Quantitative metabolic profiles of tomato flesh and seeds during fruit development: complementary analysis with ANN and PCA

Fabien Mounet,[a] Martine Lemaire-Chamley,[a] Mickaël Maucourt,[a,b] Cécile Cabasson,[a] Jean-Luc Giraudel,[c] Catherine Deborde,[a,b] René Lessire,[d] Philippe Gallusci,[a] Anne Bertrand,[a] Monique Gaudillère,[a] Christophe Rothan,[a] Dominique Rolin,[a] and Annick Moing[a],*

[a]*UMR Biologie du Fruit, Centre INRA de Bordeaux, INRA – Université Bordeaux 1 – Université Victor Ségalen Bordeaux 2, IFR BVI, BP 81, Villenave d'Ornon cedex, F-33883, France*

[b]*Pôle Métabolome de la plate-forme Génomique Fonctionnelle Bordeaux, IFR BVI, BP 81, Villenave d'Ornon cedex, F-33883, France*

[c]*Equipe Périgourdine de Chimie Appliquée, ISM, UMR 5255, CNRS-Université Bordeaux 1, rue Doyen Joseph Lajugie, Périgueux cedex, F-24019, France*

[d]*Laboratoire de Biogenèse Membranaire, UMR CNRS 5200, Université Victor Segalen Bordeaux 2, 146 rue Léo Saignat, Case 92, Bordeaux cedex, F-33076, France*

Tomato, an essential crop in terms of economic importance and nutritional quality, is also used as a model species for all fleshy fruits and for genomics of Solanaceae. Tomato fruit quality at harvest is a direct function of its metabolite content, which in turn is a result of many physiological changes during fruit development. The aim of the work reported here was to develop a global approach to characterize changes in metabolic profiles in two interdependent tissues from the same tomato fruits. Absolute quantification data of compounds in flesh and seeds from 8 days to 45 days post anthesis (DPA) were obtained through untargeted (proton nuclear magnetic resonance, ^{1}H-NMR) and targeted metabolic profiling (liquid chromatography with diode array detection (LC-DAD) or gas chromatography with flame ionization detection (GC-FID)). These data were analyzed with chemometric approaches. Kohonen self organizing maps (SOM) analysis of these data allowed us to combine multivariate (distribution of samples on Kohonen SOMs) and univariate information (component plane representation of metabolites) in a single analysis. This strategy confirmed published data and brought new insights on tomato flesh and seed composition, thus demonstrating its potential in metabolomics. The compositional changes were related to physiological processes occurring in each tissue. They pointed to (i) some parallel changes at early stages in relation to cell division and transitory storage of carbon, (ii) metabolites participating in the fleshy trait and (iii) metabolites involved in the specific developmental patterns of the seeds.

KEY WORDS: *Solanum lycopersicum*; fruit; seed; ^{1}H NMR; polar metabolites; lipids; isoprenoids; metabolic fingerprinting; metabolomics.

Abbreviations: ANN: Artificial neural networks; DPA: Days post anthesis; DW: Dry weight; FW: Fresh weight; FAME: Fatty acid methyl ester; FID: Free induction decay; GC-FID: Gas chromatography with flame ionization detection; LC-DAD: Liquid chromatography with diode array detection; NMR: Nuclear magnetic resonance; PCA: Principal component analysis; SOM: Self organizing map; TSP: (Trimethyl)propionic-2,2,3,3-*d4* acid sodium salt; VLCFA: Very long chain fatty acid

1. Introduction

Tomato is one of the most important fresh and processed vegetables in economic importance and consumption with a world production of about 122 Mt in 2005 (http://faostat.fao.org/). Tomato fruit organoleptic quality is a direct function of its metabolite content. Tomato is also a model species for studies of fruit development and ripening (Giovannoni, 2004; Stevens *et al.*, 2007) and was proposed as a model system to study seed development (Hilhorst *et al.*, 1998).

Metabolism related changes occur in response to developmental events (Tarpley *et al.*, 2005) in whole plants or individual organs such as fruit. The developmental changes, triggered by internal programming and response to environmental factors, induce metabolic transitions that result in modifications of final metabolite concentrations. Indeed, many structural and metabolic changes occur during tomato fruit flesh and seed development and maturation (Gillapsy *et al.*, 1993; White, 2002; Giovannoni, 2004; Cheniclet *et al.*, 2005). These processes allow development of both tissues in a coordinated manner (Ozga and Reinecke, 2003). On one hand, seeds influence flesh development through hormonal cross-talk (Gillapsy *et al.*, 1993). On the other

hand, fleshy tissues have been proposed to prevent precocious seed germination (Berry and Bewley, 1992). After ovule fertilization, seed development is characterized by different steps: seed formation, early embryo development with cell divisions in the peripheral integuments, endosperm development, embryo maturation involving cotyledon development, and finally root–shoot axis establishment (Gillapsy *et al.*, 1993). In parallel, after ovule fertilization, early tomato berry development is characterized by a period of cell division in fruit flesh followed by a long period of polyploidy-associated cell expansion (Gillapsy *et al.*, 1993; Cheniclet *et al.*, 2005; Lemaire-Chamley *et al.*, 2005). Cell expansion triggers the formation of large vacuolated cells associated with cell wall changes and accumulation of solutes such as sugars, organic acids and other compounds maintaining cell turgor. This cellular process greatly contributes to the fleshy trait of the fruit. Finally, during the ripening phase that occurs after embryo maturation, the tomato fruit acquires the color, texture and aroma attributes underlying its sensory quality.

To date, few metabolomics studies have been performed in tomato fruit. Gas chromatography-mass spectrometry (GC-MS) was used to characterize the fruit pericarp composition in transgenic plants (Roessner-Tunali *et al.*, 2003), assess metabolic diversity in genetic resources (Schauer *et al.*, 2005) or measure metabolic changes associated with fruit development (Carrari *et al.*, 2006). Metabolite profiling with proton NMR spectroscopy (^{1}H-NMR) was used on the whole fruit or pericarp tissue to detect unintended effects following a genetic modification (Le Gall *et al.*, 2003; Mattoo *et al.*, 2006). Targeted analyses of carbohydrates (Obiadalla-Ali *et al.*, 2004), amino acids (Boggio *et al.*, 2000), carotenoids and isoprenoids (Fraser *et al.*, 1994; Burns *et al.*, 2003) or lipids (Whitaker, 1988; Yilmaz *et al.*, 2001) have been achieved mainly during the ripening phase. Recent analyses of tomato seed composition have been focused on mature fruit seeds, consistent with the potential economic importance of tomato industry byproducts (Persia *et al.*, 2003; Knoblich *et al.*, 2005) and with fruit quality for human nutrition (Toor and Savage, 2005). Data on carbohydrate (Sunitha and Bradford, 2001), amino acid and carotenoid (Knoblich *et al.*, 2005) or lipid (Camara *et al.*, 2001) composition in seeds have been collected at the ripe fruit stage. However, data on compositional changes during seed development as published for *Arabidopsis* (Baud *et al.*, 2002) do not exist for tomato to our knowledge. Moreover, parallel analyses on both flesh and seed tissues of the same biological material are missing. The development of a metabolomics approach simultaneously analyzing tomato flesh and seeds offers new approaches to studying the relationships between metabolic changes and developmental patterns in fruit.

The aim of the present study was to characterize metabolic changes in tomato flesh and seeds in relation to crucial changes in fruit growth and development patterns. Firstly, we developed a global approach to quantify compositional changes in metabolic profiles during fruit development and ripening. We used untargeted metabolic profiling through ^{1}H-NMR, and targeted liquid chromatography with diode array detection (LC-DAD) or gas chromatography with flame ionization detection (GC-FID). We favored quantitative metabolite profiling, i.e. absolute quantification of unambiguously identified metabolites (Nielsen and Oliver, 2005), in order to facilitate the study of metabolite/metabolite or metabolites/developmental characteristics correlations. Secondly, the chemometric analyses of these data were achieved using Principal Component Analysis (PCA) and Artificial Neural Networks (ANN). This approach revealed parallel and opposing metabolic patterns in flesh and seeds, highlighted the tissue specificity of many metabolites, and provided new insights on the relationships between metabolic changes and development patterns in tomato fruit.

2. Materials and methods

2.1. Plant material

Tomato (*Solanum lycopersicum* L., cv Ailsa Craig) plants were grown in a growth chamber with a 15-h-day (25 °C)/9-h-night (20 °C) cycle with an irradiance of 400 μmol m^{-2} s^{-1} and 75–80% humidity. Individual flowers were tagged at anthesis (flower opening). The fruit number per truss was limited to six. Changes in fruit diameter were measured every 5 days on 12–36 fruits. The fruits were further selected according to size, color, and position on the truss (elimination of the first and last fruit of the truss), and collected at six times expressed in days post anthesis (DPA): division phase (8 DPA), transition between division and expansion phases (12 DPA), expansion phase (20 DPA), mature green stage (35 DPA) and red ripe fruit (45 DPA). For each stage, three pools of fruit were harvested from 15 plants: 12 fruit pools for 8 and 12 DPA, and 6 fruit pools for 20–45 DPA. Each fruit pool was collected from at least six plants. For each pool, the locular content was rapidly separated into 'seeds' and locular tissue. 'Flesh' samples were constituted of all the fruit tissues, including the locular tissue, without the seeds. All samples were rapidly frozen and ground in liquid nitrogen and stored at −80 °C until use for metabolite profiling. Frozen samples were lyophilized just before untargeted or targeted metabolite analyses.

Three other pools of 6 or 12 fruits (depending on the stage of development) cultivated in the same conditions were used to measure seed fresh weight, and flesh and seed dry matter contents. For each seed sample, seeds were collected from one fruit pool, the fresh weight was measured and seed number was counted to calculate seed fresh weight. Then for each fruit pool, part of the

flesh or seeds was weighed and lyophilized to determine dry matter content.

2.2. Extraction and ^{1}H-NMR analysis of polar compounds

Polar metabolites were extracted according to Moing et al. (2004). Briefly, for each sample, two subsamples (50 mg DW each) were extracted successively with 80% ethanol–water, 50% ethanol–water (v/v) and water at 80 °C for 15 min. The supernatants were combined, dried under vacuum and lyophilized. The dried extracts were titrated with KOD to pH 6 in 400 mM potassium phosphate buffer in D_2O, and lyophilized again. The dried titrated extracts were stored in darkness under vacuum at room temperature before ^{1}H-NMR analysis within one week.

