



# Development and testing of a pipeline of analysis methods enabling an integrative analysis of the development from cell to shoot scale.

Maryline Lievre

## ► To cite this version:

Maryline Lievre. Development and testing of a pipeline of analysis methods enabling an integrative analysis of the development from cell to shoot scale.. Morphogenesis. Institut National d'Etudes Supérieures Agronomiques de Montpellier, 2014. English. NNT : . tel-02794060

**HAL Id: tel-02794060**

**<https://hal.inrae.fr/tel-02794060>**

Submitted on 5 Jun 2020

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

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

# THÈSE

Pour obtenir le grade de  
**Docteur**

Délivré par le  
**Centre international d'études supérieures en sciences  
agronomiques – Montpellier SupAgro**

Préparée au sein de l'école doctorale SIBAGHE  
Et de l'unité de recherche UMR 759 - LEPSE

Spécialité : **Biologie Intégrative des Plantes**  
CNU : Physiologie

Présentée par **Maryline LIEVRE**

**Analyse multi-échelle et modélisation de la  
croissance foliaire chez *Arabidopsis  
thaliana*.**

—

**Mise au point et test d'un pipeline d'analyses  
permettant une analyse intégrée du  
développement de la cellule à la pousse entière.**

Soutenue le 15 décembre 2014 devant le jury composé de

|  |                     |
|--|---------------------|
| M. Teva VERNOUX, Directeur de Recherche, CNRS RDP                      | Rapporteur          |
| M. Bruno ANDRIEU, Directeur de Recherche, INRA EGC                     | Rapporteur          |
| Mme Nadia BERTIN, Directeur de Recherche, INRA PSH                     | Examinatrice        |
| M. Christian JAY-ALLEMAND, Professeur, Université Montpellier II       | Examineur           |
| Mme Christine GRANIER, Chargée de Recherche, INRA LEPSE                | Directrice de thèse |
| M. Yann GUEDON, Chercheur CIRAD, CIRAD AGAP & INRIA,<br>Virtual Plants | Co-encadrant        |





# PhD Thesis

Delivered by the  
**International center of higher education in agronomic sciences –  
Montpellier SupAgro**

Prepared in the doctoral school SIBAGHE  
And the research unit UMR 759 - LEPSE

Speciality: **Plant Integrative Biology**  
CNU: Physiology

Presented by **Maryline LIEVRE**

**Multi-scale analysis and modeling of shoot  
growth in *Arabidopsis thaliana***

–

**Development and testing of a pipeline of analysis  
methods enabling an integrative analysis of the  
development from cell to shoot scale.**

Defended on December 15<sup>th</sup>, 2014

Jury:

|   |               |
|---|---------------|
| Mr Teva VERNOUX, Senior Researcher, CNRS RDP                          | Main examiner |
| Mr Bruno ANDRIEU, Senior Researcher, INRA EGC                         | Main examiner |
| Mrs Nadia BERTIN, Senior Researcher, INRA PSH                         | Examiner      |
| Mr Christian JAY-ALLEMAND, Professor, University Montpellier II       | Examiner      |
| Mrs Christine GRANIER, Researcher, INRA LEPSE                         | Supervisor    |
| Mr Yann GUEDON, Senior Researcher, CIRAD AGAP & INRIA, Virtual Plants | Co-supervisor |



---

LABORATOIRE D'ACCUEIL

LEPSE

Laboratoire d'Ecophysiologie des Plantes sous Stress Environnementaux

UMR 759 INRA / Montpellier SupAgro

IBIP – Institut de Biologie Intégrative des Plantes

Bâtiment 7

2 place Pierre Viala

34060 Montpellier cedex 1

FRANCE

---

FINANCEMENT CIFRE

Association Nationale de la Recherche et de la Technologie – Ministère français de  
l'Enseignement supérieur et de la Recherche

&

Bayer CropScience représenté par le GIE AIFOR

---

---

## RESUME

Ce travail est basé sur le constat du manque de méthodes permettant l'analyse intégrée des processus contrôlant le développement végétatif d'*Arabidopsis thaliana* dans les études phénotypiques multi-échelles. Un phénotypage préliminaire de la croissance foliaire de 91 génotypes a permis de sélectionner 3 mutants et des variables d'intérêt pour une étude plus poussée du développement de la pousse. Un pipeline de méthodes d'analyses combinant techniques d'analyse d'images et modèles statistiques a été développé pour intégrer les mesures faites à l'échelle de la feuille et de la pousse. Des modèles multi-phasiques à changements de régime semi-markovien ont été estimés pour chaque génotype permettant une caractérisation plus pertinente des mutants. Ces modèles ont validé l'hypothèse selon laquelle le développement de la rosette peut être découpé en une suite de phases de développement, pouvant varier selon les génotypes. Ils ont aussi mis en évidence le rôle structurant de la variable «trichome abaxial», bien que les phases de développement ne puissent être entièrement expliquées par ce trait. Un 2<sup>nd</sup> pipeline d'analyses combinant une méthode semi-automatique de segmentation d'images de l'épiderme foliaire et l'analyse des surfaces de cellules par un modèle de mélange de lois gamma à paramètres liés par une loi d'échelle a été développé. Ce modèle nous a permis d'estimer la loi du nombre de cycles d'endoréduplication. Nous avons mis en évidence que cette loi dépendait du rang de la feuille. Le cadre d'analyses multi-échelles développé et testé durant cette thèse devrait être assez générique pour être appliqué à d'autres espèces végétales dans diverses conditions environnementales.

**Mots clés :** Développement foliaire, analyse multi-échelles, modèle de segmentation, hétéroblastie, *Arabidopsis thaliana*, cellules épidermiques.

