Integrative study of the proteome throughout tomato fruit development

Isma Belouah

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

École Doctorale des Sciences de la Vie et de la Terre

BIOCHIMIE

Par Isma BELOUAH

Integrative study of the proteome throughout tomato fruit development

Sous la direction de : Sophie COLOMBIÉ

Soutenue le 20 décembre 2017

Membres du jury :

M. BOURDON, Jérémie	Professeur (Université de Nantes)	Rapporteur
M. CURIEN, Gilles	Chargé de Recherche (CNRS CEA, Grenoble)	Rapporteur
Mme LEMAIRE, Martine	Chargé de Recherche (INRA BFP, Bordeaux)	Examinateur
M. ZIVY, Michel	Directeur de Recherche (CNRS GQE, Le Moulon)	Examinateur
M. MAZAT, Jean-Pierre	Professeur Emérite (Université de Bordeaux)	Président de jury
Mme. COLOMBIÉ, Sophie	Chargé de Recherche (INRA BFP, Bordeaux)	Directrice de thèse

Résumé

Étude intégrative du protéome du fruit de tomate au cours du développement

La tomate (*Solanum lycopersicum*), aujourd'hui considérée comme le modèle des fruits charnus, présente de nombreux avantages : facilité de culture, temps de génération court, génome séquencé, facilité de transformation... Le développement du fruit est un procédé complexe hautement régulé et divisible en trois étapes principales : la division cellulaire, l'expansion cellulaire et le mûrissement qui comprend une étape appelée, "mature green", "breaker'' and " turning". Chaque étape est associée à un phénotype, qui lui-même découle de changements à différents niveaux cellulaires. Ainsi l'expression des gènes, l'abondance des protéines, les activités des enzymes, les flux métaboliques et les concentrations en métabolites montrent des changements significatifs au cours de ces étapes. Grâce aux récents progrès technologiques et en particulier au développement des «techniques omiques», comme la génomique, la transcriptomique, la protéomique, la métabolomique, les principaux composants cellulaires peuvent désormais être étudiés à haute densité.

Dans ce contexte, l'objectif de mon doctorat était d'effectuer une analyse protéomique quantitative du développement du fruit de tomate puis d'intégrer les données «omiques» à la fois par des analyses statistiques et par la modélisation mathématique.

Le premier chapitre rapporte les résultats de quantification du protéome de fruit de tomate réalisé en collaboration avec la plateforme PAPPSO (INRA, Gif-sur-Yvette). Des échantillons collectés à neuf stades de développement du fruit de tomate ont été extraits et le protéome quantifié, en absence de marquage, par chromatographie liquide couplée à la spectrométrie de masse (LC-MS/MS). Ensuite, j'ai cherché la méthode la plus adaptée, testant un ensemble de filtres sur les données, pour obtenir une quantification précise des protéines à partir des intensités ioniques (XIC). Au total, j'ai pu obtenir la quantification absolue de 2494 protéines en utilisant une méthode basée sur la modélisation de l'intensité des peptides. La quantification des protéines par LC-MS/MS a finalement été validée par comparaison avec 32 capacités enzymatiques utilisées comme proxy pour l'abondance de protéines.

Le deuxième chapitre est consacré aux résultats obtenus par analyses combinées d'«omiques» au cours du développement du fruit de tomate. La transcriptomique a été réalisée en collaboration avec Genotoul GeT (Toulouse) et le groupe Usadel (RWTH Aachen University, Allemagne). Grâce

à l'ajout d'étalons internes, plus de 20000 transcrits ont été quantifiés de manière absolue à chacune des neuf étapes de développement. Cette quantification a ensuite été validée par comparaison avec des données de concentration de 71 transcrits précédemment obtenues par PCR quantitative. Enfin, nous avons cherché à intégrer les quatre niveaux de données - transcriptome, protéome, métabolome et activome- afin d'identifier les principales variables associées au développement. Pour ces quatre niveaux, les analyses ont confirmé que l'entrée en mûrissement s'accompagne de changements majeurs et révélé une grande similarité entre la fin et le début du développement, notamment au niveau du métabolisme énergétique.

Le troisième chapitre porte sur les résultats de modélisation de la traduction protéique obtenus grâce à la quantification absolue du transcriptome et du protéome. Afin d'expliquer la diminution de la corrélation observée au cours du développement entre les concentrations en protéines et celles des transcrits correspondants, nous avons résolu un modèle mathématique de la traduction protéique basé sur une équation différentielle ordinaire et impliquant deux constantes de vitesse: pour la synthèse et la dégradation de la protéine. La résolution de cette équation, validée par un critère de qualité basé sur un intervalle de confiance fermé, a conduit à l'estimation de ces constantes pour plus de 1000 protéines. Les résultats obtenus ont été comparés aux données de la littérature reportées chez des plantes et plus largement chez des cellules eucaryotes.

Enfin le dernier chapitre décrit l'ensemble du matériel et des méthodes utilisées pour obtenir les différents résultats présentés dans le manuscrit.

Dans le domaine de la biologie des systèmes, ce travail illustre comment l'intégration de multiples données «omiques» et la modélisation mécanistique basée sur la quantification absolue des «omiques» peut révéler de nouvelles propriétés des composants cellulaires.

Mots clés : Tomate, série développementale, « omiques », modélisation, traduction protéique

Biologie du Fruit et Pathologie

INRA UMR1332, Equipe Métabolisme

71, av. Edouard Bourlaux - CS 20032 - 33882 Villenave d'Ornon Cedex – France

Abstract

Integrative study of the proteome throughout tomato fruit development

The interest of the tomato (*Solanum lycopersicum*) fruit has spread in plant science where it is used as the model for fleshy fruit. The valuable advantages of the tomato fruit are numerous: an ease of culture, a short generation time, a high knowledge with important resources, a sequenced genome, an ease for transforming…. The development of tomato fruit is a complex regulated process, divided in three main steps: cell division, cell expansion, and ripening which includes phases such as, mature green, breaker, and turning. Each step is characterized by a phenotype resulting from changes at different cellular levels. Thus, gene expression, protein abundance, enzyme activities, metabolic fluxes and metabolite concentrations show significant changes during these steps. Thanks to recent technologies advances and in particular the development of "omics techniques", such as genomic, transcriptomic, proteomic, metabolomic, the main cell components can now be analyzed by high-throughput.

In this context, the objective of my PhD was to perform a quantitative proteomic analysis of the tomato fruit development and then integrate omics data both by statistical analyses and by mathematical modelling.

The first chapter focused on results obtained for the quantitative proteomic developed in collaboration with the PAPPSO platform (INRA, Gif-sur-Yvette). Samples were harvested at nine stages of tomato fruit development, total proteome was extracted and quantified by label-free LC-MS/MS. Then I searched for the most appropriate method, testing a set of filters on the data, to obtain an absolute label-free protein quantification from ion intensities (XIC). Finally, I obtained the absolute quantification of 2494 proteins using a method based on peptides intensity modelling. The quantification of proteins by LC-MS/MS was then validated by comparison with 32 enzymatic capacities used as proxy for protein abundance.

The second chapter was dedicated to the results of integrative omics analyses throughout tomato fruit development. First, transcriptomic has been performed in collaboration with Genotoul GeT (Toulouse) and Usadel'lab (RWTH Aachen University, Germany). Using spikes in the experimental design, more than 20000 transcripts have been quantitatively determined at the nine stages of development. Then, this absolute quantification of the tomato transcriptome has been

cross-validated with 71 transcripts previously measured by qRT-PCR. Finally, we integrated the four omics datasets- transcriptome, proteome, metabolome and activome – in order to identify key variables of the tomato fruit development. For the four levels, analyses confirmed that the entrance in maturation phase was accompanied by major changes, and revealed a great similarity between the end and the beginning of development, especially in the energy metabolism.

The third chapter focuses on modelling results of the protein translation based on the absolute quantification of transcriptomic and proteomic. To explain the decreasing correlation observed between proteins and transcripts concentration throughout development, we proposed a mathematical model of protein translation based on an ordinary differential equation and involving two rate constants (for synthesis and degradation of the protein). The resolution of this equation, validated by a quality criterion based on a closed confidence interval, led to the estimation of the rate constants for more than 1000 proteins. These results were then compared with previous published data reported for plants and more widely in eukaryotic cells.

Finally, the last chapter describes all the materials and methods used to obtain the results presented in the manuscript.

In the systems biology context, this work illustrates how integration of multiple omics datasets and mechanistic modelling based on absolute omics quantification can reveal new properties of cellular component.

Key words : Tomato fruit, developmental time-series, « omics », modelling, protein translation

Biologie du Fruit et Pathologie

INRA UMR1332, Equipe Métabolisme

71, av. Edouard Bourlaux - CS 20032 - 33882 Villenave d'Ornon Cedex – France

Résumé substantiel

Étude intégrative du protéome du fruit de tomate au cours du développement

La tomate (*Solanum lycopersicum*), aujourd'hui considérée comme le modèle des fruits charnus, présente de nombreux avantages : facilité de culture, temps de génération court, génome séquencé, facilité de transformation... Le développement du fruit de tomate est un procédé complexe hautement régulé et divisible en trois phases principales : la division cellulaire, l'expansion cellulaire et le mûrissement, cette dernière étant initiée par les stades "mature green", "breaker'' et " turning". Tout au long du développement, le phénotype du fruit change, résultant de modifications à tous les niveaux cellulaires. En effet l'expression des gènes, l'abondance des protéines, les activités des enzymes, les flux métaboliques et les concentrations en métabolites présentent des changements significatifs à chacune des étapes de développement. Grâce aux récents progrès technologiques et en particulier au développement des ''techniques omiques'', comme la génomique, la transcriptomique, la protéomique et la métabolomique, les principaux composants cellulaires peuvent désormais être étudiés à haut débit.

Dans ce contexte, l'objectif de mon doctorat était d'effectuer une analyse protéomique quantitative au cours du développement du fruit de tomate puis d'intégrer les différentes données ''omiques'' à la fois par des analyses statistiques et par la modélisation mathématique.

Le premier chapitre rapporte les résultats de quantification du protéome de fruit de tomate réalisée en collaboration avec la plateforme PAPPSO (INRA, Gif-sur-Yvette). Des échantillons collectés à neuf stades de développement du fruit de tomate ont été extraits et le protéome quantifié, en absence de marquage, par chromatographie liquide couplée à la spectrométrie de masse (LC-MS/MS). J'ai ensuite cherché à obtenir une quantification précise des protéines. Pour cela j'ai évalué la performance de cinq méthodes de quantification (iBAQ, TOP3, Average, Average-Log, Model) associée ou non à quatre filtres sur les données des intensités ioniques (XIC) issues d'un mélange de protéines équimolaires appelées UPS (Universal Proteomics Standard) en concentrations croissantes dans un extrait de protéines de levure. Les performances des méthodes ont été évaluées au travers de trois critères majeurs : l'exactitude absolue, l'exactitude relative et la précision. Finalement, j'ai déterminé la quantification absolue de 2494 protéines de péricarpe de fruit de tomate en utilisant la méthode Model, basée sur la modélisation de l'intensité des peptides. La quantification des protéines par LC-MS/MS a finalement été validée par comparaison avec

trente-deux capacités enzymatiques utilisées comme proxy pour l'abondance de protéines. Pour cela, dans un premier temps, nous avons réalisé une analyse de corrélation (Spearman) pour confronter les profils des concentrations en protéines à la fois quantifiées par LC-MS/MS et estimées à partir des capacités enzymatiques. Ensuite, à chaque stade de développement et pour chaque méthode de quantification (LC-MS/MS et capacités enzymatiques), nous avons exprimé les rapports entre les concentrations des 32 protéines enzymatiques. Ainsi, lorsque les coefficients de détermination significatifs (R^2, Spearman) et les rapports entre les concentrations tendant majoritairement vers la valeur attendue de un ont permis de considérer la validation acceptable.

Le deuxième chapitre est consacré aux résultats obtenus par analyses combinées d' ''omiques'' au cours du développement du fruit de tomate. La transcriptomique a été réalisée en collaboration avec Genotoul GeT (Toulouse) et le groupe Usadel (RWTH Aachen University, Allemagne). Grâce à l'ajout d'étalons internes, plus de 20000 transcrits ont été quantifiés de manière absolue à chacune des neuf étapes de développement. Cette quantification a ensuite été validée par comparaison avec des données de concentration de 71 transcrits précédemment obtenues par PCR quantitative. Enfin, nous avons cherché à intégrer les quatre niveaux de données - transcriptome, protéome, métabolome et activome- afin d'identifier les principales variables associées au développement. Pour ces quatre niveaux, les analyses ont confirmé que l'entrée en mûrissement s'accompagne de changements majeurs et révélé une grande similarité entre la fin et le début du développement, notamment au niveau du métabolisme énergétique.

Le troisième chapitre porte sur les résultats de modélisation de la traduction protéique obtenus grâce à la quantification absolue du transcriptome et du protéome. Afin d'expliquer la diminution de la corrélation observée au cours du développement entre les concentrations en protéines et celles des transcrits correspondants, nous avons résolu un modèle mathématique de la traduction protéique basé sur une équation différentielle ordinaire et impliquant deux constantes de vitesse: pour la synthèse et la dégradation de la protéine. La résolution de cette équation, validée par un critère de qualité basé sur un intervalle de confiance fermé, a conduit à l'estimation de ces constantes pour plus de 1000 protéines. La comparaison des résultats à des données de la littérature montre une similarité plus importante avec des données obtenus chez les plantes que celles obtenues chez des cellules eucaryotes. Par ailleurs, l'analyse des durées de synthèse et de dégradation des protéines selon les localisations cellulaires et les rôles fonctionnels montre que la vacuole, lieu de stockage de la cellule végétale, contient les protéines les plus stables.

Enfin le dernier chapitre décrit l'ensemble du matériel et des méthodes utilisées pour obtenir les différents résultats présentés dans le manuscrit.

Dans le domaine de la biologie des systèmes, ce travail illustre comment l'intégration de multiples données ‘‘omiques’’ et la modélisation mécanistique basée sur la quantification absolue des ‘‘omiques’’ peut révéler de nouvelles propriétés des composants cellulaires.

Remerciements

Les remerciements est la partie de la rédaction du manuscrit à la fois la plus agréable à écrire mais aussi la plus complexe. Agréable car on y est libre de s'exprimer sans devoir atteindre un certain nombre de phrases et où personne ne vous dira qu'il faut synthétiser les notions clés, mais aussi complexe car on craint toujours d'oublier quelqu'un.

Tout d'abord, je voudrais présenter mes excuses à ceux que j'aurais heurté en ne les mentionnant pas ici, surtout s'ils ont eu le sentiment d'avoir participé au succès de ma thèse.

Il faut savoir que ma thèse n'a pas commencé dans des conditions idéales mais j'aimerais remercier mon premier maître de thèse qui m'a donné la chance d'intégrer la 'MetaTeam' (G3). Je regrette de ne pas lui avoir donné envie de s'investir dans ce projet commun. Je voudrais remercier Pierre et Lucie qui m'ont grandement soutenu et amené à faire évoluer ma situation lorsque je cherchais les raisons de ma présence dans le laboratoire. Mille mercis à vous deux ! Cher Pierre, j'espère que ta bonne humeur, tellement rafraîchissante durant les journées moroses, sera éternelle.

Comme vous l'aurez compris, j'ai été amenée à changer de maître de thèse. Sophie. Je te dois vraiment le succès de cette thèse pour laquelle tu as dû passer ton HDR dans un délai relativement court. J'espère que tu superviseras encore de nombreu(ses)x doctorant(e)s car ta générosité et gentillesse te rendent accessible pour des échanges scientifiques et personnelles, ce qui pour moi constituent le socle d'une recherche scientifique fertile. Je voudrais aussi remercier Yves pour nos conversations pas toujours joviales (surtout lors de ma première année) mais aussi pour les conversations scientifiques et culturelles (non je ne compte pas atteindre ce fameux niveau d'incompétence ! cf Le principe de Peter). Malheureusement, je ne pourrais pas adresser un mot à chacun des membres de l'équipe G3 mais sachez que je garderais un très bon souvenir de chacun d'entre vous. Je n'oublie pas l'équipe de modélisation et j'espère que Jean-Pierre notera que cette lettre de remerciements est écrite en français. Christine, je tiens à te dire que tu m'auras fait apprécier les innombrables perspectives des mathématiques et ça n'a pas de prix pour une biologiste.

Maintenant, je vous emmène à Paris plus précisément à Gif-sur Yvette où j'ai rencontré Michel, Mélisande et Thierry avec qui j'ai énormément appris sur la protéomique. Merci pour votre aide, votre patience et votre pédagogie.

Constance, Jiao-Jiao, Alice, Paul et tous les autres thésards, je vous remercie pour votre soutien avant et pendant la rédaction du manuscrit. Je vous souhaite bonne chance et je pense fort à vous pour la fin de votre thèse même si je suis convaincue que vous deviendrez de brillant(e)s docteur(e)s. Constance je tiens à te remercier particulièrement et te dire que tes qualités (ta ténacité, ton courage et positivisme) m'ont beaucoup inspiré. Un merci à mes amis bordelais : Mikolaï, Romain, Margaux et tous les autres.

Cette lettre de remerciements resterait incomplète si elle ne faisait pas apparaître le nom de Matthieu Raoux. Tu auras été la personne qui m'a donné envie de faire de la recherche alors Merci pour tout !!!

Pour finir, je voudrais remercier ma famille car qu'elle soit près ou loin, elle m'a toujours soutenu. Par ailleurs, les mots n'existent pas pour exprimer la gratitude et la reconnaissance que je porte aux deux rayons de soleil de ma vie. J'espère que vous vous reconnaîtrez et que vous ressentirez tous les bons sentiments que je vous porte.

Content

List of Figures and Tables

Figures

Introduction

Chapter 1

Chapter 2

Chapter 3

Tables

Materials and methods

Abbreviation list

1,3BPG	1,3-bisphosphoglycerate
2OG	2-oxoglutarate
2PGA	2-phosphoglycerate
3PGA	3-phosphoglycerate
Ac CoA	Acetyl coenzyme A
Acid Inv	Acid invertase
ADPG	Adenosine diphosphate glucose
AGPase	ADP-glucose pyrophosphorylase
AlaAT	Alanine aminotransferase
AMP	Adenosine monophosphate
APEX	Absolute protein expression
AspAT	Aspartate aminotransferase
AU	Arbitrary unit
cFBPPase	Cytosolic Fructose-1,6-bisphosphatase
CS	Citrate synthase
CV	Coefficient of variation
DHAP	Dihydroxyacetone phosphate
DPA	Days post anthesis
F6P	Fructose-6-phosphate
FBP	Fructose 1,6-bisphosphate
FDR	False discovery rate
FK	Fructokinase
G1P	Glucose-1-phosphate
G6P	Glucose-6-phosphate
G6PDH	Glucose-6-phosphate dehydrogenase
GABA	Gamma-Aminobutyric acid
GAP	Glyceraldehyde 3-phosphate
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GK	Glucokinase
GluDH/GDH	Glutamate dehydrogenase
HPLC	High performance liquid chromatography
iBAQ	Intensity based absolute quantification
ICAT	Isotope coding affinity tags
ICPL	Isotope coded protein labeling
IDH	Isocitrate dehydrogenase
iTRAQ	Isobaric tags for relative and absolute protein quantitation
kb	Kilobase
k_{dp}	Degradation rate constant
k_{sp}	Synthesis rate constant
LC	Liquid chromatography

Lg	Gene length
Lp	Protein length
m/z	Mass-to-charge ratio
MDH	Malate dehydrogenase
mRNA	messenger RNA
MS/MS	Tandem mass spectrometers/ spectrometry
MW	Molar weight
NAD	Nicotinamide adenine dinucleotide
NAD-ME	Malic enzyme
NADP	Nicotinamide adenine dinucleotide phosphate
Neutral Inv	Neutral invertase
NMR	Nuclear magnetic resonance
NOR	Nonripening
NSAF	Normalized spectral abundance factor
OA	Oxaloacetate
ODE	Ordinary differential equation
OPP cycle	Oxidative pentose phosphate cycle
PAF	Protein abundance factor
PCA	Principal component analysis
PEP	Phosphoenolpyruvate
PEPC	Phosphoenolpyruvate carboxylase
pFBPase	Plastidial Fructose-1,6-bisphosphatase
PFK	Phosphofructokinase
PFP	Pyrophosphate phosphofructokinase
PGI	Phosphoglucose isomerase
PGK	Phosphoglycerokinase
PGM	Phosphoglucomutase
PK	Pyruvate kinase
PSY1	Phytoene synthase 1
qRT-PCR	Quantitative reverse transcription polymerase chain reation
R5P	Ribose 5-phosphate
RIN	Ripening inhibitor
RNA-Seq	RNA sequencing
RT	Retention time
Ru5P	Ribulose-5-phosphate
S7P	Sedoheptulose 7-phosphate
SBP	Sedoheptulose-1,7-bisphosphate
SC	spectral count
SILAC	Stable isotope labeling by amino acids in cell culture
SPS	Sucrose phosphate synthase
Succ-CoA ligase	Succinyl-CoA ligase
SuSy	Sucrose synthase
TASEP	Totally asymmetric simple exclusion process

TCA cycle	Tricarboxylic acid cycle
TOP3	Averaged intensities of the three most intenses peptides belonging to a protein
TPI	Triose-phosphate isomerase
UDP	Uridine diphosphate
UDPG	Uridine diphosphate glucose
UGPase	UDP-Glc pyrophosphorylase
UPS	Universal proteomics standard
Vmax	Enzyme capacity
w	Weight
X5P	D-Xylulose 5-phosphate
XIC	Extracted ion chromatogram
μ(t)	Relative growth rate

Introduction

I. Tomato fruit

1.1 An experimental model for fleshy fruits

Tomato, which originates from Central and South America (Andes), has become one of the most produced and consumed fruits world-wide. Tomato (*Solanum lycopersicum*) belongs to the Solanum genius, which includes species such as potato (*Solanum tuberosum*) and eggplant (*Solanum melongena*). There are a large number of cultivars (Moneymaker, Yellow pear…) that are distinguished by their size, color or shape. The tomato fruit production represents billions of euros yearly with more than 38 million metric tons harvested in 2017. Apart from the economical aspect, tomato fruit possesses nutritional benefits for human health. Indeed, epidemiological studies show an association between the decrease of chronic diseases and the consumption of vegetables and fruit such as tomato. Furthermore, the antioxidants present in tomato fruit such as ascorbate, carotenoid and lycopene are involved in the reduction of the oxidative stress and the cancer risk.

The interest of the tomato fruit has spread in plant science where it is used as the model for fleshy fruits. The advantages of tomato include (1) relative ease of culture, (2) short generation times, (3) a diploid genome of relatively small size and (4) good tolerance to interspecific crosses, inbreeding and transformation. Moreover, in 2012 the tomato (Heinz 1706) genome was sequenced (Sato et al., 2012) identifying more than 33 000 protein-coding genes. A vast amount of resources, such as genome sequences, genotypes and other biological data (phenotypic, molecular and biochemical data), acquired on tomato plant became increasingly available, publicly accessible databases have been implemented for their repository (Mueller and Fernandez-Pozo, 2016).

1.2 Tomato growth physiology

The tomato plant is an herbaceous plant with a vegetative and a reproduction organ. The vegetative part of the tomato plant can have a determinate or an indeterminate growth.

Indeterminate plants grow vertically like vines (Figure 1A) while determinate plants become bushy (Figure 1B).

Indeterminate tomatoes, which are usually grown in greenhouses, are used to provide fruits ready-to-eat while determinate ones, grown in open fields and mechanically harvested, are used for processed products. The tomato plant possessing hermaphrodite reproductive organs –yellow flower- offsprings can resulted from a self-fertilization (plant A x plant A) and cross-fertilization (plant A x plant B).

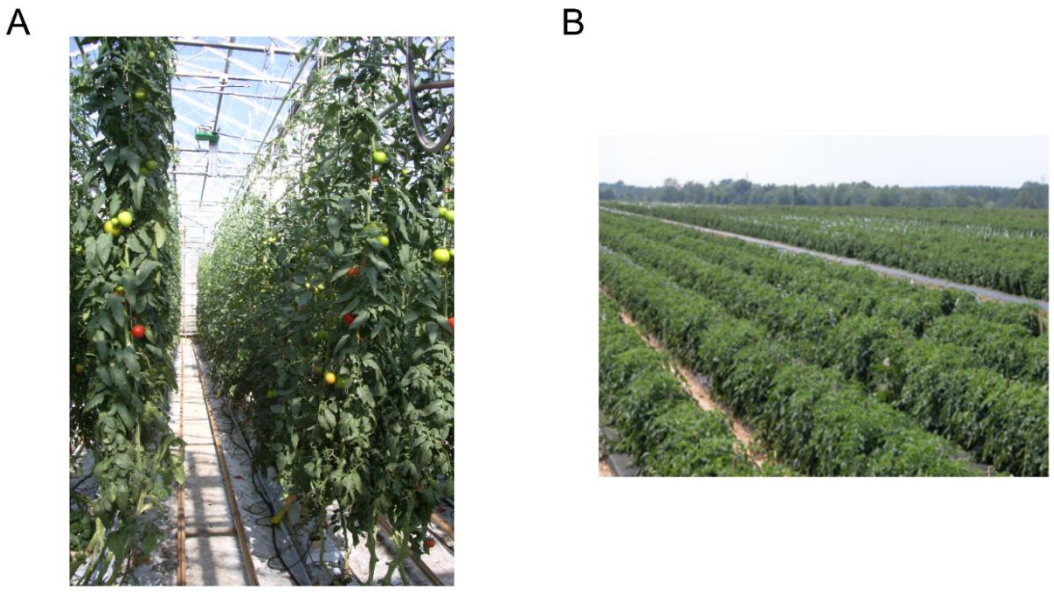


Figure 1 Photography of tomato plants with an indeterminate (A) and determinate (B) growth of the vegetative part during the development of the tomato fruit changes in size and color occurred (Figure 2).

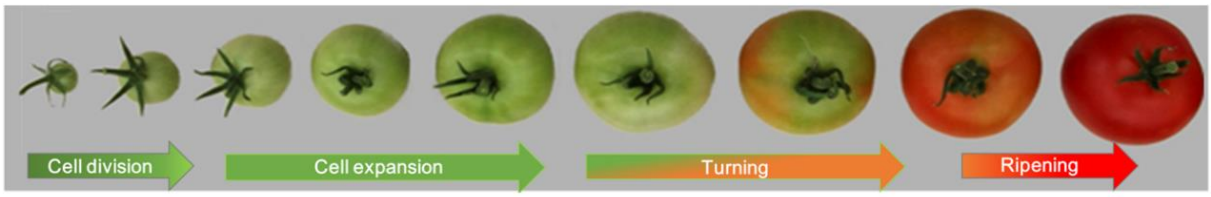


Figure 2 Time-series changes of size and color throughout tomato fruit development (Solanum lycopersicum cv. Moneymaker). Under fruits are mentioned the biological phase associated to the periods of development.

Classically, the growth of the tomato fruit is divided into three biological phases: cell division, cell expansion, and ripening which includes phases such as, mature green, breaker, and turning (Figure 2) (Gillapsi et al, 1993). The cell division stage is marked by an increase of the fruit size resulting from an intense mitosis activity, which leads to an increase of the cell number. During

the expansion phase, the volume of cells increases. In parallel, the DNA content per cell (polyploidy) changes according to the fruit age and the tissue (Bergervoet et al., 1996; Cheniclet, 2005). During the ripening phase, the tomato fruit switches from light green to orange color. In this period, chloroplasts containing chlorophyll responsible for the green color are dismantled and turn into chromoplasts. Chromoplasts conferring the red color to the fruit by the accumulation of lycopene and carotenoids (Marano et al., 1993; Carrillo-López and Yahia, 2012). The ripening phase is also marked by a change of the flavor, texture and aroma of the tomato fruit to ensure seed dispersal by its consumption.

1.3 Changes in the primary metabolism during tomato fruit development

The primary metabolism comprises all the pathways required for the plant's survival and primary metabolites are directly involved in both plant and fruit growth and maintenance while secondary metabolites are useful in long-term such as in plant defense mechanism. The metabolic composition cannot be generalized because it varies according to: (1) the genome (Robinson et al., 1988), which controls all features of the metabolic pathways, (2) the fruit age and (3) the environmental conditions (Biais et al., 2014, Yin et al., 2010). But, from a topological point of view, primary metabolism is very conserved between organs, stages of development, cell types and even between species. It is the way it is operated that makes the difference (as we reported in review submitted and given in Annex p). The main pathways, important for both the growth and quality, include, of course, central carbon metabolism, amino acid metabolism, primary cell wall metabolism and redox metabolism.

Central carbon metabolism which in fruits involves the pathways of sucrose, starch, major organic acids and respiration, provides energy and biosynthetic precursors to support fruit growth and ripening. It is also worth mentioning that most developing fleshy fruits are photosynthetic (Marcelis and Hofman-Eijer, 1995), but it is admitted they are not self-sufficient regarding carbon supply (Lytovchenko et al., 2011). Central carbon metabolism is essential for fruit quality. Indeed, sugars and organic acids, which are among the major components of most fruits, have a strong influence on fruit taste. Especially the ratio between sugars and acids is also very important for

taste. It is remarkable that tomato fruits (Causse et al., 2004) do not taste sweet although they have a relatively high sugar content of about 4%. Taste development occurring at ripening is due to increased sweetness, which is the result of a range of dramatic metabolic adjustments (Bonghi and Manganaris, 2012). These metabolic adjustment varying between tomato varieties lead to different metabolic composition explaining their different organoleptic properties (Carli et al., 2011). Starch, which in many species accumulates at high levels during fruit development, is also thought to make a major contribution to the respiration climacteric (Colombié et al., 2017). Climacteric fruits, such as apple, banana, apricot and tomato, need an increase of ethylene production and a rise of cellular respiration to ripen.

Amino acid metabolism provides precursors for protein synthesis but also for a range of secondary metabolites (Gonda et al., 2010). Major amino acids and their derivatives can have an important influence on fruit taste and quality. For example in tomato, the accumulation of large amounts of glutamate and aspartate during ripening determines the so-called umami taste, whereas gamma-Aminobutyric acid (GABA), which also accumulates at relatively high levels in growing tomato fruits, may provide interesting nutritional properties (Takayama and Ezura, 2015). Although nitrate and ammonium can be found in fruits (Sanchez et al., 2017; Horchani et al., 2008), it is generally considered that fruits do not assimilate nitrogen themselves but import amino acids from the phloem and to a lesser extent the xylem (Gourieroux et al., 2016). Similarly to the import of sugars, amino acids can take both the symplastic and apoplastic ways (Zhang X-Y et al., 2004).

Primary cell wall metabolism also belongs to primary metabolism if we consider that plant cells cannot grow or even survive without a wall. Cell wall composition is highly diverse among plant species, but the major components (cellulose, three matrix glycans composed of neutral sugars, three pectins rich in D-galacturonic acid) are usually the same (Brummell and Harpster, 2001). Cell walls are particularly important in fruits: during growth they play a major role in shaping and protecting the fruit, and imply a finely tuned trade-off with sugar metabolism while ripening is characterised by cell wall softening, a process with strong implications for fruit quality but also for shelf-life (Brummell and Harpster, 2001). Additionally, partial cell wall degradation at ripening represents a massive release of carbohydrates into central metabolism, thus providing energy and building blocks for a range of processes (e.g. protein synthesis and sugar accumulation). The cell

wall degradation is likely to make a substantial contribution to the respiration burst occuring just before ripening, 40 days after pollination in tomato (var. Moneymaker) (Colombié et al., 2017).

Redox metabolism, especially ascorbate metabolism, also connected to cell wall metabolism (Voxeur et al., 2011), represents a further important aspect of fruit metabolism. Enzyme activities, which regulate the metabolite synthesis or degradation, are also markers of the tomato fruit development (Biais et al., 2014; Steinhauser et al., 2010). Indeed, 36 enzyme activities involved in the primary metabolism, when expressed on mass of protein basis, marked the developmental stages of tomato fruit (var. Moneymaker). For instance, earliest stages were characterized by high activities of fructokinase and glucokinase, pyruvate kinase and TCA cycle enzyme, indicating a high requirement of ATP during this period. The cell expansion was more related to starch synthesis (AGPase) and involving enzymes of the Calvin-Benson cycle while enzymes of last stages were associated to metabolites accumulation, such as citrate synthase and citrate (Biais et al., 2014) (Figure 3, *Solanum lycopersicum* var. Moneymaker).

A number of studies have focused on the changes in metabolic composition of tomato fruits throughout their development and ripening. For instance, it has been found that the young fruits are characterized by highest concentrations of hexose phosphates while several amino acids and major hexoses (glucose, fructose) increase at ripening (Carrari and Fernie, 2006).

The generation of mutants and transgenic plants has allowed the identification of triggers of the tomato development. For instance, mutations of transcription factors (*RIPENING-INHIBITOR* MADS-box, *COLOURLESS NON-RIPENING* SBP-box) and ethylene receptor (*Never-ripe*) genes affecting ethylene synthesis and perception has allowed a better understanding of how ethylene participates in fruit ripening (Osorio et al., 2011 ; Giovannoni, 2007).

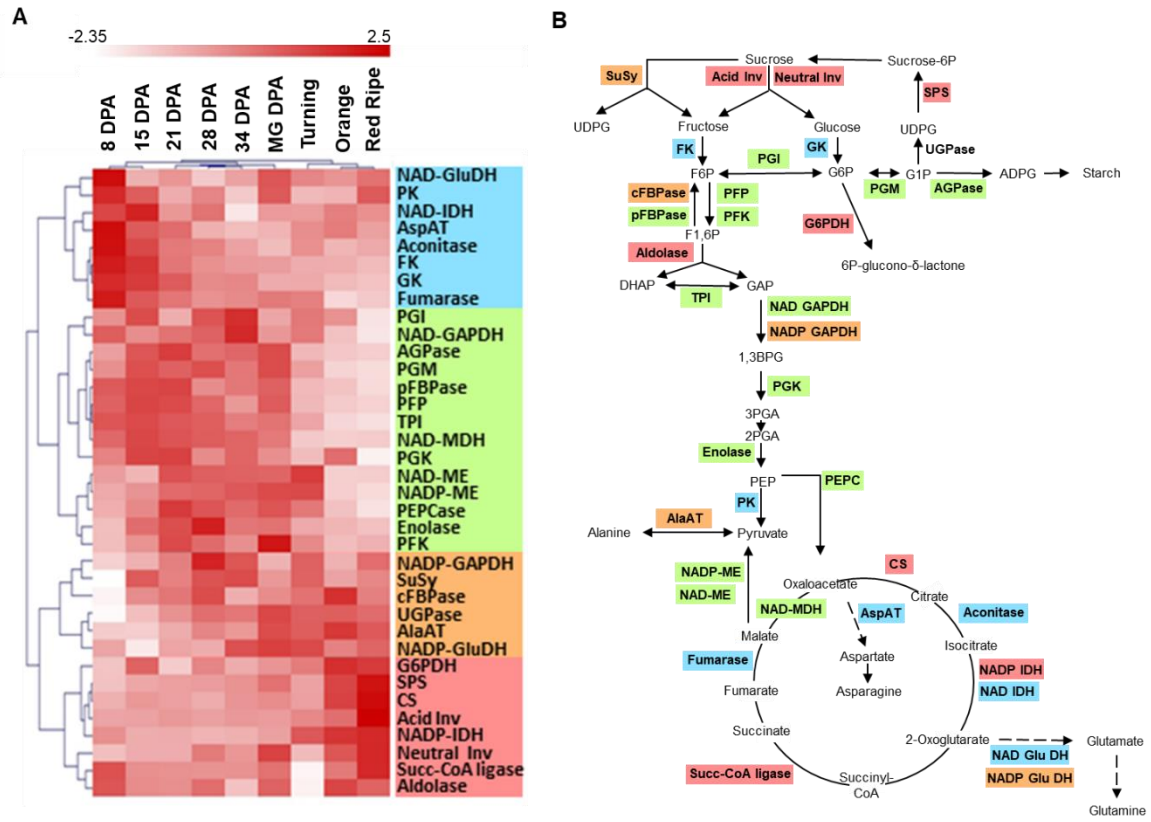


Figure 3 Hierarchical clustering analysis of 36 enzyme activity profiles throughout development of the tomato fruit (var. Moneymaker) from Biais et al., (2014). (A) The clustering analysis was performed on activities expressed on a protein basis by Pearson's correlation, mean centered, and scaled to unit data. The clustering analysis performed on activities separated enzymes in four clusters that are highlighted with a colored bar on the right of the heatmap. (B) Simplified drawing of central metabolism in plant. The color code corresponds to the clusters selected in A. Blue, activities highest during cell division and beginning of cell expansion; green, activities highest during cell expansion; orange, activity peaking at late expansion; red, activities highest at ripening.

However, the use of mutants to characterize a metabolic pathway assumed that (1) the candidate gene is directly involved in the targeted metabolism, that (2) all others mutations experimentally introduced are detected and not involved in the mutant phenotype, and that (3) the cell doesn't compensate the mutated genes by over or down-regulated others genes and thus altering the metabolic phenotype.

A complementary strategy to identify triggers genes has emerged and is based on the acquisition and integration of "omics" data such as transcriptomic, proteomic, activome, and metabolomics.

II. Omics data and fruit development

Omics designates data obtained from high-density technologies. There are genomics and transcriptomics which correspond to the study of genomes and gene expression, respectively. Then, both were further completed by proteomics and metabolomics – i.e. the study of cells' protein and metabolites, respectively. In fleshy fruits, a range of studies have dealt with transcriptomic, proteomic and metabolomic and more recently "activomics" (enzyme activity profiling) and fluxomics have emerged. One objective of such multiomics approaches is to perform integrative analyses in order to generate knowledge about interactions between biomolecular levels (Figure 4), identify candidate genes and biomarkers (developmental, pathological, and environmental). Moreover, omics data represent a real benefit for systems biology, which uses multivariate statistics and/or mathematical modelling to study biological systems in a holistic way.

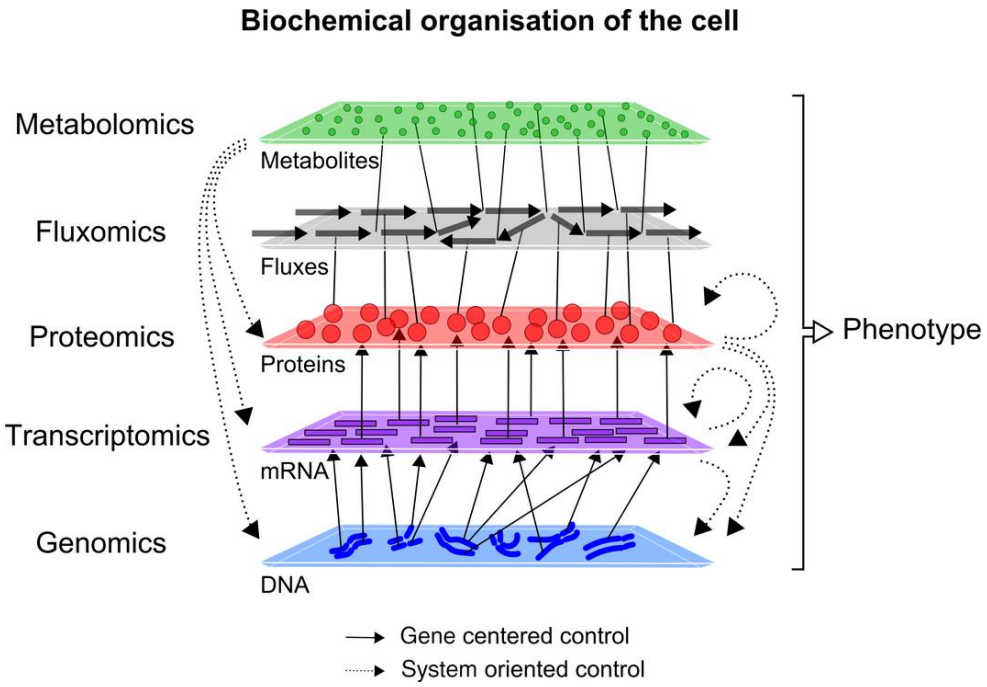


Figure 4 The biochemical nature of the cell components drives cell organization and thus methods used to analyze them. The cell organization links genes to transcripts to proteins to fluxes of metabolites and backward regulation of the gene expression via biochemical relationships. Figure is from Peyraud et al (2017).

2.1 Proteomics by LC-MS/MS

Proteomics analysis aims to collect proteins data in a largest-scale as possible to get a fingerprint of the biological system. Proteomics covers a wide range of applications such as protein structure determination, protein-protein interactions studies, studies of proteome responses to environmental variations (biotic of abiotic stresses) or genetic perturbations (*i.e.* mutations), and studies of proteome evolution in time-series.

2.1.1 Protein abundance by LC-MS/MS: principle

All proteomics studies start by protein extraction, using adapted protocols according to the organism, the tissue and also to the targeted proteome (post-translational modified proteome, cell wall proteome, sub-cellular proteome...). In the 'bottom-up' proteomics strategy, extraction is followed by protease digestion (typically trypsin). Then, the resulting peptides are separated according to hydrophobicity by liquid chromatography (LC), ionized and analysed by mass spectrometry. In proteomics, LC is usually coupled to tandem mass spectrometers (LC-MS/MS), allowing two levels of analyses called MS1 and MS2. At the first level, the mass-to-charge ratios (m/z) and intensities of the ionized peptides that entered the mass spectrometer at a given retention time are measured (Figure 5). The most intense of these ionized peptides is selected and fragmented in a collision cell. At the second level, the m/z and intensities of the product ions resulting from fragmentation are measured. These two cycles are repeated all along the chromatography (Figure 5). Together with the retention time information, the data collected from MS1 allow to produce the elution profiles of ionized peptides in what is called extracted ion chromatograms (XIC). The data obtained from MS2 are used to build fragmentation spectra (or MS2 spectra) which subsequently allow to identify peptides and proteins by comparison to theoretical spectra produced in silico from protein sequence databases. Peptide abundances, computed with or without the use of stable isotope labels, are used to infer protein abundances.

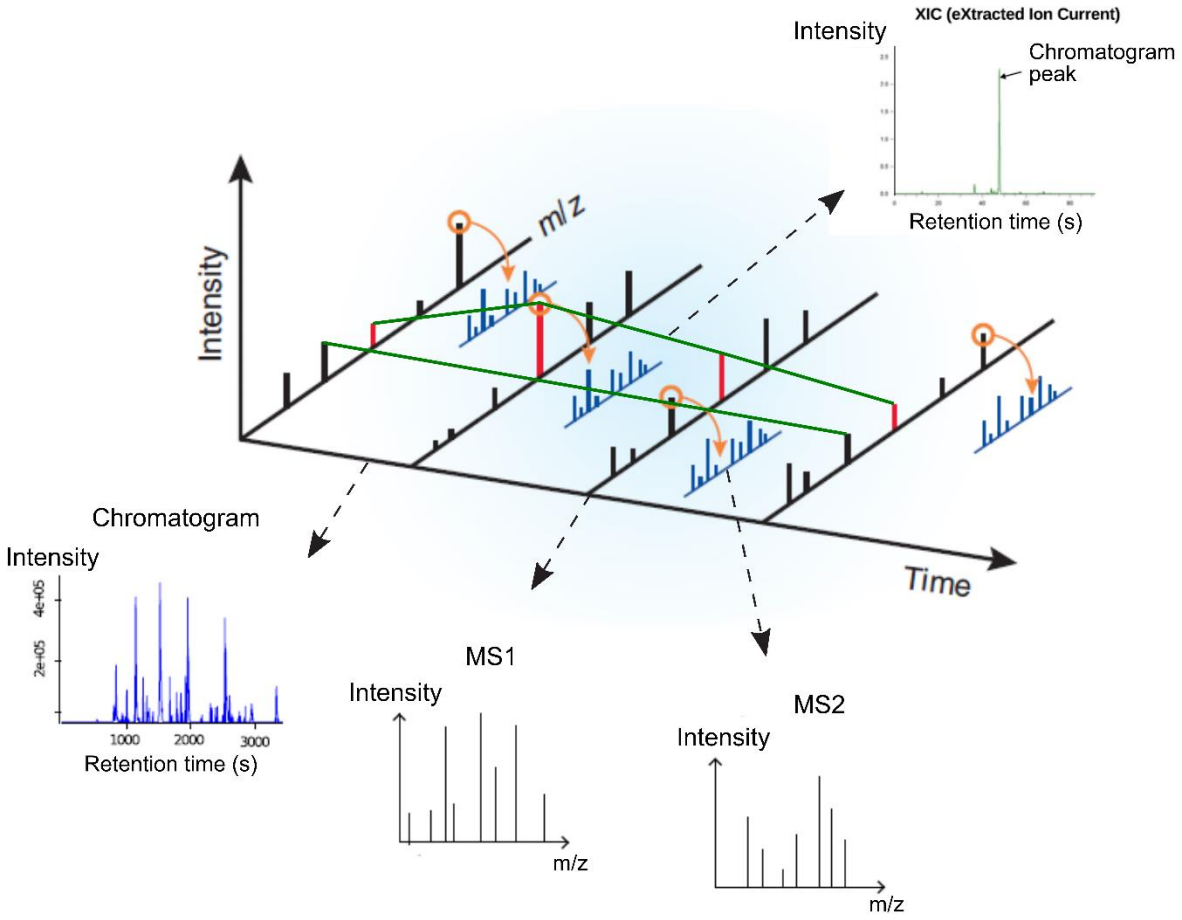


Figure 5 Data acquisition by tandem mass spectrometry (MS/MS) from McIntosh and Fitzgibbon, 2009. In MS/MS, the instrument periodically scans the mass-to-charge ratio (m/z) of eluting peptides (black peaks, MS1 scan), selects the most abundant at that time point. The selected peptides are fragmented in a collision cell in to fragment ions for which the m/z are measured (blue peaks, MS2 scan).

2.1.2 Labelling and label-free LC-MS/MS techniques

Gel-based techniques, which separate tryptic peptides using gel electrophoresis, has been successfully used in plant field (Schenkluhn et al., 2010; Sergeant et al., 2011; Suzuki et al., 2015). However in this section only gel-free techniques, which separates peptides only by LC, are considered. Two main techniques can be used for protein quantification: techniques with stable isotope or techniques in label-free. An overview of workflows using both techniques is presented in Figure 6.

a Techniques with stable-isotope labelling

Label-based techniques (Figure 6, a to f) with specific workflows allowed evaluating protein changes between two conditions by comparing labelled (condition 1) and an unlabeled (condition 2) peptides. Indeed, labelled and unlabeled peptides are identified at the same retention time but distinguishable only by a shift of m/z induced by the heavy isotopes.

Label-based techniques distinguish *in vivo* and *in vitro* labelling. In *in vivo* labelling techniques, named metabolic labelling, heavy isotopes (^{13}C, ^{15}N, or ^{18}O) are introduced in the environment of the organism and metabolized by the organism into proteins (Figure 6 a). Metabolic labelling based on ^{15}N was used to investigate the uptake and a heterogeneous distribution of nitrogen in the different plant organs such as rice plant (Oryza sativa) (Muhammad and Kumazawa, 1974), fully labeled potato plant (Solanum tuberosum) (Ippel et al., 2004) and tomato plant (Schaff et al., 2008). Metabolic labelling has also been used to determine the degradation rate of more than one thousand Arabidopsis proteins (Li et al., 2017a). Another *in vivo* technique, using labelled amino acids instead of heavy isotopes named SILAC (Ong et al., 2002) is usually used to label proteins on cell-culture. Few years ago, the SILAC protocol was efficiently adapted for Arabidopsis seedlings and lead to the identification of 215 proteins changed by a salt stress (Lewandowska et al., 2013).

In the case of *in vitro* methods (Figure 6, b to f), heavy isotopes are incorporated into peptides after the total protein extraction. This method can be done chemically or enzymatically (iTRAQ, (Ross et al., 2004)) (Figure 6 e), by cleavage in ^{18}O water (Mirza et al., 2008)(Figure 6 d) or to intact proteins (ICAT (Gygi et al., 1999), ICPL (Schmidt et al., 2005)) (Figure 6 b, c). *In vitro*

methods allowed the identification of 111 proteins up and down regulated during ripening of two strawberry varieties.

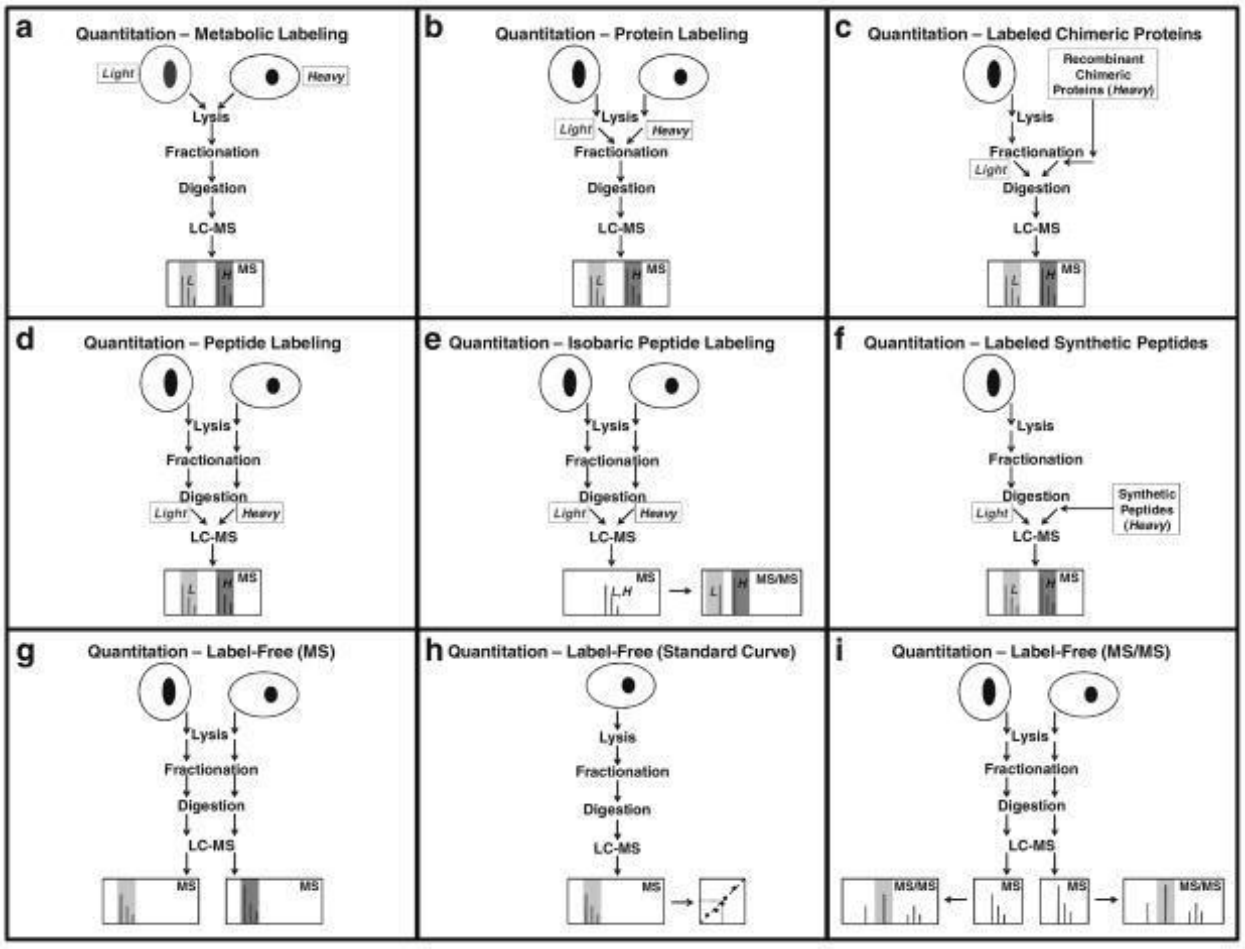


Figure 6 Workflows for mass spectrometry-based protein and peptide quantitation: (a) metabolic labeling, (b) protein labeling, (c) chimeric recombinant protein labeling, (d) peptide labeling, (e) isobaric peptide labeling, (f) synthetic peptide labeling (6), label-free quantitation using the intensity of precursor ions, (h) label-free quantitation using the intensity of precursor ions and a standard curve and (i) label-free quantitation using the intensity of fragment ions.

In parallel, label-based technique can be used for absolute quantification of proteins using known concentrations of internal standards, such as isotopically labeled synthetic peptides known as AQUA peptides (Gerber et al., 2003), similar to the targeted proteins (Figure 6 f). Artificial proteins, corresponding to signature peptides concatenated, can also be used (Beynon et al., 2005) (Figure 6 c).

Label-based techniques allow accurate, precise protein quantification but remain costly which can hamper large-scale applications. In contrast, label-free techniques are applicable to quantify a large number of proteins in any sample type at high-throughput and with minimized cost. For instance, recently, Szymanski et al., (2017) quantified by label-free LC-MS/MS more than seven thousand proteins at five developmental stages of tomato fruit (*S. lycopersicum* cv. MicroTom).

b Label-free techniques

Label-free techniques (Figure 6 g to i) have two main benefits, one related to the other. First, as none heavy isotopes are required the cost of each analysis is reduced. Second, the cost reduction per sample and their separate analysis allow to do more complex experiments. However, as sample are separately analyzed by LC-MS/MS, a high reproducibility (protein extraction, peptide ionization) between samples is required.

With label-free technique, quantification methods of protein are based either on the number of MS2 spectra, called Spectral Count (SC), or by the integration of peptides peak area called XIC.

Spectral count (SC) protein quantification is based on the number of MS2 spectra assigned to one protein. This method of protein quantification was developed more than ten years ago by Liu et al., (2004) who shown a correlation between SC and protein abundances but also between SC and protein molecular masses. Indeed, the bigger the protein, the more the numbre of spectra will be important because of highest chance to be cleaved during trypsin digestion. To consider this limitation and others resulting from the protein biochemical properties (length, sequence, peptides ionization and MS detectability), different normalizations of SC data were developed, such as the normalization by the protein length (NSAF, Zybailov et al., 2006) and the molecular mass (PAF; Powell et al., 2004). A more sophisticated method named APEX (Lu et al., 2007) predicts the number of tryptic peptides per protein and compared to experimental data to estimate the protein abundance. APEX-based protein abundance was successfully applied to proteomes of Arabidopsis (Baerenfaller et al., 2008) and rice (Laurent et al., 2010).

The popularity of using the spectral count approaches (PAF, NSAF, APEX) to get an absolute protein quantification delayed on the easiness to collect SC data, the high reproducibility. Old et al., (2005) found that due to their discrete nature, the spectral counting was more sensitive to detect

changes of proteins abundance while XIC-based methods determined more accurately protein fold-change.

- XIC-based protein quantification

Peptide abundance quantified by XIC correspond to the integrated peak area of the ion extracted chromatogram. The extracted ion being characterized by an m/z ratio and a retention time. XIC-based quantification required pre-processing (removing shared peptides, normalization of peptides intensity) to compute protein abundance.

More than ten methods have been developed to infer protein abundance from peptides intensity signal. Only some of them are mentioned here. For instance Silva et al., (2006) quantified protein by the average intensities of the three most intense peptides (TOP3) belonging to the protein. Another method consists to average all peptides intensity belonging to the protein (Higgs et al., 2005). More generally, protein abundance is quantified by the sum of all peptide ion intensities (Ning et al., 2012), such as for the iBAQ method (Schwanhäusser et al., 2011) which summed all peptides intensity and normalized by the theoretical number of tryptic peptides. Finally, the quantification based on peptides intensity statistical modelling has emerged and is now considered as the most adequate method to quantitatively compare protein abundances (Clough et al., 2009). Indeed, statistical modelling consider potential bias that might be introduced at different levels of the experiment (treatment, sampling...) in the quantification of protein from peptides intensities.

To conclude this chapter about proteomics by LC-MS/MS, we confirmed that the increasing need to conduct absolute quantification studies participate actively to the development of accurate methods of quantification, especially those based on XIC. Actually, the proteomics data add a new dimension to the existing genomic, transcriptomic and metabolomic resources and offer the opportunity to integrate several omics.

2.2 Integrative analysis of omics

Proteomics analysis represents one way to study plant model responses to changes. To obtain a more complete overview, proteomics has been integrated with others omics, such as genomics, transcriptomics and metabolomics.

One purpose of omics integrative analysis is to search for candidate genes, i.e. genes potentially involved in specific metabolism pathway, by performing large-scale correlative studies to identify relation between candidate genes expression and a trait, such as metabolites content (Usadel et al., 2009; Toubiana et al., 2013). Another purpose is to explain the reprogramming of the primary and specialized metabolism with the others biomolecular levels (Mounet et al., 2009; Bastías et al., 2014; Wong and Matus, 2017).

The combination of at least two omics has been used for the characterization of metabolic shifts during development in a range of fruit species including tomato (Osorio et al., 2011), grape berry (Dai et al., 2013), apple (Li et al., 2016 and www.transcrapple.com), melon (Guo et al., 2017) and mango (Wu et al., 2014). Omics have also been used to evaluate environmental effects on metabolism in tomato (D'Esposito et al., 2017), abiotic stress like water stress or biotic stresses induced by botrytis infection in grape berry (Agudelo-Romero et al., 2015; Ghan et al., 2015).

In apple, a comprehensive 2D gel-based proteomic analysis over five growth stages, from young fruit to maturity, coupled with targeted metabolomic profiling of soluble sugars, organic acids and amino acids provided insights into the metabolism and storage of fructose, sucrose and malate (Li et al., 2016). This analysis suggests that the decrease in amino acid concentrations during fruit development is related to a reduction in substrate flux via glycolysis. In citrus, integration of LC-MS/MS-based proteomic and metabolomic analyses showed that at the end of citrus development organic acid and amino acid accumulation shifted toward sugar synthesis and that may involve an invertase inhibitor (Katz et al., 2011). In grape exocarp, trends between metabolites and proteins revealed clear links between primary and specialized metabolism (Negri et al., 2015). For instance, several proteins involved in glycolysis, TCA cycle, and metabolic intermediates of these pathways showed a good association with anthocyanin content. By using label-free LC-MS/MS, Szymanski et al., (2017) have quantified more than seven thousand proteins in the skin and pericarp and at five developmental stages of tomato fruits. With their proteomic data, they cover 83% of all enzymatic reactions predicted in the metabolic network including primary metabolism as well as isoprenoid and carotenoid biosynthetic pathways. By relating abundance of enzyme protein to their activity, they found a significant tissue-specific reprogramming of the metabolism during fruit development.

Integrative analyses of three post-genomics datasets are less present in literature. Among the few examples found in literature, the integration of three omics approaches has been performed on grapes (Ghan et al., 2015) and on tomato fruit (Osorio et al., 2011).

In few cases, integrative omics have led to the identification of candidate gene in fruits. For instance, candidate genes involved in tomato fruit secondary metabolism (Tohge et al., 2014) and in peach fruit aroma volatiles (Sánchez et al., 2013a) have been found. When expressed in yeast, one of the peach candidate genes showed a substrate specificity that was similar to a desaturase, which might be involved in the production of precursors of aromatic volatiles.

Omics clearly represent a deep source of data to characterize and understand regulation of processes mainly by statistical approaches such as correlation analysis. Whereas omics data are most exclusively processed with statistical approaches, they can also be used to parameterize mathematical models describing biological processes, especially when the data are quantitative.

III. modelling from quantitative proteomics and transcriptomics

Recent technological advances, in particular in mass spectrometry have allowed for large-scale surveys of the proteome. Proteomics has now sufficiently advanced to obtain, an absolute abundance quantification of thousands of proteins and to complete transcriptomics approach. Globally, these large-scale studies are changing our understanding of protein-expression regulation.

Proteins are fundamental components in living cells by their structural and catalytic activity. The protein content in cells results from the equilibrium of diverse processes such as: mRNA synthesis, mRNA translation, post-translational modification and protein degradation. In eukaryotes, protein and mRNA concentrations in the cell are usually positively correlated, which suggest that the variation in protein concentration can be partially explained by the variation of the corresponding mRNA concentration.

To elucidate the mechanisms and functions that go beyond mRNA translation and protein synthesis, a systems-level understanding based on well-defined models is necessary.

3.1 Modelling translation, a universal process with a regulated efficiency

In all organisms, the translation is divided in successive regulated steps: initiation, elongation and termination. A transcript can be find bound to one or multiple ribosomes (polysomes). Studies on yeast and mammal cells described a distance from 200 to 300 nucleotides between two ribosomes in polysomes and a translation rate ranging from 3 to 10 amino acids per second (Figure 7). Basically, protein synthesis depends on the concentration of its corresponding mRNA, the availability in amino acids and ATP, which are required for synthesis.

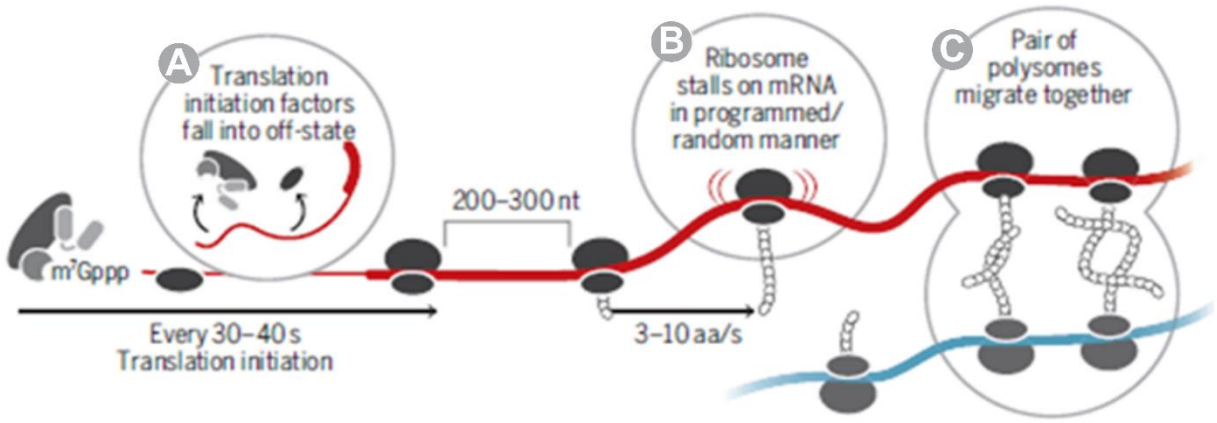


Figure 7 Translation process in vivo from Iwasaki and Ingolia 2016 modified picture. (A) Initiation. Translation typically initiates every 30 to 40 s, but this can be interrupted by a translationally silent state lasting minutes or hours. (B) Elongation. Translation usually proceeds at 3 to 10 amino acids (aa) per second, but ribosomes can stall in response to programmed signals or random events. (C) Diffusion. Most polysomes undergo free, independent diffusion, but a small fraction move together as a pair.

Translation is a complex system of biochemical reactions decoding mRNA to produce polypeptides. The complexity of this system makes it difficult to quantitatively connect its input parameters (such as translation factor or ribosome concentrations, codon composition of the mRNA, or energy availability) to output parameters (protein synthesis rates or ribosome densities on mRNAs). Since five decades, mathematical and computational models have been used to investigate translation, and to shed light on the relationship between the different reactions in the translational system (Haar, 2012). In his review, Tobias von der Haar has presented an overview of approaches, concepts and results conducted up to the current date.

The mathematical modelling of mRNA translation has a long history, and enjoys renewed interest in recent years with the development of systems and synthetic biology. Models for mRNA translation have been introduced with different formulations at various levels of abstraction, and can be divided into, roughly speaking, the Totally Asymmetric Simple Exclusion Process (TASEP) type models and the ordinary differential equations (ODEs) based models (Zhao and Krishnan, 2014).

All the TASEP models are largely based on statistical analyses of the behaviour of ribosomes on mRNA (Haar, 2012), indirectly and primarily evaluating the mRNA translation through the ribosome movement along the mRNA. This simplified transportation problem is thus modelled with TASEP to quantitatively understand the particle transport in a one-dimensional lattice. The TASEP-based models have been used for obtaining steady state information such as the average

occupancy of each codon on the mRNA, the mRNA translation rate, which are key in understanding mRNA translation.

Conversely, as mRNA translation is the outcome of several transitions, which may be conceptualized as reactions, it can be modelled with ordinary differential equations (ODEs) to directly describe mRNA translation (equivalent to protein synthesis) process in a comprehensive fashion. In that case, the rate of protein synthesis is described as a function of two main terms: (1) mRNA abundance coupled to its translation efficiency (*i.e.* the rate of mRNA translation into proteins within cells (measured in protein per mRNA per day), and (2) the protein disappearance by both the protein degradation according to the protein constant degradation (measured in protein per protein per day) and dilution of the protein abundance by growth (Dressaire et al., 2009). This simple ODE has been used to describe the ethylene biosynthesis pathway in tomato fruit from transcriptomic, proteomic and metabolic data (Van de Poel et al., 2014) and also large dataset of transcripts and proteins in yeast (Tchourine et al., 2014).

As described above, processes controlling protein synthesis and degradation are described but questions remain about their contributions to the abundance of each protein. Indeed, individual protein should be defined with degradation and synthesis rates. The synthesis rate is the rate at which the protein is produced while the degradation rate is the rate at which the protein is degraded. Both rates are expressed in time minus one.

3.2 Protein synthesis and degradation rates

The protein content of plant cells, which is constantly updated, is driven by the opposing actions of synthesis and degradation. Protein degradation is determined by the half-life of each polypeptide (Nelson and Millar, 2015).

As mentioned before, the rate of protein synthesis based on an ODE model can be schematically described as a function of two main terms: (1) mRNA abundance coupled to its translation efficiency, which regulate protein synthesis within cells, and (2) protein disappearance by both the protein degradation according to the protein turnover and dilution of the protein abundance by growth. Consequently, the protein synthesis rate is proportional to the amount of RNA and the protein degradation rate is proportional to the amount of protein.

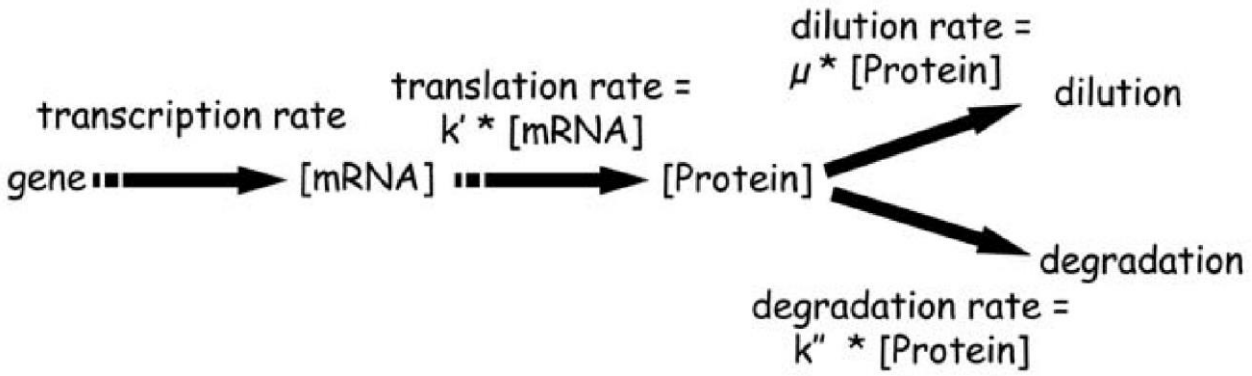


Figure 8. modelling of the cellular process from Dressaire et al. (2009). Translation, dilution and degradation rates expressed respectively by k'[mRNA], μ[protein] and k"[protein] where k' is the translation efficiency, μ the growth rate and k" the degradation rate constant.

Mathematically this concept leads to write the time-evolution of protein abundance with one simple ordinary differential equation:

$$\frac{d[protein]}{dt} = k'[mRNA] - \mu[protein] - k''[protein]$$

, where the changes of protein abundance (d[protein]/dt) results from the difference of the protein synthesis (k'[mRNA]) and the degradation rate ((μ+k")[protein]).

This model has been used to determine large datasets of protein synthesis and degradation constant rates. For instance, Dressaire et al. (2009) reported a genome-scale study analysing the various parameters influencing protein levels in bacteria cells *Lactococcus lactis* grown at different growth rates. Proteomic and transcriptomic data were thoroughly compared and modelling allowed both translation efficiencies and degradation rates to be estimated for each protein in each growth condition. These authors showed that estimated translational efficiencies and degradation rates strongly differed between proteins. Moreover, these efficiencies and degradation rates were not constant in all growth conditions and were inversely proportional to the growth rate, indicating a more efficient translation at low growth rate and also a higher rate of protein degradation. Estimated protein median half-lives of *Lactococcus lactis* bacteria cells ranged from 23 to 224 min, underlying the importance of protein degradation notably at low growth rates.

Concerning eukaryotes, two studies have been reported for yeast Tchourine et al. (2014) and more recently Lahtvee et al. (2017).

Tchourine et al. (2014) investigated the major principles of gene expression regulation in dynamic systems. They estimated protein synthesis and degradation rates from parallel time series data of mRNA and protein expression. They tested the degree to which protein expression changes can be modeled by a set of simple linear differential equations and showed that one-third of protein expression can be predicted by simple rate equations. Results showed that predictability was well determined when both protein and mRNA levels increased and unwell determined when sudden and singular shifts of expression were observed. They highlighted that the prediction quality was linked to low measurement noise and the shape of the expression profile. Finally they considered that most genes are subject to one of two major modes of regulation, which they termed synthesis- and degradation-independent regulation. These two modes, in which only one of the rates has to be tightly set while the other one can assume various values, would offer an efficient way for the cell to respond to stimuli and re-establish proteostasis.

More recently, Lahtvee et al. (2017) reported that absolute concentrations of mRNA and proteins, in combination with protein turnover measurements, give an opportunity to calculate translation efficiencies of individual proteins in yeast cultivated in ten environmental conditions. Interestingly, these authors reported (1) a 400-fold difference in translation efficiency between individual proteins and (2) a high correlation between protein and mRNA that were undergoing changes.

Finally, Schwanhaüsser et al. (2011) reported a quantitative analysis with genome-wide gene expression including the simultaneous absolute measurement of mRNAs and protein levels as well as protein turnover. These authors showed that whereas mRNA and protein levels correlated better than previously thought, no correlation between protein and mRNA half-lives was found. The quantitative model allowed genome-scale predictions of synthesis rates of mRNAs and proteins. They conclude that the cellular abundance of proteins is predominantly controlled at the level of translation. Genes with similar combinations of mRNA and protein stability shared functional properties. For instance, genes with stable mRNAs and stable proteins were associated to cellular processes like translation (that is, ribosomal proteins), respiration and central metabolism (glycolysis, citric acid cycle).

The modelling approach enabled the estimation of translational efficiencies and protein degradation rates, two biological parameters that are extremely difficult to determine

experimentally and are generally lacking. The quantitative information about all stages of gene expression provides a rich resource and helps to provide a greater understanding of the underlying organization related to the translation. While a large part of translation modelling concerns cells in culture (bacteria, yeast and mammal cells) as far as we know, there is no publication reporting similar results for plants.

Nevertheless, the degradation rate constants of proteins can be determined experimentally by labeling (Li et al., 2017b) and degradation rate constants and protein turnover measurements have been previously reported for two main plant models, barley leaves (Nelson et al., 2014) and *Arabidopsis thaliana*, in both cells (Li et al., 2012) and leaves (Ishihara et al., 2015; Li et al., 2017b).

Protein stability has been reported to play an important role in fine-tuning protein levels in cells and the enormous complexity of the shape of protein expression profiles has motivated the search for regulatory factors at the level of transcription, translation, and degradation. A way to better understand the time-dependent protein expression profiles is to search for simple relationship between the contributing protein synthesis and degradation neglecting regulatory factors. In other words, the goal is to find how many cases and what kind of protein profiles can be deduced directly from transcript profiles, without considering specific regulations (post-translational modifications such as phosphorylation, ubiquitination…).

The description of protein stability, especially when applied to enzymes, will be useful to better understand the contribution of the reprograming of metabolism to growth and further developmental events observed in plants and fruits. While in the past, modelling studies constituted a minor fraction of the enormous number of publications generated by the very active protein synthesis field, the success of Systems Biology as a new sub-discipline in life sciences has increased the trickle of modelling studies to a solid river, and it is likely that this will increase to a torrent in the not too distant future (Haar, 2012). Thus, establishing an integrated understanding of the processes that underpin changes in protein abundance under various physiological and developmental scenarios will accelerate our ability to model and rationally engineer plants (Nelson and Millar 2015).

Objectives of the PhD work

With recent technologies advances and in particular the development of 'omics techniques', especially transcriptomic (Osorio et al., 2011), proteomic (Szymanski et al., 2017) and metabolomics (Oa et al., 2009), the main cell components can now be analyzed by high-throughput. These technologies have enhanced the emergence of the systems biology research, a field that aims to understand complex interaction between the different cellular levels with computational and mathematical modelling approach (Schwanhäusser et al., 2011; Van de Poel et al., 2014).

In this context, the objective of my PhD was to perform a quantitative proteomic analysis of the tomato fruit development and then integrate quantitative omics data both by statistical analyses and by mathematical modelling.

The first chapter focused on results obtained for the quantitative proteomic developed in collaboration with the PAPPSO platform (INRA, Gif-sur-Yvette). Samples were harvested at nine stages of tomato fruit development, total proteome was extracted and quantified by label-free LC-MS/MS. Then, five methods of quantification were tested in order to select the most appropriate. In parallel, as proteome quantification based on XIC relied on the peptides quality, we tested four peptides filters to quantify their effects on the five methods performances. Finally, the method named peptides intensity modelling was used to determine the absolute quantification of 2494 proteins. The quantification of proteins by LC-MS/MS was then validated by comparison with 32 enzymatic capacities used as proxy for protein abundance. The relative accuracy of the absolute quantification provided good results.

The second chapter was dedicated to the results of integrative omics analyses throughout tomato fruit development. First, transcriptomic has been performed in collaboration with Genotoul GeT (Toulouse) and Usadel's lab (RWTH Aachen University, Germany). Using spikes in the experimental design, more than 20000 transcripts have been quantitatively determined at the nine stages of development. Then, this absolute quantification of the tomato transcriptome has been cross-validated with 71 transcripts previously measured by qRT-PCR. Finally, we integrated the four omics datasets -transcriptome, proteome, metabolome and activome obtained on the same material– in order to identify key variables of the tomato fruit development. For the four levels, analyses confirmed that the onset of ripening phase was accompanied by major changes, and

revealed a great similarity between the end and the beginning of development, especially in the energy metabolism.

The third chapter focuses on modelling results of the protein translation based on the absolute quantification of transcriptomic and proteomic. To explain the decreasing correlation observed between proteins and transcripts concentration throughout development, we proposed a mathematical model of protein translation based on an ordinary differential equation and involving two rate constants (for synthesis and degradation of the protein). To our knowledge this is the first time that translation model is applied to the tomato fruit. The resolution of this equation, validated by a quality criterion based on a closed confidence interval, led to the estimation of the rate constants for more than 1000 proteins. These results were then compared with previous published data reported for plants and more widely in eukaryotic cells. Results revealing that degradation rate constant obtained on tomato fruit were more similar to degradation rate constant obtained on plants (Arabidopsis, Barley) than yeast and mammals cells.

Chapter 1 Quantitative proteomics analysis of tomato fruit

The main objective of this section was to describe a time-series dataset of quantitative proteomics obtained throughout tomato fruit development. As none consensus emerged about methods and peptides filtering required before quantification of protein abundance, we first analyzed a yeast dataset available at PAPSO in order to evaluate the precision and accuracy limits of quantification methods associated to filtering. This work led to a paper in preparation (see Annex p) and was summarized here.

I. Evaluation of the precision and accuracy limits of different protein quantification methods

In the first section (Introduction), several methods allowing quantification of proteins by LC-MS/MS from the signal intensities of peptides were presented. These XIC-based quantification methods are used to estimate protein abundance while it is known that all peptides are not equivalent for integrity, identification and detection. Thus, to quantify proteins we have to consider these differences to analyze and compare protein abundance. As none consensus has emerged about which method and especially which peptides to use to get the most accurate quantification, we evaluated five methods of protein quantification with four filters.

The five peptide datasets included the initial peptide dataset plus four peptide datasets filtered. These filters were selected according to their capabilities in removing peptides biasing protein quantification. First, the shared peptide filter which removed peptides that are generally discarded because of the difficulty to properly deconvolve the information they carry. Second, the retention time (RT) filter, which aims to remove peptide ions showing highly variable RT potentially arising from mis-identifications. Third, the occurrence filter, which aims to remove peptide ions exhibiting many missing values, i.e peptides which are not detected in more than a threshold number of samples. Rarely observed peptide ions are indeed inadequate for statistical analysis (Webb-Robertson et al., 2010). Generally, a threshold is arbitrary chosen, e.g. a peptide ion should be observed in at least three injections (Xianyin Lai, Lianshui Wang, Haixu Tang, 2011). Fourth, the outliers filter, which aims to exclude peptide ions showing inconsistent intensity profiles.

Five XIC-based quantification methods were analyzed: (1) iBAQ (Schwanhäusser et al., 2011): the sum of peptide ion intensities was divided by the theoretical number of tryptic peptides; (2) TOP3 (Silva et al., 2006): the three most intense peptide ions in median were selected and their mean intensity was computed; (3) Average (Higgs et al., 2005): the mean of all peptide ion intensities was computed, (4) Average Log: peptide ion intensities were $\log_{10}$-transformed before their mean was computed, (5) Model (Blein-Nicolas et al., 2012): $\log_{10}$-transformed intensities were first modeled using a mixed effect model derived from Blein-Nicolas et al. (2012).

Peptides filtering effects and the five methods were evaluated through three criteria: the precision, the absolute and relative accuracy of the quantification of UPS1 proteins abundance obtained by each method. UPS1 proteins – an equimolar mix of 41 human source proteins- were spiked in a yeast proteins background at eleven concentrations (0.04, 0.09, 0.2, 0.5, 1.1, 2.2, 5.5, 12.4, 27.9, 62.8, 141.1 fmol.μl^{-1}). The serial concentration was performed in triplicates resulting to a 33 samples experiment.

The precision was determined by the coefficients of variation (CV) of UPS1 proteins across technical replicates. The lower the CV, the higher the precision. The relative accuracy was estimated by the coefficient of determination (R^2) and the slope of the linear regression between the abundances obtained experimentally for UPS1 proteins and their spiked concentrations while the absolute accuracy was estimated by the CV (%) determined between proteins abundances of equimolar proteins, such as UPS1 proteins. Otherwise, the amount of proteins removed by filters and methods was also considered.

In this part, three main results were described and the complete analysis was detailed in the in-coming publication presented in the Annex (p).

- Filters and amount of proteins

Filtering out peptides led to the exclusion of proteins more or less drastically according to the filter. The occurrence and outlier filter removed 26.6% and 32.4% of total proteins while shared peptide and RT filter removed 1.6% and 0.2% of the total proteins, respectively (Table I. 1, highlighted row). Moreover, TOP3 quantification being computed only from the three most intense peptide ions the amount of proteins was more sensitive to filters and lower than with other quantification methods which used all peptides.

Table I. 1 Number of proteins in all datasets: the normalized unfiltered dataset (No filter), after application of shared peptide, RT, occurrence and outliers filters. In parenthesis, the percentage of data removed by the filter from the previous dataset (See Annex for complete table, p).

	No filter	Shared peptide filter	RT filter	Occurrence filter	Outliers filter
Yeast + UPS1	2080	2046 (-1.6%)	2041 (-0.2%)	1491 (-26.9%)	1008 (-32.4%)
Yeast	2039	2005 (-1.7%)	2000 (-0.3%)	1455 (-21.3%)	973 (-33.1%)
UPS1	41	41 (-0%)	41 (-0%)	36 (-12.2%)	35 (-2.8%)

- Filters and precision

According to the median and dispersion of CV across technical replicates (see Figure 3A in Annex, p), the precision remained globally unchanged, indicating that neither filters nor methods can manage errors introduced during the experiment.

- Filters and accuracy

The effects of filters on the accuracy -absolute and relative- performances of the different quantification methods were synthesized in Figure I. 1. In this figure, the relative accuracy, estimated by the coefficient of determination (meaning the linearity, R^2) was plotted in x-axis while the absolute accuracy (meaning imprecision, estimated by the CV (%) between proteins abundances of equimolar UPS1 proteins) was plotted in y-axis.

Despite for iBAQ the relative and absolute accuracy (linearity and imprecision, respectively) were improved by the four filters (Figure I. 1). Average and iBAQ absolute accuracy were the most improved by the shared peptide filter. Apart for the Model, the RT filter has a slight effect on the precision but improved the linearity. The occurrence filter particularly improved the relative accuracy of Average, Average-Log and TOP3 without drastically improving the absolute accuracy, excepting for TOP3. The outliers filter improved the relative and the absolute accuracy for all the methods except for iBAQ for which the absolute accuracy decreased (increase of imprecision). Model was demonstrated to be a robust method as it achieved good performances in term of relative and absolute accuracy after only the shared peptides filter which is related to the capability of the Model to correct source of variability such as the peptides effect.

To conclude, this work was done to evaluate the filtering effect on quantification methods and was then used to evaluate rationally how to quantify proteins dataset of tomato fruit.