^{1}H-NMR spectra were recorded on each dried titrated extract solubilized in 0.5 mL D_2O containing the sodium salt of (trimethyl)propionic-2,2,3,3-d4 acid (TSP) at a final concentration of 0.01% for chemical shift calibration, at 500.162 MHz on a Bruker Avance spectrometer using a 5 mm inverse probe and the ERETIC method (Akoka et al., 1999) for quantification as described previously (Moing et al., 2004). Ethylene diamine tetraacetic acid was added at a final concentration of 20 mM to improve spectrum resolution, especially in the citrate region. Sixty-four scans of 32 K data points were acquired with a spectral width of 6000 Hz, acquisition time of 2.73 s and recycle delays of 25 s. An automation procedure (automatic shimming and automatic sample loading) requiring about 40 min per sample was used for data acquisition. Preliminary data processing was carried out with XWINNMR software (Bruker Biospin, Karlsruhe, Germany). FIDs were Fourier transformed (0.3 Hz line broadening), manually phased and baseline corrected. The resulting spectra were aligned by shifting TSP signal to zero. Metabolite concentrations in the NMR tube were calculated using the metabolite mode of AMIX software (version 3.5.6, Bruker) for calculation of resonance areas, followed by data export to Excel software. Quantification was carried out using a glucose calibration curve and the proton amount corresponding to each resonance for all compounds except fructose, glutamine and glutamate. These metabolites were quantified using specific calibration curves of the corresponding compound. Metabolite concentrations in each sample were then calculated based on concentrations in the NMR tube and sample dry weight. The concentration of each organic or amino acid was expressed as g of the acid form per weight unit. ^{1}H-^{1}H COSY (2D-Homonuclear Correlation Spectroscopy) experiments were carried out to verify the identity of known compounds and to determine if signals from unknown compounds truly corresponded to signals from different compounds. Each COSY spectrum was obtained using the cosyqf45 pulse program (COSY using a 45° read

pulse), relaxation delay of 1 s, 90° pulse of 9.75 µs, spectral width 6000 and 5000 Hz for f1 and f2 dimensions respectively, 4 K datapoints in f2, 256 in f1. The concentration of NMR unknown compounds was calculated on the assumption that the measured resonance corresponded to one proton and using an arbitrary molecular weight of 100 Da.

2.3. Starch analysis

Starch remaining in the pellet after polar compounds extraction was converted to glucose using amyloglucosidase (Moing et al., 1994) and the resulting glucose was analyzed enzymatically following NADH production at 340 nm (Kunst et al., 1984) with adaptation to a microplate spectrophotometer (Dynatech, St Cloud, France) (Velterop and Vos, 2001).

2.4. Extraction and GC analysis of lipids

Lipid profiling was achieved by the quantification of fatty acids derived from lipids and converted to the corresponding fatty acid methyl esters (FAMEs) (Browse et al., 1986). From each fruit or seed sample, two aliquots (10 mg DW each) were heated for 1 h at 80 °C in presence of 1 mL of methanolic H_2SO_4 (2.5%, v/v) and 5 µg of C17:0 fatty acid as internal standard for quantification. After cooling, 1 mL hexane and 1 mL H_2O containing 2.5% NaCl (w/v) were added. FAMEs were extracted by vigorous shaking and the hexane phase was separated by centrifugation (1500×g, 5 min, 5 °C). The hexane extraction was repeated once and the hexane extracts were pooled and evaporated to dryness under a nitrogen flux. The dry extract was solubilized in 1 mL hexane and 1 µL was injected in a Hewlett-Packard 5890 II (Wilmington, DE, USA) gas chromatographer equipped with a Carbowax column (15 m×0.53 mm, 1.2 µm) (Alltech Associates, Deerfield, IL, USA), flame ionization detection and electronic integration (Hewlett-Packard 3396 III). Temperature gradient was: 160 °C for one min, increased to 190 °C at 20 °C min^{-1}, increased to 210 °C at 5 °C min^{-1}, and then 210 °C for 5 min. FAMEs were identified by comparing their retention times with those of commercial standards (Sigma Chemical Co., St Louis, MO, USA): C16:0 (palmitic acid), C18:0 (stearic acid), C18:1 (oleic acid), C18:2 (linoleic acid), C18:3 (linolenic acid) and C20:0 (arachidic acid) and C22:0 (behenic acid). FAMEs were quantified by comparison with the C17:0 response.

2.5. Extraction and LC analysis of isoprenoids

Isoprenoids were extracted and analyzed as described (Fraser et al., 2000) with slight modification (Télef et al., 2006). Briefly, 10 mg of lyophilized sample were mixed using a hand-held homogenizer with 1 mL of methanol and then 1 mL of Tris–HCl buffer (0.05 M, pH 7.5). The

suspension was incubated at 4 °C during 10 min with punctual inversion mix and 4 mL of chloroform were added. The samples were vortexed and centrifuged 5 min at 3000×g. The lower phase was removed and two additional extractions with 4 mL of chloroform were carried out on the aqueous phase. The pooled chloroform extracts were dried under a stream of nitrogen, immediately resuspended in 200 μL of ethyl acetate and injected into the HPLC system. Isoprenoid separation, identification and quantification were as described in Télef et al. (2006).

2.6. Chemicals

D_2O (99.9%) was purchased from Eurisotop (Gif sur Yvette, France). TSP (98%) was purchased from Aldrich (Saint Quentin Fallavier, France). All the other chemicals were of reagent grade.

2.7. Data analysis

For growth parameters and individual metabolites, mean ± standard deviation (SD) were calculated from n replicates. For all biochemical analyses two extractions were completed to measure the concentration of each biological replicate, then the mean of 3 biological replicates was calculated. Mean comparison between flesh and seeds for each stage of development was done using Student's t-test with SAS software version 8.01 (SAS Institute, 1990). Pearson correlation coefficients between metabolites were calculated. Significance levels for correlation coefficients r were determined following the number of metabolite pairs n by using $t = r \cdot (n-2)^{0.5}/(1-r^2)^{0.5}$.

To explore the metabolite multidimensional data set, two unsupervised statistical methods were used: PCA (Lindon *et al.*, 2001) and the self-organizing map (SOM) algorithm (Kohonen, 2001) that is a particular application of ANN. PCA was performed, on mean-centered data scaled to unit variance, using SAS software version 8.01 (SAS Institute, 1990). To apply the SOM algorithm, the Matlab software, 1.6.1 version was used with a program file written by the authors (Giraudel and Lek, 2001). A SOM package for Matlab is available at http://www.cis.hut.fi/projects/somtoolbox/. All calculations were performed using a computer equipped with an Intel Pentium® III-2GHz processor.

SOM analysis was performed on mean-centered data scaled to unit variance. The Kohonen network consists in two layers of neurons: the first layer (input layer) is connected to each vector of the data set (i.e. one vector is a real tomato flesh or seed sample with 44 standardized metabolites) and the second layer (output layer), which is the Kohonen map, forms a two-dimensional array of neurons arranged on a hexagonal lattice (figure 1). For this purpose, in each hexagon, a reference vector is considered. The reference vectors correspond to virtual
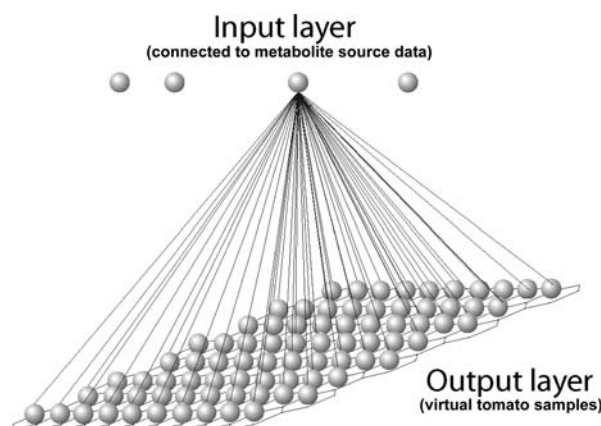

Figure 1. Representation of the unsupervised ANN approach with the Kohonen self-organizing map, showing the input neurons (observed metabolite data) and the output neurons (virtual tomato tissues) organized on a rectangular two-dimensional lattice.

tomato samples with metabolites to be computed. In the output layer, the units of the grid (virtual tomato samples) give a representation of the distribution of the tomato samples in an ordered way. The modifications of the virtual tomato samples are achieved through an ANN and computed during a training phase by iterative adjustments. For learning, only input units are used, no expected-output data is given to the system: this is referred to as an unsupervised learning. In this work, the parameters were chosen according to Kohonen's advice (Kohonen, 2001, chapter 3) as follows: flat rectangular maps were selected; the neighborhood kernel was described in terms of a Gaussian function; the number of iterations was 500 times the number of the map units with 2000 iterations for the ordering phase of the learning. Maps of different sizes were constructed to modulate their resolution capability. Then each map was read considering the relative positions of the samples and their comparison. Moreover, the chemical composition of each virtual tomato sample was used to display the distribution of each metabolite on the organized map on which real samples were plotted. This two-dimensional representation can be considered as a "component sliced" version of the SOM. A gray shade gradient was used to represent metabolite concentrations.

3. Results

3.1. Fruit growth and development

Under the growth conditions used in this study, the fruit was ripe at 45 DPA. Fruit diameter (figure 2a) increased rapidly from anthesis to 15 DPA and leveled off at mature green stage (35 DPA). Increase in fruit fresh weight (figure 2b) paralleled increase in diameter corresponding to increases of pericarp and columella thickness (figure 2c). The weight of the entire fruit reached 40 g FW at 45 DPA. Seed fresh weight
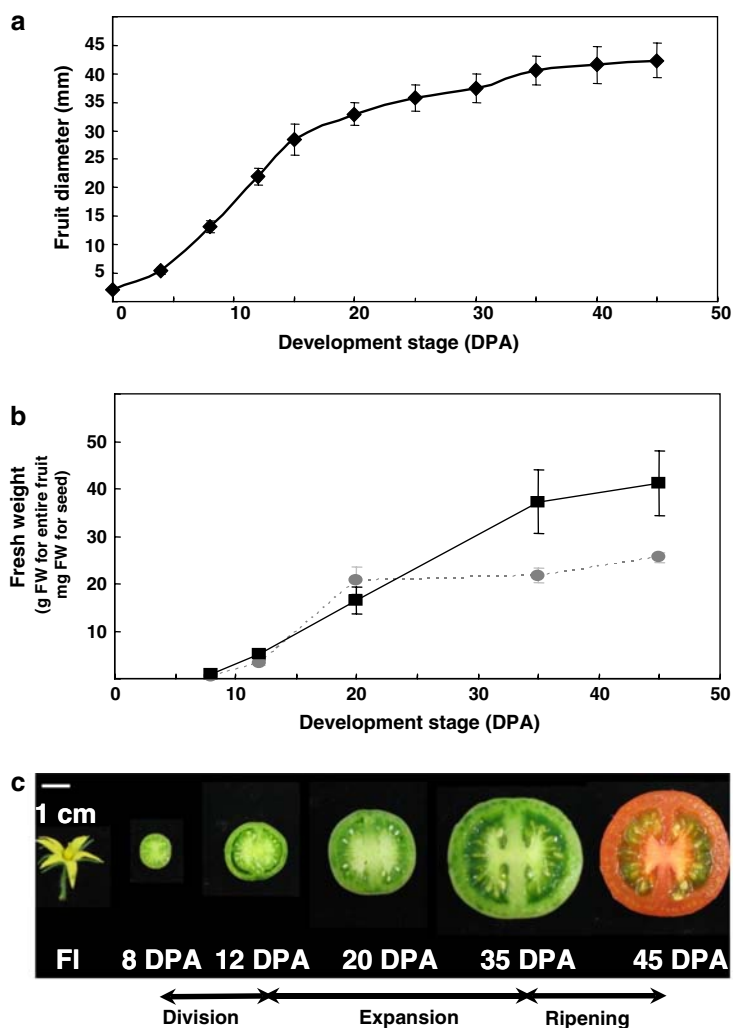
Figure 2. Growth and development of Ailsa Craig tomato fruits cultivated in a growth chamber. (a) Changes in fruit diameter. Mean of 12–36 fruits. Vertical bars represent standard deviations. (b) Changes in fruit fresh weight and seed fresh weight. Mean of 12–36 fruits. Vertical bars represent standard deviations. Black square, entire fruit; gray circle, seed. (c) Stages of fruit development from flower (Fl) to mature fruit (45 DPA) separated in 3 phases of development characterized by cell division, cell expansion and fruit ripening.

