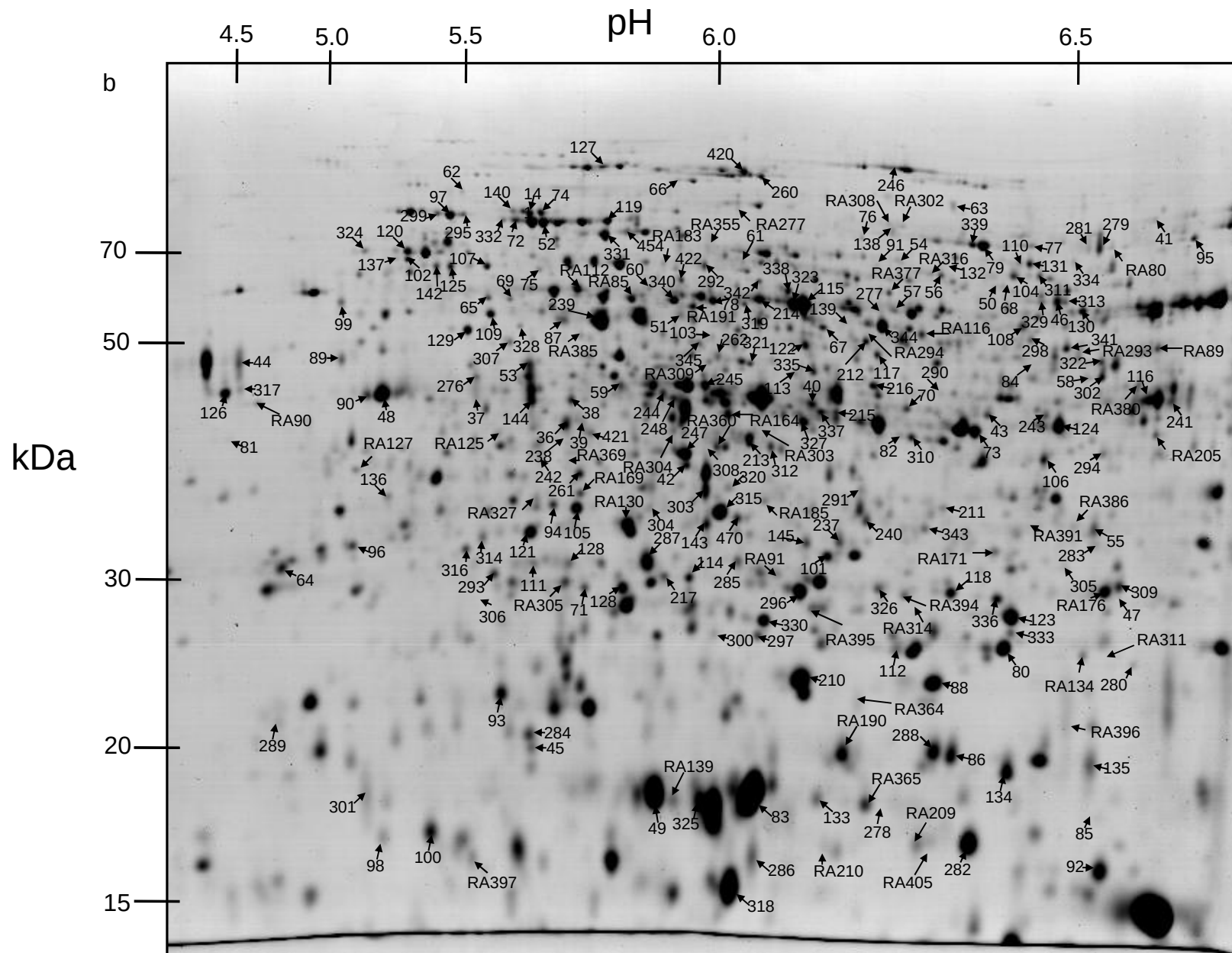
References

- [1] Weckwerth, W., Integration of metabolomics and proteomics in molecular plant physiology – coping with the complexity by data-dimensionality reduction. *Physiol. Plant.* 2008, **132**, 176–189.
- [2] Goff, S. A., A unifying theory for general multigenic heterosis: energy efficiency, protein metabolism, and implications for molecular breeding. *New Phytol.* 2011, **189**, 923–937.
- [3] Palma, J. M., Corpas, F. J., del Rio, L. A., Proteomics as an approach to the understanding of the molecular physiology of fruit development and ripening. *J. Proteomics* 2011, **74**, 1230–1243.
- [4] Giovannoni, J. J., Genetic regulation of fruit development and ripening. *Plant Cell* 2004, **16**(Suppl), S170–S180.
- [5] Faurobert, M., Iijima, Y., Aoki, K., Proteomics and metabolomics, in: Liedl, B. E., Labate, J. A., Stommel, J. R.,