---

## SUMMARY

This study is based on the observation of a lack of methods enabling the integrated analysis of the processes controlling the vegetative development in *Arabidopsis thaliana* during multi-scale phenotypic studies. A preliminary leaf growth phenotyping of 91 genotypes enabled to select 3 mutants and different variables of interest for a more in depth analysis of the shoot development. We developed a pipeline of analysis methods combining image analysis techniques and statistical models to integrate the measurements made at the leaf and shoot scales. Semi-Markov switching models were built for each genotype, allowing a more thorough characterization of the studied mutants. These models validated the hypothesis that the rosette can be structured into successive developmental phases that could change depending on the genotype. They also highlighted the structuring role of the 'abaxial trichomes' variable, although the developmental phases cannot be explained entirely by this trait. We developed a second pipeline of analysis methods combining a semi-automatic method for segmenting leaf epidermis images, and the analysis of the obtained cell areas using a gamma mixture model whose parameters of gamma components are tied by a scaling rule. This model allowed us to estimate the mean number of endocycles. We highlighted that this mean number of endocycles was function of the leaf rank. The multi-scale pipeline of analysis methods that we developed and tested during this PhD should be sufficiently generic to be applied to other plant species in various environmental conditions.

**Keywords:** Shoot development, multi-scale analysis, segmentation model, heteroblasty, *Arabidopsis thaliana*, epidermal cells.

---

## REMERCIEMENTS

La préparation d'une thèse est un travail collectif, et je tiens ici à remercier les personnes qui ont collaboré de près ou de loin à l'aboutissement de mon travail.

Je tiens tout d'abord à remercier Christine d'avoir cru en moi suite au stage de master pendant lequel elle m'a encadrée et pour m'avoir donné l'opportunité de faire cette thèse. Merci pour ton aide et ta disponibilité pendant ces trois (presque quatre) années. Tu n'as pas craint de remettre en question certaines pratiques de phénotypage pourtant bien installées et c'est grâce à ta curiosité que ce projet a pu voir le jour. Le chemin n'a pas toujours été facile mais je suis heureuse qu'on ait su mener notre collaboration à bien et dans de très bonnes conditions. Tu m'as beaucoup appris sur moi d'un point de vue professionnel. Je te remercie aussi grandement de m'avoir cédé ton bureau pour le temps de la rédaction pour me permettre de travailler dans mes conditions « vampiresques » sans qu'Anaëlle et Béatriz aient à en souffrir ! Tu apportes un vrai dynamisme dans le laboratoire, scientifiquement et personnellement. Sans les croissants suivant les victoires de Montpellier et les commentaires à la pétanque le LEPSE ne serait pas le même.

Un grand merci aussi à Yann pour son encadrement, son implication dans le projet, la patience avec laquelle il m'a expliqué les tenants et les aboutissants des modèles statistiques que nous avons dû développer. Sans tes compétences et ton intérêt nous n'aurions pas pu mener ce projet à bout. Grâce à toi j'ai vraiment pu développer de nouvelles connaissances dans ce domaine qui m'intéressait mais dans lequel je n'avais que de faibles bases, et c'était une des choses que je souhaitais en m'attelant à cette thèse. Je veux aussi te remercier pour la gentillesse dont tu as su faire preuve lors des périodes difficiles.

Je tiens aussi à remercier plus largement tous ceux qui m'ont aidé à constituer nos jeux de données (et ce ne fut pas une mince affaire !). Un énorme merci donc à Alex, Pablo, Tony, Rémi et Jessica (j'espère n'oublier personne) pour les heures passées penchés sur les tablettes graphiques à dessiner des dizaines de milliers de feuilles (je viens de regarder, plus de 45000 mesures au total sans les repassages nécessaires les gars, c'est nous les boss !) et à corriger les segmentations des empreintes d'épidermes !! On aura au moins pu enrichir mutuellement nos cultures musicales et passer des bons moments à rire et à râler en écoutant les débats à la radio en imagerie. Merci aussi à Myriam, la maman des PHENOPSIS, pour ta disponibilité et ta réactivité. N'oublions pas que sans tes « bébés » je pouvais dire adieu à mes jeux de données ! Merci à tous ceux qui ont pris part aux récoltes, Alex, Tony, Myriam et Jessica à nouveau,



---

Cris, Gaëlle aussi que je n'ai pas encore citée alors que je pouvais toujours compter sur elle, et tous ceux qui sont passés donner un coup de main ponctuel dans les moments de rush. Merci aussi à Mathilde Bettembourg pour les résultats qu'elle a produit pendant son stage.

Tous mes remerciements aussi à Christian qui a accepté de lire ce document en avant-première et m'a aidé à prendre un peu de recul sur certains aspects dans les heures intenses précédant l'envoi du manuscrit. Merci aussi à tous les chercheurs avec qui j'ai pu échanger, en particulier Denis Vile, Olivier Turc, Thierry Simonneau, Angélique Christophe et Éric Lebon pour les occasionnelles discussions scientifiques et sur la recherche en général qui m'ont beaucoup apporté.

Je tiens aussi à remercier toute l'équipe de Virtual Plants qui a fait très bon accueil lors de mes passages dans ses murs. Une attention particulière à l'égard de Léo Guignard qui a participé au développement du protocole de segmentation d'images. Ton expertise a été précieuse et les pauses clopes à la fois agréables et productives. Merci aussi à Christophe Godin pour sa participation aux discussions concernant la segmentation des images.