(figure 2b) increased rapidly from 8 DPA to 20 DPA, remained stable between 20 DPA and 35 DPA and then reached 25 mg FW per seed at 45 DPA. Fruit and seed fresh weights showed parallel changes during fruit development, except between 20 DPA and 35 DPA. During this period, the dry matter content of flesh increased only slightly from 6.5 ± 0.6 to 8.4 ± 0.9 % FW (mean $\pm$ SD of 3 replicates) whereas the dry matter content of seeds showed a strong increase from 8.1 ± 1.2 to 20.8 ± 0.7% FW (mean $\pm$ SD of 3 replicates). Therefore, between 20 DPA and 35 DPA, the seed water content decreased.

3.2. Identification of tomato fruit metabolites

For the untargeted NMR analyses, the major metabolites of each extract were identified after peak assignment using ^{1}H-NMR spectra from pure compounds associated with comparison of published data (Fan, 1996; Le Gall et al., 2003; Choi et al., 2004; Moing et al., 2004). Table 1 shows the resonances used for identification and absolute quantification of the metabolites. Among the 33 identified compounds, 25 were quantified. One compound with a 1D spectral pattern (one doublet at 5.00 ppm, 3.86 Hz, one doublet at 5.44 ppm, 3.82 Hz) similar to that of raffinose (O-α-D-galactopyranosyl-(1 → 6)-O-α-D-glucopyranosyl-(1 → 2)-O-β-D-fructofuranoside) was classified as a trisaccharide composed of glucose, fructose and galactose units. This compound was tentatively identified as planteose (O-α-D-galactopyranosyl-(1 → 6)-β-D-fructofuranosyl-(2 → 1)-α-D-glucopyranoside), a major oligosaccharide in tomato seeds (Downie et al. 2003) and referred to herein as planteose-like. Additionally, five unknown compounds were quantified in arbitrary units. Unknown compounds were named using the mid value of the chemical shift and the multiplicity of the corresponding resonance group (i.e. unknownD5.1 for a

Table 1
Compounds identified in the 1D ^{1}H NMR spectra of tomato fruit flesh or seed polar extracts

Compound[a]	Group	Multiplicity	δ[1]H (ppm)
Acetylcholine[b]	N-(C5H3)3	s	3.22
Alanine	C3H3	d	1.48
Asparagine	C3H2	m	2.92
Aspartate	½ (C3H2)	m	2.81
Chlorogenate	Caffeoyl-C7H	d	7.68
Chlorogenate derivative[b]	Caffeoyl-CxH	d	7.55
Choline	N-(C3H3)	s	3.20
Citrate	C2H2 + C4H2	dd	2.63
Formate	C1H	s	8.47
Fructose	α (C3H + C5H) + βC5H	m	4.12
Fumarate	C2H + C3H	s	6.53
GABA	C4H2	t	3.01
Galactose	βC1H	d	4.60
Glucose	αC1H	d	5.25
	βC1H	d	4.66
Glutamate	C3H2	m	2.07
Glutamine	C4H2	m	2.45
Isoleucine	C6H3	d	1.01
Lactate[b]	C3H3	d	1.33
Leucine	C5H3 + C6H3	t	0.96
Malate	C2H	dd	4.30
Mannose	C1H	d	5.2
Phenylalanine	C5H + C6H + C7H	m	7.40
Proline	C4H2	m	2.04
Pyroglutamate	C2H	dd	4.18
Planteose-like compound	Putative αC1H	d	5.00
UnknownD5.1	Sugar αC1H	d	5.10
Sucrose	Glucopyranosyl–C1H	d	5.41
Threonine	C4H3	d	1.33
Trigonelline	C3H	s	9.13
Tyrosine	C6H2	d	6.91
UDPGlucose	C2H′ ribose	m	5.61
Valine	C4H3	d	1.00
	C5H3	d	1.04
Xylose	βC1H	d	4.59

[1]H chemical shifts used for metabolite identification and quantification were determined at pH 6 in D_2O and expressed as relative values to that of TSP at 0 ppm. Groups are indicated according to Fan (1996) and Choi et al. (2004). d, doublet; dd, doublet of doublets; m, multiplet; s, singlet; t, triplet.
[a] All compounds were quantified unless indicated otherwise.
[b] Detected but not quantified.

doublet at 5.10 ppm). UnknownD5.1 belongs to the sugar family and was therefore named unkSugarD5.1. UnknownD6.2 may be an adenosine-containing compound since part of its 1D spectral pattern was similar to those of adenosine, adenosine diphosphate and adenosine triphosphate recorded under the same conditions of buffer and NMR acquisition. The coupling constant of the 6.2 ppm doublet of unknownD6.2 was 5.65 Hz while that of C1′H of the ribose of adenosine diphosphate was 5.63 Hz. The two singlets of the C2H and C8H rings expected at 8.54 (UnkS8.5) and 8.37 ppm, were present with the expected intensity ratio with the 6.2 ppm doublet. Moreover, in a COSY spectrum of a representative extract of seeds, the

resonance of the 6.2 ppm doublet showed correlation with resonances at 4.79 ppm as expected between the C1H′ and C2H′ of ribose. However, in some extracts, the intensity ratio between the 6.2 ppm doublet and the 8.54 ppm singlet was biased by resonance overlapping with another compound. Therefore unkD8.5 was considered as a mixture between an adenosine-containing compound and other compound(s). The other unknown compounds could not be integrated into a metabolite family.

Figure 3 shows the typical ^{1}H-NMR spectra obtained at 500 MHz and annotated following table 1, for flesh (a) and seed polar extracts (b) at 20 DPA. From visual inspection of these spectra, it is clear that there is a significant variation between the flesh and seed for many spectral domains, especially the phenolic region from 6 ppm to 9 ppm and the sugar region around residual water (4.5–5.7 ppm).

Isoprenoid analysis by HPLC allowed identification and quantification of 11 isoprenoids (table 2). Some of these compounds were detected only at 35 and 45 DPA, such as lycopene, phytofluene, phytoene and α and γ - tocopherol whereas ubiquinone 9 was only detected at 35 DPA (data not shown).

FAMEs analysis allowed identification and quantification of eight fatty acids (table 2), among them six (palmitic, stearic, oleic, linoleic, linolenic) are ubiquitous and three (behenic, arachidic, lignoceric) are very long chain fatty acids (VLCFAs). The global concentration range of the different fatty acids was of about three orders of magnitude and linoleic, oleic and palmitic acids presented the highest concentrations.

In an attempt to identify unknown metabolites detected through ^{1}H-NMR experiments, correlation analyses between these metabolites and all other identified and quantified polar metabolites were carried out. In order to obtain robust information, a significance level of 0.001 was chosen for both flesh and seed samples considered separately. With these criteria UnkS5.55 was highly correlated with starch ($r = 0.79$ in flesh and 0.80 in seeds). UnkD6.2 was highly correlated with glutamate ($r = 0.94$ in flesh and 0.91 in seeds). UnkS8.5 was highly correlated with aspartate ($r = -0.90$ in flesh and 0.87 in seeds), γ-aminobutyric acid (GABA, $r = -0.82$ in flesh and 0.90 in seeds), valine ($r = -0.80$ in flesh and 0.92 in seeds). The highly significant correlation between UnkS8.5 and UnkD6.2 ($r = 0.94$ for all flesh and seed samples) confirmed the assignment of UnkD6.2 as an adenosine containing compound as reported above.

3.3. Characterization of tomato fruit composition

Quantification values for 50 compounds (45 known compounds and the five unknown ones) are indicated for 8 and 45 DPA in table 2, illustrating huge changes both between these two developmental stages and between flesh and seeds. The total amount of all
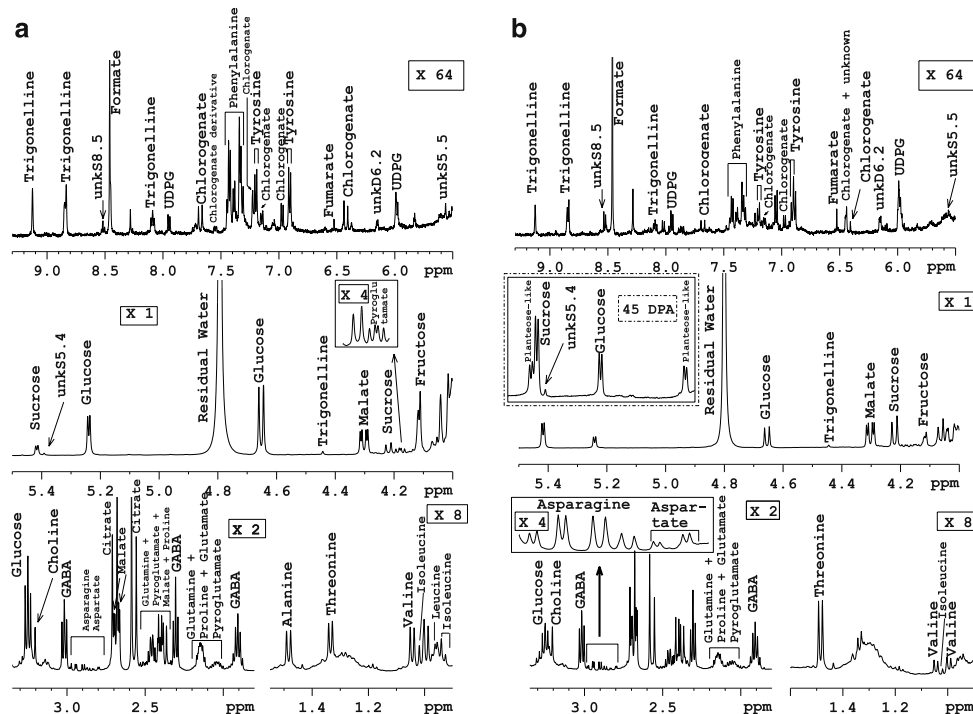
Figure 3. Representative 1-D ^{1}H NMR spectra of polar extracts from tomato. (a) Flesh i.e. fruit without seeds at 20 DPA. (b) Seeds at 20 DPA, except the dashed frame showing planteose-like compound at 45 DPA. Resonances are annotated according to Table 1.

