Metabolomics reveals organ-specific metabolic rearrangements during early tomato seedling development

Published in:
Metabolomics

DOI:
10.1007/s11306-014-0625-2

Published: 01/01/2014

Document Version
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:
• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 24. Oct. 2018
Metabolomics reveals organ-specific metabolic rearrangements during early tomato seedling development

Maria Victoria Gomez Roldan · Bas Engel · Ric C. H. de Vos · Pieter Vereijken · Laura Astola · Marian Groenenboom · Henri van de Geest · Arnaud Bovy · Jaap Molenaar · Fred van Eeuwijk · Robert D. Hall

Received: 2 September 2013 / Accepted: 20 January 2014
© Springer Science+Business Media New York 2014

Abstract Tomato seedlings (Solanum lycopersicum cv. MoneyMaker), grown under strictly controlled conditions, have been used to study alterations occurring in secondary metabolite biosynthetic pathways following developmental and environmental perturbations. Robustness and reproducibility of the system were confirmed using detailed statistical analyses of the metabolome. LCMS profiling was applied using whole germinated seeds as well as cotyledons, hypocotyls and roots from 3 to 9 days old seedlings to generate relative levels of 433 metabolites, of which 62 were annotated. Initial focus was given to the polyphenol pathway and several additional mass signals have been putatively annotated using high mass resolution fragmentation. Clear organ and developmental stage—specific differences were observed. Seeds accumulated saponin-like compounds; roots accumulated mainly alkaloids; cotyledons and hypocotyls contained mainly glycosylated flavonols and; hypocotyls contained mainly anthocyanins. For each organ, the developmental changes in metabolite profiles were described by using linear mixed models. Across three independent experiments, 85 % of the metabolites showed similar developmental trends. This tomato seedling system has given us valuable additional insights into the complexity of seedling secondary metabolism. How metabolic profiles reflect an interplay between depletion of stored molecules and de novo synthesis and how the overall picture for this important crop plant contrasts to e.g. Arabidopsis are emphasised.

Keywords Tomato seedlings · Secondary metabolism · Flavonoids · Metabolomics · Mass spectrometry

1 Introduction

Plant phenotype is the final product of differential gene expression which is under strict spatial and temporal control. Many phenotypic features are determined by the capacity of a plant to synthesize and accumulate sets of metabolites in specific organs (often also in specific cell types), at specific times of the plant life cycle and following specific (internal

Electronic supplementary material The online version of this article (doi:10.1007/s11306-014-0625-2) contains supplementary material, which is available to authorized users.

M. V. G. Roldan · L. Astola · M. Groenenboom · A. Bovy · J. Molenaar · F. van Eeuwijk · R. D. Hall
Netherlands Consortium for Systems Biology, Amsterdam, The Netherlands

M. V. G. Roldan · R. C. H. de Vos · H. van de Geest · A. Bovy · R. D. Hall
Plant Research International, PO Box 16, 6700 AA Wageningen, The Netherlands

Present Address: M. V. G. Roldan (✉) URGV - Crop Genomics, INRA Functional Genomics, CP5708, 2 rue Gaston Crémieux, 91000 Evry, France

e-mail: gomez@evry.inra.fr

B. Engel · P. Vereijken · L. Astola · M. Groenenboom · J. Molenaar · F. van Eeuwijk
Biometris, Wageningen University and Research Centre, Droevendaalsesteeg 1, Wageningen, The Netherlands

R. C. H. de Vos · R. D. Hall
Netherlands Metabolomics Centre, Einsteinweg 55, 2333 CC Leiden, The Netherlands

R. C. H. de Vos · H. van de Geest · A. Bovy · F. van Eeuwijk · R. D. Hall
Centre for Biosystems Genomics, P.O. Box 98, 6700 AB Wageningen, The Netherlands

Published online: 01 February 2014

Metabolomics DOI 10.1007/s11306-014-0625-2
and external) signals (Cook et al. 2004; Tohge et al. 2005). The complexity of plant metabolism and its strong tissue specificity implies strict genetic control.

Seed germination is a well-defined developmental process which is known to involve a paradigm shift in plant metabolism involving many different compound classes. This metabolic reprogramming is required to exit the usual dormant state and to enable the rapid development and successful establishment of a new plant. While in the past, much research has been done on seedling primary metabolism, often in relation to the transition from heterotrophic to autotrophic growth (Allen et al. 2010; Chen and Thelen 2010), distribution and biosynthesis of secondary metabolites during seed germination and early seedling growth have been poorly documented. Seedlings are at one of the most vulnerable stages in the plant life cycle and hence a seedling must already be able to protect itself rapidly from external (a)biotic stresses (De-la-Cruz Chacon et al. 2012). Some of the most rapid response mechanisms available to a seedling involve changes in secondary plant metabolism, resulting in corrective and/or protective compounds, all parts of the plant’s environmental defense arsenal [e.g. in Arabidopsis, (Milkowski and Strack 2010); brassica, (Zhou et al. 2007); tomato, (Guo and Wang 2010); sorghum, (Huang and Backhouse 2005)].

The ability of a seed to germinate and grow successfully is a major factor determining the yield potential of many crop plants. Hence detailed knowledge of this critical stage is of great importance (Weitbrecht et al. 2011). Particular environmental conditions are required by a seed to begin the process of germination (moisture, temperature, light/darkness etc.) depending on the individual species. First, the seed absorbs water (imbibition) which provides the moisture needed to promote cellular metabolism and growth of the embryo. Once the embryo expands it emerges through the seed coat to become established and initiate the autotrophic photosynthetic growth phase. All these developmental stages are associated with, and require, organ and tissue-specific changes to general metabolism. It is known that during seed germination some secondary metabolites are already available in stored form, while a significant number is synthesized de novo using the reserves present in the embryo and/or endosperm (see De-la-Cruz Chacon et al. 2012). These reserves allow the normal development and protection of the young seedling which comprises essentially three main parts: the radicle (embryonic root), the hypocotyl and the cotyledons. Each of these have their own unique metabolic requirements (De-la-Cruz Chacon et al. 2012).

There are clear advantages for using Arabidopsis for phenotypic analysis and this model has already been widely used for detailed descriptions of growth and development and after applying environmental or genetic perturbations (Bolle 2009; Fabre et al. 2011). However, the regulation of the flavonoid biosynthesis network differs between species such as tomato and Arabidopsis (Yonekura-Sakakibara et al. 2007, 2008). This is also generally true for many secondary metabolite pathways in seedlings as has been reviewed recently by De La Cruz Chacón et al. (2009). They have shown that aspects of seed storage versus de novo synthesis for many (groups of) secondary metabolites can also be species specific. Consequently, it is important to collect additional information from plants other than Arabidopsis to understand more fully flavonoid biosynthetic control in a range of organisms.

Tomato (Solanum lycopersicum) is an important crop species which is consumed worldwide (Gupta et al. 2009). Tomato fruits are an important source of nutrients and antioxidants in the human diet, including carotenoids and phenylpropanoids and especially, flavonoids. Several studies have been performed to understand the genetic regulation of the biosynthetic pathway of these secondary metabolites at the fruit level. However, despite knowing that polyphenol biosynthesis is also important in seedling (stress) physiology (Guo and Wang 2010; Saito et al. 2013), few studies have focused on phenolic biosynthesis at this early stage of tomato plant development.

