



HAL
open science

Spatiotemporal distribution of ploidy levels and ploidy specific transcriptome during Tomato fruit development

Edouard Tourdot

► **To cite this version:**

Edouard Tourdot. Spatiotemporal distribution of ploidy levels and ploidy specific transcriptome during Tomato fruit development. *Vegetal Biology*. Université de Bordeaux, 2021. English. NNT : 2021BORD0121 . tel-03230734

HAL Id: tel-03230734

<https://theses.hal.science/tel-03230734>

Submitted on 20 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE
DOCTEUR DE
L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE N°154 : SCIENCES DE LA VIE ET DE LA SANTE
SPÉCIALITÉ : BIOLOGIE VEGETALE

Par Edouard TOURDOT

**Spatiotemporal distribution of ploidy levels and ploidy
specific transcriptome during Tomato fruit
development**

Sous la direction de : Nathalie GONZALEZ

Soutenue le 29/03/2021

Membres du jury :

M. HERNOULD, Michel	Professeur Université de Bordeaux	Président du jury
Mme. BAROUX, Célia	Professeur University of Zürich	Rapporteur
M. DE VEYLDER, Lieven	Directeur de Recherche VIB-UGent Center	Rapporteur
Mme. COOKSON-WODRICH, Sarah Jane	Chargé de Recherche INRAE	Examineur
M. ZOUINE, Mohamed	Chargé de Recherche INRAE/INP-ENSAT	Examineur
Mme. GONZALEZ, Nathalie	Directrice de recherche INRAE -	Directrice de Thèse

Remerciements

Before starting my acknowledgement, I would like to thank the jury members, Célia Baroux, Sarah Cookson, Lieven De Veylder, Mohamed Zouine and Michel Hernould.

J'aimerais remercier l'équipe Organogénèse du Fruit et endoreduplication (maintenant nommée : Floraison, Développement du Fruit et Contraintes Environnementales) de m'avoir accueilli comme doctorant et aux membres de l'équipe de m'avoir accepté.

Merci Nathalie (G) d'avoir acceptée de travailler avec moi pendant cette thèse, d'avoir fait preuve de tant de patience et de bienveillance avec moi (bienveillance, qui sera je pense le mot qui dictera cette expérience). Merci d'avoir rendu possible tout le travail que j'ai pu faire et de m'avoir laissé une certaine liberté pendant cette thèse qui a été une très bonne expérience pour moi. Merci aussi pour tout ce que j'ai pu apprendre durant cette thèse aussi bien mes compétences nouvelles (je pense à la bioinfo) que personnellement (je pense à la bienveillance). Je n'ai aucun regret d'avoir choisi cette thèse même si dans l'ensemble tout a été chamboulé. J'espère que de ton côté tu ne regrettes pas trop ton choix de doctorant des trois dernières années et que les autres seront moins chiant que moi.

Avant d'arriver en thèse, il y a eu la licence et le master qui ont été des éléments déterminant pour en arriver là et beaucoup de gens sont intervenus et leur aide m'a été précieuse et je les remercie. Pour cela je souhaiterais particulièrement remercier plusieurs personnes, Michel Hernould, Didier Thoraval, Fred Gévaudant et Fred Delmas. Michel, je veux te remercier pour tous nos échanges toujours enrichissants et aux pires moments de la journée (souvent proche du départ pour rentrer à la maison) mais aussi pour m'avoir accepté en master et m'avoir soutenu pendant celui lorsque c'était difficile. Merci à Didier, tes cours en licences étaient géniaux et m'ont donné envie de continuer (je m'excuse de ne pas avoir choisi la génétique) vers le master et ton soutien m'a été précieux.

Bien sûr, merci aux deux Fred pour leurs expériences et connaissances qui ont toujours su prendre le temps de partager malgré leur emploi du temps overbooké. Merci encore à tous les deux d'avoir accepté lorsque l'on était en master de prendre le temps de faire plaisir à des gamins ayant envie de faire iGEM. Fred D, je te remercie de ta gentillesse et ta sympathie, c'était toujours un plaisir de faire « n'importe quoi » avec toi (je pense à la magnetofection entre autres). Merci à Fred G d'avoir toujours su me donner des conseils avisés peu importe le moment sur des manip ou des questions.

Merci à Norbert pour m'avoir formé, littéralement, pendant mon M2 et le début de ma thèse car je pense sincèrement avoir énormément appris aussi bien scientifiquement qu'humainement en travaillant avec toi. Merci d'avoir été présent aussi aux moments où j'en avais besoin. C'est toujours un plaisir de passé du temps avec toi et je suis très content qu'aujourd'hui nous soyons amis !

JP (Michael ou Jean-Pascal), on ne sait jamais avec toi, merci à toi aussi d'avoir été un super collègue en plus d'être un ami. Même si au départ tu faisais peur dans ton antre au bout ou avec tes centaines de boîte, finalement tu n'es pas un **si** mauvais gars (haha). Merci d'avoir été là pour le soutien pendant la thèse, la reprise du sport et les soirées, journées et moment passé ensemble, toujours du bon temps. Bon courage pour ta thèse, je serais là en soutient !

Bien sûr, merci aux autres doctorants de l'équipe, Arthur et Sabri, pour votre bonne humeur. Cela a toujours été plaisant de travailler avec vous et d'échanger sur le sport, la musique, les répliques de film... vous êtes tous les deux des gars bien et je croise fort les doigts pour que votre thèse se déroule bien et que vous ayez les best résultats ! Arthur tu n'as que le fardeau de 30 ans de recherche international à conclure (:p). Spécial dédicace à toi Gabriel qui vit bien loin de nous, on ne t'oublie pas !

Enfin je souhaitais remercier mes proches, ceux qui n'ont pas travaillé avec moi mais qui ont su être un soutien moral infaillible pendant ce parcours jusqu'aujourd'hui.

Un énorme merci à mon meilleur ami, Quentin, avec qui je partage tout depuis plus de 12 ans (4518 jours pour être exact, Facebook 2021).

Je voudrais remercier tout particulièrement ma moitié, Floriane, qui a été un soutien majeur sans qui je pense tout cela n'aurait pas été possible. Merci d'être la aujourd'hui et d'être la mère de ma fille, Freya !

Merci de toujours avoir été là, de me supporter à la maison quand je râle et que je suis de mauvaise humeur. Ton soutien est extrêmement important et je pense que je ne pourrais jamais te rendre l'équivalent.

Et pour finir, je remercie mes parents d'avoir permis que je choisisse ce qu'il me plait de faire et d'avoir soutenu mes choix quel qu'ils soient. Je dédie ce manuscrit à ma mère qui me manque et que j'aimerais toujours.

General abbreviation	4
Genes and Proteins abbreviations	6
I. Figure list	10
Chapter 1 : INTRODUCTION	13
I. Potential roles of endoreduplication in Plants	14
1. What is endoreduplication? Some generalities and definitions	14
2. Endoreduplication and plant development	16
A) Endoreduplication in the hypocotyl in Arabidopsis	16
B) Endoreduplication in the Arabidopsis thaliana leaf	18
a. Epidermal pavement cells	18
b. Leaf Epidermal specialized cells: the trichomes	18
C) Endoreduplication in the root	20
D) Endoreduplication in the flower	24
E) Endoreduplication in fruits and seeds	26
3. Endoreduplication in the model of study during the PhD: the tomato fruit pericarp	28
A) Endoreduplication during Tomato fruit development	28
B) Ploidy distribution in the pericarp and potential role of endoreduplication	32
II. Genetic regulation of endoreduplication in Plants	33
1. Dynamic of CDK-CYC complexes during cell cycle progression	33
2. Transition from the cell cycle to the endocycle	34
A) Regulation of CDK-CYC complexes by proteolysis	34
B) Regulation of CDK-CYC by phosphorylation	35
C) Binding of Cyclin kinase inhibitor proteins	35
D) Transcriptional regulation of CCS52	36
3. Maintenance and ending of the endocycle.	37
4. Genetic manipulation of endoreduplication in other plant than Arabidopsis	39
III. Effect of endoreduplication on the regulation of transcription	40
1. Global Effect on gene expression in polyploid plants	40
2. Global effect on gene expression of endoreduplication	41
3. Ploidy specific gene expression in endoreduplicated cells	44
A) Gene expression regulation mechanisms.	47
B) Regulation of gene expression by chromosome positioning in the nucleus.	47
C) Transcriptional regulation and chromatin accessibility	51
D) Role of transcription factors in the ploidy specific regulation of expression	54
E) DNA methylation and gene expression regulation	54
IV. Presentation of my PhD work	55
Chapter 2: Results	57
I. Technological development for ploidy determination in situ and for ploidy dependent transcriptomic	58
1. Introduction	58
2. Results and discussion	61
A) Use of spike-in to normalize gene expression in a global shift in expression context.	61
B) Optimizing the quantification of ploidy in situ from sample preparation to imaging using Fluorescence in situ Hybridization	64
a. Optimization of the <i>FisH</i> protocol	64
b. Semi-automated imaging	68
c. Ploidy quantification pipeline	68
C) Visualization of genomic loci in living plant cells using CRISPR/Cas9 technology	76

II. Dynamics of ploidy distribution and ploidy-dependent transcriptome during pericarp fruit development	78
1. Introduction	78
2. Results	80
A) Ploidy specific transcriptome during early Tomato fruit development: experimental setup and data-preprocessing	80
B) Gene expression is changing during early fruit development and based on ploidy levels.	80
C) Biological processes expressed as a function of ploidy.	87
a. Biological processes expressed in 6DPA fruits	87
b. Biological processes expressed in 9DPA fruits	90
c. Biological processes expressed in 12DPA fruits	93
D) Dynamics of ploidy specific gene expression in the course of early fruit development	95
E) Comparing ploidy specific transcriptome and laser microdissected pericarp transcriptome.	99
F) Quantification of ploidy on sorted nuclei by FisH: Assessing the efficiency of a 5s rDNA oligo probe.	105
G) Endoreduplication occurs early during Tomato pericarp development and ploidy levels rapidly increase in the central layers of the pericarp.	109
H) The integration of ploidy distribution with ploidy specific transcriptome reveals that ploidy specific gene expression is following specific spatio-temporal patterns.	113
3. Discussion	119
A) Endoreduplication in the pericarp induces differential gene expression.	119
B) Endoreduplication during early fruit development is associated with cell specialization.	119
C) Localization of ploidy levels in the pericarp can be inferred using ploidy transcriptome data and publicly available transcriptome.	120
D) Ploidy level distribution is dynamic during early fruit development.	120
Chapter 3: Materials and methods	123
I. Biological materials	124
1. Plant materials and Cultures	124
A) Solanum lycopersicum	124
B) Nicotiana benthamiana	124
C) Bacterial stains and culture media	124
II. Cytological analysis	124
1. Isolation of tomato nuclei and flow cytometry	124
III. Molecular biology	125
1. Methods of nucleic acid manipulation and analysis	125
A) Extraction of nucleus RNA	125
B) Quality controls of RNA	125
C) Synthesis of libraries and RNA sequencing	125
IV. Cytogenetic	125
1. FisH	125
2. CRISPR-imaging	126
3. Imaging	127
4. Image analysis	127
V. Bioinformatic	127
1. Software	127
2. Mapping and normalization of RNA-Seq data	127
3. Identification of differentially expressed genes and clustering	127
4. Gene ontology enrichment	128
5. Ploidy mapping and analysis	128
Chapter 4: Discussion and perspectives	129

I. Technological development of tools for the determination of in situ ploidy levels and ploidy specific transcriptome	130
1. Transcriptome related technologies	130
2. FisH related technologies	131
3. CRISPR-imaging	131
II. Global shift in expression upon endoreduplication.	131
III. A subset of genes is specifically regulated by ploidy levels in the pericarp.	132
IV. The ploidy distribution in tomato pericarp during early development stages is already organized as a gradient.	135
V. Perspective	136
1. How is the ploidy-specific expression regulated?	136
2. Spatiotemporal gene expression in the pericarp is correlated to ploidy levels.	138
3. Towards the identification of ploidy marker genes	138
4. Establishment progress and role of endoreduplication in the pericarp and Tomato fruit growth	139
VI. Conclusion	141
References	142
ANNEXES	0
I. Presentation	0
1. Oral presentations	0
2. Poster presentation	0
II. Disciplinary trainings	0
1. Workshop Plant Regeneration, Transformation and genome editing, April 2018.	0
2. Workshop 3D FisH and image analysis, October 2018.	1
3. Workshop From Pictures to Numbers, September 2019	1
4. Bioinformatic school AVIESAN-INSERM : Initiation to processing of NGS data.	1
III. Transversal training	2
1. “Anglais scientifique 2° session Module 1 - From Research to Publication”, “Anglais scientifique 2ème session Module 4 Conference Talk” and, “Anglais scientifique 2ème session Module 5 - From text to talk”.2	
2. PAPIRUS (Parcours d’accompagnement pour publier et communiquer en science)	2
3. EDEN (Ecole des doctorants et encadrants de l’INRA	2
4. MOOC-INTRODUCTION A LA STATISTIQUE AVEC R -SESSION 11	2
5. Find a post doc abroad	2
IV. Complementary activity	2

General abbreviation

µm	Micrometers
ABA	Abscisic Acid
AsA	Ascorbic Acid
ATAC-seq	Assay for Transposase-Accessible Chromatin with highthroughput sequencing
C	Chromatids
CAM	Crassulacean Acid Metabolism
ChIP-qPCR	Chromatin Immuno Precipitation quantitative Polymerase Chain Reaction
cm	Centimeters
COL-0	Columbia-0
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPR-GO	CRISPR-Genome Organization
DAP	Days After Pollination
DAPI	4',6-diamidino-2-phénylindole
DE	Differentially Expressed
DEG	Differentially Expressed Genes
DNA	Deoxyribonucleic Acid
DPA	Day Post Anthesis
E1, E2, E3	Epidermal layer 1, 2, 3
EBC	Epidermal Bladder Cells
FACS	Fluorescence Activated Cell Sorting
FC	Fold Change
FDR	False Decovery Rate
FisH	Fluorescence <i>in situ</i> Hybridization
G1 phase	Gap 1 phase
G2 phase	Gap 2 phase

GO	Gene Ontology
h	Hour
HG	HomoGalacturonan
Hi-C seq	Chromosome Conformation Capture Method
I1, I2	Inner epidermal cell layers 1 and 2
kb	kilobase
LM	Laser Microdissection
M phase	Mitosis phase
M.crystallinum	Mesembryanthemum crystallinum
M' and M layers	Mesocarp layer
mm	Milimeters
mRNA	Messenger RNA
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NCR	Nodule-specific Cystein-Rich
ncRNA	non-coding RNA
NOR	Nucleolus Organizer Region
OE	Overexpressed
ORC	Origin Recognition Complex
PC1	Principal Component 1
PC2	Principal Component 2
PCA	Principal Component Analysis
PCD	Programmed Cell Death
PhD	Philosophiæ doctor
pol	Polymerase
qPCR	quantity Polymerase Chain Reaction
RNA	Ribonucleic Acid
RNA seq	RNA sequencing
rRNA	Ribosome RNA
S phase	Synthesis phase
scATAC-seq	Single cells ATAC-seq

SE	Standard error
sgRNA	single guide RNA
Slxxx	<i>Solanum lycopersicum</i>
TADs	Topologically Associating Domains
TF	Transcription Factor
TFBS	Transcription Factor Binding Sites
tRNA	Transfert RNA
WT	Wild Type
WVA106	West Virginia 106

Genes and Proteins abbreviations

AAO3	ALDEHYDE OXIDASE 3
ABCG	ATP-BINDING CASSETTE (ABC) TRANSPORTER G
APC/C	ANAPHASE PROMOTING COMPLEX/CYCLOSOME
ARF	AUXIN RESPONSE FACTOR
ARP3	ACTIN RELATED PROTEIN 3
ARR	ARABIDOPSIS RESPONSE REGULATOR
AtARR2	<i>Arabidopsis thaliana</i> TWO-COMPONENT RESPONSE REGULATOR2
AtCCS52A1	<i>Arabidopsis thaliana</i> CELL CYCLE SWITCH 52 A1
ATCG32, ATCG11	ATP BINDING CASSETTE G proteins 32, 11
AtCOP1	<i>Arabidopsis thaliana</i> CONSTITUTIVE PHOTOMORPHOGENIC 1
AtDEL1	<i>Arabidopsis thaliana</i> DP-E2F-like 1
AtGTL1	GT-2 LIKE
ATHB	<i>Arabidopsis thaliana</i> HOMEBOX-LEUCINE ZIPPER proteins
AtMED16	<i>Arabidopsis thaliana</i> MEDIATOR16
AtUBP14	<i>Arabidopsis thaliana</i> UBIQUITIN-SPECIFIC PROTEASE 14
AtUVI4	ULTRAVIOLET-B INSENSITIVE4

AtWEE1	<i>Arabidopsis thaliana</i> WEE1
ATXR5/ATXR6	ARABIDOPSIS TRITHORAX-RELATED PROTEIN 5 or 6
Aux/IAA	AUXIN/INDOLE-3-ACETIC ACID
BARD1	BREAST CANCER TYPE 1
bHLH	basic HELIX-LOOP-HELIX
BRCA1	BREAST CANCER ASSOCIATED 1
CAF-1	CHROMATIN ASSEMBLY FACTOR-1
CAS9	CRISPR ASSOCIATED 9
CCS	CELL CYCLE SWITCH
CDC6	CELL DIVISION CONTROL PROTEIN 6
CDK	CYCLIN-DEPENDENT KINASE
CDKB	CYCLIN-DEPENDENT KINASE B
CER1	ECERIFERUM1
CER3	ECERIFERUM3
CKI	Inhibitors CDK
CMT3	CHROMOMETHYLASE 3
CYC	CYCLIN
CYCA1	Cyclin A1
CYCB2	Cyclin-B2
CYCB3	Cyclin-B3
DEL1	Degenerin del-1
DML2	DEMETER-LIKE 2
EGL3	ENHANCER OF GLABRA3
FBL17	F BOX LIKE PROTEIN 17
G6PDH	GLUCOSE-6-PHOSPHATE DEHYDROGENASE
GAPDH	GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE
GAUT	GALACTURONOSYLTRANSFERASE
GL1	GLABRA1
GL3	GLABRA3
GPAT4	GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE 4

GT8	GLYCOSYLTRANSFERASE 8	
H2A, H2B	HISTONE 2A, HISTONE 2B	
H2AS9ph	HISTONE2A phosphorylated on Lysine 9	
H2Aub1	HISTONE2A Monoubiquitinated	
H2Bub1	HISTONE2B Monoubiquitinated	
H3, H4	HISTONE3, HISTONE4,	
H3K27me3	HISTONE3 tri-methylated on lysine 27	
H3K36me3	HISTONE3 tri-methylated on lysine 36	
H3K4me3	HISTONE3 tri-methylated on lysine 4	
H3K9ac	HISTONE3 acetylation on lysine 9	
HPY2	HIGH PLOIDY 2	
HUB1	HISTONE MONO-UBIQUITINATION 1	
ICK/KRPs	INTERACTOR/INHIBITOR OF CDK/KIP-RELATED PROTEINs	
ILP1	INCREASE LEVEL OF POLYPLOIDY 1	
IRX9-L	GALACTOSYLGALACTOSYLXYLOSYLPROTEIN GLUCUROSYLTRANSFERASE	3-BETA-
KCS	β -KETOACYL-COA SYNTHASE	
KRP	KIP RELATED PROTEIN	
KRP5	Cyclin-dependent Kinase Inhibitor 5	
LMI1	LATE MERISTEM IDENTITY 1	
MBP3	MADS-BOX PROTEIN 3	
MCM	Mini Chromosome Maintenance	
MET1	METHYLTRANSFERASE 1	
MtCCS52A	Medicago	
MtDME	<i>Medicago truncatula</i> DEMETER	
NI	NEUTRAL INVERTASE	
nsLTP	Nonspecific LIPID TRANSPORT PROTEIN	
PDF2	PROTODERMAL FACTOR 2	
PES1	PHYTYL ESTER SYNTHASE 1	
PHS1	PHOSPHATASE PROPYZAMIDE-HYPERSENSITIVE 1	

PK	PYRUVATE KINASE
POK1	PHRAGMOPLASM ORGANISATOR KINASE 1
PSY1/2	PHYTOENE SYNTHASE 1/2
RBR1	RETINOBLASTOMA RELATED PROTEIN 1
RBR1	Retinoblastoma-Related protein 1
RLK	RECEPTOR LIKE KINASE
SIM	SIAMESE
SIWEE1	<i>Solanum lycopersicum</i> WEE1
SICCS52A	(SICCS52A1)?
SIEZ1, SIEZ2	ENHANCER OF ZESTE 1 or 2
SIGPAT4	<i>Solanum lycopersicum</i> GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE 4
SIIAA17	<i>Solanum lycopersicum</i> INDOLE-3-ACETIC ACID 17
SILAX3	<i>Solanum lycopersicum</i> LIKE-AUX 3
SIPIN9	<i>Solanum lycopersicum</i> PIN FORMED 9
SMRs	SIAMESE-RELATED PROTEINs
T6PP	TREHALOSE-6-PHOSPHATE PHOSPHATASE
TAGL1	TOMATO AGAMOUS LIKE 1
TCP15	TEOSINTE BRANCHED 1/CYCLOIDEA/PCF 15
TTG1	TRANSPARENT TESTA GLABRA1
UVI4	UV INSENSITIVE 4?
WSD1	Wax Ester Synthase /Acyl-CoA: Diacylglycerol Acyltransferase
XTH	XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE
XTH5	XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE
ZEP	ZEAXANTIN EPOXIDASE

I. *Figure list*

Figure 1. Cell cycle and endoreduplication progression and polytene chromosome illustration.

Figure 2. Endoreduplication during Arabidopsis hypocotyl.

Figure 3. Endoreduplication during Arabidopsis leaves development

Figure 4. Endoreduplication and trichome development.

Figure 5. Schematic representation of root morphology and organization.

Figure 6. Schematic representation of nodule tissue ontology.

Figure 7. Illustration of Arabidopsis thaliana sepals and epidermal giant cells.

Figure 8. Examples of endoreduplication occurrences in fruits.

Figure 9. Endoreduplication during Tomato fruit pericarp development

Figure 10. Endoreduplication occurrence through Tomato fruit pericarp development.

Figure 11. Changes of gene expression in endoreduplicated root hair cells

Figure 12. Endoreduplication can induce high expression of specialized genes in Maize, Arabidopsis and Tomato.

Figure 13. Chromosome position within the nucleus can define its transcriptional program.

Figure 14. Organization of chromosome territories in Animals and plants

Figure 15. Chromatin structure is dependent on the nucleosome composition and histone post-translational modifications.

Figure 16. CRISPR/Cas9 based methods for loci visualisation in living cells.

Figure 17. Use of spike-ins in the context of global shift of expression in tomato pericarp cells.

Figure 18. Venn diagram of non-differentially expressed genes found in 6-, 9- and 12DPA ploidy transcriptomes and their associated biological processes.

Figure 19. Repeated sequences used as targets for FisH and CRISPR imaging to quantify ploidy.

Figure20. Flowchart representation of the semi-automated procedure of ploidy quantification in the tomato pericarp cells.

Figure20bis. Illustration example of automatic nuclei assignment to cell layer for ploidy mapping

Figure 21. Flowchart representation of the steps performed by the plugin to identify and count FisH spots.

Figure 22. Example of application of the Histogram segmentation tool before and after the filtering steps.

Figure 23. Example of plugin assisted FisH spot counting and issues.

Figure 24. Illustration of CRISPR-imaging observations after transient and stable transformation.

Figure 25. Ploidy sorted nuclei from 6, 9 and 12 DPA pericarp fruits show differential gene expression depending on ploidy and developmental stage.

Figure 26. Clustered normalized expression profiles of differentially expressed genes in pericarp of 6-9-12DPA fruits

Figure 27. Clustered normalized expression profiles of differentially expressed genes in pericarp of 6-9-12DPA fruits

Figure 28. Overrepresented biological processes in the different expression clusters from 6 DPA pericarp

Figure 29. Overrepresented biological processes in the different expression clusters from 9 DPA pericarp

Figure 30. Overrepresented biological processes in the different expression clusters from 12 DPA pericarp

Figure 31. Overlaps between DE genes from the different clusters of 6DPA, 9DPA and 12DPA fruits.

Figure 32. Representation of cuticle synthesis genes and their expression in relation to ploidy levels in the pericarp

Figure 33. Significant overlaps between ploidy specific transcriptome and LM pericarp transcriptome of Shirozaki *et al.*, 2018.

Figure 34. Ploidy markers from a ploidy specific transcriptome of *Arabidopsis thaliana* root cortex (Bhosale *et al.*, 2018) compared to the pericarp ploidy transcriptome.

Figure 35. Morphology of Tomato pericarp sorted nuclei based on their ploidy and evaluation of the 5sr DNA oligo FisH probe for ploidy quantification.

Figure 36. Nuclear shape is modified with increasing ploidy levels.

Figure 37. Ploidy distribution, nuclear volumes, and cell sizes in 6, 9 and 12DPA Tomato pericarps.

Figure 38. Dynamic of nuclear volumes during early fruit development

Figure 39. Data integration of ploidy transcriptome and FisH ploidy quantification reveals a spatiotemporal ploidy specific gene expression.

Figure 40. CDKA1 and CDKB1/2 expression in relation with ploidy levels in pericarp cell layers

Figure 41. Cell cycle and endoreduplication related genes expression in relation with ploidy levels in pericarp cell layers

Figure 42. Dynamic changes of ploidy levels in the different pericarp cell layers during early fruit development from 6DPA to 12DPA

Figure 43. Summary diagram of PhD and its perspectives.

Chapter 1 : INTRODUCTION

I. Potential roles of endoreduplication in Plants

1. What is endoreduplication? Some generalities and definitions

The cell cycle is a universal process, occurring in all living organisms, that results in cell division and the transfer of the genetic information (the whole copy of the genome) from a mother cell to two daughter cells (Figure 1A).

Endoreduplication is a modified cell cycle (also referred as endoreplication or endoploidization) during which cells increase their DNA content through successive rounds of nuclear genome duplication without cell division leading to increased ploidy level (Figure 1A). Ploidy is a relative measure of DNA content of a cell referring to the haploid DNA quantity of the genome, with 1C referring to 1 chromatid copy for a chromosome. Therefore, if DNA quantity of a diploid cell is 64C, this means that each chromosome has 64 copies of chromatids in this particular cell. Endoreduplication occurs without cytokinesis and the classical mitotic DNA replication events such as, DNA condensation/decondensation phases or spindle assembly (Edgar *et al.*, 2014). In endoreduplicated cells, duplicated DNA is still attached as sister chromatids by the centromeres forming a structure called polytene chromosomes (Edgar and Orr-Weaver, 2001). These polytene chromosomes have been visualized for example in *Drosophila* salivary gland cells that serve as a model for the study of endoreduplication in insects (Figure 1B, Zhimulev and Koryakov 2009) and their existence was demonstrated in Tomato fruit pericarp cells (Bourdon *et al.*, 2012).

Endoreduplication is a widespread phenomenon found in several clades in Eukaryotes. However, ploidy levels are highly variable between species or even between different tissues of the same individual. Moreover, it is not known if during endoreduplication DNA is always totally replicated in the different species or organs in which it occurs. In animals, endoreduplication is only found in some specific cell types such as in giant neurons of Gastropods (200000C), in salivary glands of *Drosophila* (2048C) and follicle cell epithelium and in Silkworm *Bombyx mori* (up to 20 rounds of DNA replication = 1 048 576C). In Mammals, endoreduplication also occurs in trophoblast giant cells (1024C), hepatocytes, cardiac myocytes, and cancer cells (D'Amato and Durante 2001, Orr-Weaver, 2015, Shu *et al.*, 2018). The role of endoreduplication in animals is not fully elucidated but has been proposed to support rapid growth and strong metabolism of specialized organs such as nurse cells and salivary glands of larva in *Drosophila* or trophoblast in Mice (Fox and Duronio, 2013, Lee *et al.*, 2009, Orr-Weaver, 2015). Another role for endoreduplication corresponds to body size control in ants and nematodes (Scholes, 2013).

In plants, endoreduplication is found in almost 90% of Angiosperms (Joubès and Chevalier, 2000). This phenomenon occurs in several plant families and does not seem to follow phylogenetic linkage suggesting that endoreduplication results from separate evolution (homoplasmy) because of specific strategies of life rather than common ancestry (Barow and Meister, 2003). For example, annual or bi-annual plants tend, more than perennials, to perform endoreduplication as a common developmental process (Barow and Meister, 2003).

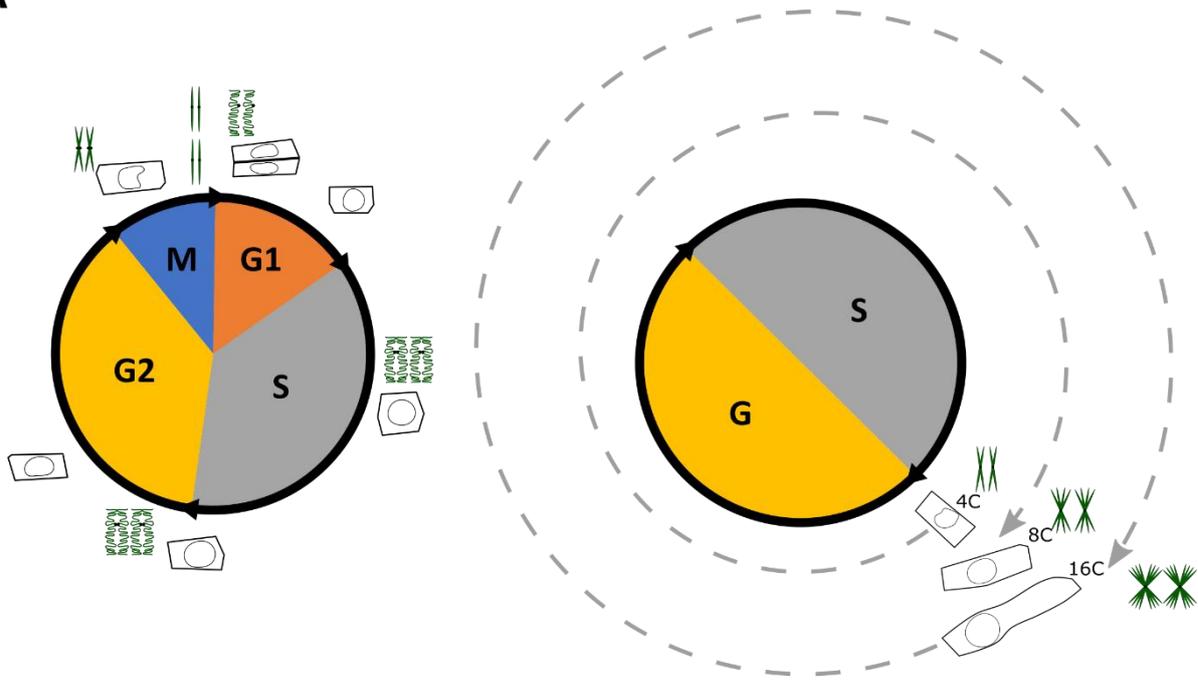
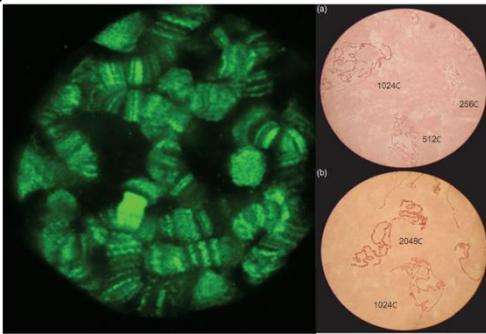
A**B**

Figure 1. Cell cycle and endoreduplication progression and polytene chromosome illustration.

(A) Illustration of the different phases of cell cycle (left panel) and endoreduplication progression (right panel). The evolution of cell size and chromosome morphology during the different phase is represented. G1: Gap phase 1; S :DNA Synthesis phase; G2: Gap2 phase; M: mitosis phase, G: Gap phase.

(B) Polytene chromosome visualization. Left panel: polytene chromosome within an intact larval salivary gland nucleus of *Drosophila melanogaster*, as revealed by green fluorescent protein (GFP) on left panel. Right panel: giant chromosomes with different degrees of polyteny visualized in light microscopy. Adapted and modified from Spradling, 2017 and Dyka *et al.*, 2016.

Thus, endoreduplication could support rapid growth needed to complete a life cycle from seed to seed in a season (Barow and Meister 2003, Bainard *et al.*, 2012). Within the same organism, endoreduplication can occur in several organs of the plant, at variable levels such as in *Arabidopsis thaliana* (Arabidopsis) leaves and sepals (up to 16C), roots (up to 8- to 16C) and trichomes (32- to 64c) or as in Tomato fruit pericarp cells (up to 512C) and leaves (up to 8C) (Cookson *et al.*, 2006, Roeder *et al.*, 2010, Bhosale *et al.*, 2018, Churchman *et al.*, 2006, Joubès *et al.*, 1999). Recent studies suggest that endoreduplication could be involved in response to environmental stresses allowing rapid cell differentiation and strong metabolic capacities to improve tolerance to stress (Scholes and Paige 2015, Bhosale *et al.*, 2019 and Lang and Schnittger 2020). But endoreduplication is also involved in different plant developmental processes. A more detailed description of the different roles of endoreduplication during plant development and in particular in our model, the Tomato fruit, is given below.

2. Endoreduplication and plant development

Endoreduplication occurs throughout plant development in many different organs and tissues: in above ground organs such as hypocotyl, leaf, and in reproductive structures such as flowers, fruits and seeds; in below ground organs such as the roots but also specialized structures like nodules. Endoreduplication occurs more frequently in organs that develop a large structure rapidly and have a need for a strong metabolic activity and storage, such as Maize endosperm cells (Leiva-Neto *et al.*, 2004) or Medicago nodules Foucher and Kondorosi, 2000). In general, DNA quantity in a cell is positively correlated with cell size (Figure 1A, Beaulieu *et al.*, 2008, Bourdon *et al.*, 2010) and most of the described cases of endoreduplicated cells are larger than non-endoreduplicated ones in the same organ (Breuer *et al.*, 2007). These observations led to link endoreduplication with a role in cell growth and organ size control (Kondorosi *et al.*, 2000, Larkins *et al.*, 2001, Barow and Meister, 2003). Endoreduplication also seems to be important for other developmental processes and hereafter, a series of organs and developmental events in which endoreduplication has a role are detailed.

A) Endoreduplication in the hypocotyl in *Arabidopsis*

In *Arabidopsis*, the hypocotyl is formed during embryogenesis and is composed of a fixed number of cells. Therefore, after germination most of the growth occurs through cell expansion (Gendreau *et al.*, 1997). Endoreduplication takes place in the hypocotyl cells very early after germination and ploidy levels can reach up to 8C in light-grown plantlets (Gendreau *et al.*, 1997). Interestingly, in this organ, light seems to inhibit endoreduplication since in the darkness, higher ploidy levels (16C) are observed (Figure 2A and 2B). The increased proportion of higher ploidy level are associated with an increase in cell length and hypocotyl length (Gendreau *et al.*, 1998; Amijima *et al.*, 2014). In this organ, the hypothesized role for endoreduplication is thus to be a strategy to grow more in a short time, in dark environment, supporting etiolated growth in order to reach light rapidly.

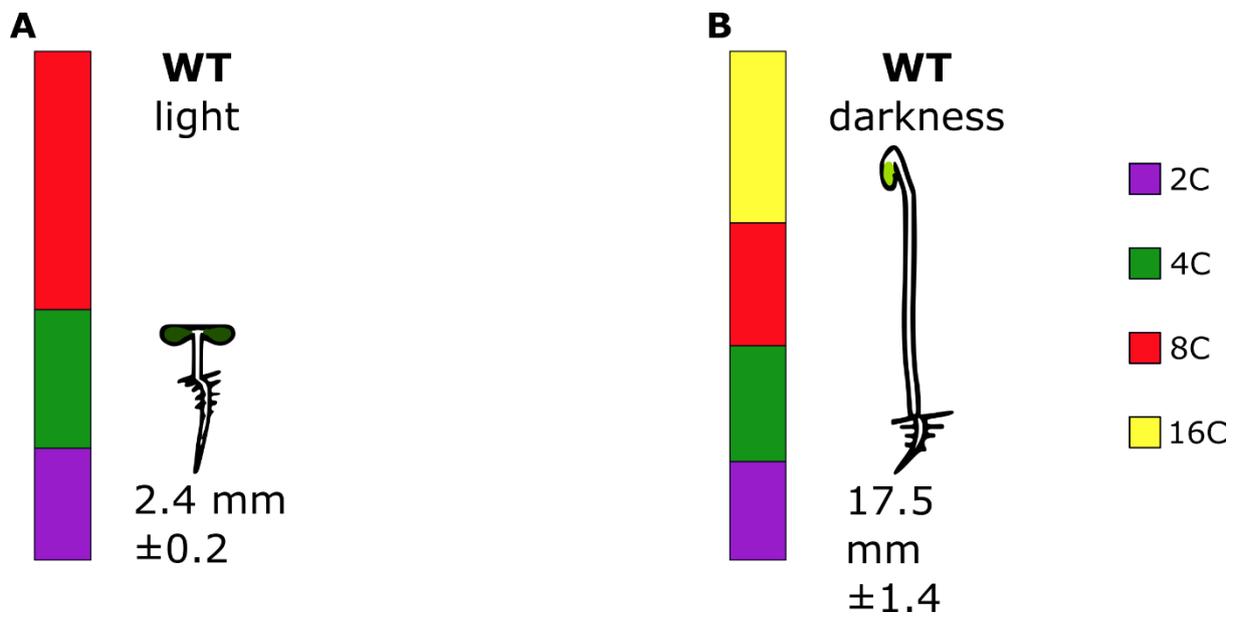


Figure 2. Endoreduplication during *Arabidopsis* hypocotyl.

(A) Endoreduplication patterns in the hypocotyl of light-grown *Arabidopsis* seedlings.

(B) Endoreduplication patterns in the hypocotyl of dark-grown *Arabidopsis* seedlings

A and B, adapted and modified from Gendreau *et al.*, 1999.

B) *Endoreduplication in the Arabidopsis thaliana leaf*

a. **Epidermal pavement cells**

In *Arabidopsis*, leaf development is divided in two phases, a cell division phase and a cell expansion phase that overlap in time. First, the leaf primordium, emerging from the shoot apical meristem, grows by cell proliferation, that occurs in the entire leaf and produces new cells (Figure 3A). In a second phase, cell proliferation progressively stops, and cells start to expand. This cell expansion phase progresses as a front from the tip to the base of the leaf (Gonzalez *et al.*, 2012, Figure 3A). This second growth phase is associated with the occurrence of endoreduplication (Melaragno *et al.*, 1993) with ploidy levels reaching up to 16C in the epidermal pavement cells (Figure 3B), the outermost cell layer of the leaf. To better understand how endoreduplication is related to increasing cell size in leaves, a mathematical model was built to simulate epidermal cell size distribution as a function of endoreduplication dynamics (Kawade and Tsukaya, 2017). This model, taking into account that endoreduplication dynamics is following a Poisson law and considering that cell size is determined according to ploidy level, recapitulated experimental data showing the Gaussian distribution of cell size. However, a meta-analysis of published data linking endoreduplication with cell growth in leaves of *Arabidopsis* mutants showed that in most cases endoreduplication is not the only factor explaining the changes in cell size (Tsukaya 2019).

Endoreduplication in the leaf is influenced by environmental factors such as day length since the transfer of *Arabidopsis* plants from short-day-period to long-day-period increases the proportion of 8C and 16C nuclei accompanied with an increase in leaf size (Del Prete *et al.*, 2019). Interestingly, this increased proportions of 8C and 16C is associated with an accumulation of the floral transition markers FLOWERING LOCUS T, CONSTANS and APETALA suggesting that endoreduplication could be involved in signaling transition from vegetative phase to reproduction (Del Prete *et al.*, 2019). In these leaves, the increase of ploidy levels has also been associated with a complex reprogramming of metabolism, including changes in sugar metabolism, cell wall synthesis and secondary metabolism (Del Prete *et al.*, 2019). In *Arabidopsis* leaf, the potential roles of endoreduplication are thus to regulate at least partially cell-size and leaf size and is also associated with flowering after changes by light conditions and metabolic changes.

b. **Leaf Epidermal specialized cells: the trichomes**

In *Arabidopsis* leaves, the epidermal pavement cells are not the only ones to undergo endoreduplication. Trichomes, large cells that differentiate from pluripotent epidermal cells called protodermal cells in young leaves, also undergo endoreduplication (Figure 4A). In *Arabidopsis*, trichomes have a three branched morphology dependent on ploidy levels (Figure 4B). The early differentiation of trichomes requires on a positive transcriptional signaling cascade of a transcription factor complex composed of GLABRA1 (GL1), TRANSPARENT TESTA GLABRA1 (TTG1), GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) (Hülkamp *et al.*, 1994; Fambrini and Pugliesi, 2019). One of the first consequence of this transcriptional activation is that the cell stops dividing and starts endoreduplicating and enlarging (Figure 4A). The cell expansion is mainly driven by the growth of the vacuole and occurs simultaneously with branching.

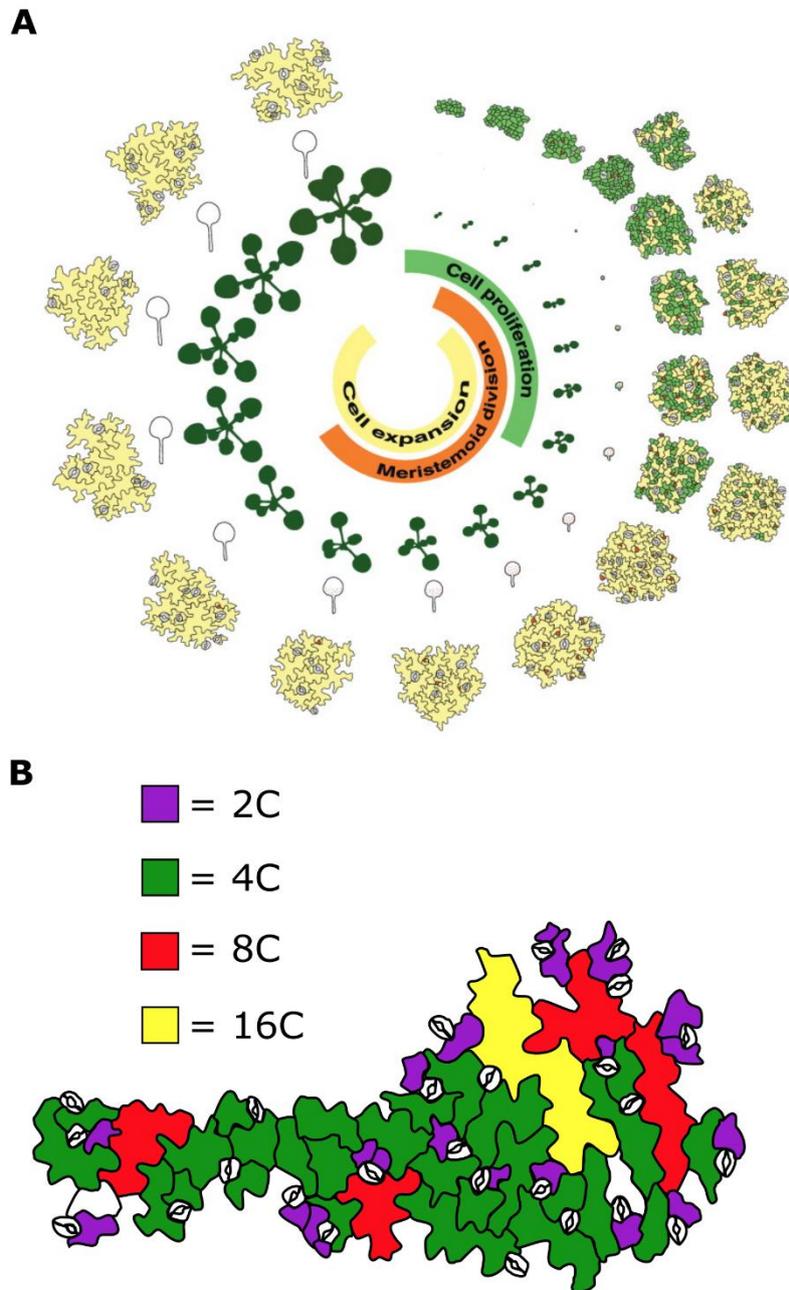


Figure 3. Endoreduplication during Arabidopsis leaves development

(A) Schematic representation of Arabidopsis developmental stages from 4 to 20 days after sowing, at rosette, leaf and cellular level. The developmental phases are represented in different colors: in green, the cell proliferation phase; in orange, the meristemoid division phase (not discussed here) and in yellow, the cell expansion phase.

(B) Representation of an epidermal cell population colored by ploidy levels in the leaf. A from Gonzalez *et al.*, 2012. B adapted and modified from Melaragno *et al.*, 1993

When fully grown, the trichome has reached 32C DNA content and has three branches (Figure 4B and 4C).

Generally, size and branching of the trichome follow a positive correlation with DNA quantity and mutants with increased endoreduplication levels show increased trichome branching (Figure 4D, Kasili *et al.*, 2011, Fambrini and Pugliesi 2019). Trichome overbranching is also observed after treatment with colchicine (Figure 4E, Robinson *et al.*, 2018), a drug able to disrupt the cell cycle, inducing whole genome duplication and thus polyploidy in plants (Robinson *et al.*, 2018). On the other hand, inhibition of endoreduplication during trichome initiation results in dedifferentiation of the cell that finally is integrated in the epithelium as an epithelial cell, suggesting a strong link between endoreduplication and the differentiation of trichome in the Arabidopsis leaf (Bramsiepe *et al.*, 2010). In Arabidopsis, endoreduplication in the trichome is thus proposed to have a role in regulating cell size and controlling cell fate.

C) *Endoreduplication in the root*

Endoreduplication also occurs during root development. The primary root can be divided in three zones: the meristem zone where cells are mostly dividing, a transition zone in which cells shift from the cell cycle to the endocycle, and the elongation zone where cells expand and endoreduplicate reaching up to 16C. In Arabidopsis, proper root development depends on the correct maintenance of the meristem size and a fine balance between dividing and differentiating (expanding and endoreduplicating) cells (Figure 5, Heyman *et al.*, 2017; Bhosale *et al.*, 2018). In Arabidopsis root, the transition between division and endoreduplication is regulated by hormones, and more specifically by auxin and cytokinin. A gradient of auxin is present in the root as a maximum in the meristem zone to a minimum in the transition zone (Figure 5, Breuer *et al.*, 2014; Mambro *et al.*, 2017). This auxin minimum is located at the place where cytokinin signaling takes place and where the transition between the cell cycle and endoreduplication is occurring (4C-8C boundary in Figure 5). The phytohormones auxin and cytokinin play an important role in this regulation. It has been shown that cytokinin controls the shift from the mitotic cycle to the endocycle in the root (Takahashi *et al.*, 2013). Plants with reduced response to cytokinin have lower endoreduplication activity in roots (Takahashi *et al.*, 2013). A screen of ethyl methanesulfonate for larger or smaller organs and change in ploidy levels in treated Arabidopsis seeds resulted in the identification of a mutant named *high ploidy 2 (hpy2)* (Ishida *et al.*, 2009). *hpy2* seedlings have higher ploidy levels in the root, very short primary root and a reduced root meristem area composed of larger cells compared to WT (Ishida *et al.*, 2009). Thus, in the primary root, endoreduplication seems to regulate growth by allowing cell expansion after the transition zone.

Some epidermal cells of the root can differentiate into root hairs cells that are long tubular shaped cells. Root hairs cells can be as long as 1 mm or more. They greatly increase the rooting area helping in nutrient acquisition, anchorage, and microbe interactions (Grierson *et al.*, 2014). It has been reported that root hair cells undergo several rounds of endoreduplication and reach 16C ploidy level when fully developed (Sliwinska *et al.*, 2012).

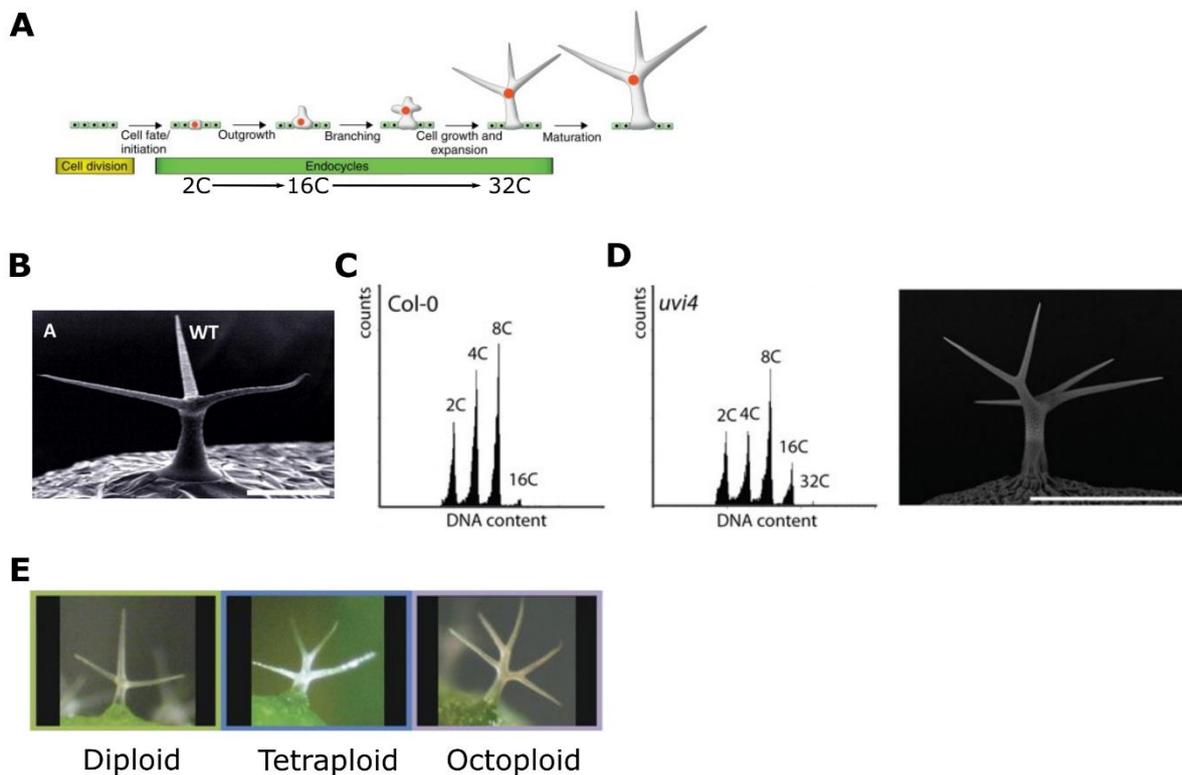


Figure 4. Endoreduplication and trichome development.

(A) Development of a trichome of *Arabidopsis* wild type. Trichome development starts with the specification of single cells within the epidermal layer of the leaf (cell fate/initiation). After this stage, the cell will start doing three rounds of endoreduplication and enlarge (outgrowth), attaining ten-times the epidermal cell size. Then, branching initiates and simultaneously cell is continuing to enlarge and undergo one more endoreduplication round (Cell growth and expansion). Finally, the trichome will eventually continue to grow until being fully grown (Maturation).

(B) Scanning electron micrograph of wild-type (Col-0) trichome showing the three branches morphology. Bar = 50 μ m. Adapted from Kasili *et al.*, 2011.

(C) Flow cytometry analysis representing the ploidy levels present in wild-type Col-0 leaves. Adapted from Heyman *et al.*, 2017.

(D) Flow cytometry analysis representing ploidy levels present in mutant line *uvi-4* having increased ploidy levels in leaves (left). Adapted from Heyman *et al.*, 2017. Scanning electron micrograph of *uvi-4* mutant trichome showing over branched phenotype (right). Bar = 300 μ m. Adapted from Heyman *et al.*, 2011

(E) Trichome phenotype observed in plants treated with colchicine. Colchicine is a drug perturbing the cell cycle and leading to the production of plants with whole genome duplication. Robinson *et al.*, 2018

It was shown that root hair cells undergo one more round of endoreduplication compared to their neighbor non hair epidermal cells (16C vs 8C, Bhosale *et al.*, 2018) indicating a potential role of endoreduplication in the specification of these root hair cells. In the *hyp7* mutant, mutated in the RHL1 protein having an essential role in the DNA topoisomerase VI complex, showing a decreased of endoreduplication results in leaf epidermal cells, trichomes and root hair cells that are smaller (Sugimoto-Shirasu *et al.*, 2005). However, in several hair mutants, the differentiation of root hair cells in the collet and root regions is impaired when there is less endoreduplication while the size of these cells is less or not affected by the decreased ploidy level, suggesting that endoreduplication might be necessary for the differentiation but that in these root hair cells, growth is independent of ploidy levels (Sliwinska *et al.*, 2015). Endoreduplication in root hair cells thus seems to play a role in the establishment of their identity (Sliwinska *et al.*, 2015; Bhosale *et al.*, 2018).

Other specialized root structures also undergo several rounds of endoreduplication such as the nodules in *Medicago* (*Medicago truncatula*) or *Lotus* (*Lotus Japonica*) (Foucher and Kondorosi, 2000, Suzaki *et al.*, 2014). Nodules are specialized structures formed after symbiotic infection of roots of leguminous plants by soil bacteria responsible of atmospheric nitrogen fixation. Nodules can be divided in four separated areas of different functions. First, a meristem zone produces new cells, through cell division, that will later be forming the different nodule cell layers (Figure 6A and 6B). The invasion zone is composed of endoreduplicated cells that can reach 16C and that can be infected by rhizobia. The next area is an interzone that contains symbiotic cells that are larger than in the meristem area or the invasion zone. In the interzone cells are endoreduplicated and can reach 16C to 32C. The last zone is the nitrogen-fixing zone where symbiotic cells are the largest with large vacuoles and ploidy levels reaching 64C. The areas containing the largest cells display higher ploidy levels indicating a positive relation between endoreduplication and cell size (Figure 6B, Nagymihály *et al.*, 2017). It has been shown using plant with impaired endoreduplication process that a proper transition from the cell cycle to the endoreduplication is necessary for correct the differentiation and functionality of the nodule structures showing that endoreduplication is an important process for the differentiation of nitrogen fixing cells (Vinardell *et al.*, 2003; Cebolla *et al.*, 1999).

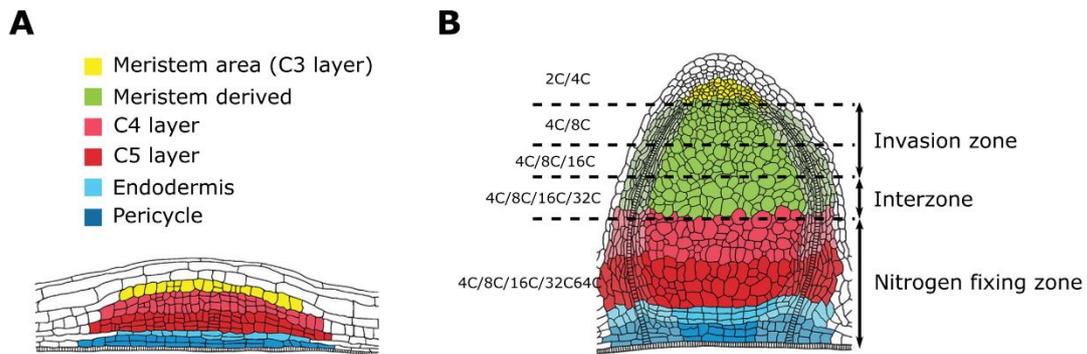


Figure 5. Schematic representation of root morphology and organization.

The dash lines represent in yellow the area considered as root meristem where cells divide, in rose the transition zone where cells enter endocycle and in green the elongation and differentiation zone where cells continue to endoreduplicate and start to grow and differentiates. The black dash lines represent an artificial frontier where cells in each side have different ploidy levels, for example the 2C-4C boundary represent areas where on the left cells are mostly 2C and on the right 4C. The root colorized areas represent in yellow the auxin gradient that is diminishing when progressing toward the transition zone (local minimum of auxin concentration) and in blue the area where cytokinin is found (transition zone).

Figure modified and adapted from Breuer *et al.*, 2014 and Bhosale *et al.*, 2018.

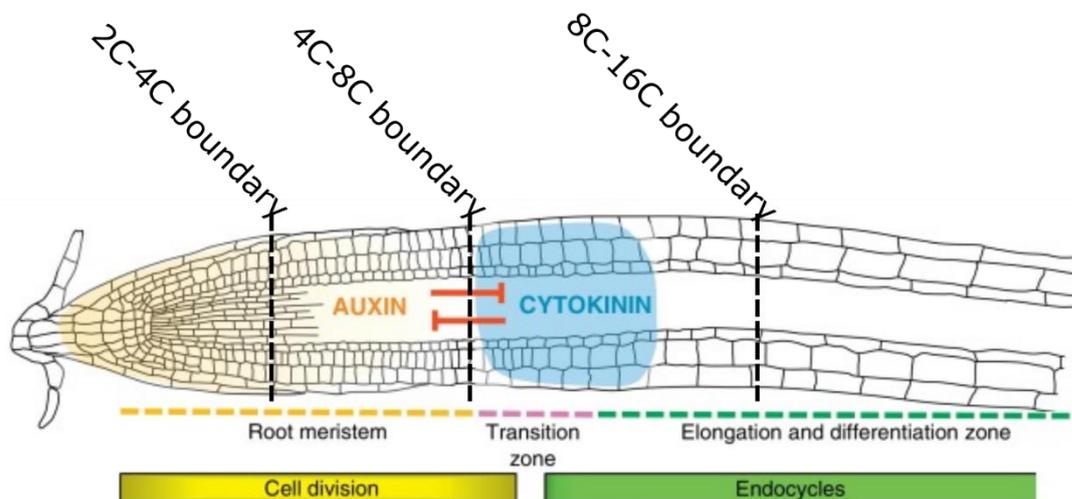


Figure 6. Schematic representation of nodule tissue ontology.

(A) Indeterminate nodule primordium with cell ontology that is indicated by the colored layers that will be involved in nodule formation. C3, C4 and C5 layers are the third, fourth and fifth layers away from epidermis.

(B) Differentiated nodule with primordium cell ontology that is indicated with same color than color used for undifferentiated primordium.

Between dashed lines are areas where indicated ploidy levels are observed. Double arrows indicate areas of nodules implicated in the nodule formation and differentiation.

Adapted and modified from Xiao *et al.*, 2016.

D) Endoreduplication in the flower

Endoreduplication takes place in flowers of many different species such as Orchids (*Phalaenopsis*, Ho *et al.*, 2016), Cabbage (*Brassica oleracea* L., Kudo and Kimura, 2002) or Red Clover (*Trifolium pratense*, Kocová *et al.*, 2016).