Table 2
Concentration (μg/g DW) of 50 individual metabolites quantified from ^{1}H-NMR, chromatography and enzymatic analyses in tomato flesh and seeds at 8 and 45 DPA

Compound	Flesh		Seed	
	8 DPA	45 DPA	8 DPA	45 DPA
Sugars				
Fructose[N]	41,012 ± 2,881	111,884 ± 3,852	15,057 ± 1,061	10,856 ± 547
Glucose[N]	60,251 ± 2,694	93,354 ± 8,206	38,369 ± 2,589	6,727 ± 775
Mannose[N]	40 ± 1	513 ± 35	46 ± 6	64 ± 8
Planteose-like[N]	1,280 ± 305	817 ± 119	745 ± 129	4,028 ± 405
Starch[E]	100,515 ± 4,251	4,583 ± 300	70,343 ± 3,810	1,319 ± 216
Sucrose[N]	18,727 ± 2,023	17,016 ± 867	26,454 ± 1,334	11,970 ± 935
UDP-glucose[N]	401 ± 14	876 ± 7	502 ± 26	0.1 ± 0
Unknown sugar UnksugarD5.1[N]	98 ± 22	189 ± 6	33 ± 1	22 ± 1
Sum of sugars	222,324 ± 4,881	229,232 ± 11,083	151,549 ± 8,460	34,987 ± 853
Amino acids				
Alanine[N]	803 ± 77	911 ± 197	449 ± 5	91 ± 10
Asparagine[N]	1,927 ± 305	2,272 ± 371	1,400 ± 138	426 ± 18
Aspartic acid[N]	956 ± 128	3,280 ± 304	990 ± 81	414 ± 13
GABA[N]	13,507 ± 884	6,304 ± 946	30,956 ± 1,622	3,324 ± 345
Glutamic acid[N]	16,985 ± 2,007	33,108 ± 5,507	14,350 ± 520	7,438 ± 781
Glutamine[N]	10,203 ± 1,091	13,381 ± 566	3,782 ± 80	672 ± 104
Isoleucine[N]	268 ± 20	309 ± 26	72 ± 6	0.1 ± 0.0
Leucine[N]	793 ± 66	804 ± 48	351 ± 13	0.1 ± 0.0
Phenylalanine[N]	773 ± 8	1,464 ± 208	527 ± 61	151 ± 21
Pyroglutamic acid[N]	2,506 ± 211	2,639 ± 749	1,622 ± 8	255 ± 15
Threonine[N]	241 ± 47	371 ± 73	77 ± 20	0.1 ± 0.0
Tyrosine[N]	261 ± 52	308 ± 35	204 ± 17	0.1 ± 0.0
Valine[N]	424 ± 40	158 ± 9	267 ± 15	20 ± 4
Sum of amino acids	49,648 ± 3,578	65,308 ± 8,383	55,048 ± 1,967	12,792 ± 951

Table 2
Continued

Compound	Flesh		Seed	
	8 DPA	45 DPA	8 DPA	45 DPA
Organic acids				
Citric acid[N]	33,634 ± 1,467	69,490 ± 3,382	30,647 ± 2,491	24,213 ± 1,717
Malic acid[N]	14,353 ± 899	18,416 ± 1,339	19,438 ± 1,546	5,352 ± 201
Fumaric acid[N]	19 ± 5	19 ± 2	67 ± 7	16 ± 2
Sum of organic acids	48,006 ± 952	87,925 ± 4,084	50,152 ± 3,987	29,581 ± 1,918
Secondary compounds or unknowns				
Chlorogenic acid[N]	2,784 ± 258	629 ± 113	6,611 ± 581	224 ± 54
Choline[N]	851 ± 28	663 ± 39	1,104 ± 50	304 ± 13
Trigonelline	504 ± 56	253 ± 2	258 ± 22	38 ± 3
unkS5.4[N]	219 ± 44	76 ± 9	97 ± 15	212 ± 32
unkS5.55[N]	27 ± 2	182 ± 13	58 ± 4	31 ± 6
unkD6.2[N]	147 ± 22	511 ± 16	176 ± 16	133 ± 18
unkS8.5[N]	89 ± 11	362 ± 12	115 ± 9	0.1 ± 0.0
Sum of secondary compounds and unknowns (without isoprenoids)	4,621 ± 306	2,675 ± 164	8,419 ± 674	942 ± 28
Isoprenoids				
Carotene[L]	32 ± 4	47 ± 3	37 ± 4	13 ± 1
Chlorophyll a[L]	960 ± 16	13 ± 1	970 ± 40	38 ± 1
Chlorophyll b[L]	243 ± 5	2 ± 0.2	288 ± 27	10 ± 0.4
Lycopene[L]	ND	248 ± 30	ND	5 ± 0.3
Lutein[L]	58 ± 6	25 ± 1	64 ± 13	6 ± 0.4
Phytoene[L]	ND	130 ± 13	ND	9 ± 1
Phytofluene[L]	ND	1608 ± 169	ND	ND
α-Tocopherol[L]	ND	61 ± 13	ND	19 ± 3
γ-Tocopherol[L]	ND	23 ± 1	ND	158 ± 20
Ubiquinone9[L]	ND	ND	ND	ND
Xanthophyll[L]	8 ± 2	0.6 ± 0.1	18 ± 3	0.4 ± 0.1
Sum of isoprenoids	1,301 ± 27	2,156 ± 217	1,377 ± 28	258 ± 19
Fatty acids				
Arachidic ac. (C20:0)[G]	503 ± 77	213 ± 132	1,250 ± 243	755 ± 67
Behenic ac. (C22:0)[G]	207 ± 71	83 ± 6	414 ± 37	240 ± 17
Linoleic ac. C18:2)[G]	14,193 ± 204	3,359 ± 59	20,642 ± 880	93,255 ± 11,195
Linolenic ac (C18:3)[G]	5,488 ± 111	2,088 ± 119	7,296 ± 415	4,035 ± 431
Lignoceric ac. (C24:0)[G]	229 ± 48	63 ± 4	421 ± 16	310 ± 26
Oleic ac. (C18:1)[G]	1,983 ± 47	391 ± 23	2,023 ± 120	30,531 ± 3,978
Palmitic ac. (C16:0)[G]	7,139 ± 201	2,512 ± 60	9,438 ± 509	23,713 ± 2,774
Stearic ac. (C18:0)[G]	1,212 ± 36	570 ± 71	1,209 ± 30	7,620 ± 1,010
Sum of FAMEs	30,950 ± 550	9,280 ± 410	42,690 ± 1,560	160,460 ± 19,360

The sum of all quantified metabolites of a compound family was calculated for sugars, amino acids, organic acids, secondary metabolites without isoprenoids plus unknown compounds, isoprenoids and FAMEs. Mean ± SD of 3 replicates. ND, not detected.

Note: [N], ^{1}H-NMR; [G], GC-FID; [L], LC-DAD; [E], enzymatic analysis.

quantified metabolites represented about 36% and 40% of flesh dry weight (DW) and about 31% and 24% of seed DW at 8 and 45 DPA, respectively (data not shown). The contribution of each family of metabolites greatly depended on the tissue and stage of development. The total amount of quantified sugars (fructose, glucose, mannose, planteose-like, sucrose, UDP-glucose, unkSugarD5.1 and starch) ranged from 15.1% DW in seeds at 8 DPA to 22.9% DW in flesh at 45 DPA. The total amount of quantified organic acids (citrate, malate and fumarate) ranged from 3.0% DW in seeds at 45 DPA to 8.8% DW in flesh at 45 DPA. The total amount of quantified amino acids (alanine, asparagine, aspartate, GABA, glutamate, glutamine, isoleucine, leucine, phenylalanine, pyroglutamate, threonine, tyrosine, valine)

varied from 1.3% DW in seeds at 45 DPA to 6.5% DW in flesh at 45 DPA. The total amount of quantified isoprenoids (carotene, chlorophylls a and b, lycopene, lutein, phytoene, α and γ tocopherols, ubiquinone 9, xanthophylls) ranged from 0.02% DW in seeds at 45 DPA to 0.3% DW in flesh at 45 DPA. The total amount of quantified fatty acids (arachidic, behenic, linoleic, linolenic, lignoceric, oleic, palmitic and stearic acids) ranged from 0.09% DW in flesh at 45 DPA to 16.0% DW in seeds at 45 DPA. Whatever the tissue and development stage, linoleic acid was always the major fatty acid. After this, in flesh and seeds, palmitic and linolenic acids were the main fatty acids at 8 DPA. At 45 DPA, the major fatty acids did not change for flesh but in seeds the picture was modified since palmitic and

oleic acids were the most abundant ones after linoleic acid.

After this overview of quantitative results at 8 and 45 DPA, multivariate analyses (PCA and SOM) were used in order to visualize and analyze the metabolite data at all developmental stages of fruit flesh and seeds.

3.4. Metabolic trajectories during development

Among the 50 quantified metabolites, 44 metabolites were detected in more than two stages of development in the flesh or seed samples, and selected for multivariate analyses. They correspond to 7 known and one unknown sugars, 13 amino acids, 3 organic acids, 8 isoprenoids, 8 fatty acids and 4 unknown compounds. The matrix containing the data of these 44 metabolites in the 30 flesh or seed samples was explored using Principal Component Analysis (PCA). PCA allowed an overall view of the differences between flesh and seed samples and between stages, and revealed discriminatory metabolites. The first four principal components explained 87% of total variability. The first two PCA scores explained 67% of total variability (figure 4a). The biological replicates were clustered. The PCA analysis clearly identified the flesh and seed samples as different for each stage of development although they followed some parallel trajectories during development. The first principal component (PC1), explaining 45% of total variability, clearly separated the seed from flesh samples. Examination of PC1 loadings (figure 4b) suggested that the difference between the seed and flesh samples involved glucose, fructose, and several amino acids (isoleucine, phenylalanine, threonine, glutamine and tyrosine) on the positive side, and fatty acids and the planteose-like compound on the negative side. The second principal component (PC2), explaining 22% of total variability, separated early from late stages of development. Examination of PC2 loadings (figure 4B) suggested that this difference between stages involved isoprenoids (lutein, chlorophyll a, xanthophyll), γ-amino-butyric acid, chlorogenate, fumarate, sucrose, starch, behenic, linolenic and lignoceric acids on the positive side, and planteose-like compound on the negative side.