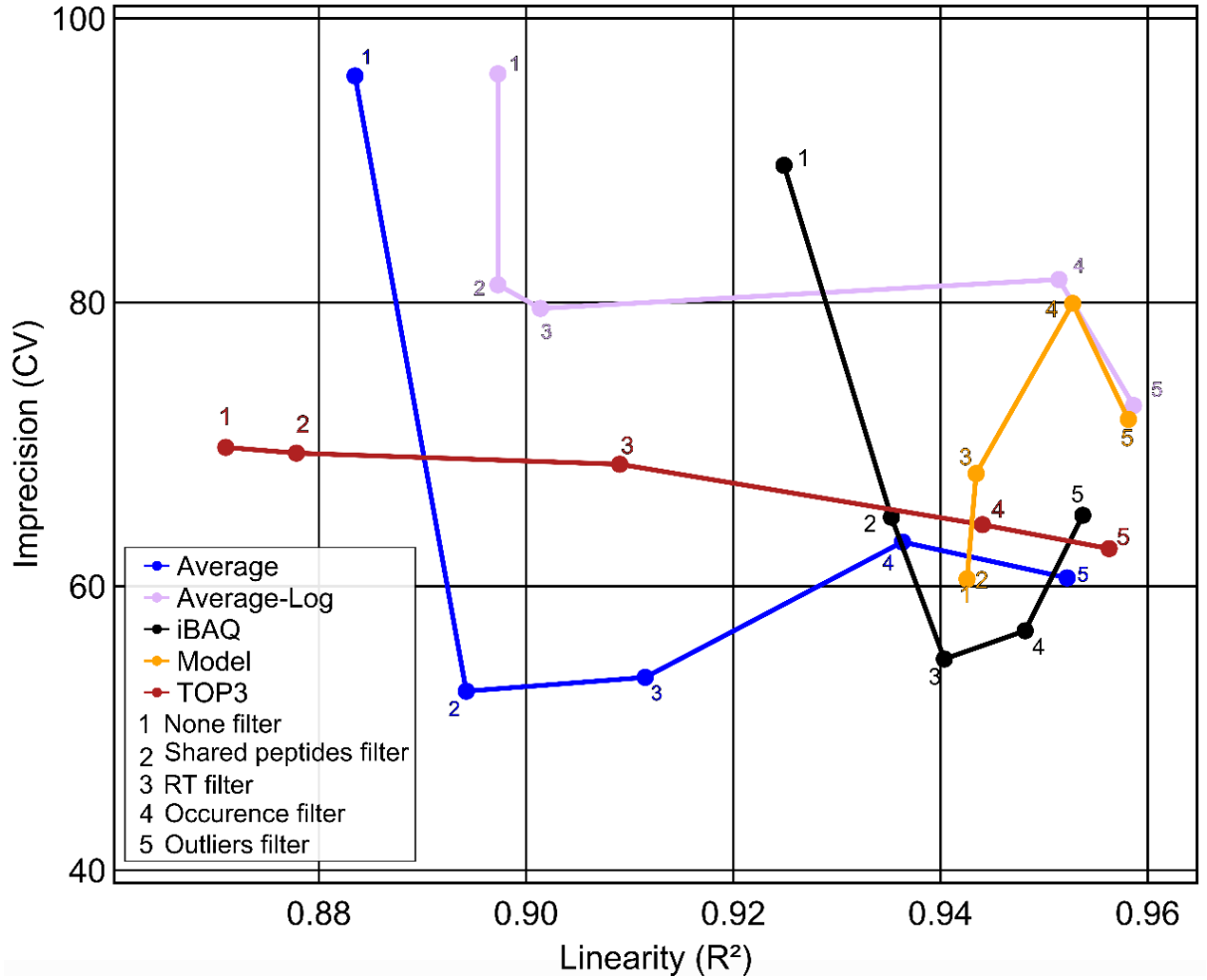


Figure I. 1 Relation between relative (R^2) and absolute accuracy (CV (%)) for each method of quantification. Only medians of CV (%) between UPS1 protein abundance versus medians of R^2 of the linear regression between estimated and spiked UPS1 protein abundance were displayed. Only UPS1 proteins detected in all filtered datasets were used. UPS1 proteins abundance was quantified by iBAQ (black line), TOP3 (red line), Average (blue line), Average-Log (purple line) and Model (orange line) methods. Numbers refer to the dataset used: 1 corresponding to the initial non-filtered normalized dataset (None filter), 2 to the shared peptides filtered dataset (Shared peptides filter), 3 to the RT filtered dataset (RT filter), 4 to the occurrence filtered dataset (Occurrence filter) and 5 to the outliers filtered dataset (Outliers filter).

II. Absolute quantification of tomato proteins from LC-MS/MS label-free proteomics

2.1 More than 2000 tomato fruit proteins quantified by peptides intensity modelling

Total proteins were extracted from tomato pericarp using adapted extraction protocol described in (Faurobert et al., 2007) at nine developmental stages: 7.7, 15, 21.7, 28, 34.3, 41.3, 48.5, 50.3 and 53 days post anthesis (DPA). Proteins were extracted and trypsin digested into peptides. Peptides were eluted, ionized by electrospray and analyzed by LC-MS/MS. Peptide ions and proteins were identified and quantified in label-free based on extracted ion chromatograms (XIC) using MassChroQ (Valot et al., 2011) program.

As shown in the previous section, the number of proteins and their respective abundance were related to peptides and to the method of quantification used. Thus, the four peptide filters and the five methods of quantification described previously (iBAQ, TOP3, Average, Average-Log, Model) were applied on tomato protein dataset in order to determine which filter and method combination represented the best compromise between quality of quantification and number of protein quantified. Here in the tomato dataset, the occurrence filter removed peptides that were not detected in at least two replicates of all developmental stages. By doing this way, we considered potential differences in peptide ions composition induced by the difference of fruit age throughout development.

As none UPS1 proteins were spiked in tomato samples, the relative and absolute accuracy could not be assessed thus only the precision between replicates and the number of proteins were investigated.

The cross-effect of the quantification methods and filters on the number of proteins are presented in Table I. 1. Note that TOP3 method quantified 10% less proteins than iBAQ, Average, Average-Log and Model which meant that at least 10% of tomato proteins were detected and identified with less than 2 peptides. The shared peptides filter removed similar proportion of proteins (~8%) between iBAQ, Average, Average-Log and Model methods while 20.7% of

proteins quantified by TOP3 were excluded. Whatever the method used, 0.3% of tomato proteins were removed by the RT filter. The most drastic effect resulted from the occurrence filter which removed around 36% of proteins for quantification based on iBAQ, Average, Average-Log, Model and more than 44.3% for TOP3 method. A less stringent threshold of the occurrence filter should be tried in a further analysis. In the same way, the proportion of excluded proteins by the outlier filter was the same for quantification based on iBAQ, Average, Average-Log, Model (21%) but higher for quantification based on TOP3 (24.1%). Whatever the filter used, these results has shown that TOP3 method quantified the lowest amount of proteins.

Effects of the peptides filters on the precision of the tomato proteins quantification was evaluated by computing, for the nine developmental stages, the CV of each tomato proteins over biological replicates (Figure I. 2). For the five methods of quantification, the precision remained globally unchanged by the shared peptides and the RT filter while the occurrence and outlier filter decreased the median and the dispersion of CVs. Besides, the dispersion and medians of CVs was the lowest for the quantification based on Model.

In summary, with two protein datasets, Yeast-UPS1 and tomato, the occurrence filter led to a significant protein exclusion thus we limited the filtering at the RT filter. At RT filter, iBAQ, Average and Average-Log quantified 16 more proteins than Model but the precision of these quantification methods remained slightly lower. Quantification based on TOP3 was removed because it was the less adequate method for quantifying a large scale proteome. Considering performances on absolute and relative accuracy estimated with UPS1 proteins (Figure I. 1) and the precision obtained on tomato proteins dataset (Figure I. 2), we quantified the 2494 tomato proteins with the Model method after the RT filter.

Table I. 2 Number of proteins quantified by the five methods (iBAQ, TOP3, Average, Average-Log, Model) in the unfiltered dataset (No filter) and after the application of shared peptides, RT, occurrence and outlier filters. In parenthesis, the percentage of proteins removed by the filter from the previous dataset.

	iBAQ	Average	Average-Log	Model	TOP3
No filter	2727	2727	2727	2707	2455
Shared peptide filter	2517 (-7.7%)	2517 (-7.7%)	2517 (-7.7%)	2502(-7.6%)	2162 (-20.7%)
RT filter	2510 (-0.3%)	2510 (-0.3%)	2510 (-0.3%)	2494 (-0.3%)	2155 (-0.3%)
Occurrence filter	1587 (-36.8%)	1587 (-36.8%)	1587 (-36.8%)	1586 (-36.4%)	1201 (-44.3%)
Outliers filter	1254 (-21%)	1254 (-21%)	1254 (-21%)	1253 (-21%)	911 (-24.1%)

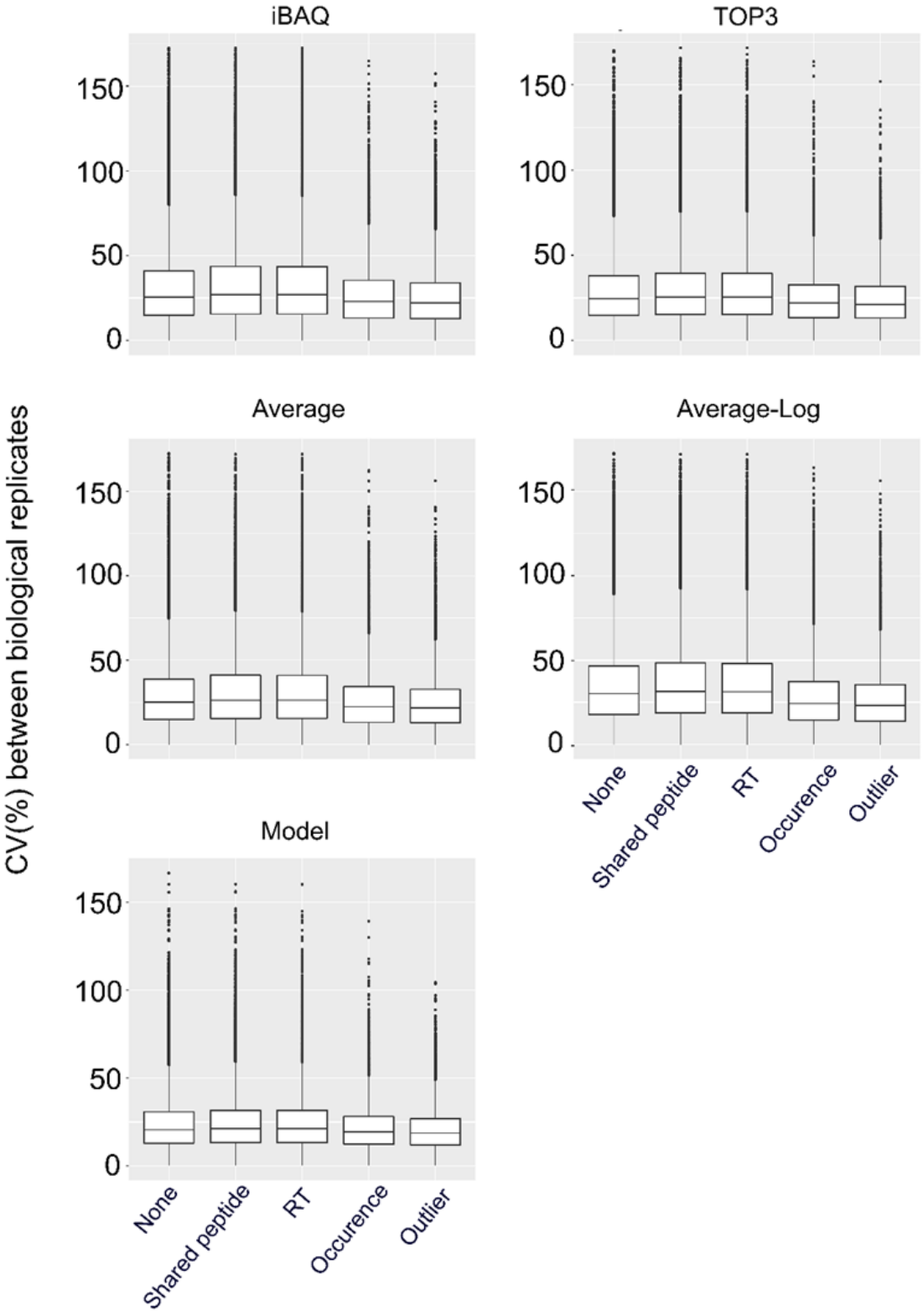


Figure I. 2 Effect of filters on the precision of the tomato proteins quantification. For each protein, CV (%) were calculated between biological replicates of the nine developmental stages and for the five methods of quantification: iBAQ, TOP3, Average, Average-Log and Model. Bolder line corresponds to aggregated outliers (black dot).

At this stage, tomato proteins were expressed in arbitrary unit based on peptides intensity signal. In order to express the protein abundance on a mole-basis (in $mol.gFW^{-1}$) we used the method called "total protein approach" (TPA) (Wiśniewski et al., 2014) represented in Equation I. 1 and Equation I. 2. First, each protein was expressed as a fraction of the total protein content in the sample ($g_{Total\ protein}.gFW^{-1}$) and second in an absolute quantification ($fmol_{protein}.gFW^{-1}$) using the protein molar weight (MW, $g_{protein}.mol_{protein}^{-1}$). By expressing the protein abundance as a fraction of the total protein content, we assumed that all the proteins were extracted and that the sum of MS signal detected proteins was equal to the total protein content MS signal. But as more than 7000 proteins have been quantified in tomato fruit by Szymanski et al., (2017), we were aware that proteins concentrations estimated here by TPA were overestimated.

$$Protein_{i,k} = \frac{abundance\ protein_{i,k}}{(\sum_1^n abundance\ protein)_k}\ x\ (Total\ protein\ content)_k \qquad \text{Equation I. 1}$$

$$[Protein_{i,k}] = Protein_{i,k} x\ \frac{1}{MW_{protein_i}} \qquad \text{Equation I. 2}$$

With $Protein_{i,k}$ the amount of each $protein_i$ (i = 1:2494) in the sample k (k=1:26) in gFW^{-1}, n the total number of protein (n=2494), $(Total\ protein\ content)_k$, the total amount of proteins in the sample k in gFW^{-1}, $[Protein_{i,k}]$ the concentration of each $protein_i$ in the sample k in $fmol.gFW^{-1}$ and $MW_{protein_i}$ the molar weight (in $g.mol^{-1}$) of the $protein_i$.

2.2 Cross-validation of protein quantification with enzyme proteins

The accuracy of the proteins concentration obtained from LC-MS/MS quantification was assessed with a subset of enzyme proteins for which the concentration was estimated from the enzyme capacities (V_{max}). V_{max} corresponds to the number of mole of substrate consumed per minute under optimal enzymatic conditions then normalized by the amount of biological sample (gram of fresh weight) ($mol_S.min^{-1}.gFW^{-1}$).

Thirty-six enzyme capacities (V_{max}), reported in Biais et al., (2014), were measured at the same nine developmental stages. To estimate the concentration of the corresponding enzymes (in

fmol.gFW^{-1}), we used the specific enzyme activity, *i.e.* the number of mole of substrate consumed per mass of purified enzyme per minute (mols.min^{-1}.g$_{protein}^{-1}$), and the molar weight (MW, g$_{protein}$.mol^{-1}) of the corresponding protein (Equation I. 3).

$$[Enzyme_{i,k}] = \frac{V_{max_{i,k}}}{Specific\ activity_{Enzyme_i}} \ x \ \frac{1}{MW_{Enzyme_i}} \qquad \text{Equation I. 3}$$

With $[Enzyme_{i,k}]$ the concentration in fmol.gFW^{-1} of the enzyme i (i =1:36) in the sample k (k = 1:26), MW the molar weight (g$_{protein}$.mol^{-1}) of the enzyme.

Most of enzyme specific activities were found in literature, but some of them could be under-estimated. Indeed, the purification of enzymes is a tedious work of successive steps resulting in partial purification and potential alteration of the purified proteins. In the bibliography of specific activity, we paid attention to select specific activity issued from model organism of plant family. Thus, as in Piques et al., (2009), the highest specific activity was preferentially used for calculation to minimize the bias. Values of enzyme specific activity, plants and references are presented Table I. 3.

Tomato proteins annotation (ITAG2.4) was used to recover and compare both concentration, estimated by LC-MS/MS and V_{max}. But enzyme proteins usually require more than one isoform (Table I. 3). Using databases such as Solgenomics (Fernandez-Pozo et al., 2015), PGSB (Spannagl et al., 2016) and Uniprot (Bateman et al., 2017), and the enzyme commission number (EC) associated to each enzyme activity, one hundred eighty-one tomato protein annotations were assigned to the 36 enzymes measured by Biais et al (2014).

In order to compare protein quantification estimated by LC-MS/MS and enzyme activity, when more than one protein isoform was detected, concentrations determined by LC-MS/MS were summed for the corresponding enzyme. By doing this way, we assumed that all isoforms detected by LC-MS/MS participated equally to the corresponding V_{max}, in other words, had the same specific activity.

Note that four enzymes (AlaAT, pF16BPase, NAD-GDH, NADP-GDH, Table I. 3) were not taken into account here because none specific activity had been found or because none protein was detected by LC-MS/MS.

Concentrations based on LC-MS/MS and V_{max} were cross-validated by two ways. First, for each enzyme-protein pair, Spearman correlation between protein concentrations estimated by LC-MS/MS and enzyme concentrations was analyzed (concentration averaged per fruit age) to evaluate the similarity of profiles throughout the tomato fruit development. Spearman correlation, a non-parametric test, was used because data were not normal distributed. Concentrations were considered as correlated when the coefficient of determination (R^2) was higher than 0.6 ($P < 0.05$). Second, at each developmental stage, we compared enzyme-to-enzyme ratios obtained for each method of quantification. This step allowed to check that molar relations between proteins were preserved with both methods of protein quantification.

Concentration profiles throughout the development and coefficient of determination are displayed in Figure I. 3. Among the 32 enzymes, a significant positive correlation ($R^2 > 0.6$ and $P < 0.05$) was found for twenty-one enzyme-protein pairs meaning that the concentrations changes were similar between the two methods of quantification for more than 68% of these enzyme proteins (Acid Inv, AGPase, Aldolase, Enolase, FK, GK, NAD-GAPDH, NAD-MDH, NAD-ME, NADP-GAPDH, NADP-ME, NADP-IDH, PGI, PGK, PGM, PK, PFP, SuccCoA Ligase, Susy, TPI, Ugpase). Among these 21 enzymes, 7 (Aldolase, FK, GK, NADP-IDH, PGI, PK and SuccCoA Ligase) were higher at the youngest stage and decreased sharply during cell division (15-21.7 DPA), tending to a plateau until the end of fruit development and maturation. For six enzymes (AGPase, Enolase, PGM, Susy, TPI and Ugpase protein), concentrations decreased almost linearly. Only one protein enzyme, the acid invertase (Acid Inv) displayed a concentration profile with an increasing concentration at the end of the development.

For eight proteins (Aconitase, AspAT, CS, Fumarase, NAD-IDH, Neutral Inv, G6PDH, PEPC), concentrations estimated by the two methods of quantification were not significantly correlated ($R^2 > 0.6$ and $P > 0.05$) and even three proteins concentrations were negatively correlated ($R^2 < 0$; SPS, cFBPase and PFK). Reasons that might explained the poor correlation between both quantification methods, by enzyme activity and LC-MS/MS, for these ten proteins could be: (1) quantification by LC-MS/MS method is more sensitive allowing for instance the detection of a peak at 34.3 DPA not really apparent with enzyme activity, (2) LC-MS/MS quantified enzyme proteins without considering if they were active (i.e. in native state), (3) all protein isoforms did not necessarily participate equally in enzyme activity (i.e. their specific activity could differ) and (4) a regulation

of enzyme activity by the environment (light sensitivity, phosphorylation and redox state of the cell) and post-translational modifications (phosphorylation…).

Table I. 3 Information about enzymes. This table summarizes information about protein specific activity (mol.g_{Enzyme} $protein^{-1}$.min^{-1}), literature sources (under the table), the number of isoforms annotated and detected by LC-MS/M, the coefficient correlation (Spearman) determined between enzyme protein concentration estimated by LC-MS/MS and Vmax and the slope of the linear regression between enzyme protein concentration estimated by LC-MS/MS and Vmax after log_{10} transformation. Significant Spearman coefficient correlation are indicated by *.

Enzyme	Specific activity (mol/min/g protein)	Plant	Annotated Isoforms	Isoforms detected by LC-MS/MS	Spearman coefficient correlation (R^2)	Slope of linear regression between V_{max} and LC-MS/MS concentrations (log10-log10)
Acid Inv	1.2	Carrot [1]	2	1	0.78*	1.49
Aconitase	0.7	Tabacco [2]	2	2	0.4	0.45
AGPase	0.156	Spinach [3]	7	5	0.9*	1.39
AlaAT	NA	NA	4	2	ND	ND
AspAT	0.3	Carrot [4]	10	6	0.55	0.49
PFK	0.06	Tomato [5]	6	1	-0.43*	-0.21
CS	0.6642	Pea [6]	4	2	0.12	0.25
Enolase	0.0103	Maize [7	5	4	0.98*	1.21
cFBPase	0.119	Spinach [8]	2	2	-0.18	0.21
pF16BPase	NA	NA	3	2	ND	ND
Aldolase	0.0263	Carrot [9]	12	7	0.88*	0.81
FK	0.025	Tomato [10]	4	3	1*	1.01
Fumarase	0.238	Pea [11]	4	1	0.08	0.35
G6PDH	0.2179	Potato [12]	5	1	0.5	0.34
GK	0.01	Potato [13]	6	2	0.98*	0.76
NAD-GAPDH	0.041	Spinach [14]	13	10	0.85*	0.98
NAD-GDH	NA	NA	4	0	ND	ND
NAD-IDH	0.008	Pea [15]	4	1	0.67*	0.49
NAD-MDH	3	Spinach [16]	10	7	0.97*	0.6

NAD-ME	0.0725	Potato [17]	2	2	0.87*	0.49
NADP GAPDH	0.123	Spinach [18]	1	1	0.8*	1.53
NADP-GDH	NA	NA	1	0	ND	ND
NADP-IDH	0.05	Tabacco [19]	3	3	0.82*	0.65
NADP-ME	0.0733	Maize [20]	6	2	0.93*	0.35
Neutral Inv	0.431	Arabidopsis [21]	9	1	0.23	0.62
PEPC	0.0496	Peanut [22]	5	5	0.53	0.49
PGI	2.456	Apple [23]	2	2	0.8*	0.45
PGK	0.914	Barley [24]	3	2	0.98*	0.95
PGM	0.48	Potato [25]	5	3	0.98*	0.57
PFP	0.0438	Pineapple [26]	6	5	0.98*	0.58
PK	0.061	Rapeseed [27]	10	8	0.88*	0.85
SPS	0.0795	Spinach [28]	4	1	-0.03	0.04
SuccCoA Ligase	0.0012	Spinach [29]	3	2	0.8*	1.02
Susy	0.0395	Tabacco [30]	6	2	0.75*	3.25
TPI	10.2	Lettuce [31]	4	3	0.98*	0.65
UGPase	1.099	Potato [32]	4	3	0.88*	3.58

Where, [1] (Unger et al., 1992), [2] (Navarre et al., 2000), [3] (Copeland and Preiss, 1981), [4] (Turano et al., 1990), [5] (Isaac and Rhodes, 1982), [6] (Unger and Vasconcelos, 1989), [7] (Lal et al., 1994), [8] (Ladror et al., 1990), [9] (Moorhead and Plaxton, 1990), [10] (Martinez-Barajas et al., 1997), [11] (Behal and Oliver, 1997), [12] (Graeve et al., 1994), [13] (Moisan and Rivoal, 2011), [14] (Scagliarini et al., 1998), [15] (Igamberdiev and Gardeström, 2003), [16] (Zschoche and Ting, 1973), [17] (Grover et al., 1981), [18] (Michels et al., 1994), [19] (Galvez et al., 1994), [20] (Thorniley and Dalziel, 1988), [21] (Tang et al., 1996), [22] (Maruyama et al., 1966), [23] (Zhou and Cheng, 2008), [24] (McMorrow and Bradbeer, 1990), [25] (Takamiya and Fukui, 1978), [26] (Tripodi and Podesta, 1997), [27] (Smith et al., 2000), 28] (Sonnewald et al., 1992), [29] (Kaufman and Alivisatos, 1995), [30] (Matic et al., 2004), [31] (Eran Pichersky, 1984), [32] (Sowokinos et al., 1993).

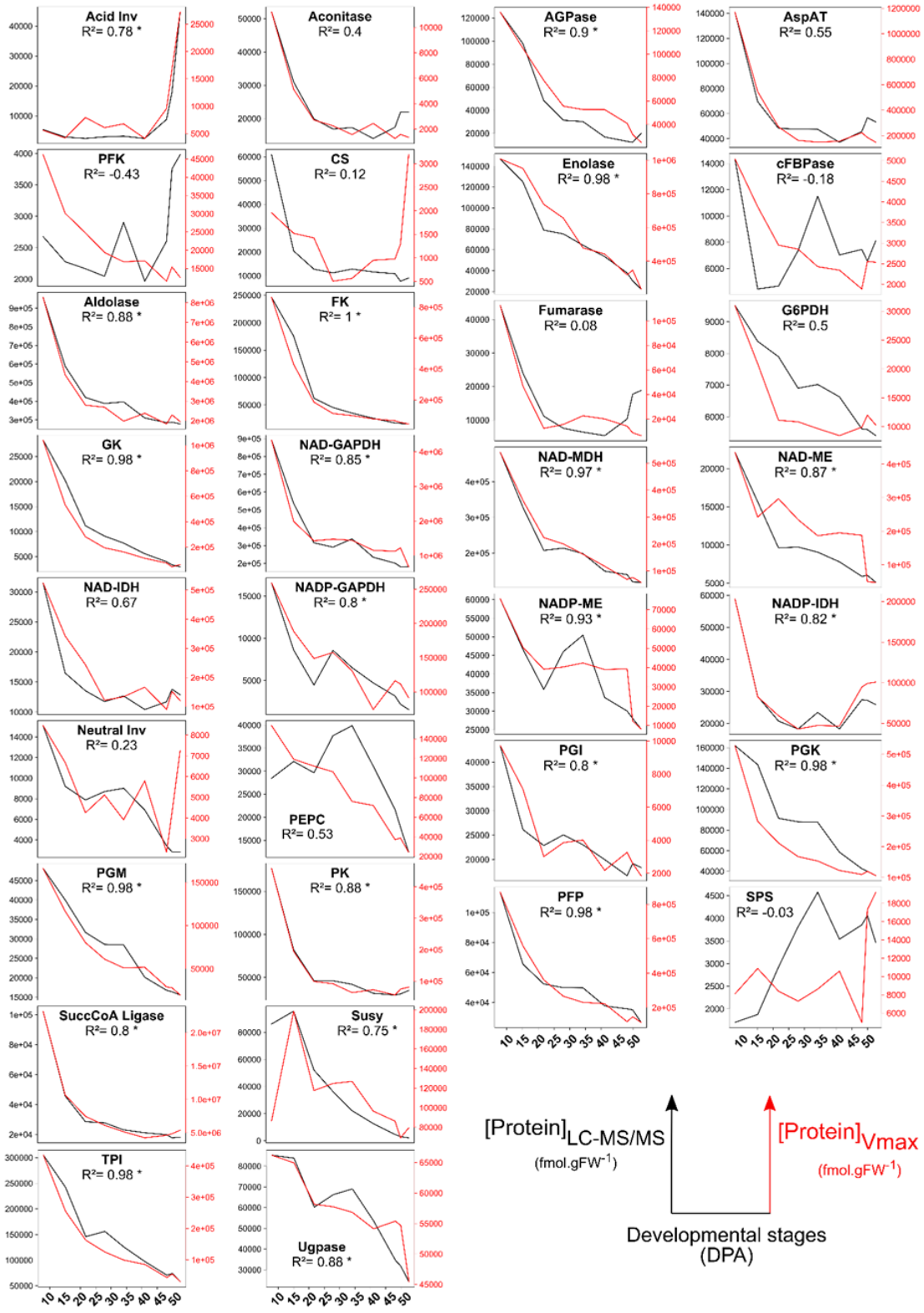


Figure I. 3 Changes in 32 enzyme proteins abundance quantified from enzyme activities (Vmax, red curves) and by LC-MS/MS (black curves), at nine developmental stages. Both concentration are expressed in fmol.gFW^{-1}.

In the second step, to evaluate the accuracy of quantification, we calculated ratios between (1) concentrations of enzyme protein estimated by both methods ($R_{i,j}$ (see Equation I. 4) expected close to one if concentrations based on LC-MS/MS and V_{max} were similar) and (2) the relative abundance between two enzyme proteins for both methods of quantification ($R_{i,o,j}$ (see Equation I. 5) also expected close to one). For both equations, we considered that ratios in the range of 0.5 to 2 as a reasonable cross-validation.

$$R_{i,j} = \frac{[Protein_{i,j}]_{LC-MS/MS}}{[Enzyme_{i,j}]_{Vmax}} \qquad \text{Equation I. 4}$$

$$R_{i,o,j} = \frac{\frac{[Protein_{i,j}]_{LC-MS/MS}}{[Protein_{o,j}]_{LC-MS/MS}}}{\frac{[Enzyme_{i,j}]_{Vmax}}{[Enzyme_{o,j}]_{Vmax}}} \qquad \text{Equation I. 5}$$

With $[protein_{i,j}]$ and $[Enzyme_{o,j}]$ the average concentrations, estimated by *LC-MS/MS* and V_{max}, of i (i =1:32) and o (o =1:32) enzyme-protein pairs at the j th developmental stage (j =1:9).

Ratios – i.e. Ri,j and Ri,o,j - were calculated on enzyme protein concentrations averaged by developmental stage (j). In Figure I. 4, we represented the distribution of Ri,j (Equation I. 4) calculated for the 32 enzyme-protein pairs at the nine developmental stages. Medians of ratios Ri,j being lowest than one at the nine stages, most of concentrations estimated for enzyme proteins by Vmax were higher than the ones estimated by LC-MS/MS (Figure I. 3, Figure I. 4). Several reasons can mutually and non-exclusively explain these discrepancies: (1) quantification by LC-MS/MS did not necessarily considered all isoforms and/or (2) post-translational modifications or protein-protein interactions modulating the enzyme activity.

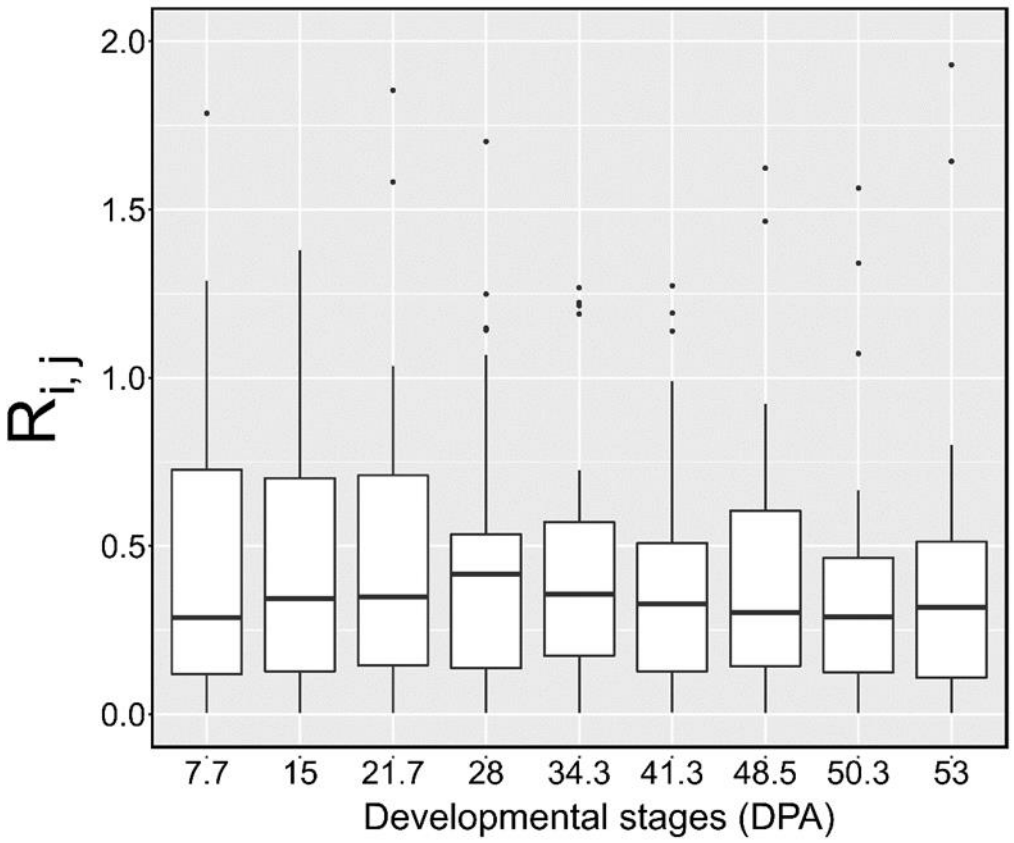


Figure I. 4 Distribution of ratio calculated from Equation I. 4 for 32 enzyme proteins. For each *i* th enzyme protein (*i*: 1:32), concentrations estimated by LC-MS/MS and Vmax were averaged per developmental stage (DPA, *j*=1:9) and used to calculate ratios (Ri,j). To gain visibility the ratio scale was set between 0 and 2.

The second calculated ratio, $R_{i,o,j}$, resulting from Equation I. 5, which estimated the relative abundance between two enzyme proteins determined with both methods of quantification, were presented in Figure I. 6. To illustrate the $R_{i,o,j}$ calculation, we described the comparison at 7.7 DPA of PGK and PGM enzyme proteins. Based on the LC-MS/MS quantification, at 7.7 DPA, PGK was found 3.37 more concentrated than PGM and, based on V_{max} quantification, PGK was found 3.18 more concentrated than PGM enzyme protein. Thus, $R_{i,o,j}$ being equal to 1.06 meant that both methods of quantification provided a similar molar relation between PGM and PGK. Enlarged to the comparison of the 32 enzyme proteins, 496 $R_{i,o,j}$ (2^{32} comparisons) were determined at each developmental stage (Figure I. 5) and summarized in Figure I. 6.

Unexpected $R_{i,o,j}$ results were obtained. Indeed, a rapid visual inspection of the nine heatmaps in Figure I. 5 did not appear mostly colored in yellow, which was expected. Besides, we noticed that a subset of 5 enzyme proteins (SuccCoA Ligase, Susy, NAD-ME, NAD-IDH, NAD-GAPDH),

appearing as a red block on the heatmaps, displayed extreme results in almost all developmental stages and whatever enzymes proteins they are compared. Extreme $R_{i,o,j}$ values resulting either from a molar relation between enzyme proteins over-estimated by LC-MS/MS or an under-estimation by V_{max} or from (2) both at the same time, an over and under estimation by LC-MS/MS and V_{max}, respectively. The $R_{i,o,j}$ calculation cumulating limitations (experimental, quantification) of both methods can explain these extreme values which then participated to the increase of the mean of $R_{i,o,j}$ values presented in Figure I. 6. In parallel, the distribution of all $R_{i,o,j}$ ($\log_{10}$ scale), i.e. with the 496 $R_{i,o,j}$ obtained at each of the nine developmental stage, displayed a distribution centered on one. Note that 23.6% of $R_{i,o,j}$ ratios were comprised between 0.5 and 2, i.e. expected values. Beside, median of $R_{i,o,j}$ calculated ratios at the nine developmental stages were close to one, especially at the six first stages (but increased up to 1.58 at 53 DPA).

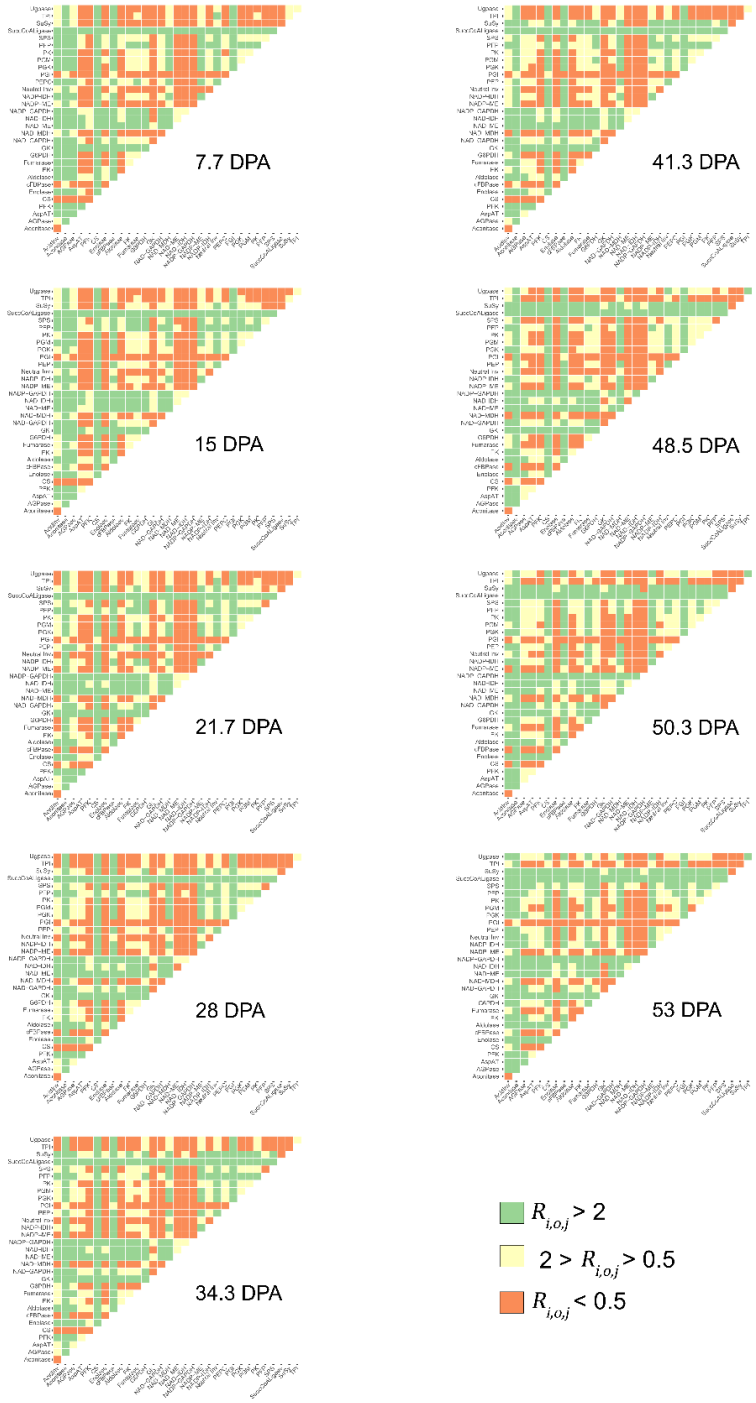


Figure I. 5 Heatmaps of the 496 $R_{i,o,j}$ obtained at the nine developmental stages. $R_{i,o,j}$ resulted from Equation I. 5. Each square represents the $R_{i,o,j}$ between the enzyme proteins heading the column and the row. $R_{i,o,j}$ higher than 2 are presented in green, lower than 0.5 are presented in orange, in the range from 0.5 to 2 are presented in yellow.

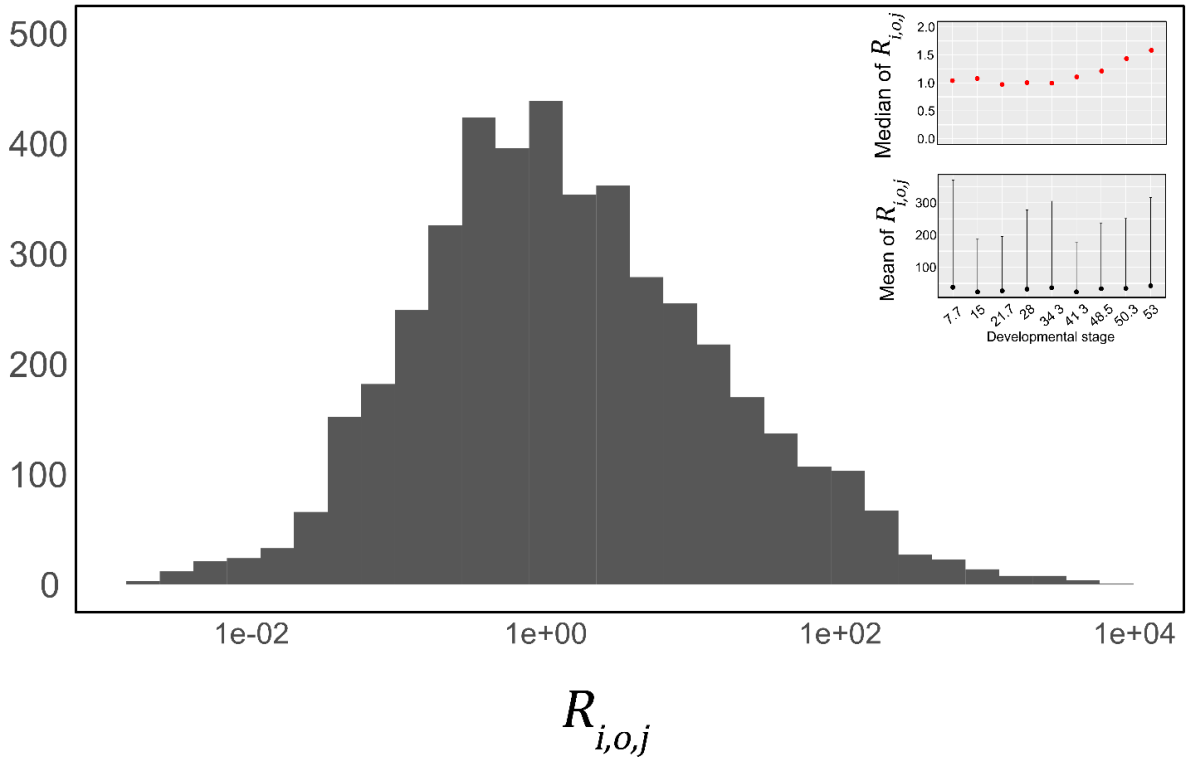


Figure I. 6 Distribution of $R_{i,o,j}$ (log_{10} scale) resulting from the Equation I. 5 presented without considering neither the developmental stage nor the couple of enzyme proteins that were compared. In inserts, median and mean +/- standard deviation of $R_{i,o,j}$ are presented per stage.

Despite some biases introduced by the two multistep methods, this analysis based on about 30 enzyme-protein pairs allowed the cross-validation of protein quantification by LC-MS / MS based on the Model method, and by extension we assume validation of the concentrations of all the proteins quantified by LC-MS / MS. To go further, it should be interesting to complete this analysis by investigating some protein complexes with known stoichiometry, such as proteasome complex (Arike et al., 2012; Fabre et al., 2014). Going further, an ideal validation should be to use AQUA peptides (See Introduction) targeted toward a subset of enzyme proteins.

At this stage we used this tomato dataset of 2494 proteins and investigated the global biological behaviors.

2.3 Changes in protein expression during tomato fruit development

In this section, we analyzed profiles of protein concentrations obtained by label-free LC-MS/MS during tomato fruit development. Figure I. 7 presented the distribution and the median of protein concentrations at the nine developmental stages. The most notable change of protein concentration occurred between 7.7 DPA and 21.7 DPA. Indeed, the protein concentration was divided by three from the first to the third developmental stage; from 18.5 pmol.gFW^{-1} at 7.7 DPA to 8.9 pmol.gFW^{-1} at 15 DPA and 5.7 pmol.gFW^{-1} at 21.7 DPA. Then the protein concentration slightly decreased from cell expansion (21.7 DPA) to ripening phase reaching 4.4 pmol.gFW^{-1} at 53 DPA (Figure I. 7).

To investigate whether changes in protein concentrations could be assigned to developmental phases, a hierarchical clustering analysis was performed on mean-centered data scaled to unit and displayed as a heat map (Figure I. 8). Protein concentrations highlighted five clusters. The first cluster grouped 263 proteins with an increase of concentration at the beginning of the ripening phase (48.5 to 53 DPA). The second cluster was characterized by 140 proteins up-regulated during cell expansion from 28 to 48.5 DPA. Conversely, the fourth cluster contained 189 proteins down-regulated in almost the same period. The third cluster contained 472 proteins with a two-time decrease, during cell division and maturation. The last cluster grouped 1430 proteins with high concentrations during the cell division (7.7 to 15 DPA) which then drastically decreased to reach a plateau from cell expansion phase until the end of the development.

These results showed that cell division phase involved more proteins than cell expansion and ripening phases.

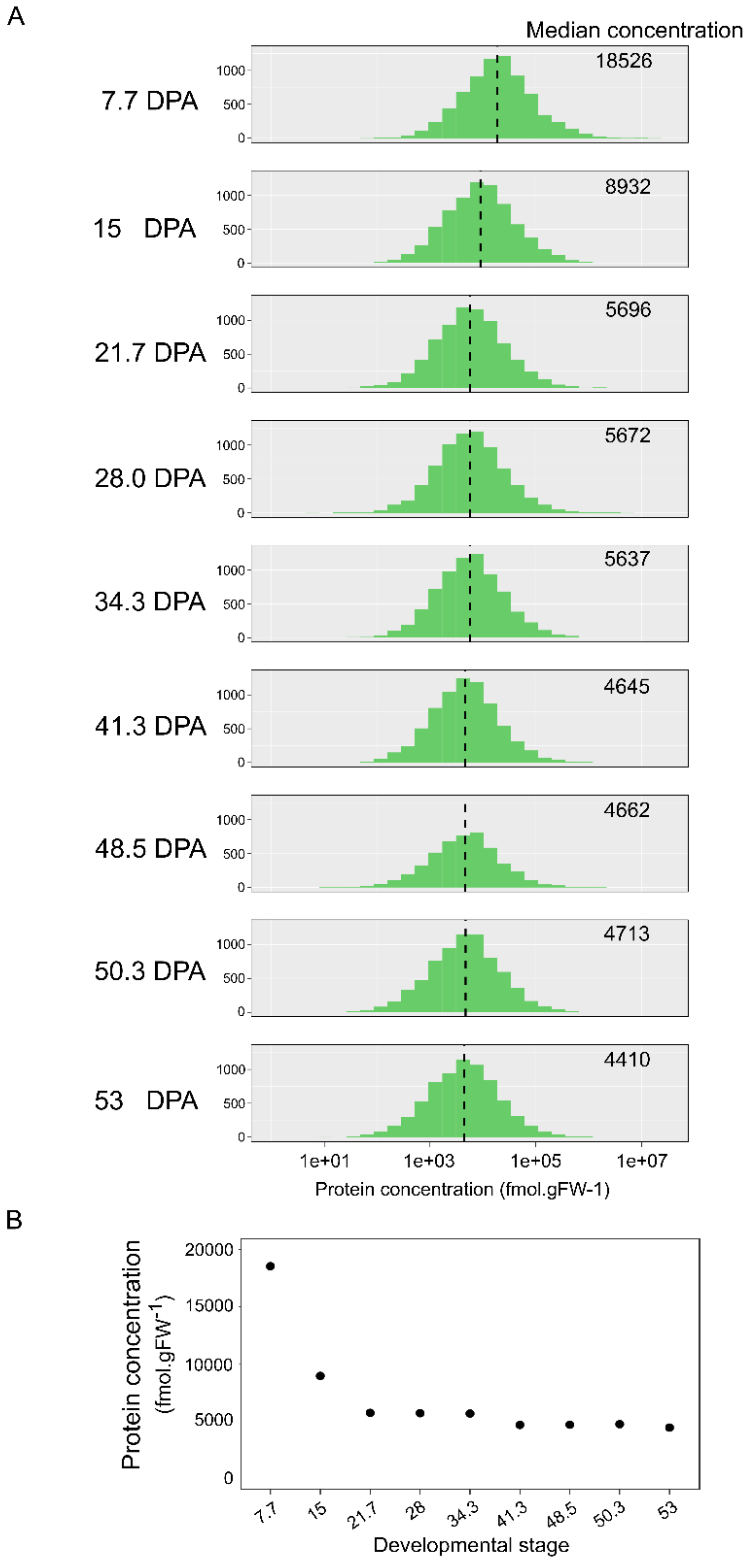


Figure I. 7 Overview of protein concentrations. (A) Distribution of protein concentrations (log_{10} scale) at the nine developmental stages, with median values mentioned at top-right corner and represented by a dashed line. (B) The median of protein concentrations at each developmental stage.

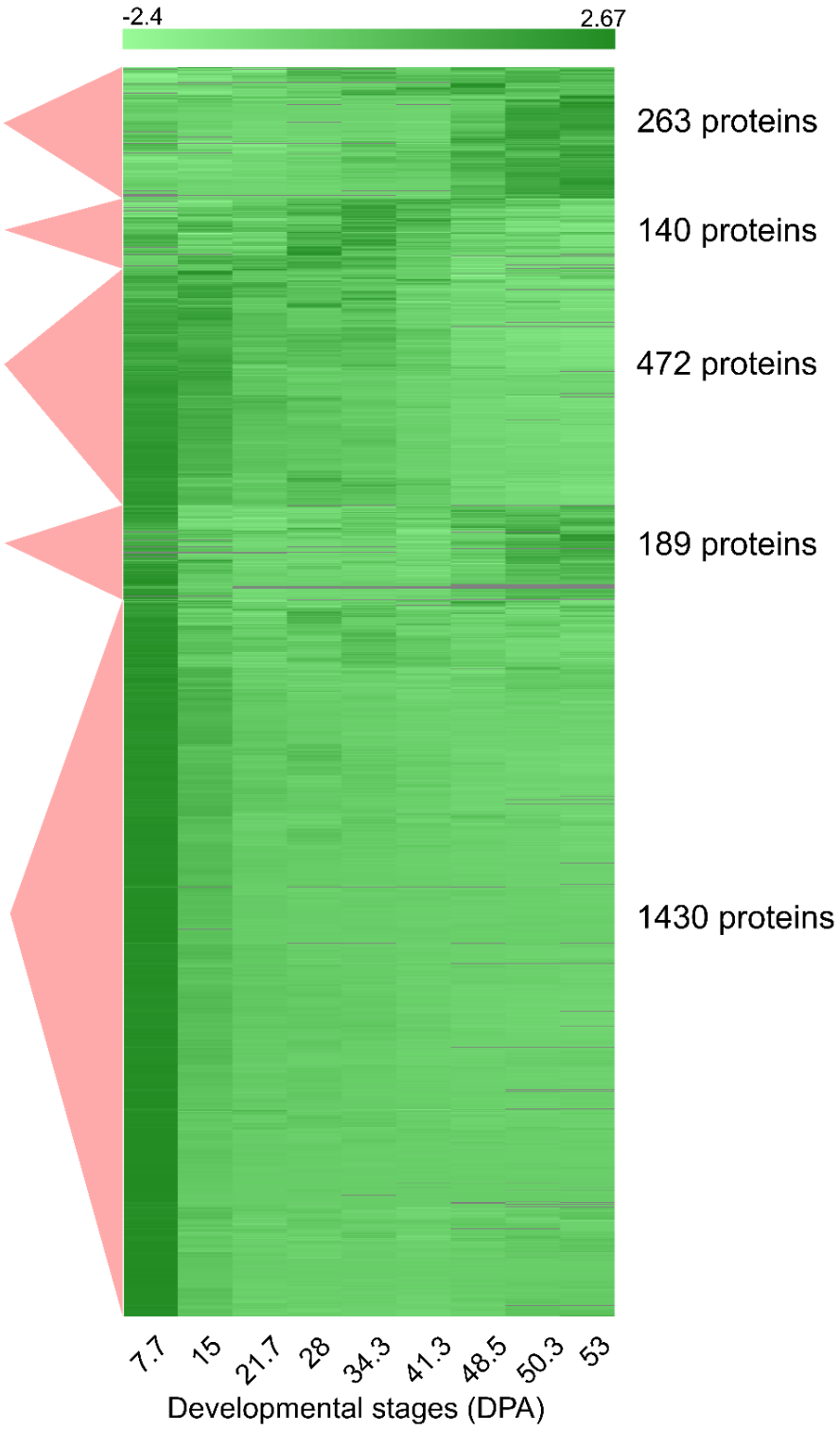


Figure I. 8 Overview of protein concentration changes. The clustering analysis was performed on protein concentrations (Pearson's correlation) mean centered and scaled to unit date. Columns correspond to the nine developmental stages, and rows correspond to protein concentrations. The number of proteins contained by the five clusters (red triangles) are indicated. Bars colored in grey are missing values.

2.4 Analysis of functional categories of 2494 tomato proteins

MapMan annotation file (Usadel et al., 2009) was used to assign a functional category to the 2494 proteins. The 35 MapMan BIN code describing functional categories were reduced to 19. For instance, the category named "Carbon metabolism" grouped carbohydrate metabolism, glycolysis, gluconeogenesis, oxidative pentose phosphate cycle, Krebs cycle and the fermentation metabolism.

The customized MapMan file was then used for two purposes: (1) to determine for the five clusters (Figure I. 8) how proteins were distributed according to their functional categories (Figure I. 9), paying attention to those that contained the most proteins, and (2) to determine the distribution of proteins concentrations in to functional categories (Figure I. 10).

In the first cluster, more than thirty percent of the 263 proteins up-regulated from the turning phase (41.3 to 48.5 DPA) were distributed in the "Miscellaneous" category (15.6%), "Stress" (11.8%) and "Protein" (10.3%) metabolism. In the second cluster, 15% of the 140 proteins were miscellaneous proteins, *i.e*. enzyme proteins, while twenty-one percent should not be assigned to any category. This latter percent suggested that most of the proteins mainly required between the cell expansion and the turning phase remained uncharacterized. Proteins with the opposite concentration profile (cluster 4) were associated to "Protein", "Miscellaneous" and "Amino acid metabolism" categories. We noticed that "Stress" category was also well represented (5.8%). Proteins belonging to the third cluster (cluster 3) were involved to the "Carbon" and "Photosynthesis metabolism" categories (11.6% and 11.2% resp.). The protein metabolism was also highly represented (14.8%). Functional categories identified above were consistent with physiological changes of tomato fruit. Indeed, similarly to Barsan et al., (2012) who described the tomato plastid proteome during the chloroplast-to-chromoplast transition we observed a (1) decrease in abundance of proteins associated to "Photosynthesis" category and "Carbon metabolism" (starch synthesis/ degradation) (cluster 3), (2) an increase of "Stress" category containing proteins such as heat shock protein (cluster 1, cluster 4). Surprisingly, cell wall metabolism, involved in the fruit firmness, was not highlighted by proteins displaying changes during ripening. By the way, the lipid metabolism was noticed in the first cluster which potentially played a role in cell membrane structure and fruit firmness at the end of development. The 1430 proteins up-regulated during the cell division and grouped in the fifth cluster phase (7.7 to 15 DPA)

were related to two functional categories: (1) Protein metabolism, sharing 28.7% of proteins and (2) the genome metabolism (DNA, RNA binding and metabolism) representing 11.5% of proteins. These two categories were expected in relation with physiological and structural (cell division, endoreduplication) occurring during this period.

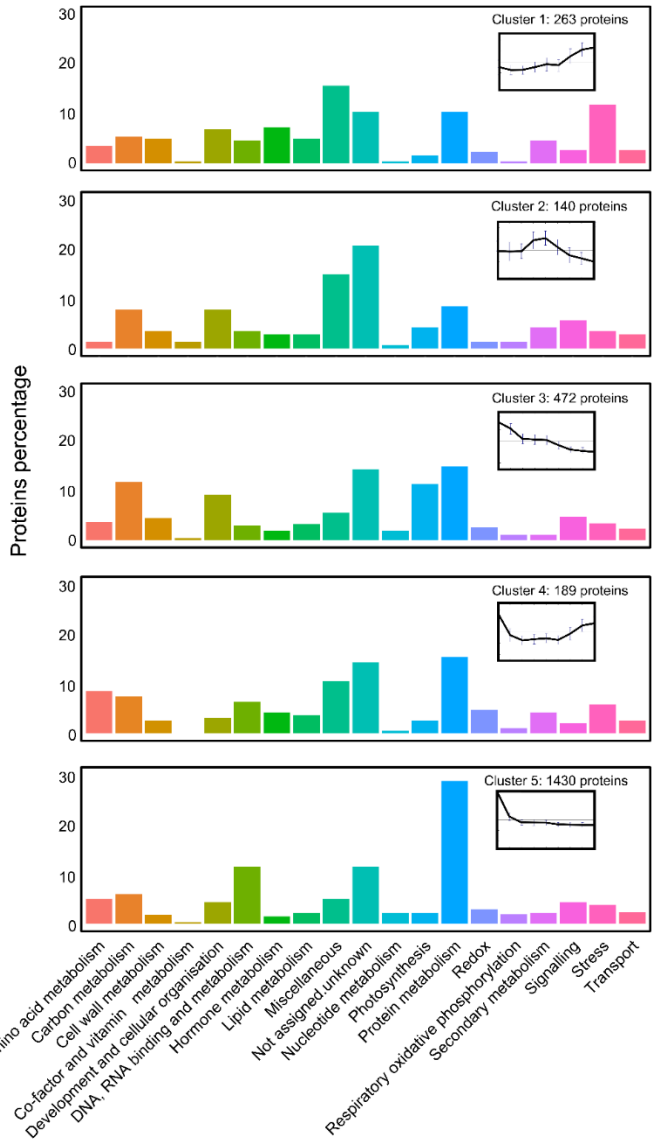


Figure I. 9 Functional categories associated to the five clusters of proteins. The 2494 quantified proteins were separated by hierarchical clustering in five clusters (Figure I. 8). Profile of proteins concentration and the numbers of proteins associated to the five clusters were presented. The 19 functional categories provided from manually summarized MapMan annotation file containing initially 35 functional categories (Usadel et al., 2009).

In the second section, we investigated the distribution of the 2494 proteins in functional category without considering clusters to determine the most "concentrated" functional categories. For this purpose, for the 2494 proteins we calculated the median concentrations throughout the development and then after being assigned to functional categories, we calculated the median concentration of proteins belonging to the 19 functional categories.

More than 30% of the 2494 proteins were shared by both "Protein metabolism" and "DNA, RNA binding and metabolism" categories. However, the "Photosynthesis", "Redox" and "Respiratory oxidative phosphorylation" categories, containing 100, 71 and 37 proteins, contained the proteins the most concentrated with median of 14.3, 13.4 and 9.4 $pmol.gFW^{-1}$, respectively. The "Respiration oxidative phosphorylation" category shared proteins involved in the mitochondrial electron transporting chain, such as F1F0 ATP synthase, NADH ubiquinone oxidoreductase, NADH dehydrogenase.

Besides, we went further detailing for the nine developmental stages, the median concentration of proteins belonging to each functional category (Figure I. 11). The categories the most concentrated, "Respiratory oxidative phosphorylation", "Redox" and "Photosynthesis", were consistent with results obtained in Figure I. 10. The visual inspection distinguished two profiles. The first, observed for a large part of functional categories, was characterized by a drastic decrease during the cell division, followed by a slight decrease or almost stabilized median proteins concentration throughout the development ("Amino acid metabolism", "Carbon metabolism", "Cell wall metabolism", "Development and cellular organization", "DNA, RNA binding and metabolism", "Lipid metabolism", "Miscellaneous", "Nucleotide metabolism", "Photosynthesis", "Protein metabolism", "Respiratory oxidative phosphorylation", "Signaling" and "Transport"). The second profile was characterized by a drastic decrease during the cell division followed by an increase during ripening, such as for "Co-factor and vitamin metabolism", "Hormone metabolism", "Redox", "Secondary metabolism" and the Stress" categories". These categories were consistent with the main processes enhanced during ripening already described in the literature (Osorio et al., 2013; Szymanski et al., 2017).

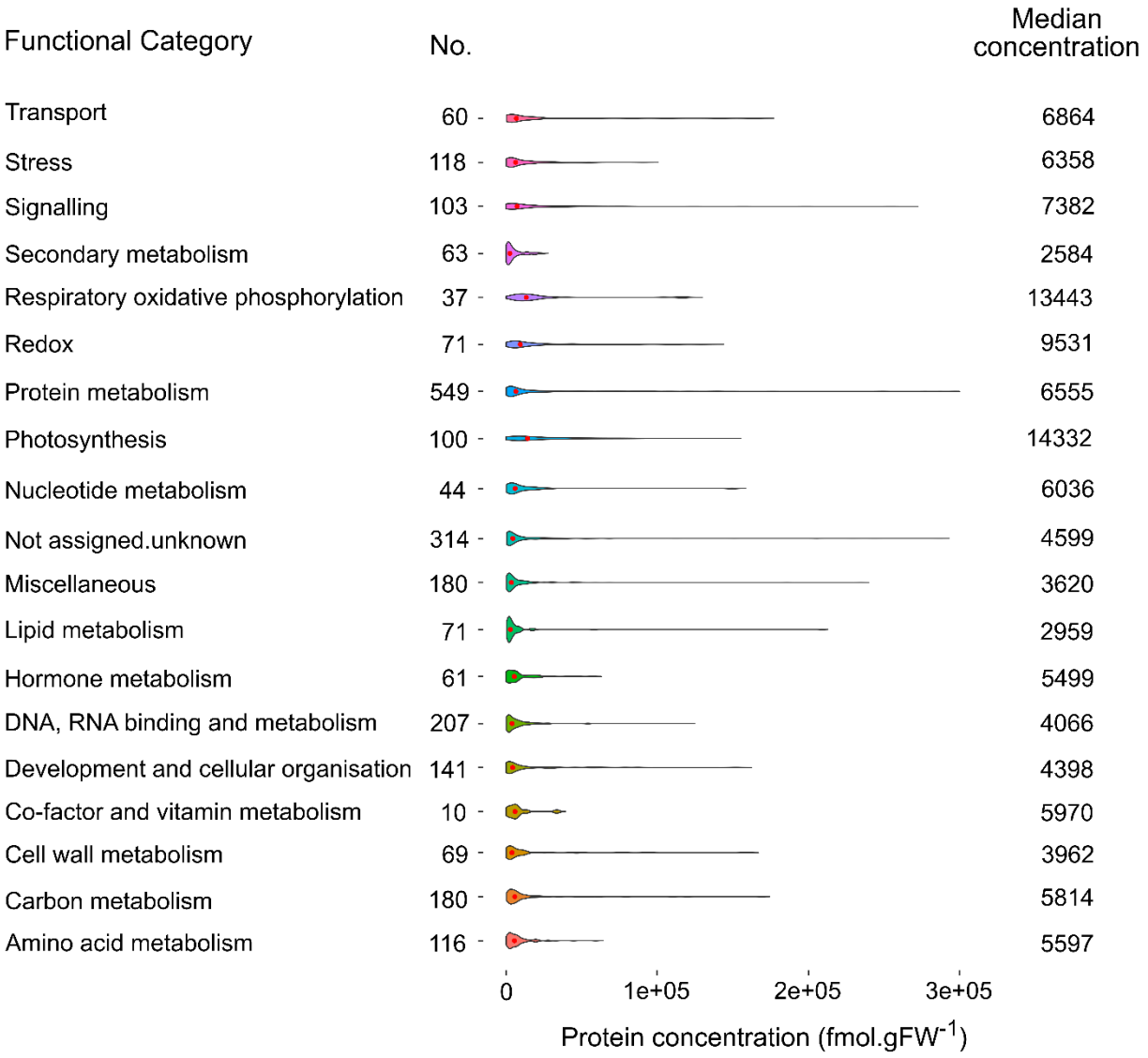


Figure I. 10 Functional categories associated the 2494 proteins according to the protein concentration. Concentration median was calculated for each protein throughout the development. For each functional category, the median of protein concentration associated was represented by a red dot. Number of proteins detected per functional category are mentioned on the left of violin plot. The violin plot is similar to box plots, except that they also show the probability density.

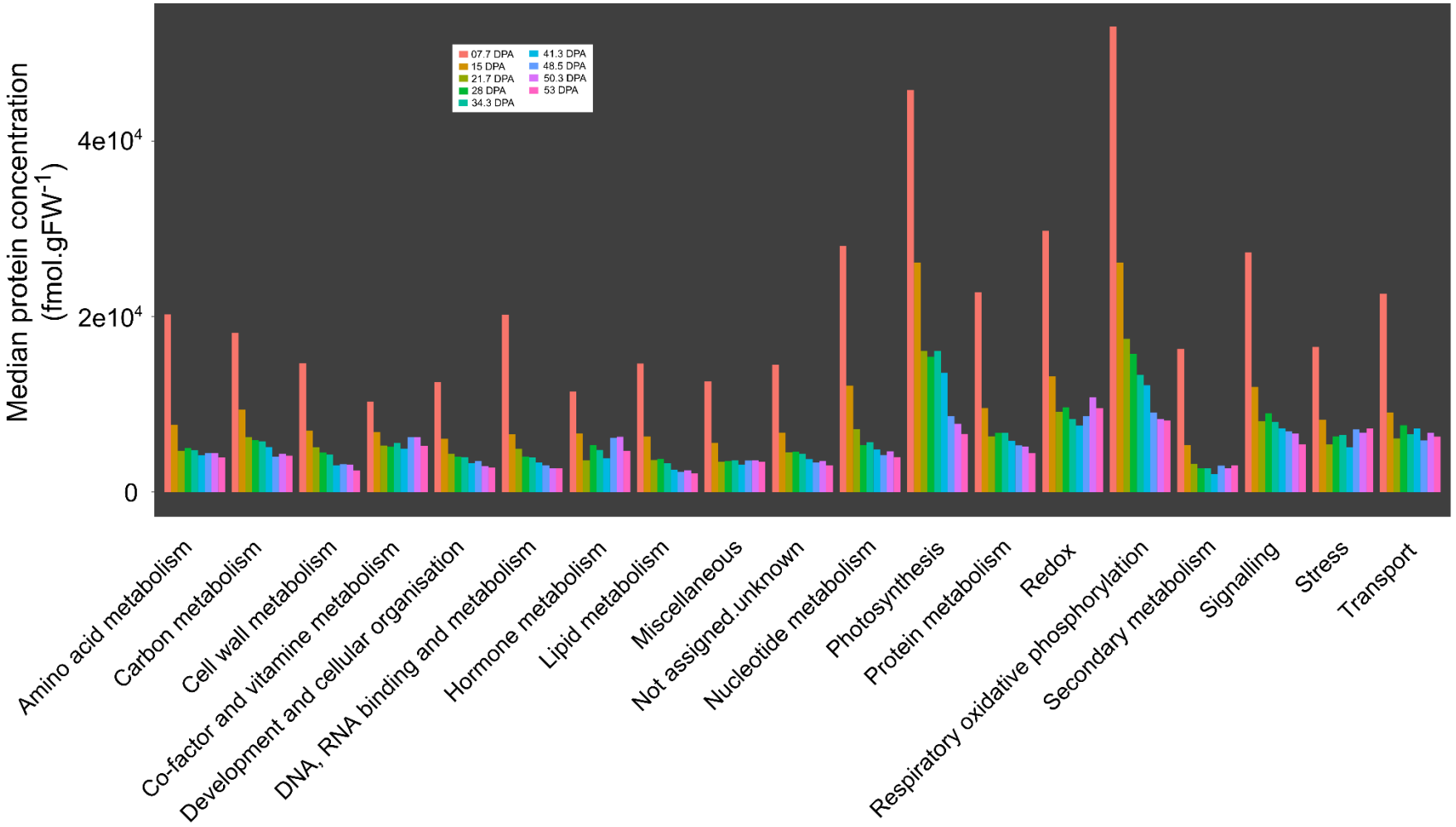


Figure I. 11 Protein concentration per functional category and at the nine developmental stages

Chapter 2 Absolute quantification of others Omics throughout tomato fruit development

I. Transcriptome of tomato fruit

In collaboration with GeT-Place (INRA Toulouse) for the sequencing and with Usadel'lab (RWTH Aachen University, Germany) for the mapping and quantification, we investigated the tomato (*S.lycopersicum* var Moneymaker) transcriptome throughout the tomato fruit development. For the clarity of the text, we used term 'transcript' referring to mRNA.

1.1 Absolute quantification of the tomato fruit transcriptome

As described in Materials and Methods section, this analysis has been performed at nine developmental stages, on the same samples than the ones used for the proteome analysis. Briefly, total RNA was extracted from frozen tissue of tomato pericarp (~100 mg) aged from 7.7 DPA to 53 DPA, cleaned-up from DNA and purified. The transcripts intactness, quantified with RIN (RNA Integrity Number), was satisfying with twenty samples with a RIN value higher than 7 and six samples with a RIN value between 5 and 7. Libraries were sequenced on Illumina sequencing machine and mapped on the ITAG 2.4 version of tomato genome (*Solanum lycopersicum* HEINZ assembly v2.40). Among the 34725 transcripts, 8403 transcripts were not detected in any of the 26 samples. Hypotheses to explain these 8403 non-detected transcripts were: (1) their concentrations were too low to be detected and quantified and/or (2) their expressions were out our developmental time-series. On the 26322 transcripts detected, we kept the transcripts that were expressed in at least the three replicates of at least one developmental stage. Thus, 3445 transcripts were removed. Finally, 22877 transcripts were considered and absolutely quantified by using spikes. As described in Material and Methods section, eight internal standards were spiked-in at the beginning of the total RNA extraction in each sample and used to calibrate the transcripts concentration ($fmol.gFW^{-1}$).

1.2 Cross-validation of absolute quantification of gene expression

The quantification by qRT-PCR of genes expression of 71 isoforms of enzymes was available in the lab. Alike the RNA-Seq protocol, eight internal standards have been used at known concentrations to determine an absolute quantification of the 71 genes expression by qRT-PCR (see Materials and methods). Thus, we compared these data with the absolute quantification of the expression of the same genes determined by RNA-Seq.

The qRT-PCR analysis was performed at the same nine developmental stages but only from samples harvested on the truss 6 (in triplicates) while RNA-Seq was performed on samples harvested on three trusses (5, 6 and 7, corresponding to three biological replicates). As already shown by Biais et al. 2014, the ANOVA and Tukey's tests on RNA-Seq data showed that transcripts concentrations were not statistically different from truss to truss. Thus we compared RNA-Seq and qRT-PCR results for 69 transcripts because two transcripts quantified by qRT-PCR were not detected by RNA-Seq (belonging to the 8403 transcripts removed). Only technique reasons were considered to explain this situation: (1) a damage of polyA tails of transcripts making impossible their amplification and detection during sequencing, (2) the mismatch of their primers used for the qRT-PCR leading to a "false quantification".

Absolute quantification determined by both RNA-Seq and qRT-PCR were compared to evaluate the quality of data, i.e. their absolute and relative accuracy. A good relative accuracy was expected with a high correlation between both methods of quantification, meaning a similar time-course (profile) of transcript as illustrated in Figure II. 1 for the gene expression of one isoform of fructokinase (Solyc06g073190.2.1). The correlation analysis (Spearman, Figure II. 2) performed on the 69 transcripts showed that the coefficients of determination were close to one (median $R^2_{spearman}$=0.87); and 81% of these coefficients were statistically validated ($P < 0.05$) (Figure II. 2). We also evaluated the relative accuracy by determining the slope – (a, expected to be close to one) - and the intercept – (b, expected to be small) - from the plot of the absolute quantifications determined by both RNA-Seq and qRT-PCR (Figure II. 1 and Figure II. 12). This analysis has been performed on not transformed data and also on log_{10}-transformed data because the first stages displayed often high values (see Figure II.1). Not surprisingly, the dispersion was lower for log_{10}-

transformed and satisfactorily the slope medians were close to one with the intercept medians close of zero (Figure II. 2C).

Then to evaluate the absolute accuracy, we calculated the ratio of the absolute quantifications determined by both RNA-Seq and qRT-PCR for the 69 transcripts and at the nine developmental stages. Ratios, displayed as a heatmap (Figure II. 3), were close to one (median ratio = 1.4) when all stages were considered, while it was clear that the absolute accuracy was altered at the last stage (53 DPA, mean ratio = 7.2).

With this analysis based on gene expression of 69 enzyme isoforms, we showed that RNA–Seq absolute quantification displayed globally similar results than qRT-PCR. While both quantification techniques have some limitations (such as the presence of identical reads biasing the transcripts quantification with a complex bioinformatic pre-analysis required to get transcript abundance for RNA-Seq and the need of gene reference for qRT-PCR), this cross-validation allowed us to use the entire quantitative dataset obtained by RNA–Seq for each gene expression throughout the tomato fruit development.

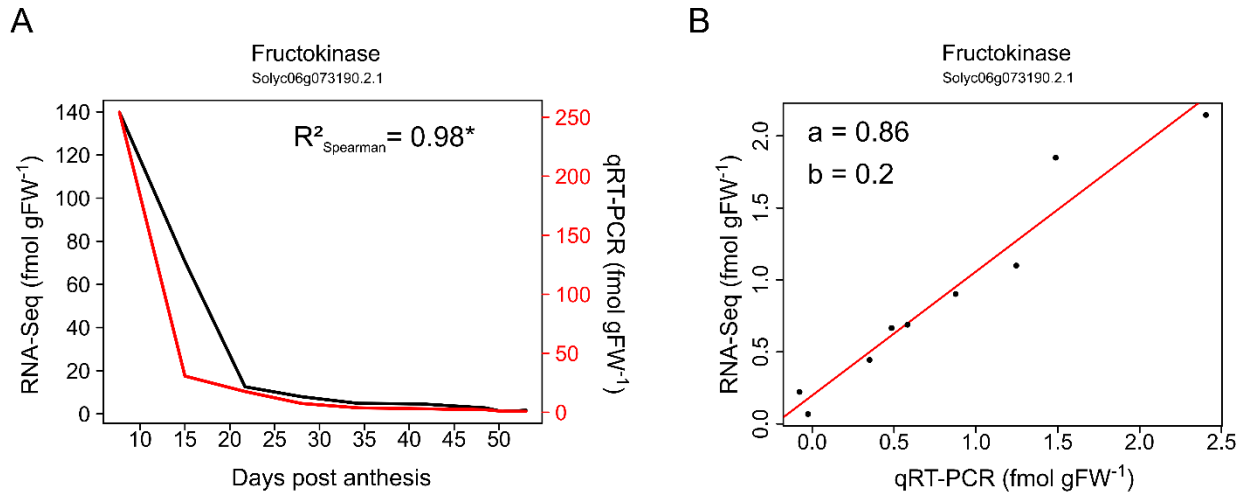


Figure II. 1 Transcript concentration of one fructokinase isoform quantified by RNA-Seq and qRT-PCR (in fmol.gFW^{-1}) represented versus time (A) displaying a significant coefficient of determination (R^2, Spearman) and (B) without time with the slope (a) and the intercept (b) determined from a linear regression.

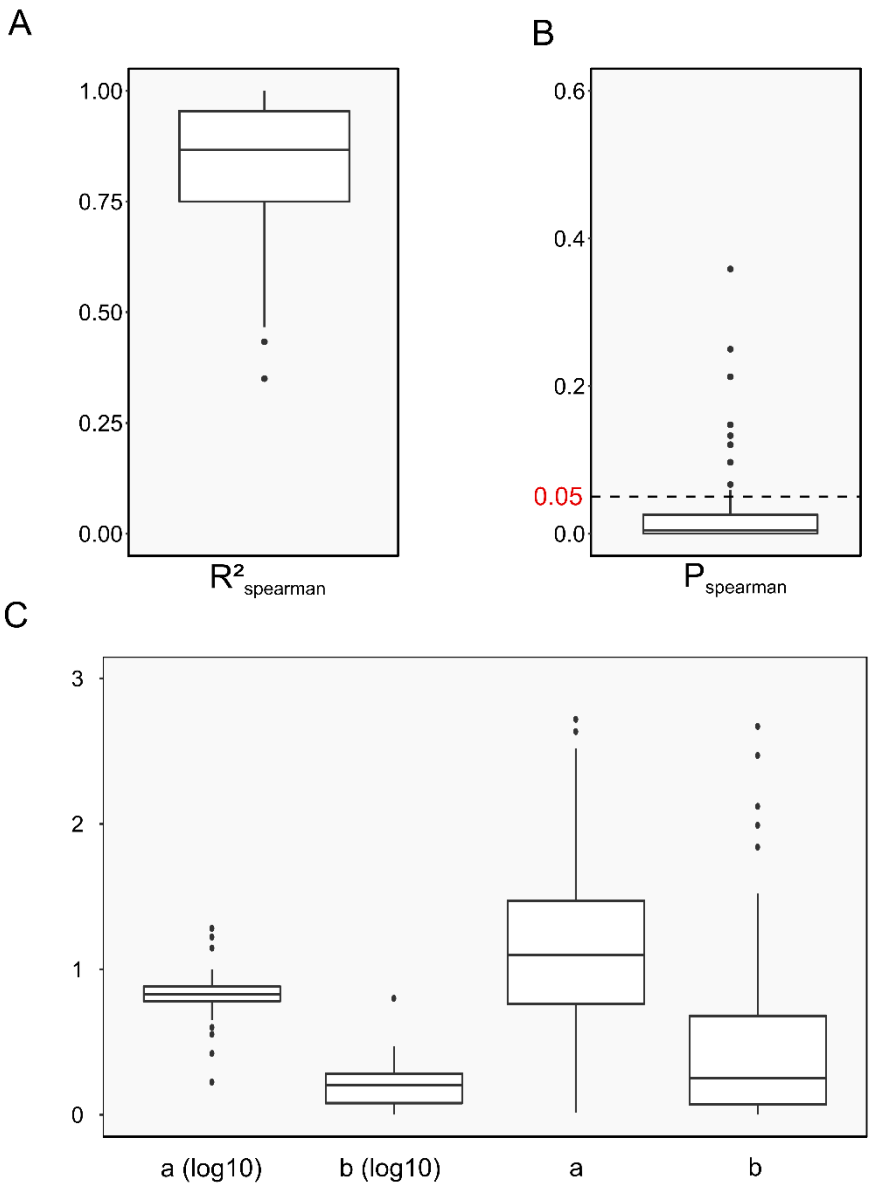


Figure II. 2 Relative accuracy of the absolute quantification determined by RNA-Seq and qRT-PCR on the 69 genes expression. The relative accuracy was first evaluated by performing a correlation analysis between gene expression determined by RNA-Seq and qRT-PCR (in fmol.gFW^{-1}). (A) Coefficient of determination (R^2, Spearman) and (B) the significance ($P < 0.05$, dashed line). Second, the relative accuracy was quantified with (C): the slope (a) and the intercept (b) of the 69 transcripts from the equation of the linear regression between concentrations quantified by RNA-Seq and qRT-PCR (in fmol.gFW^{-1}). The relative accuracy was evaluated on $\log_{10}$ transformed (a (log10), b (log10)) and not transformed (a, b) data.

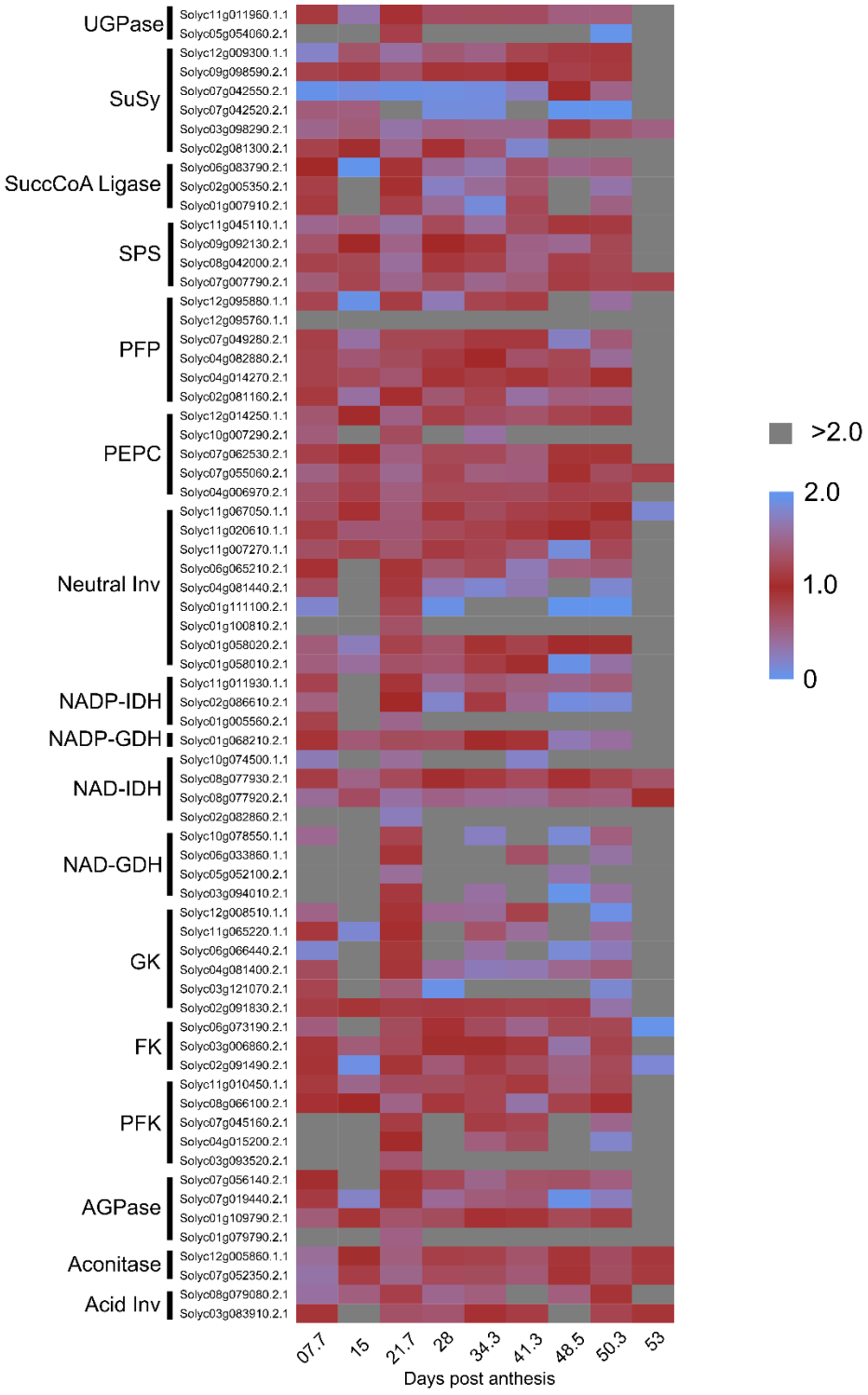


Figure II. 3 Absolute accuracy of the absolute quantification determined by RNA-Seq and qRT-PCR. The absolute accuracy was evaluated by the ratio of the absolute quantifications determined by both RNA-Seq and qRT-PCR expected close to one. Ratios were calculated for the 69 transcripts (y-axis) and at the nine developmental stages (x-axis).

