



Dynamique des corps lipidiques dans la graine d'*Arabidopsis thaliana*

Ghassen Trigui

► To cite this version:

Ghassen Trigui. Dynamique des corps lipidiques dans la graine d'*Arabidopsis thaliana*. Sciences agricoles. Université Paris Sud - Paris XI, 2014. Français. NNT : 2014PA112041 . tel-01011150

HAL Id: tel-01011150

<https://theses.hal.science/tel-01011150>

Submitted on 23 Jun 2014

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.

UNIVERSITÉ PARIS-SUD

ECOLE DOCTORALE SCIENCES DU VÉGÉTAL
UMR 1318, INSTITUT JEAN PIERRE BOURGIN

DISCIPLINE : BIOLOGIE

THÈSE DE DOCTORAT

Soutenue le 5 Mars 2014 par

Ghassen Trigui

Dynamique des corps lipidiques
dans la graine d'*Arabidopsis thaliana*

Directeur de thèse :	M. Bertrand Dubreucq	Directeur de recherche (INRA, Versailles)
Co-directeur de thèse :	M. Alain Trubuil	Ingénieur de recherche (INRA, Jouy-en-josas)

Composition du jury :

Rapporteurs :	M. Alain Zachowski	Directeur de recherche (UPMC, Paris)
	M. Charles Kervrann	Directeur de recherche (INRIA, Rennes)
Examineurs :	M. Jacques Fattaccioli	Maître de conférences (ENS, Paris)
	M. Olivier Martin	Professeur (Université Paris Sud, Orsay)

Remerciements

Je dédie cette thèse à mes chers parents et à mes deux frères qui tiennent une place immense dans mon coeur. Papa, tu es une vraie école de la vie, je ne cesse d'apprendre tous les jours avec toi. Maman, une femme aussi adorable que toi je n'en connais pas, tu as toujours été là pour moi, et à aucun moment tu n'as cessé de me couvrir de ta tendresse.

Je souhaite remercier mon directeur de thèse Bertrand Dubreucq pour avoir proposé, accepté et encadré cette thèse. Grâce à lui j'ai pu valoriser mes travaux en m'assurant les meilleures conditions de financement et en me donnant l'opportunité de participer à plusieurs manifestations scientifiques.

J'aimerais aussi remercier mon co-directeur de thèse, Alain Trubuil, pour la confiance qu'il m'a témoigné du début à la fin du travail, pour sa disponibilité à comprendre et à communiquer et sans oublier ses précieuses intuitions. Sincèrement, grâce à lui j'ai pu apprendre beaucoup de choses dont certaines fort utiles pour mes travaux académiques bien sûr, mais aussi des choses importantes sur le plan personnel.

J'aimerais également remercier Martine Miquel pour son effort considérable dans l'acquisition des données sans lesquelles la thèse n'aurait pas pu aboutir. Je voudrais la remercier aussi pour avoir suivi mes travaux pendant la durée de ma thèse. Je n'oublie pas enfin son aide précieuse dans la relecture et la correction de ma thèse. La justesse de ses critiques a été très constructive et utile. Mises à part ses contributions de fond, il a eu la pénible tâche de corriger mon français maladroît ! Je lui exprime ma très profonde gratitude.

J'aimerais également remercier Béatrice Laroche et Olivier David pour avoir suivi mes travaux à travers plusieurs réunions. Leurs conseils et remarques, toujours pertinents, m'ont permis de confronter et ajuster mes travaux à des problématiques pratiques et réelles.

J'adresse mes vifs remerciements à Monsieur Alain Zachowski et Monsieur Charles Kervrann pour avoir accepté d'être rapporteurs de ce travail. Je tiens également à remercier Monsieur Jacques Fattaccioli et Monsieur Olivier Martin d'avoir accepté de participer au jury de cette thèse. Je voudrais adresser un remerciement particulier à Monsieur Olivier Martin pour m'avoir enseigné les travaux pratiques de modélisation, une occasion à travers laquelle j'ai pu échanger avec lui autour de mon travail, et j'ai pu bénéficier de ses critiques et propositions.

Je ne peux certainement pas oublier de remercier le laboratoire MIAJ de l'INRA de Jouy-en-Josas dans son ensemble pour son soutien logistique et matériel qui ne fait jamais défaut pour les doctorants.

Table des matières

Liste des Abréviations	ix
Introduction générale	1
1 État de l'art sur les membranes biologiques	7
1.1 Introduction	7
1.2 Projet d'article	8
1.3 Conclusion	21
2 Traitement des données, extraction et analyse de l'information	23
2.1 Introduction	23
2.2 Objectif	24
2.3 Méthodologie	24
2.3.1 Acquisition des données	24
2.3.2 Traitement des données	29
2.4 Résultats	37
2.4.1 Résumé	37
2.5 Résultats complémentaires	69
2.6 Conclusion	73
3 Analyse statistique	75
3.1 Introduction	75
3.2 Objectif	76
3.3 Méthodologie	76
3.4 Résultats	78
3.4.1 Résumé	78
3.4.2 Article	79
3.5 Résultats complémentaires	86
3.6 Conclusion	87
4 Modélisation de la dynamique des corps lipidiques	89
4.1 Introduction	89
4.2 Equation d'agrégation de Smoluchowski	90
4.2.1 Existence, unicité de solution pour l'équation continue	91
4.2.2 Existence, unicité de solution pour l'équation discrète	93
4.2.3 Approximation des solutions de l'équation continue	94
4.3 Modèles d'agrégation existants	94
4.3.1 Modèles simples et solutions exactes	94
4.3.2 Applications de l'équation de Smoluchowski	95
4.4 Modèle de coalescence des corps lipidiques	95

4.5	Méthode d'estimation des paramètres	100
4.5.1	Principe de l'estimation bayésienne	100
4.5.2	Application aux modèles de dynamique de populations des corps lipidiques	101
4.6	Discrétisation et implémentation numérique	103
4.7	Résultats	106
4.7.1	Modèle constant	106
4.7.2	Modèle génotypique	110
4.7.3	Modèle multiplicatif	113
4.7.4	Modèle génotypique/spatial	116
4.7.5	Modèle génotypique/spatial modulé	119
4.8	Conclusion	123
5	Conclusion générale et perspectives	125
5.1	Résumé des principaux résultats produits	125
5.1.1	Mise en place d'un pipeline de traitement d'images	125
5.1.2	Analyse exploratoire des données ("Microscope 1")	126
5.1.3	Régression quantile ("Microscope 2")	126
5.1.4	Modélisation de la dynamique de coalescence des corps lipi- diques ("Microscope 3")	126
5.2	Qu'aurions pu nous changer pendant la thèse ?	127
5.3	Que faudrait-il faire après ce travail (perspectives) ?	128
5.3.1	Etude des caractéristiques biophysiques de la membrane des corps lipidiques	128
5.3.2	Validation du modèle de la dynamique de coalescence des corps lipidiques	129
5.3.3	Faire de nouvelles hypothèses biologiques sur la base du modèle	131
5.4	Conclusion générale	131
	Bibliographie	133
A	Annexes	147
A.1	Scripts et algorithmes	147
A.1.1	Implémentation R pour les procédures d'analyse de données	147
A.1.2	Algorithme de lignes de partage des eaux par immersion	156
A.1.3	Algorithme de la méthode d'extraction de l'enveloppe convexe	157
A.2	Régularité des solution de l'équation de Smoluchowski	157
A.3	Figures	158
A.4	Article de conférence	166

Table des figures

1	Structure et composition d'un corps lipidique	2
2	Modèle de la conformation d'une oléosine de 18 kDa à la surface du corps lipidique	3
3	Modèle de composition en acides aminés d'une oléosine de Maïs de 18-kDa à la surface du corps lipidique	4
4	Modèle de Beisson de synthèse de corps lipidiques de la plante oléagineuse	5
1.1	Structures géométriques résultantes de l'auto-assemblage des molécules lipidiques	10
1.2	Principaux comportements dynamiques des molécules de lipides dans une bicouche	12
1.3	Modes de déformations d'une membrane	13
1.4	Différents mécanismes de courbure de la membrane	14
1.5	Formes obtenues par minimisation de l'énergie de courbure.	17
1.6	Modèles de biogenèse des corps lipidiques.	18
2.1	Matériel végétal utilisé.	25
2.2	Schéma du microscope confocal	26
2.3	Spectres d'émission de la GFP et le Rouge de Nil	27
2.4	Acquisition des piles d'images à partir de différents échantillons	28
2.5	Visualisation de corps lipidiques en microscopie confocale en double marquage.	29
2.6	Principe d'immersion des bassins versants en ligne de partage des eaux	30
2.7	Images intermédiaires de la procédure de segmentation par lignes de partage des eaux.	32
2.8	Chaîne de traitement d'image (1)	32
2.9	Principe du double seuillage pour la détection d'objets	33
2.10	Chaîne de traitement d'image (2)	34
2.11	Enveloppe convexe d'une image de corps lipidiques en 2D.	35
2.12	Concept de cellules de Voronoï déterminées pour les corps lipidiques.	36
2.13	Relation lineaire entre le volume V (axe des x) et sa fraction locale ϕ (axe des y) dans l'échelle log-log	71
2.14	Suite figure 2.13	72
3.1	Valeurs des paramètres associés à l'évolution de la taille de corps lipidiques estimés par régression quantile et modèle lineaire	84
3.2	Valeurs des volumes de corps lipidiques estimés par régression quantile et modèle linéaire pour chaque génotype en fonction du jour de développement	85

3.3	Test de Kolmogorov-Smirnov de comparaison entre les distributions de volumes de corps lipidiques	87
4.1	Coéfficients de sphéricités moyens des corps lipidiques pour différents génotypes à différents stades du développement.	97
4.2	Valeurs du taux d'agrégation pour des noyaux multiplicatifs et additifs.	98
4.3	Comparaison entre solution exacte et approximation numérique pour un noyau constant.	105
4.4	Comparaison entre solution exacte et approximation numérique pour un noyau additif.	105
4.5	Comparaison entre solution exacte et approximation numérique pour un noyau multiplicatif.	106
4.6	Courbes d'ajustement du modèle constant (1)	108
4.7	Suite figure 4.6	109
4.8	Courbes d'ajustement du modèle génotypique (2).	111
4.9	Suite figure 4.8	112
4.10	Courbes d'ajustement du modèle multiplicatif (3).	114
4.11	Suite figure 4.10	115
4.12	Courbes d'ajustement du modèle génotypique/spatial (4).	117
4.13	Suite figure 4.12	118
4.14	Courbes d'ajustement du modèle génotypique/spatial modulé (5).	121
4.15	Suite figure 4.14	122
A.1	Histogrammes de volume de corps lipidiques pour chaque génotype et à chaque du stade du développement.	159
A.1	Suite figure A.1	160
A.1	Suite figure A.1	161
A.2	Histogrammes des distances des corps lipidiques au plus proche voisin selon le génotype et le temps.	162
A.3	Comparaison entre la durée de simulation du modèle de Smoluchowski utilisant des fonctions Matlab, et celle utilisant C-Mex	163
A.4	Méthodes d'imagerie en microscopie confocale	164
A.5	Thechnique d'aspiration par micropipette	165

Liste des tableaux

1.1	Evolution des connaissances de la structure des membranes biologiques	9
1.2	Exemples de systèmes biologiques modélisés en utilisant la théorie Canham-Helfrich	15
2.1	Nomenclatures des oléosines graine-spécifiques d' <i>A. thaliana</i> utilisées dans de précédentes publications.	37
3.1	Exemples d'applications de la regression quantile en biologie et écologie.	76
3.2	Paramètres estimés pour chacun des estimateur OLS et QR	86
4.1	Expressions des noyaux de coalescence pour différents mécanismes. .	95
4.2	Erreur quadratique moyenne entre la solution exacte et l'approximation numérique.	104
4.3	Différents modèles de coalescence développés pour la dynamique des corps lipidiques, ainsi que le mode d'estimation des paramètres utilisé.	107
4.4	Paramètres d'ajustement du modèle constant (Modèle 1).	107
4.5	Paramètres d'ajustement du modèle génotypique (Modèle 2).	110
4.6	Paramètres d'ajustement du modèle multiplicatif (Modèle 3).	113
4.7	Paramètres d'ajustement du modèle génotypique/spatial (Modèle 4). .	116
4.8	Paramètres d'ajustement du modèle génotypique/spatial modulé (Modèle 5).	120

Liste des Abréviations

ADE	Area Difference Elasticity
CDE	Clathrin Dependent Endocytosis
EL	Euler-Lagrange
ER	Endoplasmic Reticulum
FRAP	Fluorescence Recovery After Photobleaching
FRET	Förster Resonance Energy Transfer
GFP	Green Fluorescent Protein
GM	General Model
IJPB	Institut Jean-Pierre Bourgin
INRA	Institut National de la Recherche Agronomique
LASER	Light Amplification by Stimulated Emission of Radiation
LPE	Lignes de Partage des Eaux
LS	Level Sets
MIA	Mathématiques et Informatique Appliquées
MM	Minimal Model
OLS	Ordinary Least Squares
PF	Phase Field
PLs	Phospholipides
QR	Quantile Regression
RBCs	Red Blood Cells
SC	Spontaneous Curvature
STED	stimulated-emission-depletion
TAGs	Triacylglycérol
TCL	Tool Command Language

Introduction générale

Les huiles végétales sont largement utilisées pour l'alimentation, tant humaine qu'animale et comme sources de molécules variées pour de nombreuses industries chimiques, voire comme biocarburants. Leur similarité structurale avec les hydrocarbures d'origine fossile en fait une réelle alternative pour les produits pétroliers. De plus, il s'agit d'une ressource renouvelable. Les huiles végétales sont donc un véritable enjeu économique pour tous les acteurs impliqués dans les différentes filières, tant pour l'industrie agro-alimentaire que pour des usages non alimentaires (industries pharmaceutiques, cosmétiques, etc.). Sous nos climats, les graines oléagineuses sont la principale source d'huile végétale (plus de 2 millions d'hectares de cultures oléagineuses, colza, tournesol et soja, en 2007 en France). L'intérêt croissant pour cette ressource nécessite entre autres d'augmenter le rendement final en huile pour augmenter la rentabilité des filières ou bien de changer la qualité de l'huile pour des besoins industriels. L'effort peut donc porter sur : 1) l'amélioration du rendement en graines par hectare, 2) l'augmentation de la teneur en huile dans la graine 3) l'amélioration du rendement d'extraction de l'huile 4) la variation en contenu qualitatif de l'huile dans la graine.

- **Teneur en huile** : Actuellement, des travaux sont engagés pour mieux comprendre les bases génétiques et moléculaires de l'élaboration de la teneur en huile chez *Arabidopsis* afin d'identifier des marqueurs utilisables en sélection. Par ailleurs, l'exploration de collections de ressources génétiques devrait permettre d'identifier des génotypes à forte teneur en huile dans la graine.
- **Extraction de l'huile** : Il est souhaitable d'améliorer les procédés d'extraction de l'huile, qui comprennent de nombreuses étapes (*e.g.* aplatissage, cuisson, pressage, extraction par solvant de l'huile résiduelle à partir du tourteau de pressage, désolvatation et toastage des tourteaux, dégommage des huiles), pour les rendre moins coûteux en énergie, ne nécessitant plus de solvant et moins dommageables pour les protéines de réserve tout en préservant, voire en augmentant, leur efficacité.

Un des grands objectifs actuels est donc de mieux comprendre la formation des lipides de réserve et leur stockage afin, dans un deuxième temps, de le modifier pour faciliter l'extractibilité ou le remplissage de l'huile dans la graine.

La quasi totalité des êtres vivants, procaryotes et eucaryotes, stockent leurs réserves lipidiques dans des organelles intracellulaires spécialisés appelés adiposomes, corps lipidiques ou aussi, chez les plantes, oléosomes. Ces organelles ont une structure remarquablement conservée c'est-à-dire une matrice de lipides neutres entourée d'une monocouche de phospholipides (PLs) dans laquelle différents types de protéines peuvent être insérés. Si le type de lipides neutres est assez restreint (triacylglycérols (TAGs), stérols, esters de stérol, esters de cires), selon l'organisme le complément protéique est extrêmement varié, tant en nombre qu'en type de protéines (enzymes, protéines structurales).

Les plantes oléagineuses stockent leur huile essentiellement dans leurs graines et ces réserves lipidiques serviront de source d'énergie lors de la germination et la post-germination. Les corps lipidiques des graines matures (Figure 1) sont de petite taille (0.5 à 3 μm de diamètre), très stables et ne coalescent pas, contrairement à des préparations de liposomes artificiels¹ formés uniquement de lipides neutres qui coalescent rapidement [Leptihn 2013]. Il est admis que la stabilité est apportée par les caractéristiques de la demi-membrane du corps lipidique et, notamment, sa composition en protéines.

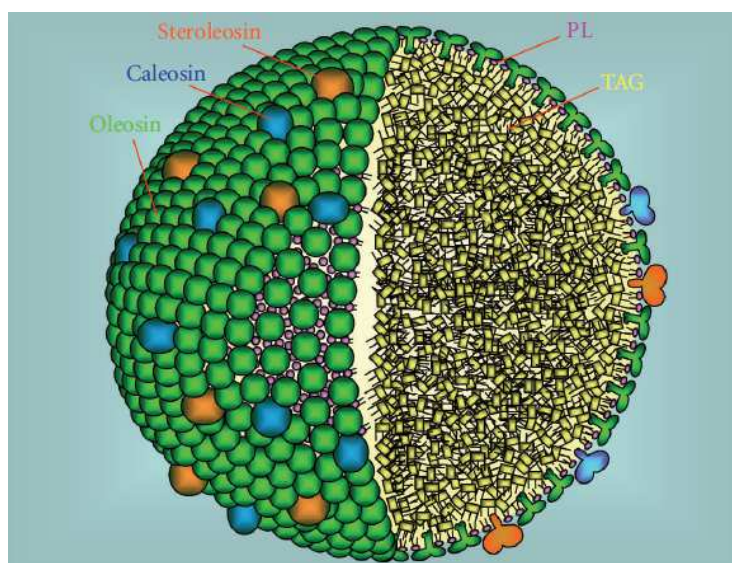


FIGURE 1 : Structure et composition d'un corps lipidique : Un corps lipidique est formé d'une matrice de TAGs entourée d'une monocouche de phospholipides dans laquelle trois types de protéines sont insérées : l'oléosine (la protéine majeure du corps lipidique), la caléosine et la stéroléosine (D'après [Tzen 2012]).

Vance et Huang [Vance 1987] ont mis en évidence chez le maïs la présence d'une protéine avec une très longue chaîne hydrophobe d'environ 70 acides aminés (la plus longue chaîne hydrophobe identifiée chez des protéines naturelles jusqu'à maintenant) associée au corps lipidique. Cette protéine a été nommée oléosine. Plus tard, d'autres isoformes d'oléosines ont été identifiées dans les corps lipidiques de diverses espèces telles que le colza, le soja, le tournesol et *A. thaliana* [Murphy 1989a, Murphy 1991, Cummins 1992, Jolivet 2009]. La séquence d'acides aminés des oléosines comprend trois parties : un domaine central hydrophobe (ou "knob") et deux domaines N- et C-terminaux amphiphiles [Tzen 1992b, Huang 1996, Katavic 2006] (Figures 2, 3).

1. Le liposome artificiel est une vésicule artificielle microscopique, à membrane lipidique, utilisée comme modèle d'étude des membranes biologiques et faisant l'objet de recherches pour l'introduction de substances dans les cellules d'un organisme. (D'après le dictionnaire LAROUSSE de la langue française)

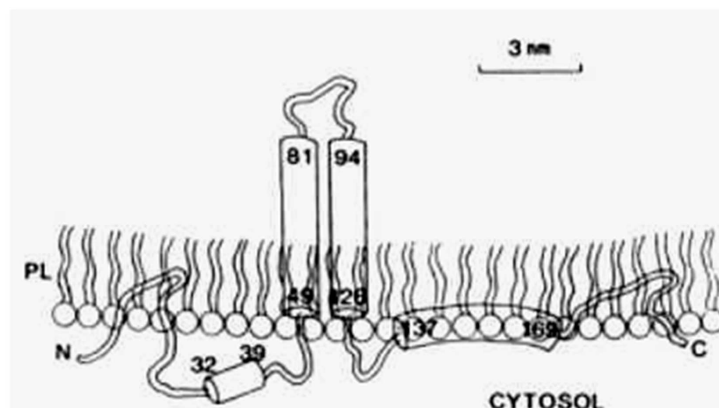


FIGURE 2 : Modèle de la conformation d'une oléosine de 18 kDa à la surface du corps lipidique (D'après [Huang 1996]).

Le domaine central, composé de 71 à 72 acides aminés enchâssés dans la monocouche de phospholipides [Lacey 1998, Tzen 1992b], est très conservé entre les différents isoformes d'oléosines avec, en particulier au centre, un motif proline très conservé [Abell 1997, Abell 2004, Peng 2007]. Les extrémités N- et C-terminales d'une cinquantaine d'acides aminés flotteraient à la surface externe de la demi-membrane de phospholipides. Contrairement au domaine central, ces extrémités sont peu conservées chez les différents isoformes d'oléosines et il est communément admis qu'elles confèrent un effet d'agent émulsifiant² interfacial à la surface du corps lipidique [Murphy 2001, Capuano 2007, Shimada 2010].

Si la structure et la composition des corps lipidiques, notamment leur protéome, est bien connue, les mécanismes sous-jacents à leur biogenèse et à leur dynamique ne sont pas encore clairement établis. Il est maintenant communément admis que les TAGs sont synthétisés dans des sous-domaines spécifiques du réticulum endoplasmique (RE) avant d'être libérés dans le cytoplasme [Wanner 1978, Murphy 1999b]. En revanche, il existe différentes hypothèses quant au mécanisme de formation du corps lipidique à la surface du réticulum endoplasmique et à sa dynamique, notamment au rôle que les protéines insérées dans la monocouche de PLs pourraient y jouer. Concernant les oléosomes, une première hypothèse suggère que la biogenèse du corps lipidique précède l'accumulation des oléosines à sa surface [Murphy 1989b, Cummins 1993, Beaudoin 2000, Banas 2000]. Une seconde hypothèse tend à montrer que les oléosines, initialement synthétisées dans le réticulum endoplasmique, diffuseraient latéralement à la surface du réticulum endoplasmique pour venir recouvrir la surface d'un corps lipidique naissant simultanément avec la synthèse des TAGs [Sarmiento 1997, Murphy 1999a, Lacey 1999]. La diffusion latérale des oléosines serait assurée par le domaine central hydrophobe de l'oléosine.

2. Les émulsifiants sont des molécules amphiphiles comprenant une partie hydrophobe et une partie hydrophile. Selon la grosseur relative des deux parties, l'émulsifiant sera favorable à la création et la stabilisation d'émulsions "huile dans eau" ou d'émulsions "eau dans huile" (D'après le dictionnaire LAROUSSE de la langue française).

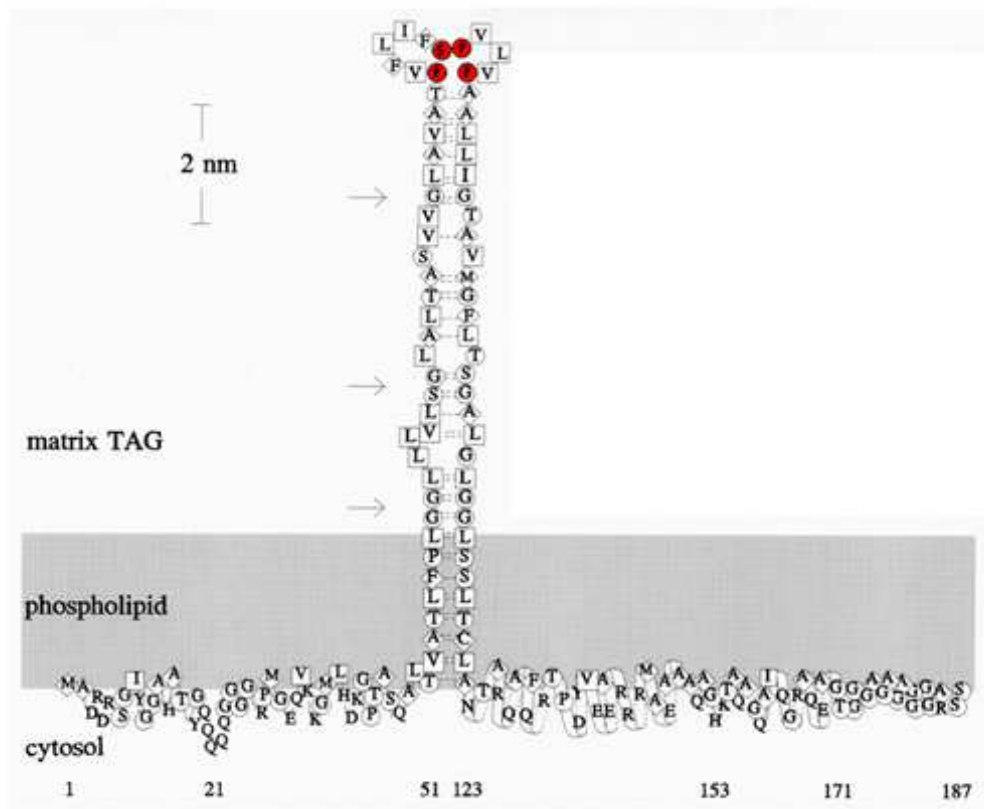


FIGURE 3 : Modèle de composition en acides aminés d'une oléosine de Maïs de 18-kDa à la surface du corps lipidique. La zone ombrée représente la monocouche de PLs avec les têtes polaires orientées vers le cytosol. Les symboles de diamant, cercles, et les lettres non entourées correspondent respectivement aux résidus amphipathiques, hydrophobes, et hydrophiles. (D'après [Tzen 1992b]).

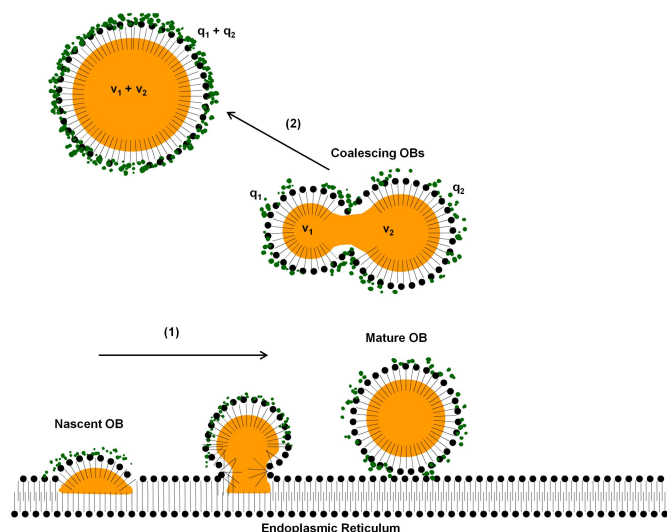


FIGURE 4 : Modèle de Beisson de synthèse de corps lipidiques de la plante oléagineuse (D'après [Beisson 1999]).

Un modèle descriptif [Beisson 1999] propose un scénario de biogenèse du corps lipidique dans la graine basé sur la seconde hypothèse : durant les premières étapes du développement de l'embryon, les corps lipidiques naissants ne seraient pas entièrement recouverts par des oléosines et coalesceraient pour former des corps lipidiques plus densément recouverts en surface et donc plus résistants à la coalescence. La vitesse de coalescence diminuerait ensuite progressivement jusqu'à avoir des corps lipidiques totalement stables (Figure 4).

Il est maintenant clairement établi que la dynamique des corps lipidiques est affectée par les protéines présentes à sa surface [Fei 2011, Cartwright 2012]. Dans la graine, Siloto et al. [Siloto 2006] ont montré qu'il existe une relation inverse entre l'accumulation de l'oléosine majeure chez *Arabidopsis*, S3 ou OLE1 (cf. 2.1), et la taille du corps lipidique, indiquant un contrôle de sa taille par les oléosines. En effet, les oléosines sont dotées de propriétés biophysiques permettant probablement la répulsion ou la coalescence des corps lipidiques mais dont la compréhension reste partielle. Il reste toutefois de nombreux aspects de la biogenèse et de la dynamique des corps lipidiques à élucider car ils mettent en jeu des facteurs génétiques, physiques et chimiques pouvant intervenir à l'échelle de la membrane du réticulum endoplasmique, de l'enveloppe du corps lipidique et de leur population dans le cytoplasme.

L'objectif de cette thèse est de modéliser la formation et la dynamique des corps lipidiques dans la graine en développement de l'espèce *Arabidopsis*, en s'appuyant sur les déterminants que sont le nombre et la variation de volume de ces organelles, en fonction de la présence de certaines des oléosines, et ce afin de mieux appréhender les mécanismes responsables de la biogenèse et la dynamique des corps lipidiques.

Ce travail est organisé comme suit :

- Chapitre 1 : Ce chapitre établit (sur la base d'un projet d'article) l'état de l'art sur les membranes biologiques, leur formation, ainsi que leurs interactions. Les modèles biophysiques sous-jacents sont rappelés.
- Chapitre 2 : Ce chapitre détaille l'ensemble des procédures qui nous ont permis d'extraire de l'information quantitative à partir des données d'imagerie confocale de différents génotypes de corps lipidiques d'*A. thaliana* au cours du développement. Pour cela, un pipeline de segmentation d'images basé sur l'algorithme de lignes de partage des eaux a été développé. Différents estimateurs qui caractérisent la taille et la dispersion spatiale des corps lipidiques ont été utilisés et validés dans le cadre d'un article accepté avec modifications mineures dans "Plant Physiology".
- Chapitre 3 : Ce chapitre aborde l'analyse statistique effectuée sur les données extraites à partir d'images (*c.f* Chapitre 2). Cette approche est basée sur l'analyse de la variance ainsi que l'analyse de quantiles et a été publié [Trigui 2013].
- Chapitre 4 : Ce chapitre propose un modèle de la dynamique des corps lipidiques au cours du développement de la graine. Après qu'un modèle simple ait été proposé (*c.f* Article de conférence : A.4), le modèle proposé ici est basé sur la théorie de coagulation développée par Smoluchowski [Von Smoluchowski 1917]. La dynamique de croissance de populations de corps lipidiques est alors simulée à travers ce modèle s'appuyant sur des hypothèses physiques proposées. Le modèle rend compte de la variabilité génotypique à travers différents paramètres qui sont estimés par une approche de Monte-Carlo par chaînes de Markov (MCMC) à partir des estimateurs extraits des données expérimentales.
- Chapitre 5 : Ce chapitre discute les principaux résultats de ce travail, leur adéquation par rapport aux hypothèses de départ, les difficultés rencontrées, et énonce des perspectives pour la poursuite de ce travail.

État de l'art sur les membranes biologiques

1.1 Introduction

Les membranes biologiques constituent les éléments de base des structures biologiques de tout organisme vivant. Ces membranes sont composées principalement de lipides et de protéines qui enveloppent les cellules, les organites ainsi que les compartiments cellulaires formant ainsi une barrière sélective. Une description en mosaïque fluide bidimensionnelle a été proposée par Singer et Nicolson en 1972 [[Singer 1972](#)]. Ce modèle suggère l'insertion et la mobilité latérale des protéines à la surface de la couche lipidique. On retrouve différentes formes géométriques d'organisation de membranes biologiques qui sont acquises en fonction des constituants de la membrane elle-même, ou de son environnement externe. Des modèles biophysiques ont été proposés, qui expliquent l'acquisition des formes membranaires et ont participé à la compréhension de divers mécanismes moléculaires comme le bourgeonnement, la fusion, la fission et le transport cellulaire. L'utilisation de ces modèles est étendue à des systèmes plus complexes, tels que les membranes multi-composants ou les radeaux lipidiques, des micro-domaines membranaires riches en cholestérol et sphingolipides et dont le rôle n'est pas encore clairement établi.

Ce projet initial d'article avait pour ambition de faire l'état de l'art des connaissances de point de vue biologique et biophysique sur les membranes biologiques et adresse les trois questions fondamentales suivantes : (i) comment les membranes se plient-elles ? comment acquièrent-elles une forme particulière ? et, enfin, comment interagissent-elles ?

N'étant pas spécialiste du domaine des membranes biologiques, nous pensons que ce projet était au delà de notre portée mais avons tout de même à coeur de restituer ce travail qui nous a occupé en début de thèse.

1.2 Projet d’article

On the concepts of biological membranes remodeling : which model for which process ?

Authors

Trigui G.^{1,2}, Miquel M.², Dubreucq B.² and Trubuil A.^{1†}

¹ INRA, MIAJ , centre de Jouy en Josas, 78352 Jouy en Josas CEDEX France..

² INRA AgroParisTech, IJPB, UMR 1318, INRA centre de Versailles, route de Saint Cyr, 78026 Versailles CEDEX, France.

[†] For correspondance (alain.trubuil@jouy.inra.fr).

Introduction

Biological membranes are key components in biological structures of almost all living organisms. These few nanometer thick structures are composed mainly of lipids and embedded proteins and act as a selective barrier delimiting cells, cell organelles and compartments from their outer environment. A revolutionarily understanding of the structure of biological membranes started with Singer and Nicolson’s model [Singer 1972], defining a two dimensional mosaic of individual inserted protein molecules drifting laterally in a lipid bilayer. Knowledge on the structure of biological membranes have increased over the last decades (Table 1.1). Cells, organelles, or closed surface vesicles may change shape depending on either membrane configuration itself or/and on the external flow, if its influence is significant. For example, the RBCs (Red Blood Cells) have initially a biconcave shape that can be changed depending on its external environment [Kaoui 2009]. This configuration of biological membranes has motivated the scientific community, particularly physicists, to assume that membranes are subject to the physical laws, and concepts coming from soft-matter physics have been applied to study different shapes and transitions for many organelle’s shapes composed of biological membrane. Models describing organelle’s shapes were first developed for general fluid-lipid vesicles and then generalized to more complex systems representing transitional mechanisms such as bending, fusion, and fission that occur in vesicle-based intracellular transport. Over the last few years, the use of vesicle-shaping models became more abundant due to a better understanding of the molecular mechanisms of cell trafficking, and was extended to the description of complex systems such as rafts and multi-components membranes. This paper explores the state of the art on the understanding of biological membrane elasticity, and then exposes the knowledge related to the growth of the lipid droplet as a particular membrane system.

Year	Finding	References
1972	Fluid mosaic model : Random distribution of proteins in a matrix of lipids.	Singer et Nicolson [Singer 1972]
1974	Cluster organization of membrane lipids.	Lee et al. 1974 [Lee 1974], Wunderlich et al. 1975 [Wunderlich 1975]
1982	Heterogeneity in the composition of membrane lipids : the existence of different phases.	Karnovsky 1982 [Karnovsky 1982]
1997	Microdomains called as rafts : the existence of the complex sphingolipid-cholesterol with membrane proteins.	Simons and Ikonen 1997 [Simons 1997], Brown and London 1998 [Brown 1998], Fratti et al. 2004 [Fratti 2004]
2001	Raft-like domains : the evidence of the coexistence of liquid ordered and liquid disordered phases.	Dietrich et al. 2001 [Dietrich 2001], Almeida et al. 2003 [de Almeida 2003]

TABLE 1.1 : Principal findings on the understanding of biological membrane structure.

How do membranes self-organize ?

Self assembly and molecular mobility. In nature, lipids exist in different forms [[Lipowsky 1995](#)], but their structure remains the same with two different domains : a hydrophilic head, and a hydrophobic tail. The hydrophobic tails of amphiphilic molecules are not soluble in water, so that they can make arrangements when in contact with aqueous solutions. Morphologically, such molecules combine to aggregate, arranging their hydrophobic domain against the water. Many shapes can be obtained such as spherical micelles, inverted spherical micelles, cylindrical aggregate, and bilayer membranes (Figure 1.1). The physical conditions determining these morphologies have been detailed by the model of Israelachvili et al. 1976 [[Israelachvili 1976](#)].

Two main geometrical constraints should be satisfied by the emerging shapes. One of them involves volume, V , and surface area, A , of aggregate, the other one avoidance of void region on the inside side of membrane. Formally, let v denote the volume of a amphiphilic molecule and a the area of the amphiphilic molecule head, then $V = Nv$ and $A = Na$ for an aggregate of N molecules, so $\frac{V}{A} = \frac{v}{a}$. This independance with respect to aggregate number of volume to surface area ratio resumes the first constraint. The second constraint is also easily translated when one evaluates constraints satisfaction for a given shape.

Hence, for a spherical micelle of radius R , satisfying these constraints we should have

$$\frac{V}{A} = \frac{R}{3} = \frac{v}{a}$$

$$R \leq l_c$$

where l_c is the characteristic length of the hydrocarbon tail, so $P \stackrel{def}{=} \frac{v}{al_c} \leq 13$. The

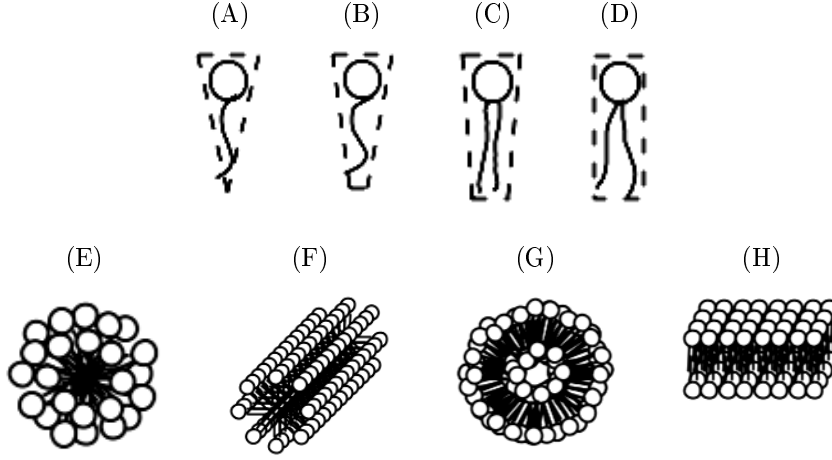


FIGURE 1.1 : Geometric structures of self-assembled molecules for a given molecular structure of lipids. For each phospholipid structure, a corresponding shape is obtained : the spherical micelles (E), the cylindrical layers (F), the vesicles (G) and the planar bilayers (H) are obtained respectively by the aggregation of the cone ($p < 1/3$) (A), the truncated cone ($1/3 < p < 1/2$) (B) and ($1/2 < p < 1$) (C) and the cylinder ($p \geq 1$) (D) shapes.

parameter P , referred to as the *packing parameter*, is dimensionless and depends only on molecule characteristics.

In the same way, a cylindrical shape of length L and radius $R \ll L$ should satisfy

$$\frac{V}{A} = \frac{R}{2} = \frac{v}{a}$$

$$R \leq l_c$$

so, $P \leq \frac{1}{2}$. For a bilayer flat membrane we would find $P \leq 1$. Actually, the rule for aggregates is :

- $P \leq \frac{1}{3}$ (spherical micelles)
- $\frac{1}{3} \leq P \leq \frac{1}{2}$ (cylindrical aggregates)
- $\frac{1}{2} \leq P \leq 1$ (planar aggregates)

The packing parameter is related to aspect ratio of amphipilic molecule. This is illustrated in Figure 1.1 where from left to right ((A)-(D)) the volume of the molecule increases with more than the product of characteristic length and area, the head area and characteristic length remaining more or less the same but the hydrophobic chain occupying more space. For figure 1.1, the ice-cream cone shape leads unsurprisingly to spherical aggregates.

Informations related to estimation of volume v , area a and characteristic length l_c are given in [Israelachvili 1976] and associated to results of Tanford [Hippel 1974]

where the following relations are proposed :

$$v = (27.4 + 26.9n)\text{\AA}^3 \text{ per hydrocarbon chain}$$

$$l_c = (1.5 + 1.265n)\text{\AA} \text{ per hydrocarbon chain}$$

where n is some value close but smaller than the number of carbon atoms per chain.

The estimation of optimal surface area per amphiphile of a spherical micelle leads to $a_0 = \sqrt{\frac{2\pi e^2 D}{\epsilon \gamma}}$ (cf. [Israelachvili 1976], page 1534) where γ is the interfacial tension of hydrocarbon-water interface, D, e and ϵ are related to repulsive electrostatic forces.

How membranes are deformed ?

Endocytosis and vesicle formation. All membranes are composed of lipids, proteins and carbohydrates, with a different composition ratios from one cell to another. Membrane proteins have regions of hydrophobic amino acids in α -helices that act as anchors. Some of membrane proteins act as channels that allow free crossing of selected ions through the membrane. Other proteins actively carry molecules across the membrane or act as a pump to send a chemical signal to the other side of the membrane. Another type of proteins is able to change mechanical characteristics of membrane for specific functions. This paragraph describes the endocytosis process mediated by proteins.

Mechanically, the curvature of a membrane is described by a transition from a planar structure to a curved one, which can either be positive (concave) or negative (convex). For a bilayer membrane, the outer monolayer bends positively, and the inner bends negatively, when looking to the headgroups of membrane lipids [Zimmerberg 2000].

Before being transported from one cellular compartment to another, cargo, lipids, trans-membrane proteins and even virus are packed inside vesicles. This function is carried out by the endocytic machinery. Endocytosis involves the budding of a membrane by modifying both the membrane tension and local curvature. These two characteristics are believed to be the most important parameters in the vesicle production rate, which may suggest the possible existence of a regulatory feed-back loop in transport mechanisms [Morris 2001, Keren 2011].

Clathrin is the major constituent of the "coat" of the clathrin coated pits and coated vesicles formed during endocytosis of materials at the surface of cells. The assembly of clathrin-coated vesicle is mediated by the recruitment of a coat component by adaptor proteins (for review see [Doherty 2009]). These proteins are also called "accessory proteins". They are able to deform the membrane and bind to the clathrin-coat protein. Concomitantly, they bend around the cargo proteins and sequester them into the clathrin bud. The best known group of accessory proteins is epsin, a family of globular soluble proteins, that act by increasing the curvature when inserted into the membrane via its amphipathic α -helix [Ford 2002, Stahelin 2003]. The endophilin, also contributes to clathrin-mediated endocytosis, particularly for the formation of synaptic vesicles. *In vitro* studies show

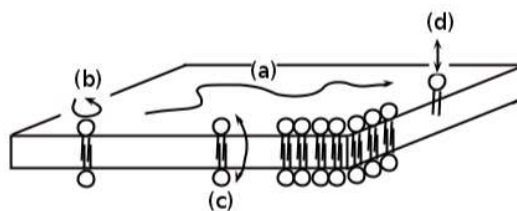


FIGURE 1.2 : Main dynamical behaviour of lipids molecules within bilayer. (a) lateral diffusion, (b) rotational diffusion, (c) flip-flop, (d) vertical vibration (D'après [Lipowsky 1995]).

that endophilin induces the generation of high curvatures in membranes and that the incubation of liposomes in the presence of the endophilin transforms them into tubes [Farsad 2001]. The CDE (Clathrin Dependent Endocytosis) is then driven by the growth of the coat domain, which deforms the membrane into a spherical cap that pinches off to form a vesicle. It is believed that dynamin, a large GTPase protein, is strongly implicated in the fission reaction of endocytic vesicles [Damke 1994, Zhang 2001], by creating a coil around the bottom of the vesicle, where it is still attached to the membrane, until its fission when dynamin coil changes conformation by GTP hydrolysis [Marks 2001].

Another example of protein interactions, which is able to change membrane curvature is that of the BAR (Bin, amphiphysin, Rvs') domain. The BAR domain is present in a large family of proteins. This proteic domain is rigid, relative to the cellular lipid bilayer and it exhibits a "crescent" shape. Upon binding, the membrane's curvature is increased by the rigid domain.

The theory of membrane elasticity. Shapes of cells and cell organelles are found in many biological complex systems [Zimmerberg 2006]. Every shape is designed either to exhibit specific physiological function, or to be adapted to specific cellular environment conditions [Scott 2003]. As seen in the previous section, biological membranes are dynamic, since their constituents, mainly lipids and proteins, are interacting permanently and varying quantitatively as well. The balance of forces created by these non-equal interactions probably modifies their structure and elastic properties.

From a molecular point of view, a membrane is defined as an isotropic two-dimensional medium. Molecules constituting the membrane can diffuse within the membrane and thus define several mobility coefficients. The diffusion coefficient can be determined by several techniques like FRAP (Fluorescence Recovery After Photobleaching) [Lipowsky 1995], Neutron scattering [König 1992], or High-frequency spectroscopy [Haibel 1998] and estimated to be $D = 10^{-12} \text{ m}^2/\text{s}$. In addition, in some cases, molecules may also move from one layer to another by a "flip-flop" movement, but with a slower diffusion, $D_f = 10^{-8} \text{ s}^{-1}$ [Alberts 2000] (Figure 1.2). These basic phenomena, implied by the microscopic properties of membranes and their coupling to the environment, such as interactions with proteins constituting the

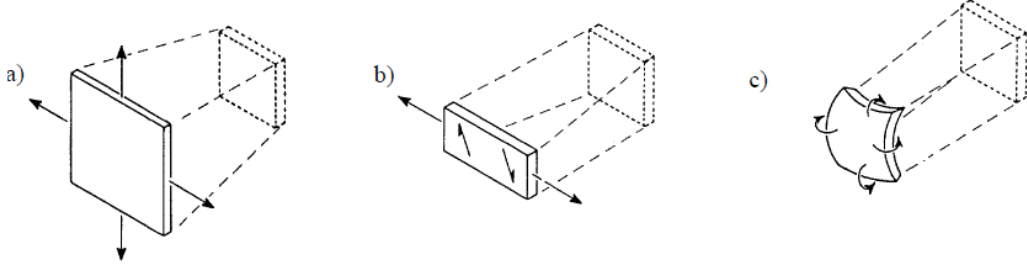


FIGURE 1.3 : Membrane deformation types : (a) stretching, (b) shearing, and (c) bending. (from [Evans 1987]).

membrane, or external membrane environment, are used to give a physical macroscopic description of the membrane based on three types of deformations : stretching, bending, changing thickness or shearing (Figure 1.3). Following Hooke's law, the energetic cost G_{area} of compressing or stretching a membrane is given by :

$$G_{area} = \chi \left(\frac{\Delta A}{A} \right) \quad (1.1)$$

Where χ is the compressibility modulus (in J/m^2), and A the area of the membrane. Shearing energy for a fluid membrane is negligible, and can be taken into account only if membrane is polymerized. The shearing energy G_{shear} is given by :

$$G_{shear} = \frac{1}{2} \mu (\bar{\lambda}^2 - \bar{\lambda}^{-2} - 2) \quad (1.2)$$

Where $\bar{\lambda}$ is the lateral extension rate and the shearing modulus (in J/m^2). The bending energy G_{bend} is given by :

$$G_{bend} = \frac{1}{2} \kappa (c_1 + c_2)^2 + \bar{\kappa} c_1 c_2 \quad (1.3)$$

Where $c_1 + c_2$ is the mean curvature and $c_1 c_2$ is the Gaussian curvature, with $r_1 = 1/c_1$ and $r_2 = 1/c_2$ are the two principal curvatures. κ and $\bar{\kappa}$ are respectively the rigidity modulus and the Gaussian curvature modulus (in J). For no topological change, the second term in 1.3 is neglected (referring to the Gauss-Bonnet theorem). For a closed surface representing the membrane, an equilibrium shape is obtained by minimizing the total energy derived from the four types of deformation with respect to the constraints of a constant area due to the incompressibility of the bi-dimensional membrane, and a constant internal volume due to the incompressibility of the fluid enclosed by the membrane. Historically, the simplest description of the bending energy without topological change was developed by Canham 1970 [Canham 1970], and was given by :

$$H_{min} = \frac{\kappa}{2} \int_s (c_1 + c_2)^2 dA \quad (1.4)$$

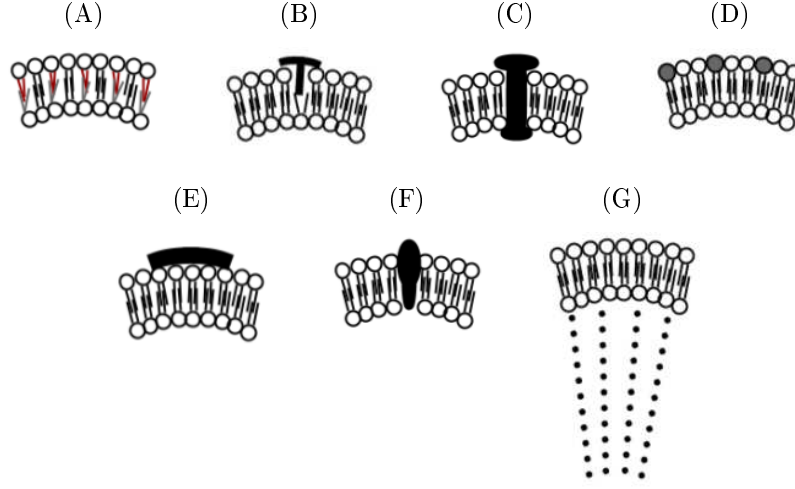


FIGURE 1.4 : Different mechanisms of molecule-induced membrane curvature : (A) Acyl-chain, (B) Helix insertion, (C) Oligomerization, (D) Head group, (E) Scaffolding, (F) Shape protein , (G) Actin cytoskeleton.

Helfrich [Helfrich 1973] improved Canham's model by adding the non-null spontaneous curvature c_0 expressing the possible dissymmetry between the two leaflets of membrane bilayers [Helfrich 1973, Deuling 1976] due to the insertion of inclusions (Figure 1.4). Another possible realization of spontaneous curvature is the exposure of the two membrane leaflets to a different chemical medium. The so called Canham-Helfrich Hamiltonian with the consideration of the spontaneous curvature model is thus expressed as follows :

$$H_{SC} = \frac{\kappa}{2} \int_s (c_1 + c_2 - c_0)^2 dA + \frac{\bar{\kappa}}{2} \int_s c_1 c_2 dA \quad (1.5)$$

Both Canham and Helfrich treated the membrane as a single elastic layer. However, the two layers of the bilayer membrane may have different areas due to the difference in the number of molecules contained by the layers [Sheetz 1974]. This idea introduced an effective non-local spontaneous curvature induced by the relaxed area difference between the layers $\Delta A_0 = A_0^{in} - A_0^{out}$, where A_0^{in} and A_0^{out} are respectively the area of the inner and the outer layer of the membrane. The ADE model introducing this constraint has the following energy functional :

$$H_{SC} = \frac{\kappa}{2} \int_s (c_1 + c_2 - c_0)^2 dA + \frac{k}{2} \frac{\pi}{Ad^2} (A_0^{in} - A_0^{out})^2 \quad (1.6)$$

Note that $k = 0$ corresponds to the original spontaneous curvature model.

The one-component vesicle constitutes the simplest model to study the different morphologies of a lipid bilayer. Equilibrium shapes are obtained by minimizing the bending energy. Several shapes can be obtained by varying different parameters *e.g.* volume and surface of the vesicle, the spontaneous curvature of the membrane or the temperature. The most widely used method for the minimization of vesicle

Biological system	Model derivation			Numerical method			Topological changes		References
	MM	SC	ADE	EL	PF	LS	yes	no	
General fluid-lipid bilayer vesicles	• • •	• •	•	• •	• •		• •	• •	[Helfrich 1973] [Seifert 1991, Miao 1991] [Du 2004] [Du 2006, Wang 2006]
RBCs	•	•	•			•		• •	[Cottet 2004, Cottet 2006] [Svetina 1989]
Experimental multi-phase bio-membranes	•	•			•		•		[Du 2006]
Golgi citernae		•						•	[Derganc 2007]
BAR dimer interaction	•	•						•	[Campelo 2008]

TABLE 1.2 : A summary of some biological systems modeled using the Canham-Helfrich theory and its derivations, the numerical method used (if developed), either the model supports topological changes or not. Abbreviations used for the model derivations refer respectively to : General Model (GM) , Minimal Model (MM), Spontaneous Curvature (SC) and Area Difference Elasticity (ADE) . Those used for numerical methods refer to : EL , PF , and LS.

energy relies on the Euler-Lagrange method. Other implicit methods like Phase-Field and Level-sets have also been used for membrane shape optimization. The use of membrane curvature models for many biological applications and within the cited methods are summarized in Table 1.2. General concepts dealing with these methods will be discussed below, but more detailed exposition about these topics will be found in references. For the general case of constraints to be imposed on a functional to be minimized, terms involving Lagrange multipliers are added to the functional. The shape of a vesicle must respect certain constraints such as the conservation of vesicle volume or conservation of vesicle surface area. Using Lagrange multipliers under the constraints of fixed enclosed volume and membrane area, the free energy is expressed as :

$$F = \frac{\kappa}{2} \int_s (c_1 + c_2 - c_0)^2 dA + \frac{\bar{\kappa}}{2} \int_s c_1 c_2 dA + \lambda \int_s dA + \Delta P \int_s dV \quad (1.7)$$

Where λ and ΔP are the Lagrangian multipliers, corresponding respectively to the surface tension of the lipid vesicles, and the pressure difference between the outer and inner sides of the vesicles. The first order variation gives the shape equation of lipid vesicles [Zhong-can 1987] :

$$F = 2\kappa \Delta H + \kappa(2H + c_0)(2H^2 - c_0H - 2H) - 2\lambda H + p = 0 \quad (1.8)$$

Where Δ is the Laplace-Beltrami operator on the surface S . The energetic phase field method is used for the treatment of interfacial problems, including topological changes. It has been firstly applied on vesicle dynamics to compute the equilibrium configuration of vesicle membrane under elastic energy and including spontaneous curvature effect [Du 2006]. The phase field approach is designed for a flow of a viscous incompressible fluid enclosed by a membrane limiting another incompressible exterior fluid. A phase variable $\varphi(x)$ is defined in a domain $\Omega \in R^3$ as the zero level sets function $\{x : \varphi(x) = 0\}$ when $x \in \Omega$, which is the compact surface representing the membrane. The phase field function is given by : $\varphi(x) = \tanh\left(\frac{d(x)}{\sqrt{x\varepsilon}}\right)$

$$E_\varphi = \int_\Omega \frac{1}{\varepsilon} \left(\varepsilon \Delta \varphi + \left(\frac{1}{\varepsilon} \varphi + c_0 \right) (1 - \varphi^2) \right)^2 dx \quad (1.9)$$

$$A_\varphi = \int_\Omega \left(\frac{1}{2} \varphi + 1 \right) dx \quad (1.10)$$

Minimizing the energy leads to the so-called phase diagram of the model. The phase diagram shows families, or branches, of shapes in different regions of the parameter space [Seifert 1991]. Limiting to closed membrane configuration, *i.e* vesicle, Seifert et al. [Seifert 1991] get the phase diagram of the minimum energy under constraints of fixed volume and area added to the curvature energy :

$$F_k = \frac{\kappa}{2} \oint dA (2C_m) \quad (1.11)$$

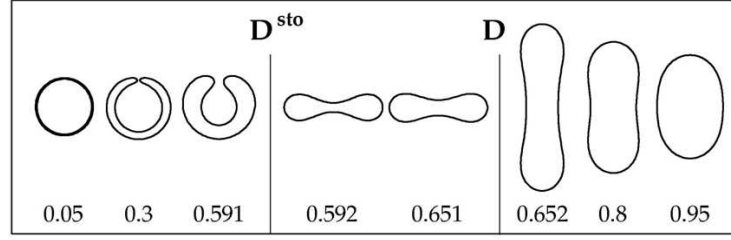


FIGURE 1.5 : The shapes of lowest energy are prolate for $v \geq 0.65$, oblate for $0.65 \geq v \geq 0.59$, and stomatocyte for $v \leq 0.59$ (from [Seifert 1991]).

This equation is given by the integral of equation 1.3 over its closed surface, where C_m is the mean curvature ($C_m = (c_1 + c_2)/2$), and A the surface area of the closed membrane. Then, using the Gauss-Bonnet theorem, the integral of the second term of equation 1.3 depends only on the topology (*i.e* The genus g) of a vesicle, *i.e* $\oint C_g dA = 4\pi(1 - g)$. This term can be discarded when using a fixed topology vesicle.

These constraints are added with Lagrange multipliers Σ and P to F_k , so :

$$F = F_k + \Sigma A + PV \quad (1.12)$$

Shapes are obtained with a simplification limiting to only axisymmetric geometries. In addition, the chemical composition and environment of both monolayers is supposed identical. The stationary shapes (Figure 1.5) depend only on one parameter which is the reduced volume $v = \frac{V}{(4\pi/3)(A/4\pi)^{3/2}}$.

The life of lipid droplets

After giving an overview of the theory associated with membrane formation and elasticity, our interest will now focus on a particular membrane system, the lipid droplets. These organelles have long been considered as a simple container of lipids on the cell of almost all living organisms. Since recently, lipid droplets were shown to be involved in lipid flux control and mobilization, as well as interacting with several cell compartments (Endoplasm Reticulum, Golgi apparatus, etc.). Indeed, from their birth to their degradation, lipid droplets are subject to several changes induced by particular events including fusion, fission, etc. Despite the development of techniques allowing the direct observation of cellular events, there is not much knowledge about the life of lipid droplets among plants, animals and bacteria. This has anyway not prevented building models and drawing hypotheses from indirect observations.

The literature on the process of birth of lipid droplets is not clear, even for the same organism and no clear description has been confirmed. Actually, the existence of a wide range of enzymes, and proteins co-localized both on the surface of lipid droplets and the surface of ER leads to believe their possible involvement in the birth process. Thus, many possible models describing the lipid droplet detachment

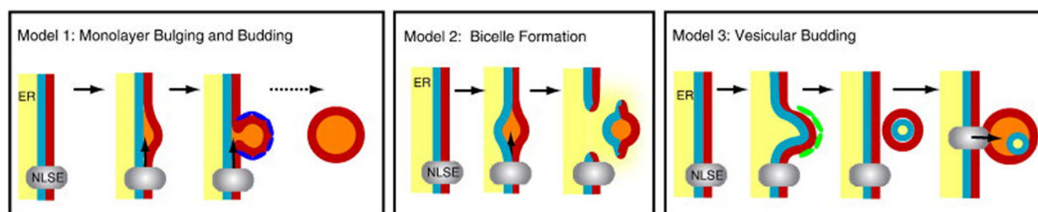


FIGURE 1.6 : Models of lipid-droplet biogenesis (D'après [Walther 2009]).

and formation have been proposed [Walther 2009] and can be summarized in the four following models :

- Model 1 : Neutral lipids accumulate within the bilayer that drives budding of a nascent droplet from the cytosolic leaflet of the membrane. A family of proteins called PAT proteins (perilipin, adipophilin, and the tail-interacting protein of 47 kDa (TIP47)), is believed to mediate budding [Goodman 2008]. It has been shown [Winfling 2013] that GPAT4 relocalizes from the ER to LD and this property is conserved for mammalian cells. In plants, *in vitro* studies with ER microsomes from sunflower (*Helianthus annuus*) have shown formation of small liquid lenses of 60-nm diameter in the ER bilayer that, similarly to LDs, contain freely mobile TG and recruit oleosins [Lacey 1999].
- Model 2 : Lipid droplets are excised from both leaflets of the ER membrane bilayer as a bicelle [Ploegh 2007] (Fig. 1.6), Model 2)
- Model 3 : In vesicular budding (1.6, Model 3), droplets are initially formed within small bilayer vesicles. In this scenario, a nascent vesicle is formed and remains tethered in close apposition to the ER membrane. Neutral lipids could fill the vesicle while it is still continuous with the ER membrane or after vesicle budding [Ploegh 2007].
- Model 4 : An "eggcup" model in which a LD grows within a concave depression of the endoplasmic reticulum (ER) through transport of neutral lipids from the ER [Robenek 2004]. The latter transfer process resembles a model for prokaryotes (28). All models posit LDs forming toward the cytosolic face of the ER membrane.

Chapman et al. [Chapman 2012] related recently to biogenesis and functions of LD in plants LD biogenesis in plants may combine several mechanisms. From biophysics arguments Zanghellini et al [Zanghellini 2010] proposed the budding of small (12 nm) droplets from the ER surface that fuse in the cytosol. In this system, the protein machinery may not have been taken enough into account, Chapman said. According to Chapman et al., alternatives models predict lipid droplets biogenesis from coatomer transport vesicles [Kalantari 2010] remaining integral part of the ER, growing or shrinking, depending on the needs [Goodman 2008].

In conclusion, it seems that the recent paper of Winfling et al. put on the foreground a model of growth of LD with or without fusion. We certainly should also examine more data considering possible growth of LD without fusion. Details on

the process could be gained with combination of inducible systems with super high light resolution microscopy. Characterizing a system for *in vitro* LD formation from ER membranes might allow molecular dissection of different steps of formation.

The storage of fat was described by two mechanisms namely hypertrophy (increasing the size), and/or the hyperplasia (increasing the number) of adipose cells [Jo 2009]. While the need of the cell regarding the use of hypertrophy is that larger lipid droplets raise storage efficiency, its use of hyperplasia is favorable for lipolysis (the maximum surface of exchange is obtained by maximizing the surface to volume ratio) [Paar 2012, Yamaguchi 2007].

A range of lipid droplets-associated proteins have been identified. In white adipose tissue, the Cide (Cell death-inducing DFF45-like effector) and Fsp27 (Fat-specific protein, 27 kDa) families are suggested to be involved in neutral lipids exchange. These proteins accumulate in the contact sites between two lipid droplets and create a bridge allowing the transfer of lipids from heavy to small droplets [Gong 2011] [Jambunathan 2011]. In addition, it has been demonstrated that the overexpression of these proteins establishes large lipid droplets while knockdown of FSP27 result in very small droplets in mammal adipose tissue [Keller 2008, Zhou 2012].

Another droplet-associated protein which deserves consideration is the so-called Perilipin, the most abundant lipid protein of white adipose tissue proteins [Londos 1999] associated with lipid droplets. This protein is a lipase storage regulator and believed to be a key component in the lipolysis of TAGs by facilitating access to HSL (Hormone-sensitive lipase) [Brasaemle 2000]. Last, another study showed an effect of decrease of lipid droplets when overexpressing perilipin. Perilipin acts by reducing the expression of Fsp27 in white adipose tissue of 3T3-L1 stain [Sawada 2010].

In seeds, a family of proteins called oleosins was identified on the surface of lipid droplets called also oil bodies or oleosoms [Vance 1987, Jolivet 2009]. These proteins seem to play an important role in the dynamics of oil bodies, and are suggested to act as "stabilizers" preventing oil bodies coalescence. Studies of oil bodies of *A. thaliana* embryos of oleosin mutants showed altered lipid and protein accumulation, accompanied with a delay of germination. In addition, abnormal over-sized OBs are observed [Siloto 2006]. On the other hand, a more recent study has demonstrated the formation of OBs of 50 nm diameter visible by electron microscopy on the surface of ER of cells of 24 kDa oleosin deficient transgenic soybean. These "micro" oil bodies coalesce with adjacent OBs forming a heterogeneous sized population regularly dispersed in cellular pool in such a manner that increased size oil bodies are held far away from the ER [Schmidt 2008].