In this paper we describe a system which is based upon analyzing distinct organs taken from seedlings of an isogenic tomato line after growth under strictly-controlled, defined conditions. Compared to Arabidopsis this is relatively easy as the larger size permits organ-level metabolomic resolution. Analytical resolution at or below organ level is of considerable importance in gaining a full understanding of how plant metabolism is locally controlled and how these control mechanisms are influenced (Keurentjes et al. 2011). Sub-organ level analysis should also become possible in the future. We have applied an LCMS-based untargeted metabolomics approach to study and statistically describe the temporal and organ-specific changes in metabolite composition after seed germination and growth, and to examine the effect of environmental perturbation by modifying day-length during early seedling growth. The results described reveal the complex specificity of metabolic reprogramming occurring upon tomato seedling development. Moreover, statistical analysis has revealed relevant patterns and trends in metabolic profiles that can be used for closer study of the dynamics of plant metabolism and organ specific trends. This will ultimately facilitate the identification of relevant structural genes and control elements.

2 Materials and methods

2.1 Plant material and growth conditions

Surface-sterilized seeds (25 per pot) of tomato (Solanum lycopersicum L. cv. Moneymaker) were sown in sterile
plastic pots filled with 70 ml half-strength Hoagland nutrition solution/0.5 % Agar. Pots were placed in a growth chamber under a cycle of 16 h light (100 μmol m−2 s−1)/8 h darkness at 25 °C (standard growth-light conditions). Samples were harvested at the same time of day (after 7 h of light) starting at day 3 after sowing and then every day until day 9. Germinated seeds from day 3 to day 4 were harvested whole, immediately frozen in liquid nitrogen, ground to a fine powder, freeze-dried for 24 h and stored at −80 °C. Seedlings from day 5 until day 9 were first separated into cotyledons, hypocotyls and roots before treating as above. For each day, three biological replicates were independently grown in separate pots. From each pot 15–20 representative seedlings were pooled as one replicate. After 1 and 2 weeks, second (B) and third experiments (C), were performed identically. Experiments, A, B and C each contained 51 samples (three biological replicates per organ per day), and were performed in order to evaluate the robustness (across experiments) and reproducibility (within experiments) of the approach. Plants subjected to modified light conditions were grown for the first 5 days under standard conditions before being transferred to either complete darkness or continuous light for the rest of the experiment (harvesting each day from day 6 to day 9). Control plants were maintained under the standard light regime throughout.

2.2 Metabolite extraction and LC-PDA-QTOF MS analysis

All samples (10 mg) were extracted with 70 volumes of 70 % aqueous-methanol acidified with 1 % (v/v) formic acid. Ten quality control samples (from a pool of tomato organ powder) were also included in the analysis as technical quality control samples. Extracts were sonicated for 15 min, filtered through a 0.45 μm inorganic membrane filter (Sarturris stedim, Biotech) and analyzed by liquid chromatography–photodiode array–electrospray ionization/ quadrupole time-of-flight mass spectrometry (LC-PDA-ESI/QTOF-MS) as described in (Moco et al. 2006). Briefly, for chromatographic separation a Luna C18(2) pre-column (4 m m2) and an analytical column (2.0 × 150 mm2, 100 Å, particle size 3 μm; Phenomenex) were used. Samples (5 μL) were separated using a gradient of formic acid:water (1:1,000, v/v; eluent A) and formic acid : acetonitrile (1:1,000, v/v; eluent B). Flow was set at 0.19 mL min−1 with the linear gradient of 5 % B to 35 % B in 45 min, after which the column was washed for 15 min and equilibrated. The column temperature was maintained at 40 °C and the samples at 20 °C. Absorbance spectra of eluting compounds were measured using a Waters 2996 PDA detector (240–600 nm) and ESI–MS analysis was performed using a QTOF Ultima V4.00.00 mass spectrometer (Waters-Corporation, MS technologies) in negative ionization mode. A collision energy of 10 eV was used for full-scan LC–MS in the m/z range 100–1,500. Leucine enkephalin, [M–H]− = 554.2620, was used for online mass calibration (lock mass). Acquisition and visualization of the LC-(PDA)-ESI/QTOF-MS data were performed using MassLynx 4.0 software (Waters).

2.3 Data processing of the LCMS profiles

The MetAlign software package (www.metAlign.nl; Lommen 2009) was used for baseline correction, noise estimation, and mass signal alignment between retention times 1.4–47.0 min, including the injection peak but excluding column washing. A data quality check was performed using MetAlign Output Transformer (METOT, Plant Research International, Wageningen). After uploading the MetAlign output files, METOT randomized noise values (replaced missing peaks), removed low and/or irreproducible peaks, evaluated/validated the filtered peak list for mass accuracy of signals across samples and analysis time (to ensure that the calculated means of accurate masses are correct), as well as variation in the intensity of detected signals in replicate samples using scatter plots and distribution plots. Signals <3 times local noise were discarded and data quality and analytical technical variation were assessed. The resulting filtered mass peak table contained 6,869 signals and was subjected to redundancy removal using MSClust software (Tikunov et al. 2012) to generate reconstructed metabolites each represented by a single, non-saturated mass signal. The resulting dataset used for further analysis was a matrix containing the relative intensities of 433 reconstructed metabolites in each sample.

Comparison and visualization of the main features of the LC–MS data was performed using principal components analysis (PCA) and hierarchical cluster analysis (HCA) using GeneMaths XT 1.6 software (www.applied-maths.com). Prior to PCA, metabolite levels were log10 transformed and standardized (subtracting the average and dividing by the standard deviation). For annotated metabolites, HCA was performed using a UPGMA method on metabolite by metabolite similarities defined by Pearson correlation coefficients between averages calculated per experiment and organ.