3.5. Similarities between flesh and seeds and between stages of development

The matrix containing the data of all 44 metabolites in flesh and seeds (30 samples) was also explored by the SOM algorithm. A first map with nine units (figure 5a) was obtained, which showed a clear separation between flesh samples on one side, and seed samples on the other side. This map evidenced also the similarity between most successive stages of development for flesh (12 and 20 DPA, 35 and 45 DPA) and seeds (8 and 12 DPA, 35 and 45 DPA), and revealed metabolic trajectories following the stages of development for each tissue.
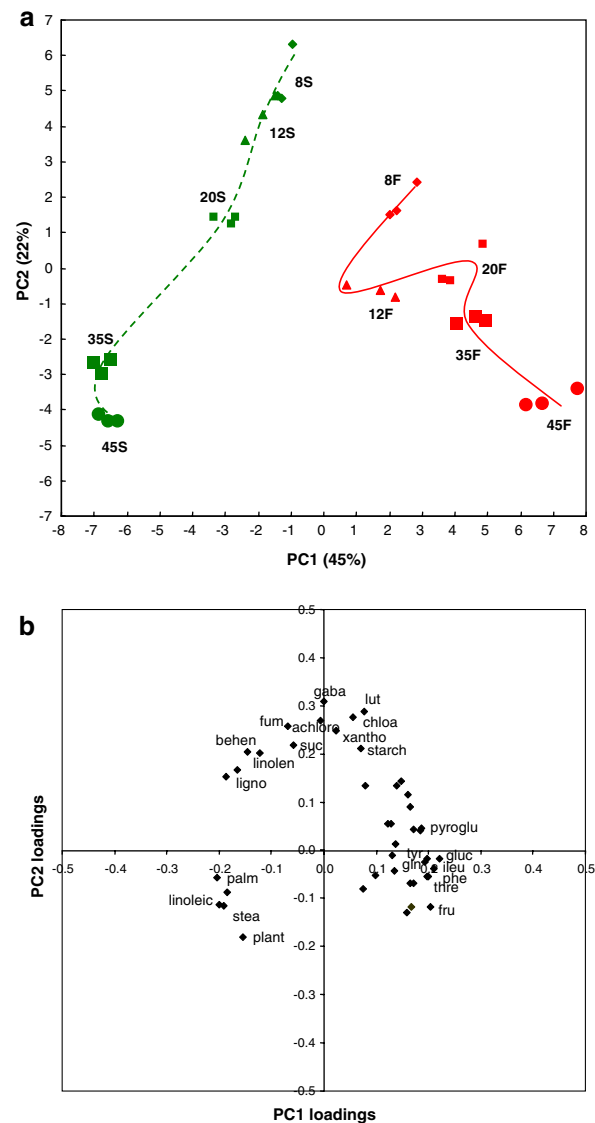

Figure 4. Principal component analysis (PCA) of absolute concentration of 44 metabolites issued from ^{1}H-NMR, GC-FID and LC-DAD analysis of tomato flesh (F) and seeds (S) at five stages of fruit development. (a) PCA scores plot. (b) PCA loadings plot. For each principal component, the 12 loadings with higher absolute value are indexed with the corresponding metabolite name. achloro, chlorogenic acid; behen, behenic acid; chloa, chlorophyll a; fruc, fructose; fum, fumaric acid; gaba, γ-aminobutyric acid; gln, glutamine; gluc, glucose; ileu, isoleucine; linoleic, linoleic acid; linolen, linolenic acid; ligno, lignoceric acid; lut, lutein; palm, palmitic acid; phe, phenylalanine; pyroglu, pyroglutamate; plant, planteose-like compound; stea, stearic acid; suc, sucrose; thre, threonin; tyr, tyrosine; xantho, xanthophyll.

However, flesh samples from 8 DPA were not mapped close to those of 12 DPA. Larger maps (15 units figure 5b and 30 units figure 5c) were constructed in order to increase the resolution capability. In the 30-unit map (figure 5c), a full discrimination of all the sample types was observed with a perfect clustering of the replicates. The 15-unit map (figure 5b) clearly revealed similarities and discriminations between the different types of samples. The fact that 8 DPA flesh and seed samples were mapped close to each other, and that 35 and
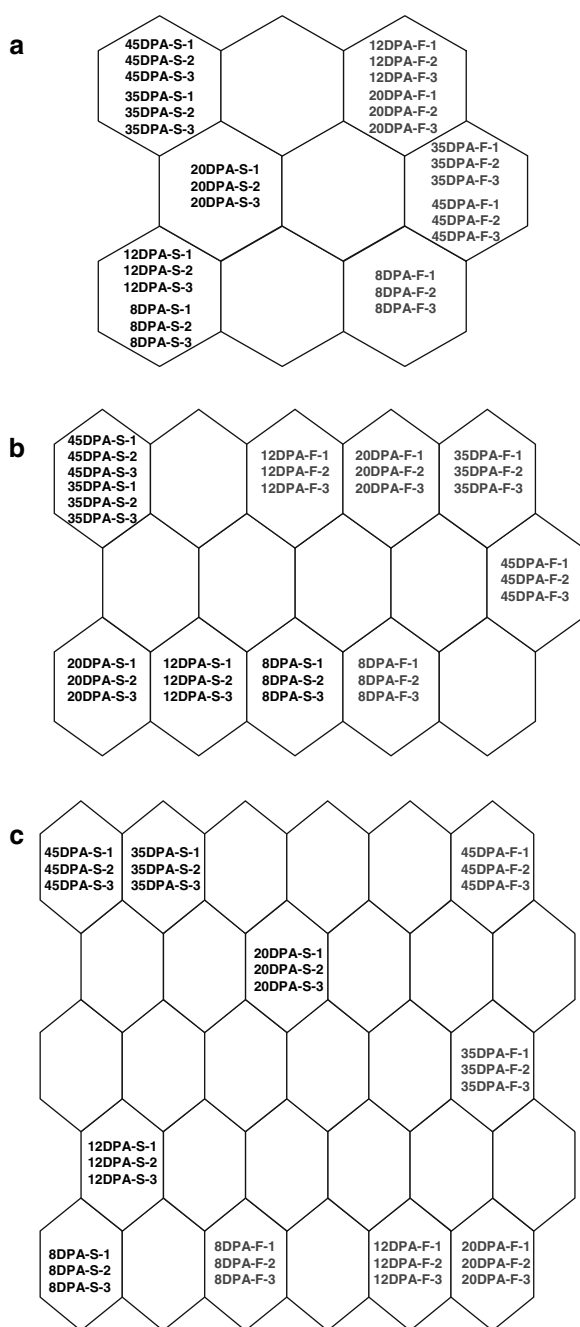
Figure 5. Distribution of flesh and seed samples at five stages of development onto a 9-unit (a), 15-unit (b) and 30-unit (c) SOM for absolute concentration of 44 metabolites issued from ^{1}H-NMR, GC-FID and LC-DAD analysis. The samples are coded with the number of DPA, F for Flesh and S for Seeds and the last number of the code refers to the replicate number.

45 DPA seed samples remained in the same unit, suggested metabolic similarities between the corresponding tissues or stages.

The concentrations of the 44 metabolites of each virtual tomato sample of the 15-unit map were used to display the concentration of each metabolite on this map. This allowed the drawing of 44 component plane representations corresponding to 44 metabolites, among

which the 39 identified metabolites are presented in figure 6a–e. By observing these maps in comparison with figure 6f, a simplified version of figure 5b, changes in concentration appearing for some metabolites could possibly be related to the stage or tissue and visually reveal discriminant metabolites. Indeed, high concentrations in starch (figure 6a), chlorogenate, choline and lutein (figure 6d) were a common feature for flesh and seeds at 8 DPA. Flesh samples were characterized by high concentration in fructose, glucose, mannose, UDP-glucose (figure 6a), aspartate, citrate and malate (figures 6b and c) at 35–45 DPA. Seed samples were characterized by high concentrations in chlorogenate and GABA (figures 6d and b) at 8–12 DPA, in fumarate (figure 6c) from 8 DPA to 20 DPA, in sucrose (figure 6a) and linolenic acid (figure 6e) at 12–20 DPA, in planteose-like compound (figure 6a) and linoleic acid (figure 6e) at 35–45 DPA. These tendencies directly derived from the virtual samples constituting the map units were validated using univariate analyses as shown below.

Examination of the concentration changes during development for both tissues (figures 7a–e for identified metabolites and data not shown for unknown compounds) confirmed the tendencies highlighted by the SOM analysis. The major differences between flesh and seeds and pattern similarities are summarized on a simplified metabolic map (figure 8). Concentrations were significantly higher in flesh than in seeds for fructose, glucose, starch (figure 7a), unknownsugarD5.1, unknownS5.4 (data not shown), glutamine, isoleucine, leucine, phenylalanine, threonine, tyrosine, valine (figure 7b), and trigonelline (figure 7d) for most stages of development. Concentrations were significantly higher in seeds than in flesh for sucrose (figure 7a), fumarate (figure 7c) and the seven fatty acids (figure 7e) for most stages of development. For some metabolites, the changes during development were opposite in flesh and seeds. For glucose, sucrose (figure 7a), aspartate (figure 7b), unknownS5.55 (data not shown), UDP-glucose (figure 7a) concentration increased in flesh during several successive stages of development while it decreased in seeds. For leucine (figure 7b) and unknownS5.4 concentration changes during development showed symmetric trends for flesh and seeds. Although absolute concentrations were different, some metabolites followed nearly parallel changes during development in flesh and seeds [framed metabolites in figure 8: starch (figure 7a), glutamate, phenylalanine and valine (figure 7b), citrate (figure 7c), chlorogenate and lutein (figure 7d), arachidic, behenic and lignoceric acids (figure 7e)].

