Exploring genetic diversity in tomato fruit with a systems biology approach

Laura Pascual, Jiaxin Xu, Nelly Desplat, Benoit B. Valot, Jean-Paul Bouchet, Dominique D. Brunel, Mathilde M. Causse

Exploring genetic diversity in tomato fruit with a systems biology approach

Laura Pascual (1), Jiaxin Xu (1), Nelly Desplat (1), Benoit Valot (2), Mireille Faubert (1), Jean-Paul Bouchet (1), Dominique Brunel (3), Mathilde Causse (1)

(1) INRA, UR1052, GAFL, BP94,F-84143 Montfavet, France (mathilde.causse@avignon.inra.fr)
(2) INRA, Plataforme d'analyse Protéomique de Paris Sud-ouest, France
(3) INRA, EPGV, Centre de Génotypage, Evry, France

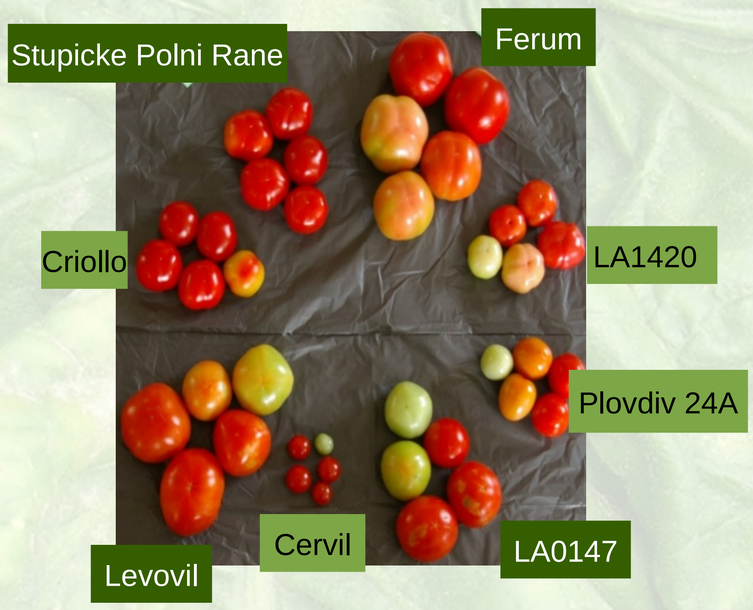
Introduction

Dissection of the genetic variation and inheritance of phenotypic trait is the first step for plant improvement. The cascade of effects from DNA variation to trait phenotype is organized in complicated biological networks (Kliebenstein 2010), where intermediate molecular phenotypes such as transcript and protein abundance also vary in populations (Rockman, 2006). However, **most of the omics studies** developed in crop species **focused on one or two genotypes** and failed to provide an insight into this natural variation.

Tomato (*Solanum lycopersicum*) is a model organism for the fleshy–fruited plants. Its genome has been fully sequenced and annotated (Tomato Genome Consortium, 2012), opening new prospects for analysing biological systems and their complex functions at different levels including genomics, transcriptomics, proteomics, and metabolomics. Here, we carried out an **extensive multi-level omic experiment** in order to dissect fruit quality at several scales.

Materials and methods

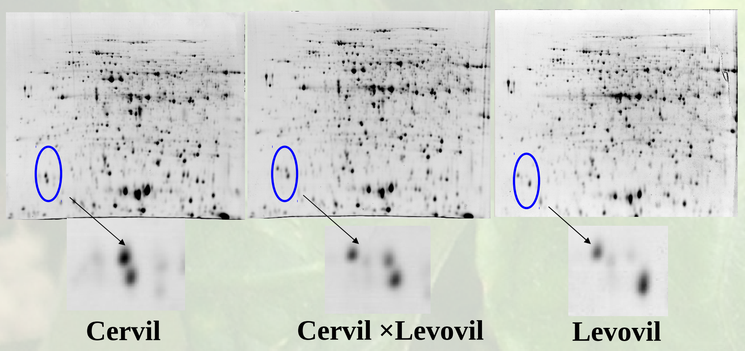
8 divergent tomato lines



The study was carried out using four *Solanum lycopersicum* lines (Levovil, Stupicke Polni Rane, LA0147, Ferum) with large fruits and four cherry type tomato (*S. l.* var *cerasiforme)* lines (Cervil, Criollo, Plovdiv 24A, LA1420), as well as four F1 hybrids between lines of the two groups. Plants were phenotyped for fruit development traits. Besides, fruits were harvested and pericarp samples analysed at 2 stages (**cell expansion and orange**) and different scales. **Proteome** profiles were revealed by 2D-PAGE, variable spots were identified and sequenced by LC-MS. **Gene expression** was analysed by Digital Gene Expression, that is, sequencing short tags (16-17 bp) located at the end of the transcripts.