1.3 Changes in transcript expression throughout tomato fruit development

With the absolute quantification of transcripts by RNA-Seq, we plotted, for each developmental stage, the distribution and the median of the 22877 transcripts concentrations (Figure II. 4). Similarly to proteins concentrations (Figure I. 7), the most notable change of transcripts concentrations occurred during cell division, between 7.7 DPA and 21.7 DPA. Indeed, during this period the median of transcripts concentrations was divided by 10, from 2.09 $fmol.gFW^{-1}$ at 7.7 DPA to 0.22 $fmol.gFW^{-1}$ at 21.7 DPA. Then, the median reached almost a plateau from cell expansion phase (21.7 DPA) to the end of the development (53 DPA) (Figure II. 4B). A slight increase of the median was observed at 48.5 DPA.

Then, a hierarchical clustering analysis was performed on transcript concentrations, using Pearson's correlation on mean centered values scaled to unit data, and displayed as a heatmap to investigate whether changes of transcripts concentrations (clusters) could be assigned to developmental phases (Figure II. 5). The 22877 transcripts were separated in seven clusters. In the first cluster (1120 transcripts), transcripts were more abundant during ripening (48.5-53 DPA). In the second cluster (760 transcripts) transcripts were more abundant during the cell division (7.7 DPA) and ripening phase (48.5-53 DPA). The third cluster (251 transcripts) was characterized by transcripts more concentrated during the cell expansion and turning phase (28-41.3DPA). The fourth cluster (407 transcripts) was determined by a "punctual" increase of transcripts concentration at 15 DPA. In the fifth cluster (2531 transcripts) transcripts followed the same profile of the third cluster but with a slighter increase during ripening (48.5-53 DPA). The sixth and seventh clusters (5291 and 12517 transcripts respectively) shared more than 77% of the transcripts. In these two clusters, transcripts were highly concentrated from cell division to cell expansion phases and decreased to reach a plateau until the end of the development.

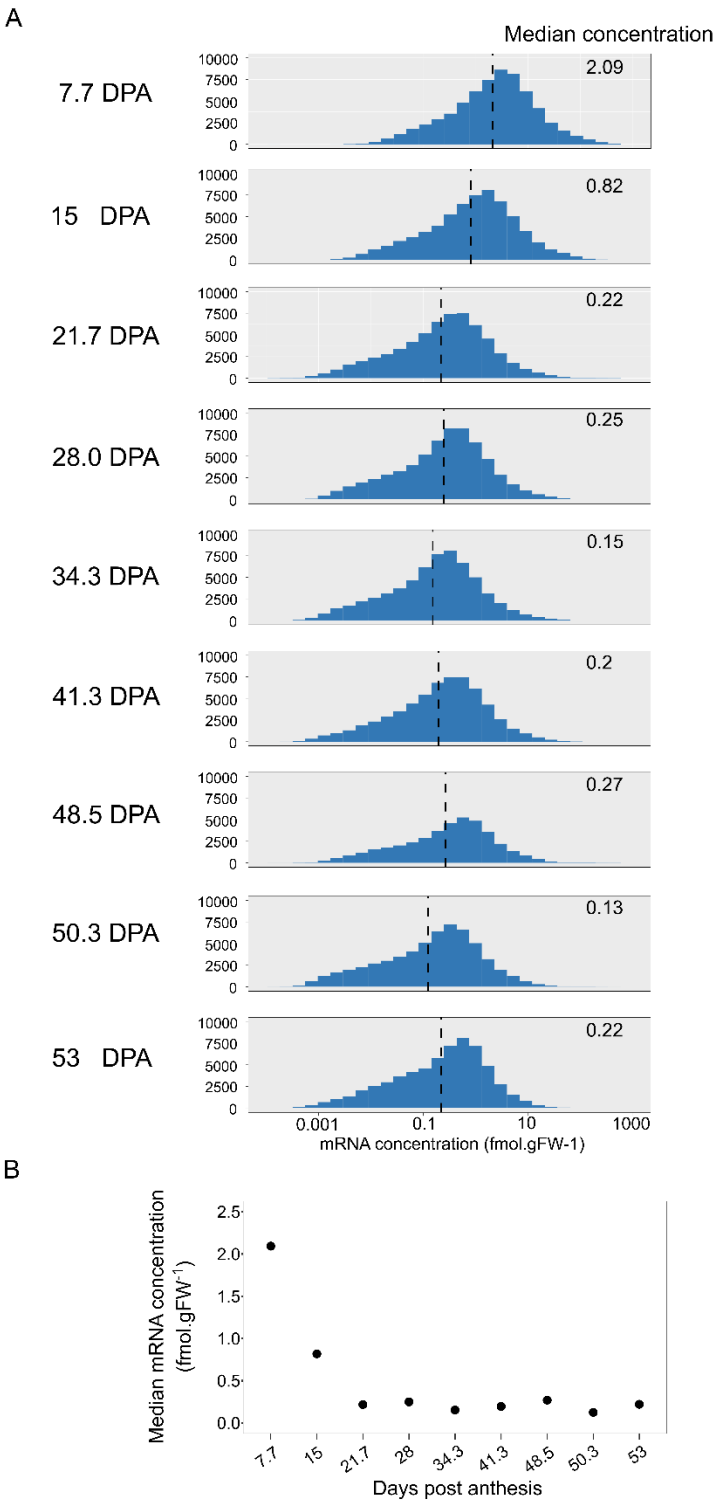


Figure II. 4 Distribution of transcripts concentrations. (A) Distribution of transcripts concentration (log_{10} scale) at the nine developmental stages. Medians of concentrations were represented by a dashed line with the value mentioned in the right corner. (B) Time-course of the median concentration of mRNA throughout the tomato fruit development.

Figure II. 5 Overview of transcripts concentration changes. The clustering analysis was performed on transcript concentrations (Pearson's correlation) mean centered and scaled to unit data. Columns correspond to the nine developmental stages, and rows correspond to transcript concentrations. The number of transcripts contained by the five clusters (red triangles) are indicated on the right of the heatmap.

1.4 Analysis of functional categories of 22877 tomato transcripts

We investigated the functional categories associated to the 22877 transcripts. For this purpose, we used an in-house reduced version of the modified MapMan BIN code containing 19 functional categories instead of the 35 initially available (Thimm et al., 2004).

First, we analyzed for each of the eight clusters described in Figure II. 5, how transcripts were distributed according their functional categories (Figure II. 6) and paid attention to those containing the most transcripts. Thus, across the eight clusters, we identified seven functional categories for which at least one cluster more than 5% of the transcripts of one cluster were assigned: "Development and cellular organization" (clusters 5, 6 and 7), "DNA, RNA binding and metabolism" (all clusters), "Miscellaneous" (clusters 1, 2, 3, 4 and 6), "Protein metabolism" (all clusters), "Signaling" (clusters 1, 2, 3, 5 and 6), "Stress" (clusters 1, 2, 3, 4), "Hormone metabolism" (cluster 1) and "Not assigned. Unknown" (all clusters) (Figure II. 6).

Not surprising, transcripts more concentrated during the cell division (clusters 2, 5, 6 and 7) were mainly associated to the "Protein metabolism" (10.5%, 16.2%, 10.4% and 17.6% of transcripts respectively). In parallel, 19% of transcripts in the clusters 5 and 19.1% of transcripts in the cluster 6 were associated to the "DNA, RNA binding and metabolism" category. This result was in agreement with the high metabolic activity associated to cell division required for biosynthesis and growth.

Transcripts up-regulated during ripening (clusters 1 and 2) were also mainly associated to the "DNA, RNA binding and metabolism" (11.5% and 13.4%, respectively) in agreement with a cell reprogramming during ripening phase where the "Stress" category (6.4% and 6.3%, respectively) reached the highest percentage among the seven clusters. Besides, transcripts in cluster 1 were allocated to the "Hormone metabolism" (5.4%) which is coherent with the essential role of hormones, such as ethylene and auxin, in the tomato fruit maturation (Gillaspy et al., 1993). The "Miscellaneous" category, which grouped a wide variety of enzyme activities, was found in most of transcript profiles (cluster 1, 2, 3, 4, 6) and more specifically with transcripts having a peak of

concentration at one stage of the development (cluster 1 with a peak at 48.5 DPA: 10.4%, cluster 3 with a peak at 41 DPA: 10.5% and cluster 4 with a peak at 15 DPA: 10.3%).

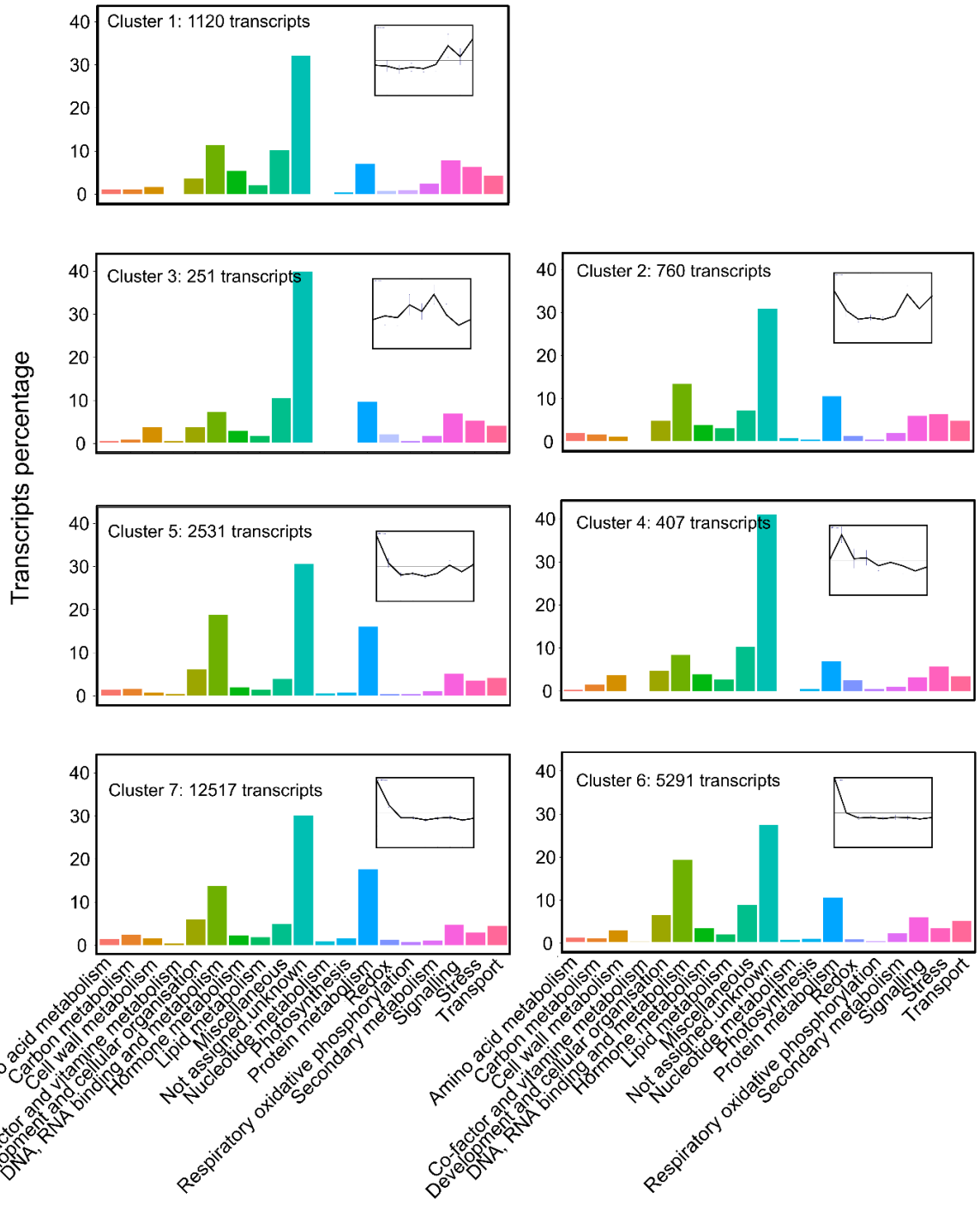


Figure II. 6 19 functional categories associated the 22877 quantified transcripts separated in seven clusters according to the hierarchical clustering (see Figure II. 5). Profiles of transcripts and numbers of transcripts associated to the cluster were mentioned on each corresponding plot.

Then we investigated the functional categories according to transcripts concentrations to determine the most "concentrated" functional categories (Figure II. 7). For this purpose, we calculated the median of concentrations, first of each transcript throughout the development and second of transcripts in each functional category.

"Protein metabolism" and "DNA, RNA binding and metabolism" categories together represented more than 27% of transcripts with 3088 and 3210 transcripts, respectively. More than a thousand transcripts were assigned to the "Signalling", "Miscellaneous" and the "Development and cellular organization" categories (1090, 1285 and 1238, respectively). Note that categories containing the most of transcripts didn't necessarily coincide with highest concentrations, such as for the last three mentioned. Indeed, the three highest median concentrations were associated to the "Respiratory oxidative phosphorylation", "Photosynthesis" and "Redox" categories (1.01, 0.94 and 1.10 $fmol.gFW^{-1}$ respectively) containing 114, 207 and 242 transcripts, respectively. Transcripts in the "Respiratory oxidative phosphorylation" category were involved for instance in the mitochondrial electron transport and ATP synthesis (Cytochrome C, F1-ATPase, NADH dehydrogenase (Complex I)…) and transcripts in the "Redox" category were related to metabolism ascorbate, thioredoxin and xenobiotic biodegradation. Conversely, the three lowest median concentrations were associated to the "Miscellaneous", Hormone metabolism" and "Secondary metabolism" (0.11, 0.11 and 0.12 $fmol.gFW^{-1}$ respectively).

A similar analysis, performed at each developmental stage (Figure II. 8) showed a slight increase of median concentration at 48.5 DPA for most of the functional categories ("Amino acid metabolism", "Carbon metabolism", "Co-factor and vitamin metabolism", "Photosynthesis", "Redox" and "Respiratory oxidative phosphorylation"). This increase was followed by a decrease at 50.3 DPA and an increase at 53 DPA.

Altogether these results clearly showed that the cell division and expansion phase required a high abundance of transcripts associated to "Photosynthesis", "Redox" and "Amino Acid" categories (Figure II. 8) in agreement with the mitosis activity and the increase of cells number and size. They also suggested a reactivation of pathways related to the energy metabolism ("Respiratory oxidative phosphorylation", "Carbon metabolism") occurring at 48.5 DPA, in agreement with important metabolic changes occurring at ripening.

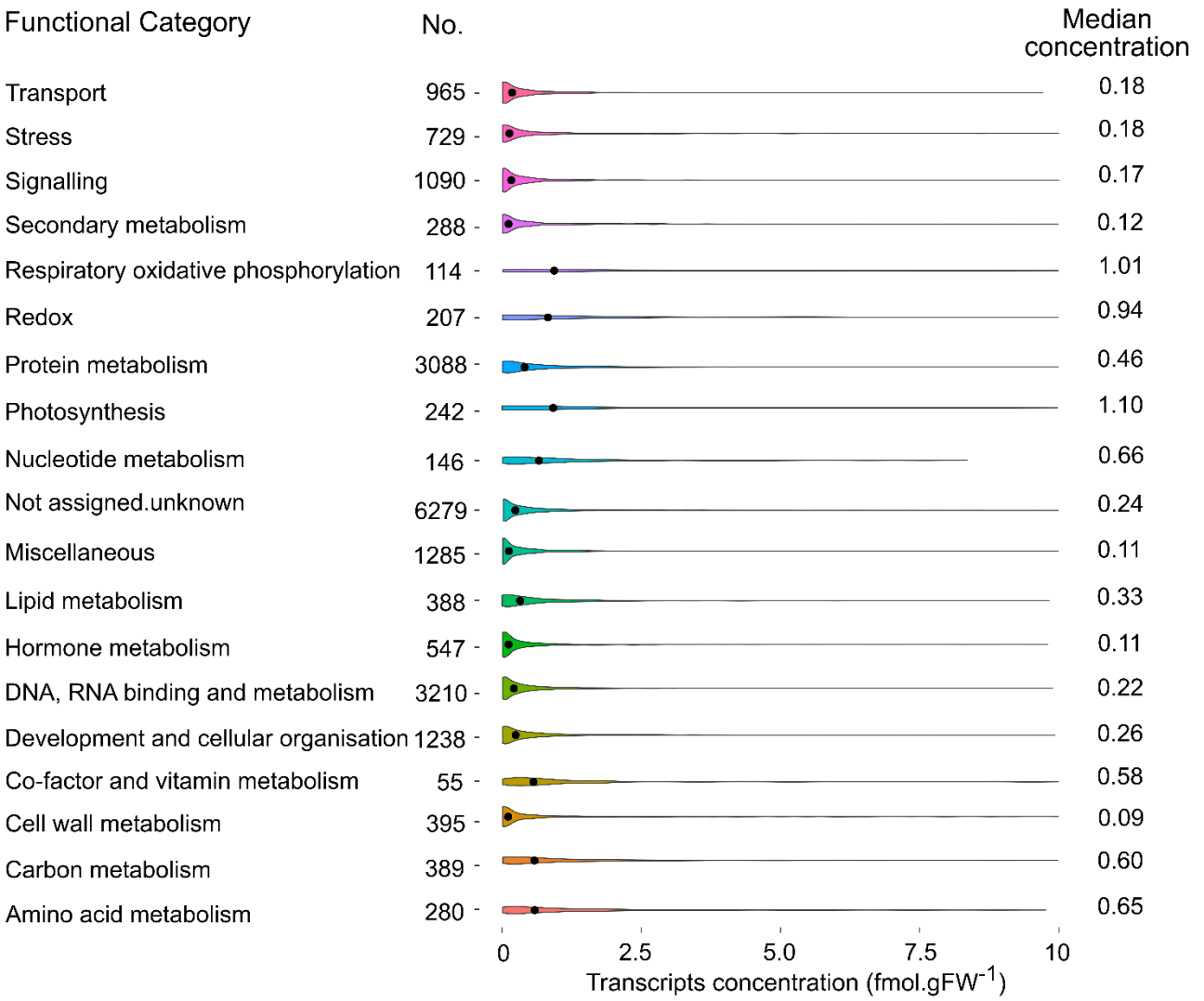


Figure II. 7 Functional categories associated the 22877 transcripts according to the transcripts concentrations. Median concentration (black dot) was calculated for each transcript throughout the development first and then for each functional category. Number of proteins detected per functional category are mentioned on the left of violin plot.

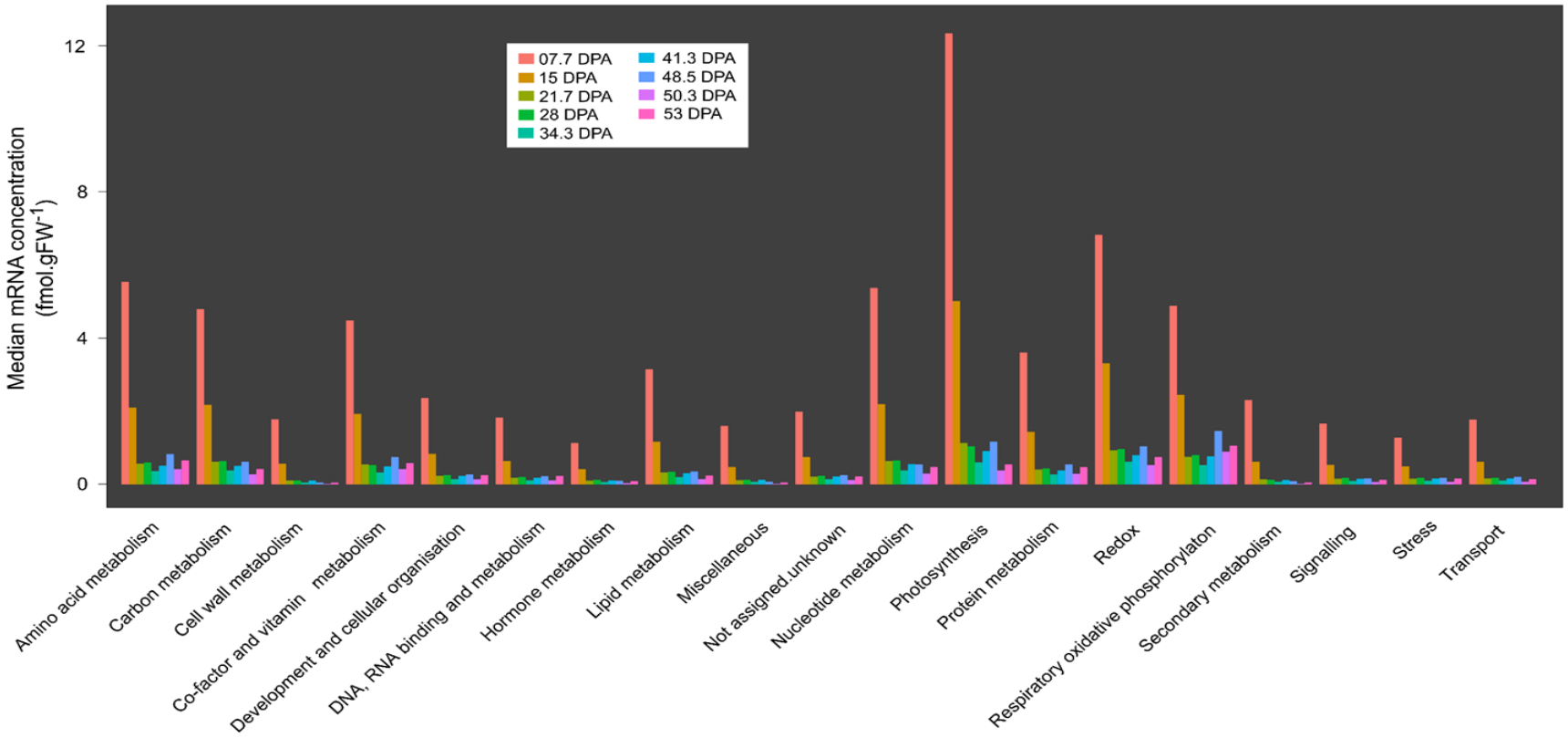


Figure II. 8 Median transcript concentrations per functional category and at the nine developmental stages.

1.5 A concentration in mole per volume of cytoplasm: a more realistic normalization for transcripts

Transcription is a cellular process ending by the export of the newly synthesized transcript from the nucleus to the cytoplasm that we assumed as being the main location of all transcripts. In order to express transcripts concentration more realistically we calculated each transcripts concentration on a cytoplasm-volume basis. Using the morphometric data and making several assumptions (on the shape of the cell and subcellular compartments, the distribution of amyloplasts in the cytoplasm, the cell-wall delimitation), Beauvoit et al. 2014 determined time-dependent functions describing changes of the vacuole volume fraction (Vvac, in $mL.mL_{tissue}^{-1}$, Equation II. 1), cytoplasm volume fraction (Vcyt in $mL.mL_{tissue}^{-1}$, Equation II. 2) and tissue density ($gFW.mL_{tissue}^{-1}$, Equation II. 3). These three equations were used to calculate Vvac, Vcyto and tissue density at the nine stages of the developmental time-series (Figure II. 9 A,B). The cytoplasm volume ($mL_{cyto}.gFW^{-1}$) was calculated by dividing the cytoplasm volume fraction by the tissue density (Figure C).

$$Vvac = 0.853 \left(1 - \exp\left(\frac{-2292 - Time}{10633}\right)\right) \qquad \text{Equation II. 1}$$

$$Vcyt = (0.933 - Vvac)/1.13 \qquad \text{Equation II. 2}$$

$$\rho = \left(0.075 \left(\frac{Time}{1440}\right) + 13\right) \Big/ \left(0.075 \left(\frac{Time}{1140}\right) + 12\right) \qquad \text{Equation II. 3}$$

, where ρ the tissue density (in $gFW.mL^{-1}$ tissue) and V_{vac} and V_{cyt}, the volume fraction (in $mL.mL_{\ tissue}^{-1}$) of vacuole and cytosol respectively.

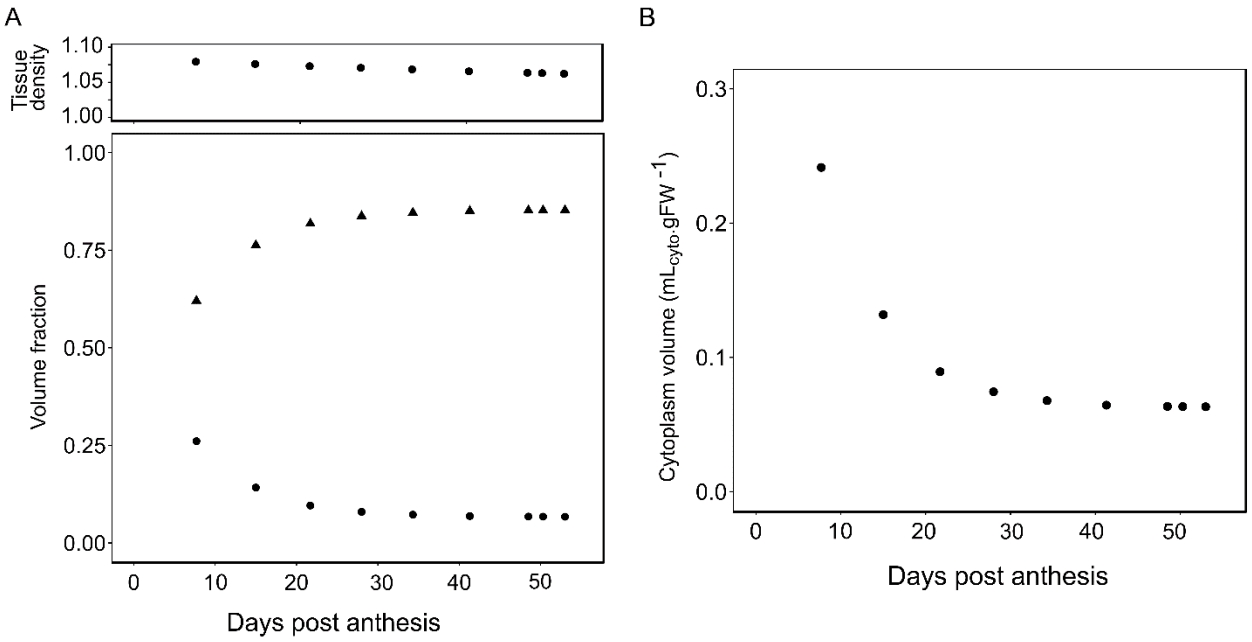


Figure II. 9 Determination of cytoplasm volume throughout the tomato fruit development From equations determined in Beauvoit et al., (2014), the pericarp density in gFW.mLtissue-1 (A), the vacuole (▲) and cytoplasm (●) volume fractions (mL.mLtissue-1) were calculated at the nine developmental stages. The cytoplasm volume (mLcyto.gFW^{-1}) was deduced by dividing the cytoplasm volume fraction by the tissue density.

Finally, transcripts concentration on cytoplasm-volume basis (fmol.mL$_{cyto}^{-1}$) was obtained by dividing transcripts concentration (in fmol.gFW^{-1}) by the cytoplasm volume (mL$_{cyto}$.gFW^{-1}).

This change of normalization lead to a global increase of the concentration, as the transcripts were more concentrated in the cytoplasm. The median of transcripts concentrations displayed a similar shape throughout the development with values decreasing from 8.38 fmol.mL^{-1} to 3.49 fmol.mL^{-1} between 7.7 DPA and 53 DPA. Interestingly, this normalization highlighted changes of concentrations, as more irregularities which appeared from 34.3 to 53 DPA as illustrated Figure II. 10 for the median of transcripts concentrations and Figure II. 11 on the fructokinase enzyme (Solyc06g073190.2.1) as an example.

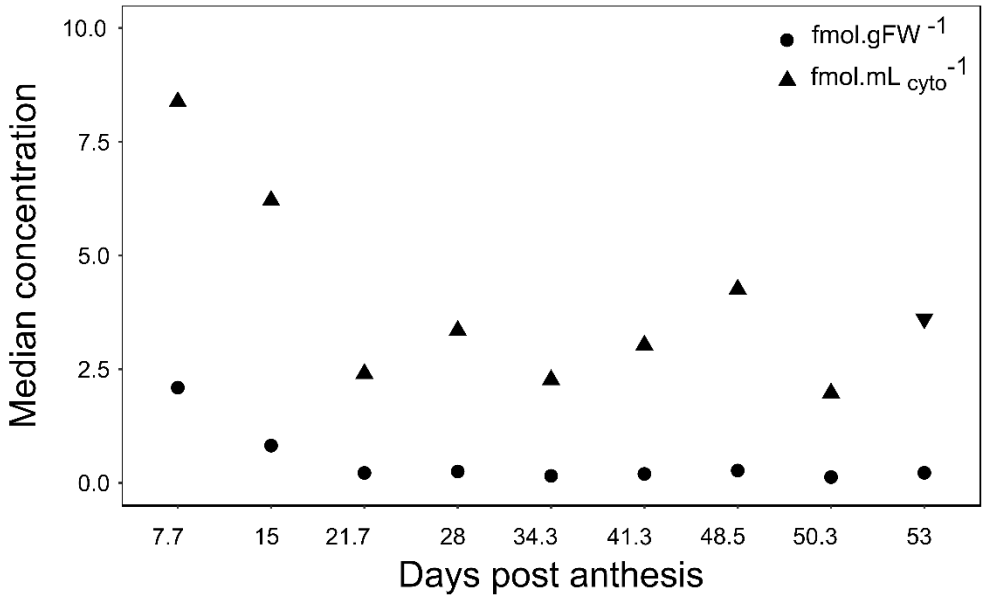


Figure II. 10 Time-course of the median of transcripts concentrations on a gFW basis (black circle) and on a cytoplasm volume basis (triangle).

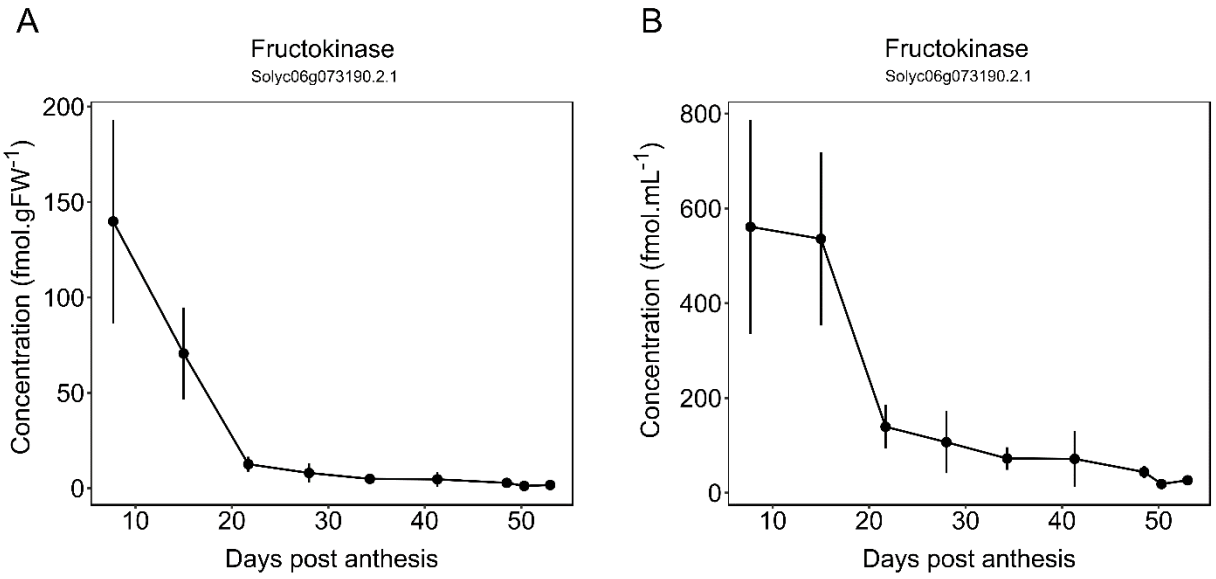


Figure II. 11 Time-course of the concentration of the fructokinase transcript (Solyc06g073190.2.1) in fmol.gFW^{-1} (A) and in fmol.mL$_{cyto}^{-1}$ (B) using the volume of cytoplasm (mL$_{cyto}$.gFW^{-1}) determined at the nine developmental stages (Figure II. 9).

II. Addition of Metabolome toward an integrative analysis

2.1 Analysis of metabolites changes in growing tomato fruit

In this section, we first described metabolomic data that completed proteomic and transcriptomic data detailed in previous sections. Then, including activome data (Biais et al 2014), we integrated four levels of omics in a descriptive analysis.

To follow metabolites changes occurring during the tomato fruit development and ripening, a quantitative metabolic profiling was carried out using four analytical techniques: enzymology, Mass Spectropmetry (MS), HPLC-DAD and NMR methods (See Materials and methods) resulting in the quantification of more than one hundred targeted metabolites. To avoid metabolite duplicates, when a metabolite was quantified by more than one technique, profiles were compared and we kept the best quantification (the most precise, *i.e.* providing the lowest CV (%) between replicates). In this study, we analyzed 77 metabolites expressed in absolute quantification ($\mu mol.gFW^{-1}$). Four metabolites, called Unknown (XX.XX), were quantified in $UA.gFW^{-1}$. Metabolites were expressed in gram fresh weight basis because of the limited knowledge about the subcellular localization of metabolites at the nine stages of tomato fruit development.

A hierarchical clustering analysis displayed as a heatmap was performed to provide an overview of metabolites changes throughout the tomato fruit development (metabolites concentrations were mean centered and scaled to unit and clustered Pearsons' correlation) (Figure II. 12). Clustering analysis distinguished four profiles. The first cluster grouped 26 metabolites accumulated more intensely after 48.5 DPA. The second and third clusters grouped 8 and 11 metabolites with opposite expressions, i.e. following "high-low-high" expression for the second cluster and "low-high-low" expression for the third cluster. The last cluster (30 metabolites) shared metabolites with a highest expression during cell division.

In the following section, metabolites were not described according to the cluster but the sub-family of metabolites they belong, such as pigment, organic acids, sugars, amino acids.

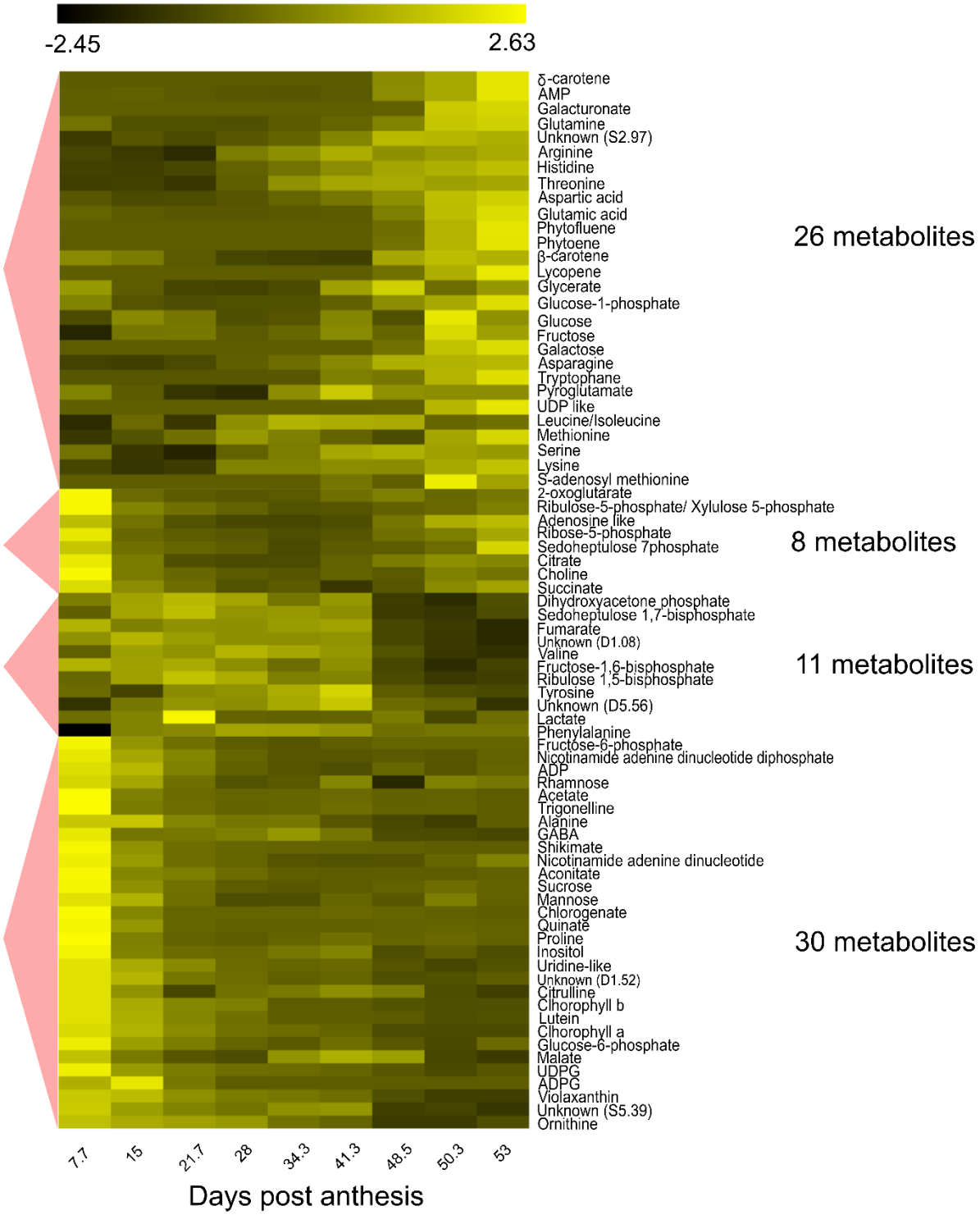


Figure II. 12 Overview of changes of metabolite concentration with a clustering analysis performed (Pearson's correlation, values mean centered and scaled to unit data) with columns corresponding to the nine developmental stages and rows to metabolites. The number of metabolites contained in the four clusters (red triangles) were indicated on the right of the heatmap. Metabolites concentration was expressed in µmol.gFW^{-1}, apart for four metabolites called Unknown (UA.gFW^{-1}).

First, pigment content was measured at the nine stages of tomato fruit development, with chlorophylls *a* and *b*, lutein, violaxanthin, β and δ carotenes and two carotenoid precursors, phytoene and phytofluene. As expected, carotenes, lycopene were accumulated at the last stages (Figure II. 12, cluster 1) and chlorophylls *a* and *b* were accumulated at earliest stages (Figure II. 12, cluster 4). Similarly to Carrari et al., (2006), chlorophyll a and b represented about 80% of the total pigment content from 7.7 PDA to 41.3 DPA where chlorophyll *a* content represented 68% (Figure II. 13). Lutein and violaxanthin were not detected after 41.3 DPA. During maturation, from 48.5 to 53 DPA, there was a drastic change in the pigment composition; chlorophyll was replaced by lycopene (67%), carotenes (9.2%) and both carotenoids precursors, phytoene (8.6%) and phytofluene (14.68%), lutein and violaxanthin becoming negligible.

About sugars, six soluble sugars (glucose, fructose, sucrose, mannose, rhamnose and galactose) and three sugar phosphates (glucose-1-phosphate (G1P), glucose-6-phosphate (G6P), and fructose-6-phosphate (F6P)) were quantified. Note that fructose and glucose were the most abundant, about ten times higher concentrated (in median throughout development) than other sugars. Fructose and glucose and also galactose were accumulated at the end of the development (Figure II. 12, cluster 1) while mannose, rhamnose and sucrose were accumulated during cell division (Figure II. 12, cluster 1). The G1P accumulation from turning phase was suggested by Biais et al., (2014) to be related to starch degradation occurring at the same period with also a net increase of sugar import.

G1P, UDP and AMP were significantly negatively correlated to UDPG ($R^2_{spearman}$ = -0.44, $R^2_{spearman}$ = -0.75 and $R^2_{spearman}$ = -0.47, respectively). In parallel, G1P was negatively correlated to ADP (Figure II. 12, cluster 3, $R^2_{spearman}$= -0.12), AGPG (Figure II. 12, cluster 3, $R^2_{spearman}$= -0.47) which is coherent with the conversion of G1P into ADPG by AGPase. The intermediate metabolites (R5P, Ru5P, X5P, DHAP, FBP, S7P and SBP) were mainly accumulated during cell division and expansion phases (Figure II. 12, cluster 2 and 3). We noticed that R5P and S7P, intermediates of the pentose phosphate pathway, were also accumulated at ripening (50.3-53 DPA).

The main organic acids, (citrate, malate, fumarate, aconitate, succinate, 2-oxoglutarate, shikimate, chlorogenate, quinate) were quantified at the nine stages of tomato fruit development. Citrate and malate were the major organic acid detected. With concentrations (average throughout the development) of 10.1 and 8.6 µmol.gFW^{-1} respectively, they were about two hundred times more abundant than the others. Aconitate, malate, chlorogenate, quinate, shikimate were grouped

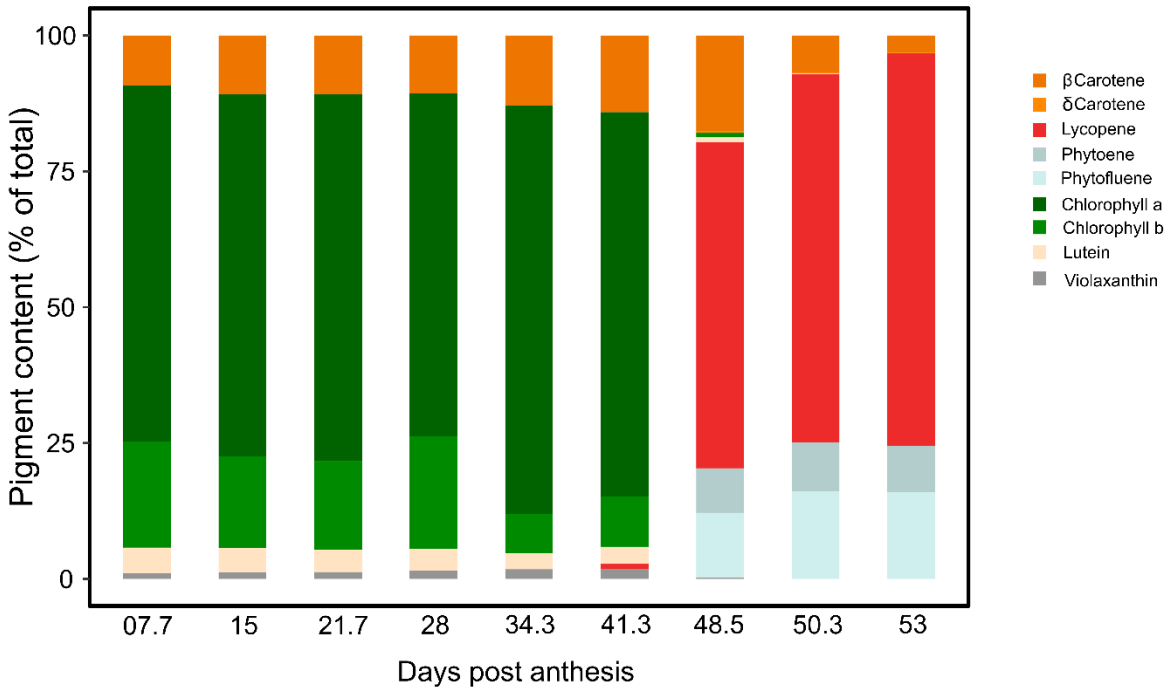


Figure II. 13 Pigment content during tomato fruit development, percentages of the total measured: chlorophylls a and b, α and δ carotenes, lutein, violaxanthin, lycopene, phytofluene and phytoene in µmol.gFW^{-1}.

in cluster 4 and accumulated during cell division while fumarate was in cluster 3 more accumulated before ripening.

Citrate, succinate and 2-oxoglutarate (2OG) were accumulated during the cell division and also during the ripening phase (cluster 2). As already suggested by Biais et al., 2014, these changes of metabolites mirrored changes in enzyme activities especially in TCA cycle pathway. For instance, citrate changes can be related to the citrate synthase enhanced during ripening phase (Biais et al., 2014).

Finally, three profiles were observed for amino acids, corresponding to clusters 1, 3 and 4 with a majority of amino acids (63%) accumulated during ripening (Figure II. 12, cluster1): tryptophane, threonine, serine, S-adenosylmethionine, methionine, lysine, histidine, glutamine, glutamate, aspartate, asparagine, arginine, leucine isoleucine and pyroglutamate. Conversely, proline, ornithine, GABA, citrulline and alanine were accumulated during cell division followed by a sharp decrease (Figure II. 12, cluster 4) while tyrosine, valine and phenylalanine were accumulated during the cell division and then slightly decreased (Figure II. 12, cluster 3). These results were in

agreement with the high activity of aminotransferase enzymes, especially during maturation (Biais et al 2014) suggesting a diversity of carbon sources required when sugar supply is too low, as suggested by Ishizaki et al., (2005).

To conclude, the metabolomics data described here were in agreement with the previous results described on Moneymaker tomato cultivar (Carrari and Fernie, 2006; Biais et al., 2014) and, with some extent, with Ailsa Craig tomato variety (Osorio et al., 2011). This dataset was then combined with the three others omics datasets of proteome, transcriptome and activome. Activome contained 36 enzyme activities involved in primary metabolism (carbohydrates metabolism, glycolysis, Calvin Benson cycle and organic acids metabolism) quantified at the same nine stages of tomato fruit development. These enzymes activities have been published (Biais et al., 2014), and previously used in Chapter 1.II.b to cross-validate the absolute quantification of protein quantified by label-free LC-MS/MS.

2.2 An integrative analysis of four omics data

In order to get an overview of what happened throughout the development of the tomato fruit, we integrated the four datasets: proteome, transcriptome, activome and metabolome. This analysis comprised 22877 transcripts, 2494 proteins, 36 enzyme activities and 77 metabolites. All variables were quantified in an absolute way. To be compared with each other, variables including enzyme activities were expressed on a gram fresh-weight basis unless transcriptome which was expressed on a cytoplasmic volume basis.

Given the large number of variables, a principal component analysis (PCA) was performed for the four datasets (Figure II. 14A) with variables averaged by developmental stage, mean centered and scaled. Interestingly, whatever the biomolecular level considered, PCA plots displayed a similar profile, schematized in Figure II. 14B. This profile was characterized by a first component explaining the highest percentage of variance and separating green stages (7.7 - 41.7 DPA) from ripening stages (48.5 - 53 DPA) while the second component segregated first and last stages (7.7 DPA, 48.5 - 53DPA). This PCA profile has already been reported to describe the development of tomato fruit (Biais et al., 2014; Szymanski et al., 2017), pear (Oikawa et al., 2015) and berry (Savoi et al., 2017). And it has been proposed that the variance of the first component was mainly linked

to developmental phases while the variance of the second component could involve metabolic transitions (Biais et al., 2014).

We focused on two main events observed on the four PCAs (Figure II. 14B): the first (named GAP1) corresponded to the gap between green and red stages, i.e. the ripening transition (between 34.3 – 41.3 DPA and 48.5 – 50.3 DPA) and the second event (named GAP2) which bring back the last stage (53 DPA) at the same level than the first stage (7.7 DPA) in the second component.

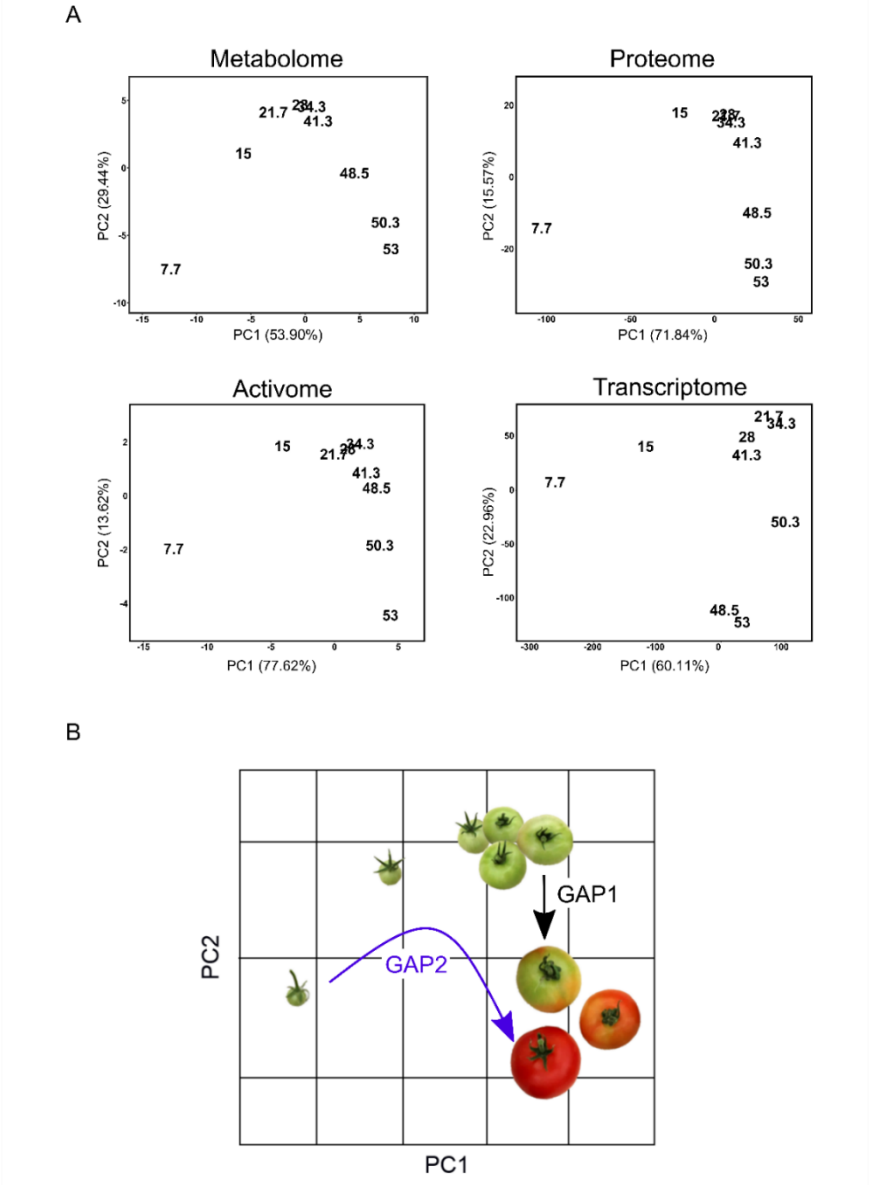


Figure II. 14 PCA performed on metabolome, proteome, enzyme activities and transcriptome datasets: (A) all variables expressed on a gFW basis unless transcriptome expressed on a cytoplasmic volume basis (B) Schematic PCA performed according to the four PCA plotted in (A).

The workflow used to analyze these two events for the four omics was the same. First, we filtered out variables not sensitive to the development (ANOVA, criteria to be filtered out $P > 0.05$). At this stage, 18327 transcripts, 2128 proteins, 78 metabolites and 35 enzyme activities remained. Then, we identified (1) variables with significant concentration changes between stages involved in the GAP1 ($P< 0.01$, FDR) and (2) variables without significant concentration changes (i.e. similarly expressed) at first and last (7.7 and 53 DPA, respectively) for the second event (GAP2). To avoid false positives variables, induced by a high dispersion between replicates, we selected only variables with a CV lower than 30% at both stages. Number of variables finally obtained for the two events and for the four omics presented in Table II. 1.

Table II. 1 Number of metabolites, proteins, activome (enzyme activities) and transcripts involved in GAP1 and GAP2.

	GAP1 (34.3+41.3 DPA vs 48.5+50.3 DPA)	GAP2 (7.7 DPA vs. 53 DPA)
Transcriptomic	1058	545
Proteomic	449	87
Metabolomic	33	14
Activome	1	0

In the following section, we described the functional analysis of variables involved in both events (GAP1 and GAP2) integrating the four omics levels.

- GAP1

PageMan was used to investigate and condense 449 proteins and 1058 transcripts involved in GAP1. PageMan used an Wilocoxon rank sum test statistic (nonparametric test statistic) which determined if the median of fold-change within a particular functional categorie group (BIN) was the same as the median fold-change of all variables not in that functional categorie. In order to diplayed p-values in PageMan, they are transformed into their respective z-values (Z-score). All p-values above 0.05 are set to a Z-score of 0 to avoid misinterpretation. The resulting values are than false color coded in a two color scale (blue-red). A highly saturated color indicates a high absolute value, whereas smaller values are indicated by a lower color saturation. Thus, blue and red

distinguished categories where the average of the signals of variables in a category increases and decreases. Proteins up and down-regulated (157 and 292 proteins, respectively) and transcripts up and down-regulated (763 and 295 transcripts, respectively) at the beginning of ripening (48.5-50.3 DPA) were separated and visualized using PageMan (Figure II. 15). PageMan diagrams used a false-color code, blue corresponded to an overrepresented category compared to the global distribution.

Coherently with previous analyses (heatmap and functional categories) performed on all proteins and transcripts datasets, the ripening transition (GAP1) was marked by a significant changes of pigments (carotenes, lycopene, chlorophyll…), proteins and hormone (ethylene, giberreline). We also noticed that the cell wall metabolism was over-represented at the proteins, transcripts and also at the metabolites level with a significant decrease ($P < 0.01$) of UDPG and UDP. At the protein level, polygalacturonase proteins (Solyc03g111690.2.1, Solyc10g080210.1.1) involved in the cell wall degradation were more than ten times more concentrated during ripening (log2FC = 6.6 and log2FC = 6.2, respectively) while proteins involved in starch metabolism, such as starch synthase protein (Solyc08g083320.2.1), were less concentrated at the beginning of the ripening phase (log2FC= -2.98). In parallel to changes in protein metabolism, seven amino acids (arginine, aspartate, asparagine, histidine, glutamine, valine, glutamate) were up regulated during this event (GAP1). At 53 DPA, the glutamate was the most concentrated amino acid (ten times higher than others, 10 ± 1.2 µmol.gFW^{-1}). The accumulation of glutamate, resulting from the starch degradation, highly participated to the "umami" taste of tomato.

The decrease of lipid synthesis metabolism at the protein level referred to changes observed for the cell wall metabolism and also the hormone metabolism. Indeed, jasmonate hormone, a derived of polyunsaturated fatty acid, is an example of the link between lipid metabolism and hormone metabolism (Koo and Howe, 2009). Almost all pigment were found significantly changed during this transition, either decreased for the chlorophyll *b* and *a* and violaxanthin or increased for lycopene, phytoene, phytofluene and both β and δ carotene.

The only enzyme activity determined in this event (GAP1) was a TCA cycle enzyme, the NADP-IDH activity (Figure II. 16, P=0.002), also determined at the protein level in PageMan analysis (Figure II. 16). This enzyme converts isocitrate into α-cetoglutarate producing reduced cofactor (here NADPH) and CO_2 while the increase of NADP-IDH activity was accompanied by

a significant accumulation of citrate and a decrease of fumarate. These results highlighted the role of NADP-IDH activity in the transition toward ripening and the involvement of TCA metabolism in the climacteric respiration, known to induce the metabolic cascade of fruit ripening.

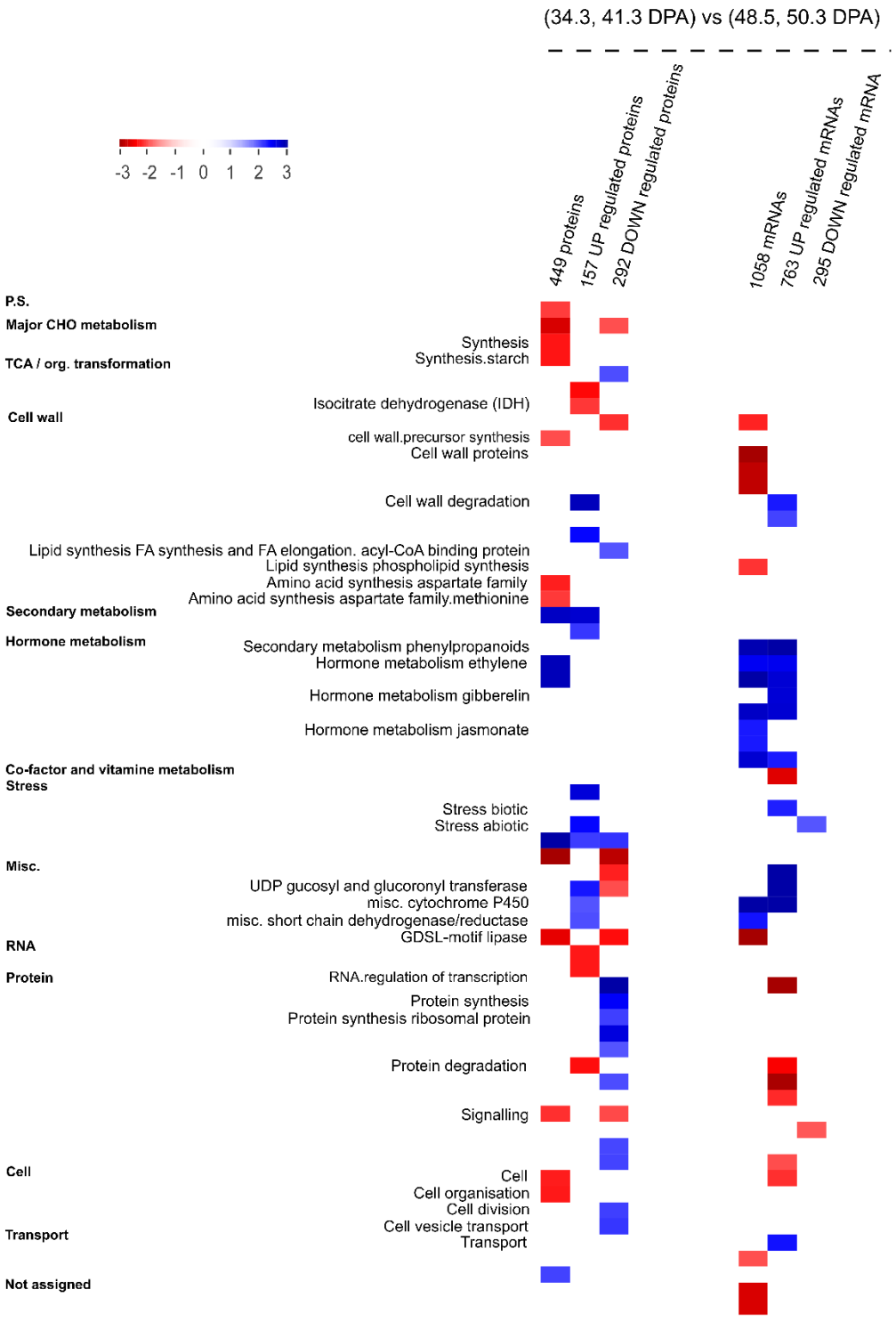


Figure II. 15 Pageman analysis of proteins and transcripts with a significant different expression from turning to ripening. Up-regulated variables were defined as variables more expressed after ripening (48.5 DPA + 50.3 DPA) than before (34.3 DPA + 41.3 DPA). The color code corresponds to the Zscore of the pvalue attributed to the category. Categories colored in red were significantly down-regulated relative to the rest of the array, whereas BINs colored in blue were up-regulated.

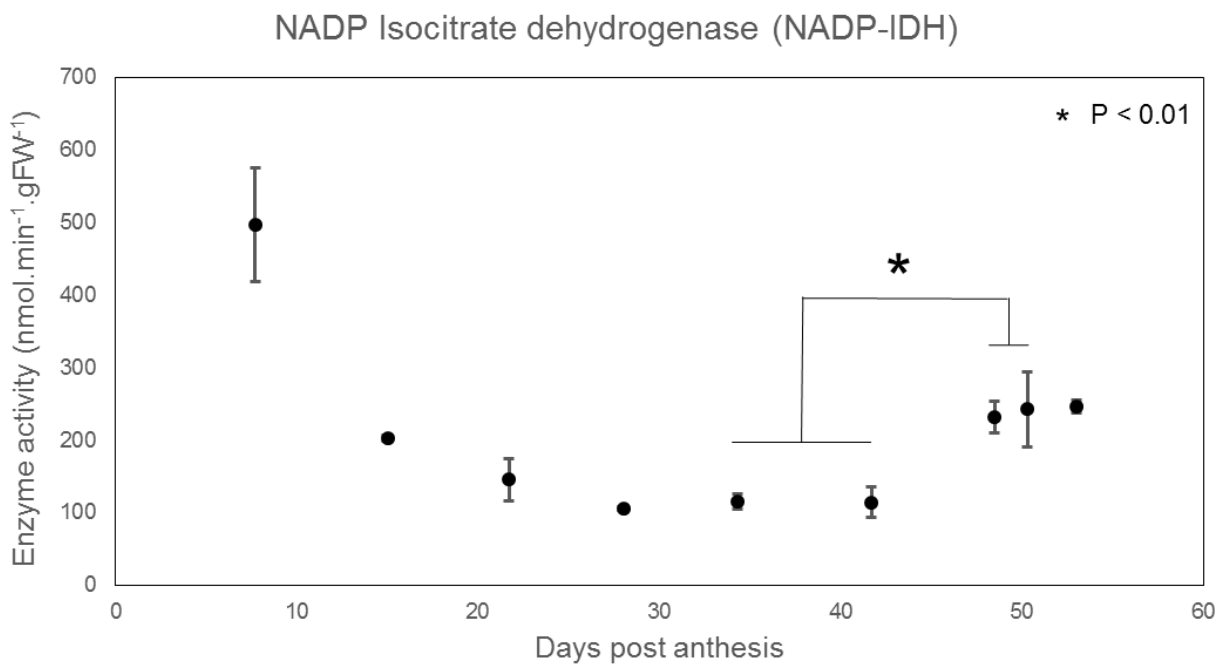


Figure II. 16 NADP-IDH enzyme activity profile from Benoit Biais et al (2010), quantified in nmol.min-1.gFW-1 showing a significant difference of activity in the transition toward ripening (i.e. between 34.3 - 41.3 DPA and 48.5-50.3 DPA).

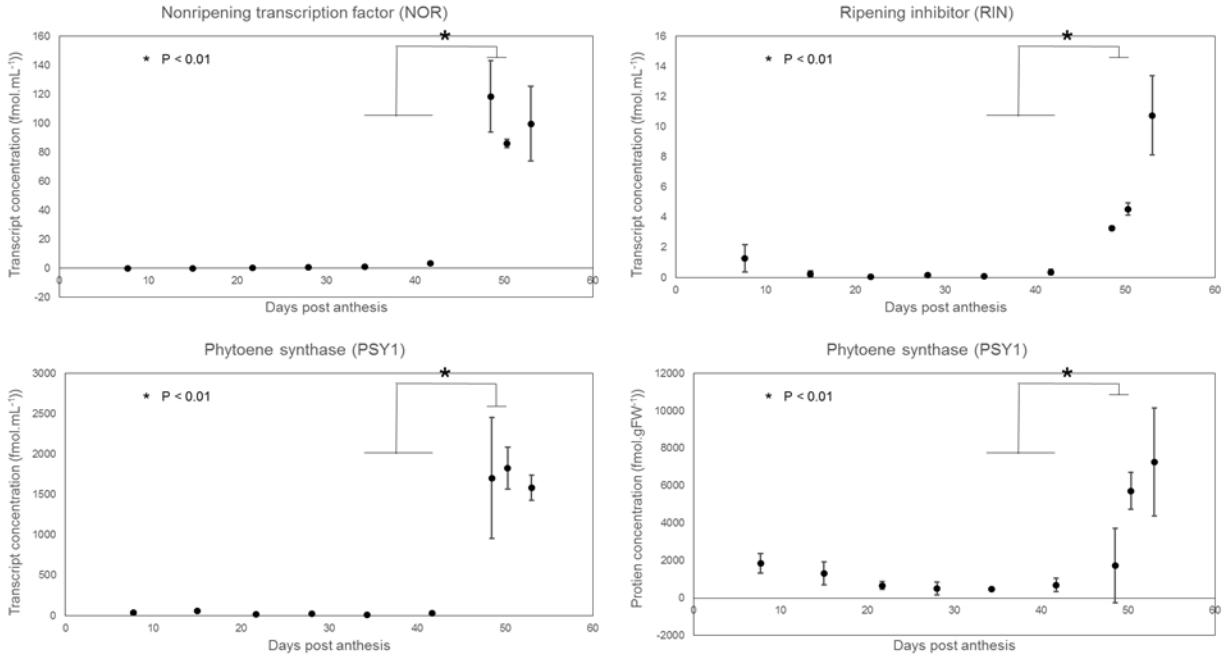


Figure II. 17 Expression of ripening markers, nonripening (NOR) and ripening inhibitor (RIN) transcription factors and phytoene synthase (PSY1), detected in the GAP1. PSY1 was detected at the transcript and protein levels. Concentration of the three transcripts and the protein were expressed in fmol.mL-1 and fmol.gFW-1, respectively.

Besides, markers of tomato ripening such as NOR (Solyc10g006880.2.1) and RIN (Solyc05g012020.2.1) transcription factors, were identified among the 545 transcripts. The phytoene synthase 1 (PSY1, Solyc03g031860.2.1), also considered as a ripening marker, was detected at both transcript and protein levels (Figure II. 17).

- GAP2

Then, variables of the second event (GAP2 in Figure II. 14) defined by similarity at the first (7.7 DPA) and the last (53 DPA) stages of the tomato fruit development, were analyzed according to "up-down-up" or "down-up-down" regulation. Mentioned above, criteria were used to select confident variables: 545 transcripts, 87 proteins and 14 metabolites were obtained. However, these variables were identified only based on a statistical analysis between the first and last stages, thus in order to consider the dynamic throughout the development we performed a hierarchical clustering analysis (Figure II. 18). The hierarchical clustering analysis proved to be not superfluous as it allowed to remove profiles not corresponding to the two targeted profiles we looked for (variables not in red squares, Figure II. 18). Finally, 6 metabolites (adenosine like, leucine/isoleucine, sedoheptulose 7-phosphate, succinate and tyrosine and one "Unknown" metabolite), 56 proteins, 75 transcripts and none enzyme activity were obtained. Note that none common variable name (SolycXXgXXXX) was found between the 56 proteins and 75 transcripts and more than 80% of variables were associated to the profile called "up-down-up". Succinate, S7P and alike adenosine metabolites were also found "up-down-up" regulated.

Then, we looked for the functional categories associated to the 56 proteins and 75 transcripts. According to their small number, "down-up-down" transcripts (10) and proteins (10) were then checked manually. The 65 transcripts and 46 proteins more concentrated at 7.7 and 53 DPA were distributed according to their functional categories (Figure II. 19) using MapMan file annotation (Usadel et al., 2009).

The 10 "down-up-down" transcripts and proteins were reduced to 7 transcripts and 8 proteins with determined molecular function. Among the seven transcripts, 6 were regulators related to the transcription (DNA-binding/ regulation of transcription, Solyc05g007890.2.1) and translation (Ribosome assembly factor (Solyc01g104470.2.1) and Eukaryotic translation initiation factor (Solyc03g115650.2.1)) and to protein (enzyme inhibitor (Solyc12g099200.1.1) and Ubiquitin protein ligase activity (Solyc04g007970.2.1)). The last transcript coded for an aspartic-type

endopeptidase activity protein. Note that among these 10 transcripts, five were ten times more concentrated than all others transcripts (Figure II. 10) with a concentration higher than 100 fmol. mL^{-1} throughout the development. Seven of 8 "down-up-down" proteins were associated to enzyme activity of the carbohydrate metabolism, such as the FBPase (Solyc02g084440.2.1), 6-phosphogluconolactonase (Solyc05g012110.2.1), alpha-L-arabinofuranosidase (Solyc12g100120.1.1), aldose 1-epimerase (Solyc02g087770.2.1). One protein, named "Auxin repressed" was also detected.

Functional categories associated to "up-down-up" variables were represented in Figure II. 12. Despite the "cell wall metabolism", all functional categories were represented with 10 categories represented by both proteins and transcripts, such as "Amino acid metabolism". As performed previously, we paid more attention to highly represented functional categories, i.e. having more than 5% of transcripts or proteins. Nine functional categories were identified with this criteria: "Protein metabolism", "DNA, RNA binding and metabolism", "Redox", "Carbon metabolism", "Amino acid metabolism", "Development and cellular organization", "Photosynthesis", "Miscellaneous" and "Not assigned".

From the Chapter 1 and Chapter 2, we showed that the absolute quantification of four omics data has allowed to describe, in coherence with literature, the tomato fruit development and ripening even if it was not yet clear to understand the role of transcripts and proteins involved in the two studied events of tomato fruit development. Thus, being confident with the protein and transcript absolute quantification, we used them to parameterize a mathematical model describing protein translation.

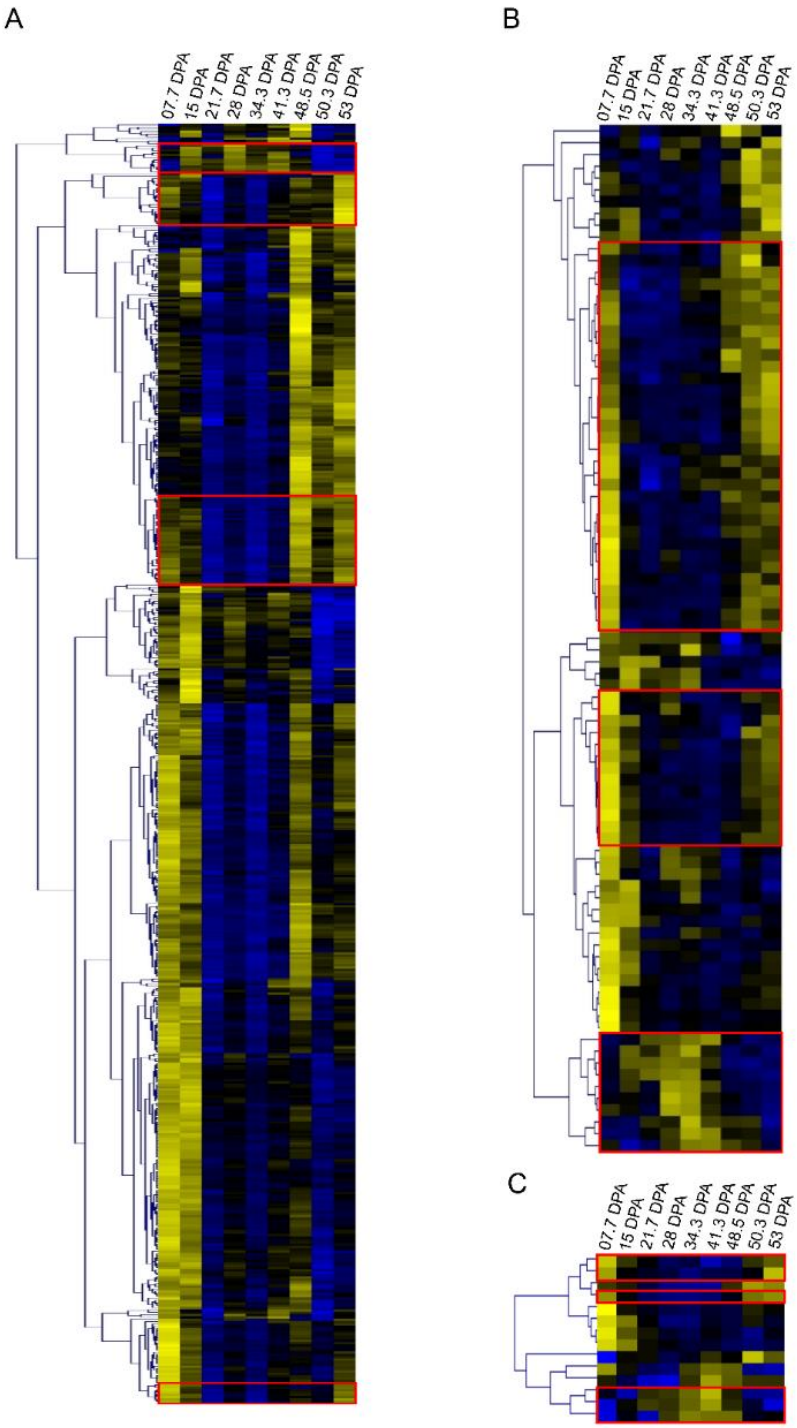


Figure II. 18 Identification of transcripts (A), proteins (B) and metabolites (C) involved in GAP2. Hierarchical clustering analysis was performed on the mean centered concentrations and scaled to unit of the selected variables (Table II. 1, GAP2), *.i.e* on the the 545 transcripts (in $fmol.mL^{-1}$), 89 proteins ($fmol.gFW^{-1}$) and 14 metabolites ($\mu mol.gFW^{-1}$ or $AU.gFW^{-1}$). Then, profiles associated to "up-down-up" and "down-up-down" regulation between 7.7 and 53 DPA were visually determined (red squares) for the three omics subsets.

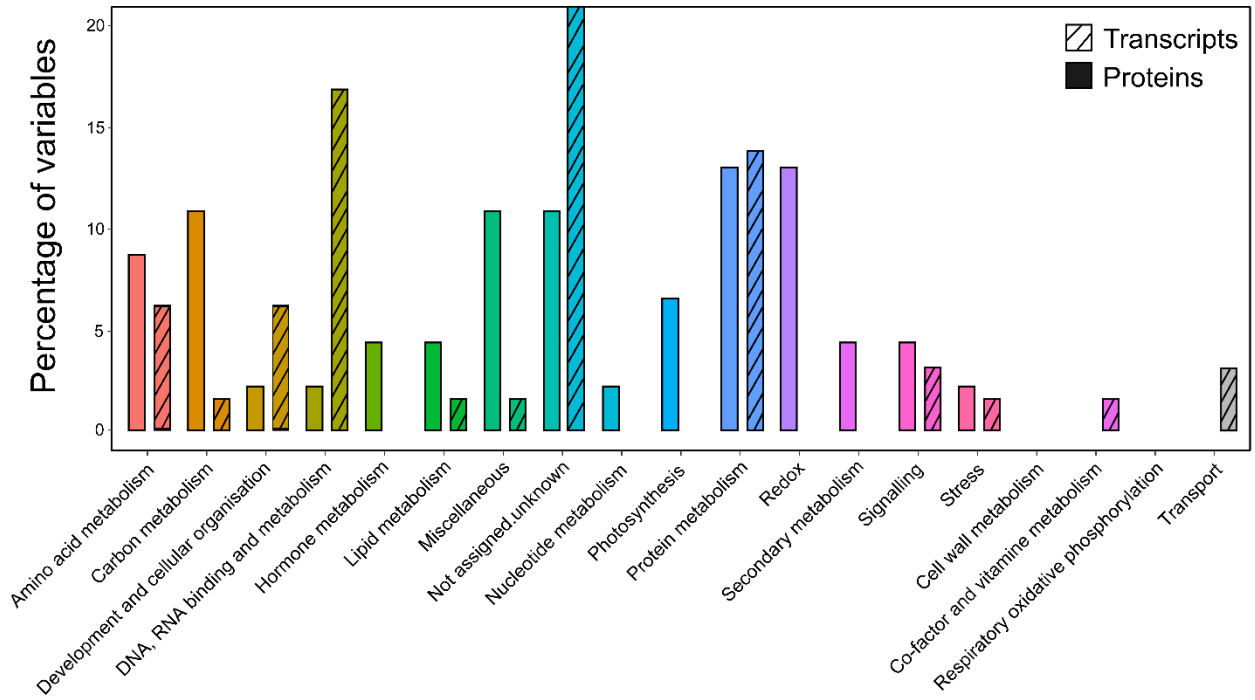


Figure II. 19 Functional categories associated to the "up-down-up" transcripts and proteins involved in GAP2. Selected with criteria (defined in the text), 46 proteins (bar not striped) and 65 transcripts (striped bar), more concentrated at 7.7 and 53 DPA were identified. These variables were distributed in the 19 functional categories deduced from the MapMan file annotation (Usadel et al., 2009).

Chapter 3 Modelling the translation from quantitative proteomic and transcriptomic data

I. How proteins and transcripts correlate during tomato fruit development?

In this chapter, proteins and transcripts data were expressed in gFW basis to allow the resolution of ODE modeling the translation. For the clarity of the text, we used term 'mRNA' referring to transcript.

Using the tomato genome ID (SolycXXgXXXX) provided in ITAG 2.4 (Sol Genomics Network, https://solgenomics.net/), we restricted our analysis to genes that were identified at both mRNA and protein levels resulting to 2490 mRNA-protein pairs. We previously showed that relatively few proteins showed a concentration lower than 100 fmol.gFW^{-1} (see Chapter 1, Figure I. 7), indicating that some proteins of low abundance escaped detection. In the subset of 2490 mRNA-protein pairs, proteins were on average 2636 times more abundant than the corresponding transcripts as illustrated by the median of the protein/mRNA ratio (Figure III. 1 right panel, Figure III. 2). Interestingly this ratio progressively increased throughout fruit development, from 1269 to 3011. This increase in the protein/mRNA ratio resulted from transcripts decreasing more than the corresponding proteins throughout fruit development, as illustrated in Figure III. 1 (left panel). This protein/mRNA ratio (2636) found here for the tomato fruit is in agreement with previously reported data. Indeed, ratios reported for other eukaryotic cells were 2800 for mouse fibroblasts (Schwanhausser et al 2011) and 748.3 yeast (Lahtvee et al. 2017) and thus in the same order of magnitude.

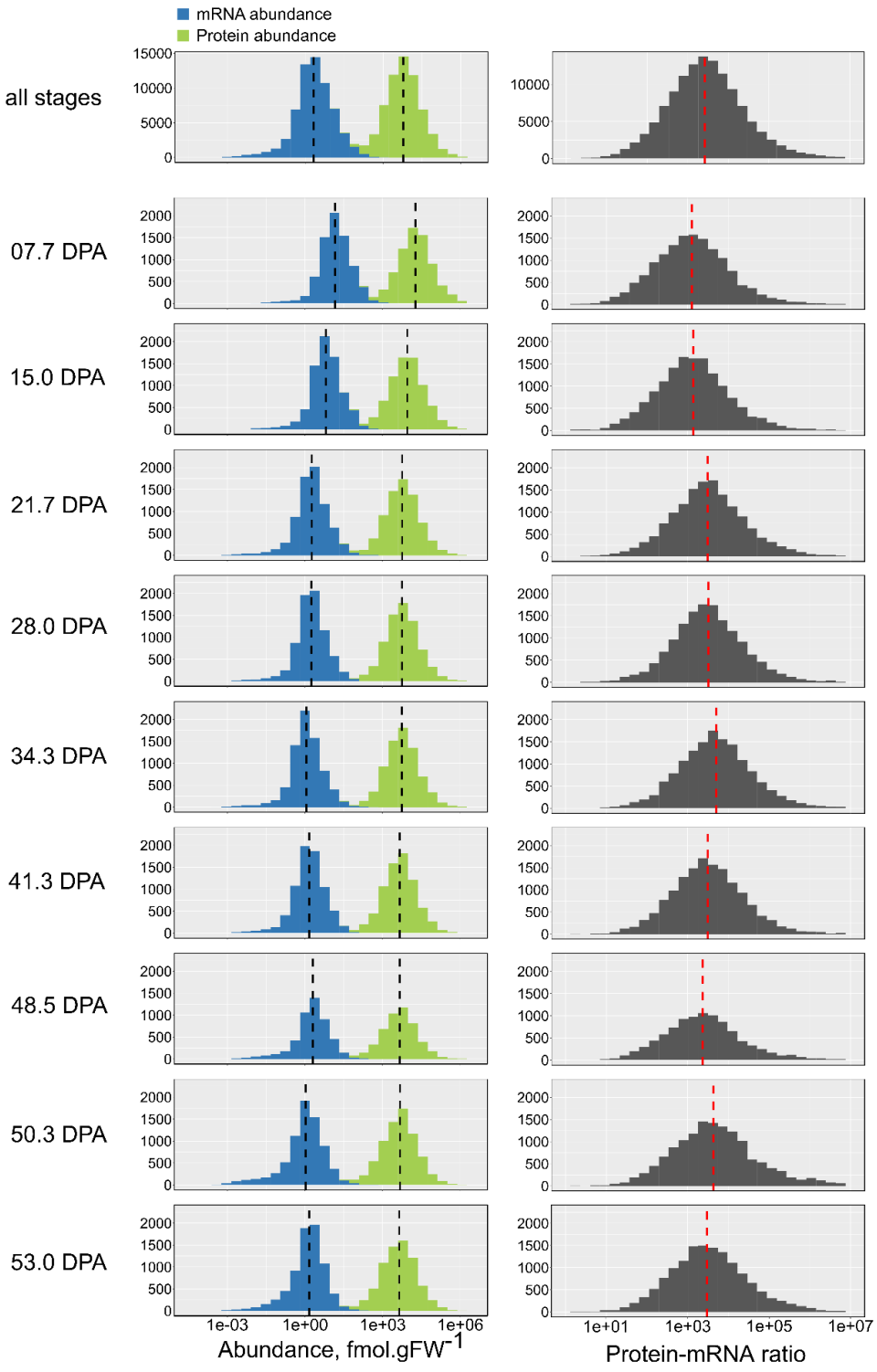


Figure III. 1 Distribution of absolute protein abundance (green) and mRNA (blue) abundance and protein-mRNA ratio (grey) for the nine stages of development (7.7, 15.0, 21.7, 28.0, 34.3, 41.3, 48.5, 50.3 and 53 DPA). Abundances of the 2490 protein (green) and corresponding 2490 mRNA (blue) were expressed in fmol.gFW^{-1}. Abundances and ratios were $\log_{10}$ scaled. Medians were represented by dashed line.