Conclusion We reported some biophysical models based on the Helfrich theory of membrane curvature. These studies are the basis for investigating more complex systems such as raft domains, fusion and fission of vesicles, and lipid and proteins mixtures systems. These systems were not been investigated in this literature. Many

applications with more detailed modelling of biological systems require simplifying assumptions. These more complex models contain more parameters making a comprehensive study more difficult and often not even desirable. A quantitative modelling of such topological changes will require more microscopic experiments and some other higher-scale modeling approaches like statistical, or population dynamics modeling in order to collect more information about shapes dynamics.

1.3 Conclusion

Dans ce chapitre, nous avons reporté l'état de l'art sur les connaissances des membranes biologiques appuyées par des modèles biophysiques. Ces modèles décrivent le changement de formes des membranes biologiques en fonction des forces appliquées à leur surface. Les modèles à deux échelles (microscopique d'auto-assemblage de membranes et macroscopique de déformation locale) sont discutés. De point de vue expérimental, l'investigation des dynamiques membranaires à ces deux échelles est délicate. D'un côté, il est difficile d'étudier des organismes *in vivo* à cause des modifications apportées par l'extraction et la préparation d'échantillons de cellules qui peuvent endommager les mécanismes qu'on désire étudier. D'un autre côté, les approches artificielles, tel que les préparations basées sur des vésicules géantes (GUVs), nécessitent un support matériel de pointe (aspiration par micro-pipettes, instruments de mesures de déformations à petite échelle (*c.f.*)). Par ailleurs, l'exploration à une échelle plus grande, notamment celle des populations cellulaires, pourra être adoptée pour comprendre certains aspects dynamiques de vésicules formées de membranes.

Traitement des données, extraction et analyse de l'information

2.1 Introduction

La microscopie confocale est une technique courante dans l'observation des processus cellulaires à une échelle subcellulaire. Elle permet de sélectionner l'information en provenance d'une région cible autour d'un plan focal contrairement à la microscopie de fluorescence conventionnelle. Par ailleurs, l'évolution des ordinateurs et du traitement numérique rapide des données permet la reconstruction 3D de sous-volumes d'intérêt à partir d'une série de coupes optiques. L'imagerie des objets cellulaires observés en microscopie confocale donne accès à leurs caractéristiques spatiales et spatio-temporelles. Pour pouvoir extraire ces caractéristiques les objets d'intérêt doivent être séparés (segmentés) dans les images afin de pouvoir analyser leurs propriétés individuelles et leur organisation. La méthode de segmentation la plus simple est le seuillage¹. Néanmoins, il se peut que deux objets voisins soient en contact, ou encore que la limite de résolution de microscope fait que l'on ne distingue

1. Segmentation d'images par seuillage : Le seuillage consiste à tester pour chaque pixel de l'image si sa valeur est supérieure ou inférieure à un certain seuil, et produit une image binaire regroupant les résultats. Il est aussi possible d'utiliser plusieurs seuils, voire de sélectionner les pixels appartenant à une plage spécifique de niveaux de gris.

- Détection automatique du seuil : Il existe plusieurs méthodes pour calculer de manière automatique la valeur du seuil à appliquer. L'une des plus répandue est la méthode d'Otsu. Elle consiste à faire l'hypothèse que l'image contient deux classes, décrites chacune par une partie de l'histogramme de l'image. La qualité du seuillage est quantifiée en mesurant la variance des niveaux de gris de chaque classe. On cherche la valeur de seuil qui minimise les variances des deux classes.
- Seuillage à hystérésis : Le seuillage à hystérésis est une méthode un peu plus perfectionnée, qui permet d'améliorer la détection des contours des zones contrastées, tout en évitant de détecter des zones du fond. Le principe est d'utiliser deux seuils : un seuil "haut" et un seuil "bas". On sélectionne d'abord l'ensemble des pixels au dessus du seuil haut, puis l'ensemble des pixels au-dessus du seuil bas. On ne garde ensuite que les composantes connexes du seuil bas qui contiennent au moins un pixel au dessus du seuil haut. En pratique, cette détection peut se faire soit par reconstruction géodésique, soit par étiquetage des composantes connexes de l'image des seuils bas et détection des étiquettes présentes dans l'image des seuils hauts. (source : Site web de la Plateforme logicielle Cépia de l'INRA, <http://www.pfl-cepia.inra.fr/>)

pas l'espace de séparation de deux objets. Dans ces deux cas, les deux objets sont considérés comme une seule entité. Ce problème a été couramment rencontré dans les données d'images de corps lipidiques acquises dans le cadre de ce travail. L'approche de segmentation utilisée pour surmonter ce problème est détaillée dans ce chapitre.

Dans ce chapitre, nous présentons dans 2.3.1 les matériels utilisés pour acquérir les données d'imagerie des corps lipidiques de l'embryon d'*A. thaliana*. Nous détaillons ensuite la procédure de traitement et d'extraction des informations à partir de ces données dans 2.3.2. L'analyse et l'interprétation de ces informations sont présentées dans 2.4.

2.2 Objectif

L'objectif de ce chapitre est de décrire l'évolution de différents estimateurs qui caractérisent le volume et la dispersion spatiale de corps lipidiques.

2.3 Méthodologie

2.3.1 Acquisition des données

Matériel biologique. Les graines d'*Arabidopsis thaliana*, écotype sauvage Columbia (Col 0) proviennent de la collection de Versailles (INRA). Les graines des mutants proviennent de différentes collections internationales. Les mutants ont été croisés afin d'obtenir les différentes combinaisons de simple, double et triples mutants qui ont été vérifiées par génotypages. Elles sont semées en serre dans des pots de 5,5 x 5,5 x 6 cm contenant un substrat de pH 5,8 composé majoritairement de tourbe de sphagnum brune et blonde, à 10 % de sable volcanique et enrichi d'engrais N17-P10-K14 à hauteur de 1 kg/m³. Elles ont été exposées à 16 heures d'éclairage artificiel d'appoint, commandé par une cellule photoélectrique qui se déclenche dès que la lumière extérieure descend en dessous le seuil programmé à 6500 lux. L'intensité de 150 micromoles photon par m² par seconde est délivrée par des lampes haute pression à halogénure métallique d'une puissance de 400 W. Les plantes sont arrosées en moyenne deux à trois fois par semaine, à l'eau en début de culture, jusqu'au stade début de floraison, puis à la solution fertilisante (N : 140 mg/L, P2O5 : 76 mg/L, K2O : 231 mg/L, CaO : 146 mg/L, MgO : 15 mg/L, pH 6,5) jusqu'à la récolte des graines. La température de consigne est de 23 °C le jour de 6h à 22h et 17 °C la nuit de 22h à 6h et l'hygrométrie est de 35 % le jour et 50 % d'humidité relative la nuit. Pour l'échantillonnage, seule la hampe florale principale est conservée (montée à fleur après 10 jours de culture). Les fleurs sont étiquetées tous les jours afin de pouvoir identifier les stades de développements nécessaires, en jours après fécondation. Le jour du prélèvement les siliques correspondantes sont prélevées, les embryons disséqués et mis à colorer dans une solution de Rouge de Nil à 2 g/L (solution stock à 1 g/mL dans l'acétone, dilution extemporanée 1000 x) dans du glycérol 50 % (vol/vol).

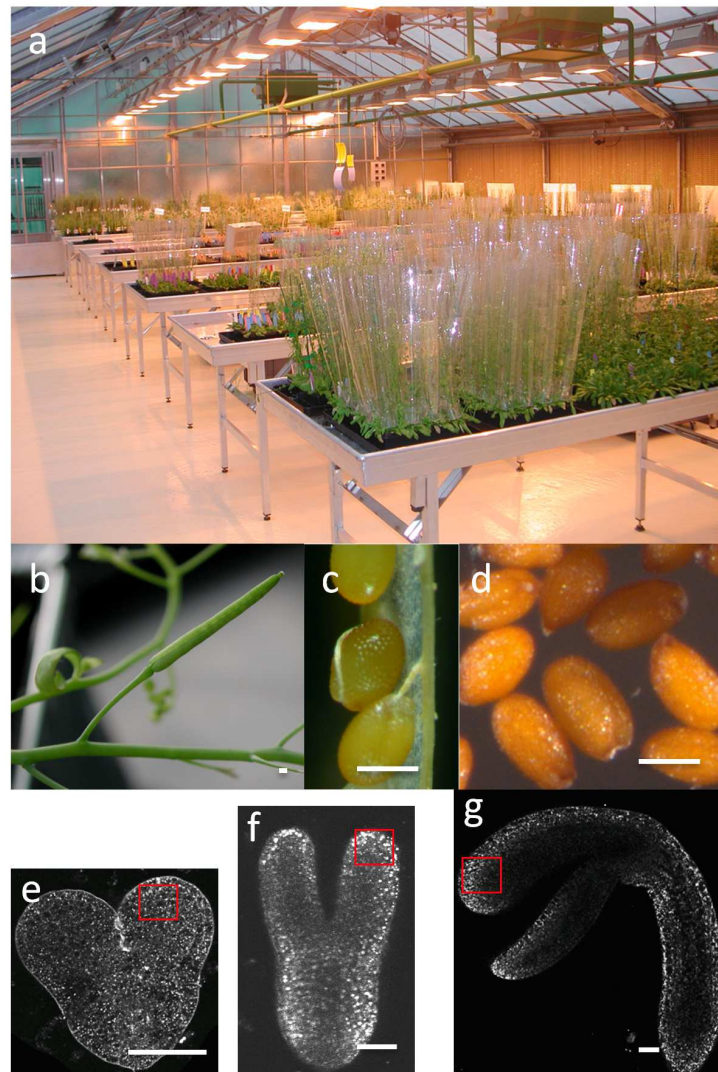


FIGURE 2.1 : Matériel végétal utilisé. Les plantes d'*Arabidopsis thaliana* ont été cultivées en serre (a). Les graines se développent après que la fécondation ait eu lieu dans des gousses (siliques - b), en s'organisant en 2 files d'une vingtaine de graines retenues au séparateur central par le funicule (c). Une fois sèches, les siliques sont déhiscentes et les graines sortent seules de la silique pour tomber au sol (d). L'embryon est entouré par les téguments de la graine qui lui assurent sa protection. Les stades de développements utilisés dans ce travail correspondent à des embryons (ici décortiqués de leurs téguments) aux stades cœur (e) torpille (f) et U-retourné (g) à respectivement 7, 9 et 11 jours après fécondation. Le carré rouge indique la zone utilisée pour réaliser les observations au microscope confocal. Bar = 1 mm b, 500 μm c,d, 30 μm e, f, g.

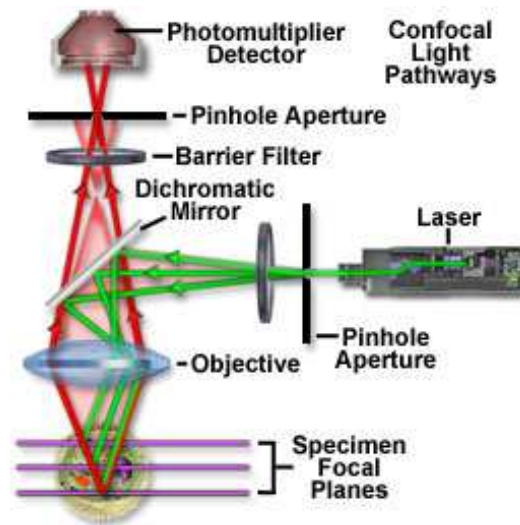


FIGURE 2.2 : Schéma du microscope confocal. Seuls les photons provenant de la zone excitée par le LASER sont détectés grâce aux pinholes. (Source : <http://www.olympusconfocal.com/theory/>).

Microscopie. Les acquisitions ont été effectuées par Bertrand Dubreucq et Martine Miquel sur la plateforme de cytologie et d'imagerie végétale² localisée sur le centre INRA de Versailles. Cette plateforme est adossée aux unités de recherche de l'IJPB (Institut Jean-Pierre Bourgin) mais a également vocation à accueillir d'autres équipes. Les activités de la plateforme couvrent la cytologie classique, l'hybridation in situ, l'immunolocalisation, la cytogénétique et la microscopie, en fond clair, en fluorescence ou confocale. Les images ont été majoritairement acquises sur un microscope confocal LEICA SP2 AOBS. Cette modalité permet d'obtenir des images de tranches optiques très fines d'un échantillon biologique grâce à sa forte résolution. Les éléments matériels de base d'un microscope confocal sont : l'optique, le LASER (Light Amplification by Stimulated Emission of Radiation), les pinhole (petit trou réglable en diamètre) et les photomultiplicateurs (Figure 2.2). Comme pour la microscopie à fluorescence conventionnelle, les éléments biologiques d'intérêt de l'échantillon à imager doivent être marqués par un ou plusieurs fluorophore (voir 2.3.1). Le fluorophore est alors excité par le rayon LASER de longueur d'onde convenable. Le faisceau du rayon LASER est balayé sur l'ensemble des points du plan confocal de l'échantillon. L'émission du fluorophore, résultant de l'excitation, est captée par le photomultiplicateur 2.3.1. Grâce au pinhole, seule la fluorescence provenant du plan focal est détectée. La quantité de photons recueillie est transformée ensuite en un signal numérique constituant un point de l'image. L'ensemble de l'image est reconstitué point par point sur l'écran. L'acquisition de plusieurs images dans le plan vertical z permet d'obtenir une reconstruction 3D des éléments marqués.

2. Site web de la plateforme : <http://www6.versailles-grignon.inra.fr/pciv>

Certains défauts dans les images peuvent résulter de l'utilisation de la microscopie confocale et parmi lesquels, nous pouvons citer :

- L'atténuation du signal en profondeur : les échantillons biologiques diffusent et absorbent des photons en provoquant ainsi une atténuation progressive du signal au fur et à mesure de la progression en profondeur selon l'axe z .
- Les aberrations sphériques : les indices de réfraction du milieu d'immersion et de l'échantillon lui-même devraient se correspondre parfaitement, ce qui n'est pas le cas en pratique. En effet, l'échantillon possède souvent un indice de réfraction inhomogène. Les rayons réfractés peuvent subir des déviations. Les objets imagés peuvent ainsi subir des distorsions de formes dans les images. Des objets de morphologie sphériques sont transformés pour devenir comprimés ou allongés le long de l'axe z de l'image par exemple.

Ces défauts sont dans certains cas incontournables, mais des corrections en post-traitement peuvent être envisagées. Le traitement par déconvolution est le plus souvent utilisé (*cf.* A.4).

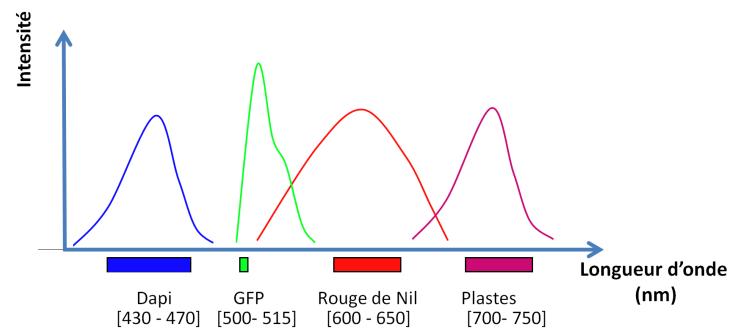


FIGURE 2.3 : Spectres d'émission de la GFP et le Rouge de Nil utilisés pour l'imagerie des corps lipidiques sur le microscope Leica SP2 AOBS. Les barres représentent les bornes de la zone de fluorescence de chaque photomultiplicateurs.

Matériel cytologique. Les organelles, ou les molécules d'intérêt, peuvent donc être marquées par un marqueur fluorescent afin de les visualiser par microscopie photonique. La fluorescence se produit suite à une absorption de photons. Elle est caractérisée par l'émission de photons qui se produit lors des transitions électroniques d'une molécule entre un état excité vers un état de relaxation. Les molécules fluorescentes sont caractérisées par leurs spectres d'absorption et d'émission et peuvent être classées par leurs spécificités aux organelles d'intérêt à marquer, par exemple, les marqueurs nucléaires (DAPI, Sytox, TOPRO), les marqueurs des mitochondries (MitoTracker), ou les marqueurs du Réticulum Endoplasmique (ER Tracker). Parmi les marqueurs spécifiques aux organelles riches en lipides, tel que les corps lipidiques, on distingue le Rouge de Nil [Greenspan 1985]. Cette sonde fluorescente hydrophobe³ produit une coloration rouge. Toutefois, dans un milieu

3. Est quenchée en milieu aqueux et fluoresce en milieu hydrophobe

très riche en lipides, la coloration devient plus intense avec des variations entre le rouge foncé et le jaune argenté. Le Rouge de Nil est excité à 485 nm et émet à 525 nm (Figure 2.3). D'autres sondes spécifiques aux milieux lipidiques existent. Un des marqueurs récemment mis sur le marché est le LipidTox [Yang 2013]. Cette sonde est connue pour être plus avantageuse que le Rouge de Nil notamment par une plus forte intensité du signal et un moindre bruit de fond, ainsi qu'un spectre d'émission plus large. Elle est cependant plus coûteuse et nécessite des rayes laser dans le rouge, mais accessibles.

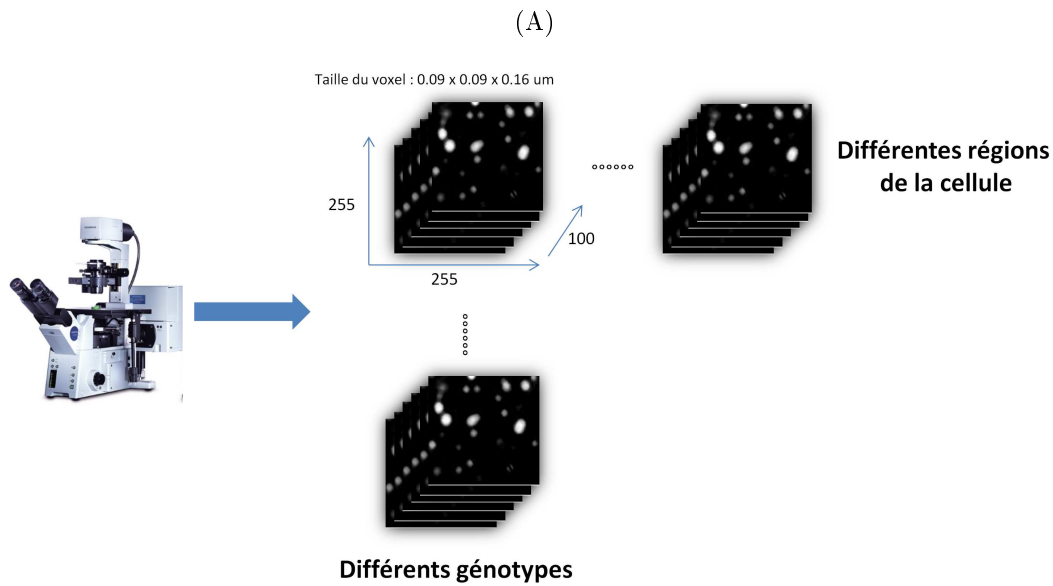


FIGURE 2.4 : Acquisition des piles d'images à partir de différents échantillons. Une pile d'images correspond à un génotype donné acquis à stade du développement donné dans une région de l'embryon donnée.

Par ailleurs, depuis la découverte de la protéine fluorescente verte, la GFP (Green Fluorescent Protein), au début des années 1960, il est devenu possible de fusionner le gène de la GFP au gène d'une protéine que l'on souhaite étudier [Chalfie 1994]. Le gène recombinant est réintroduit, sous contrôle du promoteur adéquat, dans les cellules, qui vont alors transcrire le gène et traduire l'ARNm pour produire la protéine fluorescente. On pourra ensuite l'observer, à l'aide d'un microscope à fluorescence dans les cellules vivantes. La longueur d'onde d'excitation de la GFP est entre 475 et 495 nm et celle d'émission est entre 520 et 560nm (Figure 2.3). Cette technique a permis le marquage de la membrane de corps lipidiques par fusion à un gène codant une oléosine. Il est à noter que le double marquage (oléosines et lipides neutres) permet de visualiser à la fois le contenu (lipides neutres) et le contenant (membrane du corps lipidique) sans que les deux fluorescences se mélangent (Figure 2.5). Cela valide la précision de la position des bords de corps lipidiques marqués dans le cas de l'utilisation de l'une ou l'autre technique de fluorescence.

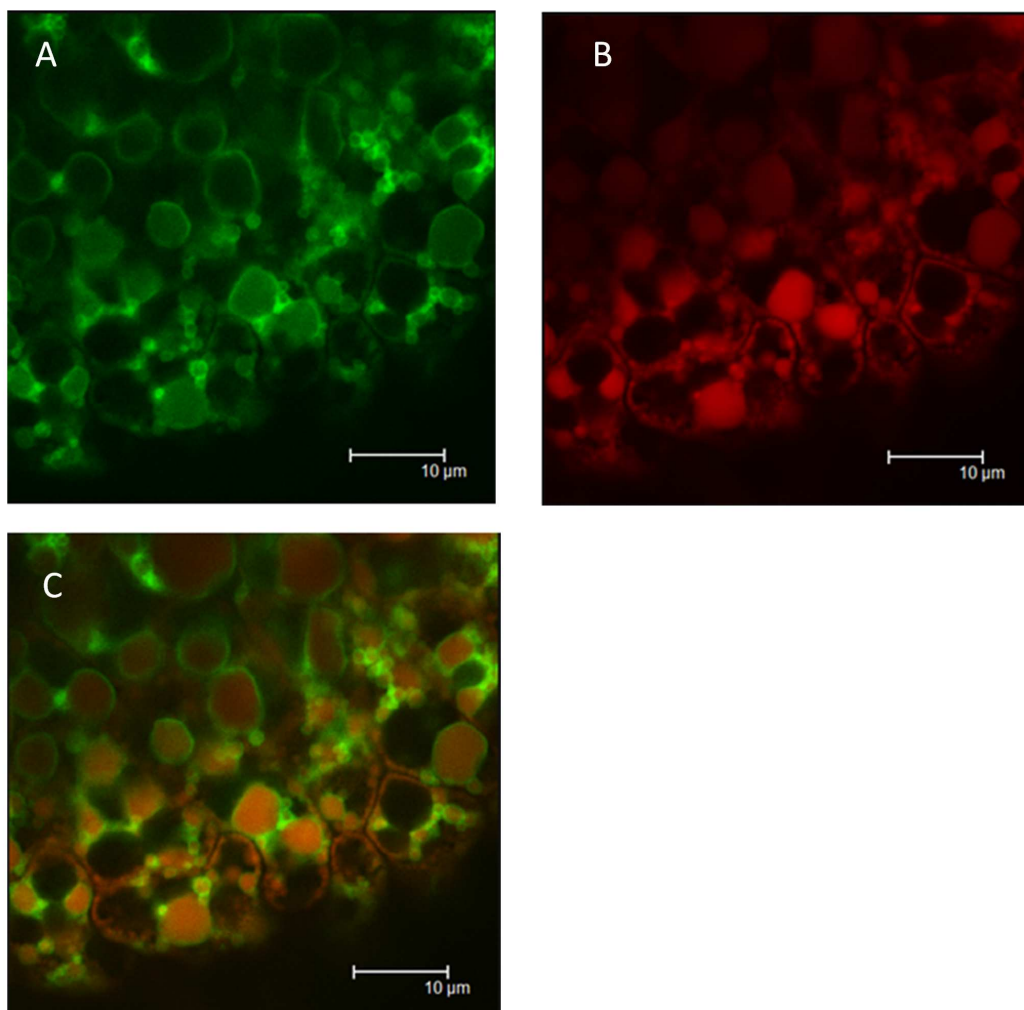


FIGURE 2.5 : Visualisation de corps lipidiques en microscopie confocale en double marquage : (A) GFP, (B) marquage au Rouge de Nil, (C) superposition.

2.3.2 Traitement des données

2.3.2.1 Segmentation d'images

La segmentation d'une image peut être définie par sa partition en différentes régions qui possèdent certaines propriétés communes (couleur, texture, homogénéité, etc.). L'unité élémentaire de référence d'une image segmentée n'est plus le pixel (ou voxel⁴), mais plutôt un ensemble de pixels (ou voxels) regroupés. A partir d'une image segmentée, les régions adjacentes, ainsi que les liens entre régions peuvent être examinés. La segmentation représente une étape cruciale vers l'interprétation

4. Voxel ("*volumetric pixel*"), est une unité de mesure d'un volume. Il représente une valeur dans un espace en trois dimensions. Le voxel est à la 3D ce qu'est le pixel à la 2D. Par extension le voxel est parfois appelé pixel 3D.

quantitative d'images numériques. D'un point de vue mathématique, la segmentation d'une image f est la division de son domaine D_f en n ensembles E appelés segments, tels que l'union de tous les segments est égale à D_f . Le choix d'un algorithme de segmentation d'une image dépend de plusieurs facteurs tels que la nature de l'image (nombre de canaux, résolution, dimensions, niveau de bruit, etc.), ainsi que la nature de l'information à extraire (taille des segments, forme, distance, etc.). Les approches de segmentation sont nombreuses, et sont classées principalement en approches basées sur la détection de discontinuités (*edge detection*) [Canny 1986], sur la détection de régions [Shih 2005], ainsi que les approches basées sur les contours actifs [Kass 1988, Xu 1998b, Xu 1998a]. L'idée de la segmentation par lignes de partage des eaux vient de la géophysique. Une goutte d'eau s'écoule sur un paysage formé de bassins et de crêtes jusqu'à atteindre le fond d'un bassin, appelé bassin versant. Les bassins versants sont séparés topographiquement par des lignes d'altitude maximale. Lorsqu'une goutte d'eau se retrouve sur des lignes d'altitude maximale elle ne peut pas être attribuée à un bassin versant. Ces points forment un paysage de lignes appelées lignes de partage des eaux (LPE). Ces définitions permettent de classer chaque point d'une surface topographique comme appartenant soit à un bassin versant, soit à une LPE. Une image numérique en échelle de gris peut également être considérée comme une surface topographique, dans laquelle on attribue une altitude à chaque pixel en fonction de sa valeur en niveaux de gris obtenant ainsi un relief. L'étape suivante consiste à immerger la surface. L'eau monte progressivement dans les bassins jusqu'à ce que les eaux issues de deux bassins versants voisins se rencontrent formant une ligne (Figure 2.6). L'ensemble des lignes constitue une frontière des objets séparés par des pixels qui appartiennent aux LPEs. Le résultat d'une segmentation avec LPE est généralement représenté par une image d'étiquettes, dans laquelle on attribue à chaque pixel appartenant à un bassin versant une étiquette (*label*) unique.

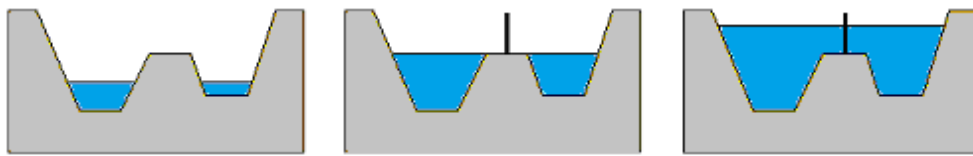


FIGURE 2.6 : Principe d'immersion des bassins versants en ligne de partage des eaux.

Dans les images à segmenter, on se confronte souvent à des zones du même niveau de gris (sous forme de plateaux en représentation en paysage) dans lesquelles la direction d'immersion ne peut pas être déterminée. Dans cette situation, les problèmes de sur-segmentation (le nombre d'étiquettes trop conséquent par rapport au nombre d'objets réel) sont fréquents [Vincent 1991]. La détermination au préalable des fonds des bassins versants peut résoudre ce problème. Une méthode de segmentation contrôlée par des marqueurs est bien décrite par Soille [Soille 2003]. Cette méthode est particulièrement utilisée pour la séparation d'objets quasi-sphériques

(*Blobs*) qui se touchent. En effet, la limite de résolution de l'instrument d'imagerie fait que l'on ne peut pas distinguer l'espace de séparation de deux objets voisins. L'idée de la segmentation par LPE contrôlée consiste à définir des marqueurs à partir desquels le processus d'immersion débute. Le paysage dans lequel le processus d'immersion se déroule est obtenu par la fonction distance de l'image originale binarisée. La fonction distance d'une image binaire est obtenue en attribuant à chaque pixel de l'image une valeur proportionnelle à sa distance au bord le plus proche de l'objet où il est contenu. Les marqueurs sont obtenus à partir des maxima de l'image distance (les maxima étendus sont obtenus par l'application d'un filtre h-maxima [Soille 2003]).

Les algorithmes de base de la morphologie mathématique permettant d'effectuer la segmentation par LPE sont implémentés dans différents logiciels de traitement d'images multiplateformes. Ces logiciels se distinguent par leur spécialisation pour chacun des domaines d'application (géophysique, science des matériaux, Imagerie biologique et médicale, etc.), ainsi que par leur performance en visualisation et manipulation d'images multidimensionnelles. AvizoFire est un logiciel commercialisé par la société VSG (Visualization Sciences Group, Mérignac, France). Il permet le traitement et la visualisation d'images 3D dans différents contextes applicatifs et notamment en sciences de matériaux et en microscopie. L'avantage de ce logiciel est que les données et les modules peuvent être connectés de façon interactive et contrôlés par plusieurs paramètres. Cette logique permet la création d'un réseau d'opérations de traitement dont la sortie est affichée dans un visualisateur 3D.

Dans le cadre de cette thèse, une licence du logiciel AvizoFire a été acquise. La procédure de segmentation d'images par LPE, ainsi que l'extraction des données relatives à l'analyse quantitative avec les opérateurs définis dans cette même section ont été implémenté sous AvizoFire. Deux modules en langage TCL⁵ (Tool Command Language) ont été développés (voir le script A.1.1.3) :

- WshedBinSeparate (Figure 2.8) : est un module qui permet d'effectuer la segmentation des images de corps lipidiques en 3D par la méthode de LPEs (figure 2.7). L'entrée est une image 3D en niveaux de gris. La procédure (Figure 2.8) commence par un filtrage médian visant à éliminer le bruit impulsionnel tout en conservant des détails suffisamment importants de l'image. Elle est ensuite poursuivie par deux opérations de seuillages, dont les valeurs seuils sont entrées par l'utilisateur (le choix s'effectue par un examen visuel du résultat de test pour différentes valeurs de seuil). Le premier seuil (haut), vise à la détection des gros objets sans les sur- ou sous-estimer, tout en négligeant la résultante élimination ou la sous-estimation des petits objets (Figure 2.9 - Seuillage haut). Pour le second seuil (bas), l'utilisateur

5. TCL ("*Tool Command Language*") est un langage de commande interprété, multiplateformes, conçu pour être facilement étendu dans une application (C, C++, etc.). Ce langage permet la définition de procédures, de variables globales, locales, etc., et dont la portée peut être accessible (de manière contrôlée) à des niveaux arbitraires. Il permet également le traitement des exceptions, la redéfinition (dynamique) des commandes, etc. Sa syntaxe est tirée à la fois du SHELL, du C et du LISP. L'utilisation de TCL dans AvizoFire est destinée à accéder à des fonctions spéciales de modules, d'automatiser les tâches de routine, pour résoudre certains problèmes en combinant des modules et des composants existants, ou pour l'écriture de données méthodes d'entrée/sortie.

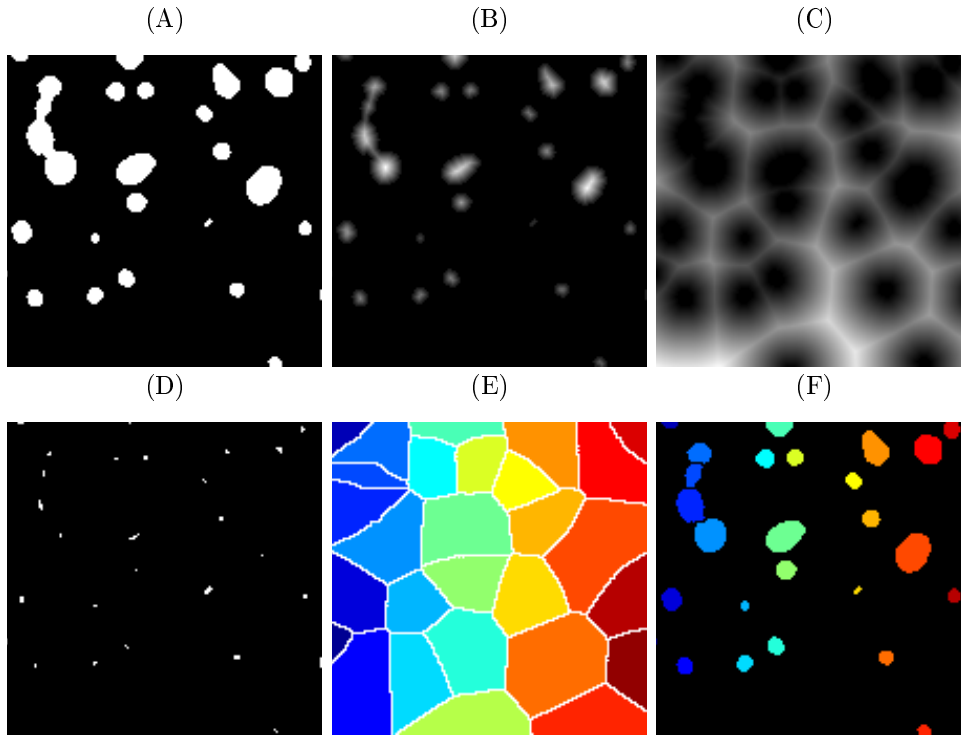


FIGURE 2.7 : Images intermédiaires de la procédure de segmentation par lignes de partage des eaux. (A) Image binaire, (B) Image distance, (C) Complément de l'image distance, (D) Image de marqueurs (maxima étendus), (E) Image de lignes de partage des eaux, (F) Image segmentée.

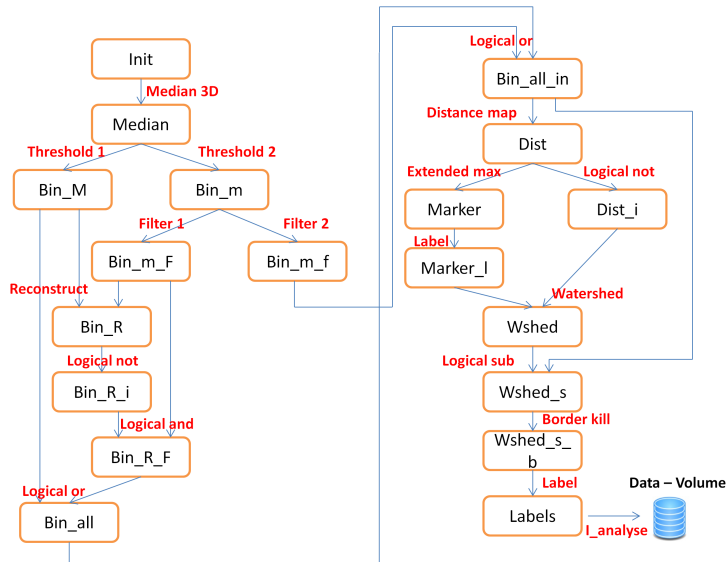


FIGURE 2.8 : Chaîne de segmentation d'image 3D de microscopie confocale de corps lipidiques marqués au Rouge de Nil. Le résultat obtenu à partir d'une image brute est un set de données de volumes individualisés de corps lipidiques

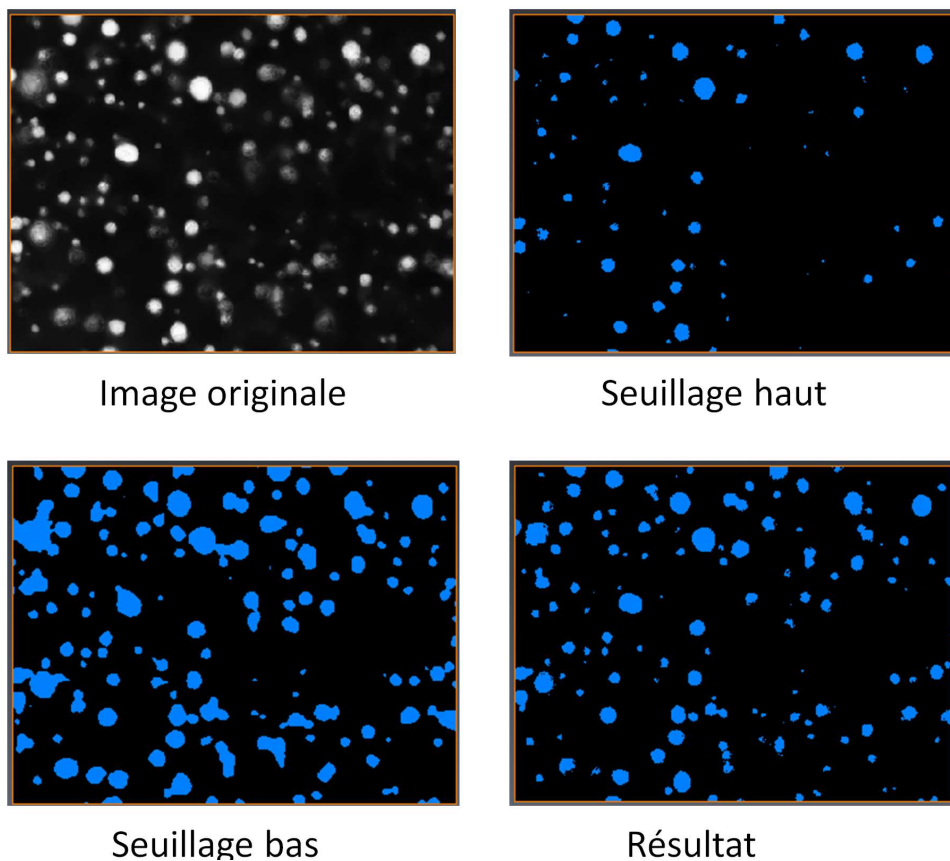


FIGURE 2.9 : Principe du double seuillage pour la détection d'objets : A partir de l'image originale en niveaux de gris, un seuillage haut permet la détection des gros objets. Le seuillage bas permet de détecter les petits objets. Le résultat (image binaire) est obtenu par combinaison des résultats des deux opérations de seuillage.

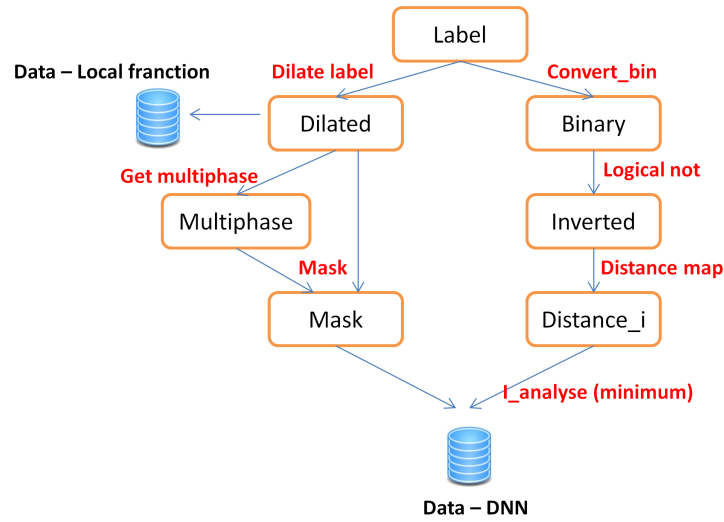


FIGURE 2.10 : Chaîne d'extraction de volumes des fractions locales des distances au plus proche voisin de chaque étiquette de corps lipidiques donnée en entrée.

s'intéresse à la détection des petits objets sans se soucier de la sur-estimation des gros objets (figure 2.9 - Seuillage bas). Il résulte de cette opération l'obtention des deux images seuillées qui correspondent aux deux valeurs de seuils. L'étape suivante consiste à l'application de deux filtres qui visent l'élimination des petits objets sous-estimés dans l'image de seuillage haut, et les gros objets sur-estimés dans l'image de seuillage bas. Une suite d'opérations binaires permet de récupérer le résultat final (figure 2.9 - Résultat). Ce résultat sert comme entrée pour la procédure de segmentation précédemment décrite dans 2.3.2.1. Le résultat final de segmentation des corps lipidiques est obtenu sous formes d'une image de labels où chaque label correspond à un objet individualisé. Le volume, la surface et les coordonnées spatiales de chaque objet sont sauvegardées dans un tableur (*datasheet*).

- ImLocalFrac (Figure 2.10) : l'entrée de ce module est une image de label des objets segmentés obtenue à la sortie du module précédent. Les cellules de Voronoï (Figure 2.12) sont obtenues par l'immersion d'un label jusqu'aux frontières de l'espace qui lui est plus proche qu'à d'autres objets. L'union du volume du corps et de l'espace occupé lors de l'opération d'immersion correspond au volume de la cellule de Voronoï. Cette valeur est ainsi sauvegardée pour chaque objet de l'image. Enfin, la mi-distance au voisin le plus proche (DNN) est calculée à partir de la valeur minimale obtenue en parcourant l'image de distance sur les frontières des cellules de Voronoï.

2.3.2.2 Extraction d'informations

Volumes et histogramme de volumes. L'histogramme est l'outil le plus simple pour caractériser la distribution d'une variable continue, où on exprime l'effectif d'une population dans chaque classe d'un critère donné pour chaque sous population. L'analyse d'images de corps lipidiques permet d'extraire un ensemble d'informations à partir de corps individualisés, à savoir leurs volumes, surfaces, etc. Ces données permettent d'effectuer un comptage de corps lipidiques (effectif par classe de volume) pour chaque classe de taille préalablement fixée. Le volume correspond au nombre de voxels multiplié par le volume unitaire d'un voxel. De telles données sont présentées en annexe A.1.

Fraction totale. Le champ photographique imagé dans l'embryon peut ne pas contenir en totalité le matériel biologique à exploiter, particulièrement quand il se situe aux bords de l'échantillon. Par ailleurs, nous n'avons pas pu colorer les parois pour s'emmener à un volume cellulaire moyen. Cette mesure résulte d'une zone inutile de l'image qui ne doit pas être prise en considération dans l'analyse quantitative. La solution étant de limiter le volume exploité aux zones contenant les objets cellulaires (dans notre cas, les corps lipidiques). Découper l'image en morceaux de façon telle à ne garder que les zones utiles est une alternative assez coûteuse en temps et résulte d'une perte d'information importante. Le découpage altère les objets dans les plans de coupe. L'enveloppe convexe permet d'extraire un emballage contenant tous les objets de l'image sans altérer les bords. En effet, elle permet d'estimer, avec une bonne précision, le bord de l'embryon qui le sépare du milieu extérieur. Ce dernier possède, la plupart du temps, une géométrie convexe dans la zone d'acquisition de l'image (Figure 2.1 - f, g). Cette enveloppe est définie comme étant le plus petit polygone convexe contenant tous les éléments T d'un ensemble fini E . Il existe plusieurs algorithmes de détermination de l'enveloppe convexe. Quickhull [Barber 1996] est un algorithme en dans l'espace à d dimensions R^d . cet algorithme est rappelé dans A.5.

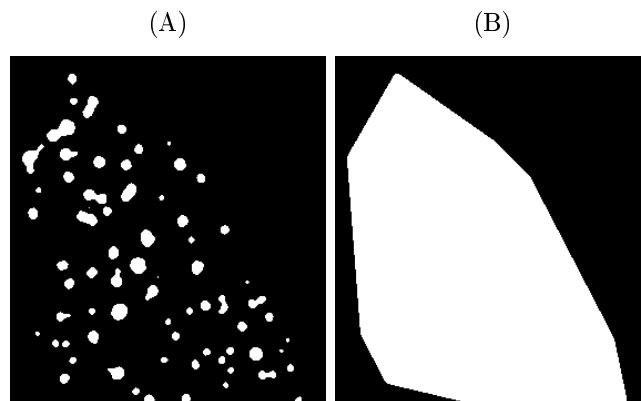


FIGURE 2.11 : Enveloppe convexe d'une image de corps lipidiques en 2D. (A) Image binaire de corps lipidique. (B) Enveloppe convexe.

L'enveloppe convexe est aussi utilisée pour caractériser la densité de corps lipidiques dans la zone imagée appelée la fraction totale (Figure 2.11). La fraction totale est définie par l'espace occupé par les corps lipidiques dans un volume donné tel que :

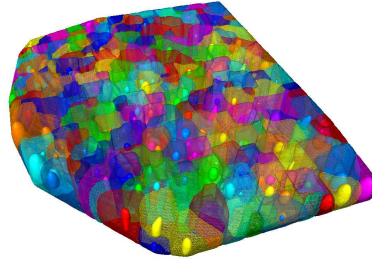
$$\phi_{tot} = \text{Volume total des corps lipidiques} / \text{Volume de l'enveloppe convexe} \quad (2.1)$$

Cet estimateur permet également d'estimer la vitesse de remplissage en corps lipidiques de l'espace cellulaire durant le développement de l'embryon et d'en déduire ainsi le volume de lipides neutres produit en fonction du temps. Tel que :

$$\text{Vitesse de remplissage (\%)} = \frac{\phi_{tot}(T + \Delta T) - \phi_{tot}(T)}{\Delta T} \quad (2.2)$$

Fraction locale. Afin d'analyser l'encombrement spatial des corps lipidiques, nous nous sommes basés sur le concept de la cellule de Voronoï. Dans notre contexte, la cellule de Voronoï d'un corps lipidique correspond à l'espace qui lui est plus proche que de tout autre corps lipidique. Dans le cas général, on définit un diagramme de Voronoï par l'ensemble des points les plus proches d'un élément p appartenant à un ensemble S de l'espace métrique E que de tout autre point. Dans le cas où les éléments sont des points discrets, on parle de tessellations de Voronoï (connues aussi sous le nom de tessellations de Dirichlet).

(A)



(B)

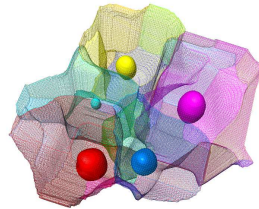


FIGURE 2.12 : Concept de cellules de Voronoï déterminées pour les corps lipidiques (A) Image 3D de corps lipidiques segmentés et entourés par leurs cellules de Voronoï (B) Extrait de corps lipidiques voisins entourés par leurs cellules de Voronoï.

Nous avons implémenté une procédure d'imagerie afin d'extraire l'information correspondante à cet estimateur) dans les images de corps lipidiques. Cette procédure consiste à assimiler l'image à un domaine constitué de deux phases : l'une associée au corps lipidique et l'autre au cytosol du milieu externe. On calcule alors une carte de distance de chaque point du domaine à la phase "corps lipidiques" (Figure 2.7-B). La valeur de chaque voxel de la phase "cytosol" est alors remplacée par sa distance au corps lipidique le plus proche. La topologie obtenue contient les bassins versants dont le fond correspond à la phase "corps lipidiques" et les crêtes ou les LPEs représentés par les points équidistants de la phase "corps lipidiques". On effectue ensuite une immersion à partir des fonds des bassins versants (Figure 2.7-D) jusqu'à atteindre les LPEs (Figure 2.7-E). La fraction locale est définie par :

$$\phi_{loc} = \text{Volume du corps lipidique} / \text{Volume de sa cellule de Voronoï} \quad (2.3)$$

2.4 Résultats

Les résultats de ce chapitre sont présentés sous forme d'un article. Cet article a été publié dans le journal *Plant Physiology*. Il est à noter que la nomenclature utilisée pour désigner les oléosines est celles données par [Shimada 2008] selon le tableau 2.1.

Locus	Nom selon Kim et al. (2001)	Nom selon Shimada et al. (2008)
Atg25140	S3	OLE1
At5g40420	S4	OLE2
At3g27660	S2	OLE3
At3g01570	S1	OLE4
At5g51210	S5	OLE5

TABLE 2.1 : Nomenclatures des oléosines graine-spécifiques d'*A. thaliana* utilisées dans de précédentes publications.

2.4.1 Résumé

Les corps lipidiques sont les organelles spécialisés dans le stockage des lipides dans la graine. Ils permettent l'accumulation de lipides neutres nécessaires pour le développement de la plante après la germination. Le corps lipidique est couvert par des protéines spécifiques incorporées dans une monocouche de phospholipides. Nous avons suivi la dynamique des corps lipidiques sur les embryons d'*Arabidopsis* pour différents stades de leur développement, utilisant des colorants fluorescents et la microscopie confocale. Les analyses ont été réalisées avec différents génotypes du type sauvage et en 3 mutants affectés dans leur accumulation de divers oléosines (OLE1, OLE2 et OLE4). L'acquisition des images a été suivie par une analyse statistique détaillée de la taille et de la distribution des corps lipidiques dans les

quatre dimensions au cours du développement de la graine (X , Y , Z , T). Nos résultats indiquent que la taille des corps lipidiques augmente fortement au cours de la maturation des graines, au moins par fusion, puis décroît jusqu'à la fin du processus de maturation. La taille ainsi que la distribution spatiale des corps lipidiques varient en fonction du génotype ce qui suggère des fonctions spécifiques de l'oléosine étudiée dans la dynamique et l'accumulation des lipides.

New roles for oil body oleosins during seed development

Bertrand DUBREUCQ, INRA – AgroParisTech, UMR 1318, Institut Jean-Pierre Bourgin, RD10, F-78026 Versailles CEDEX; +33 0130833066, bertrand.dubreucq@versailles.inra.fr

Martine MIQUEL, INRA – AgroParisTech, UMR 1318, Institut Jean-Pierre Bourgin, RD10, F-78026 Versailles CEDEX; +33 0130833320, martine.miquel@versailles.inra.fr

Research area: Cell Biology

Specialization of oleosins in OB dynamics during seed development in *Arabidopsis thaliana* seeds

Miquel M.^{1†}, Trigui G.², d'Andréa S.¹, Kelemen Z.¹, Baud S.¹, Berger A.¹, Deruyffelaere C.¹, Trubuil A.², Lepiniec L.¹ and Dubreucq B.^{1†}

¹: INRA - Institut National de la Recherche Agronomique, UMR 1318, Institut Jean-Pierre Bourgin, RD10, F-78026 Versailles CEDEX; AgroParisTech, Institut Jean-Pierre Bourgin, RD10, F-78026 Versailles CEDEX; ERL 3559 Centre National de la Recherche Scientifique, Institut Jean-Pierre Bourgin, RD10, F-78026 Versailles CEDEX.

²: INRA - Institut National de la Recherche Agronomique, UMR 0341 MIA, F-78352 Jouy en Josas CEDEX.

[†] For correspondance (bertrand.dubreucq@versailles.inra.fr, martine.miquel@versailles.inra.fr).

One sentence summary: **Oil body dynamics reveals new roles for oleosins during seed development**

Key words: *Arabidopsis thaliana*, oil body, oleosin, seed development, triacylglycerol

ABSTRACT

Oil bodies (OBs) are seed specific lipid storage organelles that allow accumulation of neutral lipids to sustain plantlet development after the onset of germination. The OB is covered with specific proteins embedded in a single layer of phospholipids. Using fluorescent dyes and confocal microscopy, we monitored the dynamics of OBs on living *Arabidopsis* embryos at different stages of their development. Analyses were carried out with different genotypes *i.e.* of the wild type and in 3 mutants affected in the accumulation of various oleosins (OLE1, OLE2 and OLE4), three major actors of the OB protein complement. Image acquisition was followed by a detailed statistical analysis of OB size and distribution in the 4 dimensions during seed development (X, Y, Z, T). Our results indicate that the OB size increases sharply during seed maturation, at least by fusion and then decreases until the end of maturation process. In simple, double and triple mutants backgrounds, the size and spatial distribution of OBs are modified affecting in turn the total lipid content, which suggests specific functions for the oleosin studied, in the dynamics of lipid accumulation.

INTRODUCTION

The seed is a complex and specific structure that allows a plant embryo in a quiescent form to cope with unfavourable germinating conditions, and/or permits dissemination of the species. To achieve these functions, seeds accumulate reserve compounds that will ensure the survival of the embryo and will fuel the growth of the plantlet after onset of germination. Lipid droplets (LDs) can be found in all eukaryotes such as yeast (Leber et al., 1994), mammals (Murphy, 2001; Hodges and Wu, 2010), *C. elegans* (Zhang et al., 2010; Mak, 2012), *Drosophila* (Beller et al., 2006; Beller et al., 2010) or plants (Hsieh and Huang, 2004) but also in prokaryotes (Wältermann et al., 2005). The basic structure of a LD is a core of neutral lipids covered by a phospholipid monolayer. LDs differ between species by the set of proteins covering their surface, by the nature of the lipids stored, and their turnover. Nevertheless they apparently ensure the same function in the cell *i.e.* energy storage (Murphy, 2012). In Brassicaceae, such as *Arabidopsis thaliana*, seed reserves are mainly composed of carbohydrates, proteins and lipids (Baud et al., 2002). The latest are primarily stored as triacylglycerols (TAGs) in LDs more commonly called oil bodies (OBs; (Hsieh and Huang, 2004; Chapman et al., 2012; Chapman and Ohlrogge, 2012) that exhibit a diameter comprised between 0.5 and 2 μm (Tzen et al., 1993).

The protein composition of seed OBs has now been determined for various plant species, including *Brassica napus* (Katavic et al., 2006; Jolivet et al., 2009) and *Arabidopsis* (Jolivet et al., 2004; D'Andrea et al., 2007; Vermachova et al., 2011). In *Arabidopsis*, 10 proteins have been identified and seed-specific oleosins represent up to 79% of the OB proteins (Jolivet et al., 2004; D'Andrea et al., 2007; Vermachova et al., 2011). Oleosins are rather small proteins of 18.5 to 21.2 kDa with a specific and highly conserved central hydrophobic domain of 72 amino acid residues surrounded by two hydrophilic domains of variable size and amino acid composition (Qu and Huang, 1990; Tzen et al., 1990; Tzen et al., 1992; Huang, 1996; Hsieh and Huang, 2004). It is admitted that oleosins cover the OB with their central hydrophobic domain inserted in the TAG through the phospholipids layer (Tzen and Huang, 1992). Beside their structural function in OB, oleosins may serve as docking stations for other proteins at its surface (Wilfling et al., 2013) and/or participate in biosynthesis and mobilization of plant oils (Parthibane et al., 2012; Parthibane et al., 2012). On the same line, oleosins are probably involved in the stability of the OB (Leprince et al., 1998; Shimada et al., 2008) and in the regulation of OB repulsion (Heneen et al., 2008), avoiding coalescence of LDs into a single organelle (Schmidt and Herman, 2008). Nevertheless, the precise function of oleosins in OB biogenesis and dynamics, if any, has not been established to date.

Global analysis of seed lipids can be performed using gas chromatography (Li et al., 2006) and allows precise determination of both lipid content and fatty acid composition. Recently, direct organelle mass spectrometry has been used to visualize lipid composition of cotton seed oil bodies (Horn et al., 2011). Nevertheless, in both cases, the methods are destructive. To observe lipid accumulation at the sub-cellular level, non-destructive techniques for lipid visualization have been available for years and adapted to seeds. Third harmonic generation microscopy (Débarre et al., 2006) and label-free coherent anti-Stokes Raman scattering microscopy (Paar et al., 2012) allow dye-less observation of LDs but require very specific equipments. Magnetic resonance imaging enables topographic analysis of lipid distribution in cereal grains (Neuberger et al., 2008) and in submillimeter seeds like tobacco ones (Fuchs et al., 2013). Nevertheless, the use of fluorescent dyes such as Nile Red (Greenspan and Fowler, 1985), BODIPY (Pagano et al., 1991) or LipidTOX™ (Invitrogen, Saint Aubin, France) associated with confocal microscopy, is also a powerful way to monitor LDs in living organisms.

Despite knowledge accumulated on this topic (Brasaemle and Wolins, 2012; Chapman et al., 2012), little is known on the dynamics of OB during seed maturation. In this article, we investigated this question by monitoring the evolution of OB in

living *Arabidopsis* embryo over time series. This analysis revealed a marked change in the size of OB at 9-10 days after pollination. We then examined simple, double and triple mutants of the major oleosins found in developing seeds (OLE1 - At4g25140, OLE2 - At5g40420 and OLE4 - At3g01570) (Jolivet et al., 2004), and analysed the OB dynamics in these mutant backgrounds. We showed that the lack of specific oleosins influences the dynamics and distribution of OBs during seed maturation that, in turn, affects lipid accumulation. These results pave the way for analysing specific function of oleosins in the synthesis, growth and evolution of OBs.

RESULTS

Oleosins and storage compounds accumulation

To gain further insight into the spatiotemporal analysis of oleosin expression pattern, several complementary techniques were used. A quantitative RT-PCR approach firstly confirmed that oleosins were highly and specifically expressed in silique material. The relative expression profiles of oleosins *OLE1*, *OLE2*, *OLE3*, *OLE4* and *OLE5* were investigated among a developmental series of siliques ranging from 4 to 22 days after flowering (DAF; Figure 1A).

If *OLE1*, *OLE2* and *OLE3* relative transcript levels were low during early silique development, a sharp peak in expression was observed from 9 to 15 DAF, corresponding to the onset of seed maturation; mRNA levels reached their maximal value at 16-18 DAF, then gradually decreased throughout the maturation phase. The transcript accumulation for *OLE4* was much lower than for *OLE1*, *OLE2* and *OLE3*, although decreasing at 18 DAF. On the contrary *OLE5* showed a low abundance of its mRNA, decreasing earlier than the other oleosins, at 13 DAF (Figure 1A).

Specific antibodies against each oleosin being available (D'Andrea et al., 2007), the temporal accumulation of oleosins was investigated by western blots (Figure 1B). Each oleosin content was quantitatively determined during seed development and expressed as percentage of its mature seed content. According to their respective transcript accumulation, *OLE5* was the first protein detected, followed by *OLE1*, *OLE2*, *OLE4* and *OLE3*. All oleosins accumulated during embryo development and were present in mature seeds. It is to be noticed that *OLE5* although present in mature seeds, represented only 2/3 of its maximal content.

Another question was the potential redundancy between oleosins. To address this question, we obtained *ole1*, *ole2* and *ole4* mutants. Immunolocalizations of oleosins *OLE1*, *OLE2* and *OLE4* in their respective mutant background (Supplemental Figure S1) were carried out and confirmed that the *ole1*, *ole2* and *ole4* mutants used in this study are null mutants. Protein extracts from wild type and the different simple, double and triple mutants were resolved by SDS-PAGE and stained with Coomassie blue (Supplemental Figure S2A). Results of image analysis (Supplemental Figure S2A) revealed that deficiency in one or several oleosins was not compensated by an increase in one of them. Storage proteins 12S and 2S were also resolved (Figure S2A). While oleosin(s) deficiency led in some cases to a reduced storage protein 12S and 2S content compared to wild type (Figure S2B), with the exception of the *ole1ole4* double mutant, the ratio between 12S and 2S storage proteins was not significantly affected (Figure S2C). On the contrary, the total fatty acid content appeared more susceptible to a deficiency in one or several oleosins as, in all cases, it was decreased compared to wild type (Figure S2B).

Oil bodies accumulation during seed formation and germination

During seed development, at early heart stage (7 DAF), WT oil bodies (OB) are scattered regularly in the embryo cells (Figure 2A). The average estimated OB volume is $0.096 \mu\text{m}^3$ (after segmentation). This volume increased during embryo growth with a marked increase after 9 DAF and at the transition between torpedo (9 DAF) and U-shape (11 DAF) stage (Figure 4A). At 11 DAF (Figure 2A) the OBs decreased in size, and were pushed to the edges of the cell, the cell center being occupied by protein bodies (PB) and other intracellular components (nucleus, Golgi network, etc...). After 11 DAF it is not possible any more to discriminate OBs with confocal microscopy as they are too close one from another. The distribution during embryo development was not affected in the single *ole1*, *ole2*, or *ole4* mutants (Figure 2B). In double mutants the OB accumulation pattern changed, especially after 9 DAF where OBs were grouped with a different spatial repartition. In the *ole1ole2ole4* mutant the OB volume increased from 9 DAF on, reaching an average $2.5 \mu\text{m}^3$ at 11 DAF. During germination, after 1-hour imbibition of dry mature seeds, rehydration was not complete and only minor changes in transcriptome were observed (Preston et al., 2009) and the OBs were still located on the edges of the cell. They were of very small size and could not be distinguished from each other in the case of the single mutants. After 24 hours of imbibition they began to increase in size and reduce in number with a population of almost distinctive OBs after a 48-hour imbibition (Figure 2B). In the double and triple mutants, the OB distribution was affected. In 1-hour imbibed seeds, the *ole2ole4* mutant containing larger oil bodies and *ole1ole4* or *ole1ole2* less compacted OBs in the cells than the wild type and the corresponding single mutants. The *ole2ole4* mutant OBs remained voluminous from dry stage to 48 h after imbibition. In the *ole1ole2ole4* mutant the average OB volume was rather constant (between 8 to $9 \mu\text{m}^3$) from 1-hour imbibed seeds until 48 hours after imbibition (Figure 2B). In the case of large OBs, the PB were pushed to the sides of cells by the OBs (Supplemental Figure S3) indicating a rather high structural flexibility of the PB when confronted to increase of OB volume

Characterizing oil body dynamics

OBs are normally scattered within plant cells, which are the larger volume for OB dispersion. We wanted to address the distribution of OBs in the plant cells in WT and oleosin mutant lines. Unfortunately, as Nile Red was not compatible with cell wall stains, no fluorescent dye allowed discriminating OBs from cell walls in living embryos (Figure 3-1 a, b). Consequently, individual plant cells could not be detected and we used estimators of the distance between each OB to its neighbors to answer the question of distribution in the 3D space. We have segmented the OB (Figure 3-2 a, b), and defined a volume containing all the OBs (the convex envelope, Figure 3-3 a, b, c). This 3D space was thus divided independently of the cellular segmentation in regions using Voronoï diagrams (Figure 3-4 a, b, c). The distance to the next neighbor or minimal distance (D_m) between OBs (Figure 4C) was computed and the Voronoï cell corresponded to junction of all the points at half the minimal distance (D_m) between the surface of two OBs. Therefore, the outer surface of all the Voronoï cells defined a global volume (the convex envelope) that resembled the surface of the organ considered (embryo cotyledon or hypocotyl) without the outer plant cell walls (Figure 3-3 b and 3-4 b).