In *Arabidopsis* developing flowers, endoreduplication is found in sepals that correspond to the green leaf like structures of the flower (Figure 7A). The role of sepals is to be protective by enclosing the developing reproductive organs. Both abaxial and adaxial epidermis cells of the *Arabidopsis* sepals undergo endoreduplication reaching up to 16C and rarely 32C (Robinson *et al.*, 2018). These cells are referred to as giant cells because of their length that is nearly a fifth of the sepal length (360µm long, Figure 7B and 7C, Roeder *et al.*, 2012). In this organ, non endoreduplicated small cells (10µm long) co-exist with these endoreduplicated giant cells. Giant cell identity is determined before endoreduplication starts. However, it seems that endoreduplication is inhibiting small cell identity since transgenic plants with increased endoreduplication have a reduced number of cells expressing small cell identity markers (Roeder *et al.*, 2012). Increase or decrease of the proportion of giant cells results in change of sepal curvature in an outward or inward manner, respectively, whereas wild type *Arabidopsis* sepals have a flat shape (Roeder *et al.*, 2010). Thus, endoreduplication is influencing the morphology of sepals. Moreover, transgenic plants with increased ploidy levels in sepal epidermal cells results present an increased proportion and size of giant cells without any alteration of sepal length or growth rate (Roeder *et al.*, 2012). In transgenics plants harboring more giant cells and thus higher ploidy levels in the sepals, due to the specific overexpression of the LGO/SMR1 gene in the sepal epidermis, an increased expression of defense genes has been observed including the up regulation of Glucosinolate biosynthesis genes (Schwarz and Roeder 2016). Glucosinolate is a secondary metabolite stored for being used as a repulsive compound against herbivores insects. The upregulation of glucosinolate biosynthesis is also upregulated in mature trichomes (Jakoby *et al.*, 2008), thus suggesting that endoreduplication could have a role in the regulation of secondary metabolism implicated in defense. Based on these data, in the *Arabidopsis* sepals, endoreduplication could have a role in regulating organ curvature, cell size, cell fate and maybe defense response pathways.

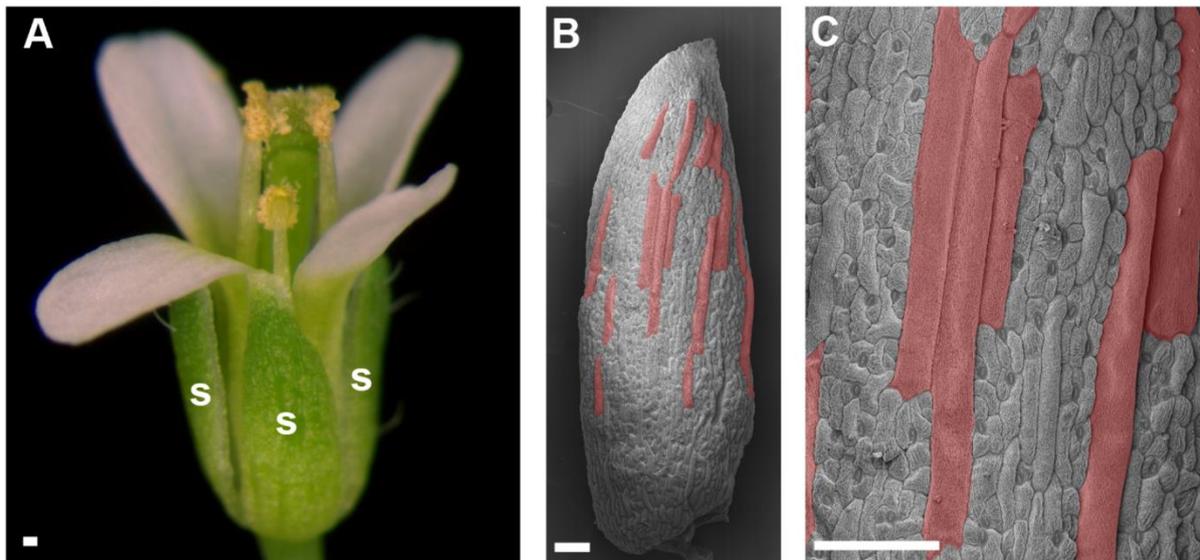


Figure 7. Illustration of *Arabidopsis thaliana* sepals and epidermal giant cells.

(A) Wild-type *Arabidopsis* flower. Sepals are indicated (s)

(B, C) Scanning electron micrograph of a mature *Arabidopsis* sepal. The typical giant cells of the sepals are colored in pink.

Adapted from Roeder *et al.*, 2010.

E) Endoreduplication in fruits and seeds

Regarding fruits and seeds, endoreduplication has also been reported in 28 species among 12 families of angiosperms for fruits and in 47 species among 27 families for seeds (D'Amato, 1984; Barow and Meister, 2003; Bourdon *et al.*, 2010). For example, in the capsule of the Foxtail Lily (*Eremurus robustus*), in cells from the pericarp tissue developing from the ovary wall, polyploid nuclei have been observed as well as polytene chromosome structures associated with increased nuclear size (Figure 8A and 8B, H Lauber 1946). Endoreduplicating cells have also been reported in fleshy fruit species such as cucurbits (Coombe, 1976; Barow, 2006) or Tomato (Joubès and Chevalier 2000). In several fruits, endoreduplication is negatively correlated with the duration of the growth period. Indeed, it was observed that fruits developing in less than 13 weeks undergo 3 to 8 rounds of endoreduplication whereas endoreduplication is absent in most of the studied fruit species with longer developmental timing (Figure 8C Bourdon *et al.*, 2010). For example, tomato developing in less than 10 weeks performs 8 endocycles whereas apple developing in a bit more than 20 weeks is only composed of non endoreduplicated cells (Figure 8C)

Endoreduplication has also been reported, for more than 60 species, in the endosperm, a differentiated structure of the seed specialized in the storage and nutrition of the developing embryo. This tissue presents a large panel of ploidy levels, up to 3072C in wild cucumber (*Echinocystis lobata*, D'Amato 1984). Maize is the most studied species regarding endoreduplication as a potential regulator of cell size, kernel size and starch content (Leiva-Neto *et al.*, 2004; Sabelli *et al.*, 2009; Sabelli *et al.*, 2013). Endosperm is formed after the double fertilization. First, one of the sperm nuclei fuses to the egg cell nucleus and forms the embryo. The second sperm nucleus fuses to the polar nucleus of the binucleate central cell resulting in the formation of a 3C cell, the primary endosperm. The endosperm then develops by several divisions and undergo up to several rounds of endoreduplication until full development (Sabelli *et al.*, 2009). The onset of endoreduplication in the Maize endosperm is associated with the production of the zein protein that accumulates as a storage of nitrogen, and with a reduction of chromatin condensation (Sabelli and Larkins 2009).

In Teosinte, the Maize wild relative and ancestor, the endosperm has been studied to understand, at the cellular level, the domesticated characters that permitted to obtain a larger endosperm in Maize (Figure 8D). Teosinte produces highly endoreduplicated nuclei in the endosperm, till 96C (Figure 8E), whereas Maize is producing nuclei with one more endoreduplication round (192C, Figure 8F, Vilhar *et al.*, 2002). However, the major difference in endoreduplication between Teosinte and Maize is the spatial distribution of highly endoreduplicated nuclei in the endosperm. In Maize, endoreduplicated nuclei are localized in the upper part of the endosperm while in Teosinte endoreduplicated nuclei are localized in all the central endosperm (Figure 8F). It was hypothesized that this modified distribution of highly endoreduplicated cells might have been selected by domestication and allowed a more efficient starch accumulation in Maize than in Teosinte (Figure 8F, Dermastia *et al.*, 2009).

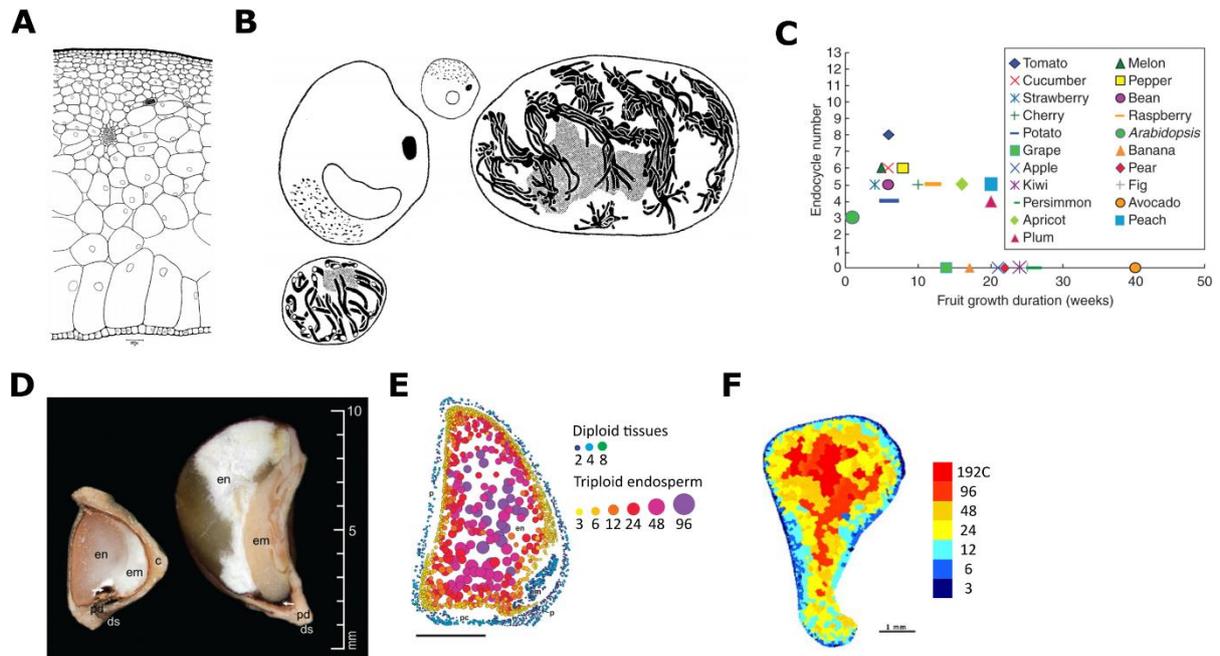


Figure 8. Examples of endoreduplication occurrences in fruits.

(A) Drawing of the cells composing the pericarp of *Ememurus robustus* fruit showing the cellular organization and cell size diversity.

(B) Drawing of nuclei observed in *Ememurus robustus* at different ploidy stages. The top left nuclei correspond to a polyplod nucleus (big nucleus) and a diploid nucleus (small nucleus). On the right and at the bottom are represented chromosome structures observed in polyploid and tetraploid nuclei, on the right and at the bottom, respectively.

(C) Occurrence of endoreduplication in fleshy fruits. The maximal number of endocycles determined in fully ripened fleshy fruits was plotted against the duration of fruit growth until ripening.

(D) Photographs of teosinte (maize ancestor, left) and cultivated maize (right) caryopses.

(E) Ploidy map of teosinte caryopsis at 11 days after pollination

(F) Ploidy map of maize caryopsis at 16 days after pollination

A and B adapted from Lauber, 1947. C from Chevalier *et al.*, 2011. D and E from Dermastia *et al.*, 2009. F adapted from Vilhar *et al.*, 2002.

In conclusion, endoreduplication in the endosperm is frequent and studies of Maize and its ancestor Teosinte shows that this process is important for kernel development and might be linked to starch accumulation in endosperm.

In conclusion, endoreduplication seems to have different roles in different tissues and organs. High endoreduplication levels are often associated with large cell size and organ size in leaves or fruits. The analysis of mutants with reduced endoreduplication allowed to observe that some mutants are able to compensate this reduction by increasing cell number to preserve the organ size (Massonnet *et al.*, 2011). Endoreduplication might have a role in enhancing/potentiating cell expansion rather than driving it because cell expansion can occur at a similar level even in absence of increased ploidy levels. Cell differentiation is also another event in which endoreduplication has a role because important impairment of endoreduplication limits differentiation in trichomes and nodules, for example. Finally, endoreduplication could have a role differentiation and regulation of metabolism in Maize kernel for starch and protein accumulation both occurring in large endoreduplicated cells.

3. Endoreduplication in the model of study during the PhD: the tomato fruit pericarp

Because Tomato fruit develops rapidly and can present large variation in size and form, it is used as a model of study to understand which roles endoreduplication and its spatial distribution could have for fleshy fruit development. Hereafter I will describe what is known about where and when endoreduplication is occurring during Tomato fruit development.

A) Endoreduplication during Tomato fruit development

In Tomato (*Solanum lycopersicum*), the fruit is a berry formed by the differentiation of the ovary. It is composed of a pericarp that derives from the walls of, at least, two fused carpels, of a jelly locular tissue surrounding the seeds and of a placenta maintaining seeds attached to a central columella (Figure 9A).

At the end of Tomato fruit development, the pericarp, the locular tissue and columella to a lesser extend are composed of cells that have undergone several rounds of endoreduplication (Figure 9B, Joubès *et al.*, 1999) and the ploidy levels varies between cultivars suggesting that the number of endoreduplication rounds is influenced by the genotype (Cheniclet *et al.*, 2005).

In Tomato, endoreduplication already occurs in the ovary wall during flower development (Cheniclet *et al.*, 2005, Renaudin *et al.*, 2017) that starts with the formation of the floral meristem and finishes with the anthesis that corresponds to the pollination and fertilization of the fully functional flower.

Before fertilization, the ovary wall grows mainly by cell division and is composed of seven to nine cell layers between stage 11 and 16 (4- and 8 mm flowers in *Solanum lycopersicum* cv. WVA106) of the flower development according to Brukhin *et al.*, 2003.

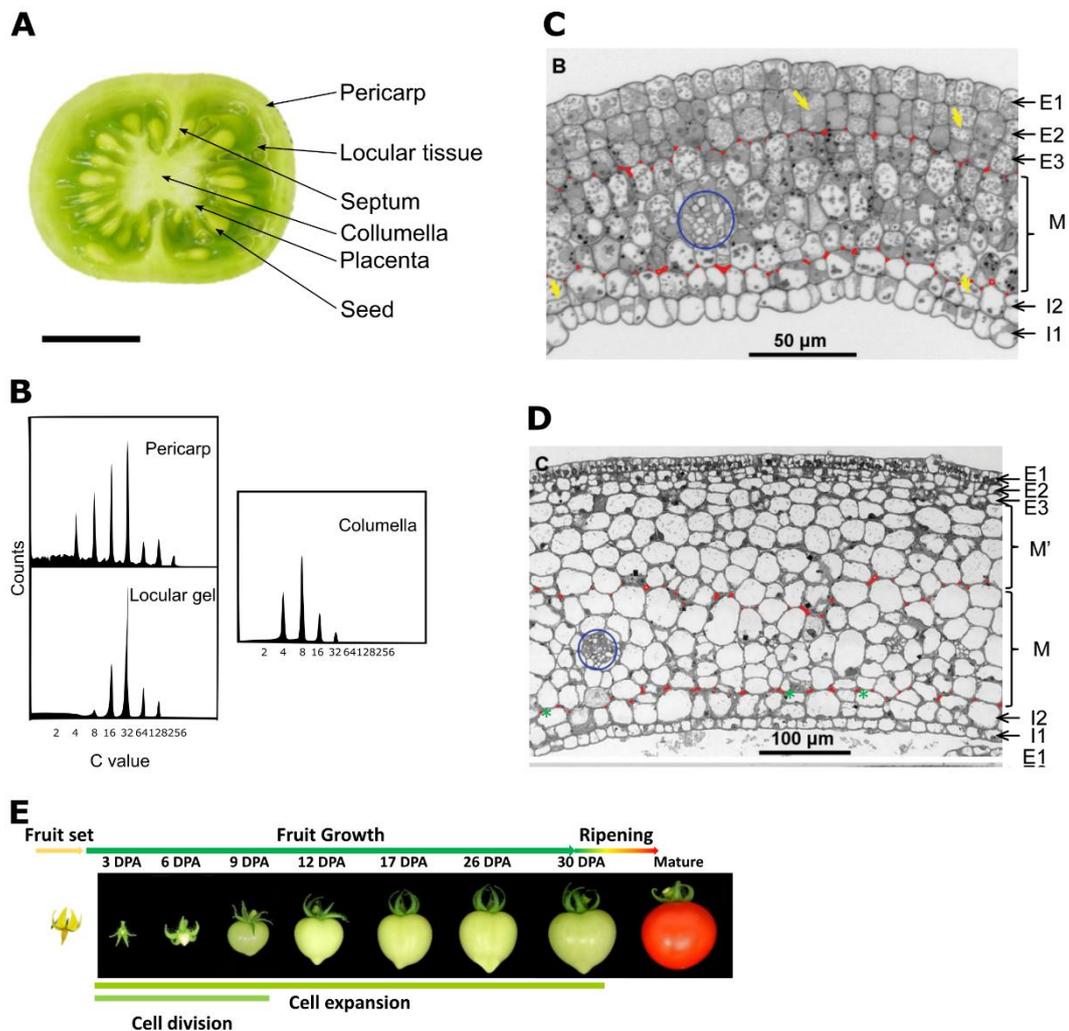


Figure 9. Endoreduplication during Tomato fruit pericarp development

(A) Anatomy of a Tomato (*Solanum lycopersicum* cv. WVA106) fruit. Equatorial section. Scale bar = 1cm.

(B) Flow cytometry histograms of DAPI stained nuclei extracted from different tissues of Tomato fruit at breaker stage.

(C and D) Structure of the pericarp of the cherry tomato WVA106 at anthesis (C) and 4DPA (D). Blue circles indicate vascular bundles, yellow arrows new cell walls and red indicates extracellular spaces delimitating the M cell layers.

(E) Illustration of different developmental stage of tomato fruit growth and the associated Phase according to Gillaspay *et al.*, 1993

Figures modified and adapted from Joubès *et al.*, 1999 (B) and Renaudin *et al.*, 2017 (C and D). (E) from Jean-Philippe Mauxion (personal communication).

At stage 18, before anthesis, 43% of the nuclei are endoreduplicated, with around 2-3% of 8C and almost 40% of 4C, considered as a first round of endoreduplication due to the absence of division structures, such as condensed chromosome of metaphase to telophase and division planes (Renaudin *et al.*, 2017). After fertilization and during the whole fruit development, the pericarp volume represents nearly half of total fruit volume and the growth rate of the fruit is the same as the growth rate of the pericarp (Renaudin *et al.*, 2017). This is indicating that the pericarp is mainly representing the tomato fruit growth, thus the next paragraph will mostly be focusing on the cellular events and endoreduplication dynamic during pericarp development.

The ovary wall and the derived structure, the pericarp, is divided in three areas (Figure 9C and 9D). First, an outer-epidermis contains three cell layers named E1 (most outer layer), E2 and E3 (the two-sub outer-epidermal layers, Figure 9C). Second, an inner part, formed after anthesis is composed of multiple layers grouped under the name of M' layers. M' cells mostly originate from layers E2 and E3 after periclinal cell divisions (Figure 9D). Third, four cell layers formed before anthesis are composing the M layer (Figure 9C). The M and M' layers form the mesocarp. Finally, the inner part of the pericarp in contact with the locular tissue is corresponding to two cell layers with the inner-epidermis layer I1 and its sub-epidermal layer I2 (Figure 9C).

Fruit growth and development is divided in three phases according to Gillaspay (1993). The first phase corresponds to the fruit set during which internal signals, corresponding to hormonal signal of gibberellins, auxins and downstream regulators, are inducing growth (Ariizumi *et al.*, 2013). This phase starts if the fertilization succeeded and lasts for nearly 20h (Figure 9D; Cheniclet *et al.*, 2005). The following phase lasts between 7 to 10 days depending on the genotype (Gillaspay *et al.*, 1993). During this period, the pericarp is characterized by the occurrence of high cell division activity. In the pericarp, cell divisions start the first day after anthesis and are mostly localized in the three epidermal layers (E1, E2 and E3, Figure 10A and 10B) and two inner-epidermal layers (I1 and I2, Figure 10E and 10F). The mitotic activity in the outer epidermis generates new cell layers whose number varies in function of the genotype (Figure 10G). However, during this period, cell expansion occurs already and increases greatly the volume of the M layer cells by 100 times compared to anthesis (Figure 10C). The increase in M layers volume participates to half the increase of the fruit volume (Cheniclet *et al.*, 2005, Renaudin *et al.*, 2017). Concomitantly, ploidy levels increase in the pericarp from a mean C value (sum of the C value weighted by their frequency) of 3.2 to 7 from anthesis to 10 days post fertilization. This change in mean ploidy levels is due to the disappearance of 2C nuclei and the appearance of new classes of ploidy (8C and 16C, Figure 10H) from anthesis to 10DPA (Days Post-Anthesis, Cheniclet *et al.*, 2005). At the end of this phase, in the pericarp, only the outer epidermis continues to divide.

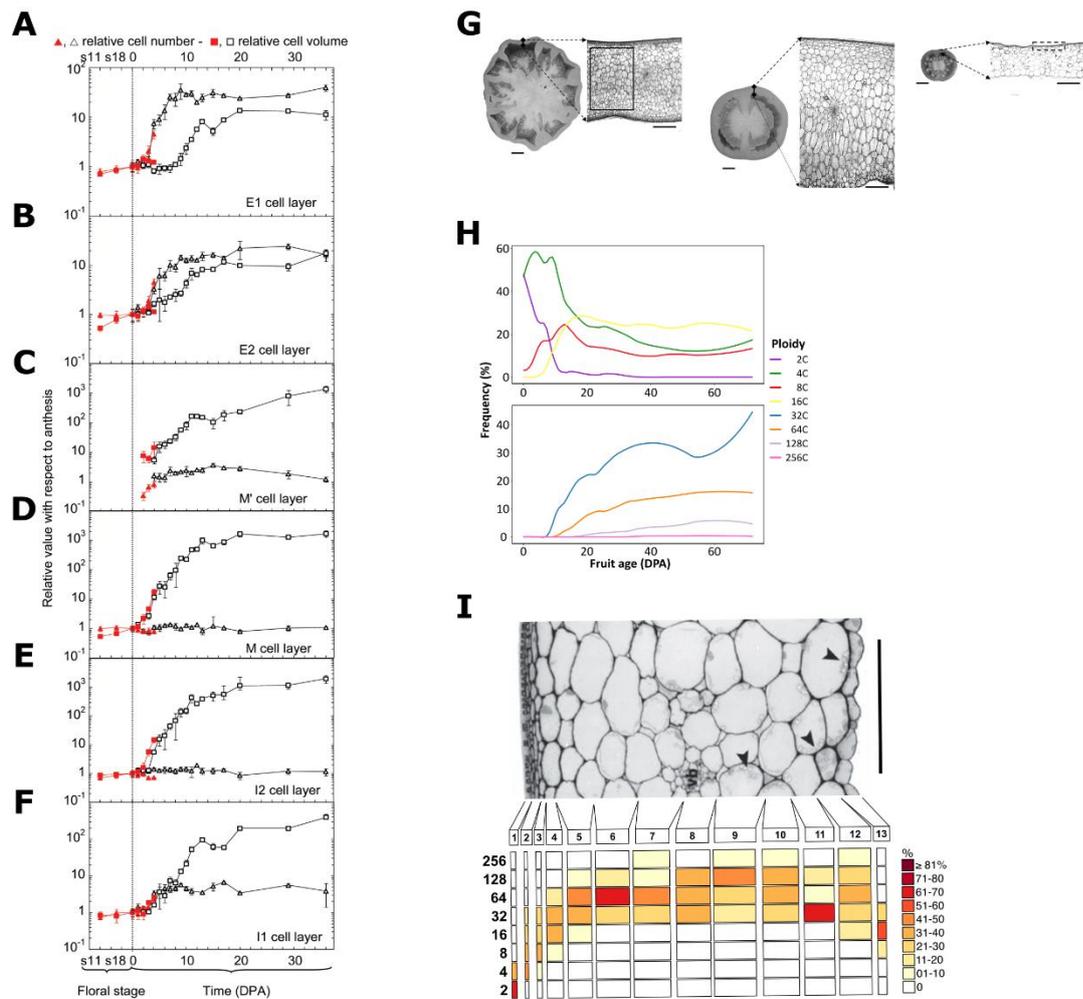


Figure 10. Endoreduplication occurrence through Tomato fruit pericarp development.

(A to F) Evolution of cell number and cell volume in the different pericarp cell layers during fruit development (WVA106). Variables are expressed as relative values with reference to the value of 1 at anthesis (dot line).

(G) Illustration of the morphological differences between three fruits and pericarp sections of different Tomato cultivars at breaker stage. From left to right is a fruit of Grosse de Gros, a fruit of Ferum 26 and a fruit of WVA106. Scale bars: left = 1cm, right = 1mm

(H) Developmental dynamic of ploidy levels in Wva106 pericarp. Top panel: evolution of 2C, 4C, 8C, and 16C ploidy levels. Bottom panel: evolution of 32C, 64C, 128C and 256C ploidy levels.

(I) Ploidy mapping in a mature green tomato pericarp.

The top picture is showing the pericarp structure of WVA106 at mature green stage. Scale bar = 500 μ m.

At the bottom is represented the ploidy map showing the frequency of nuclei by cell layer in the pericarp.

Figures modified and adapted from Cheniclet *et al.*, 2005 (A and H), Bourdon *et al.*, 2011 (I) and Renaudin *et al.*, 2017 (B to G)

The last phase starts when the fruit growth is mostly due to cell expansion in the mesocarp. Pericarp cell number does not change much from this stage. At the end of this phase the pericarp volume is ten times more important than at the start of this phase. This growth is mainly driven by cell expansion of the M' cells participating to 50% of the increase in size (Renaudin *et al.*, 2017). This cell expansion occurs simultaneously with a drastic increase in ploidy levels due to the occurrence of 2 to 3 rounds of endoreduplication (64C, 128C and 256C, Figure 10H). At the end of this phase corresponding to the mature green stage the fruit has reached its final size and the mean ploidy levels increased to 24 in cherry tomato and can be higher, up to 70, in other genotypes (Cheniclet *et al.*, 2005). After this period of growth, the fruit will start the ripening phase.

B) *Ploidy distribution in the pericarp and potential role of endoreduplication*

At mature green stage, the establishment of a ploidy map in the pericarp allowed to get a better view of the distribution of the ploidy levels inside this tissue (Figure 10I; Bourdon *et al.*, 2011). At this stage, the epidermal layers E1 and E2 are the only ones without endoreduplicated cells. In E1, 2C cells represent more than 60% of the cells whereas the other 40% are 4C cells (Figure 10I). In E2, 2C cells represent less than 10% of the cell population, 4C cells represent nearly 40% and the other 50% are half 8C and half 16C (Figure 10I). The last outer-epidermal layer, E3, is composed of 4C to 32C cells and 8C cells represent the most represented group. In the inner epidermis, I1 is composed of 8C to 32C cells and I2 of 16C to 256C cells (Figure 10I). The absence of 2C and 4C cells in the inner epidermis indicates that dividing activity has disappeared from these layers. Most of the cells in I1 layer are 16C whereas ploidy levels are nearly equally represented in I2 layer. Finally, M and M' layers are composed of the highest ploidy levels. In M', ploidy levels range from 8C to 256C but 8C cells are found only in the first M' layer and 16C only in the two first M' layers (Figure 10I). In addition, the measurement confirmed a positive correlation between ploidy and cell size (Bourdon *et al.*, 2011). Although this correlation is not perfect. It suggests that endoreduplication is more likely favoring the potential of growth by increasing the range of possible growth rather than driving cell growth (Bourdon *et al.*, 2011, Nafati *et al.*, 2011). Endoreduplication supporting rapid growth is also suggested by the presence of 8C nuclei at anthesis and rapid increase of 8C cells during the first days of development (Renaudin *et al.*, 2017). 8C cells before anthesis are hypothesized to be in the M layer because M layer attains higher ploidy levels at the end of the development (Renaudin *et al.*, 2017). The role of endoreduplication in Tomato fruit thus seems to be an enhancer of cell expansion since it is present before cell expansion starts, at anthesis, and both cell expansion and ploidy rise rapidly after anthesis in the mesocarp.

Finally, we have seen that before anthesis some cells in the ovary are already endoreduplicated that might be supporting the rapid cell expansion of mesocarp cells. In Tomato fruits, central mesocarp cells are large and highly endoreduplicated cells, are also accumulating a lot of soluble sugars and organic acids in vacuoles (Beauvoit *et al.*, 2014), thus endoreduplication could have a role in regulating metabolism of mesocarp cells to increase cell expansion through sugar accumulation in vacuoles.

II. ***Genetic regulation of endoreduplication in Plants***

1. ***Dynamic of CDK-CYC complexes during cell cycle progression***

The endocycle is a modified version of the classical mitotic cell cycle. The cell cycle in Eukaryotes is divided in four phases, Gap1 (G1), S (Synthesis), Gap2 (G2) and M (Mitosis) phases (Figure 1A). During the G1 phase the cell decides to progress into the cell cycle or not in function of environmental and internal signals. Favorable signals, such as temperature or water and hormones or sucrose allow the synthesis of the mRNA and proteins necessary for the subsequent cell cycle phases. After the G1 checkpoint and requirements are passed, the cell enters the S phase, during which the whole genome is duplicated through the production of an extra copy of each chromosome pair, thus passing from $2n$ to $4n$ (n is referring to the number of each chromosome). The S phase is followed by the G2 phase during which a verification of DNA integrity after duplication, named G2/M checkpoint, occurs and necessary components for the next step are produced. The M phase is the last step of the cell cycle and corresponds to the equivalent segregation of the chromosomes. The M phase ends with the cytokinesis leading to the formation of two daughter cells that will contain the same genetic information. The endocycle is only composed of a succession of G and S phases without M phase, resulting in an increase in cell ploidy levels and production of polytene chromosomes (Figure 1B, Vlieghe *et al.*, 2007).

During the cell cycle, the correct progression of each phase requires the action of heterodimeric complexes composed of a catalytic subunit referred to as CYCLIN-DEPENDENT KINASE (CDK) and a regulatory subunit, the CYCLIN (CYC). In CDK-CYC complexes, the CDK assures phosphorylation activities whereas the CYC defines the substrate specificity and the localization within the cell (Inzé and De Veylder, 2006). The high number of CDKs and CYCs proteins and the diversity of CDK-CYC complexes are crucial for the molecular regulation of the cell cycle and the endocycle. In Arabidopsis, the CDKs belong to a conserved family of serine/threonine protein kinases containing 12 members classified from A to F and CDKA and CDKB have been implicated to direct the regulation of the cell cycle (Gutierrez 2009) while CDKD and CDKF are CDK-activating kinases regulating small RNA processes and help to preserve mitotic activity in Arabidopsis (Hajheidari *et al.*, 2012, Takatsuka *et al.*, 2015) and CDKG are regulating mRNA splicing (Zheng *et al.*, 2014). In Arabidopsis, the CYC family is a large family of around 30 members categorized in four sub-families, CYCLINS A, B, D and H, generally involved in G1/S and S phase of the cell cycle (Gutierrez 2009) but the role of each cyclin has not clearly been established. Some have been associated to mitosis regulation such as CYCB1;1, CYCB1;2, CYCA2;3, CYCA3;1 and CYCD3;1 (De Veylder *et al.*, 2011, Gutierrez 2009) or to meiosis regulation such as for example CYCA1;2, CYCA2;1, CYCA3;2 and CYCB3;1 (Bulankova *et al.*, 2013).

CDK-CYC complexes have an essential role in the control and the maintenance of the cell cycle preventing endocycle. Consequently, the repression of their function in the cell is essential for the shift from the cell cycle to the endocycle (Edgar *et al.*, 2014).

2. Transition from the cell cycle to the endocycle

The regulation of the CDK-CYC activities is done through transcriptional regulation, phosphorylation, the binding of protein inhibitors and proteolysis (Gutierrez, 2009; Edgar *et al.*, 2014).

In plants, the main studies about the regulation of endoreduplication have been done on Arabidopsis trichomes and roots and have identified several regulating actors. Some information regarding the molecular regulation of endoreduplication is also available for other plants such as Tomato and Medicago. Here, I will present the main molecular regulators of the endocycle starting with the transition from cell cycle to endocycle, then the maintenance of the endocycle and to finish the exit from the endocycle.

To undergo endoreduplication, the cells need to exit the mitotic cycle. This transition is permitted by a reduction of the CDK-CYC complexes activities.

A) Regulation of CDK-CYC complexes by proteolysis

The degradation of the CYC by the proteasome is a first level of control of the CDK-CYC complex. This degradation is done through the action of the Ubiquitin-Proteasome system (UPS). The UPS is composed of three enzymes, the Ubiquitin-activating enzyme E1 that activates ubiquitin to be transferred to the Ubiquitin-conjugating E2 to then bind with the Ubiquitin Ligase E3 that will finally transfer the ubiquitin to target the protein to the proteasome.

The ANAPHASE PROMOTING COMPLEX/CYCLOSOME (APC/C) is an E3-ubiquitin ligase complex that recruits E2 and an activator protein responsible for substrate specificity. Examples of these activators are the CELL CYCLE SWITCH 52 (CCS52) proteins (APC/C^{CCS52} abbreviation represents an APC/C complex loaded with a CCS52 protein) involved in cell cycle and endocycle regulation. Three CCS52 proteins have been identified in Arabidopsis, CCS52A1, CCS52A2 and CCS52B (Fülop *et al.*, 2005, Lima *et al.*, 2010) and so far only CCS52A1 and CCS52B have been identified in Tomato (Mathieu-Rivet *et al.*, 2010). CCS52A are able to induce endoreduplication by targeting the mitotic CYCLINs, CYCB1, CYCB2 and CYCA2, for degradation by the UPS, as described in several plants such as Medicago, Tomato or Arabidopsis and different cell types as for example the trichomes of Arabidopsis, expanding root cells or the pericarp of tomato (Cebolla *et al.*, 1999, De Veylder *et al.*, 2011, Baloban *et al.*, 2013, Mathieu-Rivet *et al.*, 2011). In Arabidopsis, loss of function of *AtCCS52A1* results in reduced ploidy levels and accumulation of mitotic CYCLINs in non-dividing cells such as the trichomes and root hair cells. Overexpression of CCS52A in Medicago nodules (*MtCCS52A*) and in Tomato (*SICCS52A*) results in increased ploidy levels and cell size, reduction of cell number and reduced organ size compared to wild type (Larson-Rabin *et al.*, 2009, Vinardell *et al.*, 2003, Mathieu-Rivet *et al.*, 2011). *AtCCS52A1* activity may be inhibited by the activity of the Arabidopsis ULTRAVIOLET-B INSENSITIVE4 (UVI4) after forming a complex with the UBIQUITIN-SPECIFIC PROTEASE14 (UBP14) that is inhibiting endoreduplication by stabilizing the mitotic CDK/CYC complex (Heyman *et al.*, 2011 and Xu *et al.*, 2016). Loss of function plants *Atuvi4* and *Atubp14* present similar phenotypes corresponding to reduced meristem size, increased ploidy levels in roots, reduced root length and increased ploidy levels in trichomes (Heyman *et al.*,

2011, Xu *et al.*, 2016). The exact mechanism of how UBP14 and UVI4 are inhibiting the APC/C activity is still unknown.

In mature leaves, AtCCS52A2 is supporting endoreduplication onset since *ccs52a2* plants have reduced ploidy levels (Lammens *et al.*, 2008). However, the targets of CCS52A2 have not been described. In the root, CCS52A2 is not implicated in the transition into the endocycle and seems to be involved in the inhibition of cell division to assure the maintenance of the root quiescent center, a structure of mitotically inactive cells, which role is to maintain the stem cell fate of the surrounding cells (Verstraelen *et al.*, 2009). Finally, among the three CCS52 proteins, CCS52B seems to be implicated in the regulation of cell division by promoting the degradation of cyclins just before cytokinesis allowing reentering a new cell cycle (Yang *et al.*, 2017).

B) Regulation of CDK-CYC by phosphorylation

A second level of control of the mitotic CDK-CYC complexes is phosphorylation. WEE1, a protein kinase implicated in the phosphorylation of the mitotic CDK/CYC at the G2/M checkpoint, favoring the transition from the cell cycle to the endocycle. In Tomato, *S/WEE1* phosphorylates the CDKA of the mitotic CDK/CYC complexes after the S-phase (Gonzalez *et al.*, 2007). This regulation is important for Tomato plant development and its impairment results in overall plant and fruit size reduction as well as cell size and ploidy level reduction (Gonzalez *et al.*, 2007). However, *AtWEE1*-deficient plants grow normally under optimal conditions (Schutter *et al.*, 2007). In Arabidopsis, *AtWEE1* is rather implicated in more specific responses, such as response to both abiotic and biotic stresses, including DNA damage or nematode infection (Cabral *et al.*, 2020). In *late meristem identity 1 (lmi1)* mutant plants, a transcription factor from the homeodomain zipper class I family (HD-ZIP) is mutated and plants are affected in leaf growth with failure to form serration. *lmi1* plants also have cells with reduced endoreduplication and reduced expression of cell cycle related genes including *WEE1* (Vuolo *et al.*, 2019). Expression of genes favoring endoreduplication under the control of *LMI1* promoter, such as *CCS52A* or *WEE1*, was sufficient to rescue the phenotype of *lmi1*. Moreover, *LMI1* is able to induce *WEE1* expression by binding the promoter region of *WEE1* (Vuolo *et al.*, 2019). These findings might be indicating that *AtWEE1* also plays a role in leaf morphology in Arabidopsis through *LMI1* regulation (Vuolo *et al.*, 2019).

C) Binding of Cyclin kinase inhibitor proteins

Cyclin Kinase Inhibitors (CKI) mediate another control of the CDK-CYC activity by interacting with the CDKs thus inhibiting their ATP binding activity (Russo *et al.*, 1998). So far, two families of CKI have been described in plants: the INTERACTOR/INHIBITOR OF CDK/KIP-RELATED PROTEINs (ICK/KRPs), homolog of the KIP proteins in Mammals (Torres Acosta *et al.*, 2011), that are involved in regulating the cell cycle (Nakayama *et al.*, 1999), and the plant specific SIAMESE-RELATED PROTEINs (SMRs) (Kumar and Larkin 2017) involved in regulating G2 checkpoint. Both KRP and SMR are sharing a C-terminal motif that allows binding to cyclins (Churchman *et al.*, 2006) and recently N-terminal motifs have been identified as CDK binding site and as potential phosphorylation site both in SIM and SMR1 (Kumar *et al.*, 2018 and Dubois *et al.*, 2018)

KRP protein family is composed by 7 members in Arabidopsis and a kinematic analysis of cell division and cell expansion in the triple mutant *krp4/6/7* revealed that KRPs' might be more associated to the regulation of the rate of cell cycle and endocycle than in the regulation of the transition between these processes (Sizani *et al.*, 2019, see *Chapter 1.II.3 Maintenance and ending of the endocycle*).

The SMR protein family is composed of 14 members in Arabidopsis. The first described member of the SMR family is SIM, which when mutated leads to the formation of multicellular and clusters of trichomes while in the wild type trichomes are large endoreduplicated cells distant from other trichomes. On the other hand, SIM overexpressing lines have a severe dwarfism phenotype with larger leaf epidermal cells than WT plants. *sim* plants also have reduced levels of endoreduplication in leaves, hypocotyl and trichomes compared to WT (Walker *et al.*, 2000) while SIM overexpression resulted in increased of ploidy levels (Churchman *et al.*, 2006). Thus, SIM seems to promote endoreduplication and might be acting by inhibiting the cell cycle (Churchman *et al.*, 2006, Van Leene *et al.*, 2010, Kumar *et al.*, 2018) through the CDKA-CYCD complex, responsible for the G1/S transition, (Gutierrez; 2005) or CDKB complexes activities, necessary for the G2/M transition (Boudolf *et al.*, 2004).

Atsmr1 loss of function plants display reduced ploidy levels and cell sizes in leaves and sepals but not in trichomes where SIM is expressed, protecting against reduction of ploidy levels (Roeder *et al.*, 2010). SMR1 is also named LGO (LOSS OF GIANT CELLS FROM ORGANS). Overexpression of LGO under the control of a sepal epidermis specific promoter triggers the production of sepals covered with endoreduplicated giant cells (Schwarz and Roeder, 2016). Interestingly, LGO's effect on the entry into endocycles is dosage-dependent (Robinson *et al.*, 2018). *AtSMR5* and *AtSMR7* can also induce endoreduplication when overexpressed and are implicated in DNA damage response (Yi *et al.*, 2014).

D) *Transcriptional regulation of CCS52*

Lastly, the transition from the cell cycle to the endocycle is also transcriptionally regulated mainly through the control of *CCS52* genes expression.

E2F proteins form a family of transcription factors with six members in Arabidopsis named E2Fa to E2Ff and are implicated in the regulation of cell cycle progression (Gutierrez 2009). E2F family is divided in two groups with typical E2Fs, E2Fa to E2Fc, and atypical E2F proteins, E2Fd to E2Ff, also named DEL1 to DEL3, with DEL referring to DP-E2F-like (Ramirez-Parra *et al.*, 2007). Typical E2F proteins require heterodimerization with DP proteins to assure their function (Gutierrez 2009) while atypical do not need dimerization with DP to regulate gene expression (Lammens *et al.*, 2008). E2Fa, E2Fb and E2Fc can be bound by the RETINOBLASTOMA RELATED 1 (RBR1) protein that modulates E2F activities. When RBR1 binds to E2Fa, the expression of both *CCS52A1* and *CCS52A2* is repressed thus inhibiting the entry into the endocycle (Magyar *et al.*, 2012). The atypical E2F protein, *AtDEL1*, can bind to the promoter of *AtCCS52A2* and inhibit its expression resulting in the entry into the endocycle (Lammens *et al.*, 2008). *AtDEL1* interacts with a mediator subunit, *AtMED16*, to inhibit *AtCCS52A2* expression (Liu *et al.*, 2019). The role of mediator proteins is to be a bridge

between transcription factors and the RNA polymerase II (RNA pol II) and work as a coactivator of RNA-pol II to bind DNA and perform transcription. *AtMED16* does not only interact with *AtDEL1* to regulate *AtCCS52A2* expression but is also able to bind *AtCCS52A1* promoter and repress its expression. However, the partner of *AtMED16* in *AtCCS52A1* regulation is unknown (Liu *et al.*, 2019).

Interestingly, *AtDEL1* is regulated in a light dependent manner in the hypocotyl by a chain of regulation that involves transcriptional regulation by other E2F proteins, *AtE2Fb* and *AtE2Fc* and by proteolysis by the action of the *A.thaliana* CONSTITUTIVE PHOTOMORPHOGENIC 1 (*COP1*) an E3 ubiquitin ligase, that targets *E2Fb* to degradation in dark condition (Gendreau *et al.*, 1998, Berckmans *et al.*, 2011). In the dark, *AtCOP1* triggers *AtE2Fb* degradation, thus permitting *AtE2Fc* to repress *AtDEL1* transcription and consequently stimulate endoreduplication. In light conditions, *AtE2Fb* activates *AtDEL1* that can thus limit endoreduplication (Berckmans *et al.*, 2011).

Transcriptional regulation is also important for the entry into endoreduplication in the transition zone in *Arabidopsis* roots. Cells in the transition zone, as described before, have a low auxin content that is due to regulation of auxin polar transport and the synthesis of auxin degradation proteins upon positive regulation of cytokinin (Di Mambro *et al.*, 2017). In this transition zone, *AtCCS52A1* gene expression is positively regulated by a type-B ARABIDOPSIS RESPONSE REGULATOR (ARR) transcription factor, *AtARR2* that regulates transcription in response to cytokinin (Takahashi *et al.*, 2013).

In Tomato, down-regulation of the transcriptional regulator *SlIAA17* leads to an increase of the size and ploidy levels of the fruit pericarp cells resulting in increased final size of the fruit (Su *et al.*, 2014). Aux/IAA proteins form heterodimers with AUXIN RESPONSE FACTOR (ARF) proteins that are transcription factors regulating auxin dependent genes expression. When auxin content is low, then Aux/IAA binds to ARF thus blocking auxin responsive gene expression. In high auxin concentration, ARF are released from Aux/IAA binding allowing gene expression and Aux/IAA are bound by an auxin inducible E3 ubiquitin ligase triggering Aux/IAA proteolysis (Li *et al.*, 2016). The increase in mean ploidy level in pericarp cells in the *SlIAA17* mutant indicates that auxin signaling could limit the entry into the endocycle in Tomato fruit but the molecular mechanisms by which *SlIAA17* is regulating endoreduplication are unknown. (Su *et al.*, 2014).

In conclusion, altogether these data show that the exit from the cell cycle and entry into the endocycle is induced by several mechanisms of repression targeting CDK-CYC complexes. The degradation of the CDK-CYC complexes is also tightly regulated by transcriptional control of *CCS52* genes and seems to be subjected to hormonal regulation as well.

3. Maintenance and ending of the endocycle.

The molecular regulation of the transition from the cell cycle to the endocycle has been studied more extensively than the progression and maintenance of this process.

The main hypothesis regarding the regulation of the maintenance of the endocycle is a cyclic regulation inducing oscillations of the CDK-CYC activity during the G/S phases (Verkest *et al.*, 2003, Lee *et al.*, 2009, Breuer *et al.*, 2014). These oscillations in CDK-CYC activity are hypothesized to be controlled by KRP proteins (De Veylder *et al.*, 2011). A strong overexpression of AtKRP2 inhibits both cell division and endoreduplication while weaker overexpression results in an inhibition of mitosis only suggesting (Verkest *et al.*, 2005). Single *krp* knockout in Arabidopsis results in plants with no difference compared to wild type probably due to redundancy in the function of the KRP family members (Cheng *et al.*, 2013, Sizani *et al.*, 2019). However, multiple *krp* knockout plants, *krp4/6/7* and *krp1/2/3/4/7*, present an increased seed size, cotyledon size and leaf size due to an increase in the number of cells with higher number of endocycles (Sizani *et al.*, 2019, Cheng *et al.*, 2013). CDK activity is upregulated in these multiple *krp* mutants while total amount of KRP transcripts in *krp4/6/7* is around 35% less than WT indicating the importance of KRP in the cell cycle process (Sizani *et al.*, 2019). Taken together, these data led to hypothesize that KRP are global moderators of both cell cycle and endocycle process but not actors in the transition between cell division and endoreduplication (Sizani *et al.*, 2019). In Tomato, fruit specific overexpression of *SKRP1* results in a reduction of ploidy levels in the fruit showing that also in Tomato, KRP inhibits endoreduplication.

Overexpression of CYCA2 represses the progression of the endocycle by inhibiting the pre-replication complex during the S phase (Yu *et al.*, 2003 and Imai *et al.*, 2006). Interestingly, CYCA2 is repressed by ILP1 (INCREASE LEVEL OF POLYPLOIDY 1) a protein hypothesized to be implicated in the maintenance of the endocycle (Yoshizumi *et al.*, 2006). *Ilp1* plants have reduced ploidy levels whereas ILP1 overexpression induces more endoreduplication than in wild-type plants (Yoshizumi *et al.*, 2006).

In Arabidopsis trichomes, the end of the endoreduplication is controlled by the transcriptional regulation of CCS52A1 by a GT-2 like plant specific trihelix transcription factor (AtGTL1). Indeed, the loss of function of *Atgtl1* results in the overexpression of *AtCCS52A1* and *AtCCS52A2*, in increase ploidy levels (Breuer *et al.*, 2009) and trichomes sizes while a trichome specific overexpression results in the absence of trichomes (Breuer *et al.*, 2012). *AtGTL1* is expressed in the late trichome development steps and limits trichome growth and endoreduplication (Breuer *et al.*, 2009; 2012). However, the endoreduplication and growth finally stops in the *Atgtl1* mutant suggesting that other regulation steps assure the arrest of growth and endoreduplication processes.

In conclusion, the molecular regulation of endoreduplication is occurring at the gene expression and protein activity level from the onset to the end of the endocycle. CCS52 and SMR proteins seem to be central in the regulation of the entry into the endocycle, while KRP and transcriptional regulation of cyclins direct the maintenance of the endocycle. Finally, in trichome, the endocycle end is controlled by the transcriptional repression of CCS52 thus favoring the stabilization of CDK-CYC complexes that repress endocycle progression. As revealed by the previous results I presented, most of the molecular control of endoreduplication has been studied in Arabidopsis and is not yet well established in other plants.

4. Genetic manipulation of endoreduplication in other plant than *Arabidopsis*

In this part, I will present additional examples of genetic modifications of endoreduplication in different plants and tissues than *Arabidopsis* and their consequences on growth or other processes helping to better understand endoreduplication function.

In Maize endosperm, a decrease of ploidy levels obtained with a dominant negative approach overexpressing a CDKA unable to perform kinase activity, was not associated to any change in cell size or kernel size but to slightly decreasing zein accumulation and gene expression related to starch biosynthesis (Leiva-Neto *et al.*, 2004). In Maize endosperm, starch accumulation takes place in the endoreduplicated cells, but in transgenic plants with a reduced CDKA activity and lower endoreduplication levels, the accumulation of starch is unmodified. This might be due to the plasticity of the tissues and suggesting the possibility to uncouple endoreduplication and the sugar metabolism of endoreduplicated cells in the Maize endosperm (Sabelli and Larkins 2009). Moreover, the increase of endoreduplication in Maize endosperm, by the overexpression of an antisense RNA targeting RBR1 (RBR1^{AS}) does lead to change in a decrease in cell size and nuclear size as well as an increase in cell number suggesting a role of RBR1 in coupling cell size and ploidy level. But it does not lead to changes in kernel size and no changes of seed weight or Zein protein content despite a decrease in non zein protein content per DNA quantity (Sabelli *et al.*, 2013) possibly suggesting that extra DNA copies are not transcribed. Moreover, increase in Programmed Cell Death (PCD) activity has been reported at 16DAP stage of the kernel development, when PCD is normally starting (Sabelli *et al.*, 2013). However, this observation could not be linked either to a regulatory role of RBR1 on this process or an endoreduplication related regulation. These observations led to conclude that CDKA and RBR1 can influence endoreduplication in Maize endosperm without influencing the phenotype of the endosperm suggesting the absence of link between endoreduplication and cell size or the increase sugar metabolism activity in endosperm cells. However, the authors hypothesized that the absence of phenotype in the RBR1^{AS} plant could be due to compensatory mechanisms by regulation of gene expression. It has been observed in RBR1AS plant that despite higher ploidy levels measured, less protein is produced compared to the DNA quantity that is found. This is suggesting that newly formed alleles are less transcribed and since they found DNA METHYLTRANSFERASE (MET1) to be up regulated in RBR1AS plant, it has been proposed that change in chromatin conformation could be responsible of reduction of transcription per allele (Sabelli *et al.*, 2013).

In Tomato, the downregulation of Tomato ortholog of *Arabidopsis* regulators of endoreduplication, SIWEE1 and SICCS52A1 results in overall plant and fruit size reduction with reduced endoreduplication levels (Gonzalez *et al.*, 2007, Mathieu-Rivet *et al.*, 2010). Moreover, in these plants, fruit pericarp is thinner due to the presence of smaller cells in the fruit of these plants (Gonzalez *et al.*, 2007, Mathieu-Rivet *et al.*, 2010). SIKRP1 overexpression (SIKRP1-OE) specifically in fruit expanding cells resulted in reduction of endoreduplication in the pericarp. No fruit size nor cell size nor

metabolic differences have been observed probably because compensation processes took place (Nafati *et al.*, 2011). The absence of changes of cell size shows that reduction of ploidy levels can be dissociated from cell expansion in the pericarp. In addition, endoreduplication is reduced in SIKRP1-OE plants but not abolished. This might be indicating that endoreduplication is necessary for fruit growth and that the number of remaining endoreduplication rounds might be sufficient to maintain fruit size, cell size and metabolism.

The genetic manipulation of endoreduplication in Maize and Tomato allowed to hypothesize that this process could be necessary to at least endosperm development and fruit pericarp development. Finally, in Tomato, SIKRP1 overexpression is able to uncouple endoreduplication and cell size because although endoreduplication is reduced the mean cell size seems unaffected (Nafati *et al.*, 2011). This could be signifying that other regulation is affecting cell size not under the control of endoreduplication or that the reduced ploidy levels are sufficient to attain the necessary cell size for normal development of the pericarp.

Endoreduplication takes place in several tissues of multiple plant species and is often associated with specialized role of cells or enlarge cell size. These effects of endoreduplication, at cellular and organ level, could result from modifications of transcription upon genome duplication. In the following section, I will focus on what is known about the effect of endoreduplication on changes in gene expression.

III. Effect of endoreduplication on the regulation of transcription

1. Global Effect on gene expression in polyploid plants

Endoreduplication corresponds to the duplication of chromatins in some cells of an organ whereas polyploidy is the heritable condition of possessing more than two complete sets of chromosomes. Polyploidy can occur after a cross between two plants of the same species (Autopolyploidy) or of two close species (Allopolyploidy). In both cases, the number of chromosomes in the cells is changed compared to the parentals and the polyploid plant contains additional copies of each chromosome pairs while endoreduplicated cells contain only one pair of each chromosome.

In general, in a diploid plant and non-polyploid cell, we can consider that each allele of a gene is equally contributing to the absolute quantity of the corresponding transcribed mRNA (Coate and Doyle 2010). In autopolyploid and allopolyploid organisms each chromosome is duplicated while in endoreduplicated cells, each chromatid is duplicated thus increasing the number of alleles for each gene. This changes in chromosome or chromatid number raises several questions regarding transcription regulation. Are polyploid or endoreduplicated cells maintaining the same absolute mRNA quantity than diploid cells? Is each allele maintaining a diploid expression so is there a doubling of expression of genes by doubling allele number? For now, these questions have been more covered in the case of polyploidy (for review see Doyle and Coate 2018) than in endoreduplication and no clear answers can be found to these questions.

The allopolyploid *Glycine dolichocarpa*, and its diploid parents, *Glycine tomentella* and *Glycine syndetika*, are wild relative family of soybean have been studied for global expression level, transcriptome size, as a function of genome size. It was shown that transcriptome size of the tetraploid plant was 1.4x bigger than the parents but proportionally to the number of alleles, it was smaller by $\sim 0.7x$, thus corresponding to a 30% decrease of expression per diploid genome compared to the diploid parents (Coate and Doyle 2010). The same range of results have been found in synthetic tetraploids of *Arabidopsis* with an increase of the transcriptome size of 1.4x to 1.7x in tetraploid compared to the diploid and a $\sim 30\%$ decrease of the expression by diploid genome (Robinson *et al.*, 2018, MJ.Song *et al.*, 2020). In *Arabidopsis* female gametophyte the non endoreduplicated egg cells and central cells with around 10 to 30% of decrease of expression per diploid genome (Q.Song *et al.*, 2020). In polyploid plants, global transcriptome analyses thus showed that both total RNA and mRNA are increased with polyploidization and this increase tends to not follow a 1:1 ratio (doubling of expression, Robinson *et al.*, 2018). The mechanisms controlling the dosage of gene expression relative to genome duplication during polyploidization are not known.

2. Global effect on gene expression of endoreduplication

In the case of endoreduplication, the genome is duplicated through the increase of the number of chromatids per chromosome and thus the copy number of genes is increased triggering an increased transcriptional potential. Observations tend to favor the hypothesis that in the case of cell differentiation, endoreduplication is supporting an increase in metabolic activity by increasing transcription through duplication of DNA (Sabelli *et al.*, 2013 Frawley and Orr-Weaver 2015) and that expression is likely to increase with ploidy levels but not follow a linear proportionality (Doyle and Coate 2018). However, even if endoreduplication has been studied for many years and often hypothesized to increase gene expression, to date only few studies have quantified how much gene expression is changing with this process.

In tomato pericarp endoreduplicated cells, transcription of a subset of genes has been positively correlated with endoreduplication. The quantification of gene expression of *rRNA 5.8S*, *SIWEE1*, *SICCS52A*, and RNA pol II subunit (SIRPB) have shown that the relative expression of these genes increases when endoreduplication increases (Bourdon *et al.*, 2012). Moreover, for RNA pol II, the protein quantity of both active and inactive forms, quantified by immunofluorescence, doubled with the doubling of the ploidy level (Bourdon *et al.*, 2012). This increase in transcription and translation of the RNA pol II led to hypothesize that overall transcription and translation activities could be doubled with endoreduplication in Tomato pericarp cells since RNA pol II is responsible for transcription activities (Bourdon *et al.*, 2012).

Based on the results described above showing a potential doubling of expression with genome duplication (Bourdon *et al.*, 2012), a ploidy-based transcriptome of Tomato pericarp nuclei has been obtained using fluorescence assisted nucleus sorting and RNA sequencing in order to tackle the question of the proportional change of expression following increase of ploidy (Pirrello *et al.*, 2018). In the study, the four main endoreduplication levels observed in 30DPA fully grown fruit before ripening, namely

4C, 8C, 16C and 32C, were studied (Pirrello *et al.*, 2018). This study shows that the gene expression is globally increased between ploidy levels and this increase corresponds to a factor 2 between consecutive ploidy levels for most of the genes. This global shift of expression was found for 4694 genes having a proportionally increased expression between the four endoreduplication levels studied. Additional experiments implying other species or tissues could help understanding if the shift is a universal process or if it is specific to tomato pericarp.

A single cell RNA seq study of the whole *Arabidopsis* root was done to study developmental trajectories of cell lineages since whole roots contain mix of cells at varying developmental stages (Jean-Baptiste *et al.*, 2019). The analysis of single cell transcriptomic data allows to place cells in “pseudotime” order along a trajectory that describes their maturity. This positioning could be done for root hair cells that at mature stage are endoreduplicated. The analysis showed that the median total mRNA quantity profile is decreasing while pseudotime thus differentiation and endoreduplication increases. On the other hand, this median profile increases with pseudotime, thus ploidy, for hair cell specific genes (Figure 11A). Thus, differentiating root hair cells have increased the expression of root hair cell specific genes and reduced expression of unspecific genes. In addition, the authors showed that transcription rate of mRNA specific for hair cells is increasing across pseudotime thus when ploidy increases (Figure 11B, Jean-Baptiste *et al.*, 2019).

In conclusion, endoreduplication seems to be able to induce global shift of expression in Tomato and in *Arabidopsis* root hair cells. Endoreduplication could also be related to an increase in transcription rate. In both cases endoreduplication could influence mRNA transcription. However, mostly mRNA have been assayed while it is representing only 5% of total RNAs. It could be interesting to know much about the dynamic of other RNA pools (rRNA, tRNA and non-coding RNA) that could vary since 5.8S rRNA transcript abundance in Tomato pericarp is following the increase in ploidy levels (Bourdon *et al.*, 2012).

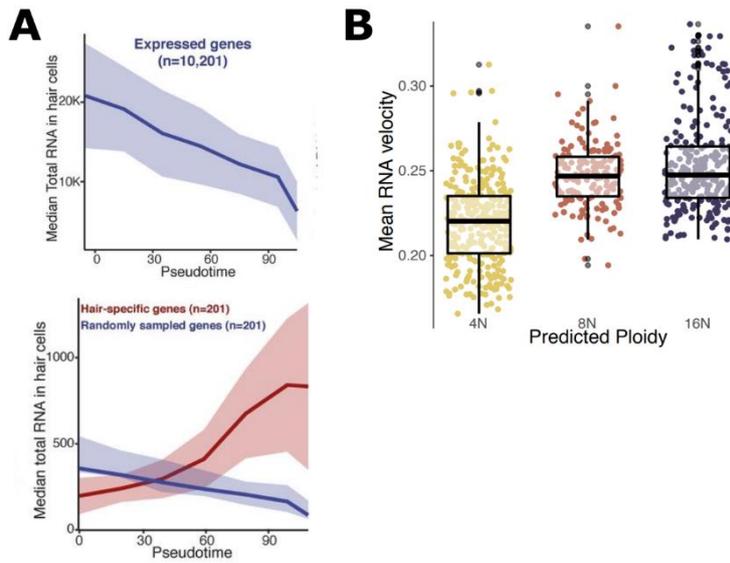


Figure 11. Changes of gene expression in endoreduplicated root hair cells

(A) Top panel: Hair cells median total mRNA is decreasing across pseudotime (predicted developmental time, for hair cell from low endoreduplicated/differentiated to highly endoreduplicated/differentiated). n is the number of genes used to get Median total mRNA.

Low panel: Comparison of median total mRNA for hair-cell-specific genes (in red) to a comparable random set of genes (in blue).

(B) Compared mRNA production rate across ploidy levels.