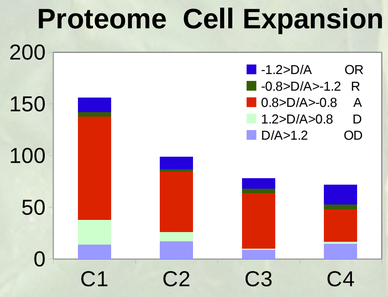
Results : Multi-scale analysis of 8 tomato lines & 4 F1

2D-PAGE + Mass spectrometry



Protein spot volume: **566 variable spots**
336 spots sequenced with known function
More variable spots between stages
Spot volumes used for inheritance analysis

Proteome Cell Expansion



405 spots
60 % A
13 % D/R
28 % OD/OR

A additive, D dominant, R recessive, OD over-dominant, OR over-recessive
C1 CervilXLevovil C2 CriolloXStupick
C3 PlovdivXLA0147 C4 LA1420XFerum

Digital gene expression cDNA tags GAII sequenced

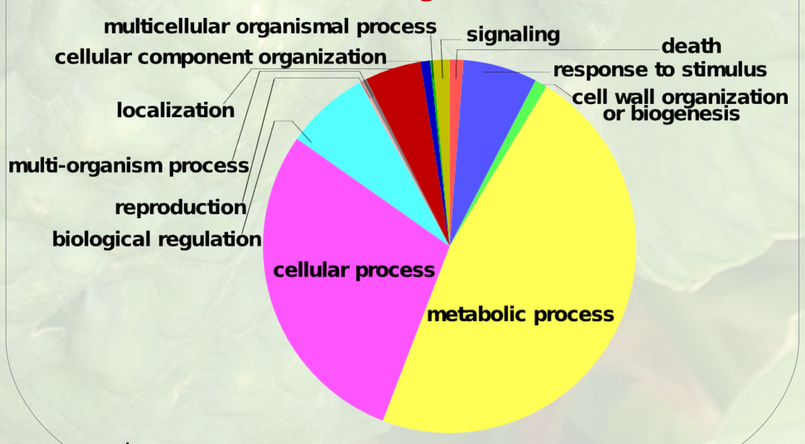
1.75 million distinct tags sequenced*
80% of the tags mapped to a unique gene

21,000 genes identified

3,919 differentially expressed genes between genotypes

	Meanr by sample
Raw tag number	60,272,383
Clean tag number	51,136,012
Distinct clean tag	391,321
% tags mapped to genes	81%
% tags mapped to genome	9%
% Non-mapped	10%

Differential genes GO terms

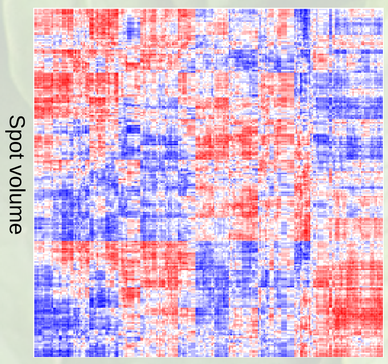


*After filtering tags sequenced just one time

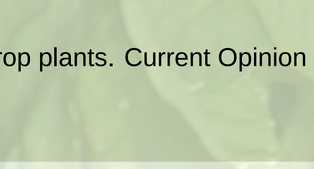
Correlations between protein and gene expression

240 genes & proteins variable at Cell Expansion corresponding to the same unigene

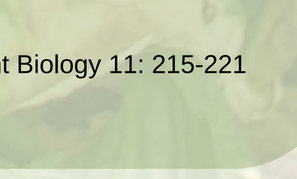
Generally no correlation between spot volume and gene expression



Some single correlations detected



* Malic enzyme



Conclusion

All these data revealed the high natural variation present in the tomato species, in spite of its low molecular diversity. They also showed that there is no clear relation between gene expression and protein amounts, except for a few genes. However, they permitted the discovery of several correlations within and between levels of analysis revealing the regulatory networks operating during fruit development.

This information will be especially useful to analyse a multiallelic genetic intercross population derived from the 8 lines, and the genome of the 8 lines, that has been already sequenced. The detailed description provided here will help us to identify the genes under the QTL and to relate the genetic polymorphisms identified at the sequence level with their expression.

Bibliography

Kliebenstein DJ. 2010. Systems biology uncovers the foundation of natural genetic diversity. Plant Physiol. 152: 480–486.
Rockman MV; Kruglvqk L. 2006 Genetics of global gene expression. Nat Rev Genet. 7(11):862-72.
Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. Nature. 485(7400):635-41.
Cavanagh C., Morell M., Mackay I. And Powell W. 2008. From mutations to MAGIC: resources for gene discovery; validation and delivery in crop plants. Current Opinion in Plant Biology 11: 215-221

The project is funded by the ANR Genomics, project MAGICTom-SNP, in collaboration with Vilmorin & Cie