2.4 Metabolite annotation

Tomato fruit metabolite databases available via http://appliedbioinformatics.wur.nl/moto (Moco et al. 2006) and http://webs2.kazusa.or.jp/komics/ (Iijima et al. 2008) were used to annotate compounds present in seedling samples, based on their corresponding accurate mass (using a threshold of 5 ppm), chromatographic retention time, UV/
<table>
<thead>
<tr>
<th>ID</th>
<th>RT (min)</th>
<th>Formula</th>
<th>UV signal</th>
<th>Av mass m/z</th>
<th>Error (ppm)</th>
<th>MS/MS</th>
<th>Putative annotation</th>
<th>Annot. level</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>C15H16O11</td>
<td>310</td>
<td>371.0639</td>
<td>5.41</td>
<td>209.0302, 191.0196 (21)</td>
<td>Caffeoylglucaric acid I</td>
<td>2</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>C15H16O11</td>
<td>310</td>
<td>371.0628</td>
<td>2.29</td>
<td>209.0303, 191.0198 (20)</td>
<td>Caffeoylglucaric acid II</td>
<td>2</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>C15H16O11</td>
<td>310</td>
<td>371.0625</td>
<td>1.47</td>
<td>209.0303, 191.0198 (22)</td>
<td>Caffeoylglucaric acid III</td>
<td>2</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>C15H18O9</td>
<td>341</td>
<td>371.0624</td>
<td>1.15</td>
<td>209.0303, 191.0198 (22)</td>
<td>Caffeoylglucaric acid III</td>
<td>2</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>C15H18O9</td>
<td>341</td>
<td>371.0623</td>
<td>1.20</td>
<td>209.0303, 191.0198 (22)</td>
<td>Caffeoylglucaric acid III</td>
<td>2</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>C15H18O9</td>
<td>341</td>
<td>371.0622</td>
<td>1.21</td>
<td>209.0303, 191.0198 (22)</td>
<td>Caffeoylglucaric acid III</td>
<td>2</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>C15H18O9</td>
<td>341</td>
<td>371.0621</td>
<td>1.22</td>
<td>209.0303, 191.0198 (22)</td>
<td>Caffeoylglucaric acid III</td>
<td>2</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>C15H18O9</td>
<td>341</td>
<td>371.0620</td>
<td>1.23</td>
<td>209.0303, 191.0198 (22)</td>
<td>Caffeoylglucaric acid III</td>
<td>2</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>C15H18O9</td>
<td>341</td>
<td>371.0619</td>
<td>1.24</td>
<td>209.0303, 191.0198 (22)</td>
<td>Caffeoylglucaric acid III</td>
<td>2</td>
<td>a</td>
</tr>
</tbody>
</table>
Table 1 continued

<table>
<thead>
<tr>
<th>ID</th>
<th>RT (min)</th>
<th>Formula</th>
<th>UV signal</th>
<th>Av mass m/z</th>
<th>Error (ppm)</th>
<th>MS/MS</th>
<th>Putative annotation</th>
<th>Annot. level</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>38</td>
<td>23.56</td>
<td>C28H32O17</td>
<td>253</td>
<td>639.1570</td>
<td>0.57</td>
<td>315.0506, 459.0926 (56), 314.0429 (45), 300.0273 (15)</td>
<td>Isorhamnetin 3-sophoroside</td>
<td>2</td>
<td>g</td>
</tr>
<tr>
<td>40</td>
<td>24.28</td>
<td>C27H30O16</td>
<td>254</td>
<td>609.1461</td>
<td>0.01</td>
<td>301.0349, 300.0271 (29), 343.0452 (6), 71.02448 (5)</td>
<td>Quercetin 3-rutinoside</td>
<td>1</td>
<td>h,e</td>
</tr>
<tr>
<td>42</td>
<td>24.50</td>
<td>C28H32O17</td>
<td>263</td>
<td>639.1567</td>
<td>0.19</td>
<td>315.0511, 300.0277 (62), 316.0544 (14), 271.0248 (12)</td>
<td>Isorhamnetin 3-gentiobioside</td>
<td>2</td>
<td>g</td>
</tr>
<tr>
<td>43</td>
<td>24.55</td>
<td>C32H38O19</td>
<td>725</td>
<td>639.1567</td>
<td>0.19</td>
<td>315.0511, 300.0277 (62), 316.0544 (14), 271.0248 (12)</td>
<td>Isorhamnetin 3-gentiobioside</td>
<td>2</td>
<td>g</td>
</tr>
<tr>
<td>45</td>
<td>25.27</td>
<td>C21H20O12</td>
<td>267</td>
<td>639.1567</td>
<td>0.19</td>
<td>315.0511, 300.0277 (62), 316.0544 (14), 271.0248 (12)</td>
<td>Isorhamnetin 3-gentiobioside</td>
<td>2</td>
<td>g</td>
</tr>
<tr>
<td>48</td>
<td>27.31</td>
<td>C27H30O15</td>
<td>263</td>
<td>593.1499</td>
<td>2.18</td>
<td>285.0399, 286.0432 (10), 257.0452 (4)</td>
<td>Kaempferol 3-glucoside</td>
<td>1</td>
<td>h,e</td>
</tr>
<tr>
<td>51</td>
<td>28.41</td>
<td>C36H36O19</td>
<td>263</td>
<td>717.1791</td>
<td>1.6</td>
<td>625.1406, 626.1432 (9), 317.0301 (1)</td>
<td>Myricetin hexose-deoxy-com</td>
<td>2</td>
<td>h</td>
</tr>
</tbody>
</table>

**Anthocyanins (7 compounds)**

<table>
<thead>
<tr>
<th>ID</th>
<th>RT (min)</th>
<th>Formula</th>
<th>UV signal</th>
<th>Av mass m/z</th>
<th>Error (ppm)</th>
<th>MS/MS</th>
<th>Putative annotation</th>
<th>Annot. level</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>18.32</td>
<td>C42H47O24</td>
<td>535</td>
<td>951.2444</td>
<td>2.8</td>
<td>807.1293, 495.2358 (40), 785.107 (35), 33.168 (10)</td>
<td>Delphinidin 3-(caffeoyl)-rutinoside-5-glucoside</td>
<td>2</td>
<td>h</td>
</tr>
<tr>
<td>29</td>
<td>20.51</td>
<td>C43H49O24</td>
<td>531</td>
<td>965.2573</td>
<td>1.98</td>
<td>807.1293, 495.2358 (40), 785.107 (35), 33.168 (10)</td>
<td>Delphinidin 3-(caffeoyl)-rutinoside-5-glucoside</td>
<td>2</td>
<td>h</td>
</tr>
<tr>
<td>30</td>
<td>20.59</td>
<td>C42H47O23</td>
<td>535</td>
<td>935.2486</td>
<td>2.02</td>
<td>755.1455, 647.1358 (60), 771.1008 (50), 501.1856 (25), 317.0301 (1)</td>
<td>Delphinidin 3-(caffeoyl)-rutinoside-5-glucoside</td>
<td>2</td>
<td>h</td>
</tr>
<tr>
<td>32</td>
<td>20.86</td>
<td>C42H47O23</td>
<td>532</td>
<td>935.2489</td>
<td>2.05</td>
<td>755.2127, 647.0863 (70), 771.1676 (60), 501.1502 (20), 317.0301 (1)</td>
<td>Delphinidin 3-(cinnamoyl)-rutinoside-5-glucoside</td>
<td>2</td>
<td>h</td>
</tr>
<tr>
<td>36</td>
<td>22.91</td>
<td>C43H49O23</td>
<td>531</td>
<td>949.2637</td>
<td>1.58</td>
<td>787.2149, 467.16 (29), 315.1787 (11)</td>
<td>Petunidin 3-(trans-coumaroyl)-rutinoside-5-glucoside</td>
<td>2</td>
<td>h</td>
</tr>
<tr>
<td>35</td>
<td>23.82</td>
<td>C44H51O23</td>
<td>537</td>
<td>979.2743</td>
<td>1.65</td>
<td>817.1396, 677.1773 (44), 495.0868 (32), 315.1904 (11)</td>
<td>Petunidin 3-(trans-coumaroyl)-rutinoside-5-glucoside</td>
<td>2</td>
<td>h</td>
</tr>
<tr>
<td>44</td>
<td>25.13</td>
<td>C44H51O23</td>
<td>532</td>
<td>963.2794</td>
<td>2.01</td>
<td>509.1215, 801.2239 (31), 347.1364 (12), 329.0881 (10)</td>
<td>Malvidin 3-(trans-coumaroyl)-rutinoside-5-glucoside</td>
<td>2</td>
<td>h</td>
</tr>
</tbody>
</table>