Use of the Voronoï cells allowed calculating the "local fraction" estimator (ϕ_{loc}) i.e. the ratio of volumes between a given OB and the Voronoï cell containing it, therefore estimating densities of OBs and compaction processes, independently of the plant cell walls (Figure 4B). To evaluate the spatial distribution of OBs during seed development, we designed a pipeline based on morphological operators for the quantitative analyses and characterization of OB size and organization. The pipeline was implemented in a 3D imaging and analysis software, Avizo-Fire (Burlington, USA). It processed LSM image-stacks of OBs, yielding a mathematical representation of segmented OB images built on water shade-based techniques. Three estimators were used: the volume of OBs (V_{OB}), the local fraction (ϕ_{loc}) and the distance of each OB to

its nearest neighbor (D_m) (Figure 4). These estimators are detailed in the material and methods. Because the population of OBs contains small and large OBs and the growing process is expected to be continuous between small and large OBs, we have used different representations of the statistical distribution of the OB populations: the mean of volume, the median of volume and the mean of the first and last quantiles (Figure 4-A).

To go further in the analysis of OB size and dynamics in the different genotypes, a linear mathematical model was developed (Trigui et al., 2012) and a statistical analysis of it using a QR (Quantile Regression) (Trigui et al., 2013) was performed (Table I), which confirmed that the volume of OBs increased significantly in function of time (DAF) regardless of the genotype ($p < 2 \times 10^{-10}$). The maximum rate of increase occurred between 9 and 10 DAF, as the Log (V_{OB}) was increased by approximately $0.35 \mu\text{m}^3$ with respect to the intercept (7 DAF) (fitted value = -1.120 ± 0.009 , Supplemental Table S1). Focusing first on OBs lacking OLE1, OLE2 and OLE4 proteins (*ole1ole2ole4* mutant), the increase of OB volume (Figure 4A, mean of volume) indicated a fusion process, even without effect of the studied oleosin. Adding OLE4 oleosin (*ole1ole2* mutant) resulted in a decrease of the OB volume compared to the *ole1ole2ole4* mutant (Figure 4A; fitted value = -0.02 ± 0.008 , T Supplemental Table S1). Similarly, OBs containing only OLE1 oleosin (*ole2ole4* mutant) were smaller than OBs of *ole1ole2ole4* mutant (Figure 4A; fitted value = -0.091 ± 0.009 , Supplemental Table S1). In addition, OLE1 acted more efficiently than OLE4 in decreasing OB volume (*ole2ole4* compared to *ole1ole2*; Figure 4A, Supplemental Table S1). On the other hand, the effect of OLE2 was shown to increase the volume of OBs (fitted value = 0.156 ± 0.007 , Supplemental Table S1): the volume of OBs in the *ole1ole4* mutant is greater than that of *ole1ole2ole4* (Figure 4A). The mutant with OLE1 and OLE4 together (*ole2*) exhibited a reduction of the mean OB volume compared to *ole1ole2ole4* but their interaction seemed to be negligible ($p = 0.031$, Supplemental Table S1) suggesting a simple additive effect. The minimal OB volume was obtained by this mutant background (Figure 4A). However, the interaction between OLE2 and OLE4 was significant (fitted value = -0.127 ± 0.012 , Supplemental Table S1) and acted by reducing the OB volume even if OLE2 acted to increase OB volume. The common presence of OLE2 with OLE1 (in mutant *ole4*) led to a reduction of the OB volume when compared to the mutant *ole1ole2ole4*. This reduction was even larger than when it was present with OLE4 (mutant *ole1*). This result is in agreement with the fact that OLE1 is more efficient than OLE4 in preventing OB fusion.

Finally, when considering the WT, the volume of OB obtained by the contribution of OLE1, OLE2 and OLE4 was reduced compared to *ole1ole2ole4* mutant and its effect was significant (fitted value = 0.126 ± 0.019 , Supplemental Table S1). The presence of these oleosins on the surface of OBs modulates the fusion process by both contributing and preventing OB growth.

Analysis of the evolution of the Voronoï cell volume showed a global decrease of this volume between 7 and 8 DAF, then an increase from 8 DAF on during maturation except for the *ole2* mutant (Figure 4B). At the beginning of the embryo development, fusion events would be rather scarce while small OBs started filling the cell, thus limiting the space occupied by each OB. As OB density increased, OBs would fuse to optimize cell space, as shown by an increase in the volume of the Voronoï cell (Figure 4B). The minimal distance between OBs decreased upon time, indicating an increase in OB density (Figure 4C) but the evolution of the curve was different for the *ole2* mutant, that exhibited a low distance between OBs, which was confirmed by the presence of OB aggregates at 11 DAF (Figure 2A, Supplemental Figure S6).

Similarly to the volume, the local fraction decreased significantly in function of time (DAF) (Figure 4B; Supplemental Table S1, $p < 2 \times 10^{-10}$) but more linearly between 8 and 11 DAF (Log (ϕ_{loc}) increased by approximately 1.5 per time step). While the effects of OLE1 and OLE4 contributed to a low decrease (fitted value = -0.078 ± 0.007 and -0.053 ± 0.006 respectively, Supplemental Table S1) of the local fraction, OLE2 contributed to its increase (fitted value = 0.155 ± 0.006 ,

Table S1). The double and triple interactions involving OLE2 (OLE2:OLE4, OLE1:OLE2 and OLE1:OLE2:OLE4) were shown to have a synergistic effect on the volume of the OB ($p = 7.81^{-9}$, Supplemental Table S1).

Oil body fusion

In order to monitor OB behavior, 3D time lapse live-cell imaging of developing embryos was performed. No specific OB movement of large amplitude was detected. Nevertheless, long acquisition of immobilized embryos at early torpedo stage showed fusions of OBs, of same or of different sizes, resulting in larger OBs (Figure 5 and Supplemental Video S1). In a wild-type embryo, a typical fusion event was completed in about 15 minutes.

DISCUSSION

Mechanisms controlling oil accumulation in OBs are yet still largely unknown (Gong et al., 2011; Krahmer et al., 2011; Sturley and Hussain, 2012; Wilfling et al., 2013). Several hypotheses have been drawn to tackle the process of OB synthesis (Walther and Farese Jr, 2009; Yang et al., 2012) but the role of the conserved family of oleosins is still not clear (Chapman et al., 2012; Murphy, 2012). The analysis presented here was made on living material and represents a comprehensive analysis of oil body dynamics in developing embryos, without any treatment but Nile Red staining for visualisation of neutral lipids. After an increase, at least by fusion, during embryo development, the size of OBs decreases during maturation and reaches a smaller size than optical resolution of standard confocal microscopes after 12 days of seed development. We did not address the question of *de novo* synthesis of OB formation itself but rather OB dynamics after they have been released. Thus our study was mainly based on image analysis and focused on both the first half of seed development and the very early seed imbibition to bring an image of the very last point in the seed development. We have chosen to model OB population dynamics to i) better estimate the volumes of the OBs in the developing embryos and ii) be able to estimate the relationships between OBs in a 3D space and evaluate their relations to their neighbours. This time course analysis of the OB variations offers thus a new perspective on the subcellular events linked to OB growth when compared to previous studies based on mature seeds. Finally, it allowed to draw conclusions and new hypotheses about the role of oleosins in the dynamics of OBs.

The Arabidopsis oleosins family counts 16 members, 5 of which being specifically expressed in seeds (Kim et al., 2002; Jolivet et al., 2004; D'Andrea et al., 2007; Liu et al., 2012). In this study we focused on the most abundant oleosins, namely OLE1, OLE2 and OLE4. At the time, only a leaky mutant (i.e. with some protein accumulated in seeds) was available for OLE3 (Shimada et al., 2008) that was not of any help to conclude about its role in OB growth. No mutant was available for OLE5, that exhibits a different expression pattern, both in terms of intensity and timing of expression when compared to other members of the family (Figure 1), although OLE2 and OLE5 originated from a duplication of OLE3 and OLE1, respectively (Liu et al., 2012). We focused our study on oleosin simple, double and triple mutants and excluded the use of antisens or RNAi approaches to reduce the risk of cross-effects between one oleosin and the others. Lack of oleosin was associated in previous studies with a germination delay (Siloto et al., 2006) or lower freezing tolerance (Shimada et al., 2008). We have investigated the germination capabilities of the mutants already characterised and available in the frame of this study and, in our culture conditions, no specific delay in germination was observed on freshly harvested seed batches.

Oil bodies exhibit volume variations during seed development

We have constructed a treatment pipeline that allowed processing of the images data and transforming them into computational objects after segmentation. Doing so we have eliminated the possible errors of diameter manual measurements and we have increased the size of the population of measured OBs. Last we have been working with measured volumes and not with extracted diameters, thus making the analysis much more robust. During early seed maturation in WT seeds, OB mean volume increases sharply and then reaches a maximum before decreasing until the end of the maturation to sizes below the resolution of conventional confocal microscopes. In the same time, the distance between OBs grows first and then decreases, thus increasing the compaction of OBs. We have shown *in planta* that the increase in the size of the OB is at least due to OB fusions. Not only from cargo originating from the endoplasmic reticulum as previously suggested (Walther and Farese Jr, 2009) but also by fusion of existing OB of different sizes. We have shown that this general phenomenon is disturbed in mutants affected in oleosin(s), either concerning the size of OBs or the distance between them. In particular, the lack of OLE2 decreases the minimal distance between OBs. This is consistent with the median size of the Voronoï volumes, which is an indicator of the spatial distribution between an OB and all the other OBs in the vicinity, that remains at the same level between 9 and 11 DAF when compared to the control. From these data we can conclude that one of the putative effects of oleosin OLE2 could be to maintain the distance between OBs, limiting too close contacts and fusions. Nevertheless, *ole2* mutants show aggregates of small sized OBs (Figure 2A and Supplemental Figure S6) that are not fused together, thus indicating that the role of OLE2 is not to prevent fusions.

Mutants for specific oleosins show different dynamics of oil body growth and affect storage compounds accumulation

Changes in OB size and distribution did perturb the flux of lipid accumulation since a small but significant decrease in total fatty acids was evidenced in the different mutants backgrounds. The quantity of storage proteins 12S and 2S accumulated was also modified in some mutants but, with the exception of the *ole1ole4* background, the ratio between those proteins was not significantly affected. Finally, lack of one or several oleosin seems to have a low impact on reserve protein accumulation in the seed. Of note is the shape of the protein bodies in the *ole1ole2ole4* mutant that is completely different from the control: In WT cells, protein bodies are found in the centre of the cells and surrounded with small OBs. When the mean OB volume increases in the *ole1ole2ole4* triple mutant to form one or several very large OBs in the cell, protein bodies are pushed to the side of the cell with a fully different deposition profile (Supplemental Figure S3 for an example) as previously observed in *ole1* mutant by (Siloto et al., 2006). This certainly indicates a relative plasticity of proteins bodies and, contrarily, a relative density of the OB that can perturb the accumulation distribution of proteins when growing in the cell.

Oleosins OLE1, OLE2 and OLE4 have different roles

On the basis of the data accumulated and the statistical analysis performed, we propose that oleosins are not only structural proteins necessary to cover the OB single membrane but also have specific role in the dynamics of the OB during seed maturation. In the light of our analyses, five rules (R) can be drawn:

- R1: Fusions can occur between OBs.
- R2: Oleosins OLE1 and OLE4 limit OB fusion.

- R3: Oleosin OLE2 promotes OB fusion.
- R4: $R2 > R3$.
- R5: Oleosin OLE1 is more efficient in preventing OB fusion than oleosin OLE4.

These 5 simple rules are sufficient to explain the phenotypes observed in OB distribution or size in the mutant backgrounds. Rule 1 was directly drawn from observations in the WT as well as mutant genotypes. Rule 2 is rather straightforward: when lacking oleosins, OBs increase in volume leading to very large OBs in the cell, pushing away the protein bodies and other compartments (Supplemental Figure S3). Rule 3 addresses a new specific function to oleosin OLE2. One can thus from these data associate oleosins with specific OB behaviour during oil accumulation. Rule 4 defines the fact that if OBs are not properly covered with oleosins, they will fuse whatever effort is made to keep a distance, probably to optimize the oleosin/surface ratio.

The main model for OB biogenesis relies on the budding of lipid droplets from the ER followed by fusion of small OBs to form larger one (Murphy and Vance, 1999; Siloto et al., 2006; Guo et al., 2009) although the mechanisms underlying the delivery of the newly synthesized triacylglycerols to the OB remains unclear. While we did not study these early stages of OB biogenesis, our data clearly showed that OB growth can occur by fusion. It is to be precised that the use of Nile Red does not allow to discriminate between a fusion of two OBs with redistribution of the oleosins and phospholipids from donor to acceptor and a net transfer of storage lipids from smaller OBs to larger ones without redistribution.

In the case of depletion of one or several oleosins, the fusion is promoted probably maintaining an average number of oleosin molecules per surface unit of OB. Recent data suggest that this simple view of OBs is probably more complex in order to allow fusions, fragmentations (fissions), organelle interactions or transport needed for the dynamic volume variations of OBs (Walther and Farese Jr, 2009). In animals, net transfer of storage lipids from smaller LDs to larger ones without redistribution of LD proteins from donor to acceptor LD has been shown (Gong et al., 2011). Also, synthesis of triacylglycerols can be directly coupled with the formation of LDs, that share common features with plant OBs, thus indicating a growing process independent of fusions with existing LDs or with the ER (Wilfling et al., 2013). In that case one can hypothesize that the role of the oleosins on the surface of the OB is not only to ensure the stability of the organelle (role devoted to OLE1 and OLE4), the distance between organelles (role of OLE2) but probably also the docking of enzymes of the lipogenic pathway, such as members of the triacylglycerol synthesis machinery (Wilfling et al., 2013) that will allow direct filling of the OB. These docking properties could be carried out by OLE1, OLE2 or OLE4 in addition to their structural function but also by other direct members of the family such as OLE5 or OLE3. The fact that OLE1 and OLE5, as well as OLE2 and OLE3, derive from duplications suggests these protein may have the same function. Moreover, other proteins found in the protein complement of the OB, such as steroleosins, caleosins, seed- and microspore-specific oleosins, could be also implicated in this docking process.

In the future, we aim at using these rules to further characterise the importance of the different oleosins in the process, including movement of OB, measure of the fusion rate and movement of oleosins on the surface of the OB. This will pave the way to construction and improvement of a mathematical model of OB dynamics that could help to a better understanding of the OB role, beside being the central player in the accumulation of oil.

MATERIAL AND METHODS

Plant Material and Growth Conditions

Seeds of *Arabidopsis thaliana*, accession Columbia (Col-0) and of oleosin-deficient mutants, were surface sterilized and stratified for 48 h at 4°C in the dark. They were then germinated on full-strength Gamborg B5 medium (Duchefa

Biochemie B.V., Haarlem, The Netherlands, G0210.0050; pH 5.6). containing 0.8% agar, 1% sucrose and the appropriate selection agent, in a growth cabinet (16/8 hours light photoperiod at $150 \mu\text{m m}^{-2} \text{s}^{-1}$; 15°C night/ 20°C day temperature) After 14 days, the plantlets were transferred to compost, grown in a greenhouse under the following conditions: 13 h of light, diurnal temperature of 25°C , and nocturnal temperature of 17°C , and irrigated twice a week with mineral nutrient solution. To harvest seeds at defined developmental stages, individual flowers were tagged on the day of opening then opened flowers and developing siliques were counted daily. Material used for RNA and protein extractions, and lipid analysis was frozen in liquid nitrogen immediately after harvest and stored at -80°C prior to analysis. Weigh determinations of seed samples were realized on a M2P balance (Sartorius, Göttingen, Germany).

Oleosin-deficient mutants

Seeds of *ole1* (SM_3_29875; At4g25140) *ole2* (SALK 072403; At5g40420) and *ole4* (SM_20767; At3g01570) were obtained from the Nottingham Arabidopsis Stock Centre (NASC). The double (*ole1ole2*, *ole2ole4* and *ole1ole4*) and triple (*ole1ole2ole4*) mutants were generated by crosses between single mutants.

Fatty acid content and composition

For total fatty acid quantity and composition analyses by gas chromatography of the corresponding fatty acyl methyl esters, pools of 20 individual seeds were placed in a glass reaction tube prior to methanol/sulfuric acid/toluene (100:2.5:30, v/v/v) with addition of C17:0 fatty acid as an internal standard (Li et al., 2006). After derivatization for 1.5 hour at 95°C , fatty acid methyl esters were extracted in hexane, separated by GC on a 30 m x 0.53 mm ECTM-WAX column (Alltech Associates Inc., Deerfield, USA) and quantified using a flame ionization detector. The gas chromatograph was programmed for an initial temperature of 160°C for 1 min, followed by an increase of $20^\circ\text{C min}^{-1}$ to 190°C and a ramp of 4°C min^{-1} to 230°C , with a 9-min hold of the final temperature.

Imaging

Developing seeds were dissected from siliques for each stage of development of interest. Seeds were spread on a glass slide, incubated with Nile Red, a neutral lipid stain, at $2 \mu\text{g ml}^{-1}$ in a 60% glycerol solution. Embryos were removed from the seed teguments by gently pressing seeds between slide and lamella and observed after 30 min of incubation in the dark. For germinating seeds, mature seeds were placed on wet Whatman paper, stratified for two days at 4°C and then placed in a growth cabinet (16/8 hours light photoperiod at $150 \mu\text{m m}^{-2} \text{s}^{-1}$; 15°C night/ 20°C day temperature) for 1, 24 or 48 hours. Seed teguments were manually removed and the embryos incubated with Nile Red as described for developing seeds.

Images of dissected Arabidopsis embryos were acquired using an inverted LEICA SP2-AOBS spectral confocal laser microscope (LEICA Microsystems, Mannheim, Germany) using an HCX PL APO CS 40 X 1.25 objective. Fluorescence labelling from Nile Red was observed with a 488 nm light wavelength generated by an argon laser and an emission band of 550-650 nm. Each image consists of the maximum projection of 90-100 optical sections. Each section is the average of two scans conducted at the resolution of 512×512 pixels with a spatial resolution of ($0.09 \mu\text{m} \times 0.09 \mu\text{m} \times 0.16 \mu\text{m}$) in the (x; y; z) referential.

For 3D time lapse live-cell imaging developing embryos were stained with Nile Red and mounted in 0.008% low-gelling-point agarose in 50% glycerol in a glass bottom dish. Images were acquired using an inverted LEICA tandem SP5-AOBS

spectral confocal laser microscope (LEICA Microsystems, Mannheim, Germany) using an HCX PL APO CS 40 X 1.25 objective. Fluorescence labelling from Nile Red was observed with a 488 nm light wavelength generated by an argon laser and an emission band of 550-650 nm. Each image consists of the maximum projection of 100 optical sections and images were acquired every 2 minutes over a 12-hr period. Each section is the average of eight scans conducted at the resolution of 512 x 512 pixels with a spatial resolution of (0.09 μ m x 0.09 μ m x 0.16 μ m) in the (x; y; z) referential. Image stacks were processed with OsiriX, a medical image processing software for 3D rendering

Immunocytochemistry

Siliques of different developmental stages were fixed in 4% paraformaldehyde in phosphate-buffered saline (PBS) as described previously (Macquet et al., 2007). Sections of 8 μ m were incubated with the primary antibody raised against oleosins (D'Andrea et al., 2007) and secondary antibody (anti-Alexa Fluor 488, Molecular Probes, Invitrogen, Cergy Pontoise, France). Samples were visualised with an inverted LEICA SP2-AOBS spectral confocal laser microscope (LEICA Microsystems, Mannheim, Germany) equipped with an argon laser.

Quantitative PCR

Frozen tissues were ground in liquid nitrogen and total RNA was extracted using the SIGMA GenElute™ Mammalian Total RNA Miniprep Kit (Sigma-Aldrich). RNA extraction of dry seeds was performed as described by (Suzuki et al., 2004). An on-column DNase treatment was performed using the Rnase-free Dnase 1 (Qiagen, Courtaboeuf, France). For RT-PCR studies, the DNA-free RNA was converted into first strand cDNA using the SuperScript™ II RT pre-amplification system for first-strand cDNA synthesis (Invitrogen, Cergy-Pontoise, France) with oligo(dT)₂₂.

Quantitative PCR was performed with a Light Cycler using SYBR Green mix (ROCHE, Tutzing, Germany). Specific primers for *OLE1*, *OLE2*, *OLE3*, *OLE4*, and *OLE5* were designed using the Light Cycler probe design software version 3.5 (see Supplemental Table S2 for oligonucleotide sequences) and tested for efficiency rates as well as sensibility to dilutions of cDNA samples. *EF1 α 1* was used as control. The parameters for quantitative PCR were initial denaturation 95°C 8 min, (2) 45 cycles as follows: 94°C 10 sec, 55°C 10 sec, 72°C 15 sec with a temperature transition rate of 20°C s⁻¹, and (3) determination of fusion curve 94°C 10 sec, 65°C 30 sec and an increase to 94°C with a temperature transition rate of 0.1°C s⁻¹.

Quantitative western blots

Oleosin contents in *A. thaliana* developing siliques and mature seeds were measured by western blot using specific sera against the 5 seed-specific oleosins, *OLE1*, *OLE2*, *OLE3*, *OLE4*, and *OLE5* (D'Andrea et al., 2007). Total protein extracts were prepared by grinding siliques (7-19 DAF) or seeds isolated from siliques (20 and 21 DAF) in 2x Laemmli buffer (125 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol) using a glass-glass conical tissue grinder, at the ratio of 1 silique/100 μ L. After centrifugation, protein extracts were separated in 12% NuPAGE gels using MOPS SDS running buffer (Novex, Life Technologies). Proteins were transferred on to PVDF membrane (Immobilon-P 0.45 μ m, Millipore) in NuPAGE transfert buffer (Life Technologies). Saturation and incubation with antibodies were carried out according to (D'Andrea et al., 2007). Luminescent signals from western blots were recorded using a LAS-3000 luminescent image analyzer (Fujifilm, Saint

Quentin en Yvelines, France) and quantified with the MultiGauge software (Fujifilm, Saint Quentin en Yvelines, France). Mature dry seed extract was used as standard to determine the linear relation between luminescent signal and oleosin content. Using this linear relation, the amount of oleosin in developing seeds was determined from the luminescent signal (provided this signal was within the linear range) and expressed as the % of mature seed content.

Quantitative analyses of OBs size distribution and dispersion

A total of 180 3D stack images were firstly de-noised using Nd Safir (Boulanger et al., 2010) before being treated within the pipeline. First, the extraction of the volumes from individualized OBs through the pipeline is based on the watershed procedure dedicated to the separation of touching objects (Soille, 2003). The case of nearest OBs is quite abundant on the image data we analysed resulting on overlapped OBs segmentation when considering a simple thresholding (Supplemental Figures S4A, B, F). The distance transform assigns a value to each pixel on an object in the threshold image corresponding to its distance from the object border. The complement of this result (Supplemental Figures S4C, G) was transformed into its topographical representation, forming the catchment basins. The new separated objects corresponding to the OBs are then labeled with a specific color (Supplemental Figure S4D, H). The volume (V_{OB}) of each OB was obtained by counting the number of voxels on an object and multiplied by the elementary volume of the voxel.

The local fraction estimator is based on the concept of Voronoï diagram. The Voronoï cells were obtained by dividing the space containing OBs into cells in such a way that every point within each cell is closest to the OB surface that that lies within cell compared to all other OB surfaces. The local fraction is so defined by:

$$\phi_{loc} = \frac{V_{OB}}{V_C} \times (100 \%)$$

Where V_{OB} is the volume of OB and V_C is the volume of the Voronoï cell.

In order to get the boundaries of the sample, and to avoid overestimation of OBs Voronoï cells on the edge of the sample, we used the convex envelope (Figure 3-3 a, b, c). Intuitively, the convex envelop is defined by the smallest convex polygon that contains all the elements T of a finite set E . The corrected Voronoï cells are then computed with respect to the convex envelope (Figure 3-4 a, b, c).

The volume of each individualized Voronoï cell (V_C) is determined from labeled components as previously explained for (V_{OB}). Finally, extracting the minimal value across the boundaries of each Voronoï cell leads to the mid-distance to the nearest neighbour (D_{NN}). (Supplemental figure S5).

In order to study the influence of each oleosin, the day factor and their interactions on both V_{OB} and V_C , a mathematical model was developed (Trigui et al., 2012) and a variance analysis of this model was performed using a quantile regression (QR) (Trigui et al., 2013).

$$\begin{aligned} y_{ijk,t}^n = & \text{intercept} + \text{Day}_t + \text{S1}_i + \text{S3}_j + \text{S4}_k \\ & + \text{S1:S3}_{ij} + \text{S1:S4}_{ik} + \text{S3:S4}_{jk} \\ & | \text{S1:S3:S4}_{ijk} | \varepsilon_{ijk,t}^n \end{aligned}$$

ACKNOWLEDGEMENTS

We thank T. Chardot for many helpful suggestions and interesting discussions. We also thank M. Anjuere for taking care of the plants, and J. Thevenin, D. Rodriguez and M. Payre for technical help. M.M. is a Chargé de Recherche at CNRS.

FIGURE LEGENDS

Figure 1. Oleosin accumulation pattern during seed development. Abundance of oleosin mRNA (A) and oleosin protein accumulation (B) were determined in *A. thaliana* siliques collected from 7 to 22 days (mature seeds) after flowering (DAF). Abundance of oleosin mRNA was assessed by quantitative RT-PCR and *EF1αA1* was used as control. Oleosin protein accumulation was measured by immunoblot using specific sera. The 5 seed-specific oleosins were quantified: OLE1 (■), OLE2 (△), OLE3 (▲), OLE4 (○) and OLE5 (●). Luminescent signals from immunoblots, recorded using the LAS-3000 imaging system and quantified with the MultiGauge software, were within the linear range, allowing quantitative determination of oleosin content. For each oleosin, data were expressed as the percentage of the mature seed content, regardless of its relative amount among oil body proteins (insert shows oil body proteins resolved by SDS-PAGE and stained with Coomassie blue (b) or silver nitrate (a), to illustrate the relative content of each oleosin in purified *A. thaliana* oil bodies.

Figure 2. Phenotypes of oil bodies stained with Nile Red in different genetic backgrounds during seed maturation (A), and imbibition (B).

Red squares on pictures of developing embryos at 7, 9, and 11 DAF stages indicate localisation of image acquisition. DAF: days after flowering; hr: hours after imbibition start. Bar = 10 µm.

Figure 3. Sequential steps of image processing and 3D volume rendering.

This example is taken from a 10 DAF *ole1ole2ole4* sample image with a resolution of 512 x 512 x 512 pixels. a. Segmentation: Crude acquired images (1), segmented OBs (2), convex envelope (3) and Voronoï cells processing (4). b. 3D volume rendering of the image stack for each step. c. Transversal section and orthoslice of the 3D reconstruction showing individual OBs (3-c) and individual OBs with the shape of the Voronoï cells (4-c).

Figure 4. Distribution of volume (A), spatial disposition (B) and minimal distance (C) of oil bodies during seed development of oleosin mutants.

Figure 5. Fusion of oil bodies. OBs in living torpedo (10 DAF) Col0 embryos immobilized in low-melting agarose were stained with Nile red and observed with a confocal laser scanning microscope during 1 hour (1 Z-stack each 2 min). Time series presented were extracted each 7.5 min. Images stacks were processed with 3D medical image processing software OSIRIX for volume rendering.

Table I. Quantile regression analysis.

A variance analysis using a quantile regression (QR) statistical model was performed on the mathematical model (see Material and Methods) developed to study the effect of oleosins, the developmental stage and the interactions of all these factors on OB volume. Fitted estimate value of OB volume, standard error and p -values (p) are given for factors affecting oil body volume. Intercept: 7 DAF. DAF: day after flowering.

	$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
	Estimate $\pm$ s.e.	p	Estimate $\pm$ s.e.	p	Estimate $\pm$ s.e.	p
(Intercept)	-1.479 $\pm$ 0.016	0.000	- 1.186 $\pm$ 0.010	0.000	0.828 $\pm$ 0.012	0.000
8DAF	0.130 $\pm$ 0.010	0.000	0.191 $\pm$ 0.010	0.000	0.226 $\pm$ 0.013	0.000
9 DAF	0.246 $\pm$ 0.009	0.000	0.355 $\pm$ 0.010	0.000	0.469 $\pm$ 0.012	0.000
10 DAF	0.538 $\pm$ 0.011	0.000	0.733 $\pm$ 0.011	0.000	0.902 $\pm$ 0.013	0.000
11 DAF	0.722 $\pm$ 0.014	0.000	0.950 $\pm$ 0.012	0.000	1.114 $\pm$ 0.014	0.000
OLE1	- 0.058 $\pm$ 0.011	0.000	- 0.056 $\pm$ 0.011	0.000	- 0.104 $\pm$ 0.013	0.000
OLE2	0.137 $\pm$ 0.011	0.000	0.195 $\pm$ 0.011	0.000	0.210 $\pm$ 0.011	0.000
OLE4	- 0.037 $\pm$ 0.011	0.001	- 0.004 $\pm$ 0.011	0.712	0.008 $\pm$ 0.012	0.490
OLE1:OLE4	- 0.038 $\pm$ 0.016	0.020	- 0.066 $\pm$ 0.017	0.000	- 0.029 $\pm$ 0.020	0.134
OLE2:OLE4	- 0.089 $\pm$ 0.017	0.000	- 0.156 $\pm$ 0.017	0.000	- 0.190 $\pm$ 0.017	0.000
OLE1:OLE2	- 0.139 $\pm$ 0.018	0.000	- 0.168 $\pm$ 0.017	0.000	- 0.126 $\pm$ 0.018	0.000
OLE1:OLE2:OLE4	0.138 $\pm$ 0.026	0.000	0.179 $\pm$ 0.026	0.000	0.106 $\pm$ 0.028	0.000

SUPPLEMENTAL DATA

The following materials are available in the online version of this article.

Supplemental Figure S1. Validation of antibodies specificity on sections of developing siliques of mutants affected in one oleosin.

Supplemental Figure S2. Effect of oleosin(s) deficiency on protein content.

Supplemental Figure S3. 3D reconstruction of storage lipid and protein accumulation in WT and *ole1ole2* mutant background in 1-hour imbibed seeds.

Supplemental Table S1. Adjusted parameters for the QR statistical analysis of oil body volume and local fraction factors.

Supplemental video S1. 3D time lapse live-cell imaging of wild-type developing embryos stained with Nile Red.

Supplemental Table S2. Sequences of qRT-PCR primers used in this study.

Supplemental Figure S4. Segmentation steps of OBs using the watershed method. **Supplemental Figure S5.** Determination of the Voronoï cells and the Distance to the Next Neighbor (DNN) estimator steps.

Supplemental Figure S6. Distribution of OBs at 12 DAF in WT (A) and *ole2* (B) embryos.

LITERATURE CITED

- Baud S, Boutin J-P, Miquel M, Lepiniec L, Rochat C (2002) An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. *Plant Physiology and Biochemistry* **40**: 151-160
- Beller M, Bulankina AV, Hsiao H-H, Urlaub H, Jäckle H, Kühnlein RP (2010) PERILIPIN-Dependent Control of Lipid Droplet Structure and Fat Storage in *Drosophila*. *Cell Metabolism* **12**: 521-532
- Beller M, Riedel D, Jänsch L, Dieterich G, Wehland J, Jäckle H, Kühnlein RP (2006) Characterization of the *Drosophila* Lipid Droplet Subproteome. *Molecular & Cellular Proteomics* **5**: 1082-1094
- Boulanger J, Kervrann C, Bouthemy P, Elbau P, Sibarita J, Salamero J (2010) Patch-based nonlocal functional for denoising fluorescence microscopy image sequences. *IEEE Trans Med Imaging* **29**: 442-454
- Brasaemle DL, Wolins NE (2012) Packaging of Fat: An Evolving Model of Lipid Droplet Assembly and Expansion. *Journal of Biological Chemistry* **287**: 2273-2279
- Chapman KD, Dyer JM, Mullen RT (2012) Biogenesis and functions of lipid droplets in plants. *Journal of Lipid Research* **53**: 215-226
- Chapman KD, Ohlrogge JB (2012) Compartmentation of Triacylglycerol Accumulation in Plants. *Journal of Biological Chemistry* **287**: 2288-2294
- D'Andrea S, Jolivet P, Boulard C, Larre C, Froissard M, Chardot T (2007) Selective one-step extraction of *Arabidopsis thaliana* seed oleosins using organic solvents. *J Agric Food Chem* **55**: 10008-10015
- Débarre D, Supatto W, Pena A-M, Fabre A, Tordjmann T, Combettes L, Schanne-Klein M-C, Beaupaire E (2006) Imaging lipid bodies in cells and tissues using third-harmonic generation microscopy. *Nat Meth* **3**: 47-53
- Fuchs J, Neuberger T, Rolletschek H, Schiebold S, Nguyen TH, Borisjuk N, Bömer A, Melkus G, Jakob P, Borisjuk L (2013) A Noninvasive Platform for Imaging and Quantifying Oil Storage in Submillimeter Tobacco Seed. *Plant Physiology* **161**: 583-593
- Gong J, Sun Z, Wu L, Xu W, Schieber N, Xu D, Shui G, Yang H, Parton RG, Li P (2011) Fsp27 promotes lipid droplet growth by lipid exchange and transfer at lipid droplet contact sites. *The Journal of Cell Biology* **195**: 953-963
- Greenspan P, Fowler SD (1985) Spectrofluorometric studies of the lipid probe, Nile red. *Journal of Lipid Research* **26**: 781-789
- Guo Y, Cordes KR, Farese RV, Jr, Walther TC (2009) Lipid droplets at a glance. *J Cell Sci* **122**: 749-752
- Heneen W, Karlsson G, Brismar K, Gummeson P-O, Marttila S, Leonova S, Carlsson A, Bafor M, Banas A, Mattsson B, Debski H, Stymne S (2008) Fusion of oil bodies in endosperm of oat grains. *Planta* **228**: 589-599
- Hodges BDM, Wu CC (2010) Proteomic insights into an expanded cellular role for cytoplasmic lipid droplets. *J. Lipid Res.* **51**: 262-273
- Horn PJ, Ledbetter NR, James CN, Hoffman WD, Case CR, Verbeck GF, Chapman KD (2011) Visualization of Lipid Droplet Composition by Direct Organelle Mass Spectrometry. *Journal of Biological Chemistry* **286**: 3298-3306
- Hsieh K, Huang AHC (2004) Endoplasmic Reticulum, Oleosins, and Oils in Seeds and Tapetum Cells. *Plant Physiology* **136**: 3427-3434
- Huang A (1996) Oleosins and Oil Bodies in Seeds and Other Organs. *Plant Physiol.* **110**: 1055-1061
- Jolivet P, Boulard C, Bellamy A, Larré C, Barre M, Rogniaux H, d'Andrea S, Chardot T, Nesi N (2009) Protein composition of oil bodies from mature *Brassica napus* seeds. *PROTEOMICS* **9**: 3268-3284
- Jolivet P, Roux E, D'Andrea S, Davanture M, Negroni L, Zivy M, Chardot T (2004) Protein composition of oil bodies in *Arabidopsis thaliana* ecotype WS. *Plant Physiol Biochem* **42**: 501-509
- Katavic V, Agrawal GK, Hajdich M, Harris SL, Thelen JJ (2006) Protein and lipid composition analysis of oil bodies from two *Brassica napus* cultivars. *Proteomics* **6**: 4586-4598
- Kim HU, Hsieh K, Ratnayake C, Huang AHC (2002) A Novel Group of Oleosins Is Present Inside the Pollen of *Arabidopsis*. *Journal of Biological Chemistry* **277**: 22677-22684
- Krahmer N, Guo Y, Wilfling F, Hilger M, Lingrell S, Heger K, Newman HW, Schmidt-Supprian M, Vance DE, Mann M, Farese RV, Walther TC (2011) Phosphatidylcholine Synthesis for Lipid Droplet Expansion Is Mediated by Localized Activation of CTP:Phosphocholine Cytidyltransferase. *Cell Metabolism* **14**: 504-515
- Leber R, Zinser E, Paltauf F, Daum G, Zellnig G (1994) Characterization of lipid particles of the yeast, *Saccharomyces cerevisiae*. *Yeast* **10**: 1421-1428
- Leprince O, van Aelst AC, Pritchard HW, Murphy DJ (1998) Oleosins prevent oil-body coalescence during seed imbibition as suggested by a low-temperature scanning electron microscope study of desiccation-tolerant and -sensitive oilseeds. *Planta* **204**: 109-119
- Li Y, Beisson F, Pollard M, Ohlrogge J (2006) Oil content of *Arabidopsis* seeds: The influence of seed anatomy, light and plant-to-plant variation. *Phytochemistry* **67**: 904-915
- Liu Q, Sun Y, Su W, Yang J, Liu X, Wang Y, Wang F, Li H, Li X (2012) Species-specific size expansion and molecular evolution of the oleosins in angiosperms. *Gene* **509**: 247-257
- Macquet A, Ralet M-C, Kronenberger J, Marion-Poll A, North HM (2007) In situ, Chemical and Macromolecular Study of the Composition of *Arabidopsis thaliana* Seed Coat Mucilage. *Plant and Cell Physiology* **48**: 984-999
- Mak HY (2012) Lipid droplets as fat storage organelles in *Caenorhabditis elegans*. *Journal of Lipid Research* **53**: 28-33
- Murphy D (2012) The dynamic roles of intracellular lipid droplets: from archaea to mammals. *Protoplasma* **249**: 541-585

- Murphy DJ (2001) The biogenesis and functions of lipid bodies in animals, plants and microorganisms. *Progress in Lipid Research* **40**: 325-438
- Murphy DJ, Vance J (1999) Mechanisms of lipid-body formation. *Trends in Biochemical Sciences* **24**: 109-115
- Neuberger T, Sreenivasulu N, Rokitta M, Rolletschek H, Göbel C, Rutten T, Radchuk V, Feussner I, Wobus U, Jakob P, Webb A, Borisjuk L (2008) Quantitative imaging of oil storage in developing crop seeds. *Plant Biotechnology Journal* **6**: 31-45
- Paar M, Jüngst C, Steiner NA, Magnes C, Sinner F, Kolb D, Lass A, Zimmermann R, Zumbusch A, Kohlwein SD, Wolinski H (2012) Remodeling of Lipid Droplets during Lipolysis and Growth in Adipocytes. *Journal of Biological Chemistry* **287**: 11164-11173
- Parthibane V, Iyappan R, Vijayakumar A, Venkateshwari V, Rajasekharan R (2012) Serine/Threonine/Tyrosine Protein Kinase Phosphorylates Oleosin, a Regulator of Lipid Metabolic Functions. *Plant Physiology* **159**: 95-104
- Parthibane V, Rajakumari S, Venkateshwari V, Iyappan R, Rajasekharan R (2012) Oleosin Is Bifunctional Enzyme That Has Both Monoacylglycerol Acyltransferase and Phospholipase Activities. *Journal of Biological Chemistry* **287**: 1946-1954
- Preston J, Tatematsu K, Kanno Y, Hobo T, Kimura M, Jikumaru Y, Yano R, Kamiya Y, Nambara E (2009) Temporal Expression Patterns of Hormone Metabolism Genes during Imbibition of *Arabidopsis thaliana* Seeds: A Comparative Study on Dormant and Non-Dormant Accessions. *Plant and Cell Physiology* **50**: 1786-1800
- Qu RD, Huang AH (1990) Oleosin KD 18 on the surface of oil bodies in maize. Genomic and cDNA sequences and the deduced protein structure. *J Biol Chem* **265**: 2238-2243
- Schmidt MA, Herman EM (2008) Suppression of soybean oleosin produces micro-oil bodies that aggregate into oil body/ER complexes. *Mol Plant* **1**: 910-924
- Shimada T, Shimada T, Takahashi H, Fukao Y, Hara-Nishimura I (2008) A novel role for oleosins in freezing tolerance of oilseeds in *Arabidopsis thaliana*. *Plant J* **55**: 798-809
- Siloto RM, Findlay K, Lopez-Villalobos A, Yeung EC, Nykiforuk CL, Moloney MM (2006) The accumulation of oleosins determines the size of seed oilbodies in *Arabidopsis*. *Plant Cell* **18**: 1961-1974
- Soille P (2003) *Morphological Image Analysis: Principles and Applications*, Ed 2nd Vol XVI. Springer-Verlag, Berlin Heidelberg
- Sturley SL, Hussain MM (2012) Lipid droplet formation on opposing sides of the endoplasmic reticulum. *Journal of Lipid Research* **53**: 1800-1810
- Suzuki Y, Kawazu T, Koyama H (2004) RNA isolation from siliques, dry seeds, and other tissues of *Arabidopsis thaliana*. *Biotechniques* **37**: 542, 544
- Trigui G, Laroche B, Miquel M, Dubreucq B, Trubuil A (2012) The dynamics of oil bodies in *A. thaliana* seeds : A mathematical model of biogenesis and coalescence. *WASET* **64**: 763-768
- Trigui G, Miquel M, Dubreucq B, David O, Trubuil A (2013) Analysis of factors affecting the growth of oil bodies in *A.thaliana* seeds : comparison of ordinary least squares and quantile regression. *TICSP - Tampere International Center for Signal Processing* **63**: 98-101
- Tzen J, Cao Y, Laurent P, Ratnayake C, Huang A (1993) Lipids, Proteins, and Structure of Seed Oil Bodies from Diverse Species. *Plant Physiology* **101**: 267-276
- Tzen J, Huang A (1992) Surface structure and properties of plant seed oil bodies. *The Journal of Cell Biology* **117**: 327-335
- Tzen JT, Lai YK, Chan KL, Huang AH (1990) Oleosin isoforms of high and low molecular weights are present in the oil bodies of diverse seed species. *Plant Physiol* **94**: 1282-1289
- Tzen JT, Lie GC, Huang AH (1992) Characterization of the charged components and their topology on the surface of plant seed oil bodies. *J Biol Chem* **267**: 15626-15634
- Vermachova M, Purkrtova Z, Santrucek J, Jolivet P, Chardot T, Kodicek M (2011) New protein isoforms identified within *Arabidopsis thaliana* seed oil bodies combining chymotrypsin/trypsin digestion and peptide fragmentation analysis. *PROTEOMICS* **11**: 3430-3434
- Wältermann M, Hinz A, Robenek H, Troyer D, Reichelt R, Malkus U, Galla HJ, Kalscheuer R, Stöveken T, Von Landenberg P, Steinbüchel A (2005) Mechanism of lipid-body formation in prokaryotes: how bacteria fatten up. *Molecular Microbiology* **55**: 750-763
- Walther TC, Farese Jr RV (2009) The life of lipid droplets. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1791**: 459-466
- Wilfling F, Wang H, Haas JT, Krahmer N, Gould TJ, Uchida A, Cheng J-X, Graham M, Christiano R, Fröhlich F, Liu X, Buhman KK, Coleman RA, Bewersdorf J, Farese RV, Walther TC (2013) Triacylglycerol Synthesis Enzymes Mediate Lipid Droplet Growth by Relocalizing from the ER to Lipid Droplets. *Developmental Cell* **24**: 384-399
- Yang H, Galea A, Sytnyk V, Crossley M (2012) Controlling the size of lipid droplets: lipid and protein factors. *Current Opinion in Cell Biology* **24**: 509-516
- Zhang SO, Box AC, Xu N, Le Men J, Yu J, Guo F, Trimble R, Mak HY (2010) Genetic and dietary regulation of lipid droplet expansion in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences* **107**: 4640-4645

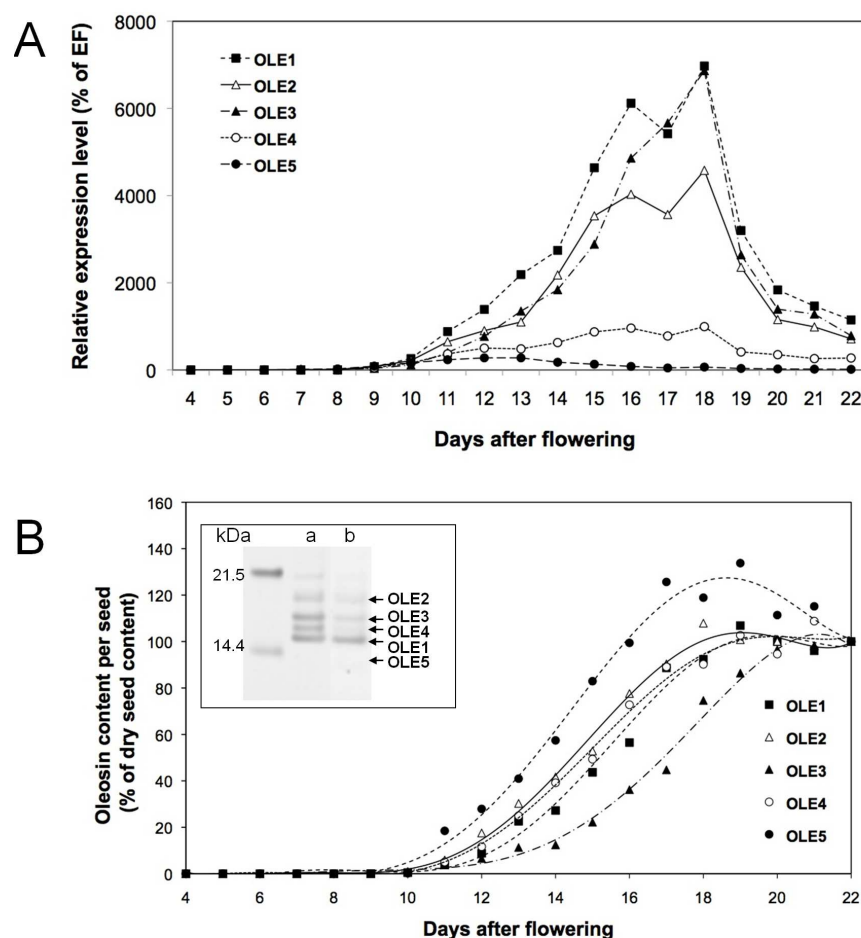


Figure 1. Oleosin accumulation pattern during seed development. Abundance of oleosin mRNA (A) and oleosin protein accumulation (B) were determined in *A. thaliana* siliques collected from 7 to 22 days (mature seeds) after flowering (DAF). Abundance of oleosin mRNA was assessed by quantitative RT-PCR and *EF1αA1* was used as control. Oleosin protein accumulation was measured by immunoblot using specific sera. The 5 seed-specific oleosins were quantified: OLE1 (■), OLE2 (△), OLE3 (▲), OLE4 (○) and OLE5 (●). Luminescent signals from immunoblots, recorded using the LAS-3000 imaging system and quantified with the MultiGauge software, were within the linear range, allowing quantitative determination of oleosin content. For each oleosin, data were expressed as the percentage of the mature seed content, regardless of its relative amount among oil body proteins (insert shows oil body proteins resolved by SDS-PAGE and stained with Coomassie blue (b) or silver nitrate (a), to illustrate the relative content of each oleosin in purified *A. thaliana* oil bodies.

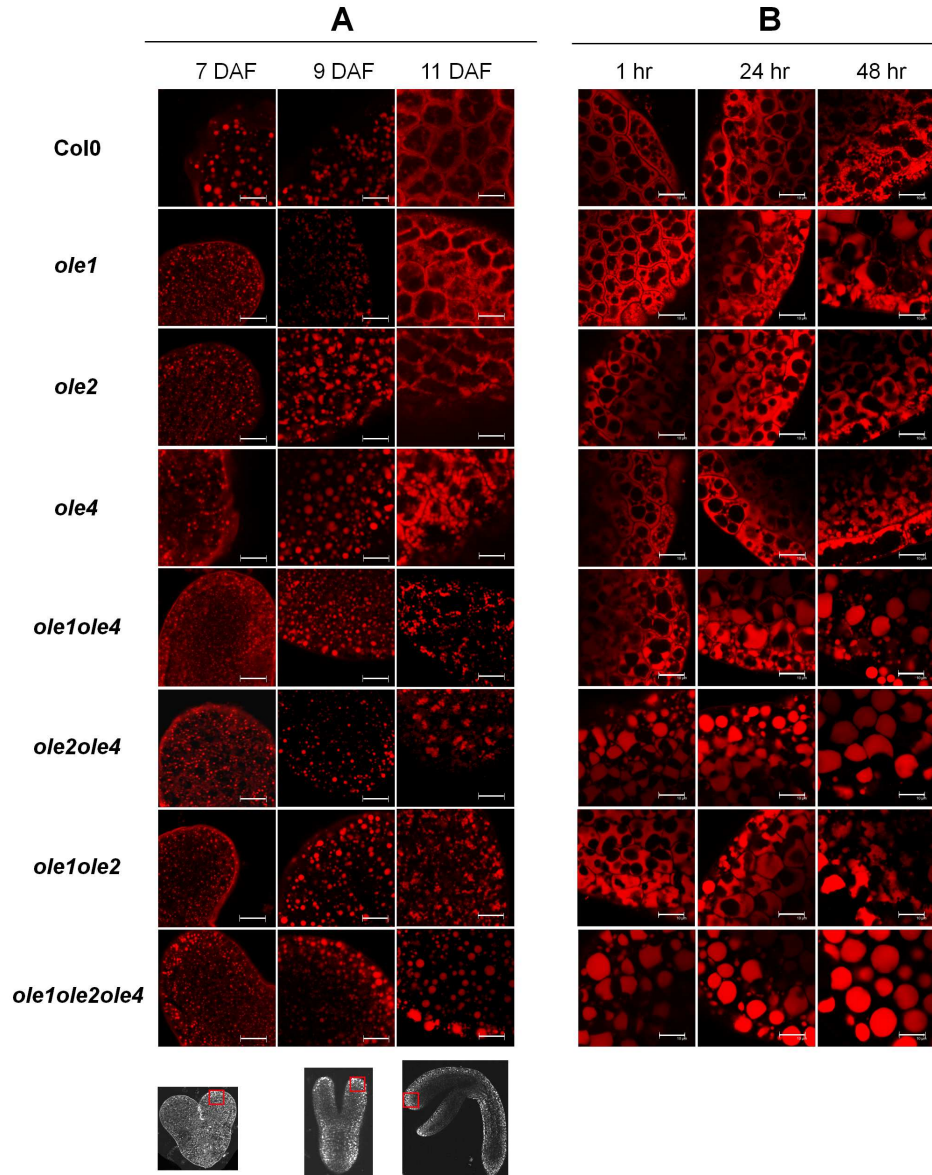


Figure 2. Phenotypes of oil bodies stained with Nile Red in different genetic backgrounds during seed maturation (A), and imbibition (B). Red squares on pictures of developing embryos at 7 9, and 11 DAF stages indicate localisation of image acquisition. DAF: days after flowering; hr: hours after imbibition start. Bar = 10 μ m.

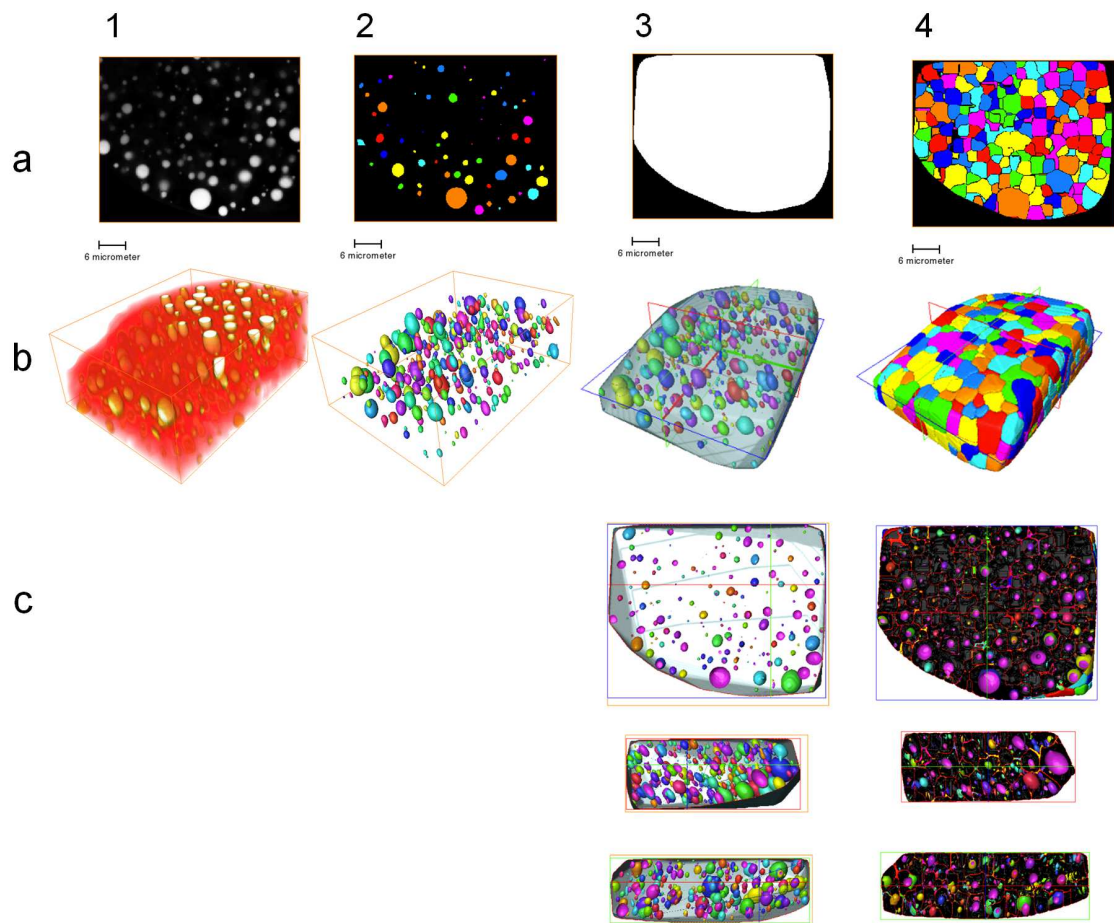


Figure 3. Sequential steps of image processing and 3D volume rendering.

This example is taken from a 10 DAF *ole1ole2ole4* sample image with a resolution of 512 x 512 x 512 pixels. a. Segmentation: Crude acquired images (1), segmented OBs (2), convex envelope (3) and Voronoi cells processing (4). b. 3D volume rendering of the image stack for each step. c. Transversal section and orthoslice of the 3D reconstruction showing individual OBs (3-c) and individual OBs with the shape of the Voronoi cells (4-c).

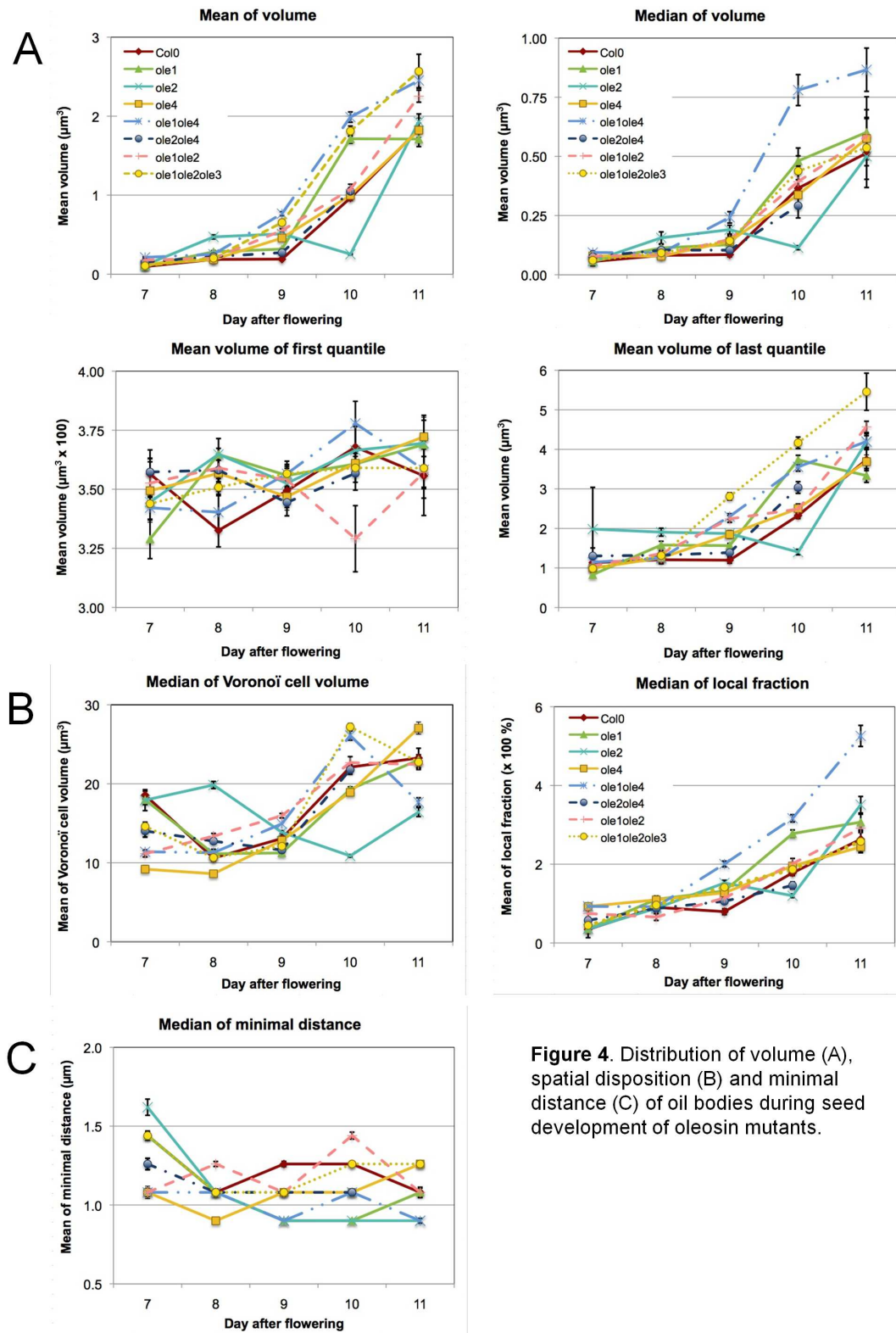


Figure 4. Distribution of volume (A), spatial disposition (B) and minimal distance (C) of oil bodies during seed development of oleosin mutants.

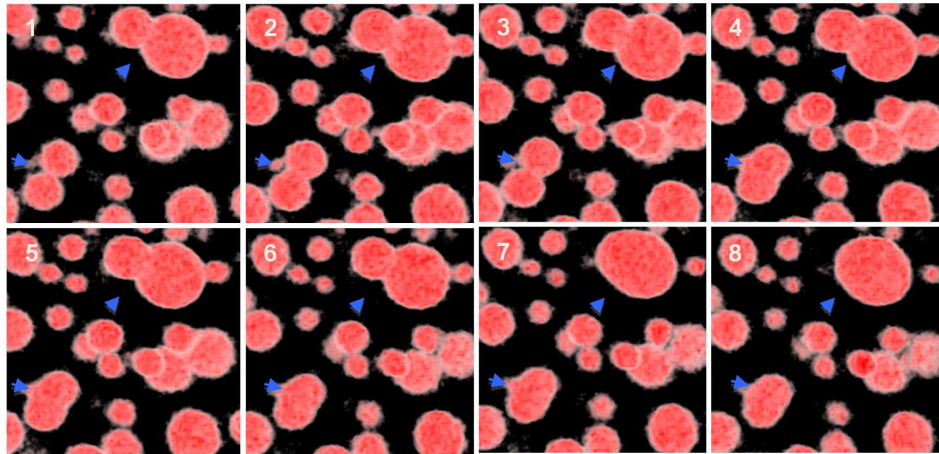
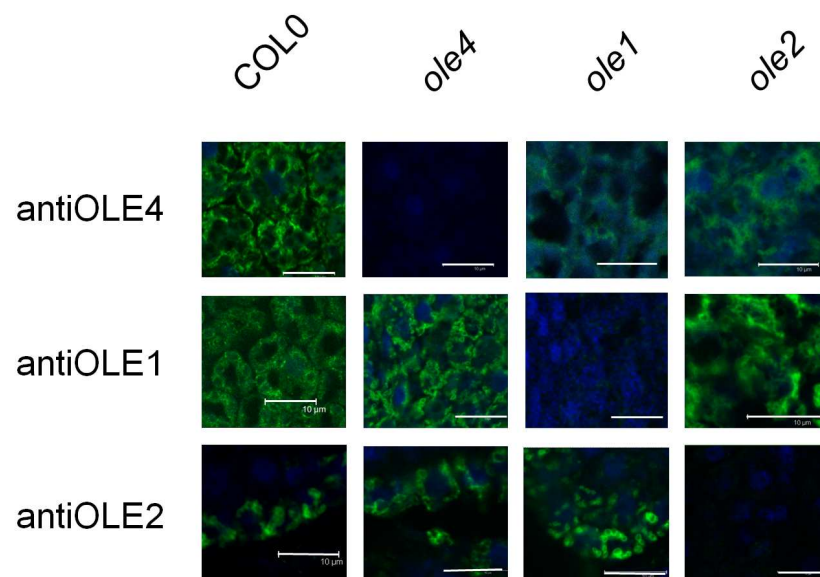
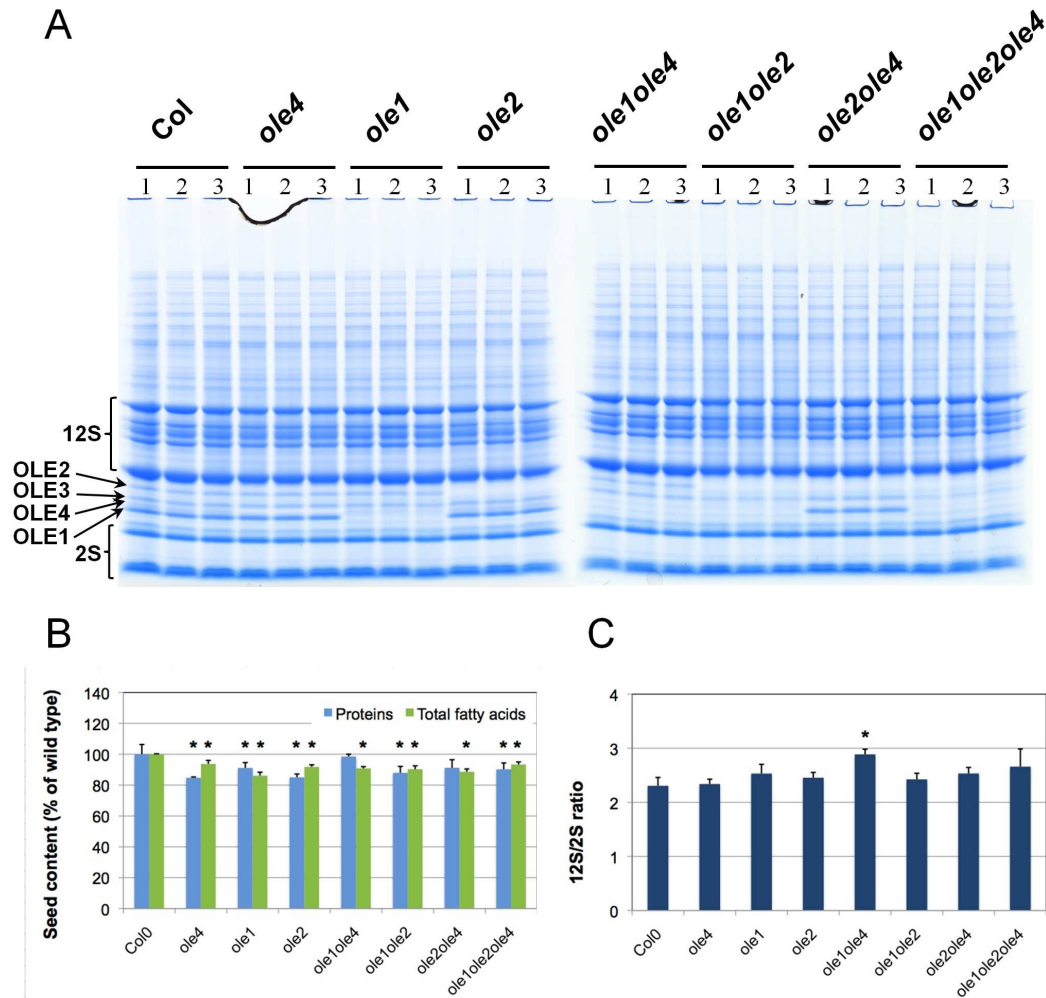


Figure 5. Fusion of oil bodies. OBs in living torpedo (10 DAF) Col0 embryos immobilized in low-melting agarose were stained with Nile red and observed with a confocal laser scanning microscope during 1 hour (1 Z-stack each 2 min). Time series presented were extracted each 7.5 min. Images stacks were processed with 3D medical image processing software OSIRIX for volume rendering.