4. Discussion

4.1. New output of SOM analysis

A global analytical and chemometrics approach was developed and used to compare tomato flesh and seed
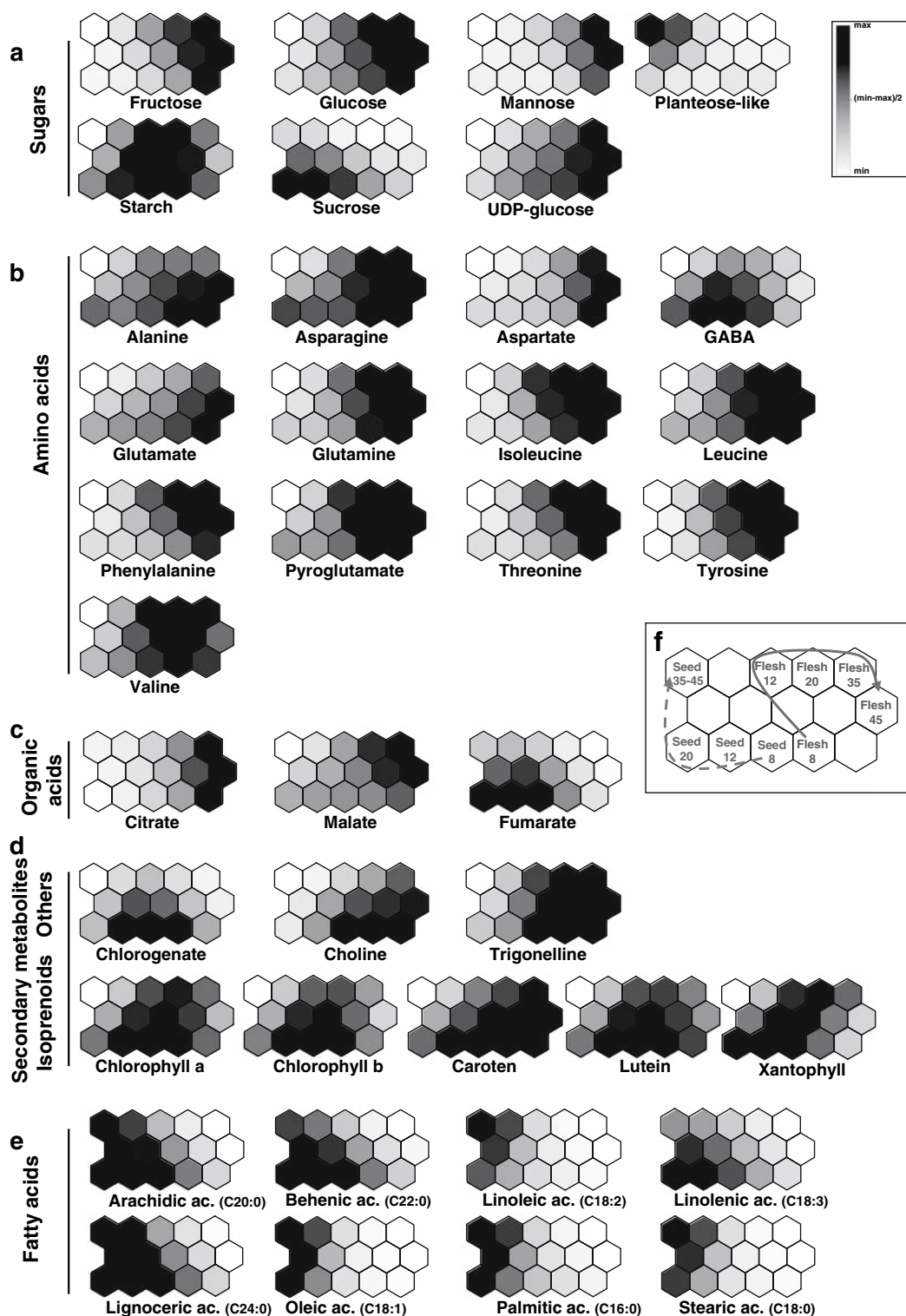
Figure 6. Distribution of 39 identified and quantified metabolites in the 15-unit SOM presented figure 5b. (a) Sugars (fructose, glucose, mannose, planteose-like, starch, sucrose, UDP-glucose). (b) Amino acids (alanine, asparagine, aspartate, GABA, glutamate, glutamine, isoleucine, leucine, phenylalanine, pyroglutamate, threonine, tyrosine, valine). (c) Organic acids (citric, malic, fumaric). (d) Secondary metabolites (chlorogenate, choline, trigonelline, chlorophyll a, chlorophyll b, carotene, lutein, xanthophylls). (e) Fatty acids (arachidic, behenic, linoleic, linolenic, lignoceric, oleic, palmitic, stearic acids). In order to simplify the figure, the concentration color scale is not indicated in μg/g DW for each metabolite but summarized for the whole figure using minimum and maximum values. (f) Simplified version of the 15-unit SOM from figure 5b showing the position of sample groups. Arrows indicate trajectories during development for flesh (solid line) and seeds (dashed line).
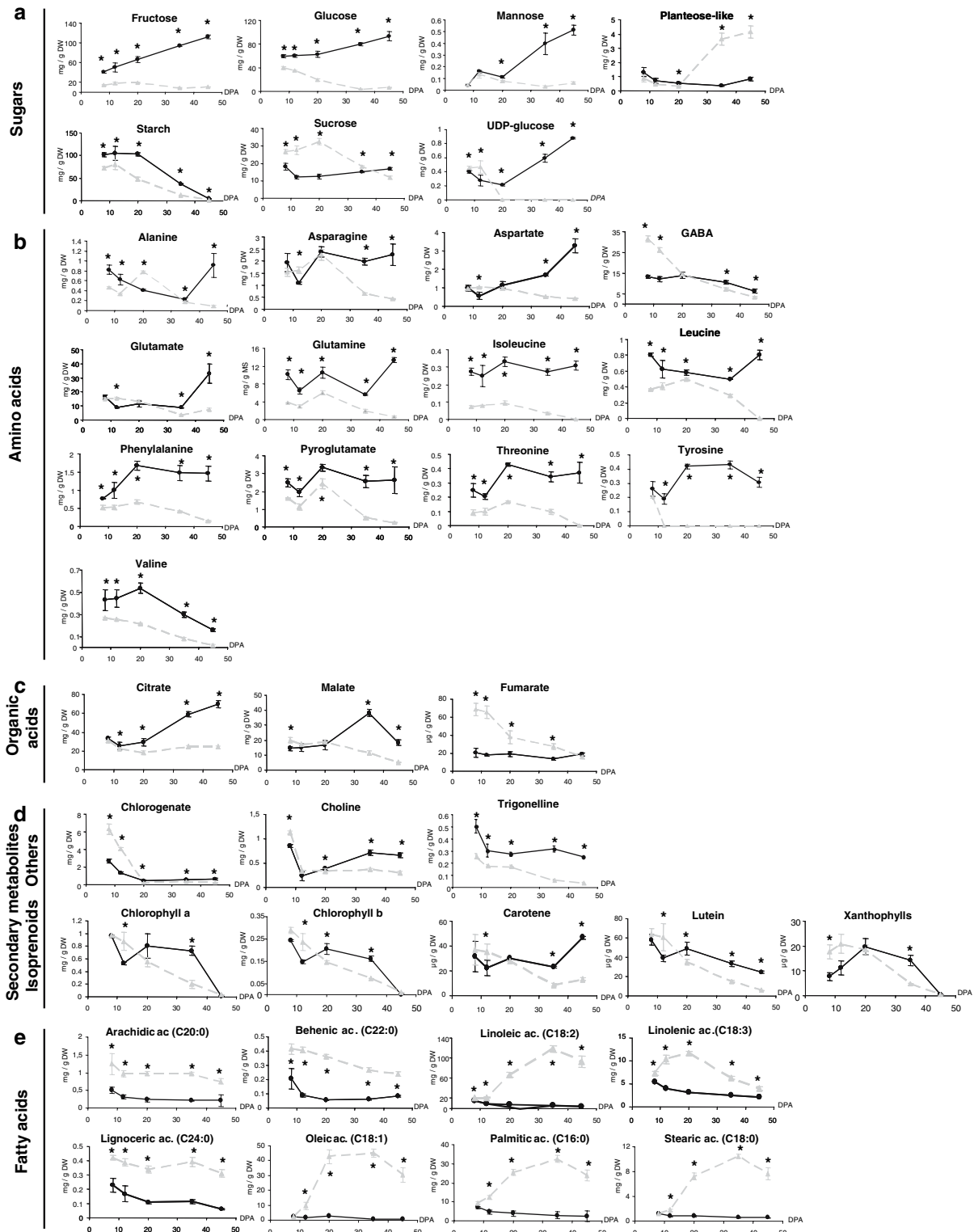
Figure 7. Changes during fruit development for 39 metabolites. (a) Soluble sugars (fructose, glucose, mannose, planteose-like, starch, sucrose, UDP-glucose). (b) Amino acids (alanine, asparagine, aspartate, GABA, glutamate, glutamine, isoleucine, leucine, phenylalanine, pyroglutamate, threonine, tyrosine, valine). (c) Organic acids (citric, malic, fumaric). (d) Secondary metabolites (chlorogenate, choline, trigonelline, chlorophyll a, chlorophyll b, carotene, lutein, xantophylls). (e) Fatty acids (arachidic, behenic, linoleic, linolenic, lignoceric, oleic, palmitic, stearic acids). For each stage of development, * indicates a significant difference between flesh and seeds according to Student's t test ($p < 0.05$).

metabolite concentrations during fruit development. The absolute concentrations of 44 metabolites were analyzed with PCA and SOM. SOM analysis of these

data allowed to combine multivariate (distribution of samples on Kohonen SOMs) and univariate information (component plane representation of each metabolite)
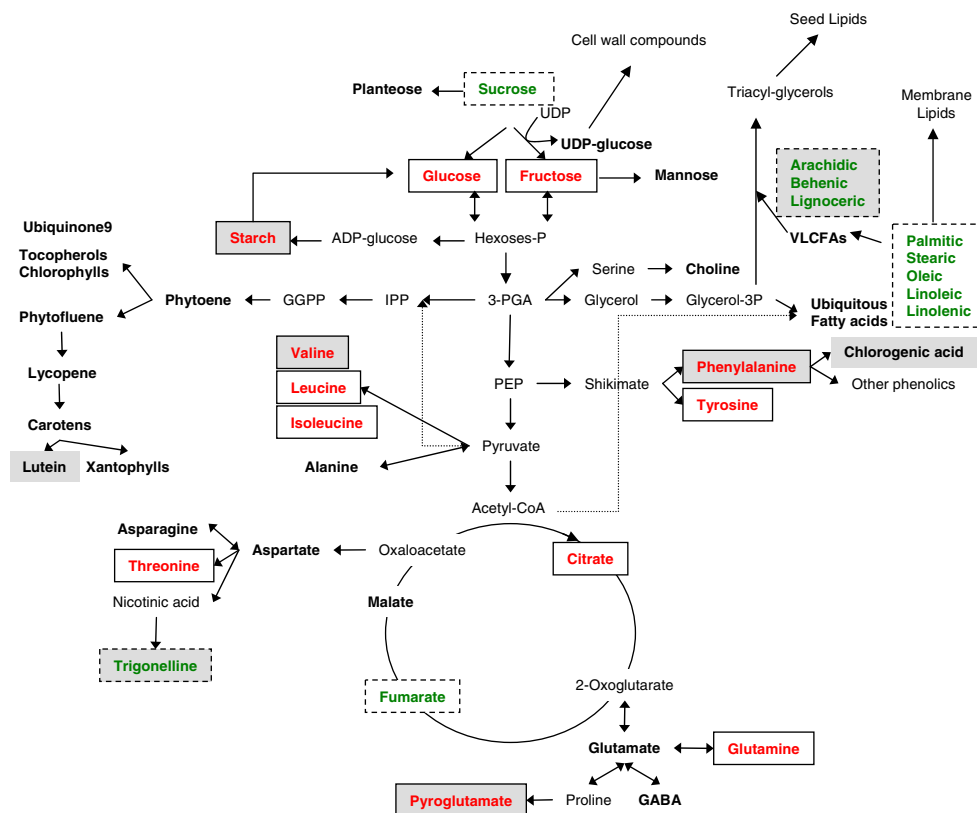
Figure 8. Schematic representation of metabolic pathways. Metabolites in bold were quantified in the present experiment. Metabolites framed with continuous line had a significantly higher concentration in flesh that in seeds for at least four stages of development. Metabolites framed with dotted line had a significantly higher concentration in seeds than in flesh for at least four stages of development. Grayed metabolites followed parallel changes in flesh and seeds during fruit development.

through an elegant way, in a single analysis. SOM and PCA analyses appeared complementary to visualize metabolic trajectories during development with SOM highlighting better punctual changes and PCA revealing better parallel trajectories during development. This strategy confirmed published data (Valle *et al.*, 1998; Rolin *et al.*, 2000; Obiadalla-Ali *et al.*, 2004; Carrari *et al.*, 2006) and brought new data on tomato flesh and seed composition, thus demonstrating its potential in metabolomics. PCA and SOM multivariate and uni-variate kinetic analyses of metabolite data revealed: (i) common features during the development of flesh and seed tissues especially at early stages and (ii) specific features for each tissue (figure 8).