**Alkaloids and saponins (11 compounds)**

<table>
<thead>
<tr>
<th>ID</th>
<th>RT (min)</th>
<th>Formula</th>
<th>UV signal</th>
<th>Av mass m/z</th>
<th>Error (ppm)</th>
<th>MS/MS</th>
<th>Putative annotation</th>
<th>Annot. level</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>26.08</td>
<td>C50H83NO22</td>
<td>1094.5439</td>
<td>1.06</td>
<td>807.1293, 495.2358 (40), 785.107 (35), 33.168 (10), 317.0301 (1)</td>
<td>Hydroxytomatine FA I</td>
<td>2</td>
<td>b,e</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>27.15</td>
<td>C50H83NO22</td>
<td>1094.5452</td>
<td>0.95</td>
<td>807.1293, 495.2358 (40), 785.107 (35), 33.168 (10), 317.0301 (1)</td>
<td>Hydroxytomatine FA II</td>
<td>2</td>
<td>b,e</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>27.91</td>
<td>C50H81NO21</td>
<td>1076.5324</td>
<td>3.86</td>
<td>868.4706, 898.4815 (52), 736.4277 (23), 574.3762 (16)</td>
<td>Hydroxytomatine FA III</td>
<td>2</td>
<td>b,e</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>28.06</td>
<td>C50H83NO22</td>
<td>1094.5438</td>
<td>1.71</td>
<td>1049.3886.4807, 916.4912 (68), 754.4388 (28), 592.3835</td>
<td>Hydroxytomatine FA IV</td>
<td>2</td>
<td>b,e</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>30.06</td>
<td>C50H83NO22</td>
<td>1094.5430</td>
<td>3.85</td>
<td>868.4706, 898.4815 (52), 736.4277 (23), 574.3762 (16)</td>
<td>Hydroxytomatine FA V</td>
<td>2</td>
<td>b,e</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>30.63</td>
<td>C50H81NO21</td>
<td>1030.5251</td>
<td>3.24</td>
<td>868.4706, 898.4815 (52), 736.4277 (23), 574.3762 (16)</td>
<td>Hydroxytomatine FA VI</td>
<td>2</td>
<td>b,e</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>34.57</td>
<td>C50H81NO21</td>
<td>1076.5308</td>
<td>2.39</td>
<td>868.4706, 898.4815 (52), 736.4277 (23), 574.3762 (16)</td>
<td>Hydroxytomatine FA VII</td>
<td>2</td>
<td>b,e</td>
<td></td>
</tr>
</tbody>
</table>

Tomato seedlings metabolism during germination
Vis absorbance spectrum and in-source fragments (if present). In addition, compounds described in the literature to be present in WT tomato fruits (Gómez-Romero et al. 2010; Tikunov et al. 2010; Vallverdu-Queralt et al. 2010), transgenic tomato fruits (Butelli et al. 2008), transgenic tomato seedlings (Bovy et al. 2002), tomato seeds (Ferreres et al. 2010) and tomato leaves (Strack et al. 1987; Lawson et al. 1997; Schmidt et al. 2011) were incorporated into the list of known tomato compounds. New metabolites were putatively annotated based on their elemental formulae calculated from the observed accurate mass and using additional information such as UV absorption spectra and high-mass resolution fragmentation data. Metabolite annotation used in Table 1 follows MSI recommendations (Sumner et al. 2007).

### 2.5 Metabolite identification using high mass resolution fragmentation

High mass resolution MS of selected molecular ions was performed on a LC-LTQ-Orbitrap FTMS hybrid mass spectrometer (Thermo Fisher Scientific), operating at 60,000 resolution, as described recently (van der Hooft et al. 2012). Xcalibur 2.07 software (Thermo) was used for data analysis and assignment of elemental formulae to the fragment ions from the parent ion masses. Chromatographic conditions were the same as those described above.

### 2.6 Statistical analysis

#### 2.6.1 Linear mixed model to study seedling profiles

Data were log10 transformed (base 10) peak intensities. For individual metabolites, linear mixed models (McCulloch et al. 2008) were fitted assuming (at most) a second degree relationship between metabolite levels and time. Thus, metabolites either remain constant in time, increase or decrease linearly, or increase or decrease with an upward or downward curvature. Models were fitted separately for cotyledon, hypocotyl and root organs.

Mass signals below 22 ion counts per second were considered as absent. For any organ where a metabolite was absent or excluded in more than 15 out of the total of 45

<table>
<thead>
<tr>
<th>ID</th>
<th>RT (min)</th>
<th>Formula</th>
<th>UV signal</th>
<th>Av mass m/z</th>
<th>Error (ppm)</th>
<th>MS/MS</th>
<th>Putative annotation</th>
<th>Annot. level</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>35.02</td>
<td>C51H85NO22</td>
<td>1108.5593</td>
<td>4.32</td>
<td>901.5009, 900.4973 (84), 576.3916 (27), 738.444 (13),</td>
<td>Tomatidine + 4 hexose FA</td>
<td>2</td>
<td>c</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>35.3</td>
<td>C50H83NO21</td>
<td>1032.5372</td>
<td>1.21</td>
<td>870.4863, 900.4971 (85), 576.3909 (14), 738.4433 (13),</td>
<td>Alpha-tomatine</td>
<td>1</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>36.61</td>
<td>C50H83NO21</td>
<td>1078.5465</td>
<td>2.41</td>
<td>870.4865, 900.4977 (62), 738.4438 (23), 576.394 (17),</td>
<td>Alpha-tomatine FA II</td>
<td>2</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>37.89</td>
<td>C51H86O24</td>
<td>1081.5468</td>
<td>1.3</td>
<td>919.4542, 757.4832 (45), 595.3728 (5), 433.4160 (1)</td>
<td>Tomatoside A</td>
<td>2</td>
<td>b</td>
<td></td>
</tr>
</tbody>
</table>