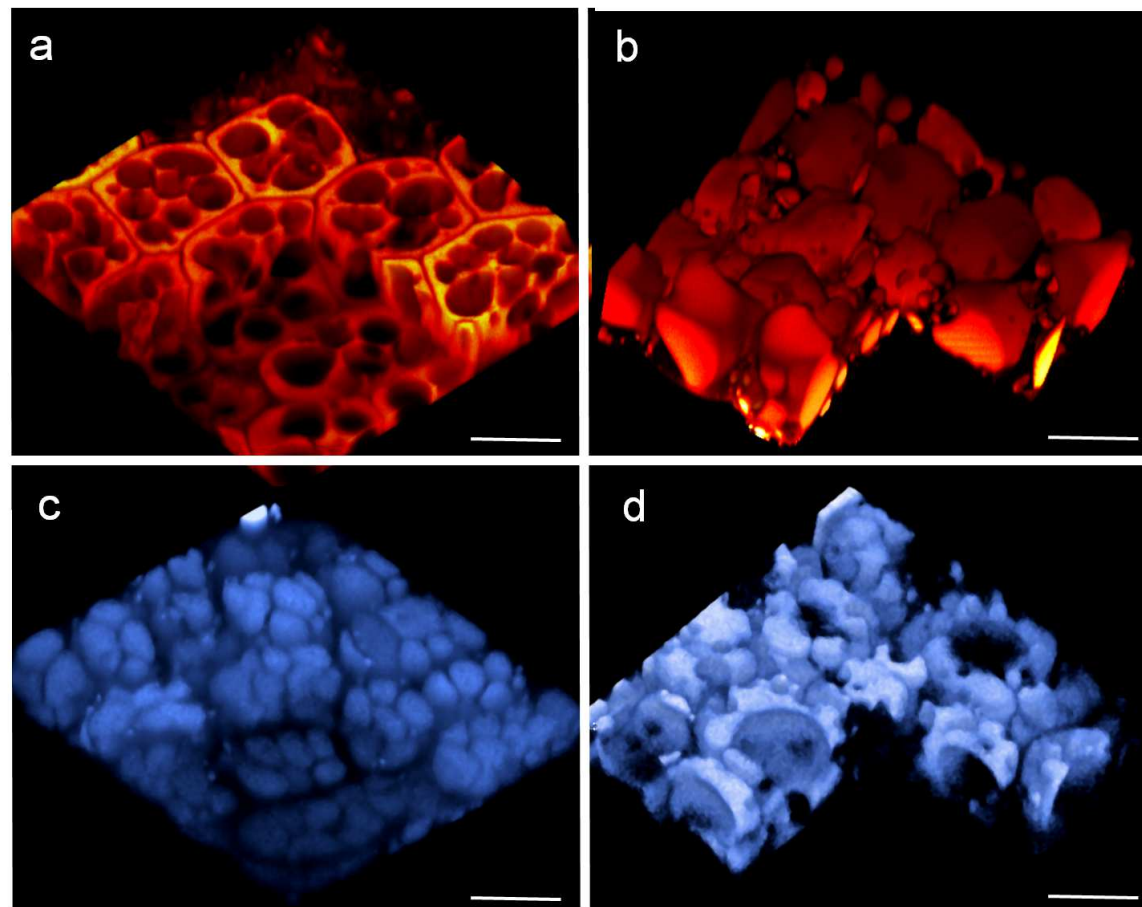
Supplemental data



Supplemental Figure S1. Validation of antibodies specificity on sections of developing siliques of mutants affected in one oleosin. Blue: DAPI signal, green: Alexa 488. Bar = 10 μ m.



Supplemental Figure S2. Effect of oleosin(s) deficiency on protein content. (A) Protein profiles of wild-type and oleosin mutants seeds. (B) Seed content in storage compounds. (C) Storage proteins 12S/2S ratio. Proteins extracted from 4 mature seeds were separated on 12% NuPAGE (Invitrogen) and stained with Coomassie blue. Total fatty acids were determined by GC following direct transmethylation. (*; $p < 0.05$).



Supplemental Figure S3. 3D reconstruction of storage lipid and protein accumulation in WT and *ole1ole2* mutant background in 1-hour imbibed seeds. In red lipids stained with Nile Red, in blue fluorescence of storage proteins. (a) Col0, Nile Red, (b) *ole1ole2*, Nile Red, (c) Col0, proteins, (d) *ole1ole2*, proteins. Bar = 10µm.

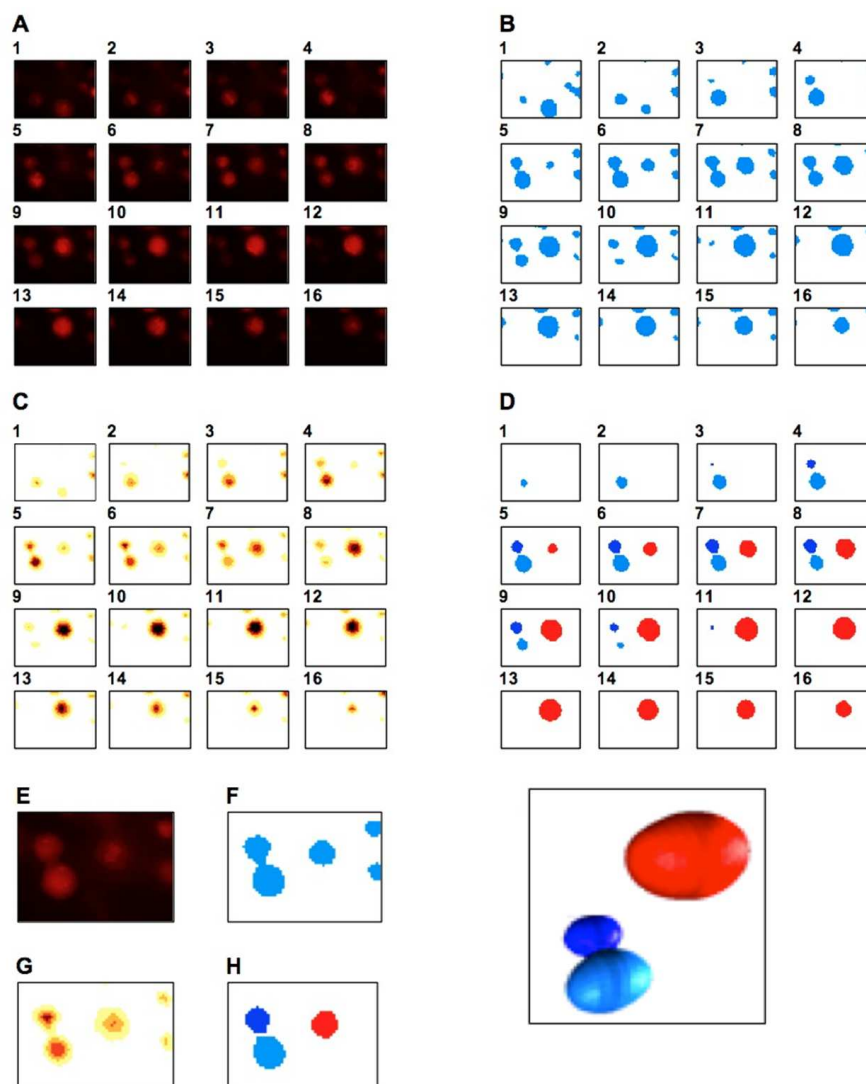
Supplemental Table S1. Adjusted parameters for the QR statistical analysis of oil body volume and local fraction factors.

Fitted estimate value of OB volume (V_{OB}) and local fraction (φ_{loc}), standard error and p -values (p) are given for factors affecting oil body volume. Intercept: 7 DAF. DAF: day after flowering.

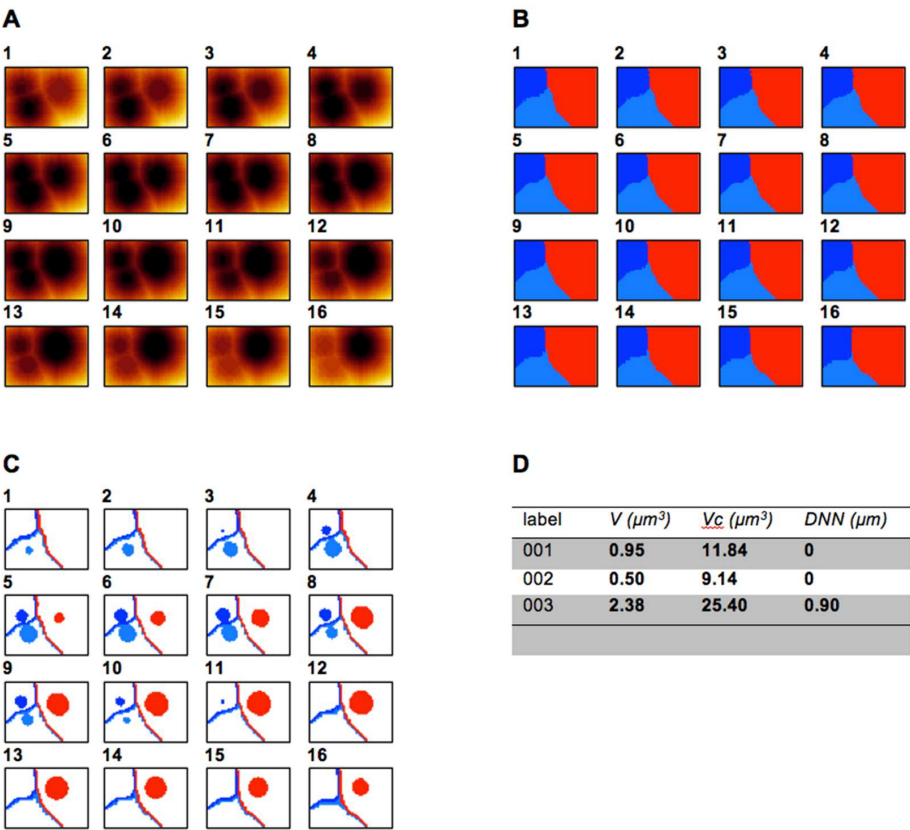
	Log (V_{OB})		Log (φ_{loc})	
	Estimate $\pm$ s.e.	P	Estimate $\pm$ s.e.	P
(Intercept)	- 1.120 $\pm$ 0.009	< 2 10^{-16}	- 2.242 $\pm$ 0.008	< 2 10^{-16}
8 DAF	0.183 $\pm$ 0.010	< 2 10^{-16}	0.230 $\pm$ 0.008	< 2 10^{-16}
9 DAF	0.361 $\pm$ 0.009	< 2 10^{-16}	0.367 $\pm$ 0.007	< 2 10^{-16}
10 DAF	0.705 $\pm$ 0.009	< 2 10^{-16}	0.505 $\pm$ 0.007	< 2 10^{-16}
11 DAF	0.892 $\pm$ 0.010	< 2 10^{-16}	0.685 $\pm$ 0.008	< 2 10^{-16}
OLE1	- 0.091 $\pm$ 0.009	< 2 10^{-16}	- 0.078 $\pm$ 0.007	< 2 10^{-16}
OLE2	0.156 $\pm$ 0.007	< 2 10^{-16}	0.155 $\pm$ 0.006	< 2 10^{-16}
OLE4	- 0.020 $\pm$ 0.008	0.005	- 0.053 $\pm$ 0.006	3.32 10^{-15}
OLE1:OLE4	- 0.028 $\pm$ 0.013	0.031	0.086 $\pm$ 0.011	2.92 10^{-15}
OLE2:OLE4	- 0.127 $\pm$ 0.012	< 2 10^{-16}	- 0.043 $\pm$ 0.010	1.40 10^{-5}
OLE1:OLE2	- 0.122 $\pm$ 0.013	< 2 10^{-16}	- 0.093 $\pm$ 0.011	< 2 10^{-16}
OLE1:OLE2:OLE4	0.126 $\pm$ 0.019	< 2 10^{-16}	- 0.094 $\pm$ 0.016	7.81 10^{-9}

Supplemental Table S2. Sequences of qRT-PCR primers used in this study

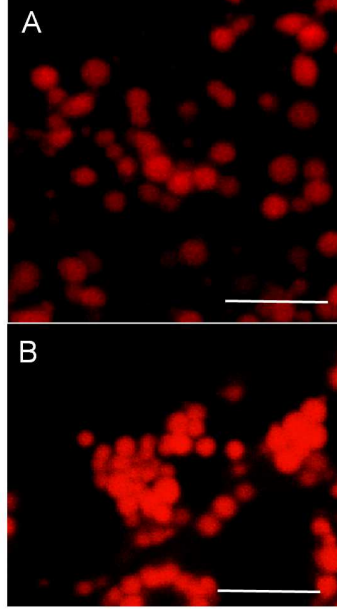
qOLE1-UP	aggcagattgctaagctgcaac
qOLE1-LOW	actgtgatgagagccggg
qOLE2-UP	aagagcatgatgcctgaaa
qOLE2-LOW	gtgaaaacacatatctaccg
qOLE3-UP	ttctgattataagagtcg
qOLE3-LOW	gatttctggttatctcaa
qOLE4-UP	ctaaagatgctggacaaa
qOLE4-LOW	tttcaaatctattacgcatcaat
qOLE5-UP	acaagaacccatcacgagatgata
qOLE5-LOW	aatgttgctgcccatactagt
EF1-ALPHA A1-UP	agacccgtgagcacgctctactt
EF1-ALPHA-A1-LOW	acggcctctgggctcgttgatct



Supplemental Figure S4. Segmentation steps of OBs using the watershed method. (A) x-y planes of 3D-stack (size: 57x40x16) filtered image I containing two nearest OBs, an isolated one and four OBs touching the borders. (B) Binarization of image I. Two values of threshold were set manually for detecting small and large OBs. (C) The complement of the distance transform of image I. Its minima (dark level) was located on the center of the OB and used as a marker for the watershed immersion. (D) Label image of OBs: each connected component of OB-object voxels was given a specific color. All objects touching borders were suppressed. (E) A selected x-y plane from the image I showing the two nearest OBs (F) and its binarization result on overlapped OBs (G), the watershed using the distance transform to perform a separation of OBs (H). (I) 3D rendering of the label image showing quasi-spherical shaped objects.



Supplemental Figure S5. Determination of the Voronoï cells and the Distance to the Next Neighbor (DNN) estimator steps. (A) The distance transform of the background of the previously segmented image I in Figure S4. (B) Label image of the Voronoï cells obtained by performing a watershed immersion on the distance transform of the background. (C) Boundaries of Voronoï cells. Each boundary has the same color level than the OB level. The mid-distance to the nearest neighbor was determined using the minimum of the distance transform (A) across the boundaries. (D) Table showing the volume of each OB, the volume of its Voronoï cell and its DNN.



Supplemental Figure S6. Distribution of OBs at 12 DAF in WT (A) and *ole2* (B) embryos. Bar = 4μm

2.5 Résultats complémentaires

En complément des résultats sur les différents estimateurs présentés dans la section précédente, nous avons exploré la relation entre deux estimateurs qui sont le volume du corps lipidique (V) et sa fraction locale (ϕ_{loc}). Cette relation reflète le pourcentage d'occupation du corps lipidique dans sa cellule de Voronoï. Une relation en loi de puissance de la forme :

$$\phi_{loc} = \theta_1 V^{\theta_2}, \quad (2.4)$$

a été mise en évidence. La transformation de l'équation 2.4 en log-log donne la loi linéaire :

$$\log \phi_{loc} = \log \theta_1 + \theta_2 \log V, \quad (2.5)$$

avec θ_2 la pente de la loi 2.5 et $\log \theta_1$ l'ordonnée à l'origine. Nous avons ensuite exploré cette tendance pour chaque génotype et à chaque jour du développement

(Figures 2.13, 2.13). La fraction locale ϕ_{loc} augmente linéairement avec le volume du corps lipidique V jusqu'à une certaine limite inférieure à la valeur maximale que peut atteindre la fraction locale égale à 1. De plus, la pente θ_2 diminue progressivement en fonction du temps. Nous avons utilisé cette relation dans la construction du modèle du chapitre 4.

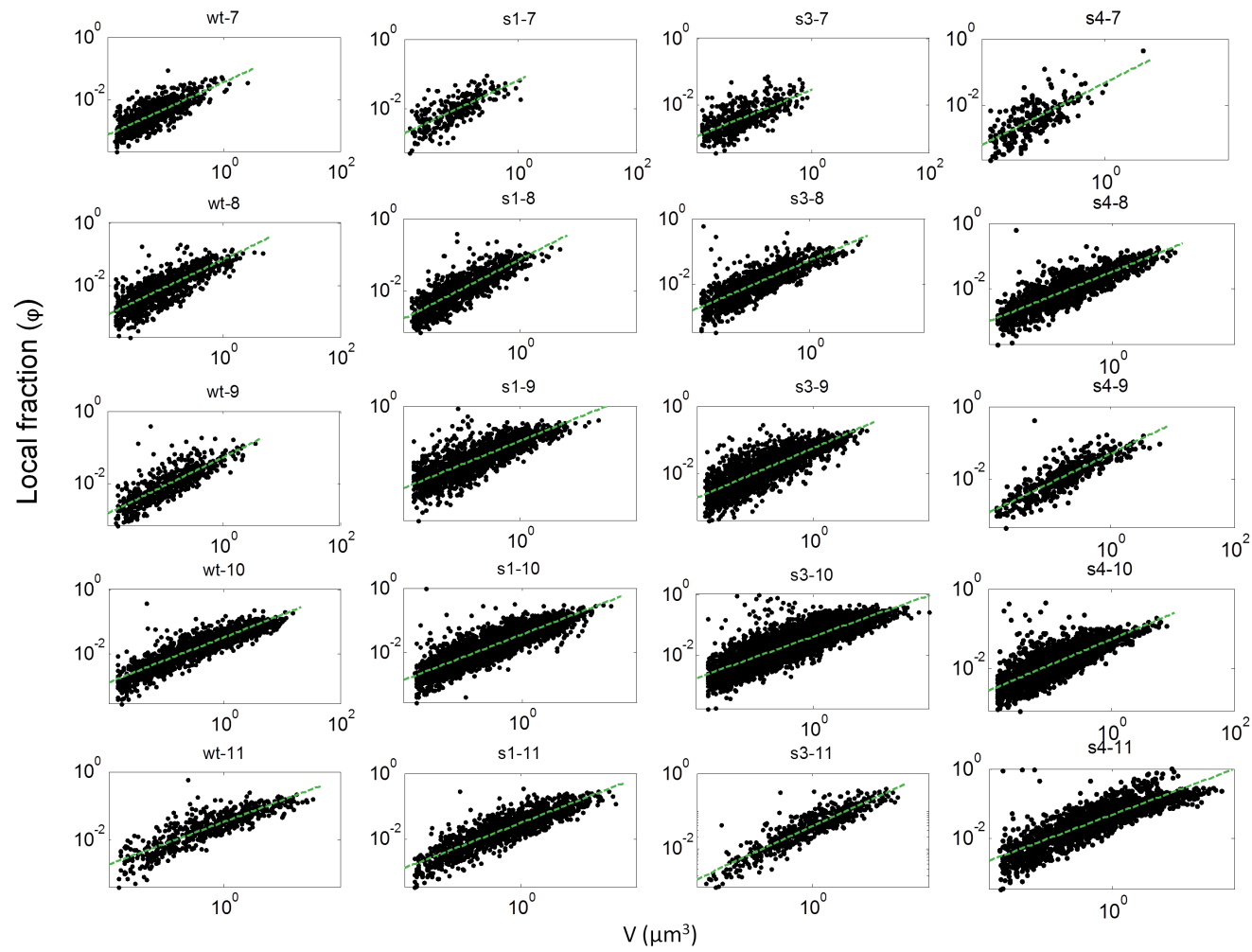


FIGURE 2.13 : Relation linéaire entre le volume V (axe des x) et sa fraction locale ϕ (axe des y) dans l'échelle log-log

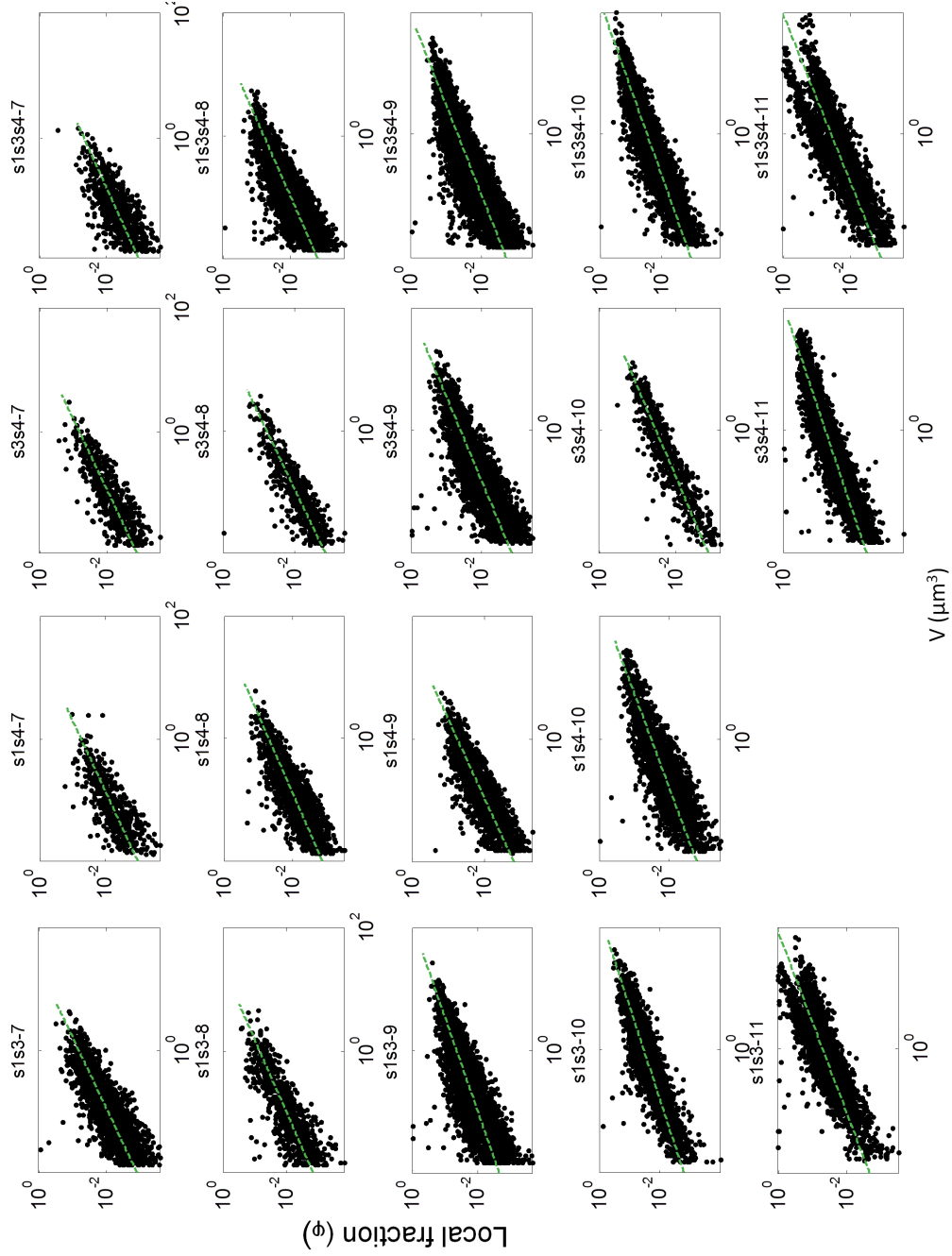


FIGURE 2.14 : Suite figure 2.13

2.6 Conclusion

La microscopie confocale est une technologie qui a révolutionné la recherche en biologie cellulaire. En effet, une reconstruction 3D rapide à partir de coupes dans le plan focal permet d'obtenir une information assez précise de la structure, la taille et la position spatiale des organites marqués par un fluorophore. Environ 200 images de microscopie confocale tridimensionnelles des corps lipidiques pour différents génotypes et à différents stades du développement d'*A. thaliana* ont été traitées. Chaque lot de 3 à 13 images est extrait de la même zone de plusieurs embryon correspondant à un génotype donné et un jour du développement donné.

Le suivi de l'évolution d'un même embryon durant toute la période de développement considérée n'a pas été possible dans le cadre de cette thèse. En effet, le suivi des embryons détachés de leurs tissus nourriciers est limitée, ce qui n'a pas permis l'obtention de données en *time lapse*. Néanmoins, quelques événements de coalescence de corps lipidiques sur des fenêtres temporelles restreintes ont pu être observés (voir 2.4). Le pipeline développé permet d'obtenir les informations extraites d'une façon semi-automatique (deux paramètres doivent être fixés au préalable par l'utilisateur, voir 2.3.2.1). En perspectives de ce travail, des données en *time lapse* d'un même échantillon seraient préférables pour permettre d'un côté, l'évaluation de la variabilité entre différents échantillons et, d'un autre côté, le suivi ainsi que la détection d'événements dynamiques (fusion, fission, naissance, etc. des corps lipidiques). Une analyse statistique plus élaborée a été développée dans le chapitre 3. Dans ce chapitre, nous étudions avec plus de détails le changement dans les distributions de tailles des corps lipidiques par l'analyse de quantiles.

Analyse statistique

3.1 Introduction

La quantification des relations liant une variable réponse aux différents co-facteurs associés est fréquemment utilisée en biologie. Parmi les outils statistiques permettant d'établir cette relation, on trouve différents modèles basés sur une approche du type régression : modèle linéaire, modèle linéaire généralisé, modèle additif, etc.

Typiquement, en régression, la variable réponse y est liée aux variables explicatives X par une relation de la forme $y = f_{\theta}(X) + \varepsilon$, où f_{θ} est une fonction, dépendante de paramètres θ à identifier et ε un bruit additif d'espérance nulle, indépendant de X . La variable réponse y est alors vue comme une réalisation d'une variable aléatoire Y et $E(Y|X = x) = f(x)$. Le plus souvent, la fonction f est identifiée par maximisation de la vraisemblance dans le cas d'un bruit gaussien *iid*¹, les paramètres estimés minimisant l'écart de la variable réponse à l'espérance conditionnelle.

Cela est particulièrement problématique pour les modèles de régression avec des variances hétérogènes. Se baser uniquement sur le changement moyen d'une distribution hétérogène peut sous-estimer, ou surestimer, des modifications non nulles apportées par des zones irrégulières dans la distribution [Cade 1999]. A l'égard de cette limite, Koenker [Koenker 2005] a proposé une méthode d'ajustement des courbes de régression sur d'autres parties de la distribution de la variable réponse, à savoir différents quantiles.

Par analogie avec les modèles de régression linéaire basés sur la minimisation de la somme des carrés des résidus, la régression quantile, quant à elle, utilise la minimisation d'un critère asymétrique avec un poids qui varie en fonction du quantile conditionnel à estimer [Cade 2003]. Contrairement à la régression linéaire, la régression quantile ne se limite pas à l'effet d'un changement marginal unique de la variable explicative par rapport à la moyenne des observations mais s'étend aux effets des co-facteurs dans d'autres endroits de la distribution, tels que le premier décile (Q10), le premier quartile (Q25) la médiane (Q50), le quantile 80e (Q80) ou le quantile 95e (Q95).

La régression quantile a d'abord été appliquée à l'économétrie, puis à d'autres domaines tels que la biologie et l'écologie (Tableau 3.1). Ce chapitre présente une analyse statistique des données extraites (cf. Chapitre 2) à partir d'images de corps lipidiques au cours du développement de l'embryon d'*A. thaliana*. L'objectif de cette

1. Indépendantes et identiquement distribuées (variables) : ce sont des variables aléatoires qui ont toutes la même loi de probabilité et sont mutuellement indépendantes.

Etude	Références
Etude des relations entre proies et prédateurs d'une population d'animaux.	[Scharf 1998]
Etude des changements de la composition des espèces végétales.	[Allen 2001]
Etude de la relation entre la concentration en oxygène dissous en eau profonde et la taille des gastéropodes.	[McClain 2001]
Relations entre la teneur en ADN nucléaire de certaines espèces végétales et les facteurs environnementaux.	[Knight 2002]
Diversité et invasibilité des espèces végétales.	[Brown 2003]
Etude des facteurs limitant le potentiel écologique des eaux courantes.	[Schmidt 2012]
Analyse des facteurs de variabilité du rendement des cultures à partir de mesures des composantes de rendement.	[Makowski 2007]

TABLE 3.1 : Exemples d'applications de la regression quantile en biologie et écologie.

étude est présenté dans la section 3.2, la méthodologie adoptée est décrite dans la section 3.3. Les résultats et la discussion sont présentés dans la section 3.4. Des analyses complémentaires sont présentées dans la section 3.5. Enfin la conclusion de ce chapitre est présentée dans la section 3.6.

3.2 Objectif

Dans le chapitre précédent, nous avons utilisé le volume comme estimateur pour caractériser l'évolution de la taille des corps lipidiques chez les différents génotypes d'*A. thaliana* durant le développement. Dans ce chapitre, nous utilisons cet estimateur comme variable d'intérêt afin d'étudier statistiquement les facteurs génotypiques associés à l'évolution de la taille (volume) des corps lipidiques. Pour cela, nous nous sommes basés sur une approche de régression quantile.

3.3 Méthodologie

La distribution de taille de corps lipidiques varie en fonction du stade de développement ainsi qu'en fonction de sa composition en oléosines. On définit la fonction quantile conditionnelle de la variable aléatoire Y conditionnellement au vecteur de variables explicatives x pour le quantile τ comme la plus petite valeur de y telle que la fonction de distribution de y conditionnellement à x soit au moins égale à τ .

Une variable aléatoire Y est caractérisée par la fonction :

$$F(y) = Pr(Y \leq y) \quad (3.1)$$

Pour chaque $\tau \in (0,1)$, le τ ème quantile de Y est défini par :

$$Q(\tau) = \inf \{y : F(y) \geq \tau\} \quad (3.2)$$

La médiane est alors $Q(1/2)$, le 1^{er} quartile est $Q(1/4)$, le 1^{er} décile est $Q(1/10)$. Les quantiles d'une variable d'intérêt Y à partir d'un échantillon $(Y_i)_{i=1..n}$ de variables

supposées *iid* peuvent être estimés en résolvant le problème d'optimisation suivant : Pour tout $\tau \in (0, 1)$, on définit la fonction :

$$\rho_\tau(u) = u(\tau - I(u < 0))u \quad (3.3)$$

Où $I(\cdot)$ est la fonction indicatrice². La solution du problème d'optimisation est alors :

$$\hat{\alpha}_\tau = \arg \min_{\xi \in \mathbb{R}} \sum_{i=1}^n \rho_\tau(y_i - \xi) \quad (3.4)$$

Considérons l'ensemble des variables explicatives x_i . Nous pouvons écrire le modèle de régression suivant :

$$y_i = x_i' \beta_\tau + u_{\tau_i} \quad (3.5)$$

où u_{τ_i} désigne le terme d'erreur. Par analogie à l'estimation de la moyenne conditionnelle,

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^*} \sum_{i=1}^n (y_i - x_i' \beta)^2 \quad (3.6)$$

La fonction du quantile conditionnel donnée par :

$$Q_y(\tau|X = x) = x_i' \beta_\tau \quad (3.7)$$

Peut être estimée par résolution de :

$$\beta_\tau = \arg \min_{\beta \in \mathbb{R}^*} \sum_{i=1}^n \rho_\tau(y_i - x_i' \beta) \quad (3.8)$$

Cette estimation peut se faire pour tout quantile d'ordre τ , où $\tau \in [0, 1]$. Il existe, en théorie, une infinité de régressions quantiles possibles. En pratique, le nombre de quantiles qu'on estime dépendra de la taille de l'échantillon. Le choix de modéliser l'ensemble des quantiles ou simplement des quantiles particuliers (déciles, quartiles) dépend du degré de précision souhaité ainsi que des données disponibles. L'interprétation des résultats de la régression quantile nécessite le choix d'un quantile et le test de significativité révèle deux situations :

- Les coefficients de la régression quantile peuvent être significativement différents de zéro.
- Les coefficients de la régression quantile peuvent être significativement différents des coefficients de la régression linéaire pour la variable à expliquer.

Sous R, un package très complet a été développé par R. Koenker : `quantreg`. La syntaxe principale s'écrit :

```
1 library(quantreg)
2 rq(y ~ x1 + x2, tau = (vecteur de quantiles), data=(table),
3 method=("br" ou "fn"))
```

2. Soit E un ensemble non vide. La fonction indicatrice d'un sous ensemble A de E notée $I(\cdot)$ est la fonction définie sur E qui vaut 1 sur A et 0 à l'extérieur de A , $I(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon} \end{cases}$

3.4 Résultats

Les résultats de ce chapitre sont présentés sous forme d'un article qui a été présenté dans une conférence internationale avec comité de lecture [Trigui 2013].

3.4.1 Résumé

Pendant le développement de la graine, les lipides sont stockés dans des structures quasi-sphériques appelées corps lipidiques. La croissance d'un corps lipidique s'effectue par la fusion avec un autre corps (données 4D d'imagerie confocale non publiées). On estime que la stabilité des corps lipidiques serait liée à leur composition qui les entoure, en protéines insérées dans la demi-membrane du corps lipidique, majoritairement des oléosines. Cinq types d'oléosines (S1 à S5) ont été identifiées [Jolivet 2009]. Nous étudions donc l'évolution du volume de corps lipidiques pendant le développement de la graine d'*A. thaliana* en fonction de la composition en oléosines. Pour cela, nous utilisons une approche modèle linéaire, complémentée par une analyse en quantile (régression quantile). Les données de volumes individualisées de corps lipidiques sont obtenues à partir d'images de microscopie confocale tridimensionnelle pour différents embryons de différents génotypes d'*A. thaliana* et en fonction du jour de développement.

3.4.2 Article

Title

Analysis of factors affecting the growth of oil bodies in *A. thaliana* seeds : the use of ordinary least squares and quantile regression

Authors

Trigui G.^{1,2}, Miquel M.², Dubreucq B.², David O.¹ and Trubuil A.^{1†}

¹ INRA, MIAJ , centre de Jouy en Josas, 78352 Jouy en Josas CEDEX France..

² INRA AgroParisTech, IJPB, UMR 1318, INRA centre de Versailles, route de Saint Cyr, 78026 Versailles CEDEX, France.

† For correspondance (alain.trubuil@jouy.inra.fr).

Abstract

The sub-cellular organelles called oil bodies (OBs) are lipid-filled quasi spherical droplets produced from the (ER) and then released into the cytoplasm during seed development. It is believed that an OB grows by coalescence with other OBs and that its stability depends on the composition in oleosins, major proteins inserted in the hemi membrane that covers OBs. Five oleosin proteins, namely S1 to S5, were discovered. The size of OBs evolves during the first steps of seed development, but is also contingent upon their protein complement. Analysis of OBs volume using ordinary least squares is limited to a point estimation of the conditional mean of the volume and so may fail to describe the evolution of OBs over all classes of volume. Thus, analysis based on quantiles can be used.

Individualized volumes of OBs were extracted from confocal microscopy images of embryos from different genotypes of *A. thaliana* seeds at different days after flowering (DAF). We used three of the five oleosins namely S1, S3, and S4 to generate null oleosin mutants. Models based on ordinary least squares (OLS) and quantile regression (QR) estimators were proposed and compared to analyze the factors associated with the growth of OBs.

Both QR and OLS indicated that the volume of OBs increases in function of time at a slower rate between 8 and 9 DAF. Whatever the estimator, S1 oleosin showed a significant effect in reducing the volume of OBs while S4 contributed to its increase. S3 was shown to act by reducing OB volume (p-values < 0.001, OLS) but only in higher ranges of volume using QR. Over all selected quantiles (in QR) and within OLS, a significant synergistic interaction between S3 and S4 was shown, while a null interaction between S1 and S4 was clearly shown within the QR in low and high (p-value = 0.69, 0.1-quantile and 0.57, 0.75-quantile respectively) volumes of OBs, as well as in OLS (p-value = 0.99).

QR provides a better understanding of the effect of factors in different ranges of data by selecting more than one quantile while OLS focuses only on the mean rather than on the tail of the distribution.

Introduction

In most eukaryotic organisms, storage lipids are deposited in stable sub-cellular structures named lipid or oil bodies (OBs). These structures are produced from the endoplasmic reticulum (ER) and then released into the cytoplasm during seed development [Murphy 1989b]. OBs differ from one species to another and between kingdoms, particularly by their composition in neutral lipids and protein complements [Huang 1996]. OBs are heterogeneous in size and in number, and exhibit growth dynamics different from one organism to another and from one cell type to another. Indeed, according to its energy needs, the cell adopts the configuration size and/or number of lipid bodies which optimizes storage capacity, production, and consumption. However, the growth mechanisms of lipid bodies are still unclear, even if different hypotheses have been proposed [Yang 2012]. In seeds, a family of proteins called oleosins have been identified on the surface of OBs [Chen 1999, Jolivet 2004]. These proteins seem to play an important role in the dynamics of OBs, and have been suggested to act as "stabilizers" preventing OBs coalescence. Studies of OBs in *A. thaliana* embryos deficient in oleosins showed impaired lipid and protein accumulation accompanied with a delay in germination, and abnormal over-sized OBs [Siloto 2006]. On the other hand, a more recent study has highlighted, by electron microscopy, the formation of OBs of 50 nm diameter on the surface of endoplasmic reticulum (ER) of transgenic soybean deficient in the 24 kDa oleosin. These "micro" OBs coalesce with adjacent OBs forming a population of heterogeneous size [Schmidt 2008]. To our knowledge, no detailed statistical analysis on the size of OBs has been carried out in the literature because of difficulties in getting size data of OB during development.

Statistical tools always utilize a model to describe experimental measurements. The conditional mean of observations is often estimated by OLS in an analysis of variance. This may not be informative enough, particularly in the case of skewed distributions, where the few values on the tails of the distribution are neglected. These values, despite their small number, represent important information when considering for example the volume of individuals in a population. In this situation, analysis of quantiles (or percentiles) can be an alternative solution to extract information from these values. Quantile regression [Koenker 1978] is a method in which one or more quantiles are used to express a linear function on a set of independent variables. The advantage of this method is that we can track the range of data in which the effect of associated covariates is impacting. Originally developed for econometrics [Levin 2001, Fitzenberger 2002], quantile regression is more and more used in biostatistics and life science fields [Makowski 2007, Gebregziabher 2011]. The aim of this paper is to study the factors involved in the growth of OBs through estimators : ordinary least squares (OLS), and quantile regression (QR).

Methods

Data acquisition *Arabidopsis thaliana* wild type and oleosin mutant plants defective for one or several oleosins were grown from seeds on soil in a greenhouse.

Upon flowering, flowers and subsequent developing siliques were tagged daily until 12 days after flower opening. Siliques for each stage of development were sampled and dissected to remove developing seeds. Seeds were spread on a glass slide, incubated with Nile Red, a neutral lipid stain at a final concentration of $1\mu\text{g}/\text{ml}$ in a 60% glycerol solution. Embryos were removed from the seed teguments by gently pressing seeds between slide and coverslip and observed after 30 min of incubation in the dark. *Arabidopsis* oleosin null mutants are available for 3 oleosins, *S1* (At3g01570), *S3* (At4g25140) and *S4* (At5g40420) (ref. NASC). Double mutants (*s1s3*, *s1s4*, *s3s4*) and a triple mutant (*s1s3s4*) have been generated in the laboratory.

3D images of dissected *Arabidopsis* embryos were acquired using a LEICA SP2(AOBS) confocal microscope with a spatial resolution of $(0.09\mu\text{m} \times 0.09\mu\text{m} \times 0.16\mu\text{m})$ in the (x, y, z) referential. Third dimension was obtained by scanning the sample through the z axis, providing a sequence of 2D images corresponding to the fluorescence emitted from the focal plane. The information obtained from one 3D-stack corresponds to the fluorescence emitted from neutral lipids in a region of the sample corresponding to one genotype after some number of days after flower opening. Images were first filtered using ND-SAFIR, a software for denoising n-dimensional images especially dedicated to microscopy image sequence analysis [Boulanger 2010], then segmented through a pipeline we designed using several algorithms from Avizo-Fire (Burlington, USA), a 3D image processing software. The pipeline was based on the topological watershed method [Beucher 1982].

Statistical analysis A total of 112 three-dimensional images from independent samples were analyzed. Each image corresponds to one of the 8 genotypes, observed at one of the 5 development stages namely day 7, 8, 9, 10, and 11 days after flowering (DAF). Two or three samples for each couple (genotype-day) were used, and from which individualized volumes of OBs were extracted. Volumes were classified on subsets of (genotype-day). Totally, 50,379 OB volume were quantified.

In the model we developed, we made the assumption that the volume of OB is affected both by the three oleosin factors (*S1*, *S3* and *S4*), and the day factor. Each oleosin factor was labeled by an index, noted i , j , and k for *S1*, *S3*, and *S4*, respectively. Each index had two levels corresponding to the presence or the absence of the oleosin. The day factor contained five levels, noted by the label t , corresponding to 7, 8, 9, 10, and 11 DAF. Volumes V of OBs were transformed to their decimal logarithm $\text{Log}_{10}(V)$ in order to verify normality assumption, and equality of variance conditions needed for OLS. The model is expressed as :

$$\begin{aligned} y_{ijk,t}^n = & \text{intercept} + \text{Day}_t + \text{S1}_i + \text{S3}_j + \text{S4}_k \\ & + \text{S1} : \text{S3}_{ij} + \text{S1} : \text{S4}_{ik} + \text{S3} : \text{S4}_{jk} \\ & + \text{S1} : \text{S3} : \text{S4}_{ijk} + \varepsilon_{ijk,t}^n \end{aligned}$$

Where $y_{ijk,t}^n$ is the value of $\text{Log}_{10}(V)$ of the n_{th} OB on the population with oleosin labels (ijk) at day level t . $\mathbf{S1}_i$, $\mathbf{S3}_j$ and $\mathbf{S4}_k$ are respectively the main effects of factors S1, S3, and S4. $\mathbf{S1}:\mathbf{S3}_{ij}$, $\mathbf{S1}:\mathbf{S4}_{ik}$, and $\mathbf{S3}:\mathbf{S4}_{jk}$ are the effects of double interactions between oleosins. $\mathbf{S1}:\mathbf{S3}:\mathbf{S4}_{ijk}$ is the effect of the triple interaction between S1, S3 and S4. Last, $\varepsilon_{ijk,t}^n$ is the error term which is supposed to follow a normal distribution with mean 0 and variance σ^2 .

The OLS estimator uses the minimization of the sum of squares to fit predictions to observations :

$$\tilde{\theta} = \underset{\theta}{\operatorname{argmin}} \left[\sum_{n=1}^N [y^n - f(\theta, n)]^2 \right] \quad (3.9)$$

Where y^n denotes the responses for observation n , θ the vector of parameters to be estimated, and $f(\theta, n)$ the model used for each observation n .

Unlike OLS, QR utilizes the minimization of :

$$\begin{aligned} \tilde{\theta}_\tau = \underset{\theta}{\operatorname{argmin}} \bigg[& \sum_{n: y^n \geq f(\theta, n)} \tau |y^n - f(\theta, n)| \\ & + \sum_{n: y^n < f(\theta, n)} (1 - \tau) |y^n - f(\theta, n)| \bigg] \end{aligned} \quad (3.10)$$

for a given quantile τ . We used QR with five selected quantiles denoted τ_1 , τ_2 , τ_3 , τ_4 and τ_5 for respectively 0.1, 0.25, 0.5 (the median), 0.75, and 0.9 quantiles.

All analyzes were done using R project. The model was implemented in the function `rq` of the `quantreg` R-package for the QR estimations.

The adjusted parameters and their p-values for both QR quantiles and OLS mean are given in table 3.2.

Results and discussion

Focusing on oleosin factors using the QR estimator (Table 3.2), the effect of S1 oleosin was statistically significant for all quantiles (p-value < 0.001) and with negative values of parameters decreasing with the increase of quantile (Figure 3.1). This reflects the contribution of S1 oleosin to the reduction of the volume of OB. Similarly, S3 oleosin factor had a significant effect of OB reduction, but only at higher OBs volume (τ_4 , τ_5) and was not significant for lower volumes (p-value = 0.99, τ_1 ; p-value = 0.64, τ_2 and p-value = 0.28, τ_3). A significant effect of reduction was also shown for S3 with the OLS. The effect of these two oleosins was clearly observed in the predictions sorted by genotype where the fitted values of OBs volumes were always higher in genotypes lacking S1 and S3 oleosins (*s1s3* and *s1s3s4*) compared to those where S1 and S3 oleosins were present (Figure 3.2). Unlike S1 and S3, S4 oleosin factor impact was significant with a positive effect (increase of OB volume) in all quantiles and within OLS estimator, except in τ_5 (p-value = 0.19). It is noteworthy that the effect of increasing OB volume brought by the presence of S4

is smaller than the effect of reduction due to the presence of S1. This is shown by the low positive parameters values of S4 factor compared to the high negative parameters values of S1 factor (Figure 3.1). A significant synergistic effect of S3 and S4 interaction was shown, thus shedding light on the effect of interactions that also participate in determining OB volume. OB volume was reduced even more when S3 and S4 were both present. This interaction effect was visible both with the QR and the OLS estimators (p-values < 0.01). The other interaction effects were less significant for S1 and S3, except in τ_4 , and τ_5 . A null interaction between S1 and S4 was clearly shown both within the QR in low (p-value = 0.69, τ_1) and high quantiles (p-value = 0.57, τ_4) as well as with the OLS (p-value = 0.99). The triple interaction between S1, S3 and S4 was significant only in middle quantiles, and had no effect for higher volumes (p = 0.89, τ_5).

An interesting scenario for the dynamics of OBs can be hypothesized from this study and the choice of making two models may be essential in order to verify resulting assumptions. The heterogeneous distribution of OBs requires the use of QR model instead of OLS model since we were able to access more detailed effects of factors, while OLS model only indicated the differences on the central portion of the distribution. Particularly, QR model had the advantage of showing the significance of factors along quantiles. Our results show that the OB volume changes in function of time (between 7 and 11 DAF) and in function of the composition in oleosin proteins. The growth rate of the OB volume is shown to be reduced between 8 and 9 DAF, then increases between 9 and 10 DAF (Figure 3.2). While S1 and S3 oleosins have an effect on reducing the volume of OBs when they are present on the surface of OBs, S4 seems to be implicated on the increase of OB volumes except for higher ranges of volumes. This role of S1 and S3 described by our analysis correlates with the hypothesis that the main function of oleosins is to prevent the coalescence of neighboring OBs. One can also hypothesize that S4 could be involved in bringing OBs close enough so that their coalescence becomes easier. The null effect of S4 in higher volumes may be explained by the fact that an increased density of oleosins at the surface of OB prevents their coalescence when two OBs get closer. Furthermore, S1 and S4 do not interact and their effects seem to be only additive. Last, S3 and S4 present a high interaction effect of OB reduction. This may reflect an opposite action of S3 when S4 acts to increase OB volume.

Conclusion

The population of OBs in the cellular pool of *A. thaliana* embryos is heterogeneous. It is mainly composed of small OBs less than $1 \mu m^3$ volume and few large OB up to $20 \mu m^3$ volume. The latter has non-negligible weight regarding the total triacylglycerols. The growth mechanism of OBs is not fully understood, but many factors are suggested to be involved. Particularly, membrane composition may act on the dynamics of OBs including production, storage and mobilization of triacylglycerols. Furthermore, elements constituting the membrane follow themselves dynamical behaviors e.g. interaction and diffusion. In plant seed cells, electrostatic and/or steric

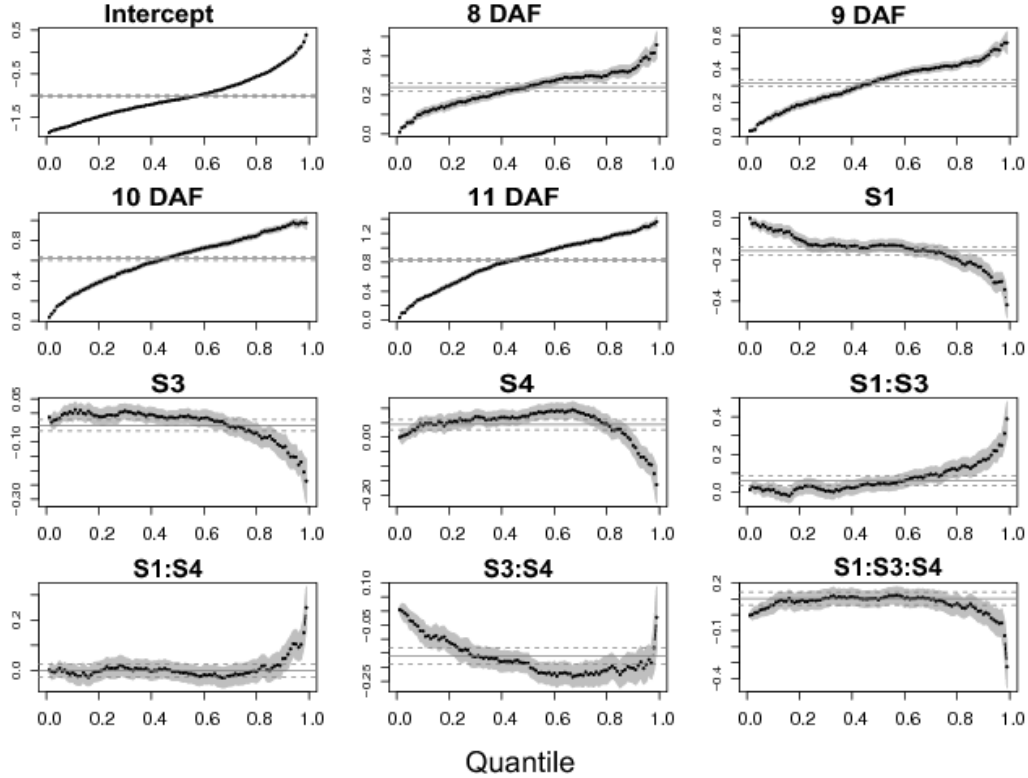


FIGURE 3.1 : OLS and QR coefficients estimation : Coefficients estimation showing the effect of day, oleosin, and interactions between oleosins. OLS estimator coefficients (horizontal solid lines) and QR estimator coefficients (dashed dotted lines) are presented with their 95% confidence interval.

interactions may occur between oleosins on the surface of OBs [Tzen 1992a]. Moreover, the coalescence process causes an increase of the surface density. Following these facts, one can suggest the probable variable effects of oleosins in function of OB size. Using OLS, we only focus on the effect of oleosins on the mean volume of OBs and disregard their variable effect due to the growth of OBs by coalescence. The QR model provides a better understanding of the effect of oleosins in different ranges of OB volumes. For example, OLS revealed that S3 oleosin reduces significantly the size of OBs (Table 3.2). However, QR showed that this effect is only significant on large OBs (τ_4, τ_5) and has no effect on small OBs, and this may be explained by the fact that S3 is only functional when reaching a certain surface coverage. The same behavior is observed with respect to S1 and S3 interactions (Table 3.2). Understanding the parameters controlling the growth of OBs may open new perspectives on oil storage and extraction techniques improvements. This statistical study will help in the design of a mechanistic model of OBs dynamics.

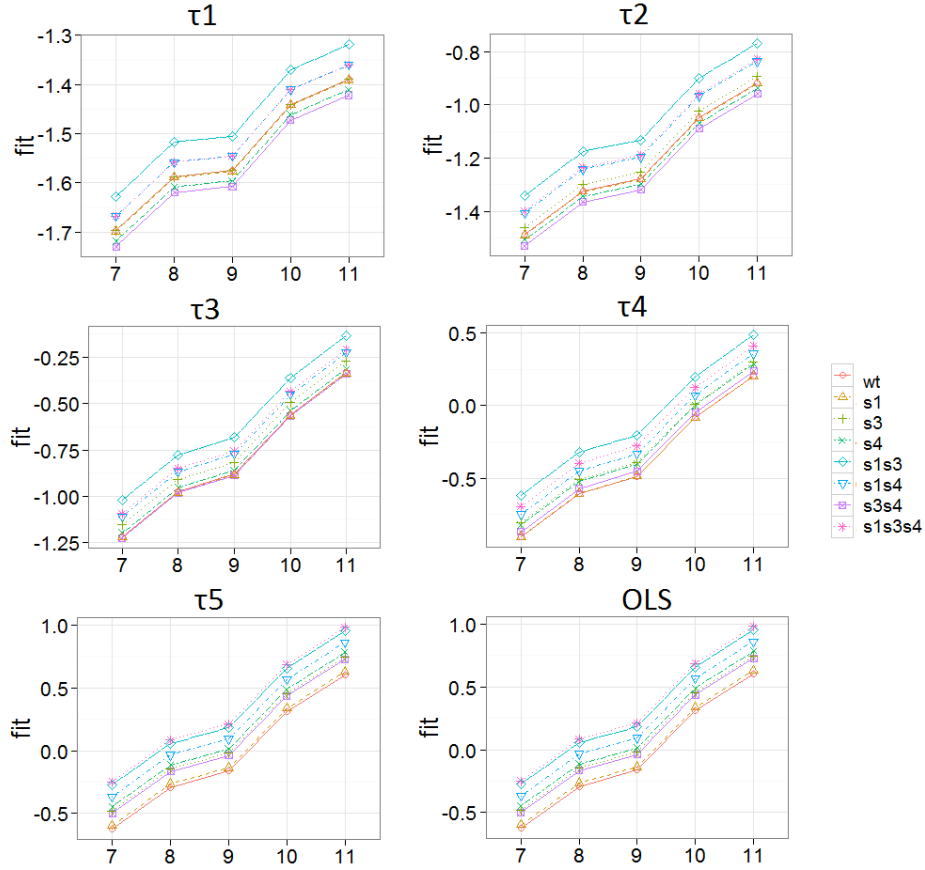


FIGURE 3.2 : Predicted values of $\text{Log}_{10}(V)$ by date sorted by genotype for each estimator (QR and the selected quantiles of OLS) : Predicted values for wt and null oleosin mutant are deduced from combining the effect of each oleosin factor and their interactions respectively : wt (the effect of S1, S3, S4, S1 :S3, S1 :S4, and S1 :S3 :S4), s1 (the effect of S3, S4, and S3 :S4), s3 (the effect of S1, S4, and S1 :S4), s4 (the effect of S1, S3, and S1 :S3), s1s3 (the effect of S4), s1s4 (the effect of S3), s3s4 (the effect of S1) and s1s3s4 (intercept).

	QR										OLS	
	τ_1		τ_2		τ_3		τ_4		τ_5			
Parameter	value	P	value	P	value	P	value	P	value	P	value	P
Intercept	-1.66 (0.01)	0.00	-1.39 (0.01)	0.00	-1.09 (0.01)	0.00	-0.69 (0.01)	0.00	-0.24 (0.02)	0.00	-1.01 (0.01)	0.00
8DAF	0.10 (0.01)	0.00	0.16 (0.01)	0.00	0.24 (0.01)	0.00	0.29 (0.01)	0.00	0.33 (0.01)	0.00	0.23 (0.01)	0.00
9DAF	0.12 (0.01)	0.00	0.20 (0.01)	0.00	0.33 (0.01)	0.00	0.41 (0.01)	0.00	0.46 (0.01)	0.00	0.31 (0.01)	0.00
10DAF	0.25 (0.01)	0.00	0.43 (0.01)	0.00	0.65 (0.01)	0.00	0.82 (0.01)	0.00	0.93 (0.01)	0.00	0.62 (0.01)	0.00
11DAF	0.30 (0.01)	0.00	0.57 (0.01)	0.00	0.88 (0.01)	0.00	1.10 (0.01)	0.00	1.22 (0.01)	0.00	0.83 (0.01)	0.00
S1	-0.06 (0.01)	0.00	-0.13 (0.01)	0.00	-0.13 (0.01)	0.00	-0.17 (0.02)	0.00	-0.25 (0.02)	0.00	-0.15 (0.01)	0.00
S3	0.00 (0.01)	0.99	0.00 (0.01)	0.64	-0.01 (0.01)	0.28	-0.05 (0.02)	0.00	-0.11 (0.02)	0.00	-0.04 (0.01)	0.00
S4	0.04 (0.01)	0.00	0.05 (0.01)	0.00	0.07 (0.01)	0.00	0.07 (0.01)	0.00	-0.02 (0.02)	0.19	0.04 (0.01)	0.00
S1 :S3	0.01 (0.02)	0.63	0.02 (0.02)	0.21	0.04 (0.01)	0.03	0.10 (0.02)	0.00	0.17 (0.03)	0.00	0.05 (0.01)	0.00
S1 :S4	0.00 (0.02)	0.69	0.00 (0.02)	0.69	0.00 (0.01)	0.91	-0.01 (0.02)	0.57	0.04 (0.02)	0.10	0.00 (0.01)	0.99
S3 :S4	-0.07 (0.02)	0.00	-0.14 (0.02)	0.00	-0.18 (0.01)	0.00	-0.22 (0.02)	0.00	-0.20 (0.03)	0.00	-0.15 (0.01)	0.00
S1 :S3 :S4	0.06 (0.03)	0.05	0.09 (0.03)	0.00	0.09 (0.01)	0.00	0.08 (0.03)	0.01	0.00 (0.04)	0.89	0.10 (0.02)	0.00

TABLE 3.2 : Adjusted parameter estimate value (standard error), and their p-values P given for each estimator (QR with selected quantiles, and OLS) : QR = quantile regression, OLS = ordinary least square, $\tau_1 = 0.1$ quantile, $\tau_2 = 0.25$ quantile, $\tau_3 = 0.5$ quantile, $\tau_4 = 0.75$ quantile, $\tau_5 = 0.9$ quantile.

3.5 Résultats complémentaires

Une approche assez simple, pour caractériser la différence génotypique, consiste à comparer les histogrammes de volume de corps lipidiques deux à deux. Le test de Kolmogorov-Smirnov à deux échantillons est un test non paramétrique qui permet de comparer les fonctions de répartition issues de deux jeux de données différents. En effet, soit $(X_1, \dots, X_n)$ un échantillon de fonction de répartition F , et $(Y_1, \dots, Y_n)$ un échantillon de fonction de répartition G . On suppose que les deux échantillons sont indépendants et que F et G sont continues. On veut tester $H_0 : F = G$ contre $H_1 : F \neq G$. Pour se faire, on mesure l'adéquation de F à G par la distance de Kolmogorov-Smirnov :

$$D_{n,m} = \sup_{x \in \mathbb{R}} |F_n(x) - G_m(x)| \quad (3.11)$$

L'hypothèse H_0 est rejetée si $D_{n,m} \geq d_{n,m,1-\alpha}$. Nous avons calculé les *p-values* du test de significativité pour chaque couple de génotype-jours à comparer. Chaque

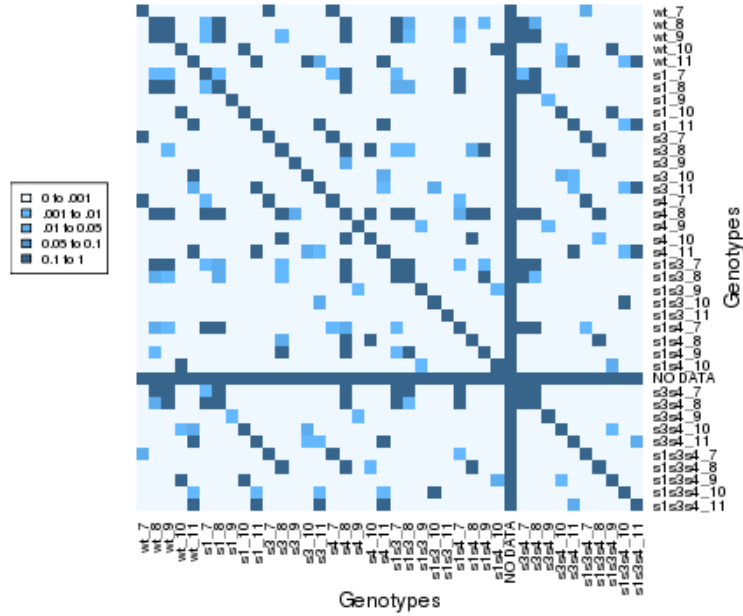


FIGURE 3.3 : Test de Kolmogorov-Smirnov de comparaison entre les distributions de volumes de corps lipidiques.

valeur est convertie en code couleur exprimant le degré de significativité (Figure 3.3). Les zones sombres indiquent que la différence entre deux distributions de corps lipidiques est non-significative. Il est particulièrement remarquable que la différence entre les distributions de corps lipidiques chez les deux génotypes *s1* et *s3s4* est non-significative. Par ailleurs, il est évident que les zones claires, qui expriment la significativité de la différence entre différentes distributions de corps lipidiques, dominant ce "paysage".