(A) and (B) are adapted and modified from Jean Baptiste *et al.*, 2019.

3. Ploidy specific gene expression in endoreduplicated cells

Endoreduplication generally occurs in specialized cells and can induce ploidy specific gene expression. Hereafter, I review examples in which increase expression in specialized genes might be necessary for differentiation and seem to be related to endoreduplication.

In Maize endosperm, endoreduplication occurs in the central part of the tissue. Endoreduplicated cells represent most of the endosperm cells at 10-12 DAP, corresponding to almost mid-time of endosperm development (Schweizer *et al.*, 1995; Leiva-Neto *et al.*, 2004). At this timepoint, more than 30% of the total transcripts correspond to around 100 of the most expressed genes, the *ZEIN* genes encoding the most important storage proteins in kernel (Figure 12A Chen *et al.*, 2014). This observation suggests that this increase of *ZEIN* genes expression is related to the increase in ploidy levels (Leiva-Neto *et al.*, 2004). Moreover, *ZEIN* transcripts, represent almost 60% of all the transcripts in the endosperm between 10 to 20DAP, when ploidy levels are increasing and endoreduplicated cells represent most of the endosperm cells (Chen *et al.*, 2014). Between 10 to 20DAP, *STARCH SYNTHASE* genes are upregulated, reinforcing the potential the role of endoreduplicated cells in storage function (Chen *et al.*, 2014, Sabelli *et al.*, 2012).

Upregulation of genes implicated in specific metabolism in endoreduplicated cells is not occurring only in Maize endosperm but also in the Arabidopsis root. In this organ, ploidy specific transcriptome was obtained by RNA-seq and the data analysis has identified genes differentially expressed as a function of ploidy (Bhosale *et al.*, 2018). Enrichment analysis of gene ontology terms related to biological processes on the identified differentially expressed genes revealed that genes highly expressed in endoreduplicated cells are related to chromosome organization and DNA methylation pattern, two processes known to influence gene expression (see further paragraphs, Bhosale *et al.*, 2018) that might be responsible of a global shift of expression. The analysis of this root ploidy specific transcriptome has also shown specific ploidy expression of hormones and stress responsive genes. For example, genes related to auxin response or to cold response are more expressed in 2C cells while ethylene response and water deprivation response genes are more expressed in 8C and 16C cells (Figure 12B, Bhosale *et al.*, 2018). These findings suggest that tolerance to stresses and hormonal signaling might be regulated and programmed by endoreduplication during root development. Moreover, expression of cell wall synthesis genes have also been found to be specifically regulated by ploidy levels, thus endoreduplication might be an answer to high demand of cell wall materials for growing cells and in response to stresses (Bhosale *et al.*, 2018, Bhosale *et al.*, 2019).

In the transcriptome of microdissected areas of different part of Medicago nodules, characterized by composition of cells with different ploidy levels, it was shown that low endoreduplicated cells express genes related to positive regulation of endoreduplication and DNA replication (Nagymihály *et al.*, 2017) and highly endoreduplicated cells express nodule-specific cysteine rich peptides (NCR) that are peptides implicated in nodule development and necessary for symbiotic cell differentiation (Maunoury *et al.*, 2010, Roux *et al.*, 2014, Nagymihály *et al.*, 2017,

Montiel *et al.*, 2017). Highly endoreduplicated cells express NCR and genes encoding ions, sugars and amino acids transporters (Nagymihály *et al.*, 2017). *Mesembryanthemum crystallinum*, a salt tolerant halophyte plant, possesses modified trichomes called epidermal bladder cells (EBC) that are specialized in salt accumulation and are highly endoreduplicated (Barkla *et al.*, 2018). In these large cells, present on the whole areal tissues, developmentally programmed endoreduplication can reach levels of 128C to 1024C. In flower buds, ploidy levels can increase up to 65536C after salt stress, corresponding to 2 to 6 supplementary rounds of endoreduplication (Barkla *et al.*, 2018). Compared to non endoreduplicated roots cells, EBC express genes related to endoreduplication progression and DNA replication such as F box-like17 (FBL17), Mini Chromosome Maintenance (MCM) and E2F target gene1 (Barkla *et al.*, 2018) and maintain low levels of transcripts of inhibitors of endoreduplication such as DEL1, UVI4 and TEOSINTE BRANCHED1/CYCLOIDEA/PCF 15 (TCP15) (Barkla *et al.*, 2018). After salt stress, upregulation of genes implicated in ion transport, ABA signaling, inositol and proline metabolism and conversion from C3 to Crassulacean acid metabolism (CAM, Oh *et al.*, 2015) was observed. This expression reprogramming is supporting the hypothesis that endoreduplication could be a general way to promote metabolism in response to environmental stress leading to production of mitigating stress metabolites. It has been shown that the increase in ploidy is often accompanied by the production of secondary metabolites in the case of stresses such as phenylpropanoid and antioxidant (phenolic compounds and NADPH, Scholes and Paige 2015) in response to stresses.

Finally, in tomato pericarp, ploidy specific expression analysis showed that endoreduplicated cells in fully grown fruit, but not yet mature express genes related to carbohydrate metabolism, transport, and cell organization (Figure 12C, Pirrello *et al.*, 2018).

A comparison of the transcriptome of nuclei of different ploidy level from tomato pericarp of fully grown fruit, revealed a total of 1211 differentially expressed genes have been found (Pirrello *et al.*, 2011). Those genes have expression profiles that are related to specific ploidy levels possibly revealing different cellular functions in the pericarp. In cells with low ploidy levels, differentially expressed genes are related to photosynthesis and flavonoids pathways whereas highly polyploid cells are expressing genes related to carbohydrate metabolism. These findings suggest that endoreduplication in tomato pericarp could be an important factor of cell differentiation (Figure 12C, Pirrello *et al.*, 2018). It has been shown that pericarp cell expansion is driven by the expansion of vacuole due to accumulation of sugars and organic acids that led to influx of water (Beauvoit *et al.*, 2014). This could be suggesting that endoreduplication in highly endoreduplicated cells is accompanying cell expansion by increasing expression of genes implicated in carbohydrate metabolism.

In conclusion, it appears that through the doubling of all alleles at each cycle endoreduplication potentialize the increase of gene expression, allowing a global shift of transcriptome and the increase of expression of specific genes related to the differentiation of the cell: storage, symbiotic development, sugar metabolism, response to stress nodule peptides.

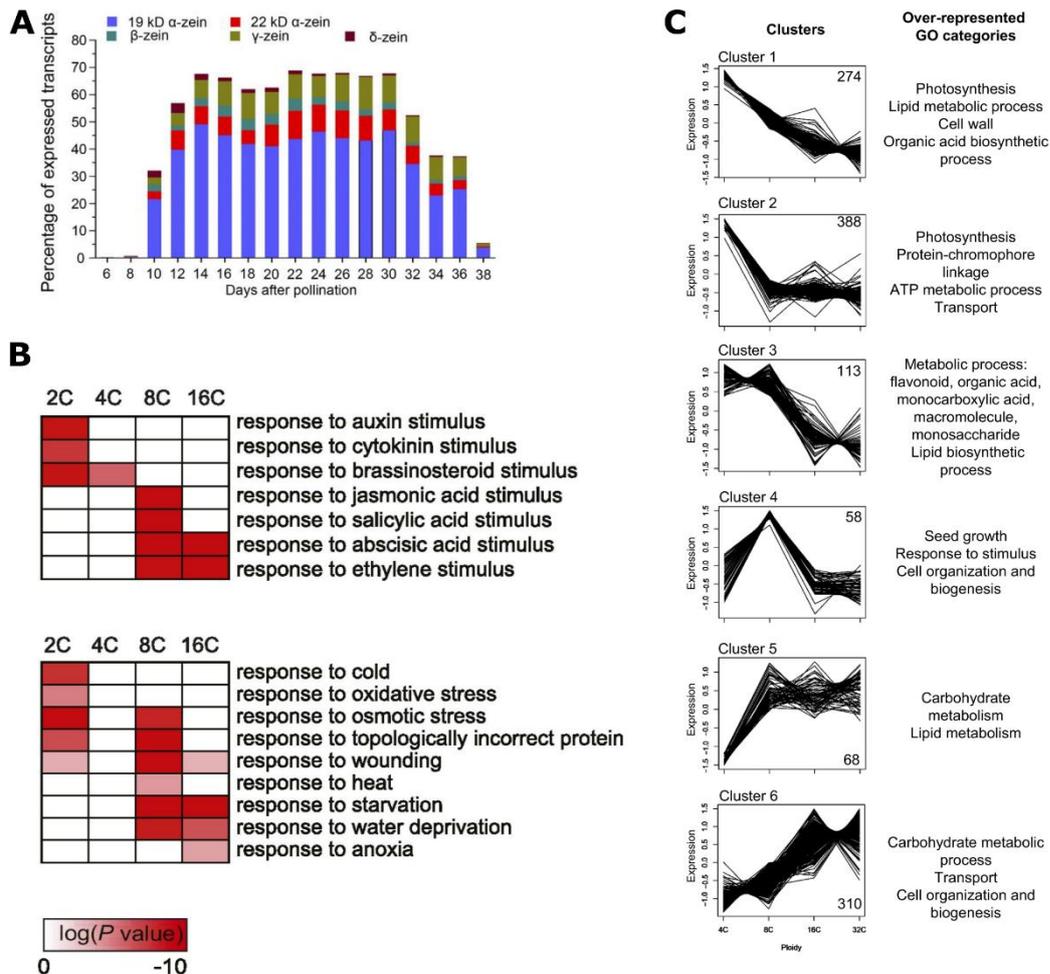


Figure 12. Endoreduplication can induce high expression of specialized genes in Maize, Arabidopsis and Tomato.

(A) Analysis of *ZEIN* expressed genes in the Maize endosperm. The expression is represented as a percentage of the total detected gene transcript.

(B) Functional enrichment of transcripts peaking at a given endoploidy level for hormone and stress response go categories in Arabidopsis root cortex cells.

(C) Expression profiles of differentially expressed genes clusters found in Tomato pericarp sorted nuclei.

(A) is adapted and modified from Leiva-Neto *et al.*, 2004. (B) is adapted and modified from Bhosale *et al.*, 2018. (C) is adapted and modified from Pirrello *et al.*, 2018

A) *Gene expression regulation mechanisms.*

Endoreduplication is associated with a general increase of gene expression in some plants and with specific patterns of expression as observed in *Arabidopsis*, *Medicago* and *Tomato* (Bhosale *et al.*, 2018, Nagymihály *et al.*, 2017, Pirrello *et al.*, 2018).

In Eukaryotes, genes expression can be regulated at the nucleus level by a specific spatial positioning of the chromosomes within the nucleus (Guerreiro and Kind 2019), and at the chromatin level with changes of chromatin condensation (Liu *et al.*, 2019), due to histones modification or nucleotide methylation (Finnegan *et al.*, 2000) impacting the binding of the final regulatory elements, the transcription factors.

During endoreduplication, chromosomes morphology is modified because the number of chromatids on a single chromosome increase from two to several hundred (Nagl, 1976). This modification of chromosome morphology might be influencing their positioning in the nucleus, the condensation of the chromatin (methylation pattern, histone composition and modification) and at the end, gene expression. The following sections will give insight into these questioning by reviewing the different mechanisms known to influence gene expression.

B) *Regulation of gene expression by chromosome positioning in the nucleus.*

Chromosomes are separated DNA molecules carrying the genetic information and having a specific organization/localization/positioning in the nucleus called chromosomes territories. Chromosome territories have been shown in animals to be involved in the regulation of gene expression (Figure 13A; Orszynowicz *et al.*, 2017). In the next paragraphs, I will describe how chromosome positioning within the nucleus can influence gene expression and when possible link this information to endoreduplication.

In non-plant systems, genes localized at the nuclear periphery tend to exhibit low transcription while those localized more internally in the nucleus often have higher expression (Van Steensel and Belmont, 2017). By using a CRISPR based tool named CRSPR-GO (referring to CRISPR-genome organization) allowing to move parts of chromosomes to specific nuclear compartments, the expression of genes adjacent to the Cas9 targeted locus was repressed when the targeted region was the nuclear periphery (Figure 13 D, Wang *et al.*, 2018). However, expression of chromatin located at the nuclear periphery is not always repressed. When Chromatin interacts with nuclear envelop proteins, transcription is in general repressed (Shaklai *et al.*, 2007) while localization to nuclear pores through interaction with nucleoporin is associated with active transcription (Liang and Hetzer, 2011)

In general, chromosomes position within nuclei is not random, but some parts of the chromosomes can move outside of the chromosome territories to intersect with other chromosomes territories (Cremer and Cremer 2010). These interphase areas are named topologically associating domains (TADs) and are visible by super-resolution microscopy and sequencing technologies (Figure 14A, Szabo *et al.*, 2018, Szabo *et al.*, 2019).

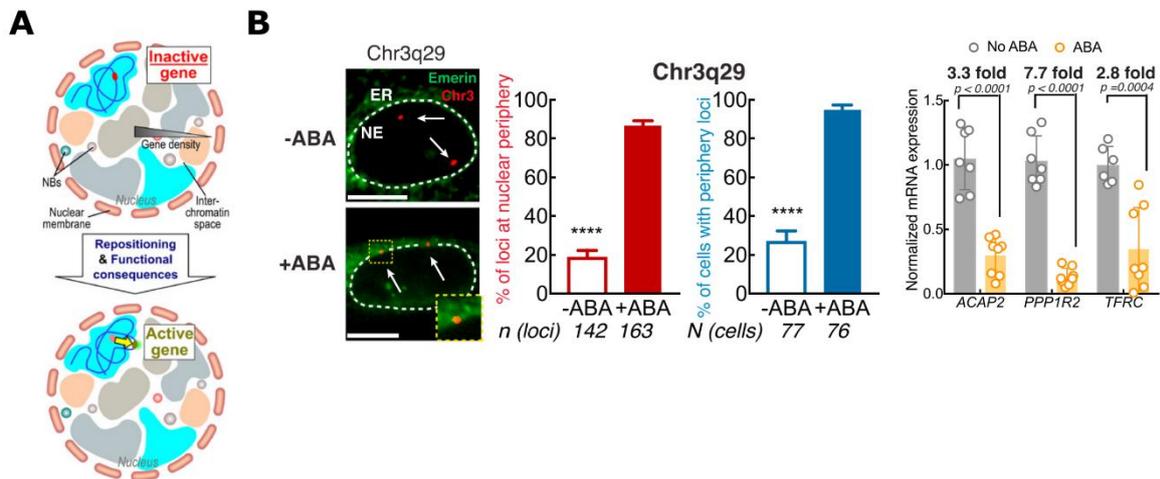


Figure 13. Chromosome position within the nucleus can define its transcriptional program.

(A) Example of human genome organization. Chromosomal function is maintained by individual chromosomes positions into chromosome territories (colored patches). Gene rich portions of chromosomes are generally positioned more internally in the nucleus. Genes are less expressed when located in the middle of the chromosome territory and are delocalized/repositioned to its periphery to be placed in interphase areas between chromosomes territories near transcriptional factories (nuclear bodies, NB) to be expressed.

(B) From left to right: fluorescence microscopy observation of a ABA inducible CRISPR based tool (named CRISPR GO) allowing to delocalize genomic loci to the nuclear periphery, here a locus name Chr3q29. Next, barplots represent the number of loci delocalized to the nuclear periphery and number of cells with delocalized loci, on the left, and gene expression of genes associated to delocalized locus Chr3q29 on the right, both after ABA induction

(A) is adapted and modified from Sawyer *et al.*, 2016. (B) is adapted and modified from Wang *et al.*, 2018.

TAD's have been shown to be related to transcriptional co-regulation of homologous or heterologous chromosome portions in Animals and *Drosophila* (Pombo and Dillon 2015, Szabo *et al.*, 2019). In plants, most studies about chromosome sub-organization within the nuclei were done on the model plant *Arabidopsis* and few studies exist in other plant species such as Maize, Wheat or Tomato. In *Arabidopsis*, chromosome territories exist in interphase nuclei but, contrary to Animals, they adopt a random positioning except for the ones carrying the Nucleolus Organizer region (NOR, Berr and Schubert 2007). Telomeres are generally found around the NOR and telomeres of sister chromatids can rarely be separated (Figure 14B, Schubert *et al.*, 2012). As in Animal, specific chromosome positioning can influence gene expression in *Arabidopsis*. For example, chromosomes regions associated with the nucleolus in *Arabidopsis*, named nucleolus associated domains (NAD) are preferentially non-expressed regions including telomeres, rDNA sequences, transposable elements and inactive coding sequences (Pontvianne *et al.*, 2016). Also, similarly to animals, the *Arabidopsis* nuclear periphery has been identified as a domain of transcriptional repression enriched in transposable elements and non-transcribed genes (Bi *et al.*, 2015). At the nuclear periphery two specific chromatin regions can be found: first, the heterochromatin that corresponds to condensed chromatin generally transcriptionally inactive and second, the chromocenters that are domains of dense heterochromatin composed of centromeres repeats and pericentromeric areas of the chromosome (Simon *et al.*, 2015). On the other hand, it has been shown that light can induce the repositioning of light inducible genes loci to the nuclear periphery and that this repositioning is associated with transcription activation (Feng *et al.*, 2014). In addition, artificial relocation of transgene to nuclear pores through the binding to nucleoporin increases its transcription (Smith *et al.*, 2015).

In *Arabidopsis*, the distribution and interaction of different chromosomal territories was analyzed in cells of different ploidy levels (Figure 14 B and C, Schubert *et al.*, 2012; Schubert *et al.*, 2014). In highly endoreduplicated cells, chromosome territories are still present and similar chromosome territories are found as in diploid cells (Figure 14C, Kato and Lam 2003, Berr and Schubert 2007). Euchromatin that corresponds to the decondensed and transcriptionally active form of the chromosome is forming loops that anchor at the chromocenters (Berr and Schubert 2007, Poulet *et al.*, 2017). In endoreduplicated nuclei, chromocenters are internalized and less heterochromatic structures are observed indicating a decondensation of the chromatin (Poulet *et al.*, 2017). This might be indicating that specific structural arrangements exist in endoreduplicated cells and could be involved in the regulation of gene expression.

Hi-C seq is a method based on sequencing technologies in which 3D chromatin conformation is frozen due to chemical crosslinking of DNA, then DNA is fractionated and sequenced and this allows-to-allows identification of TAD. In *Arabidopsis*, the analysis of chromosome conformation by sequencing has shown that TADs are not commonly observed and that chromosomes interact with each other around the heterochromatin domains (pericentromeric and telomeres, Grob *et al.*, 2014, Feng *et al.*, 2014). The absence of TAD could be related to the small genome size and low chromosome number of *Arabidopsis* (Feng *et al.*, 2014).

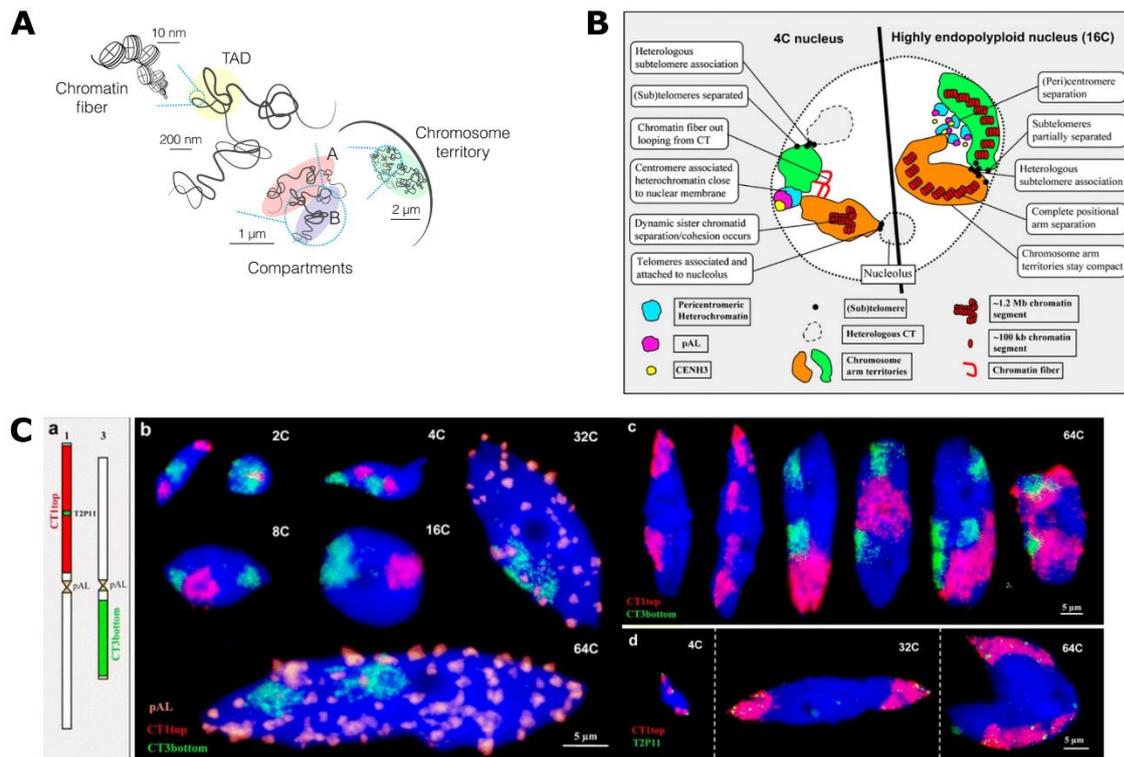


Figure 14. Organization of chromosome territories in Animals and plants

(A) Schematic representation of DNA packing in the nucleus in Animal cells.

(B) Schematic representation of chromosome organization in cells of *Arabidopsis* with different ploidy levels. A 4C cell is compared to a 16C cell showing chromatin relaxation and conservation of chromosome territories.

(C) Representation of two chromosome territories in *Arabidopsis* in nuclei of different ploidy levels (2C to 64C). In red is represented one arm of the chromosome 1 and in green one arm of the chromosome 3 and in pink are indicated the chromocenters.

(A) is adapted and modified from Szabo *et al.*, 2019. (B) and (C) are adapted and modified from Schubert *et al.*, 2012.

Arabidopsis could thus be an exception because TAD-like structures were observed in other plants with larger genomes and chromosome numbers such as Maize (n=10), Tomato (n=12), Sorghum (n=10), Foxtail millet (n=18), and Rice (n=12) (Doğan *et al.*, 2018). The borders of TAD-like structures are enriched in expressed genes, decondensed chromatin regions, and are depleted of transposable elements in Maize, Tomato, Sorghum, Foxtail millet and Rice (Dong *et al.*, 2017). Autotetraploid Arabidopsis plants have decreased intra-chromosomal interactions and higher inter-chromosomal interactions compared to diploid Arabidopsis (Zhang *et al.*, 2019) and among the 743 differentially expressed genes between these plants (436 upregulated and 307 downregulated) 72% are located in the differential chromatin interaction domains (Zhang *et al.*, 2019). However, there is no data available about chromatin interactions in relation to endoreduplicated chromosomes.

In Maize and rice endosperm cells that are endoreduplicated, heterochromatin regions are less compact compared to mesophyll cells, mostly non endoreduplicated (Rymen *et al.*, 2007, Dong *et al.*, 2020). This decreased compaction thus triggers an increased accessibility of the chromatin, reducing the local repression of some chromosomes. In addition, it has been shown that there are less repressive compartments in endoreduplicated tissues, endosperm, for example, compared to less endoreduplicated tissues such as mesophyll cells (Dong *et al.*, 2020). Moreover, TAD-like borders, enriched in groups of genes with an increased expression, are larger in endoreduplicated tissues (endosperm) (Dong *et al.*, 2020). These observations suggest that TAD-like structures could be implicated in regulating general gene expression in plants by modifying accessibility of their borders for transcription and that endoreduplicated tissues tend to have a chromatin organization favoring transcription (Dong *et al.*, 2020).

C) *Transcriptional regulation and chromatin accessibility*

In Eukaryote nuclei, DNA is not free but is compacted by proteins under the form of chromatin (Figure 14A). The conformation of the chromatin is a key element for the regulation of gene expression through DNA accessibility (Zentner and Henikoff 2013). DNA packaging in the nucleus is highly regulated and mediated by the nucleosomes composed of a protein octamer of core histones. The octamer is composed of a heterotetramer of HISTONE3 (H3) and HISTONE4 (H4) and a heterotetramer of HISTONE2A (H2A) and HISTONE2B (H2B) (Figure 15A; Klug *et al.*, 1980). Post-translational chemical modifications of the core histones include acetylation, phosphorylation and methylation and are sufficient to affect DNA compaction and thus accessibility of the transcriptional machinery. (Figure 15B, Zentner and Henikoff 2013). Several euchromatin related histone modifications have been found and are classified according to their repressing or activating role in transcription. Classical repressive marks are H3K27me3 and H2Aub1 (ubiquitination) and activating marks are H3 and H4 acetylation, H3K4me3, H3K36me3, H2Bub1 and H2AS9ph (phosphorylation, Zhao *et al.*, 2019). In Arabidopsis, for example, it has been shown that tri-methylation on lysine 27 of H3 (H3K27me3) is necessary to repress genes implicated in totipotency to maintain cell identity of guard cells forming the stomata and H3K4me3 is necessary to activate genes of aging in leaves (Lee *et al.*, 2019, Liu *et al.*, 2019).

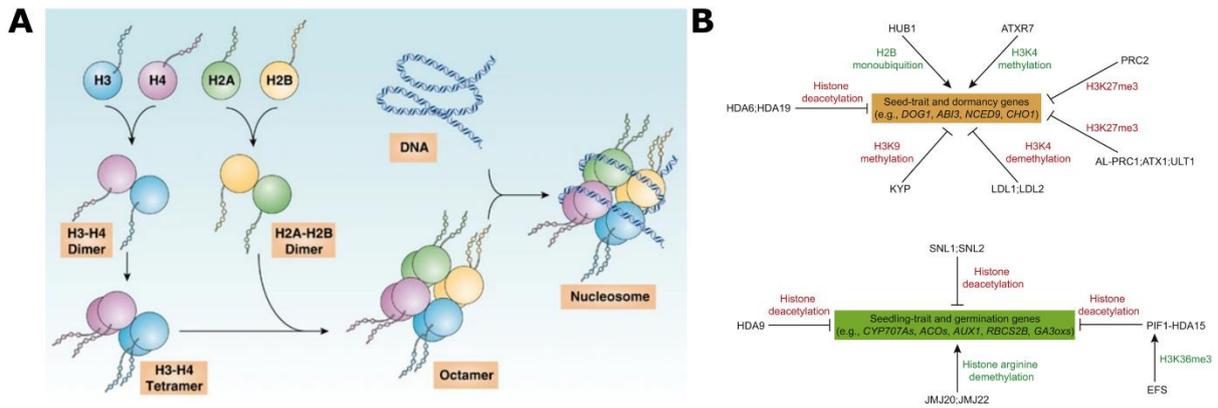


Figure 15. Chromatin structure is dependent on the nucleosome composition and histone post-translational modifications.

(A) Schematic representation of nucleosome octamer structure and formation with Histone3 (H3), Histone4 (H4) and both Histone2 A and B (H2A and H2B).

(B) Histone post-translational modifications effect on gene expression related to seed dormancy or germination. In red are indicated modifications with a repressive effect and in green active modifications.

ATXR = ARABIDOPSIS TRITHORAX-LIKE PROTEIN, HDA = HISTONE DEACETYLASE, HUB = HISTONE MONO-UBIQUITINATION, JMJ = JUMONJI DOMAIN-CONTAINING PROTEIN, KYP = KRYPTONITE, LDL = LYSINE-SPECIFIC HISTONE DEMETHYLASE, PRC = POLYCOMB REPRESSIVE CORE COMPLEX, SNL = SIN3-LIKE (histone deacetylase), ULT = ULTRAPETALA.

(A) is adapted and modified from Chen *et al.*, 2014. (B) is adapted and modified from Zhao *et al.*, 2019.

In Arabidopsis endoreduplicated cells, microscopic observations of chromatin linked fluorescent markers has shown that chromatin is occupying more space in the nuclei than in non-endoreduplicated ones, suggesting a chromatin decondensation (Kato and Lam 2003, Poulet *et al.*, 2017). Interestingly, this observed chromatin decondensation is coherent with data obtained by ATAC-seq (Assay for Transposase-Accessible Chromatin with high throughput sequencing) a method using the capacity of a transposase enzyme to insert sequencing adapter between nucleosomes where chromatin is accessible giving then the sequence of accessible chromatin regions for transcriptional machinery (Buenrostro *et al.*, 2015). Indeed, ploidy specific ATAC-seq analysis of Medicago nodules and single-cell ATAC seq (scATAC-seq) analysis of Arabidopsis roots have shown that chromatin in endoreduplicated cells is more accessible than in non endoreduplicated cells (Nagymihály *et al.*, 2017, Dorrity *et al.*, 2020).

Interestingly, in Medicago nodules, genes with a differential chromatin accessibility between different ploidy levels have been found differentially expressed in relation with these ploidy levels (Nagymihály *et al.*, 2017). For example, *s* genes found highly expressed in 16C and 32C nuclei were found in regions with higher accessibility in the ATAC-seq data (Nagymihály *et al.*, 2017). The observed increase of chromatin accessibility found by ATAC-seq in endoreduplicated cells might be indicating a reduction of nucleosome activity thus DNA packaging potentially through histone modifications (Nagymihály *et al.*, 2017). However, this correlation between dynamic accessibility and gene expression in single cells data from Arabidopsis root was poor (Dorrity *et al.*, 2020).

As mentioned previously, DNA packaging is regulated by the nucleosomes and their post-translational modifications. Interestingly, some links between endoreduplication and histone modification have been described. For example, the mutants of *CAF-1* (*CHROMATIN ASSEMBLY FACTOR-1*), encoding a protein responsible of the deposition of new H3 and H4 that are acetylated and *HUB1* (*HISTONE MONO-UBIQUITINATION 1*), encoding a protein responsible of monoubiquitylation of H2B (Desvoyes *et al.*, 2010) contain cells entering earlier in the endocycle (Ramirez-Para *et al.*, 2007, Fleury *et al.*, 2007) suggesting that histone marks could influence gene expression controlling the transition to endoreduplication. In Arabidopsis, the overexpression of *KRP5*, an inhibitor of the CDK-CYC activity, results in an increase in endoreduplication. Interestingly, *KRP5* protein is localizing to the chromocenter (Jégu *et al.*, 2013) and the overexpression phenotype is amplified in the mutant *atxr5/atxr6* (*arabidopsis trithorax-related*) mutated in genes implicated in the maintenance of repressive H3 methylation in heterochromatin (Jégu *et al.*, 2013). This finding is suggesting that *KRP5* which regulates endoreduplication could also influence the maintenance of heterochromatin through ATRX proteins regulation. Although not always linked to endoreduplication, the co-occurrence of chromatin remodeling with endoreduplication in several mutants could indicate a necessary role of histone modification and chromatin remodeling during this process.

A direct link between differential histone mark modifications and ploidy specific gene expression has been shown in only one study in Medicago nodules (Nagymihály *et al.*,

2017). ChIP-qPCR analysis (Chromatin Immuno-Precipitation) that uses the specificity of antibodies to recognize and pull down a specific protein with its associated DNA was used to specifically study modified histones H3K27me3 or H3K9ac at specific loci. This analysis has shown that *NCR* genes have a reduction in the repressive histone mark H3K27me3, an increase in the active histone mark H3K9ac in 8C, 16C and 32C nuclei compared to low ploidy level 4C and this is associated with an increase of expression of these genes (Nagy Mihály *et al.*, 2017). In Tomato, no link has been established between modification of histone marks and endoreduplication.

In conclusion, in the nodules of Medicago, specific histone modifications seem to change as a function of ploidy levels favoring ploidy specific gene expression or repression. However, more research needs to be done to study the link between ploidy specific gene expression and chromatin conformation regulation through specific histone marks.

D) *Role of transcription factors in the ploidy specific regulation of expression*

Increase of chromatin accessibility means that gene promoters become accessible for transcription factors to induce transcription. However, in an endoreduplication context, no study describes the involvement of transcription factors in the regulation of ploidy specific gene expression.

E) *DNA methylation and gene expression regulation*

Direct chemical modifications of DNA, such as the methylation of the pyrimidine base cytosine, can also influence gene expression. In coding genes, cytosine methylation occurs generally in the promoter sequences and also in coding sequence. Cytosine methylation of promoters generally has a repressing effect on gene expression (Berdasco *et al.*, 2008).

In Arabidopsis, DNA methylation occurs in all organs (Bartels *et al.*, 2018) and is implicated in multiple developmental programs such as gametogenesis, fruit development and ripening or response to stresses (Bartels *et al.*, 2018). In Tomato fruits, ripening genes promoters are actively demethylated to induce their expression at the onset of ripening (Lang *et al.*, 2017, Tang *et al.*, 2020).

Endosperms of Arabidopsis, Maize and Rice, that contain high proportion of endoreduplicated cells, have DNA globally hypomethylated and in Rice endosperm more pronounced hypomethylation is observed on specific genes related to starch accumulation (Lauria *et al.*, 2004, Hsieh *et al.*, 2009, Zemach *et al.*, 2010). In Arabidopsis, plants cultivated in short day conditions and transferred to long day show a floral transition that induces an increase in ploidy levels in the leaves (Del Prete *et al.*, 2019). This increase endoreduplication is accompanied by differential gene expression of numerous genes among which three are involved in DNA methylation modification, *METHYLTRANSFERASE 1*, *CHROMOMETHYLASE 3*, *DEMETER-LIKE 2* (MET1, CMT3 and DML2, respectively, Del Prete *et al.*, 2019). MET1 and CMT3 methylate cytosine whereas DML2 is removing methylation. The increase of expression of MET1, CMT3 and DML2 could indicate a global reprogramming of the chromosome methylation pattern in relation with endoreduplication and flowering. Lastly, in

Medicago nodules, DNA demethylation by MtDME is necessary for the differentiation of symbiotic nitrogen fixing cells (Satgé *et al.*, 2016). Moreover, genes with high expression in endoreduplicated nuclei (32C) are significantly hypomethylated (Nagy Mihály *et al.*, 2017). These data are indicating that endoreduplication is associated with a specific hypomethylation of nodule related genes thus favoring their expression.

In conclusion, we have seen several examples in which endoreduplication is associated with modifications of nuclear genome organization, chromatin compaction or DNA methylation associated with changes in gene expression. Thus, endoreduplication can be implicated in defining ploidy specific pattern of gene expression supporting differentiation of cells through the action of different regulatory mechanisms. For now, the relation between transcriptional regulation and organization of endoreduplicated chromosomes is not clearly established.

IV. Presentation of my PhD work

Tomato fruit growth is driven by two main processes, cell division and cell expansion. Cell expansion often occurs simultaneously to endoreduplication. Endoreduplication is a process resulting from the capacity of cells to modify their classical cell cycle into a partial cell cycle, or endocycle, where successive rounds of full genome replication occur in the absence of mitosis. In plants, endoreduplication is found in numerous organs and cell types and particularly in organs or tissues of agronomical high value, such as the fleshy fruit pericarp of tomato presenting high ploidy levels (Joubès and Chevalier, 2000).

As described in the introduction, endoreduplication is part of the normal development in several plants and multiple organs and is regulated through a complex genetic network organized around the CDK-CYC activity. The work of the hosting team showed that the pericarp is composed of a heterogeneous population of cells in term of ploidy levels and cell size. Endoreduplication in the pericarp is not randomly nor evenly distributed but rather distributed in a gradient toward the center of the pericarp. In addition, endoreduplication is able to induce a global shift of gene expression and specific set of genes are differentially regulated as a function of ploidy levels in mature green fruits.

Before the start of my PhD, only one developmental stage at the end of the growth of the fruit preceding ripening had been studied for ploidy distribution and ploidy-dependent gene expression. However, there was no information about the dynamic distribution of endoreduplication during early fruit development in the pericarp nor ploidy specific gene expression. My PhD work is part of a larger project aiming at understanding: i) how endoreduplication onset and progression are molecularly controlled during fruit growth, ii) why a heterogeneous population of cells co-exists within the pericarp tissue and iii) what is the functional role of endoreduplication during fruit growth.

The first objective of my PhD was thus to obtain the distribution of ploidy levels during early pericarp development, while endoreduplication is taking place, and the PhD is

integrating in a wider project with a long-term objective of defining potential rules of endoreduplication establishment and progression in the pericarp. The second objective was to identify, during early pericarp development, which genes are specifically expressed in relation to ploidy levels with the long-term goal of selecting genes that could be used as markers of ploidy levels. This was achieved by determining the ploidy distribution *in situ* and obtaining ploidy-specific transcriptional profiles during early pericarp development. First, I first determined the ploidy and cell size distribution in the pericarp of growing fruits by using the fluorescent *in situ* hybridization (FISH) method to image chromatin specific loci within the pericarp. By combining the average ploidy distribution and individual cell size during early fruit development, I have been able to study the relationships existing between ploidy levels, cell position and area and gained insight into tissue formation in relation to genome duplication. Second, to obtain a genome-wide ploidy-specific transcriptional information during fruit development, nuclei from pericarp of fruits harvested at three growth stages (6-, 9- and 12-days post anthesis), were sorted by Fluorescence Activated Cell Sorting according to their ploidy levels and used for RNA sequencing. The analysis of this datasets allowed identifying differentially expressed genes as a function of endoreduplication and developmental timing. The ploidy distribution and transcriptomic data obtained, and their integration are presented in chapter 2.

To obtain these data, I had to implement, adapt to our model or develop several methodologies. The FisH method was implemented and optimized for the quantification of ploidy levels *in situ* in the pericarp tissue. In parallel, I worked on the development of the CRISPR-imaging technique for ploidy level quantification. I also optimized the protocol for RNA-seq on sorted nuclei of different ploidy levels by the use of spike-in to normalize gene expression. The different steps of this technical optimization are presented in chapter 1.

Finally, the data obtained during my work are discussed in the discussion section.

Chapter 2: Results

I. Technological development for ploidy determination *in situ* and for ploidy dependent transcriptomic

1. Introduction

In plants, the final size of an organ is determined by cell division that sets the number of cells and cell expansion often associated to endoreduplication increasing nuclei DNA content. In the course of Tomato fruit development, cells from the fleshy part of the fruit, the pericarp, undergo endoreduplication and can attain high ploidy levels, up to 512C (Cheniclet *et al.*, 2005). Ploidy levels are often correlated with cell sizes (Melaragno *et al.*, 1993, Bourdon *et al.*, 2011, Robinson *et al.*, 2018) and endoreduplication in plants is involved in differentiation of cells in various organs but the control of the establishment and the molecular regulation of this process are still not fully known.

Ploidy levels are classically determined by flow cytometry. To do so, plant tissues are chopped in a nucleus extraction solution allowing to obtain nuclei with maintained structures and DNA is stained with a fluorescent dye such as 4',6-diamidino-2-phénylindole (DAPI) in order to estimate DNA content (Bourge *et al.*, 2018). The flow cytometry approach is precise, reliable and allows to assay many samples, but this quantification of DNA content is representing the landscape of ploidy levels of the chopped tissue and not the spatial distribution of ploidy levels in the studied tissue. Thus, cytogenetic approaches are necessary to determine this spatial distribution. For example, fluorometry which is based on the use of fluorescent dyes binding to DNA, generally similar to the ones used in flow cytometry, has been used to determine ploidy levels by quantifying the total fluorescence in the nucleus after imaging with an epifluorescence microscope (Melaragno *et al.*, 1993). This method allows the determination of ploidy levels in young tissues with small cells and nuclei, for example in *Arabidopsis* hypocotyls, but is not appropriate for large cells and nuclei such as the ones found in tomato pericarp with high ploidy levels (Bourdon *et al.*, 2011). A second method to quantify ploidy *in situ* corresponds to Fluorescence *in situ* Hybridization (FisH) and consists in the hybridization to DNA of a synthetic DNA sequence with fluorescently dyed nucleotides complementary to a target sequence from the nuclear genomic DNA. With fluorescence microscopy, dots in the nucleus representing the hybridized loci are observed and ploidy levels are determined by counting the number of these dots for an easy quantification, the synthetic DNA sequence only targets a single locus in the genome. This method was used to map ploidy levels in the pericarp of fully grown Tomato fruit pericarps (Bourdon *et al.*, 2011). This analysis showed that ploidy levels are not evenly distributed within the tissue but follow a gradient toward the center of the pericarp (Bourdon *et al.*, 2011).

The fluorometry and FisH techniques require tissue fixation, so it could be interesting to be able to quantify ploidy levels in living cells. To do so, tools based on the CRISPR/CAS9 technology have emerged for imaging chromatin in living cells. These methods called CRISPR imaging can use a deficient dCas9, mutated in the two endonuclease domains and fused with fluorescent proteins (FP) to directly visualize the targeted loci of interest (Figure 16A; Chen *et al.*, 2013). Genomic loci can also be

visualized indirectly by targeting sgRNA using virus RNA stem-loops called aptamers that are specifically recognized and bound by an aptamer binding protein fused with FP (Figure 16B, Fu *et al.*, 2016). A similar method of sgRNA recognition is done by adding repeats of 8 nucleotide long sequences at the end of the sgRNA called PUF binding sites that are specifically recognized by the Pumilio/fem-3 mRNA binding factor (PUF) proteins (Figure 16B, Cheng *et al.*, 2016). Lastly, an indirect visualization of loci can be done using a dCas9 fused with peptides specifically recognized by a small protein called single-chain variable fragments (scFV) that is an artificial fusion protein based on variable chains of antibody (Figure 16C, Tanenbaum *et al.*, 2014). In this case, scFV are fused to a FP and bind the peptides tail of the dCas9 in order to visualize the DNA. These CRISPR-imaging tools have been used multiple times in animal cells to visualize specific genomic loci (for review see Khosravi *et al.*, 2020).

In plants, CRISPR imaging using a dCas9 coupled to a FP (Figure 16A) has been used in *Nicotiana benthamiana* with a transient transformation approach to visualize and quantify telomeres and study their regulation and dynamism in leaf interphase nuclei (Dreissig *et al.*, 2017). However, plants stably expressing this system could not be obtained so far since stable transformation of *Arabidopsis* with the dCas9 fused to a FP targeting telomeres did not work, probably because the dCas9 bound to repeated sequences could interfere with the cell cycle progression in the plant (Fujimoto and Matsunaga 2017). Recently, a study showed that indirect visualization using CRISPR imaging using virus aptamers (Figure 16B) was efficient in transient transformation but did not result in any stably transformed plant when targeting the telomeres and only few plants were obtained when targeting the centromeres (Khosravi *et al.*, 2020). However, the authors reported that in the stably transformed plants targeting the centromeres, the dot like pattern observed was not as expected for interphase nuclei (Khosravi *et al.*, 2020).

Endoreduplication has been proposed to increase the metabolic activity of cells by inducing a global increase in transcription (Lee *et al.*, 2009). A transcriptomic analysis of sorted nuclei of tomato pericarp cells by Fluorescent Activated Nucleus Sorting (Bourge *et al.*, 2018) showed that endoreduplication is inducing a global shift of gene expression and that specific sets of genes, including lipid and carbohydrate metabolism genes, are differentially expressed in nuclei with different ploidy levels (Pirrello *et al.*, 2018).

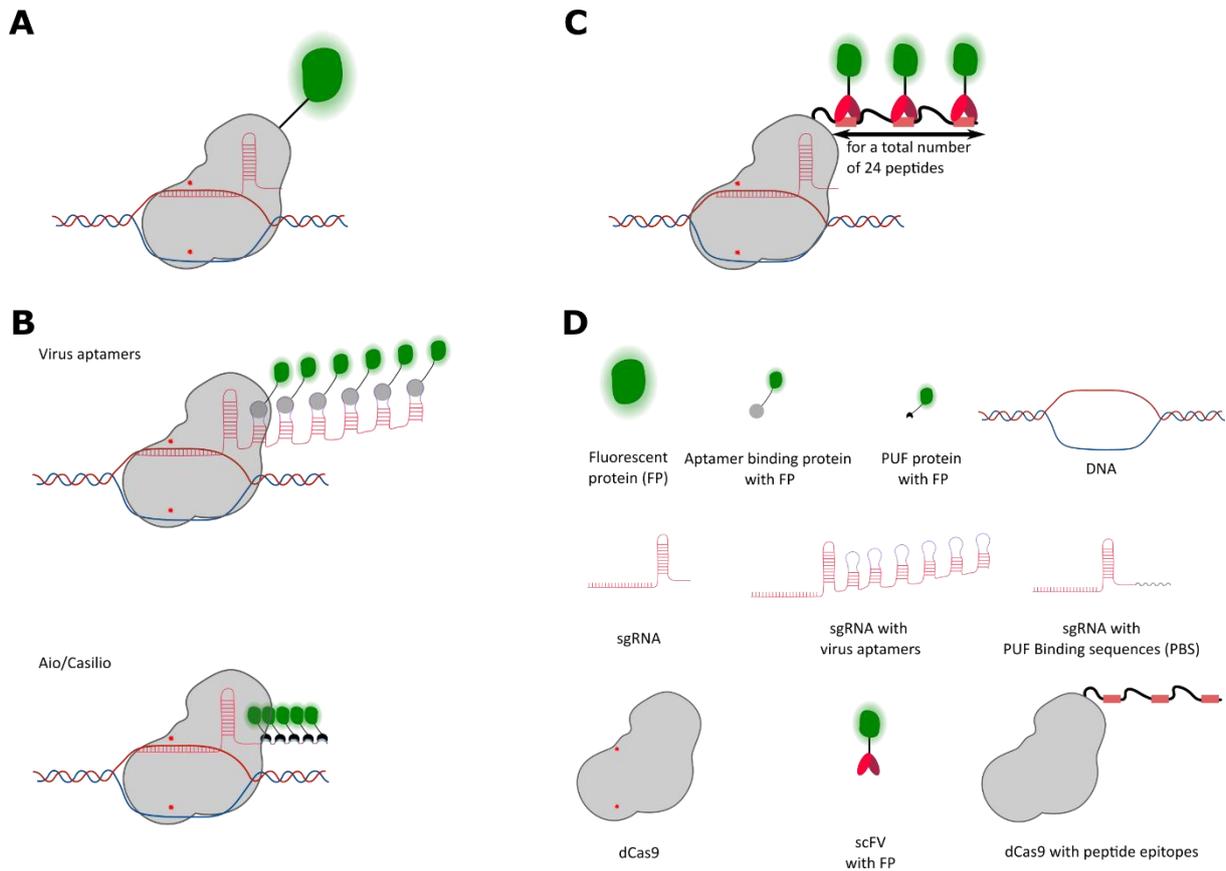


Figure 16. CRISPR/Cas9 based methods for loci visualization in living cells.

(A) Direct labelling approach using a deactivated Cas9 (dCas9). Cas9 is deactivated by two mutation in each nuclease domains (red stars) and fused to a fluorescent protein to visualize binding localization.

(B) Indirect methods using dCas9 and modified sgRNA with virus aptamers or PBS. dCas9 binds the locus of interest with help of the sgRNA and sgRNA is bound by aptamer binding proteins (purple, upper representation) or PUF protein (, lower representation).

(C) Indirect methods using modified dCas9 with peptide epitopes. A tail of peptide epitopes is added to the dCas9 that recruits simplified antibody called scFV fused to fluorescent protein that allows localization of the binding.

During my PhD, we aimed at studying how endoreduplication distribution is evolving within the pericarp during early fruit development and obtain the transcriptome of the main ploidy levels found in early developmental stages. To determine *in situ* ploidy levels of the cells in the pericarp I used *FisH* methodology and I initiated the development of a method allowing ploidy quantification in living cells based on the CRISPR-Cas9 technology. For the transcriptomic analysis, I obtained RNA sequencing data from nuclei sorted by FACS. To obtain the data presented in Chapter 2.II.2.A, I had to implement, adapt to our model or develop several methodologies. In the following paragraphs, I will present the technical and methodological optimizations that we implemented after enlightening the difficulties we tackled. These paragraphs will describe the use of a set of synthetic RNAs to properly normalize the transcriptome analyses, the improvement of the *FisH* methodology in tomato pericarp combined with a semi-automated image acquisition and analyses for ploidy quantification and finally the development of a method based on the CRISPR-Cas9 technology to quantify ploidy.

2. Results and discussion

A) Use of spike-in to normalize gene expression in a global shift in expression context.

In Tomato pericarp cells, a global shift of expression has been found between cells of different and increasing ploidy levels (Pirrello *et al.*, 2018, see *Chapter 1.III.2: Global effect on gene expression of endoreduplication*). Knowing if a shift in expression exists between different samples studied by RNA-seq is important for proper analysis and interpretation of the data. Indeed, classical RNA-seq normalization approaches (Anders and Huber, 2010; Robinson and Oshlack, 2010; Maza *et al.*, 2013) are inadequate in this case since they are generally based on the assumption that fewer than half of the genes are down-regulated or up-regulated between two conditions, which is a priori not the case with ploidy. To solve this issue, in Pirrello *et al.*, 2018, a mathematical equation that is using a factor representing the fold change induced by the global shift of expression, in the case of endoreduplication a factor2, was applied and allowed to perform classical differential expression analysis and estimate the fold change of each gene. Another solution consists in adding a known number of artificial sequences, or spike-ins, to the RNA samples to provide a scaling factor used for the normalization. To obtain the transcriptome of nuclei of different ploidy levels during my PhD, I thus implemented the use of spike-ins in our experimental setup.

We did three cultures of thirty Tomato plants each to obtain three biological repeats and randomly harvested nine fruits per plants. A mix of 20 fruits for 6DPA, and 10 fruits for 9- and 12DPA was used for the ploidy-based nuclei sorting. We collected 100k nuclei for each ploidy level from each stage. For 6DPA, we studied 2C, 4C, 8C and 16C nuclei and for 9 and 12 DPA we studied 4C, 8C, 16C and 32C nuclei. Total RNA was extracted from a fixed number of 100k nuclei for each ploidy level because RNA quantity extracted from different ploidy levels can be variable (Pirrello *et al.*, 2018). Thus, using 100k nuclei allows to compare ploidy levels. Before RNA extraction, we added to each sample a commercial solution named “*ERCC RNA Spike-In Control Mixes*” which is a mix of 92 RNAs from the ERCC plasmid reference library for which the sequences and

concentrations are known and variable allowing to assess the sensibility of the experiment. The same quantity of spike-in was added to each sample meaning that one quantity of spike-in corresponds to 100k nuclei indifferently to the ploidy level.

After sequencing and data analysis, we detected 21 of the 92 spike-in in all samples and those were among the more abundant synthetic RNAs. The lowest quantity for a spike RNA corresponded to around 70k copies in one sample indicating that theoretically we could be able to detect an RNA even if there was only one copy per nucleus since we used 100 k nuclei for the sequencing.

As in Pirrello *et al.*, 2018 we tested if a global shift in expression existed in the samples studied and if it was similar to the two-fold observed at 30DPA. Using the 21 spike-ins detected in all samples, we measured the fold change of the ratio spike-ins expression/global RNA expression between two ploidy levels at 6DPA, 9DPA and 12DPA. We found that the relative quantity of spike-ins changes between ploidy levels. For the comparisons 16C/4C, the fold changes varied between 2 and 8 at 6DPA or between 1.8 and 2.3 for 16C/8C at 9DPA (Figure 17A). If a global shift in expression between two ploidy levels correlates with the quantity of DNA, a fold change of 2 would be expected between 16C and 8C and 4 between 16C and 4C. The observed values are thus close to the expected values suggesting a shift in expression near 2-fold between consecutive ploidy levels (Figure 17A). To compare expression and visualize differences between ploidy levels, we represented the log₂ fold change between two ploidy levels and the mean of normalized counts per gene on a MA plot. In a situation with no shift of expression between conditions, the expression of most genes is expected to remain constant between the conditions studied shaping the plot symmetrically around a log fold change value of 0. In our experiment, the fold changes of the spike-ins were used to visualize, in the MA plot, the expected symmetry in case of a shift in expression. Between 16C and 4C at 6DPA, for example, we observed that the symmetry of the plot was located a bit higher than a log₂ fold change value of 2 corresponding to the expected global shift in expression of 4 (Figure 17B). These results are supporting the hypothesis that in tomato pericarp, whatever the developmental timepoint, a global shift of expression is following ploidy level increase (Pirrello *et al.*, 2018). In conclusion, the use of spike-ins in the context of a global shift of expression allowed a reliable determination of gene expression in the different ploidy level samples and stages studied in the PhD.

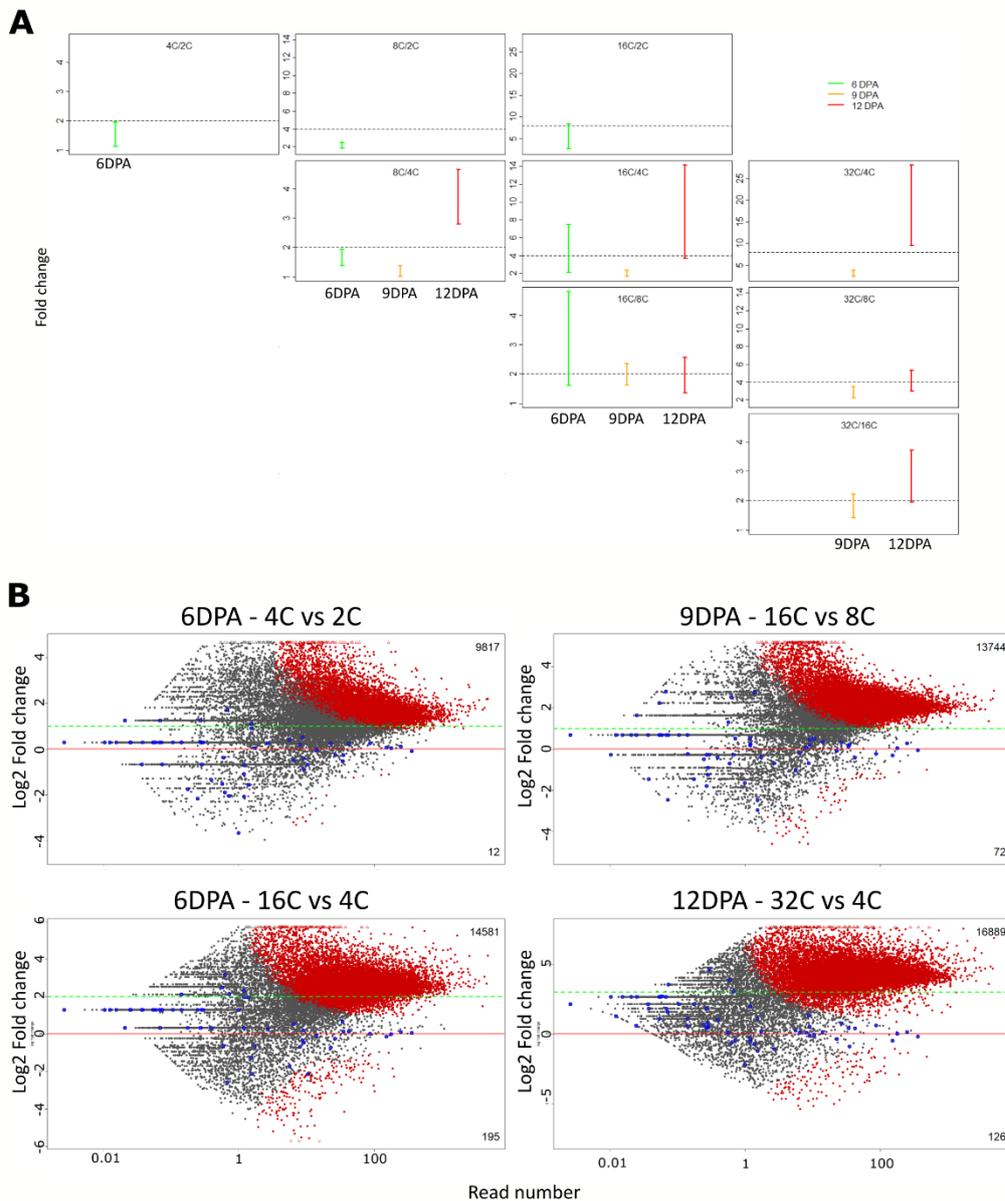


Figure 17. Use of spike-ins in the context of global shift of expression in tomato pericarp cells.

(A) Fold change of spike-ins expression compared to global RNA expression in all possible comparisons of ploidy levels at 6DPA, 9DPA and 12DPA. The dotted lines are indicating the expected fold change between ploidy levels in the case of doubling of expression between two successive ploidy levels.

(B) MA plots representing the mean number of normalized read counts in function of the log2 of fold change of reads counts between the compared ploidy levels.

The red line is indicating the absence of differential expression (Fold change=0). Most genes are expected to have a fold change value around this line if there is no shift of expression between ploidy levels. The green dotted line is indicating the expected fold change of expression if there is a shift of expression following ploidy levels.

By analyzing the data with the assumption of a global shift in expression, we found 17484, 16751 and 18596 genes expressed in at least one sample across the samples with different ploidy levels at 6, 9 and 12 DPA respectively. After removing the genes not expressed in all samples, we found 16751, 16213 and 17300 genes that were declared as being non-DE, thus having an expression that is proportional to the ploidy level at 6, 9 and 12 DPA, respectively. We found among the 18318 genes declared as non-DE across all samples, 14843 genes in common between the three stages (Figure 18A). This overlap indicates that whatever the time point the expression of a large set of genes follows the level of ploidy. This list of genes was enriched in several GO terms related to general cell function such as “*RNA processing*”, “*cell cycle*” and “*primary metabolic process*” (Figure 18B). This is in favor of hypothesize role of endoreduplication to boost general function of the cell and support growth.

In conclusion, the ploidy transcriptome with the use of spike-ins allowed to quantify that global gene expression is doubling with endoreduplication at early developmental stages and this is equivalent to the global shift observed in the later mature green stage (Pirrello *et al.*, 2018) suggesting that in the pericarp the doubling of gene expression is due to endoreduplication and not influenced by developmental stage. In our case spike-ins have helped to determine that gene expression is increasing with ploidy levels but did not allowed to determine an absolute value of increase, doubling is a likely value and to determine the exact value more sampling will be necessary to increase statistical force.

B) *Optimizing the quantification of ploidy in situ from sample preparation to imaging using Fluorescence in situ Hybridization*

a. **Optimization of the *FisH* protocol**

Previously, by using Fluorescence *in situ* Hybridization (*FisH*), the ploidy distribution in Tomato pericarp of mature green fruits, fully-grown fruits but not yet ripe, has been determined (Bourdon *et al.*, 2011). This *FisH* method is based on the use of a fluorescently dyed DNA sequence complementary to a targeted nuclear DNA sequence. In the study from Bourdon *et al.* (2011), two bacterial artificial chromosome (BAC) clones were used to target chromosomes 7 and 12 and their hybridization resulted in a single signal per targeted locus. The numbers of signals or dots were then counted and corresponded to the ploidy of the cell analyzed. However, the use of the BAC *FisH* approach can be challenging since the quality of the labelled probe can be variable. In addition, the protocol for BAC *FisH* even if efficient, required that the samples were processed in Eppendorf tubes and imposed a slow pipetting of fluids in the tube at each step. This pipetting was critical to not disrupt the pericarp sections of 150 μm thick, containing very large and turgid mesocarp cells making the tissue very fragile, and thus limited the number of samples and increased experiment duration.

For the PhD work, we thus decided to optimize the *FisH* method to decrease the duration of the experiment and increase the number of samples at once. For the quantification of ploidy *in situ*, we thus changed the method from BAC *FisH* to Oligo-*FisH* that uses oligonucleotides as probes instead of BAC because oligonucleotides are smaller and easily diffuse through the cell walls and membranes.