### Others (5 compounds)

<table>
<thead>
<tr>
<th>ID</th>
<th>RT (min)</th>
<th>Formula</th>
<th>UV signal</th>
<th>Av mass m/z</th>
<th>Error (ppm)</th>
<th>MS/MS</th>
<th>Putative annotation</th>
<th>Annot. level</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.25</td>
<td>C6H8O7</td>
<td>191.0184</td>
<td>6.85</td>
<td>111.0090, 173.0093 (12)</td>
<td>Citric acid</td>
<td>1</td>
<td>c</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5.13</td>
<td>C9H10NO2</td>
<td>164.0714</td>
<td>1.58</td>
<td>147.0451</td>
<td>Phenylalanine</td>
<td>1</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8.85</td>
<td>C11H11N2O2</td>
<td>203.0827</td>
<td>0.87</td>
<td>159.0927, 116.0507 (38)</td>
<td>Tryptophan</td>
<td>1</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>24.37</td>
<td>C13H14N2O3</td>
<td>245.0924</td>
<td>2.96</td>
<td>203.0827</td>
<td>Acetyl tryptophan</td>
<td>2</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>45.31</td>
<td>C18H31O5</td>
<td>327.2179</td>
<td>0.88</td>
<td>171.1025, 229.1443 (99), 211.1338 (63), 291.1962 (60),</td>
<td>Trihydroxy-octadecadienoic acid</td>
<td>2</td>
<td>f</td>
<td></td>
</tr>
</tbody>
</table>

Annotation level (Annot. level) was classified as recommended by the metabolomics standards initiative (MSI): 1 identified metabolites and 2 putatively annotated compounds.

samples, this metabolite was removed from the mixed model analysis. The number of metabolites analysed for each organ is indicated in Supplementary Table S4.

Per organ, for each metabolite, mixed models were fitted across the three experiments. The model comprised three fixed effects: for experiment (intercept, mean, overall level), for linear trend or increase (slope), and for quadratic trend (curvature). The trend models the dynamic behaviour of the metabolite level in relation to days after sowing. Two random terms were added, one for the random deviation from the curve (lack of fit) for the mean of the three biological replicates for a metabolite on a given day for a particular experiment, and one for the random variation occurring between the three biological replicates within a given day for a particular experiment.

With \( y_{itk} \) for the metabolite level for experiment \( i = \{A, B, C\} \), day \( t = \{5, 6, 7, 8, 9\} \) and within-day biological replicate \( k = \{1, 2, 3\} \). The model reads as follows:

\[
y_{itk} = \beta_{0i} + \beta_{1i} q_1(t) + \beta_{2i} q_2(t) + u_{it} + e_{itk},
\]

where, \( \beta_{0i} \), \( \beta_{1i} \) and \( \beta_{2i} \) represent the overall level, linear increase and curvature of the metabolite profile. Functions \( q_1(t) \) and \( q_2(t) \) are linear and quadratic orthogonal polynomials in time. Random effects \( u_{it} \) and \( e_{itk} \) (mainly) represented lack of fit and pure error. These terms are assumed to be independently, normally distributed around 0, with associated components of variance \( \sigma^2_u \) and \( \sigma^2_e \). Components of variance were estimated by restricted maximum likelihood (REML, see McCulloch et al. 2008). Calculations were performed with GenStat (International 2011). Estimates for the constant \( \beta_{0i} \), slope \( \beta_{1i} \) and curvature \( \beta_{2i} \), as obtained from a model without experiment-specific parameters were used to compare organs. Additional mixed model analyses were performed for the intercept \( \beta_{0i} \), slope \( \beta_{1i} \), and curvature \( \beta_{2i} \) separately, across metabolites. These analyses allowed us to compare variation across metabolites with variation within metabolites to judge the potential of the intercept, slope and curvature for further exploration of metabolites, e.g. indicating possible biological pathways.

### 2.6.2 Robustness and reproducibility

The structure of our linear mixed model allowed us to use \( F \) tests to evaluate the robustness of the system, i.e. whether profiles changed between experiments, where these changes could occur in the mean value, \( \beta_{0i} \), the slope \( \beta_{1i} \), and the curvature \( \beta_{2i} \). Lack of fit of the profile was tested via a likelihood ratio test for variance component \( \sigma^2_e \) being significantly different from 0 (referring to a mixture of \( \chi^2 \) distributions, see e.g. (Self and Liang 1987). Reproducibility was evaluated based on the values for the variance components \( \sigma^2_u \) and \( \sigma^2_e \). We defined a coefficient of determination \( R^2 \) and coefficient of variation (CV) as follows. \( R^2 = (1 - V_{\text{Large}}/V_{\text{Small}}) \times 100\% \), where \( V_{\text{Small}} \) is the sum of the two estimated variance components when the metabolite is assumed not to depend on experiment and time, and \( V_{\text{Large}} \) is the sum of the two components when the profile is assumed to take an experiment specific shape. CV was defined by the estimated standard deviation of a single reading as a percentage of the overall mean of the untransformed peak intensities. \( R^2 \) values are on a scale from 0 (low) to 1 (high), while CV is between 0 and 100 %.

### 2.6.3 Effects of different light conditions

Statistical analyses of the effect of light on developmentally-regulated metabolite profiles were performed per metabolite, organ and light condition, again using quadratic profiles for the development in terms of DAS. Random terms common to the three biological replicates per day were omitted from the model. Combinations of metabolites, organs and conditions with less than 9 out of a total of 12 samples were excluded from this analysis. The number of metabolites analysed on each organ/condition (Supplementary Table S6). Light conditions were compared pairwise, by a significance test based upon a normal approximation.

### 3 Results

#### 3.1 Growth characterization of the tomato seedling experiment

To study the metabolic changes occurring during tomato seed germination, sampling was scheduled to cover the earliest developmental stages of the plant. Germinated seeds were collected daily between 3 and 9 days after sowing (DAS). Per day, 3 replicates each comprising 15–20 pooled individuals were harvested. At 3 DAS, under standard growth conditions (16 h light/8 h dark; 25 °C), the seeds had just germinated and showed an emerged radicle (Fig. 1a). At 4 DAS, seedlings were characterized by the first emergence of the hypocotyl and cotyledons. This stage represents the start of the transition between heterotrophic and autotrophic growth. Seedlings 5–6 DAS had an elongating hypocotyl and the cotyledons were fully expanded. At 7–8 DAS the primary leaves had just begun to appear. At 9 DAS the first two primary leaves had developed further but were still very primordial (Fig. 1a). From 5 DAS onwards, seedlings could be reliably and reproducibly separated into three distinct organs: cotyledons, hypocotyls and roots. Independent experiments (A, B and C) enabled a proper assessment of the robustness of the system. Hypocotyl length, measured for each independent
experiment, increased consistently during seedling development from ca. 1.5 cm on 5 DAS to 2.5–3 cm on 9 DAS (Fig. 1b). In general, hypocotyl elongation was consistent during seedling development varying between 9 and 11 % of the average (see Supplementary Table S1). Percentage water content was also measured during seed germination (Supplementary Table S2) and to compensate for differences, all samples were freeze-dried prior to extraction and results calculated on a dry weight basis.

