

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

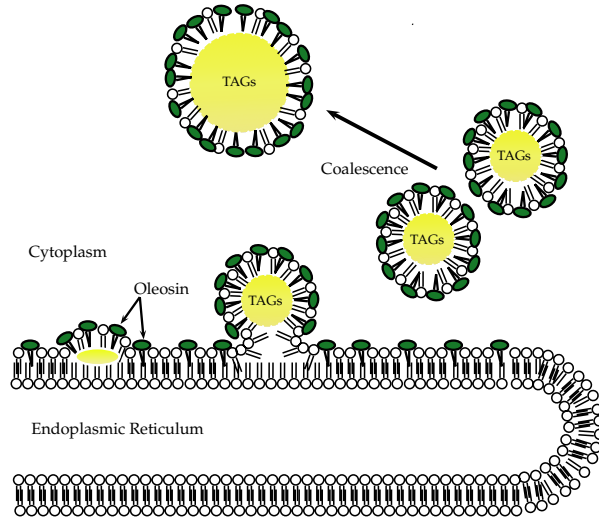


Fig. 1: Biogenesis of OBs by bending from the two leaflets of the ER. OBs that do not contain enough oleosins to coat their entire surface coalesce, forming larger OBs. Coalescence of OBs stops when the surface is completely covered with oleosins.

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:

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

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 serie $\sum_v \sum_{v'} v \alpha n_c(v', t) n_c(v-v', t)$ 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:

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

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 $100 nm$) 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:

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

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

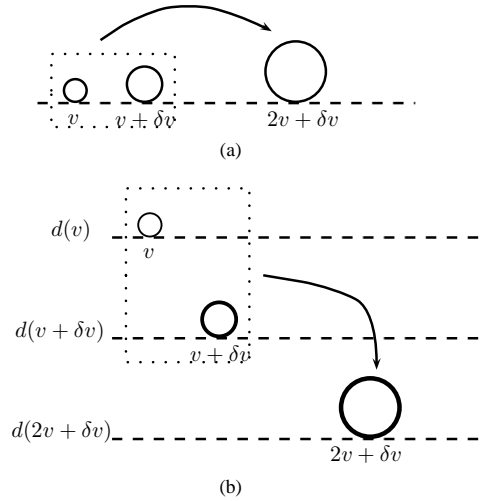


Fig. 2: Discretization methods used to discretize volume classes in the SDND model (a), and volume and coverage quantity classes in the SDD model (b).

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.

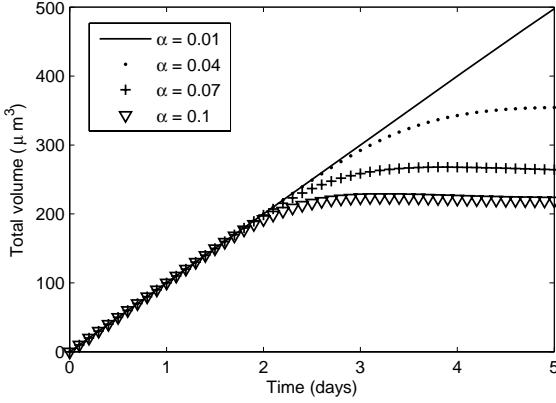


Fig. 3: The total cytoplasm volume in function of time for different values of α

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 :

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

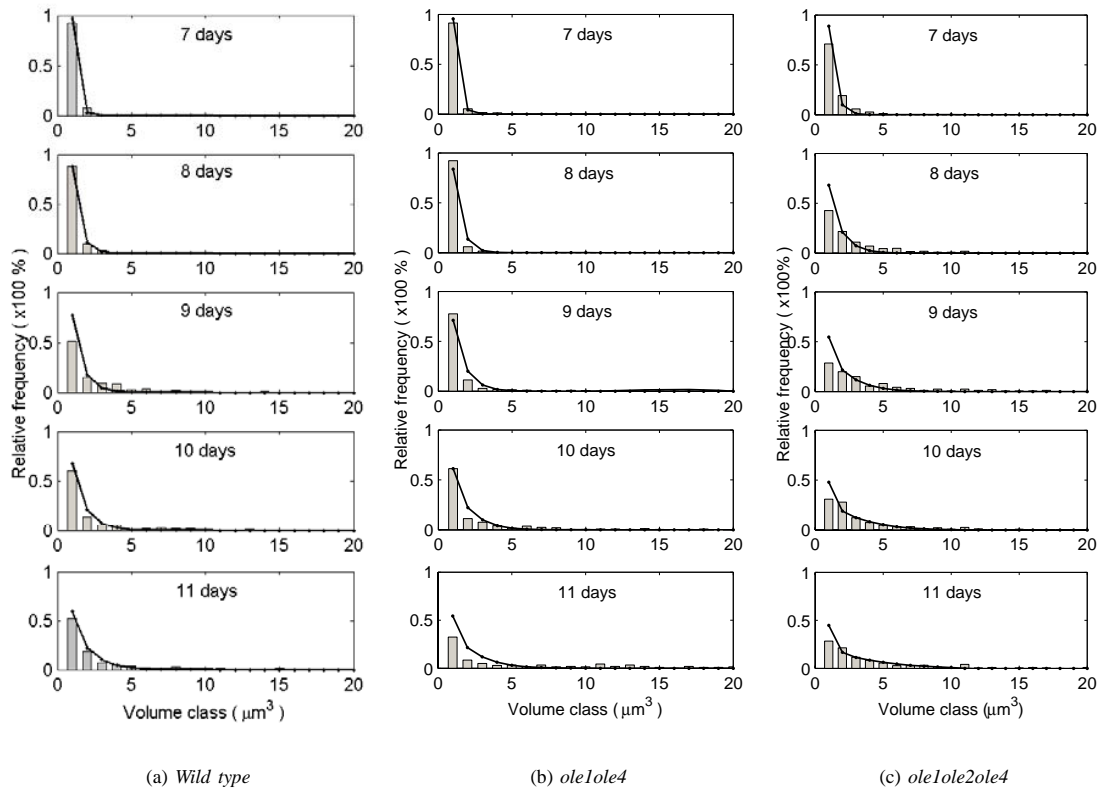


Fig. 4: OB-volume distribution at 7, 8, 9, 10 and 11 days of development for wild type and different mutants. Bars plot represent experimental data, and solid lines represent the SDD model simulations.

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

ACKNOWLEDGMENT

The authors wish to acknowledge the help of Olivier Martin, Christophe Deroulers and Silvio Franz from universit  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 PMID: 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 ectopically 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. Negrone, 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 PMID: 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.