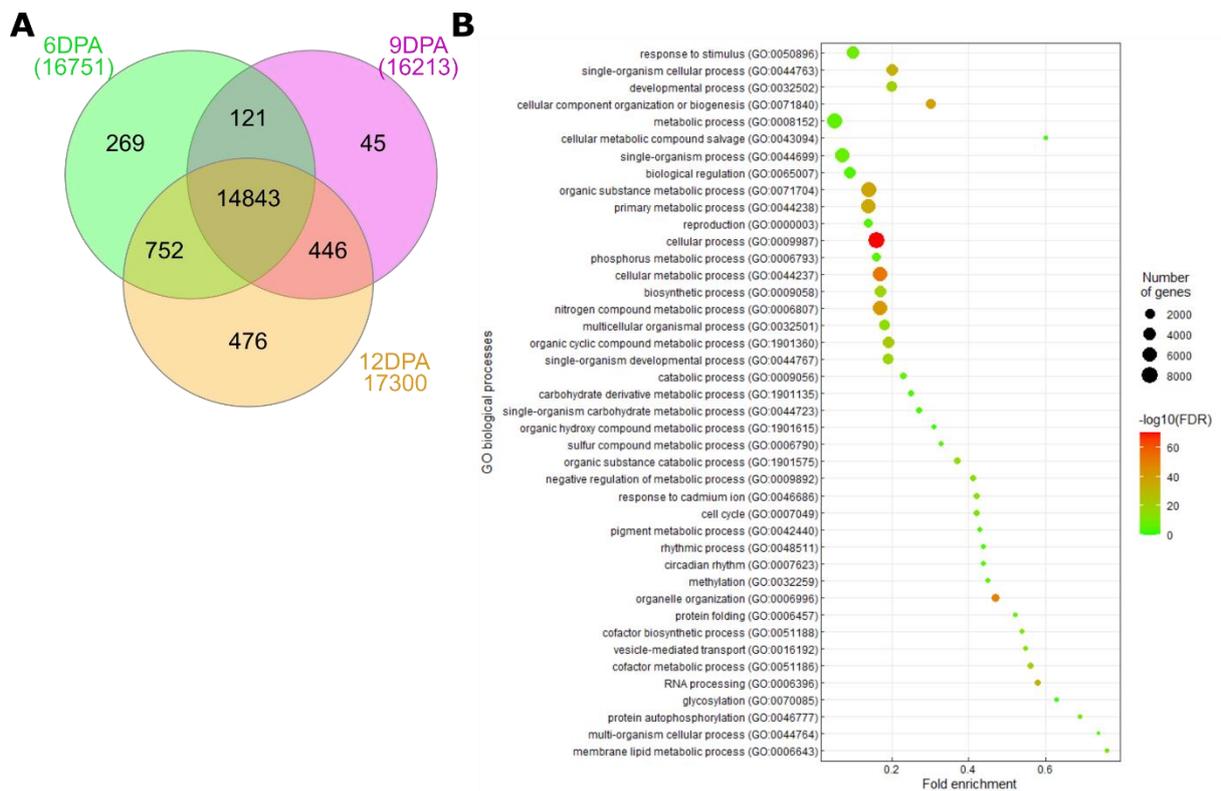


Figure 18. Venn diagram of non-differentially expressed genes found in 6-, 9- and 12DPA ploidy transcriptomes and their associated biological processes.

(A) Overlap of 14843 non-DE genes found in 6-, 9- and 12DPA ploidy transcriptome.

(B) Overrepresented biological processes in the non-DE genes found in common between stages. Dot plot showing the most representative and significant biological processes among non-DE genes. Dot size represents the number of genes in the concerned biological process and color scale is representing the fold enrichment of gene number compared to the expected number in the genome.

We also targeted another region of the genome because each oligo probe bears only one fluorophore, so Oligo-*FisH* needs to be used with a repeated sequence. To do so, we chose the 5S rDNA locus that is only located on the short arm of chromosome 1 (Figure 19) in Tomato, thus a 2C cell will have two dots so dot number will double with each endocycle.

In addition, samples were handle in 24 well plates allowing preparation and analysis of more samples at once and reducing experiment duration. Pericarp sections were placed just after slicing in baskets and change of bath was performed by transferring baskets between wells thus limiting manipulation of the pericarp sections and also reducing experiment duration. At last, the final step corresponding to the hybridization step after the addition of the fluorescent probe was performed in 0.2 μ L PCR tubes and in a thermal cycler allowing a precise and programmed management of time and temperature. After the hybridization steps (see materials and methods), slides bearing the samples were prepared and were used for imaging with a confocal.

After this optimization, the protocol duration was two half-days compared to several days for the BAC-*FisH*. In addition, we increased the number of samples that could be processed at once.

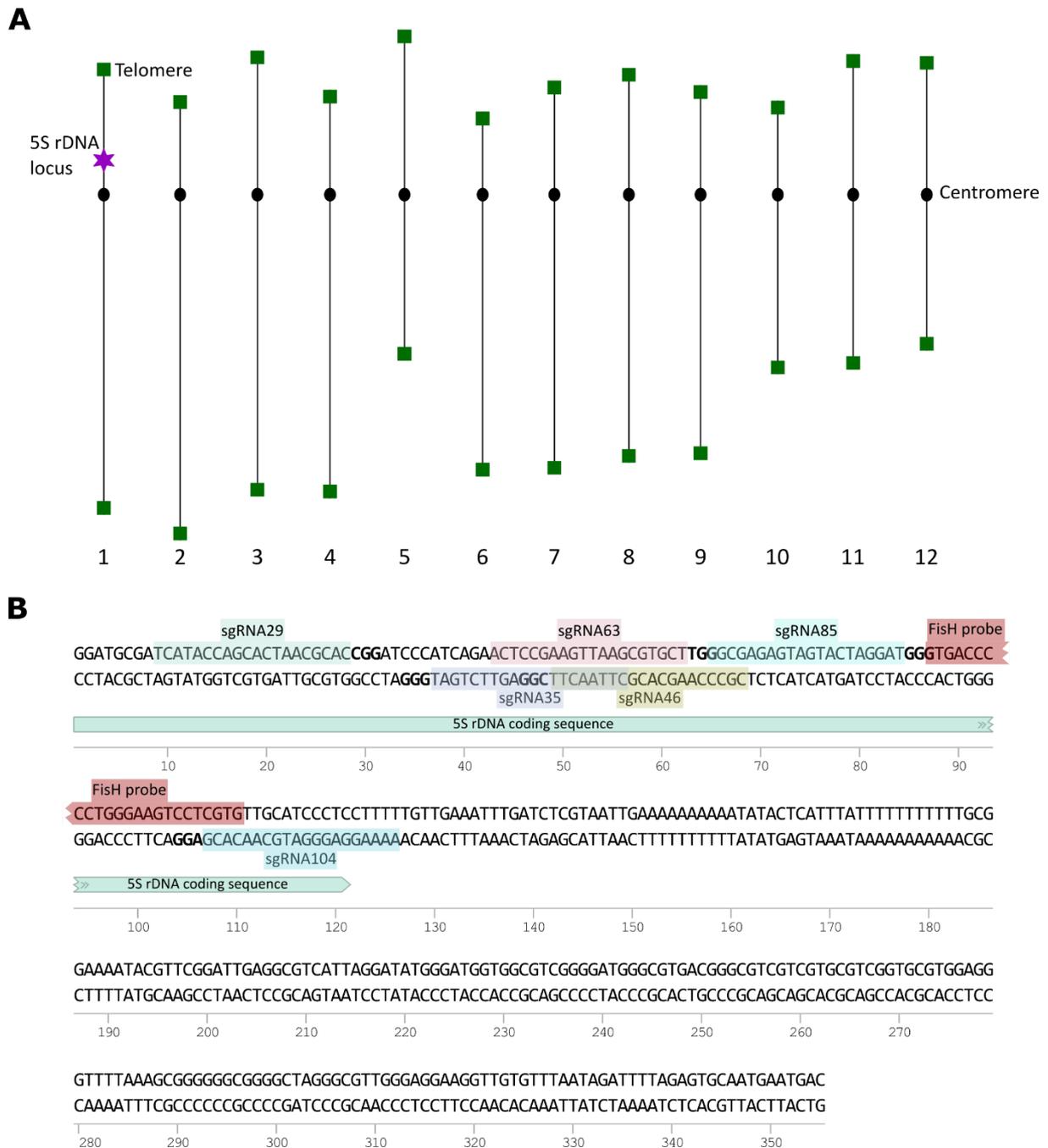


Figure 19. Repeated sequences used as targets for FisH and CRISPR imaging to quantify ploidy.

(A) Location of the targeted sequences used for FisH and CRISPR imaging on the chromosomes of Tomato. Three repeated sequences of interest for our study are represented, as green boxes for telomeres, a pink star for 5S rDNA locus and black circle for centromeres.

(B) Sequence of a repeat unit of the 5S rDNA locus in Tomato. In red is indicated the sequence targeted by the FisH probe. In pale transparent colors are indicated the different sgRNA that have been designed and tested in transient and/or stable transformation.

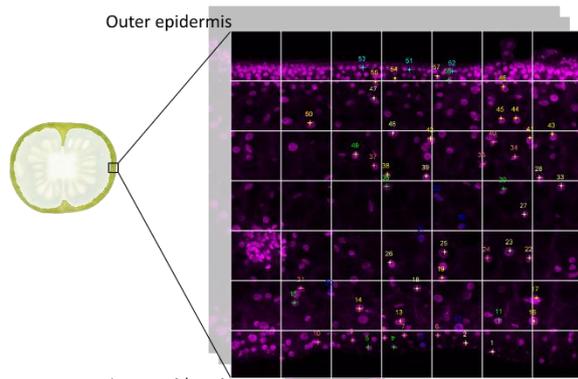
b. Semi-automated imaging

For reliable and complete ploidy quantification by *FisH* spot quantification, several nuclei per cell layer need to be imaged and all layers of the pericarp have to be assayed (Figure 2). On a pericarp sections, the objective was to determine cell size, cell position, nucleus volume and number of *FisH* dots. To do so, cell size has been estimated by dyeing the cell wall with Calcofluor-White, nucleus volume is estimated using a DAPI staining of the DNA. If we would like to obtain all the information -cell size, nuclei size and spot number- from a single acquisition it would be necessary to image large pericarp areas with the best resolution possible since this is needed for reliable *FisH* spot detection. A pericarp section represents about $\sim 0.6\text{mm}^2$ ($\sim 600000\mu\text{m}^2$, $1160\mu\text{m} \times 520\mu\text{m}$) and around 50-60 nuclei are visible in such section, without considering the small and numerous epidermal cells (Figure 20). Cells in the pericarp are large and all nuclei cannot be acquired on the same image since they are distributed in a 35-40 μm thick region over the total 150 μm thickness of the section and the *FisH* spots within the nuclei are also distributed in a 3D manner within the nuclei. This means that several Z acquisition of $\sim 0.6\text{mm}^2$ would be necessary with the minimal distance between two Z steps (140 nm) to not miss any spot. We estimated that 100 to 200 Z-acquisitions would be needed and the total acquisition time for such a pericarp sample with the best resolution would be around 48h. To reduce this acquisition time while keeping spatial information and the best resolution with x63 objective for visualizing the *FisH* dots, we designed a semi-automated pipeline to scan only the nuclei to be studied (Figure 20). First, we started scanning the region of the pericarp to be studied (about 0.6mm^2) by tile imaging with a medium resolution with the same x63 objective (Step 1 on Figure 20) allowing the identification of the nuclei and of the cell walls (Step 2 on Figure 20). For the 50 to 60 nuclei detected in the scanned regions, we defined groups of nuclei based on their relative Z positions and sizes (Step 3, Figure 20). The confocal was then programmed to use these defined groups of nuclei to image these multi-position series of nuclei in batches (Step 4 on Figure 20) with high resolution and automatically save acquisitions (Step 5 on Figure 20). All these steps allowed obtaining acquisition of DAPI and *FisH* signals for all nuclei from one pericarp section with their position at best resolution in between 3-4 hours to 18 hours for 12DPA fruits that contain bigger nuclei.

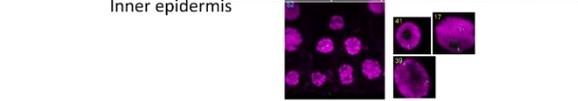
c. Ploidy quantification pipeline

For ploidy quantification, the scans obtained with the confocal allowed obtaining a 3D representation of each nucleus and of the *FisH* dots. These dots were quantified manually to obtain ploidy levels. The next step was to map the nuclei on the pericarp sections and locate them into the different cell layers. To do so, we used the confocal stage coordinate of the center of the grid acquisition that is recorded in the metadata of the image files and these coordinates were converted to pixel coordinates that can be used to retrieve positions of nuclei obtained in the same way in each file. This operation was done automatically as well as the measurement of nucleus size (Steps 6 and 7 on Figure 20). Finally, the last step was to assign each nucleus to a cell layer for obtaining the spatial distribution of ploidy levels in the pericarp.

Microscopy steps

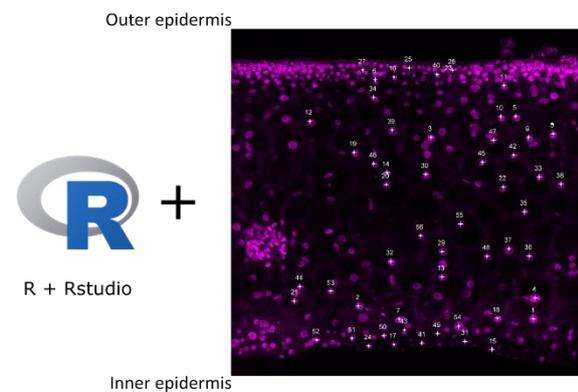


- Step 1 :** Tile imaging of pericarp area
- Step 2 :** Identification of nuclei to be imaged (cross and numbers, order have no)
- Step 3 :** Creation of groups of nuclei to do multi-position acquisition (colors)
- Step 4 :** Image acquisition parameters defined per group (multiple nuclei per image for epidermal layers, single nuclei for other)

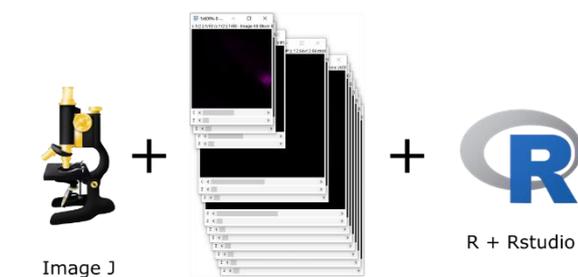


- Step 5 :** Overnight acquisition and autosave

Analyses steps



- Step 6 :** Remapping of images coordinates (order corresponds to acquisition order)



- Step 7 :** Image analyses and ploidy quantification for each position
- Step 8 :** Nuclei assignment to cell layer and generation of Ploidy maps

Figure20. Flowchart representation of the semi-automated procedure of ploidy quantification in the tomato pericarp cells.

Because identifying cell layers on thick pericarp sections was difficult, we determined the relative distances of each nucleus from the outermost cell wall of the exocarp and the endocarp to calculate a normalized distance value in the pericarp. Then, the normalized distance was compared to control normalized distances determined on previously identified cell layers from images of thin sections of the pericarp (Step 8 on Figure 20 and see Figure 20bis for example of comparison to standard images). In optimal conditions, images could be acquired in 8-18H (Steps 1 to 5, Figure 20), allowing to process 8 pericarp samples per week, and nuclei mapping and *FisH* quantification took 3-4H per pericarp (Steps 6 to 8, Figure 20).

FisH dots counting, done manually, was thus the most time-consuming step for ploidy quantification. In collaboration with engineers from the Bordeaux imaging center, a plugin for ImageJ was developed to measure nuclear volume and determine the number of *FisH* dots automatically in order to reduce analysis duration allowing and perform batch analyses.

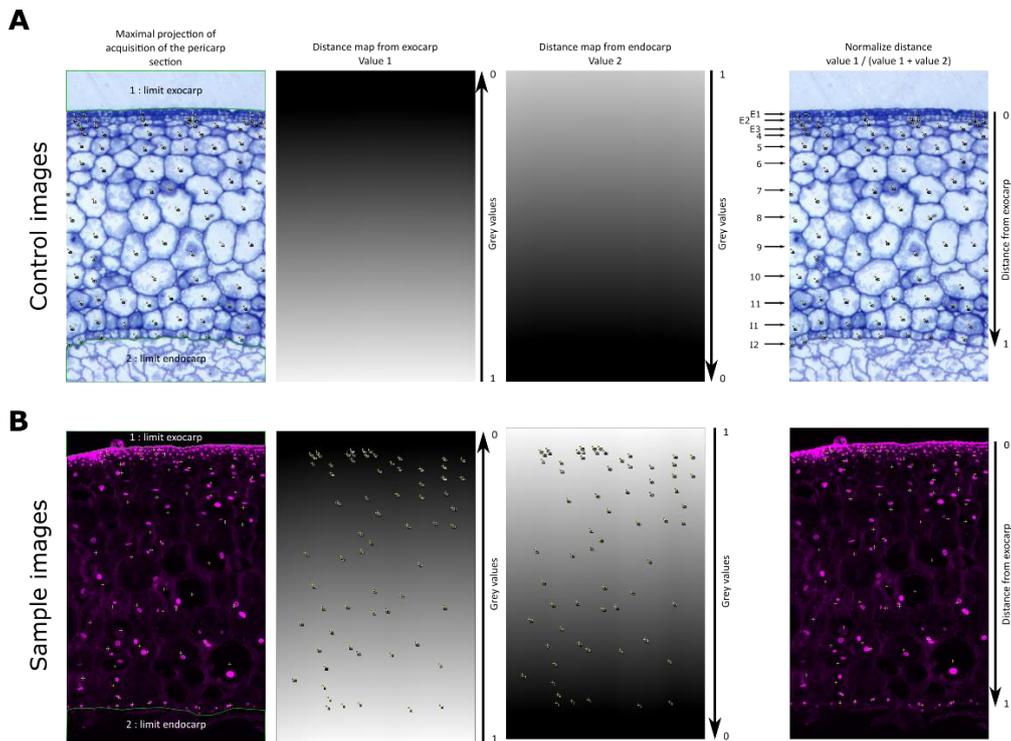


Figure 20bis. Illustration example of automatic nuclei assignment to cell layer for ploidy mapping

(A) Example attribution of range of distance to cell layers for control image of semi thin section of pericarp (Renaudin *et al.*, 2017) where cell layers are easily identified.

(B) Example of distance calculation for sample image of FisH experiments in pericarp. Crosses and numbers indicate nuclei.

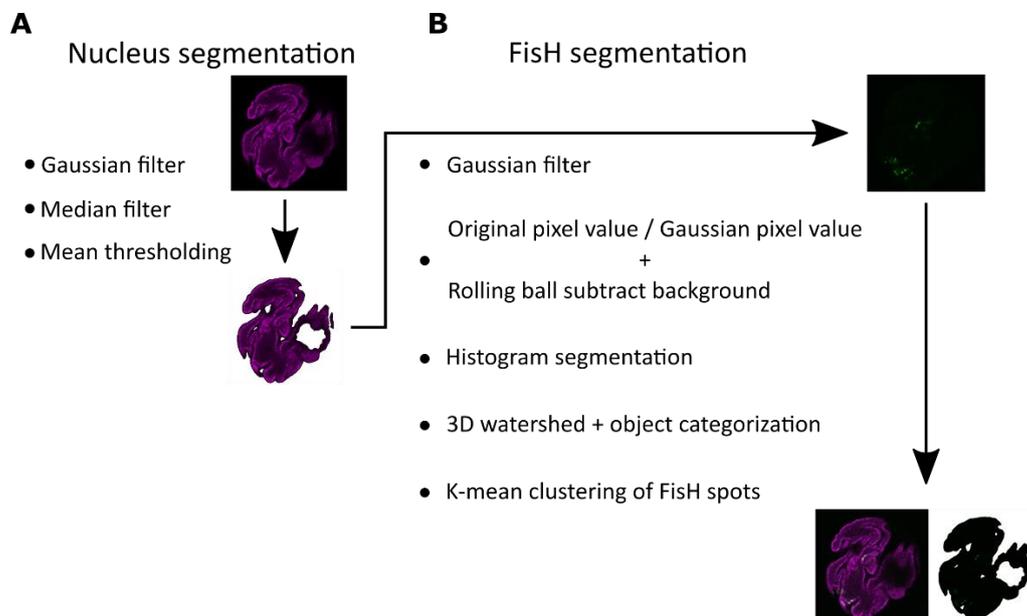


Figure 21. Flowchart representation of the steps performed by the plugin to identify and count FisH spots.

(A) Steps and tools used for nucleus segmentation that allow measuring nuclear volume and setting the boundaries where FisH spots can be found

(B) Steps and tools used for FisH spots segmentation. The tools are applied only in the area identified by nucleus segmentation to limit segmentation of artefacts outside the nucleus.

The first step of the plugin corresponds to the design of an area defined by the nucleus segmentation step (Figure 21A). This segmentation was used to limit the region for the segmentation of the *FisH* spots only inside the nucleus. Then, the same types of filters used for nuclei, gaussian and median filters, were applied for spots segmentation. Filtering was combined with a rolling ball algorithm to homogenize background in order to facilitate the segmentation of the fish spots (Figure 21B). After, *FisH* spots were identified in two steps. First all pixel values on the original image were divided by the filtered pixel values combined with a rolling ball algorithm from Image J (Figure 21B) to reduce the uneven isolated high pixel values below a determined threshold inferior to the *FisH* spots size. Rolling ball algorithm can be image as a 3D surface with the Z value corresponding to grey value of pixels of the image, so a high grey value area corresponds to a peak with small radius. Then, a ball with a determined radius is rolled beneath the surface so that it touches the surface in one or more points and it evenly adjust grey value of background as it rolls over the surface. This first method resulted in a sharper separation of the pixels between *FisH* spots and the background (Figure 22A). The second method consisted in the use of pixel values histogram (Figure 22B) and an iterative method calculating the edges of the different classes of the histogram that can then be threshold to isolate each spot as different objects (Figure 22B).

Nucleus segmentation was easily performed by using different filtering steps corresponding to both gaussian and median filters, followed by a thresholding and allowed measurement of nuclei volume. However, during the development of this plugin (Figure 21A), the main difficulty to overcome was the correct detection of the spots and their segmentation. Indeed, the Fish spots can be very close to each other and it is sometime very difficult to clearly identify one or two dots even using human eye (Figure 23).

We observe that *FisH* spots are distributed within the nucleus in two groups corresponding to the two homologous chromosomes and it was often easier to count the spots in one of the two groups because the spots were better separated. Thus, we tried to use this characteristic to correct the counting after processing by having spot numbers for each group separately. For example, if 13 spots were counted in the group A and only 3 spots in the group B it is highly probable that the 3 spots in the second group results from incorrect segmentation of spots being too close.

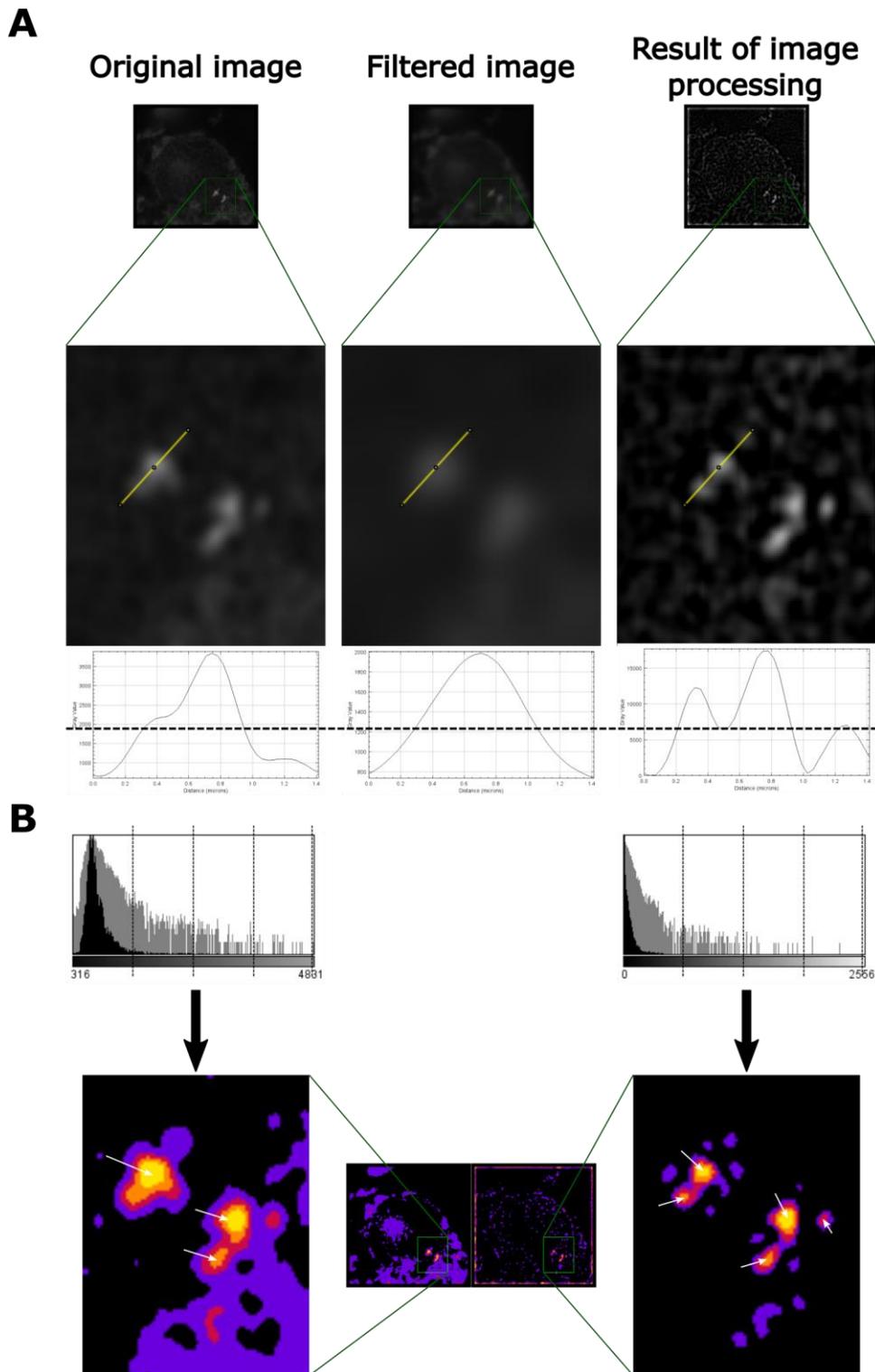


Figure 22. Example of application of the Histogram segmentation tool before and after the filtering steps.

(A) Effect of the filtering steps on an image but also on the grey values that are covered by the yellow line on the image.

(B) Histograms of the area represented in (A) with resulting segmentation of histograms in five classes. Most of the pixels are on the left of the histogram because these have low grey values while pixels of interests are localized on the right of the histogram.

Thus, after identification of the spots, they were distributed into two groups using an iterative k-mean clustering approach in order to determine two clusters of dots that could be representing the two chromosomes. The actual version of the plugin is not efficient to count ploidy levels using a 5SrDNA probe.

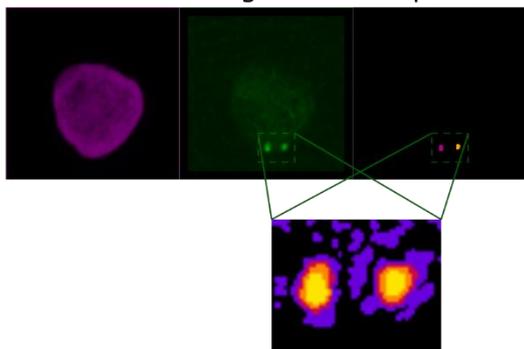
With this pipeline, the segmentation of the *FisH* spots was not efficient for all images probably because of the variable quality of the observed spot signal between samples. Indeed, often, spots were so close that even for human eyes it was complicated to determine if there were multiple spots in a single area. This limit of correct separation of *FisH* spots could be due to the nature of the 5SrDNA composed of repeated sequences that could vary in chromatin condensation. In addition, the condensation of the locus could vary in function of the status of the cell, and thus the visualization as separated dots or as closer “merging” spots could differ if the DNA is less condensed or if it is compacted.

During the PhD, we also tested another probe targeting the telomeres, but the manual quantification was too complicated because of the number of spots to count and the number of samples, for example for a 32C nuclei a number of 768 spots are expected. The spots observed with the telomere probe were clearly separated and distributed all around the nucleus, thus giving the possibility to use the developed plugin to estimate ploidy levels. This was not tried during my PhD but would be interesting to apply. Another potential solution to obtain clear dots could be to target a non-repeated sequence by using a span oligo approach that allows the production of probes for non-repeated target sequences distributed on a smaller unique chromosomal region. Targeting this type of sequencing could be eventually less sensible of the “merging” spots issue.

In conclusion, ploidy quantification by oligo-*FisH* has been optimized for tomato pericarp sections by the development of a pipeline with semi-automated image acquisition, quantification and analyses. To visualize and quantify *in situ* ploidy in living cells, we initiated the development of a method based on CRISPR/Cas9 technology.

Segmentation of FISH dots

Manual counting = 2 Computed = 2



Difficult segmentation of FISH dots

Manual counting = 17 Computed = 7

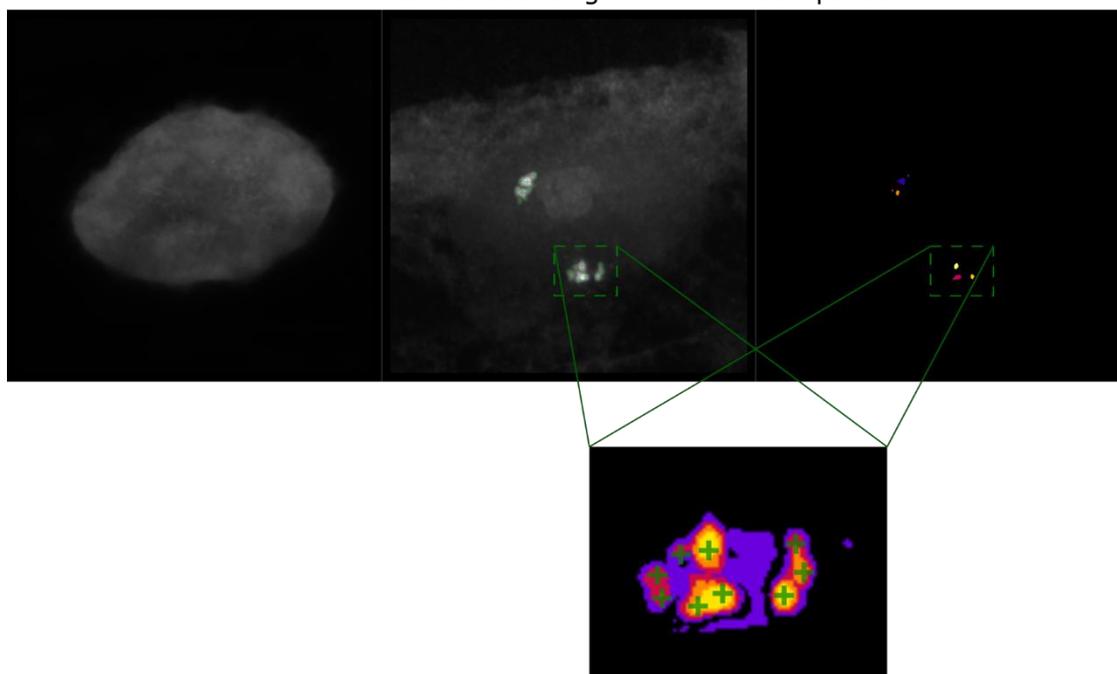


Figure 23. Example of plugin assisted FISH spot counting and issues.

Examples of an easy segmentation of a 2C nucleus with well separated spots and of a “difficult” sample with spots close to each other limiting correct identification and quantification of the spots by the plugin.

C) Visualization of genomic loci in living plant cells using CRISPR/Cas9 technology

To quantify ploidy *in situ* in Tomato, we developed a direct labelling approach based on the CRISPR/Cas9 technology by fusing the dCas9 to a FP and targeting two different genomic loci (Figure 16A). We chose two repeated loci, the telomeres and the 5S rDNA locus (Figure 19A, Chang *et al.*, 2008), in order to increase the dCas9 concentration in a relatively small area thus increasing fluorescent signal. For telomeres, a unique sgRNA has been used targeting the repeated sequence TTTAGGG (sgRNA^{telo}), similarly to Dreissig *et al.*, 2017. By using this probe, the expected number of dots is at least 24 for a 2C cell because each chromosome has two telomeres loci, and the number of dots should double if telomeres are duplicated during the endocycle. We designed six sgRNAs for the 5SrDNA locus, targeting the sequence at different positions and both DNA strands (Figure 19B, sgRNA^{5S}_{29/35/46/63/85/104}) and we used the sgRNAs alone or in combination to see which sgRNA are the more efficient and if signal is increased with simultaneous use of multiple sgRNA. Each combination of dCas9 with sgRNA targeting telomeres or 5S rDNA locus were tested in transient transformation before performing stable transformation (Figure 24A), but no signal difference depending on the sgRNA used were observed in transient transformation. Fluorescence observed in transient transformation was localized in the nuclei with a strong signal in nucleoli and the dot observed was not following the same pattern if telomeres or 5SrDNA was targeted, lot of dots or few dots, respectively (Figure 24A). The signal obtained while targeting telomere was equivalent to the one reported in literature (Dreissig *et al.*, 2017).

No stably transformed plant could be obtained with the dCas9sgRNA^{telo} and in general, only few transgenics were obtained. We obtained T1 transgenic plants for sgRNA^{5S}₃₅ (one plant), sgRNA^{5S}₄₆ (two plants), sgRNA^{5S}₈₅ (three plants) and the combination of four sgRNA^{5S}_{35/46/85/104} (one plant). Among these T1 plants, only two individual sgRNA^{5S} constructs, the sgRNA^{5S}₄₆ (2 plants) and sgRNA^{5S}₈₅ (1 plant), successfully produced fruits with seeds and showed fluorescence that could be observed at any stages of development in leaves, but they did not have dot like patterns (Figure 24B). T2 and T3 lines were the obtained, but the fluorescence in these plants was difficult to image because the signal was strong in the nucleolus and weak, near to the background, in the rest of the nucleus. Despite this weak signal, we could observe dots various types of cells: two dots in the diploid guard cells (upper panel for sgRNA^{5S}₈₅ on Figure 24B) and more than two dots in the sepal or leaf epidermal cells that undergo endoreduplication (lower panel for sgRNA^{5S}₈₅ and upper right panel for sgRNA^{5S}₄₆ Figure 24). Finally, we tried to visualize these dots in the fruit pericarp but the autofluorescence was too strong limiting the visualization of signal in nuclei.

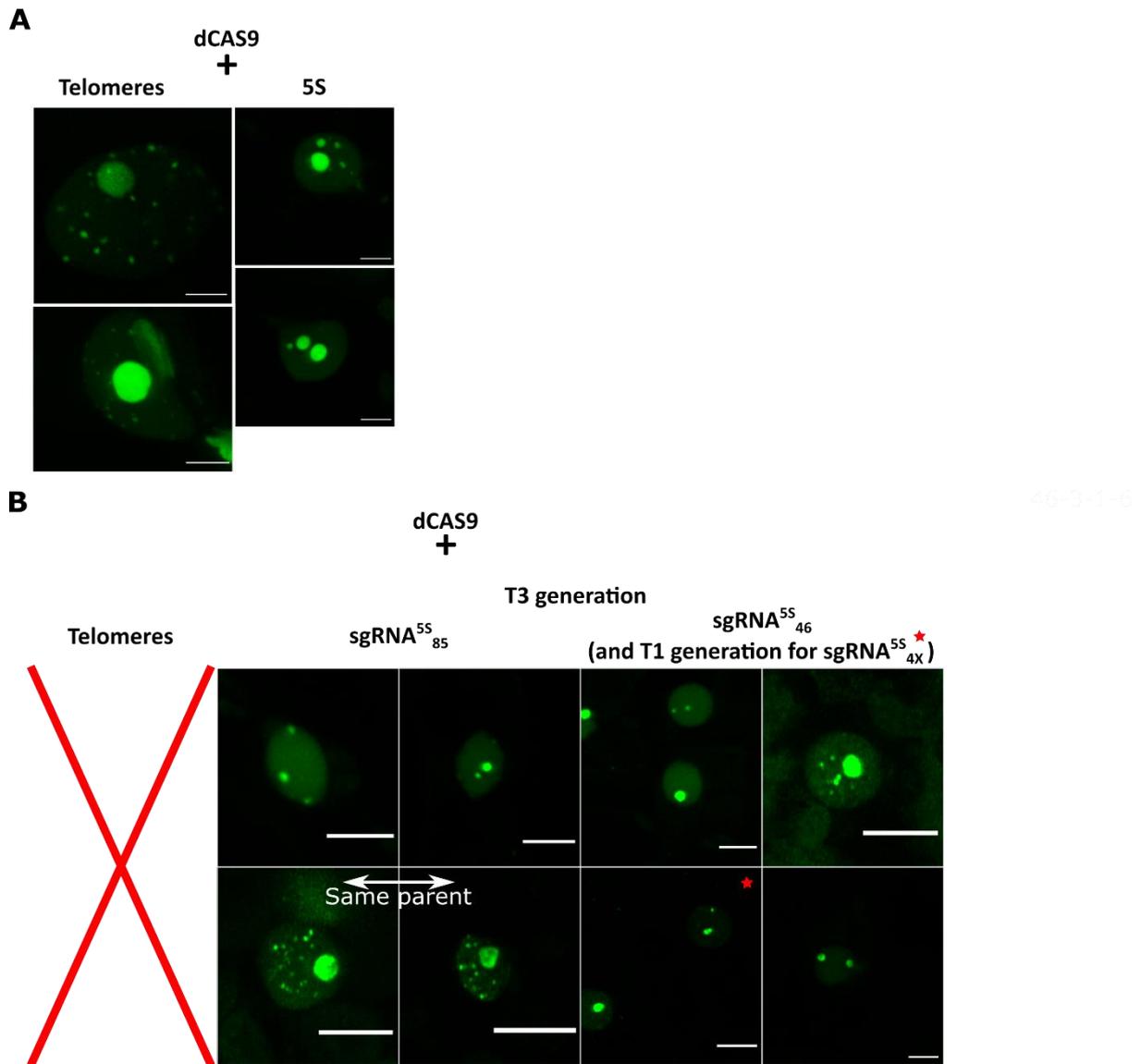


Figure 24. Illustration of CRISPR-imaging observations after transient and stable transformation.

(A) Fluorescence in agro-infiltrated three weeks old *Nicotiana benthamiana* plants. Plants have been transformed with a construct bearing a dCAS9-GFP and a sgRNA targeting either telomeres or 5S rDNA.

(B) Fluorescence in stably transformed Tomato plants with a dCAS9-GFP and one or four sgRNA targeting the 5S rDNA locus. The red cross is representing no transformant has been obtained for telomeres constructs while for the 5S rDNA three individual T3 lines were obtained with the sgRNA^{5S}₈₅ and sgRNA^{5S}₄₆. Red star: line transformed with a construct having four sgRNA targeting the 5S rDNA locus.

Finally, the CRISPR-imaging technique seems to be a promising method for *in situ* visualization of chromatin loci in living cells, however it seems that the use of the tool is not as straight forward as the gene editing tool. The direct labelling approach or targeting biologically important repeated sequences is working well in transient transformation as reported in literature (Dreissig *et al.*, 2017, Khosravi *et al.*, 2020) but targeting telomeres and 5S rDNA with a dCas9 limits plant fitness as reported in the literature (Fujimoto and Matsunaga 2017, Khosravi *et al.*, 2020). To further develop the tool, it could be interesting to explore indirect labelling approaches that is based on the labelling of another protein that recognize a peptide tail bind to the Cas9 or a fluorescent protein recognized a specific nuclei sequence of the sgRNA (see figure 16B and 16C for more detailed examples). The indirect labelling approaches seem to increase signal to noise ratio (Khosravi *et al.*, 2020) and more importantly it is necessary to be able to target non repeated sequences not implicated in genome organization such as the rDNA loci or telomeres (Simon *et al.*, 2018, Khosravi *et al.*, 2020). To target non repeated sequences assembling polycistronic sgRNA that is processed into individual sgRNAs by the tRNA-processing system could be used. By doing so, a non-repeated sequence can be targeted by using a large number of sgRNA (Xie *et al.*, 2015) to span the sequence and visualize it with direct (Figure 16A) or indirect labelling approaches (Figure 16B and 16C).

II. Dynamics of ploidy distribution and ploidy-dependent transcriptome during pericarp fruit development

1. Introduction

In all multicellular organisms, the cell cycle occurs during development, results in cell division and is responsible for transmitting the whole copy of the genome, thus the genetic information, from one mother cell to two daughter cells. In many organisms, somatic cells undergo changes in ploidy levels resulting from endopolyploidy (Orr-Weaver, 2015). In plants, endopolyploidy is mostly the result of endoreduplication (D'Amato, 1984) that occurs in several vegetative and reproductive tissues in numerous Angiosperms (Nagl 1976; D'Amato 1984).

Endoreduplication causes the cells to undergo full genome duplication without cell division (Edgar *et al.*, 2014) and without change in chromosome number and thus triggers an increase of ploidy levels solely by increasing the number of chromatids forming polytene chromosomes (Edgar and Orr-Weaver 2001). These polytene chromosomes have been observed in various plant species (Carvalho, 2000) and for example in the pericarp of Tomato fruits (Bourdon *et al.*, 2011).

Endoreduplication is associated with several developmental processes and contributes to cell differentiation and morphogenesis (Hulskamp *et al.*, 1999, Bransiepe *et al.*, 2010, Roeder *et al.*, 2010). For example, trichomes are large epidermal hair cells with a three- branches structure that require multiple rounds of endoreduplication to be functional (Hulskamp *et al.*, 1999). In addition, a positive correlation has been shown between endoreduplication and cell size such as in tomato fruits, cells that display

highly endoreduplicated nuclei, from 2C/4C to 256C (Cheniclet *et al.*, 2005). Although these observations indicate a role of endoreduplication in cell growth, the strength of the relation between endoreduplication and the increase of cell expansion is debated (Tsukaya *et al.*, 2019). Endoreduplication has also been suggested to increase the metabolic activity of the cells (Inze and De Veylder, 2006) and to be involved in the response to environmental stresses allowing building a strong metabolic capacity to improve tolerance (Scholes and Paige 2015, Bhosale *et al.*, 2019 and Lang and Schnittger 2020). Another possible role of endoreduplication is to increase gene expression since gene copy number is increased. In fully grown tomato fruits, endoreduplication leads to a global shift in expression but genes with ploidy-specific expression have also been identified revealing that cells with different ploidy levels might have different function within the pericarp (Pirrello *et al.*, 2018). In Arabidopsis, a root ploidy specific transcriptome also allowed the identification of ploidy specific gene expression (Bhosale *et al.*, 2018).

In fully grown tomato fruits, the pericarp tissue, corresponding to the fleshy part, is composed of a heterogeneous population of cells, with cell sizes ranging from 300 μm^2 to 45000 μm^2 and ploidy levels ranging from 2C to 128/256C (Bourdon *et al.*, 2012). The pericarp tissue can be divided in three parts: the exocarp that corresponds to the small cells at the external side, the mesocarp containing the larger cells in the center and the endocarp corresponding to the two cell layers in the internal side (Renaudin *et al.*, 2017). The mapping of endoreduplication in the tomato pericarp revealed that ploidy levels are distributed as a gradient toward the center of the pericarp and that all the cells inside a cell layer do not have the same ploidy level (Bourdon *et al.*, 2011). However, the reason for this heterogeneity is not understood.

Endoreduplication in the pericarp of tomato fruit is thus not randomly nor evenly distributed. In addition, endoreduplication is able to induce a global shift of gene expression and specific set of genes are differentially regulated as a function of ploidy levels in mature green fruits. However, there is no information about the dynamic distribution of endoreduplication during early fruit development in the pericarp nor ploidy specific gene expression. Ploidy mapping and transcriptome study during fruit development could help understanding how endoreduplication progresses and determines subpopulations of cells.

In this study, we obtained and analyzed the transcriptome of nuclei sorted based on their ploidy level from tomato pericarp. We studied young fruits at three stages of development, when ploidy levels rise and endoreduplication is establishing. In parallel, and for the same developmental stages, we mapped ploidy levels in the pericarp using fluorescent *in situ* hybridization (FISH). Finally, we integrated these two datasets with the aim of mapping gene expression within the pericarp as a function of ploidy.

2. Results

A) *Ploidy specific transcriptome during early Tomato fruit development: experimental setup and data-preprocessing*

To explore the variation in gene expression based on ploidy levels of cells, we used pericarp from fruits at three developmental stages: a stage at which most cells are dividing, a stage corresponding to the end of the cell division period and finally a stage at which the pericarp cells enter endoreduplication (Renaudin *et al.*, 2017). For these stages, corresponding to 6-, 9- and 12-day post anthesis (DPA), ploidy levels increased over time, and the proportion of ploidy classes also changed (Figure 25A). Indeed, ploidy levels ranged from 2C to 16C for 6DPA and from 2C to 64C for 9 and 12DPA fruits. Nuclei from the pericarp cells were collected by Fluorescent Activated Cytometry Sorting according to their DNA content (Pirrello *et al.*, 2018, *see Chapter 3: Materials and methods*). For each time-point, four populations of nuclei- 2C, 4C, 8C and 16C for 6DPA and 4C, 8C, 16C and 32C for 9 and 12DPA corresponding to the 4 major classes of ploidy at these stages, were recovered and used for RNA extraction and sequencing (Pirrello *et al.*, 2018, *see Chapter 3: Materials and methods*).

After quality control and pre-processing (Pirrello *et al.*, 2018, *see Chapter 2.1.2.A*): *Use of spike-in to normalize gene expression in a global shift in expression context*, the reads were mapped to the Sly 1.0 Tomato reference genome. After mapping, the number of uniquely mapped reads corresponding to reads aligned at only one locus was almost similar between the samples and ranged between 13k and 20k transcripts. All transcripts annotated on the Tomato genome are not represented because the sequenced RNA is extracted from pericarp cells and all genes are not expressed all the time in all organs, but the number of transcripts correspond to transcript number found in Pirrello *et al.*, 2018.

To determine the variation in expression within our dataset, we assessed intra- and inter-group variability using Principal Component Analysis (PCA), with a group corresponding to the transcriptome of nuclei of a certain ploidy level at one time-point (Figure 25B). We first observed that the biological replicates clustered together in the PCA plot, indicating a good reproducibility of the expression values. The PCA analysis also showed that 80% of the variance between samples was explained by two components. The first component, PC1, explained 56% of the variance between the samples and corresponded to the changes in ploidy levels. The second component, PC2, explained 24% of the variance and corresponded to the developmental stages (Figure 25B). We noticed that 2C nuclei at 6DPA are really close to 4C at the same stage and it is also the case of a 32C sample at 9DPA that is close to a 16C. The PCA analysis thus showed that the samples could be used to study the expression changes in function of the developmental stages and in function of the changes in ploidy levels.

B) *Gene expression is changing during early fruit development and based on ploidy levels.*

To study the changes in gene expression dependent on ploidy or developmental timing, we searched for differentially expressed genes (DEG) between the samples of

our dataset. For that, after normalization (Pirrello *et al.*, 2018), a pairwise comparison of the expression values was done and expression fold changes between two ploidy levels at a specific developmental stage were calculated. We found that 1938, 1921 and 1878 genes were differentially expressed in at least one comparison with a FDR < 0.01 and log₂ FC >1 at 6DPA, 9DPA and 12DPA respectively (Figure 25C). For each time point and as previously shown by Pirrello *et al.*, 2018 the number of DEG, up- or down-regulated increased when comparing more distant ploidy levels.

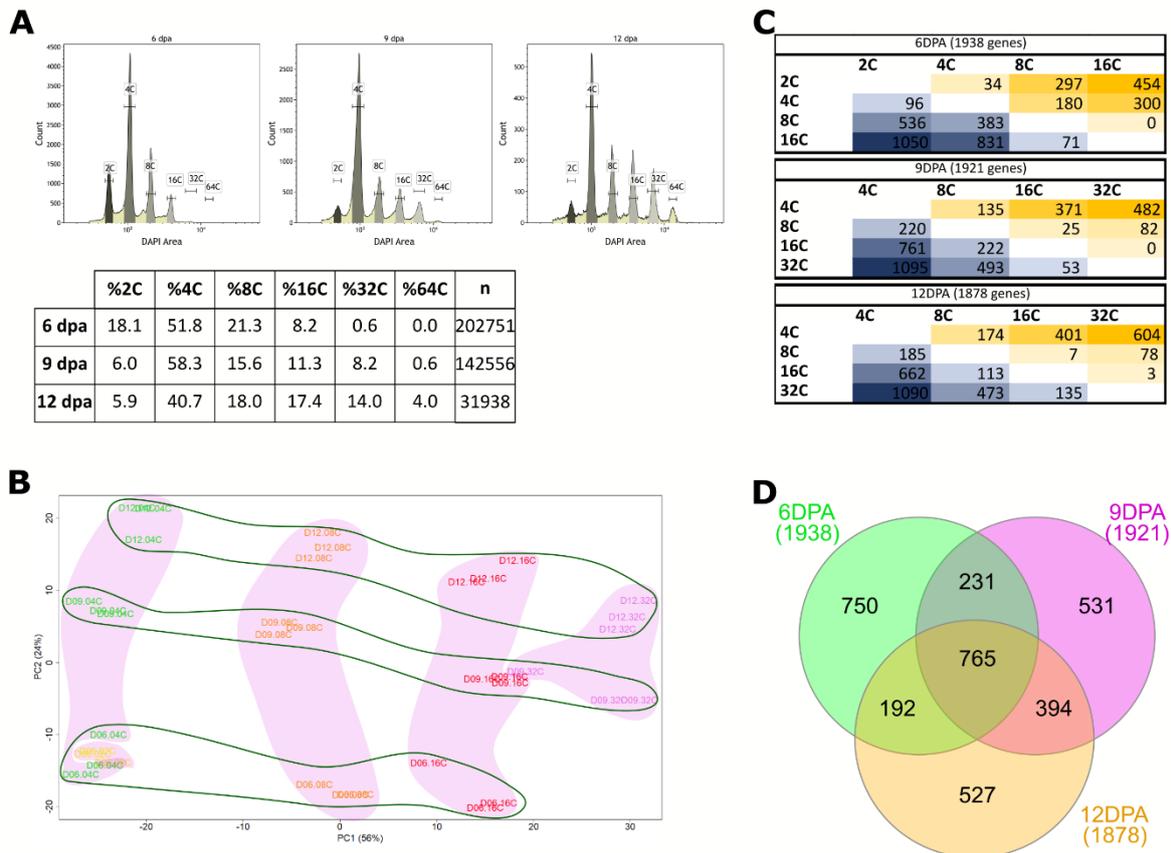


Figure 25. Ploidy sorted nuclei from 6, 9 and 12 DPA pericarp fruits show differential gene expression depending on ploidy and developmental stage.

(A) Flow cytometry histograms of DNA quantities in nuclei of Tomato pericarps at 6DPA, 9DPA and 12DPA. The table represents the distribution of ploidy classes at 6, 9 and 12 DPA (2C to 64C) and n is the number of nuclei counted in a stage to determine ploidy proportions.

(B) Principal component analysis (PCA) of ploidy specific transcriptome data. The two first components explain 80% of data variability. Samples are indicated as the corresponding developmental stage (e.g. the three D09.04C correspond to 4C nuclei of 9DPA pericarp). Font colors indicate the ploidy levels: yellow for 2C, green for 4C, orange for 8C, red for 16C and pink for 32C. Green contours indicate similar developmental stages and pink overlay indicates similar ploidy levels.

(C) Number of genes up- (yellow) and downregulated (blue) in pairwise comparisons of the expression values (false discovery rate ≤ 0.01 , FC > 2).

(D) Venn diagram representing the common genes found between ploidy specific DE genes at 6DPA, 9DPA and 12DPA.

For example, in 6DPA fruits, 34, 297 and 454 were up regulated when comparing 2C nuclei to 4C, 8C and 16C nuclei, respectively. Moreover, among the DEG from all developmental stages, there were 765 in common (Figure 25D) and 231, 394 and 192 specifically in between 6 vs 9DPA, 9 vs 12DPA and 6 vs 12DPA, respectively. The common DEG found between the different time points could indicate that some genes are specifically regulated in a ploidy specific manner during the whole development.

To identify potential specific expression profiles related to ploidy level changes, the 1938, 1921 and 1878 DE genes were clustered based on their co-expression at a specific time-point, by a hierarchical clustering method (R Core team, 2017).

In 6DPA fruit pericarps, genes with similar expression profiles were grouped into 9 clusters corresponding to expression profiles ranging from high expression in 2C nuclei to high expression in 16C nuclei (clusters 6-A to 6-I, Figure 26A). Four main expression trends were visible, corresponding to a peak of expression in 2C nuclei (Clusters A and B), in 2C and 4C nuclei (Clusters C, D), in 4C nuclei (cluster E) and in endoreduplicated (8C and 16C) nuclei (Clusters F to I).

In 9DPA and 12DPA fruit pericarps nuclei, for which expression profiles analysis was done on nuclei with ploidy levels ranging from 4C to 32C, DEG with similar expression profiles were grouped in 9 and 10 clusters (clusters 9-A to 9-I and clusters 12-A to 12-J, Figure 26B and 26C), respectively. Expression profiles in the clusters from 9 and 12DPA were very similar, and four major trends were observed. These four types of expression profiles corresponded to a peak of expression in 4C nuclei (Clusters 9-A/B and 12-A/B), in both 4C and 8C nuclei (clusters 9-C and 12-C/D), in 8C nuclei (clusters 9-D and 12-E) and finally, the last trend was represented by clusters of genes having high expression in polyploid nuclei (8C and 16C: clusters 9-E and 12-F; , 8C, 16C and 32C : clusters 9-F and 12-G; and higher ploidy levels, 16C and 32C, clusters 9-G to 9-I and 12-H to 12-J).

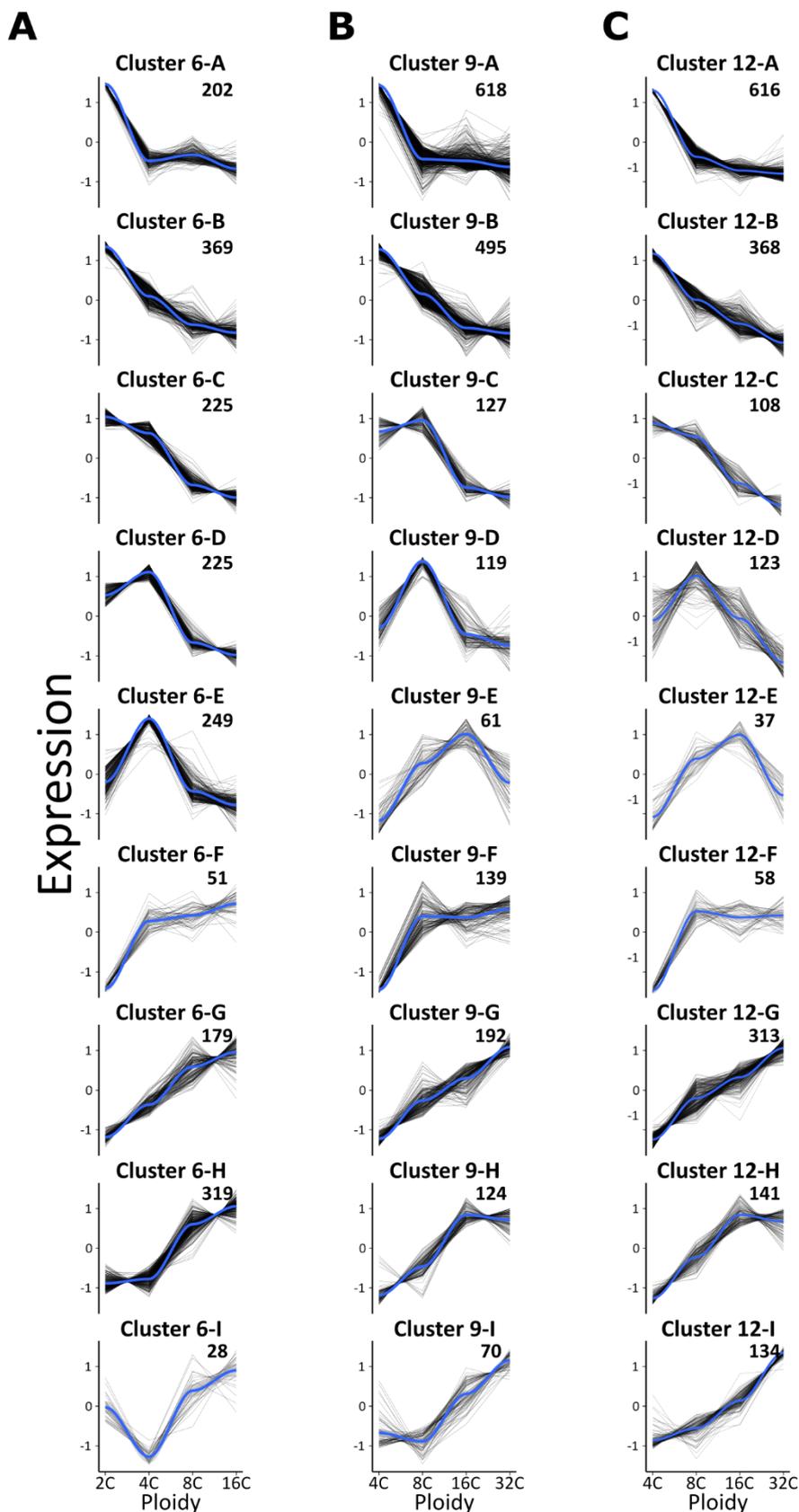


Figure 26. Clustered normalized expression profiles of differentially expressed genes in pericarp of 6-9-12DPA fruits

Nine clustered expression profiles of DE genes obtained with an ascendent hierarchical clustering for 6DPA (A), 9DPA (B) and 12DPA (C). On a cluster plot, the blue line represents the median profile and the top right number is the total number of genes.

To facilitate gene set and statistical analysis, we applied a new clustering to obtain a minimal set of clusters reporting the major expression profiles. This new clustering yielded five clusters for each time-point (Figure 27A-C) in which the main expression trends previously observed were visible: peak of expression in 2C (clusters 6-1/6-2, Figure 27A), 4C (clusters 6-3, 9-1/9-2 and 12-1/12-2, Figure 27A-C), 8C (clusters 9-3 and 12-3, Figure 27B-C), 16C (cluster 6-5, Figure 27A) or 32C (clusters 9-5 and 12-5, Figure 27B-C) nuclei; high expression in low ploidy levels nuclei, 2C and 4C, (cluster 6-2, Figure 27A) and high expression in polyploid nuclei, 8C, 16C and 32C (clusters 6-4, 9-4 and 12-4, Figure 27A-C).

For each time-point, these trends thus corresponded to specific expression at a particular ploidy level, in nuclei with low ploidy levels or with high ploidy levels. During fruit development, transcriptional dynamics are thus associated with changes in ploidy levels.