4.2. Flesh and seed compositional similarities

Surprisingly, flesh and seed tissues that clearly differ in structure and function showed some compositional similarities at early stages of development. At 8 DPA, multivariate analyses showed that flesh and seed samples were mapped close to each other. For instance, concentrations in mannose, choline, oleic, stearic, linoleic, linolenic and palmitic acids were similar in flesh and seeds at this stage. Moreover, parallel trends in flesh and

seeds were observed for several metabolites during development. Thus, parallel decreases during early development were observed for starch, chlorogenate, trigonelline, arachidic, and lignoceric acids. This may reflect common features related to cell division, growth or differentiation in fleshy and seed tissues (Gillapsy *et al.*, 1993; Hilhorst *et al.*, 1998).

At early stages (8 DPA), the fatty acid compositions of seed and flesh were similar and typical of the lipid composition generally encountered in membranes from higher plants (Millar *et al.*, 2000). The fatty acids mainly synthesized were ubiquitous phospholipids components (with 16 and 18 carbon atoms) or specific VLCFAs for other membrane lipids such as sphingolipids and cera-mides. These results supported the fact that at these stages the fatty acids were largely involved in membrane synthesis. In the same way, at 8 DPA the high concentration of choline, a precursor of phosphatidylcholine [the most important phospholipid in plant membranes (Kent, 1995)], could reflect high membrane synthesis. So, we can assume that membrane fatty acids and choline patterns reflect membrane synthesis related to active cell divisions occurring in flesh and seeds. Besides, the highest concentrations of trigonelline (N-methyl nico-tinamide), a secondary metabolite that may play a role

in cell cycle regulation (Minorsky, 2002), were observed at early stages in flesh and seeds. More generally, the compositional similarities at 8 DPA may reflect common features of both tissues: low differentiation associated with active cell proliferation (Gillapsy *et al.*, 1993).

Starch was transitorily accumulated in flesh and seeds at early stages of development. In fruit, the high starch concentrations from 8 DPA to 20 DPA corresponded to transitory storage of imported carbon reutilized during the expansion phase for soluble sugar synthesis (Schaffer and Petreikov, 1997; Obiadalla-Ali *et al.*, 2004). In seeds, the high starch concentrations at 8 and 12 DPA corresponded to transitory storage of imported carbon reutilized later during a high rate of lipid synthesis or for the synthesis of sugar contributing to the acquisition of desiccation tolerance as described for Brassicaceae (da Silva *et al.*, 1997; Baud *et al.*, 2002). This early development of a capacity for starch synthesis and storage may be related to the establishment of a high sink strength in both tissues.

4.3. Flesh specificity

For fruit flesh, there is a general agreement between our data and previous data on polar extracts of entire fruit or pericarp polar compounds (Rolin *et al.*, 2000; Schauer *et al.*, 2005; Carrari *et al.*, 2006), starch (Yelle *et al.*, 1988), amino acids (Valle *et al.*, 1998) and carotenoids (Bino *et al.*, 2005) despite some differences possibly due to genetic and/or environmental effects. Compositional similarities between two successive stages of development, estimated from the visualization of metabolic trajectories using PCA or SOMs, showed that flesh samples from 8 DPA were not mapped close to those of 12 DPA. This indicates high metabolic changes between 8 DPA and 12 DPA in this tissue, which may reflect the developmental transition from cell division to cell expansion (Cheniclet *et al.*, 2005), in agreement with shifts in gene expression data (Lemaire-Chamley *et al.*, 2005).

During the cell expansion phase (12–35 DPA), flesh was characterized by the accumulation of soluble sugars and organic acids that contributed to the acquisition of the fleshy trait associated with fruit cell expansion. Cell enlargement mainly depends on the increase in turgor pressure, which is itself driven by osmolytes (mostly soluble sugars, organic acids and potassium) and water accumulation inside the vacuoles. During the ripening phase (35–45 DPA) the well-known increase in major pigments, lycopene and carotene characterized flesh and its color changes. In parallel, glucose, fructose and citrate concentrations, participating in the sugar/acid ratio, a major determinant of fruit taste, continued to increase. The high concentration in mannose [a constituent of cell-wall polymers (Greve and Labavitch, 1991)] and UDP-glucose [a precursor for cell-wall biosynthesis (Seifert, 2004)] in flesh may reflect changes in cell wall structure linked with fruit softening.

4.4. Seed specificity

Published data on tomato seed composition remain rather scarce in comparison with data on entire fruit or fruit pericarp. In the present experiment, concentration patterns in tomato seeds were in general agreement with that of cucumber seeds (Handley *et al.*, 1983) and Brassicaceae seeds (Baud *et al.*, 2002) for carbohydrates. Compositional similarities between two successive stages of development, estimated from the visualization of metabolic trajectories using PCA or SOMs, showed that 35 and 45 DPA seed samples remained in the same SOM unit, suggesting low metabolic changes in the seeds during this period of the fruit development. This is in agreement with the fact that the development of the embryo is finished at the beginning of fruit ripening (Hilhorst *et al.*, 1998).

Seeds were characterized by the accumulation of several metabolites, including γ-aminobutyric acid, fatty acids, sucrose and the planteose-like compound. At early stages, the high concentration of γ-aminobutyric acid observed in seeds may be related with a high flux of glutamine or glutamate import into the seed metabolized through glutamate decarboxylase, or reflects hypoxia (Shelp *et al.*, 1995) in relation with high metabolic activity and low oxygen diffusion. The fatty acid concentration was always high in seeds and strongly increased from 8 DPA to 45 DPA. At the earlier stage, lipids composition reflected their involvement in membranes as indicated above. After 12 DPA the high modification of FAMEs profiles probably reflects a specific accumulation of fatty acids in seeds as triacylglycerols (Voelker and Kinney, 2001) that constituted the energy storage for a further germination. This implies that lipid synthesis pathways changed in seeds during development. The decrease in VLCFA concentration may be related to a decrease in the activity of stearoyl-CoA elongation, with stearate being essentially desaturated successively into oleate, linoleate and linolenate. Finally, seeds at 35–45 DPA were characterized by high concentrations in linoleic and oleic acids. Linoleic acid represented more than 50% of the total fatty acids, and VLCFAs only 0.7%, in good agreement with the data previously reported (Camara *et al.*, 2001). High concentrations of the planteose-like compound characterized seeds at 35–45 DPA. The presence of galactooligosaccharides in seeds, is in agreement with data in several plants species (Bailly *et al.*, 2001; Karner *et al.*, 2004). In tomato seeds, where planteose is the major galactooligosaccharide, expression of a galactinol synthase gene correlated with dry weight deposition and desiccation tolerance. Therefore, planteose-like compound accumulation in tomato seeds may be related to the protection of cellular structures during desiccation

and to the constitution of carbon reserves for early germination (Downie *et al.*, 2003).

5. Concluding remarks

In conclusion, a global analytical and chemometrics approach was developed and used to compare compositional changes in tomato flesh and seeds from the same fruits of the same plants. It was validated since it allowed to point to known metabolic changes. In addition, it revealed similarities and differences between the two tissues during fruit development. A map of the different metabolite concentrations at different stages of development for two tissues (flesh and seed) of tomato fruit was established. This approach and this map will be useful for further analyses of genotype and environment effects on fruit or seed quality, and for the characterization of metabolic modifications induced in fruit of transformant plants for candidate genes in functional genomics. The map will be refined separating the different tissues constituting the flesh and those constituting the seeds and using parallel transcriptomics data for some selected stages and tissues.

Acknowledgments

References

Akoka, S., Barantin, L. and Trierweiler, M. (1999). Concentration measurement by proton NMR using the ERETIC method. *Anal. Chem.* **71**, 2554–2557.

Bailly, C., Audigier, C., Ladonne, F., Wagner, M.H., Coste, F., Corbineau, F. and Come, D. (2001). Changes in oligosaccharide content and antioxidant enzyme activities in developing bean seeds as related to acquisition of drying tolerance and seed quality. *J. Exp. Bot.* **52**, 701–708.

Baud, S., Boutin, J.P., Miquel, M., Lepiniec, L. and Rochat, C. (2002). An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. *Plant Physiol. Biochem.* **40**, 151–160.

Berry, T. and Bewley, J.D. (1992). A role for the surrounding fruit tissues in preventing the germination of tomato (*Lycopersicon esculentum*) seeds. *Plant Physiol.* **100**, 951–957.

Bino, R.J., De Vos, C.H.R., Lieberman, M., Hall, R.D., Bovy, A., Jonker, H.H., Tikunov, Y., Lommen, A., Moco, S. and Levin, I. (2005). The light-hyperresponsive high pigment-2(dg) mutation of tomato: alterations in the fruit metabolome. *New Phytol.* **166**, 427–438.

Boggio, S.B., Palatnik, J.F., Heldt, H.W. and Valle, E.M. (2000). Changes in amino acid composition and nitrogen metabolizing enzymes in ripening fruits of *Lycopersicon esculentum* Mill. *Plant Sci.* **159**, 125–133.

Browse, J., McCourt, P.J. and Somerville, C.R. (1986). Fatty acid composition of leaf lipids determined after combined digestion and fatty acid methyl ester formation from fresh tissue. *Anal. Biochem.* **152**, 141–145.

Burns, J., Fraser, P.D. and Bramley, P.M. (2003). Identification and quantification of carotenoids, tocopherols and chlorophylls in commonly consumed fruits and vegetables. *Phytochemistry* **62**, 939–947.

Camara, M., Del Valle, M., Torija, M.E. and Castilho, C. (2001). Fatty acid composition of tomato pomace. *Acta Hort.* **542**, 175–180.

Carrari, F., Baxter, C.J., Usadel, B., Urbanczyk-Wochniak, E., Zanor, M.I., Nunes-Nesi, A., Nikiforova, V., Centero, D., Ratzka, A., Pauly, M., Sweetlove, L.J. and Fernie, A.R. (2006). Integrated analysis of metabolites and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Physiol.* **142**, 1380–1396.