3.2 Metabolomics data processing and metabolic profile analysis

Samples were analyzed using an accurate mass LC-(PDA)-ESI/QTOF-MS system. Raw data files were pre-processed in an untargeted manner using MetAlign. The overall technical variation of the untargeted analysis workflow applied, including all sample preparation and analysis steps such as weighing, extraction, injection, ionization and mass signal detection, as well as all data pre-processing steps such as unbiased peak picking, noise correction, mass signal alignment and data filtering, was evaluated using the METOT software (Supplementary Fig. S1). Technical variation and mass accuracy was markedly better at a peak intensity threshold of 110 (ion counts per second) as compared to the threshold of 22, indicating that the lower abundant signals, mostly derived from isotopes and fragments, showed more overall technical variation. At a threshold of 110, i.e. signal to noise ratio of ca. 5, which is still below the intensity of the molecular ions of all 433 metabolites, the variation in mass accuracy was <5 ppm for ca. 40 % and <10 ppm for >50 % of the mass signals. Based on the peak list data from the 10 quality control (QC) samples, which were independently prepared from a pool of all organs and injected at both the beginning and end of the sample series and at equal intervals between the experimental samples, the overall technical variation in signal intensity was <18 with 95 % of the signals showing <30 % intensity variation and >60 % of the signals being present in all 10 QC samples.

After data pre-processing into reconstructed metabolites, a PCA was performed to compare the complete tomato LC–MS profiles (Fig. 1c). The QC clusters together near the origin of the PCA scores plot, indicating that these samples behaved neutrally to a system that was able to define the different developmental stages/organs in terms of metabolic differences, thereby confirming the reliability of the metabolite extraction method, the analytical system and the entire data pre-processing procedure. The first and second dimensions of the PCA explain 30 and 20 % of the total variability in the data, respectively. These two components reveal clear differences between germinated seeds, root, cotyledon and hypocotyl organs without clear effects between the three independent experiments (A, B or C, Fig. 1c).

3.3 Organ metabolite profiles of seedlings

For annotation of the metabolites, their retention times, UV/Vis spectra and MS data (accurate mass and fragmentation pattern in negative mode) were compared to metabolite data previously reported for tomato plants.
Using this approach, 62 metabolites could be directly annotated (Table 1), comprising a saponin-like compound, several alkaloids, phenolic acids and flavonoids (flavonols and anthocyanins). Based on these 62 annotated metabolites, clear dynamic changes were observed between the different tomato organs when performing a hierarchical clustering analysis (Fig. 2). Anthocyanins abundantly accumulated in hypocotyls (ids. 27, 29, 30, 32, 36, 39 and 44, Figs. 2, 3). Some of these anthocyanins had previously been observed in transgenic purple tomato fruits Ros/Del (Butelli et al. 2008). Three myricetin derivatives, predicted as anthocyanins precursors, were also detected in hypocotyls samples (ids. 33, 55 and 62). Several flavonols and phenolic acid compounds were particularly detected in cotyledons (ids. 3, 4, 5, 7, 8, 10, 16, 22, 24, 25, 45, 48, 51 and 54; Table 1; Fig. 2). In germinated seeds also certain flavonols (sophoroside derivatives) were highly abundant (ids. 31, 37 and 38). Several phenolic acids and alkaloidal compounds were found in radicals/roots (ids. 9, 11, 17, 19, 20 and 60).
To increase the number of annotated semi-polar metabolites showing particular accumulation in specific tomato organs, the averages of each peak signal from each organ (independent of developmental stage, days) were compared between organs. Signals 3× more abundant in one specific organ, were considered to be particularly accumulating in that organ. Selected signals with intensities ≥1,000 counts were extracted for further characterization. Supplementary MS$^n$ analyses using a high-mass resolution system (LTQ-Orbitrap-FTMS) were performed to provide elemental composition and putative annotation of these unknown metabolites (Supplementary Table S3). Three additional anthocyanins, with RTs of 23.5, 25.8 and 26.1 min and with masses of $m/z$ 949.2636, 949.2641 and 963.2778, respectively, were identified which were highly abundant in hypocotyls (but absent in the Ros/Del tomato fruits, Fig. 3; Supplementary Table S3). Based on their observed MS$^n$ fragments and accurate mass, the first two peaks were putatively identified as isomers of petunidin-(coumaroyl)-rutinoside-glucoside (metabolites h1 and h2, C$_{43}$H$_{49}$O$_{23}$, mass deviation ±1.5 ppm) and the third as an isomer of malvidin-(coumaroyl)-rutinoside-glucoside (h3, C$_{44}$H$_{51}$O$_{23}$, mass deviation ±0.8 ppm, Fig. 3). In addition, a compound with $m/z$ 493.0995 (RT 28.31 min) was selected for MS$^n$ fragmentation and was putatively annotated as methylated myricetin-hexoside (h4, C$_{22}$H$_{22}$O$_{13}$, mass deviation ±1.4 ppm, Supplementary Fig. S2a; Table S3).

In roots, several mass signals were putatively annotated as being glycosylated benzyl and terpenoid derivatives. One was annotated as a rishitin-like derivative (r11, Supplementary Fig. S2b; Table S3), a phytoalexin previously reported to be induced in potato roots during vesicular–arbuscular mycorrhization and which inhibits growth of *Rhizoctonia solani* (Yao et al. 2002; Aliferis and Jabaji 2012).

Two glycoalkaloids (s1, s2) and three tomatoside-like compounds were putatively annotated in gminated seeds (s3, s4 and s5, Supplementary Fig. S2c; Table S3). Saponine compounds (Tomatoside A) had been previously described by Moco et al. (2006) as being a characteristic compound of the jelly part of tomato fruits, which surrounds the seeds. Similar compounds have also been shown to be predominant in unripe tomato fruits (Yamanaka et al. 2009).

In total, of the 433 reconstructed MS signals (representing 433 putative metabolites) extracted from the tomato seedling metabolic profiles, 14 % could be identified based on metabolites previously reported in the literature as being present in tomato organs (Table 1). In addition, we have putatively annotated another 27 new metabolites (6 %) using a high-mass resolution system (Supplementary Table S3). Most of these newly annotated metabolites were identified in the hypocotyl and root organs.

### 3.4 Robustness and reproducibility of the independent seedling experiments

To test whether the metabolic time-profiles from the tomato seedling samples (only cotyledons, hypocotyls and roots) differed between the three experiments, F-tests were constructed for all the metabolites. Results indicated that the majority of these, i.e. 81, 87 and 85 % of the metabolites in cotyledons, hypocotyls and roots, respectively, showed the same pattern upon seedling development (decreasing, increasing or unchanging) between the three experiments (Supplementary Table S4). These results indicate that the development-dependent metabolic changes detected were robust across independent experiments (e.g. see quercetin 3-O-rutinoside slope per organ per experiment in Supplementary Fig. S3). Although 62 % of the metabolites showed significant differences in intercept between experiments, these differences were negligibly small compared to variation in intercept (mean level) across metabolites. The variation between metabolites with regards to their mean level was 5, 72 and 18 times as large as the variation between the three measurements within metabolites, for cotyledons, hypocotyls and roots respectively. Thus, for their mean levels, metabolites can best be differentiated in hypocotyls. For curvature, variation across metabolites was a factor 0.33, 0.08 and 0.09 smaller than...
variation between experiments within metabolites for cotyledons, hypocotyls and roots respectively. Therefore, curvature was discarded as a useful measure for development, and attention was restricted to the constant and slope, giving emphasis to the slopes, as these reflect the dynamics of the system.