A ce sujet je veux aussi dire merci à Ullrich Koethe et Niko Krasowski pour les échanges autour de l'utilisation d'ilastik. J'espère que cette expérience aura des retombées positives pour toutes les parties impliquées. Je salue également J. L. Micol qui nous a fourni les graines des 90 mutants SALK que son équipe avait préliminairement screenés.

Je remercie aussi Bayer CropScience pour avoir soutenu (scientifiquement et financièrement) ma thèse. Je suis particulièrement reconnaissante à Rajendra Bari et Korneel Vandenbroucke pour la liberté qu'ils nous ont laissé face à l'orientation du sujet. Merci également à Bernard Pelissier pour avoir bien voulu remplir le rôle de correspondant auprès du GIE AIFOR.

Merci aussi à Nadine Hilgert, Jean-Luc Regnard et Patrick Laufs pour avoir participé à mes comités de thèse. Les discussions que nous avons eues à ces occasions ont été très enrichissantes et ont bien aidé à cadrer le travail à faire. Merci particulièrement à Patrick pour nous avoir fourni des graines pour notre dernière manip. J'espère que les résultats obtenus vous seront utiles.

Je tiens également à remercier les membres du jury pour avoir bien voulu analyser et étudier mon travail. Merci donc à Christian Jay-Allemand pour avoir accepté d'être le président du jury, et à Bruno Andrieu et Teva Vernoux pour avoir accepté la lourde tâche des rapporteurs. Merci également à Nadia Bertin son intérêt pour avoir accepté de faire partie du

---

jury de thèse au titre d'examinatrice.

De manière générale je remercie l'ensemble du LEPSE pour la bonne ambiance qui règne dans ce labo. Merci à Anaëlle, Emilie, Nathalie, Aude, Erwann, Justine, Sébastien, Benoît V. et les joueurs occasionnels pour les parties de tarots du midi (Emilie, tu auras finalement réussi à tenir malgré mes chansonnettes, merci d'avoir résister). Merci à Pauline, Stéphane, Tony, Nico, Alex, Jonathan, Antonin, Romain, Benoît S., Vincent N. et les autres pour les pétanques. Merci à tous pour les pauses de quatre heure, les apéros et barbecues de fin de semaine qui permettaient de relâcher un peu la pression. Merci aussi d'avoir supporté mes discours sans fin sur la politique et la société, les débats ont souvent été enrichissants ! Un immense merci à tous ceux qui ont bien voulu me dépanner multiples cigarettes ces derniers mois, surtout à Alex (fournisseur officiel !). Je vous dois une fière chandelle (et bien plus...). Je tiens d'ailleurs à remercier plus particulièrement Alex. Ta gentillesse, ta bonne humeur et ton écoute attentive m'ont sauvé bien des journées et je t'en serai toujours reconnaissante.

J'aimerais aussi faire des remerciements plus personnels à tous mes proches qui ont vécu cette thèse en même temps que moi. Sandra, tu as été mon soutien indéfectible, merci mille fois pour les heures passées sur Skype ou au téléphone à parler de la recherche et de la vie en général. Tu as toujours su me rassurer et ton regard extérieur a permis de me débloquer plusieurs fois, j'espère pouvoir te retourner la faveur dans l'avenir. Merci aussi bien sûr à celui qui partage ma vie depuis plus de 2 ans, Sam. Tu as toujours été là pour moi, tu as su me soutenir, t'impliquer et me proposer ton aide aux moments clef et même me remettre sur pieds quand j'en ai eu besoin. Je te remercie aussi pour les multitudes de choses que tu m'as fait découvrir, et surtout pour ta façon de me faire rire, toujours. Sans toi, ces années auraient été bien plus difficiles à vivre. Enfin, un grand merci « aux copains », Pablo, Clem, Tom, François, Simon, Alex M., Ian, Sandra C., pour les heures passées à « coincher », à s'enflammer dans des débats sans fin, à refaire le monde, à chanter. Vous m'avez aidé à élargir ma vision du monde et des possibles, et les moments passés avec vous ont souvent été des bouffées d'air bien nécessaires. Merci enfin à ma famille, mes parents et ma petite sœur qui m'ont accompagné de loin dans cette aventure un peu folle qu'est la thèse. Votre soutien et vos encouragements m'ont été précieux.