Cheniclet, C., Rong, W.Y., Causse, M., Frangne, N., Bolling, L., Carde, J.P. and Renaudin, J.P. (2005). Cell expansion and endoreduplication show a large genetic variability in pericarp and contribute strongly to tomato fruit growth. *Plant Physiol.* **139**, 1984–1994.

Choi, Y.H., Casas Tapias, E., Kim, H.K., Lefeber, A.W.M., Erkelens, C., Verhoeven, J.T.J., Brzin, J., Zel, J. and Verpoorte, R. (2004). Metabolic discrimination of *Catharanthus roseus* leaves infected by phytoplasma using ^{1}H-NMR spectroscopy and multivariate analysis. *Plant Physiol.* **135**, 1–13.

Da Silva, P., Eastmond, P.J., Hill, L.M., Smith, A.M. and Rawsthorne, S. (1997). Starch metabolism in developing embryos of oilseed rape. *Planta* **203**, 480–487.

Downie, B., Gurusinghe, S., Dahal, P., Thacker, R.R., Snyder, J.C., Nonogaki, H., Yim, K., Fukanaga, K., Alvarado, V. and Bradford, K.J. (2003). Expression of a galactinol synthase gene in tomato seeds is up-regulated before maturation desiccation and again after imbibition whenever radicle protrusion is prevented. *Plant Physiol.* **131**, 1347–1359.

Fan, T.W.M. (1996). Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. *Prog. Nucl. Magn. Reson. Spectrosc.* **28**, 161–219.

Fraser, P.D., Pinto, E.S., Holloway, D.E. and Bramley, P.M. (2000). Application of high-performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids. *Plant J.* **24**, 551–558.

Fraser, P.D., Truesdale, M., Bird, C.R., Schuch, W. and Bramley, P.M. (1994). Carotenoid biosynthesis during tomato fruit development. *Plant Physiol.* **105**, 405–413.

Gillapsy, G., Ben-David, H. and Gruissem, W. (1993). Fruits: a developmental perspective. *Plant Cell* **5**, 1439–1451.

Giovannoni, J.J. (2004). Genetic regulation of fruit development and ripening. *Plant Cell* **16**, S170–S180.

Giraudel, J.L. and Lek, S. (2001). A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. *Ecol. Modell.* **146**, 329–339.

Greve, L.C. and Labavitch, J.M. (1991). Cell wall metabolism in ripening fruit. V Analysis of cell wall synthesis in ripening tomato pericarp tissue using a D-[U-^{13}C] glucose tracer and gas chromatography-mass-spectrometry. *Plant Physiol.* **97**, 1456–1461.

Handley, L.W., Pharr, M. and McFeeters, F. (1983). Carbohydrate changes during maturation of cucumber seeds. *Plant Physiol.* **72**, 498–502.

Hilhorst, H.W.M., Groot, S.P.C. and Bino, R.J. (1998). The tomato seed as a model system to study seed development and germination. *Acta Bot. Neerl.* **47**, 169–183.

Karner, U., Peterbauer, T., Raboy, V., Jones, D.A., Hedley, C.L. and Richter, A. (2004). Myo-inositol and sucrose concentrations affect the accumulation of raffinose family oligosaccharides in seeds. *J. Exp. Bot.* **55**, 1981–1987.

Kent, C. (1995). Eukaryotic phospholipid biosynthesis. *Annu. Rev. Biochem.* **64**, 315–343.

Knoblich, M., Anderson, B. and Latshaw, D. (2005). Analyses of tomato peel and seed byproducts and their use as a source of carotenoids. *J. Sci. Food Agric.* **85**, 1166–1170.

Kohonen, T (2001). *Self Organizing Maps.* Springer, Berlin.

Kunst A., Draeger B., Ziegenhorn J. (1984). Carbohydrates, U.V. methods with hexokinase and glucose-6-phosphate dehydrogenase in Bergmeyer H.U. (ed) Methods in Enzymatic Analysis, (Third ed.), VI, Metabolites I. Verlag Chemie, Basel. pp. 163–172.

Le Gall, G., Colquhoun, I.J., Davis, A.L., Collins, G.J. and Verhoeyen, M.E. (2003). Metabolite profiling of tomato (*Lycopersicon esculentum*) using H-1 NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. *J. Agric. Food Chem.* **51**, 2447–2456.

Lemaire-Chamley, M., Petit, J., Garcia, V., Just, D., Baldet, P., Germain, V., Fagard, M., Mouassite, M., Cheniclet, C. and Rothan, C. (2005). Changes in transcriptional profiles are associated with early fruit tissue specialization in tomato. *Plant Physiol.* **139**, 750–769.

Lindon, J.C., Holmes, E. and Nicholson, J.K. (2001). Pattern recognition methods and applications in biomedical magnetic resonance. *Prog. Nucl. Magn. Reson. Spectrosc.* **39**, 1–40.

Mattoo, A.K., Sobolev, A.P., Neelam, A., Goyal, R.K., Handa, A.K. and Segre, A.L. (2006). Nuclear magnetic resonance spectroscopy-based metabolite profiling of transgenic tomato fruit engineered to accumulate spermidine and spermine reveals enhanced anabolic and nitrogen–carbon interactions. *Plant Physiol.* **142**, 1769–1770.

Millar, A.A., Smith, M.A. and Kunst, L. (2000). All fatty acids are not equal: discrimination in plant membrane lipids. *Trends Plant Sci.* **5**, 95–101.

Minorsky, P.V. (2002). Trigonelline: a diverse regulator in plants. *Plant Physiol.* **128**, 7–8.

Moing, A, Escobar-Gutierrez, A and Gaudillère, J.P. (1994). Modeling carbon export out of mature peach leaves. *Plant Physiol.* **106**, 591–600.

Moing, A., Maucourt, M., Renaud, C., Gaudillère, M., Brouquisse, R., Lebouteiller, B., Gousset-Dupont, A., Vidal, J., Granot, D., Denoyes-Rothan, B., Lerceteau-Kohler, E. and Rolin, D. (2004). Quantitative metabolic profiling through one-dimensional ^{1}H-NMR analyses: application to plant genetics and functional genomics. *Funct. Plant Biol.* **31**, 889–902.

Nielsen, J. and Oliver, S. (2005). The next wave in metabolome analysis. *Trends Biotechnol.* **23**, 544–546.

Obiadalla-Ali, H., Fernie, A.R., Kossmann, J. and Lloyd, J.R. (2004). Developmental analysis of carbohydrate metabolism in tomato (*Lycopersicon esculentum* cv. Micro-Tom) fruits. *Physiol. Plant.* **120**, 196–204.

Ozga, J.A. and Reinecke, D.M. (2003). Hormonal interactions in fruit development. *J. Plant Growth Regul.* **22**, 73–81.

Persia, M.E., Parsons, C.M., Schang, M. and Azcona, J. (2003). Nutritional evaluation of dried tomato seeds. *Poultry Sci.* **82**, 141–146.

Roessner-Tunali, U., Hegemann, B., Lytovchenko, A., Carrari, F., Bruedigam, C., Granot, D. and Fernie, A.R. (2003). Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development. *Plant Physiol.* **133**, 84–99.

Rolin, D., Baldet, P., Just, D., Chevalier, C., Biran, M. and Raymond, P. (2000). NMR study of low subcellular pH during the development of cherry tomato fruit. *Aust. J. Plant Physiol.* **27**, 61–69.

Sas Institute (1990). *SAS/STAT User's Guide*, Version 6. SAS Institute Inc, Cary, NC.

Schaffer, A.A. and Petreikov, M. (1997). Sucrose-to-starch metabolism in tomato fruit undergoing transient starch accumulation. *Plant Physiol.* **113**, 739–746.

Schauer, N., Zamir, D. and Fernie, A.R. (2005). Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. *J. Exp. Bot.* **56**, 297–307.

Seifert, G.J. (2004). Nucleotide sugar interconversions and cell wall biosynthesis: how to bring the inside to the outside. *Curr. Opin. Plant Biol.* **7**, 277–284.

Shelp, B., Walton, C., Snedden, W., Tuin, L., Oresnik, I. and Layzell, D. (1995). Gaba shunt in developing soybean seed is associated with hypoxia. *Physiol. Plant.* **94**, 219–228.

Stevens, R, Bouzayen, M, Causse, M, Etienne, C and Rothan, C (2007). Tomato: a model species for Solanaceae genomics in Morot-Gaudry in Lea, J.F. and Briat, J.F. (Eds), *Functional Plant Genomics*. INRA, Paris, pp. 443–461.

Sunitha, G. and Bradford, K.J. (2001). Galactosyl-sucrose oligosaccharides and potential longevity of primed seeds. *Seed Sci. Res.* **11**, 121–133.

Tarpley, L., Duran, A.L., Kebrom, T.H. and Sumner, L.W. (2005). Biomarker metabolites capturing the metabolite variance present in a rice plant developmental period. *BMC Plant Biol.* **5**, doi:10.1186/1471-2229-5-8.

Télef, N., Stammitti-Bert, L., Mortain-Bertrand, A., Maucourt, M., Carde, J.P., Rolin, D. and Gallusci, P. (2006). Sucrose deficiency delays lycopene accumulation in tomato fruit pericarp discs. *Plant Mol. Biol.* **62**, 453–469.

Toor, R.K. and Savage, G.P. (2005). Antioxidant activity in different fractions of tomatoes. *Food Res. Intern.* **38**, 487–494.

Valle, E.M., Boggio, S.B. and Heldt, H.W. (1998). Free amino acid composition of phloem sap and growing fruit of *Lycopersicon esculentum. Plant Cell Physiol.* **39**, 458–461.

Velterop, J.S. and Vos, F. (2001). A rapid and inexpensive microplate assay for the enzymatic determination of glucose, fructose, sucrose, L-malate and citrate in tomato (*Lycopersicon esculentum*) extracts and in orange juice. *Phytochem. Anal.* **12**, 299–304.

Voelker, T. and Kinney, A.J. (2001). Variations in the biosynthesis of seed-storage lipids. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 335–361.

Whitaker, B.D. (1988). Changes in the steryl lipid content and composition of tomato fruit during ripening. *Phytochemistry* **27**, 3411–3416.

White, P.J. (2002). Recent advances in fruit development and ripening: an overview. *J. Exp. Bot.* **53**, 1995–2000.

Yelle, S., Hewitt, J.D., Robinson, N.L., Damon, S. and Bennett, A. (1988). Sink metabolism in tomato fruit. III. Analysis of carbohydrate assimilation in a wild species. *Plant Physiol.* **87**, 737–740.

Yilmaz, E., Tandon, K.S., Scott, J.W., Baldwin, E.A. and Shewfelt, R.L. (2001). Absence of a clear relationship between lipid pathway enzymes and volatile compounds in fresh tomatoes. *J. Plant Physiol.* **158**, 1111–1116.