To determine how much of the variability of the metabolic patterns could be explained by the quadratic model, the \( R^2 \) was calculated for each organ and metabolite. Median \( R^2 \) values across metabolites varied between 60, 50 and 44 % for cotyledons, hypocotyls and roots, respectively, showing that the unexplained random variation is higher in roots that in the other two organs analysed. The median CV across metabolites was 24 % for cotyledons, 29 % for hypocotyls and 39 % for roots (Supplementary Table S4). So, across metabolites, the cotyledons showed the highest reproducibility, closely followed by the hypocotyls, while roots showed lowest reproducibility.

3.5 Comparing developmental effects on seedling metabolite profiles

Using the preceding mixed model analysis; the slopes of each metabolite were compared to observe organ-specific changes between cotyledon versus hypocotyls, cotyledon versus roots and hypocotyls versus roots (Fig. 4, Supplementary Fig. S4a, b). Slopes were plotted for each metabolite time pattern on four outer edges (Fig. 4), representing the following patterns: upper-left edge: trend with positive slope for hypocotyls (y-axis) and negative slope for cotyledons (x-axis), upper-right edge: trend with positive slope for both organs, lower-right: trend with negative slope on hypocotyls and positive slope on cotyledons, lower-left: trend with negative slope on both organs. Most flavonoids appear to change significantly in both organs. The flavonols quercetin 3-O-rutinoside (id_40) and methyl-myricetin hexose (h4, Fig 4b) significantly increased across time in both organs, whereas some mono and di-glycosylated flavonols (ids_31, 38 and 51) decreased. Several anthocyanins (ids_30, 36 and 44) significantly decreased in both organs (Fig. 4c). New metabolites specific for each organ (cotyledons or hypocotyls) or appearing to change significantly during seedling development were selected for further identification (Supplementary Table S5).

3.6 Metabolic changes upon continuous light and dark growth

In addition to developmental changes, several environmental factors such as light, pathogen infection and nutritional conditions can perturb flavonoid biosynthesis (Shirley 1996). We have determined the effects of environmental perturbations on metabolite pathways by exposing seedlings to three light regimes. These were control conditions (16 h light/8 h dark), and two stress conditions, continuous darkness and continuous light which were applied between 5 and 9 DAS. Hypocotyl elongation increased rapidly in seedlings subjected to continuous darkness, whereas under continuous light the hypocotyl length was slightly reduced compared to control seedlings (Supplementary Fig. S5). Non-targeted LC–MS analyses were performed on both cotyledon and hypocotyl samples followed by PCA on cotyledon and hypocotyl data separately revealed clear differences between the metabolic profiles of seedlings grown in darkness and continuous light (Fig. 5a, b).

For both organs, the effect of light on time-dependent changes in metabolic profiles were analysed by comparing the trends of the slopes for metabolite accumulation (Supplementary Fig. S6 a–d). Comparisons revealed 47.2 and 19.4 % of the cotyledon metabolites showed significant differences between control and dark, and control and 24 h light treatments, respectively, while in hypocotyls these percentages were 18.1 and 13.4 %, respectively (Supplementary Table S6).

Cotyledons in 24 h light showed a significant accumulation of flavonol rutinoside conjugates (Fig. 5c, ids_31 and 37), whereas sophoroside conjugates were significantly increased in cotyledon samples grown in complete darkness (Fig. 5c, ids_40 and 25). In hypocotyls, rutinosides likewise increased in time in the 24 h light samples, while under darkness, diglucoside conjugates of isorhamnetin decreased over time, as compared to the control samples (Supplementary Table S7). Other, yet unknown, metabolite peaks also significantly changed in both hypocotyls and cotyledons upon light perturbation. Additional experiments to annotate these metabolites are needed in order to understand more fully the biochemical response of tomato seedlings upon light perturbation.

4 Discussion

Genomics, in combination with metabolomics approaches have already allowed us to increase our understanding of important biochemical processes of tomato fruit development (Carrari et al. 2006; Moco et al. 2007; Capanoglu et al. 2008; Carl et al. 2009; Osorio et al. 2011; Rohrmann et al. 2011). A vast complexity of primary metabolites and secondary metabolites has been identified as being important factors involved in fruit ripening, quality and flavor. Nevertheless, similarly-detailed studies on the early stages of tomato seed germination are only sporadically described in the literature. Here, we present results from a fast and readily-adaptable system to study metabolic changes
occurring during tomato seedling development, primarily focusing on secondary metabolism. These results reveal the spatio-temporal dynamism of both previously and newly identified metabolites. This approach might also be effectively used in a more systems biology-type approach to study metabolic networks and to find candidate genes for their control points which are of specific relevance to seedling establishment.

Flavonoids (including flavonols and anthocyanins) were found to be the group of secondary metabolites (26 unequivocally identified compounds) predominantly detected at the early stages of tomato seedling development. They were observed in most of the samples analyzed (seeds, cotyledons, hypocotyls) with the exception of the root samples where alkaloids and other phenolic compounds were more prevalent (Fig. 2). Flavonoids are important plant secondary metabolites which are known to accumulate in tomato (Slimestad et al. 2008). In tomato fruit, three flavonoids are typically present in tomato peel; chalcone, naringenin chalcone as well as the flavonol, quercetin-3-rutinoside (Bovy et al. 2010). Tomato seeds and seedlings do not accumulate naringenin chalcone probably due to its rapid conversion into the subsequent precursors of flavonols and anthocyanins. In tomato fruits, the flavonol quercetin-3-O-rutinoside (rutin) accumulates highly during fruit ripening. The same metabolite was readily detected in tomato cotyledons, as well as other rutinoside derivatives of kaempferol and myricetin. None of these compounds was however, detectable in tomato seeds, indicating that de novo biosynthesis must occur early in germination. The accumulation of rutinoside flavonols in seedlings, rather than in seeds, may be relevant as they have a known role in defense against environmental factors such as UV, mechanical wounding or pathogen attack. These (a)biotic factors are of particular importance in the early stages of tomato plant establishment and defense (Dixon et al. 2002).

Anthocyanins have generally only been found to accumulate in vegetative organs and not in fruits of cultivated tomato plants. With only a few exceptions (such as certain mutants and transgenic tomatoes ectopically expressing two transcription factors, Rosea and Delila (RosDel), were
high concentrations of anthocyanins observed in fruit tissues (Butelli et al. 2008). In our study, similar anthocyanins to those synthesized in the transgenic RosDel tomatoes as well as additional ones (Fig. 3), were found to accumulate specifically in tomato hypocotyls. This indicates that under standard cultivation conditions, a positive up-regulation of the anthocyanin pathway seems to occur only in tomato hypocotyls through the activity of transcription factors, which are presumably not expressed in fruits. More detailed investigations of the internal controlling factors inducing this anthocyanin accumulation in tomato seedlings may open up new opportunities to improve the flavonoid content in different tissues of tomato fruits.