---

---

---

# TABLE OF CONTENT

|   |               |
|---|---------------|
| <b>CHAPTER I. CONTEXT, AIM AND STRATEGY OF THE PHD</b>  | <b>15</b>     |
| <b>1. Presentation of different processes involved in <i>Arabidopsis thaliana</i> rosette development</b>   | <b>15</b>     |
| <b>2. Interactions between leaf growth related processes occurring at different scales</b>  | <b>18</b>     |
| 2.1. Contribution of cell division and expansion to final leaf size   | 18            |
| 2.2. Impact of flowering transition on leaf area development and its cellular components  | 19            |
| <b>3. Practices and pitfalls in leaf growth phenotyping</b>   | <b>20</b>     |
| 3.1. Interest of phenotyping leaf growth over time  | 21            |
| 3.2. Phenotyping leaf growth over time: late differences can be due to early processes  | 22            |
| 3.3. Comparing growth of successive leaves along the shoot  | 23            |
| <b>4. Aim and strategy of the PhD</b>   | <b>25</b>     |
| <br><b>CHAPTER II. USE AND LIMITS OF EXISTING PIPELINE OF ANALYSES FOR COMPARING LEAF GROWTH PHENOTYPES IN <i>A. THALIANA</i>: A TEST ON 90 T-DNA INSERTION LINES</b> | <br><b>27</b> |
| <b>1. Introduction</b>  | <b>27</b>     |
| <b>2. Material and Methods</b>  | <b>29</b>     |
| 2.1. Biological material  | 29            |
| 2.2. Growth conditions  | 29            |
| 2.3. Growth measurements  | 30            |
| 2.4. Statistical analyses   | 33            |
| <b>3. Results</b>   | <b>33</b>     |
| 3.1. Lines with contrasted leaf phenotypes could not be identified robustly when leaf growth variables were compared 18 days after sowing                             | 33            |
| 3.2. Many lines were affected on the different growth variables measured at the opening of the first flower when compared to the wild-type                            | 35            |
| 3.3. Selection of 4 lines with contrasted leaf growth phenotypes for a more in depth phenotypical study   | 38            |
| 3.4. In depth phenotyping analysis of the 4 contrasted lines  | 40            |
| <b>4. Discussion</b>  | <b>46</b>     |
| 4.1. Early stage leaf growth phenotyping does not reflect the final leaf growth phenotype   | 46            |
| 4.2. Considerations regarding the results of the more in depth phenotyping analysis   | 47            |
| <br><b>CHAPTER III. INTEGRATIVE MODELS FOR ANALYZING <i>A. THALIANA</i> ROSETTE DEVELOPMENTAL PHASES</b>  | <br><b>51</b> |
| <b>1. Introduction</b>  | <b>51</b>     |
| <b>2. Material &amp; Methods</b>  | <b>53</b>     |
| 2.1. Plant material and growth conditions   | 53            |
| 2.2. Growth measurements – Image analyses and variables selection   | 54            |
| 2.3. Integrative multi-scale models for analyzing <i>A. thaliana</i> rosette developmental phases   | 56            |
| <b>3. Results</b>   | <b>61</b>     |
| 3.1. Validation of the assumption of segmentation in developmental phases   | 61            |

---

|   |  |            |
|---|--|------------|
| 3.2.  | Comparison of the segmentations into developmental phases between Col-0 and each T-DNA insertion line  | 61         |
| 3.3.  | Effect of the mutations on the values of leaf blade length-to-width ratio, final leaf area and characteristic growth duration                                      | 68         |
| 3.4.  | Analysis of global plant shoot development: effects of the mutations on the duration of the vegetative phase, mean leaf apparition rate and final number of leaves | 71         |
| <b>4.</b>   | <b>Discussion</b>  | <b>71</b>  |
| 4.1.  | The successive developmental phases structuring <i>A. thaliana</i> rosette are determined by a combination of leaf growth traits                                   | 72         |
| 4.2.  | Our analysis method gives new insights concerning the effect of the mutations in T-DNA insertions lines.   | 74         |
| <br><b>CHAPTER IV. A NEW PIPELINE OF METHODS COMBINING MACHINE LEARNING APPROACHES FOR IMAGE SEGMENTATION AND GAMMA MIXTURE MODELS FOR ANALYZING CELL DIMENSIONS IN <i>A. THALIANA</i> LEAF EPIDERMIS</b> |  | <b>77</b>  |
| <b>1.</b>   | <b>Introduction</b>  | <b>77</b>  |
| <b>2.</b>   | <b>Methodological developments</b>   | <b>80</b>  |
| 2.1.  | Plant material and growth conditions   | 80         |
| 2.2.  | Image acquisition  | 81         |
| 2.3.  | Image segmentation using ilastik v.0.5.12  | 81         |
| 2.4.  | Gamma mixture models incorporating a scaling rule between components   | 85         |
| <b>3.</b>   | <b>Results</b>   | <b>88</b>  |
| 3.1.  | Identification of the scaling rule and selection of the number of components of the mixture model  | 88         |
| 3.2.  | Estimation of the endoreduplication factor in the epidermis using the 4-component mixture model  | 92         |
| <b>4.</b>   | <b>Discussion</b>  | <b>94</b>  |
| 4.1.  | Semi-automatic method for measuring cell areas in <i>A. thaliana</i> epidermis imprints: an improvement for analyzing cell size diversity                          | 94         |
| 4.2.  | Analyzing cell size diversity in <i>A. thaliana</i> epidermis using a 4-component gamma mixture model  | 94         |
| <br><b>CHAPTER V. CONCLUSIONS AND PERSPECTIVES</b>  |  | <b>97</b>  |
| <b>1.</b>   | <b>Added value of our pipeline for the analysis of <i>A. thaliana</i> shoot development</b>  | <b>97</b>  |
| <b>2.</b>   | <b>Added value of our pipeline for the analysis of cell area diversity in <i>A. thaliana</i> leaf epidermis</b>  | <b>99</b>  |
| <b>3.</b>   | <b>Integrating the estimated endoreduplication factor in our pipeline to analyse shoot structuring</b>   | <b>100</b> |
| <b>4.</b>   | <b>What next? Incorporating new variables in the multivariate pipeline and using targeted genotypes for specific questions</b>                                     | <b>104</b> |
| <br><b>REFERENCES</b>   |  | <b>107</b> |
| <br><b>APPENDIX I . LIST OF T-DNA INSERTION LINES GROWN IN EXPERIMENTS 1 &amp; 2</b>  |  | <b>121</b> |