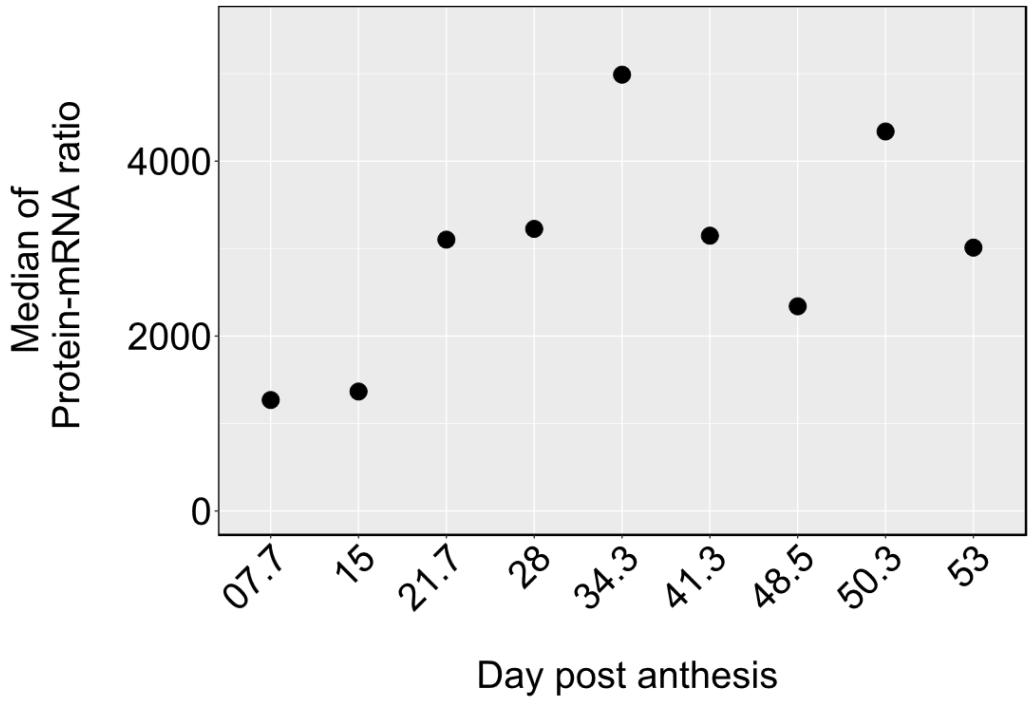


Figure III. 2 Changes of median of protein-mRNA ratio throughout tomato fruit development.

For these 2490 mRNA-protein pairs, despite a huge spread, mRNA and protein abundances were clearly positively correlated (Figure III. 3A) with a Pearson coefficient of determination equal to 0.61. This result indicates that more than half of the variation in protein content can be explained by transcript level. This result is consistent with data reported in other organisms (reviewed in Maier et al. 2009), such as mammals (R^2=0.59), yeast ($0.36 < R^2 < 0.76$) and bacteria ($0.50 < R^2 < 0.57$). In plants, the correlation coefficients calculated for different sections of growing maize leaves were also found higher than 0.5 (Ponnala et al., 2014). The authors suggested a contribution of post-translational regulations for about half of proteins in the cell system.

By examining each stage of development (Figure III. 3B) we see that the protein-mRNA correlation decreased throughout fruit development. Indeed, proteins and their encoding mRNA were highly correlated ($R^2 \sim 0.6$) until the fourth stage, *i.e.* during cell division and the beginning of cell expansion (from 7.7 DPA to 28 DPA), then the correlation decreased until the 53 DPA stage, reaching an $R^2 \sim 0.5$ (Figure III. 1B).

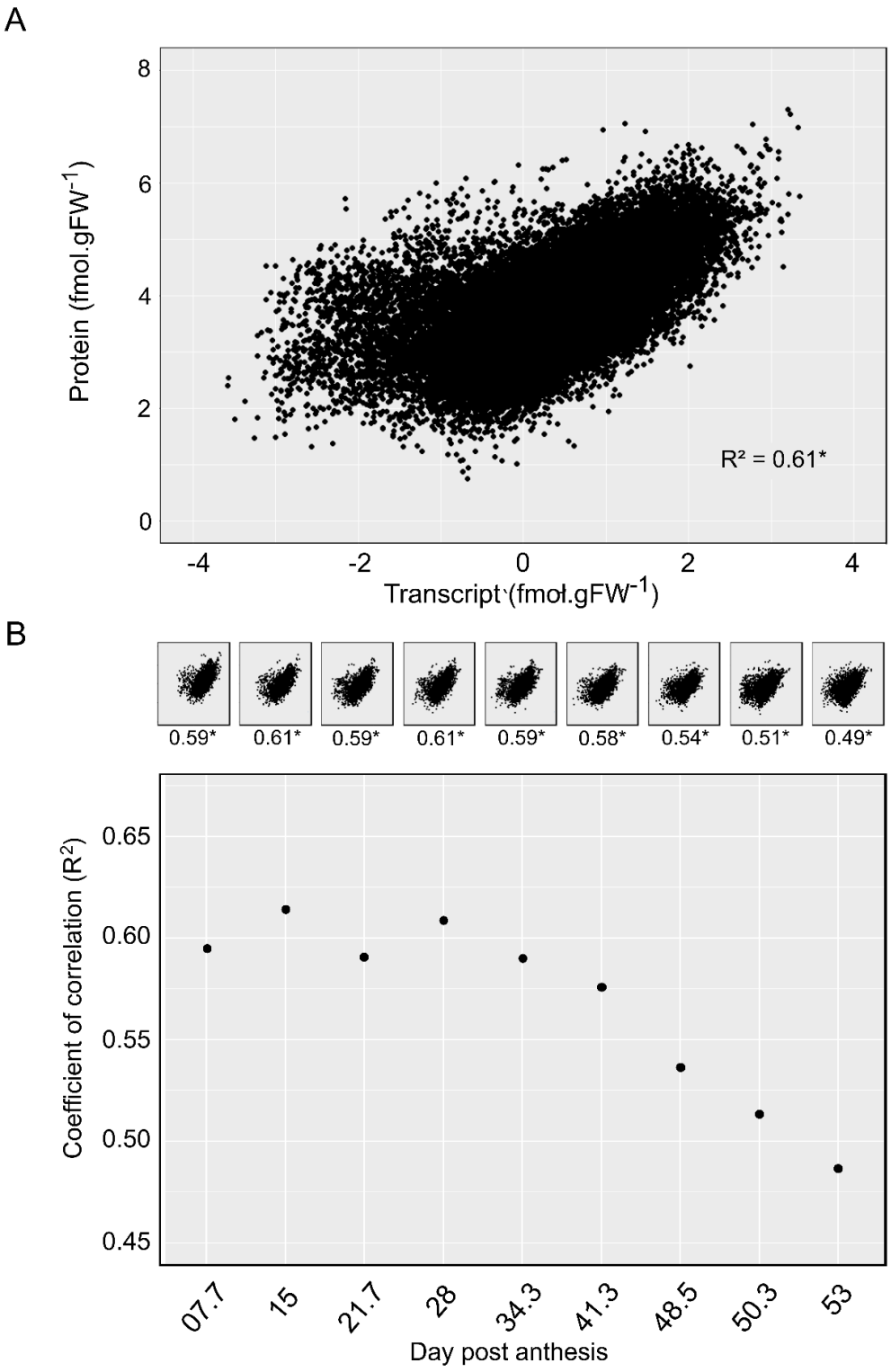


Figure III. 3 Correlation between protein and mRNA abundance. Correlation (Pearson) was estimated on 2490 protein-mRNA pairs with all data (A) and at each of the nine stage (B). Abundances were expressed in fmol.gFW^{-1}. Significant Pearson correlation ($P < 0.05$) was significant are annotated by *.

Assuming that the abundance of proteins is conditioned by both their synthesis and degradation rates, one hypothesis is that the correlation decreased with fruit age because the proteins are more stable than the transcripts encoding them.

Since 2012, the lab-group I worked in intended to model protein translation with a set of transcripts quantified by qRT-PCR and of enzyme activities used as proxy of protein concentrations (same data used in Chapter 1.II.b p). However this dataset was too small to allow the resolution of the translation model. Thus, with quantitative data obtained for more than 2000 pairs of transcripts and proteins, we had the opportunity to properly solve the model. The next section describes the mathematical model of translation and its resolution. The resolution of the model involving the estimation of synthesis and degradation rate constants for each protein. Finally, these rate constants were analyzed and compared to literature data for validation purpose.

II. The translation model

2.1 The translation model: a differential equation involving two constants k_{sp} and k_{dp}

To investigate the major principles of gene expression regulation in dynamic systems, we estimated protein synthesis and degradation rates from time series data of mRNA and protein expression. By that way, we tested the degree to which expression changes can be modelled by a differential equation. Indeed, among the existing models presented in the introduction (third section, p), we selected and implemented the simple mathematical model based on only one ordinary differential equation (ODE) describing the synthesis and degradation of one protein from its corresponding mRNA.

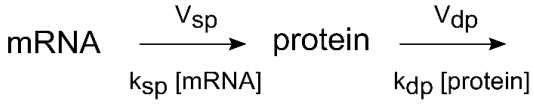


Figure III. 4 Schema of the translation model with the protein synthesis rate (Vsp) proportional to the abundance of the corresponding mRNA (fmol.gFW^{-1}) according to a synthesis rate constant (k_{sp}, day-1) and the protein degradation rate (Vdp) proportional to the protein abundance (fmol.gFW^{-1}) according to the degradation rate constant (k_{dp}, day^{-1}).

This model has been already described for global dataset of human cells (Dressaire et al., 2009; Tchourine et al., 2014) and also to describe the ethylene biosynthesis in tomato fruit (Van de Poel and Van Der Straeten, 2014). Mathematically, the model has been written with a differential equation describing the evolution of the protein synthesis rate (dP/dt) as the result of two mains terms: the rate of synthesis from its corresponding mRNA (V_{sp}) and the rate of degradation of the protein itself (V_{dp}) (Figure III. 4, Equation III. 1). In this equation, the synthesis rate (V_{sp}), *i.e.* the translational rate operated by ribosomes was considered as proportional to the abundance of the corresponding mRNA according to a synthesis rate constant (k_{sp}, day^{-1}). The degradation rate (V_{dp}) was considered proportional to the abundance of the protein according to the degradation rate constant (k_{dp}, day^{-1}) (Equation III. 1).

$$\frac{dP(t)}{dt} = k_{sp}\ R(t) - k_{dp}\ P(t) \qquad \text{Equation III. 1}$$

With $k_{sp}\ and\ k_{dp}$ the rate constants of synthesis and degradation respectively (> 0, in day^{-1}).

Equation III. 1 takes into account the abundances of transcripts (*R(t)*) and proteins (*P(t)*) in the whole system, *i.e.* the fruit, throughout its development (*R(t)* and *P(t)*in $fmol.fruit^{-1}$).

At each time *t*, abundances of mRNA and protein per fruit (*R(t)*) and (*P(t)*) resulted from their respective concentration on a gram FW-basis (*r(t)* and *p(t)*in $fmol.g^{-1}FW$) multiplied by the fruit weight (*w(t)* in $gFW.fruit^{-1}$) according the Equation III. 2 and Equation III. 3.

$$R(t) = r\ (t) * w\ (t) \qquad \text{Equation III. 2}$$

$$P(t) = p\ (t) * w\ (t) \qquad \text{Equation III. 3}$$

From Equation III. 2 and Equation III. 3, Equation III. 1 became:

$$\boldsymbol{\frac{dp(t)}{dt} = k_{sp}\ r(t) - \left(k_{dp}\ + \mu(t)\right) p(t)} \qquad \text{Equation III. 4}$$

With $\mu(t) = (\frac{1}{w(t)}) * \frac{dw(t)}{dt}$ defined as the relative growth rate (in day^{-1}) describing the fruit growth. From Equation III. 4, we showed that the protein dilution due to growth contributed to protein disappearance in addition to protein degradation.

Note that the degradation rate constant (k_{dp}) is tightly related to the half-life of the protein ($t_{1/2}$), which is usually experimentally determined by isotope labelling (Introduction). The relation linking the degradation rate constant (k_{dp}) and half-life ($t_{1/2}$) is given by Equation III. 5 (Claydon et al., 2012):

$$t_{1/2} = \frac{\ln(2)}{k_{dp}} \qquad \text{Equation III. 5}$$

The degradation rate constant (k_{dp}) should ideally be the parameter reported (Claydon et al., 2012). Indeed, while the conversion of the degradation rate constant to a half-life ($t_{1/2}$) is often used to express turnover rates, this is not ideal when used analytically or in comparative studies as the relationship between k_{dp} and $t_{1/2}$ is nonlinear. According to (Claydon et al., 2012), the most appropriate parameter is the first-order rate constant for degradation.

Finally, in a particular case of the steady state the protein pool was considered constant, so that the rate of change of the protein pool *dP/dt* was null and Equation III. 4 reduced to:

$$k_{sp}\, r(t) = \left(k_{dp} + \mu(t)\right) p(t) \qquad \text{Equation III. 6}$$

2.2 Resolution of the translation model

The model described by Equation III. 4 has been solved for each of the 2490 mRNA-protein pairs. For that, time functions were required for both the relative growth rate and the mRNA content.

To estimate a time function of the relative growth rate ($\mu(t)$), we fitted the time-course of tomato fruit weight ($w(t)$) (Figure III. 5). For that, several growth models have been tested including classical growth models (Logistic, Contois, Gompert etc.) and polynomial regressions with or without a log transformation. Classical growth models often generated wrong estimations at the

beginning of growth, when fruit weight is very low, whereas polynomial regressions sometimes lead to negative values and log transformation to exaggerate waves as well as too high values at the end of development. Finally, the sigmoid and especially the double sigmoid was the best appropriate fit according to the lowest calculated error between experimental and fitted values of tomato fruit weight. The double sigmoid also showed the advantage to reach an expected plateau at the end of development (Figure III. 5).

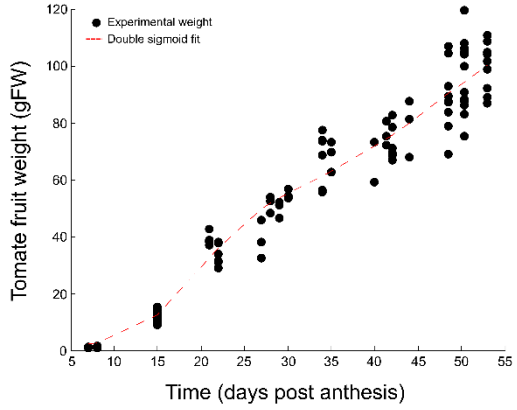


Figure III. 5 Time course of the tomato fruit weight (●) and the double sigmoid fit (red dashed line).

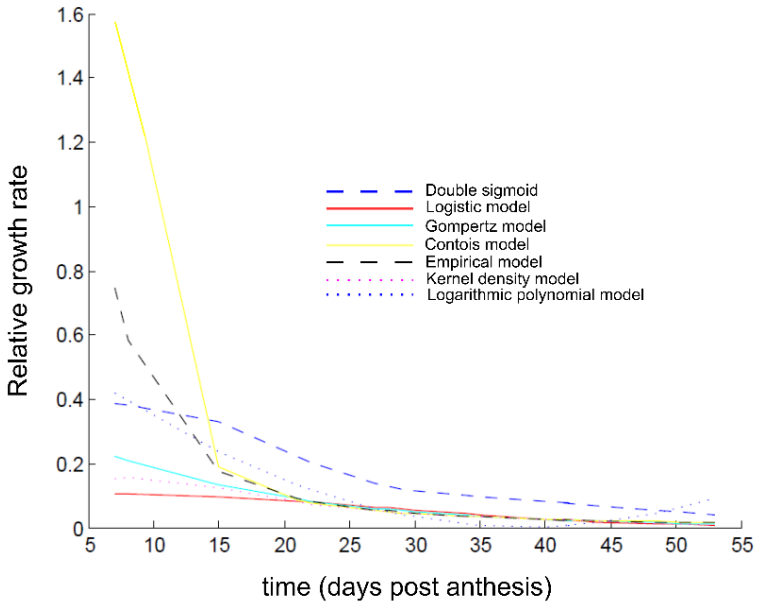


Figure III. 6 Time-course of the relative growth rate (μ(t)) calculated throughout tomato fruit development from a double-sigmoid fit.

With this double-sigmoid fit, the relative growth rate ($\mu(t)$) was calculated throughout fruit development (Figure III. 6) and used to solve the model for each mRNA-protein pair.

To solve the ODE (Equation III. 4), a time function was also required for mRNA ($r(t)$ in fmol.gFW^{-1}). While the mRNA values were all positive, a polynomial regression fitting tends to become negative when mRNA values were close to zero. To avoid this pitfall, a log transformation was done before fitting the data with a polynomial regression. Among the several degrees tried for the polynomial regression, the degree three was found, with a training dataset of about 30 mRNA profiles, as the most appropriated, as illustrated for Solyc01g005560.2 (Figure III. 7A).

Then, to solve the model, both mRNA and protein data had to be in the same order of magnitude. Thus, both transcript and protein datasets ($r(t)$ and $p(t)$) were normalized by their respective average values calculated over the nine stages (Figure III. 7B). Normalization by the first stage and intermediate stage (34.3 DPA) was also tested but we noticed that these two normalizations affected the dynamic protein and transcript expression. Furthermore, these normalization, being highly dependent on the variability at these stages complicated the polynomial regression fitting and the resolution.

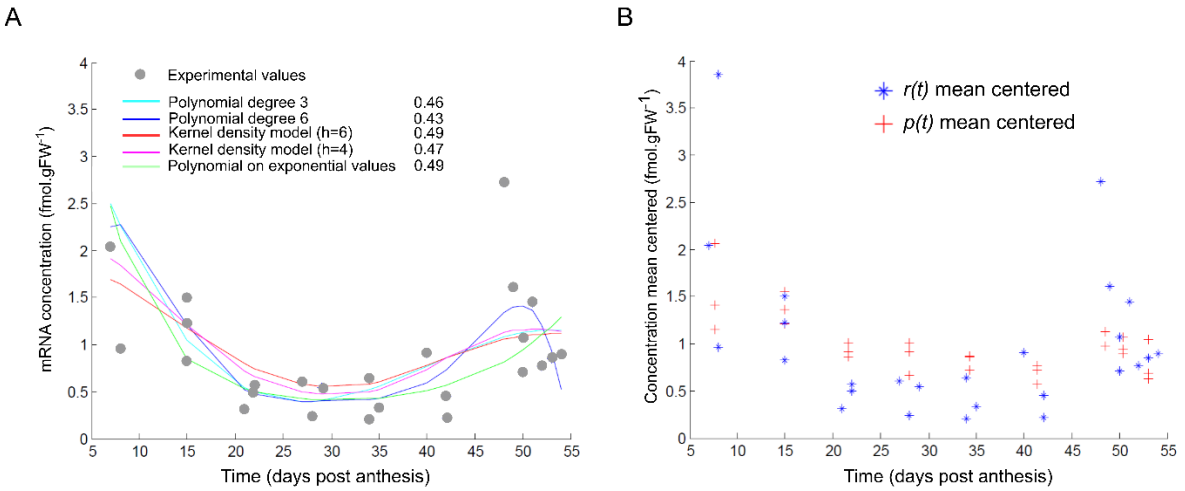


Figure III. 7 Data processing before solving the resolution of the ODE. (A) Five models (Polynomial, Kernel density) tried to fit experimental mRNA values (●). The best scores (lowest error relative) were obtained with polynomial model (degree 3: 0.46, degree 6: 0.43). (B) Protein (+) and mRNA (✱) values were respectively mean centered which was necessary to solve the ODE.

Finally, the resolution of the ODE was performed with the MATLAB software to determine both k_{sp} and k_{dp} applying the least square method: at each time t_i (DPA$_i$) the sum of the square

deviations between the solution of the ODE P(DPAi; *ksp*; *kdp*) and the experimental protein content Pi, noted S(*ksp*; *kdp*) was calculated according to Equation III. 7 and minimized.

$$S(ksp\ ;\ kdp\ ) = \sum_{i} (P(DPAi\ ;\ ksp\ ;\ kdp\ ) - Pi\ )^2 \qquad \text{Equation III. 7}$$

The resolution has been performed with the help of Segolène Augé who did her master 1 (Bioinformatique et Biologie des Systèmes, Université Toulouse) volunteer internship from June to August 2017.

2.3 Two protein groups distinguished by the quality of the resolution

Three criteria were used to evaluate the quality of the resolution: (1) a score on the mRNA fit, (2) the reliability of optimization and (3) a statistical evaluation of constant quality (see Materials and methods section).

To statistically evaluate the quality of the k_{sp} and k_{dp} constants, we calculated a confidence region. This mathematical verification allowed validating the resolution with a right determination of both constants k_{sp} and k_{dp} associated to one mRNA-protein pair. For that we used a numerical method to calculate an approximate value of the area delimited by the contour of the confidence region. In the case of an unclosed confidence region, the resolution of the model was considered as unsatisfying (Figure III. 8). Conversely, when the confidence region was closed (Figure III. 9) the resolution was acceptable and the calculated rate constants can be further analyzed.

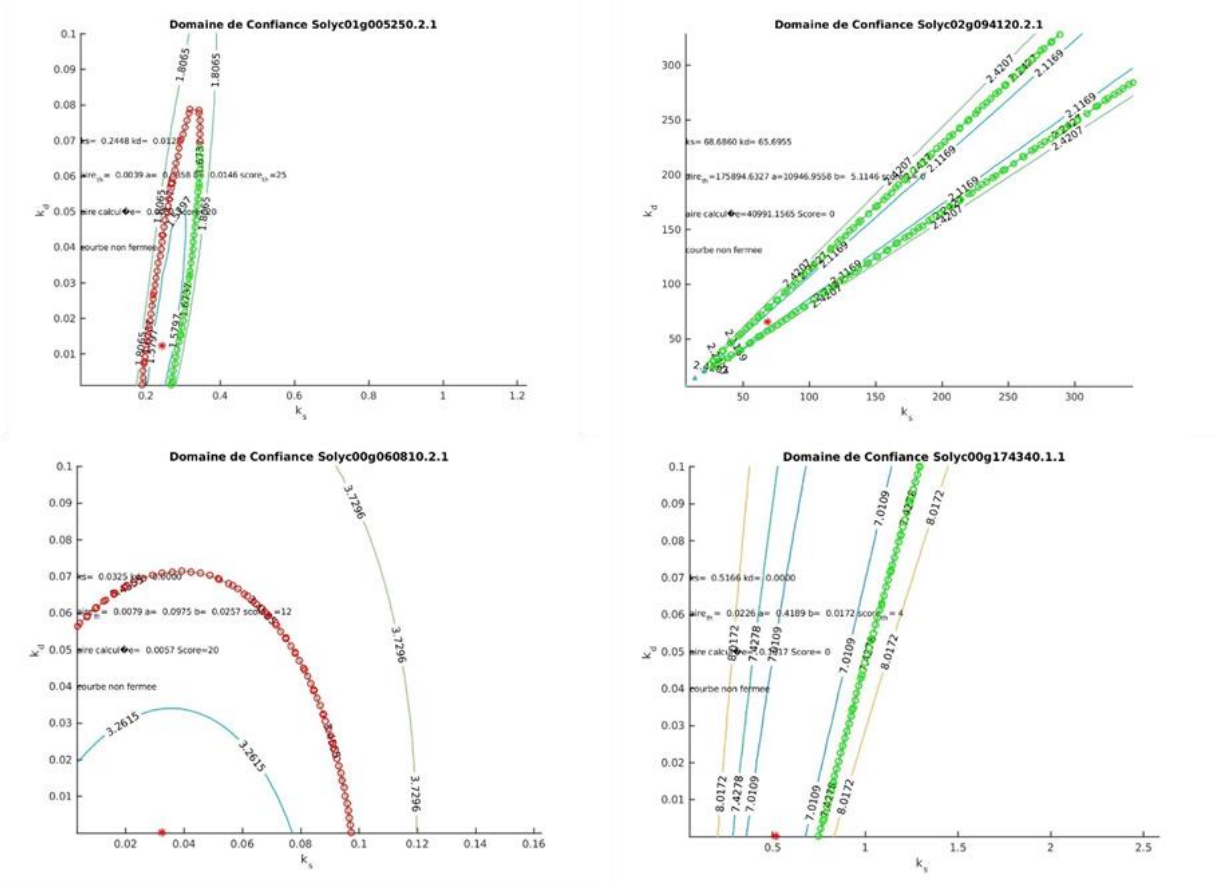


Figure III. 8 Examples of the unsatisfying confidence region calculated from the two rate constants ksp and kdp after resolution of the translation model with a percentage of confidence of 10% (blue), 25% (cyan) and 50% (brown).

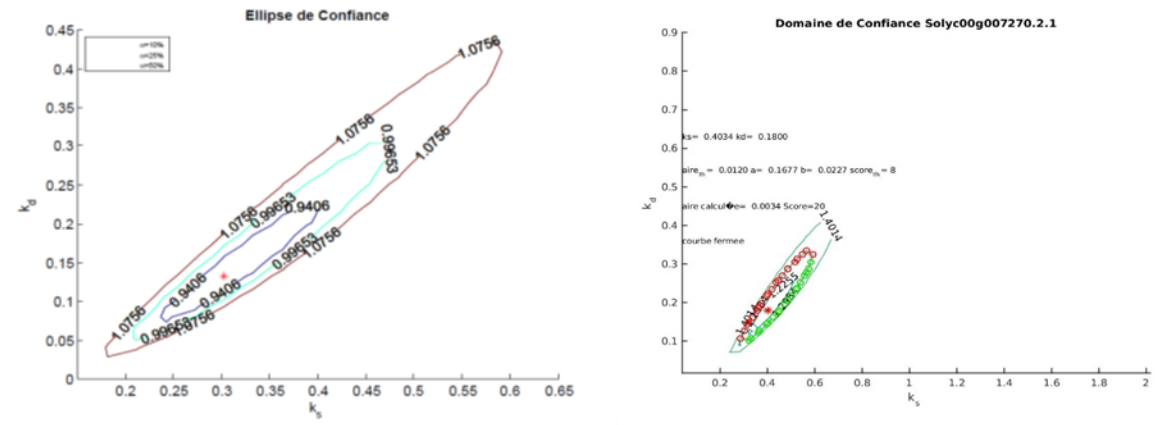


Figure III. 9 Example of the confidence region calculated from the two rate constants k_{sp} and k_{dp} after resolution of the translation model with a percentage of confidence of 10% (blue), 25% (cyan) and 50% (brown).

III. Analysis of k_{sp} and k_{dp}

The objective of this section was to globally analyze the calculated rate constants and to compare the results with constants reported in the literature.

The resolution could not be carried out for 119 mRNA-protein pairs because too many values of protein concentrations were missing (unaffected). Also, to keep the rate constants for analysis i.e. to consider a satisfying resolution, we used the quality of the resolution evaluated with the confidence region criteria. Thus, the results have been manually split into two groups: The first group of "closed confidence region" corresponds to a satisfying resolution, thus both constants k_{sp} and k_{dp} were further analyzed. This group was the biggest and contained 1247 mRNA-protein pairs. The second group called "unclosed confidence region" contained the '1128 rejected' mRNA-protein pairs from modelling, thus both constants k_{sp} and k_{dp} have not been analyzed so far as they were considered as badly estimated. Some hypotheses were proposed to explain the poor quality of the resolution in the next part.

3.1 Rate constants determined by an unclosed confidence region

The "unclosed confidence region" group contained the 1128 "unsatisfactory" mRNA-protein pairs. The optimization score, which summarized the reliability of the mRNA fit optimization, was investigated to determine if the optimization can result to the "rejection" of ODE. For this "unclosed confidence region" group, the optimization score was clearly lower (Figure III. 10). Moreover, while k_{sp} and k_{dp} distributions were almost similar (Figure III. 11), finding more outliers (higher dispersion) for both constants suggests that "mistakes" occurred during the resolution (Figure III. 12).

Several assumptions were proposed to find an explanation of the bad quality of the resolution.

(1)- An unsatisfying mRNA fit could lead to a bad resolution. But the median of the scores calculated for mRNA fitting were similar in both closed and unclosed confidence region groups. More dispersed optimization scores were found for the unclosed confidence region group (Figure III. 13).

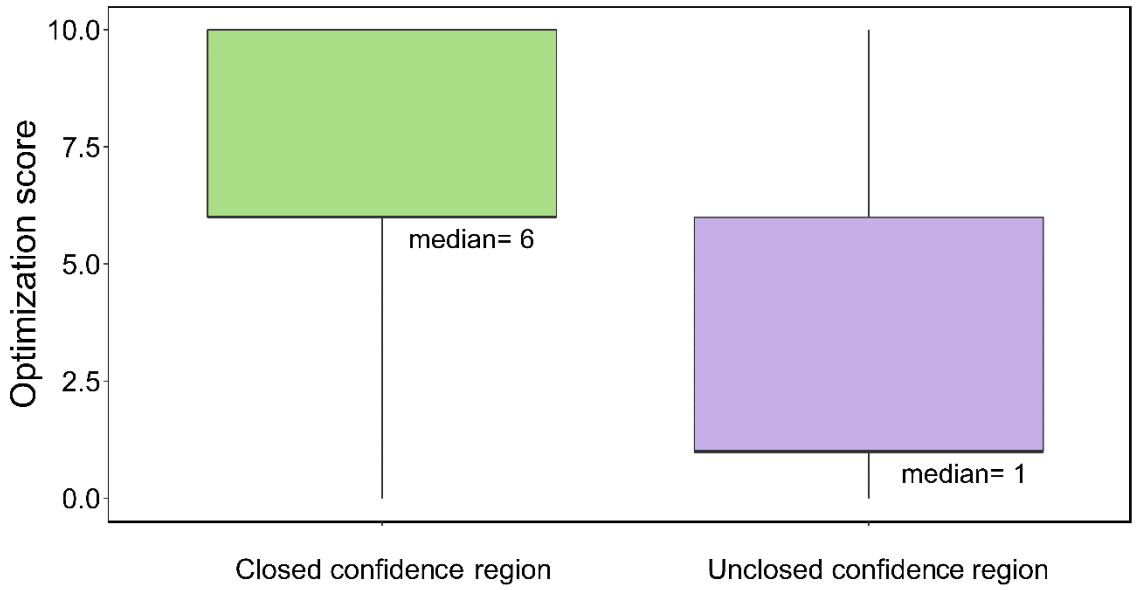


Figure III. 10 Optimization scores (from 0 to 10) characterizing the resolution of the model for both the unclosed (purple) and closed (green) confidence region.

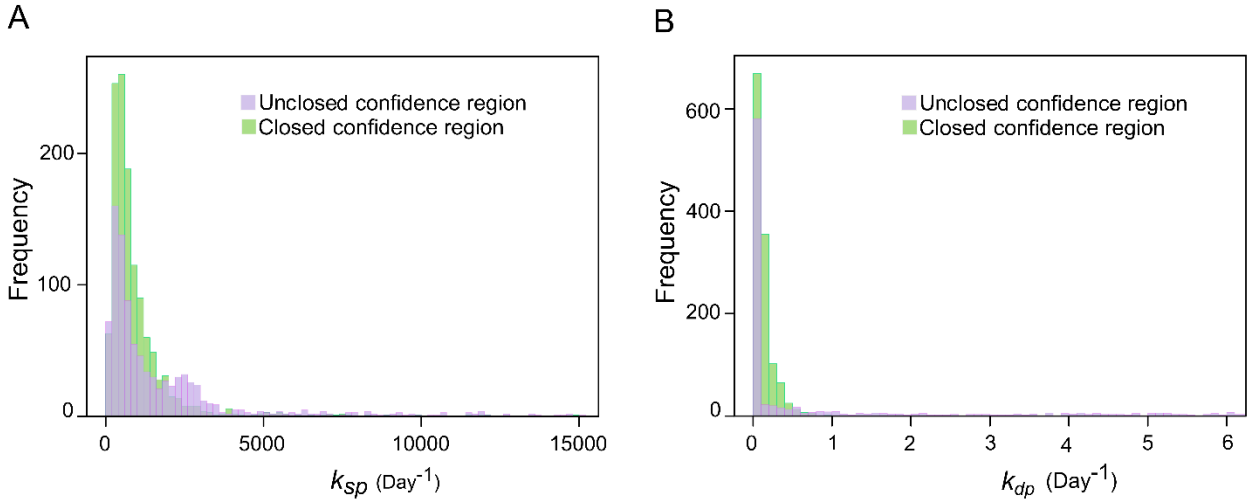


Figure III. 11 Distribution of k_{sp} (day^{-1}) (A) and k_{dp} (day^{-1}) (B) for both the unclosed (purple) and closed confidence region (green) groups.

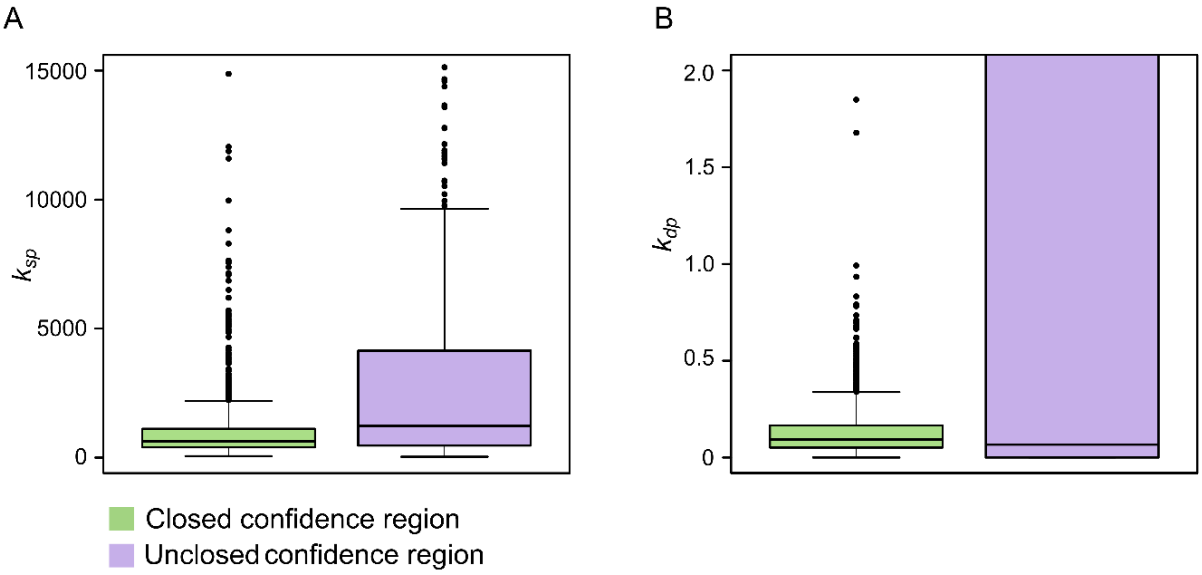


Figure III. 12 Repartition of k_{sp} (day^{-1}) (A) and k_{dp} (day^{-1}) (B) for both the unclosed (purple) and closed confidence region (green) groups.

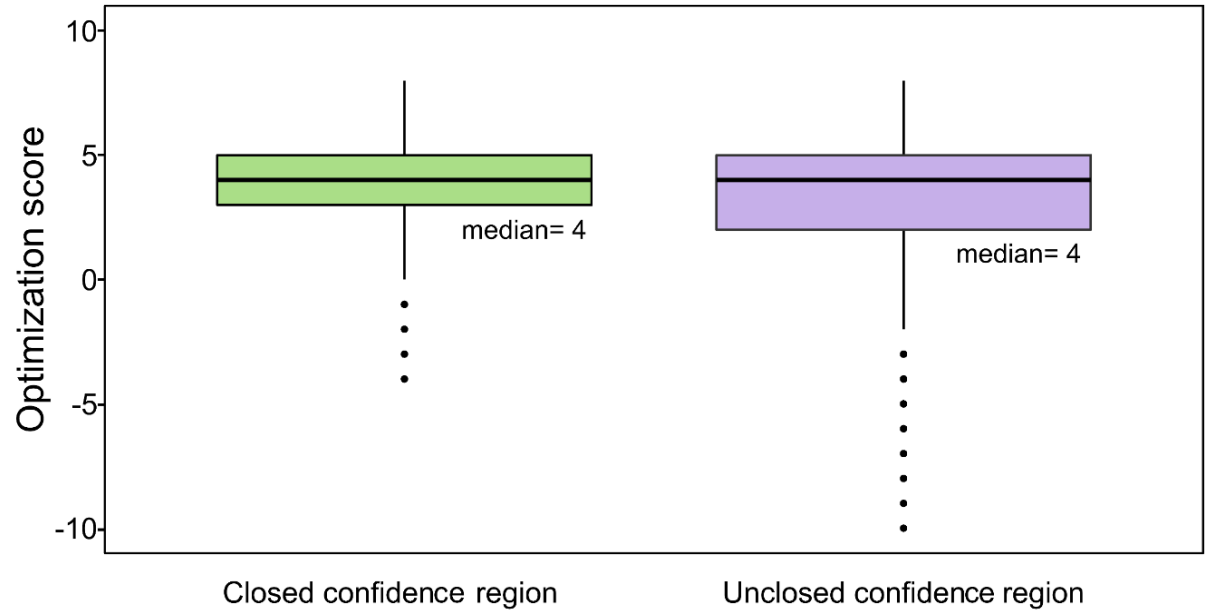


Figure III. 13 Optimization scores for mRNA fitting for both the unclosed (purple) and closed confidence region (green).

(2)- An absence of correlation between the protein and its corresponding transcript, which suggested an increase in mRNA without increased protein synthesis or conversely increased protein synthesis without increased transcript, could be a potential explanation for the unsatisfying resolution. Spearman correlation analysis performed on the 2375 proteins and mRNAs concentration could not explain the "unclosed confidence region", as similar results were obtained with "closed confidence region" (Figure III. 14). A correlation analysis led to the same conclusion, with non-significant difference of correlation between the two groups (P>0.05).

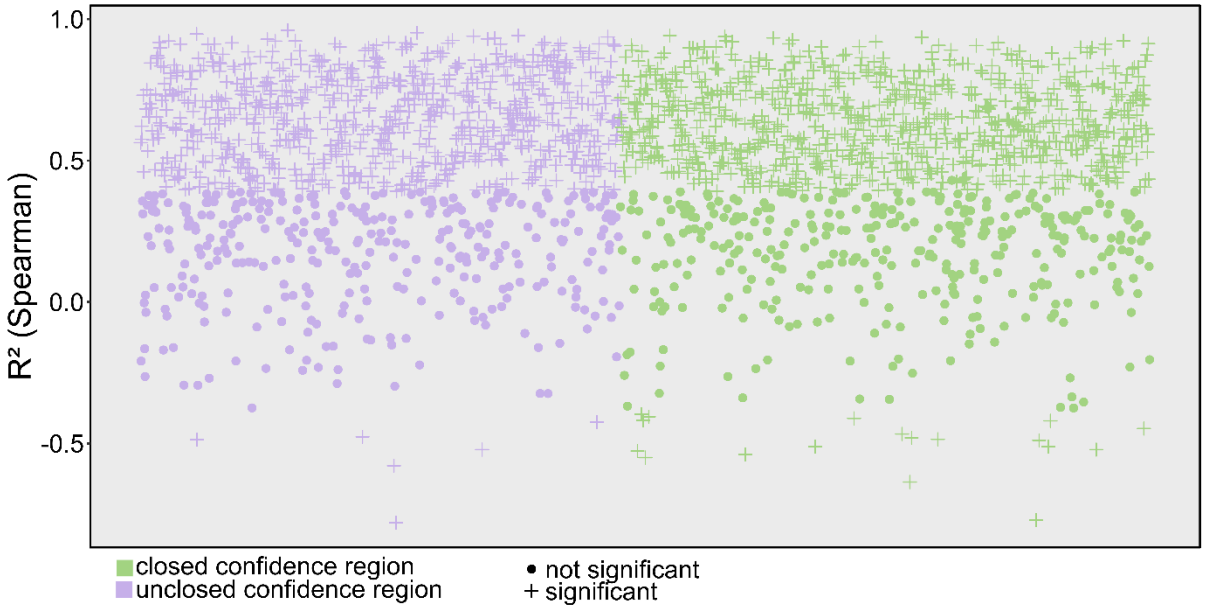


Figure III. 14 Spearman correlation calculated between protein and transcript concentration (in fmol.gFW^{-1}, log10 transformed). Coefficients of determination (R^2) were separated according to the unclosed confidence region (purple) and the closed confidence region (green). Significant ($P < 0.05$) and non-significant correlation are indicated by + and ●, respectively.

(3)- A high number of missing values could have penalized the resolution, especially for the proteins dataset (there was no missing values in the transcript dataset). Indeed, when at least one value was missing, more mRNA-protein pairs belonged to "unclosed confidence region". Also, the proportion of satisfying resolution was higher when mRNA-protein pairs did not contain missing values (Figure III. 15).

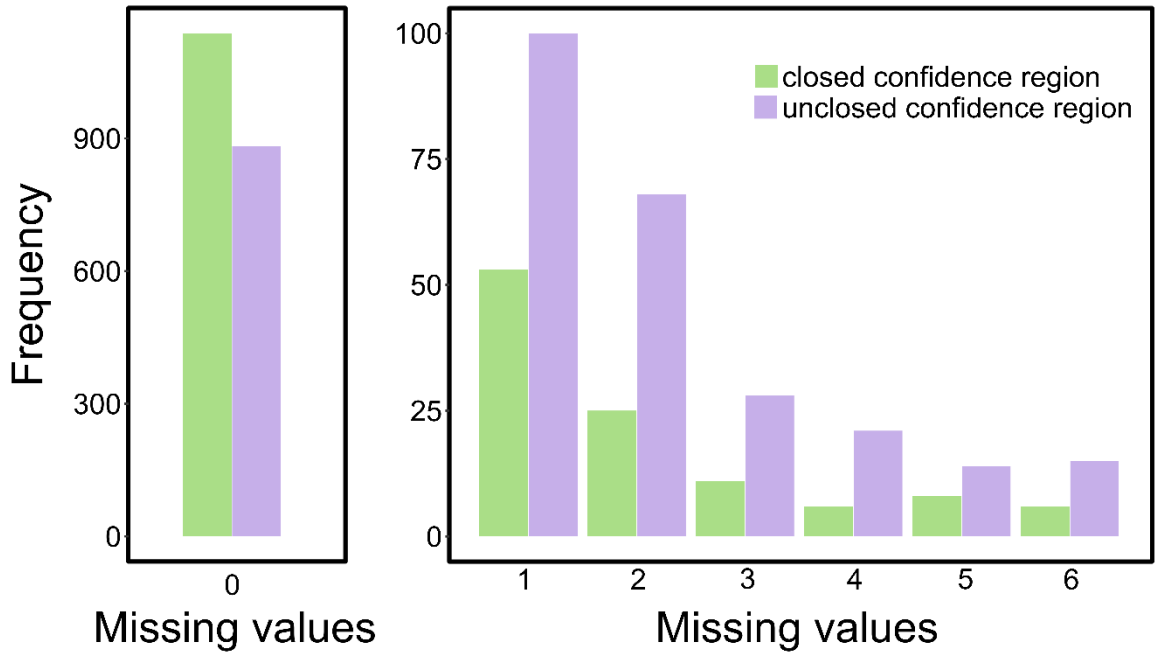


Figure III. 15 Impact of protein missing values on the model resolution. Number of mRNA-protein pairs solved of both unclosed (purple) and closed confidence region (green) according to the number of missing values in the protein dataset among the 26 samples (0: no missing value, 1-6: one to six missing values, at more than six missing values the model didn't solve the equation).

(4)- The last hypothesis to explain "unclosed confidence region" was that the model was not well-adapted to the data. The resolution was unsatisfying because the simple model described with one ODE cannot match with the data. For instance, a delay for protein synthesis could be required, the assumption of first order for the synthesis and degradation rates is unappropriated or the rate constants k_{sp} and/or k_{dp} could depend on the stage of development.

With the same model, Tchourine et al 2014 described protein expression profiles for yeast and concluded than one third of dynamic protein expression can be predicted by the model. However, they also observed low and high predictabilities of protein expression depending on genes with well-predicted profiles often monotonically increasing or decreasing. They mentioned that low predictability was often associated with drastic expression changes due to reasons other than noise. Such profiles often look smooth except for two or three consecutive outliers in the protein time-series data, these possible outliers may be due to technical artefacts or systematic errors rather than noise.

Finally, both rate constants belonging to the "closed confidence region" group were less correlated ($R^2_{spearman}$= 0.24 ($P < 0.05$), Figure III. 16B) meaning that synthesis and degradation have independent regulation. Note that they were more correlated for the "unclosed confidence region" group ($R^2_{spearman}$ = 0.72 ($P < 0.05$), Figure III. 16A). In agreement with Tchourine et al 2014, the synthesis and degradation rates do not correlate within one treatment, consistent with their independent regulation.

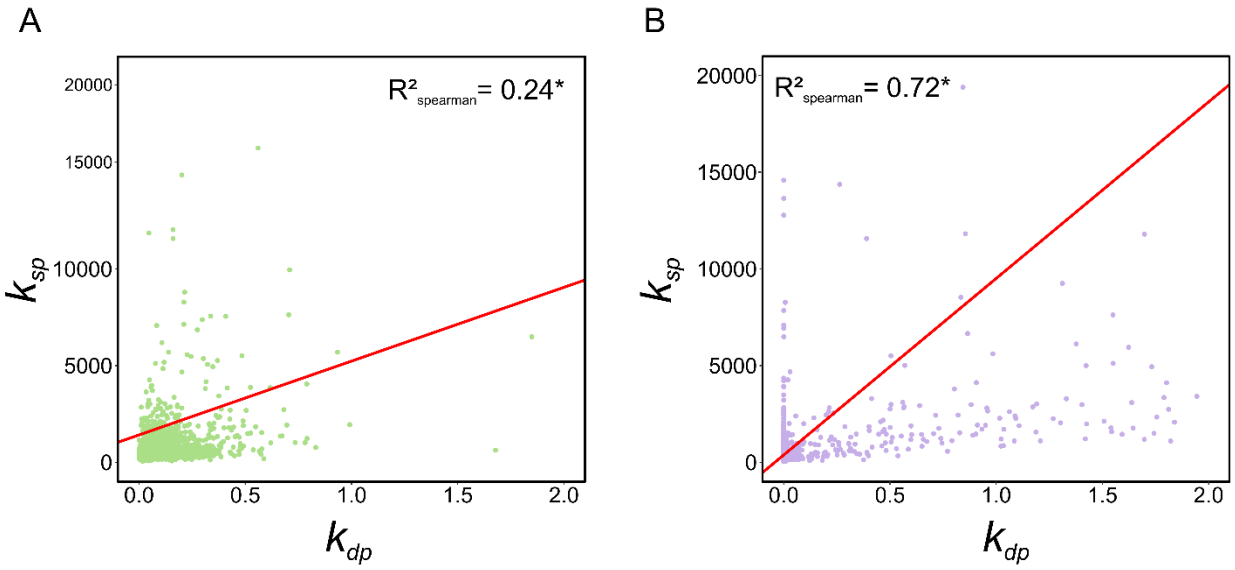


Figure III. 16 Spearman correlation analysis were performed between kdp and ksp separated according to the confidence region: closed (A) and unclosed (B). All coefficients of determination were significant (*). Linear regression was displayed by red line.

All together, these results showed that this group of unclosed confidence region contained protein for which the profiles could not be properly estimated from the mRNA data with the model. The main suspected reason was the missing values in the proteins dataset.

3.2 Analysis of well determined synthesis and degradation rate constants

We considered here the group of "closed confidence region", containing 1247 mRNA-protein pairs and we examined both rate constants k_{sp} and/or k_{dp} calculated after the model resolution. In this section we intended to understand the global meaning of these constants and we searched for

biological relevance of the results. We also compared our results with the rate constants published in the literature.

The median values obtained for both rate constants were 0.093 and 639.8 day^{-1} for degradation and synthesis, respectively, thus the degradation rate constants were about 6400 times lower than the synthesis rate constants (Figure III. 17, Figure III. 12).

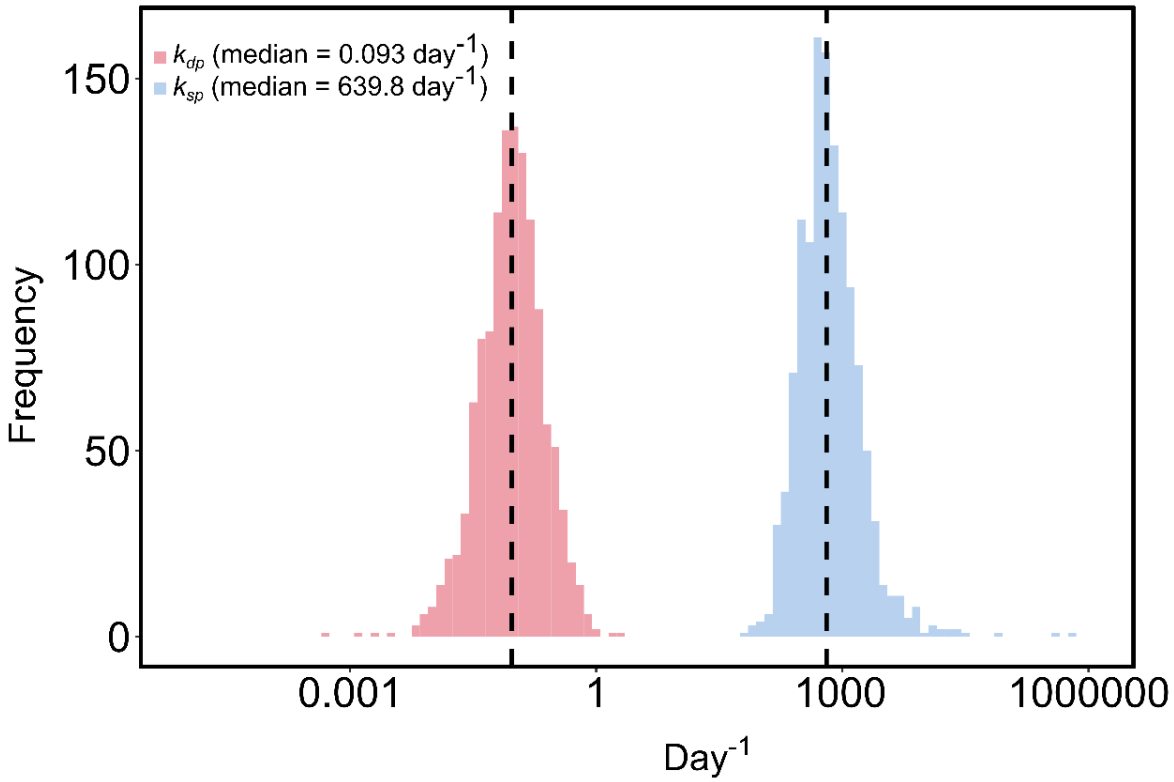


Figure III. 17 Distribution of rate constants: kdp (red) and ksp (blue). Medians (dashed line) of kdp and ksp were determined at 0.093 day-1 and 639.8 day-1, respectively.

Synthesis rate constant (k_{sp})

The synthesis or translation rate constant corresponds to 'how many proteins are made from each mRNA template per day'. Thus, this synthesis rate constant is tightly related to the protein/transcript ratio ($R^2_{spearman} = 0.91$ ($P < 0.05$), Figure III. 18).

The correlation between k_{sp} and the protein/mRNA ratio can be explained by an increased translation efficiency and/or high protein stability. Indeed, a high k_{sp} did not necessarily lead to an abundant protein, as the abundance also depends on the degradation rate constant.

Actually, this correlation traduced a pseudo-steady state, with no net protein synthesis (*dP/dt=0*), thus with synthesis and degradation rates similar, as described by the Equation III. 8.

$$k_{sp}\, r(t) = \left(k_{dp} + \mu(t)\right) p(t) \qquad \text{Equation III. 8}$$

From the linear regression, the slope (0.3239) represented the term (k_{dp}+ $\mu(t)$) and intercept was close to zero.

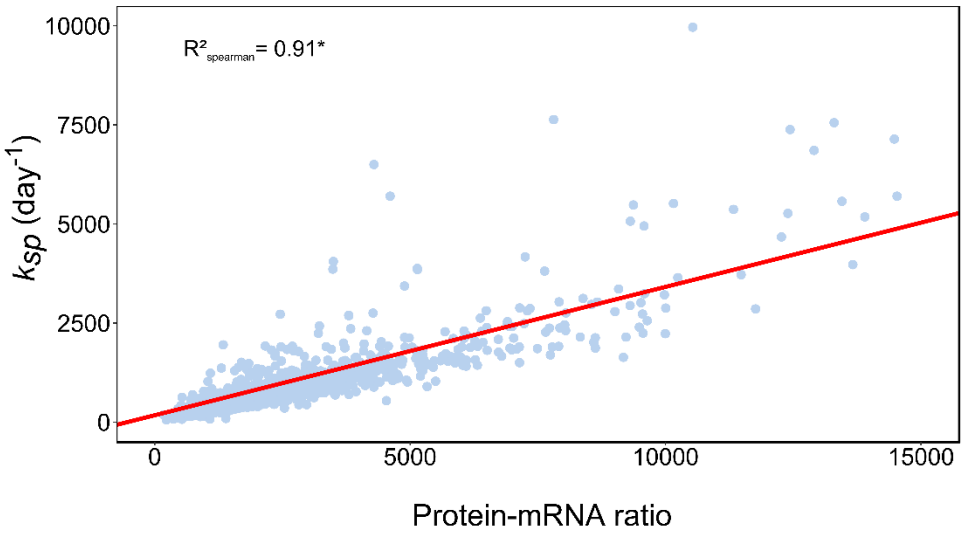


Figure III. 18 Correlation analysis between protein/mRNA ratio and ksp (day^{-1}). The 1247 protein concentrations were averaged over the nine fruit developmental stages and divided by the corresponding mRNA averaged concentrations, resulting in 1247 protein/mRNA ratios. The correlation (Spearman) between ratios and ksp was found significant ($R^2 = 0.91$, $P < 0.05$).

From a biological point of view, the 1247 synthesis rate constants were not distributed similarly in the main functional categories (Figure III. 19). The highest k_{sp} median (882 day^{-1}) was observed for the 61 "signalling" proteins in agreement with the high synthesis rate expected for these proteins, while the lowest k_{sp} median was observed for the 25 "secondary metabolism" proteins (443 day^{-1}).

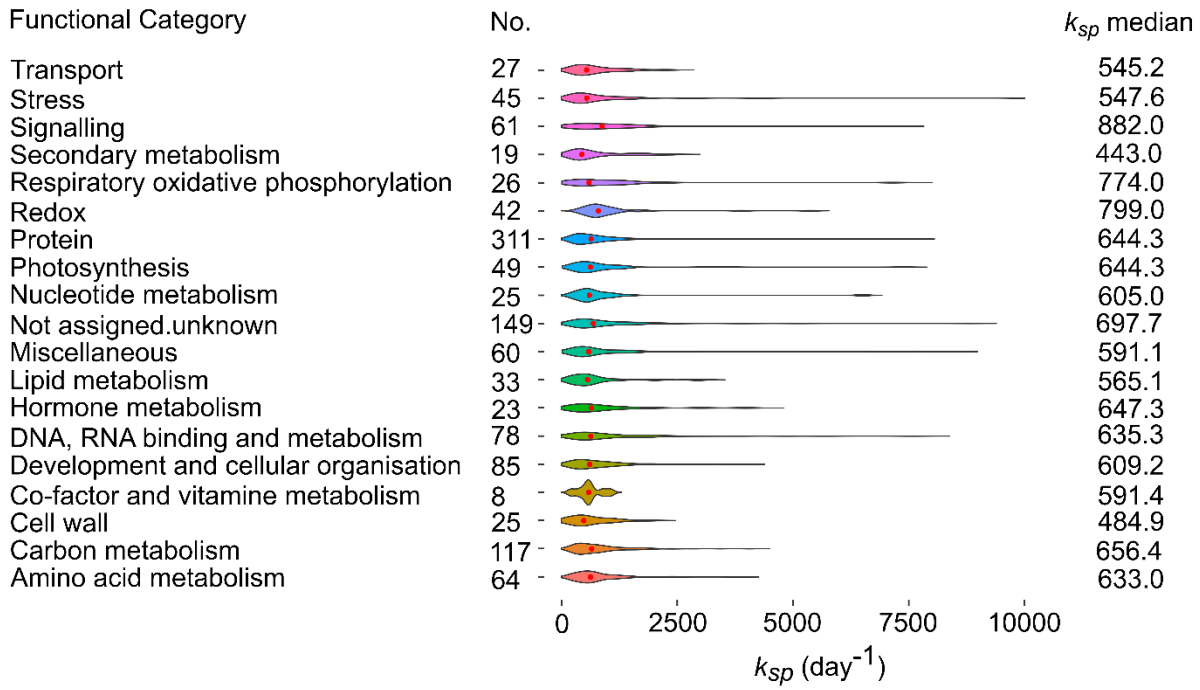


Figure III. 19 Functional categories and k_{sp} (day^{-1}) associated to the 1247 protein-mRNA pairs. The 1247 protein-mRNA pairs were assigned to functional category using the simplified MAPMAN file (Thimm et al., 2004) and the k_{sp} distribution and median (●) associated to the 19 functional categories. Number of protein-mRNA pairs (No.) and k_{sp} median determined per functional category were presented.

Then, we searched if subcellular localization was relevant to discern the synthesis rate constants. The 1247 k_{sp} were not distributed similarly in the 10 subcellular localizations provided by the MultiLoc2 prediction program (Blum et al., 2009) (Figure III. 20). More than half of the proteins were localized in the cytoplasm. The highest k_{sp} median (775.7 day^{-1}) was observed for the 24 extracellular proteins while the lowest k_{sp} median was observed for the 21 Golgi located proteins (357.5 day^{-1}).

As protein translation is a universal process, especially highly conserved in eukaryotes cells (see Introduction p) results were compared to published rate constants data. Unfortunately, to date only few papers report comparable datasets of synthesis rate constants. We picked two papers describing large sets of synthesis rate constants determined in mammalian cells (fibroblasts, (Schwanhäusser et al., 2011)) and more recently in yeast (Lahtvee et al., 2017) and superimposed the distributions (Figure III. 21).

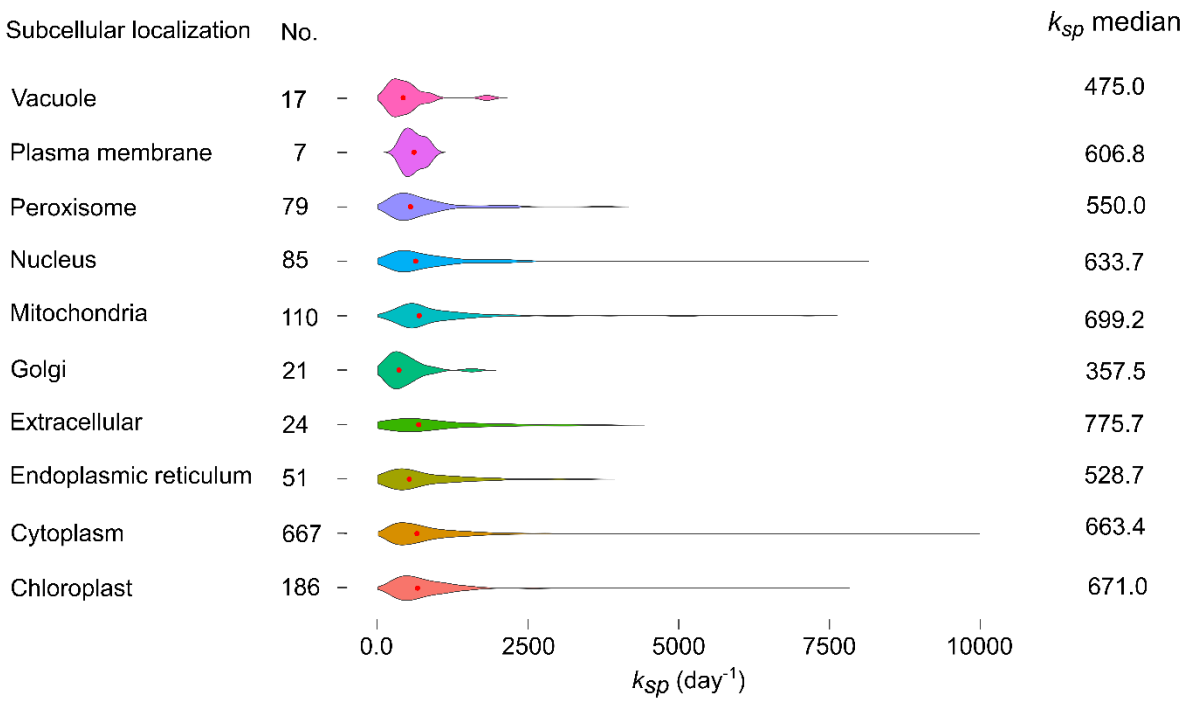


Figure III. 20 Subcellular localization and ksp (day-1) associated to the 1247 protein-mRNA pairs. The 1247 protein-mRNA pairs were assigned to the most probable subcellular localization using MultiLoc2. Number of protein-mRNA pairs (No.) and ksp median (●) determined per subcellular localization were presented.

In the case of yeast (Lahtvee et al., 2017), the translation efficiency was estimated at steady state (the growth rate (µ) equal to the dilution rate (D) equal to 0.1 h^{-1}) according to the Equation III. 6: k_{sp} named k_{TL} was calculated as following

$$k_{TL} = C_{prot}\left(k_{deg} + \mu\right)/C_{mRNA}$$

, where C_{prot} and C_{mRNA} refer to the measured absolute protein and mRNA abundances.

A set of 1115 values reported in yeast was expressed on a day basis prior to the comparison with the k_{sp} values found for the tomato fruit pericarp.

In the case of mammal cells, the translation efficiency was estimated for more than 4200 proteins and was converted from h^{-1} to day^{-1}. Figure III. 21 showed higher median values for yeast (4930.3 day^{-1}) and mammal cells (2981.0 day^{-1}) than for tomato (data were log_{10}-scaled distributed). Indeed, the median synthesis rate constant of tomato (640.4 day^{-1}) was about five (4.65) times lower than the one of yeast and about eight (7.7) times lower than the one of fibroblast (mammal).

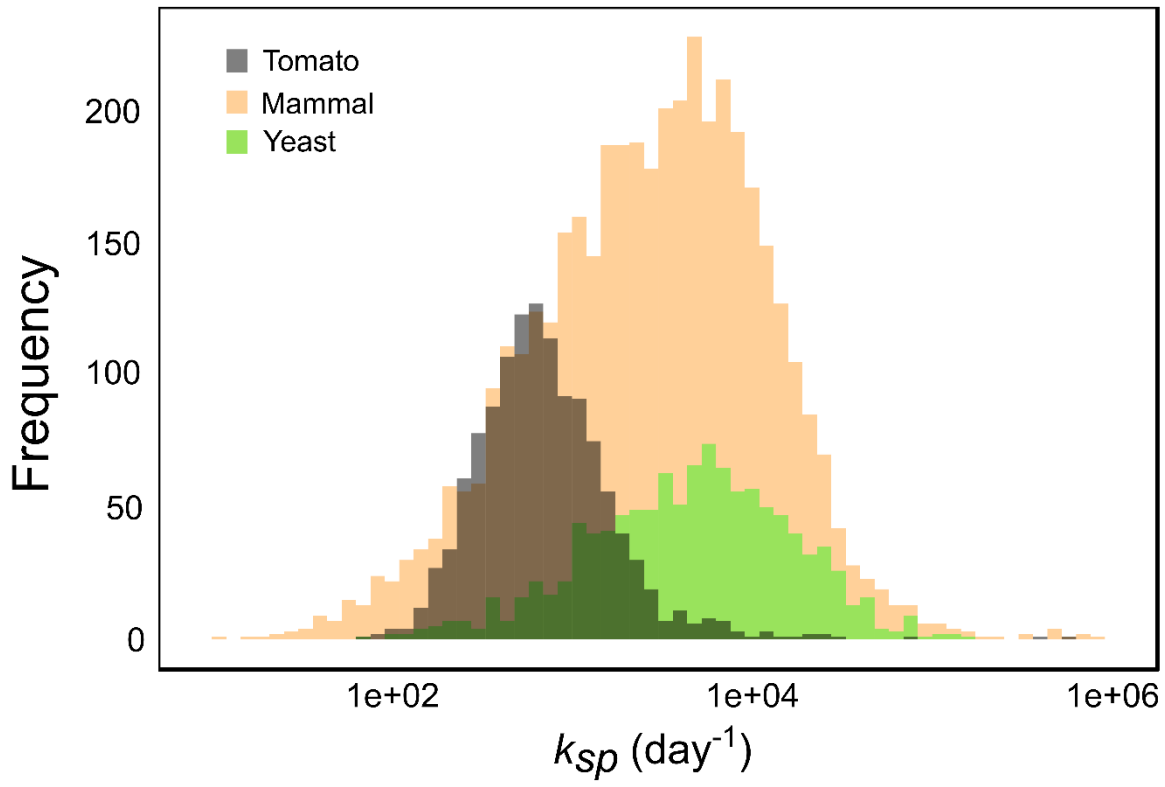


Figure III. 21 Comparison of ksp (day-1) between organisms: 1115 yeast ksp (green, Lahtvee et al., 2017); 4247 mammals ksp (yellow, Schwanhäusser et al., 2011) and 1247 tomato pericarp cells ksp (grey). ksp values were log10-scaled.

To refine the comparison, we searched for Arabidopsis homologous of yeast, mammal and tomato proteins based on proteins sequences. A threshold of 60% identity between homologous proteins sequences was used to safely filter out unsure alignments, resulting in 1091 tomato, 263 mammal and 85 yeast proteins. In order to compare yeast and mammal to tomato k_{sp}, we selected yeast and mammal proteins corresponding to tomato proteins. Finally, 100 human-tomato k_{sp} pairs and 47 yeast-tomato k_{sp} pairs were identified. As in Figure III. 21, k_{sp} medians were higher for yeast (11392.6 day^{-1}) and mammal (6618.5 day^{-1}) than for tomato (635.6 day^{-1} and 727.2 day^{-1}, resp). These results make sense with the fact that despite that the translation is a universal process, the regulation of protein synthesis can distinguish organisms.

Finally, to quantify the translation, we went further concerning the k_{sp} rate constant (Equation III. 9) inspired from the equation reported by Piques et al., (2009) where the rate of protein synthesis was dependant of (1) the ribosome density on transcripts in the polynosomial fraction (number of ribosomes per transcript) and (2) the rate of ribosome progression/elongation (number

of amino acids added per second and per ribosome). While several polysomial fractions (large, small….) can be measured, for sake of simplicity we assumed here only one fraction with a same ribosomal density per transcript, thus the equation was:

$$Velong = \frac{k_{sp}\ Lp}{Nrib\ Lg}$$ Equation III. 9

, where $Velong$ the overall speed of ribosome elongation assumed to be determined by the rates of its three major steps - initiation, elongation and termination (in amino acids / ribosome / day), Lp the protein length (amino acids), $Nrib$ the number of ribosomes per transcript (ribosomes / kb) and Lg the gene length (kb).

We estimated the elongation rate with a known ribosomal density from 4 to 6 ribosomes per kb, as Iwasaki and Ingolia (2016) reported that ribosomes could be separated by 200 or 250 pb along the transcript. Assuming a ribosome density of 4 or 6 ribosomes/kb and Lg/Lp ratio of 3.10^{-3} (as three nucleotides are required for one amino acid, here in kb), the elongation rate $Velong$ estimated from the median k_{sp} (640 day^{-1}) was 0.62 or 0. 42 amino acids / ribosome / sec. This elongation rate appeared to be lower than the one reported by Iwasaki and Ingolia (2016), which ranges from 3 to 10 amino acids / ribosome / sec for eukaryote cells or by Piques et al. (2009), which ranges from 1 to 8 amino acids / ribosome / sec. Conversely, an elongation rate of 3 amino acids / ribosome / sec with a ribosomal density equal to 4 ribosomes / kb lead to a synthesis rate constant of 3110 day^{-1}, five times higher than the median calculated by the model and the tomato dataset.

Degradation rate constant (k_{dp})

We still considered the group of "closed confidence region", containing 1247 mRNA-protein pairs and we examined the degradation rate constants k_{dp} determined by the model resolution. The k_{dp} median value obtained was 0.093 day^{-1} (Figure III. 17, red) which corresponds approximatively to a lifetime of 1/0.093= 10.8 days and a half-life of the protein ($t_{1/2}$) of 7.45 days or 180 hours according to Equation III. 5.

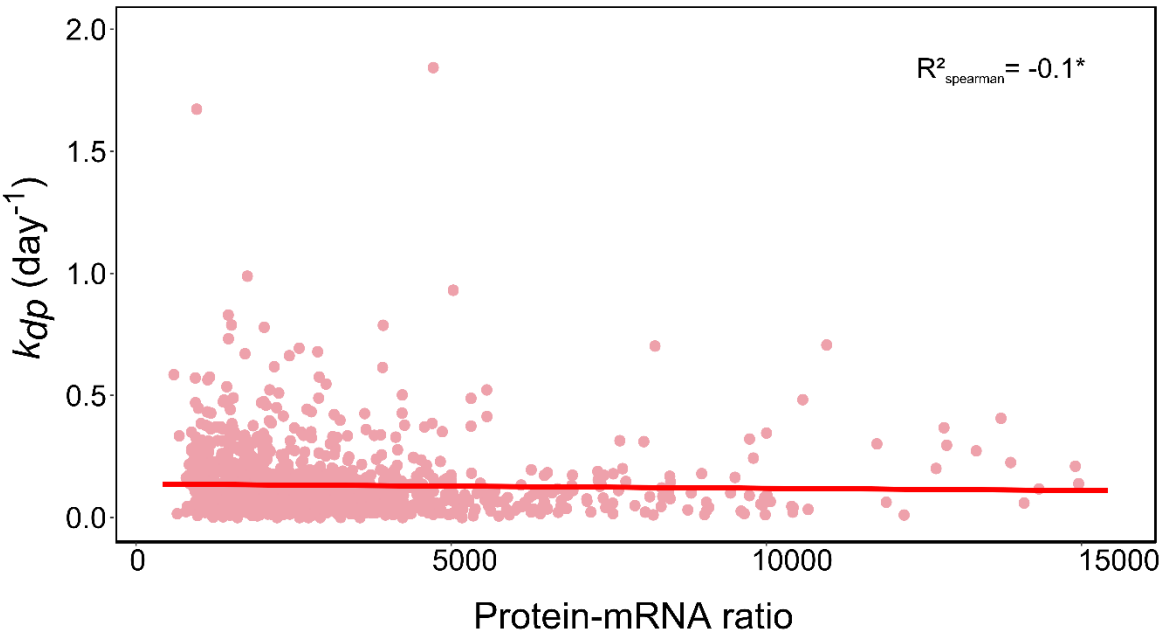


Figure III. 22 Correlation analysis between protein-mRNA ratio and k_{dp} (day^{-1}). The 1247 proteins concentration were averaged over the nine stages and divided by the mRNA averaged concentration, resulting in 1247 protein-mRNA ratios. The correlation (Spearman) between ratios and kdp was found significant ($R^2 = -0.1$, $P < 0.05$).

Contrary to k_{sp}, no correlation has been found between the degradation rate constant and the protein abundance ($R^2_{spearman} = -0.06$ ($P < 0.05$)), mRNA abundances ($R^2_{spearman} = -0.019$) and protein-mRNA ratios (Figure III. 22).

As performed with the synthesis rate constants, the 1247 degradation rate constants were differently distributed in the main functional categories (Figure III. 23). The highest medians were observed for the 78 proteins associated with "DNA-RNA binding and metabolism", the 45 proteins associated to "Stress metabolism" and the 23 proteins of "Hormone metabolism" with k_{dp} medians equal to 0.14, 0.13 and 0.12 day^{-1}, respectively. Conversely, lowest medians were observed for the 8 proteins associated to co-factor and vitamin metabolism, the 19 proteins associated to secondary metabolism and the 27 proteins involved in transport with k_{dp} equal to 0.04, 0.05 and 0.05 day^{-1}, respectively.

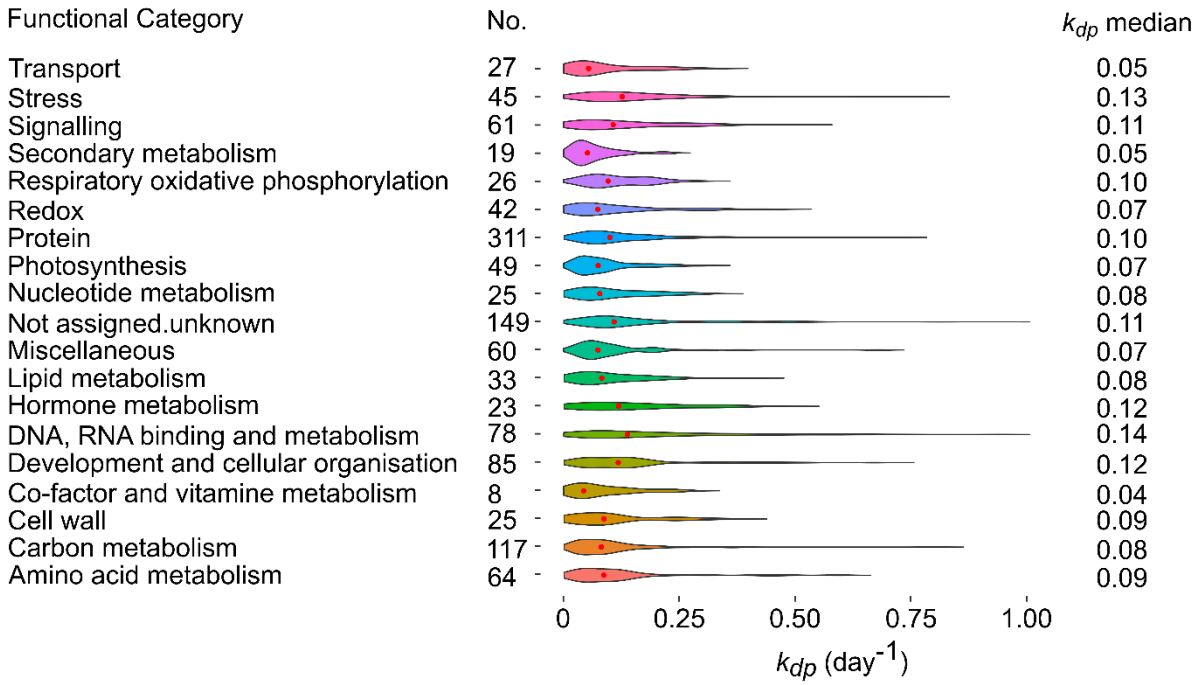


Figure III. 23 Functional categories and k_{dp} (day^{-1}) associated to the 1247 protein-mRNA pairs. The 1247 protein-mRNA pairs were assigned to functional categories using the simplified MAPMAN file (Thimm et al., 2004). Number of protein-mRNA pairs (No.) and k_{dp} median (●) determined per functional category were presented.

Using the Plant and Alga-Protein Annotation Suite (PrAS), 19 physicochemical and structural properties of tomato proteins were obtained. As Arabidopsis but not tomato plant database was in PrAS resources, Arabidopsis homologues of the 1247 tomato proteins were selected. Without filtering on percentage of identity between Arabidopsis and tomato proteins sequence, 1375 identifiers were matched to the 1247 tomato proteins. Then, we searched to what extent a subset of the protein properties, for instance the protein length, the degree of ubiquitination, hydrophobicity or the amino acid composition could influence the magnitude of degradation rate constants (Table III. 1). Based on correlation analysis (Spearman) and non-parametric analysis (Kruskal Wallis), no clear relation has been established. Note that slight negative coefficients of determination ($P < 0.05$) were determined between k_{dp} and nonpolar amino acid and hydropathy protein property while a slightly positive correlation ($P < 0.05$) was determined with protein disorder and ubiquitylation site. Also, to perform an exhaustive analysis, protein properties should be confirmed by different predictive software using different predictive algorithms.

Table III. 1 Evaluation of the influence of properties values on k_{dp} (day^{-1}). Spearman correlation analysis were performed when protein property was quantitative and non-parametric test was performed when protein property was qualitative (Solubility, Subcellular location, Cleavage sites).

	Protein properties (PrAS)	Spearman R^2 (* for $P < 0.05$)/ Kruskal-Wallis test
Physicochemical parameters	Length (aa)	0.01
	Charged amino acid	0.19*
	Nonpolar amino acid	-0.23*
	Acidic amino acid	0.06*
	Basic amino acid	0.21*
	Isoelectric point	0.07*
	Hydropathy (GRAVY)	-0.26*
Secondary structure	β sheet	0.19*
	Intrinsic disorder	0.18*
	Protein cleavage sites	$P_{kruskal} < 0.05$
	Transmembrane helices	-0.02
	S-S bond	-0.3
Others	Ubiquitylation site	0.13*
	N-glycosylation site	0.02
	O-glycosylation	0.02
	Protein solubility	$P_{kruskal} < 0.05$
	Subcellular location	$P_{kruskal} < 0.05$

The 1247 degradation rate constants were plotted according to their subcellular compartments (Figure III. 24). As expected and coherently with the results obtained for the synthesis rate constant (Figure III. 20), more than half of proteins were located in the cytoplasm, 15% in the chloroplast, 9% in mitochondria and less than 7% were located in the nucleus. All the medians associated to the subcellular compartments were close to the median unless for the 85 proteins associated to the nucleus and the 24 extracellular proteins displaying higher k_{dp} values (median values 0.15 and 0.12 day^{-1}, respectively) suggesting less stability associated to these compartments. Surprisingly, the k_{dp} value median associated to the vacuole was the lowest (0.06 day^{-1}) suggesting that the 17 vacuolar proteins adapted to an acidic environment are particularly stable. It should be interesting to describe protein properties of these 17 vacuolar proteins to characterize parameters associated to their high stability.

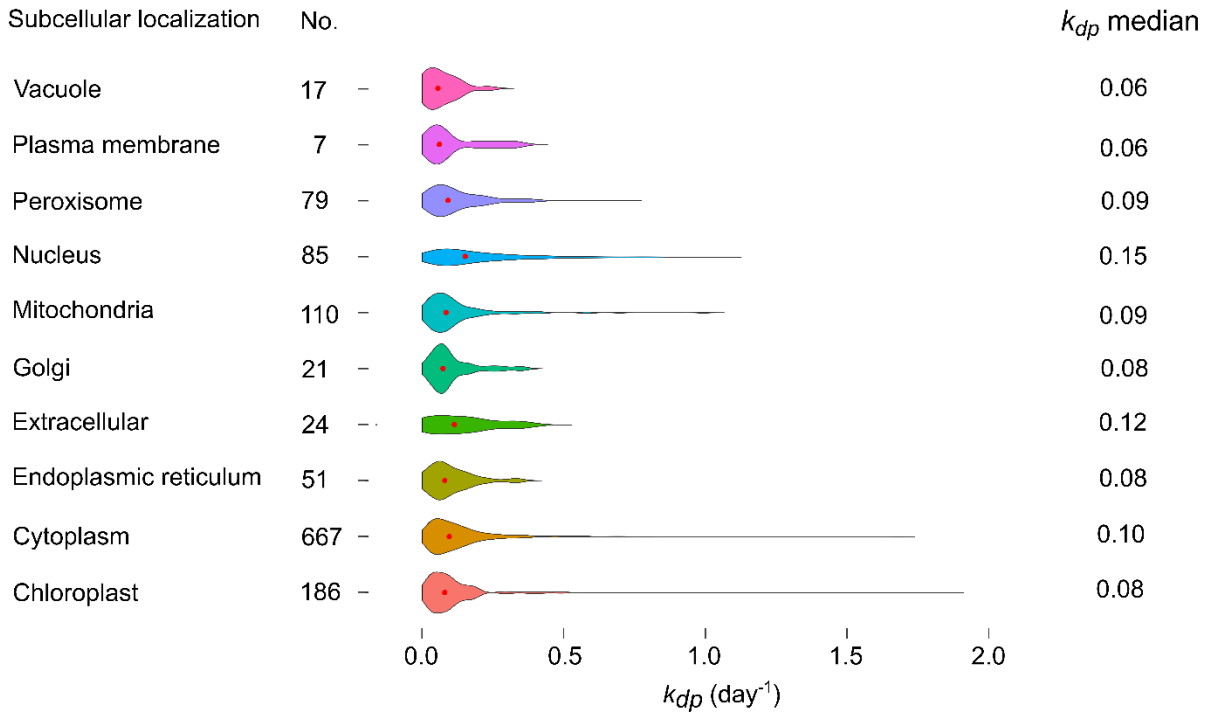


Figure III. 24 Subcellular localization and k_{dp} (day^{-1}) associated to the 1247 protein-mRNA pairs. The 1247 protein-mRNA pairs were assigned to the most probable subcellular localization using MultiLoc2. Number of protein-mRNA pairs (No.) and k_{sp} median (●) determined per subcellular localization were presented

We then compared our results with published degradation rate constants obtained by Harvey Millars' group with two plant species: barley leaves (Nelson et al., 2014) and *Arabidopsis thaliana*

leaves (Li et al., 2017b). For these experiments using ^{15}N labelling, 508, 1011 and 1127 rate constants were respectively obtained.

The distributions of the degradation rate constants determined with these plant species/tissues presented were in the same range as those found in the present work (Figure III. 25A).

We also compared our results with published degradation rate constants picked in the previously cited papers reporting data obtained with mammal cells (fibroblasts, Schwanhaüsser et al. 2011) and yeast (Lahtvee et al. 2017) (Figure III. 25B).

In the case of yeast, the degradation rate constant that had been estimated for a set of 1384 proteins had been expressed on a daily basis to be comparable to our tomato k_{dp} values.

In the case of mammal cells, the constants that had been estimated for more than 4200 proteins and had been converted from h^{-1} to day^{-1}.

Although, the estimated tomato k_{dp} values were in the range of those published, higher median values were found for yeast (1.03) and mammal cells (0.35) than for tomato, which was at 0.093 day^{-1} (data were log_{10}-scaled distributed, Figure III.26B). This suggested that plant proteins were more stable.