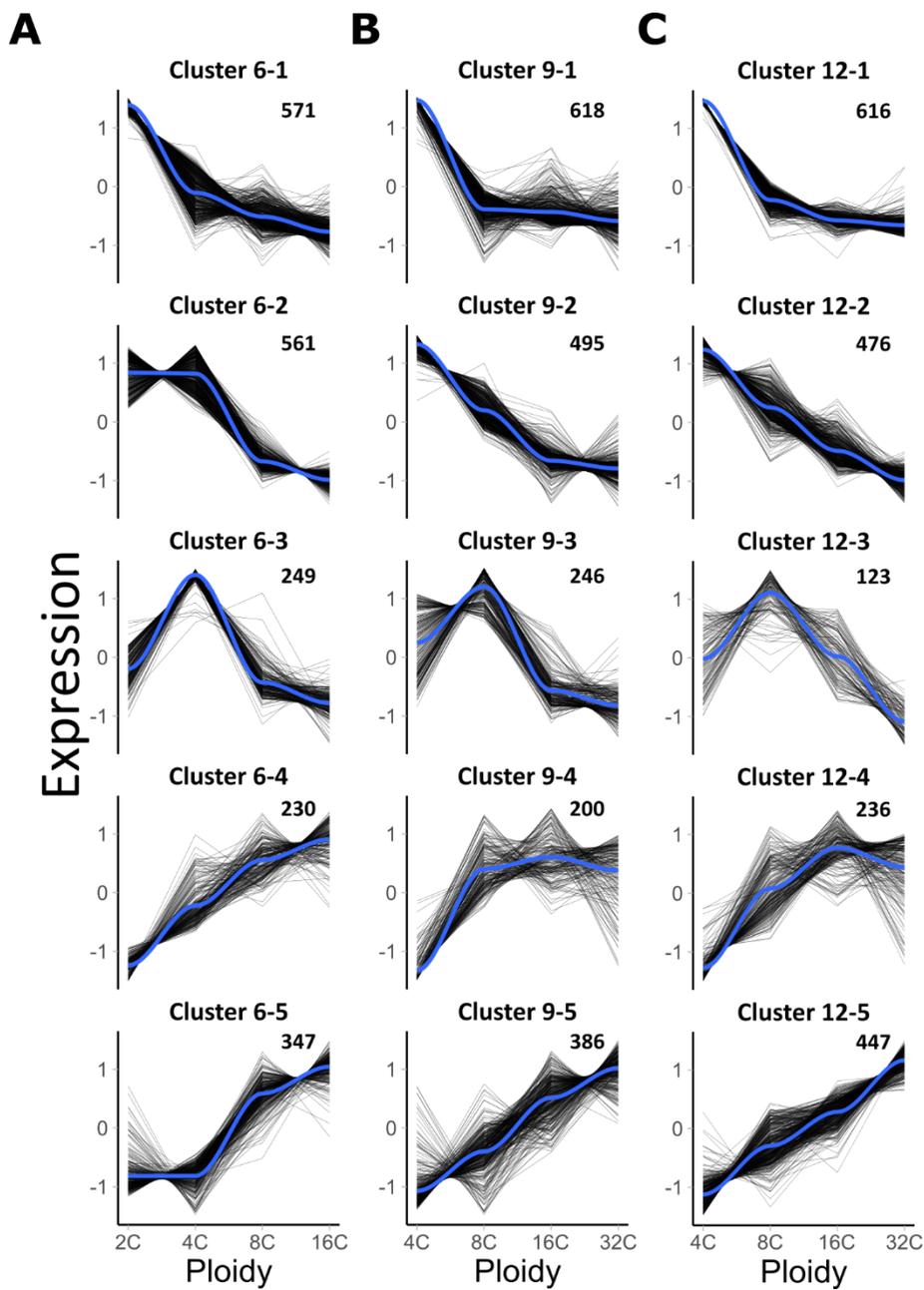


Figure 27. Clustered normalized expression profiles of differentially expressed genes in pericarp of 6-9-12DPA fruits

Five clustered expression profiles of DE genes obtained with an ascendent hierarchical clustering for 6DPA (A), 9DPA (B) and 12DPA (C). On a cluster plot, the blue line represents the median profile and top right number is the total number of genes.

C) *Biological processes expressed as a function of ploidy.*

To identify the biological processes associated to ploidy and developmental stage specific expression profiles, we estimated over-representation of functional categories by using the Gene Ontology (GO) enrichment tool from PLAZA 4.5 (Van Bel *et al.*, 2017) combined with the REVIGO web tool that allows to compute GO term enrichment lists after removing redundant terms (Supek *et al.*, 2011, Bonnot *et al.*, 2017). This combined use of GO enrichment analysis and REVIGO tool permits to highlight enrichment in specific GO terms by considering the hierarchy of the GO terms, thus making data easier to interpret. Briefly, the GO enrichment tool is attributing to a set of genes, GO terms and an associated p-value resulting from an enrichment test assaying the representativity of the GO term in the sample compared to the total number of genes associated to this GO term in the genome. In REVIGO, the list of GO terms and their associated p-value are computed to determine for each term a redundancy value (dispensability) and a value representing how the term is an outlier compared to the input list (uniqueness). Then the computed REVIGO values can be used to reduce large GO lists containing redundant enriched GO terms.

By combining the GO terms enriched in all clusters from the three time points studied, we ended up with 250 unique terms reduced to 43 after the REVIGO analysis. We found 30, 19 and 15 significantly enriched GO terms in the different clusters from 6DPA, 9DPA and 12DPA fruits, respectively. With this analysis, we found that the number of enriched GO terms varied between clusters but was not following the number of genes within a cluster, which varied from 123 to 618 (Figures 28-30). For example, in cluster 6-3 containing 249 genes seven terms grouping 120 genes were enriched while in cluster 6-5 composed of 347 genes only 3 terms grouping 68 genes were found enriched.

a. **Biological processes expressed in 6DPA fruits**

In 6DPA fruit pericarps, cluster 6-1 containing genes highly expressed in 2C cells and cluster 6-3 composed of genes with specific expression in 4C (Figure 28) nuclei were enriched in GO terms mainly associated to the classical cell cycle process. For cluster 6-1, the terms “*chromosome organization*” (55 genes), “*cell cycle*” (42 genes), “*DNA replication*” (45 genes) were found and included *RBR1* and *E2Fb* genes, encoding proteins important for the regulation of cell division by modulating the expression of many genes involved in cell cycle progression (Figure 28A, Gutierrez 2009, Abraham and del Pozo 2012, Ószi *et al.*, 2020). The target genes of *RBR1* and *E2Fb* were also found in the term “*DNA replication*” and include the subunits of the *ORIGIN RECOGNITION COMPLEX (ORC)*, *ORC4/5/6* (Gutierrez 2009). In addition, the *MINI CHROMOSOME MAINTENANCE (MCM)* genes, *MCM2* to *7* were found in the term “*chromosome organization*”. Together, *ORC* and *MCM* proteins are implicated in the initiation of DNA replication between the Gap1 (G1) phase at which the cell decides to continue into cell cycle or stops in function of environmental and internal signals, and the S phase, DNA replication phase of the cell cycle (DePamphilis 2005, Tujeta *et al.*, 2012).

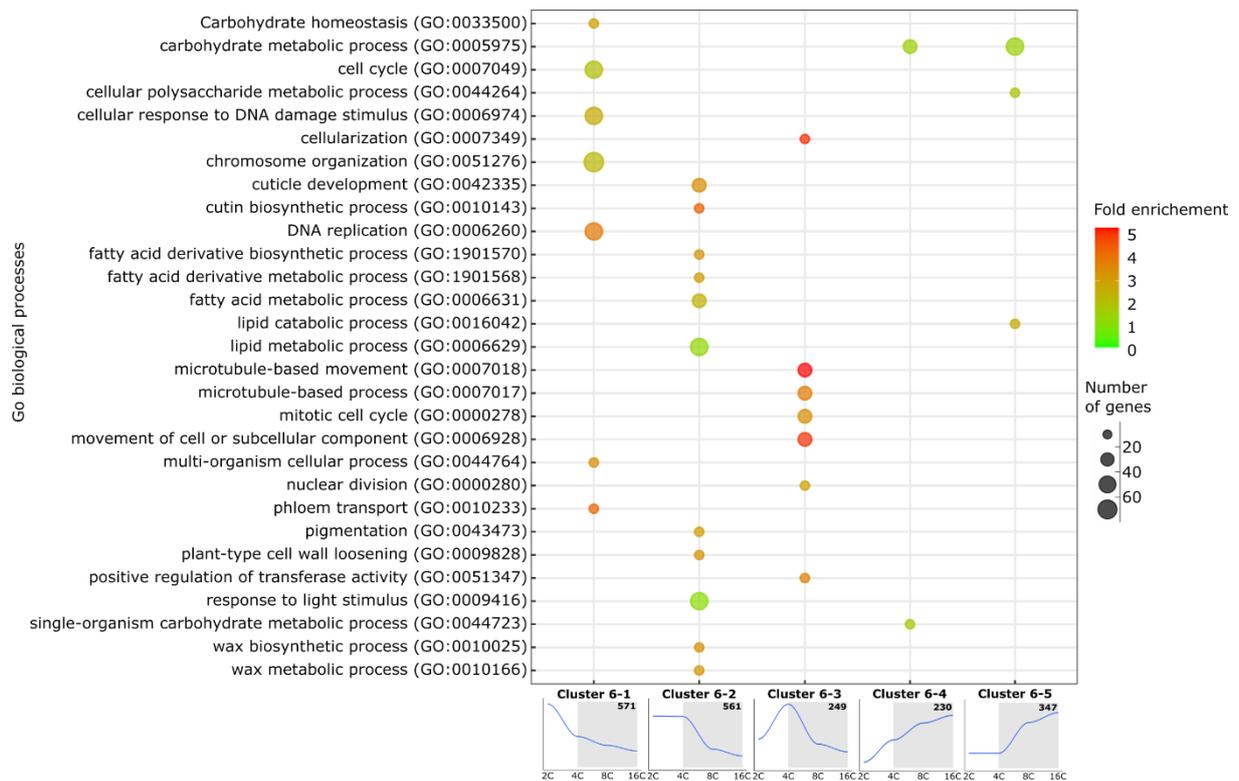


Figure 28. Overrepresented biological processes in the different expression clusters from 6 DPA pericarp

(A) Dot plot showing the most representative and significant biological processes in each cluster. Dot size represents the number of genes in the concerned biological process and color scale the fold enrichment of gene number compared to the expected number in the genome. Below the dotplot are represented views of 6DPA ploidy specific clusters. The grey square indicates the ploidy levels analyzed at all developmental stages (4C to 16C).

In the cluster 6-3, the enriched terms corresponded to "*microtubule-based movement*" (25 genes), "*mitotic cell cycle*" (22 genes), "*movement of cell or subcellular component*" (25 genes) and "*nuclear division*" (11 genes). Related to these terms, *CYCLIN* genes such as *CYCA1*, *CYCB2* having a role in the mitotic cycle and the G2/M transition (Menges *et al.*, 2005, Sabelli *et al.*, 2014), *CYCB3* and the *CYCLIN DEPENDENT KINASE*, *CDKB*, were found in this cluster. *CDKB* is involved in the G2 to M transition in Tomato fruit (Czerednik *et al.*, 2012). Moreover, several kinesin genes (25 genes) were found in this cluster, with for example the centromeric kinesin *PHRAGMOPLAST ORIENTATING KINASE*, *POK1*, encoding a protein responsible for phragmoplast orientation during the M phase of the cell cycle (Müller *et al.*, 2006). The phragmoplast is a plant specific structure that forms during the cytokinesis, division of the cell, and serves as a scaffold for the formation of the new cell wall that will separate the daughter cells.

The genes found in clusters 6-1 and 6-3 thus suggest that at 6DPA; the 2C nuclei could correspond to cells in late G1 phase (G1/S transition) whereas the 4C nuclei would correspond to cells in late G2 phase, maybe at the G2/M transition.

In cluster 6-2 representing highly expressed genes in 2C and 4C nuclei, we found 11 enriched terms mostly associated with cuticle formation such as "*lipid metabolic process*" (47 genes), "*fatty acid metabolic process*" (29 genes), "*fatty acid derivative metabolic process*" (11 genes), "*cuticle development*" (17 genes), "*cutin biosynthetic process*" (9 genes), "*wax metabolic process*" (11 genes) and "*plant type cell wall loosening*" (8 genes) (Figure 28A). Cuticle is a protective layer localized at the surface of plant organs and produced by the first cell layers of the organ (Isaacson *et al.*, 2009). The role of cuticle is to protect organs primarily from water loss (Kenrick and Crane 1997) but also from pathogens (Fu *et al.*, 2018). Cuticle is composed of a cutin membrane composed of insoluble polyesters covered by cuticular waxes that are a mainly very long chain fatty acid (Fernandez *et al.*, 2016). In Tomato, the fruit is covered by a thick cuticle layer having an important protective role. Cuticle synthesis and deposition at the fruit surface represent important events occurring during fruit development and regulated by a complex set of genes (Vogg *et al.*, 2004, Isaacson *et al.*, 2009, Hen-Avivi *et al.*, 2014). Several genes associated to cuticle synthesis in plants (Figure 32) were thus found in the cluster 6-2 and include very long chain fatty acid synthesis genes such as β -*KETOACYL-COA SYNTHASE* (*KCS 3/6/10*), *ECERIFERUM1* and *3* (*CER 1* and *CER3*) and *GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE 4* (*GPAT4*) that encode enzymes responsible for fatty acid elongation, alkane formation (wax biosynthesis) and cutin polyester formation, respectively. Wax and cutin are the principal constituents of cuticle and need to be exported in the extracellular space (Bernard *et al.* 2012, Bernard and Joubes 2013, Lashbrooke *et al.*, 2015). Transportation of cuticle polymers is performed by ATP BINDING CASSETTE G proteins, *ATCG32* for cutin and *ATCG11* for wax (Bessire *et al.*, 2011, Lashbrooke *et al.*, 2015). Interestingly, we found both *ATCG32* and *ATCG11* in cluster 6-2.

Clusters 6-4 and 6-5, containing genes with high expression in 8C and 16C nuclei (Figure 28B), had only 2 and 3 enriched GO terms, respectively. In cluster 6-4, 15 genes belonged to the GO category "*carbohydrate metabolic process*" that is associated with central metabolism and are particularly related to sucrose metabolism, glycolysis and

pentose phosphate pathway (Figure 28A). Among these genes, the *SUCROSE SYNTHASE (SUSY)* and *NEUTRAL INVERTASE (NI)*, responsible for fruit imported sucrose cleavage into fructose and glucose, were found. The *PYRUVATE KINASE (PK)* and the *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (GAPDH)* involved in glycolysis and the *GLUCOSE-6-PHOSPHATE DEHYDROGENASE (G6PDH)*, involved in the first step of the pentose phosphate pathway, were also found. In cluster 6-5, terms related to catabolism and synthesis of complex sugars were enriched: "*carbohydrate metabolic process*" (35 genes), "*cellular polysaccharide metabolic process*" (15 genes). These terms correspond mainly to 15 genes among which 10 encode proteins involved in starch synthesis, secondary cell wall synthesis and modifications. For example, we found the *GLUCOSE-1-PHOSPHATE ADENYLYLTRANSFERASE* which encodes the large subunit of the ADP-GLUCOSE PYROPHOSPHORYLASE catalyzing the first rate-limiting step in starch synthesis (Balicora *et al.*, 2004) and the *1-4-ALPHA-GLUCAN BRANCHING ENZYME* encoding an enzyme responsible for starch branching. Finally, for cell wall modifications, we found four *XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE (XTH)* and a *GLYCOSYLTRANSFERASE* participating in the modification of existing cell wall and linkage in the cell wall matrix (Terao *et al.*, 2013, Amos and Mohmen 2019). In conclusion, at 6DPA, non endoreduplicated cells are actively dividing and involved in cuticle synthesis (clusters 6-1 to 6-3). Genes found in cluster 6-4, highly expressed in 8C and 16C cells, indicate that at 6DPA endoreduplicated cells are specialized in sugar metabolism and this might be related to metabolic fluxes in young Tomato fruit during cell proliferation phase. Moreover, in endoreduplicated cells at 6DPA, the intense cell wall related activity could be in relation with cell growth occurring in the mesocarp (Terao *et al.*, 2013; Renaudin *et al.*, 2017).

b. Biological processes expressed in 9DPA fruits

In 9DPA fruit pericarp, cluster 9-1 in which expression peaked in 4C nuclei (Figure 29B), was enriched with 5 GO terms (Figure 29A), including "*carbohydrate catabolic process*" (14 genes) and "*wax biosynthetic process*" (8 genes). Among these genes, several are associated with cuticular wax synthesis (*CER1* and *WAX ESTER SYNTHASE/ACYL-COA DIACYLGLYCEROL ACYLTRANSFERASE: WSD1*) and glycolysis (*ENOLASE* and *PK*). In the "carbohydrate catabolic process" category, we found genes involved in sugar transportation and cell wall related genes. For example, we found a *BETA-FRUCTOFURANOSIDASE* responsible for sucrose import from the vascular tissues (Friedman and Zamir 2003, Shen *et al.*, 2019) and an *ALPHA-L-ARABINOFURANOSIDASE*. In *Arabidopsis*, the *ALPHA-L-ARABINOFURANOSIDASE*, by hydrolyzing arabinose branching of pectins, plays a role in secondary wall synthesis of vascular tissue cells (Montes *et al.*, 2008). Arabinose modification is important for cell adhesion and is necessary for maintenance of stomatal closure (Jones *et al.*, 2003). Arabinose modification has also been implicated in the fruit firmness of the mutant *Cnr* (*colorless non-ripening*) that stay green (Orfila *et al.*, 2001).

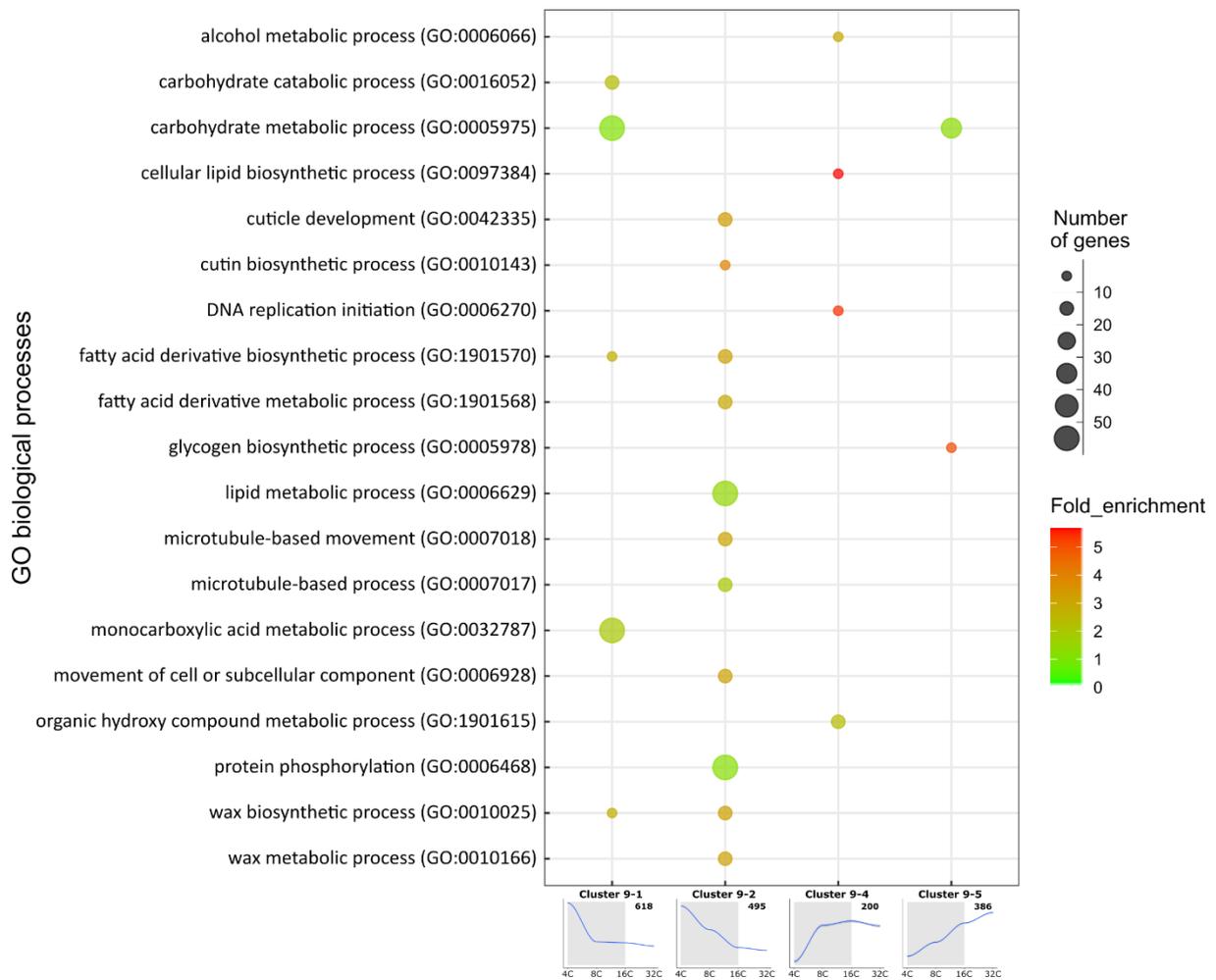


Figure 29. Overrepresented biological processes in the different expression clusters from 9 DPA pericarp

(A) Dot plot showing the most representative and significant biological processes in each cluster. Dot size represents the number of genes in the concerned biological process and color scale is representing the fold enrichment of gene number compared to the expected number in the genome. Below the dotblot are represented views of 9DPA ploidy specific clusters. The grey square indicates the ploidy levels analyzed at all developmental stages (4C to 16C).

In cluster 9-2, genes are also highly expressed in 4C cells, as in cluster 9-1, and 11 GO terms were found to be enriched (Figure 29). A large part of those terms are related to cuticle synthesis pathways such as for example “*cuticle development*” (12 genes), “*cutin biosynthesis process*” (5 genes) or “*lipid metabolic process*” (44 genes), but we also found a term associated to microtubule, “movement of cell or subcellular component”. The “cuticle related terms” contained between 12 and 44 genes including *CER1* and *NON-SPECIFIC LIPID-TRANSFER PROTEIN (nsLTP)* genes involved cutin and wax secretion in the apoplast for the formation of cuticle (Liu *et al.*, 2015, Petit *et al.*, 2017). The “microtubule related terms” contained 11 genes with myosin (2 genes) and kinesin (9 genes) genes, usually involved in microtubule motor transport of large structure in cells (Li *et al.*, 2012).

Cluster 9-4 and 9-5 are clusters with high gene expression in endoreduplicated nuclei, 8C, 16C and 32C (Figure 29B). The cluster 9-4 was enriched in terms associated with hydroxylation of molecules such as “*organic hydroxy compound metabolic process*” (12 genes), “*alcohol metabolic process*” (8 genes) and “*cellular lipid biosynthetic process*” (3 genes) (Figure 29A). These terms corresponded to 12 genes with 4 related to the carotenoid pathway such as the *PHYTOENE SYNTHASE 1/2 (PSY1/2)*, the *ZEAXANTIN EPOXIDASE (ZEP)* that encode genes synthesizing precursors of tomato pigments and abscisic acid (Kachanovsky *et al.*, 2012, Ariizumi *et al.*, 2014, Karniel *et al.*, 2020). Moreover, we found an *ALDEHYDE OXIDASE* gene (*AAO3*) responsible for the last step of the synthesis of abscisic acid (ABA), important hormone implicated in Tomato fruit cell expansion (Nitsch *et al.*, 2012), that derives from the carotenoid pathway. In addition, two genes related to the inositol pathway, responsible for ascorbic acid (AsA) synthesis (Munir *et al.*, 2020) were found. Finally, the term “DNA replication initiation” (5 genes) was also enriched in this cluster with genes encoding MINI CHROMOSOME MAINTENANCE proteins (MCM 2/4/6) and a CELL DIVISION CONTROL PROTEIN 6 HOMOLOG (*CDC6*). Some MCM proteins have been associated with endocycle in mammals and maize endosperm (Edgar *et al.*, 2001, Sabelli *et al.*, 2013) whereas *CDC6* proteins have been associated both with cell cycle and endocycle replication processes (Castellano *et al.*, 2001, Gutierrez 2009). Genes usually associated with DNA repair are also found under the enriched term “DNA metabolic process” such as a *DNA REPAIR PROTEIN MUTS* homolog and *BREAST CANCER TYPE 1 (BARD1)* and its partner *BREAST CANCER SUSCEPTIBILITY1 (BRCA1)*.

These expression data from cluster 9-4 and 9-5 might indicate that at 9DPA, endoreduplicated cells (8C, 16C and 32C) would be already engaged into secondary metabolism pathways responsible for color of Tomato fruits and in ABA production in the pericarp.

The last cluster, 9-5, was enriched in only two GO terms associated with metabolism of sugars and starch: “*carbohydrate metabolic process*” (32 genes) and “*glycogen biosynthetic process*” (4 genes). These two terms contain the same genes also found in cluster 6-5 that are related to secondary cell wall synthesis and starch synthesis. Although cluster 9-3 contained 243 genes, no enrichment in biological process terms was found.

The terms related to metabolism of sugars and starch indicate that 16C and 32C cells

at 9DPA contain genes that are related to sugar metabolism. The presence of *G6PDH* is indicating an activation of the pentose phosphate pathway for the biosynthesis of nucleotides, consistent with needs of endoreduplicated cells. Finally, sugar metabolism related terms are giving an insight at cellular level of the pericarp metabolism that is potentially driven by endoreduplicated cells (Biais *et al.*, 2014, Colombié *et al.*, 2014).

c. Biological processes expressed in 12DPA fruits

At 12DPA, clusters with genes showing decreasing expression with ploidy, clusters 12-1 and 12-2 (Figure 30B), were enriched in GO terms associated with cuticle development and microtubule dynamic for cluster 12-1 ("*lipid metabolic process*" (57 genes), "*cutin biosynthetic process* (8 genes)" and "*microtubule-based movement*" (20 genes)) and modification of protein activities and developmental process for cluster 12-2 ("*lipid metabolic process*" (35 genes), "*protein phosphorylation*" (46 genes) and "*anatomical structure morphogenesis*" (50 genes)) (Figure 30A).

Lipid and cutin related terms in cluster 12-1 mostly corresponded to the cuticle biosynthetic pathway genes and microtubule associated terms corresponded to genes encoding kinesin proteins and kinesin proteins are implicated in several cellular event such as meiosis, mitosis movement of chloroplast or in the morphogenesis of trichomes (Li *et al.*, 2012). In cluster 12-2, the lipid related term contained a set of genes associated with cuticle synthesis and inositol mediated signaling. The inositol mediated signaling is represented by three *INOSITOL 1 4 5-TRISPHOSPHATE 5-PHOSPHATASE*, one *GLYCOSYLPHOSPHATIDYLINOSITOL-ANCHORED WALL TRANSFER PROTEIN 1* and one *PHOSPHOINOSITIDE-SPECIFIC PHOSPHOLIPASE C* that have been described in Arabidopsis to be involved in ABA signaling (Gunesequera *et al.*, 2007, Sanchez *et al.*, 2001, Carland and Nelson 2004). In the cluster 12-2, we found terms related to organ morphogenesis and protein phosphorylation among which 29 genes corresponded to *RECEPTOR LIKE KINASE* (RLK). RLKs belong to a large protein family involved in various plant developmental processes such as growth, stress (abiotic or biotic) and cellular signaling and response (Jose *et al.*, 2020). The 50 genes associated to organ morphogenesis were mainly associated to hormones, mostly auxin. Seven genes related to auxin signaling and transport were found with for example, *AUXIN RESPONSE FACTOR 9* involved in inhibiting cell division (Wang *et al.*, 2005), and auxin transporters *AUXIN1/LIKE-AUX* (*SILAX3*) and polar transporter *PIN FORMED 9* (*SIPIN9*). Six genes associated to organ morphogenesis belonged to class IV Homeodomain-Leucine Zipper Gene Family (HD-Zip IV) and include *PROTODERMAL FACTOR2 PDF2* and *GLABRA2*, that are found in the six genes, both necessary for differentiation of epidermal cells and trichomes (Abe *et al.*, 2003, Nakamura *et al.*, 2006). Since cluster 12-2 contains genes with high expression in 4C nuclei, the enrichment in the described GO terms might indicate that the 4C cells analyzed would still be engaged in cell division and cuticle synthesis and could be localized in the outer-epidermis layers in 12DPA fruits. However, these 4C cells or a sub-population of these cells could be engaged in differentiation and integration of developmental signals through hormone signaling and RLK activity.

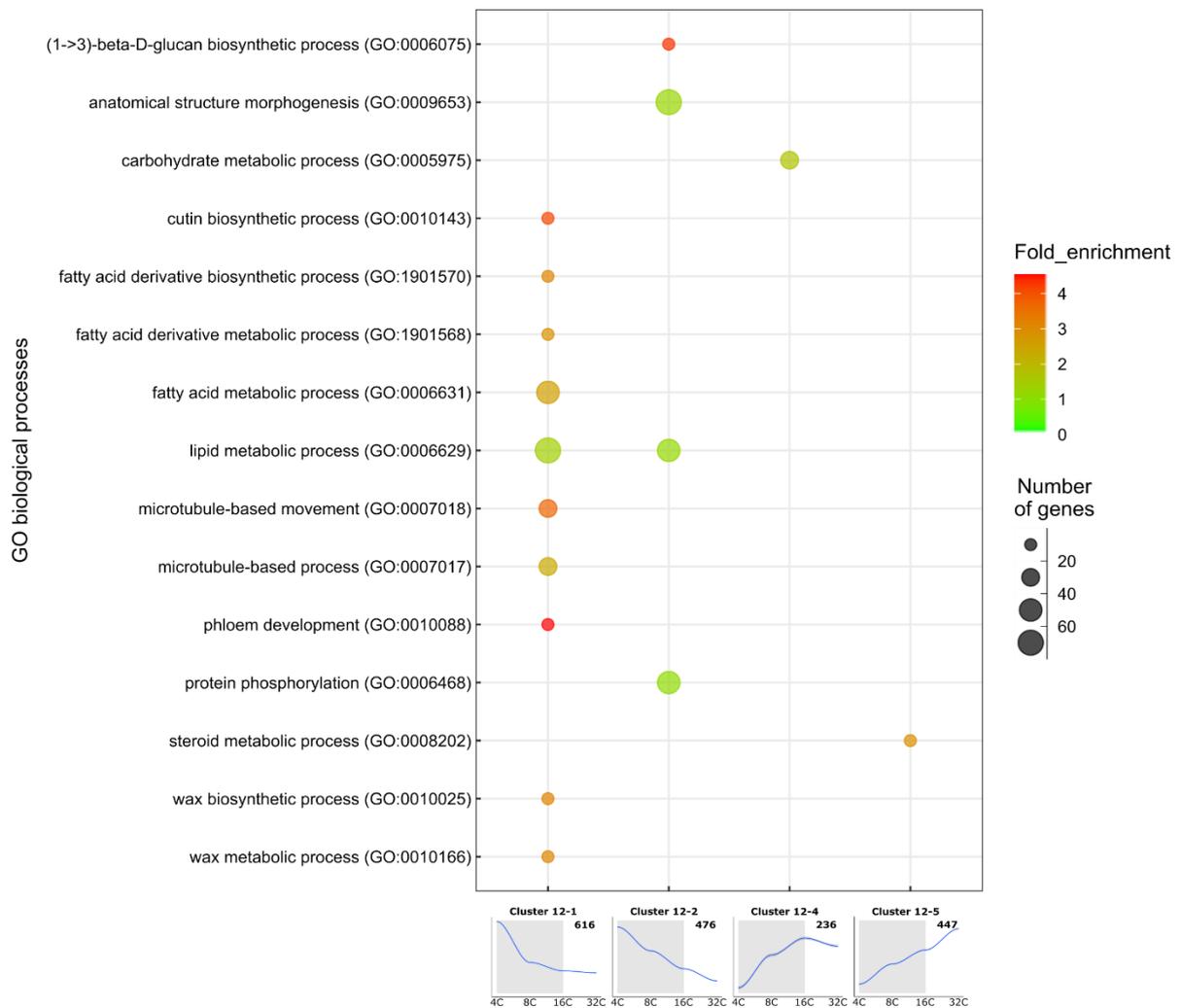


Figure 30. Overrepresented biological processes in the different expression clusters from 12 DPA pericarp

(A) Dot plot showing the most representative and significant biological processes in each cluster. Dot size represents the number of genes in the concerned biological process and color scale is representing the fold enrichment of gene number compared to the expected number in the genome. Below the dotplot are represented views of 12DPA ploidy specific clusters. The grey square indicates the ploidy levels analyzed at all developmental stages (4C to 16C).

The cluster 12-4 was only enriched in one metabolic related term, “*carbohydrate metabolic process*” (28 genes) with genes implicated in secondary cell wall synthesis modifications and sugar metabolism (Figure 30A). Cell wall related genes were for example a *GLYCOSYLTRANSFERASE FAMILY GT8 PROTEIN (GAUT)* responsible for the synthesis of homogalacturonan (HG), the most represented polysaccharide in pectin (Atmotjo *et al.*, 2011) and one of the XTH found in cluster 6-5 and 9-5. Sugar metabolism genes were related to glycolysis and interestingly a gene encoding an enzyme responsible for the dephosphorylation of Trehalose-6-phosphate, the *TREHALOSE-6-PHOSPHATE PHOSPHATASE* was found. Trehalose-6-phosphate is able to inhibit SNF1-related/AMPK protein kinases activity and SnRK1 is responsible of sensing energy homeostasis to regulate pathways activities (Hardie 2007), is a positive regulator of starch synthesis and promote cell and organ size growth (Li and Sheen 2016). *TREHALOSE-6-PHOSPHATE PHOSPHATASE* turns T6P into a non-inhibiting form for snRK1 and might be responsible for removing the inhibition of starch synthesis in crop plants (Zhang *et al.*, 2009, Paul *et al.*, 2018). In Tomato mechanisms of sugar sensing are still unknown, however transgene experiment expressing the Apple SnRK1 in tomato resulted in reducing the duration of fruit development and in an increased nitrogen assimilation (Wang *et al.*, 2012).

The cluster 12-5 was only enriched in “*steroid metabolic process*” (9 genes) with three genes implicated in brassinosteroid synthesis or signaling (Figure 30A). Brassinosteroids are a group steroidal phytohormones with roles in plant development and stress responses (Zhu *et al.*, 2013). Moreover, Brassinosteroids are implicated in enhancing sugar content and carotenoid accumulation in Tomato and inducing parthenocarpy in cucumber fruit (Nie *et al.*, 2017, Hu *et al.*, 2020, Fu *et al.*, 2008). Finally, the cluster 12-3 was not enriched in any biological process GO term.

To summarize, in 6-9-12DPA fruits cells with low ploidy levels (2-4C) are mainly associated with cuticle synthesis and transport and cell divisions and cells with higher ploidy levels (8-16-32C) are mainly associated to central metabolism (6-9-12DPA) and secondary metabolism (9-12DPA).

D) *Dynamics of ploidy specific gene expression in the course of early fruit development*

To study the evolution of the genetic programs related to ploidy levels in the course of fruit development, we compared the expression clusters from the different time-points, 6, 9 and 12DPA. To do so, the overlap between two clusters was estimated by determining the number of common genes and by calculating the significance of the overlap (Figure 31).

For each cluster, except 6-1, we found at least one significant overlapping cluster. The lack of significant overlap for cluster 6-1 might be explained by the fact that it contains genes with high expression in 2C nuclei, not studied at 9 and 12DPA. : In general, when comparing clusters with similar ploidy levels, significant overlaps were found between clusters presenting similar expression trends, such as when comparing 9-1 and 12-1 (Figure 31).

A

Cluster name		6-1	6-2	6-3	6-4	6-5	9-1	9-2	9-3	9-4	9-5	
	Nb of genes in cluster	571	561	249	230	347	618	495	246	200	386	
	Profile											
9-1	618		77	153	53	2						
9-2	495		61	146	40	1						
9-3	246		25	73	41	1						
9-4	200		23			28						38
9-5	386		2			45						135
12-1	616		99	166	77		256	142	27		1	
12-2	476		62	130	37	2	1	95	132	48		
12-3	123		8	42	23	1		7	20	61	1	
12-4	236		2	1	2	32	48			8	37	61
12-5	447		2	1		49	129	1			46	173

	No significant overlap
	Significant overlap (FDR <0.05)
	Significant overlap (FDR <0.001)

Figure 31. Overlaps between DE genes from the different clusters of 6DPA, 9DPA and 12DPA fruits.

(A) Overlap between genes present in the different clusters at 6DPA, 9DPA and 12DPA. The number of genes in each overlap is shown for each comparison between two clusters. Empty boxes correspond to comparisons with no overlap.

Significant overlap for $P < 0.05$ is highlighted by light blue and for $P < 0.001$ by dark blue.

To identify the biological processes in which genes with common expression profiles over time take part, we compared GO terms from clusters showing common expression trend and significant overlap in the pairwise comparison. Of note, the dataset from 6DPA includes 2C nuclei not studied in 9 and 12DPA, but not 32C nuclei, so attention has to be pay when analyzing the expression trends. We thus compared the GO terms for clusters showing decreasing expression with ploidy levels (6-2, 9-1, 12-1) and the opposite trending clusters with increase expression in higher ploidy levels (6-4,9-4,12-4 and 6-5,9-5,12-5).

In the clusters 6-2, 9-1 and 12-1 in which high expression is found in 4C nuclei, we observed common enrichment in GO-terms related to cuticle development such as “fatty acid derivative biosynthetic process”, “cutin biosynthetic process” and “wax metabolic process”. We also found additional genes associated to cuticle synthesis in other “low ploidy” clusters (clusters 6-1, 6-2, 6-3, 9-1, 9-2 or 12-1 and 12-2). Due to the high representation of cuticle related genes, we searched in literature for key genes implicated in cuticle synthesis, transport and degradation in Tomato fruits (Figure 32, Lashbrooke *et al.*, 2015, Petit *et al.*, 2017, Trivedi *et al.*, 2019). We found 50 genes involved in these different steps among which 40 were DE in our dataset. In 6-9-12DPA fruits, 4C (2C at 6DPA as well) cells expressed all known genes necessary for the synthesis of both cutin and wax (Figure 32) such as *SIGPAT4* (in cluster 6-2/9-1 = 2C and 4C) and *three tomato WSD1 orthologs* (in clusters 6-2/6-3/9-1/9-2/12-1/12-3 = 2C/4C/8C, Li-Beisson *et al.*, 2009, Petit *et al.*, 2016) and some of their transcriptional regulators such as for example *SISHINE1* and *SISHINE3* (Clusters 6-2/9-1/9-3/12-1/12-2 = 2C to 8C, Figure 32, Shi *et al.*, 2013). *SISHINE 1/3* are master activators of most of the cutin and wax synthesis genes and their expression has been associated to epidermal cells (Shi *et al.*, 2013 and Al-Abdallat *et al.*, 2014). Transporters ABCG’s and nsLTP’s necessary to export bricks of cuticle to the extracellular space (Fabre *et al.*, 2016, Lashbrooke *et al.*, 2015) were found in the clusters 6-2, 9-1, 9-2, 9-3, 12-1, 12-2 and 12-3 indicating expression in 2C to 8C cell. Interestingly, 8C cells (Cluster 3) of 9 and 12DPA fruits also expressed genes encoding enzymes necessary for the synthesis and exportation of cutin (Figure 32). Cuticle is synthesized in the epidermal layers of organs (Fernández *et al.*, 2016), and Tomato pericarp is no exception since the three outer-epidermal layers are the main location of cuticle production (Dominguez *et al.* 2008, Segado *et al.*, 2016). Thus, this could indicate that the epidermal layers contain mainly 2C and 4C cells at 6DPA and up to 8C at 9DPA and 12DPA, with 8C cells specialized in cutin production.

Clusters with higher expression in 4C nuclei at 6DPA and 8C cells in 9DPA and 12DPA fruits shared a significant number of genes (6-3x9-3 = 41, 9-3x12-3 = 61 and 6-3x12-3 = 23) however, none of these overlaps was enriched in any GO terms. Among the genes from these overlaps, we found genes important for Tomato fruit development such as *TOMATO AGAMOUS LIKE (TAGL1 and MBP3)* genes, *OVATE LIKE* (annotated *LONGIFOLIA PROTEIN*), and *SUN-like*s (*SUN-like 16* and *SUN-like 30*).

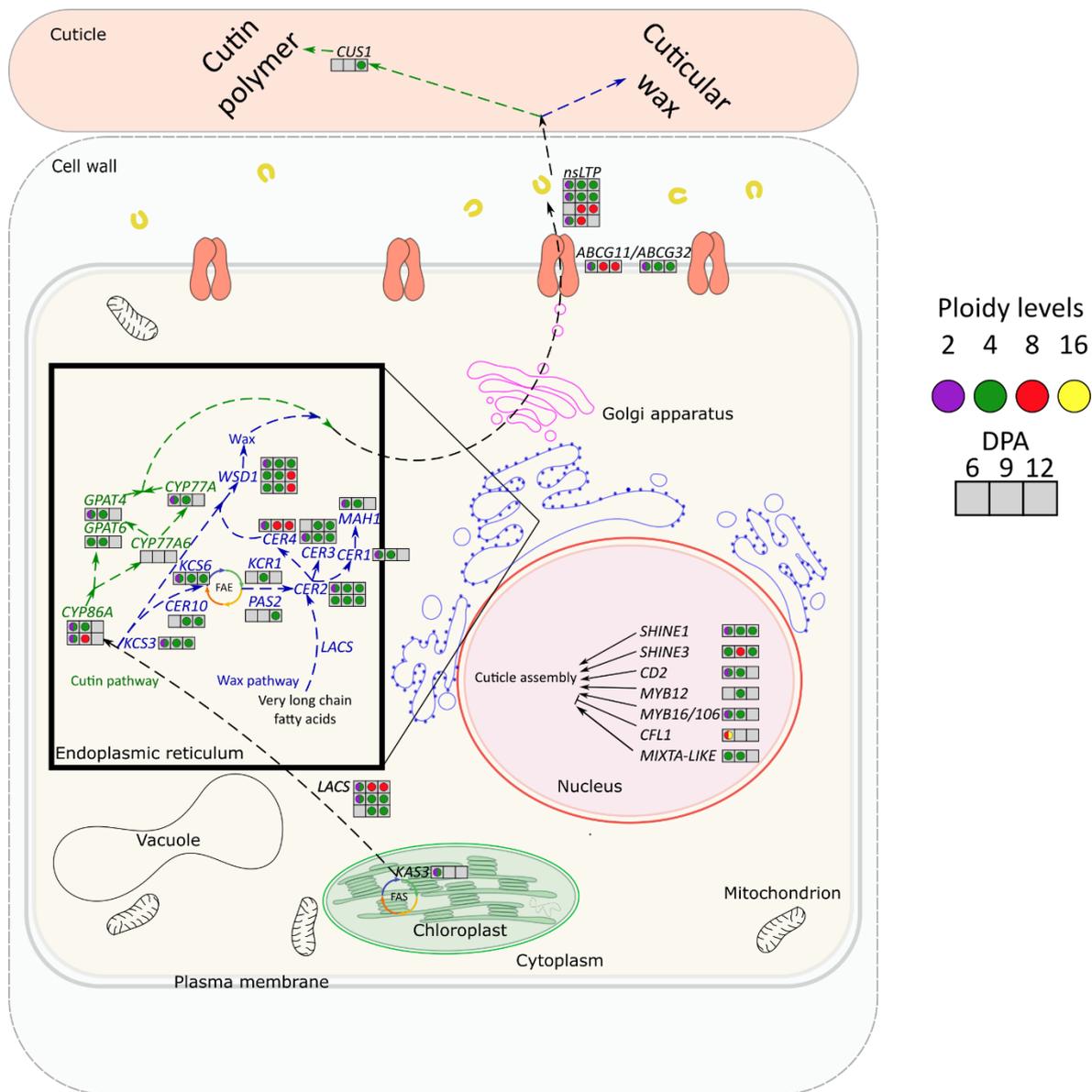


Figure 32. Representation of cuticle synthesis genes and their expression in relation to ploidy levels in the pericarp

(A) Cutin and wax synthesis genes pathways are represented with only transcriptional regulators found differentially expressed in the clusters.

Expression peaking at a specific ploidy level is symbolized by a colored circle (or two-colored half circles for expression peaking in two ploidy levels) in a box corresponding to a specific stage (6DPA, 9DPA and 12DPA). For example, cluster 6-2 would be represented by a purple/green circle for expression peak in 2C/4C levels whereas clusters 9-1, 9-2, 12-1 and 12-2 are represented by a green circle for an expression peak in 4C and clusters 9-3 and 12-3 are represented by a red circle for 8C peak of expression.

Abbreviations: KCS = KETOACYL-CoA SYNTHASE, CYP = CYTOCHROME P450, nsLTP = nonspecific LIPID TRANSPORT PROTEIN, CD2 = CUTIN DEFICIENT 2, MAH = MIDCHAIN ALKANE HYDROXYLASE, KCR = β -ketoacyl reductase, GPAT = GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE, ABCG = ATP Binding Cassette G, CER = ECERIFERUM, WSD + WAX ESTER SYNTHASE, LACS = LONG-CHAIN ACYL-CoA SYNTHETASE, CFL1 = CURLY FLAG LEAF 1, PAS = PASTICCINO, CUS = CUTIN SYNTHASE

TAGL1 and MBP3 are both necessary for normal development of Tomato fruit, starch metabolism and cell wall synthesis (Zhao *et al.*, 2018, Zhang *et al.*, 2018), and OVATE and SUN are necessary for the determination of the round fruit shape (Rodriguez *et al.*, 2011, Wu *et al.*, 2018). Moreover, the Tomato ortholog of Arabidopsis *PERIANTHA* was found in the overlap and in Arabidopsis *PERIANTHA* is a direct activator of *AGAMOUS* expression (Maier *et al.*, 2009). OVATE, SUN and *PERIANTHA* indicate that 4C and 8C cells express to signal the regulation of fruit shape and organ identity (Monniaux and Vandebussche 2018).

The multiple processes found in 8C cells could be explained by the fact that several cell types are composing the 8C cell population or that 8C cells are having roles of interface between non-endoreduplicated cells and highly endoreduplicated cells.

For clusters of genes highly expressed in endoreduplicated nuclei (6-4, 6-5, 9-4, 9-5 and 12-4 12-5), few GO-enrichments were detected. Similar enrichment was found between 6-4x9-4 and 9-4x12-4 and corresponds to “*Sorbitol response*”. Two additional enriched GO terms were found between 6-5x9-4 and 6-5x12-4 and correspond to “*Carbohydrate metabolic process*” and “*response to reactive oxygen species*”, respectively. Among the 38 and 48 genes in the overlap between 6-5x9-4 and 6-5x12-4, we found representation (not enrichment) of terms associated to Actin filament, terpenoid biosynthesis (6-4x9-4,6-4x12-4 and 9-4x12-4). The term “*Terpenoid biosynthesis*” was represented by the *TERPENE SYNTHASE* (Beta-phellandrene synthase, PHS1) and the *PHYSTOENE SYNTHASE* (*PSY1*) involved in carotenoids pathway precursor of the synthesis of several hormones, Brassinosteroids, ABA, or Gibberellins (Gutensohn *et al.*, 2014, Fraser *et al.*, 2007).

The comparison of the clusters with high expression in high ploidy levels (6-5,9-5 and 12-5) contained terms related to “*ion transports*”, “*cytoskeletal organization*”, “*energy reserve metabolism*” and “*cell wall modifications*” (only 9-5x12-5). The term “*cytoskeletal organization*” contains the *ACTIN RELATED PROTEIN3* (*ARP3*) gene, important for normal cell expansion in highly endoreduplicated trichomes (Yanagisawa *et al.*, 2015), indicating a potential role in cell expansion in endoreduplicated pericarp cells. The term “*energy reserve metabolism*” contains an encoding gene for a starch degrading enzyme *ALPHA-GLUCAN PHOSPHORYLASE 2* (Emes *et al.*, 2018) and cell wall modification genes (*PECTINESTERASE*, Phan *et al.*, 2007).

Finally, most of the genes involved in cuticle and cell division in low ploidy levels are shared during early fruit development. However central metabolism shifted from sucrose degradation to glycolysis when passing from 6DPA to 12DPA and secondary metabolism is enriched in 9DPA and 12DPA fruits.

E) Comparing ploidy specific transcriptome with laser microdissected pericarp transcriptome and ploidy transcriptome of Arabidopsis root.

After nuclei sorting, we lost the spatial information of cells within the pericarp tissue. We have seen that specific biological processes are expressed in a ploidy-specific manner and associated to low or high ploidy levels. Thus, to retrieve spatial

information and try to localize the cells within the pericarp based on the ploidy specific gene expression, we compared our transcriptome data to genome wide pericarp cell type/tissue specific gene expression obtained by Laser Microdissection (LM) or hand-microdissection of five pericarp cell types/tissues from the M82 Tomato cultivar (Shinozaki *et al.*, 2018). To identify specific expression patterns among the Tomato tissues and the pericarp cell/tissue types, the authors clustered the expressed genes in 43 modules. For the comparison, we used the 1977 so-called “Hub genes” from the LM dataset that correspond to the genes having the strongest correlation with their modules (Shinozaki *et al.*, 2018). We used expression data from 5 and 10DPA samples that approximately correspond to the developmental stages studied here, 6-9 and 12DPA.

We found a total of 337 unique genes in common between the ploidy transcriptome (3330 genes) and the selected “Hub genes” (2645 genes) after the comparison. More specifically, we calculated the significance of the overlap between ploidy specific clusters and modules from the LM transcriptome dataset and found overlap enrichment for 13 of the ploidy clusters and 9 of the LM modules (Figure 33A). Five of the nine modules correspond to genes related to only one zone of the pericarp: the vascular tissue for M14, M17 and M19, and the outer epidermis for M20 and M21. DPA. The rest of the modules contain genes expressed in two zones of the pericarp: genes from M2 are mainly expressed in the outer epidermis and the vascular tissues; genes from M7 and M8 are expressed in the collenchyma and the parenchyma, corresponding to the upper cell layers and inner cell layers surrounding vascular tissue of mesocarp, respectively. Genes from M26 are lowly expressed in the vascular tissues and the most in collenchyma. We used this spatial information and the ploidy specific expression to infer/propose a ploidy identity to the different zones of the pericarp tissue and thus to locate the nuclei with specific ploidy level (Figure 33B).

Clusters with genes expressed in low ploidy levels nuclei (6-1,6-2,6-3,9-1,9-2,12-1,12-2) showed significant overlap with modules containing genes mainly expressed in the outer epidermis (M2, M21), in the vascular tissues (M14, M17, M19) and the inner epidermis (M20, Figure 33A). High ploidy level clusters (6-4,6-5,9-4,9-5,12-4,12-5) were significantly overlapping with gene modules mainly expressed in collenchyma and parenchyma (M7 and M8) and cluster 12-5 was also overlapping with a module expressed in every tissue except vascular tissues (M26, Figure 33A). No overlap was significant for the comparisons with clusters 9-3 or 12-3.

This comparison between ploidy-specific transcriptome and LM transcriptome is indicating that cells with lower ploidy levels (2C and 4C at 6DPA and 4C at 9-12DPA) are found in the epidermis (outer and inner, Figure 33B) and the vasculature. This is in accordance with the GO enrichment analysis highlighting epidermal related processes (cuticle synthesis and cell division) and vasculature related processes (transport of sugar from vascular cells). We found enrichment between high ploidy clusters and collenchyma and parenchyma related clusters for 6 and 9DPA, but, at 12DPA we could also associate some high ploidy related gene expression with an expression in all the pericarp except the vascular tissue. This finding might indicate that endoreduplication is not restricted to collenchyma and parenchyma after 12DPA.

A

		6-1	6-2	6-3	6-4	6-5	9-1	9-2	9-3	9-4	9-5	12-1	12-2	12-3	12-4	12-5		
		553	542	242	227	358	618	460	241	205	376	606	449	128	246	432		
M2	156	24	10	31			19	11	2	4	1	22	5	1	1	1		
M7	41			1	2	8				2	6					1	8	
M8	66				9	5				4	5					6	8	
M14	29	2	4	1			3	3				6	3					
M17	36	5	8				6	2	1			12	6					
M19	42	4	2	2			11	1	2			9	8					
M20	49	1	23	9			28	13				39						
M21	17	2	4	1			6	4				3	5					
M26	36				1	1				1	2						5	



B

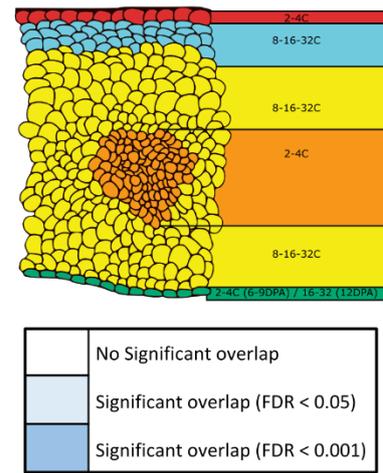


Figure 33. Significant overlaps between ploidy specific transcriptome and LM pericarp transcriptome of Shirozaki *et al.*, 2018.

(A) Overlap between genes from in the different clusters at 6DPA, 9DPA and 12DPA and Laser Microdissection Pericarp (LM) transcriptome. The number of genes in each overlap is shown for each comparison between two clusters. Empty boxes correspond to comparisons with no overlap.

Significant overlap for $P < 0.05$ is highlighted by light blue and for $P < 0.001$ by dark blue.

(B) Ploidy map inferred from the comparison between the ploidy-specific transcriptome and the laser microdissected pericarp transcriptome. Colors are indicating LM parts of the pericarps corresponding to outer-epidermis in red, collenchyma in blue, parenchyma in yellow, vascular bundles in orange and inner-epidermis in green. Adapted and modified from Shirozaki *et al.*, 2018.

Shared ploidy marker genes across different tissues of *Arabidopsis thaliana* and *Solanum lycopersicum* could be universal ploidy markers.

To search for potential universal ploidy markers, we compared our ploidy specific gene expression data from tomato pericarp to the ploidy specific gene expression data from the *Arabidopsis* root cortex (Bhosale *et al.*, 2018). This later transcriptome data was obtained from sorted nuclei based on their ploidy level ranging from 2C to 16C and the authors found 3737 genes that were differentially expressed between the studied ploidy levels.

For the comparison, we used a set of 954 *Arabidopsis* genes associated to a specific ploidy level: 407 genes have a peak of expression at 2C, 162 at 4C, 302 at 8C, and 83 at 16C (Bhosale *et al.*, 2018). To compare the datasets, we identified Tomato orthologous genes of *Arabidopsis* markers by using integrative orthology method provided by Plaza4.5 (Van Bel *et al.*, 2018). 1329 orthologous Tomato genes were found to correspond to the 954 input *Arabidopsis* ploidy markers. We overlapped these 1329 orthologous genes with the genes from the tomato ploidy clusters and found 237 of the *Arabidopsis* marker orthologs have a peak of expression at one ploidy level.

Even if not all *Arabidopsis* ploidy markers showed coherent expression profiles in Tomato, for each Tomato cluster, we found markers that match the ploidy expression pattern. For example, in cluster 6-1, in which genes are peaking in 2C nuclei, most of the common genes with the *Arabidopsis* dataset are 2C markers (35/50, Figure 34A). For cluster 6-5, 10 over 17 genes were found to be 8C or 16C markers, the rest of the genes corresponding to 2C (6/17) and 4C (1/17) markers.

To test the universality of a ploidy marker in the two species, we focused on *Arabidopsis* ploidy markers expressed in tomato clusters with a matching expression pattern, for example 4C markers present in clusters 9-1, 9-2 or 12-1, 6-1 and searched for their presence in in a least two developmental stages. Among the clusters with high expression in low ploidy levels (6-1,6-2,6-3,9-1,9-2,12-1 and 12-2), we first studied cluster 6-1 since expression in this cluster is high in 2C nuclei, not studied at 9 and 12DPA. In this cluster, we found 35 genes that are 2C marker genes and correspond to DNA replication related genes. In clusters 9-3 and 12-3, with expression peaking in 8C, only two *Arabidopsis* 8C marker genes were found one coding for a gene of unknown function and the other GPAT1 that is involved in glycerolipid metabolism in *Arabidopsis* flowers, siliques, and seeds.

In clusters with high expression in high ploidy levels nuclei (6-4, 6-5, 9-4, 9-5, 12-4 and 12-5) we searched for 8C markers in clusters 6-4, 6-5, 9-4 and 12-4 and 16C markers in clusters 9-4, 9-5, 12-4 and 12-5DPA. We found in these clusters 10 common markers for 8C (4 genes) and 16C (6 genes). The genes are associated to cell wall synthesis or modification (4 genes), ion transport (2 genes), transport of molecules (a kinesin of unknown function), carotenoid biosynthesis (PSY1) and transcription regulation with a bHLH of unknown function in Tomato. In conclusion, by comparing our dataset to the ploidy specific transcriptome from *Arabidopsis thaliana* root cortex we have been able to identify genes that could be ploidy markers from 2C to 16C in different species, Tomato and *Arabidopsis*, and tissues, fruit pericarp and root cortex.

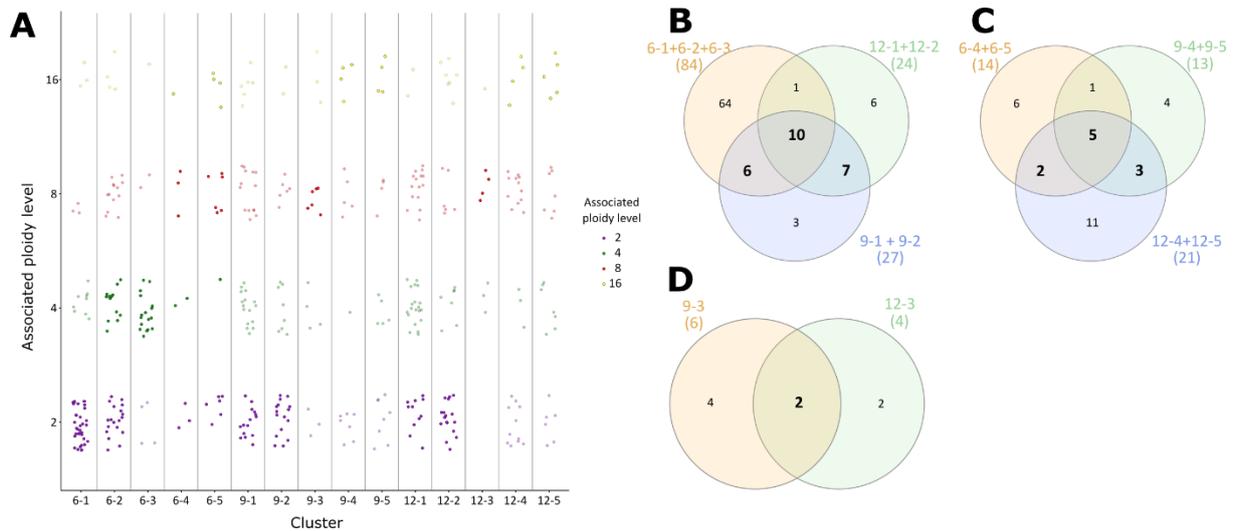


Figure 34. Ploidy markers from a ploidy specific transcriptome of *Arabidopsis thaliana* root cortex (Bhosale *et al.*, 2018) compared to the pericarp ploidy transcriptome.

(A) Dot plot representing each overlapping gene between the ploidy dependent transcriptome of *Arabidopsis* root cortex and the ploidy dependent transcriptome of Tomato pericarp. A stronger opacity of dots is associated to the ploidy level a gene is supposed to be marker of, according to Bhosale *et al.*, 2018. 2C markers are selected for 6-1 and 6-2. 4C markers are selected for 6-2,6-3,9-1,9-2,12-1 and 12-2. 8C markers are selected for 9-3 and 12-3 and 16C markers are selected for 9-4,9-5,12-4 and 12-5.