3.6 Conclusion

La distribution de volume de corps lipidiques possède un aspect hétérogène (figure A.1). Au-delà de la différence en effectif entre classes de volumes, l'hétérogénéité concerne la différence génotypique et le stade de développement. Se baser uniquement sur l'évolution du volume moyen introduit une perte d'information sur le changement de comportement des distributions et peut ainsi conduire à de fausses conclusions sur l'effet de chaque oléosine sur le volume des corps lipidiques. Nous avons étendu notre analyse aux différents classes de sous-populations (quantiles).

Cette approche nous a permis de tirer des conclusions sur l'effet des cofacteurs (oléosines) sur le changement de distribution de volumes de corps lipidiques et de proposer ainsi des hypothèses sur le rôle de chaque oléosine. Ce travail a montré que les oléosines, bien que structurellement proches, ne jouent pas le même rôle dans la dynamique du corps lipidique et de confirmer certaines relations énoncées à la fin du chapitre 2.

Modélisation de la dynamique des corps lipidiques

La modélisation statistique développée dans les chapitres 2 et 3 ne fait pas d'hypothèse sur les mécanismes sous-jacents à l'observation des populations de corps lipidiques qui évoluent dans l'espace et le temps pour les différents génotypes-stades de développement étudiés. Elle se contente de supposer de possibles effets des génotypes et des stades et propose un résumé des données (observations) autour d'une valeur moyenne construite en fonction de la présence, ou non, de l'une ou l'autre des oléosines. Dans le cas de la régression quantile, ce sont un ou plusieurs quantiles qui sont expliqués. Finalement, beaucoup de l'information présente dans les données est alors oubliée. Afin d'étudier plus finement des populations qui évoluent dans l'espace et le temps, au moins deux approches peuvent être envisagées. Toutes deux reposent sur une description des mécanismes qui conduisent à l'évolution des populations. L'approche multi-agents consiste à représenter chaque élément (agent) de la population par un vecteur de variables d'états et à faire évoluer cette population en fonction de processus appliqués à chacun de ces agents. Des statistiques pourront alors résumer les propriétés essentielles de cette population. Une autre approche consiste à modéliser directement l'évolution de la distribution d'effectifs de la population éventuellement structurée selon un faible nombre de variables. C'est le point de vue que nous avons adopté par la suite. Il permet de ne pas prendre en compte explicitement les variables d'espace.

4.1 Introduction

La modélisation de dynamiques de populations au travers de leur effectif est abondamment utilisée depuis fort longtemps. En 1202, Fibonacci fut le premier à décrire la croissance d'une population par une suite d'entiers. L'introduction du taux d'accroissement des fonctions [Euler 1767], ainsi que des bases mathématiques du calcul différentiel [Malthus 1798, Lotka 1925, Volterra 1926] ont ouvert de nouveaux horizons pour la dynamique des populations [Hanski 1999], notamment pour des populations structurées en espèces [Lotka 1925], en âge [Leslie 1945], en état sanitaire [Hethcote 2000], etc. En physique aussi, de nombreuses situations sont appréhendées sous cet angle, notamment avec l'équation d'agrégation-désagrégation par fusion-fission de Smoluchowski [Von Smoluchowski 1917], qui rend compte de l'évolution d'une population d'agrégats structurée en taille et pouvant évoluer par

fusion de deux agrégats ou fission. Cette équation est à la base du modèle développé dans ce chapitre pour rendre compte de la dynamique des corps lipidiques.

Le chapitre commence, en section 4.2, par un résumé des connaissances autour de l'équation de Smoluchowski. Quelques modèles d'agrégation existants sont alors présentés en section 2. En Section 3, nous proposons un modèle d'agrégation dans le cadre de la biogenèse des corps lipidiques. La discrétisation du modèle et son implémentation numérique sont décrites en section 4. En section 5, nous présentons la méthode d'estimation des paramètres. Enfin, l'estimation des paramètres et l'interprétation des résultats est faite en section 6.

4.2 Equation d'agrégation de Smoluchowski

Les phénomènes d'agrégation sont très abondants dans la nature (cf. 4.3.2). Ils sont aussi parfois accompagnés de phénomènes de fission ainsi que d'autres phénomènes, de croissance par exemple.

L'équation de coagulation de Smoluchowski [Von Smoluchowski 1917] décrit l'évolution temporelle d'une densité de nombre de particules, de caractéristique $x \in \mathbb{R}^q$, qui s'agrègent au cours du temps. Une caractéristique usuelle est la taille d'une particule mais parfois la caractéristique englobe plusieurs composantes comme la taille, la composition chimique, etc. L'équation de coagulation de Smoluchowski, dans sa version continue, est de la forme :

$$\frac{\partial n(x, t)}{\partial t} = \frac{1}{2} \int_0^x K(x-y, y) n(x-y, t) n(y, t) dy - \int_0^\infty K(x, y) n(x, t) n(y, t) dy \quad (4.1)$$

où $n(x, t) = \frac{N(x, t)}{V}$ et $N(x, t)$ représente le nombre de particules de taille x contenues dans un volume de référence V au temps t . L'opérateur linéaire de coalescence (ou noyau) $K : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ est le taux de coagulation entre des particules de taille x et y ¹. Le noyau K est non négatif, $K(x, y) \geq 0, \forall (x, y)$ et symétrique, $K(x, y) = K(y, x), \forall (x, y)$.

Le premier terme du membre de droite décrit la formation d'une particule de taille x à partir de l'agrégation de deux particules de tailles respectives $x-y$ et y . Le second terme décrit la disparition d'une particule de taille x suite à l'agrégation avec une particule de taille quelconque, positive. On suppose donc que l'agrégation de deux particules de taille respective x et y conduit à une particule de taille $x+y$. On s'attend donc à ce que la taille totale de l'ensemble des particules soit conservée au cours du temps, soit :

$$\int_0^\infty x n(x, t) dx = \int_0^\infty x n(x, t=0) dx, \quad \forall t \geq 0 \quad (4.2)$$

1. $n(x, t)V \times n(y, t)V$ est le nombre de couples de particules de taille x et y qui peuvent s'agréger, on en retient seulement une fraction $\tilde{K} \leq 1$, ce qui en densité revient à faire varier $n(x+y)$ de $\frac{K(\tilde{x}, y)n(x, t)V \times n(y, t)V}{V} = \tilde{K}n(x, t) \times n(y, t)V$ et donc $K(x, y) = K(\tilde{x}, y)V$. $K(x, y)$ représente donc la limite, lorsque δt tend vers 0 du nombre de particules de taille x et y agrégées dans le volume V dans le temps Δt .

Cependant, il est connu que pour certains noyaux d'agrégation augmentant suffisamment vite pour des tailles x et y grandes² il peut y avoir des pertes de taille totale en temps fini; ce phénomène est connu sous le nom de gélation³; il a été très étudié par des physiciens [Dongen 1988] et mathématiciens [Escobedo 2002].

Ce modèle d'agrégation ignore donc la localisation spatiale des particules et les éventuelles fluctuations spatiales de la densité.

Lorsque les tailles des particules appartiennent à un ensemble discret de tailles $x_i, i \in \mathbb{N}$, l'équation de coagulation sous sa forme discrète est :

$$\frac{\partial n(x_i, t)}{\partial t} = \frac{1}{2} \sum_{j=1}^{i-1} K(x_i - x_j, x_j) n(x_i - x_j, t) n(x_j, t) \quad (4.3)$$

$$- \sum_{j=1}^{\infty} K(x_i, x_j) n(x_i, t) n(x_j, t) \quad (4.4)$$

Ces équations ont donné, et continuent de donner, lieu à de nombreux travaux relatifs à l'existence et l'unicité de solutions, dans un cadre plus ou moins général ainsi qu'à des travaux relatifs aux propriétés des solutions, ou encore à l'approximation numérique des solutions. On rappellera ici ces propriétés essentielles.

4.2.1 Existence, unicité de solution pour l'équation continue

On peut trouver des preuves d'existence de solution dans les articles suivants [Ball 1990, Dubovskii 1996].

Dubovskii et al [Dubovskii 1996] considèrent l'équation en variables continues d'agrégation-fragmentation, soit :

$$\frac{\partial n(x, t)}{\partial t} = \frac{1}{2} \int_0^x K(x - y, y) n(x, t) dy + \int_0^\infty K(x, y) n(x, t) n(y, t) dy \quad (4.5)$$

$$- \frac{1}{2} \int_0^x F(x - y, y) n(x - y, t) n(y, t) dy + \int_0^\infty F(x, y) n(x + y, t) dy$$

$$n(x, 0) = n_0(x) \geq 0 \quad (4.6)$$

où $F(x, y)$ est le taux de fragmentation d'une particule de taille $x + y$ en deux particules de taille respective x et y . Nous rappellerons ici quelques théorèmes essentiels, d'une façon précise mais renvoyons le lecteur intéressé aux articles cités pour une meilleure compréhension des hypothèses sous-jacentes.

Théorème 4.2.1. *Soient K et F des noyaux continus, non négatifs et symétriques. Supposons que K vérifie :*

$$K(x, y) \leq k(1 + x + y), \quad k > 0$$

2. $K(x, y) \geq (xy)^\alpha$ avec $\alpha > \frac{1}{2}$

3. Le nom vient de la solidification d'un liquide, c'est-à-dire du groupement de toutes les molécules en un seul bloc de très grande masse.

et qu'il existe des constantes m , m_1 et b telles que

$$\int_0^x F(x-y, y)dy \leq b(1+x^{m_1}), F(x-y', y) \leq b(1+x^m) \\ 0 \leq y \leq y' \leq x, \quad x \in [0, \infty).$$

Supposons que la valeur initiale vérifie soit :

1. $c_0 \in \Omega_{0,r}^+$, $r > \max\{m, 1\}$, et $r \geq m_1$ ou
2. $c_0 \in \Omega^+(0)$

Alors le problème 4.5 a respectivement, soit :

- au moins une solution dans $\Omega_{0,r}^+(T)$, ou
- une solution dans $\Omega^+(T)$

avec $\Omega_\lambda(T) = \{f \in C^0(0, \infty), \|f\|_\lambda \stackrel{\text{Def}}{=} \sup_{0 \leq t \leq T} \int_0^\infty \exp \lambda x |f(x, t)| dx\}$, $\Omega_{0,r}(T) = \{f \in C^0(0, \infty), \|f\|_{0,r} \stackrel{\text{def}}{=} \sup_{0 \leq t \leq T} \int_0^\infty x^r |f(x, t)| dx\}$, $\Omega(T) = \cup_{\lambda > 0} \Omega_\lambda(T)$. $\Omega_{0,r}^+(T)$ (resp. $\Omega_\lambda^+(T)$, $\Omega^+(T)$) est le cône des fonction non négatives de $\Omega_{0,r}(T)$ (resp. $\Omega_\lambda(T)$, $\Omega(T)$).

Théorème 4.2.2. [Dubovskii 1996] (conservation de la taille) On suppose que les conditions du théorème 4.2.1 sont remplies, que $r \geq 2$ et qu'en plus :

$$\int_0^x yF(x-y, y)dy \leq \text{const.}(1+x^r)$$

alors il y a conservation de la taille totale au cours du temps.

Théorème 4.2.3. [Dubovskii 1996] (unicité) On suppose que les hypothèses du cas 2 du théorème 4.2.1 sont vérifiées et qu'en outre $m_1 \leq 1$, alors le problème 4.5 a une solution unique dans $\Omega(T)$.

Pour certains noyaux, il est possible de caractériser davantage la solution du problème 4.5. Ainsi Fournier et al [Fournier 2005] ont montré l'existence de solutions autosimilaires,

$$n(x, t) = \frac{1}{s(t)^2} \psi\left[\frac{x}{s(t)}\right]$$

Les solutions construites dans [Fournier 2005] ont une décroissance rapide en x pour des temps longs pour trois classes de noyaux d'agrégation :

$$K_1(x, y) = (x^\alpha + y^\alpha)(x^{-\beta} + y^{-\beta}), \quad \alpha \in [0, 1], \quad \beta \in [0, \infty], \quad \lambda = \alpha - \beta \in (-\infty, 1) \\ K_2(x, y) = (x^\alpha + y^\alpha)^{-\beta}, \quad \alpha \in [0, \infty), \quad \beta \in [0, \infty], \quad \lambda = \alpha\beta \in [0, 1) \\ K_3(x, y) = x^\alpha y^\beta + x^\beta y^\alpha, \quad \alpha \in [0, 1], \quad \beta \in [0, 1], \quad \lambda = \alpha + \beta \in [0, 1)$$

4.2.2 Existence, unicité de solution pour l'équation discrète

On se réfère aux travaux de Ball et Carr [Ball 1990] qui concernent l'équation discrète d'agrégation-fragmentation :

$$\begin{aligned} \frac{\partial n(x_i, t)}{\partial t} = & \frac{1}{2} \sum_{j=1}^{i-1} K(x_i - x_j, x_j) n(x_i - x_j, t) n(x_j, t) - \sum_{j=1}^{\infty} K(x_i, x_j) n(x_i, t) n(x_j, t) \\ & - \frac{1}{2} \sum_{j=1}^{i-1} F(x_i - x_j, x_j) n(x_i, t) + \sum_{j=1}^{\infty} F(x_i, x_j) n(x_i + x_j, t) \quad \forall i = 1, 2, \dots \end{aligned} \quad (4.7)$$

où F est un noyau de fragmentation. Comme dans le cas de l'équation continue, on s'attend à ce qu'il y ait conservation de la taille, soit :

$$\sum_{i=1}^{\infty} x_i n(x_i, t) = \sum_{i=1}^{\infty} x_i n(x_i, 0), \quad \forall t \geq 0$$

ou encore

$$\lim_{n \rightarrow \infty} \int_0^t \sum_{i=1}^n \sum_{j=n-i+1}^{\infty} W_{i,j}(n(s)) ds = 0$$

avec $W_{i,j}(n) \stackrel{\text{def}}{=} K_{i,j} n_i n_j - F_{i,j} n_{i+j}$.

Cependant, pour certains noyaux, la densité de taille totale n'est pas conservée au cours du temps. Il en va ainsi, y compris pour des modèles sans fragmentation, avec des noyaux d'agrégation multiplicatifs de la forme $K(x_i, x_j) = (x_i x_j)^\alpha$ avec $\alpha > \frac{1}{2}$ [Leyvraz 1981, Hendriks 1983].

Théorème 4.2.4. [Ball 1990] (*Existence de solution*) Supposons que $K_{i,j} \leq \text{const.}(i+j)$, $\forall i, j \geq 1$. Soit $n_0 \in X^+$. Alors le problème 4.7 admet une solution telle que $n(0) = n_0$.

avec $X \stackrel{\text{def}}{=} \{y = (y_r) : \|y\| < \infty\}$, $\|y\| \stackrel{\text{def}}{=} \sum_{r=1}^{\infty} r|y_r|$ et $X^+ \stackrel{\text{def}}{=} \{y \in X : y \geq 0\}$.

Théorème 4.2.5. [Ball 1990] (*Conservation de la taille*) Sous les hypothèses H_1 et H_2 :

$H1 : K_{i,j} = r_i + r_j + \alpha_{i,j}$ ou (r_i) est une suite non négative et $\alpha_{i,j} \geq 0 \quad \forall i, j$

et $\alpha_{i,j} \leq \text{const.}(i+j) \quad \forall i, j \geq k_0$

$H2 : \sum_{i=k_0}^{h(r)} i F_{r-i, i} \leq \text{const.} r \quad \forall r \geq 2k_0 \quad :$

soit n une solution du problème 4.7 sur $[0, T)$ avec $\rho_0 = \sum_{i=1}^{\infty} i n_i(0)$ alors

$$\sum_{i=1}^{\infty} i n_i(t) = \rho_0, \quad \forall t \in [0, T)$$

Théorème 4.2.6. [*Ball 1990*] (*Unicité*) Soit $C > 0$ et $0 \leq \alpha \leq \frac{1}{2}$ et les hypothèses suivantes :

$$(i) \quad K_{i,j} \leq C(ij)^\alpha \quad \forall i, j$$

$$(ii) \quad \sum_{i=1}^{h(r)} i^{1-\alpha} F_{r-i,i} \leq Cr^{1-\alpha} \quad \forall r \geq 2$$

Soit $n_0 \in X^+$ et $T > 0$, alors il existe une unique solution du problème 4.7 sur $[0, T)$

Par ailleurs, des théorèmes relatifs à la régularité des solutions sont énoncés en annexe A.2.

4.2.3 Approximation des solutions de l'équation continue

On se place sous les hypothèses (cf. Théorème 4.2.1) qui permettent d'affirmer que le problème 4.5 admet au moins une solution conservant la taille. Pour tout entier $k \geq 1$, on définit alors [*Filbet 2004*] :

$$n_0^k(x) = n_0(x) \mathbb{I}_{[0,k]}(x), \quad K^k(x, y) = K(x, y) \mathbb{I}_{[0,k]}(x + y)$$

et on considère le problème approché :

$$\frac{\partial \bar{n}(x, t)}{\partial t} = \frac{1}{2} \int_0^x K(x - y, y) \bar{n}(x - y, t) \bar{n}(y, t) dy \quad (4.8)$$

$$- \int_0^{k-x} K(x, y) \bar{n}(x, t) \bar{n}(y, t) dy(x, t) \in (0, k) \times \mathbb{R}^+$$

$$\bar{n}_0(., 0) = n_0^k$$

On peut alors montrer [*Filbet 2004*] que la taille totale des particules de taille maximale égale à k est conservée, soit :

$$\int_0^k x \bar{n}(x, t) dx = \int_0^k x n_0(x, t) dx, \quad t \geq 0$$

et aussi une convergence faible de cette suite d'approximations vers la solution du problème.

4.3 Modèles d'agrégation existants

4.3.1 Modèles simples et solutions exactes

Les solutions exactes de l'équation de Smoluchowski sont connues pour des conditions initiales ainsi que des formes de noyaux bien particuliers [*Aldous 1999*]. Dans

le cas d'une distribution initiale homogène (*i.e* $n(1, t) = 1, n(x, t) = 0 \quad \forall x \geq 2$) et un noyau constant $K(x, y) = 1$, la solution est de la forme :

$$\forall x \in \mathbb{N}^*, \forall t \geq 0, n(x, t) = \left(\frac{2}{2+t} \right)^2 \left(\frac{t}{2+t} \right)^{x-1} \quad (4.9)$$

Dans le cas d'un noyau additif $K(x, y) = x + y$ respectivement multiplicatif $K(x, y) = xy$ les solutions sont, respectivement :

$$\forall x \in \mathbb{N}^*, \forall t \geq 0, n(x, t) = \frac{x(1 - e^{-t})^{x-1}}{x!} e^{-t} e^{-x(1-e^{-t})} \quad (4.10)$$

et

$$\forall x \in \mathbb{N}^*, n(x, t) = \begin{cases} \frac{1}{x} \frac{(nt)^{x-1}}{x!} e^{-nt} & \text{si } t \in [0, 1] \\ \frac{1}{t} \frac{x^{x-2}}{x!} e^{-x} & \text{si } t > 1. \end{cases} \quad (4.11)$$

pour la condition initiale mentionnée ci-dessus. Ces expressions analytiques sont utiles pour valider et comparer les schémas d'approximation numériques déterministes, ou stochastiques.

4.3.2 Applications de l'équation de Smoluchowski

Les principales applications de l'équation de Smoluchowski concernent notamment la production d'aérosols [Drake 1972], l'astrophysique, la formation et la croissance des planètes [Lee 2000], la biologie (polymérisation [Lushnikov 1976], la fusion des globules rouges [Samsel 1984]). L'équation de Smoluchowski a été également utilisée, avec des modifications de sa formulation générale, pour inclure d'autres phénomènes tels que la fragmentation, ainsi que l'apport ou le prélèvement continu de matière [Hendriks 1984, Davies 1999]. Certaines de ces applications [Vanni 2000] sont basées sur des modèles d'agrégation dans un contexte de mouvement brownien, laminaire, etc. (cf. Tableau 4.1).

Noyau	K	Auteur
Mouvement brownien	$K_{i,j} = \frac{2kT}{3\mu} \frac{(a_i + a_j)^2}{a_i a_j}$	[Von Smoluchowski 1917]
Mouvement laminaire	$K_{i,j} = \frac{4}{3} G(a_i + a_j)^3$	[Von Smoluchowski 1917]
Mouvement de turbulence isotropique	$K_{i,j} = \frac{4}{3} \left(\frac{3\pi}{10} \right)^2 \sqrt{\frac{\varepsilon}{\nu}} (a_i + a_j)^3$	[Saffman 1956], [Adachi]
Sédimentation différentielle	$K_{i,j} = \left(\frac{2\pi g}{9\mu} \right) (\rho_s - \rho) (a_i + a_j)^4$	[Schumann 1940]

TABLE 4.1 : Expressions des noyaux de coalescence pour différents mécanismes.

4.4 Modèle de coalescence des corps lipidiques

Nous faisons donc l'hypothèse que la fusion est l'un des principaux processus en œuvre lors de l'évolution des tailles des corps lipidiques (cf. Chapitre Introduction

générale). A notre connaissance, la biogenèse des corps lipidiques chez la plante n'a jamais été étudiée sous l'angle d'une dynamique de populations dans un contexte d'agrégation. Nous proposons donc un modèle d'agrégation selon la formulation de Smoluchowski. Ce modèle sera défini au travers d'un noyau d'agrégation, K , et d'un taux de production⁴, P , de corps lipidiques au cours du temps, soit :

$$\frac{\partial n(x, t)}{\partial t} = \frac{1}{2} \int_0^x K(x - y, y) n(x - y, t) n(y, t) dy \quad (4.12)$$

$$- \int_0^\infty K(x, y) n(x, t) n(y, t) dy + P(x, t) \quad (4.13)$$

$$n(x, 0) = n_0(x) \quad (4.14)$$

Le taux de production ne sera pas modélisé, mais estimé à partir des données. Le taux d'agrégation sera basé sur la densité de charge de la demi-membrane. Nous partons pour cela d'observations et de connaissances bibliographiques que nous reconstituons sous forme d'hypothèses de travail argumentées. Nous avons alors essayé de conserver des formulations simples et parcimonieuses en paramètres. Nous noterons respectivement $Vol(CL)$, $A(CL)$, $Q(CL)$ et $\sigma(CL) = \frac{Q(CL)}{A(CL)}$, le volume, l'aire de la membrane, la charge totale de la membrane et la densité de charge du corps lipidique CL .

Nous listons donc tout d'abord un certain nombre d'hypothèses relatives au corps lipidique et à son environnement :

H1 : *Les corps lipidiques sont sphériques.* Par conséquent, nous supposons que deux corps qui s'agrègent reforment un corps sphérique une fois l'agrégation terminée. On rappelle que la sphère est la surface d'aire minimale englobant un volume donné quelconque. En réalité, on observe bien des corps lipidiques sphériques comme indiqué sur la figure 4.1 représentant les distributions des coefficients de sphéricité, $C_S = 36\pi \frac{Vol(CL)^2}{A(CL)^3}$.

H2 : *La distribution de charge est uniforme sur la membrane.* On sait que la charge globale à la surface du corps lipidique est négative [Tzen 1992a]. Un champ électrique de symétrie radiale et donc d'amplitude constante est alors engendré par le corps lipidique. Du point de vue de l'agrégation, dans l'environnement spatial immédiat du corps lipidique, il n'y a pas de région privilégiée par le corps lipidique.

H3 : *Le contenu du corps lipidique est électriquement neutre.* On suppose que le corps lipidique est rempli de lipides neutres.

H4 : *L'agrégation est binaire.* On suppose qu'à un même instant et pour un corps lipidique en cours d'agrégation, l'agrégation se fait avec un seul autre corps lipidique.

H5 : *Il y a conservation du volume des CL lors de l'agrégation, soit $Vol(CL_1 \cup CL_2) = Vol(CL_1) + Vol(CL_2)$.*

H6 : *Il y a conservation de la charge des CL lors de l'agrégation, soit $Q(CL_1 \cup CL_2) = Q(CL_1) + Q(CL_2)$.*

4. Dans le volume élémentaire V considéré.

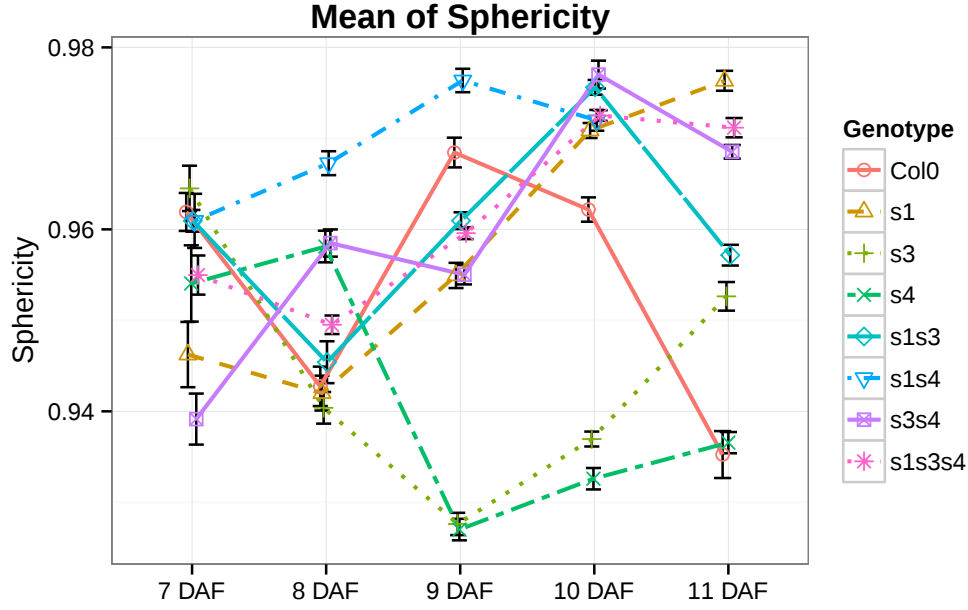


FIGURE 4.1 : Coefficients de sphéricités moyens des corps lipidiques pour différents génotypes à différents stades du développement.

H7 : *Le pH du cytoplasme est constant durant le développement.* Le pH du cytoplasme pourrait cependant jouer un rôle car certaines protéines de la membrane peuvent être ionisées et sensibles au pH.

Nous faisons alors les hypothèses de modélisation suivantes :

HMOD1 : *Le taux d'agrégation est inversement proportionnel à la densité de charge.* Plus la densité de charge est faible et plus la fréquence d'agrégation est forte.

HMOD2 : *Le taux d'agrégation est multiplicatif.* Ce sont donc les agrégations "faible densité de charge"- "faible densité de charge" qui dominent contrairement à un taux d'agrégation additif où les interactions "faible densité de charge"- "faible densité de charge" sont du même ordre de fréquence que les interactions "faible densité de charge"- "forte densité de charge" (figure 4.2).

Le taux d'agrégation prend alors la forme suivante :

$$K(CL_1, CL_2) \propto \frac{1}{\sigma(CL_1)} \times \frac{1}{\sigma(CL_2)} \quad (4.15)$$

En raison de l'hypothèse **H1**, on a aussi :

$$K(CL_1, CL_2) \propto \frac{Vol(CL_1)^{2/3}}{Q(CL_1)} \times \frac{Vol(CL_2)^{2/3}}{Q(CL_2)} \quad (4.16)$$

Afin de pouvoir s'affranchir de la variable Q , nous émettons une autre hypothèse (HMOD3) sur le processus de production des corps lipidiques. Cette hypothèse va nous permettre d'exprimer Q en fonction du volume du corps lipidique.

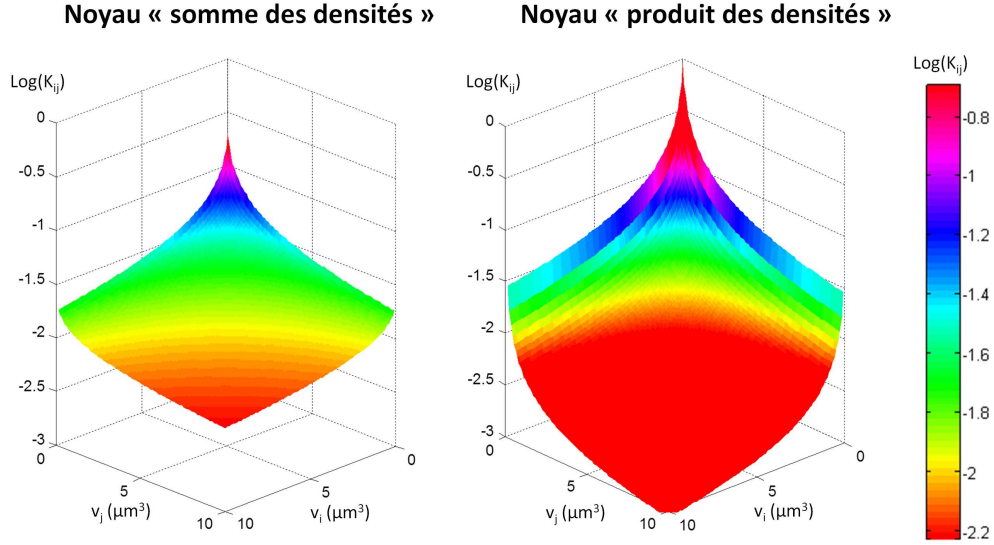


FIGURE 4.2 : Valeurs du taux d'agrégation pour des noyaux multiplicatifs et additifs.

HMOD3 : Le processus de production au niveau du RE produit une population de CL homogène dans le rapport $\frac{Q(CL)}{\text{Vol}(CL)}$. En d'autres termes, soit sont produits des corps lipidiques tous identiques en volume et charge, soit sont produites des populations hétérogènes en volume et charge mais, dans ce cas, un corps lipidique de volume p fois plus important qu'un autre a aussi une charge p fois plus importante.

En s'appuyant sur **H5** et **H6**, on en déduit alors que :

$$Q(CL) = \text{Const.} \times \text{Vol}(CL) \quad (4.17)$$

et

$$K(CL_1, CL_2) = K(\text{Vol}(CL_1), \text{Vol}(CL_2)) \propto \frac{\text{Vol}(CL_1)^{2/3}}{\text{Vol}(CL_1)} \times \frac{\text{Vol}(CL_2)^{2/3}}{\text{Vol}(CL_2)} \quad (4.18)$$

$$\propto (\text{Vol}(CL_1)\text{Vol}(CL_2))^{-1/3} \quad (4.19)$$

Ainsi lorsque les corps lipidiques deviennent gros, leur densité de charge est forte et leur agrégation avec des gros corps lipidiques est peu fréquente. Pour éviter de faire l'hypothèse **HMOD3**, nous aurions pu considérer la charge comme une variable, comme dans le contexte d'aérosols avec des compositions chimiques variable [Dergaoui 2013]. Cependant, cela complique cependant le calcul des solutions numériques car il faut discrétiser l'espace produit des volumes et des charges.

Le noyau 4.18 ne prend en compte ni les aspects spatiaux ni les aspects génétiques. On voit cependant clairement que les oléosines (cf. chapitres Introduction générale et 3) influent les distributions de volumes au cours du développement ; nous pouvons alors faire l'hypothèse qu'elles influent sur le processus d'agrégation. Nous

ferons aussi l'hypothèse qu'elles n'influent pas sur le processus de production⁵.

Pour prendre en compte le facteur génétique, nous supposons qu'au delà de la charge intervient aussi un facteur abstrait d'efficacité de la charge et que cette efficacité peut dépendre de la présence des différentes oléosines qui auront une influence positive ou négative sur le taux d'agrégation. Ce facteur d'efficacité pondère alors la charge et nous définissons le taux d'agrégation par :

$$K(CL_1, CL_2) \propto \frac{1}{e(G)\sigma(CL_1)} \times \frac{1}{e(G)\sigma(CL_2)} \quad (4.20)$$

où l'efficacité $0 < e(G) \leq 1$ est reliée au génotype G par un modèle linéaire qualitatif et ne dépend pas du volume, soit :

$$\begin{aligned} e(G) = & \gamma + \alpha_1 \mathbb{I}_{\{S_1=1\}}(G) + \alpha_2 \mathbb{I}_{\{S_3=1\}}(G) + \alpha_3 \mathbb{I}_{\{S_4=1\}}(G) \\ & + \alpha_{1,2} \mathbb{I}_{\{S_1=1, S_3=1\}}(G) + \alpha_{1,3} \mathbb{I}_{\{S_1=1, S_4=1\}}(G) + \alpha_{2,3} \mathbb{I}_{\{S_3=1, S_4=1\}}(G) \\ & + \alpha_{1,2,3} \mathbb{I}_{\{S_1=1, S_3=1, S_4=1\}}(G) \end{aligned}$$

Ici, $S_1 = 1$ (resp. $S_3 = 1$, $S_4 = 1$), signifie que l'oléosine S1 (resp. S3, S4) est présente dans le génotype G .

Nous avons choisi un modèle linéaire pour sa simplicité fonctionnelle et sa facilité d'interprétation. Les paramètres α_i , $i = 1, \dots, 3$ sont les effets principaux associés à la présence des différentes oléosines, les paramètres $\alpha_{i,j}$, $1 \leq i < j \leq 3$ sont les effets d'interaction de paire et $\alpha_{1,2,3}$ est un effet d'interaction de triplet. Le paramètre γ est un effet moyen, ici il sera associé au triple mutant qui sera, en quelque sorte, notre référence pour interpréter les paramètres du modèle. En définitive, le noyau d'agrégation aura donc la forme suivante :

$$K_G(CL_1, CL_2) \propto \frac{(Vol(CL_1)Vol(CL_2))^{-1/3}}{(\gamma + \sum_{i=1}^3 \alpha_i \mathbb{I}_{\{S_i=1\}}(G) + \sum_{1 \leq i < j \leq 3} \alpha_{i,j} \mathbb{I}_{\{S_i=1, S_j=1\}}(G) + \alpha_{1,2,3} \mathbb{I}_{\{S_1=1, S_2=1, S_3=1\}}(G))^2} \quad (4.21)$$

Remarque : Le modèle est basé ici seulement sur la présence/absence des oléosines et pas sur leur disponibilité en plus ou moins grande quantité. On pourra cependant examiner l'influence de ces quantités au travers des paramètres du modèle grâce à l'utilisation de plantes transgéniques surexprimant certaines oléosines, puis à la mesure des volumes des corps lipidiques au travers du pipeline d'analyse. Si le processus de fusion est sensible à la quantité d'une oléosine particulière, on s'attend à ce que les coefficients relatifs à cette oléosine varient dans notre modèle.

Au chapitre 2 nous avons mis en évidence une relation entre la proximité spatiale des corps lipidiques d'une part et le génotype, stade de développement et volume d'autre part. Cette relation est alors approximée par :

$$\ln \phi(V, t, G) = \theta_2(t, G) \ln V + \log \theta_1(t, G) \quad (4.22)$$

5. A notre connaissance, cette information n'est pas disponible dans la littérature, et la microscopie électronique semble révéler une population de petits corps lipidiques de volumes (de l'ordre de 50 nm) assez homogènes au voisinage du RE [Schmidt 2008].

où ϕ désigne la fraction locale.

Il est alors tentant de proposer un noyau qui combine les aspects de charge, les facteurs génétiques et une information de proximité en fonction du volume. Nous proposons d'opter pour un produit du noyau précédent (cf. 4.21) et d'un noyau multiplicatif portant sur la fraction locale, soit :

$$K_{GS}(CL_1, CL_2) \propto K_G(CL_1, CL_2) \times (\phi(Vol(CL_1)) \times \phi(Vol(CL_2))) \quad (4.23)$$

$$\propto K_G(CL_1, CL_2) \times ((Vol(CL_1))^{\theta_2} \times (Vol(CL_2))^{\theta_2}) \quad (4.24)$$

où θ_2 est un paramètre supplémentaire du modèle.

Si on développe K_{GS} en fonction du volume, on s'aperçoit que le noyau est de la forme $(Vol(CL_1))^{\theta_2 - \frac{1}{3}} Vol(CL_2)^{\theta_2 - \frac{1}{3}}$, soit :

$$K_{GS} = \frac{\theta_1(t, G)^2 (Vol(CL_1) Vol(CL_2))^{(-\frac{1}{3} + \theta_2(t, G))}}{e(G)^2} \quad (4.25)$$

4.5 Méthode d'estimation des paramètres

Nous serons amenés à considérer plusieurs modèles avec un nombre variable de paramètres (1 à 9). Nous avons opté pour une méthode d'estimation bayésienne de ces paramètres : la méthode DRAM [Haario 2006], que nous décrivons en 4.5.1 et expliciterons dans le cadre de notre application en 4.5.2. La difficulté dans notre cas vient de l'équation d'état de notre modèle, qui une fois discrétisée, consiste en un système différentiel de taille $N > 100$.

4.5.1 Principe de l'estimation bayésienne

On se place dans un cadre général pour lequel un ensemble de paramètres $\theta \in \mathbb{R}^p$ est associé à un modèle de données $\mathcal{M}(x, \theta)$ où x désigne un ensemble de variables explicatives, soit alors :

$$y = \mathcal{M}(x, \theta) + \varepsilon$$

où y est un vecteur de variables aléatoires dont on peut observer des réalisations et ε un vecteur aléatoire de même dimension représentant les erreurs au modèle. D'après la formule de Bayes, on a alors :

$$\mathbb{P}(\theta/y) \propto \mathbb{P}(y/\theta) \times \mathbb{P}(\theta)$$

où $\mathbb{P}(\theta/y)$ désigne la distribution *a posteriori* des paramètres et $\mathbb{P}(\theta)$ la distribution *a priori*.

L'estimation des paramètres θ est alors basée sur celle de la loi *a posteriori*. L'algorithme standard classique de Metropolis-Hastings [Tierney 1994] pour simuler une loi, consiste à simuler une chaîne de Markov qui va converger vers cette loi en évitant le recours à la constante de normalisation de la loi.

Un estimateur, dit MCMC, de la moyenne selon cette loi d'une fonction f sera alors aisément calculé en s'appuyant sur une réalisation, $X_1, \dots, X_N$, de cette chaîne de Markov, soit : $\frac{1}{N}f(X_i)$.

L'algorithme DRAM est une amélioration de l'algorithme de Metropolis-Hastings, il combine deux stratégies : le rejet différé et le Metropolis adaptatif, pour augmenter la vitesse de convergence d'estimateurs MCMC obtenus par l'algorithme classique de Metropolis-Hastings.

La stratégie de rejet différé [Mira 2002] consiste à différer un rejet dans la boucle MCMC. Ainsi au lieu de conserver en cas de rejet l'état proposé, on propose un nouvel état que l'on accepte avec une probabilité qui permet de garantir la convergence de l'algorithme vers la distribution visée. Le nouvel état peut dépendre des états proposés précédemment, ce qui permet d'adapter la distribution instrumentale à l'intérieur d'une itération de l'algorithme MCMC.

La seconde stratégie du Metropolis adaptatif [Haario 2001] consiste à adapter la distribution instrumentale⁶ au cours des itérations en s'appuyant sur la trajectoire passée de la chaîne. On choisit alors une distribution instrumentale gaussienne, centrée sur l'état courant et de matrice de covariance :

$$C_t = \begin{cases} C_0, & t \leq t_0 \\ s_d \text{Cov}(X_0, \dots, X_{t-1}) + s_d \varepsilon I_d, & t > t_0 \end{cases}$$

avec l'expression suivante pour la covariance empirique :

$$\text{Cov}(X_0, \dots, X_k) = \frac{1}{k} \left(\sum_{i=0}^k X_i X_i^T - (k+1) \bar{X}_k \bar{X}_k^T \right)$$

Cet algorithme est accessible à partir du site <http://helios.fmi.fi/~lainema/mcmc/>

4.5.2 Application aux modèles de dynamique de populations des corps lipidiques

Dans notre contexte de modélisation, le vecteur des données élaborées est constitué des quantités $y_r^g(x_i, t_j) \stackrel{\text{def}}{=} n_r^g(x_i, t_j)$, $i = 1, \dots, N(g, t_j)$, $j = 1, \dots, 5$, $r = 1, \dots, r(g, t_j)$, $g = 1, \dots, 8$ représentant les effectifs observés, rapportés au volume élémentaire V lorsque les volumes x_i sont répartis dans $N(g, t_j)$ classes pour la répétition r du génotype g au temps t_j :

$$n_r^g(x_i, t_j) = \widehat{n_r^g(x_i, t_j)} + (\varepsilon_r^g)_{i,j}$$

où $\widehat{n_r^g(x_i, t_j)} = \mathcal{M}(x_i, t_j, g, r, \theta)$ désigne l'approximation de la solution de l'équation d'état (cf. 4.12) pour la classe de volume x_i , le temps t_j , le génotype g et la répétition

6. La distribution instrumentale Q dans la table est la distribution selon laquelle on effectue un nouveau tirage des paramètres lors de la construction de la chaîne de Markov. Idéalement, cette distribution doit être choisie de façon à limiter le nombre de rejets.

r . Sous l'hypothèse d'indépendance des termes d'erreur au modèle, on peut alors écrire⁷ :

$$\mathbb{P}(y|\theta) = \mathbb{P}(n|\theta) = \prod_{j=1}^5 \prod_{g=1}^8 \prod_{r=1}^{r(g)} \prod_{i=1}^N \mathbb{P}((\varepsilon_r^g)_{i,j} = n_r^g(x_i, t_j) - \widehat{n_r^g(x_i, t_j)})$$

A ce stade, il faut choisir la loi des erreurs et donc évaluer s'il y a lieu ou pas, étant donné le modèle considéré, d'adopter une loi avec une variance qui dépend du volume et, éventuellement, du génotype ainsi que la date. En particulier, on peut se demander si le modèle est moins précis sur certaines classes : les petits volumes de corps lipidiques, par exemple, du fait de la difficulté à les mesurer, ou les gros volumes, du fait de leur interception avec le bord de la région échantillonnée.

Ceci conduit cependant à une complexité numérique importante car il faut effectuer $\sum_{g=1}^8 \prod_{j=1}^5 r(g, t_j)$ résolutions de l'équation d'état à chaque itération de l'algorithme DRAM. Nous avons donc opté pour un allègement de l'estimation en ne retenant qu'une observation moyennée par génotype et par date, soit :

$$n^g(x_i, t_j) = \frac{\sum_{r=1}^{r(g, t_j)} n_r^g(x_i, t_j)}{r(g, t_j)}$$

Cela permet d'économiser $\sum_{g=1}^8 \prod_{j=1}^5 (r(g, t_j) - 1)$ résolutions de l'équation d'état à chaque itération de l'algorithme DRAM.

Pour ce qui concerne le terme d'erreur, on peut s'appuyer sur l'image de l'erreur sur la moyenne pour décider du modèle d'erreur et choisir un modèle gaussien de la forme $(\varepsilon^g)_{i,j} \sim \mathcal{N}(0, \sigma^M(x_i, g, t_j)^2)$, avec $\sigma^M(x_i, g, t_j) = \frac{\sigma(x_i, g, t_j)}{\sqrt{r(g, t_j)}}$.

Le terme de vraisemblance aux données devient alors :

$$\begin{aligned} \mathbb{P}(d|\theta) &= \prod_{j=1}^5 \prod_{i=1}^N \prod_{g=1}^8 \mathbb{P}((\varepsilon^g)_{i,j} = n^g(x_i, t_j) - \widehat{n^g(x_i, t_j)}) \\ &\propto \exp - \sum_{j=1}^5 \sum_{i=1}^N \sum_{g=1}^8 \frac{(n^g(x_i, t_j) - \widehat{n^g(x_i, t_j)})^2}{\sigma^M(i, g, j)^2} \end{aligned}$$

Il reste à choisir les distributions *a priori* et les contraintes de bornes pour procéder à l'estimation.

Contraintes de positivité. Pour le modèle 4.23, nous restreignons l'espace des paramètres pour satisfaire à certaines contraintes de positivité. On rappelle que le terme $(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_{12} + \alpha_{13} + \alpha_{23} + \alpha_{123} + \gamma)$ présent au dénominateur du noyau d'agrégation doit être interprété comme une pondération de la "fitness" d'une membrane pour fusionner. C'est donc un terme positif. Par ailleurs, on veut

7. Les taux de production correspondant aux différentes répétitions peuvent changer. Il est alors *a priori* légitime de vouloir ajouter toutes les répétitions plutôt que des mesures moyennées.

pouvoir interpréter les différents coefficients présents dans ce dénominateur comme des adaptations de cette pondération, suite à la présence de l'une ou plusieurs des trois oléosines S1, S3 et S4. Ceci conduit aux contraintes suivantes :

$$\gamma > 0 \text{ [triple mutant S1, S3, S4. (g=1)]} \quad (4.26)$$

$$\gamma + \alpha_1 > 0 \text{ [double mutant S3, S4. (g=2)]} \quad (4.27)$$

$$\gamma + \alpha_2 > 0 \text{ [double mutant S1, S4. (g=3)]} \quad (4.28)$$

$$\gamma + \alpha_3 > 0 \text{ [double mutant S1, S3. (g=4)]} \quad (4.29)$$

$$\gamma + \alpha_1 + \alpha_2 + \alpha_{12} > 0 \text{ [simple mutant S4. (g=5)]} \quad (4.30)$$

$$\gamma + \alpha_1 + \alpha_3 + \alpha_{13} > 0 \text{ [simple mutant S3. (g=6)]} \quad (4.31)$$

$$\gamma + \alpha_2 + \alpha_3 + \alpha_{23} > 0 \text{ [simple mutant S1. (g=7)]} \quad (4.32)$$

$$\gamma + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_{12} + \alpha_{13} + \alpha_{23} + \alpha_{123} \text{ [sauvage (wt) (g=8)]} \quad (4.33)$$

Distributions *a priori*. Nous allons aussi respecter les contraintes 4.26 et profiter des découplages entre certaines données et certains paramètres pour effectuer une estimation progressive des paramètres. Avec les données relatives au triple mutant (g=1), nous estimerons le paramètre γ . Nous choisirons alors une distribution *a priori* de type $\Gamma(u_1, u_2)$, les hyperparamètres u_1 et u_2 seront fixés : $u_1 = 1.01$, $u_2 = 2000$. On procède exactement de la même façon pour l'estimation des paramètres $\gamma + \alpha_i$, $i = 1, \dots, 3$.

Estimation des paramètres. Pour le modèle 4.23, on effectue une estimation progressive des paramètres. On commence par faire une estimation de γ en s'appuyant uniquement sur la donnée "triple mutant" (aucune des trois oléosines présente). Il s'agit alors d'optimiser par rapport à γ et, sous la contrainte de positivité, le critère d'adéquation aux données.

On poursuit l'estimation des paramètres en estimant tout d'abord successivement α_1 , α_2 , α_3 puis les coefficients d'interaction simple $\alpha_{1,2}$, $\alpha_{1,3}$, $\alpha_{2,3}$ et, enfin, le paramètre d'interaction triple $\alpha_{1,2,3}$. Ainsi, pour l'estimation du paramètre α_1 , on utilisera les jeux de données pour lesquels seule S1 est présente, soient les doubles mutants de S3 et S4. On obtient donc des estimateurs $\hat{\delta}_1 \stackrel{Def}{=} \widehat{\gamma + \alpha_1}$, (resp. $\hat{\delta}_2 \stackrel{Def}{=} \widehat{\gamma + \alpha_2}$) et finalement un estimateur de α_1 , (resp. α_2) par $\hat{\alpha}_1 = \hat{\delta}_1 - \hat{\gamma}$, (resp. $\hat{\alpha}_2 = \hat{\delta}_2 - \hat{\gamma}$)

Dans le cadre d'une généralisation du modèle 4.23 (cf. 4.7), il est nécessaire d'utiliser l'ensemble des données pour l'estimation des différents paramètres.

4.6 Discrétisation et implémentation numérique

Nous avons utilisé l'approche déterministe afin de résoudre numériquement 4.12 sur un domaine fini $\Omega = [1, 2, \dots, N]$. Il est donc nécessaire de tronquer les classes de volumes sur Ω . Pour cela, on définit v_m , v_M , et v_s respectivement la classe volume minimale, la classe de volume maximale, et la largeur de classes de volume. Dans

Constant			
N	10	100	1000
MSE	$2.00 \cdot 10^{-11}$	$1.81 \cdot 10^{-17}$	$1.51 \cdot 10^{-18}$
CPU(sec)	0.017	0.025	5.870
Additif			
N	10	100	1000
MSE	$2.79 \cdot 10^{-5}$	$5.77 \cdot 10^{-9}$	$1.22 \cdot 10^{-10}$
CPU(sec)	0.018	0.028	2.18
Multiplicatif			
N	10	100	1000
MSE	$2.29 \cdot 10^{-5}$	$3.06 \cdot 10^{-8}$	$6.24 \cdot 10^{-9}$
CPU(sec)	0.015	0.033	13.298

TABLE 4.2 : Erreur quadratique moyenne entre la solution exacte et l'approximation numérique pour différents types de noyaux dont la solution exacte est connue (constant, additif et multiplicatif), pour différentes valeurs de N. La complexité numérique est évaluée en temps de calcul CPU (en seconde) pour chaque situation.

les simulations effectuées nous avons fixé v_m et v_s à $0.1 \mu m^3$. Nous avons ensuite adopté une méthode classique de résolution des équations différentielles (Runge Kutta d'ordre 4) pour résoudre l'équation discrète 4.4. Pour cela, nous avons utilisé une procédure C++ appelée depuis Matlab (mex-file) afin d'accélérer le temps de calcul (cf. Annexes A.3). Cette méthode est rapide et possède de bonnes propriétés de stabilité. Les solutions analytiques de l'équation de Smoluchowski énoncées dans 4.3.1 ont été utilisées afin de valider notre implémentation numérique. Pour cela, nous avons comparé, ces solutions à notre approximation numérique (Tableau 4.2) pour les différents noyaux, respectivement, constant (Figure 4.3), additif (Figure 4.4) et multiplicatif (Figure 4.5).

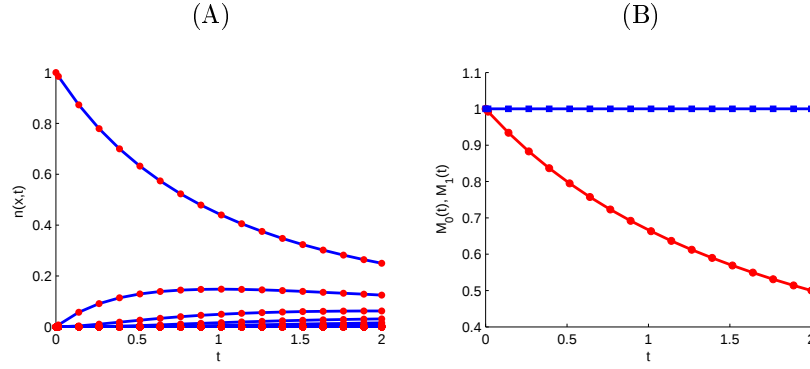


FIGURE 4.3 : Comparaison entre solution exacte et approximation numérique pour un noyau constant (A) Solution exacte (ligne continue en bleu) obtenue pour t de 0 à 2 comparée à l'approximation numérique (pointillé en rouge) obtenue dans les mêmes conditions. (B) Moments d'ordre 0 (rouge) et d'ordre 1 (bleu) obtenus pour un noyau constant, par solution exacte (ligne continue) et approximation numérique (ligne pointillée).

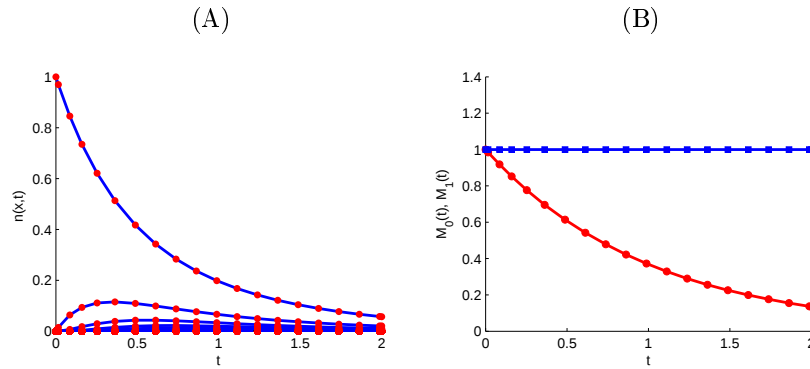


FIGURE 4.4 : Comparaison entre solution exacte et approximation numérique pour un noyau additif (A) Solution exacte (ligne continue en bleu) obtenue pour t de 0 à 2 comparée à l'approximation numérique (pointillé en rouge) obtenue dans les mêmes conditions. (B) Moments d'ordre 0 (rouge) et d'ordre 1 (bleu) obtenus pour un noyau additif, par solution exacte (ligne continue) et approximation numérique (ligne pointillée).

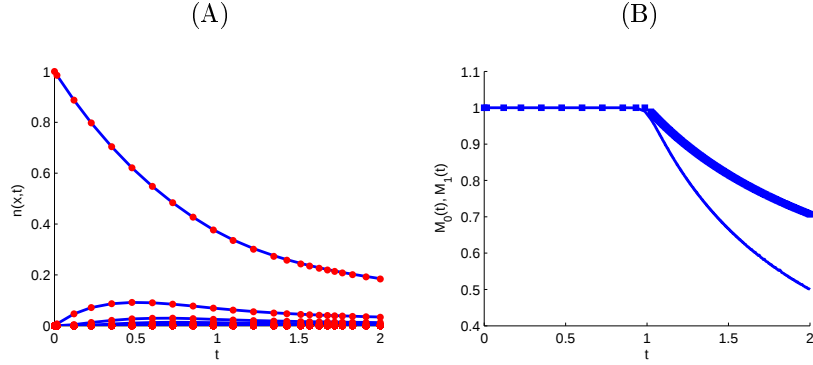


FIGURE 4.5 : Comparaison entre solution exacte et approximation numérique pour un noyau multiplicatif (A) Solution exacte (ligne continue en bleu) obtenue pour t de 0 à 2 comparée à l'approximation numérique (pointillé en rouge) obtenue dans les mêmes conditions. (B) Moments d'ordre 0 (rouge) et d'ordre 1 (bleu) obtenus pour un noyau multiplicatif, par solution exacte (ligne continue) et approximation numérique (ligne pointillée).

4.7 Résultats

Nous avons considéré 5 modèles différents (cf. Tableau 4.3) pour le noyau d'agrégation, le modèle 4.23, une généralisation de ce dernier, mais aussi des modèles plus simples, et, pour finir, un modèle englobant le modèle proposé qui s'est avéré satisfaisant sur plusieurs génotypes. Dans le tableau 4.3 nous distinguons deux types d'estimations : une estimation génotype par génotype d'une manière indépendante. Cela concerne les modèles 1 à 4 et une estimation basée sur l'utilisation simultanée de tous les génotypes pour le modèle 5. Pour les modèles 2 et 4, en principe, les estimations devraient être très proches (cf. Section 4.5.2). Les résultats présentés dans cette section comprennent une évaluation des différents paramètres des modèles développés (cf. Tableaux 4.4, 4.5, 4.6, 4.7 et 4.8) ainsi qu'une évaluation de la qualité d'ajustements de chaque modèle par rapport aux données expérimentales. La qualité d'ajustements est mesurée par l'erreur quadratique moyenne :

$$\frac{1}{5 \times 8 \times N} \sum_{j=1}^5 \sum_{g=1}^8 \sum_{i=1}^N \frac{(F^g(x_i, t_j) - \widehat{F^g(x_i, t_j)})^2}{\sigma^M(i, g, j)^2}$$

avec $F^g(x_i, t_j) = \sum_{k < i} x_k n^g(x_k, t_j)$ désignant le volume total des corps lipidiques de volumes $< x_i$ au jour t_j .

4.7.1 Modèle constant

Le taux de coalescence est maximum en absence d'oléosines (cf. Tableau 4.4), ce qui se traduit chez le triple mutant avec un taux de 297.05 ± 13.43 : L'ajout de S1, S3 ou S4 réduit significativement le taux de coalescence respectivement chez les mutants $s3s4$, $s1s4$, et $s1s3$. Le taux de coalescence le plus réduit est celui avec la présence de S4 chez le mutant $s1s3$ (136.65 ± 5.07). Par ailleurs, les simples génotypes ($s1$, $s3$ et

No	Modèle	Expression	Mode d'estimation
1	Constant	$K(v_1, v_2) = \eta$	Indépendant
2	génotypique	$K(v_1, v_2) = \frac{(v_1 v_2)^{-\frac{1}{3}}}{e(G)^2}$	Indépendant
3	Multiplicatif	$K(v_1, v_2) = \eta(v_1 v_2)^\beta$	Indépendant
4	génotypique/spatial 4.23	$K(v_1, v_2) = K_{GS}$	Indépendant
5	génotypique/spatial modulé	$K(v_1, v_2) = (\theta_1(G, t))^2 \frac{(v_1 v_2)^{\beta - \frac{1}{3} + \theta_2(G, t)}}{(e(G))^2}$	Simultané

TABLE 4.3 : Différents modèles de coalescence développés pour la dynamique des corps lipidiques, ainsi que le mode d'estimation des paramètres utilisé.

$s4$) présentent un taux de coalescence globalement réduit par rapport aux doubles génotypes. La réduction est remarquable chez le simple mutant $s1$ (présence de S3 et S4) qui est de 93.78 ± 2.03 . Cependant, on peut observer sur les figures 4.6 et 4.7 que ce modèle ne s'ajuste pas toujours très bien, en particulier pour les génotypes $s3$, $s4$, $s1s4$, et $s3s4$. L'erreur moyenne de ce modèle est de 0.0165 ± 0.0124 . Par ailleurs, on observe aussi, la plupart de temps, qu'il y a sur-estimation du taux de fusion puisque la courbe de volume cumulé estimée est souvent au-dessous de la courbe observée dans la première partie de la distribution.

Génotype	Paramètre	Erreur
	η	
wt	166.43 ± 3.50	0.0253
$s1$	93.78 ± 2.03	0.0193
$s3$	187.16 ± 1.50	0.0077
$s4$	145.87 ± 3.74	0.0032
$s1s3$	136.65 ± 5.07	0.0094
$s1s4$	214.12 ± 4.10	0.0245
$s3s4$	210.04 ± 10.23	0.0384
$s1s3s4$	297.05 ± 13.43	0.0046

TABLE 4.4 : Paramètres d'ajustement du modèle constant (Modèle 1).

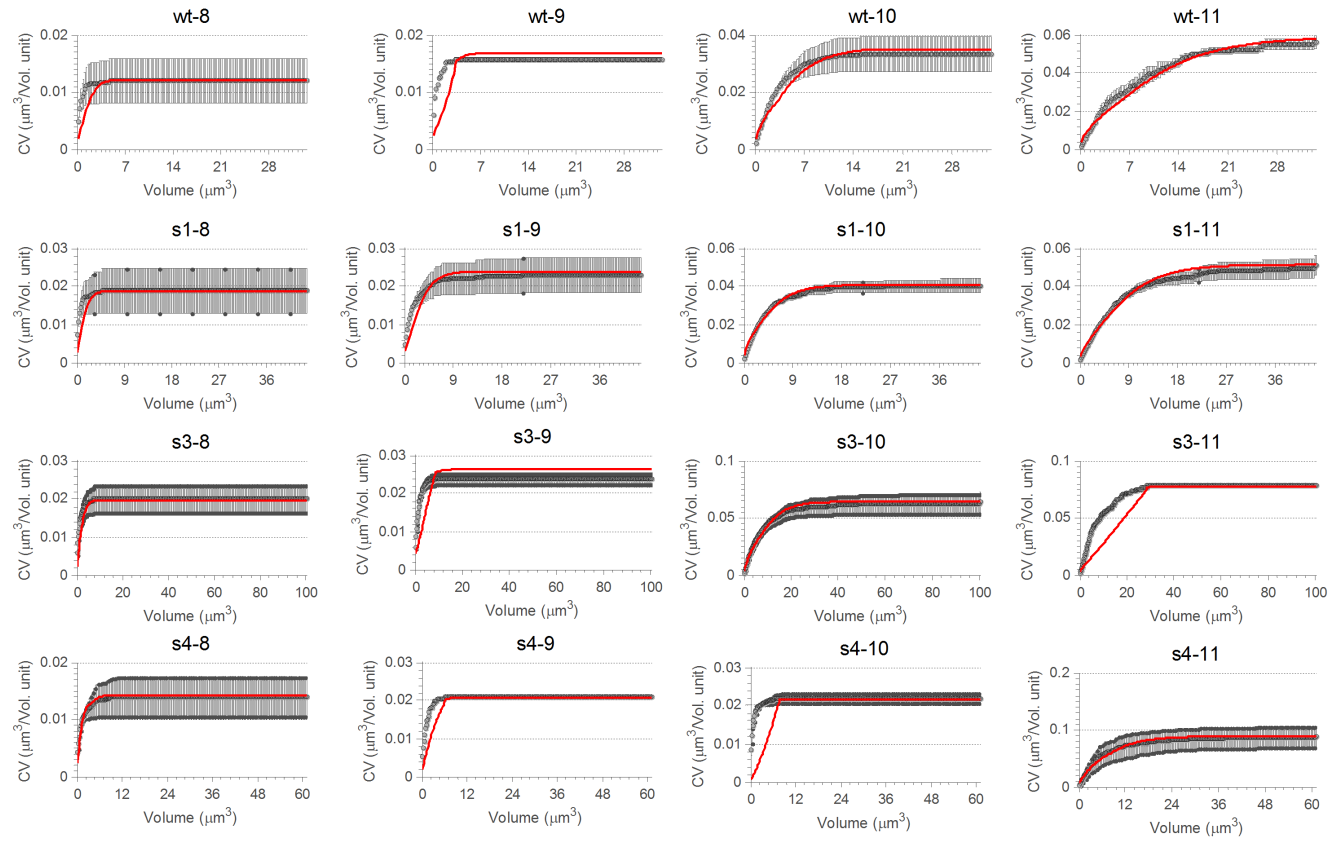


FIGURE 4.6 : Courbes d'ajustement du modèle constant (1) : volume de corps lipidiques cumulé par unité de volume de référence (CV). En noir la densité CV observée avec erreur, en rouge la densité estimée par ajustement des paramètres.