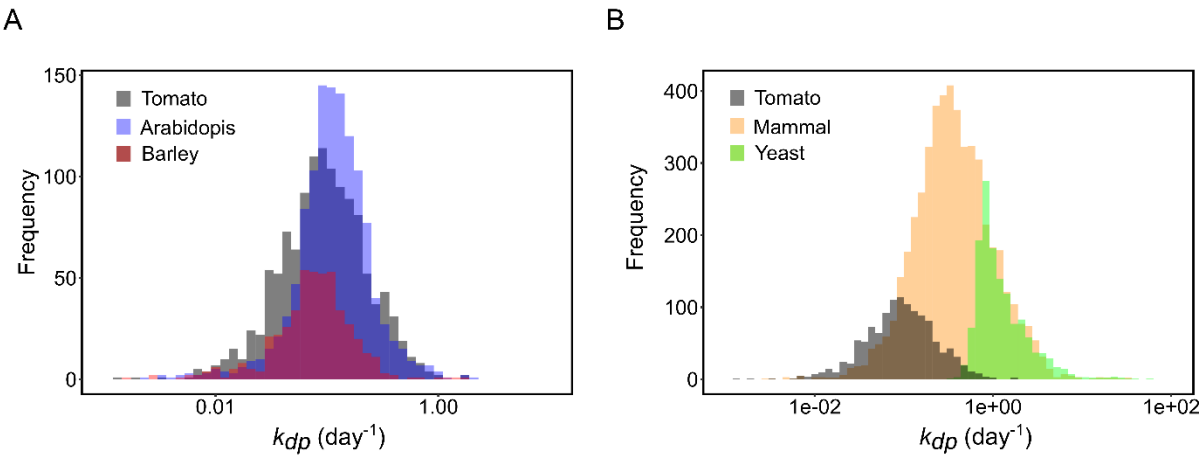


Figure III. 25 Comparison of kdp (day-1) between plant models and other organisms. 1247 tomato kdp were compared to kdp of plant organisms (A): 1228 ksp from Arabidopsis leaf (red, Li et al., 2017), 505 ksp from barley leaf (blue, Nelson et al., 2014) and to kdp of mammal and yeast (B): 1384 yeast ksp (green, Lahtvee et al., 2017); 5028 mammals ksp (yellow, Schwanhäusser et al., 2011). All kdp were expressed in day minus one.

To go further, blasts have been searched in all datasets and orthologous *Arabidopsis* genes have been found for yeast, and human cells (mammals) and tomato. In the case of barley leaves, orthologous *Arabidopsis* genes were already mentioned in the paper (Nelson et al., 2014). The results were filtered according to the homology (% identity > 60%) with Arabidopsis sequences. This significantly reduced the number of variables. Spearman coefficients of determination were higher when tomato k_{dp} were compared to Barley and Arabidopsis k_{dp} than to mammal and yeast (Figure III. 26). Despite disappointing results of the correlation analysis, we noted that few k_{dp} were almost equal between species. Most of these similar k_{dp} proteins were obtained with plant comparison (barley vs tomato and Arabidopsis vs tomato). One perspective should be to identity these subsets of proteins and determine their functions and properties.

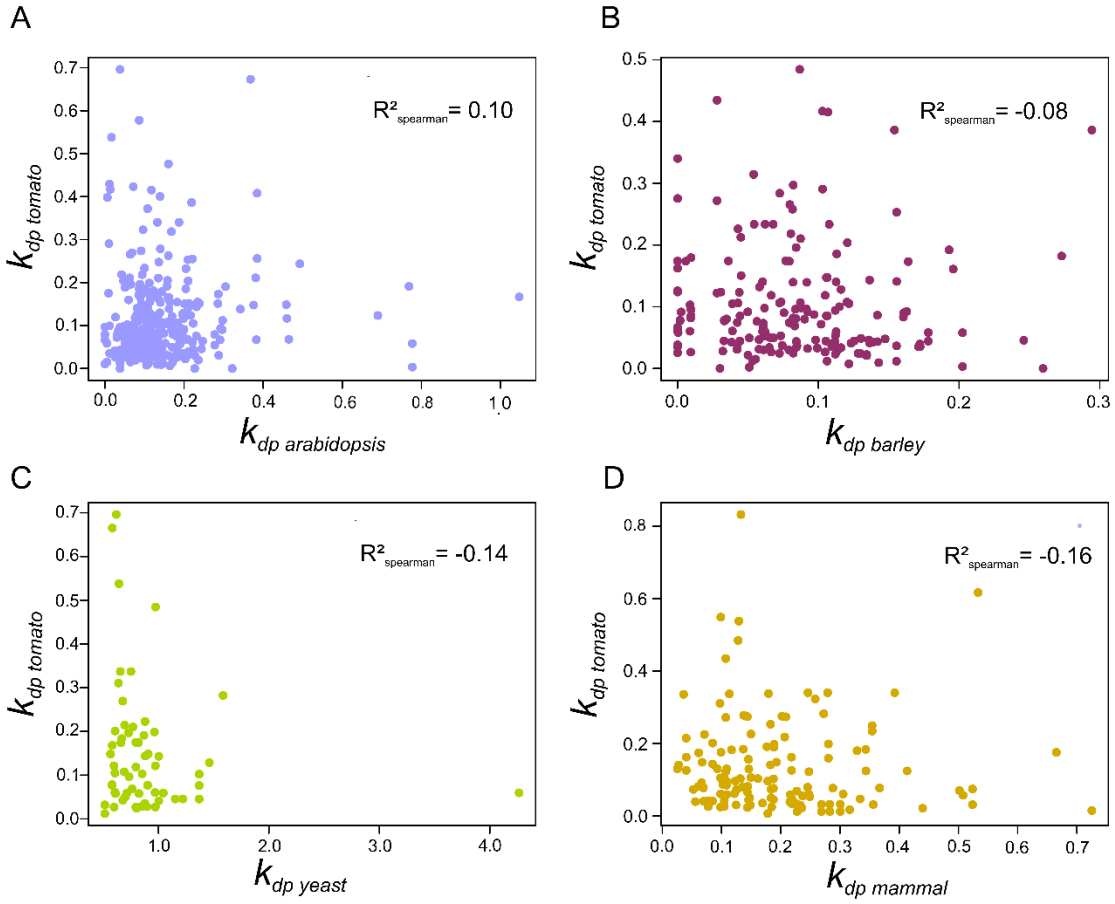


Figure III. 26 Correlation analysis between k_{dp} (day^{-1}) between plant models and other organisms. Arabidopsis identifies were used to determine tomato proteins homologous in barley (180 proteins), mammal (134 proteins), yeast (61 proteins) and Arabidopsis (362 proteins) data. Then, Spearman correlation analysis were performed.

Conclusions and perspectives

With the recent sequencing of its genome (*S. lycopersicum* HEINZ assembly v2.40; ITAG2.4), tomato fruit, the model for fleshy fruit, could benefit from large scale analyses such as proteomic, transcriptomic and genomic throughout its development.

In this study, these four omics data –transcriptomic, metabolomic, proteomic and activomic – were acquired on tomato (var. Moneymaker) and analyzed in a developmental time-series (9 stages, from 7.7 DPA to 53 DPA). To our knowledge, this is the first time that such quantitative data set was produced representing an extensive source of information. Moreover, an absolute quantification was searched for the four omics data set, using internal standards in the case of the metabolome and transcriptome or using mathematical /statistical approach for the proteome. The LC-MS/MS label-free absolute quantification of the proteome was cross-validated with 32 enzymes activities and similarly the absolute quantification of transcriptomic data obtained by RNA-Seq has been cross-validated using qRT-PCR of about 70 genes expression.

The analysis of fruit development with these four omics has characterized the cell division by a high concentration of chlorophyll, sugars, mainly imported from leaves by the phloem (Osorio et al., 2014), and proteins involved in the photosynthesis, proteins and amino acid metabolism. In parallel, the ripening phase was characterized by an increase of phosphate-sugars, organic acids involved in the "umami" taste of tomato such as the glutamate and pigment (carotene, phytoene..). Proteins and transcripts involved in the redox, amino acid and vitamin metabolism were enhanced. Among proteins and transcripts especially enhanced at the beginning of ripening (48.5-50.3 DPA) classic ripening markers (RIN and NOR transcription factor, PSY) have been found. TCA cycle metabolism appeared also to be improved especially the NADP-IDH pointed here at the protein and enzyme activity level. Thus, the integrative analysis of these four omics data set confirmed changes observed in previous publications on tomato fruit development (Carrari and Fernie, 2006; Osorio et al., 2011; Biais et al., 2014). However, functional analysis of the proteome and transcriptome data presented here, and elsewhere, were depending of the genome annotation, which requires to be completed and continuously updated. Moreover, an enrichment analysis, *.i.e* the relative proportion of selected genes associated to functional categories compared to the genome,

could be perform with gene ontology classifications available on the Gene Ontology Consortium website (http://www.geneontology.org/), such as Panther.

The integrative analysis, from PCA highlighted the complexity of large-scale analysis. Software especially developed to integrate several omics, such as MixOmics could be tried further than conventional correlative analysis (Rohart et al., 2017). The identification and characterization of candidate genes being tedious, the integrative analysis of "N-levels" omics should be a great help as it has been in Tohge et al., (2014) and Sánchez et al., (2013).

After the integrative analysis, the quantitative transcriptomic and proteomic data were used to model the process of protein translation based on one ordinary differential equation (ODE). In this model, the rate of change of protein pool over the time was explained by the balance between the rates of synthesis and degradation of the protein itself which were dependent of the synthesis (k_{sp}) and degradation (k_{dp}) rate constants, respectively. Finally, the resolution of the equation has been confidently performed for more than one thousand tomato proteins (~50% detected proteins). A global comparison of the obtained results showed that medians of tomato synthesis rate constants k_{sp} were more similar to the ones of barley and arabidopsis than to the ones of mammal and yeast. Moreover, the amino acid sequence seems to influence both rate constants k_{dp} (Dressaire et al., 2009) and k_{sp} (Table III. 1). In the same way, Li et al., (2012) obtained different rate constants k_{dp} between isoforms like mitochondrial malate dehydrogenases (At3g15020 and At1g53240), suggesting that k_{dp} of each protein is regulated by more than its amino acid sequence. To go further, it would be interesting to carry out the same analysis on: (1) different tissues (tomato leaf, fruit and root), (2) other varieties of tomato (MicroTom, Ailsa Craig…) and *Solanacea* species (pepper, eggplant...) to determine if the range of degradation rate constants k_{dp} can be explained by the tissue and phylogenetic distance. The differences between species can also be related to the division cell rate or the difference of temperature, which promotes chemical reactions, between culture cell and greenhouse conditions. The next step should be to refine the synthesis rate constant k_{sp} by considering parameters that were fixed in this study, such as, the ribosome density per transcript, the translation initiation rate, the codon usage.

To conclude, with this study we hope to have convinced and confirmed the interest of the absolute quantification of omics both for statistical and descriptive analysis and in the field of system biology.

Materials and methods

I. Plant material

The samples were provided from Biais et al. (2014) experiment. Briefly, the tomato plant (*Solanum lycopersicum* cv. 'Moneymaker') were cultivated in a greenhouse at Sainte-Livrade (France) in commercial practice conditions between June and October of 2010. Lateral stems were systematically removed and trusses were pruned to six fruits to limit fruit size heterogeneity. For sample preparation, locular tissue, seeds and placenta were removed, and the pericarp of each fruit was cut into small pieces and immediately deep frozen in liquid nitrogen (). Frozen samples were then ground into a fine powder with liquid nitrogen and stored at -80°C.

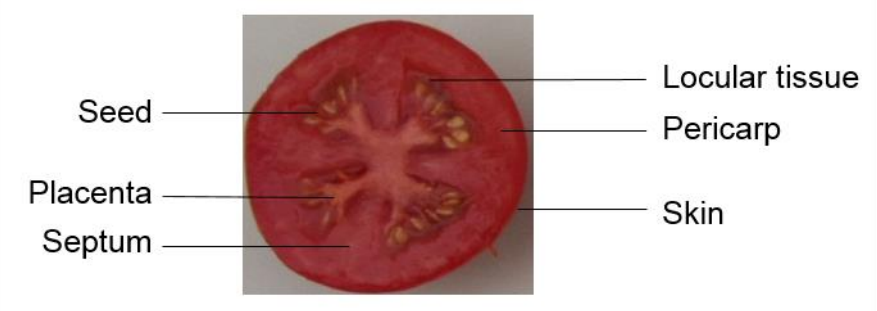


Figure MM. 1Transversal section of ripen tomato fruit (cv. Moneymaker)

Based on the age and color (OECD color gauge), fruits were harvested at nine different days post anthesis (DPA) from green to red fruit (07.7 DPA, 15 DPA, 21.7 DPA, 28 DPA, 34.3 DPA, 41.3 DPA, 48.5 DPA, 50.3 DPA, 53 DPA) and on three different trusses (trusses 5-7). A biological replicate was constituted of several fruits. The nine fruit developmental ages (in DPA) correspond to the average of replicate fruits ages.

In the first-generation of samples, used in Biais et al (2014) to quantify enzymatic activities, three biological replicates (each constituted of at least four fruits) per truss were used for trusses 5, 6 and 7. In order to run a range of analyses on the same sampling, a second-generation of samples was produced. This second-generation corresponded to the pool of the three first-generation replicates for each truss. Metabolites, proteome and transcriptome analyses were performed on the second-generation samples. Besides, for transcriptome, proteome and metabolite analyses, one

biological replicate (Truss 6) at 48.5 DPA was missing. In addition, one biological replicate (Truss 5) at 50.3 DPA was also missing only for metabolites.

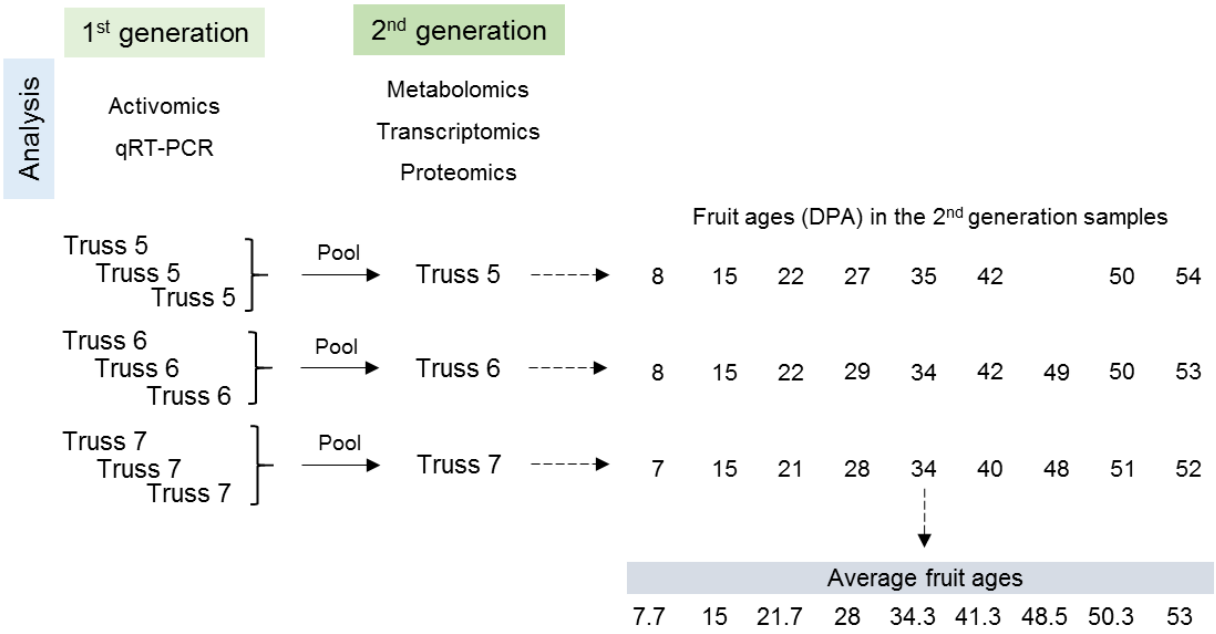


Figure MM. 2 Organization of the samples production used for analysis.

II. Proteins

2.1 Protein quantification by LC-MS/MS

2.1.1 Total protein extraction and digestion

Total tomato proteins were extracted by phenol extraction using a modified protocol described by Faurobert et al., (2007). Frozen powder of pericarp tissue (100 mg) were suspended in 10 mL of extraction buffer (0.5M Tris-HCl pH 7.5, 0.7 M sucrose, 50 mM EDTA, 0.1 M KCl, 10 mM thiourea, 2 mM phenylmethane sulfonyl fluoride, 2% β-mercaptoethanol). Then an equal volume of Water-Saturated Phenol pH 8 (Ambion) was added and the solution was incubated with steel beads on a shaker for 30 min at 4°C. After a 30-min centrifugation step (12 000 *g* at 4°C), the phenol phase was recovered and transferred into a new tube with 10 mL of extraction buffer followed by shaking without steel beads and centrifugation steps (30 min, 12 000 g, 4°C). The phenol phase was recovered, proteins were precipitated by adding the equivalent of five phenolic phase volume of cold methanol, 0.1 M of acetate ammonium, and overnight incubated at -20°C. After centrifugation (30 min, 10 000 *g*, 4°C), protein pellets were washed with methanol and cold acetone before drying under the hood. Proteins were then solubilized (6 M urea, 2 M thiourea, 30 mM Tris HCl pH 8.8, 10 mM dithiotreitol (DTT), 0.1% Zwitterionic acid labile surfactant I (Protea)) and quantified using the Plusone 2D Quant kit (GE Healthcare). Bovine Serum Albumin solution (2mg/mL) serially diluted was used to create protein assay standard curves and accurately measure protein concentration. Proteins were incubated at room temperature for 30 min and alkylated by iodoacetamide (50 mM) during incubation (60 min, dark room, RT). Proteins were diluted ten times in ammonium bicarbonate buffer (50 mM) to decrease the total urea and thiourea concentration, trypsin (800 ng) digested and incubated overnight at 37°C. Trypsin digestion was stopped by acidification (1% total volume trifluoroacetic acid). The resulting peptides were purified on solid phase extraction using a polymeric C18 column (Phenomenex) with a washing solution containing 0.06% acetic acid and 3% acetonitrile (ACN). After elution with 0.06% acetic acid and 40% ACN, peptides were dried under vacuum (Speedvac).

Concerning the extraction, digestion and purification of yeast and UPS1proteins (Sigma-Aldrich), materials and methods are described in Annex (p).

2.1.2 Protein LC-MS/MS analyses

The mass-spectrometer, associated parameters and software used to analyze tomato proteins were the same as described in the "Materials and methods" in the Annex (p). A bulk of samples was passed at the beginning, middle and end of the LC-MS/MS analysis to check the detection and retention time repeatability.

2.1.3 Protein identification

Protein identification was performed using the protein sequence database of *S. lycopersicum* HEINZ assembly v2.40 (ITAG2.4) downloaded from https://solgenomics.net/ (34725 entries). A contaminant database containing the sequences of standard contaminants was also interrogated (58 entries, trypsin, keratin, serum albumin.…). The decoy database comprised the reverse sequences of tomato proteins. Database search was performed with X!Tandem (version 2015.04.01.1; http://www.thegpm.org/TANDEM/) with the following settings. Carboxyamidomethylation of cysteine residues was set to static modification. Oxidation of methionine residues, acetylation or deamination of glutamine and cystein residues were set to possible modifications. Precursor mass precision was set to 10 ppm. Fragment mass tolerance was 0.02 Th. Only peptides with a E-value smaller than 0.05 were reported.

Identified proteins were filtered and sorted by using X!TandemPipeline (version 3.3.4 , http://pappso.inra.fr/bioinfo/xtandempipeline/). Criteria used for protein identification were (1) at least two different peptides identified with an E-value smaller than 0.01, and (2) a protein E-value (product of unique peptide E-values) smaller than 10^{-5}.

2.1.4 Peptide and protein quantification

Peptide ions were quantified based on extracted ion chromatograms (XIC) using MassChroQ software (Valot et al., 2011) version 2.2 with the following parameters: "ms2_1" alignment method, tendency_halfwindow of 10, MS1 smoothing halfwindow of 0, MS2 smoothing

halfwindow of 15, "quant1" quantification method, XIC extraction based on max, min and max ppm range of 10, anti-spike half of 5, mean filter half hedge, minmax_half_edge and maxmin_half_edge respectively set to 2 4 3. Detection thresholds on min and max at 30 000 and 50 000, respectively, peak post-matching mode.

Peptide intensities of each sample were normalized by peptide intensities obtained on the pool of the 26 samples. The most appropriate method of quantification and peptide filters were selected following the same procedure as in Annex (p). Briefly, peptides were submitted to four filters - shared peptide filter, retention time filter, occurrence filter and outlier filter- and five methods were used to compute protein abundance – iBAQ (Schwanhäusser et al., 2011); TOP3 (Silva et al., 2006); Average (Higgs et al., 2005), Average Log, Model (Blein-Nicolas and Zivy, 2016).

2.2 Enzyme activities

Protocols used for enzyme activities assays were described in (Biais et al., 2014)

III. Transcripts

3.1 RNA-Seq

3.1. 1 Library preparation

Total RNA was isolated from frozen tissue powder of tomato pericarp using Plant RNA Reagent (PureLink kit, Invitrogen^{TM}) followed by DNase treatment (DNA-free kit, Invitrogen^{TM}) and purification over RNeasy Mini spin columns (RNeasy Plant Mini kit, QIAGEN), following manufacturer's instruction. The concentration of total RNA was determined by spectrophotometry (260 nm) considering that an absorbance of 1 unit was equal to 44 µg of RNA per mL. The RNA quality was determined by quantifying the RIN value (RNA integrity number) using an RNA 6000 Nano kit (Agilent) and Agilent 2100 Bioanalyzer. A RIN of '10' standing for a total RNA without any degradation, whereas RIN of '1' marked a total RNA completely degraded. A subsample of at least 5 µg of total RNA from each of 26 RNA extracts was sent to the Get-Plage GenoTOUL facility in Toulouse (France). Transcripts were absolutely quantified using eight internal standards

spiked-in at the beginning of the total RNA extraction (in mole, $3.97x10^{-14}$ (spike 1), $4.01x10^{-15}$ (spike 2), $4.01x10^{-16}$ (spike 3), $4.02x10^{-17}$ (spike 4), $4.08x10^{-18}$ (spike 5), $4.04x10^{-19}$ (spike 6), $3.82x10^{-20}$ (spike 7), $3.82x10^{-21}$ (spike 8)).

Spike1

gtggagaaagaaatggctcgtctggcagcatttgatatggatggcactttattgatgcccgaccatcatttaggtgagaaaaccctctctactttggcgc
gactgcgtgaacgcgacattaccctcacttttgccacggggcgtcatgcgctggagatgcagcatattctcggggcgctatcgctggatgcgtatttgat
taccggcaacggaacgcgcgtgcattctctggaaggtgaacttttacatcgtgatgatttacctgcggatgtcgcggagctggtgctgtatcagcaatg
ggatacccgagccagcatgcatatcttcaatgacgacggttggtttaccgggaaagagatccctgcgttgttgcaggcatttgtctatagcggttttcgtt
atcagataatcgatgtcaaaaaaatgccactcggcagcgtcaccaagatctgcttctgtggcgatcacgacgatcttacacgcttgcagatccagcta
tacgaagcattaggcgagcgtgcacatttgtgtttttccgccacggattgcctcgaagtgctgccggtgggctgcaataaaggcgctgcattgacggtg
ctgacccaacatttaggtttatcgttgcgcgattgcatggcctttggtgatgcgatgaacgatcgcgaaatgttagtcagcgtcggtagcggatttattatg
ggcaatgcgatgccgcaactgcgcgcggagctcccgcatttaccggtgattaaaaaaaaaaaaaaaaaaaaaaaaa

Spike2

cttcgattctgttttgctacccgttgttgcgccggaagatgcttttccgctgcctgttcaatggtcattgcgctcgccatatacaccagattcagacagccaat
cacccgttgttcactgcgcagcggtacggcgatagaggcgatcttctcctcctgatcccagccgcggtagttctgtccgtaaccctctttgcgcgcgcgc
gccagaatggcttccagctttaaaggttcccgtgccagttgatagtcatcaccggggcgggaggctaacatttcgattaattccttgcggtcttgttccgg
gcaaaaggccagccaggtcaggcccgaggcggttttcagaagcggcaaacgtcgcccgaccattgcccggtgaaaggataagcggctgaaac
ggtgagtggtttcgcgtaccaccattgcatcaacatccagcgtggacacatctgtcggccataccacttcgcgcaacagatcgcccagcagtggggc
cgccagtgcagaaatccactgttcgtcacgaaatccttcgcttaattgccgcactttgatggtcagtcgaaaactatcatcggaggggctacggcgga
catatccctcttcctgcagcgtctccagcagtcgccgcacagtggtgcgatgcaggccgctgagttccgccagcagcccgacgctggcaccgccat
caagtttatttaacatatttaataacattagaccgcgggttaccgcgcacggttttgtatagaaaaaaaaaaaaaaaaaaaaaaaaa

Spike3

cttcggcaacattaactggttgatgctgaaaaacatcgaactgacggcggtgatgggcagcatttatcagtatatccacgtggcgtttcagggatcgttt
gcctgcattaccgtcggcttgatagttggggcgctggcggaacgaatccgcttctcagctgtgttgattttcgtggtggtatggctgacgctctcttacattc
cgattgcgcatatggtgtggggcggtggtttgctggcttctcacggtgcgctggatttcgcgggtggcaccgtggtgcacattaacgccgcaatcgccg
gtctggtgggcgcgtatctgataggaaaacgcgtgggcttcggtaaagaggcgtttaaaccgcacaacctgccgatggtcttcaccgggactgccat
tctctatatcggttggtttggctttaacgccgggtcagcgggcacggcgaatgaaatcgcggcactggcatttgtgaatactgtggtcgcaacggcggc
ggcaattcttggctggatcttcggtgaatgggcgctgcgtggtaagccttcactgctgggggcgtgttctggcgcgattgccggtctggtcggcgtgacg
ccagcctgcggctacattggggttggcggcgcgttgattatcggcgtggtagctggtctggcgggcttgtggggcgttaccatgctcaaacgcttgctgc
gggtggatgatccctgcgatgtcttcggtgtgcacggcgtttgtggcattgtcggctgtatcatgaccgggatttttgccgccagctcgctgggcggcgtg
ggcttcgctgaaggtgtgacgatgggccatcagttgctggtacagctggaaagcatcgccattacgatcgtctggtccggtgttgtggcatttatcggct
acaaattggcggatctgacggttggtctgcgtgtaccggaagagcaggagcgagaagggctggatgtcaacagccacggcgagaatgcctataa
cgcgtaaaaaaaaaaaaaaaaaaaaaaaaaaaa

Spike4

attcatctgcgtggcgaagaggtggcagccgtctcgttgcaatgcgtcgggccggggcattacgtcgcacaaccacaatcagaatacgcattcatgc
gtagataacattcaggcggagaataaaatggcaagagctgtacaccgtagtgggttagtggcgctgggcattgcgacagcgttgatggcatcttgtg
cattcgctgccaaagatgtggtggtggcggtaggatcgaatttcaccacgctcgatccgtatgacgcaaatgacacgttatctcaggccgtagcgaa
atcgttttaccaggggctgttcggtctggataaagagatgaaactgaaaaacgtgctggcggagagttataccgtttccgatgacggcattacttacac
cgtgaaattgcgggaaggcattaaattccaggatggcaccgatttcaacgccgcggcggtgaaagcgaatctggaccgggccagcgatccggcg
aatcatcttaaacgctataacctgtataagaatattgctaaaacggaagcgatcgatccgacaacggtaaagattaccctcaaacagccgttctcag
cgtttattaatattcttgcccatccggcgaccgcgatgatttcaccggcagcgctggaaaaatatggcaaggagattggtttttatccggtgggaaccgg
accgtatgaactggatacctggaatcagaccgattttgtgaaggtgaaaaaattcgcgggttactggcagccaggattgcccaaactggacagcata
acctggcgtccggtggcggataAcaacacccgcgcggcaatgctgcaaaccggtgaagcgcagtttgctttccccattccttacgagcaggccaca
ctgctggagaaaaacaaaaatatcgaGttgatggccagtccgtcaattatgcagcgttatatcagtatgaacgtgacgcaaaagccgttcgataacc
cgaaggtccgtgaggcgctgaattacgccaaaaaaaaaaaaaaaaaaaaaaaaaaa

Spike5

gtggcactggctggtttcgctaccgtagcgcaggccgctccgaaagataacacctggtacactggtgctaaactgggctggtcccagtaccatgaca
ctggtttcatcaacaacaatggcccgacccatgaaaaccaactgggcgctggtgcttttggtggttaccaggttaacccgtatgttggctttgaaatggg
ttacgactggttaggtcgtatgccgtacaaaggcagcgttgaaaacggtgcatacaaagctcagggcgttcaactgaccgctaaactgggttaccca

atcactgacgacctggacatctacactcgtctgggtggcatggtatggcgtgcagacactaaatccaacgtttatggtaaaaaccacgacaccggcg
tttctccggtcttcgctggcggtgttgagtacgcgatcactcctgaaatcgctacccgtctggaataccagtggaccaacaacatcggtgacgcacaca
ccatcggcactcgtccggacaacggcatgctgagcctgggtgtttcctaccgtttcggtcagggcgaagcagctccagtagttgctccggctccagct
ccggcaccggaagtacagaccaagcacttcactctgaagtctgacgttctgttcaacttcaacaaagcaaccctgaaaccggaaggtcaggctgct
ctggatcagctgtacagccagctgagcaacctggatccgaaagacggttccgtagttgttctgggttacaccgaccgcatcggttctgacgcttacaa
ccagggtctgtccgagcgccgtgctcagtctgttgttgattacctgatctccaaaggtatcccggcagacaagatctccgcacgtggtatgggcgaatc
caacccggttactggcaacacctgtgacaacgtgaaacagcgtgctgcactgatcgactgcctggctccggatcgtcgcgtagagatcgaagttaa
aggtatcaaagacgttgtaactcagccgcaggcttaagttctcgtctggtagaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

Spike6

gggctggagatcatcctacaagggcgcgacccgcgcgatgcgcgggcgttcgttgaacgtatctgcggcgtctgtactggcgtacacgccctggctt
cggtttacgccatcgaagatgctatcggtattaaagtgccggacaacgccaatatcatccgcaacattatgctggcaacgctctggtgccacgatcat
ctggtgcacttctatcagcttgccgggatggactggatcgatgtgttagatgcgctgaaagccgacccgcggaaaacctccgaactggcgcaaagtc
tctcctcttggccgaaatcatcccctggctatttcttcgacgtacaaaaccgcctgaaaaaatttgttgaaggcgggcagttggggatcttccgcaatgg
ctactgggggcacccgcagtacaaactgccgccagaagctaacctgatgggctttgcccactatctcgaagctctcgatttccagcgtgaaattgtca
aaatccacgcggtctttggcggtaaaaacccgcatccaaactggattgtcggcgggatgccttgcgccatcaacattgacgaaagcggcgcggtcg
gggcagtcaatatggaacgcctgaacctggtgcagtcaattatcacccgcacggcggacttcattaacaacgtgatgatccccgacgccttagccat
cggtcagttcaacaaaccgtggagcgaaatcggcactggtctttctgataaatgcgttctcagctacggcgcattcccggatattgccaacgactttgg
cgagaaaagtctgctgatgcctggcggcgcggtgattaacggcgacttcaacaatgtgctgccagtggatttggttgatccgcagcaggtgcaggag
tttgtcgaccacgcctggtatcgatatcccaacgatcaggtcgggcgtcatccgttcgatggcatcaccgacccgtggtacaaccccggcgatgtcaa
aggcagcgataccaacattcagcagctgaatgaacaggaacgctactcgtggatcaaagcgccacgctggcgcggtaacgcgatggaagtggg
gccgctggcgcgcacgttaatcgcttatcacaaaggcgatgctgcgaccgttgagtcggtcgatcgcatgatgtcggcgttgaacctgccgctttccg
gtatccagtcaacgttaggccgcattttgtgccgcgcgcacgaagcgcagtgggccgcaggtaagttgcagtatttcttcgaaaaaaaaaaaaaaaa
aaaaaaaaaaaa

Spike7

ggcccggaaaccctgcgtcaggtcacccaacatgccgagcacgtcgttaatgcgctgaatacggaagcgaaactgccctgcaaactggtgttgaa
accgctgggcaccacgccggatgaaatcaccgctatttgccgcgacgcgaattacgacgatcgttgcgctggtctggtggtgtggctgcacaccttct
ccccggccaaaatgtggatcaacggcctgaccatgctcaacaaaccgttgctgcaattccacacccagttcaacgcggcgctgccgtgggacagt
atcgatatggactttatgaacctgaaccagactgcacatggcggtcgcgagttcggcttcattggcgcgcgtatgcgtcagcaacatgccgtggttacc
ggtcactggcaggataaacaagcccatgagcgtatcggctcctggatgcgtcaggcggtctctaaacaggatacccgtcatctgaaagtctgccgat
ttggcgataacatgcgtgaagtggcggtcaccgatggcgataaagttgccgcacagatcaagttcggtttctccgtcaatacctgggcggttggcgat
ctggtgcaggtggtgaactccatcagcgacggcgatgttaacgcgctggtcgatgagtacgaaagctgctacaccatgacgcctgccacacaaatc
cacggcaaaaaacgacagaacgtgctggaagcggcgcgtattgagctggggatgaagcgtttcctggaacaaggtggcttccacgcgttcacca
ccacctttgaagatttgcacggtctgaaacagcttcctggtctggccgtacagcgtctgatgcagcagggttacggctttgcgggcgaaggcgactgg
aaaactgccgccctgcttcgcatcatgaaggtgatgtcaaccggtctgcagggcggcacctcctttatggaggactacacctatcacttcgagaaag
gtaatgacctggtgctcggctcccatatgctggaagtctgcccgtcgatcgccgcagaagagaaaccgatcctcgacGTTCAGCATCTCG
GTATTGGTGGTAAGGACGATCCTGCCCGCCTGATCTTCAATACCCAAACCGGCCCAGcgattgtcgccagct
tgattgatctcggcgatcgttaccgtctactggttaactgcatcgacacggtgaaaacaccgcactccctgCcgaaactgccggtggcgaatgcgctg
tggaaagcgcaaccggatctgccaactgcttccgaagcgtggatcctcgctggtggcgcgcaccataccgtctTcagccatgcactgaacctcaac
gatatgcgccaattcgccgagatgcacgacattgaaatcacggtgattgataacgacacacgcctgccagcgtttaaagacgcgctgcgctggaac
gaagtgtattacgggtttcgtcgctaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

Spike8

Aaggtctgctggcgaccggaatcattaaagcggaaggcaatatgacctccggcgatgcgcacctggcggtgaatttcccgctgctgctggaaaaa
gggcttgatggtctgcgcgaggaagtagcggaacgtcgctcgcgcatcaacctgacggtgctggaagatttacacggtgagcaattcctgaaagcg
attgatatcgtgctggtggcagtcagtgaacacattgaacgtttcgctgccctggcgcgtgaaatggccgcgaccgaaacccgcgaaagccgtcgc
gatgaactgctggcgatggcagaaaactgcgatcttatcgcccaccagccgccgcagactttctggcaggcgctgcaactgtgttacttcatccagtt
gattttgcagatcgaatctaacggtcactcagtatcgtttggtcgtatggaccagtatctctacccgtactatcgccgcgacgttgaactcaaccagacg
ctggatcgcgaacacgccatcgagatgctgcatagctgctggctgaaactgctggaagtgaacaagatccgctccggctcacactcaaaagcctct
gcgggaagtccgctgtatcagaacgtcactattggcgggcaaaatctggttgatggtcaaccaatggacgcggtgaatccactctcttacgcgatcct
cgaatcctgcggtcgcctgcgttccactcagcctaacctcagcgtgcgttaccatgcaggaatgagcaacgatttcctcgacgcctgcgtacaggtga
tccgttgcggcttcgggatgccggcgttcaacaacgacgaaatcgtgatcccggaatttattaaactcggtattgaaccgcaggacgcttatgactacg
cagcgattggttgtatagaaaccgccgtcggtggcaaatggggctatcgctgtaccggcatgagctttatcaacttcgcccgcgtgatgctggcggcg
ctggaaggcgggcatgatgccaccagcggcaaagtgttcctgccacaagaaaaagcgttgtcggcaggtaacttcaacaacttcgatgaagtgat

ggacgcgtgggatacgcaaatccgttactacacccgcaaatcaatcgaaatcgaatatgtcgtcgacaccatgctggaagagaacgtgcacgata
ttctctgctcggcgctggtggatgactgtattgagcgagcgaaaagtatcaagcaaggcggcgcgaaatatgactgggtttctggcctgcaggtcggc
attgccaacctcggcaacagcctggcggcagtgaagaaactggtgtttgaacaaggtgcgattggtcagcaacagcttgctgccgcactggcagat
gacttcgacggcctgactcacgagcagctgcgtcagcggctgattaacggtgcgccgaagtacggcaacgacgatgatactgtcgatacgctgctg
gctcgcgcttatcagacctatatcgacgaactgaaacagtaccataatccgcgctacggtcgtggtccggttggcggcaactattacgcgggtacgtc
atcaatctccgctaacgtaccgtttggcgcgcagactatggcaacaccggacgggcgtaaagcccacaccccgctggcagaaggcgcaagccc
ggcctccggtactgaccatcttggccctactgcggtcattggctcagtgggtaaactgcctacggcagcgattctcggcggcgtgttgctcaaccagaa
actgaatccggcaacgctggagaacgaatctgacaagcagaaactgatgatcctgctgcgtaccttctttgaagtgcataaaggctggcatattcagt
acaacatcgtttcccgcgaaacgctgctggatgcgaaaaaacatcccgatcagtatcgcgatctggtagtgcgtgtcgcgggctattccgcgttcttca
ccgcgctctctccagacgctcaggacgatatcatcgcccgtactgaacatatgctgtaaaaaaaaaaaaaaaaaaaaaaaaaaa

3.1. 2 Illumina sequencing

RNAseq was performed at the GeT-PlaGe core facility, INRA Toulouse. RNA-seq libraries have been prepared according to Illumina's protocols on a Tecan EVO200 liquid handler using the Illumina TruSeq Stranded mRNA sample prep kit to analyze mRNA. Briefly, mRNA were selected using poly-T beads. Then, RNA were fragmented to generate double stranded cDNA and adaptors were ligated to be sequenced. 10 cycles of PCR were applied to amplify libraries. Library quality was assessed using a Agilent Bioanalyzer and libraries were quantified by QPCR using the Kapa Library Quantification Kit. RNA-seq experiments have been performed on an Illumina HiSeq2000 or HiSeq2500 sequencer using a paired-end read length of 2x100 pb with the Illumina TruSeq SBS sequencing kits v3.

3.1. 3 Transcriptome analysis

Genes were mapped to the *Solanum lycopersicum* HEINZ assembly v2.40, concatenated with the chloroplast (gi|544163592|ref|NC_007898.3|) and mitochondrial genomes (gi|209887431|gb|FJ374974.1|), and an "artificial chromosome" containing the 8 spike sequences. Genome data was downloaded for *S. lycopersicum* from *S. lycopersicum* 2.5 and the corresponding ITAG2.4 gene models were downloaded from https://solgenomics.net/ (34725 entries). The quality of libraries was checked with FastQC (Andrews, 2010). Quality and adapter trimming was performed with Trimmomatic v0.32 (Bolger et al., 2014). Trimmed reads were mapped to their respective genomes with Star v2.4.2a (Dobin et al., 2015) and the unique counts per locus were quantified with HTSeq v0.6.1 (Anders et al., 2015); transcripts per million (TPM) was calculated from the unique counts and gene length. Normalized FPKM (fragments per kilobase per million) was calculated with cufflinks v2.2.1. Briefly, quantification based on FPKM corresponds to the

normalization of data by depth sequencing (summed fragment per sample) divided per one million followed by a normalization by the gene length. Non-default parameters that were used are presented below. FPKM were then converted in TPM quantification (transcript per million) which takes into account gene length to get relative transcript abundance, prior normalized to per million. Spikes were quantified in the same way as all the transcripts. A standard curve was used per sample to estimate the concentration (fmol.gFW^{-1}) from the TPM values. Non-default parameters used for Trimmomatic v0.32 and Star v2.4.2a are presented in Table MM. 1.

Table MM. 1 Non-default parameters

Trimmomatic (v0.32)
ILLUMINACLIP:TruSeq2-SE/PE.fa:2:30:10
LEADING:3
TRAILING:3
SLIDINGWINDOW:4:15
MINLEN: 25
Star (v2.4.2)
--outSAMstrandField intronMotif --outSAMattributes All
HTSeq (v0.6.1)
htseq-count –stranded=no
Cufflinks (v2.2.1)
-G

3.2 Quantitative real-time PCR assay

Performed in 2012 by Virginie Mengin at the MPIMP (Potsdam, Germany), the total RNA was extracted with the TRIzol kit (Invitrogen; www.thermofisher.com) using a slightly modified version of the manufacturer's instructions. Briefly, frozen tissue powder from tomato pericarp (100 mg FW) were mixed to 1 mL TRIzol and centrifuged (10 min, 12 000 g, 4 °C). Then 0.4 mL of chloroform was added to the supernatant, incubated (at least 2 min), and centrifuged (15 min, 12 000 *g*, 4°C). The total RNA contained in the aqueous phase was transferred to a new tube and precipitated by adding 0.5 volume of isopropanol per volume of TRIzol, incubated 30 minutes (RT) and then centrifuged (10 min, 12 000 *g*, 4°C). After discarding the supernatant, the pellet was washed with 1/1 volume of 75% ethanol per volume of TRIzol, vortexed, centrifuged (15 min,

7500 *g*, 4°C) and air dried under a fume hood. Each dried pellet was then resuspended in 40 µl of RNase-free water and incubated (10 min, 55°C). DNA contaminants were removed using TURBO DNA-free kit (Invitrogen; www.thermofisher.com) following manufacturer's instruction. RNA integrity number (RIN) was determined to control the RNA quality using the Agilent Technologies RNA 6000 NANO and measured with the Agilent 2100 Bioanalyzer (Agilent Technologies, www.genomics.agilent.com)). RIN values above 7 indicated no degradation of the RNA. cDNAs were constructed from RNA using SuperScript III Reverse Transcriptase kit (Invitrogen). Quantitative real-time PCR reactions were performed and run in 384 well plates pipetted using robot Evolution P3 (PerkinElmer Life Science; http://www.perkinelmer.com) and measured on a 7900 HT Fast Real-Time PCR System (Applied Biosystems; www.thermofisher.com). PCR reaction mix was consisted of 5 µL of Master Mix SYBR Green (Life Technologies; www.thermofisher.com), 1 µl of 1:20 diluted template cDNA and 4 µl of primers (0.5 µM each) in a total volume of 10 µl per PCR reaction. PCR cycling was performed as follows: 2 min at 50°C, 10 min at 95°C followed by 40 cycles of 15 s at 94°C, and 1 min at 60 °C. Melting curve cycle consisted in a progressive heating from 60 to 95 °C with 1.9°C/min rate. Data acquisition and analysis were performed using the software SDS 2.4 (Applied Biosystems; www.thermofisher.com) and Microsoft Excel. Transcripts were absolutely quantified using internal standards added in each sample prior the total RNA extraction.

IV. Metabolite measurements

4.1 Intermediate metabolites by selected reaction monitoring mass spectrometry

Twenty-four metabolites (glutamate, aspartate, glycerate, glucose-1-phosphate, glucose-6-phosphate, fructose-6-phosphate, succinate, malate, 2-oxogluatarate, uridine diphosphate glucose, ribose-5-phosphate, mixture of ribulose-5-phosphate and xylose-5-phosphate, dihydroxyacetone phosphate, adenosine monophosphate and adenosine diphosphate, adenosine diphosphate glucose, fructose bisphosphate, shikimate, ribulose 1,5-bisphosphate, sedoheptulose-7-phosphate, nicotinadenine dinucleotide and nicotinadenine dinucleotide phosphate, sucrose bisphosphate, aconitate) were analysed by ion pair reversed-phase chromatography coupled to a Finnigan TSQ

Quantum Discovery (ThermoScientific, San Jose, CA, USA) triple quadrupole MS equipped with an ESI source as previously described by Arrivault et al., (2009) with slight modifications. Aliquots of frozen tissue powder from tomato pericarp (20 mg FW) were extracted with chloroform-methanol with phase partitioning as previously described by Lunn et al., (2006). The polar phase was lyophilized and the lyophilized extracts were reconstituted in 250 µL of water before analysis. Data were acquired in negative mode by selected reaction monitoring (SRM). Quantification was performed using external calibrations curves using authentic standard compounds. ^{13}C-labelled internal standards, when available, were added to correct for matrix effects.

4.2 Polar metabolites by liquid chromatography coupled to mass spectrometry

The targeted analysis of 17 polar metabolites (leucine, phenylalanine, alanine, methionine, serine, gamma aminobutyric acid, arginine, lysine, ornithine, S-adenosyl methionine, histidine, valine, citrulline, threonine, pyroglutamic acid, aspartic acid, glutamic acid) was performed by hydrophilic interaction liquid chromatography (HILIC) coupled to mass spectrometry as previously described (Berton et al., submitted for publication) using an Acclaim Mixed-Mode HILIC-1 column (2.1 x 150 mm; 3 µm, Dionex-Thermo Scientific, Courtaboeuf, France) and an LTQ-Orbitrap MS (Thermo Scientific, Brême, Germany) equipped with an electrospray interface. Aliquots of frozen tissue powder from tomato pericarp (50 mg FW) were extracted with 300 µL ethanol/water (80:20, v:v) at 80°C in a water bath during 20 min. Acquisition was performed in positive and negative modes, in full-scan mode with a resolving power of 120 000 FWHM in the scan range of m/z 50-1000. Polar metabolites were extracted with a window tolerance of 10 ppm. Quantification was performed using ^{13}C and ^{15}N labelled internal standards to correct for matrix effects. When labelled internal standards were not available, compounds with similar chemical properties were used. Internal standards were added before extraction.

4.3 Polar metabolites by ^{1}H-NMR

Polar metabolites were extracted from 20 mg of lyophilised tomato pericarp powder with an ethanol–water series at 80°C (adapted from Moing et al., (2004)) using an automated liquid handling workstation (Hamilton, Bonaduz, Switzerland). The supernatants were combined, dried under vacuum and lyophilized. Each lyophilized extract was solubilized in 500 µL of 200 mM deuterated potassium phosphate buffer solution pH 6, containing 2 mM ethylene diamine tetraacetic acid disodium salt (EDTA), pH-adjusted with KOD solution to apparent pH 6.00 when necessary, and lyophilized again. The lyophilized titrated extracts were stored in darkness under vacuum at room temperature, before ^{1}H-NMR analysis was completed within one week. Before ^{1}H-NMR analysis, 500 µL of D_2O with sodium trimethylsilyl [2,2,3,3-d4] propionate (TSP, 0.01% w/v final concentration for chemical shift calibration) were added to each lyophilized pH-adjusted extract. The mixture was centrifuged at 17700 *g* for 5 min at room temperature. The supernatant was then transferred into a 5 mm NMR tube for acquisition. Quantitative ^{1}H-NMR spectra were recorded at 500.162 MHz and 300 K on an Avance III spectrometer (Bruker Biospin, Wissembourg, France) using a 5-mm ATMA broadband inverse probe, a 90° pulse angle and an electronic reference for quantification (Digital ERETIC, Bruker TopSpin 3.0). The assignments of metabolites in the NMR spectra were made by comparing the proton chemical shifts with literature (Mounet et al., 2007; Bénard et al., 2015) and databases values (MeRy-B, HMDB, BMRB), and by comparison with spectra of authentic compounds. For absolute quantification, four calibration curves (glucose and fructose: 1.25 to 50 mM, glutamate and glutamine: 0 to 15 mM) were prepared and analysed under the same conditions. The glucose calibration was used for the quantification of all compounds, as a function of the number of protons of selected resonances, except fructose, glutamine and glutamate that were quantified using their own calibration curve. The metabolite concentrations were calculated using AMIX (version 3.9.14, Bruker) and Excel (Microsoft, Redmond, WA, USA) softwares. Representative ^{1}H-NMR spectra of the dataset have been deposited into the Metabolomics Repository of Bordeaux MeRy-B (http://services.cbib.u-bordeaux.fr/MERYB/public/PublicREF.php?REF=T10004).

4.4 Isoprenoids

Isoprenoids were analysed by High-Performance Liquid Chromatography-Diode Array Detector (HPLC-DAD) from frozen tissue powder (100 mg FW for green fruits, 50 mg FW for turning and ripening fruits) using the extraction protocol described in Fraser et al., (2000) and modified by Mortain-Bertrand et al., (2008). Whenever possible, all subsequent manipulations were carried out on ice and shielded from light. Briefly, samples were first extracted using methanol (1 mL) and buffer (0.05 M Tris-HCl, pH 7.5), incubated with chloroform. The pooled chloroform extracts were dried upon a stream of nitrogen and stored at -20°C before analysis. Dried extracts were dissolved in ethyl acetate (200 µl for green fruits, 400 µl for turning and ripening fruits). Chromatography was performed on a Spectra system (Dionex DX 600) with an UV–vis Diode Array Detector (DAD-3000 (RS) Dionex) optimized for colored and non-colored isoprenoids (290, 330 and 460 nm). Isoprenoids were separated using a 3 µm (21 x 250 mm) reverse-phase C30 column (YMC Inc. Europe GmbH, Germany) and eluted with a 0.3 $mL.min^{-1}$ gradient of (A) methanol, (B) water/methanol (5:1) containing 1% ammonium acetate and (C) *tert*-methyl butyl ether. The volume injection was 20 µL and the column was kept at constant temperature (30°C). Data were collected and processed using Chromeleon software v.6.80 (Dionex Co., Sunnyvale, USA). Identification and absolute quantification were performed by using standards. Lycopene, β-carotene, α-carotene, lutein, chlorophyll a, chlorophyll b were purchased from Sigma-Aldrich (France). Phytoene was obtained from *Escherichia coli* harbouring the plasmids pAC-DELTA, pAC-EPSILON, pAC-PHYT kindly provided by Francis Cunningham (University of Maryland, USA). Violaxanthin was isolated from tomato leaf tissue. When standards were not available, contents were expressed as all-trans-beta-carotene or lutein equivalents depending on chromophores and spectra similarities. To check the detection and retention time repeatability, one blank and one purchased standard –lycopene or β-carotene- were injected each three and ten samples, respectively. Ten samples maximum were analyzed daily. Each biological sample was repeated three times.

V. Translation model

The resolution of mathematical model based on one ordinary differential equation was implemented with the MATLAB software (Mathworks, http://www.mathworks.fr/).

To perform the resolution, the relative growth rate ($\mu(t)$) has been estimated by fitting the growth curve throughout the tomato fruit development with known growth models (such as Logistic, Contois, Gompertz, sigmoid etc.) or polynomial regression. The benefit of a log transformation has been evaluated and the best appropriate fit was selected according to the lowest calculated error between experimental and fitted values of tomato fruit weight.

A time function was also required to describe the profile of transcripts throughout the tomato fruit development. While the mRNA values were all positives, a polynomial regression fitting tended to become negative when mRNA values were close to zero. To avoid this pitfall, a log transformation was done before fitting the data with a polynomial regression. For the polynomial regression, a degree from 2 to 6 has been tried and a degree three was found as the most appropriated for a training dataset of about 30 mRNA profiles.

To improve the numerical accuracy of the computations, the mRNA and proteins data which scales differed by several orders of magnitude (from 10^2 to 10^5) were normalized by their respective average calculated over the nine stages.

Finally, the resolution of the ODE was performed to determine both rate constants *ksp* and *kdp* applying the least-square method (*lscurvefit* function): at each time *ti* (DPAi) the sum of the square deviations between the solution of the ODE and the experimental protein content was calculated and minimized.

After the resolution, three criteria were used to evaluate the quality of the estimation: (1) a score on the mRNA fit, (2) the reliability of optimization and (3) a statistical evaluation of the quality of the rate constants.

The score on the mRNA fit was based on the percent accuracy calculated between the fit and the experimental data of mRNA: six levels of quality were attributed according to the error: < 0.1 'excellent fit'; < 0.15 'very good fit'; < 0.20 'good fit'; < 0.30 'good enough fit'; < 0.4 'poor fit'; and else > 0.4, 'bad fit'.

The reliability of optimization was given by a score reporting the quality of three optimizations. The score was a number between 0 and 10, 10 if the resolution converged three times to almost the same value starting from different initializations; 8 is if the resolution converged three times to closed values starting from different initializations; 6 if the resolution converged two times to the same value; 4 if the resolution converged two times to closed values, and 1 if the resolution converged to different values.

To evaluate statistically the quality of the constants, we calculated a confidence region for the parameters estimation:

Considering that the errors in the observations are independently distributed and that the standard deviations of the errors are all equal, choosing a significance level α, statistical considerations allow us to determine a $100(1-\alpha)\%$ confidence region for the estimators *ksp* and *kdp*. For that, in practice, on a rectangular grid around the best estimators (*ksp**,*kdp**) and for each parameter (*ksp_i*,*kdp_i*), we compared the least square values (of the errors) with the boundary values of the confidence region and a contour was plotted. In linear cases, the confidence region was delimited by an ellipse centered at (*ksp**,*kdp**) but in our case, various shapes occurred. When the second derivatives of the model were not very large, the confidence region might lead to a closed shape region, similar to an ellipse. In the case of an unclosed domain, the resolution of the model was considered as unsatisfying. Conversely, the resolution was acceptable if the domain was closed, thus the calculated rate constants were further analyzed. The area of the closed domain gives an indication of the level of accuracy. For that we used a numerical method to calculate an approximate value of the area.

VI. Statistical analyses

All statistical analyses were performed using R studio Software (http://www.rstudio.com/.) or BioStatFlow web application (http://biostatflow.org/), except for hierarchical clustering. Hierarchical clustering and heat maps were performed on mean-centered data scaled to unit variance using MEV software v4.8.1.with Pearson's correlations and complete linkage. Functional protein annotation has been acquired from MapMan (Thimm et al., 2004). The PageMan software package (Usadel et al., 2006) was used to select and display biologically relevant biological category (default parameters).

References

Agudelo-Romero P, Erban A, Rego C, Carbonell-Bejerano P, Nascimento T, Sousa L, Martínez-Zapater JM, Kopka J, Fortes AM (2015) Transcriptome and metabolome reprogramming in Vitis vinifera cv. Trincadeira berries upon infection with Botrytis cinerea. J Exp Bot **66**: 1769–85

Anders S, Pyl PT, Huber W (2015) HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics **31**: 166–169

Andrews S (2010) A quality control tool for high throughput sequence data.

Arike L, Valgepea K, Peil L, Nahku R, Adamberg K, Vilu R (2012) Comparison and applications of label-free absolute proteome quantification methods on Escherichia coli. J Proteomics **75**: 5437–5448

Arrivault S, Guenther M, Ivakov A, Feil R, Vosloh D, Van Dongen JT, Sulpice R, Stitt M (2009) Use of reverse-phase liquid chromatography, linked to tandem mass spectrometry, to profile the Calvin cycle and other metabolic intermediates in Arabidopsis rosettes at different carbon dioxide concentrations. Plant J **59**: 824–839

Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, Zimmermann P, Grossniklaus U, Gruissem W, Baginsky S (2008) Genome-Scale Proteomics Reveals Arabidopsis thaliana Gene Models and Proteome Dynamics. Science (80-) **320**: 938–941

Barsan C, Zouine M, Maza E, Bian W, Egea I, Rossignol M, Bouyssie D, Pichereaux C, Purgatto E, Bouzayen M, et al (2012) Proteomic Analysis of Chloroplast-to-Chromoplast Transition in Tomato Reveals Metabolic Shifts Coupled with Disrupted Thylakoid Biogenesis Machinery and Elevated Energy-Production Components. Plant Physiol **160**: 708–725

Bastías A, Yañez M, Osorio S, Arbona V, Gómez-Cadenas A, Fernie AR, Casaretto JA (2014) The transcription factor AREB1 regulates primary metabolic pathways in tomato fruits. J Exp Bot **65**: 2351–2363

Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, Bely B, Bingley M,

Bonilla C, Britto R, et al (2017) UniProt: The universal protein knowledgebase. Nucleic Acids Res **45**: D158–D169

Beauvoit BP, Colombie S, Monier A, Andrieu M-H, Biais B, Benard C, Cheniclet C, Dieuaide-Noubhani M, Nazaret C, Mazat J-P, et al (2014) Model-Assisted Analysis of Sugar Metabolism throughout Tomato Fruit Development Reveals Enzyme and Carrier Properties in Relation to Vacuole Expansion. Plant Cell **26**: 3224–3242

Behal RH, Oliver DJ (1997) Biochemical and Molecular Characterization of Fumarase from Plants : Purification and Characterization of the Enzyme — Cloning , Sequencing , and Expression of the Gene 1. **348**: 65–74

Bénard C, Bernillon S, Biais B, Osorio S, Maucourt M, Ballias P, Deborde C, Colombié S, Cabasson C, Jacob D, et al (2015) Metabolomic profiling in tomato reveals diel compositional changes in fruit affected by source–sink relationships. J Exp Bot **66**: 3391–3404

Bergervoet JHW, Berhoeven H a., Gilissen LJW, Bino RJ (1996) High amounts of nuclear DNA in tomato (Lycopersicon esculentum Mill.) pericarp. Plant Sci **116**: 141–145

Beynon RJ, Doherty MK, Pratt JM, Gaskell SJ (2005) Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. Nat Methods **2**: 587–589

Biais B, Benard C, Beauvoit B, Colombie S, Prodhomme D, Menard G, Bernillon S, Gehl B, Gautier H, Ballias P, et al (2014) Remarkable Reproducibility of Enzyme Activity Profiles in Tomato Fruits Grown under Contrasting Environments Provides a Roadmap for Studies of Fruit Metabolism. Plant Physiol **164**: 1204–1221

Blein-Nicolas M, Xu H, de Vienne D, Giraud C, Huet S, Zivy M (2012) Including shared peptides for estimating protein abundances: A significant improvement for quantitative proteomics. Proteomics **12**: 2797–2801

Blein-Nicolas M, Zivy M (2016) Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics. Biochim Biophys Acta - Proteins Proteomics **1864**: 883–895

Blum T, Briesemeister S, Kohlbacher O (2009) MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. BMC Bioinformatics **10**: 274

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics **30**: 2114–2120

Bonghi C, Manganaris G (2012) Systems Biology Approaches Reveal New Insights into Mechanisms Regulating Fresh Fruit Quality. Omi. Technol. CRC Press, pp 201–226

Brummell DA, Harpster MH (2001) Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. Plant Mol Biol **47**: 311–339

Carli P, Barone A, Fogliano V, Frusciante L, Ercolano MR (2011) Dissection of genetic and environmental factors involved in tomato organoleptic quality. BMC Plant Biol **11**: 58

Carrari F, Baxter C, Usadel B, Urbanczyk-Wochniak E, Zanor M-I, Nunes-Nesi A, Nikiforova V, Centero D, Ratzka A, Pauly M, et al (2006) Integrated Analysis of Metabolite 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

Carrari F, Fernie AR (2006) Metabolic regulation underlying tomato fruit development. J Exp Bot **57**: 1883–1897

Carrillo-López A, Yahia EM (2012) Changes in color-related compounds in tomato fruit exocarp and mesocarp during ripening using HPLC-APcI+-mass Spectrometry. J Food Sci Technol **51**: 2720–2726

Causse M, Duffe P, Gomez MC, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-Chamley M, Rothan C (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. J Exp Bot **55**: 1671–1685

Cheniclet C (2005) Cell Expansion and Endoreduplication Show a Large Genetic Variability in Pericarp and Contribute Strongly to Tomato Fruit Growth. Plant Physiol **139**: 1984–1994

Claydon AJ, Thom MD, Hurst JL, Beynon RJ (2012) Protein turnover: Measurement of

proteome dynamics by whole animal metabolic labelling with stable isotope labelled amino acids. Proteomics **12**: 1194–1206

Clough T, Key M, Ott I, Ragg S, Schadow G, Vitek O (2009) Protein Quantification in Label-Free LC-MS Experiments. J Proteome Res **8**: 5275–5284

Colombié S, Beauvoit B, Nazaret C, Bénard C, Vercambre G, Le Gall S, Biais B, Cabasson C, Maucourt M, Bernillon S, et al (2017) Respiration climacteric in tomato fruits elucidated by constraint-based modelling. New Phytol **213**: 1726–1739

Copeland L, Preiss J (1981) Purification of Spinach Leaf ADPglucose Pyrophosphorylase. Plant Physiol **68**: 996–1001

D'Esposito D, Ferriello F, Molin AD, Diretto G, Sacco A, Minio A, Barone A, Di Monaco R, Cavella S, Tardella L, et al (2017) Unraveling the complexity of transcriptomic, metabolomic and quality environmental response of tomato fruit. BMC Plant Biol **17**: 66

Dai ZW, Léon C, Feil R, Lunn JE, Delrot S, Gomès E (2013) Metabolic profiling reveals coordinated switches in primary carbohydrate metabolism in grape berry (Vitis vinifera L.), a non-climacteric fleshy fruit. J Exp Bot **64**: 1345–1355

Dobin A, Gingeras TR, Spring C (2015) Mapping RNA-seq Reads with STAR. Curr Protoc Bioinforma 11.14.1-11.14.19

Doerflinger FC, Miller WB, Nock JF, Watkins CB (2015) Variations in zonal fruit starch concentrations of apples – a developmental phenomenon or an indication of ripening? Hortic Res **2**: 15047

Dressaire C, Gitton C, Loubière P, Monnet V, Queinnec I, Cocaign-Bousquet M (2009) Transcriptome and proteome exploration to model translation efficiency and protein stability in Lactococcus lactis. PLoS Comput Biol **5(12)**: e1000606

Eran Pichersky LDG (1984) Plant Triose Phosphate Isomerase Isozymes. Plant Physiol **74**: 340–347

Fabre B, Lambour T, Bouyssié D, Menneteau T, Monsarrat B, Burlet-Schiltz O, Bousquet-Dubouch MP (2014) Comparison of label-free quantification methods for the determination

of protein complexes subunits stoichiometry. EuPA Open Proteomics **4**: 82–86

Faurobert M, Pelpoir E, Chaïb J (2007) Phenol extraction of proteins for proteomic studies of recalcitrant plant tissues. Methods Mol Biol **355**: 9–14

Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Tecle IY, Strickler SR, Bombarely A, Fisher-York T, Pujar A, Foerster H, et al (2015) The Sol Genomics Network (SGN)-from genotype to phenotype to breeding. Nucleic Acids Res **43**: D1036–D1041

Fraser PD, Pinto ME, Holloway DE, Bramley PM (2000) Technical advance: application of high-performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids. Plant J **24**: 551–558

Galvez S, Bismuth E, Sarda C, Gadal P (1994) Purif ication and Characterization of Chloroplastic NADP- lsocitrate Dehydrogenase from Mixotrophic Tobacco Cells Comparison with the Cytosolic lsoenzyme. Plant phy 593–600

Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci **100**: 6940–6945

Ghan R, Van Sluyter SC, Hochberg U, Degu A, Hopper DW, Tillet RL, Schlauch KA, Haynes PA, Fait A, Cramer GR (2015) Five omic technologies are concordant in differentiating the biochemical characteristics of the berries of five grapevine (Vitis vinifera L.) cultivars. BMC Genomics **16**: 946

Gillaspy G, Ben-David H, Gruissem W (1993) Fruits: A Developmental Perspective. PLANT CELL ONLINE **5**: 1439–1451

Giovannoni JJ (2007) Fruit ripening mutants yield insights into ripening control. Curr Opin Plant Biol **10**: 283–289

Gonda I, Bar E, Portnoy V, Lev S, Burger J, Schaffer AA, Tadmor Y, Gepstein S, Giovannoni JJ, Katzir N, et al (2010) Branched-chain and aromatic amino acid catabolism into aroma volatiles in Cucumis melo L. fruit. J Exp Bot **61**: 1111–1123

Gourieroux AM, Holzapfel BP, Scollary GR, McCully ME, Canny MJ, Rogiers SY (2016)

The amino acid distribution in rachis xylem sap and phloem exudate of Vitis vinifera "Cabernet Sauvignon" bunches. Plant Physiol Biochem **105**: 45–54

Graeve K, von Schaewen a, Scheibe R (1994) Purification, characterization, and cDNA sequence of glucose-6-phosphate dehydrogenase from potato (Solanum tuberosum L.). Plant J **5**: 353–361

Grover SD, Canellas PF, Wedding RT (1981) Purification of NAD malic enzyme from potato and investigation of some physical and kinetic properties. Arch Biochem Biophys **209**: 396–407

Guo X, Xu J, Cui X, Chen H, Qi H (2017) iTRAQ-based Protein Profiling and Fruit Quality Changes at Different Development Stages of Oriental Melon. BMC Plant Biol **17**: 28

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol **17**: 994–999

Haar T von der (2012) Mathematical and Computational Modelling of Ribosomal Movement and Protein Synthesis: an overview. Comput Struct Biotechnol **1(1)**: e201204002

Higgs RE, Knierman MD, Gelfanova V, Butler JP, Hale JE (2005) Comprehensive label-free method for the relative quantification of proteins from biological samples. J Proteome Res **4**: 1442–1450

Hill SA, ap Rees T (1993) Fluxes of carbohydrate metabolism in ripening bananas. Planta **192**: 52–60

Igamberdiev AU, Gardeström P (2003) Regulation of NAD- and NADP-dependent isocitrate dehydrogenases by reduction levels of pyridine nucleotides in mitochondria and cytosol of pea leaves. Biochim Biophys Acta - Bioenerg **1606**: 117–125

Ippel JH, Pouvreau L, Kroef T, Gruppen H, Versteeg G, van den Putten P, Struik PC, van Mierlo CPM (2004) In vivo uniform15N-isotope labelling of plants: Using the greenhouse for structural proteomics. Proteomics **4**: 226–234

Isaac JE, Rhodes MJC (1982) Purification and properties of phosphofructokinase from fruits of Lycopersicon Esculentum. Phytochemistry **21**: 1553–1556

Ishihara H, Obata T, Sulpice R, Fernie AR, Stitt M (2015) Quantifying Protein Synthesis and Degradation in Arabidopsis by Dynamic $^{13}CO_2$ Labeling and Analysis of Enrichment in Individual Amino Acids in Their Free Pools and in Protein. Plant Physiol **168**: 74–93

Ishizaki K, Larson TR, Schauer N, Fernie AR, Graham IA, Leaver CJ (2005) The Critical Role of Arabidopsis Electron-Transfer Flavoprotein:Ubiquinone Oxidoreductase during Dark-Induced Starvation. Plant Cell Online **17**: 2587–2600

Jourda C, Cardi C, Gibert O, Giraldo Toro A, Ricci J, Mbéguié-A-Mbéguié D, Yahiaoui N (2016) Lineage-Specific Evolutionary Histories and Regulation of Major Starch Metabolism Genes during Banana Ripening. Front Plant Sci **7**: 1778

Katz E, Boo KH, Kim HY, Eigenheer RA, Phinney BS, Shulaev V, Negre-Zakharov F, Sadka A, Blumwald E (2011) Label-free shotgun proteomics and metabolite analysis reveal a significant metabolic shift during citrus fruit development. J Exp Bot **62**: 5367–5384

Kaufman S, Alivisatos SPA (1995) Purification and properties of the phosphorylating enzyme from spinach. J Biol Chem **216**: 141–152

Koo AJK, Howe GA (2009) The wound hormone jasmonate. Phytochemistry **70**: 1571–1580

Ladror US, Latshaw SP, Marcus F (1990) Spinach cytosolic fructose-1,6-bisphosphatase: Purification, enzyme properties and structural comparisons. Eur J Biochem **189**: 89–94

Lahtvee PJ, Sánchez BJ, Smialowska A, Kasvandik S, Elsemman IE, Gatto F, Nielsen J (2017) Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast. Cell Syst **4**: 495–504.e5

Lal SK, Kelley PM, Elthon TF (1994) Purification and differential expression of enolase from maize. Physiol Plant **91**: 587–592

Laurent JM, Vogel C, Kwon T, Craig SA, Boutz DR, Huse HK, Nozue K, Walia H, Whiteley M, Ronald PC, et al (2010) Protein abundances are more conserved than mRNA abundances across diverse taxa. Proteomics **10**: 4209–4212

Lewandowska D, ten Have S, Hodge K, Tillemans V, Lamond AI, Brown JWS (2013) Plant SILAC: Stable-Isotope Labelling with Amino Acids of Arabidopsis Seedlings for

Quantitative Proteomics. PLoS One **8**: 1–8

Li L, Nelson CJ, Solheim C, Whelan J, Millar AH (2012) Determining Degradation and Synthesis Rates of *Arabidopsis* Proteins Using the Kinetics of Progressive ^{15}N Labeling of Two-dimensional Gel-separated Protein Spots. Mol Cell Proteomics **11**: M111.010025

Li L, Nelson CJ, Trösch J, Castleden I, Huang S, Millar AH (2017a) Protein Degradation Rate in *Arabidopsis thaliana* Leaf Growth and Development. Plant Cell **29**: 207–228

Li L, Nelson CJ, Trösch J, Castleden I, Huang S, Millar AH (2017b) Protein Degradation Rate in *Arabidopsis thaliana* Leaf Growth and Development. Plant Cell **29**: 207–228

Li M, Li D, Feng F, Zhang S, Ma F, Cheng L (2016) Proteomic analysis reveals dynamic regulation of fruit development and sugar and acid accumulation in apple. J Exp Bot **67**: 5145–5157

Liu H, Sadygov RG, Yates III JR (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem **76**: 4193–4201

Lu P, Vogel C, Wang R, Yao X, Marcotte EM (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol **25**: 117–124

Lunn JE, Feil R, Hendriks JHM, Gibon Y, Morcuende R, Osuna D, Scheible W-R, Carillo P, Hajirezaei M-R, Stitt M (2006) Sugar-induced increases in trehalose 6-phosphate are correlated with redox activation of ADPglucose pyrophosphorylase and higher rates of starch synthesis in *Arabidopsis thaliana*. Biochem J **397**: 139–148

Lytovchenko A, Eickmeier I, Pons C, Osorio S, Szecowka M, Lehmberg K, Arrivault S, Tohge T, Pineda B, Anton MT, et al (2011) Tomato Fruit Photosynthesis Is Seemingly Unimportant in Primary Metabolism and Ripening But Plays a Considerable Role in Seed Development. Plant Physiol **157**: 1650–1663

Maier T, Güell M, Serrano L (2009) Correlation of mRNA and protein in complex biological samples. FEBS Lett **583**: 3966–3973

Marano MR, Serra EC, Orellano EG, Carrillo N (1993) The path of chromoplast development

in fruits and flowers. Plant Sci **94**: 1–17

Marcelis LFM, Hofman-Eijer LRB (1995) The contribution of fruit photosynthesis to the carbon requirement of cucumber fruits as affected by irradiance, temperature and ontogeny. Physiol Plant **93**: 476–483

Martinez-Barajas E, Krohn BM, Stark DM, Randall DD (1997) Purification and characterization of recombinant tomato fruit (Lycopersicon esculentum Mill.) fructokinase expressed in Escherichia coli. Protein Expr Purif **11**: 41–46

Maruyama H, Easterday L, Chang HC, Lane D (1966) The Enzymatic Carboxylation of Phosphoenolpyruvate. **241**: 2405–2412

Matic S, Åkerlund HE, Everitt E, Widell S (2004) Sucrose synthase isoforms in cultured tobacco cells. Plant Physiol Biochem **42**: 299–306

McIntosh M, Fitzgibbon M (2009) Biomarker validation by targeted mass spectrometry. Nat Biotechnol **27**: 622–623

McMorrow EM, Bradbeer JW (1990) Separation, Purification, and Comparative Properties of Chloroplast and Cytoplasmic Phosphoglycerate Kinase from Barley Leaves. Plant Physiol **93**: 374–383

Michels S, Scagliarinl S, Setaa F Della, Caries C, Riva M, Trostb P, Branlanta G (1994) Arguments against a close relationship between non-phosphorylating and phosphorylating glyceraldehyde-3-phosphate dehydrogenases. FEBS Lett **339**: 97–100

Mirza SP, Greene AS, Olivier M (2008) 18O labeling over a coffee break: a rapid strategy for quantitative proteomics. J Proteome Res **7**: 3042–8

Moing A, Maucourt M, Renaud C, Gaudillère M, Brouquisse R, Lebouteiller B, Gousset-Dupont A, Vidal J, Granot D, Denoyes-Rothan B, et al (2004) Quantitative metabolic profiling by 1-dimensional 1H-NMR analyses: Application to plant genetics and functional genomics. Funct Plant Biol **31**: 889–902

Moisan M-C, Rivoal J (2011) Purification to homogeneity and characterization of nonproteolyzed potato (*Solanum tuberosum*) tuber hexokinase 1. Botany **89**: 289–299

Moorhead GB, Plaxton WC (1990) Purification and characterization of cytosolic aldolase from carrot storage root. Biochem J **269**: 133–139

Mortain-Bertrand A, Stammitti L, Telef N, Colardelle P, Brouquisse R, Rolin D, Gallusci P (2008) Effects of exogenous glucose on carotenoid accumulation in tomato leaves. Physiol Plant **134**: 246–256

Mounet F, Lemaire-Chamley M, Maucourt M, Cabasson C, Giraudel J-L, Deborde C, Lessire R, Gallusci P, Bertrand A, Gaudillère M, et al (2007) Quantitative metabolic profiles of tomato flesh and seeds during fruit development: complementary analysis with ANN and PCA. Metabolomics **3**: 273–288

Mounet F, Moing A, Garcia V, Petit J, Maucourt M, Deborde C, Fruit IB, Bordeaux C De, Ornon FV, France FM (2009) Gene and Metabolite Regulatory Network Analysis of Early Developing Fruit Tissues Highlights New Candidate Genes for the Control of Tomato Fruit Composition and Development. **149**: 1505–1528

Mueller L, Fernandez-Pozo N (2016) Tomato Databases. *In* M Causse, J Giovannoni, M Bouzayen, M Zouine, eds, Tomato Genome. Compend. Plant Genomes. Springer, Berlin. Heidelberg., pp 245–255

Muhammad S, Kumazawa K (1974) Assimilation and transport of nitrogen in rice I. 15N-labelled ammonium nitrogen. Plant Cell Physiol **15**: 747–758

Navarre D a, Wendehenne D, Durner J, Noad R, Klessig DF (2000) Nitric oxide modulates the activity of tobacco aconitase. Plant Physiol **122**: 573–582

Negri AS, Prinsi B, Failla O, Scienza A, Espen L (2015) Proteomic and metabolic traits of grape exocarp to explain different anthocyanin concentrations of the cultivars. Front Plant Sci **6**: 603

Nelson CJ, Alexova R, Jacoby RP, Millar AH (2014) Proteins with High Turnover Rate in Barley Leaves Estimated by Proteome Analysis Combined with in Planta Isotope Labeling. Plant Physiol **166**: 91–108

Nelson CJ, Millar AH (2015) Protein turnover in plant biology. Nat Plants **1**: 15017

Ning K, Fermin D, Nesvizhskii AI (2012) Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. J Proteome Res **11**: 2261–71

Oikawa A, Otsuka T, Nakabayashi R, Jikumaru Y, Isuzugawa K, Murayama H, Saito K, Shiratake K (2015) Metabolic profiling of developing pear fruits reveals dynamic variation in primary and secondary metabolites, including plant hormones. PLoS One **10**: 1–18

Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG (2005) Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics. Mol Cell Proteomics **4**: 1487–1502

Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics **1**: 376–86

Osorio S, Alba R, Damasceno CMB, Lopez-Casado G, Lohse M, Zanor MI, Tohge T, Usadel B, Rose JKC, Fei Z, et al (2011) Systems Biology of Tomato Fruit Development: Combined Transcript, Protein, and Metabolite Analysis of Tomato Transcription Factor (nor, rin) and Ethylene Receptor (Nr) Mutants Reveals Novel Regulatory Interactions. Plant Physiol **157**: 405–425

Osorio S, Ruan Y-L, Fernie AR (2014) An update on source-to-sink carbon partitioning in tomato. Front Plant Sci **5**: 1–11

Osorio S, Scossa F, Fernie AR (2013) Molecular regulation of fruit ripening. Front Plant Sci **4**: 1–8

Peyraud R, Dubiella U, Barbacci A, Genin S, Raffaele S, Roby D (2017) Advances on plant-pathogen interactions from molecular toward systems biology perspectives. Plant J **90**: 720–737

Piques M, Schulze WX, Höhne M, Usadel B, Gibon Y, Rohwer J, Stitt M (2009) Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis. Mol Syst Biol **5**: 314

Van de Poel B, Bulens I, Hertog ML, Nicolai BM, Geeraerd AH (2014) A transcriptomics-

based kinetic model for ethylene biosynthesis in tomato (Solanum lycopersicum) fruit: development, validation and exploration of novel regulatory mechanisms. New Phytol **202**: 952–963

Van de Poel B, Van Der Straeten D (2014) 1-aminocyclopropane-1-carboxylic acid (ACC) in plants: more than just the precursor of ethylene! Front Plant Sci **5**: 640

Ponnala L, Wang Y, Sun Q, van Wijk KJ (2014) Correlation of mRNA and protein abundance in the developing maize leaf. Plant J **78**: 424–440

Powell DW, Weaver CM, Jennings JL, McAfee KJ, He Y, Weil PA, Link AJ (2004) Cluster Analysis of Mass Spectrometry Data Reveals a Novel Component of SAGA. Mol Cell Biol **24**: 7249–7259

Robinson NL, Hewitt JD, Bennett AB, Damon SUE, Hewitt JD, Nieder M, Bennett AB, Damon SUE (1988) Sink Metabolism in Tomato Fruit1. Plant Physiol **87**: 727–730

Rohart F, Gautier B, Singh A, Cao K-A Le (2017) mixOmics: an R package for 'omics feature selection and multiple data integration. bioRxiv 108597

Rohrmann J, Tohge T, Alba R, Osorio S, Caldana C, McQuinn R, Arvidsson S, van der Merwe MJ, Riaño-Pachón DM, Mueller-Roeber B, et al (2011) Combined transcription factor profiling, microarray analysis and metabolite profiling reveals the transcriptional control of metabolic shifts occurring during tomato fruit development. Plant J **68**: 999–1013

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, et al (2004) Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. Mol Cell Proteomics **3**: 1154–1169

Sánchez G, Venegas-Calerón M, Salas JJ, Monforte A, Badenes ML, Granell A (2013a) An integrative "omics" approach identifies new candidate genes to impact aroma volatiles in peach fruit. BMC Genomics **14**: 343

Sánchez G, Venegas-Calerón M, Salas JJ, Monforte A, Badenes ML, Granell A (2013b) An integrative "omics" approach identifies new candidate genes to impact aroma volatiles in peach fruit. BMC Genomics **14**: 343

Sato S, Tabata S, Hirakawa H, Asamizu E, Shirasawa K, Isobe S, Kaneko T, Nakamura Y, Shibata D, Aoki K, et al (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635–641

Savoi S, Wong DCJ, Degu A, Herrera JC, Bucchetti B, Peterlunger E, Fait A, Mattivi F, Castellarin SD (2017) Multi-Omics and Integrated Network Analyses Reveal New Insights into the Systems Relationships between Metabolites, Structural Genes, and Transcriptional Regulators in Developing Grape Berries (Vitis vinifera L.) Exposed to Water Deficit. Front Plant Sci **8**: 1124

Scagliarini S, Trost P, Pupillo P (1998) The non-regulatory isoform of NAD(P)-glyceraldehyde-3-phosphate dehydrogenase from spinach chloroplasts. J Exp Bot **49**: 1307–1315

Schaff JE, Mbeunkui F, Blackburn K, Bird DM, Goshe MB (2008) SILIP: A novel stable isotope labeling method for in planta quantitative proteomic analysis. Plant J **56**: 840–854

Schenkluhn L, Hohnjec N, Niehaus K, Schmitz U, Colditz F (2010) Differential gel electrophoresis (DIGE) to quantitatively monitor early symbiosis- and pathogenesis-induced changes of the Medicago truncatula root proteome. J Proteomics **73**: 753–768

Schmidt A, Kellermann J, Lottspeich F (2005) A novel strategy for quantitative proteomics using isotope-coded protein labels. Proteomics **5**: 4–15

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. Nature **473**: 337–342

Sergeant K, Spieß N, Renaut J, Wilhelm E, Hausman JF (2011) One dry summer: A leaf proteome study on the response of oak to drought exposure. J Proteomics **74**: 1385–1395

Silva JC, Gorenstein M V., Li G-Z, Vissers JPC, Geromanos SJ (2006) Absolute Quantification of Proteins by LCMS ^{E}. Mol Cell Proteomics **5**: 144–156

Smith CR, Knowles VL, Plaxton WC (2000) Purification and characterization of phosphoenolpyruvate carboxylase from Brassica napus (rapeseed) suspension cell cultures and the integration of glycolysis with nitrogen assimilation. Eur J Biochem **267**: 4477–4485

Sonnewald AU, Quick WP, Macrae E, Krause K, Stitt M, Sonnewald U, Quick WP, Macrae

E, Krause K, Stitt M (1992) Planta Purification , cloning and expression. **189**: 174–181

Sowokinos JR, Spychalla JP, Desborough SL (1993) Pyrophosphorylase in Solanum tuberosum. Plant Physiol **101**: 1073–1080

Spannagl M, Nussbaumer T, Bader KC, Martis MM, Seidel M, Kugler KG, Gundlach H, Mayer KFX (2016) PGSB plantsDB: Updates to the database framework for comparative plant genome research. Nucleic Acids Res **44**: D1141–D1147

Steinhauser MC, Steinhauser D, Koehl K, Carrari F, Gibon Y, Fernie AR, Stitt M (2010) Enzyme Activity Profiles during Fruit Development in Tomato Cultivars and Solanum pennellii. Plant Physiol **153**: 80–98

Suzuki M, Takahashi S, Kondo T, Dohra H, Ito Y, Kiriiwa Y, Hayashi M, Kamiya S, Kato M, Fujiwara M, et al (2015) Plastid Proteomic Analysis in Tomato Fruit Development. PLoS One **10**: e0137266

Szymanski J, Levin Y, Savidor A, Breitel D, Chappell-Maor L, Heinig U, Töpfer N, Aharoni A (2017) Label-free deep shotgun proteomics reveals protein dynamics during tomato fruit tissues development. Plant J **90**: 396–417

Takamiya S, Fukui T (1978) Purification and multiple forms of phosphoglucomutase from potato tubers. Plant Cell Physiol **19**: 759–768

Takayama M, Ezura H (2015) How and why does tomato accumulate a large amount of GABA in the fruit? Front Plant Sci. doi: 10.3389/fpls.2015.00612

Tang X, Ruffner H-P, Scholes JD, Rolfe SA (1996) Purification and characterisation of soluble invertases from leaves of Arabidopsis thaliana. Planta **198**: 17–23

Tchourine K, Poultney CS, Wang L, Silva GM, Manohar S, Mueller CL, Bonneau R, Vogel C (2014) One third of dynamic protein expression profiles can be predicted by a simple rate equation. Mol Biosyst **10**: 2850–62

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M (2004) mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J **37**: 914–939

Thorniley MS, Dalziel K (1988) NADP-linked malic. **254**: 229–233

Tohge T, Alseekh S, Fernie AR (2014) On the regulation and function of secondary metabolism during fruit development and ripening. J Exp Bot **65**: 4599–4611

Toubiana D, Fernie AR, Nikoloski Z, Fait A (2013) Network analysis: tackling complex data to study plant metabolism. Trends Biotechnol **31**: 29–36

Tripodi K, Podesta FE (1997) Purification and Structural and Kinetic Characterization of the Pyrophosphate:Fructose-6-Phosphate 1-Phosphotransferase from the Crassulacean Acid Metabolism Plant, Pineapple. Plant Physiol **113**: 779–786

Turano FJ, Wilson BJ, Matthews BF (1990) Purification and Characterization of Aspartate Aminotransferase Isoenzymes from Carrot Suspension Cultures. 587–594

Unger C, Hofsteenge J, Sturm A (1992) Purification and characterization of a soluble beta-fructofuranosidase from Daucus carota. Eur J Biochem **204**: 915–21

Unger EA, Vasconcelos AC (1989) Purification and Characterization of Mitochondrial Citrate Synthase. Plant Physiol **89**: 719–723

Usadel B, Nagel A, Steinhauser D, Gibon Y, Bläsing OE, Redestig H, Sreenivasulu N, Krall L, Hannah MA, Poree F, et al (2006) PageMan: an interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. BMC Bioinformatics **7**: 535

USADEL B, OBAYASHI T, MUTWIL M, GIORGI FM, BASSEL GW, TANIMOTO M, CHOW A, STEINHAUSER D, PERSSON S, PROVART NJ (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. Plant Cell Environ **32**: 1633–1651

Usadel B, Poree F, Nagel A, Lohse M, Czedik-Eysenberg A, Stitt M (2009) A guide to using MapMan to visualize and compare Omics data in plants: A case study in the crop species, Maize. Plant, Cell Environ **32**: 1211–1229

Valot B, Langella O, Nano E, Zivy M (2011) MassChroQ: A versatile tool for mass spectrometry quantification. Proteomics **11**: 3572–3577

Voxeur A, Gilbert L, Rihouey C, Driouich A, Rothan C, Baldet P, Lerouge P (2011) Silencing

of the GDP-d-mannose 3,5-Epimerase Affects the Structure and Cross-linking of the Pectic Polysaccharide Rhamnogalacturonan II and Plant Growth in Tomato. J Biol Chem **286**: 8014–8020

Webb-Robertson BJM, McCue LA, Waters KM, Matzke MM, Jacobs JM, Metz TO, Varnum SM, Pounds JG (2010) Combined statistical analyses of peptide intensities and peptide occurrences improves identification of significant peptides from MS-based proteomics data. J Proteome Res **9**: 5748–5756

Wiśniewski JR, Hein MY, Cox J, Mann M (2014) A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. Mol Cell Proteomics **13**: 3497–506

Wong DCJ, Matus JT (2017) Constructing Integrated Networks for Identifying New Secondary Metabolic Pathway Regulators in Grapevine: Recent Applications and Future Opportunities. Front Plant Sci **8**: 505

Wu H, Jia H, Ma X, Wang S, Yao Q, Xu W, Zhou Y, Gao Z, Zhan R (2014) Transcriptome and proteomic analysis of mango (Mangifera indica Linn) fruits. J Proteomics **105**: 19–30

Xianyin Lai, Lianshui Wang, Haixu Tang and FAW (2011) A Novel Alignment Method and Multiple Filters for Exclusion of Unqualified Peptides To Enhance Label-Free Quantification Using Peptide Intensity in LC—MS/MS. J Proteome Res **10**: 759–785

Ye J, Hu T, Yang C, Li H, Yang M, Ijaz R, Ye Z, Zhang Y (2015) Transcriptome profiling of tomato fruit development reveals transcription factors associated with ascorbic acid, carotenoid and flavonoid biosynthesis. PLoS One **10**: 1–25

Yin YG, Kobayashi Y, Sanuki A, Kondo S, Fukuda N, Ezura H, Sugaya S, Matsukura C (2010) Salinity induces carbohydrate accumulation and sugar-regulated starch biosynthetic genes in tomato (Solanum lycopersicum L. cv. "Micro-Tom") fruits in an ABA-and osmotic stress-independent manner. J Exp Bot **61**: 563–574

Zhang X-Y, Wang X-L, Wang X-F, Xia G-H, Pan Q-H, Fan R-C, Wu F-Q, Yu X-C, Zhang D-P (2004) A shift of phloem unloading from symplasmic to apoplasmic pathway is involved in developmental onset of ripening in grape berry. PLANT Physiol **136**: 2475–2482

Zhao Y-B, Krishnan J (2014) mRNA translation and protein synthesis: an analysis of different

modelling methodologies and a new PBN based approach. BMC Syst Biol **8**: 25

Zhou R, Cheng L (2008) Competitive inhibition of phosphoglucose isomerase of apple leaves by sorbitol 6-phosphate. J Plant Physiol **165**: 903–910

Zschoche WC, Ting P (1973) Purification and Properties from of Microbody Malate Dehydrogenase from Spinacia oleracea Leaf Tissue. Arch Biochem Biophys **159**: 767–776

Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP (2006) Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces c erevisiae*. J Proteome Res **5**: 2339–2347

Annexes

Primary metabolism into perspective for better fruits

Thank you for agreeing to review this paper for *Annals of Botany*.