(B) Venn diagram representing the overlap between selected markers in clusters of each stages for low ploidy clusters (2C and 4C).

(C) Venn diagram representing the overlap between selected markers in clusters of each stages for high ploidy clusters (8C to 16C).

(D) Venn diagram representing the overlap between kept markers in clusters of 9DPA and 12DPA for 8C clusters.

F) *Quantification of ploidy on sorted nuclei by FisH: Assessing the efficiency of a 5s rDNA oligo probe.*

To complete and precise the spatial information on ploidy distribution predicted from the ploidy-specific transcriptome data (Figure 33), we aimed at determining the dynamic of ploidy changes *in situ* at 6DPA, 9DPA and 12DPA. To quantify ploidy *in situ*, we decided to use Oligo-FisH which is based on the use of synthetic oligonucleotides coupled to fluorescent dyes already used successfully to image different DNA loci in plant (Simon *et al.*, 2018; Easterling *et al.*, 2018). Oligo-FisH is particularly powerful when used for targeting repeated sequences. Therefore, we chose to target the 5s rDNA locus of Tomato because this sequence is organized in tandem repeats and localized on a peri-centromeric locus of only one chromosome, chromosome 1 (Chang *et al.*, 2008; Lapitan *et al.*, 1991). To test if the 5s rDNA oligo-FisH probe was suitable for ploidy quantification, we determined ploidy on nuclei with known ploidy level. To do so, Fluorescence Activated Cell Sorting (FACS) was used to sort nuclei based on their ploidy levels. Nuclei ranging from 2C to 128C were collected on microscope slides and hybridized with the 5s rDNA oligo FisH probe.

By using this probe, we observed clear dot signals whose number increased with ploidy levels (Figure 35A). The number of dots was then counted for the different sorted nuclei. First, we found that the number of dots was significantly different between two consecutive ploidy levels (Figure 35B). Second, the number of dots counted was always close to the expected ploidy level studied, for example 3.36 ± 0.49 dots and 26 ± 4.21 for 4C and 32C nuclei, respectively. Third, the number of counted dots was more variable in higher ploidy levels but still significantly different between ploidy levels (Figure 35B). These results thus show that the use of the 5s rDNA oligo-FisH probe allows correct quantification of ploidy *in situ* since the number of dots allows classifying nuclei into specific ploidy level classes.

We also observed that nuclear volume was increasing with ploidy levels allowing to distinguish each ploidy class except for the 16C and 32C nuclei that did not differ in sizes (Figure 35C). Nuclear volume was small and relatively similar for low ploidy levels, 2C and 4C, 62 and $103 \mu\text{m}^3$, respectively, and then strongly increased in 8C, around 6x. For higher ploidy levels, the change in size between ploidy level was around 1.46x between 16C and 8C, 2.05x between 32C and 16C, 3x between 64C and 32C and 2x between 128C and 64C. The coefficient of determination R^2 between ploidy level and nuclear volume was 0.87 indicating a good correlation between these two traits.

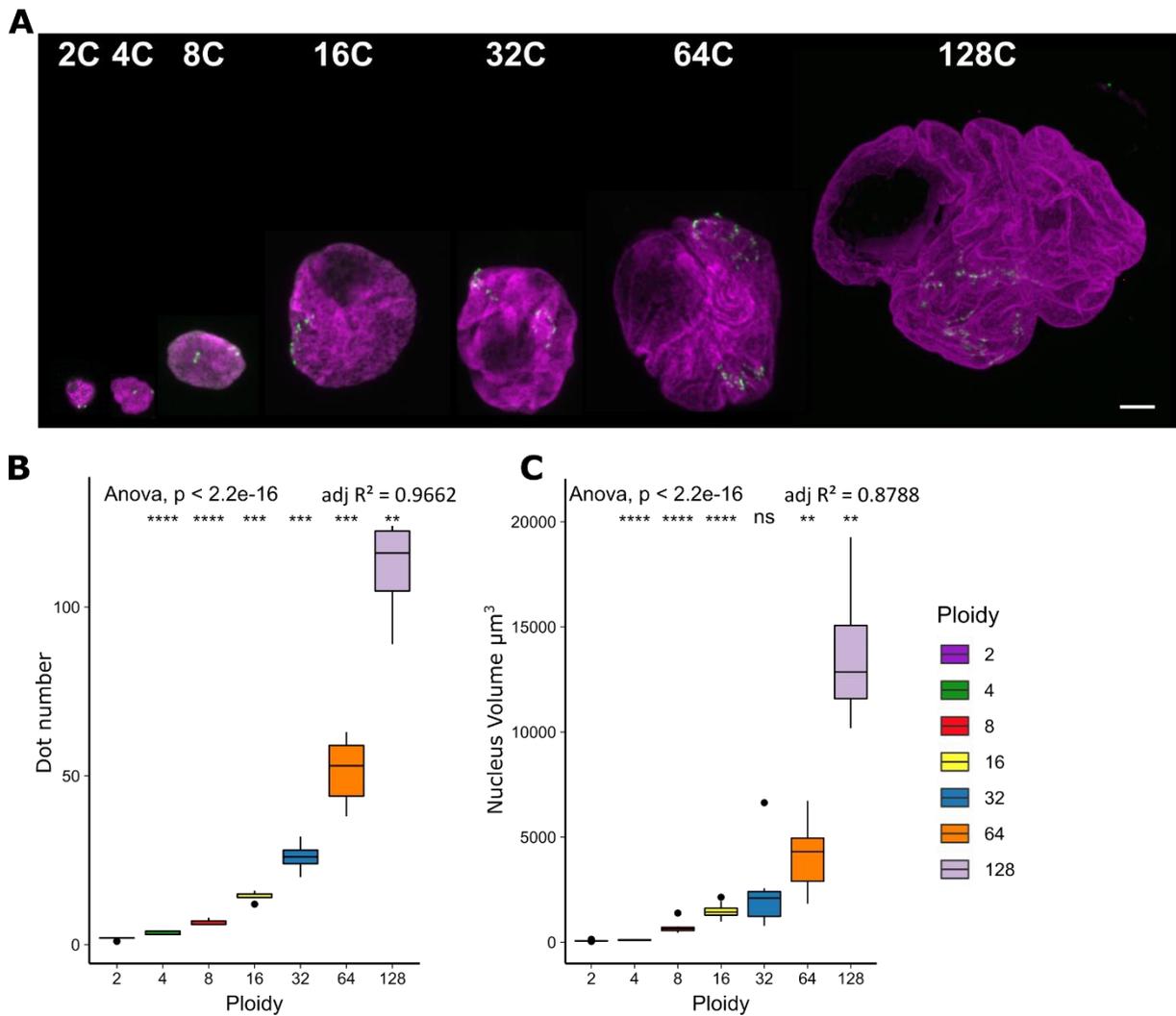


Figure 35. Morphology of Tomato pericarp sorted nuclei based on their ploidy and evaluation of the 5SrDNA oligo FisH probe for ploidy quantification.

(A) Representative illustration of nuclear morphology of different ploidy levels. The green dots correspond to the hybridization signal of the 5sr DNA oligo FisH probe. Scale bar = $5\mu\text{m}$.

(B) Boxplot representing the counted dots in nuclei sorted based on their ploidy level.

(C) Boxplot representing nuclear volumes as a function of ploidy level

(B) and (C) A Wilcoxon test was used to compare foci counts and nuclear size in relation to ploidy; number of nuclei studied per ploidy level = (14,22,13,9,9,13,4).

The associated significance codes are: 0 '*****' 0.001 '***' 0.01 '**' 0.05 '*'

Although nuclear volume increased from 2C to 16C, nuclear morphology stayed quite similar with a round shape. However, we noticed, as previously described (Bourdon *et al.*, 2012) a major difference in the 32C nuclei shape compared to 16C nuclei: nuclear shape changed from round in 16C to invaginated nuclei in 32C (Figure 36A and 36B, Bourdon *et al.*, 2012). Invagination were also visible in some 16C nuclei but less often. Compared to 32C nuclei, invaginations were more complex in 64C nuclei and 128C (Figure 36A and 36B).

In conclusion, Oligo-FISH using a synthetic probe targeting the 5s rDNA locus allowed ploidy quantification on sorted nuclei, thus the method could be transferred to analyze ploidy distribution in the pericarp of young developing tomato.

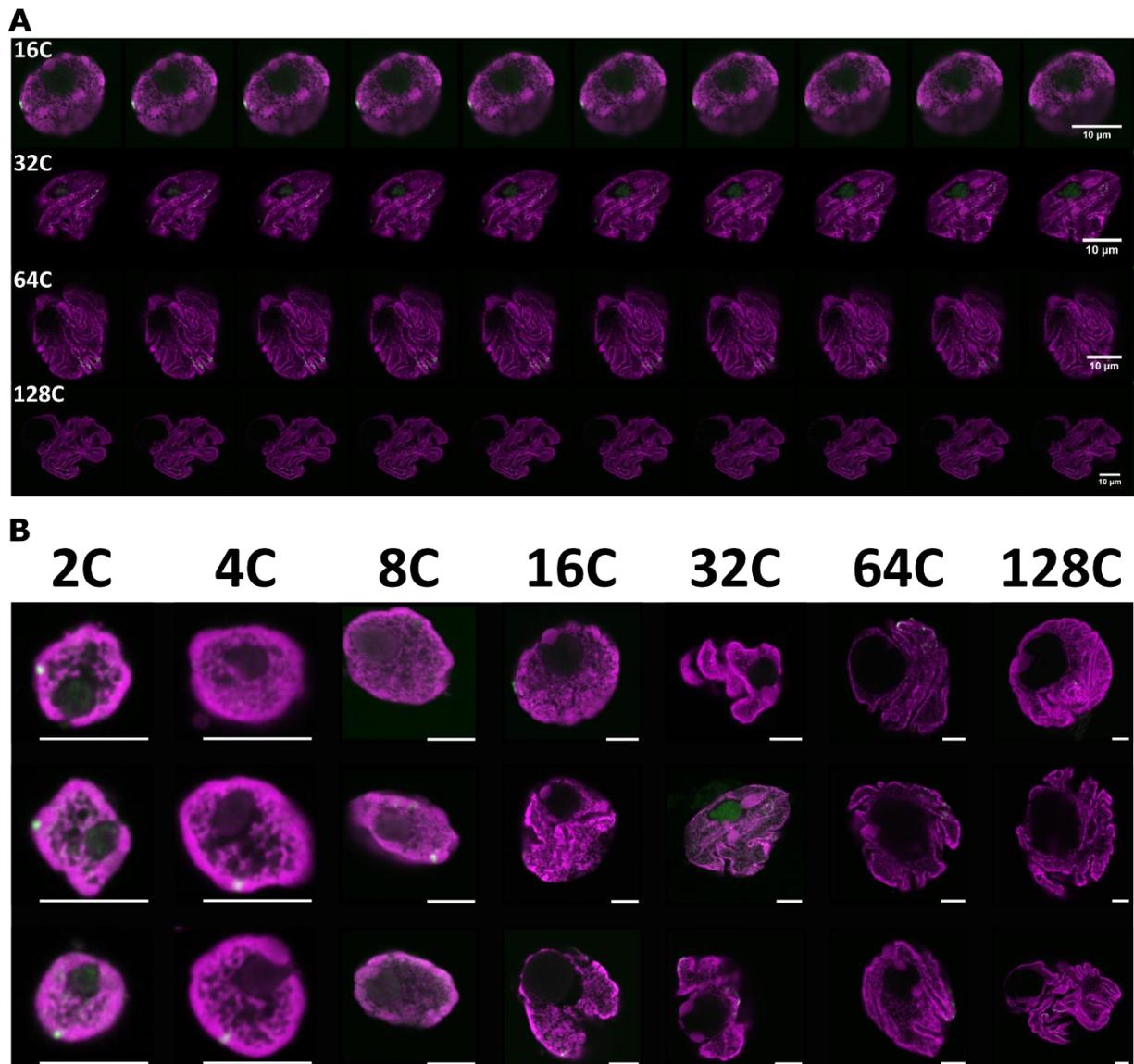


Figure 36. Nuclear shape is modified with increasing ploidy levels.

(A) Successive images following Z axis representative of nuclear morphology at different ploidy levels. Z step = $0.14\mu\text{m}$. Scale bar = $10\mu\text{m}$

(B) Illustration of nuclear morphology of three nuclei per ploidy levels illustrating shape complexification with increasing ploidy levels. Scale bar = $5\mu\text{m}$

G) *Endoreduplication occurs early during Tomato pericarp development and ploidy levels rapidly increase in the central layers of the pericarp.*

To understand the dynamic of endoreduplication in young Tomato pericarp, we quantified *in situ* ploidy levels and generated ploidy maps of 6, 9 and 12DPA fruit pericarps. We also measured nuclei volumes and cell area in these samples.

To generate ploidy maps, we needed to acquire nuclei on the whole pericarp image and locate them into the tissue. By measuring the distance between the cells and the border of the pericarp, we were able to attribute ploidy levels of the cells to cell layers (see *Chapter 2.1.2.B*) *c.: Semi automated imaging*).

These cell layers were defined using the nomenclature from Renaudin *et al*, 2017. Cell layers were distributed in 7 classes based on pericarp structure at anthesis. At this stage, the pericarp is composed of nine layers including, from outside to inside the fruit, the outer epidermal layer E1, two-sub outer epidermal layers, E2 and E3, four rows of mesocarp cells grouped under the name of M layer and lastly two inner layers, the sub inner epidermal layer I2 and the inner epidermal cell layer I1. The latest class corresponds to the new layers formed after anthesis by periclinal cell division and forming M-like cell layers named M'. In our experiment, we found between 14 and 15 layers at 6, 9 and 12DPA, thus the M' layer contained five to six mesocarp layers localized below the E3 layer.

At 6DPA, ploidy levels ranged from 2C to 32C (figure 37A). 2C and 4C cells were found in the epidermal layers, E1, E2, I2 and I1. In E1 layer, there were 50% of 2C and 4C nuclei, while in the E2 layer there were 33% of 2C and 48% of 4C nuclei (Figure 37A). 8C cells were found in nearly all layers except E1 and represented nearly half of the pericarp cells at 6DPA. The proportion of 8C nuclei was nearly 50% in E3, M', I2 and I1, 34% in M layers and 19% in E2 (Figure 37A). 16C cells were absent from the outer and inner epidermal layers and were found from E3 layer to M layers and represented 36% of E3, 44% in M' cells and most of the M cells with >50%. In the M region, the most internal layers of the pericarp, layers 8, 9 and 10, had the highest proportion of 16C with 62%, 70%, 54%, respectively. Finally, 32 C cells were also found in 6DPA pericarp fruits, but only in very low proportions <10% in M' and most of the M layers. Only layer 11 contained more than 10% of 32C, with 17% (Figure 37A).

In the 9DPA fruit pericarps, ploidy levels were also ranging from 2C to 32C (Figure 37B). The low ploidy levels, 2C and 4C were found in the outer and inner epidermal layers. The proportion of 2C was of 47% in E1, ~26% in E2 and E3 and ~20% in I2 and I1 while the proportion of 4C was of 55% in E1, E2 and E3 and 60% in I2 and I1 (Figure 37B). The decrease of 2C cells compared to 6DPA was accompanied by an increase of 8C proportion in the outer and inner epidermal layers with 15% in E2 and E3 and 20% in I2 and I1. In the mesocarp layers, the proportion of 8C decreased compared to 6DPA with 28% in M' and ~24% in the layers 10, 11 and 12 (Figure 37B).

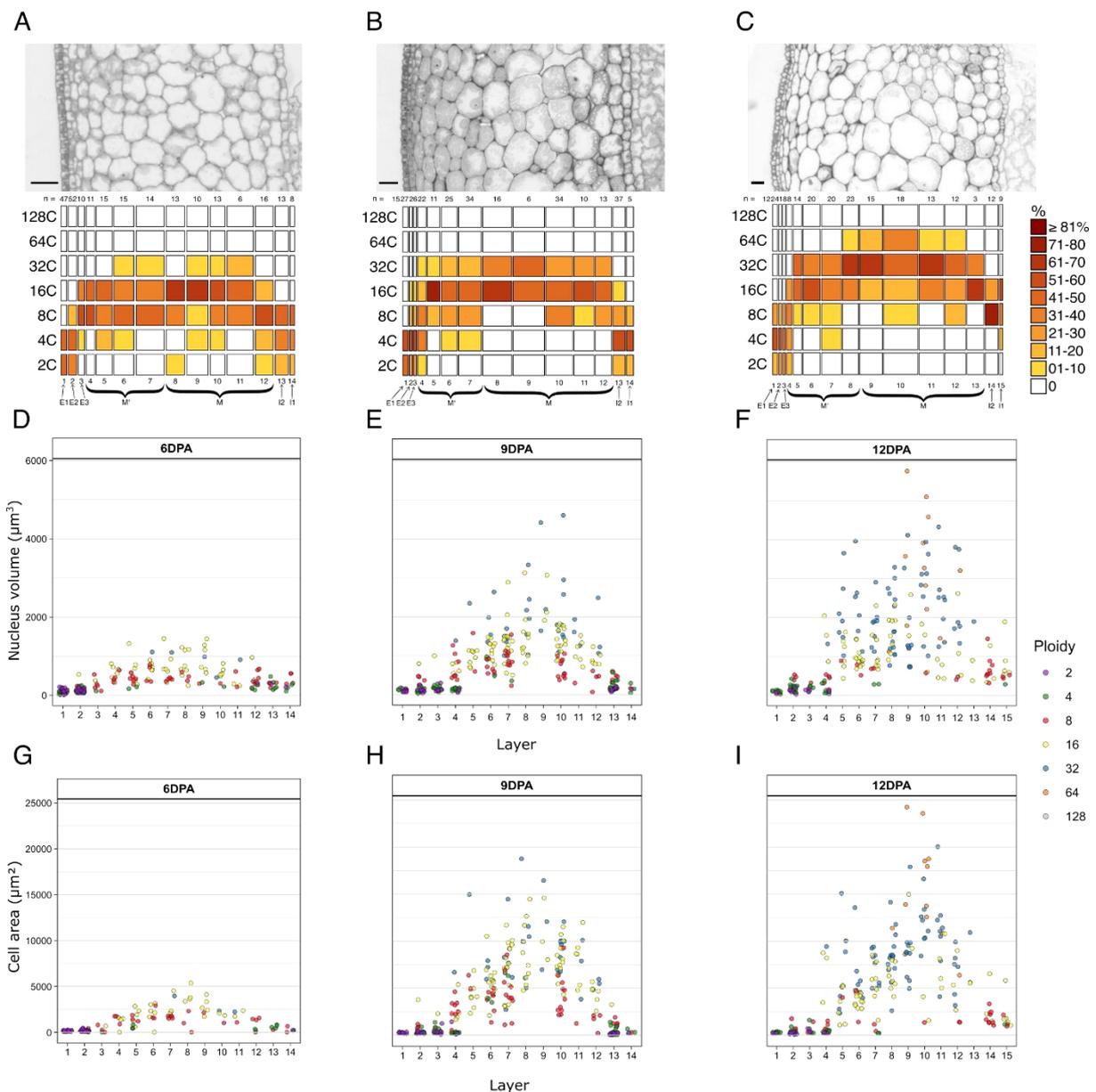


Figure 37. Ploidy distribution, nuclear volumes, and cell sizes in 6, 9 and 12DPA Tomato pericarps.

(A) to (C) Ploidy maps associated with a representative pericarp section of 6DPA (A), 9DPA (B) and 12DPA (C) fruit. The proportion of each ploidy level per layer is associated with a color scale. n is the number of observed nuclei per layer. E1, E2 and E3 correspond to the outer epidermis (layers 1 to 3). M and M' layer correspond to distinct mesocarp layers (layers 4 to 12 or 13). I2 and I1 correspond to the inner epidermis (layers 13 and 14 or 14 and 15).

(D) to (F) Nuclear volume of each observed nucleus per layer colored by its associated ploidy level (color scale on the right) at 6 (D), 9 (E) and 12DPA (F).

(G) to (I) Cell area of each observed nucleus per layer colored by its associated ploidy level (color scale on the right) at 6 (G), 9 (H) and 12DPA (I).

The absence of 8C cells in layers 8 and 9 was accompanied by an increase in 16C and 32C nuclei. 16C represented almost the majority of M' layers cells with a percentage of 47 and in M layers with 51% of the cells. Finally, 32C cells were found in all mesocarp layers with proportion of 12% in M' and 33% in M layers with a maximum of 50% in the layer 9 (Figure 37B).

In the latest developmental stage studied, 12DPA, ploidy levels in the pericarp increased compared to the previous time points and ranged from 2C to 64C (Figure 37C). The outer epidermal cell layers, as in 9DPA fruit, were composed of 2C and 4C in majority with ~20% of 2C and 55% of 4C cells. 4C cells were also found in I1 with a proportion of 11% and 2C cells were absent from the I2 and I1 layers (Figure 37C). Finally, cells in the inner epidermal layers started to endoreduplicate since 8C and 16C nuclei represented nearly all cells with a percentage of ~90-100%. The ploidy levels observed in M' and M layers were differentially distributed compared to 6- or 9DPA pericarps. 8C cells were more present in the outer epidermal layers (21%), in I2 layer with 75% of the cells and in I1 with 33% of the cells. In the other layers, M' and M, the proportion of 8C cells was strongly decreased compared to 6 and 9DPA with only ~10% of the cells (Figure 37C). In the M' layers, ~40% of the cells were 16C and the other half were 32C (~37%) with the exception of the inner M' layer (layer 8) that had more 32C (61%) and even only one 64C nucleus was found (Figure 37C). In the M layers, cells were mostly 32C (~57%) and proportions of 32C and 64C together could even reach 70. 64C cells represented < 10% of the cells in the M layers but the higher proportion was recorded for the inner M layer, layer 10, with 33% of 64C.

Often, an increase in ploidy levels has been associated with an increase in nuclear size and cell size (Cheniclet *et al.*, 2005, Kawade and Tsukaya 2017). To test if this correlation exists in our samples, we measured nuclei volumes and cell area.

If we compare same ploidy levels at different time-points, average nuclear volumes and cell areas were smaller in 6DPA fruit pericarps compared to 9DPA and 12DPA (Figure 37 D/E/F and G/H/I). 2C and 4C nuclear volumes and cell areas were similar between the different stages and were smaller than those of the other ploidy levels. 2C median nuclear volume were comprised between 120 and 130 μm^3 (Figure 38) and the median cell surface ranged between 142 and 252 μm^2 (Figure 37 G/H/I) whereas for 4C median nuclear volumes was between 115 and 163 μm^3 (Figure 38) and cell surface between 151 and 288 μm^2 (Figure 37 G/H/I). From 8C to 32C, nuclear volumes and cell surfaces of 6DPA pericarp cells were smaller than for 9 and 12DPA, but at 9DPA, sizes were similar to those of 12DPA cells. For example, 16C nuclei volumes at 6DPA were comprised between 485 and 812 μm^3 (Figure 37 D) whereas 9DPA and 12DPA nuclear volumes were comprised between 877 and 1443 μm^3 at 9DPA (Figure 37 E) and 748 and 1533 μm^3 at 12DPA (Figure 37 F). Similar trends were found for cells surfaces. Moreover, nuclear volume as well as cellular area was increasing in relation with ploidy levels. However, this increase distributed inside a gaussian-like distribution whose tails are overlapping between ploidy levels meaning that for example the population of 16C is smaller than the 32C population even if a given 16C nuclei can be larger than a 32C. Most of the increase in nuclear size and cell size was observed in the mesocarp where endoreduplication is taking place.

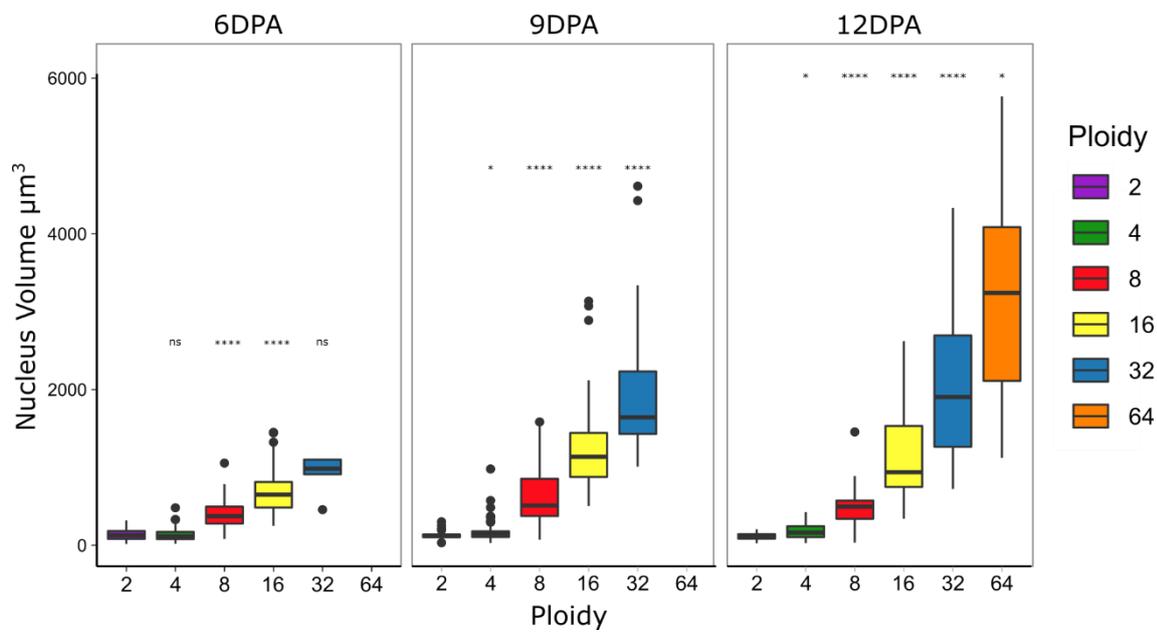


Figure 38. Dynamic of nuclear volumes during early fruit development

Boxplot representing nuclear volumes as a function of ploidy level.

A Wilcoxon test was used to compare nuclei volumes between consecutive ploidy levels.

Number of nuclei studied per ploidy level =

- 6DPA: 2C = 45, 4C = 65, 8C = 72, 16C = 56, 32C = 5;
- 9DPA: 2C = 29, 4C = 73, 8C = 60, 16C = 83, 32C = 36;
- 12DPA: 2C = 13, 4C = 34, 8C = 34, 16C = 58, 32C = 70, 64C = 12.

The associated significance codes are: 0 '****' 0.001 '***' 0.01 '**' 0.05 '*'

The construction of ploidy maps using oligo-FISH thus allowed studying the dynamic of endoreduplication establishment during early Tomato fruit development the construction of ploidy maps using oligo-FISH thus allowed studying the dynamic of endoreduplication establishment during early Tomato fruit development showing that ploidy is increasing importantly in the center of the pericarp.

H) *The integration of ploidy distribution with ploidy specific transcriptome reveals that ploidy specific gene expression is following specific spatio-temporal patterns.*

We combined the ploidy specific transcriptome data and the ploidy mapping of the pericarp to obtain spatial information on the ploidy specific gene expression in the pericarp *in situ*. To represent gene expression based on the ploidy distribution across the pericarp tissue, for each layer, we weighted gene expression values for each ploidy level by the proportion of cells of a specific ploidy level, and then we summed up the expressions in each layer and normalized this expression to all layers for a given cluster.

In clusters 6-1, 6-2 and 6-3, genes had high expression in 2C and 4C and would be mainly expressed in the outer epidermis (Figure 39A), in the layers 1,2 and 3. Genes from clusters 6-2 and 6-3 would also be expressed in the inner epidermis, layers 13 and 14, while genes from clusters 6-4 and 6-5 mainly would localize to the mesocarp (Figure 39A), layers 4 to 12.

Genes from clusters 9-1 and 9-2 can be found in the outer and inner epidermis (Figure 39B). In cluster 9-3, genes would be mainly expressed in the outer and inner epidermis but also some M' mesocarp layers, the 5th and 6th layers, and one M mesocarp layer, layer 10 (Figure 39B). Lastly for 9DPA, cluster 9-4 contains genes expressed in cells from M' and M layers (layers 4 to 12), while cluster 9-5 contains genes mainly expressed in M' layer 5 and M layers (Figure 39B).

At 12DPA, the clusters 12-1 and 12-2 contains genes that would be expressed in the outer epidermis and less in the mesocarp (Figure 39C). Genes from cluster 12-3, as for 9DPA the cluster 9-3, would localise in the outer and inner epidermal cells but also in some mesocarp cells (Figure 39C). For clusters 12-4 and 12-5, genes would be expressed in the mesocarp cells (Figure 39C and Figure 39A). At 12DPA, it is important to remind that the highest ploidy level assayed in the transcriptome was 32C while 64C are present in 5 of the 11 mesocarp layers and represent between 3% to 25% so for these layers interpretation has to be done with care.

Next, we studied in more detail the expression of selected genes having a role in pericarp development by inferring their spatial expression in the pericarp from the integration of our datasets. CYCLIN DEPENDENT KINASE proteins (CDK) are implicated in the regulation of several cellular events in the tomato pericarp such as regulation of cell number, cell size and indirectly cuticle development because of fewer cells (Czerednik *et al.*, 2012). Fruit specific overexpression of *SICDKB1* and *SICDKB2* resulted in increased expression of *SICDKA1* and fruit specific down regulation of *SICDKA1* resulted in the same phenotype: smaller fruits than WT, alteration of fruit shape and reduction of pericarp thickness due to less cell layers and smaller cells.

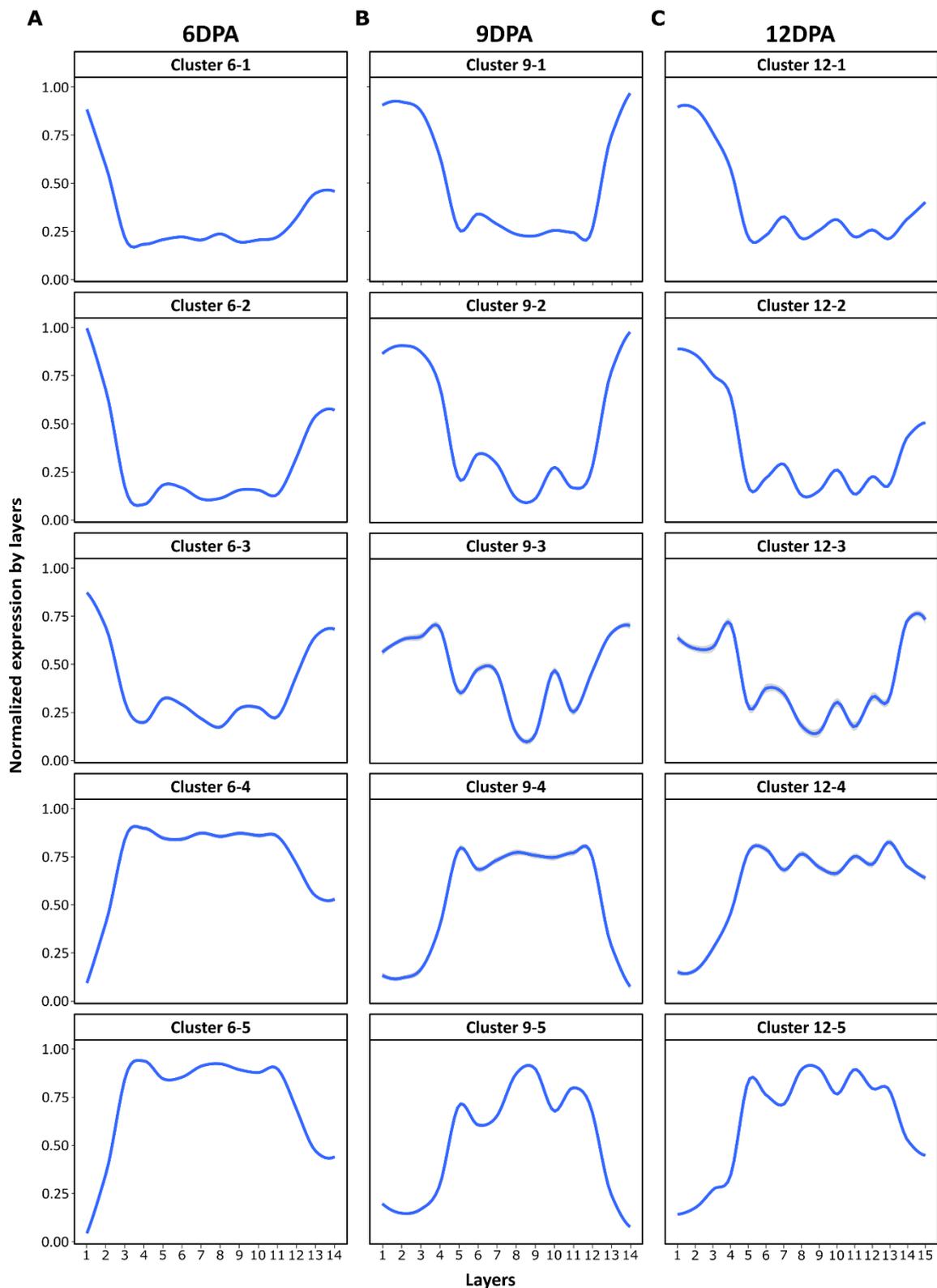


Figure 39. Data integration of ploidy transcriptome and FisH ploidy quantification reveals a spatiotemporal ploidy specific gene expression.

Spatiotemporal gene expression in relation with the ploidy levels in each layer for the five clusters at 6DPA (A), 9DPA (B) and 12DPA (C).

Spatiotemporal gene expression is obtained from ploidy dependent gene expression weighed by the proportion of each ploidy level in the cell layer for each cluster of genes and each stage.

This phenotype was also observed in plants overexpressing *SICDKB* indicating an antagonistic mechanism of regulation between *SICDKA* and *SICDKB* in the pericarp (Czerednik *et al.*, 2012). In the ploidy transcriptome *SICDKB1* was found in cluster 6-3 and *SICDKB2* in clusters 6-3 and 12-1 while *SICDKA* was not differentially expressed. The prediction of the spatial expression in the pericarp based on the ploidy distribution revealed that *SICDKB2* would be expressed in the outer and inner epidermis from 6DPA to 12DPA, *SICDKB1* in the outer and inner epidermis at 6DPA, in the mesocarp at 9DPA and again in epidermis at 12DPA while *SICDKA1* would be expressed in the mesocarp cells at 6DPA, 9DPA and in M' and inner epidermal layers at 12DPA (Figure 40). This analysis thus provides new insights on the potential expression localization of genes acting during pericarp development and shows expression profile that seem to be coherent with the antagonist regulation between *SICDKB* and *SICDKA*.

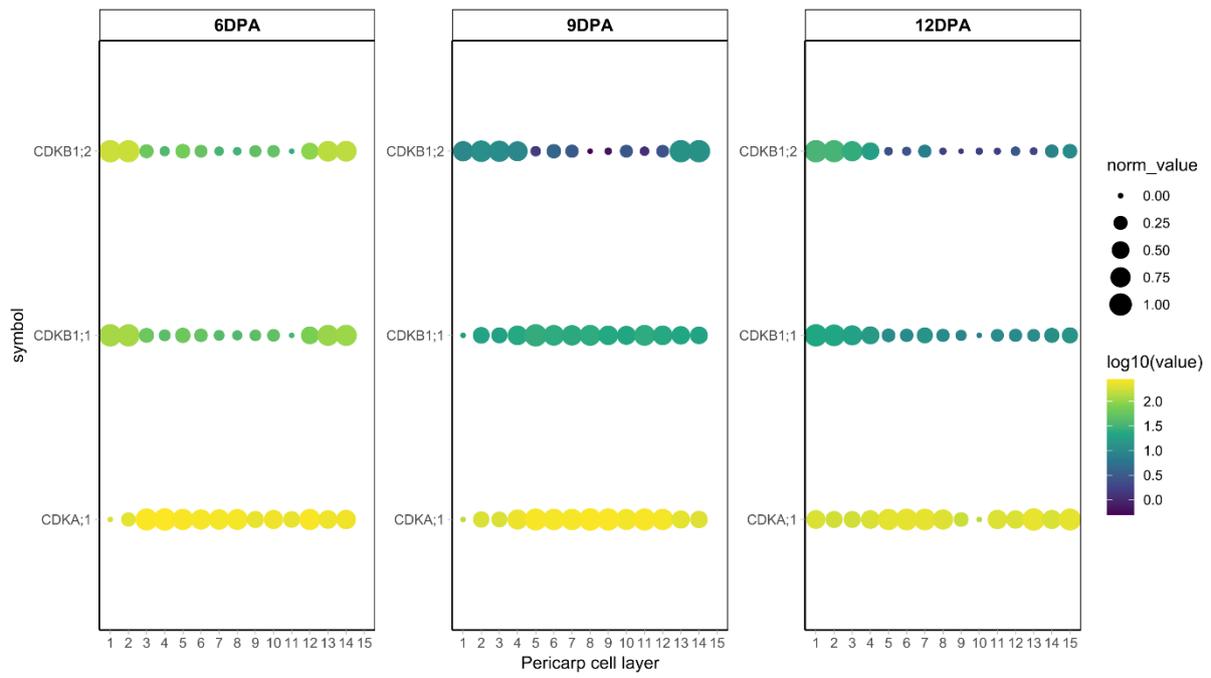


Figure 40. CDKA1 and CDKB1/2 expression in relation with ploidy levels in pericarp cell layers

We also predicted the spatial expression of Tomato orthologous of positive or negative regulators of the cell cycle or endoreduplication in Arabidopsis in function of the observed ploidy levels in the cell layers (Figure 41). Some of these regulators of the cell cycle or endoreduplication were found differentially expressed with ploidy levels such as SIM, HPY2 and SIKRP1 and were generally associated with low ploidy levels except for *CCS52A1* associated to high ploidy levels (Figure 41). Among the 23 genes analyzed, only *SICC52A1*, *SIKRP1* and *SIWEE1* have been shown to be able to regulate endoreduplication in Tomato fruit (Mathieu-Rivet *et al.*, 2010, Nafati *et al.*, 2010, Gonzalez *et al.*, 2007). I will here first describe the prediction of expression for these genes. *SICC52A1* was differentially expressed in cluster 6-5 while *KRP* was found DE in cluster 6-1. *CCS52A1* was thus predicted to be preferentially expressed in layers containing endoreduplicated cells at all developmental stages (Figure 41A). *KRP1* would be preferentially expressed in layers 1-2 and 13-14 at 6DPA, then in mesocarp layers at 9DPA and more evenly at 12DPA since most cell layers contain endoreduplicated cells (Figure 41A). Lastly, *WEE1* would be first expressed preferentially the outer and inner epidermal layers, then at 9DPA and 12DPA in the mesocarp (Figure 41A). These expression patterns of *CCS52A1* and *WEE1* (Mathieu-Rivet *et al.*, 2010, Gonzalez *et al.*, 2007) are consistent with their roles in promoting endoreduplication. For *KRP1*, the interpretation is more complicated because the overexpression of *KRP1* is inducing a reduction of ploidy levels in the pericarp (Nafati *et al.*, 2011) but this does not necessarily mean that the role of *KRP1* is to repress endoreduplication. Moreover, *KRP* proteins have been hypothesized to regulate the rate of cell cycles and endoreduplication cycles (Sizani *et al.*, 2019) and *KRP1* does not necessarily mean that protein is accumulated since KRPs are actively regulated at protein level (Li *et al.*, 2016). In addition, in general, the positive cell cycle regulators, such as *CDKB1*, *SIE2Fa* and *CYCD3;1* had a stronger expression in 6DPA fruits and would locate in the epidermal cell layers while their expression decreases later during fruit development. The transcripts of negative regulators of the cell cycle also had a stronger expression in 6DPA fruits that would locate in the epidermal layers, but they also stayed expressed at 9DPA and 12DPA at high levels and their expression would be more evenly distributed inside the pericarp. The positive regulators of endoreduplication were expressed at a similar level at the three developmental stages and their transcripts would mostly accumulate in the mesocarp layers while the expression of the negative regulators of endoreduplication would be more dispersed (Figure 41A). Moreover, some exceptions are found such as *ILP1* and *CYCD5;1*. *ILP1* normally positively regulate, in Arabidopsis, the expression of a mitotic cyclin (Yoshizumi *et al.*, 2006) and here is found expressed in endoreduplicated cells while *CYCD5;1* expression favorize increase of ploidy level in cells (Sterken *et al.*, 2011) but here *CYCD5;1* is expressed in dividing cells at 6DPA and evenly in the pericarp later.

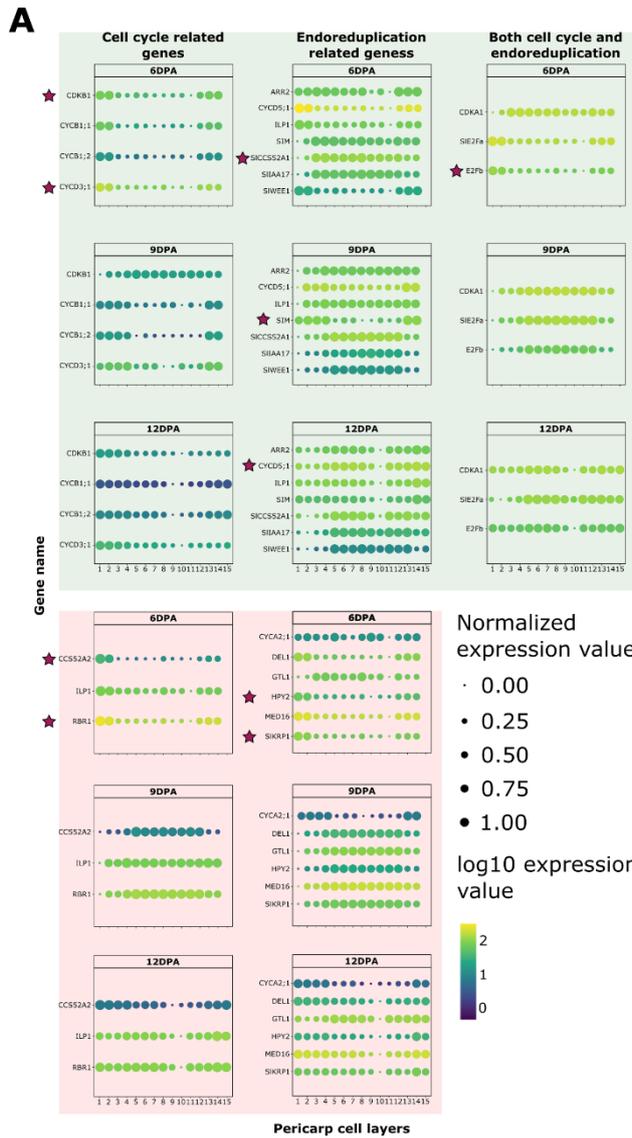


Figure 41B: Table of genes used for this figure with the associated references showing their function to regulate cell cycle, endoreduplication or both.

Solyc	symbol	cycle	effect	reference
Solyc10g074720	CDKB1	cell_cycle	+	Boudolf <i>et al.</i> , 2004
Solyc09g091280	RBR1	cell_cycle	-	Gutierrez 2009, Sabelli <i>et al.</i> , 2013
Solyc01g107730	CYCD3;1	cell_cycle	+	Schnittger <i>et al.</i> , 2002b
Solyc12g056490	CCS52A2	cell_cycle	-	Vanstraelen <i>et al.</i> , 2009, Willems <i>et al.</i> , 2020
Solyc10g078330	CYCB1;1	cell_cycle	+	Schnittger <i>et al.</i> , 2002a
Solyc06g073610	CYCB1;2	cell_cycle	+	Schnittger <i>et al.</i> , 2002a
Solyc05g009850	ILP1	endoreduplication	+	Yoshizumi <i>et al.</i> , 2006
Solyc08g066330	CDKA1	both_cell_cycle_and_endoreduplication	+	Czerednik <i>et al.</i> , 2015
Solyc11g068800	E2Fb	both_cell_cycle_and_endoreduplication	+	Abraham and del Pozo 2012
Solyc07g062780	HPY2	endoreduplication	-	ishida <i>et al.</i> , 2009
Solyc03g113760	DEL1/E2Fe	endoreduplication	-	Berckmans <i>et al.</i> , 2011, Heyman <i>et al.</i> , 2017
Solyc01g065540	ARR2	endoreduplication	+	Takahashi <i>et al.</i> , 2013
Solyc03g071710	SIM	endoreduplication	+	Walker <i>et al.</i> , 2000
Solyc08g080080	SICCS2A1	endoreduplication	+	Mathieu-Rivet <i>et al.</i> , 2010
Solyc03g025890	med16	endoreduplication	-	Liu <i>et al.</i> , 2019
Solyc09g074830	SIWE1	endoreduplication	+	Gonzalez <i>et al.</i> , 2007
Solyc09g091780	SIKRP1	endoreduplication	+	bisbis <i>et al.</i> , 2006
Solyc04g076040	cycD5;1	endoreduplication	+	Sterken <i>et al.</i> , 2012
Solyc03g120440	CYCA2;1	endoreduplication	-	Roudier <i>et al.</i> , 2003
Solyc11g005380	GTL1	endoreduplication	-	Breuer <i>et al.</i> , 2009
Solyc01g007760	SIE2Fa	both_cell_cycle_and_endoreduplication	+	De Veylder <i>et al.</i> , 2002
Solyc09g008810	LMI1	endoreduplication	+	Vuolo <i>et al.</i> , 2018
Solyc06g008590	SIAA17	endoreduplication	+	Su <i>et al.</i> , 2014

Figure 41. Cell cycle and endoreduplication related genes expression in relation with ploidy levels in pericarp cell layers

(A) Prediction of spatial expression of genes encoding positive (green area) or negative (red area) regulators of cell cycle, endoreduplication or both. Red stars indicate genes that are found DE.

(B) Table of genes used for this figure with the associated references showing their function to regulate cell cycle, endoreduplication or both.

3. Discussion

A) Endoreduplication in the pericarp induces differential gene expression.

From the ploidy dependent transcriptome, we found 1938, 1921 and 1878 genes differentially expressed at 6DPA, 9DPA and 12DPA when comparing gene expression between ploidy levels. The PCA analysis of gene expression in the samples showed that most of the variability was due to the change in ploidy levels suggesting that endoreduplication is probably an important regulatory component of pericarp development. Interestingly, the PCA analysis showed that 2C nuclei and 4C nuclei at 6DPA were very close to each other in the graph probably meaning that 4c nuclei are closer to be dividing nuclei than early endoreduplicated nuclei.

B) Endoreduplication during early fruit development is associated with cell specialization.

Our transcriptome data showed that cells with different ploidy levels express specific biological pathways suggesting that endoreduplication permits the establishment of specific developmental programs important for fruit growth. For example, in endoreduplicated nuclei, we found genes involved in sugar metabolism such as *PK*, *G6PDH* and *GAPDH* and these three genes encode enzymes responsible for the major metabolite fluxes in Tomato fruit during the cell proliferation phase thus favoring fruit biomass increase (Beauvoit *et al.*, 2014, Colombié *et al.*, 2015) suggesting an important role of endoreduplicated cells in the metabolic activities measured in the pericarp during early fruit development. We also found genes involved in cell wall synthesis in endoreduplicated nuclei such as the *UDP-GLUCOSE 4 EPIMERASE* encoding an enzyme converting UDP-glucose into UDP-galactose and the *GALACTOSYL GALACTOSYLXYLOSYLPROTEIN 3-BETA-GLUCURONYLTRANSFERASE* (*IRX9-L*) involved in hemicellulose synthesis (Rösti *et al.*, 2007, Meents *et al.*, 2018). The potential relation between endoreduplication and cell wall synthesis and modification has already been described in Arabidopsis root cells or in response to stresses (Bhosale *et al.*, 2018, Scholes and Paige 2018). Interestingly, in clusters 9-4 and 9-5, we found a *MUTS* homolog and Tomato orthologous of *BARD1* and *BRCA1*, genes generally related to DNA repair. *SIBRCA1* and *SIBARD* contain a plant homeodomain, the PHD domain, not found in animals suggesting a potential new role in plants during development (Trapp *et al.*, 2011). *MUTS*, *BARD1* and *BRCA1* participate in DNA damage response (Yoshiyama *et al.*, 2013) but *BARD1* and *BRCA1* are also expressed in plant tissues without stress (Trapp *et al.*, 2011) and *BRCA1* is overexpressed in high ploidy mutant of Arabidopsis *e2fa* and *flb17* without DNA damage stress (Ramirez-Parra *et al.*, 2007, Noir *et al.*, 2015). These findings might be suggesting a potential role of these genes in developmentally controlled endoreduplication without DNA damage or in the control of DNA damage following replication of chromatids ensuring DNA integrity during endoreduplication. Thus, during fruit development, endoreduplication seems to be important for cell specialization.

C) *Localization of ploidy levels in the pericarp can be inferred using ploidy transcriptome data and publicly available transcriptome.*

The ploidy specific transcriptome is obtained from sorted nuclei that came from a population of pericarp cells and while tissue is disrupted and nuclei sorted, we lost the spatial information of cells. Thus, using data from literature and other transcriptome data we are able to determine a localization for ploidy sorted sequenced nuclei.

We found in our dataset genes involved in cuticle synthesis in 2C nuclei at 6DPA and 4C cells at all developmental stages. Since it is known that in general cuticle is formed in the first layers of the fruit (Dominguez *et al.*, 2008), these 2C and 4C cells might form the outer epidermis. We were interested in the spatial distribution of ploidy levels in the pericarp and we have seen that comparing our ploidy transcriptome with transcriptome data obtained after microdissection of the pericarp allowed to draw a preliminary map of ploidy level distribution in the pericarp. We found that a gradient distribution of ploidy levels toward the center of the pericarp is taking place already early during fruit development, in accordance with previously published ploidy map at 30DPA (Bourdon *et al.*, 2011). The obtained map was not precise nor absolute at the cell layer scale but was indicating that all ploidy levels are not found everywhere and that the pull of sorted nuclei at a specific ploidy level might belong to multiple tissues inside the pericarp. The mapping of ploidy in the pericarp at 6-, 9- and 12DPA using *FisH* allowed to more precisely attribute ploidy levels to the pericarp cell layers and confirmed the inferred map based on transcriptome data.

D) *Ploidy level distribution is dynamic during early fruit development.*

The ploidy maps obtained by *FisH* on young fruits reinforce the data obtained fully grown fruits (Bourdon *et al.*, 2011) that showed a gradual increase of ploidy levels towards the central cell layers. Overtime, we found that the ploidy levels increased in each cell layer but with a different dynamic (Figure 42). The outer epidermal layers were composed of equal proportion of 2C and 4C cells at 6DPA and at 12DPA, the layers contained 20% of 2C and 55% of 4C and almost a fifth of cells were 8C. From 6DPA to 12DPA, the decrease of 2C cells thus indicates that less cells are likely to divide, consistent with the nearly absent cell division activities recorded after 8DPA in the pericarp (Renaudin *et al.*, 2017). The inner epidermal layers were already engaged in endoreduplication at 6DPA and were composed of 50% of 2C+4C cells and 50% of 8C cells while at 12DPA the inner epidermal layers were totally endoreduplicated with few 4C cells in I2 layer and >90% of 8C and 16C. The mesocarp layers M, formed before anthesis, were also the more endoreduplicated layers with already some 32C at 6DPA that became the majority of cells later. Concerning the other mesocarp layers, M', formed after anthesis, the proportion of 16C was relatively stable between 6DPA and 12DPA while 8C were largely decreasing from and 32C increased. Thus, we found that ploidy levels are increasing globally but older mesocarp layers become more endoreduplicated than M' formed later. This evolution suggests that endoreduplication might be a continuous process occurring at the same speed in each cell layer with older cells becoming more endoreduplicated.

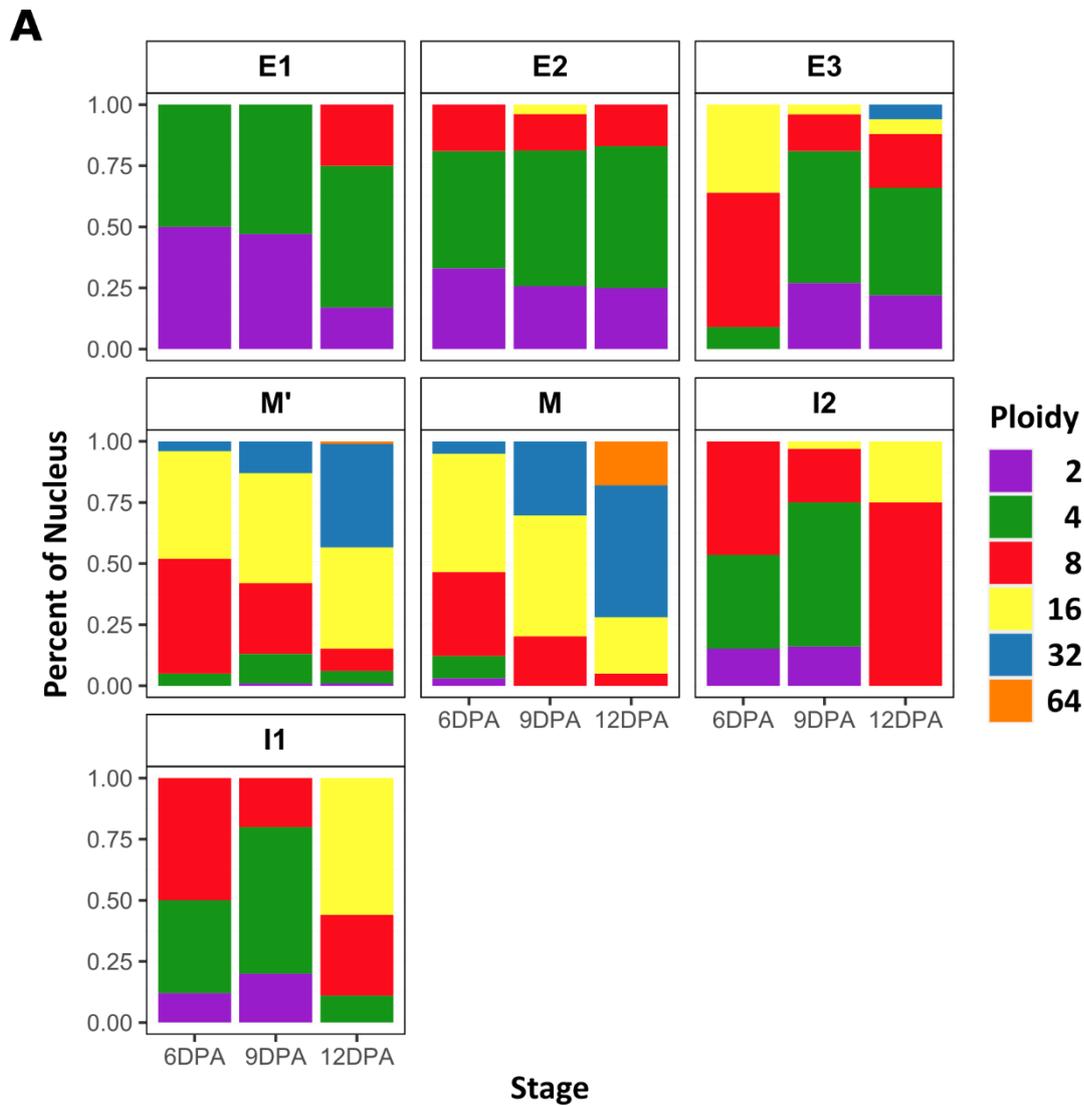


Figure 42. Dynamic changes of ploidy levels in the different pericarp cell layers during early fruit development from 6DPA to 12DPA

(A) Percentage of nuclei for each ploidy level in each cell layer for 6DPA, 9DPA and 12DPA pericarps. Color scale indicates ploidy levels.

This phenomenon could be similar to what happens in epidermal pavement cells of *Arabidopsis*, where endoreduplication is stochastic and older cells become more endoreduplicated (Kawade *et al.*, 2017). Finally, we also observed a heterogeneity of ploidy levels within the same cell layer suggesting that endoreduplication is an asynchronous process inside a layer.

A positive correlation has been shown between ploidy and cell size (Cheniclet *et al.*, 2005, Bourdon *et al.*, 2011). Here we confirmed this correlation at early stages of fruit development. But we also observed that the mean area of cells with similar ploidy levels is different at different time-points and increases over time. This was also the case between cells at 12 DPA and 30 DPA (Bourdon *et al.*, 2011), with, for example, 8600 μm^2 and 15000 μm^2 for 16C, 32C and 64C cells, respectively at 12DPA, compared to 15000 μm^2 and 25000 μm^2 in 32C and 64C cells at 30DPA. The larger mean cell area at 30DPA might be indicating that endoreduplication could be giving a potential to grow bigger as already described for trichome, hypocotyl or sepal epidermal cells (Breuer *et al.*, 2007, Barkla *et al.*, 2018, Gendreau *et al.*, 1998, Roeder *et al.*, 2012).

By integrating the transcriptome data as a function of ploidy together with the ploidy distribution information in the pericarp, we can now propose a spatial distribution of the expression of all genes thus obtaining a cartography of gene expression within the pericarp overtime. Using this information, we assayed expression of known regulators of cell cycle or endoreduplication from *Arabidopsis* and Tomato and found for example that *SICC52A1*, a positive regulator of endoreduplication in Tomato is expressed in mesocarp cells favoring their endoreduplication. This cartography could be confirmed by performing *in situ* hybridization of set of genes or by spatial transcriptomic experiments.

Chapter 3: Materials and methods

I. Biological materials

1. Plant materials and Cultures

A) *Solanum lycopersicum*

Tomato plants *Solanum lycopersicum* Mill. Cv. West Virginia 106 (WVA106) were cultivated in greenhouse in 5L pots with fertilization. Plants were cultivated with a photoperiod of 16H light condition at 25°C and 8h night at 20°C and a relative humidity between 70 and 75%.

For RNA-seq and *FisH*, individual flowers were tagged on the day of anthesis, fully opened flowers that is wilt the next day. Fruits were harvested at specific stages 6DPA, 9DPA and 12DPA.

B) *Nicotiana benthamiana*

Tobacco plants *Nicotiana benthamiana* were cultivated in greenhouse in Ø13cm pots with subirrigation. Plants were cultivated with a photoperiod of 16H day at 25°C and 8h night at 20°C and a relative humidity between 70 and 75%.

C) Bacterial stains and culture media

Escherichia coli (*E.coli*) TOP10 strain from INVITROGEN (catalog number: C404052) have been used for the cloning of the plasmid used in CRISPR-imaging experiments.

Agrobacterium tumefaciens GV3101 and C58C1 stains have been used for stable and transient plant transformations of Tomato and Tobacco.

TOP10 bacteria are cultivated at 37°C and *Agrobacteria* are cultivated at 28°C. Bacterial culture medium for both *E.coli* and *Agrobacteria* are commercial LB media from DUCHEFA BIOCHEMIE (catalog number: L1704) containing Bactotryptone 1% w/v, yeast extract 0.5% w/v and NaCl 1% w/v with appropriate antibiotics when necessary.

II. Cytological analysis

1. Isolation of tomato nuclei and flow cytometry

Nuclei were prepared 20 sections of tomato pericarp fruits at 6DPA and 10 sections for 9DPA and 12DPA. Nuclei isolation is performed as described in Bourge *et al.*, 2018 and resumed hereafter. Pericarp sections are chopped using a razor blade in 2mL of Gif Nuclear Buffer (GNB) and the suspension was filtered twice through a 48 µm nylon mesh before adding DAPI (4',6-diamidino-2-phenylindole).