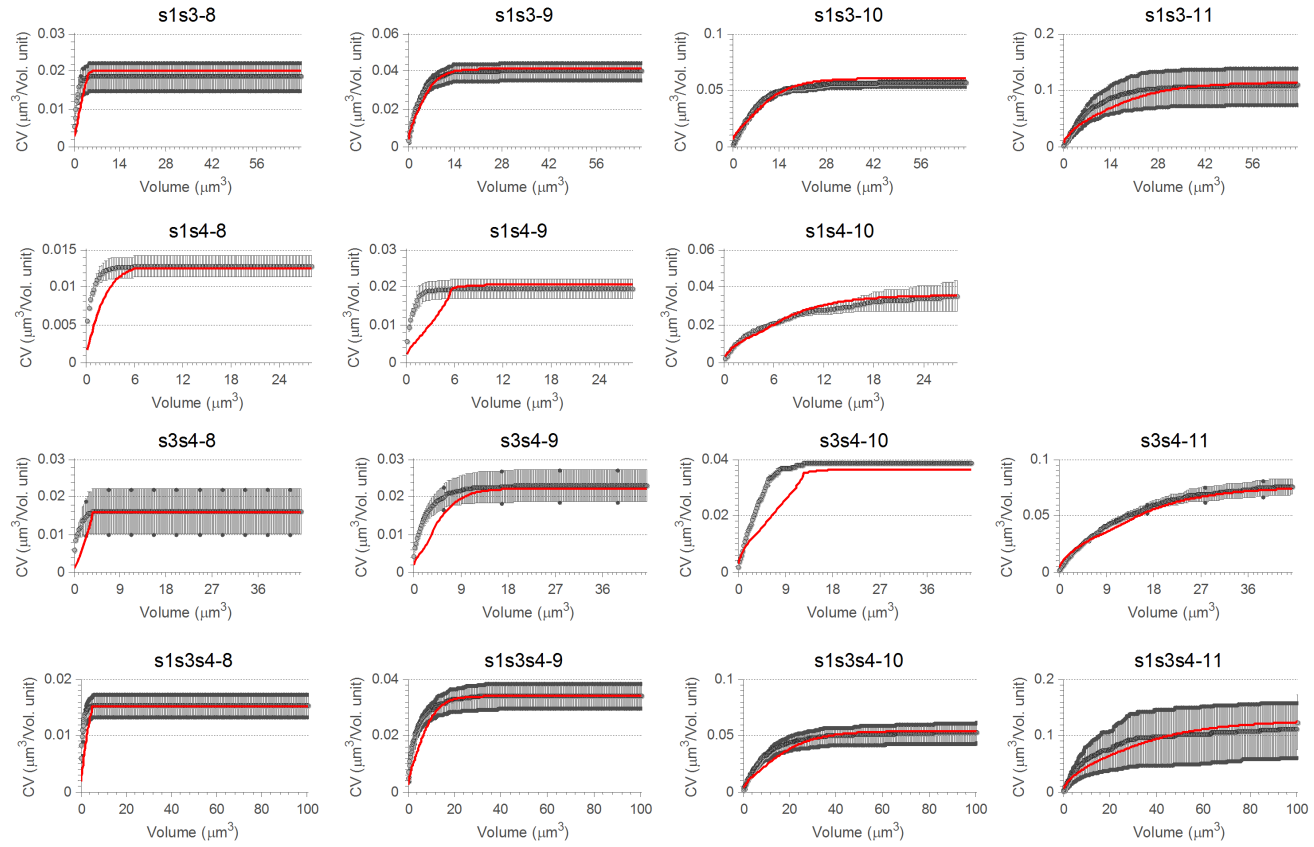


FIGURE 4.7 : Suite figure 4.6

4.7.2 Modèle génotypique

Ce modèle est à la base des hypothèses développées dans 4.4, mais sans la prise en compte de la composante spatiale (cf. Tableau 4.1). Lors de la simulation de ce modèle, nous avons directement constaté une surestimation de la fusion pour les corps lipidiques de petit volume dans les courbes de volume cumulé estimées, et ce, pratiquement chez tous les génotypes (cf. Figures 4.8 et 4.9). L'erreur moyenne de l'estimation sur tous les génotypes, par rapport aux données expérimentales, est de 0.0327 (cf. Tableau 4.5). Nous avons donc décidé de moduler la puissance du processus de coalescence par un paramètre supplémentaire β , que nous avons estimé, initialement, pour chacun des génotypes dans 4.7.3.

Paramètres								Erreur
α_1	α_2	α_3	α_{12}	α_{13}	α_{23}	α_{123}	γ	
0.0004	0.0109	0	-0.0003	0	0.0110	0.1339	2.6223	0.0327

TABLE 4.5 : Paramètres d'ajustement du modèle génotypique (Modèle 2).

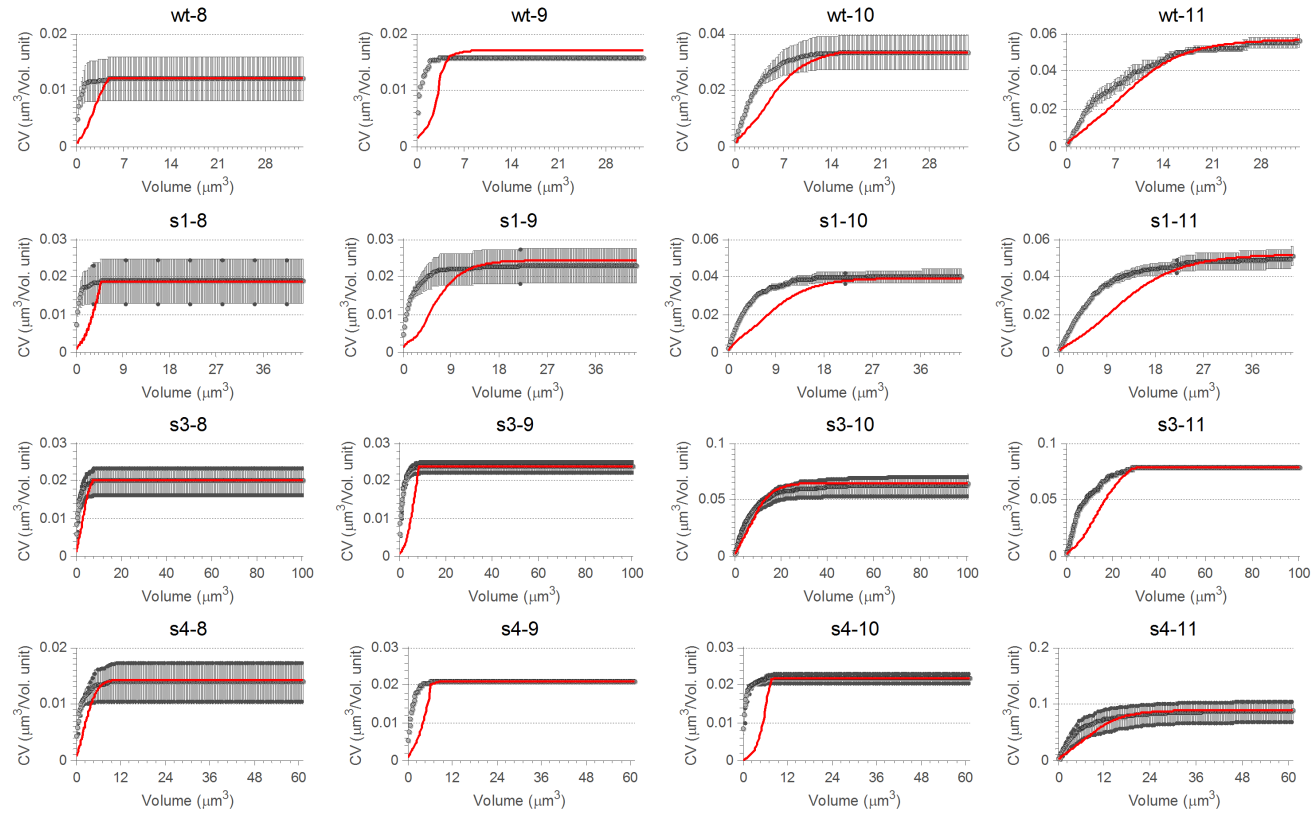


FIGURE 4.8 : Courbes d'ajustement du modèle génotypique (2) : volume de corps lipidiques cumulé par unité de volume de référence (CV). En noir, la densité CV observée avec erreur, en rouge, la densité estimée par ajustement des paramètres.

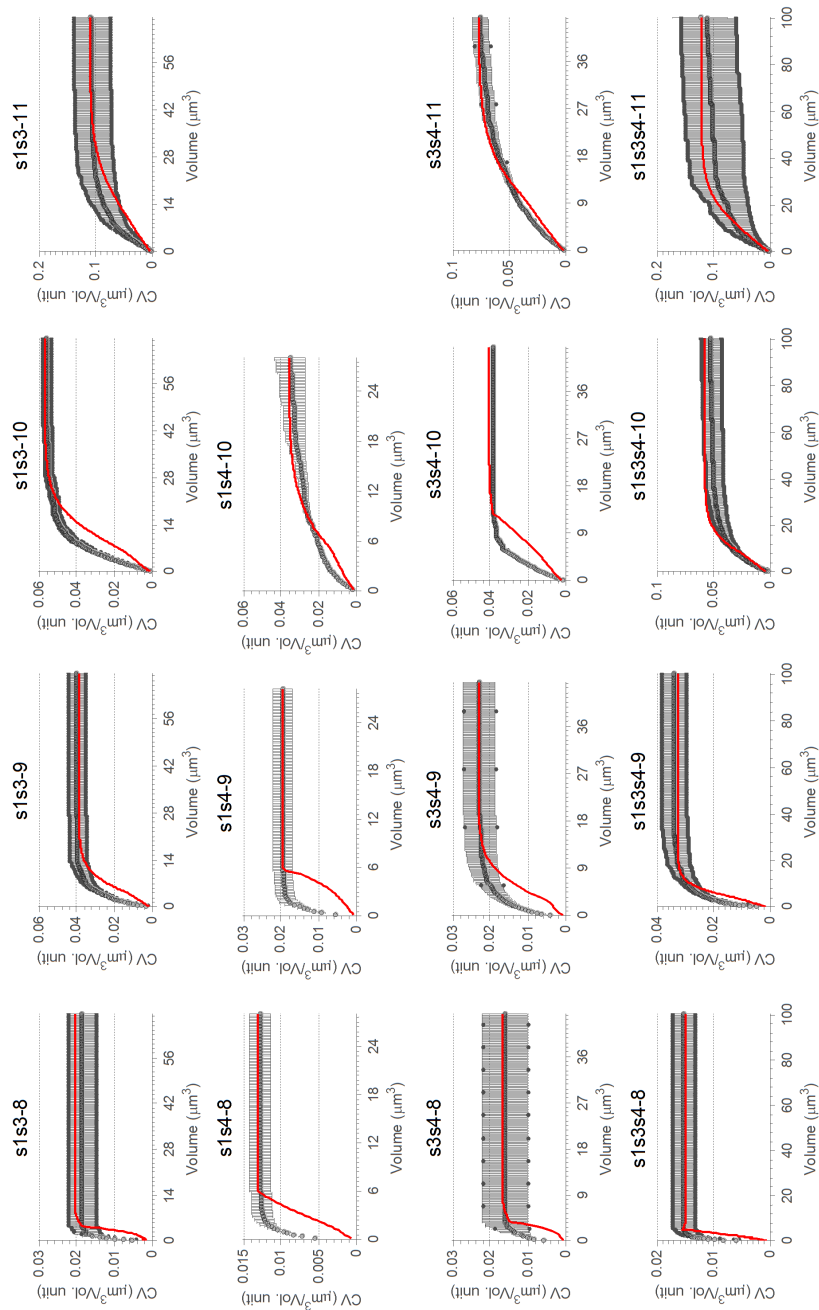


FIGURE 4.9 : Suite figure 4.8

4.7.3 Modèle multiplicatif

Le paramètre β estimé avec le noyau multiplicatif (cf. Tableau 4.4) est de valeurs différentes en fonction du génotype avec une valeur moyenne de -0.0531 ± 0.0562 globalement négative chez tous les génotypes sauf pour les génotypes $s4$ et $s1s4$. Sans la présence d'un autre processus, cela suggère une décroissance de la fusion proportionnelle aux volumes de corps lipidiques concernés par la fusion. Avec le paramètre β libre, il n'est donc pas judicieux de comparer les valeurs du paramètre η puisque β est différent d'un génotype à un autre ce qui modifie le taux de fusion. L'effet purement génotypique sur la coalescence n'est plus valable. Il est donc nécessaire d'estimer le paramètre β avec la même valeur pour tous les génotypes. Par ailleurs, l'ajout de ce paramètre (β) a permis une amélioration de la qualité d'ajustements (cf. Figures 4.10 et 4.11), avec une erreur moyenne de 0.0132 ± 0.0074 .

Génotype	Paramètre		Erreur
	η	β	
wt	345.98 ± 12.97	-0.142 ± 0.007	0.0169
$s1$	150.94 ± 6.50	-0.037 ± 0.010	0.0161
$s3$	88.66 ± 3.50	-0.105 ± 0.010	0.0161
$s4$	131.26 ± 5.72	0.020 ± 0.0101	0.0032
$s1s3$	197.46 ± 7.85	-0.098 ± 0.001	0.0069
$s1s4$	146.76 ± 15.31	0.087 ± 0.024	0.0221
$s3s4$	466.89 ± 20.90	-0.174 ± 0.007	0.0203
$s1s3s4$	471.92 ± 19.46	-0.104 ± 0.013	0.0041

TABLE 4.6 : Paramètres d'ajustement du modèle multiplicatif (Modèle 3).

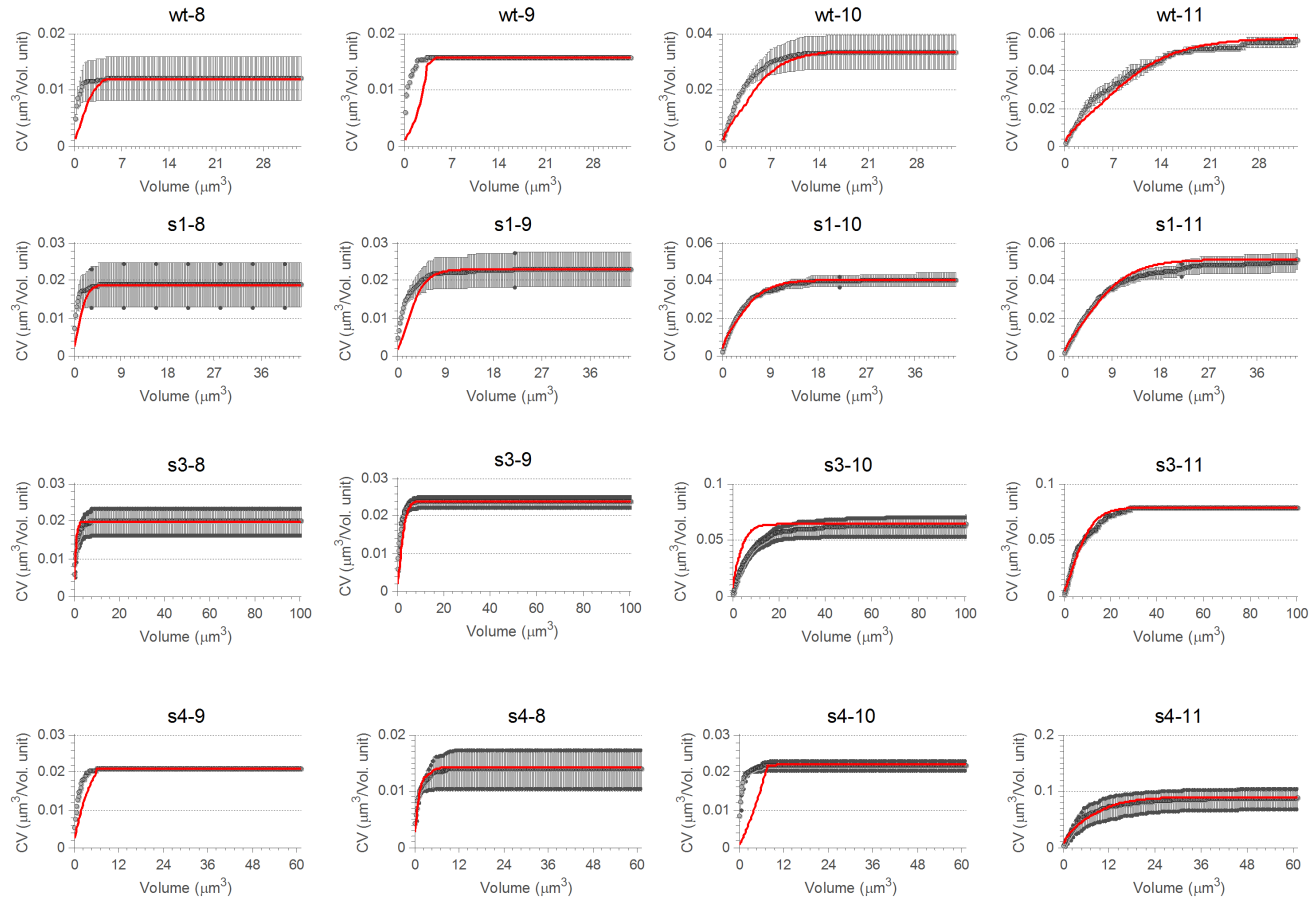


FIGURE 4.10 : Courbes d'ajustement du modèle multiplicatif (3) : volume de corps lipidiques cumulé par unité de volume de référence (CV). En noir, la densité CV observée avec erreur, en rouge, la densité estimée par ajustement des paramètres.

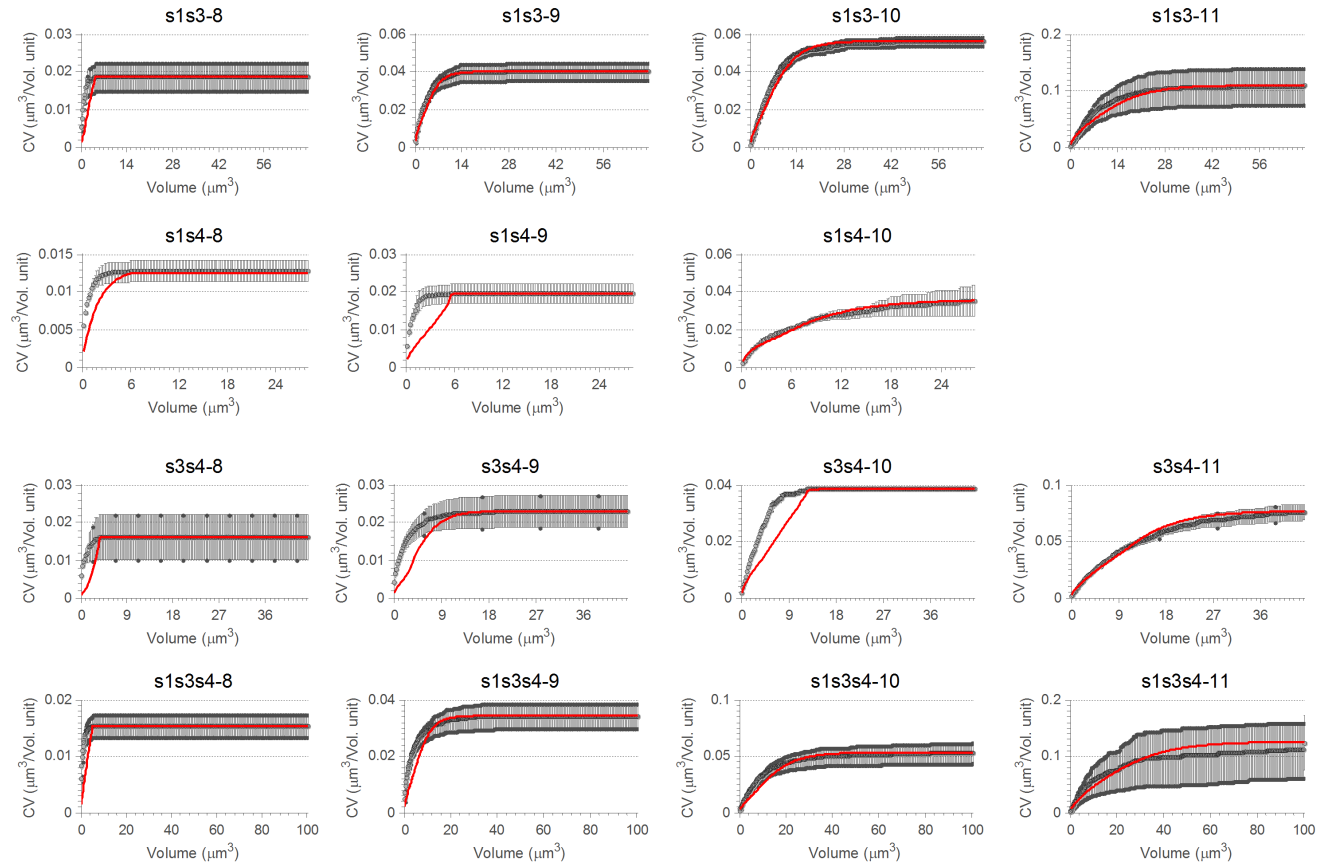


FIGURE 4.11 : Suite figure 4.10

4.7.4 Modèle génotypique/spatial

Dans ce modèle, nous avons pris en compte la composante spatiale. Nous avons donc souhaité tester son influence sur la qualité d'estimations. Le nombre de paramètres à estimer reste le même que pour le modèle 4.7.2, mais en rajoutant l'influence de la composante spatiale avec les paramètres θ_1 et θ_2 . Ce modèle a permis d'avoir une amélioration de la qualité d'estimations (cf. Figures 4.12 et 4.13). L'erreur est diminuée d'environ 30 % par rapport celle du modèle 4.7.2.

Paramètres										Erreur
α_1	α_2	α_3	α_{12}	α_{13}	α_{23}	α_{123}	γ	θ_1	θ_2	
0.5746	0.8117	0.0002	0.0056	0	0.0096	0.5402	2.6567	θ_1	θ_2	0.0248

TABLE 4.7 : Paramètres d'ajustement du modèle génotypique/spatial (Modèle 4).

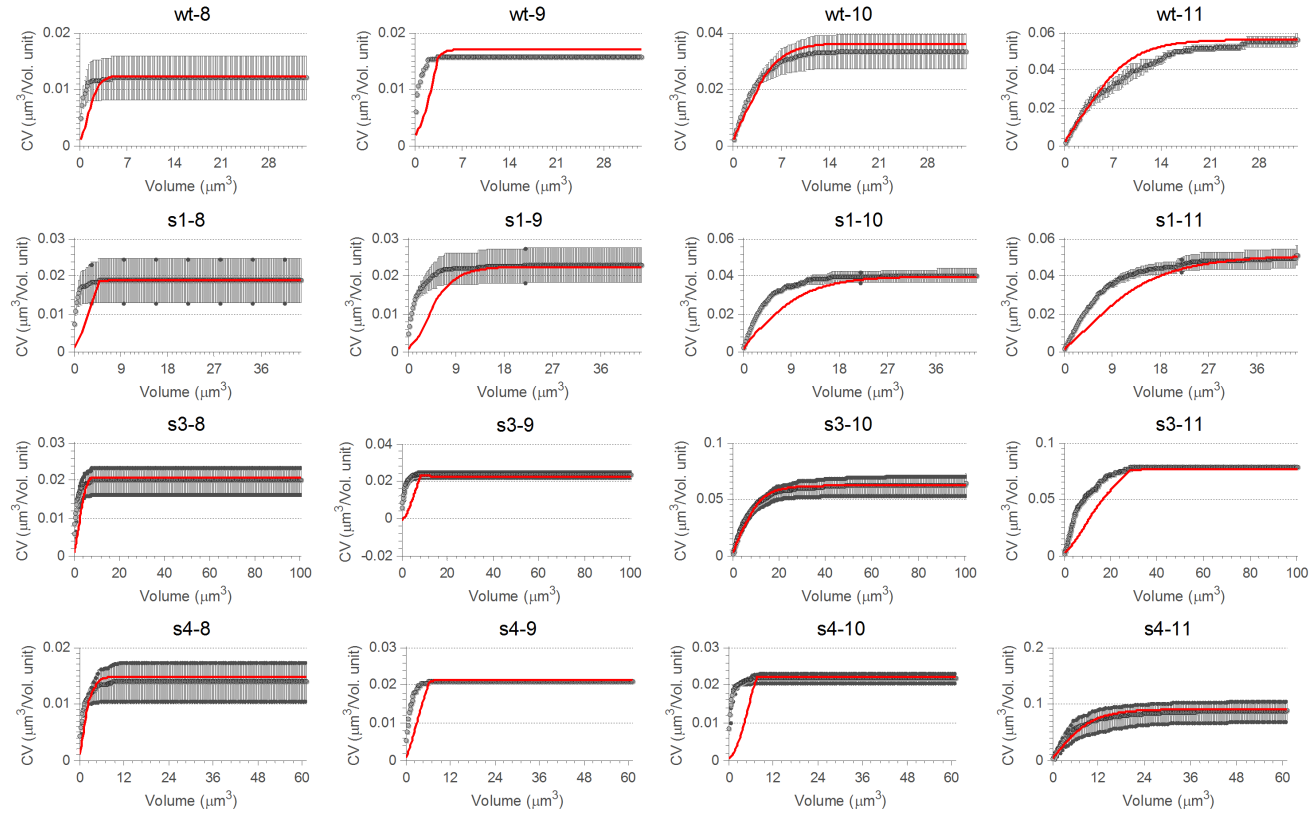


FIGURE 4.12 : Courbes d'ajustement du modèle génotypique/spatial (4) : volume de corps lipidiques cumulé par unité de volume de référence (CV). En noir, la densité CV observée avec erreur, en rouge, la densité estimée par ajustement des paramètres.

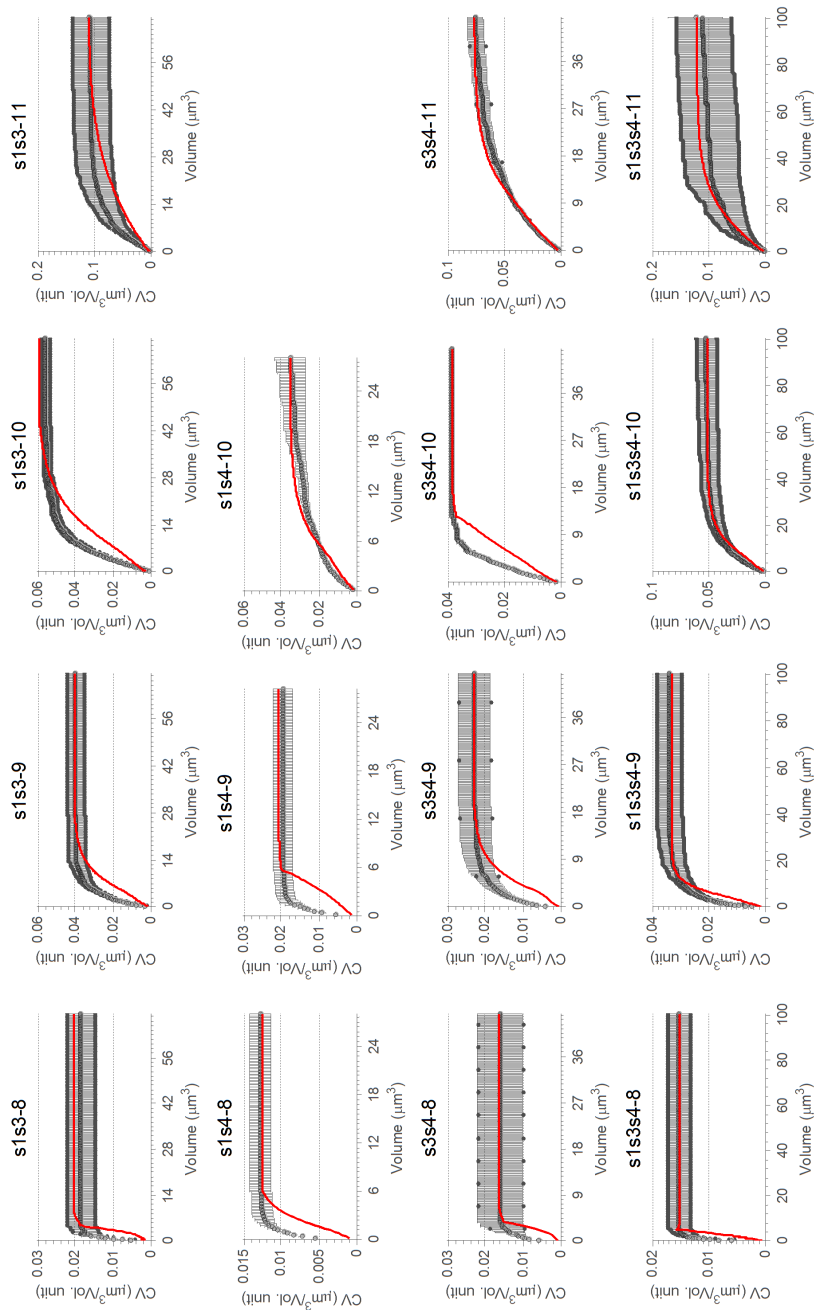


FIGURE 4.13 : Suite figure 4.12

4.7.5 Modèle génotypique/spatial modulé

Après avoir estimé d'une part, le paramètre β génotype par génotype (cf. 4.7.3), puis, d'autre part, analysé l'effet d'une composante spatio-temporelle dont les paramètres sont fixés à partir des données (cf. 4.7.4), nous libérons cette fois la partie fixe $\beta = -1/3$ de ce modèle, et estimons un paramètre β unique modulé par une composante spatio-temporelle fixée (modèle 5). Ce modèle tient compte à la fois de l'efficacité de charge (ou "fitness") $e(G)$, de la dispersion spatiale, ainsi que la modulation de la fusion par le paramètre β estimé sur tous les génotypes. Le critère d'évaluation par rapport aux données expérimentales est de la forme :

$$\varepsilon = \sum_{j=1}^5 \sum_{g=1}^8 \sum_{i=1}^N \frac{(F^g(x_i, t_j) - \widehat{F^g(x_i, t_j)})^2}{\sigma^M(i, g, j)^2}$$

avec $F^g(x_i, t_j) = \sum_{k < i} x_k n^g(x_k, t_j)$ désignant le volume total des corps lipidiques de volumes $< x_i$ au jour t_j .

Les paramètres estimés qui reflètent l'influence génotypique (cf. Tableau 4.8) sont alors discutés pour ce modèle :

- En absence d'oléosines (triple mutant), le taux de coalescence est maximal (i.e. efficacité de charge $e(G)$ minimale), soit $e(G) = \gamma = 7.5895$
- L'efficacité de charge en présence de l'oléosine S4 (génotype $s1s3$) est : $e(G) = \gamma + \alpha_3 = 10.7625$. Celle-ci est augmentée de 41% par rapport au triple mutant.
- L'efficacité de charge en présence de l'oléosine S3 (génotype $s1s4$) est : $e(G) = \gamma + \alpha_2 = 8.0335$. Celle-ci est augmentée de 5.8% par rapport au triple mutant.
- L'efficacité de charge en présence de l'oléosine S1 (génotype $s3s4$) est : $e(G) = \gamma + \alpha_1 = 7.8425$. Celle-ci est augmentée de 3.3% par rapport au triple mutant.
- L'efficacité de charge en présence des oléosines S3 et S4 (génotype $s1$) est : $e(G) = \gamma + \alpha_2 + \alpha_3 + \alpha_{23} = 11.9117$. L'ajout de S4 au mutant $s1s4$ contribue à une augmentation de l'efficacité de charge de $(\alpha_3 + \alpha_{23})/(\gamma + \alpha_2) = 48.2\%$. L'ajout de S3 au mutant $s1s3$ contribue à une augmentation de l'efficacité de charge de 10.6%.
- L'efficacité de charge en présence des oléosines S1 et S4 (génotype $s3$) est : $e(G) = \gamma + \alpha_1 + \alpha_3 + \alpha_{13} = 8.0293$. L'ajout de S4 au mutant $s3s4$ contribue à une augmentation de l'efficacité de charge de 2.3%. L'ajout de S1 au mutant $s1s3$ contribue à une réduction de l'efficacité de charge de 34%.
- L'efficacité de charge en présence des oléosines S1 et S3 (génotype $s4$) est : $e(G) = \gamma + \alpha_1 + \alpha_2 + \alpha_{12} = 15.0464$. L'ajout de S3 au mutant $s3s4$ contribue à une augmentation de l'efficacité de charge de 91.8%. L'ajout de S1 au mutant $s1s4$ contribue à une augmentation de l'efficacité de charge de 87.3%.
- L'efficacité de charge quand toutes les oléosines sont présentes (wt) est $e(G) = 16.8748$.

Il est clairement remarquable que toutes les oléosines contribuent à l'augmentation de l'efficacité de charge de la membrane du corps lipidique. Globalement, l'oléosine S4 est celle qui augmente le plus l'efficacité de charge, mais son effet semble être largement diminué par l'interaction avec l'oléosine S1 ($\alpha_{13} = -2.9863$

cf. 4.8). En revanche, l'interaction de S1 et S3 à un effet significatif sur l'augmentation de l'efficacité de charge ($\alpha_{12} = 6.7599$ cf. 4.8), et donc, la diminution du taux de coalescence.

L'erreur moyenne de l'ajustement par rapport aux données expérimentales est diminuée à 0.00041, ce qui fait un rapport d'environ 60 par rapport au modèle précédent. Visuellement, les ajustements sont acceptables, à l'exception de wt à 10 et 11 jours, $s3$ à 11 jours, $s1s4$ à 9 jours, et $s3s4$ à 10 jours (cf. 4.14 et 4.15).

Paramètres									Erreur
α_1	α_2	α_3	α_{12}	α_{13}	α_{23}	α_{123}	γ	β	
0.253	0.4440	3.1731	6.7599	-2.9863	0.7051	0.9365	7.5895	-0.0344	0.00041

TABLE 4.8 : Paramètres d'ajustement du modèle génotypique/spatial modulé (Modèle 5).

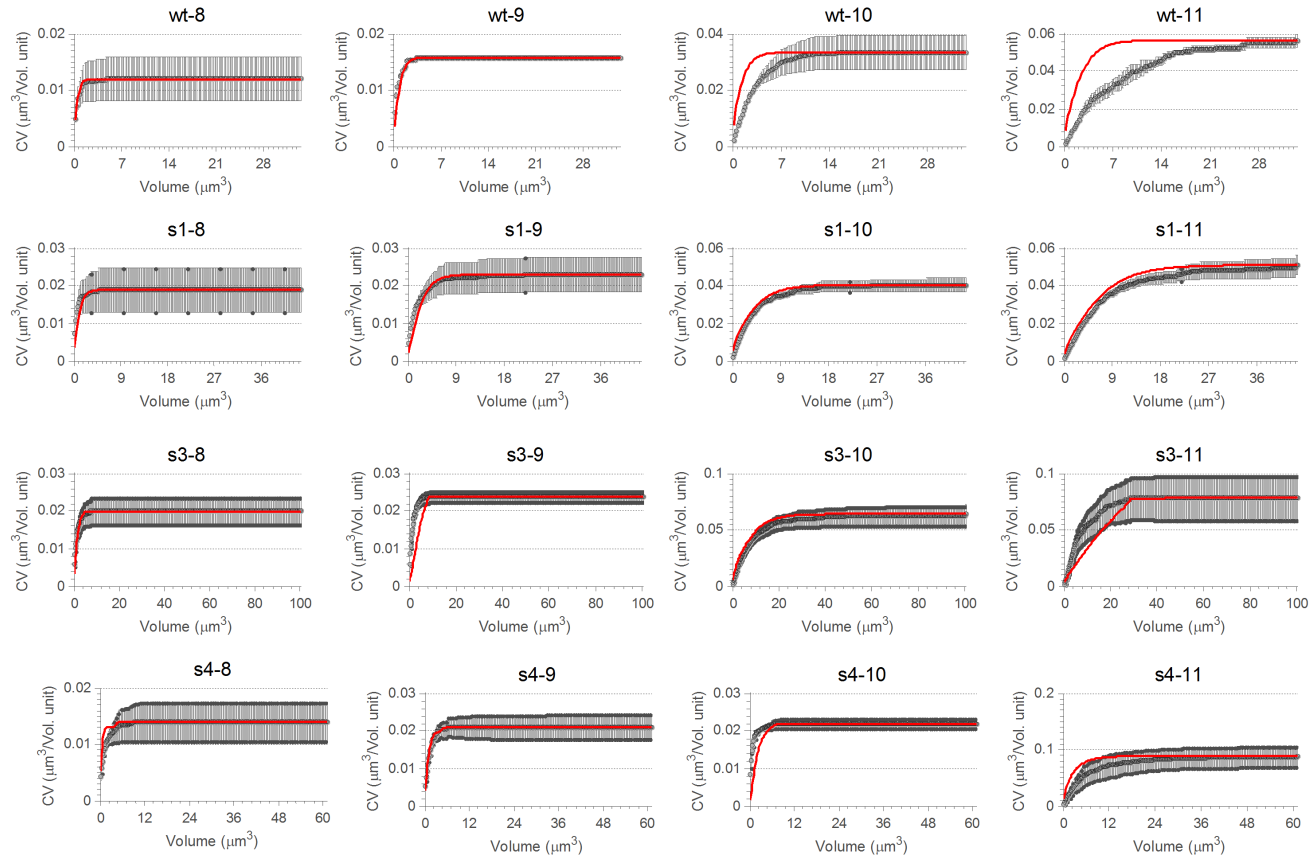


FIGURE 4.14 : Courbes d'ajustement du modèle génotypique/spatial modulé (5) : volume de corps lipidiques cumulé par unité de volume de référence (CV). En noir, la densité CV observée avec erreur, en rouge, la densité estimée par ajustement des paramètres.

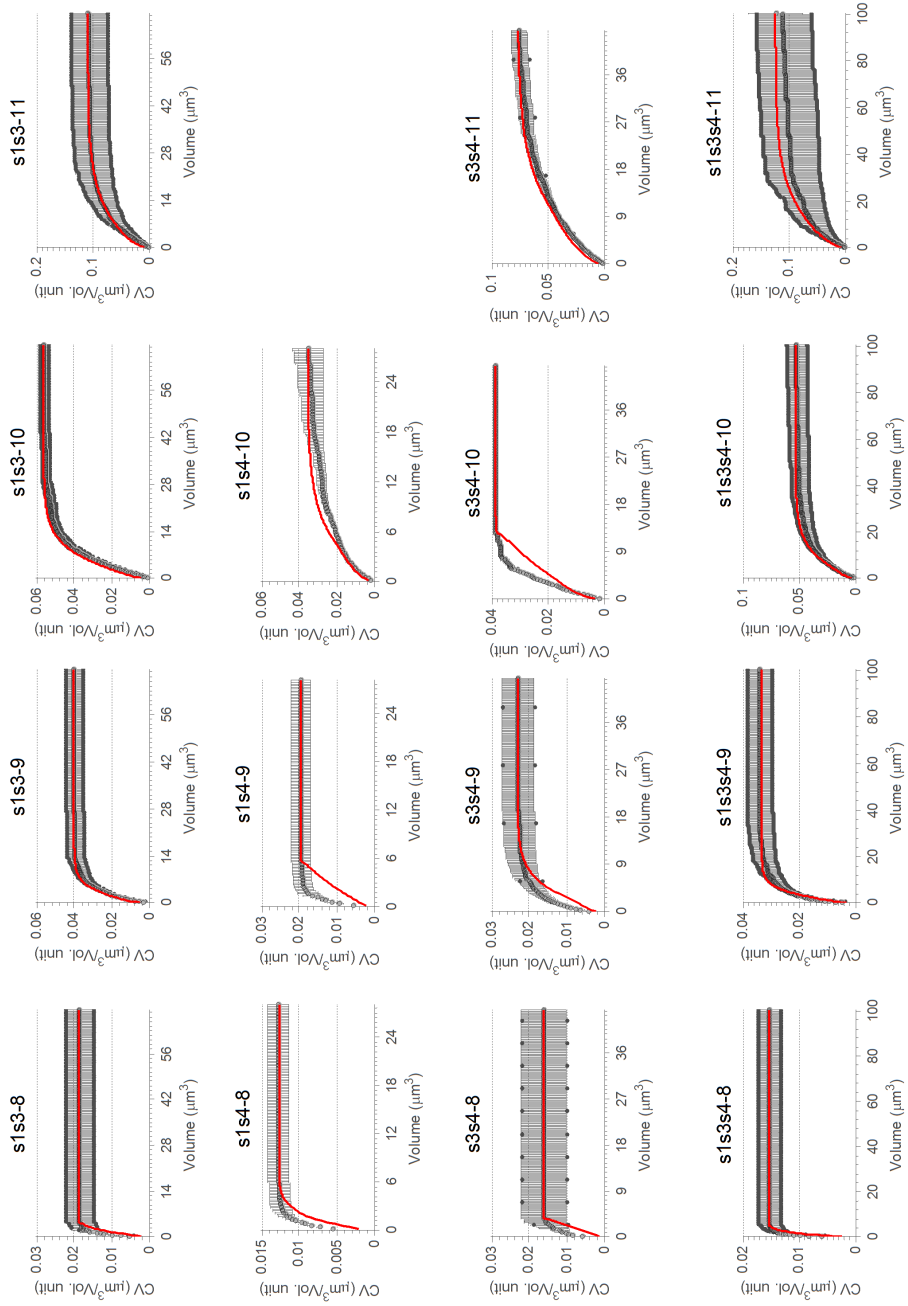


FIGURE 4.15 : Suite figure 4.14

4.8 Conclusion

Dans notre approche de modélisation, nous nous sommes placés à l'échelle de la population des corps lipidiques. Ce choix a été adopté en fonction des données de microscopie confocale tridimensionnelle disponibles, ainsi que la variable d'intérêt, qui est le volume individualisé de corps lipidiques, extraite à partir de l'analyse d'images. Les deux processus qui ont été modélisés sont la production, et la coalescence de corps lipidiques. Notre construction du modèle du noyau de coalescence est basée sur le peu de connaissances disponibles quant à l'effet de l'insertion des oléosines aussi nous avons du faire des hypothèses très fortes pour aboutir au modèle proposé. Ce modèle tient compte, à la fois, de la composante génotypique et la composante spatiale, sans avoir besoin à un recours explicite à la dimension spatiale dans la formulation (EDP en temps et volume). La composante spatiale est introduite via des données de dispersion spatiale de corps lipidiques dans l'espace cellulaire. Nous avons implémenté un algorithme déterministe pour simuler notre modèle et procéder à l'estimation des paramètres à l'aide d'un algorithme stochastique (MCMC). Les paramètres estimés ont montré différents effets des oléosines dans la contribution à l'efficacité de la membrane vis-à-vis de la coalescence. De plus, nous avons montré la diminution du processus de coalescence en fonction du temps. En conclusion, le modèle rend assez bien compte d'une assez grande proportion de génotype-date, même si certains ajustements demeurent mauvais avec parfois une sous-estimation apparente de la fusion chez le sauvage, et une surestimation de la fusion chez certain mutants.

Conclusion générale et perspectives

Dans ce travail, nous nous sommes placés à une échelle plus grande que celle de la déformation de la membrane tant pour des problèmes expérimentaux (limite de la résolution des microscopes) que conceptuels (mécanismes sous-jacents). Nous avons cherché à comprendre la dynamique de biogenèse des corps lipidiques chez la plante *A. thaliana*, en tenant compte de sa composition membranaire en trois isoformes d'oléosines : S1, S3 et S4. Notre démarche s'est effectuée autour de trois axes principaux : le traitement et l'analyse de données issues de microscopie confocale de corps lipidiques, l'analyse statistique des estimateurs qui les caractérisent et, enfin, la modélisation mathématique de leur biogenèse et dynamique. Dans la suite de ce chapitre, nous exposons les principaux résultats de ces trois axes, les limites, ainsi que les perspectives futures.

5.1 Résumé des principaux résultats produits

5.1.1 Mise en place d'un pipeline de traitement d'images

La visualisation par microscopie confocale des corps lipidiques au cours du développement de la graine d'*A. thaliana* a permis de mettre en évidence l'évolution de leur volume *in vivo*. A cet effet, les données des TAGs des corps lipidiques colorés au Rouge de Nil, obtenues sous forme de piles d'images 3D pour chaque génotype et chaque stade de développement étudiés (c.f. 2.3.1). Afin de pouvoir exploiter l'information apportée par ces données, nous avons développé un pipeline de traitement d'images qui exécute un ensemble d'opérations conduisant à l'étiquetage des corps lipidiques et au calcul de leurs volume et organisation spatiale (cf. 2.3.2.1). La procédure de segmentation d'images est basée sur l'algorithme de lignes de partage des eaux. Les principaux problèmes rencontrés concernent, comme souvent en biologie cellulaire, la proximité des objets à identifier, la large gamme des volumes des objets. La procédure est semi-automatique et nécessite la donnée de deux seuils. Les segmentations ont été validées une à une. Les principaux paramètres extraits à partir de ces segmentations sont : le volume individuel, la fraction locale et la distance au plus proche voisin. Les objets touchant le bord de la région échantillonnée ne sont pas pris en compte, le domaine extérieur à l'embryon est approximé à partir de l'enveloppe convexe des corps lipidiques. Le pipeline pourrait sans doute être appliqué à la segmentation d'autres organelles cellulaires (*e.g.* noyaux des cellules). Il a

été développé sous Matlab™ et AvizoFire™. Les données extraites ont alors été analysées à l'aide de trois "microscopes mathématiques" opérant à trois "résolutions" différentes comme expliqué ci-dessous :

5.1.2 Analyse exploratoire des données ("Microscope 1")

Nous avons d'abord réalisé une analyse de variance assez classique de données de volumes et de fraction locale. Pour cela, nous avons considéré un modèle linéaire comprenant des effets principaux pour chaque facteur de présence/absence des trois oléosines considérées, des effets d'interaction de paires, et un effet d'interaction triple ainsi qu'un effet principal sur le stade de développement. Nous avons tenu compte des disparités d'effectifs des corps lipidiques en fonction des génotypes et des stades. Nous avons considéré une transformation (en log) afin de stabiliser la variance. Nous avons mis en évidence des effets significatifs pour la plupart des facteurs aussi bien principaux que d'interactions. Cette analyse explique essentiellement la moyenne des volumes conditionnellement à un modèle linéaire choisi par avance.

5.1.3 Régression quantile ("Microscope 2")

Cette analyse ne rend pas toujours bien compte des données, notamment lorsque les queues de distributions portent une information importante. Nous avons donc utilisé une approche de régression quantile qui ne se limite pas à une modélisation de la moyenne, mais concerne plutôt différents quantiles. Nous avons donc estimé différents quantiles conditionnellement au modèle linéaire. Ce modèle linéaire, à nouveau, prend en compte des effets principaux de présence/absence, d'interaction de paires, de triple interaction ainsi qu'un effet principal sur la date. En considérant des quantiles suffisamment élevés, on a donc une chance de mettre en évidence des différences dues à la présence ou non de gros corps lipidiques. Nous avons pu observer des effets significatifs des facteurs, variables selon les quantiles.

5.1.4 Modélisation de la dynamique de coalescence des corps lipidiques ("Microscope 3")

Les hypothèses que nous avons adoptées pour la construction du modèle de croissance des corps lipidiques (cf. 4.4) étaient principalement basées sur le modèle proposé par Beisson (cf. Introduction générale).

Nous avons décidé d'essayer d'expliquer non pas la moyenne, ou encore quelques quantiles de la distribution de volumes, mais la distribution dans sa totalité. Pour cela, nous avons adopté une approche de modélisation de la dynamique de populations de corps lipidiques. Nous avons fait l'hypothèse que les deux processus les plus importants à prendre en compte étaient les processus de production et de coalescence des corps lipidiques. Nous avons très tôt retenu l'hypothèse de coalescence car elle était évoquée dans plusieurs articles, sans pour autant avoir été démontrée chez la plante (c.f Introduction générale). Cette hypothèse s'est avérée exacte comme nous l'avons montré en 2.4. Nous avons adopté le formalisme de l'équation d'agrégation

de Smoluchowski et proposé un noyau de coalescence multiplicatif basé sur la densité de charge sur la demi-membrane. Ce noyau prend en compte des facteurs génétiques et spatiaux. Nous avons implémenté un algorithme permettant la simulation des densités en nombres de corps lipidiques par volume de référence, selon le modèle de coalescence proposé. Nous avons alors estimé les paramètres du modèle proposé et de différentes variantes à partir des données que nous avons extraites (cf. Chapitre 2). Avec le modèle proposé, intégrant à la fois des facteurs génétiques et spatiaux, nous avons obtenu des ajustements¹ de bonne qualité pour un grand nombre de génotypes et stades. Si l'on considère uniquement la partie génétique du noyau, on trouve une dépendance en puissance -0.36 du noyau par rapport au volume. Ce qui est assez proche de notre construction initiale (cf. 4.3.2) pour laquelle la dépendance est en $-1/3$. Quand on couple la partie génétique et la partie spatiale, cette dépendance devient positive, décroissante en fonction du temps.

Toutefois, certains ajustements ne sont pas satisfaisants. Les causes peuvent être diverses : (i) l'approximation numérique de l'équation d'état (Smoluchowski), (ii) le choix du noyau, (iii) l'estimation des paramètres, (iv) la variabilité des données, (v) le modèle de production, (vi) la censure sur les corps lipidiques de petits volumes et pour finir (vii) l'oubli d'autres processus. On peut certainement mieux contrôler l'erreur d'approximation numérique, qui est en partie due à la variabilité du taux de production au cours du temps (régularisation du taux de production). Pour ce qui est du choix du noyau et du taux de production, cela sera discuté en section suivante. Pour l'estimation des paramètres, deux principaux ingrédients entrent en ligne de compte : le choix de la variable aléatoire sur laquelle on base l'estimation (ici le volume cumulé) et la prise en compte de la variabilité des données. Davantage d'observations en microscopie électronique, ou à forte résolution devrait permettre d'en savoir plus sur les corps lipidiques de faible volume. Enfin, la possibilité d'existence d'autres processus est discutée en section suivante.

5.2 Qu'aurions pu nous changer pendant la thèse ?

Dans les données acquises, nous avons eu peu de répétitions sur certains génotypes, et il aurait donc fallu acquérir plus de données sur ces derniers. Par ailleurs, à l'acquisition d'un nouveau microscope confocal intervenu au cours de ce projet, nous avons pu visualiser des fusions de corps lipidiques et nous aurions pu essayer d'acquérir des données time-lapse avec un pas temporel suffisamment faible. Plus tôt, quelques données temporelles avaient été acquises avec le matériel disponible en début de thèse, mais avec un pas d'échantillonnage trop long. Nous aurions pu alors observer des événements de fusion non pas au cours des derniers mois mais un peu plus tôt dans le déroulement de la thèse. Ceci reste à faire et sera développé dans la section suivante. Il n'est pas inutile de rappeler que notre étude est basée sur des données acquises sur des échantillons différents et cela complique clairement l'analyse qui peut être faite de la biogénèse. Aussi, il n'est pas exclu que, notamment, dès

1. des données simulées aux données observées.

le jour 10 du développement, il puisse y avoir des événements de fission des corps lipidiques. Nous n'avons pas encore observé ces événements mais nous aurions pu ajouter un processus de fission dans notre modèle sans difficulté particulière hormis que cela complexifie le modèle et donc l'estimation des paramètres.

Nous aurions pu essayer d'évaluer l'effet quantitatif des différentes oléosines en considérant des plantes surexprimant une ou plusieurs oléosines. Nous avons produit ces plantes et tout cela est une question de temps et de moyens d'acquisition.

Avec les données dont nous disposions, nous pensons avoir choisi la bonne échelle pour étudier la dynamique. Nous avons fait le choix d'un modèle de population basé sur l'équation de Smoluchowski, nous aurions pu adopter un formalisme multi-agent spatio-temporel pour lequel chaque corps lipidique aurait été un agent soumis à différents processus de naissance, de fusion, de coalescence, de mouvement, etc. Néanmoins, la durée de simulation pourrait être assez lourde dans une approche d'estimation de paramètres.

5.3 Que faudrait-il faire après ce travail (perspectives) ?

5.3.1 Etude des caractéristiques biophysiques de la membrane des corps lipidiques

Dans le but de comprendre les aspects théoriques de la déformation de la membrane du corps lipidique lors du processus de biogenèse, nous avons commencé dans ce travail par explorer l'état de l'art des mécanismes de déformation des membranes biologiques, ainsi que les modèles biophysiques développés pour divers systèmes biologiques. Ces modèles sont particulièrement développés pour des vésicules formées de deux demi-membranes de phospholipides et dont le contenu est supposé être un fluide incompressible. L'application de ces modèles pour un système vésiculaire formé d'une demi-membranes requiert une certaine connaissance des caractéristiques mécaniques de la demi-membranes. En ce qui concerne l'application au corps lipidique de la plante, ces connaissances comprennent l'étude de la tension de surface induite par les protéines caractéristiques, notamment les oléosines. L'accès à ces paramètres permettra de déterminer le volume maximal que peut atteindre un corps lipidique avant la rupture de sa membrane du RE, les conditions de bourgeonnement des corps lipidiques, etc. Parmi les expériences qui peuvent être envisagées dans le cadre de l'étude de la demi-membrane du corps lipidique, nous pouvons citer l'analyse de la tension de surfaces et le module d'élasticité des vésicules géantes artificielles dont la surface serait couverte par différentes combinaisons d'oléosines. Les protocoles expérimentaux pour l'étude de cette caractéristique sont assez courants dans le domaine de la biophysique des membranes biologiques et parmi lesquels on peut citer la technique d'aspiration par micropipette (cf. Annexes A.5).

5.3.2 Validation du modèle de la dynamique de coalescence des corps lipidiques

Un peu tard dans ce projet, des événements de coalescence binaire (deux à deux) ont été observés en microscopie confocale à balayage rapide et en "time-lapse" (cf. 2.4). Ces observations sont liées à l'utilisation d'un microscope de toute dernière génération permettant des acquisitions très rapides en x, y, z et t et disponible depuis peu au laboratoire. En effet, les nouvelles techniques de marquage, d'imagerie rapide et à haute résolution sont de plus en plus maîtrisées à l'IJPB dans le contexte du développement de la plante et de l'imagerie des corps lipidiques en particulier. Des acquisitions en "time-lapse" de l'embryon des différents génotypes d'*A. thaliana* sur de longues périodes (2 à 3 jours du développement) sont envisagées. Une fois disponibles, ces données pourraient être traitées par une procédure de segmentation d'images capable de détecter et quantifier les événements de coalescence. En effet, la procédure que nous avons développée en 2.3.2.1 pourrait être améliorée et adaptée à cette fin. Ainsi, le taux de fusion du modèle 4.4 pourrait être déduit à partir des données réelles et pourrait servir comme moyen de validation² des simulations numériques. Ces approches reposent cependant sur l'hypothèse que les mécanismes de fusion mis en oeuvre dans l'embryon restent actifs dans des embryons isolés pendant de longues périodes ce qui n'a jamais été observé à ce jour. En outre, les données de surexpression sont un autre moyen de validation. Hormis la coalescence, il existe plusieurs hypothèses concernant la dynamique de croissance des corps lipidiques (cf. Chapitre 1). Ces hypothèses comprennent notamment la croissance par synthèse locale, par remplissage et/ou par le processus dit de "mûrissement d'Ostwald". En outre, la décroissance par fission de corps lipidiques est aussi évoquée, à un instant avancé du développement, dans la littérature et beaucoup plus tôt selon nos données. Un ensemble de modèles tenant compte de l'un ou plusieurs de ces processus peut être évalué.

En outre, plusieurs approches expérimentales, notamment génétique et d'imagerie du vivant, peuvent être envisagées dans le cadre de l'étude de la dynamique du corps lipidiques, à savoir :

- Les gènes qui codent pour une oléosine donnée, peuvent être surexprimés *in vivo* de façon à augmenter le niveau intracellulaire de cette protéine modifiée. Cette méthode permettra d'étudier l'effet sur l'accumulation des lipides, ainsi que sur la taille des corps lipidiques dont la composition membranaire en termes de densité de l'oléosine surexprimée, est modifiée par rapport au type sauvage. D'un autre côté, il est possible, toujours par des techniques génétiques, d'inter-changer les domaines N, C terminaux, et le domaine central (Knob) entre oléosines et de créer ainsi des oléosines chimériques. Par exemple

2. Généralement, la validation d'un modèle se repose sur des données expérimentales différentes de celles utilisées pour la construction ou l'identification des paramètres du modèle. Les données de validation comprennent une série d'expérience avec des systèmes particuliers tel des données soumises à des perturbations, des composantes biologiques sur- ou sous-exprimées, etc. Si le modèle simulé avec les données d'apprentissage reproduit correctement ces données de validation, le modèle est considéré comme valide.

nous pouvons imaginer des couplages pour les oléosines S3 et S4 du genre : 4N-3K-3C, 3N-4K-3C, 3N-3K-4C, 4N-4K-3C, etc., où N et C représentent les parties N- et C- terminales, et K la partie centrale des oléosines. Il sera ainsi possible d'observer l'effet distinct des parties constituant les oléosines sur l'accumulation des lipides et la taille des corps lipidiques.

- Il est possible de suivre, à l'échelle de la demi-membrane, les éventuels changements de courbures lors d'un événement de fusion. En effet, les oléosines peuvent être fusionnées à un marqueur fluorescent (e.g. GFP) traçant plus finement la surface des corps lipidiques et permettant ainsi le suivi les conformations de la demi-membrane aux voisinages des sites de fusion entre deux corps lipidiques. Le choix a été fait de ne pas introduire de marqueurs fluorescents de la demi-membrane dans le cadre de ce travail, mais le matériel biologique est disponible.
- L'utilisation du double marquage peut être envisagée pour étudier la façon par laquelle les oléosines sont distribuées sur la membrane. On pourra, par exemple, utiliser deux marqueurs différents pour deux types d'oléosines dans le même génotype. Diverses questions peuvent ainsi être posées : (i) Est-ce que les oléosines sont distribuées de la même façon à la surface des corps lipidiques ?, (ii) Est-ce que cette distribution peut changer lors, ou après un événement de fusion de deux corps lipidiques ? (iii) Est-ce qu'il existe une relation entre la taille du corps lipidique et la distribution d'oléosines ?, etc.
- Avec un marqueur fluorescent photoactivable, il est possible d'étudier la mobilité des oléosines. On peut photoactiver des oléosines avant fusion et regarder si, après fusion, les oléosines sont très mobiles ou restent globalement au point ou autour du point de fusion. Dans la même optique, il est possible de faire des analyses FRAP (Fluorescence Recovery After Photobleaching) (cf. A.4) pour photoinactiver le marqueur fluorescent et regarder à quelle vitesse la fluorescence revient, là, aussi à des fins d'analyse de fluidité des protéines dans la demi-membrane et d'estimation de leur diffusion latérale. Il est aussi possible d'investiguer d'éventuelles interactions entre oléosines par FRET (Förster Resonance Energy Transfer) (cf. A.4). Enfin, l'utilisation de la microscopie super-résolution de type STED (stimulated-emission-depletion ou déplétion par émission stimulée) permettrait d'obtenir un gain de l'ordre de x4 dans la résolution des images et, de là avoir plus de finesse dans l'observation des mécanismes de fusion/fission.

Les connaissances sur la biogenèse et la dynamique du corps lipidique n'ont cessé de croître ces dernières années. Il est aussi évident que les approches théoriques de la déformation des membranes biologiques, ainsi que les études expérimentales sur des systèmes artificiels, ont dévoilé plusieurs réalités sur les propriétés biophysiques des corps lipidiques. Il faut cependant noter que les tentatives de modélisation restent limitées.

5.3.3 Faire de nouvelles hypothèses biologiques sur la base du modèle

Nous avons évoqué en 5.3.2 différents autres processus notamment la synthèse locale et le « mûrissement d'Ostwald ». S'il s'avère que ces processus pourraient être observés dans les données, nous pourrions les intégrer dans le modèle de population. Ainsi, la prise en compte d'un processus de croissance consiste à modifier l'équation de Smoluchowski (Équation 4.8) dans laquelle apparaît un taux de croissance qu'il faudrait modéliser en tenant compte de composante spatiale et génotypique. Le modèle pourrait alors avoir la forme suivante :

$$\begin{aligned} \partial_t n(x, t) + \partial_s G(v) n(x, t) = & \frac{1}{2} \int_0^x K(x-y, y) n(x-y, t) n(y, t) dy \\ & - \int_0^\infty K(x, y) n(x, t) n(y, t) dy \end{aligned}$$

où, ici, $G(v)$ désigne le taux de croissance qui ne dépend que du volume v à l'instant t .

5.4 Conclusion générale

Les corps lipidiques ont longtemps été considérés comme de simples réservoirs inertes de stockage et de mobilisation des réserves. Les nouvelles technologies de marquage et d'imagerie cellulaire ont récemment été utilisées dans le but d'étudier ces structures. Grâce à notre analyse quantile, de nouveaux aspects du corps lipidique ont été révélés et ont mis en évidence la richesse de sa composition et l'organisation de sa structure.

Chez la plante, les corps lipidiques fournissent des molécules, dont les oléosines, dotées de propriétés physiques activement recherchées par différentes industries. Les connaissances du fonctionnement de ces molécules dans leur milieu naturel (la cellule vivante) est l'un des éléments clé dans la recherche de nouvelles applications et l'approche de modélisation de cette dynamique que nous avons mise en place, permettra sans doute, une fois complétée et certains paramètres estimés, de contribuer à une meilleure compréhension des phénomènes sous-jacents.

Bibliographie

- [Abell 1997] B M Abell, L A Holbrook, M Abenes, D J Murphy, M J Hills et M M Moloney. *Role of the proline knot motif in oleosin endoplasmic reticulum topology and oil body targeting*. The Plant Cell, vol. 9, no. 8, Août 1997. (Cité en page 3.)
- [Abell 2004] Ben M Abell, Margaret Hahn, Larry A Holbrook et Maurice M Moloney. *Membrane topology and sequence requirements for oil body targeting of oleosin*. The Plant Journal, vol. 37, no. 4, Février 2004. (Cité en page 3.)
- [Adachi] Y. Adachi, M. a. C. Stuart et R. Fokkink. *Kinetics of turbulent coagulation studied by means of End-over-End rotation*. Journal of colloid and interface science, vol. 165, no. 2, pages 310–317. (Cité en page 95.)
- [Alberts 2000] Bruce Alberts, Julian Lewis et Dennis Bray. *Molecular biology of the cell*, 2000. (Cité en page 12.)
- [Aldous 1999] David J. Aldous. *Deterministic and stochastic models for coalescence (aggregation and coagulation) : a review of the mean-field theory for probabilists*. Bernoulli, vol. 5, no. 1, pages 3–48, Février 1999. Mathematical Reviews number (MathSciNet) : MR1673235 ; Zentralblatt MATH identifier : 0930.60096. (Cité en page 94.)
- [Allen 2001] A. W. Allen, B. S. Cade et M. W. Vandever. *Effects of emergency haying on vegetative characteristics within selected conservation reserve program fields in the northern great plains*. Journal of Soil and Water Conservation, vol. 56, no. 2, pages 120–125, Janvier 2001. (Cité en page 76.)
- [Ball 1990] J. M. Ball et J. Carr. *The discrete coagulation-fragmentation equations : Existence, uniqueness, and density conservation*. Journal of Statistical Physics, vol. 61, no. 1-2, pages 203–234, Octobre 1990. (Cité en pages 91, 93, 94 et 157.)
- [Banas 2000] A Banas, A Dahlqvist, H Debski, P O Gummeson et S Stymne. *Accumulation of storage products in oat during kernel development*. Biochemical Society Transactions, vol. 28, no. 6, pages 705–707, Décembre 2000. (Cité en page 3.)
- [Barber 1996] C. Bradford Barber, David P. Dobkin et Hannu Huhdanpaa. *The Quickhull Algorithm for Convex Hulls*. ACM Trans. Math. Softw., vol. 22, no. 4, pages 469–483, Décembre 1996. (Cité en page 35.)
- [Beaudoin 2000] Frédéric Beaudoin et Johnathan A. Napier. *The targeting and accumulation of ectopically expressed oleosin in non-seed tissues of Arabidopsis thaliana*. Planta, vol. 210, pages 439–445, Février 2000. (Cité en page 3.)
- [Beisson 1999] Frédéric Beisson. Thèse de doctorat : Etude des oléosomes de plantes et de leur lipolyse : Méthodes de dosage de l'activité des lipases. Université Aix-Marseille 2 : 1999, 1999. (Cité en page 5.)