- Slade, A., Kole, C. (Eds.), *Genetics, Genomics and Breeding of Tomato*, CRC Press, Florida, USA, 2012, pp. 403–427.
- [6] Iwahashi, Y., Hosoda, H., Effect of heat stress on tomato fruit protein expression. *Electrophoresis* 2000, **21**, 1766–1771.
- [7] Page, D., Gouble, B., Valot, B., Bouchet, J. P. et al., Protective proteins are differentially expressed in tomato genotypes differing for their tolerance to low-temperature storage. *Planta* 2010, **232**, 483–500.
- [8] Manaa, A., Ben Ahmed, H., Valot, B., Bouchet, J. P. et al., Salt and genotype impact on plant physiology and root proteome variations in tomato. *J. Exp. Bot.* 2011, **62**, 2797–2813.
- [9] Marjanovic, M., Stikic, R., Vucelic-Radovic, B., Savic, S. et al., Growth and proteomic analysis of tomato fruit under partial root-zone drying. *OMICS* 2012, **16**, 343–356.
- [10] Sheoran, I. S., Olson, D. J., Ross, A. R., Sawhney, V. K., Proteome analysis of embryo and endosperm from germinating tomato seeds. *Proteomics* 2005, **5**, 3752–3764.
- [11] Rocco, M., D'Ambrosio, C., Arena, S., Faurobert, M. et al., Proteomic analysis of tomato fruits from two ecotypes during ripening. *Proteomics* 2006, **6**, 3781–3791.
- [12] Faurobert, M., Mihr, C., Bertin, N., Pawlowski, T. et al., Major proteome variations associated with cherry tomato pericarp development and ripening. *Plant Physiol.* 2007, **143**, 1327–1346.
- [13] Sato, S., Tabata, S., Hirakawa, H., Asamizu, E. et al., The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 2012, **485**, 635–641.
- [14] Ranc, N., Munos, S., Santoni, S., Causse, M., A clarified position for *solanum lycopersicum* var. *cerasiforme* in the evolutionary history of tomatoes (solanaceae). *BMC Plant Biol.* 2008, **8**, 130.
- [15] Chaib, J., Devaux, M. F., Grotte, M. G., Robini, K. et al., Physiological relationships among physical, sensory, and morphological attributes of texture in tomato fruits. *J. Exp. Bot.* 2007, **58**, 1915–1925.
- [16] Faurobert, M., Pelpoir, E., Chaib, J., Phenol extraction of proteins for proteomic studies of recalcitrant plant tissues. *Methods Mol. Biol.* 2007, **355**, 9–14.
- [17] Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J. et al., Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 2005, **21**, 3674–3676.
- [18] Sung, D., Vierling, E., Guy, C., Comprehensive expression profile analysis of the Arabidopsis hsp70 gene family. *Plant Physiol.* 2001, **126**, 789–800.
- [19] Rappsilber, J., Ryder, U., Lamond, A. I., Mann, M., Large-scale proteomic analysis of the human spliceosome. *Genome Res.* 2002, **12**, 1231–1245.
- [20] Wang, W. X., Vinocur, B., Shoseyov, O., Altman, A., Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends Plant Sci.* 2004, **9**, 244–252.
- [21] Neta-Sharir, I., Isaacson, T., Lurie, S., Weiss, D., Dual role for tomato heat shock protein 21: protecting photosystem II from oxidative stress and promoting color changes during fruit maturation. *Plant Cell* 2005, **17**, 1829–1838.
- [22] Vizcaino, J. A., Côté, R. G., Csordas, A., Dianes, J. A. et al., The PRoteomics IDentifications (PRIDE) database and associated tools: Status in 2013. *Nucleic Acids Res.* 2013, **41**, D1063–D1069.