We are aiming to be among the very top plant science journals, which currently means an Impact Factor greater than 4.5. We receive over 1000 submissions every year and we only have room to publish a limited number of these.

We therefore need to be very selective in deciding which papers we can publish, so in making your assessment please consider the following points.

- **We want to publish papers where our reviewers are enthusiastic about the science: is this a paper that you would keep for reference, or pass on to your colleagues?**
  If the answer is "no" then please enter a low priority score when you submit your report.

- **We want to publish papers with novel and original content that move the subject forward, not ones that report incremental advances or findings that are already well known in other species.**
  Please consider this when you enter a score for originality when you submit your report.

Notes on categories of papers

Research papers should demonstrate an important advance in the subject area, and the results should be clearly presented, novel and supported by appropriate experimental approaches. The Introduction should clearly set the context for the work and the Discussion should demonstrate the importance of the results within that context. Concise speculation, models and hypotheses are encouraged, but must be informed by the results and by the authors' expert knowledge of the subject.

Reviews should place the subject in context, include the most up-to-date references available and add significantly to previous reviews in the topic. An idea review will move forward research in the topic.

Research in Context should combine a review/overview of a subject area with original research that moves the topic forward; i.e. it is a hybrid of review/research papers.

Viewpoints should present clear, concise and logical arguments supporting the authors' opinions, and in doing so help stimulate discussions within the topic.

Special Issue/Highlight papers should be judged by the same standards as other papers in terms of the strength of the work they contain. They are allowed a more narrow focus within the topic of the issue in which they will appear. Special Issue papers should still make the topic of interest to a wide audience.

Primary metabolism into perspective for better fruits

Bertrand Beauvoit[1], Isma Belouah[1], Nadia Bertin[2], Coffi Belmys Cakpo[2], Sophie Colombié[1], Zhanwu Dai[3], Hélène Gautier[2], Michel Génard[2], Annick Moing[1], Léa Roch[1], Gilles Vercambre[2], Yves Gibon[1*]

[1]UMR 1332 BFP, INRA, Univ Bordeaux, F33883 Villenave d'Ornon, France

[2]UR 1115 PSH, INRA, F84914 Avignon Cedex 9, France.

[3]UMR 1287 EGFV, INRA, Univ Bordeaux, Bordeaux Sci Agro, F33883 Villenave d'Ornon, France

[*]For correspondence. E-mail yves.gibon@inra.fr

Abstract

One of the key goals of fruit biology is to understand the factors that influence fruit growth and quality, ultimately with a view to manipulating these levels for improvement of fruit traits. Primary metabolism, which is not only essential for growth, but also a major component of fruit quality, is an obvious target for improvement. However, metabolism is a moving target that undergoes dramatic changes throughout fruit growth and ripening. Agricultural practice and breeding have been successfully used to improve fruit metabolic traits, but both face the complexity of the interplay between development, metabolism and environment. Thus, more fundamental knowledge is needed to identify further strategies for the manipulation of fruit metabolism. Nearly two decades of post genomics approaches integrating transcriptomics, proteomics and/or metabolomics have generated considerable information about the behaviour of fruit metabolic networks. Today, the emergence of an ensemble of modelling tools is giving the opportunity to turn this information into mechanistic understanding of fruits, and ultimately to design better fruits. Also, because gathering high quality data represent a key step for modelling, a range of must-have parameters and variables is proposed.

Context

Fruits are a huge success in the evolution of plants. Within 150 million years, the organ of angiosperms dedicated to seed dissemination has been declined in a myriad of forms, tastes and properties, sometimes to protect the seeds by becoming impregnable or toxic, sometimes to help their spread by becoming winged, floatable, explosive or even desirable. Man has long enjoyed this profusion, first as a consumer, then as a farmer and eventually as a breeder. Today, fruit production, which is essential in human nutrition, is under significant pressure from environmental stresses but also by changes in consumer demand for taste and nutritional value, resulting in a constantly renewed need for improved varieties meeting this demand. Yields are presently reaching a plateau in an increasing number of crops including fruit crops, indicating that new breeding strategies are urgently

needed (Raines, 2010; Rossi et al., 2015). A further problem is that major breeding companies, who have the capacity to experiment new strategies, restrict their investments to leading crops for economic reasons (Stamp and Visser, 2012). Who will take care of the vast majority of other ones? The possibility to come up with unified strategies for improvement therefore represents a good opportunity for both major and minor crops.

Metabolism is an obvious target for unified strategies, especially in fleshy fruits, our main source of vitamins and antioxidants, and understanding the mechanisms linking it to fruit phenotypes will help to focus breeding strategies (Giovannoni 2006). Indeed, traits such as pathogen and abiotic stress resistance during growth, as well as flavour, nutritional value and health benefits are all affected by the composition of metabolites in fruit tissues. One key goal is therefore to understand the factors that affect metabolite concentrations in cells and tissues and how they are balanced with growth, ultimately for manipulating these levels for the improvement of crop traits. Metabolism can be subdivided into primary and specialised metabolism, depending on absolute requirement for cell survival and growth. Importantly, reactions involved in primary metabolism are highly conserved whereas those involved in specialised pathways show much higher diversity between fruit species. It is nevertheless striking that a large part of fruit diversity involves primary metabolism.

The aim of the present review is to focus on primary metabolism, its contribution to fruit growth and quality, and how to influence it to improve quality and biomass production in fruits. After a brief description of fruit primary metabolism and its reprogramming throughout fruit growth and ripening, we will discuss the different approaches that have been taken to manipulate fruit metabolism: agricultural practice, breeding, and the search for metabolic targets. We will emphasise the modelling of fruit development and metabolism, as an ensemble of emerging tools that could be used in any species, lead to a better understanding of fruits and ultimately to better fruits.

Fruit primary metabolism

From a topological point of view, primary metabolism (Figure 1) is not very different between organs, stages of development or cell types and as mentioned above it is highly conserved between species. It is the way it operates that makes the difference. We will focus on pathways that are particularly important for both the growth and quality of most fleshy fruits.

Central carbon metabolism, which in fruits involves the pathways of sucrose, starch, major organic acids and respiration, provides energy and biosynthetic precursors to support fruit growth and maturation. In most species, the major source of carbon for the fruit is sucrose, which is imported from leaves via the phloem. In some species, carbon traffic is enhanced by the transport of additional sugars, such as stachyose and raffinose in Cucurbitaceae (Haritatos et al., 1996) or sorbitol in Rosaceae (Noiraud et al., 2001). It is also worth mentioning that most developing fleshy fruits are

photosynthetic, but it is now admitted that they are not self-sufficient regarding carbon supply (Lytovchenko et al., 2011). Central carbon metabolism is essential for fruit quality. Indeed, sugars and organic acids, which are among the major components of most fruits, have a strong influence on fruit taste. For example, sugars represent about 8% of the fruit fresh matter weight at maturity in peach (Desnoues et al., 2014) and 15% in grapevine (Davies and Robinson, 1996). Organic acids, especially citrate and malate, represent further large metabolic pools with citrate reaching 5% of the fresh pulp in lemon (Albertini et al., 2006). The ratio between sugars and acids is also very important for taste. It is remarkable that lemon (Albertini et al., 2006) or tomato fruits (Causse et al., 2004) do not taste sweet although they both have a relatively high sugar content of about 4%. In most fruits, taste development occurring at ripening is due to increased sweetness, which is the result of a range of dramatic metabolic adjustments (Bonghi and Manganaris, 2012). Among those, the degradation of starch occurring at the beginning of ripening is often mentioned as being a major source of sugars (e.g., Jourda et al., 2017; Hill and Ap Rees, 1994). Starch, which in many species accumulates at high levels during fruit development, is also thought to make a major contribution to the respiration climacteric (Colombié et al., 2017).

Amino acid metabolism provides precursors for protein synthesis but also for a range of specialised metabolites (Gonda et al., 2010). Major amino acids and their derivatives can have an important influence on fruit taste and quality. For example in tomato, the accumulation of large amounts of glutamate and aspartate during ripening determines the so-called umami taste, whereas GABA, which also accumulates at relatively high levels in growing tomato fruits, may provide interesting nutritional properties (Takayama and Ezura, 2015). Although nitrate and ammonium can be found in fruits (Sanchez et al., 2017; Horchani et al., 2008), it is generally considered that fruits do not assimilate nitrogen themselves but import amino acids from the phloem and to a lesser extent the xylem (Gourieroux et al., 2016). Similarly to the import of sugars, amino acids can take both the symplastic and apoplastic routes (Zhang et al., 2015).

Primary cell wall metabolism also belongs to primary metabolism if we consider that plant cells cannot grow or even survive without a wall in nature. Cell wall composition is highly diverse among plant species, but the major components (cellulose, three matrix glycans composed of neutral sugars, three pectins rich in D-galacturonic acid) are usually the same (Brummell and Harpster, 2001). Cell walls are particularly important in fruits: during growth they play a major role in shaping and protecting the fruit, and imply a finely tuned trade-off with sugar metabolism while ripening is characterised by cell wall softening, a process with strong implications for fruit quality but also for shelf-life (Brummell and Hapster, 2001). Additionally, partial cell wall degradation at ripening represents a massive release of carbohydrates into central metabolism, thus providing energy and building blocks for a range of processes (e.g. protein synthesis and sugar accumulation) and is likely to make a substantial contribution to the respiration burst in climacteric fruits (Colombié et al., 2017).

Redox metabolism, especially ascorbate metabolism, also connected to cell wall metabolism (Voxeur et al. 2011), represents a further important aspect of fruit metabolism. Fruits are considered to be our major source of antioxidants but domestication tended to reduce their concentrations, suggesting that there is a trade-off with growth, and thus productivity (Gest et al., 2013). Thus in cultivated kiwifruit, ascorbate has been found to be down to 20 times lower than in wild relatives. Lower ascorbate content is also thought to have implications for stress resistance in fruits (Gest et al., 2013) and the inability to recycle ascorbate is lethal at high metabolic activity (Eastmond, 2007; Gallie, 2013). Strikingly, the induction of blossom end rot, a necrosis usually attributed to calcium deficiency which can cause up to 50% losses in tomato production, has been attributed to an alteration of the recycling of glutathione (Mestre et al., 2012). A further interesting crosstalk exists between the biosynthesis pathways of ascorbate and primary cell wall, which share GDP-D-mannose epimerase (Mounet-Gilbert et al., 2016). Finally, tartrate, which is a degradation product of ascorbate, is a major organic acid in several fruits including citrus (Albertini et al., 2006) and grape berries where it plays a major role in winemaking (de Bolt, 2006).

To summarise, primary metabolism involves pathways that are mostly common to all fruits from a topological point of view, but flux distributions, levels of intermediates and products as well as the contribution to growth and further sinks (e.g., specialised metabolites) show a huge diversity among fruits.

Metabolism undergoes profound reprogramming throughout fruit development

The development of fleshy fruits is characterised by 3 partly overlapping phases: cell division, cell expansion and maturation, which each time involve a profound reprogramming of metabolism (Figure 2).

The involvement of hormones in fruit growth and ripening has been known for a long time and hormonal treatments are common in fruit production (Ginzberg and Stern, 2016). Briefly, cytokinins (reviewed in Jameson and Song, 2016), auxins (reviewed in Pattison et al., 2014) and gibberellins (Serrani et al., 2007) are involved in the early events following pollination. Cytokinin levels are high in ovaries and are believed to promote auxin synthesis whereas pollination results in increased levels of gibberellins (Olimpieri et al., 1999). Auxins and gibberellins promote cell division and/or cell expansion (Pattison et al., 2014) and there is accumulating evidence that they are able to induce parthenocarpy (Ding et al., 2013; Shinozaki et al., 2015). It is thought that in very young fruits, auxins are mainly produced by the seeds and that seed number is correlated with fruit size, which implies that cell number represents a major parameter regarding fruit size (Frary et al., 2000). Noteworthy, brassinosteroids have also been found to be involved in early fruit growth (Fu et al., 2008). Fruit ripening occurs after growth stops. Ethylene but also abscisic acid (Leng et al., 2014) are considered as

major factors controlling fruit ripening. Whereas the role of abscisic acid in ripening remains poorly known (Jia et al., 2016) the role of ethylene is getting well known (Giovannoni, 2004; Giovannoni et al., 2017). Climacteric fruits (e.g., tomato, banana, mango…) show a respiratory peak and a concomitant rise in ethylene, which initiates a range of ripening processes. In non-climacteric fruits (e.g. strawberry, grape, citrus…), there is no respiratory peak and ethylene remains relatively low. Interestingly, transcription factors acting upstream of ethylene signalling have been found in both climacteric and non-climacteric fruits (Giovannoni et al., 2017). However, the nature of the prime signals initiating ripening remains mysterious. Is the completion of fruit or seed growth sensed or does decreased sink demand lead to metabolic signals? Whereas a number of results indicate that hormones can trigger metabolic changes, there is also emerging evidence that metabolic signals are involved in the control of fruit development and ripening. Thus, a link between sucrose metabolism, ethylene biosynthesis and ripening has recently been found in tomato (Qin et al., 2016). It is important to note that the mode of action of hormones varies between species. Thus, hormones interact with a range of transcription factors, which leads to many possible combinations regarding the coordination of gene expression.

In tomato, several transcriptomic studies indicate that pollination, for a large part via gibberellins, has a strong effect on gene expression, including genes involved in primary metabolism (Ruiu et al., 2015). Although this suggests that major changes occur at the level of the cellular machinery, the proteome and the metabolome have hardly been investigated in ovaries and very young fruits. Then, in young tomato fruits the capacities of enzymes involved in energy metabolism (i.e. enzymes involved in glycolysis and TCA cycle), including enzymes catalysing irreversible reactions (fructokinase, glucokinase and pyruvate kinase) have been found to be very high whereas later on, during cell expansion, anaplerotic enzymes are becoming more abundant (Biais et al., 2014). At ripening, the capacities of a number of enzymes involved in energy metabolism are rising again, suggesting an increased demand in energy to support the dramatic changes occurring at that stage. Strikingly, these changes in enzymes are mirrored by strong variations in the content of numerous metabolites (Carrari et al., 2006). Furthermore, an integrative study combining transcriptomics, proteomics and metabolomics conducted with mutants impacted in the production or the sensing of ethylene has shown that a range of metabolic events are mediated by ethylene (Osorio et al., 2011). Although most integrative studies have been conducted in tomato, which is considered as the model system for fleshy fruits, it will be important to consider further fruit species. Indeed, the profiles of proteins, enzyme capacities and/or metabolites have been found to behave differently throughout fruit development in grape berries (Hawker, 1969), kiwifruit (Nardozza et al., 2013), peach (Desnoues et al., 2014) and apple (Li et al., 2016a), thus reinforcing the idea that changing enzyme capacities and properties would affect metabolite concentrations and fluxes.

The way metabolites, exported from source leaves, enter fruits represents an important point of control. Based on few reports (Ruan and Patrick, 1995; Zhang et al., 2006) it is thought that in most cases sugar import is mainly symplastic at initial stages of fruit development and becomes mainly apoplastic at later stages. The signification of such shift could be due to the fact that breaking the symplastic continuum enables the accumulation of metabolites at very high concentrations inside the fruit (e.g. molar sugar concentrations in grape berries), as the apoplastic transport does not require a favourable water potential difference between fruit and phloem (Patrick, 1997). In contrast, symplastic transport could be associated with a strong requirement in terms of incoming carbon flux. Thus, the carbon demand of tomato young fruits is the highest on a fresh weight basis (Colombié et al., 2015), which corroborates the massive abortion of young fruits when carbon supply drops (Jean and Lapointe, 2001). A further striking point is that the flux capacity of the petiole and pedicel (expressed as the proportion of phloem vessels) has been found to be correlated to fruit growth rates and size (Savage et al., 2015).

What strategies to manipulate fruit metabolism

Crop management

Changes in agricultural practices have been mostly driven by their potential to increase yield or reduce pest attacks, and it is only recently that the idea of using agronomic levers has emerged to manipulate fruit composition, especially the levels of antioxidant metabolites (Poiroux-Gonord et al., 2010). The composition of ripe fruits in soluble sugars, acids, phenolic compounds, vitamins and carotenoids have been assessed under varying crop management, for instance in response to water deficit or salinity stress (Ripoll et al., 2016), partial root-drying (Zegbe et al., 2006), temperature (Gautier et al., 2008), light intensity (Biais et al., 2014), fertilizers (Bénard et al., 2009) or grafting (Rouphael et al., 2010). Effects resulted either from dilution/concentration due to changes in the fruit water content, from changes in carbohydrate supply to the fruit, or from modifications of fruit primary and specialised metabolisms.

Under high salinity or moderate water deficit, fruit size is inversely related to treatment intensity while the fruit contents in dry matter, soluble sugars and organic acids increase in a range which depends on genotypes (Ripoll et al. 2014; 2016). In tomato, fruit hexose content also increases in response to high temperature and light intensity, but interactions between environmental conditions and plant source:sink ratio or genotype have been reported (Gautier et al., 2005; Truffault et al. 2015).The effects of crop management on fruit acidity are more confused in literature. For instance water deficit tends to increase the sugar:acid ratio although the response is genotype-dependent (Ripoll et al., 2016). Several reports show that the fruit metabolite composition depends on metabolic fluxes

and enzyme activities (Beckles et al., 2012), which unfortunately have been seldom investigated in response to crop management. During tomato fruit development under control, shaded or water limited conditions, it has been found that metabolite levels are more sensitive to the environment than enzyme capacities (Biais et al., 2014). Conversely, it has been suggested that under water deficit an increase in the activity of the apoplastic invertase facilitates sugar import into fruits (Osorio et al., 2014).

Concerning antioxidants, ascorbate is generally accumulated at higher levels at relatively low temperatures during the growth period, in contrast to carotenoids which decrease (Gautier et al., 2008). Light also strongly affects the biosynthesis of antioxidants. Thus, ascorbate accumulation strongly depends on the fruit irradiance itself, which may be increased by leaf pruning (Massot et al., 2010). It was recently shown that light and temperature interact to regulate the ascorbate pool size in relation with biosynthesis gene expression and ascorbate oxidation and recycling (Massot et al. 2013). This likely explains large seasonal variations in fruit ascorbate content (Massot et al., 2010). Carotenoid accumulation is also positively regulated by light exposure (Fanciullino et al., 2014; Truffault et al., 2015) or by an increase in the red to far-red ratio (Alba et al., 2000). Regarding the effects of water and mineral supply, high salinity has a globally positive effect on the accumulation of ascorbate, lycopene and beta-carotene (Frary et al. 2010), with strong genotype by environment interactions (Gautier et al., 2009). Under nitrogen depletion ascorbate slightly increases, possibly because more light reaches the fruits. Many studies report positive effects of water deficit on ascorbate. However the potential benefits of drought on fleshy fruit quality might be exacerbated or mitigated depending on genotype, seasonal factors or on intensity and duration of treatment (Ripoll et al., 2014). Crop management and in particular water deficit or high salinity may influence fruit metabolism first, through an effect on net photosynthesis and supply of precursors for biosynthesis, second through an oxidative stress signalling, which may trigger some biosynthetic pathways. In tomato, there is much evidence that the synthesis of carotenoids and ascorbate is linked to oxidative stress (Poiroux-Gonord et al., 2010). On the contrary carbohydrate availability does not limit the synthesis and accumulation of ascorbate in fruits (Poiroux-Gonord et al., 2013).

The manipulation of plant fruit load via flower, fruit, leaf and/or or shoot pruning, which is often used to regulate or increase fruit size, may induce a parallel increase in the content of individual metabolites expressed on a fresh weight basis (Kromdijk et al., 2014). However, several exceptions have also been reported (e.g., Massot et al. 2010; Fanwoua et al. 2012). In tomato most of the water enters into the fruit via the phloem, together with assimilates, which explains that sugar and acid content hardly increase at low plant fruit load (Ho, 1996). In contrast carotenoid and ascorbate contents can be significantly altered by fruit load and carbon availability (e.g., Gautier et al. 2005; Massot et al. 2010; Poiroux-Gonord et al. 2013).

We have seen that factors such as salinity, water stress, high light intensity, heat and sub- or supra-mineral nutrition can have positive impacts on fruit growth and/or quality. However, they can

also result in oxidative stress, and ultimately cell death. Blossom end rot is a necrosis appearing at the blossom end of the fruit (in tomato, pepper, apple…). Although usually attributed to calcium deficiency, it may rather result from complex interactions between environmental factors and involve secondary oxidative stress (Saure, 2014). The fact that solutions found so far to prevent the appearance of such disease are largely empirical indicates that more mechanistic studies integrating metabolism and growth conditions are needed.

Breeding

Plant domestication has resulted in considerable phenotypic modifications from wild species to modern varieties. For instance in tomato, a study combining gene expression and population genetics in wild and crop tomato showed that domestication globally modified expression levels for hundreds of genes, acting on entire gene networks, including genes involved in carbohydrate metabolism (Sauvage et al. 2017). Breeding based on molecular markers and quantitative genetics still has a lot to offer (Grandillo and Cammareri 2016; Tomason et al. 2013, Kumar et al. 2014) and is moving to genomics-assisted breeding (Kinkade et al. 2013). Genetic diversity, the motor of breeding, continues to be searched in wild relative or ancestral varieties as done for decades for tomato (Knapp and Peralta, 2016) or melon (Burger et al. 2006). Diversity of genetic resources including natural mutants has been shared for tomato, for instance through the Charles M. Rick Tomato Genetic Resource Centre (http://tgrc.ucdavis.edu/, Rick 1986). This Centre remains a central source of tomato wild species germplasm, various true-breeding populations and monogenic mutants (Giovanonni 2016). Diversity has been induced by EMS mutagenesis on Targeting Induced Local Lesions In Genomes (TILLING) platforms (Okabe and Ariizumi 2016). Such collections can also be used in forward genetics approaches, as a rapid identification of causal mutations in tomato EMS populations is possible using mapping-by-sequencing (Garcia et al. 2016). Furthermore, the use of TILLING for the discovery of candidate gene function is presently being replaced by genome editing techniques, which are easily applied in several fruit species (Malnoy et al. 2016).

Fruit traits of interest can easily be detected and selected, even if underlying mechanisms might be highly complex. Quantitative trait loci (QTLs) of fruit traits have been largely studied in a number of fruit species after the pioneering work by Paterson et al. (1988) for tomato soluble solid content and pH related with the content of soluble sugars and organic acids. Metabolite QTLs (mQTLs) remain largely used, together with recombinant inbred lines (RILs), and only a few recent representative examples are listed here. In melon, a map-based cloning strategy based on natural genetic variability for fruit acidity allowed identifying a gene family encoding membrane proteins responsible for acidity in fruit (Cohen et al. 2014). For example, QTLs controlling individual soluble sugars and organic acids have been mapped in tomato in relation with water deficit response (Albert et al. 2016). In peach, co-

281 locations between annotated genes, QTLs for enzyme activities and QTLs controlling major soluble
282 sugar or organic acid concentrations were observed (Desnoues et al. 2016). This dynamic QTL
283 approach revealed changing effects of alleles during fruit growth. The QTL approach has also been
284 used for the identification of loci affecting the accumulation of specialised metabolites, for example in
285 tomato (Ballester et al. 2016; Bauchet et al., 2017) or in apple (Khan et al. 2012). In melon, single-
286 gene resolution QTL mapping achieved using 81 recombinant inbred lines, genotyping conducted
287 using almost 60,000 SNPs of the flesh tissue of mature fruit, phenotyping and metabolic profiling has
288 been reported (Katzir 2015). Interestingly, a recent genetic study of sugar metabolism suggests that the
289 maximal capacity of sucrose accumulation has been reached in melon (Argyris et al., 2017).

290 Metabolite-based genome-wide association studies (mGWAS) are progressing (Luo, 2015). In
291 tomato, a core collection of 163 tomato accessions was used to map loci controlling variation in fruit
292 metabolites including amino acids, sugars, and ascorbate and the accessions were genotyped with
293 about 6000 single-nucleotide polymorphism markers (Sauvage et al. 2014). This GWA study
294 confirmed cell wall invertase as a candidate gene for the control of soluble sugar content (Fridman et
295 al. 2000), and provided a list of other candidate loci including loci underlying the genetic architecture
296 of fruit malate and citrate levels. However, it is now admitted that classical breeding will inevitably
297 reach a plateau in a given species and it has been proposed many times that new strategies involving
298 more fundamental knowledge will be needed. More recently, it also appeared that epialleles may
299 determine the content of compounds of interest in fruits (Quadrana et al. 2014). Therefore, epigenetic
300 differences may provide new targets for breeding and crop improvement (Gallusci et al. 2017).

301

302 ***A priori approaches***

303 A large body of literature shows the importance of genetic factors in the control of fruit quality,
304 and manipulating the expression and properties of pathway enzymes is an obvious approach to
305 manipulate fruit metabolism. Variations in properties of an enzyme can indeed have spectacular
306 effects on fruit phenotypes. For example, the introgression of a gene encoding regulatory subunit of
307 ADP-glucose pyrophosphorylase from *Solanum hirsutum* into cultivated tomato results in a
308 stabilisation of the activity of this enzyme during early stages of fruit growth, which supports
309 increased starch accumulation, and ultimately leads to higher soluble solids (Schaffer et al., 2000). The
310 introgression of an apoplastic invertase with a higher affinity for sucrose from *Solanum pennellii* also
311 leads to higher soluble solids, probably by increasing sink strength (Fridman et al., 2004).

312 Topological knowledge of metabolism has motivated a range of a priori approaches, in which
313 given enzymes were targeted with the hope of improving fruits. However, there are many examples
314 indicating that manipulating enzymes does not necessary lead to improvements of both fruit biomass
315 and/or quality. Thus in tomato, the down regulation of the expression of the vacuolar acid invertase

increases sucrose but decreases hexoses and fruit growth rate and size (Klann et al., 1996); hexokinase overexpression results in lower sugar and starch, and impaired fruit growth (Menu et al., 2004); fruit specific overexpression of a bacterial pyrophosphatase leads to a significant increase in ascorbate content but also to a decrease in fruit size (Osorio et al., 2013); the manipulation of malate concentrations via down regulation of fumarase or mitochondrial malate dehydrogenase results in dramatic alterations of the metabolome, although fruit size is only marginally impaired (Centeno et al., 2011).

Subcellular compartmentation is a further important point to take into account when studying the control of metabolic fluxes and concentrations, in particular the vacuole (Beauvoit et al., 2014). Indeed, in fleshy fruits most of the cell volume is occupied by a large central vacuole, which is assumed to participate to fruit growth via its enlargement driven by the accumulation of large amounts of osmolytes such as organic acids and sugars (Ho, 1996) and thus happens to be of major importance for fruit quality. Although it is assumed that the transport of sugars and organic acids into the vacuole is active (Shiratake and Martinoia, 2007) very little is known about the properties of fruit tonoplast transporters, and *in vitro* experiments can hardly be extrapolated within the framework of metabolic changes that underlie fruit development. Recently, the overexpression of SlCAT9, a tonoplastic amino acid exchanger, resulted in increased levels of GABA, aspartate and glutamate paralleled by a decrease in citrate in tomato fruits (Snowden et al., 2015). Also in tomato, the down regulation of the proton-pumping ATPase has been shown to increase the sucrose-to-hexose ratio but to decrease the fruit growth rate and size (Amemiya et al., 2006).

There are many more examples indicating that the manipulation of enzymes or transporters involved in primary metabolism hardly results in fruit and/or yield improvement. Among the rare successful approaches, the manipulation of the sucrose sensing machinery led to tomato fruits with increased sweetness without affecting plant or fruit growth (Sagor et al., 2016). It is striking that the manipulation of specialised metabolism has been more successful (Lewinsohn et al., 2001; Tohge et al., 2015).

Post genomics

Post genomics, which can be defined as the shift in biology observed in the early 2000, once the first genomes had been sequenced, has brought the possibility to perform untargeted and multidisciplinary studies including transcriptomics, proteomics, metabolomics and bioinformatics. One aim was to search for “better” candidate genes by performing large-scale correlative studies identifying “suspects by association” (Usadel et al. 2009, Toubiana et al. 2013). About ten years ago, Carrari and Fernie (2006) reviewed earlier works using targeted approaches, as well as pioneering studies in which metabolic or transcriptional profiling aimed at identifying candidate genes for

modifying metabolite content. They included primary metabolites and several specialised metabolites considered as important with respect to fruit quality. We will focus here on exemplary works of the past few years.

The combination of at least two omics has contributed to the characterization of metabolic shifts during development in a range of fruit species including tomato (Osorio et al. 2011), grape berry (Dai et al. 2013), apple (Li et al. 2016a; see also www.transcrapple.com), melon (Guo et al. 2017) and mango (Wu et al. 2014). Metabolic shifts during post-harvest storage have also been characterized, for instance in litchi (Yun et al. 2016) or citrus (Ding et al. 2015). Moreover, omics approaches have been used to describe effects of the environment on fruit metabolism in tomato (D'Esposito et al. 2017) and of abiotic or biotic stresses such as water stress or botrytis infection in grape berry (Agudelo-Romero et al. 2015, Ghan et al. 2015). In addition, omics have allowed characterizing cultivars and mutants. An example of the characterization of mutants concerns a study about low citrate accumulation in orange (Guo et al. 2016). Nowadays, a crucial aim is to elucidate the major biochemical and signal transduction pathways that are active for primary (Mounet et al. 2009, Bastias et al. 2014), as well as specialised metabolism (Wong and Matus 2017), including the identification of transcription factors (Rohrmann et al. 2011, Ye et al. 2015), and their targets as done recently for tomato (Fernandez-Moreno et al. 2016) or citrus (Li et al. 2016b) fruit.

In apple, a comprehensive 2D gel-based proteomic analysis over five growth stages, from young fruit to maturity, coupled with targeted metabolomic profiling of soluble sugars, organic acids and amino acids provided insights into the metabolism and storage of fructose, sucrose and malate (Li et al. 2016a). Another output of the latter study was the hypothesis that the decrease in amino acid concentrations during fruit development was related to a reduction in substrate flux via glycolysis. In parallel with the improvement of proteomic technologies, LC-MS/MS-based shotgun proteomic studies are exploding in fruits. In citrus, integration of LC-MS/MS-based proteomic and metabolomic analyses showed that organic acid and amino acid accumulation shifted toward sugar synthesis during the later stage of citrus fruit development, and that an invertase inhibitor may be involved during maturation (Katz et al. 2011). In grape exocarp, related trends between metabolites and proteins revealed clear links between primary and specialised metabolisms (Negri et al. 2015). For instance several proteins involved in glycolysis, TCA cycle, and metabolic intermediates of these pathways showed a good association with anthocyanin content. In tomato, changes in protein abundance were measured in skin and flesh during development, including for 61 differentially expressed transcription factors (Szymanski 2017). These large-scale proteomic data were used to estimate metabolic activity by employing the LycoCyc pathway annotations, local topology of the pathways and protein expression values. This approach revealed a significant tissue-specific reprogramming of metabolism during fruit development.

The combination of three omics levels was performed in grapes in a study involving a comparison between five cultivars at maturity (Ghan et al. 2015). The omic technologies were consistent in distinguishing cultivar variation. This integration of multiple omic datasets revealed complex biochemical variation amongst the cultivars including for amino acid metabolism. Mineral elements may be inhibitors or activators of enzyme or take part in complex regulation cascades. However, integration of ionomics and metabolomics in fruit remains rare. In melon such a combination (Moing et al. 2011) enabled the identification of co-regulated hubs, including aspartic acid and 2-isopropylmalic acid besides several specialised metabolites, in metabolic association networks, and of links of primary and specialised metabolism to key mineral and volatile fruit complements. For instance in the latter study, potassium was highly correlated with pyruvic acid and copper was associated with 14 amino compounds including proline. A particular category of metabolites involved in the regulation of development and metabolism are hormones. The development of 'hormonomics' in parallel with the analysis of primary metabolites and other omics is of special interest for the study of the metabolic regulations linked with fruit set or maturation (Oikawa et al. 2015).

If so far the candidate genes approach proved to be complicated for central metabolism, it has been more successful for specialised metabolism (Tohge et al., 2015). For instance in grapes a recent study for the search for berry-specific regulators of the phenylpropanoid pathway (Wong and Matus 2017) used overlaying maps of co-expression between structural and transcription factor genes, integrated with the presence of promoter cis-binding elements, microRNAs, and long non-coding RNAs. This strategy revealed new uncharacterized transcription factors and several microRNAs potentially regulating different steps of the phenylpropanoid pathway, and one particular long non-coding RNAs was shown to compromise the expression of nine stilbene synthase genes. In peach a combination of volatile compound and gene expression analysis revealed a set of genes that are highly associated with fruit volatiles, which could prove useful in breeding or for biotechnological purposes. As a proof of concept, one peach fruit candidate gene was cloned and expressed in yeast to show that it may be involved in the production of a precursor of lactones/esters (Sanchez et al. 2013).

After less than two decades in the era of post-genomics it is probably too early to conclude about their contribution to the improvement of the performance of fruit crops. However, they have exposed the complexity of metabolic networks. Factors limiting the accumulation of metabolites in fruits have recently been reviewed, revealing that the constraints shaping the responses of metabolic systems to manipulation are mass conservation, cellular resource allocation and, most prominently, energy supply, particularly in heterotrophic tissues (Morandini 2013). Modelling represents a promising way to link such factors with the complexity of metabolism.

Towards fruit integrative modelling

Life sciences have reached a point where many aspects of the genotype-phenotype relationship can be quantified and used to construct mechanistic models of metabolism that allow for meaningful biological predictions (Bordbar et al., 2014). We will discuss three types of models that have been adopted in fruit research: enzyme-based (i.e. kinetic), reaction-based (i.e. stoichiometric) and process-based (i.e. biophysical) models, which may prove highly complementary and enable us to cope with the complexity of fruit metabolism.

Kinetic modelling

It has been frequently assumed that certain enzymes are rate limiting (Krebs, 1957), a concept that has been challenged in the light of results from metabolic control analysis (Kacser and Burns, 1973; Heinrich and Rapoport, 1974). Briefly, it is now accepted that the control of a metabolic flux is distributed between the different steps in the relevant pathway and that this distribution can vary with the physiological state. One consequence of this is that it is almost impossible to predict the effect of an alteration of a given activity on the flux and metabolite concentrations of the corresponding pathway without implementing a kinetic model (Morandini, 2009). An enzyme-based kinetic model consists in sets of ordinary differential equations (ODEs) describing reactions of a metabolic network. When the reactions are adequately parameterised, ideally with experimental data, the computation of fluxes and concentrations becomes possible, as well as the estimation of so-called control coefficients for enzymes, which may allow the identification of candidate enzymes that could be manipulated to modify metabolism in a desired manner (Rohwer, 2012). High quality experimental data about enzymes and metabolites are critical for building kinetic models, but they have usually been hardly available to modelling projects, mainly due to technical and organisational limitations (Kettner, 2007). Although such models were already used more than 60 years ago to describe biochemical processes, the number of validated and available kinetic models remains astonishingly low, especially in plants (Rohwer, 2012; see also http://jjj.mib.ac.uk/ and http://www.ebi.ac.uk/biomodels-main/) and despite their great potential for discovery. Thus, a model describing sucrose metabolism in sugarcane stems has revealed that fructose and glucose uptake, vacuolar sucrose import and cytosolic neutral invertase are the most critical steps in determining the rate of sucrose accumulation (Rohwer and Botha, 2001; Uys et al., 2007). Then, the importance of neutral invertase as exerting a strong control over the hexose-to-sucrose ratio has been demonstrated with transgenic sugarcane in which this enzyme was downregulated (Rossouw et al., 2010). Later on, the transfer of this model to the tomato fruit has been made possible by implementing the vacuole, indicating that transferring a model to another species is much more than a confirmatory procedure (Beauvoit et al., 2014).

The prerequisite to build and parameterise an enzyme-based and compartmented kinetic model is based on three kinds of knowledge: (i) the cellular reactions (i.e. network topology and enzymology), (ii) the cellular composition (i.e. biomass compounds, cofactors and accumulated metabolites) and (iii) the cell compartmentation (i.e. subcellular volume fractions) (Figure 3) and for reviews, see Schallau and Junker (2010) and Rohwer et al. (2012). In this framework, fluxes are expressed as a function of reactant concentrations and kinetic properties using enzyme kinetic rate laws, such as Michaelis-Menten or other *ad hoc* kinetics (Cornish-Bowden, 2004, Liebermeister and Klipp, 2006). Since enzyme capacities (i.e. maximal enzyme activities measured at substrate saturation) may vary during fruit development as a consequence of metabolic reprogramming (e.g. Biais et al., 2014), they must be experimentally determined. However, kinetic constants can be taken from previous literature or from experimental measurements. The set of ODEs is solved assuming that the growing fruit is at metabolic steady state, thus allowing modellers to perform a sensitivity analysis of the model which, in turn, pinpoints the most influential parameters whose values must be properly set. Ultimately, a model parameterization refinement is performed, based on the comparison of simulated and experimentally measured metabolites. Using optimisation algorithms, the least-square fit of the data provides estimates for unknown parameters that are biologically relevant (for review Tummler et al. 2010), such as the carbon input flux or tonoplastic carrier capacities throughout tomato fruit development (Beauvoit et al., 2014). Finally, independent datasets obtained for instance with transgenic lines (Beauvoit et al., 2014), can be used for validation purpose, allowing the model analysis to be established with high confidence

An important benefit of kinetic modelling is the possibility to implement the model with isoenzymes that catalyse the same reaction but display distinct kinetic properties and subcellular localization. An enzyme-based model of sucrose metabolism has been able to discriminate the functioning of the various sucrose degrading enzymes in developing tomato fruit (Beauvoit et al., 2014). For instance, sucrose cleavage was mainly sustained by acid invertase during cell division and then was relayed by neutral invertase and sucrose synthase during cell expansion. Meanwhile, the sucrose phosphate synthase activity remained at a low level. All together, these results indicated that each cleaving enzyme contributes to fruit sink strength, in contrast to previous findings, and that the sucrose synthesis-breakdown cycle was less active than previously hypothesized. Strikingly, the vacuolar sucrose carrier and acid invertase were found to exert a strong control over sugar composition, a prediction that has also been validated with data obtained from transgenic plants. Indeed, the transport of sucrose into the vacuole and its subsequent hydrolysis, drive the osmotic potential of this organelle and, in turn, are likely to control vacuole expansion during early fruit growth.

An additional layer of information provided from kinetic modelling relies on the possibility to test the physiological relevance of regulatory features that have been previously biochemically

characterized *in vitro*. For instance, retro-inhibition of acid invertase and glucokinase, on the one hand, and proton-coupling mechanism of tonoplastic carriers, on the other, have been found to be essential to accommodate the experimentally measured sugar content through tomato fruit development (Beauvoit et al., 2014).

Admittedly, the kinetic analysis is usually restricted to small and medium scale networks, not exceeding tens of reactions and transporters (Zhu et al., 2007). Pioneering approaches aimed to account for spatiotemporal specificity of sucrose metabolism, especially during the maturation of culm nodes of sugarcane in close interactions with phloem (Rohwer, 2012). However, the detailed biochemical description of the network becomes challenging when scaling up kinetic models so that the essential features captured by the model do not increase in proportion. One of the challenges in constructing realistic kinetic models is the scarcity of enzyme data (especially capacities within compartments, post-translational modifications…) and of validation data sets. A further challenge will be to integrate information from high-throughput transcriptomics, proteomics and metabolomics into mechanistic models, since such data sets are becoming more readily available for a growing number of fleshy fruits.

Stoichiometric modelling

Over the last 30 years, several hundreds of stoichiometric models, also called constraint-based models (CBM), have been published (Bordbar et al., 2014), including an increasing number of models describing plant metabolism. Reasons for such success include that stoichiometric models are amenable to the genome scale, do not necessitate massive computing resources, and overcome experimental difficulties encountered with other modelling approaches (Shi and Schwender 2016). Thus, unlike kinetic models, stoichiometric models do not require detailed knowledge about enzyme amounts and properties, which remain very difficult to measure, especially when dealing with large metabolic network. In turn, stoichiometric models do not enable predictions of metabolite concentrations, but they equally provide the possibility to predict fluxes, which is a valuable option when the use of isotopically-labelled precursors is difficult. This is of great interest in fruits, which are very difficult to label (Sweetlove and Ratcliffe, 2011).

Stoichiometric modelling is based on a metabolic network description through stoichiometric equations of reactions and on the assumption of pseudo-steady state. This network consists in coupled chemical conversions (reactions) that are mostly catalysed by enzymes. Nutrients are converted into building blocks, such as nucleotides, fatty acids, lipids, amino acids and free-energy carriers, which enable the synthesis of macromolecules such as DNA, proteins or cellulose. These macromolecules are required for the maintenance of cellular integrity and formation of new cells. In a single reaction, substrates are converted into products and the number of atoms of a given type, such as C, H, O, N and

the net charge should balance on each side of the equation. These balancing principles are followed in genome-scale metabolic reconstructions. Stoichiometric models have been widely used to estimate the metabolic flux distribution in the cell on the basis of some optimality hypothesis (flux balance analysis). Up to genome-scale metabolic networks can be converted to stoichiometric matrices, which enable constraint-based modelling when they are associated to e.g., input and/or output fluxes, minimal and/or maximal reaction rates (Bordbar et al., 2014). Once parameterised with such boundaries, these models can be used to generate a solution space for steady-state flux distributions. Then, objective functions can be used to narrow the solution space. Commonly used objective functions include flux minimisation, maximisation of biomass production per unit substrate and maximised ATP-yield. Stoichiometric models have proven very useful in biochemical industry, by enabling the optimisation of the production of high-value molecules such as vanillin in yeast (Brochado et al., 2010) or lycopene in *E. coli* (Alper et al., 2005). In plant research, stoichiometric models are still exploratory, facing challenges such as tissue- and cell metabolic specificities and subcellular compartmentation. Thus, metabolic reconstructions will necessitate a more unified way of representation to make models comparable. In particular, cofactor specificity will be needed to be carefully addressed during reconstruction steps (Pfau at al. 2016).

With a medium-scale knowledge-based stoichiometric model describing central metabolism fluxes have been determined throughout the development of the tomato fruit (Colombié et al 2015). This model has subsequently been implemented with a detailed description of the respiratory pathway including alternative oxidase and uncoupling proteins, which enabled the investigation of respiration and energy dissipation (Colombié et al. 2017). With a large metabolic dataset transformed into constraints the model has then been solved on a daily basis throughout the fruit development. It detected a peak of CO_2-release as well as an excess of energy dissipation just before the onset of ripening, which coincided with the respiration climacteric. The unbalanced carbon allocation, which resulted from the simultaneous slowdown of biomass construction on the one hand and the degradation of starch and cell wall polysaccharides on the other hand, was found to explain the excess of energy that has to be dissipated. Additionally, constraint-based modelling might appeared as a promising tool for estimating fruit respiration, which is difficult to measure on fruits still attached on the mother plant. Therefore, it will be important to confront predicted- and experimental data for respiration in fruits.

The most critical point regarding stoichiometric modelling is that flux predictions are highly dependent on the choice of the objective function used in the analysis. This function has to appropriately describe the metabolic ‘purposes’ even if cells are dedicated to several functions. While growth-based objective functions seem to be more appropriate to study individual cells in culture, flux minimization is thought to be more adequate to describe complex metabolic networks of plant cells. The principle of flux minimization (Holzhutter, 2004) based on an assumption that evolution selects

562 for cells able to fulfil vital functions (growth, DNA repair, etc.) by adjusting metabolic inputs
563 stipulates that stationary metabolic fluxes attain minimum values based on the availability of external
564 substrates (i.e., substrates of the network under study). This principle has been shown to agree with the
565 global behaviour of *in vivo* cellular systems, and yield biological flux values (Grafahrend-Belau et al.,
566 2013; Colombié et al. 2015, 2017).

567 While the 'enzyme cost', i.e. the amount of protein needed for a given metabolic flux is crucial
568 for the metabolic choices cells have to make, it has generally been ignored by constraint-based
569 metabolic models, probably because information about protein amounts and/or enzyme activities was
570 not available. A better description of the costs of protein synthesis and degradation (turnover) will be
571 needed to refine the energy (ATP) and carbon demand at the level of whole metabolism. Recently
572 Noor et al. 2016, by developing a method for computing enzyme amounts needed to support a given
573 metabolic flux at minimal protein costs, showed that the minimization of enzyme cost is a meaningful
574 optimality principle for exponentially growing *E. coli* cells. In contrast, the modelling of fruit
575 metabolism by using kinetic and stoichiometric approaches revealed the paradox that on the one hand
576 most enzyme capacities always exceeded the fluxes of the reactions they catalyse (Beauvoit et al.,
577 2014; Colombié et al., 2015), which suggests that changing capacities would have a limited effect on
578 fluxes and their distributions, and that on the other hand all enzyme capacities measured throughout
579 fruit development were found to undergo major reproducible and stage-dependent changes, suggesting
580 that the control of capacities still plays an important role during development (Biais et al., 2014).
581 Consequently, given the fact that highly conserved metabolic networks such as central and primary
582 metabolism may operate very differently between species, organs, tissues and cell-types but also
583 between growing and steady cells, or depending on the environment, stoichiometric modelling
584 provides the opportunity to compare such diversity with relative ease. Thus, flux analysis and
585 modelling of a range of plant systems has pointed the importance of the supply of metabolic inputs
586 and demand for end products as key drivers of metabolic behaviour (Sweetlove et al., 2013). Thus, in
587 fruits, the transposition of the heterotrophic model from tomato to other fruit species might prove very
588 useful to improve our understanding of the links between metabolism and fruit phenotypes such as
589 sweetness, acidity, growth rate or occurrence of a respiration climacteric. Then, such models could be
590 further developed in order to be able to describe metabolic diversity within species, by taking
591 advantage of the genetic diversity existing within species. This could for example enable the
592 identification of loci associated to fluxes (flux QTL), which could lead to the identification of genes
593 involved in flux control and ultimately in new breeding strategies. Given the fact that several species
594 comprise cultivars exhibiting climacteric or non-climacteric behaviour (Barry and Giovannoni, 2007),
595 it will also be interesting to compare flux maps obtained for climacteric or non-climacteric genotypes,
596 in order to achieve a better understanding of the physiological meaning of the respiration climacteric.

597

Process-based modelling

Fruit quality is *per se* the result of a complex chain of biological processes. Let us consider sweetness: it results from hundreds of processes involved in sugar production in the leaves, loading and translocation in the phloem, unloading in the fruit cells, metabolism in the fruit cells and dilution by the water accumulated in the fruit. The technical operations applied to the plant influence these processes in a complex way. It is clear that all the processes involved in the quality of fruits cannot be integrated in models. But some degree of complexity is needed to consider quality and the effect of agricultural practices.

Most plant simulation models were originally developed for agronomic applications (van Ittersum and Donatelli, 2003). Their success in such applications is largely due to their robust empirical description of the relationship between plant growth, environmental conditions and management practices. However with the increase of knowledge, models with more processes and less empiricism have emerged during the last 20 years. Those process-based models offer a theory describing how the components of the system causally interact with one another to produce a given outcome. Simulations can be seen as the creation of a possible world that is constructed *in silico* using computer programs to formally represent relevant aspects of the real system under investigation. Process-based models decompose plant traits into various processes subjected to environmental variations, and enable the quantification of plant responses to genetic, environmental, and management factors within a mathematical framework that allows dynamic simulation of the physical, biophysical and physiological processes, with parameters independent of the environment and characteristic of a genotype or group of genotypes.

Prediction of fruit growth and composition requires an integrated view of plant functioning, with a formalisation of interactions between resources, between organs, and between processes. Indeed, the environment and agricultural practices are affecting several processes, with many interactions between them. Such processes, which include organ emergence, growth and resource acquisition, do not have the same sensitivity to the environmental, thus resulting in large variations in source and/or sink phenotypes. As the plant is the main source of water, carbohydrates and minerals for the fruit, there is a need to link fruit growth with plant development, and to take into account various organisational levels and the way they interact (Baldazzi et al., 2012). For example, the contribution of fruit photosynthesis to the accumulation of carbohydrate in the fruit is marginal whereas the position of a given fruit on the plant has a strong effect on the inflow of water and carbohydrates (Fishman and Génard, 1998). In order to model fruit growth and its variability within the plant, some functional-structural plant-models have been developed. They explicitly describe the architecture of the plant and its functioning by formalising the processes of development, growth and acquisition of resources at the level of the organ. Such models allow the simulation of plant phenotypic plasticity with various environmental conditions (de Jong et al., 2011) and agricultural

practices (Louarn et al., 2008) and hence are useful to investigate their effects on yield and fruit composition. A functional-structural plant model linking plant and fruit growth (see the fruit growth model hereafter, Fishman and Génard, 1998) has been already developed for tomato (Baldazzi et al., 2013). Estimations of resource acquisition (photosynthesis module), transpiration (radiative balance model), carbohydrate loading and leakage along the phloem pathway and transfer within the plant enable the simulation of water and carbohydrate availability at various locations within the plant. The water flow between the plant and the fruit is driven by the water potential gradient of the xylem and the phloem, and the carbohydrate import into the fruit is related to the phloem carbohydrate concentration through active uptake and mass flow. The model is able to simulate variations in leaf photosynthesis and transpiration with plant age and season, and hence to simulate carbohydrate concentration as well as water potential and their variability within the plant. Therefore, depending on plant age at anthesis and on the fruit position on the plant, variations in fresh and dry masses can be simulated. Thus, the model showed that fruits of the first truss are smaller because the leaf area is not fully developed, inducing lower carbohydrate availability. It also showed that within a given truss the distal fruits are smaller because of the progressive decrease of water potential along the truss rachis (Baldazzi et al., 2013).

In the early 1980s, modelling fruit growth was mainly limited to the accumulation of dry matter. Even to date, there are only a few models that simulate water accumulation. Models considering (1) water uptake and transpiration per unit of fruit area as a constant (Lee, 1990) or as a variable (Génard and Huguet, 1996), (2) the driving force resulting from the difference in water potential between the stem and the fruit, and (3) the role of fruit anatomy (Bussières, 1994) have been proposed. Then, a model of fruit growth integrating both dry matter and water accumulation within the fruit has been developed (Fishman and Génard, 1998; Liu et al., 2007, de Swaef et al., 2014). This model is based on a biophysical representation of the fruit as one big cell, in which sugars are transported from the plant phloem by mass flow, diffusion and active transport. Incoming water flows are regulated, in particular, by differences in water potential, and growth is effective only when the flow balance induces a sufficient turgor pressure on the cell walls. Fruit turgor pressure depends on carbon partitioning between soluble and insoluble solids. Soluble solids such as sugars and organic acids have rarely been subjected to modelling work. However, a model for sugar accumulation (Génard and Souty, 1996) and two models for the accumulation of citrate (Lobit et al., 2003; Etienne et al.,2015) and malate (Lobit et al., 2006; Etienne et al., 2014) have been developed. The “Sugar” dynamic model represents the transformation of phloemic sugars into different sugars accumulating in the fruit pulp (mainly sucrose, glucose and fructose), a part of which is used for synthesising compounds other than sugars and for respiration. In this model, a simplified view of sugar metabolism relies on the ‘‘rate law’’ of chemical kinetics, which state that the carbon flow between two compounds is proportional to the quantity of carbon in the source compound. Thermodynamic considerations of how cells function led to infer that

variations in mitochondrial metabolism explain citric acid concentrations, whereas vacuole storage would explain variations in malic acid (Etienne et al., 2013). The citrate model is based on a simplified representation of the TCA cycle, in which pyruvate, malate and citrate are the only metabolites considered because they are at branch points between several reactions and they are exchanged between the cytosol and the mitochondria. The model is able to simulate both seasonal variations in citric acid production and degradation. The malate model assumes that malate accumulation in fleshy fruits is mainly determined by the conditions of vacuolar storage in cells. The transport of malate is passive and occurs by facilitated diffusion of the di-anion form through specific ion channels and transporters. It follows the electrochemical potential gradient of the di-anion across the tonoplast, which is mainly controlled by the di-anion malate activity across the tonoplast and the electric potential gradient across the tonoplast.

A "Virtual Fruit Model" has been proposed (Lescourret and Génard, 2005; Martre et al., 2011) that integrates the main processes involved in fruit quality development into one system. This type of model has interesting complex behaviours. For example, according to the model, the application of a water stress after a period of optimal irrigation results in a strong decrease in growth, whereas fruits grown on plants under continuous stress grow normally. This suggests that fruits can adapt to stressful situations. In real plants, this kind of adaptation has been called a memory effect (Trewavas 2004). The model also predicts that enhanced unloading of sugars into the fruit leads to an increase in the amount of water accumulated in the fruit and, consequently, to an increase in fruit size. It also predicts an increase in the concentration of sugars in the fruit. Also, an increase of water supply leads to an increase in the amount of water accumulated in the fruit and, consequently, to an increase in fruit size, but the concentration of sugars decreases. The quality traits are therefore affected differently according to the factor (C or water) considered, with either positive or negative correlations between fruit mass or sugar concentrations.

The "Virtual Fruit Model" has been used to study intra-specific genetic variability of fruit growth, dry matter content and sugar concentration (Quilot, et al., 2005). Fruit species diversity, which is high regarding traits such as size, sweetness, acidity, starch accumulation, skin transpiration, xylem fluxes and growth rates, could be advantageously analysed with this modelling approach.

The Virtual Fruit model could also be improved by refining the coupling between cell division and cell expansion and by integrating endoreduplication (Fanwoua et al., 2013), for which an independent model is available (Bertin et al., 2007). Despite their importance, the interactions between cell growth processes (division, endoreduplication, expansion) during fruit development are still unclear and subjected to debate (Beemster et al., 2003; Breuninger and Lenhard, 2010; Sugimoto-Shirasu and Roberts, 2003; John and Qi, 2008). To overcome this problem, *in silico* analyses of different coupling strategies could help to clarify the debate, providing insights into the control of organ development. In parallel, recent models describing cell growth and resource allocation

developed for unicellular systems could also be used as a benchmark to better investigate the links among cell growth, metabolism and ploidy, in a general theory of cell economy (Molenaar et al., 2009; Weiße et al., 2015; Scott et al., 2010).

Considering that most parameters are usually fitted in process-based models, the search for their genetic bases is only possible by forward genetics approaches such as QTL-mapping, in which co-localisations between QTL for traits and QTL for model parameters are searched (e.g., Yin et al., 1999; Reymond et al., 2003; Quilot et al., 2005; Prudent et al., 2011). Although such approach is very promising, it is relatively slow and work intensive, especially in species in which genetic resources and tools are limited. Now, the integration of process-based models with more mechanistic models might represent an easier way to identify those parameters having the strongest control over a trait of interest.

Integrative modelling

Fruit growth and quality are a result of an integrative system that functions at different levels of the plant and combines metabolic networks and biophysical processes. For example, fruit size is a function of cell number and cell expansion (Bertin et al., 2007), where the former is tightly related to cell division and the latter largely depends on the biophysical properties of water transport that cannot be predicted solely from metabolic reactions. As discussed above, stoichiometric and enzyme-based kinetic models focused at subcellular or cellular levels can capture a clear picture of metabolic fluxes, but often overlook the dependencies and coordination between different compartments of a whole-plant (Grafahrend-Belau et al., 2013; Rennenberg and Herschbach, 2014). On the other hand, the process-based dynamic models are often too simplified to have direct links to biological processes. Linking process-based models (Figure 4) to the genetic basis of metabolism could lead to powerful tools to manipulate fruit biomass and quality (Struik et al., 2005). The interest is twofold (Baldazzi et al. 2012). From the point of view of molecular biology, the existence of an integrated, multi-scale model could offer a useful framework to interpret omics data, in relation to environmental factors, developmental stages and agricultural practices. From an ecophysiological perspective, the integration of cellular and molecular levels can help refine plant models, shedding light onto the complex interplay between different spatial and temporal scales in the emerging system response (Chew et al., 2014). In particular, the integration of an enzyme-based kinetic model (Beauvoit et al., 2014) into a process-based model (Fishman and Génard, 1998) would enable the identification of those enzymes and/or transporters having the strongest control over a trait of interest (e.g., fruit size or sugar concentration), thus opening the possibility to manipulate this trait.

Despite those advantages, an integrated fruit model linking detailed fruit metabolism with biophysical fruit growth is, to our knowledge, not available. However, active initiatives are running in

741 the crop research community in attempting to create an integrative and multilevel 'crop *in silico*'
742 platform (Marshall-Colon et al., 2017). A model covering various organisation levels (subcellular,
743 cellular, organ, or whole plant) will provide a holistic view of the system regulation and coordination
744 that cannot be reached with a model that is specific for a single level. Moreover integrating models at
745 multi-scales will pave a way to exploiting trade-offs in configuration of metabolism between
746 organisation levels (Sweetlove and Fernie, 2013). Multiscale and combined metabolic models are
747 required to be able to use flux-balance models as a framework for metabolic engineering especially to
748 improve crop yield and quality (Baghalian et al., 2014).

749 Model integration can be done by different strategies, from manual and loose integration to tight
750 and automatic integration, which will also affect the efficiency and performance of the integrated
751 model (Borgdorff et al., 2012; Zhu et al., 2016). Several platforms have been developed to facilitate
752 model integration with different frameworks but they are still rarely used by plant modellers (see
753 detailed review in Marshall-Colon et al., 2017). Process-based simulation models have been
754 successfully integrated into a so-called virtual peach fruit by manually recoding and connecting
755 several existing models (Lescourret and Génard, 2005), a process that turned out to be time-
756 consuming. Flux balance analysis (FBA) models have also been integrated with other types of models
757 to provide an organ or even whole-plant view. Multiscale and combined metabolic models are required
758 to be able to use flux-balance models as a framework for metabolic engineering, especially to improve
759 crop yield and quality (Baghalian et al. 2014). For instance the role of photorespiration during the
760 evolution of C4 photosynthesis has been studied by coupling the genome-scale FBA model C4GEM
761 (de Oliveira Dal'Molin et al., 2010) with a mechanistic model of carbon fixation. The same authors
762 also applied the FBA model of metabolism for leaf, stem and root systems across a day and night
763 cycle to investigate how the metabolism of a given tissue is coordinated within the whole-plant and to
764 assess the effect of translocation costs on tissue metabolism (de Oliveira Dal'Molin et al., 2015).

765 In addition to spatial integration, it is also possible to extend the static FBA into dynamic mode
766 (dFBA), by integrating the simulated outputs at an earlier step to update the substrate and product
767 amounts of the metabolic network, which will then be used as inputs for the next time step
768 (Mahadevan et al., 2002). Grafahrend-Belau et al. (2013) developed FBA models for leaf, stem, ear,
769 and root of a barley plant and integrated each of them with a dynamic whole-plant function-structure
770 model. The resulting integrated model revealed source-to-sink shifts during plant development and
771 provided a novel approach for *in silico* analysis of whole-plant metabolism. Chew et al. (2014; 2017a)
772 achieved another elegant model in Arabidopsis, from gene regulation via metabolism to whole-plant
773 growth, by integrating several existing models in a modular way with minimal modifications of the
774 original model. Recently, it has been proposed that epigenetic regulation, gene expression, and
775 metabolism could be integrated to simulate lycopene biosynthesis in growing tomato fruit (Gallusci et
776 al., 2017).

Although there are successful examples of model integration, it still remains very challenging to achieve (Baldazzi et al., 2012). For example FBA models often provide a population of solutions with equal goodness-of-fit for the objective function, while a unique solution will be needed for the following iterations when it is integrated into a dynamic growth model. This may result in important derivations of model simulation and novel algorithms will have to be developed to solve this problem (Martins Conde et al., 2016). Then, integrating a detailed metabolic model with a process-based biophysical fruit growth model will dramatically increase the number of parameters, which can cause difficulties in parameterisation of the integrated model. Thus, model reduction during model integration might be necessary to obtain combined models with a reasonable number of parameters (Baldazzi et al., 2012). The challenge here will be to perform large numbers of simulations, in which parameters would be merged and environmental factors removed or simplified. To this end, the following steps seem to be crucial for model integration: (1) Standardising data collection and organisation for creating a comprehensive data depository accessible to the public. It will be crucial to have a database with sufficient quality and scope covering the various organisation levels for model integration (Zhu et al., 2016); (2) Perform model cross-validations by comparing common variables. This will also open up a range of possibilities regarding the analysis of metabolism; (3) Reducing model complexity. As mentioned above, integrating models might dramatically increase the number of parameters to estimate or determine experimentally, and thus strongly increase the need for phenotypic data. Therefore, a compromise between performance and complexity could be searched by excluding dispensable components, i.e. parameters that have little influence on the simulations. Finally, we anticipate that integrated models will enable *in silico* analyses of the interactions between fruit biophysical properties and the distribution of metabolic fluxes, and ultimately provide valuable clues for potential targets of metabolic engineering.

Overall, with the development of high-performance computing, progresses in FBA and enzyme-based kinetic models, expansion of process-based dynamic models, it is timely to integrate isolated models into a multiscale model framework covering gene regulatory networks, activities and properties of enzymes, metabolic pathways and their compartmentation, and plant growth. Such multiscale models, both for crops and fruits, which will gain from multidisciplinarity within plant sciences and above (Zhu et al., 2016; Chew et al., 2017b; Marshall-Colon et al., 2017), could lead to ideotype design by picking the right parameters, and eventually accelerate breeding (Long et al., 2015; Constantinescu et al., 2016; Zhu et al., 2016; Chenu et al., 2017).

Conclusion

Considerable knowledge about fruit metabolism has been accumulated. So far, progress in manipulating fruit quality and biomass production has mainly resulted from forward approaches, i.e. in

which the phenotype has been used to select the best genotypes and/or agricultural practices. The fact that reverse approaches have been less successful implies that the right targets for improvement remain to be found. Indeed, we have seen above that increasing or decreasing the activity of enzymes or transporters does not necessarily lead to desired phenotypes. Then, despite the considerable work that has been required to collect and interpret post-genomic data, our understanding of the functioning of central and primary metabolism remains patchy. Trade-offs between metabolic pools on the one hand and between quality and growth on the other hand are often invoked although rarely expected, confirming that understanding what determines the size and composition of fruits is challenging. Indeed, these traits result from a range of processes that are controlled at different levels of organisation, with subtle interactions occurring inside or between these levels. They are determined through successive phases of development including cell division, cell expansion with potential endoreduplication, carbon storage and accumulation of specialised metabolites, and finally maturation, which can be seen as sinks in competition. Furthermore, fruit traits are not only a matter of molecular and biochemical events, biophysical processes also need to be taken into account, in particular to understand what is behind the trade-offs mentioned above. Modelling represents a great hope to cope with such complexity. When combined to experimentation through an iterative progression, it takes advantage of the presently available resources in computation and analytics to simulate biological processes. Experimentation on fruit producing crops is usually costly and time consuming, especially when slow growing fruits are studied. In consequence, anticipating as much as possible future needs in terms of modelling might prove very useful. Tables I and II propose a range of parameters and variables that are needed in the modelling approaches presented above. We estimate that all analyses mentioned in Table II could be performed with samples of 2-3 g of fresh material. Sampling would be best performed under cryogenic conditions and throughout fruit growth and development. It can indeed be anticipated that fruit modelling will increasingly benefit from high quality data, especially data about biomass composition.

Acknowledgements

All authors acknowledge funding from ANR (ANR-15-CE20-0009-01 FRIMOUSS).

References

Agudelo-Romero P, Erban A, Rego C, Carbonell-Bejerano P, Nascimento T, Sousa L, Martínez-Zapater JM, Kopka J, Fortes AM. 2015. Transcriptome and metabolome reprogramming in Vitis vinifera cv. Trincadeira berries upon infection with *Botrytis cinerea*. *Journal of Experimental Botany* 66: 1769-1785.

Alba R, Cordonnier-Pratt MM, Pratt LH. 2000. Fruit-localized phytochromes regulate lycopene accumulation independently of ethylene production in tomato. *Plant Physiology* 123: 363-370.

Albert E, Segura V, Gricourt J, Bonnefoi J, Derivot L, Causse M. 2016. Association mapping reveals the genetic architecture of tomato response to water deficit: focus on major fruit quality traits. *Journal of Experimental Botany* 7, 6413-6430.

Albertini MV, Carcouet E, Pailly O, Gambotti C, Luro F, Berti L. 2006. Changes in organic acids and sugars during early stages of development of acidic and acidless citrus fruit. *Journal of Agricultural and Food Chemistry* 54: 8335-8339.

Alper H, Jin YS, Moxley JF, Stephanopoulos G. 2005. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli. *Metabolic Engineering* 7: 155-164.

Amemiya T, Kanayama Y, Yamaki S, Yamada K, Shiratake K. 2006. Fruit-specific V-ATPase suppression in antisense transgenic tomato reduces fruit growth and seed formation. *Planta* 223: 1272–1280.

Argyris JM, Díaz A, Ruggieri V, Fernandez M, Jahrmann T, Gibon Y, Pico B, Martín-Hernández AM, Monforte AJ, Garcia-Mas J. 2017. QTL analyses in multiple populations employed for the fine mapping and identification of candidate genes at a locus affecting sugar accumulation in melon (Cucumis melo L.). *Frontiers in Plant Science* 8: 1679.

Baghalian K, Hajirezaei M-R, Schreiber F. 2014. Plant Metabolic Modeling: Achieving New Insight into Metabolism and Metabolic Engineering. *The Plant Cell* 26: 3847-3866.

Baldazzi V, Bertin N, de Jong H, Génard M. 2012. Towards multiscale plant models: integrating cellular networks. *Trends in Plant Science* 17, 728-736.

Baldazzi V, Pinet A, Vercambre G, Bénard C, Biais B, Génard M. 2013. In-silico analysis of water and carbon relations under stress conditions. A multi-scale perspective centered on fruit. *Frontiers in Plant Science* 4: 495.

Ballester A-R, Tikunov Y, Molthoff J, Grandillo S, Viquez-Zamora M, de Vos R, de Maagd RA, van Heusden S, Bovy AG. 2016. Identification of Loci Affecting Accumulation of Secondary Metabolites in Tomato Fruit of a Solanum lycopersicum × Solanum chmielewskii Introgression Line Population. *Frontiers in Plant Science* 7: 1428.

874 Barry CS, Giovannoni JJ. 2007. Ethylene and fruit ripening. *Journal of Plant Growth Regulation* 26:
875 143-159.

876 Bauchet G, Grenier S, Samson N, Segura V, Kende A, Beekwilder J, Cankar K, Gallois J-L Gricourt J,
877 Bonnet J, Baxter C, Grivet L, Causse M. 2017. Identification of major loci and genomic regions
878 controlling acid and volatile content in tomato fruit: implications for flavor improvement. *New*
879 *Phytologist* 215: 624–641.

880 Beauvoit B, Colombié S, Monier A, Andrieu MH, Biais B, Bénard C, Cheniclet C, Dieuaide-
881 Noubhani M, Nazaret C, Mazat JP, Gibon Y. 2014. Model-Assisted Analysis of Sugar Metabolism
882 throughout Tomato Fruit Development Reveals Enzyme and Carrier Properties in Relation to
883 Vacuole Expansion. *The Plant Cell* 26: 3224-3242.

884 Bénard C., Gautier H., Bourgaud F., Grasselly D., Navez B., Caris-Veyrat C., Weiss M., Genard M.
885 2009. Effects of low nitrogen supply on tomato (solanum lycopersicum) fruit yield and quality with
886 special emphasis on sugars, acids, ascorbate, carotenoids, and phenolic compounds. *Journal of*
887 *Agricultural and Food Chemistry* 57:4112-4123.

888 Bertin N, Lecomte A, Brunel B, Fishman S, Génard M. 2007. A model describing cell
889 polyploidization in tissues of growing fruit as related to cessation of cell proliferation. *Journal of*
890 *Experimental Botany* 58, 1903-1913.

891 Biais B, Bénard C, Beauvoit B, Colombié S, Prodhomme D, Ménard G, Bernillon S, Gehl B, Gautier
892 H, Ballias P, Mazat J-P, Sweetlove L, Génard M, Gibon Y. 2014. Remarkable reproducibility of
893 enzyme activity profiles in tomato fruits grown under contrasting environments provides a roadmap
894 for studies of fruit metabolism. *Plant Physiology* 164: 1204–1221.

895 Bonghi C, Manganaris GA. 2012. Systems Biology Approaches Reveal New Insights into
896 Mechanisms Regulating Fresh Fruit Quality. In OMICs Technologies - Tools for Food Science,
897 Benkeblia N (Ed) CRC Press, pp 201-226.

898 Bordbar A, Monk JM, King ZA, Palsson BO. 2014. Constraint-based models predict metabolic and
899 associated cellular functions. *Nature Reviews Genetics* 15: 107-120.

900 Borgdorff J, Bona-Casas C, Mamonski M, Kurowski K, Piontek T, Bosak B, Rycerz K, Ciepiela E,
901 Gubala T, Harezlak D, Bubak M, Lorenz E, Hoekstra AG. 2012. A distributed multiscale
902 computation of a tightly coupled model using the multiscale modeling language. *Procedia*
903 *Computer Science* 9: 596-605.

904 Both AJ, Benjamin L, Franklin J, Holroyd G, Incoll LD, Lefsrud MG, Pitkin G. 2015. Guidelines for
905 measuring and reporting environmental parameters for experiments in greenhouses. *Plant Methods*
906 11: 43.

907 Brochado AR, Matos C, Moller BL, Hansen J, Mortensen UH, Patil KR. 2010. Improved vanillin
908 production in baker's yeast through in silico design. *Microbial Cell Factories* 9: 84.

909 Brummell DA, Harpster MH. 2001. Cell wall metabolism in fruit softening and quality and its
910 manipulation in transgenic plants. *Plant Molecular Biology* 47: 311-340.

911 Burger J, Sa'ar U, Paris HS, Lewinsohn E, Katzir N, Tadmor Y, Schaffer AA. 2006. Genetic
912 variability for valuable fruit quality traits in *Cucumis melo*. *Israel Journal of Plant Sciences* 54:
913 233-242.

914 Bussières P. 1994. Water Import Rate in Tomato Fruit: A Resistance Model. *Annals of Botany* 73: 75-
915 82.

916 Carrari F, Baxter C, Usadel B, Urbanczyk-Wochniak E, Zanor MI, Nunes-Nesi A, Nikiforova V,
917 Centero D, Ratzka A, Pauly M, Sweetlove LJ, Fernie AR. 2006. Integrated analysis of metabolite
918 and transcript levels reveals the metabolic shifts that underlie tomato fruit development and
919 highlight regulatory aspects of metabolic network behaviour. *Plant Physiology* 142: 1380-1396.

920 Carrari F, Fernie AR. 2006. Metabolic regulation underlying tomato fruit development. *Journal of*
921 *Experimental Botany* 57, 1883-1897.

922 Causse M, Duffe P, Gomez MC, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-
923 Chamley M, Rothan C. 2004. A genetic map of candidate genes and QTLs involved in tomato fruit
924 size and composition. *Journal of Experimental Botany* 55: 1671-1685.

925 Centeno DC, Osorio S, Nunes-Nesi A, Bertolo ALF, Carneiro RT, Araujo WL, Steinhauser MC,
926 Michalska J, Rohrmann J, Geigenberger P, Oliver SN, Stitt M, Carrari F, Rose JKC, Fernie AR.
927 2011. Malate plays a crucial role in starch metabolism, ripening, and soluble solid content of
928 tomato fruit and affects postharvest softening. *The Plant Cell* 23: 162-184.

929 Chenu K, Porter JR, Martre P, Basso B, Chapman SC, Ewert F, Bindi M, Asseng S. 2017.
930 Contribution of crop models to adaptation in wheat. *Trends in Plant Science* 22, 472-490.

931 Chew YH, Seaton DD, Mengin V, Flis A, Mugford ST, Smith AM, Stitt M, Millar AJ. 2017a. Linking
932 circadian time to growth rate quantitatively via carbon metabolism. *bioRxiv* 105437.

933 Chew YH, Seaton DD, Millar AJ. 2017b. Multi-scale modelling to synergise Plant Systems Biology
934 and Crop Science. *Field Crops Research* 202: 77-83.

935 Chew YH, Wenden B, Flis A, Mengin V, Taylor J, Davey CL, Tindal C, Thomas H, Ougham HJ, de
936 Reffye P, Stitt M, Williams M, Muetzelfeldt R, Halliday KJ, Millar AJ. 2014. Multiscale digital
937 Arabidopsis predicts individual organ and whole-organism growth. *Proceedings of the National*
938 *Academy of Sciences* 111, E4127-E4136.

Cohen S, Itkin M, Yeselson Y, Tzuri G, Portnoy V, Harel-Baja R, Lev S, Sa‘ar U, Davidovitz-Rikanati R, Baranes N, Bar E, Wolf D, Petreikov M, Shen S, Ben-Dor S, Rogachev I, Aharoni A, Ast T, Schuldiner M, Belausov E, Eshed R, Ophir R, Sherman A, Frei B, Neuhaus HE, Xu Y, Fei Z, Giovannoni J, Lewinsohn E, Tadmor Y, Paris HS, Katzir N, Burger Y, Schaffer AA. 2014. The PH gene determines fruit acidity and contributes to the evolution of sweet melons. *Nature Communications* 5: 4026.

Colombié S, Nazaret C, Bénard C, Biais B, Mengin V, Sole M, Fouillen L, Dieuaide-Noubhani M, Mazat JP, Beauvoit B, Gibon Y (2015) Modelling central metabolic fluxes by constraint-based optimization reveals metabolic reprogramming of developing *Solanum lycopersicum* (tomato) fruit. *The Plant Journal* 81: 24-39.

Colombié S, Beauvoit B, Nazaret C, Bénard C, Vercambre G, Le Gall S, Biais B, Cabasson C, Maucourt M, Bernillon S, Moing A, Dieuaide-Noubhani M, Mazat JP, Gibon Y. 2017. Respiration climacteric in tomato fruits elucidated by constraint-based modelling. *New Phytologist* 213: 1726-1739.

Constantinescu D, Memmah M-M, Vercambre G, Génard M, Baldazzi V, Causse M, Albert E, Brunel B, Valsesia P, Bertin N. 2016. Model-assisted estimation of the genetic variability in physiological parameters related to tomato fruit growth under contrasted water conditions. *Frontiers in Plant Science* 7, 1841.

Cornish-Bowden A. 2004. Fundamentals of enzyme kinetics, 3rd edn. Portland Press, London

D’Esposito D, Ferriello F, Dal Molin A, Diretto G, Sacco A, Minio A, Barone A, Di Monaco R, Cavella S, Tardella L. 2017. Unraveling the complexity of transcriptomic, metabolomic and quality environmental response of tomato fruit. *BMC Plant Biology* 17: 66.

Dai ZW, Léon C, Feil R, Lunn JE, Delrot S, Gomès E. 2013. Metabolic profiling reveals coordinated switches in primary carbohydrate metabolism in grape berry (*Vitis vinifera* L.), a non-climacteric fleshy fruit. *Journal of Experimental Botany* 64: 1345-1355.

Dal'Molin CGD, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK. 2010. AraGEM, a Genome-Scale Reconstruction of the Primary Metabolic Network in Arabidopsis. *Plant Physiology* 152: 579-589

Davies C, Robinson SP. 1996. Sugar accumulation in grape berries - Cloning of two putative vacuolar invertase cDNAs and their expression in grapevine tissues. Plant Physiol. 111: 275-283.

de Bolt S, Cook DR, Ford CM. 2006. L-Tartaric acid synthesis from vitamin C in higher plants. PNAS 103: 5608-5613.

de Jong TM, Favreau R, Grossman YL, Lopez G. 2011. Using Concept-Based Computer Simulation Modeling to Study and Develop an Integrated Understanding of Tree Crop Physiology. *Acta Horticulturae* 903: 751-757.

de Oliveira Dal'Molin CG, Quek L-E, Palfreyman RW, Brumbley SM, Nielsen LK. 2010. AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. *Plant Physiology* 152: 579-589.

de Oliveira Dal'Molin CG, Quek L-E, Saa PA, Nielsen LK. 2015. A multi-tissue genome-scale metabolic modeling framework for the analysis of whole plant systems. *Frontiers in Plant Science* 6: 4.

de Swaef T, Mellisho CD, Baert A, de Schepper V, Torrecillas A, Conejero W, Steppe K. 2014. Model-assisted evaluation of crop load effects on stem diameter variations and fruit growth in peach. *Trees* 28: 1607-1622.

Desnoues E, Gibon Y, Baldazzi V, Signoret V, Génard M, Quilot-Turion B. 2014. Profiling sugar metabolism during fruit development in a peach progeny with different fructose-to-glucose ratios. *BMC Plant Biology* 14: #336.

Desnoues E, Baldazzi V, Génard M, Mauroux J-B, Lambert P, Confolent C, Quilot-Turion B. 2016. Dynamic QTLs for sugars and enzyme activities provide an overview of genetic control of sugar metabolism during peach fruit development. *Journal of Experimental Botany* 67: 3419-3431.