- [Beucher 1982] S. Beucher. *Watersheds of functions and picture segmentation*. In Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82., volume 7, pages 1928 – 1931, Mai 1982. (Cit  en page 81.)
- [Boulanger 2010] J r me Boulanger, Charles Kervrann, Patrick Bouthemy, Peter Elbau, Jean-Baptiste Sibarita et Jean Salamero. *Patch-based nonlocal functional for denoising fluorescence microscopy image sequences*. IEEE transactions on medical imaging, vol. 29, no. 2, F vrier 2010. (Cit  en page 81.)
- [Brasaemle 2000] D L Brasaemle, B Rubin, I A Harten, J Gruia-Gray, A R Kimmel et C Londos. *Perilipin A increases triacylglycerol storage by decreasing the rate of triacylglycerol hydrolysis*. The Journal of biological chemistry, vol. 275, no. 49, pages 38486–38493, D cembre 2000. (Cit  en page 19.)
- [Brown 1998] D. A. Brown et E. London. *Functions of Lipid Rafts in Biological Membranes*. Annual Review of Cell and Developmental Biology, vol. 14, no. 1, 1998. (Cit  en page 9.)
- [Brown 2003] Rebecca L. Brown et Robert K. Peet. *Diversity and invasibility of southern appalachian plant communities*. Ecology, vol. 84, no. 1, pages 32–39, Janvier 2003. (Cit  en page 76.)
- [Cade 1999] Brian S. Cade, James W. Terrell et Richard L. Schroeder. *Estimating Effects of Limiting Factors with Regression Quantiles*. Ecology, vol. 80, no. 1, page 311, Janvier 1999. (Cit  en page 75.)
- [Cade 2003] Brian S. Cade et Barry R. Noon. *A gentle introduction to quantile regression for ecologists*. Frontiers in Ecology and the Environment, vol. 1, no. 8, pages 412–420, Octobre 2003. (Cit  en page 75.)
- [Campelo 2008] Felix Campelo, Harvey T McMahon et Michael M Kozlov. *The hydrophobic insertion mechanism of membrane curvature generation by proteins*. Biophysical journal, vol. 95, no. 5, Septembre 2008. (Cit  en page 15.)
- [Canham 1970] P.B. Canham. *The minimum energy of bending as a possible explanation of the biconcave shape of the human red blood cell*. Journal of Theoretical Biology, vol. 26, no. 1, pages 61–81, Janvier 1970. (Cit  en page 13.)
- [Canny 1986] J Canny. *A Computational Approach to Edge Detection*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 8, no. 6, page 679’698, Juin 1986. (Cit  en page 30.)
- [Capuano 2007] Floriana Capuano, Fr d ric Beaudoin, Johnathan A Napier et Peter R Shewry. *Properties and exploitation of oleosins*. Biotechnology advances, vol. 25, no. 2, Avril 2007. (Cit  en page 3.)
- [Cartwright 2012] Bethany R Cartwright et Joel M Goodman. *Seipin : from human disease to molecular mechanism*. Journal of lipid research, vol. 53, no. 6, pages 1042–1055, Juin 2012. PMID : 22474068 PMCID : PMC3351812. (Cit  en page 5.)

- [Chalfie 1994] M Chalfie, Y Tu, G Euskirchen, W W Ward et D C Prasher. *Green fluorescent protein as a marker for gene expression*. Science (New York, N.Y.), vol. 263, no. 5148, Février 1994. (Cité en page 28.)
- [Chapman 2012] Kent D Chapman, John M Dyer et Robert T Mullen. *Biogenesis and functions of lipid droplets in plants : Thematic Review Series : Lipid Droplet Synthesis and Metabolism : from Yeast to Man*. Journal of lipid research, vol. 53, no. 2, pages 215–226, Février 2012. PMID : 22045929. (Cité en page 18.)
- [Chen 1999] Jeff C. F. Chen, Corinne C. Y. Tsai et Jason T. C. Tzen. *Cloning and Secondary Structure Analysis of Caleosin, a Unique Calcium-Binding Protein in Oil Bodies of Plant Seeds*. Plant and Cell Physiology, vol. 40, no. 10, pages 1079–1086, Janvier 1999. (Cité en page 80.)
- [Cottet 2004] Georges-Henri Cottet et Emmanuel Maitre. *A level-set formulation of immersed boundary methods for fluid-structure interaction problems*. Comptes Rendus Mathématique, vol. 338, no. 7, pages 581–586, Avril 2004. (Cité en page 15.)
- [Cottet 2006] Georges-Henri Cottet et Emmanuel Maitre. *A level set method for fluid-structure interaction with immersed surfaces*. Mathematical Models and Methods in Applied Sciences, vol. 16, no. 03, pages 415–438, Mars 2006. (Cité en page 15.)
- [Cummins 1992] Ian Cummins et Denis J. Murphy. *cDNA sequence of a sunflower oleosin and transcript tissue specificity*. Plant Molecular Biology, vol. 19, no. 5, pages 873–876, Août 1992. (Cité en page 2.)
- [Cummins 1993] Ian Cummins, Matthew J. Hills, Joanne H. E. Ross, Douglas H. Hobbs, Martin D. Watson et Denis J. Murphy. *Differential, temporal and spatial expression of genes involved in storage oil and oleosin accumulation in developing rapeseed embryos : implications for the role of oleosins and the mechanisms of oil-body formation*. Plant Molecular Biology, vol. 23, pages 1015–1027, Décembre 1993. (Cité en page 3.)
- [Damke 1994] H. Damke, T. Baba, D. E. Warnock et S. L. Schmid. *Induction of mutant dynamin specifically blocks endocytic coated vesicle formation*. The Journal of Cell Biology, vol. 127, no. 4, Novembre 1994. (Cité en page 12.)
- [Davies 1999] Susan C. Davies, John R. King et Jonathan A. D. Wattis. *The Smoluchowski coagulation equations with continuous injection*. Journal of Physics A : Mathematical and General, vol. 32, no. 44, page 7745, Novembre 1999. (Cité en page 95.)
- [de Almeida 2003] Rodrigo F. M. de Almeida, Aleksandre Fedorov et Manuel Prieto. *Sphingomyelin/Phosphatidylcholine/Cholesterol Phase Diagram : Boundaries and Composition of Lipid Rafts*. Biophysical Journal, vol. 85, no. 4, Octobre 2003. (Cité en page 9.)

- [Derganc 2007] Jure Derganc. *Curvature-driven lateral segregation of membrane constituents in Golgi cisternae*. Physical Biology, vol. 4, no. 4, page 317, Décembre 2007. (Cité en page 15.)
- [Dergaoui 2013] Hilel Dergaoui, Karine N. Sartelet, Edouard Debry et Christian Seigneur. *Modeling coagulation of externally mixed particles : Sectional approach for both size and chemical composition*. Journal of Aerosol Science, vol. 58, pages 17–32, Avril 2013. (Cité en page 98.)
- [Deuling 1976] H J Deuling et W Helfrich. *Red blood cell shapes as explained on the basis of curvature elasticity*. Biophysical Journal, vol. 16, no. 8, Août 1976. (Cité en page 14.)
- [Dietrich 2001] C Dietrich, L A Bagatolli, Z N Volovyk, N L Thompson, M Levi, K Jacobson et E Gratton. *Lipid rafts reconstituted in model membranes*. Biophysical Journal, vol. 80, no. 3, Mars 2001. (Cité en page 9.)
- [Doherty 2009] Gary J. Doherty et Harvey T. McMahon. *Mechanisms of Endocytosis*. Annual Review of Biochemistry, vol. 78, no. 1, 2009. (Cité en page 11.)
- [Dongen 1988] P. G. J. van Dongen et M. H. Ernst. *Scaling solutions of Smoluchowski's coagulation equation*. Journal of Statistical Physics, vol. 50, no. 1-2, pages 295–329, Janvier 1988. (Cité en page 91.)
- [Drake 1972] RL Drake. *A general mathematical survey of the coagulation equation*. Topics in Current Aerosol Research (Part 2), pages 201–376, 1972. (Cité en page 95.)
- [Du 2004] Qiang Du, Chun Liu et Xiaoqiang Wang. *A phase field approach in the numerical study of the elastic bending energy for vesicle membranes*. Journal of Computational Physics, vol. 198, no. 2, pages 450–468, Août 2004. (Cité en page 15.)
- [Du 2006] Qiang Du, Chun Liu et Xiaoqiang Wang. *Simulating the deformation of vesicle membranes under elastic bending energy in three dimensions*. Journal of Computational Physics, vol. 212, no. 2, pages 757–777, Mars 2006. (Cité en pages 15 et 16.)
- [Dubovskii 1996] P. B. Dubovskii et I. W Stewart. *Existence, Uniqueness and Mass Conservation for the Coagulation-Fragmentation Equation*. Mathematical Methods in the Applied Sciences, vol. 19, no. 7, pages 571–591, 1996. (Cité en pages 91 et 92.)
- [Escobedo 2002] M. Escobedo, S. Mischler et B. Perthame. *Gelation in Coagulation and Fragmentation Models*. Communications in Mathematical Physics, vol. 231, no. 1, pages 157–188, Août 2002. (Cité en page 91.)
- [Euler 1767] Leonhard Euler. *Recherches générales sur la mortalité et sur la multiplication du genre humain / par m. euler*. Deutsche Akademie der Wissenschaften zu, 1767. (Cité en page 89.)
- [Evans 1987] Evan. Evans et David. Needham. *Physical properties of surfactant bilayer membranes : thermal transitions, elasticity, rigidity, cohesion and*

- colloidal interactions*. The Journal of Physical Chemistry, vol. 91, no. 16, pages 4219–4228, Juillet 1987. (Cité en page 13.)
- [Farsad 2001] K Farsad, N Ringstad, K Takei, S R Floyd, K Rose et P De Camilli. *Generation of high curvature membranes mediated by direct endophilin bi-layer interactions*. The Journal of cell biology, vol. 155, no. 2, Octobre 2001. (Cité en page 12.)
- [Fei 2011] Weihua Fei, Ximing Du et Hongyuan Yang. *Seipin, adipogenesis and lipid droplets*. Trends in endocrinology and metabolism : TEM, vol. 22, no. 6, pages 204–210, Juin 2011. PMID : 21497513. (Cité en page 5.)
- [Filbet 2004] Francis Filbet et Philippe Laurençot. *Mass-conserving solutions and non-conservative approximation to the Smoluchowski coagulation equation*. Archiv der Mathematik, vol. 83, no. 6, pages 558–567, Décembre 2004. (Cité en page 94.)
- [Fitzenberger 2002] Bernd Fitzenberger, Roger Koenker et Jose A. F. Machado. *Economic applications of quantile regression*. Springer, Février 2002. (Cité en page 80.)
- [Ford 2002] Marijn G J Ford, Ian G Mills, Brian J Peter, Yvonne Vallis, Gerrit J K Praefcke, Philip R Evans et Harvey T McMahon. *Curvature of clathrin-coated pits driven by epsin*. Nature, vol. 419, no. 6905, Septembre 2002. (Cité en page 11.)
- [Fournier 2005] Nicolas Fournier et Philippe Laurençot. *Existence of Self-Similar Solutions to Smoluchowski's Coagulation Equation*. Communications in Mathematical Physics, vol. 256, no. 3, pages 589–609, Juin 2005. (Cité en page 92.)
- [Fratti 2004] Rutilio A Fratti, Youngsoo Jun, Alexey J Merz, Nathan Margolis et William Wickner. *Interdependent assembly of specific regulatory lipids and membrane fusion proteins into the vertex ring domain of docked vacuoles*. The Journal of cell biology, vol. 167, no. 6, Décembre 2004. (Cité en page 9.)
- [Gebregziabher 2011] Mulugeta Gebregziabher, Cheryl P. Lynch, Martina Mueller, Gregory E. Gilbert, Carrae Echols, Yumin Zhao et Leonard E. Egede. *Using quantile regression to investigate racial disparities in medication non-adherence*. BMC Medical Research Methodology, vol. 11, no. 1, page 88, Juin 2011. (Cité en page 80.)
- [Gong 2011] Jingyi Gong, Zhiqi Sun, Lizhen Wu, Wenyi Xu, Nicole Schieber, Dijin Xu, Guanghou Shui, Hongyuan Yang, Robert G Parton et Peng Li. *Fsp27 promotes lipid droplet growth by lipid exchange and transfer at lipid droplet contact sites*. The Journal of cell biology, vol. 195, no. 6, pages 953–963, Décembre 2011. (Cité en page 19.)
- [Goodman 2008] Joel M. Goodman. *The Gregarious Lipid Droplet*. Journal of Biological Chemistry, vol. 283, no. 42, pages 28005–28009, Octobre 2008. PMID : 18611863. (Cité en page 18.)

- [Greenspan 1985] P. Greenspan, E. P. Mayer et S. D. Fowler. *Nile red : a selective fluorescent stain for intracellular lipid droplets*. The Journal of cell biology, vol. 100, no. 3, Mars 1985. (Cité en page 27.)
- [Haario 2001] H. Haario, E. Saksman et J. Tamminen. *An adaptive Metropolis algorithm*. Bernoulli, vol. 7, pages 223–242, 2001. (Cité en page 101.)
- [Haario 2006] Heikki Haario, Marko Laine, Antonietta Mira et Eero Saksman. *DRAM : efficient adaptive MCMC*. Statistics and Computing, vol. 16, no. 4, pages 339–354, 2006. (Cité en page 100.)
- [Haibel 1998] A. Haibel, G. Nimtz, R. Pelster et R. Jaggi. *Translational diffusion in phospholipid bilayer membranes*. Physical Review E, vol. 57, no. 4, pages 4838–4841, Avril 1998. (Cité en page 12.)
- [Hanski 1999] Ilkka Hanski. Metapopulation ecology. Oxford University Press, Mars 1999. (Cité en page 89.)
- [Helfrich 1973] W. Helfrich. *Elastic properties of lipid bilayers : theory and possible experiments*. Zeitschrift für Naturforschung. Teil C : Biochemie, Biophysik, Biologie, Virologie, vol. 28, no. 11, Décembre 1973. (Cité en pages 14 et 15.)
- [Hendriks 1983] E. M. Hendriks, M. H. Ernst et R. M. Ziff. *Coagulation equations with gelation*. Journal of Statistical Physics, vol. 31, no. 3, pages 519–563, Juin 1983. (Cité en page 93.)
- [Hendriks 1984] E. M. Hendriks. *Exact solution of a coagulation equation with removal term*. Journal of Physics A : Mathematical and General, vol. 17, no. 11, page 2299, Août 1984. (Cité en page 95.)
- [Hethcote 2000] Herbert W. Hethcote. *The Mathematics of Infectious Diseases*. SIAM Review, vol. 42, no. 4, pages 599–653, Janvier 2000. (Cité en page 89.)
- [Hippel 1974] Peter H. von Hippel. *The Hydrophobic Effect. Formation of Micelles and Biological Membranes*. Charles Tanford. Wiley-Interscience, New York, 1973. viii, 200 pp., illus. \$12.50. Science, vol. 184, no. 4136, pages 559–560, Mars 1974. (Cité en page 10.)
- [Huang 1996] Aihc Huang. *Oleosins and Oil Bodies in Seeds and Other Organs*. Plant Physiology, vol. 110, no. 4, pages 1055–1061, Janvier 1996. (Cité en pages 2, 3 et 80.)
- [Israelachvili 1976] Jacob N. Israelachvili, D. John Mitchell et Barry W. Ninham. *Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers*. Journal of the Chemical Society, Faraday Transactions 2 : Molecular and Chemical Physics, vol. 72, no. 0, pages 1525–1568, Janvier 1976. (Cité en pages 9, 10 et 11.)
- [Jambunathan 2011] Sriarthika Jambunathan, Jun Yin, Waheed Khan, Yoshikazu Tamori et Vishwajeet Puri. *FSP27 Promotes Lipid Droplet Clustering and Then Fusion to Regulate Triglyceride Accumulation*. PLoS ONE, vol. 6, no. 12, page e28614, Décembre 2011. (Cité en page 19.)

- [Jo 2009] Junghyo Jo, Oksana Gavrilova, Stephanie Pack, William Jou, Shawn Mullen, Anne E. Sumner, Samuel W. Cushman et Vipul Periwal. *Hypertrophy and/or Hyperplasia : Dynamics of Adipose Tissue Growth*. PLoS Computational Biology, vol. 5, no. 3, Mars 2009. (Cit   en page 19.)
- [Jolivet 2004] Pascale Jolivet, Emeline Roux, Sabine D'Andrea, Marl    ne Davanture, Luc Negroni, Michel Zivy et Thierry Chardot. *Protein composition of oil bodies in Arabidopsis thaliana ecotype WS*. Plant physiology and biochemistry : PPB / Soci  t   fran  aise de physiologie v  g  tale, vol. 42, no. 6, Juin 2004. (Cit   en page 80.)
- [Jolivet 2009] Pascale Jolivet, C  line Boulard, Annick Bellamy, Colette Larr  , Marion Barre, H  l    ne Rogniaux, Sabine d'Andr  a, Thierry Chardot et Nathalie Nesi. *Protein composition of oil bodies from mature Brassica napus seeds*. Proteomics, vol. 9, no. 12, Juin 2009. (Cit   en pages 2, 19 et 78.)
- [Kalantari 2010] Fariba Kalantari, John J M Bergeron et Tommy Nilsson. *Biogenesis of lipid droplets  how cells get fatter*. Molecular membrane biology, vol. 27, no. 8, pages 462  468, Novembre 2010. PMID : 21142874. (Cit   en page 18.)
- [Kaoui 2009] Badr Kaoui, George Biro  s et Chaouqi Misbah. *Why Do Red Blood Cells Have Asymmetric Shapes Even in a Symmetric Flow ?* Physical review letters, vol. 103, no. 18, 2009. (Cit   en page 8.)
- [Karnovsky 1982] M J Karnovsky, A M Kleinfeld, R L Hoover et R D Klausner. *The concept of lipid domains in membranes*. The Journal of cell biology, vol. 94, no. 1, Juillet 1982. (Cit   en page 9.)
- [Kass 1988] Michael Kass, Andrew Witkin et Demetri Terzopoulos. *Snakes : Active contour models*. International Journal of Computer Vision, vol. 1, no. 4, pages 321  331, Janvier 1988. (Cit   en page 30.)
- [Katavic 2006] Vesna Katavic, Ganesh Kumar Agrawal, Martin Hajd  ch, Stefan L Harris et Jay J Thelen. *Protein and lipid composition analysis of oil bodies from two Brassica napus cultivars*. Proteomics, vol. 6, no. 16, Ao  t 2006. (Cit   en page 2.)
- [Keller 2008] Pernille Keller, John T. Petrie, Paul De Rose, Isabelle Gerin, Wendy S. Wright, Shian-Huey Chiang, Anders R. Nielsen, Christian P. Fischer, Bente K. Pedersen et Ormond A. MacDougald. *Fat-specific Protein 27 Regulates Storage of Triacylglycerol*. The Journal of Biological Chemistry, vol. 283, no. 21, pages 14355  14365, Mai 2008. (Cit   en page 19.)
- [Keren 2011] Kinneret Keren. *Cell motility : the integrating role of the plasma membrane*. European Biophysics Journal, vol. 40, no. 9, Septembre 2011. (Cit   en page 11.)
- [Knight 2002] Charles A. Knight et David D. Ackerly. *Variation in nuclear DNA content across environmental gradients : a quantile regression analysis*. Ecology Letters, vol. 5, no. 1, pages 66  76, 2002. (Cit   en page 76.)
- [Koenker 1978] Roger Koenker et Gilbert Bassett. *Regression Quantiles*. Econometrica, vol. 46, no. 1, page 33, Janvier 1978. (Cit   en page 80.)

- [Koenker 2005] Roger Koenker. Quantile regression. Cambridge University Press, Mai 2005. (Cité en page 75.)
- [König 1992] S. König, W. Pfeiffer, T. Bayerl, D. Richter et E. Sackmann. *Molecular dynamics of lipid bilayers studied by incoherent quasi-elastic neutron scattering*. Journal de Physique II, vol. 2, no. 8, pages 1589–1615, Août 1992. (Cité en page 12.)
- [Lacey 1998] D J Lacey, N Wellner, F Beaudoin, J A Napier et P R Shewry. *Secondary structure of oleosins in oil bodies isolated from seeds of safflower (*Carthamus tinctorius* L.) and sunflower (*Helianthus annuus* L.)*. The Biochemical journal, vol. 334 (Pt 2), Septembre 1998. (Cité en page 3.)
- [Lacey 1999] Dominic J Lacey, Frédéric Beaudoin, Christopher E Dempsey, Peter R Shewry et Johnathan A Napier. *The accumulation of triacylglycerols within the endoplasmic reticulum of developing seeds of *Helianthus annuus**. The Plant Journal, vol. 17, no. 4, pages 397–405, Février 1999. (Cité en pages 3 et 18.)
- [Lee 1974] A. G. Lee, N. J. M. Birdsall, J. C. Metcalfe, Penelope A. Toon et G. B. Warren. *Clusters in lipid bilayers and the interpretation of thermal effects in biological membranes*. Biochemistry, vol. 13, no. 18, pages 3699–3705, Août 1974. (Cité en page 9.)
- [Lee 2000] Man Hoi Lee. *On the Validity of the Coagulation Equation and the Nature of Runaway Growth*. Icarus, vol. 143, no. 1, pages 74–86, Janvier 2000. (Cité en page 95.)
- [Leptihn 2013] Sebastian Leptihn, Oliver K Castell, Brid Cronin, En-Hsin Lee, Linda C M Gross, David P Marshall, James R Thompson, Matthew Holden et Mark I Wallace. *Constructing droplet interface bilayers from the contact of aqueous droplets in oil*. Nature protocols, vol. 8, no. 6, Juin 2013. (Cité en page 2.)
- [Leslie 1945] P. H. Leslie. *On the use of matrices in certain population mathematics*. Biometrika, vol. 33, pages 183–212, Novembre 1945. PMID : 21006835. (Cité en page 89.)
- [Levin 2001] Jesse Levin. *For whom the reductions count : A quantile regression analysis of class size and peer effects on scholastic achievement*. Empirical Economics, vol. 26, no. 1, pages 221–246, 2001. (Cité en page 80.)
- [Leyvraz 1981] F. Leyvraz et H. R. Tschudi. *Singularities in the kinetics of coagulation processes*. Journal of Physics A : Mathematical and General, vol. 14, no. 12, page 3389, Décembre 1981. (Cité en page 93.)
- [Lipowsky 1995] R. Lipowsky et E. Sackmann. Structure and dynamics of membranes : I. from cells to vesicles / II. generic and specific interactions. Elsevier, Juin 1995. (Cité en pages 9 et 12.)
- [Londos 1999] C Londos, D L Brasaemle, C J Schultz, J P Segrest et A R Kimmel. *Perilipins, ADRP, and other proteins that associate with intracellular neutral*

- lipid droplets in animal cells*. Seminars in cell & developmental biology, vol. 10, no. 1, pages 51–58, Février 1999. (Cité en page 19.)
- [Lotka 1925] A. J. Lotka. Elements of physical biology. Williams and Wilkins Company, 1925. (Cité en page 89.)
- [Lushnikov 1976] A.A Lushnikov. *Evolution of coagulating systems : III. Coagulating mixtures*. Journal of Colloid and Interface Science, vol. 54, no. 1, pages 94–101, Janvier 1976. (Cité en page 95.)
- [Makowski 2007] David Makowski, Thierry Doré et Hervé Monod. *A new method to analyse relationships between yield components with boundary lines*. Agromony for Sustainable Development, vol. 27, no. 2, pages 119–128, Juin 2007. (Cité en pages 76 et 80.)
- [Malthus 1798] T. R. Malthus. An essay on the principle of population, as it affects the future improvement of society. with remarks on the speculations of mr. godwin, m. condorcet and other writers. London, J. Johnson, 1798. (Cité en page 89.)
- [Marks 2001] Bruno Marks, Michael H. B. Stowell, Yvonne Vallis, Ian G. Mills, Adele Gibson, Colin R. Hopkins et Harvey T. McMahon. *GTPase activity of dynamin and resulting conformation change are essential for endocytosis*. Nature, vol. 410, no. 6825, pages 231–235, Mars 2001. (Cité en page 12.)
- [McClain 2001] C. McClain et M. Rex. *The relationship between dissolved oxygen concentration and maximum size in deep-sea turrid gastropods : an application of quantile regression*. Marine Biology, vol. 139, no. 4, pages 681–685, Octobre 2001. (Cité en page 76.)
- [Miao 1991] Ling Miao, Bertrand Fourcade, Madan Rao, Michael Wortis et R. K. P. Zia. *Equilibrium budding and vesiculation in the curvature model of fluid lipid vesicles*. Physical Review A, vol. 43, no. 12, pages 6843–6856, Juin 1991. (Cité en page 15.)
- [Mira 2002] A. Mira. *Ordering and improving the performance of Monte Carlo Markov C*. Statistical Science, vol. 16, pages 340–350, 2002. (Cité en page 101.)
- [Morris 2001] C E Morris et U Homann. *Cell surface area regulation and membrane tension*. The Journal of membrane biology, vol. 179, no. 2, Janvier 2001. (Cité en page 11.)
- [Murphy 1989a] D. J. Murphy et D. M. Y. Au. *A new class of highly abundant apolipoproteins involved in lipid storage in oilseeds*. Biochemical Society Transactions, vol. v. 17(4) p. 682-683, 1989. (Cité en page 2.)
- [Murphy 1989b] Denis J. Murphy et Ian Cummins. *Seed oil-bodies : Isolation, composition and role of oil-body apolipoproteins*. Phytochemistry, vol. 28, no. 8, pages 2063–2069, 1989. (Cité en pages 3 et 80.)
- [Murphy 1991] D J Murphy, J N Keen, J N O’Sullivan, D M Au, E W Edwards, P J Jackson, I Cummins, T Gibbons, C H Shaw et A J Ryan. *A class of amphipathic proteins associated with lipid storage bodies in plants. Possible*

- similarities with animal serum apolipoproteins.* Biochimica et biophysica acta, vol. 1088, no. 1, Janvier 1991. (Cité en page 2.)
- [Murphy 1999a] D J Murphy. *Production of novel oils in plants.* Current Opinion in Biotechnology, vol. 10, no. 2, pages 175–180, Avril 1999. (Cité en page 3.)
- [Murphy 1999b] D J Murphy et J Vance. *Mechanisms of lipid-body formation.* Trends in biochemical sciences, vol. 24, no. 3, pages 109–115, Mars 1999. (Cité en page 3.)
- [Murphy 2001] D J Murphy. *The biogenesis and functions of lipid bodies in animals, plants and microorganisms.* Progress in lipid research, vol. 40, no. 5, Septembre 2001. (Cité en page 3.)
- [Paar 2012] Margret Paar, Christian Jüngst, Noemi A. Steiner, Christoph Magnes, Frank Sinner, Dagmar Kolb, Achim Lass, Robert Zimmermann, Andreas Zumbusch, Sepp D. Kohlwein et Heimo Wolinski. *Remodeling of Lipid Droplets during Lipolysis and Growth in Adipocytes.* Journal of Biological Chemistry, vol. 287, no. 14, pages 11164–11173, Mars 2012. (Cité en page 19.)
- [Peng 2007] Chi-Chung Peng, Viola S Y Lee, Meei-Yn Lin, Hsin-Yi Huang et Jason T C Tzen. *Minimizing the central hydrophobic domain in oleosin for the constitution of artificial oil bodies.* Journal of agricultural and food chemistry, vol. 55, no. 14, Juillet 2007. (Cité en page 3.)
- [Ploegh 2007] Hidde L. Ploegh. *A lipid-based model for the creation of an escape hatch from the endoplasmic reticulum.* Nature, vol. 448, no. 7152, pages 435–438, Juillet 2007. (Cité en page 18.)
- [Robenek 2004] Mirko J Robenek, Nicholas J Severs, Karin Schlattmann, Gabriele Plenz, Klaus-Peter Zimmer, David Troyer et Horst Robenek. *Lipids partition caveolin-1 from ER membranes into lipid droplets : updating the model of lipid droplet biogenesis.* FASEB journal : official publication of the Federation of American Societies for Experimental Biology, vol. 18, no. 7, pages 866–868, Mai 2004. PMID : 15001554. (Cité en page 18.)
- [Saffman 1956] P. G. Saffman et J. S. Turner. *On the collision of drops in turbulent clouds.* Journal of Fluid Mechanics, vol. 1, no. 01, pages 16–30, 1956. (Cité en page 95.)
- [Samsel 1984] R W Samsel et A S Perelson. *Kinetics of rouleau formation. II. Reversible reactions.* Biophysical Journal, vol. 45, no. 4, pages 805–824, Avril 1984. PMID : 6426540 PMCID : PMC1434900. (Cité en page 95.)
- [Sarmiento 1997] C Sarmiento, J H Ross, E Herman et D J Murphy. *Expression and subcellular targeting of a soybean oleosin in transgenic rapeseed. Implications for the mechanism of oil-body formation in seeds.* The Plant journal : for cell and molecular biology, vol. 11, no. 4, Avril 1997. (Cité en page 3.)
- [Sawada 2010] Takashi Sawada, Hideaki Miyoshi, Kohei Shimada, Akira Suzuki, Yuko Okamatsu-Ogura, James W. Perfield, Takuma Kondo, So Nagai, Chikara Shimizu, Narihito Yoshioka, Andrew S. Greenberg, Kazuhiro Kimura

- et Takao Koike. *Perilipin Overexpression in White Adipose Tissue Induces a Brown Fat-Like Phenotype*. PLoS ONE, vol. 5, no. 11, page e14006, Novembre 2010. (Cité en page 19.)
- [Scharf 1998] Frederick S. Scharf, Francis Juanes et Michael Sutherland. *Inferring ecological relationships from the edges of scatter diagrams : comparison of regression techniques*. Ecology, vol. 79, no. 2, pages 448–460, Mars 1998. (Cité en page 76.)
- [Schmidt 2008] Monica A. Schmidt et Eliot M. Herman. *Suppression of Soybean Oleosin Produces Micro-Oil Bodies that Aggregate into Oil Body/ER Complexes*. Molecular Plant, vol. 1, no. 6, pages 910–924, Novembre 2008. (Cité en pages 19, 80 et 99.)
- [Schmidt 2012] Travis S. Schmidt, William H. Clements et Brian S. Cade. *Estimating risks to aquatic life using quantile regression*, Juin 2012. (Cité en page 76.)
- [Schumann 1940] T. E. W. Schumann. *Theoretical aspects of the size distribution of fog particles*. Quarterly Journal of the Royal Meteorological Society, vol. 66, no. 285, pages 195–208, 1940. (Cité en page 95.)
- [Scott 2003] Ian C. Scott et Didier Y. R. Stainier. *Developmental biology : Twisting the body into shape*. Nature, vol. 425, no. 6957, pages 461–463, Octobre 2003. (Cité en page 12.)
- [Seifert 1991] Udo Seifert, Karin Berndt et Reinhard Lipowsky. *Shape transformations of vesicles : Phase diagram for spontaneous- curvature and bilayer-coupling models*. Physical Review A, vol. 44, no. 2, pages 1182–1202, Juillet 1991. (Cité en pages 15, 16 et 17.)
- [Sheetz 1974] Michael P. Sheetz et S. J. Singer. *Biological Membranes as Bilayer Couples. A Molecular Mechanism of Drug-Erythrocyte Interactions*. Proceedings of the National Academy of Sciences, vol. 71, no. 11, Janvier 1974. (Cité en page 14.)
- [Shih 2005] Frank Y. Shih et Shouxian Cheng. *Automatic seeded region growing for color image segmentation*. Image Vision Comput., vol. 23, no. 10, page 877–886, Septembre 2005. (Cité en page 30.)
- [Shimada 2008] Takashi L Shimada, Tomoo Shimada, Hideyuki Takahashi, Yoichiro Fukao et Ikuko Hara-Nishimura. *A novel role for oleosins in freezing tolerance of oilseeds in Arabidopsis thaliana*. The Plant journal : for cell and molecular biology, vol. 55, no. 5, Septembre 2008. (Cité en page 37.)
- [Shimada 2010] Takashi L Shimada et Ikuko Hara-Nishimura. *Oil-body-membrane proteins and their physiological functions in plants*. Biological & pharmaceutical bulletin, vol. 33, no. 3, 2010. (Cité en page 3.)
- [Siloto 2006] Rodrigo M. P. Siloto, Kim Findlay, Arturo Lopez-Villalobos, Edward C. Yeung, Cory L. Nykiforuk et Maurice M. Moloney. *The Accumulation of Oleosins Determines the Size of Seed Oilbodies in Arabidopsis*. The Plant Cell Online, vol. 18, no. 8, Janvier 2006. (Cité en pages 5, 19 et 80.)

- [Simons 1997] K Simons et E Ikonen. *Functional rafts in cell membranes*. Nature, vol. 387, no. 6633, Juin 1997. (Cité en page 9.)
- [Singer 1972] S J Singer et G L Nicolson. *The fluid mosaic model of the structure of cell membranes*. Science (New York, N.Y.), vol. 175, no. 4023, Février 1972. (Cité en pages 7, 8 et 9.)
- [Soille 2003] Pierre Soille. *Morphological image analysis : principles and applications*. Springer, Berlin ; New York, 2003. (Cité en pages 30 et 31.)
- [Stahelin 2003] Robert V Stahelin, Fei Long, Brian J Peter, Diana Murray, Pietro De Camilli, Harvey T McMahon et Wonhwa Cho. *Contrasting membrane interaction mechanisms of AP180 N-terminal homology (ANTH) and epsin N-terminal homology (ENTH) domains*. The Journal of biological chemistry, vol. 278, no. 31, Août 2003. (Cité en page 11.)
- [Svetina 1989] S Svetina et B Zeks. *Membrane bending energy and shape determination of phospholipid vesicles and red blood cells*. European biophysics journal : EBJ, vol. 17, no. 2, 1989. (Cité en page 15.)
- [Tierney 1994] Luke Tierney. *Markov Chains for Exploring Posterior Distributions*. The Annals of Statistics, vol. 22, no. 4, pages 1701–1728, Décembre 1994. Mathematical Reviews number (MathSciNet) MR 1329166, Zentralblatt MATH identifier 0829.62080. (Cité en page 100.)
- [Trigui 2013] Ghassen Trigui, Martine Miquel, Bertrand Dubreucq, Olivier David et Alain Trubuil. *Analysis of factors affecting the growth of oil bodies in A. Thaliana seeds : use of ordinary least squares and quantile regression*. In Proceedings of The 10th International Workshop on Computational Systems Biology, June 10-12, Tampere, Finland, pages 98–101. TICSP - Tampere International Center for Signal Processing, Finland, Juin 2013. (Cité en pages 6 et 78.)
- [Tzen 1992a] J T Tzen, G C Lie et A H Huang. *Characterization of the charged components and their topology on the surface of plant seed oil bodies*. The Journal of biological chemistry, vol. 267, no. 22, Août 1992. (Cité en pages 84 et 96.)
- [Tzen 1992b] JT Tzen et AH Huang. *Surface structure and properties of plant seed oil bodies*. The Journal of Cell Biology, vol. 117, no. 2, pages 327–335, Avril 1992. (Cité en pages 2, 3 et 4.)
- [Tzen 2012] Jason T. C. Tzen. *Integral Proteins in Plant Oil Bodies*. ISRN Botany, vol. 2012, pages 1–16, 2012. (Cité en page 2.)
- [Vance 1987] V B Vance et A H Huang. *The major protein from lipid bodies of maize. Characterization and structure based on cDNA cloning*. The Journal of biological chemistry, vol. 262, no. 23, Août 1987. (Cité en pages 2 et 19.)
- [Vanni 2000] Marco Vanni. *Approximate Population Balance Equations for Aggregation-Breakage Processes*. Journal of Colloid and Interface Science, vol. 221, no. 2, pages 143–160, Janvier 2000. (Cité en page 95.)

- [Vincent 1991] L. Vincent et P. Soille. *Watersheds in digital spaces : an efficient algorithm based on immersion simulations*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 6, pages 583–598, 1991. (Cité en page 30.)
- [Volterra 1926] Vito Volterra. *Fluctuations in the Abundance of a Species considered Mathematically*. Nature, vol. 118, no. 2972, pages 558–560, Octobre 1926. (Cité en page 89.)
- [Von Smoluchowski 1917] Marian Von Smoluchowski. *Versuch einer mathematischen Theorie der Koagulationskinetik kolloidaler Lösungen*. Zeitschrift für physikalische Chemie, vol. 92, no. 2, pages 129–168, 1917. (Cité en pages 6, 89, 90 et 95.)
- [Walther 2009] Tobias C Walther et Jr Farese Robert V. *The life of lipid droplets*. Biochimica et biophysica acta, vol. 1791, no. 6, pages 459–466, Juin 2009. PMID : 19041421. (Cité en page 18.)
- [Wang 2006] Xiaoqiang Wang, Qiang Du, Xiaoqiang Wang et Qiang Du. *Modelling and simulations of multi-component lipid membranes and open membranes via diffusive interface approaches*. 2006. (Cité en page 15.)
- [Wanner 1978] Gerhard Wanner et Roland R. Theimer. *Membranous appendices of spherosomes (oleosomes)*. Planta, vol. 140, pages 163–169, 1978. (Cité en page 3.)
- [Wilfling 2013] Florian Wilfling, Huajin Wang, Joel T Haas, Natalie Krahmer, Travis J Gould, Aki Uchida, Ji-Xin Cheng, Morven Graham, Romain Christiano, Florian Frohlich, Xinran Liu, Kimberly K Buhman, Rosalind A Coleman, Joerg Bewersdorf, Jr Farese Robert V et Tobias C Walther. *Triacylglycerol synthesis enzymes mediate lipid droplet growth by relocating from the ER to lipid droplets*. Developmental cell, vol. 24, no. 4, pages 384–399, Février 2013. PMID : 23415954. (Cité en page 18.)
- [Wunderlich 1975] F Wunderlich, A Ronai, V Speth, J Seelig et A Blume. *Thermotropic lipid clustering in tetrahymena membranes*. Biochemistry, vol. 14, no. 17, Août 1975. (Cité en page 9.)
- [Xu 1998a] Chenyang Xu et Jerry L. Prince. *Generalized Gradient Vector Flow External Forces for Active Contours*. Signal Processing, vol. 71, pages 131–139, 1998. (Cité en page 30.)
- [Xu 1998b] Chenyang Xu et Jerry L. Prince. *Snakes, Shapes, and Gradient Vector Flow*. IEEE TRANSACTIONS ON IMAGE PROCESSING, vol. 7, no. 3, page 359–369, 1998. (Cité en page 30.)
- [Yamaguchi 2007] Tomohiro Yamaguchi, Naoto Omatsu, Emi Morimoto, Hiromi Nakashima, Kanki Ueno, Tamotsu Tanaka, Kiyoshi Satouchi, Fumiko Hirose et Takashi Osumi. *CGI-58 facilitates lipolysis on lipid droplets but is not involved in the vesiculation of lipid droplets caused by hormonal stimulation*. Journal of lipid research, vol. 48, no. 5, pages 1078–1089, Mai 2007. (Cité en page 19.)

- [Yang 2012] Hongyuan Yang, Anne Galea, Vladimir Sytnyk et Merlin Crossley. *Controlling the size of lipid droplets : lipid and protein factors*. Current Opinion in Cell Biology, vol. 24, no. 4, pages 509–516, Août 2012. (Cité en page 80.)
- [Yang 2013] H. Robert Yang et Peng Li. Lipid droplets : Methods in cell biology. Academic Press, Octobre 2013. (Cité en page 28.)
- [Zanghellini 2010] J Zanghellini, F Wodlei et H H von GrÄnberg. *Phospholipid demixing and the birth of a lipid droplet*. Journal of theoretical biology, vol. 264, no. 3, pages 952–961, Juin 2010. PMID : 20184900. (Cité en page 18.)
- [Zhang 2001] P Zhang et J E Hinshaw. *Three-dimensional reconstruction of dynamin in the constricted state*. Nature cell biology, vol. 3, no. 10, Octobre 2001. (Cité en page 12.)
- [Zhong-can 1987] Ou-Yang Zhong-can et W. Helfrich. *Instability and Deformation of a Spherical Vesicle by Pressure*. Physical Review Letters, vol. 59, no. 21, pages 2486–2488, Novembre 1987. (Cité en page 16.)
- [Zhou 2012] Linkang Zhou, Li Xu, Jing Ye, De Li, Wenshan Wang, Xuanhe Li, Lizhen Wu, Hui Wang, Feifei Guan et Peng Li. *Cidea promotes hepatic steatosis by sensing dietary fatty acids*. Hepatology, vol. 56, no. 1, pages 95–107, 2012. (Cité en page 19.)
- [Zimmerberg 2000] Joshua Zimmerberg. *Are The Curves in all the Right Places*. Traffic, vol. 1, no. 4, pages 366–369, 2000. (Cité en page 11.)
- [Zimmerberg 2006] Joshua Zimmerberg et Michael M. Kozlov. *How proteins produce cellular membrane curvature*. Nature Reviews Molecular Cell Biology, vol. 7, no. 1, pages 9–19, Janvier 2006. (Cité en page 12.)

Annexes

A.1 Scripts et algorithmes

A.1.1 Implémentation R pour les procédures d'analyse de données

A.1.1.1 Script R du modèle de regression quantile

```

1  # Ordinary Least Squares model applied on OB data
2  # Loading data
3  DataOleo = read.csv("../DataOleoOB.csv")
4  DataOleo$Volume = log10(DataOleo$Volume) # Log-transformed data
5  # Setting covariates
6  DataOleo$Day = as.factor(DataOleo$Day)
7  S1 ← factor(DataOleo$Genotype == 1 | DataOleo$Genotype == 3 | DataOleo$
   Genotype == 4 | DataOleo$Genotype == 7)
8  S3 ← factor(DataOleo$Genotype == 1 | DataOleo$Genotype == 2 | DataOleo$
   Genotype == 4 | DataOleo$Genotype == 6)
9  S4 ← factor(DataOleo$Genotype == 1 | DataOleo$Genotype == 2 | DataOleo$
   Genotype == 3 | DataOleo$Genotype == 5)
10 # Ordinary Linear Regression Model on log-transformed data
11 DataOleo.lm ← lm(Volume ~ Day + (S1* S3 * S4 ), DataOleo)
12 drop1(DataOleo.lm, scope=DataOleo.lm$call, test="F")
13 summary(DataOleo.lm) # Summary of the predicted parameters
14 # Predicted values of Log10(V) sorted by Genotype and day
15 newdat ← with(DataOleo, expand.grid(Day = levels(Day ), S1= levels(S1),S3=
   levels(S3),S4= levels(S4)))
16 (newdat ← cbind(newdat, predict(DataOleo.lm, newdat, se = TRUE)))
17 colnames(newdat) ← c("Day", "S1", "S3", "S4", "fit ")
18 gen ← rep(1:8, rep(5,8))
19 gen ← as.factor(gen)
20 ord ← c("s1s3s4", "s1s3s4", "s1s3s4", "s1s3s4", "s1s3s4", "s3s4", "s3s4", "s3s4",
   "s3s4", "s3s4", "s1s4", "s1s4", "s1s4", "s1s4", "s1s4", "s4", "s4", "s4", "s4", "
   s4", "s1s3", "s1s3", "s1s3", "s1s3", "s1s3", "s1s3", "s3", "s3", "s3", "s3", "s3", "s1", "
   s1", "s1", "s1", "s1", "wt", "wt", "wt", "wt", "wt", "wt")
21 ord ← as.factor(ord)
22 newdata ← data.frame(newdat, gen)
23 newdata ← data.frame(newdata, ord)

```

Listing A.1: Implémentation sous R du modèle de regression lineaire

A.1.1.2 Script R du test de comparaison d'histogrammes

```

1  # compare distributions
2  rm(list=ls(all=TRUE))
3  setwd("../KS-test/")
4
5  # Charger les édonnes
6  DataOleo = read.csv("../DataOleoSpecial.csv")
7
8  #Test de KS

```

```

9  pvalues <- matrix(100,40,40, dimnames = list(
10 c("wt_7","wt_8","wt_9","wt_10","wt_11",
11 "s1_7","s1_8","s1_9","s1_10","s1_11",
12 "s3_7","s3_8","s3_9","s3_10","s3_11",
13 "s4_7","s4_8","s4_9","s4_10","s4_11",
14 "s1s3_7","s1s3_8","s1s3_9","s1s3_10","s1s3_11",
15 "s1s4_7","s1s4_8","s1s4_9","s1s4_10","NO DATA",
16 "s3s4_7","s3s4_8","s3s4_9","s3s4_10","s3s4_11",
17 "s1s3s4_7","s1s3s4_8","s1s3s4_9","s1s3s4_10","s1s3s4_11"),
18 c("wt_7","wt_8","wt_9","wt_10","wt_11",
19 "s1_7","s1_8","s1_9","s1_10","s1_11",
20 "s3_7","s3_8","s3_9","s3_10","s3_11",
21 "s4_7","s4_8","s4_9","s4_10","s4_11",
22 "s1s3_7","s1s3_8","s1s3_9","s1s3_10","s1s3_11",
23 "s1s4_7","s1s4_8","s1s4_9","s1s4_10","NO DATA",
24 "s3s4_7","s3s4_8","s3s4_9","s3s4_10","s3s4_11",
25 "s1s3s4_7","s1s3s4_8","s1s3s4_9","s1s3s4_10","s1s3s4_11"))
26 );
27 for (i in 1:8) {
28   for (j in 1:5) {
29     for (k in 1:8) {
30       for (l in 1:5) {
31         x <- subset(DataOleo , DataOleo$Genotype == i & DataOleo$Day == j)
32           [ ,4]
33         y <- subset(DataOleo , DataOleo$Genotype == k & DataOleo$Day == l)
34           [ ,4]
35         A <- ks.test(x,y)
36         A$p.value
37         pvalues[5*i-5+j,5*k-5+l] <- A$p.value
38       }
39     }
40   }
41   rm(i)
42   rm(j)
43   for (i in 1:40){
44     for (j in 1:40){
45       if ((pvalues[i,j] >= 0)&&(pvalues[i,j] < 0.001)){
46         pvalues[i,j] <- 10}
47     }
48   }
49   rm(i)
50   rm(j)
51   for (i in 1:40){
52     for (j in 1:40){
53       if ((pvalues[i,j] >= 0.001)&&(pvalues[i,j] < 0.01)){
54         pvalues[i,j] <- 20}
55     }
56   }
57   rm(i)
58   rm(j)
59   for (i in 1:40){
60     for (j in 1:40){
61       if ((pvalues[i,j] >= 0.01)&&(pvalues[i,j] < 0.05)){
62         pvalues[i,j] <- 30}
63     }
64   }
65   rm(i)
66   rm(j)
67   for (i in 1:40){
68     for (j in 1:40){
69       if ((pvalues[i,j] >= 0.05)&&(pvalues[i,j] < 0.1)){

```

```

69     pvalues[i,j] ← 50}
70   }
71 }
72 rm(i)
73 rm(j)
74 for (i in 1:40){
75   for (j in 1:40){
76     if ((pvalues[i,j] ≥ 0.1)&&(pvalues[i,j] ≤ 1)){
77       pvalues[i,j] ← 50}
78     }
79   }
80 png(filename="KStest")
81 library(gplots)
82
83 #Build the matrix data to look like a correlation matrix
84 x ← pvalues
85 x ← (x - min(x))/(max(x) - min(x)) #Scale the data to be between 0 and 1
86 for (i in 1:8) x[i, i] ← 1.0 #Make the diagonal all 1's
87 myCol ← c("AliceBlue", "SteelBlue1", "SteelBlue2", "SteelBlue3", "SteelBlue4"
88 )
89 #Format the data for the plot
90 xval ← ?formatC(x, format="f", digits=2)
91 pal ← colorRampPalette(c(rgb(0.96,0.96,1), rgb(0.1,0.01,0.9)), space = "
92   rgb")
93 mypal = rgb(0.1,0.01,0.9)
94 #Plot the matrix
95 x_hm ← heatmap.2(x, Rowv=FALSE, Colv=FALSE,dendrogram="none", xlab="
96   Genotypes",
97   ylab="Genotypes",key=FALSE, col=myCol, tracecol="#303030", trace="none",
98   keysize = 1.5,
99   margins=c(5, 5))
100 legend(0, 0.5,cex=0.6, fill = myCol,
101 legend = c("0 to .001", ".001 to .01", ".01 to 0.05", "0.05 to 0.1", "0.1
to 1"))
dev.off()
graphics.off()

```

Listing A.2: Implémentation sous R du modèle de regression lineaire

A.1.1.3 Scripts TCL du module AVIZO pour la segmentation d'images de corps lipidiques

```

1  # AVIZO-Script for OB image processing
2  # Trigui G, Trubuil A
3  # December 2012
4
5  # constructor is called when scro is instantiated
6  $this proc constructor {} {
7
8  remove Quantification
9
10 #define ports
11
12 #Low threshold
13 $this newPortFloatTextN thresholdL 1
14 $this thresholdL setLabel "Low threshold value"
15 $this thresholdL setValues 1
16
17 # High threshold

```

```

18 $this newPortFloatTextN thresholdH 1
19 $this thresholdH setLabel "High threshold value"
20 $this thresholdH setValues 40
21
22 # Processing button
23 $this newPortButtonList action 1
24 $this action setLabel 0 "Executer !"
25 $this action setCmd 0 {$this computeDoIt}
26 }
27
28
29 # destructor is called when scro is removed
30 $this proc destructor {} {
31
32 }
33
34 $this proc computeDoIt {} {
35
36 viewer redraw
37
38 set data [$this data source]
39
40 if {$data == ""} {return}
41
42 # Create a Visilog Quantification module
43 if {![exists Quantification]} {
44 create HxVisilog {Quantification}
45 # Quantification module could be hidden:
46 Quantification hideIcon
47 Quantification Execution setValue 0
48 # Make sure colormaps are available
49 Quantification createColormaps
50 }
51
52
53 # Median filter
54 echo Applying median filter
55 set LineCmd "cmd=medianfilter3d input=$data param=3 output=ImMed3D"
56
57 Quantification exeCommand $LineCmd
58 ImMed3D fire
59
60 # Low thresholding
61 echo Applying low thresholding
62 set threshold [$this thresholdL getValue]
63 set LineCmd "cmd=threshold input=ImMed3D param={$threshold,255} output=
    ImThreshL"
64 Quantification exeCommand $LineCmd
65 ImThreshL fire
66
67
68 # High thresholding
69 echo Applying high thresholding
70 set threshold [$this thresholdH getValue]
71 set LineCmd "cmd=threshold input=ImMed3D param=\{$threshold,255\} output=
    ImThreshH"
72 Quantification exeCommand $LineCmd
73 ImThreshH fire
74
75
76 # filtrage  $\leq 0.05$ 
77 echo Creating filter  $\leq 0.05$ 

```

```

78 Quantification exeCommand {cmd=createfilter param=(Volume3d≤0.05) outflr=
    filterVol3D}
79 echo Applying filter ≤0.05
80 Quantification exeCommand {cmd=l_filter_image input=ImThreshL input=
    ImThreshL param=filterVol3D param=adjust output=ImFitlL1 }
81 ImFitlL1 fire
82
83 # filtrage >0.05
84 echo Creating filter >0.05
85 Quantification exeCommand {cmd=createfilter param=(Volume3d>0.05) outflr=
    filterVol3D2}
86 echo Applying filter >0.05
87 Quantification exeCommand {cmd=l_filter_image input=ImThreshL input=
    ImThreshL param=filterVol3D2 param=adjust output=ImFitlL2 }
88 ImFitlL2 fire
89
90 # Reconstruct
91 echo reconstruct
92 Quantification exeCommand {cmd=reconstruct input=ImFitlL2 input=ImThreshH
    output=ImReconst }
93 ImReconst fire
94
95 # Reconstruct inverse
96 echo reconstruct inverse
97 Quantification exeCommand {cmd=logical_not input=ImReconst output=
    ImReconstInv }
98 ImReconstInv fire
99
100 # Merge reconstructed and filtered
101 echo Merging reconstructed and filtered
102 Quantification exeCommand {cmd=logical_and input=ImFitlL2 input=
    ImReconstInv output=ImMerged }
103 ImMerged fire
104
105 # Recupere detected
106 echo Recupere detected
107 Quantification exeCommand {cmd=logical_or input=ImThreshH input=ImMerged
    output=ImRecupered }
108 ImRecupered fire
109
110 # Recupere all
111 echo Recupere all binary
112 Quantification exeCommand {cmd=logical_or input=ImFitlL1 input=ImRecupered
    output=ImBinAll }
113 ImBinAll fire
114
115 # save memory
116 remove ImRecupered
117 remove ImThreshH
118 remove ImThreshL
119 remove ImMed3D
120 remove ImMerged
121 remove ImReconstInv
122 remove ImReconst
123 remove ImFitlL2
124 remove ImFitlL1
125
126
127 # Image distance
128 echo Applying Image distance
129 set LineCmd "cmd=distxxx input=ImBinAll param= 1.00 param= 1.41 param= 2
    .45 output=ImDist "

```

```

130 Quantification exeCommand $LineCmd
131 ImDist fire
132
133
134 # Matlab Module
135 echo Applying Image distance
136 create HxCalculus "CalculusMatlab"
137 # "CalculusMatlab" hideIcon
138 "CalculusMatlab" A connect "ImDist"
139 "CalculusMatlab" MatlabBuffer setState %% Matlab script example for
    HxLattice3 input.
140 if exist( ) == 1 && length(size(field.data))==3,
141 % colormap(gray);
142 Result = imextendedmax(field.data,0.2);
143 U=uint8(Result);
144 ImExtMax.data=U;
145 ImExtMax.bbox=field.bbox;
146 ImExtMax.type=field.type;
147 ImExtMax.coordsType=field.coordsType;
148 % imagesc(U(:, :, 1));
149 clear A field Result U;
150 end;
151 }
152 "CalculusMatlab" Options setValue 1 1
153 "CalculusMatlab" Options setToggleVisible 1 1
154 "CalculusMatlab" Apply snap 0 1
155 "CalculusMatlab" applyTransformToResult 1
156 "CalculusMatlab" fire
157 "CalculusMatlab" setViewerMask 65535
158 "CalculusMatlab" select
159 "CalculusMatlab" setPickable 1
160
161 set hideNewModules 0
162 "ImExtMax" setIconPosition 20 213
163 "ImExtMax" master connect "CalculusMatlab" 1
164 "ImExtMax" fire
165
166
167 # Binary
168 echo Applying binary
169 set LineCmd "cmd=convert input=ImExtMax param=binary output=ImExtMaxBin"
170 Quantification exeCommand $LineCmd
171 ImExtMaxBin fire
172
173
174 # Label on markers
175 echo Making labels on markers
176 set LineCmd "cmd=label input=ImExtMaxBin output=ImMarkers"
177 Quantification exeCommand $LineCmd
178 ImMarkers fire
179
180 # inverse of Distance
181 echo Inverse of distance image
182 set LineCmd "cmd=logical_not input=ImDist output=ImDistInv"
183 Quantification exeCommand $LineCmd
184 ImDistInv fire
185
186 # watershed
187 echo Applying watershed
188 set LineCmd "cmd=fastwatershed input=ImDistInv input=ImMarkers output=
    ImWatershed"
189 Quantification exeCommand $LineCmd

```

```

190 ImWatershed fire
191
192 # Logical sub
193 echo Applying Logical sub
194 set LineCmd "cmd=logical_sub input=ImBinAll input=ImWatershed output=
    ImObjE"
195 Quantification exeCommand $LineCmd
196 ImObjE fire
197
198 # filtrage >0.013
199 echo Filtering 10 voxels objects
200 Quantification exeCommand {cmd=createfilter param=(Volume3d>0.013) outflr=
    filterVol3D2}
201 echo Applying filter >0.013
202 Quantification exeCommand {cmd=I_filter_image input=ImObjE input=ImObjE
    param=filterVol3D2 param=adjust output=ImObj }
203 ImObj fire
204
205 # label
206 echo Applying Label
207 set LineCmd "cmd=label input=ImObj output=ImLabels"
208 Quantification exeCommand $LineCmd
209 ImLabels fire
210
211 # Border kill
212 echo Border kill
213 set LineCmd "cmd=border_kill input=ImLabels output=ImLabelsK"
214 Quantification exeCommand $LineCmd
215 ImLabelsK fire
216
217
218
219 # save memory
220 remove ImObj
221 remove ImDist
222 remove ImExtMaxBin
223 remove ImDistInv
224 remove ImLabels
225 remove ImWatershed
226 remove ImMarkers
227 remove ImExtMaxBin
228 remove ImExtMax
229 remove ImBinAll
230 remove ImObjE
231
232 # binary
233 echo binary
234 set LineCmd "cmd=convert input=ImLabelsK param=binary output=BinIm"
235 Quantification exeCommand $LineCmd
236 BinIm fire
237
238
239 # distance
240 echo Applying Image distance
241 set LineCmd "cmd=distxxx input=BinIm param= 1.00 param= 1.41 param= 2.45
    output=DistIm"
242 Quantification exeCommand $LineCmd
243 DistIm fire
244
245 # inverse of Distance
246 echo Inverse of distance image
247 set LineCmd "cmd=logical_not input=DistIm output=DistImInv"

```

```

248 Quantification exeCommand $LineCmd
249 DistImInv fire
250
251 # labels
252 echo labels
253 set LineCmd "cmd=label input=BinIm output=LabIm"
254 Quantification exeCommand $LineCmd
255 LabIm fire
256
257 # Basin catchement
258 echo Basin catchement
259 set LineCmd "cmd=catchbasin input=DistImInv input=LabIm output=BasinIm"
260 Quantification exeCommand $LineCmd
261 BasinIm fire
262
263
264 set hideNewModules 0
265 create HxGMC "Generate Surface"
266 "Generate Surface" setIconPosition 160 139
267 "Generate Surface" data connect "LabIm"
268 "Generate Surface" fire
269 "Generate Surface" smoothing setIndex 0 0
270 "Generate Surface" options setValue 0 1
271 "Generate Surface" options setToggleVisible 0 1
272 "Generate Surface" options setValue 1 0
273 "Generate Surface" options setToggleVisible 1 1
274 "Generate Surface" border setValue 0 1
275 "Generate Surface" border setToggleVisible 0 1
276 "Generate Surface" border setValue 1 0
277 "Generate Surface" border setToggleVisible 1 1
278 "Generate Surface" minEdgeLength setMinMax 0 0 0.800000011920929
279 "Generate Surface" minEdgeLength setValue 0 0
280 "Generate Surface" materialList setIndex 0 0
281 "Generate Surface" action snap 0 1
282 "Generate Surface" applyTransformToResult 1
283 "Generate Surface" fire
284 "Generate Surface" setViewerMask 65535
285 "Generate Surface" setPickable 1
286
287 set hideNewModules 0
288 [ "Generate Surface" create "LabIm.surf"
289 ] setLabel "LabIm.surf"
290 "LabIm.surf" setIconPosition 20 200
291 "LabIm.surf" master connect "Generate Surface"
292 "LabIm.surf" fire
293 "LabIm.surf" LevelOfDetail setMinMax -1 -1
294 "LabIm.surf" LevelOfDetail setButtons 1
295 "LabIm.surf" LevelOfDetail setEditButton 1
296 "LabIm.surf" LevelOfDetail setIncrement 1
297 "LabIm.surf" LevelOfDetail setValue -1
298 "LabIm.surf" LevelOfDetail setSubMinMax -1 -1
299 "LabIm.surf" setTransform 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1
300 "LabIm.surf" fire
301 "LabIm.surf" setViewerMask 65535
302
303 set hideNewModules 0
304 create HxConvexHull "Convex Hull"
305 "Convex Hull" setIconPosition 160 200
306 "Convex Hull" data connect "LabIm.surf"
307 "Convex Hull" fire
308 "Convex Hull" MaxHullPoints setMinMax 0 -2147483648 2147483648
309 "Convex Hull" MaxHullPoints setValue 0 4096

```

```

310 "Convex Hull" Epsilon setMinMax 0 -3.40282346638529e+038 3.40282346638529e
    +038
311 "Convex Hull" Epsilon setValue 0 0.00100000004749745
312 "Convex Hull" UseSkinWidth setState item 0 0 item 1 0.01
313 "Convex Hull" action snap 0 1
314 "Convex Hull" applyTransformToResult 1
315 "Convex Hull" fire
316 "Convex Hull" setViewerMask 65535
317 "Convex Hull" setPickable 1
318
319 # Get input image characteristics
320 set DimsIm [LabIm getDims]
321 set Dims0 [lindex $DimsIm 0]
322 set Dims1 [lindex $DimsIm 1]
323 set Dims2 [lindex $DimsIm 2]
324 set BboxIm [LabIm getBoundingBox]
325 set Bbox0 [lindex $BboxIm 0]
326 set Bbox1 [lindex $BboxIm 1]
327 set Bbox2 [lindex $BboxIm 2]
328 set Bbox3 [lindex $BboxIm 3]
329 set Bbox4 [lindex $BboxIm 4]
330 set Bbox5 [lindex $BboxIm 5]
331
332 set hideNewModules 0
333 create HxScanConvertSurface "Scan Surface To Volume"
334 "Scan Surface To Volume" setIconPosition 160 235
335 "Scan Surface To Volume" data connect "LabIm-convexHull"
336 "Scan Surface To Volume" fire
337 "Scan Surface To Volume" bbox setMinMax 0 -3.40282346638529e+038 3
    .40282346638529e+038
338 "Scan Surface To Volume" bbox setValue 0 $Bbox0
339 "Scan Surface To Volume" bbox setMinMax 1 -3.40282346638529e+038 3
    .40282346638529e+038
340 "Scan Surface To Volume" bbox setValue 1 $Bbox1
341 "Scan Surface To Volume" bbox setMinMax 2 -3.40282346638529e+038 3
    .40282346638529e+038
342 "Scan Surface To Volume" bbox setValue 2 $Bbox2
343 "Scan Surface To Volume" bbox setMinMax 3 -3.40282346638529e+038 3
    .40282346638529e+038
344 "Scan Surface To Volume" bbox setValue 3 $Bbox3
345 "Scan Surface To Volume" bbox setMinMax 4 -3.40282346638529e+038 3
    .40282346638529e+038
346 "Scan Surface To Volume" bbox setValue 4 $Bbox4
347 "Scan Surface To Volume" bbox setMinMax 5 -3.40282346638529e+038 3
    .40282346638529e+038
348 "Scan Surface To Volume" bbox setValue 5 $Bbox5
349 "Scan Surface To Volume" dimensions setMinMax 0 -2147483648 2147483648
350 "Scan Surface To Volume" dimensions setValue 0 $Dims0
351 "Scan Surface To Volume" dimensions setMinMax 1 -2147483648 2147483648
352 "Scan Surface To Volume" dimensions setValue 1 $Dims1
353 "Scan Surface To Volume" dimensions setMinMax 2 -2147483648 2147483648
354 "Scan Surface To Volume" dimensions setValue 2 $Dims2
355 "Scan Surface To Volume" materials setIndex 0 0
356 "Scan Surface To Volume" options setValue 0 0
357 "Scan Surface To Volume" options setToggleVisible 0 1
358 "Scan Surface To Volume" action snap 0 1
359 "Scan Surface To Volume" applyTransformToResult 1
360 "Scan Surface To Volume" fire
361 "Scan Surface To Volume" setViewerMask 65535
362 "Scan Surface To Volume" setPickable 1
363
364 # Basin catchement

```

```

365 echo Basin catchement
366 set LineCmd "cmd=mask input=BasinIm input=LabIm-convexHull.scanConverted
      output=BasinImMask"
367 Quantification exeCommand $LineCmd
368 BasinImMask fire
369
370 # Volume extraction
371 echo Volume extraction
372 set LineCmd "cmd=I_analyze input=LabIm input=LabIm param=adjust param=
      Basic outanl=anl1"
373 Quantification exeCommand $LineCmd
374
375 # cells Volume extraction
376 echo Volume extraction
377 set LineCmd "cmd=I_analyze input=BasinImMask input=BasinImMask param=
      adjust param=Basic outanl=anl2"
378 Quantification exeCommand $LineCmd
379
380 # binary
381 echo binary
382 set LineCmd "cmd=convert input=LabIm param=binary output=LabImB"
383 Quantification exeCommand $LineCmd
384 LabImB fire
385
386 # binary
387 echo binary
388 set LineCmd "cmd=convert input=BasinImMask param=binary output=
      BasinImMaskB"
389 Quantification exeCommand $LineCmd
390 BasinImMaskB fire
391
392 # save data
393 # Labels: OB labels image
394 set OutputN $data
395 set pattern ".tif"
396 regsub $pattern $OutputN "_Labels.tif" OutputN
397 LabImB save "3D Tiff" $OutputN
398
399 # save data
400 # LabelsCells: voronoi cell labels image
401 set OutputN $data
402 set pattern ".tif"
403 regsub $pattern $OutputN "_LabelsCells.tif" OutputN
404 BasinImMaskB save "3D Tiff" $OutputN
405 }

```

Listing A.3: Implémentation sous R du modèle de regression lineaire

A.1.2 Algorithme de lignes de partage des eaux par immersion

```

1 ENTREE # IMAGE DE MARQUEURS (LISTE DE MARQUEURS "Region") ET GRAPHE DE
      NIVEAUX (LISTE DES NIVEAUX "Level")
2 SORTIE # IMAGE DE LIGNE DE PARTAGE DES EAUX
3 POUR{ # "Level" de 0 à "LevelMax"}
4     POUR{ # "Region" DANS LA LISTE "RegionList"}
5         FAIRE # FAIRE CROITRE "Region" JUSQU'A L'ALTITUDE "Level"
6     FINPOUR
7     POUR{ # "Pixel" A L'ALTITUDE "Level"}
8         SI{ # "Pixel" N'EST PAS ASSOCIE A UNE REGION}
9             FAIRE # CREER UNE NOUVELLE REGION "Region" DANS LA LISTE "RegionList"
            "

```

```

10      FAIRE # AJOUTER "Pixel" A LA REGION "Region"
11      FAIRE # FAIRE CROITRE "Region" JUSQU'A L'ALTITUDE "Level"
12      FINSI
13  FINPOUR
14  FINPOUR

```

Listing A.4: Algorithme simplifié de lignes de partage des eaux par immersion

A.1.3 Algorithme de la méthode d'extraction de l'enveloppe convexe

```

1  ENTREE # ENSEMBLE DE POINTS DANS L'ESPACE
2  SORTIE # POLYGONE D'ENVELOPPE CONVEXE
3  FAIRE # CREER UN SIMPLEX d+1 POINTS
4  POUR{ # POUR CHAQUE FACETTE F}
5      POUR{ # CHAQUE POINT p NON AFFECTE}
6          SI{ # p EST AU DESSUS DE F }
7              FAIRE # AFFECTER p A L'ENSEMBLE DES POINTS EXTERIEURS A F
8              FINSI
9      FINPOUR
10 FINPOUR
11 POUR{ # POUR CHAQUE FACETTE F AVEC UN ENSEMBLE DE POINTS EXTERIEURS NON
      NULS}
12 FAIRE # SELECTIONNER p EXTERIEUR A F LE PLUS PROCHE
13 FAIRE # INITIALISER UN ENSEMBLE V A F
14 FINPOUR
15 POUR{ # POUR TOUT LES VOISINS N AU-DESSUS DES FACETTES EN V}
16 FAIRE # AJOUTER N A V
17 FAIRE # LES BORDS DE V EST L'ENSEMBLE D'ARRETES H
18     POUR{ # POUR CHAQUE ARRETE R DE H}
19         FAIRE # CREER UNE NOUVELLE FACETTE FORMEE DE R ET p
20         FAIRE # LIER LA NOUVELLE FACETTE A SES VOISINS
21     FINPOUR
22 FINPOUR
23 POUR{ # POUR CHAQUE NOUVELLE FACETTE F'}
24     POUR{ # POUR CHAQUE POINT q NON AFFECTE A L'ENSEMBLE DE POINTS
        EXTERIEURES A F'}
25         FAIRE # AFFECTER q A L'ENSEMBLE DES POINTS EXTERIEURES A F'
26     FINPOUR
27 FINPOUR

```

Listing A.5: Algorithme de la méthode d'extraction de l'enveloppe convexe (QUICKHULL)

A.2 Régularité des solution de l'équation de Smoluchowski

Theorem 1. [Ball 1990] (Continuité des solutions) Soit n une solution, conservant la taille, du problème 4.7 sur $[0, T)$, alors $n : [0, T) \rightarrow X^+$ est continue et les séries $\sum_{i=1} in_i(t)$ sont uniformément convergentes sur les intervalles compacts de $[0, T)$.