Supplemental Figure S1 Representative two-dimensional electrophoresis gel of tomato pericap proteins of Criollo at cell expansion stage (a) and orange red stage (b). The positions and numbers of 424 identified protein spots are indicated by arrows.



Supplemental Figure S1 Representative two-dimensional electrophoresis gel of tomato pericarp proteins of Criollo at cell expansion stage (a) and orange red stage (b). The positions and numbers of 424 identified protein spots are indicated by arrows.

Supplemental text S1: Method for protein identification by nano-LC-MS/MS

In-gel digestion was performed with the Progest system (Genomic Solution) according to a standard trypsin protocol. Gel pieces were washed twice by successive separate baths of 10% acetic acid, 40% ethanol, and acetonitrile (ACN). They were then washed twice with successive baths of 25 mM NH_4CO_3 and ACN. Digestion was subsequently performed for 6 h at 37°C with 125 ng of modified trypsin (Promega) dissolved in 20% methanol and 20 mM NH_4CO_3 . The peptides were extracted successively with 2% trifluoroacetic acid (TFA) and 50% ACN and then with ACN. Peptide extracts were dried in a vacuum centrifuge and suspended in 20 μL of 0.05% TFA, 0.05% HCOOH , and 2% ACN.

HPLC was performed on an NanoLC-Ultra system (Eksigent). A 4 μL sample was loaded at 7.5 $\mu\text{L}/\text{min}^{-1}$ on a precolumn cartridge (stationary phase: C18 Biosphere, 5 μm ; column: 100 μm i.d., 2 cm; Nanoseparations) and desalted with 0.1% HCOOH . After 3 min, the precolumn cartridge was connected to the separating PepMap C18 column (stationary phase: C18 Biosphere, 3 μm ; column: 75 μm i.d., 150 mm; Nanoseparations). Buffers were 0.1% HCOOH in water (A) and 0.1% HCOOH in ACN (B). The peptide separation was achieved with a linear gradient from 5 to 30% B for 11 min at 300 nL/min⁻¹. Including the regeneration step at 95% B and the equilibration step at 95% A, one run took 25 min.

Eluted peptides were analysed on-line with a LTQ XL ion trap (Thermo Electron) using a nanoelectrospray interface. Ionization (1.5 kV ionization potential) was performed with liquid junction and a noncoated capillary probe (10 μm i.d.; New Objective). Peptide ions were analysed using Xcalibur 2.07 with the following data-dependent acquisition steps: (1) full MS scan (mass-to-charge ratio (m/z) 300 to 1400, centroid mode) and (2) MS/MS ($q_z = 0.25$, activation time = 30 ms, and collision energy = 35%; centroid mode). Steps 2 was repeated for the three major ions detected in step 1. Dynamic exclusion was set to 30 s.

A database search was performed with XTandem (version 2010.12.01.1) (<http://www.thegpm.org/TANDEM/>). Enzymatic cleavage was declared as a trypsin digestion with one possible misscleavage. Cys carboxyamidomethylation and Met oxidation were set to static and possible modifications, respectively. Precursor mass and fragment mass tolerance were 2.0 and 0.5, respectively. A refinement search was added with similar parameters except that semi-trypsinic peptide and possible N-ter proteins acetylation were searched. The International Tomato Annotation Group (ITAG) Release 2.3 predicted proteins (SL2.40) database (<http://solgenomics.net/>, containing 34727 protein coding genes, 30855 of them

supported by RNA sequencing data) and a contaminant database (trypsin, keratins, ...) were used. Only peptides with a E value smaller than 0.1 were reported.

Identified proteins were filtered and grouped using XTandem Pipeline (<http://pappso.inra.fr/bioinfo/xtandempipeline/>) according to : (1) A minimum of two different peptides was required with a E value smaller than 0.03, (2) a protein E value (calculated as the product of unique peptide E values) smaller than 10^{-3} . In case of identification with only two or three MS/MS spectra, similarity between the experimental and the theoretical MS/MS spectra was visually checked. To take redundancy into account, proteins with at least one peptide in common were grouped. This allowed to group proteins of similar function. Within each group, proteins with at least one specific peptide relatively to other members of the group were reported as sub-groups.