Ding JG, Chen BW, Xia XJ, Mao WH, Shi K, Zhou YH, Yu JQ. 2013. Cytokinin-Induced Parthenocarpic Fruit Development in Tomato Is Partly Dependent on Enhanced Gibberellin and Auxin Biosynthesis. *PLoS One* 8: e70080

Ding Y, Chang J, Ma Q, Chen L, Liu S, Jin S, Han J, Xu R, Zhu A, Guo J, Luo Y, Xu J, Xu Q, Zeng Y, Deng X, Cheng Y. 2015. Network analysis of postharvest senescence process in citrus fruits revealed by transcriptomic and metabolomic profiling. *Plant Physiology* 168: 357-376.

Eastmond PJ. 2007. MONODEHYROASCORBATE REDUCTASE4 is required for seed storage oil hydrolysis and postgerminative growth in Arabidopsis. *The Plant Cell* 19: 1376-1387.

Etienne A., Génard M., Bugaud C. 2015. A process-based model of TCA cycle functioning to analyze citrate accumulation in pre- and post-harvest fruits. *PLoS One* 10: e0126777.

Etienne A., Génard M., Lobit P., Bugaud C. 2014. Modeling the vacuolar storage of malate shed lights on pre- and post-harvest fruit acidity. *BMC Plant Biology* 14: 310.

Etienne A., Génard M., Lobit P., Mbeguie-A-Mbeguie D., Bugaud C. 2013. What controls fleshy fruit acidity? A review of malate and citrate accumulation in fruit cells. *Journal of Experimental Botany* 64:1451-1469.

Fanciullino AL, Bidel LPR, Urban L. 2014. Carotenoid responses to environmental stimuli: integrating redox and carbon controls into a fruit model. *Plant Cell and Environment* 37: 273-289.

Fanwoua J, de Visser PHB, Heuvelink E, Yin XY, Struik PC, Marcelis LFM. 2013. A dynamic model of tomato fruit growth integrating cell division, cell growth and endoreduplication. *Functional Plant Biology* 40: 1098-1114.

Fernandez-Moreno J-P, Tzfadia O, Forment J, Presa S, Rogachev I, Meir S, Orzaez D, Aharoni A, Granell A. 2016. Characterization of a new pink fruit tomato mutant result in the identification of a null allele of the SlMYB12 transcription factor. *Plant Physiology* 171: 1821-1836.

Fishman S, Génard M. 1998. A biophysical model of fruit growth: simulation of seasonal and diurnal dynamics of mass. *Plant Cell and Environment* 21: 739-752.

Frary A, Nesbitt TC, Frary A, Grandillo S, van der Knaap E, Cong B, Liu JP, Meller J, Elber R, Alpert KB, Tanksley SD. 2000. fw2.2: A quantitative trait locus key to the evolution of tomato fruit size. *Science* 289: 85-88.

Frary A, Göl D, Keleş D, Ökmen B, Pınar H, Şığva HÖ, Yemenicioğlu A, Doğanlar S. 2010. Salt tolerance in Solanum pennellii: antioxidant response and related QTL. *BMC Plant Biology* 10: 58-74.

Fridman E, Pleban T, Zamir D. 2000. A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. *Proceedings of the National Academy of Sciences* 97, 4718-4723.

Fridman E, Carrari F, Liu YS, Fernie AR, Zamir D. 2004. Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* 305: 1786-1789.

Fu FQ, Mao WH, Shi K, Zhou YH, Asami T, Yu JQ. 2008. A role of brassinosteroids in early fruit development in cucumber. *Journal of Experimental Botany* 59: 2299–2308

Galindo A, Cruz ZN, Rodríguez P, Collado-González J, Corell M. 2016. Jujube fruit water relations at fruit maturation in response to water deficits. *Agricultural Water Management* 164: 110-117.

Gallie DR 2013. The role of ascorbic acid recycling in responding to environmental stress and in promoting plant growth. *Journal of Experimental Botany* 64: 433-443.

Gallusci P, Hodgman C, Teyssier E, Seymour GB. 2016. DNA Methylation and Chromatin Regulation during Fleshy Fruit Development and Ripening. *Frontiers in Plant Science* 7: 807.

Gallusci P, Dai Z, Génard M, Gauffretau A, Leblanc-Fournier N, Richard-Molard C, Vile D, Brunel-Muguet S. 2017. Epigenetics for plant improvement: Current knowledge and modeling avenues. *Trends in Plant Science* 22: 610-623.

1036 Garcia V, Bres C, Just D, Fernandez L, Wong Jun Tai F, Mauxion J-P, Le Paslier M-C, Bérard A,
1037 Brunel D, Aoki K, Alseekh S, Fernie AR, Fraser PD, Rothan C. 2016. Rapid identification of
1038 causal mutations in tomato EMS populations via mapping-by-sequencing. *Nature Protocols* 11:
1039 2401-2418.

1040 Garcia V, Stevens R, Gil L, Gilbert L, Gest N, Petit J, Faurobert M, Maucourt M, Deborde C, Moing
1041 A, Poessel J-L, Jacob D, Bouchet J-P, Giraudel J-L, Gouble B, Page D, Alhagdow M, Massot C,
1042 Gautier H, Lemaire-Chamley M, de Daruvar A, Rolin D, Usadel B, Lahaye M, Causse M, Baldet P,
1043 Rothan C. 2009. An integrative genomics approach for deciphering the complex interactions
1044 between ascorbate metabolism and fruit growth and composition in tomato. *Comptes Rendus*
1045 *Biologie* 332, 1007-1021.

1046 Gary C, Lebot J, Frossard JS, Andriolo JL. 1998. Ontogenic Changes in the Construction Cost of
1047 Leaves, Stems, Fruits, and Roots of Tomato Plants. *Journal of Experimental Botany* 49: 59-68.

1048 Gautier H, Rocci A, Buret M, Grasselly D, Causse M. 2005. Fruit load or fruit position alters response
1049 to temperature and subsequently cherry tomato quality. *Journal of the Science of Food and*
1050 *Agriculture* 85: 1009-1016.

1051 Gautier H, Diakou-Verdin V, Bénard C, Reich M, Buret M, Bourgaud F, Poessel JL, Caris-Veyrat C,
1052 Génard M. 2008. How does tomato quality (sugar, acid, and nutritional quality) vary with ripening
1053 stage, temperature, and irradiance? *Journal of Agricultural and Food Chemistry* 56: 1241-1250.

1054 Gautier H., Massot C., Stevens R., Serino S., Genard M. 2009. Regulation of tomato fruit ascorbate
1055 content is more highly dependent on fruit irradiance than leaf irradiance. *Annals of Botany*
1056 103:495-504.

1057 Gest N, Gautier H, Stevens R. 2013. Ascorbate as seen through plant evolution: the rise of a successful
1058 molecule? *Journal of Experimental Botany* 64: 33-53.

1059 Génard M, Huguet JG. 1996. Modeling the response of peach fruit growth to water stress. *Tree*
1060 *Physiology* 16: 407-415.

1061 Génard M, Souty M. 1996. Modeling the peach sugar contents in relation to fruit growth. *Journal of*
1062 *the American Society for Horticultural Science* 121: 1122–1131.

1063 Ghan R, Van Sluyter SC, Hochberg U, Degu A, Hopper DW, Tillet RL, Schlauch KA, Haynes PA,
1064 Fait A, Cramer GR. 2015. Five omic technologies are concordant in differentiating the biochemical
1065 characteristics of the berries of five grapevine (*Vitis vinifera* L.) cultivars. *BMC Genomics* 16: 946.

1066 Gibert C, Lescourret F, Génard M, Vercambre G, Pérez Pastor A. 2005. Modelling the effect of fruit
1067 growth on surface conductance to water vapour diffusion. *Annals of Botany* 95: 673-683.

1068 Ginzberg I, Stern RA. 2016. Strengthening fruit-skin resistance to growth strain by application of plant
1069 growth regulators. *Scientia Horticulturae* 198: 150-153.

1070 Giovannoni JJ. 2004. Genetic regulation of fruit development and ripening. *The Plant Cell* 16: S170-
1071 S180.

1072 Giovannoni JJ. 2006. Breeding new life into plant metabolism. *Nature Biotechnology* 24: 418-419.

1073 Giovannoni JJ. 2016. Prospects: The Tomato Genome as a Cornerstone for Gene Discovery. In:
1074 Causse M, Giovannoni J, Bouzayen M, Zouine M, eds. The Tomato Genome. Berlin, Heidelberg:
1075 Springer Berlin Heidelberg, 257-259.

1076 Giovannoni JJ, Nguyen C, Ampofo B, Zhong SL, Fei ZJ. 2017. The Epigenome and Transcriptional
1077 Dynamics of Fruit Ripening. *Annual Review of Plant Biology* 68: 61-84.

1078 Gonda I, Bar E, Portnoy V, Lev S, Burger J, Schaffer AA, Tadmor Y, Gepstein S, Giovannoni JJ,
1079 Katzir N, Lewinsohn E. 2010. Branched-chain and aromatic amino acid catabolism into aroma
1080 volatiles in Cucumis melo L. fruit. *Journal of Experimental Botany* 61: 1111-1123.

1081 Gourieroux AM, Holzapfel BP, Scollary GR, McCully ME, Canny MJ, Rogiers SY. 2016. The amino
1082 acid distribution in rachis xylem sap and phloem exudate of *Vitis vinifera* 'Cabernet Sauvignon'
1083 bunches. *Plant Physiology and Biochemistry* 105: 45-54.

1084 Grafahrend-Belau E, Junker A, Eschenroder A, Muller J, Schreiber F, Junker BH. 2013. Multiscale
1085 metabolic modeling: dynamic flux balance analysis on a whole-plant scale. *Plant Physiology* 163:
1086 637-647.

1087 Grandillo S, Cammareri M. 2016. Molecular Mapping of Quantitative Trait Loci in Tomato. In:
1088 Causse M, Giovannoni J, Bouzayen M, Zouine M, eds. The Tomato Genome. Berlin, Heidelberg:
1089 Springer Berlin Heidelberg, 39-73.

1090 Grossman YL, de Jong TM. 1994. Peach: a simulation-model of reproductive and vegetative growth in
1091 peach trees. *Tree Physiology* 14: 329-345.

1092 Guo LX, Shi CY, Liu X, Ning DY, Jing LF, Yang H, Liu YZ. 2016. Citrate accumulation-related gene
1093 expression and/or enzyme activity analysis combined with metabolomics provide a novel insight
1094 for an orange mutant. *Scientific Reports* 6: 29343.

1095 Guo X, Xu J, Cui X, Chen H, Qi H. 2017. iTRAQ-based protein profiling and fruit quality changes at
1096 different development stages of oriental melon. *BMC Plant Biology* 17: 28.

1097 Haritatos E, Keller F, Turgeon R. 1996. Raffinose oligosaccharide concentrations measured in cell and
1098 tissue types in *Cucumis melo* L. leaves: implications for phloem loading. *Planta* 198: 614–622.

1099 Hawker JS. 1969. Changes in the activities of enzymes concerned with sugar metabolism during the
1100 development of grape berries. *Phytochemistry* 8: 9-17.

1101 Heinrich R, Rapoport TA. 1974. A linear steady-state treatment of enzymatic chains. General
1102 properties, control and effector strength. *European Journal of Biochemistry* 42: 89-95.

1103 Hill SA, Ap Rees T. 1994. Fluxes of carbohydrate-metabolism in ripening bananas. *Planta* 192: 52-60.

1104 Ho LC. 1996. The mechanism of assimilate partitioning and carbohydrate compartmentation in fruit in
1105 relation to the quality and yield of tomato. *Journal of Experimental Botany* 47: 1239-1243.

1106 Holzhutter HG. 2004. The principle of flux minimization and its application to estimate stationary
1107 fluxes in metabolic networks. *European Journal of Biochemistry* 271: 2905-2922.

1108 Horchani F, Gallusci P, Baldet P, Cabasson C, Maucourt M, Rolin D, Aschi-Smiti S, Raymond P.
1109 2008. Prolonged root hypoxia induces ammonium accumulation and decreases the nutritional
1110 quality of tomato fruits. *Journal of Plant Physiology* 165: 1352-1359.

1111 Jameson PE and Song JC. 2016. Cytokinin: a key driver of seed yield. *Journal of Experimental Botany*
1112 67: 593-606.

1113 Jean D, Lapointe L. 2001. Limited carbohydrate availability as a potential cause of fruit abortion in
1114 *Rubus chamaemorus*. *Physiologia Plantarum* 112: 379-387.

1115 Jia H, Jiu S, Zhang C, Wang C, Tariq P, Liu Z, Wang B, Cui L, Fang J. 2016. Abscisic acid and
1116 sucrose regulate tomato and strawberry fruit ripening through the abscisic acid-stress-ripening
1117 transcription factor. *Plant Biotechnology Journal* 14: 2045-2065.

1118 Jourda C, Cardi C, Gibert O, Toro AG, Ricci J, Mbeguie-A-Mbeguie D, Yahiaoui N. 2017. Lineage-
1119 Specific Evolutionary Histories and Regulation of Major Starch Metabolism Genes during Banana
1120 Ripening. *Frontiers in Plant Science* 7: 1778.

1121 Kacser H, Burns JA. 1973. The control of flux. *Symposia of the Society for Experimental Biology* 27:
1122 65-104.

1123 Katz E, Boo KH, Kim HY, Eigenheer RA, Phinney BS, Shulaev V, Negre-Zakharov F, Sadka A,
1124 Blumwald E. 2011. Label-free shotgun proteomics and metabolite analysis reveal a significant
1125 metabolic shift during citrus fruit development. *Journal of Experimental Botany* 62: 5367-5384.

1126 Katzir N. 2015. Ultra-High Resolution QTL Mapping of Fruit Quality Traits in Melon. Plant and
1127 Animal Genome XXIII Conference: Plant and Animal Genome.

1128 Kettner C. 2007. Good publication practice as a prerequisite for comparable enzyme data? *In Silico*
1129 *Biology* 7: S57-S64.

Khan SA, Chibon P-Y, de Vos RC, Schipper BA, Walraven E, Beekwilder J, van Dijk T, Finkers R, Visser RG, van de Weg EW. 2012. Genetic analysis of metabolites in apple fruits indicates an mQTL hotspot for phenolic compounds on linkage group 16. *Journal of Experimental Botany* 63: 2895-2908.

Klann EM, Hall B, Bennett AB. 1996. Antisense acid invertase (TIV1) gene alters soluble sugar composition and size in transgenic tomato fruit. *Plant Physiology* 112: 1321–1330.

Kinkade MP, Foolad MR. 2013. Genomics-Assisted Breeding for Tomato Fruit Quality in the Next-Generation Omics Age. In: Varshney RK, Tuberosa R, eds. Translational Genomics for Crop Breeding. Chichester, UK: John Wiley & Sons Ltd, 193-210.

Knapp S, Peralta IE. 2016. The Tomato (*Solanum lycopersicum* L., Solanaceae) and Its Botanical Relatives. In: Causse M, Giovannoni J, Bouzayen M, Zouine M, eds. The Tomato Genome. Berlin, Heidelberg: Springer Berlin Heidelberg, 7-21.

Krebs HA. 1957. Control of metabolic processes. *Endeavour* 16: 125-132.

Kromdijk J., Bertin N., Heuvelink E., Molenaar J., de Visser P.H.B., Marcelis L.F.M., Struik P.C.. 2013. Crop management impacts the efficiency of QTL detection and use - case study of fruit load x QTL interactions. *Journal of Experimental Botany* 65: 11-22.

Kumar S, Volz RK, Chagné D, Gardiner S. 2014. Breeding for Apple (Malus × domestica Borkh.) Fruit Quality Traits in the Genomics Era. In: Tuberosa R, Graner A, Frison E, eds. Genomics of Plant Genetic Resources: Volume 2. Crop productivity, food security and nutritional quality. Dordrecht: Springer Netherlands, 387-416.

Lee DR. 1990. A unidirectional water flux model of fruit growth. *Canadian Journal of Botany* 68: 1286-1290.

Lechaudel M, Vercambre G, Lescourret F, Normand F, Genard M. 2007. An Analysis of Elastic and Plastic Fruit Growth of Mango in Response to Various Assimilate Supplies. *Tree Physiology* 27: 219-230.

Leng P, Yuan B, Guo YD. 2014. The role of abscisic acid in fruit ripening and responses to abiotic stress. *Journal of Experimental Botany* 65: 4577-4588.

Lescourret F, Génard M. 2005. A virtual peach fruit model simulating changes in fruit quality during the final stage of fruit growth. *Tree Physiology* 25: 1303-1315.

Leterme P, Buldgen A, Estrada F, Londo AM. 2006. Mineral Content of Tropical Fruits and Unconventional Foods of the Andes and the Rain Forest of Colombia. *Food Chemistry* 95: 644-652.

Lewinsohn E, Schalechet F, Wilkinson J, Matsui K, Tadmor Y, Nam KH, Amar O, Lastochkin E, Larkov O, Ravid U, Hiatt W, Gepstein S, Pichersky E. 2001. Enhanced levels of the aroma and flavor compound S-linalool by metabolic engineering of the terpenoid pathway in tomato fruits. *Plant Physiology* 127: 1256-1265.

Li M, Chen M, Zhang Y, Fu C, Xing B, Li W, Qian J, Li S, Wang H, Fan X, Yan Y, Wang Y, Yang X.2015. Apple fruit diameter and length estimation by using the thermal and sunshine hours approach and its application to the digital orchard management information system. *PLoS One* 10: e0120124.

Li M, Li D, Feng F, Zhang S, Ma F, Cheng L. 2016a. Proteomic analysis reveals dynamic regulation of fruit development and sugar and acid accumulation in apple. *Journal of Experimental Botany* 67: 5145–5157.

Li SJ, Yin XR, Xie XL, Allan AC, Ge H, Shen SL, Chen KS. 2016b. The Citrus transcription factor, CitERF13, regulates citric acid accumulation via a protein-protein interaction with the vacuolar proton pump, CitVHA-c4. *Scientific Reports* 6, 20151.

Liebermeister W, Klipp E. 2006. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theoretical Biology and Medical Modelling* 3:41

Liu HF, Génard M, Guichard S, Bertin N. 2007. Model-assisted analysis of tomato fruit growth in relation to carbon and water fluxes. *Journal of Experimental Botany* 58: 3567-3580.

Lobit P, Génard M, Wu BH, Soing P, Habib R. 2003. Modelling citrate metabolism in fruits: response to growth and temperature. *Journal of Experimental Botany* 54: 2489-2501.

Lobit P, Génard M, Soing P, Habib R. 2006. Modelling malic acid accumulation in fruits: relationships with organic acids, potassium, and temperature. *Journal of Experimental Botany* 57: 1471–1483.

Long SP, Marshall-Colon A, Zhu X-G. 2015. Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell* 161: 56-66.

Luo J. 2015. Metabolite-based genome-wide association studies in plants. *Current Opinion in Plant Biology* 24: 31-38.

Louarn G, Lecoeur J, Lebon E. 2008. A three-dimensional statistical reconstruction model of grapevine (*Vitis vinifera*) simulating canopy structure variability within and between cultivar/training system pairs. *Annals of Botany* 101: 1167-1184.

Lytovchenko A, Eickmeier I, Pons C, Osorio S, Szecowka M, Lehmberg K, Arrivault S, Tohge T, Pineda B, Anton MT, Hedtke B, Lu YH, Fisahn J, Bock R, Stitt M, Grimm B, Granell A, Fernie

AR. 2011. Tomato Fruit Photosynthesis Is Seemingly Unimportant in Primary Metabolism and Ripening But Plays a Considerable Role in Seed Development. *Plant Physiology* 157: 1650-1663.

Mahadevan R, Edwards JS, Doyle FJ. 2002. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophysical Journal* 83, 1331-1340.

Malnoy M, Viola R, Jung M-H, Koo O-J, Kim S, Kim J-S, Velasco R, Nagamangala Kanchiswamy C. 2016. DNA-Free Genetically Edited Grapevine and Apple Protoplast Using CRISPR/Cas9 Ribonucleoproteins. *Frontiers in Plant Science* 7: 1904.

Marshall-Colon A, Long SP, Allen DK, Allen G, Beard DA, Benes B, von Caemmerer S, Christensen AJ, Cox DJ, Hart JC, Hirst PM, Kannan K, Katz DS, Lynch JP, Millar AJ, Panneerselvam B, Price ND, Prusinkiewicz P, Raila D, Shekar RG, Shrivastava S, Shukla D, Srinivasan V, Stitt M, Turk MJ, Voit EO, Wang Y, Yin X, Zhu X-G. 2017. Crops in silico: generating virtual crops using an integrative and multi-scale modeling platform. *Frontiers in Plant Science* 8: 786.

Martins Conde PdR, Sauter T, Pfau T. 2016. Constraint based modeling going multicellular. *Frontiers in Molecular Biosciences* 3: 3.

Martre P, Bertin N, Salon C, Génard M. 2011. Modelling the size and composition of fruit, grain and seed by process-based simulation models. *New Phytologist* 191: 601-618.

Massot C, Génard M, Stevens R, Gautier H. 2010. Fluctuations in sugar content are not determinant in explaining variations in vitamin C in tomato fruit. *Plant Physiology and Biochemistry* 48:751-757.

Massot C, Bancel D, Lauri F, Truffault V, Baldet P, Stevens R, Gautier H. 2013. High temperature inhibits ascorbate recycling and light stimulation of the ascorbate pool in tomato despite increased expression of biosynthesis genes. *PLoS One* 8: e84474.

McAtee P, Karim S, Schaffer R, David K. 2013. A dynamic interplay between phytohormones is required for fruit development, maturation, and ripening. *Frontiers in Plant Science* 4: 79.

Menu T, Saglio P, Granot D, Dai N, Raymond P, Ricard B. 2004. High hexokinase activity in tomato fruit perturbs carbon and energy metabolism and reduces fruit and seed size. *Plant Cell and Environment* 27: 89-98.

Mestre TC, Garcia-Sanchez F, Rubio F, Martinez V, Rivero RM. 2012. Glutathione homeostasis as an important and novel factor controlling blossom-end rot development in calcium-deficient tomato fruits. *Journal of Plant Physiology* 169: 1719-1727.

Moing A, Aharoni A, Biais B, Rogachev I, Meir S, Brodsky L, Allwood JW, Erban A, Dunn WB, Kay L, de Koning S, de Vos RCH, Jonker H, Mumm R, Deborde C, Maucourt M, Bernillon S, Gibon Y, Hansen TH, Husted S, Goodacre R, Kopka J, Schjoerring JK, Rolin D, Hall RD. 2011. Extensive

metabolic cross-talk in melon fruit revealed by spatial and developmental combinatorial metabolomics. *New Phytologist* 190, 683-696.

Morandini P. 2009. Rethinking metabolic control. *Plant Science* 176: 441-451.

Morandini P, 2013. Control limits for accumulation of plant metabolites: brute force is no substitute for understanding. *Plant Biotechnology Journal* 11: 253–267.

Mounet F, Moing A, Garcia V, Petit J, Maucourt M, Deborde C, Bernillon S, Le Gall G, Colquhoun I, Defernez M, Giraudel J-L, Rolin D, Rothan C, Lemaire-Chamley M. 2009. Gene and metabolite regulatory network analysis of early developing fruit tissues highlights new candidate genes for the control of tomato fruit composition and development. *Plant Physiology* 149, 1505-1528.

Mounet-Gilbert L, Dumont M, Ferrand C, Bournonville C, Monier A, Jorly J, Lemaire-Chamley M, Mori K, Atienza I, Hernould M, Stevens R, Lehner A, Mollet JC, Rothan C, Lerouge P, Baldet, P. 2016. Two tomato GDP-D-mannose epimerase isoforms involved in ascorbate biosynthesis play specific roles in cell wall biosynthesis and development. *Journal of Experimental Botany* 67: 4767-4777.

Nardozza S, Boldingh HL, Osorio S, Hohne M, Wohlers M, Gleave AP, MacRae EA, Richardson AC, Atkinson RG, Sulpice R, Fernie AR, Clearwater MJ. 2013. Metabolic analysis of kiwifruit (*Actinidia deliciosa*) berries from extreme genotypes reveals hallmarks for fruit starch metabolism. *Journal of Experimental Botany* 64: 5049-5063.

Negri AS, Prinsi B, Failla O, Scienza A, Espen L. 2015. Proteomic and metabolic traits of grape exocarp to explain different anthocyanin concentrations of the cultivars. *Frontiers in Plant Science* 6: 603.

Noiraud N, Maurousset L, Lemoine R. 2001. Transport of polyols in higher plants. *Plant Physiology and Biochemistry* 39: 717-728.

Noor E, Flamholz A, Bar-Even A, Davidi D, Milo R, Liebermeister W. 2016. The Protein Cost of Metabolic Fluxes: Prediction from Enzymatic Rate Laws and Cost Minimization. *PLoS Computational Biology* 12: e1005167.

Oikawa A, Otsuka T, Nakabayashi R, Jikumaru Y, Isuzugawa K, Murayama H, Saito K, Shiratake K. 2015. Metabolic profiling of developing pear fruits reveals dynamic variation in primary and secondary metabolites, including plant hormones. *PLoS One* 10, e0131408.

Okabe Y, Ariizumi T. 2016. Mutant Resources and TILLING Platforms in Tomato Research. In: Ezura H, Ariizumi T, Garcia-Mas J, Rose J, eds. Functional Genomics and Biotechnology in Solanaceae and Cucurbitaceae Crops. Berlin, Heidelberg: Springer Berlin Heidelberg, 75-91.

1258 Olimpieri I, Siligato F, Caccia R, Mariotti L, Ceccarelli N, Soressi GP, Mazzucato A. 2007. Tomato
1259 fruit set driven by pollination or by the parthenocarpic fruit allele are mediated by transcriptionally
1260 regulated gibberellin biosynthesis. *Planta* 226: 877-888.

1261 Osorio S, Alba R, Damasceno CMB, Lopez-Casado G, Lohse M, Zanor MI, Tohge T, Usadel B, Rose
1262 JKC, Fei Z, Giovannoni JJ, Fernie AR. 2011. Systems Biology of Tomato Fruit Development:
1263 Combined Transcript, Protein, and Metabolite Analysis of Tomato Transcription Factor (nor, rin)
1264 and Ethylene Receptor (Nr) Mutants Reveals Novel Regulatory Interactions. *Plant Physiology* 157,
1265 405-425.

1266 Osorio S, Nunes-Nesi A, Stratmann M, Fernie AR. 2013. Pyrophosphate levels strongly influence
1267 ascorbate and starch content in tomato fruit. *Frontiers in Plant Science* 4: #308.

1268 Palmer LJ, Palmer LT, Pritchard J, Graham RD, Stangoulis JC. 2013. Improved techniques for
1269 measurement of nanolitre volumes of phloem exudate from aphid stylectomy. *Plant Methods* 9: 18.

1270 Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD. 1988. Resolution of
1271 quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment
1272 length polymorphisms. *Nature* 335: 721-726.

1273 Patrick JW, 1997. Phloem unloading: sieve element unloading and post-sieve element transport.
1274 *Annual Review of Plant Physiology and Plant Molecular Biology* 48: 191-222.

1275 Pattison RJ, Csukasi F and Catala C. 2014. Mechanisms regulating auxin action during fruit
1276 development. *Physiologia Plantarum* 151: 62-72.

1277 Pfau T, Pires Pacheco M and Sauter T. 2016. Towards improved genome-scale metabolic network
1278 reconstructions: unification, transcript specificity and beyond. *Briefings in Bioinformatics* 17:
1279 1060-1069.

1280 Poiroux-Gonord F, Bidel LPR, Fanciullino AL, Gautier H, Lauri-Lopez F, Urban L. 2010. Health
1281 benefits of vitamins and secondary metabolites of fruits and vegetables and prospects to increase
1282 their concentrations by agronomic approaches. *Journal of Agricultural and Food Chemistry* 58:
1283 12065-12082.

1284 Poiroux-Gonord F, Fanciullino AL, Poggi I, Urban L 2013. Carbohydrate control over carotenoid
1285 build-up is conditional on fruit ontogeny in clementine fruits. *Physiologia Plantarum* 147: 417-431.

1286 Qin GZ, Zhu Z, Wang WH, Cai JH, Chen Y, Li L, Tian SP. 2016. A Tomato Vacuolar Invertase
1287 Inhibitor Mediates Sucrose Metabolism and Influences Fruit Ripening. *Plant Physiology* 172:
1288 1596-1611.

1289 Quadrana L, Almeida J, Asís R, Duffy T, Dominguez PG, Bermúdez L, Conti G, Corrêa da Silva JV,
1290 Peralta IE, Colot V, Asurmendi S, Fernie AR, Rossi M, Carrari F. 2014. Natural occurring
1291 epialleles determine vitamin E accumulation in tomato fruits. *Nature Communications* 5: 4027.

1292 Quilot B, Kervella J, Génard M, Lescourret F. 2005. Analysing the genetic control of peach fruit
1293 quality through an ecophysiological model combined with a QTL approach. *Journal of*
1294 *Experimental Botany* 56: 3083-3092.

1295 Raines C. 2010. Increasing photosynthetic carbon assimilation in C3 plants to improve crop yield:
1296 current and future strategies. *Plant Physiology* 155: 36-42.

1297 Rennenberg H, Herschbach C. 2014. A detailed view on sulphur metabolism at the cellular and whole-
1298 plant level illustrates challenges in metabolite flux analyses. *Journal of Experimental Botany* 65,
1299 5711-5724.

1300 Rick CM. 1986. Tomato mutants, freaks, anomalies, and breeders' resources. *HortScience* 21: 918-
1301 919.

1302 Ripoll J, Urban L, Staudt M, Lopez-Lauri F, Bidel LPR and Bertin N. 2014. Water shortage and
1303 quality of fleshy fruits-making the most of the unavoidable. *Journal of Experimental Botany* 65:
1304 4097-4117.

1305 Ripoll J, Urban L, Brunel B, Bertin N. 2016. Water deficit effects on tomato quality depend on fruit
1306 developmental stage and genotype. *Journal of Plant Physiology* 190: 26-35.

1307 Rohrmann J, Tohge T, Alba R, Osorio S, Caldana C, McQuinn R, Arvidsson S, van der Merwe MJ,
1308 Riaño-Pachón DM, Mueller-Roeber B, Fei Z, Nesi AN, Giovannoni JJ, Fernie AR. 2011.
1309 Combined transcription factor profiling, microarray analysis and metabolite profiling reveals the
1310 transcriptional control of metabolic shifts occurring during tomato fruit development. *The Plant*
1311 *Journal* 68: 999-1013.

1312 Rohwer JM, Botha FC. 2001. Analysis of sucrose accumulation in the sugar cane culm on the basis of
1313 in vitro kinetic data. *Biochemistry Journal* 358: 437-445.

1314 Rohwer J.M. 2012. Kinetic modelling of plant metabolic pathways. *Journal of Experimental Botany*
1315 63:2275-2292.

1316 Rossi M. Bermudez L. Carrari F. 2015. Crop yield: challenges from a metabolic perspective. *Current*
1317 *Opinion in Plant Biology* 25: 79:89.

1318 Rouphael Y, Schwarz D, Krumbeinb A, Colla G. 2010. Impact of grafting on product quality of fruit
1319 vegetables. Scientia *Horticulturae* 127: 172-179.

Rossouw D, Kossmann J, Botha FC, Groenewald JH. 2010. Reduced neutral invertase activity in the culm tissues of transgenic sugarcane plants results in a decrease in respiration and sucrose cycling and an increase in the sucrose to hexose ratio. *Functional Plant Biology* 37: 22-31.

Ruan YL, Patrick JW. 1995. The cellular pathway of post-phloem sugar transport in developing tomato fruit. *Planta* 196: 434-444

Ruiu F, Picarella ME, Imanishi S, Mazzucato A. 2015. A transcriptomic approach to identify regulatory genes involved in fruit set of wild-type and parthenocarpic tomato genotypes. *Plant Molecular Biology* 89: 263-278.

Sagor GHM, Berberich T, Tanaka S, Nishiyama M, Kanayama Y, Kojima S, Muramoto K, Kusano T. 2016. A novel strategy to produce sweeter tomato fruits with high sugar contents by fruit-specific expression of a single bZIP transcription factor gene. *Plant Biotechnology Journal* 14: 1116-1126.

Sanchez MT, Perez-Marin D, Torres I, Gil B, Garrido-Varo A, de la Haba MJ. 2017. Use of NIRS technology for on-vine measurement of nitrate content and other internal quality parameters in intact summer squash for baby food production. *Postharvest Biology and Technology* 125: 122-128.

Sanchez G, Venegas-Calerón M, Salas JJ, Monforte A, Badenes ML, Granell A. 2013. An integrative "omics" approach identifies new candidate genes to impact aroma volatiles in peach fruit. *BMC Genomics* 14: 343.

Saure MC. 2014. Why calcium deficiency is not the cause of blossom-end rot in tomato and pepper fruit - a reappraisal. *Scientia Horticulturae* 174: 151-154

Sauvage C, Rau A, Aichholz C, Chadoeuf J, Sarah G, Ruiz M, Santoni S, Causse M, David J, Glémin S. 2017. Domestication rewired gene expression and nucleotide diversity patterns in tomato. *The Plant Journal* 91: 631–645.

Sauvage C, Segura V, Bauchet G, Stevens R, Do PT, Nikoloski Z, Fernie AR, Causse M. 2014. Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant Physiology* 165: 1120-1132.

Savage JA, Haines DF, Holbrook NM. 2015. The making of giant pumpkins: how selective breeding changed the phloem of Cucurbita maxima from source to sink. *Plant Cell and Environment* 38:1543-1554.

Schaffer AA, Levin I, Oguz I, Petreikov M, Cincarevsky F, Yeselson Y, Shen S, Gilboa N, Bar M, 2000. ADPglucose pyrophosphorylase activity and starch accumulation in immature tomato fruit: the effect of a Lycopersicon hirsutum-derived introgression encoding for the large subunit. *Plant Science* 152: 135-144.

1353 Schallau K, Junker BH. 2010. Simulating plant metabolic pathways with enzyme-kinetic models.
1354 *Plant Physiology* 152:1763-1771.

1355 Scholander PF, Hammel HT, Hemmingsen E, Bradstreet ED. 1965. Hydrostatic Pressure and Osmotic
1356 Potential in Leaves of Mangroves and Some Other Plants. *Proceedings of the National Academy of*
1357 *Sciences of the United States of America* 52: 119-125.

1358 Serrani JC, Sanjuán R, Ruiz-Rivero O, Fos M, García-Martínez JL. 2007. Gibberellin regulation of
1359 fruit set and growth in tomato. *Plant Physiology* 145: 246-257.

1360 Shi H, Schwender J. 2016. Mathematical models of plant metabolism. *Current Opinion in*
1361 *Biotechnology* 37: 143-152.

1362 Shinozaki Y, Hao SH, Kojima M, Sakakibara H, Ozeki-Iida Y, Zheng Y, Fei ZJ, Zhong SL,
1363 Giovannoni JJ, Rose JK, Okabe Y, Heta Y, Ezura H, Ariizumi T. 2015. Ethylene suppresses tomato
1364 (*Solanum lycopersicum*) fruit set through modification of gibberellin metabolism. *The Plant*
1365 *Journal* 83: 237-251.

1366 Shiratake K, Martinoia E. 2007. Transporters in fruit vacuoles. *Plant Biotechnology* 24: 127–133.

1367 Snowden CJ, Thomas B, Baxter CJ, Smith JAC, Sweetlove LJ. 2015. A tonoplast Glu/Asp/GABA
1368 exchanger that affects tomato fruit amino acid composition. *The Plant Journal* 81: 651-60.

1369 Stamp P, Visser R. 2012. The twenty-first century, the century of plant breeding. *Euphytica* 186:585–
1370 591.

1371 Struik PC, Yin X, de Visser P. 2005. Complex quality traits: now time to model. *Trends in Plant*
1372 *Science* 10: 513-516.

1373 Sweetlove LJ, Ratcliffe RG. 2011. Flux-balance modeling of plant metabolism. *Frontiers in Plant*
1374 *Science* 2: 38

1375 Sweetlove LJ, Fernie AR. 2013. The spatial organization of metabolism within the plant cell. *Annual*
1376 *Review of Plant Biology* 64: 723-746.

1377 Sweetlove LJ, Williams TCR, Cheung CYM and Ratcliffe RG 2013. Modelling metabolic CO_2
1378 evolution - a fresh perspective on respiration. *Plant Cell and Environment* 36: 1631-1640.

1379 Szymanski J, Levin Y, Savidor A, Breitel D, Chappell-Maor L, Heinig U, Töpfer N, Aharoni A. 2017.
1380 Label-free deep shotgun proteomics reveals protein dynamics during tomato fruit tissues
1381 development. *The Plant Journal* 90, 396-417.

1382 Takayama M, Ezura H. 2015. How and why does tomato accumulate a large amount of GABA in the
1383 fruit? *Frontiers in Plant Science* 6: 612.

1384 Tohge T, Scossa F, Fernie AR. 2015. Integrative Approaches to Enhance Understanding of Plant
1385 Metabolic Pathway Structure and Regulation. *Plant Physiology* 169: 1499-1511.

1386 Tomason Y, Nimmakayala P, Levi A, Reddy UK. 2013. Map-based molecular diversity, linkage
1387 disequilibrium and association mapping of fruit traits in melon. *Molecular Breeding* 31: 829-841.

1388 Toubiana D, Fernie AR, Nikoloski Z, Fait A. 2013. Network analysis: tackling complex data to study
1389 plant metabolism. *Trends in Biotechnology* 31: 29-36.

1390 Trewavas A. 2004. Aspects of plant intelligence: an answer to firn. *Annals of Botany* 93: 353-357.

1391 Truffault V., Fifel F., Longuenesse JJ., Gautier H. 2015. Impact of temperature integration under
1392 greenhouse on energy use efficiency, plant growth and development and tomato fruit quality
1393 depending on cultivar rootstock combination. *Acta Horticulturae* 1099: 95-100.

1394 Tummler K, Lubitz T, Schelker M, Klipp E. 2010. New types of experimental data shape the use of
1395 enzyme kinetics for dynamic network modelling. *The FEBS Journal* 281: 549-571.

1396 Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A, Steinhauser D,
1397 Persson S, Provart NJ. 2009. Co-expression tools for plant biology: opportunities for hypothesis
1398 generation and caveats. *Plant Cell and Environment* 32: 1633-1651.

1399 Uys L, Botha FC, Hofmeyr JHS, Rohwer JM. 2007. Kinetic model of sucrose accumulation in
1400 maturing sugarcane culm tissue. *Phytochemistry* 68: 2375-2392.

1401 van Ittersum MK, Donatelli M. 2003. Modelling cropping systems - highlights of the symposium and
1402 preface to the special issues. *European Journal of Agronomy* 18: 187-197.

1403 Voxeur A, Gilbert L, Rihouey C, Driouich A, Rothan C, Baldet P, Lerouge P. 2011. Silencing of the
1404 GDP-D-mannose 3,5-Epimerase Affects the Structure and Cross-linking of the Pectic
1405 Polysaccharide Rhamnogalacturonan II and Plant Growth in Tomato. *Journal of Biological*
1406 *Chemistry* 286: 8014-8020.

1407 Walton EF, Dejong TM. 1990. Estimating the Bioenergetic Cost of a Developing Kiwifruit Berry and
1408 Its Growth and Maintenance Respiration Components. *Annals of Botany* 66: 417-424.

1409 Wong DCJ, Matus JT. 2017. Constructing integrated networks for identifying new secondary
1410 metabolic pathway regulators in grapevine: recent applications and future opportunities. *Frontiers*
1411 *in Plant Science* 8: 505.

1412 Wu HX, Jia HM, Ma XW, Wang SB, Yao QS, Xu WT, Zhou YG, Gao ZS, Zhan RL. 2014.
1413 Transcriptome and proteomic analysis of mango (*Mangifera indica* Linn) fruits. *Journal of*
1414 *Proteomics* 105: 19-30.

Ye J, Hu T, Yang C, Li H, Yang M, Ijaz R, Ye Z, Zhang Y. 2015. Transcriptome profiling of tomato fruit development reveals transcription factors associated with ascorbic acid, carotenoid and flavonoid biosynthesis. *PLoS One* 10, 1-25.

Yun Z, Qu H, Wang H, Zhu F, Zhang Z, Duan X, Yang B, Cheng Y, Jiang Y. 2016. Comparative transcriptome and metabolome provides new insights into the regulatory mechanisms of accelerated senescence in litchi fruit after cold storage. *Scientific Reports* 6: 19356.

Zegbe JA, Behboudian MH, Clothier BE. 2006. Yield and fruit quality in processing tomato under partial rootzone drying. *European Journal of Horticultural Science* 71: 252-258.

Zhang LZ, Garneau MG, Majumdar R, Grant J, Tegeder M. 2015. Improvement of pea biomass and seed productivity by simultaneous increase of phloem and embryo loading with amino acids. *The Plant Journal* 81: 134-146.

Zhang X-Y, Wang X-L, Wang X-F, Xia G-H, Pan Q-H, Fan R-C, Wu F-Q, Yu X-C, Zhang D-P. 2006. A shift of phloem unloading from symplasmic to apoplasmic pathway is involved in developmental onset of ripening in grape berry. *Plant Physiology* 142: 220-232.

Zhang M, Yuan B, Leng P. 2009. The role of ABA in triggering ethylene biosynthesis and ripening of tomato fruit. *Journal of Experimental Botany* 60: 1579-1588.

Zhu X-G, Lynch JP, LeBauer DS, Millar AJ, Stitt M, Long SP. 2016. Plants in silico: why, why now and what?—an integrative platform for plant systems biology research. *Plant Cell and Environment* 39: 1049-1057.

Zhu X-G, de Sturler E, Long SP. 2007.K Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthesis rate: a numerical simulation using an evolutionary algorithm. *Plant Physiology* 145:513-526.

Figure 1. Simplified representation of fruit primary metabolism. Major primary pathways and compounds involved in fruit growth and quality are represented: orange for glycolysis, green for the pentose phosphate pathway (PPP) and fatty acids synthesis, red for the TCA cycle associated to respiration, yellow for redox, purple for the synthesis of structural compounds (proteins, lipids and nucleotides), blue for vacuolar storage, and grey for sugar import.

Figure 2. Hormonal, enzymatic and metabolic changes occurring in tomato fruit pericarp during development and ripening. Hormone levels are expressed in arbitrary units, metabolite levels in $\mu mol.g^{-1}$fresh weight, protein content in $mg.g^{-1}$fresh weight. Enzyme capacities expressed in $units.mg^{-1}$protein have been normalised, grouped into 4 clusters and averaged. Cluster 1: fructokinase, glucokinase, pyruvate kinase, aconitase, NAD-isocitrate dehydrogenase, fumarase, NAD-glutamate dehydrogenase and aspartate aminotransferase; Cluster 2: phosphoglucose isomerase, phosphoglucomutase, ADP-glucose pyrophosphorylase, ATP-phosphofructokinase, PPi-phosphofructokinase, plastidial fructose bisphosphatase, triose phosphate isomerase, NAD-glycerldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase, enolase, phosphoenolpyruvate carboxylase, NAD-malate dehydrogenase, NAD-malic enzyme, NADP-malic enzyme; Cluster 3: sucrose synthase, UDP-glucose pyrophosphorylase, cytosolic fructose bisphosphatase, NADP-glyceraldehyde-3-phosphate dehydrogenase, NADP-glutamate dehydrogenase and alanine aminotransferase; Cluster 4: acid invertase, neutral invertase, sucrose phosphate synthase, aldolase, glucose-6-phosphate dehydrogenase, citrate synthase, NADP-isocitrate dehydrogenase and succinyl-coenzyme A ligase. Adapted from Zhang et al. 2009 and McAtee et al. 2013 for changes in hormone levels and from Biais et al. 2014 for changes in enzyme activities and metabolite concentrations.

Figure 3. Schematic representation of data integration pipeline during construction and refinement of an enzyme-based kinetic model. Chemical information gives a structural framework, which is implemented with enzyme data and further realistically constrained by metabolomic and cytological data to calculate local metabolite concentrations and reaction fluxes.

Figure 4. Fruit model comparison and integration. The comparison of common variables enables cross-validation. The arrows indicate further potential benefits that will be obtained from comparing or coupling kinetic, stoichiometric and/or process-based models describing fruit growth and metabolism. Common variables are summarized in the circuits between the models.

Table I: Experimental variables and parameters for kinetic, stoichiometric and process-based modelling measured in the field, greenhouse or growth chamber. Type refers to the modelling approach (A, all models; K, kinetic, P, process-based; S, stoichiometric). Note that parameters that are very difficult or impossible to measure can be fitted (model calibration).

Variable / parameter	Type	Biological material	Methodology	Purpose
Fruit age	A	Flowers and fruits	Tagging flowers at fertilisation	Determination of the time course of development (Li et al., 2015)
Fruit size	A	Whole fruit	Metric scales	Plotting growth curve (Li et al., 2015)
Fruit surface area	P			Calculate of fruit transpiration and mass flow of phloem and xylem water
Fruit fresh mass	A	Ovaries, whole fruits, or specific fruit tissues	Weighing scales	Plotting of the growth curve and calculation of relative growth rate
Stone or seed proportion	A	Fruit stone or seed		Estimation of flesh proportion
Air temperature	P, K	Ambient air around fruit	Thermometer	Calculation of transpiration, respiration, water and sugar flow (Both et al., 2015)
Air relative humidity	P		Hygrometer	
Stem water potential	P	Plant stem	Pressure chamber[1]	Calculation of water mass flow from xylem into fruit (Scholander et al., 1965)
Fruit maintenance respiration	P	Whole fruit or specific fruit tissues	CO_2 gas analyser	Calculation of maintenance respiration coefficient and Q_{10} temperature coefficient (Walton and Dejong 1990)

Table II: Experimental variables and parameters for kinetic, stoichiometric and process-based modelling measured in the laboratory. The analyses are performed on whole fruits, phloem samples or fruit samples that have been shock frozen in liquid nitrogen when collected. Type refers to the modelling approach (A, all models; K, kinetic, P, process-based; S, stoichiometric).

Variable / parameter	Type	Biological material	Methodology	Purpose
Fruit dry weight	A	Ovaries, whole fruits, or specific fruit tissues	Lyophilisation or oven at 70°C Weighing scale	Calculation of relative growth rate and fresh to dry weight ratio (Gary et al. 1998)
Initial fruit hydrostatic pressure (turgor)	P	Whole fruit or specific tissue of fresh fruit	Pressure probe or calculated from fruit water potential and osmotic pressure	Model initialisation (Lechaudel et al. 2007)
Fruit water potential	P		Chilled mirror hygrometer	Calculation of hydrostatic pressure (turgor) in fruit initialisation (Lechaudel et al. 2007)
Fruit surface conductance to water	P	Whole fruit	Mass loss registered using weighing scales	Calculation of fruit transpiration (Gibert et al. 2005)
Fruit hydrostatic pressure (turgor)	P		Pressure probe or calculated from water potential and osmotic pressure	Estimation of cell wall extensibility/elasticity and yield threshold (Lechaudel et al. 2007)
Fruit osmotic pressure	P	Fruit juice	Freezing point (osmometer)	Calculation of hydrostatic pressure (turgor) in fruit (Galindo et al., 2016)
Fruit pH	P, K		pH meter	Parameterisation of vacuolar H^+-coupled transport (Etienne et al., 2016; Beauvoit et al., 2014)
Fruit growth respiration	P	Whole fruit or specific tissue of fruit	Estimated from carbon and nitrogen content of fruit ashes	Calculation of growth respiration coefficient (Gary et al., 1988)
Stem phloem sugar concentration	P	Stem apex, cut stem or petioles	Aphid stylectomy or phloem exudation Analysis of phloem metabolic composition	Calculation of sugar mass flow from phloem into fruit and active uptake of sugars (Grossman and DeJong 1994; Palmer et al. 2013)
Osmotic pressure of other solutes in stem phloem	P			
Fruit mineral concentrations	P	Fruit ash	Atomic absorption spectrophotometry	Calculation of the contribution of minerals to fruit osmotic pressure and vacuolar transport of acids (Leterme et al. 2006)
Intermediary metabolites	K	Fresh fruit frozen powder or lyophilized powder	Mass spectrometry (IC-Qtrap-MS)	Fitting unknown parameters and model validation (Dai et al., 2013; Beauvoit et al., 2014)
Accumulated metabolites	K, S		Spectrophotometry	Calculation of outfluxes towards accumulated metabolites and biomass compounds (Colombié et al., 2017; Beauvoit et al., 2014)
Total soluble proteins	K, S			
Starch	K, S			
Nucleic acids	K, S			
Starch	K, S			
Lipids	K, S		GC-FID	
Cell wall proportion	K, S	Whole fruit or specific tissue of fruit	Calculated from fruit dry mass, total soluble carbohydrate content and starch content	
Cell wall polysaccharides	S	Dry fruit powder	GC-MS	
Enzymes capacities	K	Fresh fruit frozen powder	Spectrophotometry	Model parameterisation (Beauvoit et al., 2014)
Estimation of subcellular volumes	K	Fixed fruit tissue	Photonic microscopy	

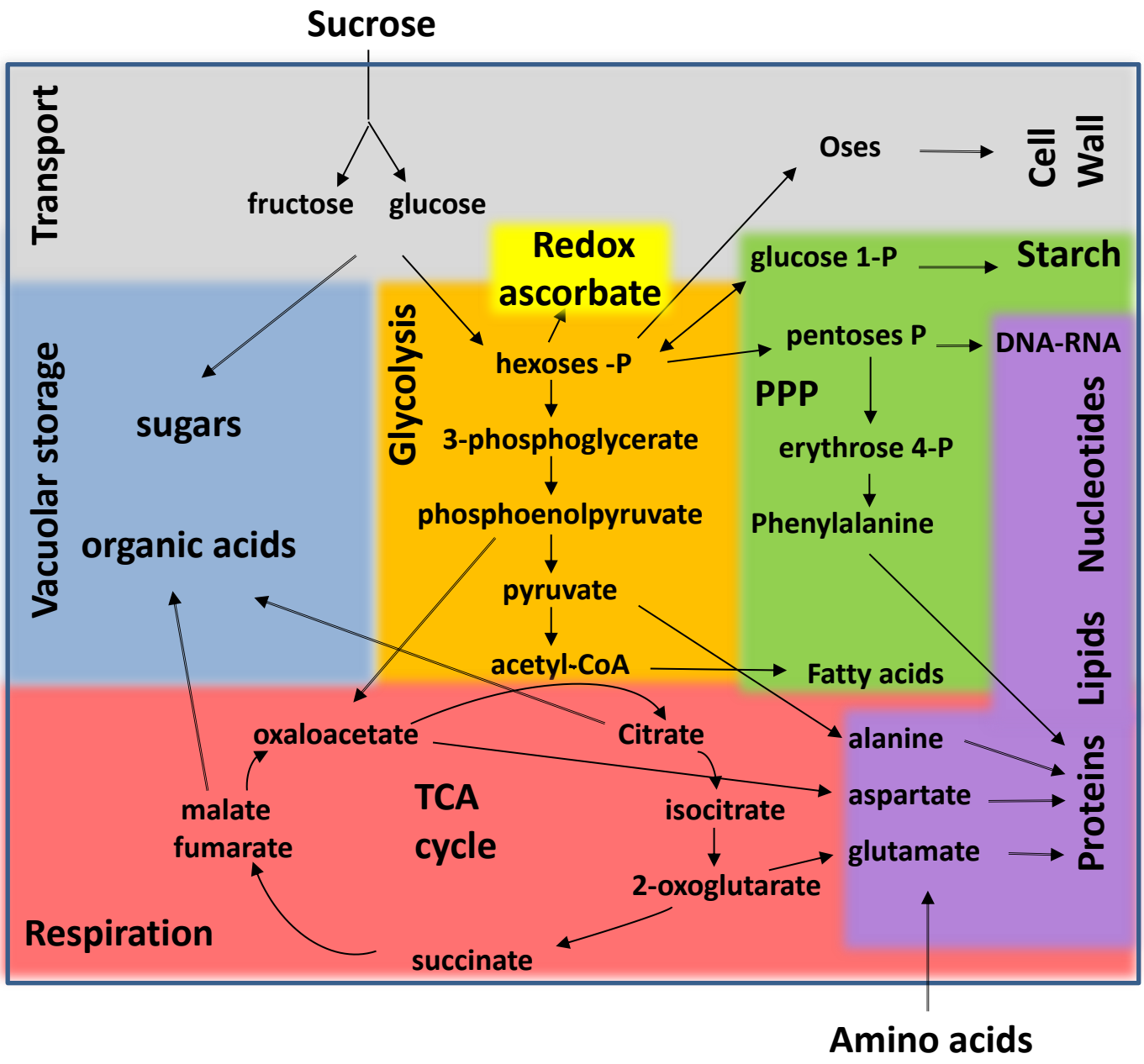

Sucrose
Transport
fructose
glucose
Oses
Cell Wall
Redox
ascorbate
glucose 1-P
Starch
Vacuolar storage
sugars
Glycolysis
hexoses -P
pentoses P
DNA-RNA
PPP
3-phosphoglycerate
erythrose 4-P
Nucleotides
phosphoenolpyruvate
Phenylalanine
organic acids
pyruvate
Lipids
acetyl-CoA
Fatty acids
oxaloacetate
Citrate
alanine
Proteins
aspartate
malate
fumarate
TCA
cycle
isocitrate
glutamate
2-oxoglutarate
Respiration
succinate
Amino acids

Figure 1

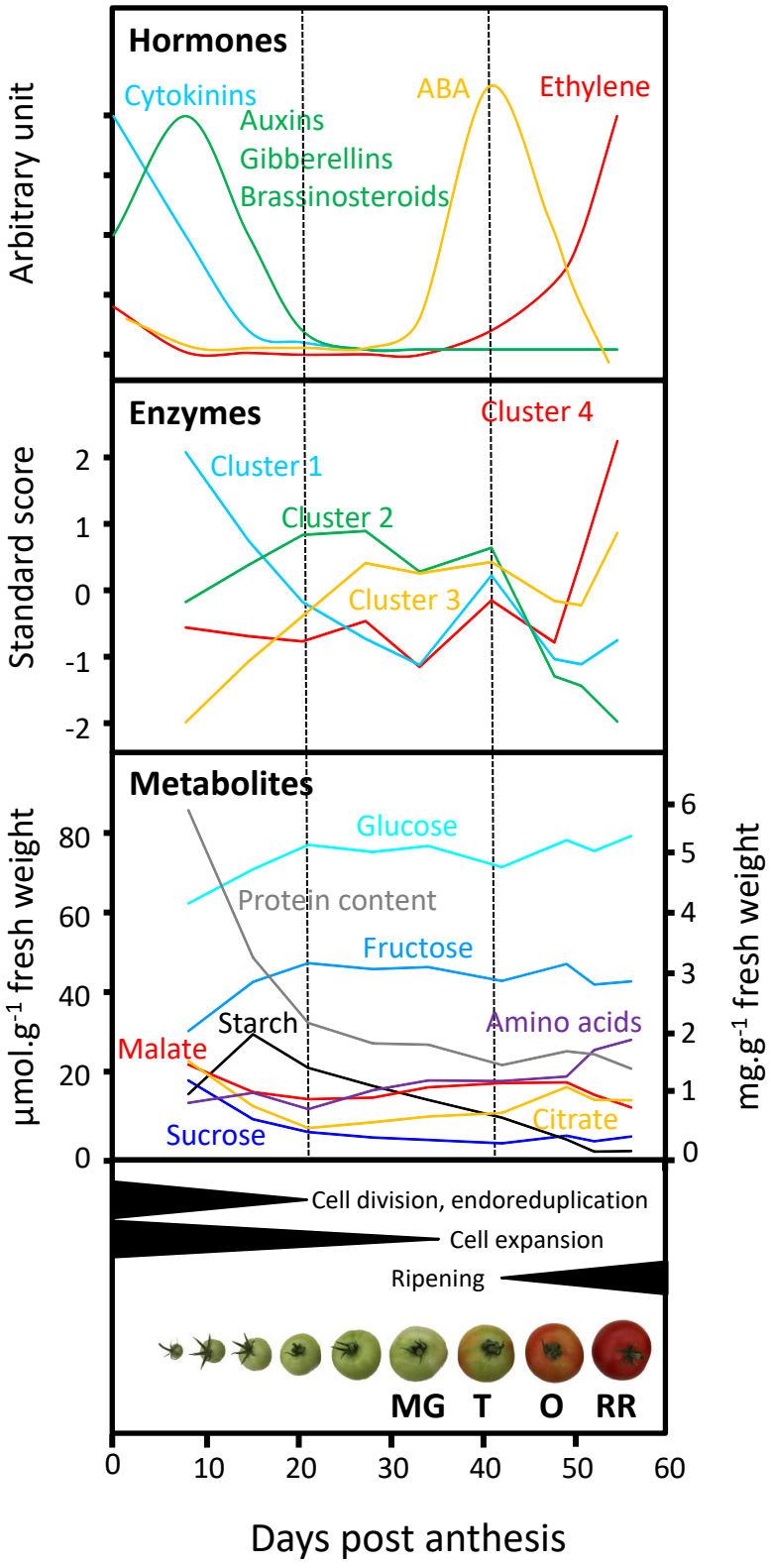

Hormones
Cytokinins
Auxins
Gibberellins
Brassinosteroids
ABA
Ethylene
Arbitrary unit
Enzymes
Cluster 4
Cluster 1
Cluster 2
Cluster 3
Standard score
2
1
0
-1
-2
Metabolites
Glucose
Protein content
Fructose
Starch
Amino acids
Malate
Citrate
Sucrose
µmol.g-1 fresh weight
80
60
40
20
0
6
5
4
3
2
1
0
mg.g-1 fresh weight
Cell division, endoreduplication
Cell expansion
Ripening
MG
T
O
RR
0
10
20
30
40
50
60
Days post anthesis

Figure 2

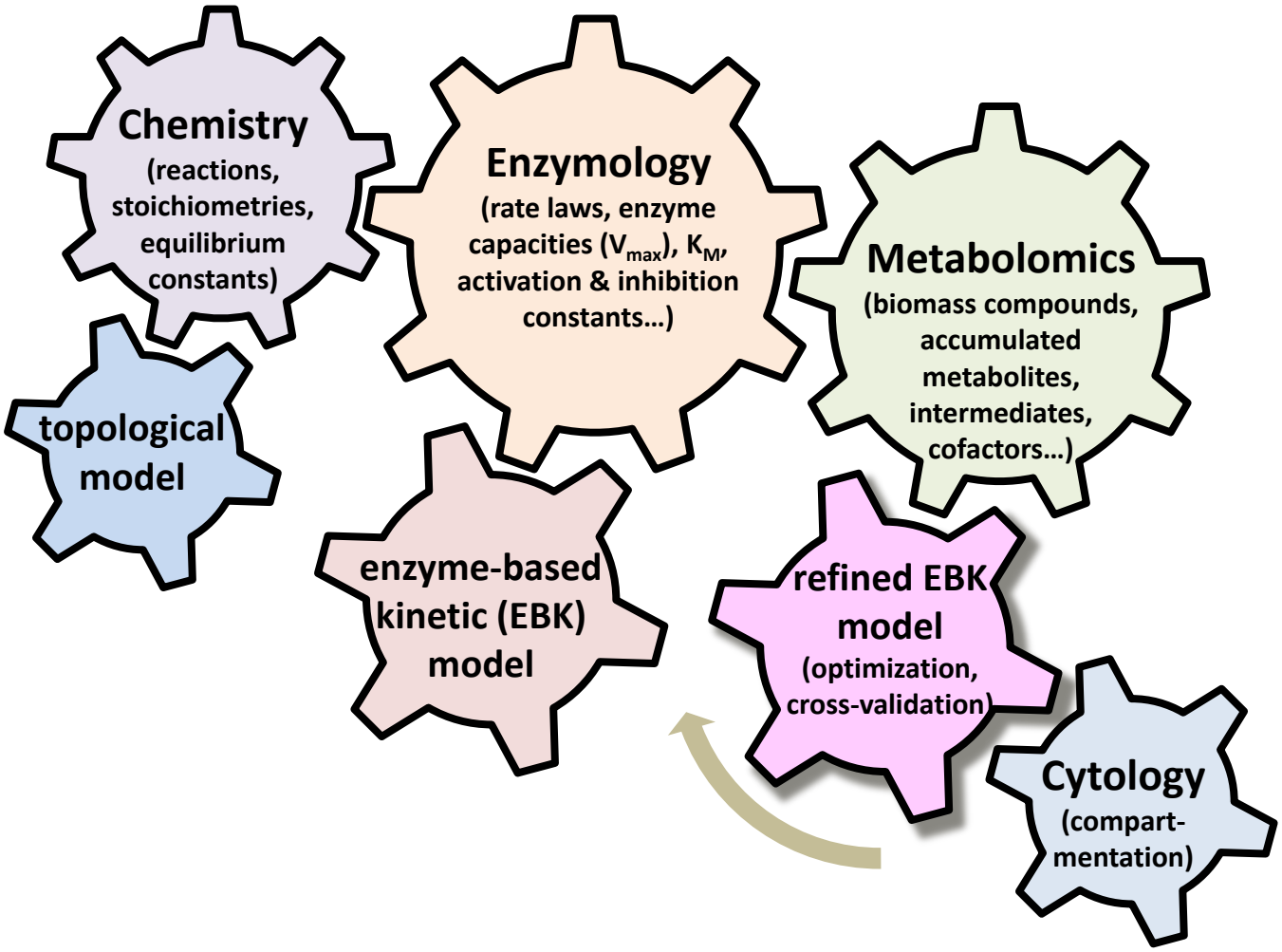

Chemistry
(reactions, stoichiometries, equilibrium constants)
Enzymology
(rate laws, enzyme capacities (V_{max}), K_M, activation & inhibition constants…)
Metabolomics
(biomass compounds, accumulated metabolites, intermediates, cofactors…)
topological model
enzyme-based kinetic (EBK) model
refined EBK model
(optimization, cross-validation)
Cytology
(compart-mentation)

Figure 3

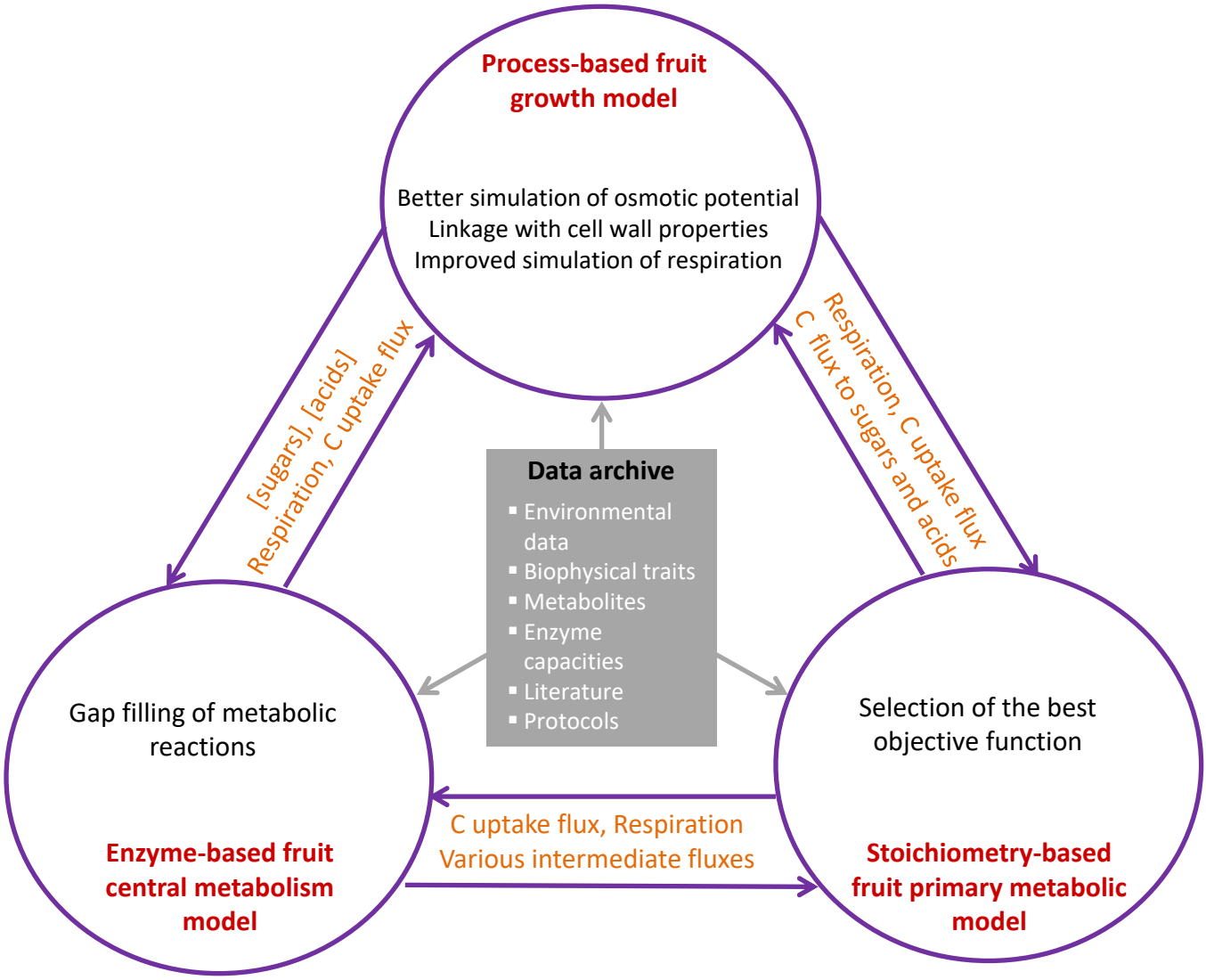

Process-based fruit growth model
Better simulation of osmotic potential
Linkage with cell wall properties
Improved simulation of respiration
[sugars], [acids]
Respiration, C uptake flux
Respiration, C uptake flux
C flux to sugars and acids
Data archive
Environmental data
Biophysical traits
Metabolites
Enzyme capacities
Literature
Protocols
Gap filling of metabolic reactions
Enzyme-based fruit central metabolism model
Selection of the best objective function
Stoichiometry-based fruit primary metabolic model
C uptake flux, Respiration
Various intermediate fluxes

Figure 4

Paper in preparation

Title:

Peptide filtering differently affects the performances of XIC-based quantification methods

Authors:

Belouah Isma, Blein-Nicolas Melisande, Baillau Thierry, Zivy Michel, Gibon Yves, Colombié Sophie

Note: Sections underlined in grey color need to be attested.

INTRODUCTION

In bottom-up proteomics, proteins are digested into peptides which are subsequently separated by liquid chromatography (LC), ionized by electrospray and analyzed by tandem mass spectrometry (MS/MS). Peptide ions, and consequently the proteins from which they originate, can be quantified by integrating the signal intensities obtained from extracted ion currents (XIC; Voyksner and Lee, 1999; Chelius and Bondarenko, 2002). This protein quantification approach, referred to as XIC-based quantification, is highly sensitive. However, it provides as many measurements as there are quantified peptide ions, which presents two main disadvantages. Firstly, peptide intensities have to be summed up into protein abundances. In the last fifteen years, several quantification methods have been proposed to do so, some based on quantitative summaries, other based on statistical modeling (reviewed in Blein-Nicolas and Zivy, 2016). Their relative performances have been evaluated repeatedly, but no clear consensus has emerged so far. The second main disadvantage is that all the peptide intensities associated to a protein are not equivalent because i) all the peptides do not bear the same information (*e.g.* peptides shared by several proteins *vs* proteotypic peptides), ii) the ionization potential varies according to the peptide, such that peptides belonging to a same protein will display different intensity levels (Daly et al., 2008), iii) some peptide ions are incorrectly identified and iv) some peptide ions are incorrectly quantified due to mis-cleavages and/or modifications. Therefore, if not properly considered, peptide ions can introduce errors when computing protein abundances.

To reduce these errors, several authors proposed to filter the peptide data before computing protein abundances. Four types of filter can be distinguished. First, the shared peptide filter. Although they constitute a valuable source of information (Blein-Nicolas et al., 2012), shared peptides are generally discarded because of the difficulty to properly deconvolve the information they carry. Second, the retention time (RT) filter, which aims to remove peptide ions showing highly variable RT potentially arising from mis-identifications. Different methods have been used, based on the standard deviation of RT (Blein-Nicolas et al., 2015) or on RT clustering (Lai et al., 2011). Third, the occurrence filter, which aims to remove peptide ions exhibiting many missing values. Rarely observed peptide ions are indeed inadequate for statistical analysis (Webb-Robertson et al., 2010). Generally, a threshold is arbitrary chosen, *e.g.* a peptide ion should be observed in at least three injections (Lai et al., 2011). More refined approaches have also been proposed, based on a model filtering routine to select peptide ion sets that produce optimal information content (Karpievitch et al., 2009) or taking into account experimental groups such that statistical tests can be properly performed (Webb-Robertson et al.,

2010). Fourth, the outliers filter, which aims to exclude peptide ions showing inconsistent intensity profiles. Several approaches have been proposed, based on the Grubbs' test (Polpitiya et al., 2008), the coefficient of variation (Lai et al., 2011), the peptide ions' correlation (Forshed et al., 2011; Lai et al., 2011) or covariation (Zhang et al., 2017).

These filters have been shown to improve protein quantification (Forshed et al., 2011; Lai et al., 2011; Zhang et al., 2017). However, as quantification methods have different properties related to the computation mode used to estimate protein abundances, we may expect that the relative benefits of filters vary from one quantification method to another. To see how true this is, we performed a spike-in experiment using UPS1 standard to evaluate the effects of the four filter types described above on the performances of five methods of protein quantification. Four of them are commonly used in bottom-up proteomics: i) iBAQ (Schwanhäusser et al., 2011), ii) TOP3 (Silva et al., 2006), iii) Average (Higgs et al., 2005) and iv) intensity modeling (Clough et al., 2009). TOP3 and iBAQ were developed for absolute quantification while Average is widely used for relative quantification. Intensity modeling is recommended by some authors as the most adequate method to infer and quantitatively compare protein abundances (Clough et al., 2009). We included a fifth method, thereafter called Average-Log, to examine the influence of log-transformation of peptide intensities. To our knowledge, this method has never been reported previously.

Materials and Methods

Yeast growth

Saccharomyces cerevisiae strain S288C was inoculated in 5 ml YPD (Yeast extract Peptone Dextrose) medium containing yeast extract (10 $g.l^{-1}$; Difco Laboratories, Detroit, Michigan), bacteriological peptone (20 $g.l^{-1}$; Difco) and glucose (20$g.l^{-1}$). After 24 h of growth at 30°C under agitation, the culture medium was centrifuged (2750 g, 10°C, 3 min) and the supernatant was discarded. The remaining yeast cells pellet was rinsed twice with 5 ml cold distilled water, frozen in liquid nitrogen and stored at -80°C for subsequent protein extraction.

Yeast protein extraction

Proteins were extracted by suspending the pellet of yeast cells in 500 µl of an ice-cold extraction/precipitation solution of acetone containing trichloroacetic acid (10%) and β2-mercaptoethanol (0.07%). To promote cell wall disruption, cells were ground 5 min with 200 µl of glass beads. The protein extract was then shortly vortexed for homogenization and immediately transferred in to new vials to remove glass beads. 750 µl of the extraction/precipitation solution were added to the protein extract before incubation (-20°C for 90 min) and centrifugation (19283 g, 0°C, 15 min). The supernatant was removed and the remaining protein extract was re-suspended in 1.8 ml cold washing acetone solution containing 0.07% β2-mercaptoethanol, incubated (1 h at -20°C) and then centrifuged (19283 g, 0°C, 10 min). This step was repeated twice. After the last washing, the protein pellet was dried in a vacuum centrifuge, weighted and solubilized by adding 15 µl per mg of pellet of a solubilization buffer (6M urea, 2M thiourea, 10mM dithiothreitol (DTT), 30mM Tris-HCl at pH 8.8, 0.1% zwitterionic acid labile surfactant (ZALS)). Remaining cellular debris were segregated from soluble proteins by centrifugation (15000 g, 25 °C, 25 min). Protein concentration was determined using the PlusOne 2-D Quant Kit (GE Healthcare, Little Chalfont, UK) and adjusted with the solubilization buffer to 0.887 $µg.µl^{-1}$.

Spike-in UPS1 preparation

Dried UPS1 proteins (Sigma-Aldrich) were solubilized in the buffer containing yeast proteins to a final concentration of 0.75 $\mu g.\mu l^{-1}$ (0.625 $fmol.\mu l^{-1}$ of each UPS1 protein) such that the total protein (yeast + UPS) concentration was 1.637 $\mu g.\mu l^{-1}$. Proteins were incubated one hour at room temperature for reduction by the 10 mM DTT present in the buffer. Thereafter, proteins were alkylated one hour in presence of 20 mM iodoacetamide and diluted with 50 mM ammonium bicarbonate to decrease total urea and thiourea concentration to 3.6 M before being twice digested. A first 4 hour digestion was performed with 1/32 (w/w) rLysC protease (Promega). After dilution with a solution of 50mM ammonium bicarbonate to decrease total urea and thiourea concentration to 0.77 M, a second overnight digestion was performed with 1/32 (w/w) trypsin (Promega). Both rLysC and trypsin digestion were performed at 37°C. Trypsin digestion was stopped by acidification (1% total volume trifluoroacetic acid). The resulting peptides were purified on solid phase extraction using polymeric C18 column (Phenomenex) with a washing solution containing 0.06% acetic acid and 3% acetonitrile (ACN). After elution with 0.06% acetic acid and 40% ACN, peptides were speedvac-dried and suspended in a solution containing 2% ACN, 0.06% trifluoroacetic acid and 0.06% formic acid so that the concentration of each UPS1 peptide was 141.1 $fmol.\mu l^{-1}$ and the total concentration of yeast peptides was 200 $ng.\mu l^{-1}$. A serial 2.25-fold dilution was prepared by mixing 6.7 µl of UPS1-yeast peptide mix with 8.3 µl of solubilized yeast peptides at 200 $ng.\mu l^{-1}$ until reaching a UPS1 peptides concentration of 0.04 $fmol.\mu l^{-1}$. Eleven samples were thus obtained, containing 141.1, 62.8, 27.9, 12.4, 5.5, 2.2, 1.1, 0.5, 0.2, 0.09 and 0.04 $fmol.\mu l^{-1}$ of each UPS1 peptide. This serial dilution was performed in three replicates from aliquots of the same yeast culture thus leading to a 33 samples experiment.

LS-MS/MS analyses

LC-MS/MS analyses were performed using a NanoLC-Ultra System (nano2DUltra, Eksigent, Les Ulis, France) connected to a Q-Exactive mass spectrometer (Thermo Electron, Waltham, MA, USA). For each sample, 4 µl of protein digest were loaded onto a Biosphere C18 precolumn (0.1 × 20 mm, 100Å, 5 µm; Nanoseparation) at 7.5 $\mu l.min^{-1}$ and desalted with 0.1% formic acid and 2% ACN. After 3 min, the pre-column was connected to a Biosphere C18 nanocolumn (0.075 × 300 mm, 100Å, 3 µm; Nanoseparation). Electrospray ionization was performed at 1.3 kV with an uncoated capillary probe (10 µm tip inner diameter; New Objective,

Woburn, MA, USA). Buffers were 0.1% formic acid in water (A) and 0.1% formic acid and 100% ACN (B). Peptides were separated using a linear gradient from 5 to 35% buffer B for 110 min at 300 nl.min^{-1}. One run took 120 min, including the regeneration step at 95% buffer B and the equilibration step at 100% buffer A.

Peptide ions were analyzed using Xcalibur 2.1 (Thermo Electron) with the following data-dependent acquisition steps: (1) MS scan (mass-to-charge ratio (m/z) 300 to 1.400, 70.000 resolution, profile mode), (2) MS/MS (17.500 resolution, normalized collision energy of 30, profile mode). Step 2 was repeated for the eight major ions detected in step (1). Dynamic exclusion was set to 30 seconds. Xcalibur raw datafiles were transformed to mzXML open source format using msconvert software in the ProteoWizard 3.0.3706 package (Chambers et al., 2012). During conversion, MS and MS/MS data were centroided. The raw MS output files were deposited on-line using PROTICdb database (Ferry-Dumazet et al., 2005; Langella et al., 2007; Langella et al., 2013) at the following URL: http://moulon.inra.fr/protic/XXX. They are currently available with the following username: XXX and password: XXX. They will be made freely available after publication.

Protein identification

Protein identification was performed using the protein sequence database of *S. cerevisiae* strain S288c downloaded from the Saccharomyces Genome Database (SGD project, http://www.yeastgenome.org/, version dated 13/01/2015) completed with the sequences of UPS1 proteins available at http://www.sigmaaldrich.com/content/dam/sigma-aldrich/life-science/proteomics-and-protein/ups1-ups2-sequences.fasta. A contaminant database containing the sequences of standard contaminants was also interrogated. The decoy database comprised the reverse sequences of yeast and UPS1 proteins. Database search was performed with X!Tandem (version 2015.04.01.1; http://www.thegpm.org/TANDEM/) with the following settings. Carboxyamidomethylation of cysteine residues was set to static modification. Oxidation of methionine residues, acetylation or deamination of glutamine and cystein residues were set to possible modifications. Precursor mass precision was set to 10 ppm. Fragment mass tolerance was 0.02 Th. Only peptides with a E-value smaller than 0.05 were reported.

Identified proteins were filtered and sorted by using X!TandemPipeline (version 3.3.0, http://pappso.inra.fr/bioinfo/xtandempipeline/). Criteria used for protein identification

were (i) at least two different peptides identified with an E-value smaller than 0.01 and (ii) a protein E-value (product of unique peptide E-values) smaller than 10^{-5}.

Peptide ion quantification and intensity data filtering

Peptide ions were quantified based on extracted ion chromatograms (XIC) using MassChroQ software version 2.2 (Valot et al., 2011) with the following parameters: "ms2_1" alignment method, tendency_halfwindow of 10, MS1 smoothing halfwindow of 0, MS2 smoothing halfwindow of 15, "quant1" quantification method, XIC extraction based on max, min and max ppm range of 10, anti-spike half of 5, background half median of 5, background half min max of 20, detection thresholds on min and max at 30 000 and 50 000, respectively, peak post-matching mode, ni min abundance of 0.1. The peptide intensities thus obtained constituted the initial dataset (Dataset 0), which was used to derive five differently filtered datasets (Figure 1).

In the first dataset (Dataset 1), no filtering was applied. Yeast peptide intensities were normalized to take into account possible global quantitative variations between LC-MS runs. For this, we used a local normalization method adapted from Lyutvinskiy et al., (2013) and described in Millan-Oropeza et al., (2017). In the second dataset (Dataset 2), yeast peptide intensities were normalized as described above and shared peptides were subsequently removed (shared peptide filter). In the third dataset (Dataset 3), peptides with a standard deviation of retention time higher than 30 seconds were removed (RT filter). Since these peptides are considered as dubious, this filter was applied before normalization of yeast peptide intensities. Then, shared peptides were removed. To derive the fourth dataset (Dataset 4), an occurrence filter was applied to Dataset 3, which resulted in the selection of peptide ions quantified in at least 28 samples, with no more than one missing value per UPS1 concentration. In this way, a maximum of 15.15% of missing values per peptide ion was tolerated and the selected peptide ions were quantified in at least two replicates for each UPS1 concentration. Not to degrade the quality of normalization, which depends on the number of peptide ions quantified both in a sample chosen as reference and in a sample to be normalized, we decided to apply this filter after normalization. A number of peptide ions removed by the occurrence filter are indeed good quality peptides whose intensities can fall below the detection threshold because their ionization potential is low. To derive the fifth dataset (Dataset 5), an outliers filter was applied to Dataset 4. Pearson correlations between log10-transformed intensities were computed for each pair of

peptide ions belonging to the same protein. The peptide ion with the highest number of coefficients of correlation superior or equal to the mean of the positive coefficients of correlation was chosen as a reference for the protein. The peptide ions showing non-significant correlation to the reference (p-value>=0.01) or whose coefficients of correlation to the reference was inferior to 0.8 were considered as outliers and were removed. In order for correlations between peptides to be based on biological and not technical variations, this filter was applied after normalization. Proteins quantified by less than two peptide ions were removed from all the datasets.

Protein quantification

For each protein, five methods were used to compute abundances: i) iBAQ (Schwanhäusser et al., 2011): the sum of peptide ion intensities was divided by the theoretical number of tryptic peptides; ii) TOP3 (Silva et al., 2006): the three most intense peptide ions in median were selected and their mean intensity was computed; iii) Average (Higgs et al., 2005): the mean of all peptide ion intensities was computed, iv) Average Log: peptide ion intensities were log10-transformed before their mean was computed, v) Model: log10-transformed intensities were first modeled using a mixed effect model derived from Blein-Nicolas et al., (2012):

$$I_{ijk} = \mu + A_i + R_j + P_k + \theta_{ijk} + \varepsilon_{ijk}$$

$$where \theta_{ijk} \sim N(0, \sigma_\theta^2),$$

$$\epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$$

, where I_{ijk} is the intensity measured for peptide ion k in replicate j at UPS1 concentration i; A_i represents the effect due to UPS1 concentration i; R_j represents the effect due to replicate j; P_k represents the effect due to the ionization potential of peptide k (also called peptide effect); Θ_{ijk} represents the technical variation due to sample handling and injection in the mass spectrometer; $\boldsymbol{\varepsilon}_{ijk}$ is the residual error. Model was fitted with sum contrasts by maximizing the restricted log-likelihood. Estimated effects of P_k and Θ_{ijk} were subtracted from log10-transformed intensities before their mean is computed. Log-abundances obtained by Average-Log and Model were converted to abundances for further analyses. All data analyses and graphical representations were performed using R version 3.3.2.

RESULTS AND DISCUSSION

In this paper, we evaluated the crossed-effects of peptide filters and quantification methods on protein quantification using a spike-in experiment where UPS1 proteins were mixed at different concentrations to a constant yeast background. Five datasets containing normalized yeast peptide intensities were produced from an initial raw dataset by cumulating five filtering procedures: i) no filter, ii) shared peptide filter, iii) RT filter, iv) occurrence filter and v) outliers filter (Figure 1). For each of these datasets, five quantification methods, referred to as iBAQ, TOP3, Average, Average-Log and Model, were used to compute protein abundances.