---

---

|  |            |
|--|------------|
| <b>APPENDIX II . SUPPLEMENTARY METHODS FOR INTEGRATIVE MODELS FOR ANALYZING <i>A. THALIANA</i> ROSETTE DEVELOPMENTAL PHASES</b>                                    | <b>122</b> |
| <b>APPENDIX III . SUPPLEMENTARY FIGURES FOR INTEGRATIVE MODELS FOR ANALYZING <i>A. THALIANA</i> ROSETTE DEVELOPMENTAL PHASES – MACROSCOPIC SCALE</b>               | <b>125</b> |
| <b>APPENDIX IV . COMPARISON BETWEEN MANUAL AND SEMI-AUTOMATIC IMAGE SEGMENTATION METHODS</b>   | <b>131</b> |
| <b>APPENDIX V . SUPPLEMENTARY FIGURES FOR INTEGRATIVE MODELS FOR ANALYZING <i>A. THALIANA</i> ROSETTE DEVELOPMENTAL PHASES</b>                                     | <b>135</b> |
| <b>APPENDIX VI . PRELIMINARY ANALYSES ON MUTANT LINES INVESTIGATING THE CROSSTALK BETWEEN MIRNA-DEPENDENT PATHWAYS AND SHOOT DEVELOPMENT IN <i>A. THALIANA</i></b> | <b>137</b> |
| <b>APPENDIX VII . POSTER OF M. LIÈVRE ET AL. (ICAR 2012)</b>   | <b>139</b> |
| <b>APPENDIX VIII . POSTER OF J.L. MICOL ET AL. (ICAR 2012)</b>   | <b>140</b> |

---

## CHAPTER I. CONTEXT, AIM AND STRATEGY OF THE PhD

### 1. PRESENTATION OF DIFFERENT PROCESSES INVOLVED IN *ARABIDOPSIS THALIANA* ROSETTE DEVELOPMENT

The shoot system is an essential unit of plant architecture as it bears leaves that play important roles in photosynthesis, respiration and photorespiration. Therefore, understanding how leaves develop during shoot formation is an important subject in plant biology research. Because of its small size, short life cycle and genetic resources available in different stock centers, *Arabidopsis thaliana* (*A. thaliana*) has been selected as a model plant for the systematic characterization of growth and development (Alonso et al. 2003; Koornneef and Meinke 2010). There has been extensive recent progress in understanding mechanisms controlling growth in this species, even if the circuitry that links the different mechanisms remains unclear. Shoot establishment results from the combination of developmental changes – i.e. the succession of events that contribute to the increase in organ number – with growth – i.e. the irreversible increase in organ dimensions over time. Developmental changes are often abrupt and take the form of a series of discrete events such as leaf initiation, emergence, or end of expansion (see also Boyes et al. 2001) for other developmental changes at the whole plant scale). In contrast, growth, such as the dynamic change in leaf area is gradual and often takes the form of a trend.

During shoot development leaves are initiated at regular intervals as primordia at specific sites on the shoot apical meristem (SAM). This is the first manifestation of the leaf that is not visible by naked eye. The rhythm at which leaf primordia are produced can affect the final number of leaves as shown in different mutants or different environmental conditions (Cookson et al. 2005). It is often but not strictly related to the size of the shoot apical meristem (Clark et al. 1993; Mauseth 2004; Boucheron et al. 2005; Skirycz et al. 2010). Leaf primordia localization is specified by auxin maxima in a flanking region of the SAM (Reinhardt et al. 2000), following a spiral phyllotaxy. The genetic, hormonal and mathematical rules defining this phyllotaxy have been extensively studied (de Reuille et al. 2006; Jönsson et al. 2006; Peaucelle and Laufs 2007; Mandel et al. 2014).