Theorem 2. [Ball 1990] (Différentiabilité des solutions) Soit n une solution, conservant la taille, du problème 4.7 sur $[0, T)$. Si on suppose en outre que $K_{i,j} \leq C.(i + j)$, $F_{i,j} \leq C.ij$ alors n_i est continûment différentiable sur $[0, T)$.

A.3 Figures

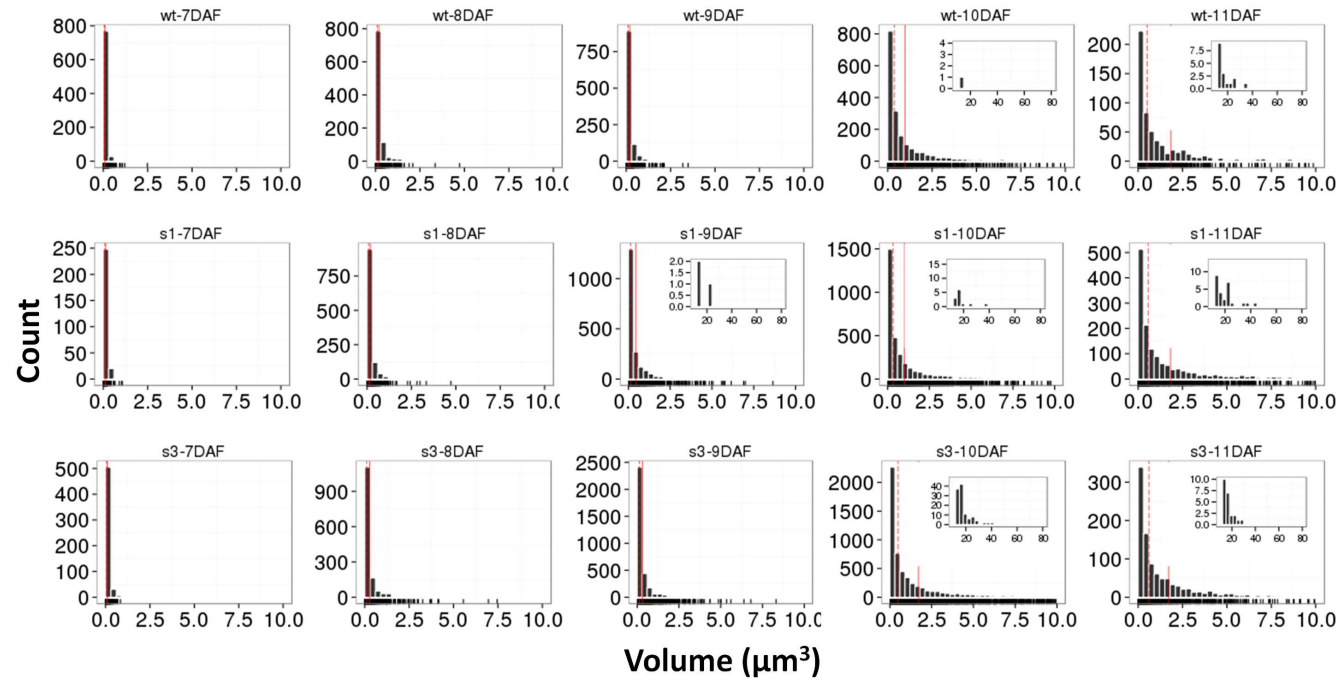


FIGURE A.1 : Histogrammes de volume de corps lipidiques pour chaque génotype et à chaque du stade du développement. L'histogramme principal est tracé pour les volumes $\leq 10 \mu\text{m}^3$ et avec un pas de $0.3 \mu\text{m}^3$. L'histogramme joint concerne les volumes $> 10 \mu\text{m}^3$ et avec un pas de $3 \mu\text{m}^3$. La moyenne et la médiane sont représentées avec des lignes verticales respectivement en ligne continue et discontinue.

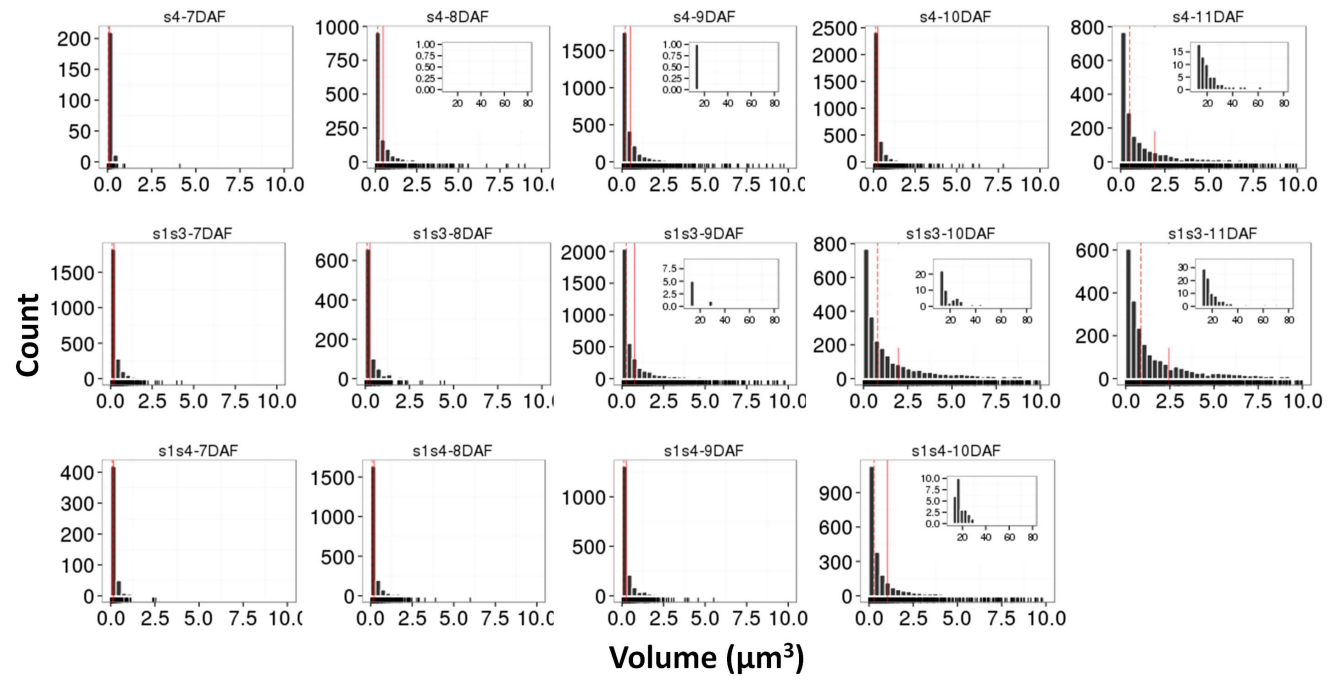
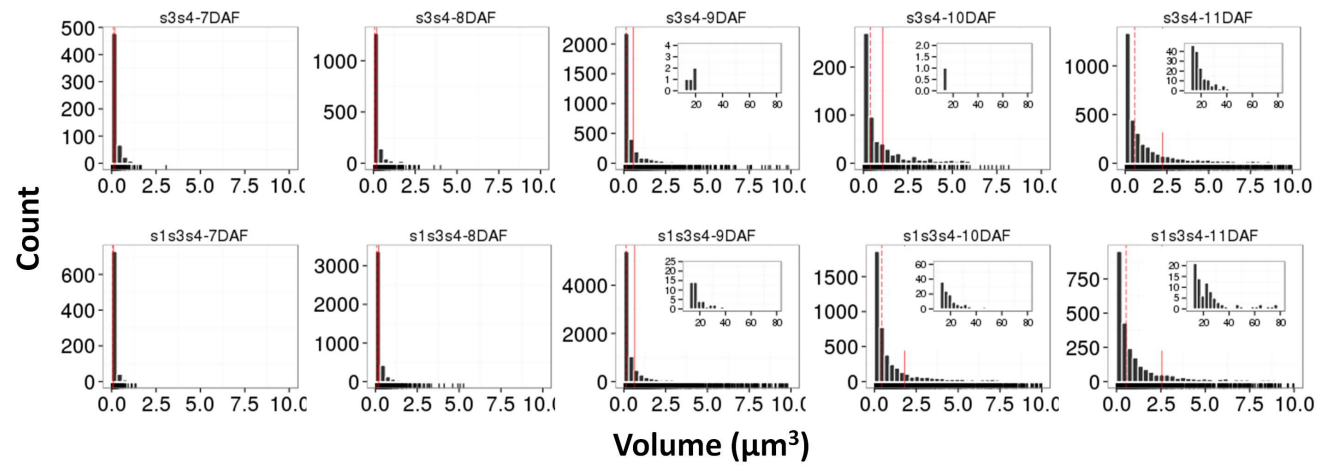


FIGURE A.1 : Suite figure A.1

FIGURE A.1 : Suite figure [A.1](#)

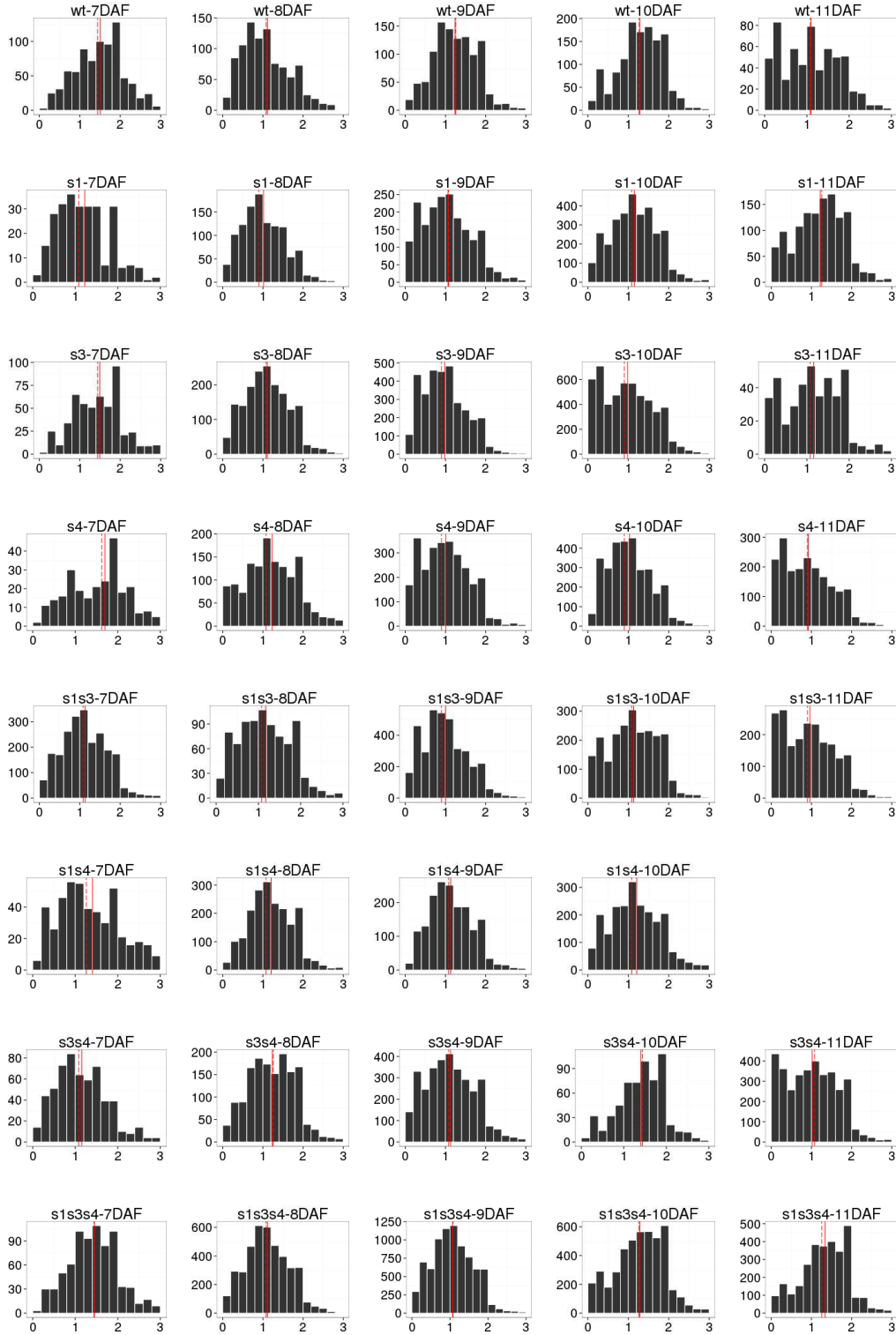


FIGURE A.2 : Histogrammes des distances des corps lipidiques au plus proche voisin selon le génotype et le temps. En abscisses, la distance (en μm^3) et en ordonnées l'effectif.

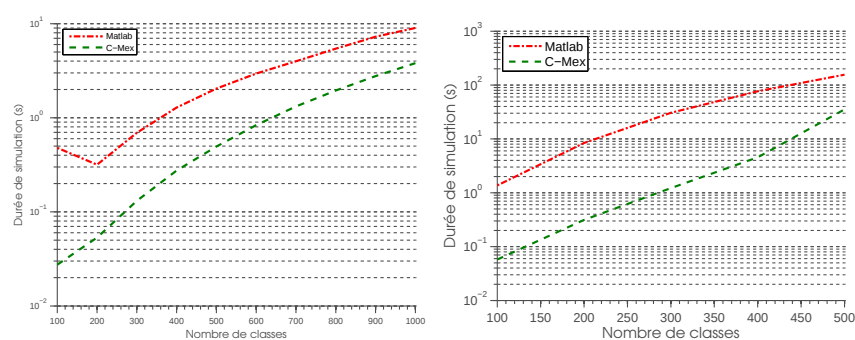


FIGURE A.3 : Comparaison entre la durée de simulation du modèle de Smoluchowski utilisant des fonctions Matlab et celle utilisant C-Mex respectivement avec un noyau constant (figure à gauche) et un noyau multiplicatif (figure à droite).

A. La déconvolution

La déconvolution est une technique utilisée pour corriger les distorsions optiques (diffraction, aberration des lentilles, ...) liés à la limitation du système optique. En microscopie, l'image d'un point n'est pas un point mais plutôt une tâche, qui définira la résolution du microscope. Connaissant la réponse d'un système à une source ponctuelle (un point) permet de définir une fonction mathématique appelée PSF (Point Spread Function). Cette fonction va donc être le reflet des déformations optiques engendrées par l'équipement et/ou le spécimen. Ce phénomène est plus présent en microscopie conventionnelle, où toutes les structures marquées par fluorescence d'un échantillon émettent de la lumière lorsqu'elles sont excitées par une lumière de longueur d'onde appropriée, qu'elles soient ou non dans le plan de mise au point. En microscopie confocale, une grande partie des photons provenant des plans défocalisés sont éliminés grâce à un dispositif appelé pinhole. La résolution en profondeur d'un microscope confocal est donc meilleure que celle d'un microscope conventionnel.

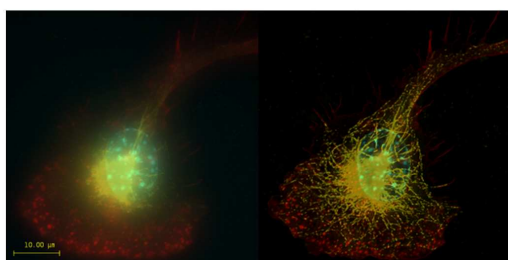


Figure A : Microscopie à champ large du macrophage dont les éléments tubulin, actine et le noyau sont respectivement marqués en jaune, rouge et bleu. A gauche : l'image originale. A droite la même image après déconvolution (Source : Extraite de la galerie photo du site : <http://www.svi.nl>)

B. FRAP (Fluorescence Recovery After Photobleaching):

Le FRAP est une méthode utilisée en microscopie à fluorescence afin de mesurer la vitesse de diffusion moléculaire.

L'excitation intense de courte durée par un laser d'une petite région de l'échantillon (quelques micromètres de diamètre) présentant une population des molécules fluorescente produit une réduction de l'intensité de fluorescence. Cette procédure s'appelle photoblanchiment. Suite à cette excitation, deux populations de molécules distinctes sont ainsi créées : une qui a gardé sa fluorescence, et l'autre qui l'a perdue.

Le retour à l'état initial (récupération de la fluorescence) est assuré par le déplacement latéral de molécules fluorescentes du milieu adjacent pour homogénéiser la fluorescence dans la région excitée. Le suivi du temps de retour à la fluorescence initiale permet de déterminer les paramètres de mobilité des molécules étudiées, notamment le coefficient de diffusion latérale.

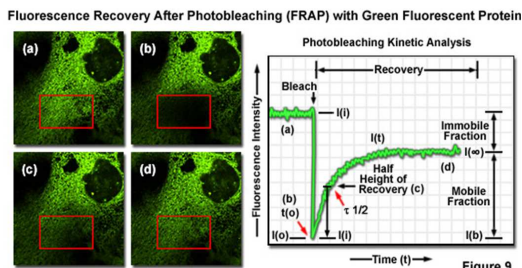


Figure B : Imagerie FRAP de protéines marquées à la GFP. Après photoblanchiment d'une zone spécifique (a), la fluorescence disparaît (b). Le retour progressif par diffusion de protéines fluorescentes pour couvrir la zone photoblanche (c,d) peut être mesuré par le suivi de l'évolution de la fluorescence en fonction du temps l'intensité en l'intensité (figure à droite)

(Source : Extraite du site : <http://zeisscampus.magnet.fsu.edu/articles/livecellimaging/techniques.html>)

C. FRET (Förster Resonance Energy Transfer)

Le FRET est utilisé en imagerie cellulaire pour rechercher des interactions moléculaires.

Cette technique permet de prédire la distance entre deux molécules portant chacune un groupe fluorophore, et de conclure ainsi que les deux molécules sont en interaction ou pas. La technique repose sur le transfert de fluorescence de l'émetteur à l'accepteur. L'émetteur (donneur), étant excité, transmet de l'énergie vers l'accepteur. Pour que ce transfert puisse avoir lieu, il faut que les deux molécules soient à une distance inférieure à 100Å. Il se peut que les deux molécules soient distantes, mais interagissent sans que le FRET ait lieu, ce qui représente la limite de cette technique.

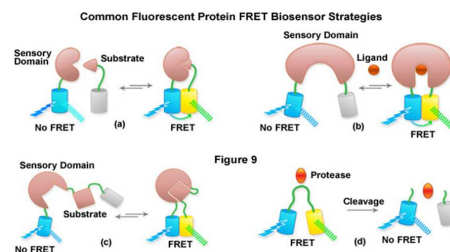


FIGURE A.4 : Méthodes d'imagerie en microscopie confocale

Principe d'aspiration par micropipette

L'aspiration par micropipette est une méthode qui permet de mesurer les propriétés mécaniques des membranes biologiques (1,2 et 3)

Comment les vésicules, tel que les globules rouges, peuvent circuler à travers les petits vaisseaux sanguin, comment les éléments de la surface d'une membrane peuvent influencer son comportement mécanique et comment les stimuli mécaniques et chimiques pourraient modifier leur réponse.

La surface d'une vésicule est aspirée dans un petit tube en verre (Figure 1). Au même temps, le bord entrant dans le tube est suivi par microscopie. Le diamètre d'une micropipette varie entre 1 et 10 μm^3 , avec un bord qui entre dans le tube d'une dizaine de nanomètres.

Le protocole expérimental comprend une vésicule suspendue dans une solution saline et dont la membrane externe est mise en contact avec l'extrémité du tube de la micropipette. L'autre extrémité de la micropipette est reliée à un tuyau plongée dans un réservoir rempli de fluide dont la hauteur est variable. La pression appliquée à la vésicule est de l'ordre de 0.1-0.2 pN / μm^2 réglable par la hauteur du fluide dans le réservoir. Connaissant la pression appliquée, nous pouvons déterminer la tension de surface de la région de la membrane aspirée.

Lors de l'aspiration par micropipette, on distingue principalement 3 états : un état initial (figure 2A) , un état critique (figure 2B) et un état final (figure 2C).

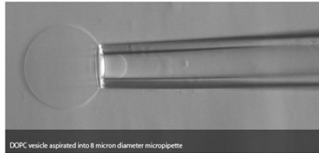


Figure 1 : Principe d'aspiration par micropipette

Considérons les paramètres reportés dans la figure 3, où est la pression atmosphérique, P_{cell} est la pression à l'intérieur de la vésicule, R_{cell} est le rayon de la vésicule à l'extérieur de la micropipette, R_p le rayon de la micropipette, et r le rayon de courbure du bord entrant dans la micropipette.

Nous nous plaçons dans la situation où $l/R_p = 1$. Dans ce cas, en utilisant la loi de Laplace, on déduit que :

$$P_{cell} - P_{atm} = \frac{2N}{R_{cell}}$$

Où N est la tension de surface de la vésicule. Pour la portion de la vésicule à l'intérieur de la micropipette, on a :

$$P_{cell} - P_{pip} = \frac{2N}{r}$$

Et on aboutit à :

$$P_{atm} - P_{pip} = \Delta p = 2N \left(\frac{1}{r} - \frac{1}{R_{cell}} \right)$$

$$\Delta p = 2N \left(\frac{1}{R_p} - \frac{1}{R_{cell}} \right)$$

Qui relie la différence de pression entre l'intérieur et l'extérieur de la micropipette avec le rayon de la vésicule à l'intérieur et à l'extérieur de la micropipette pour une tension de surface donnée

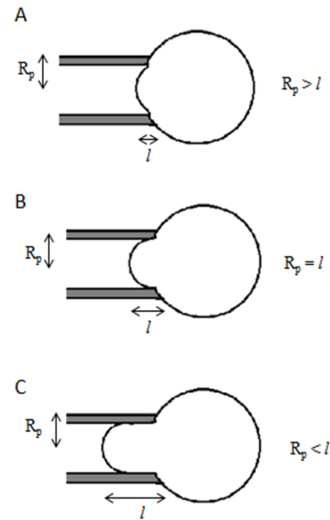


Figure 2 : Différents états possibles d'aspiration d'une vésicule par micropipette

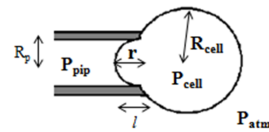


Figure 3 : Paramètres de pression extérieure, à l'intérieur de la vésicule, et à l'intérieur du tube .

1. Guo Q, Park S, Ma H: **Microfluidic micropipette aspiration for measuring the deformability of single cells.** *Lab Chip* 2012, **12**:2687–2695.
2. Kee Y-S, Robinson DN: **Micropipette Aspiration for Studying Cellular Mechanosensory Responses and Mechanics.** In *Dictyostelium discoideum Protocols*. edited by c L, Rivero F Totowa, NJ: Humana Press; 2013, **983**:367–382.
3. Hochmuth RM: **Micropipette aspiration of living cells.** *J Biomech* 2000, **33**:15–22.

FIGURE A.5 : Thechnique d'aspiration par micropipette

A.4 Article de conférence

Publié dans : Proceedings of World Academy of Science, Engineering and Technology, International Conference on Mathematical Biology, April 25-26 ; 2012, Paris, issue 64 p 763-768.

The dynamics of oil bodies in *A. thaliana* seeds : A mathematical model of biogenesis and coalescence

G. Trigui, B. Laroche, M. Miquel, B. Dubreucq, A. Trubuil

Abstract—The subcellular organelles called oil bodies (OBs) are lipid-filled quasi-spherical droplets produced from the endoplasmic reticulum (ER) and then released into the cytoplasm during seed development. It is believed that an OB grows by coalescence with other OBs and that its stability depends on the composition of oleosins, major proteins inserted in the hemi membrane that covers OBs. In this study, we measured the OB-volume distribution from different genotypes of *A. thaliana* after 7, 8, 9, 10 and 11 days of seed development. In order to test the hypothesis of OBs dynamics, we developed a simple mathematical model using non-linear differential equations inspired from the theory of coagulation. The model describes the evolution of OB-volume distribution during the first steps of seed development by taking into consideration the production of OBs, the increase of triacylglycerol volume to be stored, and the growth by coalescence of OBs. Fitted parameters values show an increase in the OB production and coalescence rates in *A. thaliana* oleosin mutants compared to wild type.

Keywords—Biogenesis, coalescence, oil body, oleosin, population dynamics.

I. INTRODUCTION

IN seeds, storage lipids are deposited in stable subcellular structures named oil bodies (OBs). These organelles have a quasi-spherical shape with a diameter of 0.2 to 3 μm and are constituted of 95 % neutral lipids mostly in the form of triacylglycerols (TAGs) [1], [2], [3] surrounded by a monolayer of phospholipids (PLs), in which different proteins, mainly so-called oleosins, are embedded [4], [5]. Mechanisms for OBs biogenesis and dynamics are not yet fully understood. OBs differ from one species to another and between kingdoms, particularly by their composition in neutral lipids and protein complements [2], [6]. However, it is largely admitted that TAGs are synthesized by specific enzymes in specialized sub-domains, between the membrane leaflets of the ER, and that OBs are then individualized by bending from the ER before being released in the cytoplasm [6], [7]. Two alternative hypotheses about how and when

oleosins are targeted to the surface of OBs are proposed. In the first one, there would be a time lag between the lipid accumulation and oleosin insertion, causing the oleosins to insert only into already-formed OBs [8], [9], [10], [11]. In the second hypothesis, oleosins would be synthesized within the ER and then would diffuse to cover the surface of a nascent OB concomitantly with the accumulation of TAGs [2], [12], [13], [14]. Van Rooijen et al. [15], when using oleosins lacking their central hydrophobic domain, observed a reduction in targeting to OBs, suggesting that an oleosin is transported to the OB via its central hydrophobic domain, which would act as a targeting signal. In the second hypothesis [16], during the first steps of seed reserves accumulation, less oleosins would accumulate compared to TAGs, and nascent OBs would not be entirely covered with oleosins (Fig. 1). These OBs would then coalesce to form larger OBs totally covered with oleosins exhibiting a decreased surface to volume ratio. The aggregation process would decrease progressively upon time and the OB size would stabilize. Oleosins are proteins of about 15 to 26 kDa , believed to play a central role in OB stability. Many isoforms of oleosin proteins have been identified [17], suggesting other functions for oleosins, in particular in membrane bending processes, when elastic properties of biological membrane may change due to the curvature induced by proteins insertion [18], [19]. We aim with the model developed here to give some insights about the growth of OBs by coalescence. In particular, we wish to analyze how the production and coalescence rates depend on the composition in oleosins, since these parameters may change throughout seed development. This paper is organized as follows: In Section II, we expose the different materials and methods used to acquire, treat and exploit data. In Section III two versions of the model of biogenesis and coalescence, respectively the surface density-non-dependent (SDND) model and the surface density-dependent (SDD) model are presented. In section IV we present numerical methods and simulations which allow us to fit the model predictions to a set of experimental data. We also interpret the simulation results. Discussion and conclusion are presented in section V.

II. MATERIALS AND METHODS

A. Biological material

Arabidopsis thaliana wild type and oleosin mutant plants were grown from seeds on soil in a greenhouse. Upon flowering, flowers and subsequent developing siliques were tagged

G. Trigui is with the Institut National de la Recherche Agronomique, Unité Mathématiques et Informatique Appliquées, Domaine de Vilvert, 78352 Jouy-en-Josas, France. (phone: +33(0)134652219 ; email: Ghassen.Trigui@jouy.inra.fr)

B. Laroche and A. Trubuil are with the Institut National de la Recherche Agronomique, Unité Mathématiques et Informatique Appliquées, Domaine de Vilvert, 78352 Jouy-en-Josas, France. (emails: Beatrice.Laroche@jouy.inra.fr ; Alain.Trubuil@jouy.inra.fr)

M. Miquel and B. Dubreucq are with the Institut National de la Recherche Agronomique, Institut Jean-Pierre Bourgin, Centre de Versailles-Grignon, Route de Saint-Cyr, 78026 Versailles, France. (emails: Martine.Miquel@versailles.inra.fr ; Bertrand.Dubreucq@versailles.inra.fr)

daily until 12 days after flower opening. Siliques for each of development were sampled and dissected to remove developing seeds. Seeds were spread on a glass slide, incubated with Nile Red, a neutral lipid stain, at $2\mu\text{g}/\text{ml}$ in a 60 % glycerol solution. Embryos were removed from the seed teguments by gently pressing seeds between slide and lamella and observed after 30 min of incubation in the dark. Arabidopsis oleosin null mutants are available for 3 oleosins, OLE1 (At4g25140), OLE2 (At5g40420) and OLE4 (At3g01570) (ref. NASC). Double mutants (*ole1ole2*, *ole1ole4*, *ole2ole4*) and a triple mutant (*ole1ole2ole4*) have been generated in the laboratory.

B. Data acquisition and measurements

3D images of dissected *Arabidopsis* embryos were acquired using a LEICA SP2(AOBS) confocal microscope with a spatial resolution of $(0.09\mu\text{m} \times 0.09\mu\text{m} \times 0.16\mu\text{m})$ in the (x, y, z) referential. Third dimension was obtained by scanning the sample through the z axis, providing a sequence of 2D images corresponding to the fluorescence emitted from the focal plane. The microscope is equipped with a $40\times$ magnification oil-immersion lens with a numerical aperture of 1.25. The fluorescent stain (Nile Red) is excited with a 488nm laser and the emitted light is detected using a photomultiplier between 550 and 650nm . These settings are appropriate for the excitation of the stain used to label neutral lipids constituting OBs. The information obtained from one 3D-stack corresponds to the fluorescence emitted from neutral lipids in a region of the sample corresponding to one genotype after some number of days of development. We relied on Avizo®Fire (Burlington, USA), a 3D image processing software, to extract volumes of individualized OBs. A watershed segmentation method [20] was used to this end. A histogram of volumes for every stack was obtained. The total OB-volume in a stack was obtained by summing on all individual OB-volumes after normalizing to a referential stack size of $(300 \times 300 \times 300)$ voxels. This corresponds approximately to a $3.5 \cdot 10^4 \mu\text{m}^3$ volume obtained by multiplying the stack size by the spatial resolution.

III. THE MODEL OF BIOGENESIS AND COALESCENCE

A. The SDND-model

1) *Description*: Define $n_c(v, t)$ as the population size of OBs of volume v contained in the cytoplasm at time t . The evolution of the OB-volume distribution in the cytoplasm is modelled by the following non-linear discrete differential equation:

$$\frac{\partial n_c(v, t)}{\partial t} = \mu \delta_{v, v_m} + \frac{1}{2} \sum_{v'=1}^{v-1} \alpha n_c(v', t) n_c(v-v', t) - n_c(v, t) \sum_{v'=1}^{\infty} \alpha n_c(v', t) \quad (1)$$

The initial conditions are : $n_c(v=0, t) = 0$, and $n_c(v, t=0) = 0$.

The right hand side of Eq. 1 comprises three components of the OBs growth process. The first term assumes that the

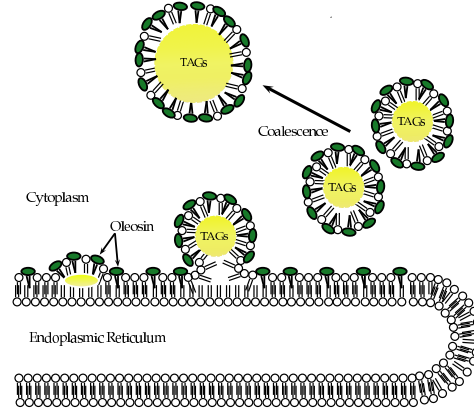


Fig. 1: Biogenesis of OBs by budding from the two leaflets of the ER. OBs that do not contain enough oleosins to coat their entire surface coalesce, forming larger OBs. Coalescence of OBs stops when the surface is completely covered with oleosins.

OB-volume distribution varies as a result of a stream of newly born OBs with a minimum volume $v = v_m$ which are detached from the ER to join the cytoplasm with a rate μ . This process is modeled by a simple kronecker delta function. The second term supposes that, once in the cytoplasm, OBs of volume v' coalesce with a rate α with other OBs of volume $v - v'$ to form OBs of volume v . The third term reflects the OBs leaving the volume class v when coalescing with other OBs volumes. Globally, the OB-volume distribution reshapes to larger unbounded volume classes when advancing in time.

2) *Volume preservation law*: From a physical point of view, it is important to note that the total volume produced by the ER should be preserved since nothing is lost throughout the coalescence process. So the condition:

$$\partial_t \sum_{v=v_m}^{\infty} v n_c(v, t) = \mu v_m \quad \forall t, \quad (2)$$

should be verified. Performing the calculation from Eq. 1 (Eq. 8 in appendix A) we indeed obtain the condition in Eq 2. But this reasoning is true only if the series $\sum_v \sum_{v'} v \alpha n_c(v', t) n_c(v-v', t)$ converges. However, it is not always the case and we can get cases where the total volume is not conserved. This phenomena is called gelation and has been studied in several literatures [21], [22]. Gelation may be interpreted as the appearance of infinite mass clusters (e.g. molecules, drops, etc.) in a finite time.

B. The SDD-model

1) *Description*: In the second version of the model, we considered the fact that coalescence process decreases the surface to volume ratio which implies that coalescing OBs with an initial density of proteins on the surface will reach a

maximum protein density after a finite number of coalescence events. Considering q the number of protein packages corresponding to a number of molecules of proteins associated in a package, we called "protein package" (pp) the unit of q . We made the coalescence rate α dependent on the protein density on the surface of coalescing OBs (which is adequate with the biological hypothesis described previously). Considering $n_c(v, q, t)$ as the population size of OBs of volume v and having q protein packages contained in the cytoplasm at time t . Generalizing Eq.1, the evolution of the OB-volume distribution in the cytoplasm is modelled by the following non-linear differential equation:

$$\begin{aligned} \frac{\partial n_c(v, q, t)}{\partial t} = & \mu \delta_{v_m, q_m} + \\ & \frac{1}{2} \sum_{v'=1}^{v-1} \sum_{q'=1}^{q-1} \varepsilon \alpha(v', v-v'; q', q-q') n_c(v', q', t) n_c(v-v', q-q', t) \\ & - n_c(v, q, t) \sum_{v'=1}^{\infty} \sum_{q'=1}^{\infty} \varepsilon \alpha(v, v'; q, q') n_c(v', q', t) \end{aligned} \quad (3)$$

The initial conditions are: $n_c(v = 0, q, t) = 0$, $n_c(v, q = 0, t) = 0$ and $n_c(v, q, t = 0) = 0$.

The coalescence rate of OBs of volumes respectively v' and $v-v'$ and proteins coverage q' and $q-q'$ to get OBs of volume v and coverage q is taken to be:

$$\alpha(v', v-v'; q', q-q') = \frac{(v'(v-v'))^{\frac{2}{3}(1+\beta)}}{(q'(q-q'))^{\beta}} \quad (4)$$

α is taken to be dependent on OBs volume and protein density such that α increases when the volume of coalescing OBs increases and decreases with the decrease of their coverage density $d = q/v^{2/3}$, where the surface of an OB is $v^{2/3}$ up to an irrelevant constant factor. The value of β is set to 3 in order to give more weight for the density dependence in the value of α . The value of ε in this case is characteristic of the composition in proteins, which can be dependent on the genotype. By this approach, OBs with initial uniform density will lead to a higher surface density when coalescing. The coalescence rate decreases to zero when the protein density on the surface increases to a maximum value. So, very large OBs will not be produced.

2) *Description of SDD-model parameters:* For a given population at early stage of seed development, the OB-volume distribution is believed to be monomodal with OB-volume less than $1\mu m^3$. Very little is known about the dynamics of OBs production within ER. It is believed that very small OBs (less than $100nm$) are produced. Besides, very small OBs, in the limit of the microscope resolution, are also observed throughout the five days of seed development starting at day 7. For these reasons, we make the simple assumption that OBs of minimum volume v_m and coverage q_m are continuously produced. The production rate μ is obtained by a regression on total experimental OBs volume $V(t)$ observed at 7, 8, 9, 10, and 11 days after development.

IV. NUMERICAL SIMULATIONS

A. Discretization methods

We solved Equation 1 using a simple linear grid with volume classes ranging between $v_m = 1\mu m^3$ and $v_M = 20\mu m^3$ separated by $1\mu m^3$ steps. In fact, every coalescence occurring between two OBs of volumes v and $v + \delta v$ gives an OB of volume $2v + \delta v$ (Fig. 2a). The number of differential equations to be solved is equal to the number of volume classes $v \in [1, 2, \dots, v_M]$. The equations are of the form:

$$\frac{\partial n_c(v, t)}{\partial t} = f_v(n_c(1, t), \dots, n_c(v_M, t)) \quad (5)$$

We used the same discretization method scheme to solve Equation 3 but by adding a linear jump from a coverage quantity class to another in every coalescence occurring between two OBs of volumes respectively v and $v + \delta v$ and coverage quantities respectively q and $q + \delta q$ to give an OB of volume $2v + \delta v$ and a coverage quantity $2q + \delta q$ (Fig. 2b). The coverage quantity classes range between $q_m = 1pp$ and $q_M = 10pp$. So the equations to be solved are of the form:

$$\begin{aligned} \frac{\partial n_c(v, d, t)}{\partial t} = & f_{v,d}(n_c(1, 1, t), \dots, n_c(v_M, 1, t), n_c(1, 2, t), \\ & \dots, n_c(v_M, 2, t), \dots, n_c(v_M, q_M, t)) \end{aligned} \quad (6)$$

We solve these equations numerically using the 4th order Runge-Kutta method within the solver in Matlab 7.0 (Natick, USA).

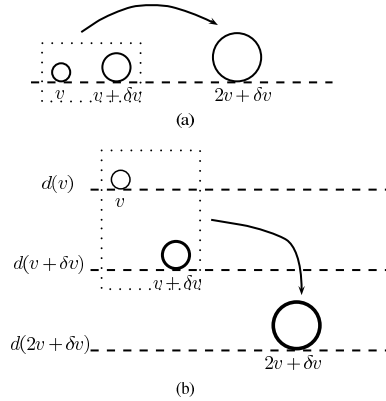


Fig. 2: Discretization methods used to discretize volume classes in the SDND model (a), and volume and coverage quantity classes in the SDD model (b).

B. Simulation test for the volume conservation problem

To test the validity of the model on respecting the volume conservation law and within the simple discretization used for

the model, we simulated Eq. 1 of the SDND model with a production rate of 10 OBs (of volume $1\mu m^3$) per 0.1 days time-step. We examined the total cytoplasm volume of OBs expressed as $V_c(t) = \sum_{v=1}^{v_{max}} v n_c(v, t)$ as a function of time, ranging from 0 to 5 days, and for different values of α (Fig. 3). In the case of $\alpha = 0.01$, the cytoplasm volume in every time step still equal to the total produced volume, whereas for increasing values of α , an effective loose in the cytoplasmic volume is observed.

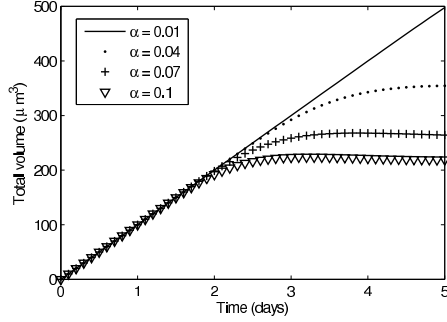


Fig. 3: The total cytoplasm volume in function of time for different values of α

C. Data fitting with optimized parameters

We used the minimization of the sum-of-square function (ss) which computes the deviation between the numerical simulations and the experimental results to optimize parameters $X = \{\mu, \alpha\}$ or $X = \{\mu, \varepsilon\}$ so that they can describe, as closely as possible, the experimental results. The ss function is defined as :

$$ss = \sum_{\tau=1}^5 \sum_{v=1}^{v_M} \sum_{q=1}^{q_M} \frac{[g(v, \tau) - \hat{g}(v, d, \tau; X)]^2}{5}, \quad (7)$$

Here $g(v, \tau) = n(v, \tau)/N(\tau)$ is the experimental normalized OB-volume distribution at $\tau = (1, 2, 3, 4, 5)$ corresponding to the observations after 7, 8, 9, 10 and 11 days of seed development respectively, and $N(\tau)$ is the total number of OBs at time τ . We used a Markov-Chain-Monte-Carlo method from a Matlab toolbox [23] to find the minimum of Equation 7. The optimized parameters of the model are given in Table I. The coalescence fitting parameters α and ε are increased for oleosin mutants compared to wild type (Tab I). More oleosins are suppressed (*ole1ole2ole4*) more OBs coalesce. This result is coherent with the hypothesis suggesting the function of oleosin proteins as a surfactant. Likewise, optimized μ shows an increase of OBs production in mutant. The fitting efficiency does not change significantly between the two versions of models. However, the advantage of the proposed SDD model is that the coalescence rate is controlled in contrast to the SDND model where for increased values of constant coalescence rate, infinite volumes may be obtained in a finite time.

TABLE I: Estimated production rate (μ), coalescence rate (α) given by the SDND model, and coalescence index (ε) for wild type and different mutants of *A.thaliana*. The values are expressed as (the mean $\pm$ the standard deviation) of the estimated parameters obtained by experimental data fitting when minimizing the ss-error.

	Wild type	<i>ole1ole4</i>	<i>ole1ole2ole4</i>
μ	218	234	273
SDND model			
α ($\cdot 10^{-3}$)	1.02 ± 0.56	1.80 ± 0.33	2.081 ± 0.278
SDD model			
ε ($\cdot 10^{-3}$)	1.2 ± 0.25	1.52 ± 0.25	2.13 ± 0.3

TABLE II: Settings used to run the SDD model.

SDD-model parameter setting	value
Initial time	0 Days
Final time	5 Days
Time step	Determined by the solver
v_m	$1 \mu m^3$
v_M	$20 \mu m^3$
v step	$1 \mu m^3$
q_m	1 pp
q_M	10 pp
q step	1 pp

We simulated Eq. 3 of the SDD model with the settings given in Tab II. The simulations fitted on the experimental data are presented in Fig. 4.

V. CONCLUSION

Our model therefore still requires improvement. First, to enhance OB-volume discretized classes, particularly for volumes $< 1\mu m^3$. Second, to take into consideration the dynamics of OB fragmentation occurring from 11th day of OB development. Third, to integrate more experimental data.

APPENDIX A

ANALYTICAL VERIFICATION OF THE TOTAL CYTOPLASM VOLUME CONSERVATION

The total volume of OBs in the cytoplasm should depend only on the produced volume during the coalescence process. From Eq. 1 and 2 :

$$\begin{aligned} \partial_t \sum_{v=v_m}^{\infty} v n_c(v, t) - \mu v_m = \\ \frac{1}{2} \sum_{v=1}^{\infty} \sum_{v'=1}^{v-1} v \alpha n_c(v', t) n_c(v - v', t) \\ - \sum_{v=1}^{\infty} \sum_{v'=1}^{\infty} v \alpha n_c(v', t) n_c(v, t) \end{aligned}$$

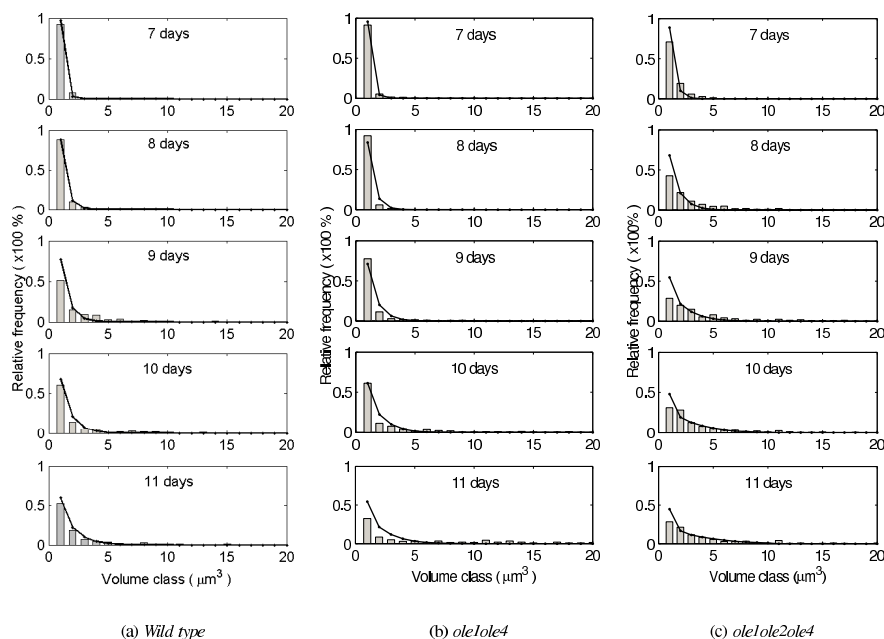


Fig. 4: OB-volume distribution at 7, 8, 9, 10 and 11 days of development for wild type and different mutants. Bars plot represent experimental data, and solid lines represent the SDD model simulations.

$$\begin{aligned}
 &= \frac{1}{2} \sum_{v'=1}^{\infty} \sum_{v=v'+1}^{\infty} v \alpha n_c(v', t) n_c(v - v', t) \\
 &\quad - \sum_{v=1}^{\infty} \sum_{v'=1}^{\infty} v \alpha n_c(v', t) n_c(v, t) \\
 &= \frac{1}{2} \sum_{v'=1}^{\infty} \sum_{\tilde{v}=1}^{\infty} (\tilde{v} + v') \alpha n_c(v', t) n_c(\tilde{v}, t) \\
 &\quad - \sum_{v=1}^{\infty} \sum_{v'=1}^{\infty} v \alpha n_c(v', t) n_c(v, t) \quad (8) \\
 &= 0
 \end{aligned}$$

ACKNOWLEDGMENT

The authors wish to acknowledge the help of Olivier Martin, Christophe Deroulers and Silvio Franz from université Paris XI. This work has been funded by the Institut National de la Recherche Agronomique.

REFERENCES

- [1] A. H. Huang, "Oleosins and oil bodies in seeds and other organs," *Plant Physiology*, vol. 110, pp. 1055–1061, Apr. 1996. PMID: 8934621 PMCID: 160879.
- [2] J. Tzen, Y. Cao, P. Laurent, C. Ratnayake, and A. Huang, "Lipids, proteins, and structure of seed oil bodies from diverse species," *Plant Physiology*, vol. 101, pp. 267–276, Jan. 1993.
- [3] J. C. Chen, C. C. Tsai, and J. T. Tzen, "Cloning and secondary structure analysis of caleosin, a unique Calcium-Binding protein in oil bodies of plant seeds," *Plant and Cell Physiology*, vol. 40, pp. 1079–1086, Jan. 1999.
- [4] D. J. Murphy, "Storage lipid bodies in plants and other organisms," *Progress in Lipid Research*, vol. 29, no. 4, pp. 299–324, 1990. PMID: 2135719.
- [5] J. Tzen and A. Huang, "Surface structure and properties of plant seed oil bodies," *The Journal of Cell Biology*, vol. 117, pp. 327–335, Apr. 1992.
- [6] D. J. Murphy and J. Vance, "Mechanisms of lipid-body formation," *Trends in Biochemical Sciences*, vol. 24, pp. 109–115, Mar. 1999.
- [7] G. Wanner and R. R. Theimer, "Membranous appendices of spherosomes (oleosomes)," *Planta*, vol. 140, pp. 163–169, 1978.
- [8] D. J. Murphy and I. Cummins, "Seed oil-bodies: Isolation, composition and role of oil-body apolipoproteins," *Phytochemistry*, vol. 28, no. 8, pp. 2063–2069, 1989.
- [9] I. Cummins, M. J. Hills, J. H. E. Ross, D. H. Hobbs, M. D. Watson, and D. J. Murphy, "Differential, temporal and spatial expression of genes involved in storage oil and oleosin accumulation in developing rapeseed embryos: implications for the role of oleosins and the mechanisms of oil-body formation," *Plant Molecular Biology*, vol. 23, pp. 1015–1027, Dec. 1993.

- [10] F. Beaudoin and J. A. Napier, “The targeting and accumulation of ecotopically expressed oleosin in non-seed tissues of *arabidopsis thaliana*,” *Planta*, vol. 210, pp. 439–445, Feb. 2000.
- [11] A. Bana, A. Dahlqvist, H. Debski, P. O. Gummesson, and S. Stymne, “Accumulation of storage products in oat during kernel development,” *Biochemical Society Transactions*, vol. 28, pp. 705–707, Dec. 2000. PMID: 11171178.
- [12] C. Sarmiento, J. H. E. Ross, E. Herman, and D. J. Murphy, “Expression and subcellular targeting of a soybean oleosin in transgenic rapeseed. implications for the mechanism of oil body formation in seeds,” *The Plant Journal*, vol. 11, pp. 783–796, Apr. 1997.
- [13] D. J. Murphy, “Production of novel oils in plants,” *Current Opinion in Biotechnology*, vol. 10, pp. 175–180, Apr. 1999. PMID: 10209131.
- [14] D. J. Lacey, F. Beaudoin, C. E. Dempsey, P. R. Shewry, and J. A. Napier, “The accumulation of triacylglycerols within the endoplasmic reticulum of developing seeds of *Helianthus annuus*,” *The Plant Journal*, vol. 17, pp. 397–405, Feb. 1999.
- [15] G. van Rooijen and M. M. Moloney, “Structural requirements of oleosin domains for subcellular targeting to the oil body,” *Plant Physiology*, vol. 109, pp. 1353–1361, Dec. 1995.
- [16] F. Beisson, *Etude des oléosomes de plantes et de leur lipolyse : Méthodes de dosage de l’activité des lipases*. PhD thesis, 1999.
- [17] P. Jolivet, E. Roux, S. d’Andrea, M. Davanture, L. Negroni, M. Zivy, and T. Chardot, “Protein composition of oil bodies in *arabidopsis thaliana* ecotype WS,” *Plant Physiology and Biochemistry*, vol. 42, pp. 501–509, June 2004.
- [18] J. Zimmerberg and M. M. Kozlov, “How proteins produce cellular membrane curvature,” *Nat Rev Mol Cell Biol*, vol. 7, pp. 9–19, Jan. 2006.
- [19] L. J. Pike, “The challenge of lipid rafts,” *Journal of Lipid Research*, vol. 50, pp. S323–S328, Apr. 2009. PMID: 18955730 PMCID: 2674732.
- [20] S. Beucher and C. Lantuejoul, “Use of watersheds in contour detection,” in *International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation*, Rennes, France., 1979.
- [21] P. G. J. Dongen and M. H. Ernst, “On the occurrence of a gelation transition in Smoluchowski’s coagulation equation,” *Journal of Statistical Physics*, vol. 44, pp. 785–792, Sept. 1986.
- [22] R. C. Ball, C. Connaughton, T. H. M. Stein, and O. Zaboronski, “Instantaneous gelation in Smoluchowski’s coagulation equation revisited,” *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 84, July 2011. PMID: 21867117.
- [23] H. Haario, M. Laine, A. Mira, and E. Saksman, “DRAM: efficient adaptive MCMC,” *Statistics and Computing*, vol. 16, pp. 339–354, Dec. 2006.

Dynamique des corps lipidiques dans la graine d'*Arabidopsis thaliana*

Résumé : Chez les végétaux, les lipides de réserve sont stockés dans des structures subcellulaires, les corps lipidiques (CL). Ces organelles quasi-sphériques sont constituées d'un coeur de triacylglycérols (TAGs), entourés d'une monocouche de phospholipides (PLs) et sont produites à partir du réticulum endoplasmique avant d'être libérés dans le cytoplasme cellulaire. Les oléosines, dont il existe 5 isoformes graine spécifiques (S1 à S5) chez *Arabidopsis thaliana*, sont des protéines majeures du CL, insérées à la surface de sa demi-membrane. La dynamique du CL (chargement/déchargement en huile) est complexe et reste largement mal comprise. L'objectif de ce travail est de modéliser la formation et la dynamique des corps lipidiques dans la graine en développement de l'espèce *Arabidopsis thaliana*, afin de mieux appréhender les mécanismes responsables de la biogenèse et la dynamique des CLs. L'utilisation de colorants des lipides neutres constituant les CLs, couplée à la microscopie confocale, a permis l'obtention de piles d'images de CLs d'embryons à différents jours du développement, en contexte sauvage et en contexte déplétif pour une, deux ou trois oléosines (S1, S3 et S4).

- Un pipeline de segmentation d'images a tout d'abord été développé pour extraire différents estimateurs caractérisant la taille et la dispersion spatiale des corps lipidiques. Les estimateurs ont permis d'analyser l'évolution de la taille et de la dispersion spatiale des corps lipidiques en fonction du temps du développement, et de mettre en évidence la variabilité entre génotypes.
- Ces données ont ensuite été analysées et étudiées statistiquement par des approches utilisant des modèles linéaires et des modèle quantile qui ont permis de conclure sur l'effet de chacune des oléosines étudiées, ainsi que celui de leurs interactions, sur la distribution des corps lipidiques.
- Enfin, un modèle décrivant la dynamique de coalescence de la population des corps lipidiques a été proposé, simulé numériquement, puis comparé aux données expérimentales. Ce modèle a permis de tester différentes hypothèses de la dynamique de biogenèse et de croissance par coalescence du corps lipidique formalisées dans le modèle mathématique. Différents effets de la composition du corps lipidique en oléosines sur la vitesse de coalescence des corps lipidiques ont été mis en évidence. Les résultats de ces trois axes ont permis de proposer et discuter des rôles associés à chacune oléosine dans une perspective de compréhension des mécanismes mis en oeuvre dans la dynamique du corps lipidique.

Mots clés : Corps lipidiques, oléosine, dynamique, images, modélisation

The dynamics of oil bodies in *Arabidopsis thaliana* seed

Abstract : In plants, lipid reserves are stored in subcellular structures called lipid bodies (LB). These virtually spherical organelles consist of a core of triacylglycerols (TAG), surrounded by a monolayer of phospholipids (PLs), are produced from the endoplasmic reticulum and then released into the cell cytoplasm. Oleosins, composed of five seed-specific isoforms (S1 to S5) in *Arabidopsis thaliana*, are major proteins of the LB, inserted on the surface of the half-membrane. The dynamics of LB (charging / uncharging oil) is complex and remains largely misunderstood. The objective of this work is to model the formation and dynamics of lipid bodies in the developing seed of *Arabidopsis thaliana*, to better understand the mechanisms responsible for the biogenesis and dynamics of LBs . The use of dyes staining neutral lipids constituting the LD, coupled with confocal microscopy, allowed obtaining image stacks of LB from embryos at different days of development, in a wild-type or depleted (mutant) context for one, two or three oleosins (S1, S3 and S4).

- An image segmentation pipeline has been first developed, enabling extraction of various estimators for characterizing the size and spatial dispersion of the lipid bodies. Estimators were used to analyse the evolution of the size and spatial dispersion of lipid bodies as a function of stage of development, and to highlight the variability between genotypes.
- These data were then processed and statistically analysed by approaches using linear as well as quantile model that concluded on the effect of each of oleosins investigated as well as their interactions on the distribution of lipid bodies.
- Last, a model describing the coalescence dynamics of LB populations has been proposed, digitally simulated and compared to experimental data sets. This model was used to test various hypotheses on the dynamics of biogenesis and coalescence-based growth of lipid bodies as formalized according to the mathematical model. Several effects of oleosin composition on LB coalescence rate have been highlighted. The results of these three axes allowed to propose and discuss the roles associated with each oleosin in the broader perspective of understanding the mechanisms involved in the lipid bodies dynamics.

Keywords : Lipid body, oleosin, dynamics, imaging, modeling