Alkaloids were mainly observed in tomato roots and hypocotyls. This class of triterpene-derived metabolites have been found in several members of the Solanaceae family. Tomato fruits accumulate high levels of steroidal glycoalkaloids (SGAs). These provide a chemical barrier to pathogens but may also be an anti-feeding agents—also for humans (Itkin et al. 2011). The most abundant SGAs in tomato fruits are α-tomatine, generally only present in unripe green tissues, and esculeoside A (a potentially related molecule with an additional glucose), which is highly abundant in red-ripe fruit (Moco et al. 2007). In seedlings, no esculoside derivatives were detected whereas several tomatine derivatives were found in both hypocotyls and roots (but absent in cotyledons and seeds). Various steroidic saponins without nitrogen, of the furostanol type (tomatoside A), were also initially observed in seeds samples, but these rapidly disappeared during germination.

The antifungal activity of tomatoside A and the putative function of alkaloids as nitrogen reservoirs in tomato seedlings (De-la-Cruz Chacon et al. 2012), makes this group of metabolites of particular relevance regarding the
study of the unidentified enzymes and genes contributing to their biosynthesis.

4.1 Diversity of secondary metabolites in tomato seedlings

With the detailed statistical analysis having shown that both the experimental design and the results obtained were robust and reproducible, it was justified to proceed with a detailed examination of the discriminatory metabolites and those showing greatest potential relevance. Using our LC–MS analysis we have identified using public data from known tomato metabolites, 20 phenolic acids, 11 alkaloids, 19 flavonols and 7 anthocyanins in tomato seedlings (in addition to several putatively annotated compounds). Similar metabolic profiling has previously also been performed on the model plant Arabidopsis thaliana where induced perturbations have been used to study metabolic dynamics (Bolle 2009; Fabre et al. 2011). A major list of flavonoid molecules has been identified in Arabidopsis using MS and NMR studies (including 35 flavonols, 11 anthocyanins and 8 pro-anthocyanidins) (Saito et al. 2013). Flavonols are widely localised in several tissues such as flowers, mature seeds and seedlings whereas anthocyanins have only be detected in leaves and roots. Based on flavonoid profiles both similarities and differences are observed between Arabidopsis and tomato. For example, proanthocyanidins (PA) have not been previously identified in tomato seeds. PAs are complex flavonoids commonly found in the Arabidopsis seed coat which are predicted to help control seed permeability. Two enzymes, leucoanthocyanidin reductase (LAR) and anthocyanidin reductase (ANR) are involved in PA biosynthesis but seem to be absent in tomato plants.

The major anthocyanins accumulating in tomato seedlings are dihydromyricetin derivatives (delphinidin, petunidin and malvidin). In contrast, in Arabidopsis only cyanidin type anthocyanins (derived from dihydroquercetin) have been identified. While both the enzymes dihydroflavonol reductase (DFR) and anthocyanin synthase (ANS) are present in both species, the substrate specify of each enzyme must differ in order to generate the observed differences in anthocyanin profile. The regulation of the flavonoid biosynthesis network in the two species has clear differences (Yonekura-Sakakibara et al. 2007, 2008). Consequently, it is important to collect additional information from plants other than Arabidopsis in order to understand more fully the flavonoid biosynthetic control mechanisms in a range plants.

Knowledge on the genetics and genomics of tomato is rapidly increasing and has received a major recent boost through the completion of the tomato genome sequencing project (Tomato Genome 2012). Most of the key structural genes of the main flavonoid pathway have already been identified in tomato, as well as several regulatory factors determining overall pathway activity (Bovy et al. 2007; Ballester et al. 2010). After synthesis of their C15 aglycone skeleton, flavonoids are generally subjected to a diverse set of modifying enzymes (glycosyltransferases, methyl/acyl transferases etc.). These decorate this basic molecular skeleton with one to several moieties. These modifications are of great biological relevance as they are known to influence molecular stability, solubility, sequestration, bioavailability etc. (Gomez et al. 2009; Peel et al. 2009; Zhao and Dixon 2010). However, many of the genes involved in these modifications are still unknown or bear only ‘putative’ annotations. We aim to use this seedling system to follow the spatio-temporal dynamics of early plant metabolism and as a tool to find and characterize such genes to enable us to link gene/enzyme with substrate and product. Consequently, in this initial study we have focused on describing the flavonoid changes observed during development as a starting point for future studies of flavonoid regulatory networks in tomato.

4.2 Regulation of flavonoid accumulation in tomato seedlings during development and under light perturbed conditions

Flavonols conjugated with a sophoroside (diglucoside) group were seen to increase in the cotyledons grown in darkness, whereas in contrast, rutinoside derivatives increased in the 24 h light samples. Sophoroside flavonols widely occur in the vegetative organs and seeds but have been rarely observed in tomato fruits. Glycosylation of flavonoids is carried out by uridine diphosphate (UDP) glycosyltransferases (UGTs). UGTs are involved in the transfer of different sugar moieties (glucose, rhamnose, apiose, etc.) to OH groups of flavonoid aglycones. Subsequently, additional sugars can be attached onto already glycosylated flavonoids, most likely catalyzed by another set of (specialized) UGT enzymes. This whole process of glycosylation is a key mechanism determining the specific activity, function and location of different flavonoids in plants (Offen et al. 2006). From the last official annotation of the tomato genome (provided by the International Tomato Annotation Group, ITAG) 228 unigenes were annotated as putative UDP-glycosyltransferases (Tomato Genome 2012). This not only reveals the complexity of the network of modifying enzymes but also defines a clear need to functionally characterize these genes and their spatial and temporal biosynthetic function. While several known UGTs have been shown to have a broad substrate specificity in vitro, recent studies have revealed that UGTs can also be highly regio- and stereo-selective (Tikunov et al. 2010; Itkin et al. 2011). Tomato seedlings, being rich in flavonoid glycosides, should prove a valuable additional tool with which to study further this aspect of the polyphenolic metabolic network and
to expand our gene annotation potential. Similar reports using time-course analysis to infer regulatory/metabolic networks have been reported in recent years for different plant species such as *Arabidopsis* and rice (Kim et al. 2007; Sato et al. 2008; Ozfidan et al. 2013) in relation to different biotic and abiotic conditions. Exploiting deep metabolomics and metabolite annotation approaches, together with e.g. genome wide gene expression analysis using the same plant materials, will help us link specific UGT genes to specific glycosylation patterns of defined glycosylated flavonoids accumulating in tomato. Further identification of the unknown metabolites is therefore still required to elucidate fully their involvement in tomato. Further identification of the unknown metabolites will help us link specific UGT genes to specific glycosylation switch-points. The experimental design and the flavonoid profiles presented here represent a useful resource in specific organs and/or at specific developmental stages. By exploiting the tomato genome and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. Plant Physiology, 152, 71–84.


References


Ozfidan, C., Turkan, I., Sekmen, A. H., & Seckin, B. (2013). Time course analysis of ABA and non-ionic osmotic stress-induced changes in water status, chlorophyll fluorescence and osmotic adjustment in *Arabidopsis thaliana* wild-type (Columbia) and ABA-deficient mutant (aba2). *Environmental and Experimental Botany*, 86, 44–51.


Tomato seedlings metabolism during germination