GNB is composed of 45 mM MgCl₂, 30 mM Sodium-Citrate and 60 mM MOPS pH 7.0, 1% PVP 10.000, 0.1% Triton X-100 and 10 mM sodium metabisulfite (S2O₅Na₂).

Nuclei were sorted according to their ploidy levels with a MoFlo ASTRIOS EQ (BECKMAN COULTER) and harvested either in a screw cap 2mL tube containing 900µL of TRIzol™ Reagent (THERMOFISHER SCIENTIFIC) or on a microscope glass slides (Superfrost) prepared with a 10 µl of a cushion made of 2M sucrose, 50% GNB and 1% paraformaldehyde.

100k nuclei were sorted in each tube for RNA-seq, mixed by inversions and frozen in

liquid nitrogen before being stored at -80°C while for *FisH*, slides were stored in -20°C freezer after sorting 2000 nuclei.

III. Molecular biology

1. Methods of nucleic acid manipulation and analysis

A) Extraction of nucleus RNA

The ploidy specific transcriptome was obtained from 90 tomato plants that were cultivated in three cultures between April and June 2017 for a final harvest of 135 fruits pericarp for each culture (total = 405 fruits) at 3 different developmental stages, 6DPA, 9DPA and 12DPA. Pericarps from 20 fruits were used for ploidy sorting of 200k nuclei of 2C or 4C or 100k nuclei for other ploidy levels

A 2mL screw cap tube contains 100k nuclei sorted in 900µL of TRIzol™ Reagent resulting in a final volume of around 1.2mL. Before extraction 1µL of a dilution 1:1000 of ERCC RNA Spike-In Mix (THERMOFISHER SCIENTIFIC) was added as a control of RNA quantification (more details in Chapter 2.I.A): Use of spike-in to normalize gene expression in a global shift in expression context).

The mix of nuclei and spike-ins were shortly vortexed and rested for 5mn at room temperature before adding 200µL of Chloroform then tubes are agitated by hand for 15sec and incubated for 3min at room temperature. After the incubation, tubes are centrifuged for 15mn at 12000g and 4°C, this centrifugation allow the creation of phases and the supernatant (aqueous phase) is pipetted (around 840µL) and transferred and mixed in a new tube that contains 420µL of 100% ethanol. The next steps follow the protocol of the commercial solution from RNeasy Micro Kit (QIAGEN) from step 3. A final volume of 14µL of RNA is obtained from the extraction and due to low input of material at first, low concentrations of RNA are obtained.

B) Quality controls of RNA

The quality control of nuclei RNA has been performed on the transcriptomics platform of Neurocentre Magendie (University of Bordeaux, Bordeaux, Nouvelle Aquitaine) on an Bioanalyzer 2100 (AGILENT) using an RNA 6000 Nano chip (AGILENT).

C) Synthesis of libraries and RNA sequencing

The synthesis of the RNA libraries and the sequencing have been performed by the GeT-Genotoul platform (INRAE Occitanie-Toulouse). After the synthesis of the libraries, we performed a Duplex-specific nuclease (DSN) treatment of the libraries to remove the maximum amount of rRNA that are present in nuclei RNA that are not yet matured. The DSN treatment has been performed using the commercial enzyme from EVROGEN following manufacturer's instructions. Following DSN treatment, the Nuclei RNA have been paired-end sequenced with a HiSeq3000 (ILLUMINA).

IV. Cytogenetic

1. *FisH*

The *FisH* protocol is adapted from Bey *et al.*, 2018.

Fruit is sectioned transversally around equator of the fruit and pericarp section are then 100 to 150 μm sections were obtained using a vibrating blade microtome Microm HM 650V (Microm MicroTech France, <http://mm-france.fr/>).

Buffer that are used to wash or handle sample in these protocol are classical cytological solution for biological sample such as Phosphate buffer saline (PBS, NaCl: 137 mM, KCl: 2.7 mM, Na_2HPO_4 : 10 mM, KH_2PO_4 : 1.8 mM and pH7) and saline-sodium citrate (SSC, NaCl: 3M, trisodium citrate : 300mM, pH7).

Two to three slices are placed in a well in a 24 well plate containing a fixative fixation (Formaldehyde 4%, DMSO 10% in PBS1X) and samples are place 5min under vacuum infiltration at 20kDA and let stand in the solution at room pressure and room temperature (RT). After, samples are put in successive solvent bath for 5min each (2x methanol 100%, 2x ethanol 100%). In the last ethanol bath samples rest for at least 45mn at 4°C and samples are then transferred in a solution of 50%ethanol/50%xylylene for 30mn at RT. Before proceeding to next steps, extra care is taken to clean the samples of xylene solution before successive solvent wash bath for 5min at RT each (2x ethanol). Then samples will be rehydrated with successive wash bath in water-based buffer for 5min at RT in each bath (2x PBS1X+0.1% v/v Tween20 and 2x SSC2X) and a last SSC2X bath is performed in presence of 100 $\mu\text{g}/\text{mL}$ of RNaseA for 1h at 37°C, then an equal volume of Hybridation buffer (HB, 50 mM Sodium phosphate pH 7.0, 2x SSC, 50% deionized formamide) is added and samples rests for 30mn at RT. Following step is critical because it required to physically manipulate the samples to transfer it in a PCR 0.2mL microtube containing 100 μL of HB and rest for 30min before changing the solution with the probe solution (HB + 0.5 to 3 μL of labelled oligonucleotide, here the dye was a TexasRed). Tubes are then incubated in a thermal cycle for 1H at 37°C, the 4mn at 88°C followed by a ramp down to 37°C in 3min and samples rest at 37°C for 12-15h. The next day, the samples are washed in a stringent bath of HB for 30mn at 42°C followed by 5 water-based baths for 10min in PBS1X, at RT 15min in PBS1X+10 μg DAPI at RT and 3x 5min at RT in PBS1X. Finally, the samples are transferred to superfrost slides in VECTASHIELD containing DAPI (VECTOR LABORATORIES, catalog number: H-2000) under stereomicroscope to assure the flatness of the samples, then samples are recovered and covered with a coverslip (20x20, #1.5) and sealed with nail hardener. Samples can be imaged right after or stored several weeks at 4°C.

2. CRISPR-imaging

All the constructs needed for the development of the CRISPR imaging were obtained using the Goldengate cloning method (Engler *et al.*, 2008) that allowed the assembly of the multiple transcriptional units that were necessary for our work.

First, the dCAS9 coding sequence was amplified and mutated from the pDe-CAS9-D10A plasmid (Fauser *et al.*, 2014) by PCR and transferred into an entry GoldenGate vector. Then, this coding sequence was assembled with the 35S promoter, t the GFP at the N terminal or C terminal end and the Octopine synthase (OCS) terminator, constituting a transcription unit, into a plasmid named. The dCAS9 was tagged at its N-terminal or C-terminal part to verify if the position of the fluorescent protein influences the fluorescent signal.

In parallel, the sgRNA was cloned by PCR and assembled into a level1 plasmid under the control of the AtU6 promoter (AtU6) of shRNA (small hairpin RNA). Finally, a final plasmid was generated by assembling the transcriptional units containing the dCAS9 plasmids for multigenic expression. This final plasmid contained the dCAS9 tagged with N-GFP or C-GFP, the sgRNA targeting telomeres or 5SrDNA (sgRNA^{telo or 5S}) and the selective marker nptII for Kanamycin resistance in plants.

Each assembled plasmid has been used to transform *E.coli*. From the bacteria resistant to the selective antibiotic, PCR on single colony has been done to verify the correct construction of the plasmids by checking the size of the PCR product. If the size was corresponding to the one expected, then purified plasmids from *E.coli* were sent for sequencing to verify the sequence of the extremities of the inserted cassette.

These plasmids were used for the transformation of tomato protoplasts, tobacco leaves, and to perform Tomato stable.

3. Imaging

All microscopy experiment has been realized on a confocal microscope Zeiss LSM880 with airyscan piloted with Zen Black 2.1. FisH acquisition were performed using x63 objective with oil immersion in airyscan mode.

4. Image analysis

Airyscan raw images were saved on the computer used to pilot the microscope and processed using ZenBlue 2.3 with automatic parameters.

Then images are analyzed using Fiji (Schindelin *et al.*, 2012) and using homemade macro and plugin developed by Fabrice Cordelière (Bordeaux Imaging Center).

V. Bioinformatic

1. Software

R version used for this work was 4.0.1 (Team R Core 2017) and Rstudio 1.3. Statistical tests were performed with native R functions.

All plots were generated with the ggplot2 R-package (Wickham, 2016).

Figures was assembled using Inkscape 1.0.

2. Mapping and normalization of RNA-Seq data

Sequence quality checking, read trimming and mapping was performed as in Pirrello *et al.*, 2018. The Tomato reference genome used is the Sly2.40 tomato reference genome with STAR. ERCC spike-ins has been used to normalize gene expression.

3. Identification of differentially expressed genes and clustering

The differential expression analyses were performed with the DESeq2 package as performed and described in Pirrello *et al.*, 2018. Normalized expression values were then used to identify DEG. The data sets for all stages were processed the same way. First, log₂ fold changes (Log₂FC) and false discovery rates (FDRs) for each gene were filtered. We used a filter for Log₂FC to be > 2 and <2 and FDR <0,01. Then DE genes

were clustered according to hierarchical clustering of standardized expression values $((\text{expression value}) - (\text{mean expression of gene}) / (\text{standard deviation of expression of gene X}))$. Distances between genes were calculated with the Euclidean distance measure, and Ward method applied to square values was used as agglomeration method.

4. Gene ontology enrichment

Gene ontology enrichment were done using Plaza 4.5 (Van Bel *et al.*, 2018) using p-value corrected using Bonferroni correction <0.05 and default parameters. Then gene ontology has been cleaned from redundancy using REVIGO (Supek *et al.*, 2011). REVIGO files were then filtered on R following protocol from Bonnot *et al.*, 2019.

5. Ploidy mapping and analysis

Data were obtained from segmented nuclei from images, after filtering step detailed in chapter X nuclei are segmented and the voxel contained in the segmented nucleus are quantified to measure the volume. Cell areas have been determined manually by surrounding cells on their largest diameter. Ploidy counting has also been done manually checking in 3D in successions of image if close spot is a unique or a duplicated spot.

Chapter 4: Discussion and perspectives

In Tomato fruit, endoreduplication is an important process taking place during development and ploidy levels can attain up to 256C and even 512C in some varieties (Cheniclet *et al.*, 2005). A strong correlation has been found between endoreduplication level and the mean cell size in the pericarp and also fruit weight (Cheniclet *et al.*, 2005). However, this correlation between mean cell size and mean ploidy level is weaker when cell size and ploidy level are measured at the cellular level instead of the tissue level (Bourdon *et al.*, 2011). Nevertheless, the largest cells are always highly endoreduplicated in the pericarp (Bourdon *et al.*, 2011). In the pericarp of fully grown fruit, endoreduplication is not evenly distributed but organized as a gradient of increasing ploidy levels toward the center of the tissue (Bourdon *et al.*, 2011). Endoreduplication might thus act as a growth factor for pericarp cells. In addition, in a fully grown fruit, an increase of ploidy levels is associated with a global gene expression shift but also to specific regulation of the expression of sets of genes (Pirrello *et al.*, 2018).

Based on these observations, mainly obtained in fully grown fruits, we wanted to know i) if endoreduplication is already distributed as a gradient in young tomato fruit pericarps as observed in fully grown fruit ii) if the endoreduplication is correlated with cell growth in young fruits and iii) if a specific regulation of gene expression occurs in these young fruits. Thus, my PhD work aimed at determining the ploidy distribution in the pericarp during early fruit development while endoreduplication is taking place. The objective was also to determine the genes that have a ploidy specific expression and to identify potential molecular markers of ploidy levels. To do so, we studied during early fruit development how endoreduplication distribution progresses in the different cell layers of the pericarp and if cell specialization is already established in these endoreduplicated cells. We mapped endoreduplication in the pericarp of tomato fruit using *FisH* and analyzed ploidy specific transcriptome of ploidy sorted nuclei at three developmental stages, 6DPA, 9DPA and 12DPA. These stages correspond to an important step in endoreduplication progression since they correspond to the transition between a stage where numerous cells are 2C and 4C, thus dividing, at 6DPA, to a stage where most cells that are endoreduplicating and are 8C or more. The data obtained could help to better understand how endoreduplication is taking place in the pericarp and how it is regulating genetic programs in cells during early development. The results obtained pave the way for a wider project questioning the functional role of endoreduplication in the pericarp and how entering endoreduplication is genetically controlled in cells of the pericarp.

I. *Technological development of tools for the determination of in situ ploidy levels and ploidy specific transcriptome*

1. *Transcriptome related technologies*

The transcriptome of sorted nuclei was easily obtained and only few adjustment of the protocol was needed to correctly extract RNA to obtaining RNA-seq data, most of the more time spent was due to logistical problem. The use of spike-ins allowed to quantify the global shift in the case of ploidy level variation and was a good quality control of the sample after sequencing because we can evaluate the sensibility of the

experiment, here we were theoretically able to detect a single RNA per nucleus. However, I encounter more difficulties with the differential analysis because of the poor annotation of the tomato genome in the public databases, for example it was difficult to clearly identified in datasets genes related to central metabolism, that has been extensively studied.

2. *FisH* related technologies

The use of Oligo Fish allowed to map ploidy levels in the pericarp and optimization of sample handling and imaging allowed to map 6DPA 9DPA and 12DPA fruits. Oligo *FisH* is working only when multiple dyed oligonucleotides are bound on the same area, so we wanted a repeated sequence and the choice of 5SrDNA was motivated by the low number of known repeated sequence on a single chromosome in Tomato genome. We also did use a probe targeting telomere repeats, but the number of dots and lack of time did not allow ploidy quantification in these samples for now. The automatization of ploidy quantification by the help of the Image J plugin could help with the used of telomere probe to see if it can be used in ploidy quantification. However, we have seen that the plugin is not yet functional, but it should be rapidly able to be functional and will way facilitate ploidy quantification in a lot of sample.

3. *CRISPR*-imaging

CRISPR-imaging is a promising tool that need to be adapted in the use of living organisms such as plants and not cellular model as in animals because in complex organisms CAS9 is interfering with cellular activities and inhibit normal development of plants when targeting repeated sequences. However, we only tested a single approach while there are at least three other popular approaches, unfortunately transient transformation is useless to assay the “toxicity” of the method. On the point of CRIPSR imaging, it could be interesting to communicate with other team working on this specific theme to coordinate a project in order to tackle the methods and find the one that would allow to assay chromatin organization in living cells.

II. *Global shift in expression upon endoreduplication.*

A global shift in expression has been reported in the case of allopolyploidy and autopolyploidy and resulted 1.6 to 2.1 more gene expression while comparing polyploid and diploid plants (Coate and Doyle 2010, Robinson *et al.*, 2018). Endoreduplication has also been shown to induce a global transcription increase in fully grown tomato pericarp by measuring and comparing gene expression in the different ploidy levels and this shift corresponds to a doubling of gene expression between successive ploidy levels (Pirrello *et al.*, 2018).

In this PhD work, we used spike-ins, which is a mix of RNAs from a reference library for those the sequences and concentrations are known allowing to scale gene expression. The spike-ins allowed showing that there is a global shift in gene expression corresponding to a doubling with endoreduplication at early developmental stages. This change in global expression is similar to the global shift observed in the later mature green stage (Pirrello *et al.*, 2018) suggesting that in the pericarp the doubling of gene expression is due to endoreduplication and not dependent on the

developmental stage. The spike-ins have helped to determine that gene expression is increasing with ploidy levels but did not allow to determine an absolute value of increase, doubling is an approximate value. To determine the exact value more sampling will be necessary to increase statistical force of our analysis.

III. A subset of genes is specifically regulated by ploidy levels in the pericarp.

The global shift in expression related to the increase of ploidy levels in the pericarp is accompanied with ploidy specific regulation of subsets of genes (Chapter 2.II.2.B) Gene expression is changing during early fruit development and based on ploidy levels). We found that five trends of expression were sufficient to represent the ploidy specific expressions with genes mostly expressed in non endoreduplicated nuclei, genes expressed in newly or low endoreduplicated nuclei, and genes expressed in endoreduplicated nuclei or higher ploidy levels. Ploidy specific gene expression related to ploidy levels has also been found in Arabidopsis ploidy specific transcriptome (Bhosale *et al.*, 2018). To better understand which types of genes are associated with expression trends, we performed a GO enrichment analysis of the genes showing these different trends of expression. We have found different enriched functional categories such as those related to cell cycle and cuticle metabolism mostly in non endoreduplicated cells and genes related to sugar metabolism in endoreduplicated cells. This finding is suggesting that endoreduplication could be specifying cell differentiation in the pericarp and that sugar metabolism and storage is occurring specifically in endoreduplicated cells (Beauvoit *et al.*, 2014). While non endoreduplicated cells are responsible of the establishment of the protection of the fruit via the accumulation of cuticle (Hen-Avivi *et al.*, 2014) and the production of new cells to assure growth flexibility to the pericarp (Liu *et al.*, 2003).

Interestingly, in the transcriptome data, we have found that this cell type specific expression is also influenced by the developmental stage since, for cuticle synthesis, we found that as early as 6DPA cuticle synthesis is important in non endoreduplicated cells, with the expression of genes related to Cutin and Wax synthesis pathways and only wax synthesis pathway is more importantly represented later in the development (Chapter 2.II.2.D) Dynamics of ploidy specific gene expression in the course of early fruit development).

The few enriched GO terms found in endoreduplicated cells are suggesting that cells are engaged in very specific pathways related to carbohydrate metabolism, cell wall synthesis and modification or hormone synthesis and response. The enrichment of GO terms in clusters with high expression in higher ploidy levels were generally associated with central metabolism, more specifically to glycolysis and pentose phosphate pathway and this is consistent with the expression of enzymes necessary to produce the measured metabolites accumulating in the fruit at these stages of fruit development (Beauvoit *et al.*, 2014, Colombié *et al.*, 2015). In the genes associated with carbohydrate metabolism we also found some related with cell wall at all stages studied. We found genes involved in both synthesis and relaxation such as hemicellulose synthesis genes (Rösti *et al.*, 2007, Meents *et al.*, 2018) and cell wall

modifying genes (Terao *et al.*, 2013, Amos and Mohmen 2019). The expression of these genes in endoreduplicated cells would be suggesting that both new cell wall production and modification of cell wall are supporting cell growth. Similar observations have been done for cell wall regulation in endoreduplicated cell of Arabidopsis roots (Bhosale *et al.*, 2018) and it has been proposed that endocycle could be an important factor in regulating cell wall production and modification to sustain rapidly expansion of cells (Bhosale *et al.*, 2019). Interestingly, we found in the same clusters biological categories related to hormonal regulation with auxin, brassinosteroid or abscisic acid related genes, suggesting that ploidy level can modulate the response to these hormones. Similarly, in Arabidopsis endoreduplicated root cells, a ploidy specific expression of genes implicated in hormonal responses was found (Bhosale *et al.*, 2018). Further exploration of hormone accumulation using reporting lines or fruit specific hormone synthesis mutant analyses would be helpful to understand how endoreduplication is regulated in the pericarp. Indeed, it could be interesting to use fluorescent reporter lines for auxin or cytokinin (Liao *et al.*, 2015, Steiner *et al.*, 2020) in order to study the distribution of hormones in relation to the ploidy levels in the pericarp. While fruit specific hormone synthesis mutants might be interesting to study the modification of distribution of ploidy levels in the pericarp.

We found specific set of genes expressed either in non-endoreduplicated or in endoreduplicated cells, but we also identified a singular behavior for a group of genes expressed specifically in 8C cells at 9- and 12DPA and not at 6DPA. We looked at gene ontology enrichment, however, we did not find any enriched functional categories. Specific expression of genes in 8C cells can also be noticed in data of fully grown fruit and contrary to our results, gene ontology enrichment was found for this category (Pirrello *et al.*, 2018). The lack of enrichment might be due to an “undifferentiated” nature of the 8C cells that are just entering endoreduplication or because the pool of 8C sorted nuclei correspond to a heterogeneous population assuring different functions. If each cell type expresses a specific functional category, in a pool of these different cells, these categories might be hidden in our analysis.

Lastly, by comparing our transcriptome data to public dataset obtained from micro-dissected tissues of the pericarp (Shinozaki *et al.*, 2018), we showed that ploidy levels are not randomly distributed within the pericarp but correspond more or less to specific types of cells. We found, for example that epidermal cells express genes found highly expressed in low endoreduplicated nuclei and that mesocarp cells express genes found in highly endoreduplicated nuclei (Chapter 2.II.2.E Comparing ploidy specific transcriptome and laser microdissected pericarp transcriptome). The use of a pool of cells sorted based on ploidy level for transcriptomic has allowed identifying ploidy specific expression patterns. However, if cells with similar ploidy level have different functions, our analysis might not be able to properly distinguish between genes expressed in each cell type. Thus, in the future, it would be interesting to perform single cell RNA-seq (scRNA-seq) or single nuclei RNA-seq (snRNA-seq) that will allow better understanding the spatial resolution of gene expression, identifying potential difference in ploidy populations, and determining cell lineage trajectories (Figure 43A).

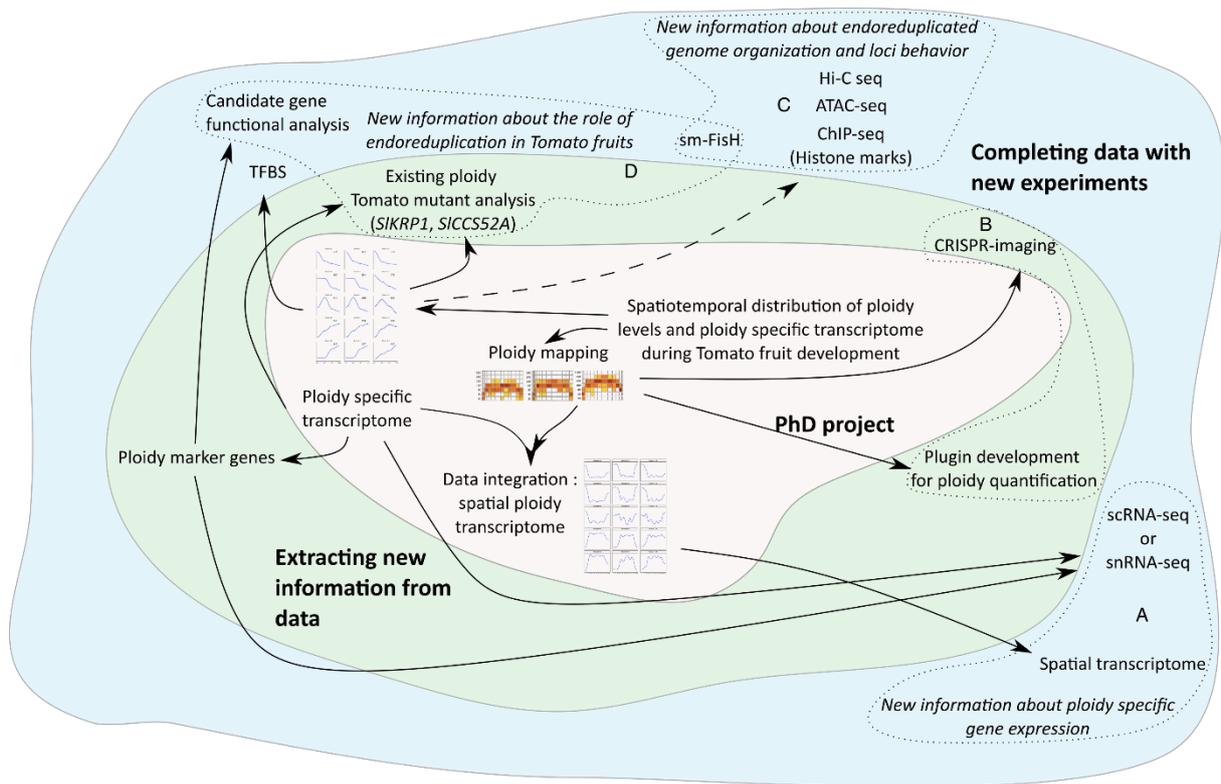


Figure 43. Summary diagram of PhD and its perspectives.

Going from the center to the outside: PhD project in pink, information that can be obtained from the PhD data in green and new information linked to the PhD that need new experiments to be obtained in blue.

(A) Increasing knowledge of ploidy specific gene expression

(B) Dynamic of loci behavior during endoreduplication

(C) Experiments giving new information about genome organization and loci behavior during endoreduplication

(D) Experiments giving new information about the role of endoreduplication in Tomato fruit

Plain arrows represent direct links between information while dotted arrow represent inferred information.

IV. The ploidy distribution in tomato pericarp during early development stages is already organized as a gradient.

In fully grown fruit, ploidy levels are following a gradient from the epidermis (outer and inner) to the internal part of the pericarp with low levels on the outside and high levels in the inside (Bourdon *et al.*, 2011). Here we studied ploidy distribution and cell size in the pericarp at early stages of fruit development. We found that between 6 and 12DPA, the ploidy levels distribution within the pericarp is also following a similar gradual increase of ploidy levels towards the central part of the and that the distribution is highly dynamic at these early stages. The epidermal layers are composed of dividing cells at 6DPA and cells are not all engaged in endoreduplication in the outer epidermis at 12DPA contrary to the inner epidermis where almost all cells became endoreduplicated. The drastic reduction of 2C cells in the epidermal layers is consistent with the measure of cell division activities that are nearly absent in the pericarp after 8DPA (Renaudin *et al.*, 2017). The M Mesocarp layers, formed before anthesis, were also the more endoreduplicated layers with only few non-endoreduplicated cells at 6DPA. These layers contained the largest proportion of higher endoreduplication levels, 32C and 64C at 12DPA. Concerning the other mesocarp layers, M', formed after anthesis, we observed a progressive increase of endoreduplicated cells toward higher ploidy levels but only few 64C were found at 12DPA. Ploidy levels were increasing globally but older mesocarp layers, M layers, were more endoreduplicated than M' probably indicating that endoreduplication is a continuous process and that older cells become more and more endoreduplicated over time. Renaudin *et al.*, 2017 hypothesized that 8C cells already present before anthesis would be localized in the center of the mesocarp, corresponding to M cells and our data support the hypothesis because at 12DPA, M cells, that are older than M' cells are also more endoreduplicated. Thus, this difference in ploidy levels in the mesocarp cell layers would be suggesting that the age of the cell is related to its final size. However, in fully grown fruit, endoreduplication levels are almost similar between M' and M layers (Bourdon *et al.*, 2011) potentially indicating that the difference observed between 6DPA and 12DPA tends to reduce with the age of the fruit. This observed reduction of difference in ploidy levels might be due to a reduction of endoreduplication speed or exit from endoreduplication in M layers or more rapid endoreduplication in M' layers. Moreover, a heterogeneity of ploidy levels was observed within a layer suggesting that endoreduplication is occurring in an asynchronous manner inside the same layer.

The ploidy distribution in the pericarp was determined in a wild-type plant and in optimal growth conditions. It would be interesting to study how distribution could be affected after impairment of the endoreduplication process or cell growth. For example, in the case of the overexpression of *SIKRP1*, it might be interesting to determine ploidy distribution using *FisH* because fruit have a strong reduction of ploidy levels in the pericarp without any visible growth phenotype (Figure 43D, Nafati *et al.*, 2011).

We also observed that mean cell areas in 30dpa fruit pericarps were larger than in our study when comparing the same ploidy level. The larger mean cell area at 30DPA might

be indicating that endoreduplication appears before the cell expand completely and could thus be giving potential to the cell to grow bigger and in SIKRP1-OE plant, even with reduce endoreudplciation in the pericarp, the normal cell size has been maintained in fully grown fruit suggesting the possibility to uncouple endoreduplication and cell growth (Nafati *et al.*, 2011). Moreover, modelling approaches using Arabidopsis as a model have shown that endoreduplication process and cells growth could be stochastic in an organ (Kawate and Tsukaya 2017) and in Tomato pericarp cell age and ploidy are able to influence cell size (Baldazzi *et al.*, 2019), it would be then possible that a minimal ploidy level is sufficient for cells to grow normally and assure their functions. We also observed that the size of endoreduplicated cells was increased with ploidy levels at all stages and that cell size of the same ploidy level in an older fruit was larger. This size difference could be due to cell expansion that can be more important or because the population of cells with the sameploidy level at 12DPA could be composed of newly formed cells but also older cells that have enlarged more.

We demonstrated that Oligo-*FisH* combined with a semi-automated image acquisition process can be efficiently used to determine ploidy levels in three developmental stages and plugin development would help to make this method a powerful cytogenetic approach to evaluate change of ploidy. This would be valuable to enrich the data for other time points such as earlier developmental stages for example to spot cell newly endoreduplicated in the forming M' layers in order to compare cell size and determine if cell size is increasing before or after endoreduplication. Obtaining ploidy distribution in the ovary would help confirming the hypothesis that the most central cells of the pericarp are 8C. Ploidy mapping in the developmental stages between 12DPA and 30DPA could be also interesting data to gather in order for example to see if indeed cell size in an older fruit for a given ploidy levels is higher. Lastly, the study of other zones of the fruit such as above or under the equatorial sections could help to see if endoreduplication is distributed and progressing in the same way compared to the equatorial region. This information could then be used to model the spatial and temporal organization of endoreduplication in the pericarp.

Ploidy mapping using *FisH* corresponds to static images of ploidy levels at different stages. Ploidy determination in living cells by CRIPSR imaging method (Chapter 2.I.C Visualization of genomic loci in living plant cells using CRISPR/Cas9 technology) that uses the capability of a deficient Cas9 to specifically target a locus with the help of the sgRNA could give new information on dynamic of endoreduplicated loci. For example, this method could help to measure distance and map localization of several loci whose expression is differentially regulation in function of endoreduplication level or in response to abiotic stresses for example (osmotic stress, heat stress or cold stress *etc.*) at different ploidy levels.

V. Perspective

1. How is the ploidy-specific expression regulated?

Our ploidy-dependent transcriptome data are underlying the existence of ploidy specific regulation of gene expression that is different than the global shift of

expression. In the introduction (“Endoreduplication effect on genome regulation of transcription”), we have seen that chromosome positioning, chromatin condensation, DNA methylation and transcription factor binding are the different mechanisms that might allow a differential gene expression. In the case of Tomato pericarp, transcriptomic data and ploidy mapping are now available for several early developmental stages, 6DPA, 9DPA and 12DPA and at the end of the fruit growth before ripening, at 30DPA (Bourdon *et al.*, 2011). Using these data, we can now understand how the transcriptional dynamics during endoreduplication are regulated, particularly by searching regulating transcription factors. It is known that co-expressed genes can have common elements in their promoter allowing their co-regulation (Babu *et al.*, 2004). To identify the potential regulators of ploidy-specific expression, we could first identify the putative Transcription Factor Binding Sites (TFBS) that would be overrepresented in the promoters of co-expressed genes during endoreduplication. Based on the enriched TFBS identified, the next step of the analysis would be to identify TF that are binding these sites for different sets of co-expressed genes. We would mainly focus on TF implicated in favoring expression of genes in higher ploidy levels to know how endoreduplication controls expression. Such approach has allowed to uncover conservation of regulatory elements implicated in leaf senescence in Wheat (Borrill *et al.*, 2019). The last step in the identification of potential TF implicated in these differential expressions would be to perform electrophoretic mobility shift assays and transient expression assay to confirm the ability of the TFs to bind the promoters of targets and ability to induce activation or repression of gene expression.

It would also be also interesting to study if, in the pericarp, epigenetic changes, including chromatin conformation, chromosome organization or DNA methylation, occur as a function of endoreduplication, and influence the ploidy specific gene expression. It has been shown that endoreduplication could be associated with more accessible chromatin (Nagymihály *et al.*, 2017, Dorrity *et al.*, 2020). To study chromatin conformation, new datasets need to be produced, ideally obtained from sorted nuclei of the same ploidy levels at the same developmental stages of fruit development. These data would correspond to ChIP-seq analysis targeting active or repressive histone marks in ploidy sorted nuclei allowing to identify genes associated with changes in chromatin condensation in each ploidy levels. Hi-C would be employed to assay the 3-D chromatin conformation according to ploidy levels (Figure 43C). Overall, these data could help to better understand the effect of endoreduplication in modifying genome organization potentially related to changes in gene expression in tomato pericarp. To gain insight into genome organization as a function of endoreduplication imaging approaches could also be used. Genes found as DE in the ploidy transcriptome could be visualized and localized within the nucleus using oligo *FisH* at different developmental stages. By measuring their distances to nucleus elements such as nucleoli, nucleus membrane and other loci, the organization of the loci in the nucleus can be determined. Oligo *FisH* could be used as a mapping tool to localized structure that could be detected with Hi-C, as for example TAD-like structures (Figure 43B). This will bring new information about localization of DE genes in the nucleus in function of ploidy levels potentially revealing localization pattern for genes in relation to endoreduplication level or a co-localization of the co-expressed genes.

Lastly, concerning the ploidy specific gene expression regulation, it would be interesting to study the activity of the newly formed alleles by endoreduplication, for example to know if all copies of the genes are expressed or to assay distribution of mRNA in the cell in function of ploidy levels. To do so, gene expression could be assayed in cells using single molecule *FisH* (smFISH). smFISH is a variation of *FisH* similar to Oligo *FisH* using similar DNA oligonucleotides to target RNA molecules and thus allows tracking of single molecules of RNA. By a combination of two colors for the fluorochrome, one for exons and introns, the transcription loci can be visualized (Figure 43D). Sm*FisH* has been successfully used to study single cell regulation of transcription and mRNA dynamic of flowering genes under cold influence in meristem of Arabidopsis (Rosa *et al.*, 2016), the authors shows that a cold responsive gene is able to inhibit FLOWERING LOCUS C, a gene responsible for the positive regulation of flowering in plants. The data obtained with sm*FisH* would allow to determine if all copies of endoreduplicated genes are transcribed for DE genes and non-DE genes for example.

2. Spatiotemporal gene expression in the pericarp is correlated to ploidy levels.

By combining ploidy distribution data and transcriptomic data, we have shown that by weighting gene expression as a function of ploidy level proportion in a cell layer, we were able to estimate the spatiotemporal expression of genes in the pericarp. We observed in the representation of expression clusters in function of cell layers that genes highly expressed in high ploidy levels in the mesocarp have an oscillating expression pattern from one layer to another. This oscillation is creating local minima of gene expression that could be important because it means that whole set of genes are differentially expressed between neighbor cells. These variations could be artefactual because we know that 32C is not the maximal ploidy level at 12DPA and we know that there is also 4-5% of 64C (Figure 25A) that we have identified in the pericarp of 12DPA fruit in M layers (Figure 37C). These drops of expression can probably be due to missed expression data from 64C that we can expect to be higher than 32C for genes in clusters 12-4 and 12-5. However, as the ploidy mapping is done from the outer epidermis to the inner epidermis in the pericarp, “vertically” by cell layer, the “horizontal” information, mapping within a layer, is not accessible. It could be interesting to gather absolute spatial gene expression information in the pericarp to measure the spatial effect of endoreduplication on gene expression inside a cell layer where two neighbor cells can have different ploidy levels. To obtain this absolute spatial information in the pericarp the recently developed transcriptomic method named “spatial transcriptomic” (Figure 43A) could be used. Spatial transcriptomic allows to measure the gene expression *in situ* within a tissue sample at around 2µm of resolution (Vickovic *et al.*, 2019) thank to a chip that have areas with multiple wells containing unique barcoded probe allowing to map each sequencing read.

3. Towards the identification of ploidy marker genes

We used the tomato ploidy transcriptome in association with the Arabidopsis root ploidy transcriptome (Bhosale *et al.*, 2018) to identify genes that could be specific markers of ploidy levels. We have been able to identify common markers for 2C to 16C

ploidy levels. Further experiments would be required to determine if markers are indeed specific to ploidy levels. To do so, the marker genes could be tested by qPCR in a large set of nuclei, sorted based on ploidy level, from different organs and developmental times. In addition, the promoters of these potential marker could be used to produce fluorescent transgenic reporter plants and verify their expression specifically in the expected ploidy level.

The marker genes can be useful in several cases:

- to determine the ploidy levels of cells for which we would have the transcriptome. This type of analysis could be applied in the case of single cell RNA seq or spatial transcriptome analysis. Marker genes found in Arabidopsis ploidy root transcriptome have been used to predict ploidy map of the root (Bhosale *et al.*, 2018).
- to use promoters to express reporter genes in cells of specific ploidy levels. Using several fluorescent colors, we could have plants expressing different fluorescent colors to visualize different ploidy levels simultaneously in the tissues. These reporter plants could be used to visualize ploidy distribution in mutant plants for example after crossing reporter lines with mutant lines.
- Using ploidy specific marker gene promoters can be useful to express transgenes having a known or potential role in endoreduplication regulation to increase or limit endoreduplication such as *SICCS52A* (Mathieu-Rivet *et al.*, 2010) or *SIKRP* (Nafati *et al.*, 2011), respectively. This approach will limit pleiotropic effect of overexpressing a gene in the plant or overexpressing a knock down construct. *SICCS52A* for example was expressed using a 35S promoter and resulted in pleiotropic effects giving dwarf plants limiting interpretation of phenotype on fruits.

Ploidy specific promoters could be used in a cell specific genome editing approach (Decaestecker *et al.*, 2019, Bollier *et al.*, 2020), to knock-out a gene or more to induce precise changes in endoreduplication. Those new plants could be inspected in a ploidy specific manner using Oligo-FisH methods and/or ploidy specific transcriptome to see the influence of ploidy distribution in the pericarp.

4. Establishment progress and role of endoreduplication in the pericarp and Tomato fruit growth

The ploidy mapping at 6DPA revealed that endoreduplication is already well established in the pericarp at this stage and most of mesocarp cells are already 16C. Moreover, the expression profile of ploidy DE genes at 6DPA showed that 4C nuclei are identified as nuclei from cells in division but not endoreduplication, thus do not correspond to 4C endoreduplicated nuclei. Both ploidy map and transcriptome data suggest that at 6DPA we already missed the genetic switch to endoreduplication in the cell of pericarp or at least with our analysis. It could be valuable to have information about earlier developmental stage such as the day after anthesis and/or around 3DPA where normally cell expansion is highly active and maybe endoreduplication as well in

the pericarp (Renaudin *et al.*, 2017), this information could give new insight about signaling of endoreduplication in M layers and newly formed M' cells. Also, it will be highly important for future project to dissect and study the role of the ploidy specifically expressed genes in endoreduplication progress and control in the pericarp. Moreover, ploidy specific transcriptome could be supported by equivalent work in stress condition or different genetic background (other tomato varieties or mutant background) in order to assay the effect of variation of endoreduplication in the pericarp and further increase knowledge about its exact role in the Tomato fruit development.

VI. Conclusion

During the PhD we optimized *in situ* methods for ploidy determination in the pericarp using Oligo-FISH and the use of spike-ins for absolute gene expression determination in ploidy sorted nuclei having a global shift of gene expression. The ploidy map obtained at early stage of development allowed to determine that the gradient distribution of ploidy levels in the pericarp was already established in 6DPA fruits. We also found that cell size is increasing with endoreduplication and the developmental stage suggesting that endoreduplication is not the sole responsible of cell growth in the pericarp. Lastly, the ploidy transcriptome revealed that a global shift of expression, doubling, exists in early stage of development and is due to endoreduplication. The integration of the ploidy map data and the transcriptome data shows that gene expression controlled in a ploidy specific manner in early stage of development inducing spatialization of genetic programs in the pericarp.

The data obtained during the PhD prepares for future projects that could aim at determining the potential rules of establishment and progress of endoreduplication in the pericarp, the identification of transcriptional regulators responsible for the ploidy specific expression and the study of the role of ploidy specifically expressed genes in endoreduplication control.

References

- 1. Abe M, Katsumata H, Komeda Y, Takahashi T.** 2003. Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in Arabidopsis. *Development* **130**, 635–643.
- 2. Abraham Z, del Pozo JC.** 2012. Ectopic Expression of E2FB, a Cell Cycle Transcription Factor, Accelerates Flowering and Increases Fruit Yield in Tomato. *Journal of Plant Growth Regulation* **31**, 11–24.
- 3. Abudayyeh OO, Gootenberg JS, Essletzbichler P, et al.** 2017. RNA targeting with CRISPR–Cas13. *Nature* **550**, 280–284.
- 4. Al-Abdallat AM, Al-Debei HS, Ayad JY, Hasan S.** 2014. Over-Expression of SISHN1 Gene Improves Drought Tolerance by Increasing Cuticular Wax Accumulation in Tomato. *International Journal of Molecular Sciences* **15**, 19499–19515.
- 5. Amijima M, Iwata Y, Koizumi N, Mishiba K.** 2014. The polar auxin transport inhibitor TIBA inhibits endoreduplication in dark grown spinach hypocotyls. *Plant Science* **225**, 45–51.
- 6. Amos RA, Mohnen D.** 2019. Critical Review of Plant Cell Wall Matrix Polysaccharide Glycosyltransferase Activities Verified by Heterologous Protein Expression. *Frontiers in Plant Science* **10**.
- 7. Ariizumi T, Kishimoto S, Kakami R, et al.** 2014. Identification of the carotenoid modifying gene PALE YELLOW PETAL 1 as an essential factor in xanthophyll esterification and yellow flower pigmentation in tomato (*Solanum lycopersicum*). *The Plant Journal* **79**, 453–465.
- 8. Ariizumi T, Shinozaki Y, Ezura H.** 2013. Genes that influence yield in tomato. *Breeding Science* **63**, 3–13.
- 9. Atmodjo MA, Sakuragi Y, Zhu X, Burrell AJ, Mohanty SS, Atwood JA, Orlando R, Scheller HV, Mohnen D.** 2011. Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex. *Proceedings of the National Academy of Sciences* **108**, 20225–20230.
- 10. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA.** 2004. Structure and evolution of transcriptional regulatory networks. *Current Opinion in Structural Biology* **14**, 283–291.
- 11. Bainard JD, Bainard LD, Henry TA, Fazekas AJ, Newmaster SG.** 2012. A multivariate analysis of variation in genome size and endoreduplication in angiosperms reveals strong phylogenetic signal and association with phenotypic traits. *New Phytologist* **196**, 1240–1250.

- 12. Baldazzi V, Valsesia P, Génard M, Bertin N.** 2019. Organ-wide and ploidy-dependent regulation both contribute to cell-size determination: evidence from a computational model of tomato fruit. *Journal of Experimental Botany* **70**, 6215–6228.
- 13. Ballicora MA, Iglesias AA, Preiss J.** 2004. ADP-Glucose Pyrophosphorylase: A Regulatory Enzyme for Plant Starch Synthesis. *Photosynthesis Research* **79**, 1–24.
- 14. Baloban M, Vanstraelen M, Tarayre S, Reuzeau C, Cultrone A, Mergaert P, Kondorosi E.** 2013. Complementary and dose-dependent action of AtCCS52A isoforms in endoreduplication and plant size control. *New Phytologist* **198**, 1049–1059.
- 15. Barkla BJ, Rhodes T, Tran K-NT, Wijesinghege C, Larkin JC, Dassanayake M.** 2018. Making Epidermal Bladder Cells Bigger: Developmental- and Salinity-Induced Endopolyploidy in a Model Halophyte. *Plant Physiology* **177**, 615–632.
- 16. Barow M.** 2006. Endopolyploidy in seed plants. *BioEssays* **28**, 271–281.
- 17. Barow M, Meister A.** 2003. Endopolyploidy in seed plants is differently correlated to systematics, organ, life strategy and genome size. *Plant, Cell & Environment* **26**, 571–584.
- 18. Bartels A, Han Q, Nair P, et al.** 2018. Dynamic DNA Methylation in Plant Growth and Development. *International Journal of Molecular Sciences* **19**, 2144.
- 19. Beaulieu JM, Leitch IJ, Patel S, Pendharkar A, Knight CA.** 2008. Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytologist* **179**, 975–986.
- 20. Beauvoit BP, Colombié S, Monier A, et al.** 2014. Model-Assisted Analysis of Sugar Metabolism throughout Tomato Fruit Development Reveals Enzyme and Carrier Properties in Relation to Vacuole Expansion. *The Plant Cell* **26**, 3224–3242.
- 21. Berckmans B, Lammens T, Daele HVD, Magyar Z, Bögre L, Veylder LD.** 2011. Light-Dependent Regulation of DEL1 Is Determined by the Antagonistic Action of E2Fb and E2Fc. *Plant Physiology* **157**, 1440–1451.
- 22. Berdasco M, Alcázar R, García-Ortiz MV, et al.** 2008. Promoter DNA Hypermethylation and Gene Repression in Undifferentiated Arabidopsis Cells. *PLOS ONE* **3**, e3306.
- 23. Bernard A, Domergue F, Pascal S, Jetter R, Renne C, Faure J-D, Haslam RP, Napier JA, Lessire R, Joubès J.** 2012. Reconstitution of Plant Alkane Biosynthesis in Yeast Demonstrates That Arabidopsis ECERIFERUM1 and

ECERIFERUM3 Are Core Components of a Very-Long-Chain Alkane Synthesis Complex. *The Plant Cell* **24**, 3106–3118.

- 24.** Bernard A, Joubès J. 2013. Arabidopsis cuticular waxes: Advances in synthesis, export and regulation. *Progress in Lipid Research* **52**, 110–129.
- 25.** Berr A, Schubert I. 2007. Interphase Chromosome Arrangement in Arabidopsis thaliana Is Similar in Differentiated and Meristematic Tissues and Shows a Transient Mirror Symmetry After Nuclear Division. *Genetics* **176**, 853–863.
- 26.** Bessire M, Borel S, Fabre G, *et al.* 2011. A Member of the PLEIOTROPIC DRUG RESISTANCE Family of ATP Binding Cassette Transporters Is Required for the Formation of a Functional Cuticle in Arabidopsis. *The Plant Cell* **23**, 1958–1970.
- 27.** Bhosale R, Boudolf V, Cuevas F, *et al.* 2018. A Spatiotemporal DNA Endoploidy Map of the Arabidopsis Root Reveals Roles for the Endocycle in Root Development and Stress Adaptation. *The Plant Cell* **30**, 2330–2351.
- 28.** Bi X, Cheng Y-J, Hu B, Ma X, Wu R, Wang J-W, Liu C. 2017. Nonrandom domain organization of the Arabidopsis genome at the nuclear periphery. *Genome Research* **27**, 1162–1173.
- 29.** Biais B, Bénard C, Beauvoit B, *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 Physiology* **164**, 1204–1221.
- 30.** Bollier N, Buono RA, Jacobs TB, Nowack MK. Efficient simultaneous mutagenesis of multiple genes in specific plant tissues by multiplex CRISPR. *Plant Biotechnology Journal* n/a.
- 31.** Bonnot T, Gillard MB, Nagel DH. 2019. A Simple Protocol for Informative Visualization of Enriched Gene Ontology Terms. *Bio-protocol*, e3429–e3429.
- 32.** Borrill P, Harrington SA, Simmonds J, Uauy C. 2019. Identification of Transcription Factors Regulating Senescence in Wheat through Gene Regulatory Network Modelling. *Plant Physiology* **180**, 1740–1755.
- 33.** Boudolf V, Lammens T, Boruc J, *et al.* 2009. CDKB1;1 Forms a Functional Complex with CYCA2;3 to Suppress Endocycle Onset. *Plant Physiology* **150**, 1482–1493.
- 34.** Bourdon M, Coriton O, Pirrello J, Cheniclet C, Brown SC, Poujol C, Chevalier C, Renaudin J-P, Frangne N. 2011. In planta quantification of endoreduplication using fluorescent in situ hybridization (FISH). *The Plant Journal* **66**, 1089–1099.

- 35. Bourdon M, Frangne N, Mathieu-Rivet E, Nafati M, Cheniclet C, Renaudin J-P, Chevalier C.** 2010. Endoreduplication and Growth of Fleshy Fruits. In: Lüttge U,, In: Beyschlag W,, In: Büdel B,, In: Francis D, eds. Progress in Botany. Progress in Botany 71. Berlin, Heidelberg: Springer, 101–132.
- 36. Bourdon M, Pirrello J, Cheniclet C, et al.** 2012. Evidence for karyoplasmic homeostasis during endoreduplication and a ploidy-dependent increase in gene transcription during tomato fruit growth. *Development* **139**, 3817–3826.
- 37. Bourge M, Brown SC, Siljak-Yakovlev S.** 2018. Flow cytometry as tool in plant sciences, with emphasis on genome size and ploidy level assessment. *Genetics & Applications* **2**, 1–12.
- 38. Bramsiepe J, Wester K, Weinl C, Roodbarkelari F, Kasili R, Larkin JC, Hülskamp M, Schnittger A.** 2010. Endoreplication Controls Cell Fate Maintenance. *PLoS Genetics* **6**.
- 39. Breuer C, Braidwood L, Sugimoto K.** 2014. Endocycling in the path of plant development. *Current Opinion in Plant Biology* **17**, 78–85.
- 40. Breuer C, Kawamura A, Ichikawa T, Tominaga-Wada R, Wada T, Kondou Y, Muto S, Matsui M, Sugimoto K.** 2009. The Trihelix Transcription Factor GTL1 Regulates Ploidy-Dependent Cell Growth in the Arabidopsis Trichome. *The Plant Cell* **21**, 2307–2322.
- 41. Breuer C, Morohashi K, Kawamura A, Takahashi N, Ishida T, Umeda M, Grotewold E, Sugimoto K.** 2012. Transcriptional repression of the APC/C activator CCS52A1 promotes active termination of cell growth. *The EMBO Journal* **31**, 4488–4501.
- 42. Breuer C, Stacey NJ, West CE, Zhao Y, Chory J, Tsukaya H, Azumi Y, Maxwell A, Roberts K, Sugimoto-Shirasu K.** 2007. BIN4, a Novel Component of the Plant DNA Topoisomerase VI Complex, Is Required for Endoreduplication in Arabidopsis. *The Plant Cell* **19**, 3655–3668.
- 43. Breuninger H, Lenhard M.** 2010. Chapter Seven - Control of Tissue and Organ Growth in Plants. In: Timmermans MCP, ed. *Plant Development. Current Topics in Developmental Biology*. Academic Press, 185–220.
- 44. Bulankova P, Akimcheva S, Fellner N, Riha K.** 2013. Identification of Arabidopsis Meiotic Cyclins Reveals Functional Diversification among Plant Cyclin Genes. *PLOS Genetics* **9**, e1003508.
- 45. Cabral D, Banora MY, Antonino JD, et al.** 2020. The plant WEE1 kinase is involved in checkpoint control activation in nematode-induced galls. *New Phytologist* **225**, 430–447.
- 46. Carland FM, Nelson T.** 2004. COTYLEDON VASCULAR PATTERN2–Mediated Inositol (1,4,5) Triphosphate Signal Transduction Is Essential for

Closed Venation Patterns of Arabidopsis Foliar Organs. *The Plant Cell* **16**, 1263–1275.

- 47.** **Castellano MM, del Pozo JC, Ramirez-Parra E, Brown S, Gutierrez C.** 2001. Expression and Stability of Arabidopsis CDC6 Are Associated with Endoreplication. *The Plant Cell* **13**, 2671–2686.
- 48.** **Cebolla A, María Vinardell J, Kiss E, Oláh B, Roudier F, Kondorosi A, Kondorosi E.** 1999. The mitotic inhibitor ccs52 is required for endoreduplication and ploidy-dependent cell enlargement in plants. *The EMBO Journal* **18**, 4476–4484.
- 49.** **Chang S-B, Yang T-J, Datema E, et al.** 2008. FISH mapping and molecular organization of the major repetitive sequences of tomato. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology* **16**, 919–933.
- 50.** **Chen B, Gilbert LA, Cimini BA, et al.** 2013. Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **155**, 1479–1491.
- 51.** **Chen J, Zeng B, Zhang M, Xie S, Wang G, Hauck A, Lai J.** 2014. Dynamic Transcriptome Landscape of Maize Embryo and Endosperm Development. *Plant Physiology* **166**, 252–264.
- 52.** **Cheng Y, Cao L, Wang S, et al.** 2013. Downregulation of multiple CDK inhibitor ICK/KRP genes upregulates the E2F pathway and increases cell proliferation, and organ and seed sizes in Arabidopsis. *The Plant Journal* **75**, 642–655.
- 53.** **Cheng AW, Jillette N, Lee P, Plaskon D, Fujiwara Y, Wang W, Taghbalout A, Wang H.** 2016. Casilio: a versatile CRISPR-Cas9-Pumilio hybrid for gene regulation and genomic labeling. *Cell Research* **26**, 254–257.
- 54.** **Cheniclet C, Rong WY, Causse M, Frangne N, Bolling L, Carde J-P, Renaudin J-P.** 2005. Cell expansion and endoreduplication show a large genetic variability in pericarp and contribute strongly to tomato fruit growth. *Plant Physiology* **139**, 1984–1994.
- 55.** **Chew W, Hrmova M, Lopato S.** 2013. Role of Homeodomain Leucine Zipper (HD-Zip) IV Transcription Factors in Plant Development and Plant Protection from Deleterious Environmental Factors. *International Journal of Molecular Sciences* **14**, 8122–8147.
- 56.** **Churchman ML, Brown ML, Kato N, et al.** 2006. SIAMESE, a Plant-Specific Cell Cycle Regulator, Controls Endoreplication Onset in Arabidopsis thaliana. *The Plant Cell* **18**, 3145–3157.

- 57. Coate JE, Doyle JJ.** 2010. Quantifying Whole Transcriptome Size, a Prerequisite for Understanding Transcriptome Evolution Across Species: An Example from a Plant Allopolyploid. *Genome Biology and Evolution* **2**, 534–546.
- 58. Colombié S, Nazaret C, Bénard C, et al.** 2015. Modelling central metabolic fluxes by constraint-based optimization reveals metabolic reprogramming of developing *Solanum lycopersicum* (tomato) fruit. *The Plant Journal* **81**, 24–39.
- 59. Cookson SJ, Radziejwoski A, Granier C.** 2006. Cell and leaf size plasticity in *Arabidopsis*: what is the role of endoreduplication? *Plant, Cell & Environment* **29**, 1273–1283.
- 60. Coombe BG.** 1976. The Development of Fleshy Fruits. *Annual Review of Plant Physiology* **27**, 207–228.
- 61. Cremer T, Cremer M.** 2010. Chromosome Territories. *Cold Spring Harbor Perspectives in Biology* **2**.
- 62. Czerednik A, Busscher M, Bielen BAM, Wolters-Arts M, de Maagd RA, Angenent GC.** 2012. Regulation of tomato fruit pericarp development by an interplay between CDKB and CDKA1 cell cycle genes. *Journal of Experimental Botany* **63**, 2605–2617.
- 63. D’Amato F.** 1984. Role of Polyploidy in Reproductive Organs and Tissues. In: Johri BM, ed. *Embryology of Angiosperms*. Berlin, Heidelberg: Springer, 519–566.
- 64. D’Amato F, Durante M.** 2003. Polyploidy. eLS. American Cancer Society, .
- 65. De Veylder L, Larkin JC, Schnittger A.** 2011. Molecular control and function of endoreplication in development and physiology. *Trends in Plant Science* **16**, 624–634.
- 66. Decaestecker W, Buono RA, Pfeiffer ML, Vangheluwe N, Jourquin J, Karimi M, Isterdael GV, Beeckman T, Nowack MK, Jacobs TB.** 2019. CRISPR-TSKO: A Technique for Efficient Mutagenesis in Specific Cell Types, Tissues, or Organs in *Arabidopsis*. *The Plant Cell* **31**, 2868–2887.
- 67. Del Prete S, Molitor A, Charif D, Bessoltane N, Soubigou-Taconnat L, Guichard C, Brunaud V, Granier F, Fransz P, Gaudin V.** 2019. Extensive nuclear reprogramming and endoreduplication in mature leaf during floral induction. *BMC Plant Biology* **19**, 135.
- 68. DePamphilis ML.** 2005. Cell Cycle Dependent Regulation of the Origin Recognition Complex. *Cell Cycle* **4**, 70–79.
- 69. Dermastia M, Kladnik A, Koce JD, Chourey PS.** 2009. A Cellular Study of Teosinte *Zea mays* subsp. *parviglumis* (Poaceae) Caryopsis Development

Showing Several Processes Conserved in Maize. *American Journal of Botany* **96**, 1798–1807.

- 70.** Desvoyes B, Sanchez MP, Ramirez-Parra E, Gutierrez C. 2010. Impact of nucleosome dynamics and histone modifications on cell proliferation during Arabidopsis development. *Heredity* **105**, 80–91.
- 71.** Doğan ES, Liu C. 2018. Three-dimensional chromatin packing and positioning of plant genomes. *Nature Plants* **4**, 521–529.
- 72.** Domínguez E, López-Casado G, Cuartero J, Heredia A. 2008. Development of fruit cuticle in cherry tomato (*Solanum lycopersicum*). *Functional Plant Biology* **35**, 403–411.
- 73.** Dong P, Tu X, Chu P-Y, Lü P, Zhu N, Grierson D, Du B, Li P, Zhong S. 2017. 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. *Molecular Plant* **10**, 1497–1509.
- 74.** Dong P, Tu X, Li H, Zhang J, Grierson D, Li P, Zhong S. 2020. Tissue-specific Hi-C analyses of rice, foxtail millet and maize suggest non-canonical function of plant chromatin domains. *Journal of Integrative Plant Biology* **62**, 201–217.
- 75.** Dorrity MW, Alexandre C, Hamm M, Vigil A-L, Fields S, Queitsch C, Cuperus J. 2020. The regulatory landscape of *Arabidopsis thaliana* roots at single-cell resolution. *bioRxiv*, 2020.07.17.204792.
- 76.** Doyle JJ, Coate JE. 2018. Polyploidy, the Nucleotype, and Novelty: The Impact of Genome Doubling on the Biology of the Cell. *International Journal of Plant Sciences*.
- 77.** Dreissig S, Schiml S, Schindele P, Weiss O, Rutten T, Schubert V, Gladilin E, Mette MF, Puchta H, Houben A. 2017. Live-cell CRISPR imaging in plants reveals dynamic telomere movements. *The Plant Journal* **91**, 565–573.
- 78.** Dubois M, Selden K, Bediée A, Rolland G, Baumberger N, Noir S, Bach L, Lamy G, Granier C, Genschik P. 2018. SIAMESE-RELATED1 Is Regulated Posttranslationally and Participates in Repression of Leaf Growth under Moderate Drought. *Plant Physiology* **176**, 2834–2850.
- 79.** Easterling KA, Pitra NJ, Jones RJ, Lopes LG, Aquino JR, Zhang D, Matthews PD, Bass HW. 2018. 3D Molecular Cytology of Hop (*Humulus lupulus*) Meiotic Chromosomes Reveals Non-disomic Pairing and Segregation, Aneuploidy, and Genomic Structural Variation. *Frontiers in Plant Science* **9**.
- 80.** Edgar BA, Orr-Weaver TL. 2001. Endoreplication Cell Cycles: More for Less. *Cell* **105**, 297–306.
- 81.** Edgar BA, Zielke N, Gutierrez C. 2014. Endocycles: a recurrent evolutionary innovation for post-mitotic cell growth. *Nature Reviews Molecular Cell Biology* **15**, 197–210.