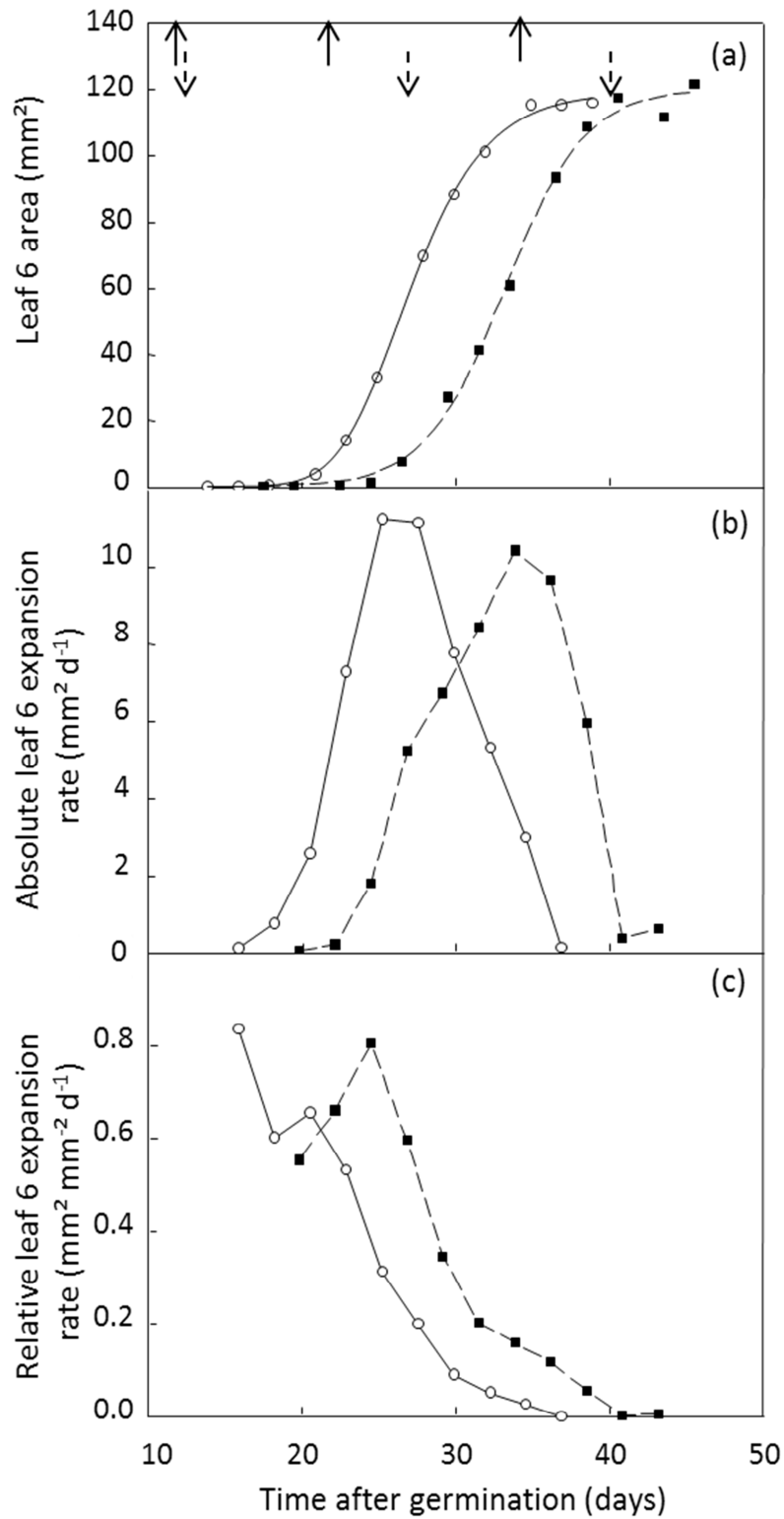


Figure I-1. Dynamics of leaf 6 area (a) and corresponding changes in absolute (b) and relative (c) leaf expansion rate for two *A. thaliana* genotypes, *Ler* (○) and *elo-1* (■) ( $n = 5$  or  $6$ ). The curves fitted in (a) are 3-parameter sigmoids for *Ler*  $y = 116.56/[1 + \exp - \{(t - 17.11)/2.25\}]$  and *elo-1*  $y = 120.22/[1 + \exp - \{(t - 23.18)/2.58\}]$ . For each genotype, leaf development was characterized by three successive stages shown in (a): dates of initiation, emergence and end of expansion. They are indicated from left to right by solid upward and dashed downward arrows for *Ler* and *elo-1* respectively. From Lièvre et al. (2013).

The leaf primordium is defined as a group of founder cells that develop further into a leaf after a series of coordinated division and expansion (Horiguchi et al. 2006a; Vanhaeren et al. 2010; Gonzalez et al. 2012). In dicotyledons, these processes co-exist spatially during the first stages of leaf development. One usually refers to the separation between the cell-proliferative area and the area where cells stop dividing and begin expanding and differentiating as the ‘cyclic arrest front’ (Nath et al. 2003). It was initially believed that cell proliferation stopped following a tip-to-base longitudinal gradient, but a recent study demonstrated that, in fact, the cyclic arrest front remains located at an almost constant distance from the leaf base during a certain period and then disappears relatively quickly progressing toward the base (Kazama et al. 2010). Numerous positive and negative regulators of cell proliferation in leaf primordia have been identified (see Tsukaya 2013 for review).

At the individual leaf level, leaf expansion corresponds to a gradual but non-linear gain of surface (Figure I-1 (a)). Individual leaves grow at a maximal relative expansion rate – i.e. a maximal area formed per unit area and unit of time – during the hidden phase (Figure I-1 (c)) whereas absolute leaf expansion rate – i.e. the area formed per unit of time – is low during this phase (Figure I-1 (b)) (Granier and Tardieu 1998). Upon emergence, relative leaf expansion rate decreases over time until the cessation of expansion, whereas absolute leaf expansion rate increases until a maximal value and decreases afterwards (Granier and Tardieu 2009). These trends are similar in eudicots and monocots with differences in the spatial distribution of expansion rate over the lamina resulting in a long phase with linear expansion, i.e. constant absolute expansion rate, in monocots that does not occur in eudicots (see Granier and Tardieu 2009 for details).

In *A. thaliana* and most flowering plants the cessation of rosette leaf initiation is due to a change in the SAM identity. This transition is the switch from vegetative to reproductive development. Environmentally and genetically induced changes in the timing of floral transition have been extensively studied. We thus start to have a clear knowledge of molecular networks controlling this transition in *A. thaliana*, as well as of the role of environmental variables such as photoperiod and temperature, as evidenced by recent reviews on the subject (Amasino 2010; Huijser and Schmid 2011; Srikanth and Schmid 2011; Andrés and Coupland 2012).

Floral transition in *A. thaliana* is visible because of the elongation of the primary axis internodes and the production of novel structures such as flowers. However, as in other higher plants, *A. thaliana* undergoes another, earlier, less apparent transition. Its vegetative

development can be divided into two main phases – a juvenile and an adult vegetative phase (Poethig 1990; Huijser and Schmid 2011). The transition between these phases, called vegetative phase change, corresponds to the acquisition of reproductive competence by the plant. It also involves changes, sometimes quite subtle, in leaves morphological traits. In *A. thaliana*, these changes include leaf size, leaf blade length-to-width ratio (Steynen et al. 2001; Cookson et al. 2007), number and depth of serrations (Röbbelen 1957), patterns of trichome production (Chien and Sussex 1996; Telfer et al. 1997) and epidermal cell area (Cookson et al. 2007; Usami et al. 2009). Recent progresses have been made in understanding the molecular pathways controlling the vegetative phase change (Poethig 2010, 2013; Yang et al. 2011, 2013; Willmann and Poethig 2011; Matsoukas et al. 2013), but the regulation of this transition remains largely unclear.