Filters differently affect yeast and UPS1 data

The consequences of filters on the amount of observations are presented in Table 1, showing that yeast and UPS1 data are differently affected by the shared peptide filter, the occurrence filter and the outliers filter. The proportion of shared peptides removed was indeed much higher for yeast than for the UPS1 standard (-4.2% *vs* -0.8%, respectively). Shared peptides are related the evolutionary history of organisms. They are particularly common when genes are duplicated and can represent over 50% of the peptides (Podwojski et al., 2010). The occurrence and outliers filters were those that most drastically reduced the whole dataset (-38% and -64% peptide ions, respectively; -26.9% and -32.4% proteins, respectively). At the peptide level, the occurrence filter removed twice more UPS1 peptide ions than yeast peptide ions (77.1% *vs* 35.9%, respectively). This can be explained by the fact that many UPS1 peptide ions were quantified at the highest by not at the lowest UPS1 concentrations. At the protein level, the occurrence filter had also a high impact on the number of quantified UPS1 proteins (-12.2%), mainly excluding small proteins quantified with few peptide ions (Figure S-1). The outliers filter reduced yeast data more drastically than UPS1 data, both at the peptide level (-65% yeast peptide ions *vs* -12.6% UPS1 peptide ions, respectively) and at the protein level (-33.1% yeast proteins *vs* -2.8% UPS1 proteins, respectively). This was expected because the outliers filter is based on the correlation between peptide ions: yeast peptide ions being in constant amounts across samples, they necessarily exhibited poor correlations. Since the outliers filter implicitly allowed to select for proteins showing abundance variations across samples, we could have expected all yeast proteins to be removed. This was not the case because the relative proportion of yeast proteins in the total protein pool actually decreased with increasing UPS1 concentration. However, this variation in the total abundance of yeast proteins was subtle and barely detectable until the

highest concentration of UPS1 (Figure S-2). Altogether, these results show that the effects of filters on the amount of data are highly depend on the experimental design. In particular, the effect of the outliers filter depends on the factors driving protein abundance variations.

Filters effects on estimated protein abundances highlight specific properties of quantification methods

For each UPS1 protein, peptide intensities and protein abundances obtained in the five datasets are presented in Figure S-3. Four of these proteins were used as cases study to illustrate the effects of filters on peptide data and on estimated protein abundances (Figure 2) and to highlight specific properties of the different quantification methods.

The shared peptide filter could change the estimation of protein abundances by several orders of magnitude. In the example illustrated on Figure 2A, six peptide ions were shared between a human ubiquitin and two yeast proteins of high abundance. As the intensities of shared peptides correspond to the sum of abundances of the proteins they belong to (Bukhman et al., 2008), these peptides lead to over-estimate the ubiquitin abundance, especially at the lowest UPS1 concentrations (Figure 2A). Over-estimation was higher for TOP3, iBAQ and Average than for Average-Log and Model because these three quantification methods give more weight to high intensities than to low intensities. TOP3 is indeed computed only from the three most intense peptide ions. If one of them is not representative of the protein it belongs to, it will necessarily affect abundance estimation. Regarding iBAQ and Average, both are more strongly affected by high than by low intensities: iBAQ because it is based on the sum; Average because it is based on the mean of intensities that are log-normally distributed (Podwojski et al., 2010). Mean is indeed known to be highly influenced by extreme values and in the case of log-normally distributed data, there are no extremely low values to counterbalance extremely high values.

The RT filter proved to be efficient to remove peptide ions with inconsistent intensity profiles (Figure 2B), supporting the hypothesis that peptide ions exhibiting high RT variations across samples result from mis-identifications. In the example shown on Figure 2B, the peptide ions removed by the RT filter were among the three most intense. As previously observed for shared peptides, they lead to strongly over-estimate the protein abundances computed by TOP3, iBAQ and Average.

Many peptide ions removed by the occurrence filter presented nice linear responses to increasing UPS1 concentrations, but due to their low ionization potential, they exhibited missing

values at the lowest UPS1 concentrations (Figure 2C). Missing values introduced between-samples variations in the number of peptide ions used to compute protein abundances. As they are mean-based, TOP3, Average and Average Log should be independent from the number of peptide ions quantified in the samples. However, Figure 2C shows that the effect of the occurrence filter on Average and Average Log values increased with the number of peptide ions removed. This is related to the peptide ionization potential, which was on average lower at the highest than at the lowest UPS1 concentrations. In the case of TOP3, the effect of the occurrence filter as illustrated on Figure 2C was not the same as for Average and Average Log because the peptide ion removed by the filter was replaced by another one exhibiting a different ionization potential. Model is also mean-based, but contrarily to Average and Average Log, its values changed uniformly across the concentration range after application of the occurrence filter (Figure 2C). This is related to the P_k term declared in the mixed effects model, which allowed to adjust means of intensities according to the estimated ionization potentials of the peptide ions. Altogether, these results show that due to the unequal peptide ionization potential, missing values can be an important source of between-samples variability for TOP3, Average and Average Log. Of note, as the peptides ions removed by the occurrence filter were generally among the least intense, TOP3, iBAQ and Average were less affected by the occurrence filter than Average Log and Model.

Finally, the outliers filter removed some, but not all the peptide ions exhibiting inconsistent intensity profiles (Figure 2D). To improve the efficiency of this filter, we could have used more stringent filtering criteria. But by doing this, we also took the risk to remove a number of valuable peptide ions. We also could have used a more elaborate algorithm, such as the one recently developed by (Zhang et al., 2017). However, filters optimization was outside the scope of the present study. In the example shown on Figure 2D, TOP3 was not affected by the outliers filter because the removed peptide ions were not among the three most intense. This result shows that TOP3 can be less susceptible to filters than the other quantification methods because it is based on a reduced set of peptide ions that does not include the irrelevant ones.

Relative benefits of filters on precision of protein quantification

Precision is determined by the dispersion around the mean value. It can be enhanced by the implementation of appropriate experimental designs including replicates (Oberg and Vitek, 2009), but it can also be altered by multiple sources of variability, including irrelevant peptides.

Therefore, to evaluate the relative effects of filters on the performances of quantification methods, we first analysed the precision reached by each quantification method in the different datasets. To do so, we computed, at each UPS1 concentration, the coefficients of variation (CV) of each UPS1 protein across technical replicates. The lower the CV, the higher the precision.

Results show that median CVs remained globally unchanged (Figure 3), indicating that filters had only poor effects on precision. This is probably specific to our experiment since the variation among our three technical replicates was very low. Nonetheless, in some cases, extreme CV values decreased with filters, indicating that precision was particularly enhanced for proteins showing high abundance variations across replicates. This was especially true in the case of TOP3, when the occurrence filter was applied. As previously mentioned, because of the unequal peptide ionization potential, missing values can be an important source of variability between samples and thus between replicates for TOP3, Average and Average Log. The occurrence filter was thus expected to improve precision for these three quantification methods. In fact, TOP3 precision was more particularly affected by the occurrence filter because TOP3 is based on a reduced number of peptide ions. More generally, these results indicate that regarding precision, filters will be more beneficial to proteins quantified by few peptides ions than to proteins quantified by a high number of peptide ions.

Filters effects on accuracy depend on the quantification method

Accuracy is determined by the difference between observed and theoretical values. In the framework of absolute quantification, it is crucial to reach high accuracy to reliably estimate intracellular protein concentrations. Therefore, to evaluate the relative benefits of peptide filters on the performances of the different quantification methods in relative quantification, we examined accuracy of protein quantification. To do so in absence of a reference indicating the theoretical protein abundances expected at each UPS1 concentration, we used the property of equimolarity of UPS1 proteins and of yeast ribosomal proteins (50 of them were quantified in this study). We assumed equimolarity of ribosomal proteins based on the previous observation that the proteins involved in ribosomal complexes are present in one copy per isolated subunit (Kruiswijk et al., 1978). If accuracy is high, the estimated abundances within these two groups of proteins should present few dispersion. We therefore computed the CVs of protein abundances across UPS1 proteins and ribosomal proteins in each sample and used it as a proxy for accuracy. Results are presented in Figure 4.

In the case of UPS1 proteins, the shared peptide filter and the RT filter allowed to decrease the CVs especially for iBAQ, Average and TOP3 (Figure 4A). This is in agreement with our previous observation that the high intensity peptides removed by these two filters lead to overestimate protein abundances more strongly for iBAQ, Average and TOP3 than for Average Log and Model (Figure 2A, B). This result indicates that in terms of accuracy, Average Log and Model are less sensitive to irrelevant high intensity peptides than the other quantification methods. Regarding TOP3, the shared peptide filter did not particularly affect the CVs of abundances across UPS1 proteins because shared peptides were not always among the three most intense peptide ions. As a consequence, the number of proteins affected by the shared peptide filter was lower for TOP3 than for the other quantification methods. In the case of ribosomal proteins, we surprisingly observed that the shared peptide filter increased the CVs of abundances across proteins (Figure 4B), which indicates that taking into account shared peptides did not degrade equimolarity of ribosomal proteins. To explain this result, we relied on the fact that ribosomal proteins have highly conserved sequences (Lee and Traut, 1984) to assume that the peptides removed by the shared peptide filter were in fact all shared between ribosomal proteins in theoretically equal amounts. Under this hypothesis, the errors introduced by shared peptides on estimated abundances were the same for all ribosomal proteins.

The occurrence filter increased the CVs of abundances across UPS1 proteins (Figure 4A), indicating a detrimental effect on accuracy. This result was unexpected since in the particular case of our experimental design, the occurrence filter allowed to select peptide ions with high ionization potentials (Figure 2C). These peptides are indeed commonly admitted to be the most representative of the protein abundances (e.g. Worboys et al., 2014) based on the observation that the average intensity of the three most intense peptides per mole of protein was constant within a CV less than 10% (Silva et al., 2006). This observation has led to the development of TOP3 for absolute quantification (Silva et al., 2006). By contrast, the CVs of abundances across ribosomal proteins decreased, especially for Model (Figure 4B). In the case of ribosomal proteins, peptide ions with low ionization potential were not as massively removed as for UPS1 proteins. Thus, we supposed that these peptide ions were involved in UPS1 accuracy. To confirm this hypothesis, we restrained our experimental design to a UPS1 concentration range that was more representative of a natural dynamic range (0.5 to 27.9 fmole.μl^{-1}). In these conditions, the UPS1 peptides with low ionization potentials had much less missing values, such that they were no more removed by the occurrence filter. This time, we observed, as for ribosomal proteins, that the CVs of abundances across UPS1 proteins decreased after the occurrence filter,

and more particularly for Model (Figure 4A). Altogether, these results show that decreasing the number of valuable peptide ions to compute protein abundance negatively affects accuracy. They also indicate that the benefit of the occurrence filter on accuracy was higher for Model than for the other quantification methods. This is probably because for peptide ions showing many missing values, the amount of data is too low to robustly estimate the P_k term (see Materials and Methods).

When applied on UPS1 peptide ions, the outliers filter had contrasting effects depending on the quantification method (Figure 4A). No particular effect was observed for iBAQ and Average, while accuracy was clearly improved for Average Log and Model in both the whole and the restrained experimental designs. This is due to the fact that the peptide ions removed by the outliers filter were generally of low intensity (Figure S-3). As previously mentioned, peptide ions of low intensity have less weight in iBAQ and Average than in Average Log and Model. The outliers filter was not applied for ribosomal proteins because this filter is not relevant when all the proteins are in constant amounts across samples.

Relative benefits of filters on relative protein quantification

In the framework of relative quantification, accuracy can be neglected as long as the errors between observed and theoretical values are similar in all samples. If this is not the case, the response of UPS1 abundances to increasing UPS1 concentration would be affected. This response is expected to be linear of the type $y_i = ax_i$ where y_i is the estimated protein abundance at UPS1 concentration x_i and a is a constant. For convenience of representation, data can be log-transformed. In this case, the response is expected to be linear of the type $\log(y_i) = \log(a) + \log(x_i)$, with a slope equal to one. To evaluate to which extend peptide filters improved the performances of the quantification methods in relative quantification, we examined the estimated values of slope (Figure 5A) and the coefficients of determination (R^2, Figure 5B) of linear regressions between the log-transformed abundances obtained experimentally for UPS1 proteins and their spiked log-transformed concentrations in both the whole and restrained experimental designs.

The three filters, shared peptide, RT and outliers, all improved the slope and R^2 regardless the quantification method (Figure 5). This was expected given that these three filters removed peptide ions displaying non-linear responses to increasing UPS1 concentrations (Figures 2A, B, C). The RT filter improved more particularly TOP3 linearity which, in absence of any filter,

was worse than for the other quantification methods because TOP3 is based on a reduced number of peptide ions.

The occurrence filter improved the response to increasing UPS1 concentrations especially for Average and Average Log, which indicates that in terms of relative quantification, these two methods were more susceptible to missing values than the other quantification methods. This result is in agreement with the previous observation that in the case of Average and Average Log, the occurrence filter lead to reduce undesired between-samples variability (Figure 2C). This was not the case for Model because the peptide ionization potential was taken into account in the abundance computation. Model has therefore a great advantage over Average and Average Log for relative quantification.

Conclusions

Altogether, these results illustrate that filters can have significant effects on protein abundances, even if only a few peptide ions are removed, and that filters can have contrasting effects depending on the quantification method. They also show that filters have to be carefully think since valuable information may be unintentionally lost. In the present study, we indeed showed that applying the occurrence filter in the particular case of our experimental design lead to remove many peptide ions with low ionization potential that correctly responded to increasing UPS1 concentrations. These peptide ions could be worth considering for protein quantification, provided that missing data are appropriately handled.Because TOP3 is based on a lower number of peptide ions used than the other quantification methods, its precision will be more affected by missing values. This result indicates that in terms of accuracy, Average Log and Model are less sensitive to irrelevant high intensity peptides than the other quantification methods.

Altogether, these results show that in terms of accuracy, quantification methods based on log-transformed intensities are less sensitive to irrelevant high intensity peptides and that iBAQ and TOP3 are less sensitive to missing values. They also highlight that decreasing the number of valuable peptide ions to compute protein abundance negatively affects accuracy.

We synthesized the effects of filters on the performances of the different quantification methods in absolute and relative quantification in Figure 6.

Despite for iBAQ, in the full range, the relative and absolute accuracy were improved by the four filters (Figure 6A). Average and iBAQ absolute accuracy were the most improved by the shared peptide filter certainly because these methods are more sensitive to highest intense

peptides. Apart for the Model, the RT filter has a slight effect on the precision but improved the linearity. A less permissive threshold than 30 sec, the one we used in this analysis, should lead to a more drastic effect of the RT filter. The occurrence filter particularly improved the relative accuracy of Average, Average-Log and TOP3 without drastically improving the absolute accuracy, excepting for TOP3. The outliers filter improved the relative and the absolute accuracy for all the methods except for iBAQ for which the absolute accuracy decreased.

In the narrow range, the absolute accuracy was reduced without filtering (Figure 6B) compared to the full-range. Furthermore, the absolute and relative accuracy were less drastically improved by filters than in the full-range. In the narrow range, less peptides ions with unequal ionization potential were removed underlying their valuable role on both absolute and relative accuracy. However, as in the full range an excessive filtering damage the precision of quantification based on iBAQ (Figure 6).

The experimental design should be considered before applying the occurrence and outliers filters. Indeed, in time-series and silencing experiments the occurrence filter will lead to the remove of valuable peptide ions and probably proteins because of their lowest and shut-down detection, respectively. Otherwise, the outliers filter should be carefully used in the case of proteins with constant expression.

In this paper, we described five methods of quantification according to the peptides dataset used. For each method of quantification, limitations with more or less impact on the accuracy were identified confirming why none consensus to which method is the best to use. However, Model was demonstrated to be a robust method as it achieved good performances in term of relative and absolute accuracy after only the shared peptides filter (Figure 6). This results from the unique capability of the Model to correct source of variability such as the peptides effect.

In perspective, it could be interesting to redo this analysis with modifying the thresholds used on the RT, occurrence and the outliers filters. It could also be interesting to evaluate the effect of the five filters each one separated from others (on independent analysis).

Table 1 Data composition: peptides ions and total proteins of the normalized unfiltered dataset (No filter) and after application of shared peptide, RT, occurrence and outliers filters. Numbers in parenthesis indicate the percentage of data removed by the filter from the previous dataset.

	Peptide ions			Total proteins		
Filter	Total	Yeast	UPS1	Total	Yeast	UPS1
None	22950	21820	1138	2080	2039	41
Shared peptide	22044 (-3.9%)	20915 (-4.2%)	1129 (-0.8%)	2046 (-1.6%)	2005 (-1.7%)	41 (-0%)
RT	21857 (-0.8%)	20778 (-0.7%)	1079 (-4.4%)	2041 (-0.2%)	2000 (-0.3%)	41 (-0%)
Occurrence	13561 (-38.0%)	13314 (-35.9%)	247 (-77.1%)	1491 (-26.9%)	1455 (-21.3 %)	36 (-12.2%)
Outliers	4882 (-64.0%)	4666 (-65.0%)	216 (-12.6%)	1008 (-32.4%)	973 (-33.1%)	35 (-2.8%)

Figure 1 Schema of the experimental design. Dataset 1 derived from the normalization of the raw dataset (Dataset 0), Dataset 2 derived from normalized Dataset 0 without shared peptides (Shared peptide filter). In Dataset 3, peptides with a standard deviation of retention time higher than 30 seconds were removed before normalization and then shared peptides were removed (RT filter). In Dataset 4, peptide ions presenting more than 15.15% of missing values were filtered out (Occurrence filter) from Dataset 3. In Dataset 5, uncorrelated peptide ions (Pearson, $R^2>0.8$, p-value < 0.01) were filtered out (Ooutliers filter)

Figure 2 Examples of effect of filters on peptides selection (left panel) and quantification of protein abundance (right panel). Filters were illustrated on four UPS1 proteins. Filter on shared peptides on P62988 protein (A), RT filter on P63165 protein (B), occurrence filter on P02144 protein (C) and outliers filter on P02787 protein (D). Peptides filtered out are colored in red (left panel). Five methods of protein abundance quantification based on the integration of peptides intensity are used (right panel): iBAQ (black), TOP3 (red), Average (blue), AverageLog (purple) and Model (orange). Full and dashed lines indicate protein quantification obtained with all peptides (blue and red lines in left panel) and with kept peptides after filtering (blue lines in left panel). At each concentration, and for each method the average of three technical replicates were shown (sd not shown). Peptide ions intensity are log10 transformed and concentrations of UPS (fmol.µl-1) are log10 scaled.

Figure 3 Effect of the four filters on the precision of the five methods of quantification of UPS proteins (iBAQ, TOP3, Average, AverageLog and Model). Precision was estimated by the CV (%) of protein abundance between the three technical replicates for each UPS1 proteins concentration. Only UPS1 proteins detected in all filtered datasets were used. Precision was estimated on the initial non-filtered normalized dataset (None) and after the shared peptides (Shared peptide), RT (RT), occurrence (Occurrence) and outliers (Outliers) filter.

Figure 4 Effect of the four filters on the absolute accuracy. Absolute accuracy was estimated by the CV (%) determined between proteins abundances of UPS1 proteins (A) and ribosomal yeast (B) proteins. Proteins abundance was quantified by iBAQ, TOP3, Average, Average-Log and Model methods. Only UPS1 and ribosomal proteins detected in all filtered datasets were used. Absolute accuracy was estimated on the initial non-filtered normalized dataset (None) and after the shared peptides (Shared peptide), RT (RT), occurrence (occurrence) and outliers (Outliers) filter. For UPS1 proteins, the absolute accuracy was calculated on the full range

(0.04-141.1 fmol. μl^{-1}, red) and on a narrow range of UPS1 concentrations (0.5-27.9 fmol. μl^{-1}, blue).

Figure 5 Effect of the four filters on the relative accuracy. Relative accuracy was estimated by the coefficient of determination (R^2) (A) and the slope (B) of the linear regression between the abundances obtained experimentally for UPS1 proteins and their spiked concentrations. UPS1 proteins abundance was quantified by iBAQ, TOP3, Average, Average-Log and Model methods. Only UPS1 proteins detected in all filtered datasets were used. Linear regressions were performed on the initial non-filtered normalized dataset (None) and after the shared peptides (Shared peptide), RT (RT), occurrence (occurrence) and outliers (Outliers) filter. The relative accuracy was calculated on the full range (0.04-141.1 fmol.µl-1, red boxplots) and on a narrow range of UPS1 concentrations (0.5-27.9 fmol.µl-1, blue boxplots).

Figure 6 Relation between relative (R^2) and absolute accuracy (CV (%)) for each method of quantification. Only medians of CV (%) between UPS1 protein abundance versus medians of R^2 of the linear regression between estimated and spiked UPS1 protein abundance are displayed. Only UPS1 proteins detected in all filtered datasets were used. UPS1 proteins abundance was quantified by iBAQ (black line), TOP3 (red line), Average (blue line), Average-Log (purple line) and Model (orange line) methods. Numbers refer to the dataset used: 1 corresponding to the initial non-filtered normalized dataset (None filter), 2 to the shared peptides filtered dataset (Shared peptides filter), 3 to the RT filtered dataset (RT filter), 4 to the occurrence filtered dataset (Occurrence filter) and 5 to the outliers filtered dataset (Outliers filter). Relation between relative (R^2) and absolute accuracy (CV (%)) was performed with results obtained on the full range (0.04-141.1 fmol. μl^{-1}, A) and on the narrow range (0.5-27.9 fmol.µl-1, B).

References

Blein-Nicolas M, Xu H, de Vienne D, Giraud C, Huet S, Zivy M (2012) Including shared peptides for estimating protein abundances: A significant improvement for quantitative proteomics. Proteomics **12**: 2797–2801

Blein-Nicolas M, Zivy M (2016) Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics. Biochim Biophys Acta - Proteins Proteomics **1864**: 883–895

Bukhman Y V, Dharsee M, Ewing ROB, Chu P, Topaloglou T, Bihan TLE, Goh T, Duewel H, Stewart IANI, Wisniewski JR, et al (2008) DESIGN AND ANALYSIS OF QUANTITATIVE DIFFERENTIAL PROTEOMICS INVESTIGATIONS USING LC-MS TECHNOLOGIES. J Bioinform Comput Biol **6**: 107–123

Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, et al (2012) A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol **30**: 918–920

Chelius D, Bondarenko P V (2002) Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. J Proteome Res **1**: 317–23

Clough T, Key M, Ott I, Ragg S, Schadow G, Vitek O (2009) Protein Quantification in Label-Free LC-MS Experiments. J Proteome Res **8**: 5275–5284

Daly DS, Anderson KK, Panisko EA, Purvine SO, Fang R, Monroe ME, Baker SE (2008) Mixed-Effects Statistical Model for Comparative LC−MS Proteomics Studies. J Proteome Res **7**: 1209–1217

Forshed J, Johansson HJ, Pernemalm M, Branca RMM, Sandberg A, Lehtiö J (2011) Enhanced Information Output From Shotgun Proteomics Data by Protein Quantification and Peptide Quality Control (PQPQ). Mol Cell Proteomics **10**: M111.010264

Higgs RE, Knierman MD, Gelfanova V, Butler JP, Hale JE (2005) Comprehensive label-free method for the relative quantification of proteins from biological samples. J Proteome Res **4**: 1442–1450

Karpievitch Y, Stanley J, Taverner T, Huang J, Adkins JN, Ansong C, Heffron F, Metz TO, Qian W, Yoon H, et al (2009) A statistical framework for protein quantitation in bottom-up MS-based proteomics. **25**: 2028–2034

Kruiswijk T, Planta RJ, Mager WH (1978) Quantitative analysis of the protein composition of yeast ribosomes. Eur J Biochem **83**: 245–252

Lai X, Wang L, Tang H, Witzmann FA (2011) A Novel Alignment Method and Multiple Filters for Exclusion of Unqualified Peptides To Enhance Label-Free Quantification Using Peptide Intensity in LC—MS/MS. J Proteome Res **10**: 759–785

Lee JC, Traut RR (1984) Proximity of 5.8 S RNA-binding proteins and A-site proteins in yeast ribosomes inferred from cross-linking. J Biol Chem **259**: 9971–4

Lyutvinskiy Y, Yang H, Rutishauser D, Zubarev RA (2013) *In Silico* Instrumental Response Correction Improves Precision of Label-free Proteomics and Accuracy of Proteomics-based Predictive Models. Mol Cell Proteomics **12**: 2324–2331

Millan-Oropeza A, Henry C, Blein-Nicolas M, Aubert-Frambourg A, Moussa F, Bleton J, Virolle M-J (2017) Quantitative Proteomics Analysis Confirmed Oxidative Metabolism Predominates in *Streptomyces coelicolor* versus Glycolytic Metabolism in *Streptomyces lividans*. J Proteome Res **16**: 2597–2613

Oberg AL, Vitek O (2009) Statistical Design of Quantitative Mass Spectrometry-Based Proteomic Experiments. J Proteome Res **8**: 2144–2156

Podwojski K, Eisenacher M, Kohl M, Turewicz M, Meyer HE, Rahnenführer J, Stephan C (2010) Peek a peak: a glance at statistics for quantitative label-free proteomics. Expert Rev Proteomics **7**: 249–261

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. Nature **473**: 337–342

Silva JC, Gorenstein M V., Li G-Z, Vissers JPC, Geromanos SJ (2006) Absolute Quantification of Proteins by LCMS ^{E}. Mol Cell Proteomics **5**: 144–156

Valot B, Langella O, Nano E, Zivy M (2011) MassChroQ: A versatile tool for mass spectrometry quantification. Proteomics **11**: 3572–3577

Voyksner RD, Lee H (1999) Investigating the use of an octupole ion guide for ion storage and high-pass mass filtering to improve the quantitative performance of electrospray ion trap mass spectrometry. Rapid Commun Mass Spectrom **13**: 1427–1437

Webb-Robertson BJM, McCue LA, Waters KM, Matzke MM, Jacobs JM, Metz TO,

Varnum SM, Pounds JG (2010) Combined statistical analyses of peptide intensities and peptide occurrences improves identification of significant peptides from MS-based proteomics data. J Proteome Res **9**: 5748–5756

Worboys JD, Sinclair J, Yuan Y, Jørgensen C (2014) Systematic evaluation of quantotypic peptides for targeted analysis of the human kinome. Nat Methods **11**: 1041–1044

Zhang B, Pirmoradian M, Zubarev R, Käll L (2017) Covariation of Peptide Abundances Accurately Reflects Protein Concentration Differences. Mol Cell Proteomics 1–42

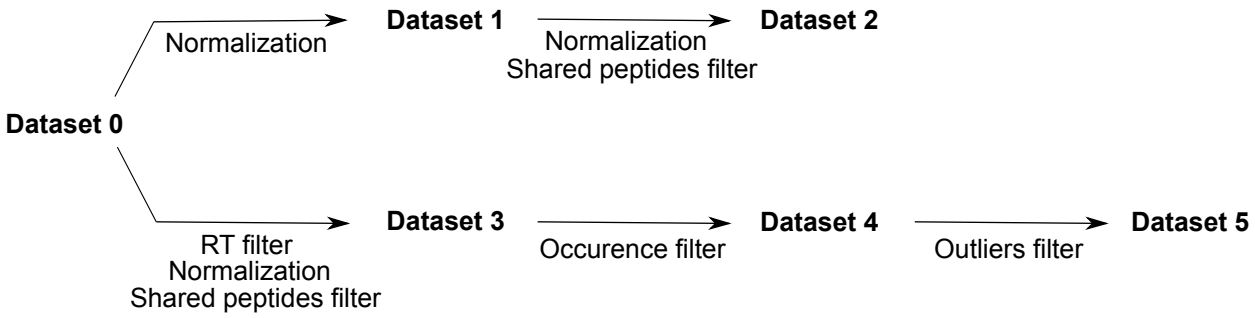

Dataset 0
Normalization
Dataset 1
Normalization
Shared peptides filter
Dataset 2
RT filter
Normalization
Shared peptides filter
Dataset 3
Occurence filter
Dataset 4
Outliers filter
Dataset 5

Figure 1

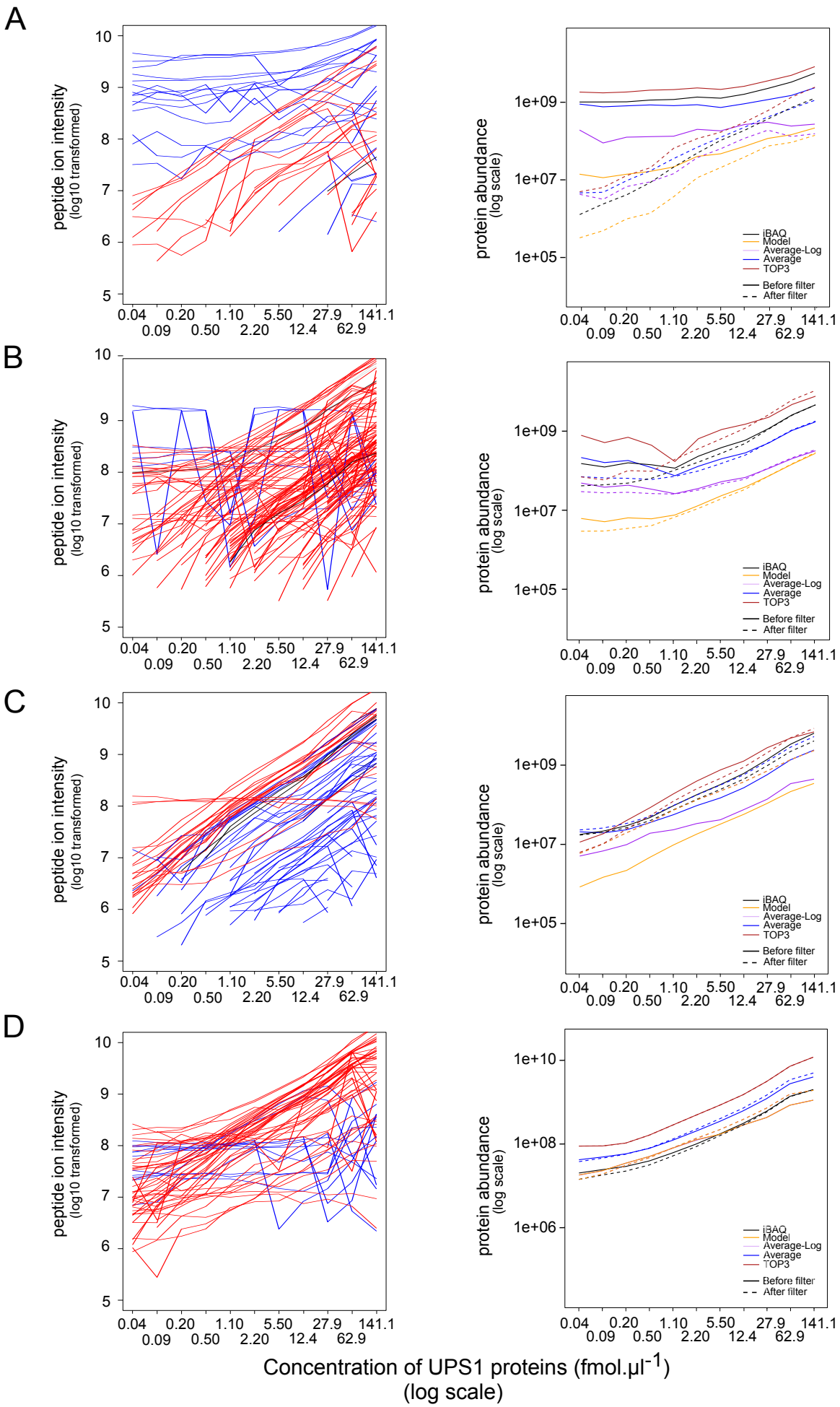

A
B
C
D
peptide ion intensity
(log10 transformed)
protein abundance
(log scale)
iBAQ
Model
Average-Log
Average
TOP3
Before filter
After filter
0.04 0.09 0.20 0.50 1.10 2.20 5.50 12.4 27.9 62.9 141.1
Concentration of UPS1 proteins (fmol.µl-1)
(log scale)

Figure 2

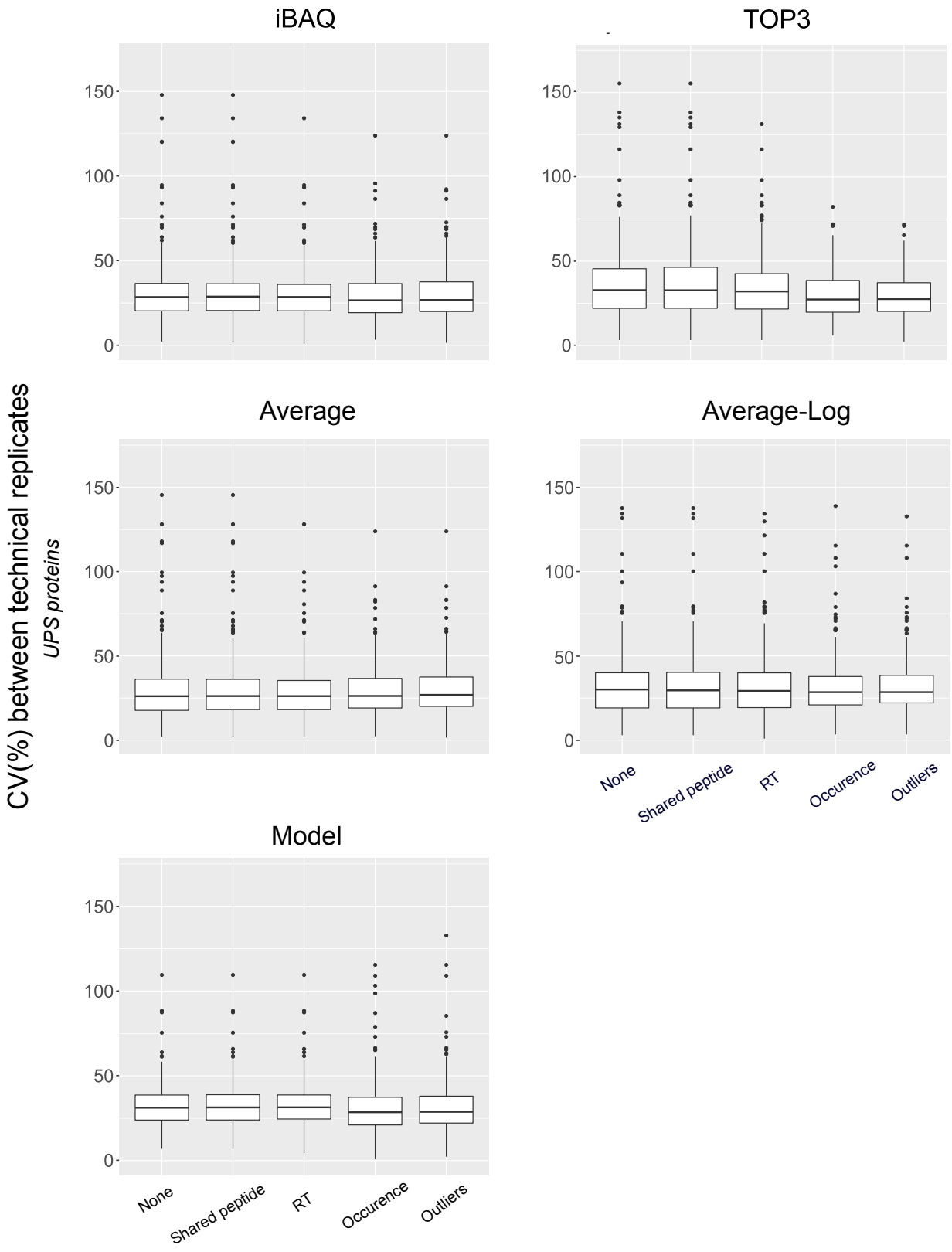

iBAQ
TOP3
Average
Average-Log
Model
CV(%) between technical replicates
UPS proteins
150
100
50
0
None
Shared peptide
RT
Occurence
Outliers

Figure 3

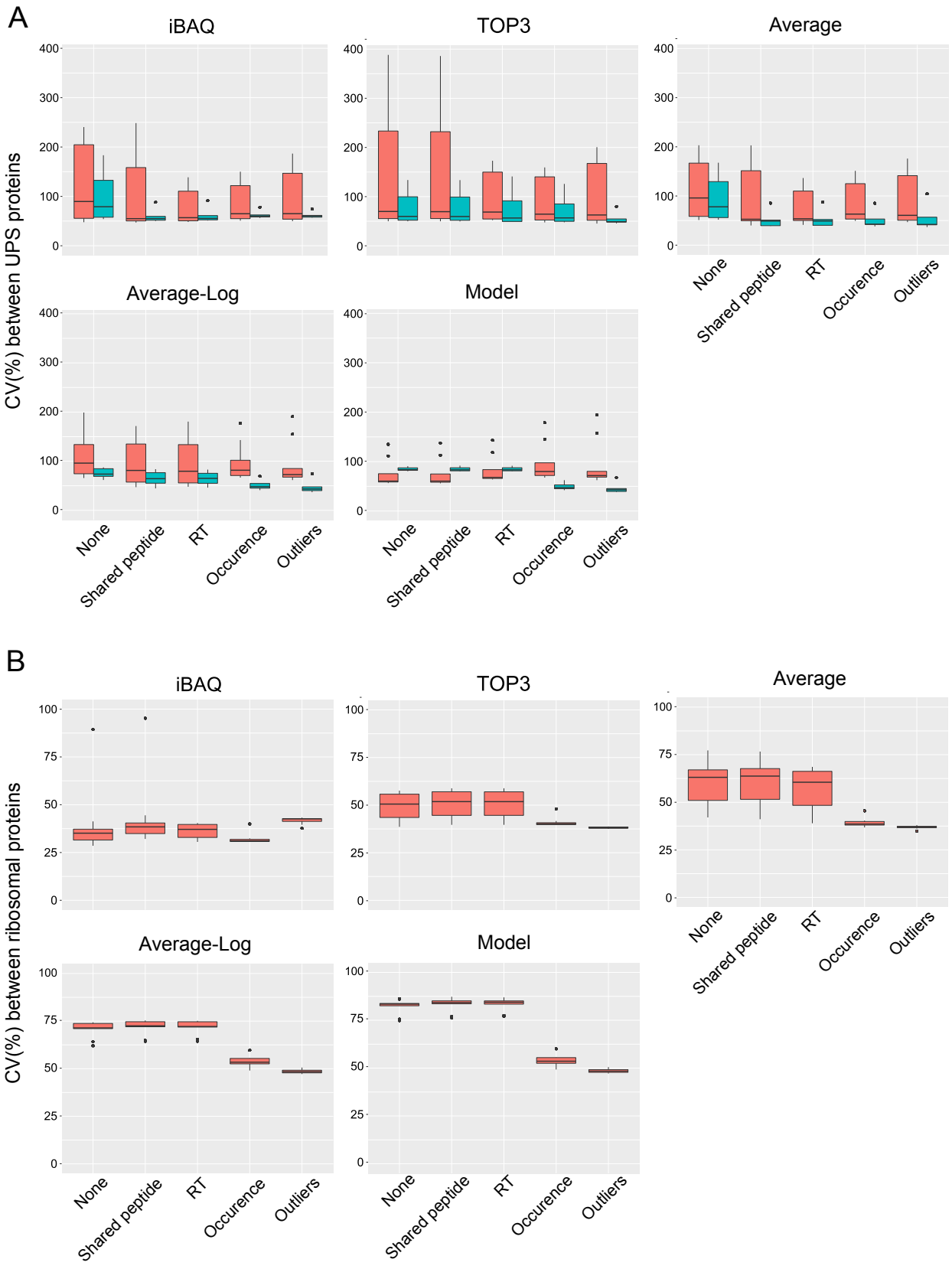

A
iBAQ
TOP3
Average
Average-Log
Model
CV(%) between UPS proteins
0
100
200
300
400
None
Shared peptide
RT
Occurence
Outliers
B
iBAQ
TOP3
Average
Average-Log
Model
CV(%) between ribosomal proteins
0
25
50
75
100
None
Shared peptide
RT
Occurence
Outliers

Figure 4

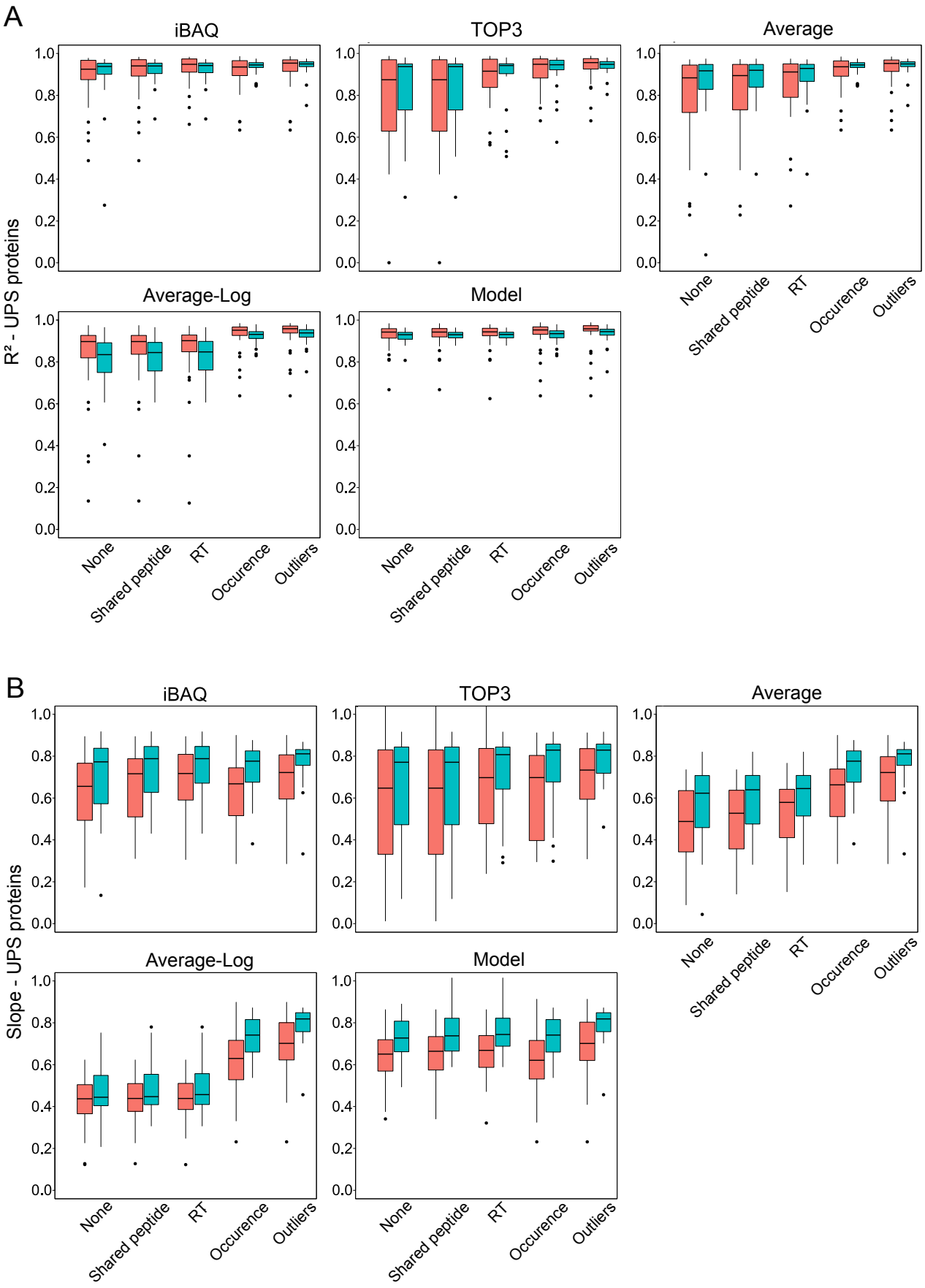
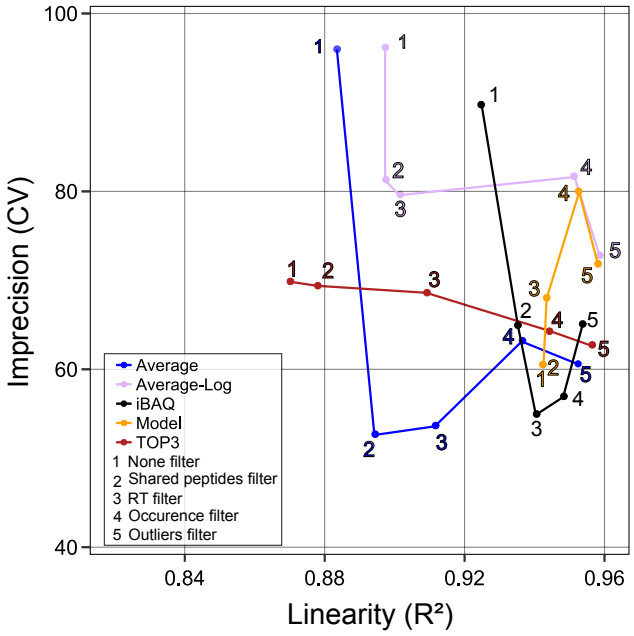

A
iBAQ
TOP3
Average
Average-Log
Model
R² - UPS proteins
1.0
0.8
0.6
0.4
0.2
0.0
None
Shared peptide
RT
Occurence
Outliers
B
iBAQ
TOP3
Average
Average-Log
Model
Slope - UPS proteins
1.0
0.8
0.6
0.4
0.2
0.0
None
Shared peptide
RT
Occurence
Outliers

Figure 5

A


Imprecision (CV)
Linearity (R²)
100
80
60
40
0.84
0.88
0.92
0.96
Average
Average-Log
iBAQ
Model
TOP3
1 None filter
2 Shared peptides filter
3 RT filter
4 Occurence filter
5 Outliers filter

B

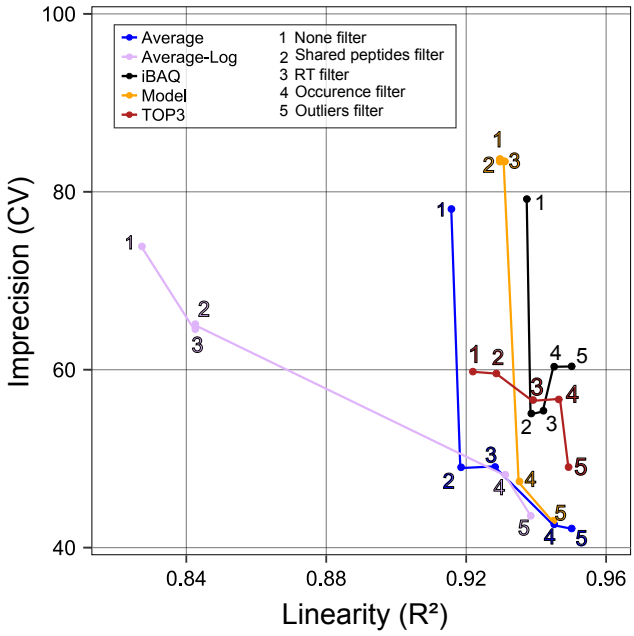

Imprecision (CV)
Linearity (R²)
100
80
60
40
0.84
0.88
0.92
0.96
Average
Average-Log
iBAQ
Model
TOP3
1 None filter
2 Shared peptides filter
3 RT filter
4 Occurence filter
5 Outliers filter

Figure 6

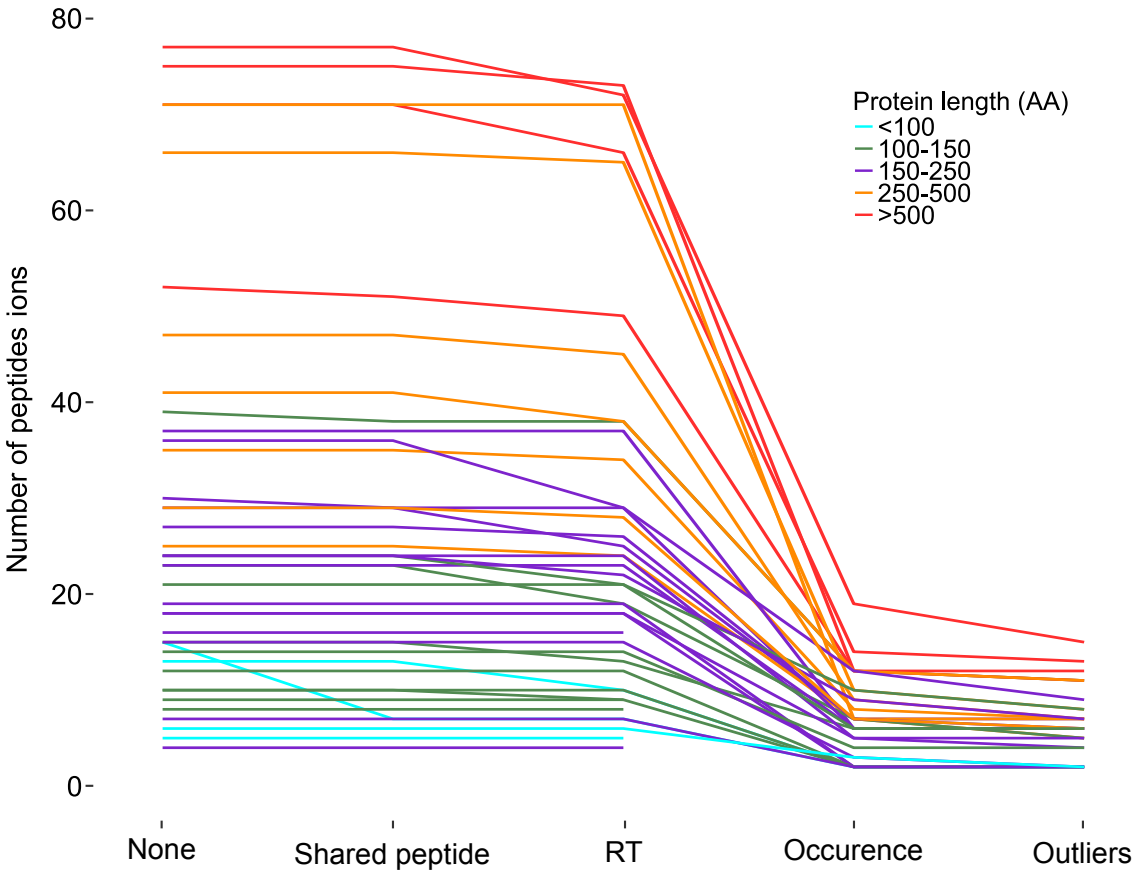


Figure S-1 Longest UPS1 proteins displayed the largest number of detected peptides. Length of UPS1 proteins (in Amino Acid, AA) were seperated in 5 groups: <100 AA (blue-colored curve), 100-150 AA (green-colored curve), 150-250 AA (purple-colored curve), 250-500 AA (orange-colored curve) and >500 AA (red-colored curve). The total number of peptides per protein was counted in the normalized unfiltered dataset (None) and after each filter: shared peptide filter, RT filter, occurence filter and outliers filter.

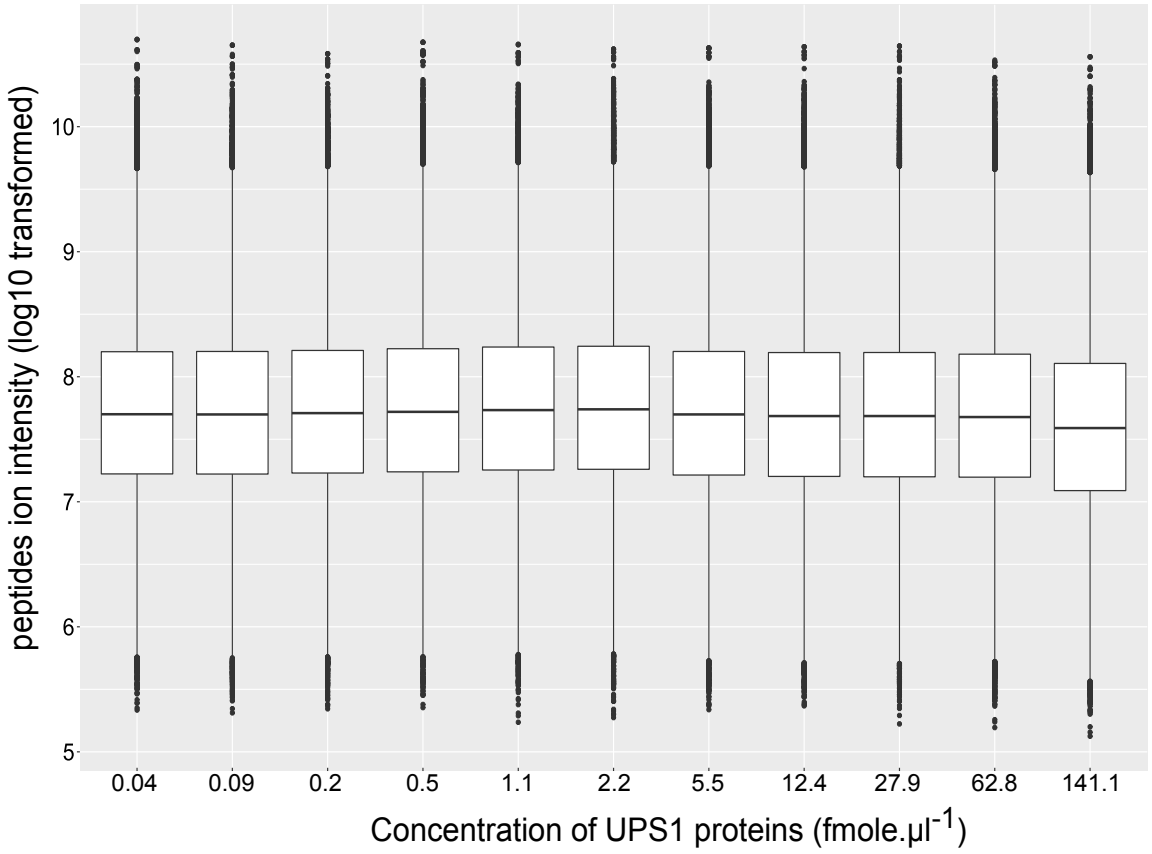
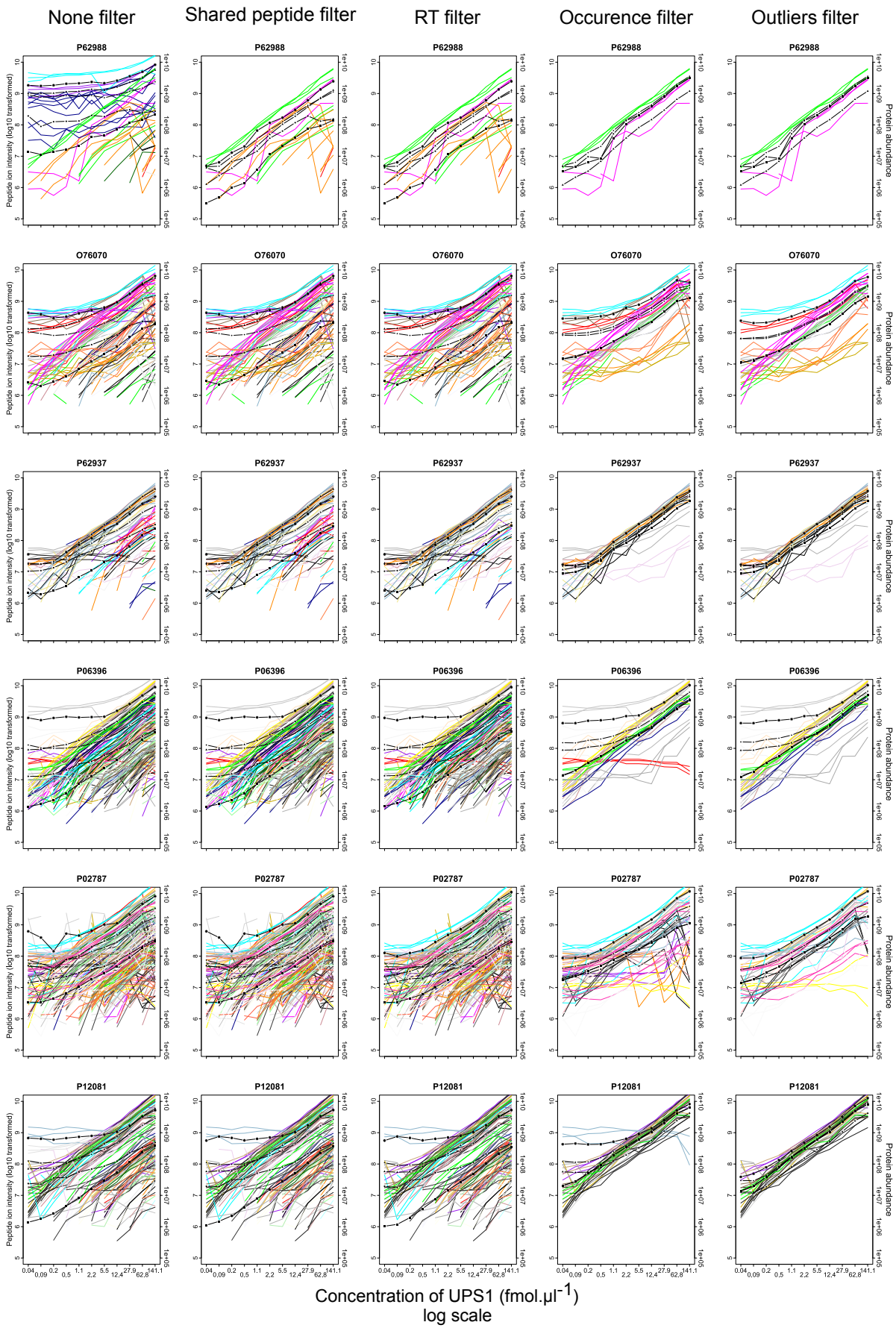
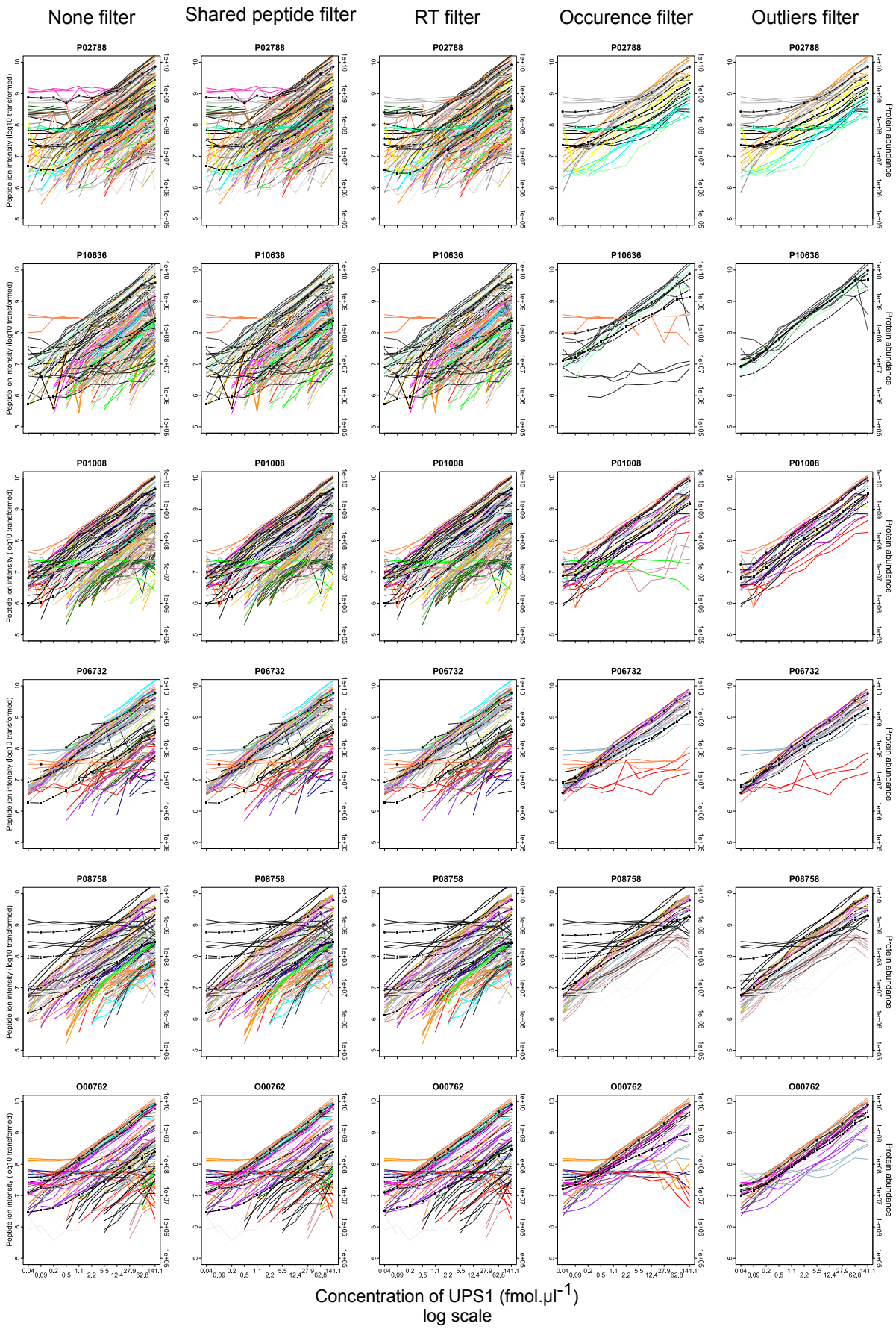
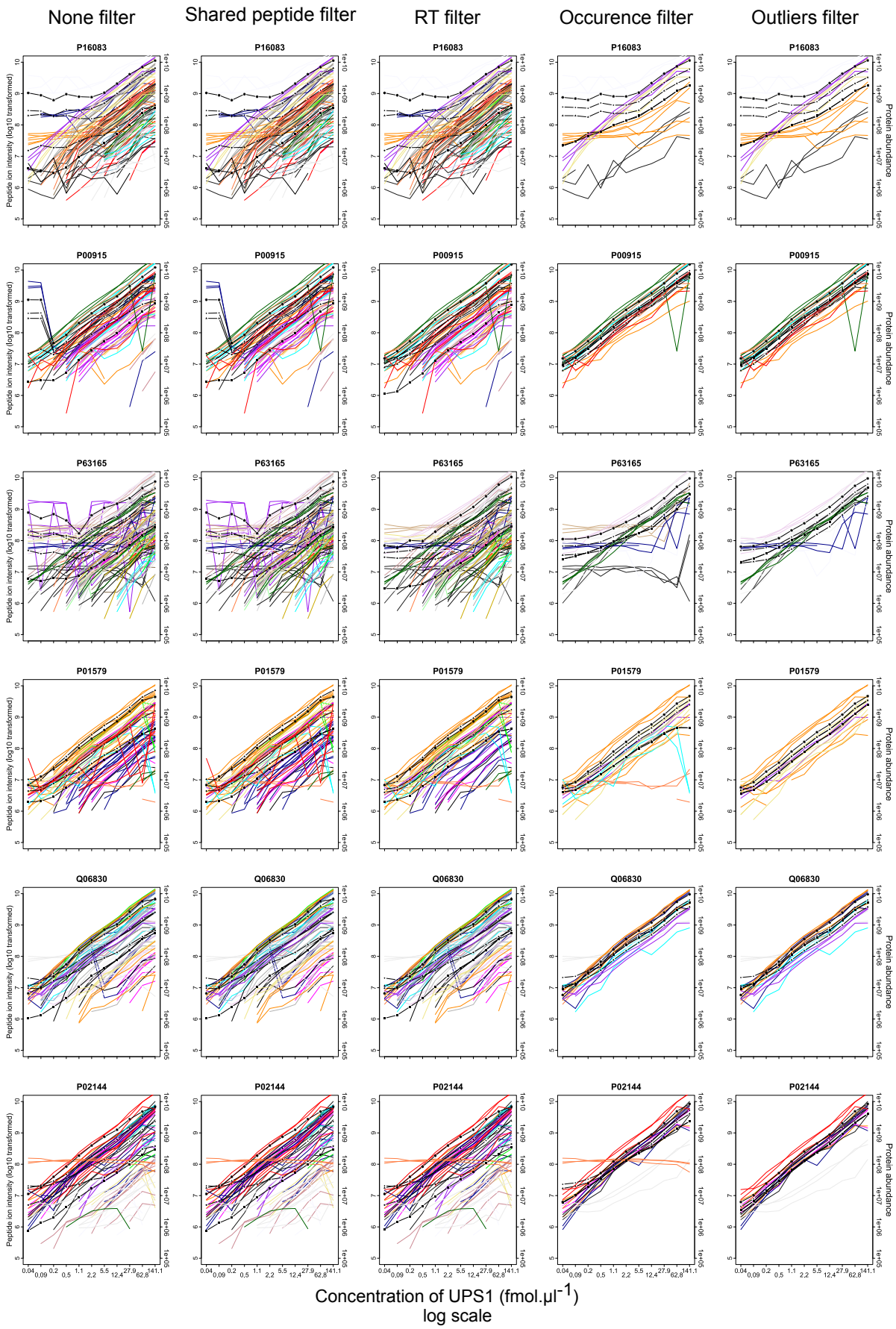
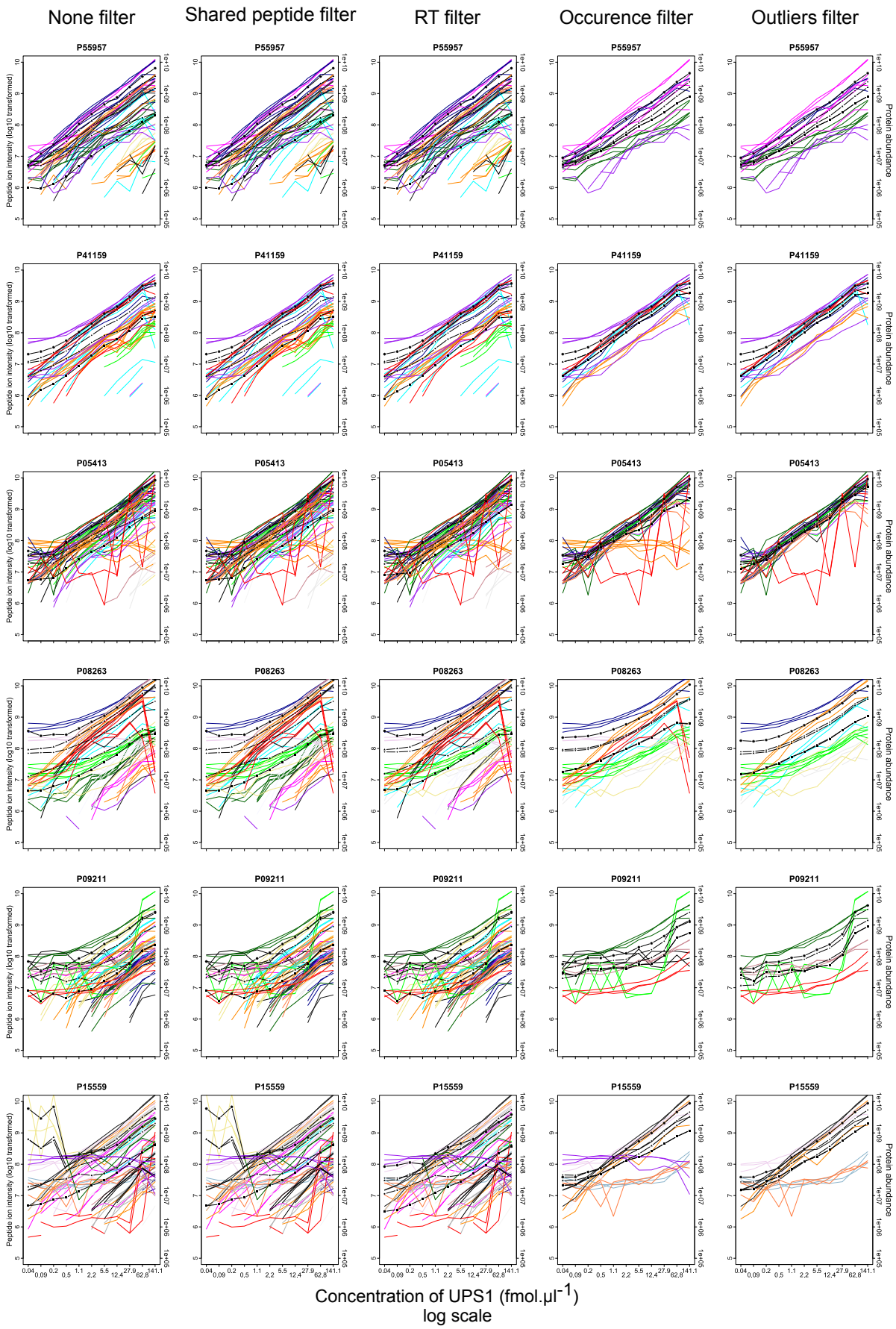


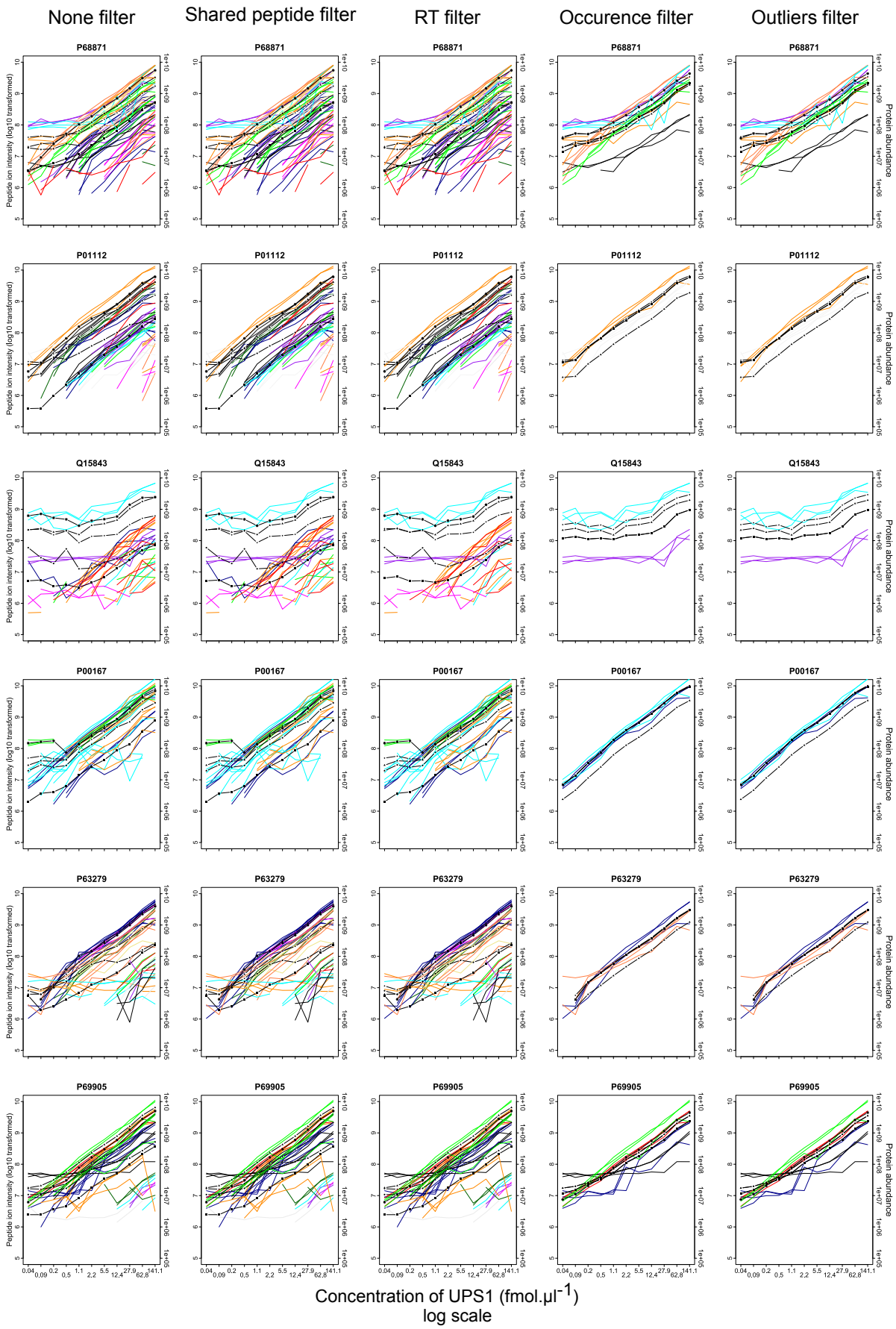
Figure S-2: Distribution of yeast peptide ions intensity at each concentration of UPS1 proteins. Yeast peptides ions intensity used for this plot provided from the Dataset 0.

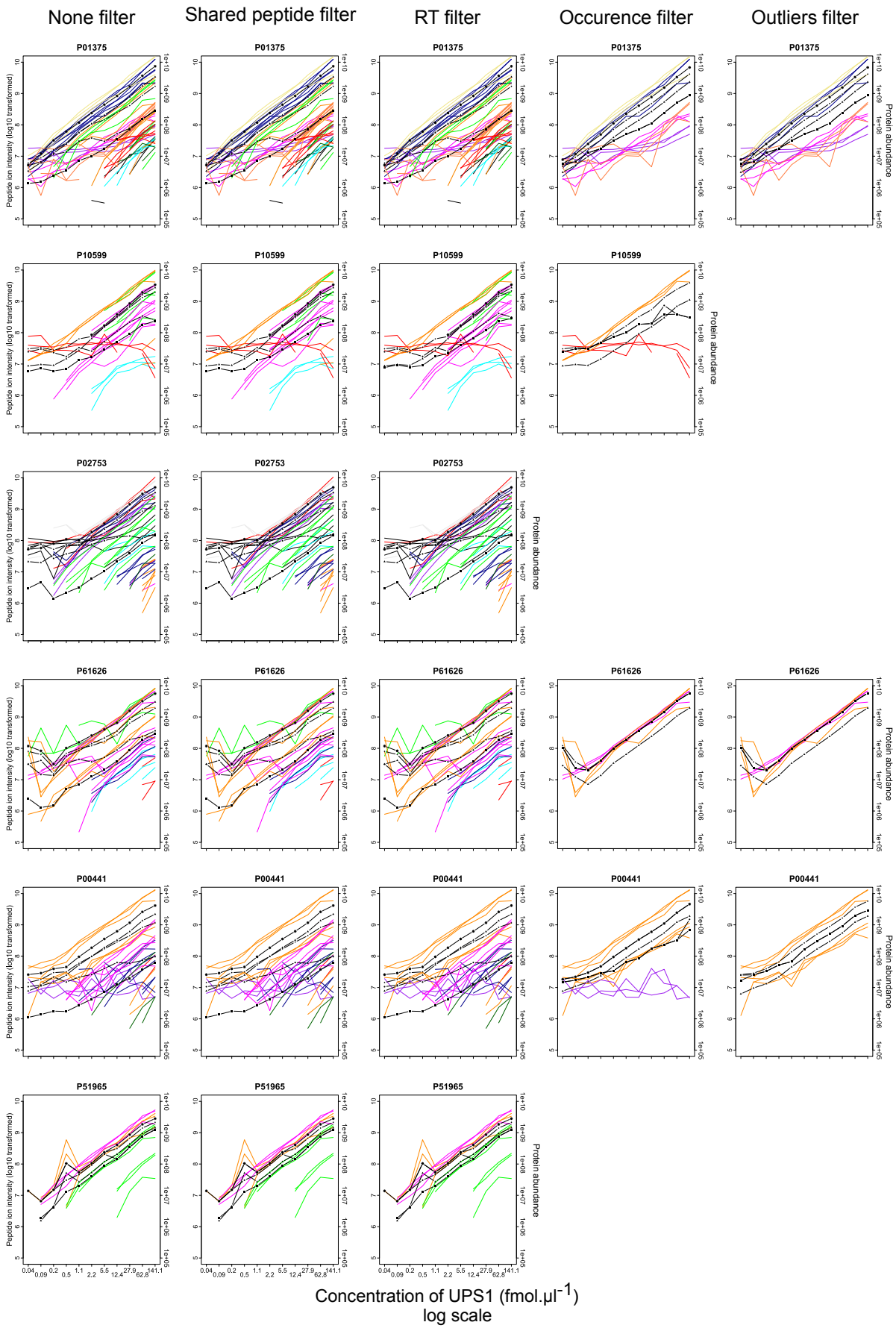

None filter
Shared peptide filter
RT filter
Occurence filter
Outliers filter
P62988
O76070
P62937
P06396
P02787
P12081
Peptide ion intensity (log10 transformed)
Protein abundance
Concentration of UPS1 (fmol.µl-1)
log scale


None filter
Shared peptide filter
RT filter
Occurence filter
Outliers filter
P02788
P10636
P01008
P06732
P08758
O00762
Peptide ion intensity (log10 transformed)
Protein abundance
Concentration of UPS1 (fmol.µl-1)
log scale


None filter
Shared peptide filter
RT filter
Occurence filter
Outliers filter
P16083
P00915
P63165
P01579
Q06830
P02144
Peptide ion intensity (log10 transformed)
Protein abundance
Concentration of UPS1 (fmol.µl-1)
log scale


None filter
Shared peptide filter
RT filter
Occurence filter
Outliers filter
P55957
P41159
P05413
P08263
P09211
P15559
Peptide ion intensity (log10 transformed)
Protein abundance
0.04
0.09
0.2
0.5
1.1
2.2
5.5
12.4
27.9
62.8
141.1
Concentration of UPS1 (fmol.µl-1)
log scale


None filter
Shared peptide filter
RT filter
Occurence filter
Outliers filter
P68871
P01112
Q15843
P00167
P63279
P69905
Peptide ion intensity (log10 transformed)
Protein abundance
0.04
0.09
0.2
0.5
1.1
2.2
5.5
12.4
27.9
62.8
141.1
Concentration of UPS1 (fmol.µl-1)
log scale


None filter
Shared peptide filter
RT filter
Occurence filter
Outliers filter
P01375
P10599
P02753
P61626
P00441
P51965
Peptide ion intensity (log10 transformed)
Protein abundance
0.04
0.09
0.2
0.5
1.1
2.2
5.5
12.4
27.9
62.8
141.1
Concentration of UPS1 (fmol.µl-1)
log scale

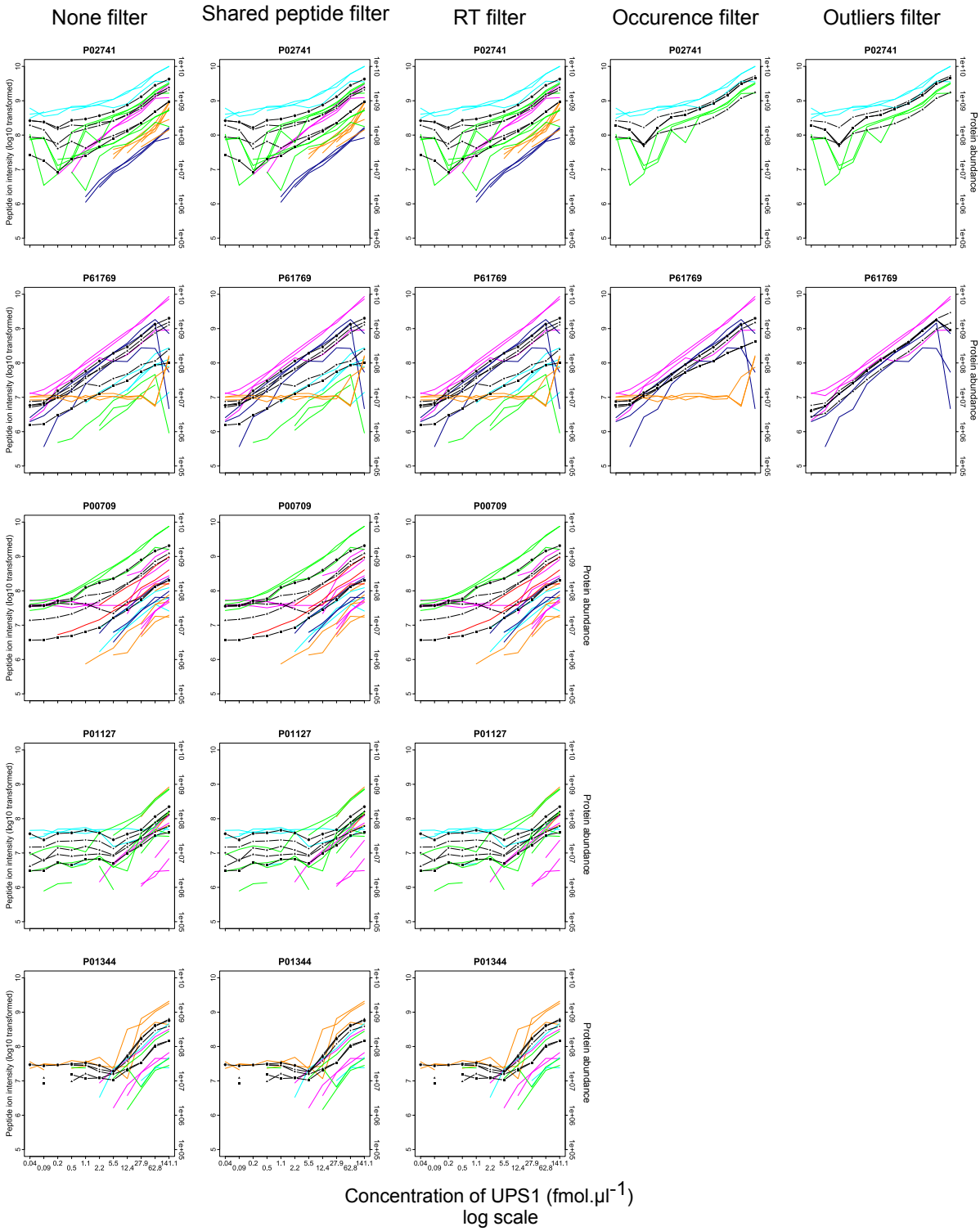


Figure S-3 Influence of peptide ions filtering on the 41 UPS1 protein abundances. Five methods of quantification of protein abundance (right y-axis) based on the integration of peptides intensity (log10 transformed, left y-axis) are used: iBAQ (▲), Top3 (*), Average (♦), AverageLog (●) and Model (■). Each row corresponds to one UPS1 protein named on the top of the row' plots. Each column correspond to one filter starting with the none-filter situation and then shared peptides filter, RT filter, occurence filter and outliers filter. At the each concentration, and for each method a symbol is an average of three technical replicates (sd not shown). Concentrations of UPS (fmol.µl-1) are log10 scaled. For each protein, random colours distinguish peptides.