- 82. Fabre G, Garroum I, Mazurek S, Daraspe J, Mucciolo A, Sankar M, Humbel BM, Nawrath C.** 2016. The ABCG transporter PEC1/ABCG32 is required for the formation of the developing leaf cuticle in Arabidopsis. *New Phytologist* **209**, 192–201.
- 83. Fambrini M, Pugliesi C.** 2019. The Dynamic Genetic-Hormonal Regulatory Network Controlling the Trichome Development in Leaves. *Plants* **8**.
- 84. Feng C-M, Qiu Y, Van Buskirk EK, Yang EJ, Chen M.** 2014. Light-regulated gene repositioning in Arabidopsis. *Nature Communications* **5**, 3027.
- 85. Fernández V, Guzmán-Delgado P, Graça J, Santos S, Gil L.** 2016. Cuticle Structure in Relation to Chemical Composition: Re-assessing the Prevailing Model. *Frontiers in Plant Science* **7**.
- 86. Finnegan EJ, Peacock WJ, Dennis ES.** 2000. DNA methylation, a key regulator of plant development and other processes. *Current Opinion in Genetics & Development* **10**, 217–223.
- 87. Fleury D, Himanen K, Cnops G, et al.** 2007. The Arabidopsis thaliana Homolog of Yeast BRE1 Has a Function in Cell Cycle Regulation during Early Leaf and Root Growth. *The Plant Cell* **19**, 417–432.
- 88. Foucher F, Kondorosi E.** 2000. Cell cycle regulation in the course of nodule organogenesis in Medicago. In: Inzé D, ed. *The Plant Cell Cycle*. Dordrecht: Springer Netherlands, 229–242.
- 89. Fox DT, Duronio RJ.** 2013. Endoreplication and polyploidy: insights into development and disease. *Development* **140**, 3–12.
- 90. Fraser PD, Enfissi EMA, Halket JM, Truesdale MR, Yu D, Gerrish C, Bramley PM.** 2007. Manipulation of Phytoene Levels in Tomato Fruit: Effects on Isoprenoids, Plastids, and Intermediary Metabolism. *The Plant Cell* **19**, 3194–3211.
- 91. Frawley LE, Orr-Weaver TL.** 2015. Polyploidy. *Current Biology* **25**, R353–R358.
- 92. Fridman E, Zamir D.** 2003. Functional Divergence of a Syntenic Invertase Gene Family in Tomato, Potato, and Arabidopsis. *Plant Physiology* **131**, 603–609.
- 93. Fu FQ, Mao WH, Shi K, Zhou YH, Yu JQ.** 2010. Spatio-temporal changes in cell division, endoreduplication and expression of cell cycle-related genes in pollinated and plant growth substances-treated ovaries of cucumber. *Plant Biology* **12**, 98–107.
- 94. Fu Y, Rocha PP, Luo VM, Raviram R, Deng Y, Mazzoni EO, Skok JA.** 2016. CRISPR-dCas9 and sgRNA scaffolds enable dual-colour live imaging of satellite

sequences and repeat-enriched individual loci. *Nature Communications* **7**, 11707.

- 95.** Fujimoto S, Matsunaga S. 2016. Chromatin Live Imaging with Genome Editing Techniques: Switching from Scissors to a Lamp. *Cytologia* **81**, 359–362.
- 96.** Fülöp K, Tarayre S, Kelemen Z, Horváth G, Kevei Z, Nikovics K, Bakó L, Brown S, Kondorosi A, Kondorosi E. 2005. Arabidopsis Anaphase-Promoting Complexes: Multiple Activators and Wide Range of Substrates Might Keep APC Perpetually Busy. *Cell Cycle* **4**, 4084–4092.
- 97.** 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**.
- 98.** Gendreau E, Höfte H, Grandjean O, Brown S, Traas J. 1998. Phytochrome controls the number of endoreduplication cycles in the Arabidopsis thaliana hypocotyl. *The Plant Journal* **13**, 221–230.
- 99.** Gendreau E, Orbovic V, Höfte H, Traas J. 1999. Gibberellin and ethylene control endoreduplication levels in the Arabidopsis thaliana hypocotyl. *Planta* **209**, 513–516.
- 100.** Gendreau E, Traas J, Desnos T, Grandjean O, Caboche M, Hofte H. 1997. Cellular Basis of Hypocotyl Growth in Arabidopsis thaliana. *Plant Physiology* **114**, 295–305.
- 101.** Gillaspay G, Ben-David H, Gruissem W. 1993. Fruits: A Developmental Perspective. *The Plant Cell* **5**, 1439–1451.
- 102.** Gonzalez N, Gévaudant F, Hernould M, Chevalier C, Mouras A. 2007. The cell cycle-associated protein kinase WEE1 regulates cell size in relation to endoreduplication in developing tomato fruit. *The Plant Journal* **51**, 642–655.
- 103.** Gonzalez N, Vanhaeren H, Inzé D. 2012. Leaf size control: complex coordination of cell division and expansion. *Trends in Plant Science* **17**, 332–340.
- 104.** Grierson C, Nielsen E, Ketelaarc T, Schiefelbein J. 2014. Root Hairs. *The Arabidopsis Book / American Society of Plant Biologists* **12**.
- 105.** Grob S, Schmid MW, Grossniklaus U. 2014. Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. *Molecular Cell* **55**, 678–693.
- 106.** Guerreiro I, Kind J. 2019. Spatial chromatin organization and gene regulation at the nuclear lamina. *Current Opinion in Genetics & Development* **55**, 19–25.

- 107. Gunesekera B, Torabinejad J, Robinson J, Gillaspay GE.** 2007. Inositol Polyphosphate 5-Phosphatases 1 and 2 Are Required for Regulating Seedling Growth. *Plant Physiology* **143**, 1408–1417.
- 108. Gutensohn M, Nguyen TTH, McMahon RD, Kaplan I, Pichersky E, Dudareva N.** 2014. Metabolic engineering of monoterpene biosynthesis in tomato fruits via introduction of the non-canonical substrate neryl diphosphate. *Metabolic Engineering* **24**, 107–116.
- 109. Gutierrez C.** 2005. Coupling cell proliferation and development in plants. *Nature Cell Biology* **7**, 535–541.
- 110. Gutierrez C.** 2009. The Arabidopsis Cell Division Cycle. *The Arabidopsis Book / American Society of Plant Biologists* **7**.
- 111. Hajheidari M, Farrona S, Huettel B, Koncz Z, Koncz C.** 2012. CDKF;1 and CDKD Protein Kinases Regulate Phosphorylation of Serine Residues in the C-Terminal Domain of Arabidopsis RNA Polymerase II. *The Plant Cell* **24**, 1626–1642.
- 112. Hardie DG.** 2007. AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. *Nature Reviews Molecular Cell Biology* **8**, 774–785.
- 113. Hen-Avivi S, Lashbrooke J, Costa F, Aharoni A.** 2014. Scratching the surface: genetic regulation of cuticle assembly in fleshy fruit. *Journal of Experimental Botany* **65**, 4653–4664.
- 114. Heyman J, Daele HV den, Wit KD, Boudolf V, Berckmans B, Verkest A, Kamei CLA, Jaeger GD, Koncz C, Veylder LD.** 2011. Arabidopsis ULTRAVIOLET-B-INSENSITIVE4 Maintains Cell Division Activity by Temporal Inhibition of the Anaphase-Promoting Complex/Cyclosome. *The Plant Cell* **23**, 4394–4410.
- 115. Heyman J, Polyn S, Eekhout T, De Veylder L.** 2017. Tissue-Specific Control of the Endocycle by the Anaphase Promoting Complex/Cyclosome Inhibitors UVI4 and DEL11[OPEN]. *Plant Physiology* **175**, 303–313.
- 116. Ho T-T, Kwon A-R, Yoon Y-J, Paek K-Y, Park S-Y.** 2016. Endoreduplication level affects flower size and development by increasing cell size in *Phalaenopsis* and *Doritaenopsis*. *Acta Physiologiae Plantarum* **38**, 190.
- 117. Hsieh T-F, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D.** 2009. Genome-Wide Demethylation of Arabidopsis Endosperm. *Science* **324**, 1451–1454.
- 118. Hu S, Liu L, Li S, et al.** 2020. Regulation of fruit ripening by the brassinosteroid biosynthetic gene SICYP90B3 via an ethylene-dependent pathway in tomato. *Horticulture Research* **7**, 1–13.
- 119. Hülkamp M.** 2004. Plant trichomes: a model for cell differentiation. *Nature Reviews Molecular Cell Biology* **5**, 471–480.

- 120.** Hülskamp M, Miséra S, Jürgens G. 1994. Genetic dissection of trichome cell development in Arabidopsis. *Cell* **76**, 555–566.
- 121.** Imai KK, Ohashi Y, Tsuge T, Yoshizumi T, Matsui M, Oka A, Aoyama T. 2006. The A-Type Cyclin CYCA2;3 Is a Key Regulator of Ploidy Levels in Arabidopsis Endoreduplication. *The Plant Cell* **18**, 382–396.
- 122.** Inzé D, De Veylder L. 2006. Cell Cycle Regulation in Plant Development. *Annual Review of Genetics* **40**, 77–105.
- 123.** Isaacson T, Kosma DK, Matas AJ, *et al.* 2009. Cutin deficiency in the tomato fruit cuticle consistently affects resistance to microbial infection and biomechanical properties, but not transpirational water loss. *The Plant Journal* **60**, 363–377.
- 124.** Ishida T, Fujiwara S, Miura K, Stacey N, Yoshimura M, Schneider K, Adachi S, Minamisawa K, Umeda M, Sugimoto K. 2009. SUMO E3 Ligase HIGH PLOIDY2 Regulates Endocycle Onset and Meristem Maintenance in Arabidopsis. *The Plant Cell* **21**, 2284–2297.
- 125.** Jakoby MJ, Falkenhan D, Mader MT, Brininstool G, Wischnitzki E, Platz N, Hudson A, Hülskamp M, Larkin J, Schnittger A. 2008. Transcriptional Profiling of Mature Arabidopsis Trichomes Reveals That NOECK Encodes the MIXTA-Like Transcriptional Regulator MYB106. *Plant Physiology* **148**, 1583–1602.
- 126.** Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, Bubb KL, Trapnell C, Fields S, Queitsch C, Cuperus JT. 2019. Dynamics of Gene Expression in Single Root Cells of Arabidopsis thaliana. *The Plant Cell* **31**, 993–1011.
- 127.** Jégu T, Latrasse D, Delarue M, *et al.* 2013. Multiple functions of Kip-related protein5 connect endoreduplication and cell elongation. *Plant Physiology* **161**, 1694–1705.
- 128.** Jones L, Milne JL, Ashford D, McQueen-Mason SJ. 2003. Cell wall arabinan is essential for guard cell function. *Proceedings of the National Academy of Sciences* **100**, 11783–11788.
- 129.** Jose J, Ghantasala S, Roy Choudhury S. 2020. Arabidopsis Transmembrane Receptor-Like Kinases (RLKs): A Bridge between Extracellular Signal and Intracellular Regulatory Machinery. *International Journal of Molecular Sciences* **21**, 4000.
- 130.** Joubès J, Chevalier C. 2000. Endoreduplication in higher plants. In: Inzé D, ed. *The Plant Cell Cycle*. Dordrecht: Springer Netherlands, 191–201.
- 131.** Joubès J, Chevalier C, Dudits D, Heberle-Bors E, Inzé D, Umeda M, Renaudin J-P. 2000. CDK-related protein kinases in plants. In: Inzé D, ed. *The Plant Cell Cycle*. Dordrecht: Springer Netherlands, 63–76.

- 132.** Joubès J, Phan T-H, Just D, Rothan C, Bergounioux C, Raymond P, Chevalier C. 1999. Molecular and Biochemical Characterization of the Involvement of Cyclin-Dependent Kinase A during the Early Development of Tomato Fruit. *Plant Physiology* **121**, 857–869.
- 133.** Kachanovsky DE, Filler S, Isaacson T, Hirschberg J. 2012. Epistasis in tomato color mutations involves regulation of phytoene synthase 1 expression by cis-carotenoids. *Proceedings of the National Academy of Sciences* **109**, 19021–19026.
- 134.** Karniel U, Koch A, Zamir D, Hirschberg J. 2020. Development of zeaxanthin-rich tomato fruit through genetic manipulations of carotenoid biosynthesis. *Plant Biotechnology Journal* **18**, 2292–2303.
- 135.** Kasili R, Huang C-C, Walker JD, Simmons LA, Zhou J, Faulk C, Hülskamp M, Larkin JC. 2011. BRANCHLESS TRICHOMES links cell shape and cell cycle control in *Arabidopsis thaliana* trichomes. *Development (Cambridge, England)* **138**, 2379–2388.
- 136.** Kato N, Lam E. 2003. Chromatin of endoreduplicated pavement cells has greater range of movement than that of diploid guard cells in *Arabidopsis thaliana*. *Journal of Cell Science* **116**, 2195–2201.
- 137.** Kawade K, Tsukaya H. 2017. Probing the stochastic property of endoreduplication in cell size determination of *Arabidopsis thaliana* leaf epidermal tissue. *PLOS ONE* **12**, e0185050.
- 138.** Kenrick P, Crane PR. 1997. The origin and early evolution of plants on land. *Nature* **389**, 33–39.
- 139.** Khaled A-S, Vernoud V, Ingram GC, Perez P, Sarda X, Rogowsky PM. 2005. Engrailed-ZmOCL1 fusions cause a transient reduction of kernel size in maize. *Plant Molecular Biology* **58**, 123–139.
- 140.** Khosravi S, Schindele P, Gladilin E, Dunemann F, Rutten T, Puchta H, Houben A. 2020. Application of Aptamers Improves CRISPR-Based Live Imaging of Plant Telomeres. *Frontiers in Plant Science* **11**.
- 141.** Klug A, Rhodes D, Smith J, Finch JT, Thomas JO. 1980. A low resolution structure for the histone core of the nucleosome. *Nature* **287**, 509–516.
- 142.** Kocová V, Straková N, Kolarčík V, Rákai A, Mártonfi P. 2016. Endoreduplication as a part of flower ontogeny in *Trifolium pratense* cultivars. *Botanical Studies* **57**, 34.
- 143.** Kondorosi E, Roudier F, Gendreau E. 2000. Plant cell-size control: growing by ploidy? *Current Opinion in Plant Biology* **3**, 488–492.

- 144.** Kudo N, Kimura Y. 2002. Nuclear DNA endoreduplication during petal development in cabbage: relationship between ploidy levels and cell size. *Journal of Experimental Botany* **53**, 1017–1023.
- 145.** Kumar N, Dale R, Kemboi D, Zeringue EA, Kato N, Larkin JC. 2018. Functional Analysis of Short Linear Motifs in the Plant Cyclin-Dependent Kinase Inhibitor SIAMESE. *Plant Physiology* **177**, 1569–1579.
- 146.** Kumar N, Larkin JC. 2017. Why do plants need so many cyclin-dependent kinase inhibitors? *Plant Signaling & Behavior* **12**.
- 147.** Lammens T, Boudolf V, Kheibarshekan L, *et al.* 2008. Atypical E2F activity restrains APC/CCCS52A2 function obligatory for endocycle onset. *Proceedings of the National Academy of Sciences* **105**, 14721–14726.
- 148.** Lang L, Schnittger A. 2020. Endoreplication — a means to an end in cell growth and stress response. *Current Opinion in Plant Biology* **54**, 85–92.
- 149.** Lapitan NLV, Ganai MW, Tanksley SD. 1991. Organization of the 5S ribosomal RNA genes in the genome of tomato. *Genome* **34**, 509–514.
- 150.** Larkins BA, Dilkes BP, Dante RA, Coelho CM, Woo Y, Liu Y. 2001. Investigating the hows and whys of DNA endoreduplication. *Journal of Experimental Botany* **52**, 183–192.
- 151.** Larson-Rabin Z, Li Z, Masson PH, Day CD. 2009. FZR2/CCS52A1 Expression Is a Determinant of Endoreduplication and Cell Expansion in Arabidopsis. *Plant Physiology* **149**, 874–884.
- 152.** Lashbrooke J, Adato A, Lotan O, *et al.* 2015. The Tomato MIXTA-Like Transcription Factor Coordinates Fruit Epidermis Conical Cell Development and Cuticular Lipid Biosynthesis and Assembly. *Plant Physiology* **169**, 2553–2571.
- 153.** Lauber H. 1947. Untersuchungen über das Wachstum der Früchte einiger Angiospermen unter endomitotischer Polyploidisierung. *Österreichische botanische Zeitschrift* **94**, 30–60.
- 154.** Lauria M, Rupe M, Guo M, Kranz E, Pirona R, Viotti A, Lund G. 2004. Extensive Maternal DNA Hypomethylation in the Endosperm of Zea mays. *The Plant Cell* **16**, 510–522.
- 155.** Lee HO, Davidson JM, Duronio RJ. 2009. Endoreplication: polyploidy with purpose. *Genes & Development* **23**, 2461–2477.
- 156.** Lee LR, Wengier DL, Bergmann DC. 2019. Cell-type-specific transcriptome and histone modification dynamics during cellular reprogramming in the Arabidopsis stomatal lineage. *Proceedings of the National Academy of Sciences* **116**, 21914–21924.

- 157.** Leiva-Neto JT, Grafi G, Sabelli PA, Dante RA, Woo Y, Maddock S, Gordon-Kamm WJ, Larkins BA. 2004. A Dominant Negative Mutant of Cyclin-Dependent Kinase A Reduces Endoreduplication but Not Cell Size or Gene Expression in Maize Endosperm. *The Plant Cell* **16**, 1854–1869.
- 158.** Li Z-Y, Li B, Dong A-W. 2012a. The Arabidopsis Transcription Factor AtTCP15 Regulates Endoreduplication by Modulating Expression of Key Cell-cycle Genes. *Molecular Plant* **5**, 270–280.
- 159.** Li L, Sheen J. 2016. Dynamic and diverse sugar signaling. *Current Opinion in Plant Biology* **33**, 116–125.
- 160.** Li Q, Shi X, Ye S, Wang S, Chan R, Harkness T, Wang H. 2016a. A short motif in Arabidopsis CDK inhibitor ICK1 decreases the protein level, probably through a ubiquitin-independent mechanism. *The Plant Journal: For Cell and Molecular Biology* **87**, 617–628.
- 161.** Li S-B, Xie Z-Z, Hu C-G, Zhang J-Z. 2016b. A Review of Auxin Response Factors (ARFs) in Plants. *Frontiers in Plant Science* **7**.
- 162.** Li J, Xu Y, Chong K. 2012b. The novel functions of kinesin motor proteins in plants. *Protoplasma* **249**, 95–100.
- 163.** Liang Y, Hetzer MW. 2011. Functional interactions between nucleoporins and chromatin. *Current Opinion in Cell Biology* **23**, 65–70.
- 164.** Liao C-Y, Smet W, Brunoud G, Yoshida S, Vernoux T, Weijers D. 2015. Reporters for sensitive and quantitative measurement of auxin response. *Nature methods* **12**, 207–210.
- 165.** Li-Beisson Y, Pollard M, Sauveplane V, Pinot F, Ohlrogge J, Beisson F. 2009. Nanoridges that characterize the surface morphology of flowers require the synthesis of cutin polyester. *Proceedings of the National Academy of Sciences* **106**, 22008–22013.
- 166.** Lima M deF, Eloy NB, Pegoraro C, *et al.* 2010. Genomic evolution and complexity of the Anaphase-promoting Complex (APC) in land plants. *BMC Plant Biology* **10**, 254.
- 167.** Liu J, Cong B, Tanksley SD. 2003. Generation and Analysis of an Artificial Gene Dosage Series in Tomato to Study the Mechanisms by Which the Cloned Quantitative Trait Locus fw2.2 Controls Fruit Size. *Plant Physiology* **132**, 292–299.
- 168.** Liu F, Zhang X, Lu C, Zeng X, Li Y, Fu D, Wu G. 2015. Non-specific lipid transfer proteins in plants: presenting new advances and an integrated functional analysis. *Journal of Experimental Botany* **66**, 5663–5681.
- 169.** Liu P, Zhang S, Zhou B, *et al.* 2019. The Histone H3K4 Demethylase JMJ16 Represses Leaf Senescence in Arabidopsis. *The Plant Cell* **31**, 430–443.

- 170.** Magyar Z, Horváth B, Khan S, Mohammed B, Henriques R, De Veylder L, Bakó L, Scheres B, Bögre L. 2012. Arabidopsis E2FA stimulates proliferation and endocycle separately through RBR-bound and RBR-free complexes. *The EMBO Journal* **31**, 1480–1493.
- 171.** Maier AT, Stehling-Sun S, Wollmann H, Demar M, Hong RL, Haubeiß S, Weigel D, Lohmann JU. 2009. Dual roles of the bZIP transcription factor PERIANTHIA in the control of floral architecture and homeotic gene expression. *Development* **136**, 1613–1620.
- 172.** Mambro RD, Ruvo MD, Pacifici E, *et al.* 2017. Auxin minimum triggers the developmental switch from cell division to cell differentiation in the Arabidopsis root. *Proceedings of the National Academy of Sciences* **114**, E7641–E7649.
- 173.** Massonnet C, Tisné S, Radziejowski A, Vile D, De Veylder L, Dauzat M, Granier C. 2011. New Insights into the Control of Endoreduplication: Endoreduplication Could Be Driven by Organ Growth in Arabidopsis Leaves1[W]. *Plant Physiology* **157**, 2044–2055.
- 174.** Mathieu-Rivet E, Gévaudant F, Sicard A, *et al.* 2010. Functional analysis of the anaphase promoting complex activator CCS52A highlights the crucial role of endo-reduplication for fruit growth in tomato. *The Plant Journal* **62**, 727–741.
- 175.** Maunoury N, Redondo-Nieto M, Bourcy M, *et al.* 2010. Differentiation of Symbiotic Cells and Endosymbionts in Medicago truncatula Nodulation Are Coupled to Two Transcriptome-Switches. *PLOS ONE* **5**, e9519.
- 176.** Meents MJ, Watanabe Y, Samuels AL. 2018. The cell biology of secondary cell wall biosynthesis. *Annals of Botany* **121**, 1107–1125.
- 177.** Melaragno JE, Mehrotra B, Coleman AW. 1993. Relationship between Endopolyploidy and Cell Size in Epidermal Tissue of Arabidopsis. *The Plant Cell* **5**, 1661–1668.
- 178.** Menges M, Jager SMD, Grisse W, Murray JAH. 2005. Global analysis of the core cell cycle regulators of Arabidopsis identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control. *The Plant Journal* **41**, 546–566.
- 179.** Monniaux M, Vandenbusche M. 2018. How to Evolve a Perianth: A Review of Cadastral Mechanisms for Perianth Identity. *Frontiers in Plant Science* **9**.
- 180.** Montes RAC, Ranocha P, Martinez Y, *et al.* 2008. Cell Wall Modifications in Arabidopsis Plants with Altered α -L-Arabinofuranosidase Activity. *Plant Physiology* **147**, 63–77.

- 181.** Montiel J, Downie JA, Farkas A, Bihari P, Herczeg R, Bálint B, Mergaert P, Kereszt A, Kondorosi É. 2017. Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. *Proceedings of the National Academy of Sciences* **114**, 5041–5046.
- 182.** Müller S, Han S, Smith LG. 2006. Two Kinesins Are Involved in the Spatial Control of Cytokinesis in *Arabidopsis thaliana*. *Current Biology* **16**, 888–894.
- 183.** Munir S, Mumtaz MA, Ahiakpa JK, Liu G, Chen W, Zhou G, Zheng W, Ye Z, Zhang Y. 2020. Genome-wide analysis of Myo-inositol oxygenase gene family in tomato reveals their involvement in ascorbic acid accumulation. *BMC Genomics* **21**.
- 184.** Nafati M, Cheniclet C, Hernould M, Do PT, Fernie AR, Chevalier C, Gévaudant F. 2011. The specific overexpression of a cyclin-dependent kinase inhibitor in tomato fruit mesocarp cells uncouples endoreduplication and cell growth. *The Plant Journal* **65**, 543–556.
- 185.** Nagl W. 1976. DNA endoreduplication and polyteny understood as evolutionary strategies. *Nature* **261**, 614–615.
- 186.** Nagymihály M, Veluchamy A, Györgypál Z, Ariel F, Jégu T, Benhamed M, Szűcs A, Kereszt A, Mergaert P, Kondorosi É. 2017. Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression. *Proceedings of the National Academy of Sciences* **114**, 4543–4548.
- 187.** Nakamura M, Katsumata H, Abe M, Yabe N, Komeda Y, Yamamoto KT, Takahashi T. 2006. Characterization of the Class IV Homeodomain-Leucine Zipper Gene Family in *Arabidopsis*. *Plant Physiology* **141**, 1363–1375.
- 188.** Nakayama H, Scott IC, Cross JC. 1998. The Transition to Endoreduplication in Trophoblast Giant Cells Is Regulated by the mSNA Zinc Finger Transcription Factor. *Developmental Biology* **199**, 150–163.
- 189.** Nie S, Huang S, Wang S, Cheng D, Liu J, Lv S, Li Q, Wang X. 2017. Enhancing Brassinosteroid Signaling via Overexpression of Tomato (*Solanum lycopersicum*) SIBRI1 Improves Major Agronomic Traits. *Frontiers in Plant Science* **8**.
- 190.** Nitsch L, Kohlen W, Oplaat C, *et al.* 2012. ABA-deficiency results in reduced plant and fruit size in tomato. *Journal of Plant Physiology* **169**, 878–883.
- 191.** Noir S, Bömer M, Takahashi N, Ishida T, Tsui T-L, Balbi V, Shanahan H, Sugimoto K, Devoto A. 2013. Jasmonate Controls Leaf Growth by Repressing Cell Proliferation and the Onset of Endoreduplication while Maintaining a Potential Stand-By Mode. *Plant Physiology* **161**, 1930–1951.
- 192.** Oh D-H, Barkla BJ, Vera-Estrella R, Pantoja O, Lee S-Y, Bohnert HJ, Dassanayake M. 2015. Cell type-specific responses to salinity – the epidermal

bladder cell transcriptome of *Mesembryanthemum crystallinum*. *New Phytologist* **207**, 627–644.

- 193.** Orfila C, Seymour GB, Willats WGT, Huxham IM, Jarvis MC, Dover CJ, Thompson AJ, Knox JP. 2001. Altered Middle Lamella Homogalacturonan and Disrupted Deposition of (1→5)- α -l-Arabinan in the Pericarp of Cnr, a Ripening Mutant of Tomato. *Plant Physiology* **126**, 210–221.
- 194.** Orr-Weaver TL. 2015. When bigger is better: the role of polyploidy in organogenesis. *Trends in Genetics* **31**, 307–315.
- 195.** Orsztynowicz M, Lechniak D, Pawlak P, Kociucka B, Kubickova S, Cernohorska H, Madeja ZE. 2017. Changes in chromosome territory position within the nucleus reflect alternations in gene expression related to embryonic lineage specification. *PLoS ONE* **12**.
- 196.** Paul MJ, Gonzalez-Uriarte A, Griffiths CA, Hassani-Pak K. 2018. The Role of Trehalose 6-Phosphate in Crop Yield and Resilience. *Plant Physiology* **177**, 12–23.
- 197.** Petit J, Bres C, Mauxion J-P, Bakan B, Rothan C. 2017. Breeding for cuticle-associated traits in crop species: traits, targets, and strategies. *Journal of Experimental Botany* **68**, 5369–5387.
- 198.** Petit J, Bres C, Mauxion J-P, Tai FWJ, Martin LBB, Fich EA, Joubès J, Rose JKC, Domergue F, Rothan C. 2016. The Glycerol-3-Phosphate Acyltransferase GPAT6 from Tomato Plays a Central Role in Fruit Cutin Biosynthesis. *Plant Physiology* **171**, 894–913.
- 199.** Phan TD, Bo W, West G, Lycett GW, Tucker GA. 2007. Silencing of the Major Salt-Dependent Isoform of Pectinesterase in Tomato Alters Fruit Softening. *Plant Physiology* **144**, 1960–1967.
- 200.** Pirrello J, Deluche C, Frangne N, *et al.* 2018. Transcriptome profiling of sorted endoreduplicated nuclei from tomato fruits: how the global shift in expression ascribed to DNA ploidy influences RNA-Seq data normalization and interpretation. *The Plant Journal* **93**, 387–398.
- 201.** Pombo A, Dillon N. 2015. Three-dimensional genome architecture: players and mechanisms. *Nature Reviews Molecular Cell Biology* **16**, 245–257.
- 202.** Pontvianne F, Carpentier M-C, Durut N, *et al.* 2016. Identification of Nucleolus-Associated Chromatin Domains Reveals a Role for the Nucleolus in 3D Organization of the *A. thaliana* Genome. *Cell Reports* **16**, 1574–1587.
- 203.** Poulet A, Duc C, Voisin M, Desset S, Tutois S, Vanrobays E, Benoit M, Evans DE, Probst AV, Tatout C. 2017. The LINC complex contributes to heterochromatin organisation and transcriptional gene silencing in plants. *Journal of Cell Science* **130**, 590–601.

- 204.** Ramirez-Parra E, Pozo JC del, Desvoyes B, Sanchez M de la P, Gutierrez C. 2007. E2F–DP Transcription Factors. Annual Plant Reviews Volume 32: Cell Cycle Control and Plant Development. John Wiley & Sons, Ltd, 138–163.
- 205.** Renaudin J-P, Deluche C, Cheniclet C, Chevalier C, Frangne N. 2017. Cell layer-specific patterns of cell division and cell expansion during fruit set and fruit growth in tomato pericarp. *Journal of Experimental Botany* **68**, 1613–1623.
- 206.** Robinson DO, Coate JE, Singh A, Hong L, Bush M, Doyle JJ, Roeder AHK. 2018. Ploidy and Size at Multiple Scales in the Arabidopsis Sepal. *The Plant Cell* **30**, 2308–2329.
- 207.** Rodríguez GR, Muños S, Anderson C, Sim S-C, Michel A, Causse M, Gardener BBM, Francis D, van der Knaap E. 2011. Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity. *Plant Physiology* **156**, 275–285.
- 208.** Roeder AHK, Chickarmane V, Cunha A, Obara B, Manjunath BS, Meyerowitz EM. 2010. Variability in the Control of Cell Division Underlies Sepal Epidermal Patterning in Arabidopsis thaliana. *PLOS Biology* **8**, e1000367.
- 209.** Roeder AHK, Cunha A, Ohno CK, Meyerowitz EM. 2012. Cell cycle regulates cell type in the Arabidopsis sepal. *Development*, dev.082925.
- 210.** Rosa S, Duncan S, Dean C. 2016. Mutually exclusive sense–antisense transcription at FLC facilitates environmentally induced gene repression. *Nature Communications* **7**, 13031.
- 211.** Rösti J, Barton CJ, Albrecht S, Dupree P, Pauly M, Findlay K, Roberts K, Seifert GJ. 2007. UDP-Glucose 4-Epimerase Isoforms UGE2 and UGE4 Cooperate in Providing UDP-Galactose for Cell Wall Biosynthesis and Growth of Arabidopsis thaliana. *The Plant Cell* **19**, 1565–1579.
- 212.** Roux B, Rodde N, Jardinaud M-F, *et al.* 2014. An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *The Plant Journal* **77**, 817–837.
- 213.** Rymen B, Fiorani F, Kartal F, Vandepoele K, Inzé D, Beemster GTS. 2007. Cold Nights Impair Leaf Growth and Cell Cycle Progression in Maize through Transcriptional Changes of Cell Cycle Genes. *Plant Physiology* **143**, 1429–1438.
- 214.** Sabelli PA. 2012. Replicate and die for your own good: Endoreduplication and cell death in the cereal endosperm. *Journal of Cereal Science* **56**, 9–20.
- 215.** Sabelli PA, Dante RA, Nguyen HN, Gordon-Kamm WJ, Larkins BA. 2014. Expression, regulation and activity of a B2-type cyclin in mitotic and endoreduplicating maize endosperm. *Frontiers in Plant Science* **5**.

- 216.** Sabelli PA, Larkins BA. 2009. The contribution of cell cycle regulation to endosperm development. *Sexual Plant Reproduction* **22**, 207–219.
- 217.** Sabelli PA, Liu Y, Dante RA, *et al.* 2013. Control of cell proliferation, endoreduplication, cell size, and cell death by the retinoblastoma-related pathway in maize endosperm. *Proceedings of the National Academy of Sciences* **110**, E1827–E1836.
- 218.** Sanchez J-P, Chua N-H. 2001. Arabidopsis PLC1 Is Required for Secondary Responses to Abscisic Acid Signals. *The Plant Cell* **13**, 1143–1154.
- 219.** Saprunauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Research* **39**, 9275–9282.
- 220.** Satgé C, Moreau S, Sallet E, *et al.* 2016. Reprogramming of DNA methylation is critical for nodule development in *Medicago truncatula*. *Nature Plants* **2**, 1–10.
- 221.** Sawyer IA, Sturgill D, Sung M-H, Hager GL, Dundr M. 2016. Cajal body function in genome organization and transcriptome diversity. *BioEssays : news and reviews in molecular, cellular and developmental biology* **38**, 1197–1208.
- 222.** Scholes DR, Paige KN. 2015. Plasticity in ploidy: a generalized response to stress. *Trends in Plant Science* **20**, 165–175.
- 223.** Scholes DR, Suarez AV, Paige KN. 2013. Can endopolyploidy explain body size variation within and between castes in ants? *Ecology and Evolution* **3**, 2128–2137.
- 224.** Schutter KD, Joubès J, Cools T, *et al.* 2007. Arabidopsis WEE1 Kinase Controls Cell Cycle Arrest in Response to Activation of the DNA Integrity Checkpoint. *The Plant Cell* **19**, 211–225.
- 225.** Schwarz EM, Roeder AHK. 2016. Transcriptomic Effects of the Cell Cycle Regulator LGO in Arabidopsis Sepals. *Frontiers in Plant Science* **7**.
- 226.** Segado P, Domínguez E, Heredia A. 2016. Ultrastructure of the Epidermal Cell Wall and Cuticle of Tomato Fruit (*Solanum lycopersicum* L.) during Development. *Plant Physiology* **170**, 935–946.
- 227.** Shaklai S, Amariglio N, Rechavi G, Simon AJ. 2007. Gene silencing at the nuclear periphery. *The FEBS journal* **274**, 1383–1392.
- 228.** Shen S, Ma S, Liu Y, Liao S, Li J, Wu L, Kartika D, Mock H-P, Ruan Y-L. 2019. Cell Wall Invertase and Sugar Transporters Are Differentially Activated in Tomato Styles and Ovaries During Pollination and Fertilization. *Frontiers in Plant Science* **10**.

- 229.** Shi JX, Adato A, Alkan N, *et al.* 2013. The tomato SISHINE3 transcription factor regulates fruit cuticle formation and epidermal patterning. *New Phytologist* **197**, 468–480.
- 230.** Shinozaki Y, Nicolas P, Fernandez-Pozo N, *et al.* 2018. High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening. *Nature Communications* **9**, 364.
- 231.** Shu Z, Row S, Deng W-M. 2018. Endoreplication: The Good, the Bad, and the Ugly. *Trends in Cell Biology* **28**, 465–474.
- 232.** Simon L, Rabanal FA, Dubos T, *et al.* 2018. Genetic and epigenetic variation in 5S ribosomal RNA genes reveals genome dynamics in *Arabidopsis thaliana*. *Nucleic Acids Research* **46**, 3019–3033.
- 233.** Simon L, Voisin M, Tatout C, Probst AV. 2015. Structure and Function of Centromeric and Pericentromeric Heterochromatin in *Arabidopsis thaliana*. *Frontiers in Plant Science* **6**.
- 234.** Sizani BL, Kalve S, Markakis MN, *et al.* 2019. Multiple mechanisms explain how reduced KRP expression increases leaf size of *Arabidopsis thaliana*. *New Phytologist* **221**, 1345–1358.
- 235.** Sliwinska E, Mathur J, Bewley JD. 2012. Synchronously developing collet hairs in *Arabidopsis thaliana* provide an easily accessible system for studying nuclear movement and endoreduplication. *Journal of Experimental Botany* **63**, 4165–4178.
- 236.** Sliwinska E, Mathur J, Bewley JD. 2015. On the relationship between endoreduplication and collet hair initiation and tip growth, as determined using six *Arabidopsis thaliana* root-hair mutants. *Journal of Experimental Botany* **66**, 3285–3295.
- 237.** Smirnova J, Fernie AR, Steup M. 2015. Starch Degradation. In: Nakamura Y, ed. *Starch: Metabolism and Structure*. Tokyo: Springer Japan, 239–290.
- 238.** Smith S, Galinha C, Desset S, Tolmie F, Evans D, Tatout C, Graumann K. 2015. Marker gene tethering by nucleoporins affects gene expression in plants. *Nucleus* **6**, 471–478.
- 239.** Song Q, Ando A, Jiang N, Ikeda Y, Chen ZJ. 2020. Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in *Arabidopsis* female gametophytes. *Genome Biology* **21**, 178.
- 240.** van Steensel B, Belmont AS. 2017. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780–791.
- 241.** Steiner E, Israeli A, Gupta R, *et al.* 2020. Characterization of the cytokinin sensor TCSv2 in *Arabidopsis* and tomato. *Plant Methods* **16**, 152.

- 242.** Sterken R, Kiekens R, Boruc J, *et al.* 2012. Combined linkage and association mapping reveals CYCD5;1 as a quantitative trait gene for endoreduplication in Arabidopsis. *Proceedings of the National Academy of Sciences* **109**, 4678–4683.
- 243.** Su L, Bassa C, Audran C, Mila I, Cheniclet C, Chevalier C, Bouzayen M, Roustan J-P, Chervin C. 2014. The Auxin SI-IAA17 Transcriptional Repressor Controls Fruit Size Via the Regulation of Endoreduplication-Related Cell Expansion. *Plant and Cell Physiology* **55**, 1969–1976.
- 244.** Sugimoto-Shirasu K, Roberts GR, Stacey NJ, McCann MC, Maxwell A, Roberts K. 2005. RHL1 is an essential component of the plant DNA topoisomerase VI complex and is required for ploidy-dependent cell growth. *Proceedings of the National Academy of Sciences* **102**, 18736–18741.
- 245.** Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE* **6**, e21800.
- 246.** Suzuki T, Ito M, Yoro E, Sato S, Hirakawa H, Takeda N, Kawaguchi M. 2014. Endoreduplication-mediated initiation of symbiotic organ development in *Lotus japonicus*. *Development* **141**, 2441–2445.
- 247.** Szabo Q, Bantignies F, Cavalli G. 2019. Principles of genome folding into topologically associating domains. *Science Advances* **5**, eaaw1668.
- 248.** Szabo Q, Jost D, Chang J-M, *et al.* 2018. TADs are 3D structural units of higher-order chromosome organization in *Drosophila*. *Science Advances* **4**, eaar8082.
- 249.** Takahashi N, Kajihara T, Okamura C, Kim Y, Katagiri Y, Okushima Y, Matsunaga S, Hwang I, Umeda M. 2013. Cytokinins Control Endocycle Onset by Promoting the Expression of an APC/C Activator in Arabidopsis Roots. *Current Biology* **23**, 1812–1817.
- 250.** Takatsuka H, Umeda-Hara C, Umeda M. 2015. Cyclin-dependent kinase-activating kinases CDKD;1 and CDKD;3 are essential for preserving mitotic activity in Arabidopsis thaliana. *The Plant Journal* **82**, 1004–1017.
- 251.** Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD. 2014. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell* **159**, 635–646.
- 252.** Team RC. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- 253.** Terao A, Hyodo H, Satoh S, Iwai H. 2013. Changes in the distribution of cell wall polysaccharides in early fruit pericarp and ovule, from fruit set to early

fruit development, in tomato (*Solanum lycopersicum*). *Journal of Plant Research* **126**, 719–728.

- 254.** Torres Acosta JA, Fowke LC, Wang H. 2011. Analyses of phylogeny, evolution, conserved sequences and genome-wide expression of the ICK/KRP family of plant CDK inhibitors. *Annals of Botany* **107**, 1141–1157.
- 255.** Trapp O, Seeliger K, Puchta H. 2011. Homologs of Breast Cancer Genes in Plants. *Frontiers in Plant Science* **2**.
- 256.** Trivedi P, Nguyen N, Hykkerud AL, Häggman H, Martinussen I, Jaakola L, Karppinen K. 2019. Developmental and Environmental Regulation of Cuticular Wax Biosynthesis in Fleshy Fruits. *Frontiers in Plant Science* **10**.
- 257.** Tsukaya H. 2019. Has the impact of endoreduplication on cell size been overestimated? *New Phytologist* **223**, 11–15.
- 258.** Tuteja N, Tran NQ, Dang HQ, Tuteja R. 2011. Plant MCM proteins: role in DNA replication and beyond. *Plant Molecular Biology* **77**, 537–545.
- 259.** Van Bel M, Diels T, Vancaester E, Kreft L, Botzki A, Van de Peer Y, Coppens F, Vandepoele K. 2018. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Research* **46**, D1190–D1196.
- 260.** Van Leene J, Hollunder J, Eeckhout D, *et al.* 2010. Targeted interactomics reveals a complex core cell cycle machinery in *Arabidopsis thaliana*. *Molecular Systems Biology* **6**, 397.
- 261.** Vanstraelen M, Baloban M, Ines OD, Cultrone A, Lammens T, Boudolf V, Brown SC, Veylder LD, Mergaert P, Kondorosi E. 2009. APC/CCCS52A complexes control meristem maintenance in the *Arabidopsis* root. *Proceedings of the National Academy of Sciences* **106**, 11806–11811.
- 262.** Verkest A, Manes C-L de O, Vercruyse S, Maes S, Schueren EVD, Beeckman T, Genschik P, Kuiper M, Inzé D, Veylder LD. 2005. The Cyclin-Dependent Kinase Inhibitor KRP2 Controls the Onset of the Endoreduplication Cycle during *Arabidopsis* Leaf Development through Inhibition of Mitotic CDKA;1 Kinase Complexes. *The Plant Cell* **17**, 1723–1736.
- 263.** Veylder LD, Larkin JC, Schnittger A. 2011. Molecular control and function of endoreplication in development and physiology. *Trends in Plant Science* **16**, 624–634.
- 264.** Vickovic S, Eraslan G, Salmén F, *et al.* 2019. High-definition spatial transcriptomics for in situ tissue profiling. *Nature Methods* **16**, 987–990.
- 265.** Vilhar B, Kladnik A, Blejec A, Chourey PS, Dermastia M. 2002. Cytometrical Evidence That the Loss of Seed Weight in the miniature1 Seed

Mutant of Maize Is Associated with Reduced Mitotic Activity in the Developing Endosperm. *Plant Physiology* **129**, 23–30.

- 266.** Vinardell JM, Fedorova E, Cebolla A, *et al.* 2003. Endoreduplication Mediated by the Anaphase-Promoting Complex Activator CCS52A Is Required for Symbiotic Cell Differentiation in *Medicago truncatula* Nodules. *The Plant Cell* **15**, 2093–2105.
- 267.** Vlieghe K, Inzé D, Veylder L de. 2007. Physiological Relevance and Molecular Control of the Endocycle in Plants. *Annual Plant Reviews Volume 32: Cell Cycle Control and Plant Development*. John Wiley & Sons, Ltd, 227–248.
- 268.** Vogg G, Fischer S, Leide J, Emmanuel E, Jetter R, Levy AA, Riederer M. 2004. Tomato fruit cuticular waxes and their effects on transpiration barrier properties: functional characterization of a mutant deficient in a very-long-chain fatty acid β -ketoacyl-CoA synthase. *Journal of Experimental Botany* **55**, 1401–1410.
- 269.** Vuolo F, Kierzkowski D, Runions A, *et al.* 2018. LMI1 homeodomain protein regulates organ proportions by spatial modulation of endoreduplication. *Genes & Development* **32**, 1361–1366.
- 270.** Walker JD, Oppenheimer DG, Concienne J, Larkin JC. 2000. SIAMESE, a gene controlling the endoreduplication cell cycle in *Arabidopsis thaliana* trichomes. *Development* **127**, 3931–3940.
- 271.** Wang H, Jones B, Li Z, Frasse P, Delalande C, Regad F, Chaabouni S, Latché A, Pech J-C, Bouzayen M. 2005. The Tomato Aux/IAA Transcription Factor IAA9 Is Involved in Fruit Development and Leaf Morphogenesis. *The Plant Cell* **17**, 2676–2692.
- 272.** Wang X, Peng F, Li M, Yang L, Li G. 2012. Expression of a heterologous SnRK1 in tomato increases carbon assimilation, nitrogen uptake and modifies fruit development. *Journal of Plant Physiology* **169**, 1173–1182.
- 273.** Wang H, Xu X, Nguyen CM, Liu Y, Gao Y, Lin X, Daley T, Kipniss NH, La Russa M, Qi LS. 2018. CRISPR-Mediated Programmable 3D Genome Positioning and Nuclear Organization. *Cell* **175**, 1405-1417.e14.
- 274.** Wu S, Zhang B, Keyhaninejad N, *et al.* 2018. A common genetic mechanism underlies morphological diversity in fruits and other plant organs. *Nature Communications* **9**.
- 275.** Xiao TT, Schilderink S, Moling S, Deinum EE, Kondorosi E, Franssen H, Kulikova O, Niebel A, Bisseling T. 2014. Fate map of *Medicago truncatula* root nodules. *Development* **141**, 3517–3528.
- 276.** Xie K, Minkenberg B, Yang Y. 2015. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences* **112**, 3570–3575.

- 277.** Xu Y, Jin W, Li N, Zhang W, Liu C, Li C, Li Y. 2016. UBIQUITIN-SPECIFIC PROTEASE14 Interacts with ULTRAVIOLET-B INSENSITIVE4 to Regulate Endoreduplication and Cell and Organ Growth in Arabidopsis. *The Plant Cell* **28**, 1200–1214.
- 278.** Yanagisawa M, Desyatova AS, Belteton SA, Mallery EL, Turner JA, Szymanski DB. 2015. Patterning mechanisms of cytoskeletal and cell wall systems during leaf trichome morphogenesis. *Nature Plants* **1**, 1–8.
- 279.** Yang W, Wightman R, Meyerowitz EM. 2017. Cell Cycle Control by Nuclear Sequestration of CDC20 and CDH1 mRNA in Plant Stem Cells. *Molecular Cell* **68**, 1108-1119.e3.
- 280.** Yi D, Kamei CLA, Cools T, *et al.* 2014. The Arabidopsis SIAMESE-RELATED Cyclin-Dependent Kinase Inhibitors SMR5 and SMR7 Regulate the DNA Damage Checkpoint in Response to Reactive Oxygen Species. *The Plant Cell* **26**, 296–309.
- 281.** Yoshiyama KO, Sakaguchi K, Kimura S. 2013. DNA Damage Response in Plants: Conserved and Variable Response Compared to Animals. *Biology* **2**, 1338–1356.
- 282.** Yoshizumi T, Tsumoto Y, Takiguchi T, Nagata N, Yamamoto YY, Kawashima M, Ichikawa T, Nakazawa M, Yamamoto N, Matsui M. 2006. INCREASED LEVEL OF POLYPLOIDY1, a Conserved Repressor of CYCLINA2 Transcription, Controls Endoreduplication in Arabidopsis. *The Plant Cell* **18**, 2452–2468.
- 283.** Yu Y, Steinmetz A, Meyer D, Brown S, Shen W-H. 2003. The Tobacco A-Type Cyclin, Nicta;CYCA3;2, at the Nexus of Cell Division and Differentiation. *The Plant Cell* **15**, 2763–2777.
- 284.** Zemach A, Kim MY, Silva P, Rodrigues JA, Dotson B, Brooks MD, Zilberman D. 2010. Local DNA hypomethylation activates genes in rice endosperm. *Proceedings of the National Academy of Sciences* **107**, 18729–18734.
- 285.** Zentner GE, Henikoff S. 2013. Regulation of nucleosome dynamics by histone modifications. *Nature Structural & Molecular Biology* **20**, 259–266.
- 286.** Zhang H, Lang Z, Zhu J-K. 2018. Dynamics and function of DNA methylation in plants. *Nature Reviews Molecular Cell Biology* **19**, 489–506.
- 287.** Zhang Y, Primavesi LF, Jhurrea D, Andralojc PJ, Mitchell RAC, Powers SJ, Schluepmann H, Delatte T, Wingler A, Paul MJ. 2009. Inhibition of SNF1-Related Protein Kinase1 Activity and Regulation of Metabolic Pathways by Trehalose-6-Phosphate. *Plant Physiology* **149**, 1860–1871.

- 288.** Zhang H, Zheng R, Wang Y, Zhang Y, Hong P, Fang Y, Li G, Fang Y. 2019. The effects of Arabidopsis genome duplication on the chromatin organization and transcriptional regulation. *Nucleic Acids Research* **47**, 7857–7869.
- 289.** Zhao Y, Garcia BA. 2015. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harbor Perspectives in Biology* **7**, a025064.
- 290.** Zhao X, Yuan X, Chen S, Meng L, Fu D. 2018. Role of the tomato TAGL1 gene in regulating fruit metabolites elucidated using RNA sequence and metabolomics analyses. *PLOS ONE* **13**, e0199083.
- 291.** Zhao T, Zhan Z, Jiang D. 2019. Histone modifications and their regulatory roles in plant development and environmental memory. *Journal of Genetics and Genomics* **46**, 467–476.
- 292.** Zheng T, Nibau C, Phillips DW, Jenkins G, Armstrong SJ, Doonan JH. 2014. CDKG1 protein kinase is essential for synapsis and male meiosis at high ambient temperature in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 2182–2187.
- 293.** Zhimulev IF, Koryakov DE. 2009. Polytene Chromosomes. eLS. American Cancer Society, .
- 294.** Zhu J-Y, Sae-Seaw J, Wang Z-Y. 2013. Brassinosteroid signalling. *Development* **140**, 1615–1620.
- 295.** Ziv C, Zhao Z, Gao YG, Xia Y. 2018. Multifunctional Roles of Plant Cuticle During Plant-Pathogen Interactions. *Frontiers in Plant Science* **9**.

ANNEXES

I. *Presentation*

1. *Oral presentations*

Edouard Tourdot, Norbert Bollier, Julien Pirrello, Frederic Gevaudant, Nathalie Frangne, Elie Maza, Anis Djari, Mickaël Bourge, Mohammed Zouine, Christian Chevalier, Nathalie Gonzalez : Understanding the transcriptional regulation of endoreduplication for tomato fruit growth regulation

Tsukuba Global Science Week, Tsukuba (Japan) September 2018.

2. *Poster presentation*

Edouard Tourdot, Norbert Bollier, Julien Pirrello, Frederic Gevaudant, Nathalie Frangne, Elie Maza, Anis Djari, Mickaël Bourge, Mohammed Zouine, Christian Chevalier, Nathalie Gonzalez : Understanding the transcriptional regulation of endoreduplication for tomato fruit growth regulation

Tsukuba Global Science Week, Tsukuba (Japan) September 2018.

Edouard Tourdot, Norbert Bollier, Julien Pirrello, Elie Maza, Anis Djari, Mickaël Bourge, Nathalie Gonzalez : *Spatiotemporal distribution of ploidy levels and ploidy specific transcriptome during Tomato fruit development*

Plant Organ Growth Symposium 2019, Bordeaux (France) April 2019.

Edouard Tourdot, Norbert Bollier, Julien Pirrello, Elie Maza, Anis Djari, Mickaël Bourge, Nathalie Gonzalez : *Spatiotemporal distribution of ploidy levels and ploidy specific transcriptome during Tomato fruit development*

Doctoral school days, Bordeaux (France) June 2019.

II. *Disciplinary trainings*

1. *Workshop Plant Regeneration, Transformation and genome editing, April 2018.*

Program :

Lectures :

- Transformation of monocot and dicot crops
- Protoplast transformation and regeneration
- CRISPR/Cas9 methods, vectors, cloning, gRNA design
- Analysis of edited alleles phenotypes in multiploid species
- Sequencing of Mutated genes
- Rules and regulations of New Breeding Technologies

Practical course:

- Preparation of explants; protoplasts, cotyledons, immature and mature embryos
- GFP analyses in live tissues, GUS staining
- Agrobacterium transformation, biolistics
- Detection of CRISPR/Cas9 mutations

2. Workshop 3D FisH and image analysis, October 2018.

This Training school included practical courses in DNA- and RNA-Fluorescent In Situ Hybridization (FISH) as well as a training in the softwares OMERO, ImageJ and Imaris dedicated to image analysis. It also included formal lecture from Pr. Paul Fransz (Faculty of Science, Swammerdam Institute for Life Sciences, Netherland), Dr. Stefanie Rosa (Swedish University of Agricultural Sciences, Uppsala, Sweden), Pr. Hank Bass (Department of Biological Science, Florida State University, USA), Dr. Monica Pradillo (Complutense University of Madrid, Spain), Elise Bertrand (Zeiss, France) and Dr. Christophe Chamot (France).

3. Workshop From Pictures to Numbers, September 2019

The aim of this course is to lower the threshold for researchers who use microscopy on a routine basis and/or demand image analysis for their own research by:

- a) providing fundamental knowledge on the most important image properties, basic processing algorithms and typical analyses, and
- b) by giving direct hands-on training on how to perform very common tasks in open source software ranging from simple image annotation to more advanced phenotypic analyses.

An overview of different image analysis platforms and an introduction to basic scripting capabilities should further enable the participants to select or devise apt solutions for their image analysis needs. The primary goal is to familiarize PhD students, post-docs and lab technicians, who work with microscopy images of biological samples and have limited or no experience with the fundamentals of biological image analysis.

After this course participants should be able to:

- Know important image properties and processing algorithms
- Annotate and present images in a publication-ready format
- Know which analytical routine to select and apply
- Compose basic image analysis workflows
- Gain proficiency in basic scripting language

The organizing committee is composed of the following members:

Prof. Winnok De Vos (University of Antwerp), Prof. Sebastian Munck (University of Leuven), Dr. Bob Asselbergh (University of Antwerp), Prof. Kevin Braeckmans (University of Ghent), Dr. Herlinde De Keersmaecker (University of Ghent), Prof. Jelle Hendrix (University of Hasselt).

4. Bioinformatic school AVIESAN-INSERM: Initiation to processing of NGS data.

This training is for Biologist implicated in project that imply NGS. The school focused on new technology and three practical courses parallelly (RNA-seq, CHIP-seq/ATAC-seq, Variant DNA-seq). Introduction to data integration, single cell genomic and long reads technology will be dispensed. The aim of the school is to give the bases to be

able to use bioinformatic tool and understand the results. The student has a personal training on their project in order to be able to do alone the first step of the analysis.

III. *Transversal training*

- 1. *“Anglais scientifique 2° session Module 1 - From Research to Publication”, “Anglais scientifique 2ème session Module 4 Conference Talk” and, “Anglais scientifique 2ème session Module 5 - From text to talk”.***

I am concentrating my transversal training on that point by doing three trainings named “Anglais scientifique 2ème session Module 4 Conference Talk”, “Anglais scientifique 2° session Module 1 - From Research to Publication” and “Anglais scientifique 2ème session Module 5 - From text to talk”. These three trainings should help me to learn about rhetorical methods, vocabulary, grammar and also about participating to debates, knowing how to ask a question without hesitations and for networking sessions.

- 2. *PAPIRUS (Parcours d’accompagnement pour publier et communiquer en science)***

- 3. *EDEN (Ecole des doctorants et encadrants de l’INRA)***

Together with Nathalie, we participated to EDEN14 (école des doctorants et encadrants de l'INRA). This training helped us to identify the difficulties presented before and how to improve team work together with Nathalie but also other co-workers.

- 4. *MOOC-INTRODUCTION A LA STATISTIQUE AVEC R -SESSION 11***

I refreshed my statistical basis from an INRA training in March 2018. This was coupled with another training by following a MOOC (end of March 2019) done by two professors from “Université Paris Sud” (Introduction à la statistique avec R). Taken together, these two trainings helped me understanding and learning on how to use R to perform statistical analysis in order to analyze my data with more reliability.

- 5. *Find a post doc abroad***

This was a presentation of the different type of post-doc and options that exists in order to find a post-doc.

IV. *Complementary activity*

Teaching with a pedagogic team in practical course for first year students and animated a course about genome editing both theoretical and practical for Master2 students.

Titre : Distribution spatio-temporelle des niveaux de ploïdie et du transcriptome associés durant le développement du fruit de Tomate

Résumé :

L'endoréduplication, qui est la source majeure d'endopolyploïdie chez les Plantes, est un processus cellulaire durant lequel la quantité d'ADN nucléaire (ploïdie) est augmentée par une succession de réplifications du génome sans division cellulaire. L'endoréduplication est le résultat de la capacité des cellules à transformer le classique cycle cellulaire en un cycle cellulaire partiel, l'endocycle, où la synthèse de l'ADN intervient indépendamment de la mitose. Chez les plants, l'endoréduplication est présente dans de nombreux organes et types cellulaires et notamment ceux ayant un fort intérêt agronomique, comme la partie charnue (péricarpe) du fruit de Tomate dans lequel les niveaux de ploïdie sont important. A la fin du développement du fruit de tomate, le péricarpe est composé d'une population hétérogène de cellules dont les niveaux de ploïdie et la tailles varient. De plus, à ce stade, des gènes présentent des profils d'expression spécifiquement lié aux niveaux de ploïdie. Cependant, peu de données sont disponible à propos de l'engagement et la progression de l'endoreduplication dans les cellules durant le développement du fruit de tomate ou son effet sur l'expression des gènes. Afin d'expliquer pourquoi une population hétérogène de cellule co-existent dans le péricarpe et comme base de description des régulations transcriptionnelles par l'endoréduplication dans le péricarpe, la thèse a eu d'étudier cette population hétérogène de cellule afin d'établir les bases de détermination des programmes génétiques mis en place par l'endoreduplication. Donc, j'ai cartographié le dynamisme des niveaux de ploïdie et de l'expression des gènes en fonction des niveaux de ploïdie dans le péricarpe au cours du développement précoce du fruit, à 6-, 9- et 12 jours après anthèse. Les données ont montré qu'un doublement de l'expression était présent en relation avec le doublement de ploïdie et l'endoreduplication est déjà bien en place, organisée en gradient dans le péricarpe à 6DPA. En utilisant les données produites, une carte virtuelle des niveaux de ploïdie et de l'expression des gènes sera établie au cours de du développement du péricarpe du fruit de Tomate.

Mots clés : Endoreduplication, Tomate, Transcriptome, Croissance, Fish

Title : Spatiotemporal distribution of ploidy levels and ploidy specific transcriptome during Tomato fruit development

Abstract :

Endoreduplication, during which cells increase their DNA content through successive rounds of full genome replication without cell division, is the major source of endopolyploidy in higher plants. Endoreduplication results from the capacity of cells to modify their classical cell cycle into a partial cell cycle, or endocycle, where DNA synthesis occurs independently from mitosis. In plants, endoreduplication is found in numerous organs and cell types and particularly in organs or tissues of agronomical high value, such as the fleshy fruit (pericarp) of tomato presenting high ploidy levels. At the end of tomato fruit development, the pericarp is composed of a heterogeneous population of cells in term of ploidy levels and cell sizes. In addition, at this stage, ploidy-specific gene expression profiles have been found. However, little is known about the onset and progression of endoreduplication during tomato fruit growth and its consequences on the regulation of cell size and gene expression. the aim of my PhD project was to study the heterogeneous population of cells that co-exists within the pericarp tissue and as a basis to dissect the genetic regulation of pericarp endoreduplication. Thus, I mapped the ploidy and transcriptome dynamics of the tomato fruit pericarp during early growth at three stages, 6-, 9- and 12 days post-anthesis (dpa) during fruit growth. The data showed a global shift of expression in relation to ploidy level as well as ploidy specific gene expression and that ploidy levels are already gradiently distributed in the pericarp. Using this data, a virtual map of ploidy distribution and gene expression will be done for the early development of Tomato fruit pericarp.

Keywords : Endoreduplication, Tomato, Transcriptome, Growth, Fish

UMR1332

Biologie du Fruit et Pathologie, 1332, 71 Avenue Edouard Bourlax, 33140 Villenave-d'Ornon