The different processes that have been presented here have mainly been studied individually, but interactions have been highlighted between processes even if, for now, our knowledge on this field is still limited. The following section presents some examples of these interactions found in the literature.

## **2. INTERACTIONS BETWEEN LEAF GROWTH RELATED PROCESSES OCCURRING AT DIFFERENT SCALES**

### **2.1. Contribution of cell division and expansion to final leaf size**

The majority of mutants displaying smaller leaves compared to their wild-type also have similar decreases in both their number and size of cells in the leaf (Horiguchi et al. 2006a). Some mutations increasing both number and size of cells also result in larger organs (Feng et al. 2011). Both the size and number of leaf cells seem to affect and be determinant in the control of the final sizes of leaves. However, the observation of compensation phenomenon in leaf morphogenesis, i.e. an abnormal increase in cell volume triggered by a decrease in cell number, revealed the existence of integration systems linking levels of cell proliferation and cell expansion in leaves (Tsukaya 2002; Ferjani et al. 2007). A study suggesting that in order to the compensation phenomenon to be triggered cell proliferation has to decrease below some threshold (Fujikura et al. 2009) led to the proposal of the existence of a cell-cell communication system (Kawade et al. 2010). The existence of this threshold could indicate that the leaf size is to some extent uncoupled from the size and number of cells by the compensatory system. Moreover, using QTLs analysis and structural equation models, Tisé

et al. (2008) showed that, at least in some cases, leaf expansion could be controlled at the level of the leaf itself, and that it can be a driving force for cell proliferation in the leaf. There is a large debate on that subject in the literature and depending on the studies, cellular processes are reported or not as engines driving organ growth (Cookson et al. 2005; John and Qi 2008; Massonnet et al. 2011).

## **2.2. Impact of flowering transition on leaf area development and its cellular components**

Factors controlling the timing of flowering transition can also affect vegetative development. Numerous studies have shown that nearly all of the conditions that affect the timing of floral transition also alter the total number of leaves (Koornneef et al. 1991; Cookson et al. 2007; Méndez-Vigo et al. 2010; Franks 2011; Itoh and Shimizu 2012). It was then suggested that the leaf growth phenotype observed in late flowering genotypes could be due, at least partly, to the delay in flowering time itself. Comparing the dynamics of leaf production and expansion in a few mutants in the *Ler* background, *ROUNTUNDA2* mutant (*ron2-1*) was identified as the only one with an increased leaf size compared to its wild-type (Cnops et al. 2004; Cookson et al. 2005). The final size of the 6<sup>th</sup> leaf formed on the rosette was increased in *ron2-1* (Figure I-2 (a)) as previously observed on the three first leaves (Cnops et al. 2004). Overall, rosette leaf number was significantly increased in *ron2-1* in comparison with its wild-type (Figure I-2 (b)). Transferring plants to long day conditions during early development triggered a synchronization of flowering time in both genotypes and resulted in similar phenotypes in terms of leaf number and individual leaf areas (Figure I-2 (c) and (d)). Thus, even though there is no doubt that *ron2-1* mutation did affect individual leaf growth, it seemed to be an indirect effect primarily due to an effect on leaf production and/or floral transition.

An elegant demonstration of some of these interactions has been made by Willmann and Poethig (2011) using genetics to specifically and directly alter flowering time. The authors used the polymorphism of *FRIGIDA* (*FRI*) and *FLOWERING LOCUS C* (*FLC*) genes to induce a delay in flowering time. *FRI* is a positive regulator of *FLC*, *FLC* being itself an inhibitor of the floral inductors *FT* and *SOC1*. Therefore, if a plant genome contains functional alleles for both of these genes (*FRI;FLC*), a strong repression is applied on *FT* and *SOC1* and floral transition will occurred later than if one of the genes *FRI* and *FLC* is present as a non-functional allele (*fri;FLC* or *FRI;flc*). This study has highlighted that a delay in flowering time triggers an increase in epidermal cell area, in leaf size and in the final number

of rosette leaves. It has also shown that a delayed floral transition does not affect the onset of vegetative phase change but delays it. The duration of the adult vegetative phase was also extended. The same effects on the leaf area, duration of expansion and epidermal cell area in individual leaves have been shown when flowering time was delayed by a shortening of day-length (Cookson et al. 2007).

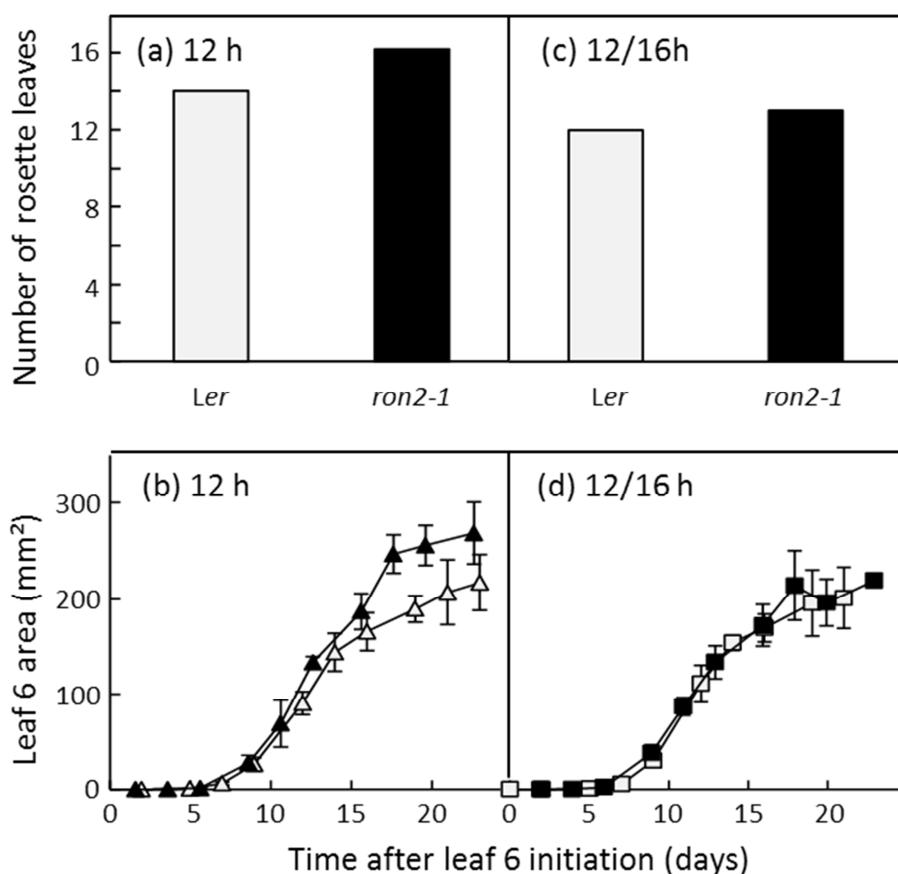


Figure I-2. Final number of rosette leaves of *Ler* (light grey) and *ron2-1* (black) plants grown under a constant day length of 12h (a) or transferred from 12 to 16 h after 15 days following leaf initiation (c). Changes with time in leaf 6 area of *Ler* (light grey) and *ron2-1* (black) grown under the photoperiods of 12h (b) or transferred from 12 to 16 h after 15 days following leaf initiation (d). Means with 95 % confidence intervals are shown (n = 5 plants per genotype and treatment). From Lièvre et al. 2013.

### 3. PRACTICES AND PITFALLS IN LEAF GROWTH PHENOTYPING

Frameworks of analysis as well as protocols to assess plant development are shared in the plant science community (Boyes et al. 2001; Cookson et al. 2010; Rymen et al. 2010).

Identifying the genetic control of shoot growth and the developmental regulators involved requires a precise characterization of growth phenotypes using robust descriptors of shoot

organ size and overall geometry, production and growth rates. Because shoot development is the sum of discrete morphogenetic structures that appear at regularly spaced positions at the shoot apical meristem flanks over time, spatial and temporal descriptors are commonly used to measure shoot growth. In most studies, morphological traits are recorded in adult plants but in a few cases, these datasets are completed with temporal indicators, i.e., specific time-points at which visible events contributing to final shoot morphology take place.

### **3.1. Interest of phenotyping leaf growth over time**

In large collections of natural variants or mutants, hundreds of genotypes have been classified according to leaf size, shape or number (Berná et al. 1999; Serrano-Cartagena et al. 1999; Juenger et al. 2005; Tisné et al. 2008; Massonnet et al. 2011; Pérez-Pérez et al. 2011). In some cases the authors drew conclusions on the genetic controls of ‘leaf growth’ or ‘leaf production’ from these datasets (Tisné et al. 2008; Ghandilyan et al. 2009). In these studies, leaf growth traits were measured at a given date after sowing or at a given developmental stage. However, several examples illustrate that a static ‘picture’ of a growth trait at a given date or stage does not necessarily reflect what will be the final value of this trait (Cockcroft et al. 2000). Genotypes with high initial relative leaf expansion rate have generally a shorter duration of leaf expansion (Cockcroft et al. 2000; Cookson et al. 2005). Conclusions drawn on traits measured at a given date cannot be generalized to the whole growing period and another analysis at another date or stage may lead to different conclusions, whatever the scale of growth analysis.

As illustrated in the section 1 of this chapter (Figure I-1 (a)), individual leaf expansion is a gradual but non-linear increase in area. Several examples highlighted that measuring final leaf area only can lead to the omission of underlying alterations in the dynamics of expansion. Indeed, alterations in final leaf area can be explained by changes in the duration of leaf expansion and/or the leaf expansion rate, and measuring these traits can give another dimension to phenotyping studies. It has been reported that environmental cues can affect individual leaf dynamics in several species. For instance, shading and reduced soil water content cause a decrease in the maximal leaf expansion rate whereas the duration of leaf expansion is either maintained or increased (Cookson et al. 2005; Pereyra-Irujo et al. 2008; Tisné et al. 2010). In a few examples, the increase in duration of leaf expansion compensates the decrease in maximal leaf expansion rate, in such a way that comparison of final leaf areas between the different environmental cues or genotypes does not show any alteration (Cookson

et al. 2006; Aguirrezabal et al. 2006). This is also illustrated in Figure I-1 (a), showing two genotypes (*Ler* and *elo-1*) reaching the same final leaf 6 area but with different dynamics of leaf expansion. In a similar way, tobacco (*Nicotiana tabacum*) plants overexpressing the D-type cyclin, CycD2, display no visible leaf phenotypic differences when observed after completion of vegetative growth, but only because they have higher rates and shorter durations of expansion (Cockcroft et al. 2000). In such cases, only the measurement of leaf area over time can highlight the incidence of the environmental change or of the mutation on leaf growth, hence their importance in order to collect information as complete as possible.

### **3.2. Phenotyping leaf growth over time: late differences can be due to early processes**

One limit to the comparison of leaf growth curves is due to the difficulty of measuring early stages of leaf expansion. Precisely measuring the number of initiated leaves requires the careful dissection of the rosettes using a binocular or a microscope at high magnification, removal of the cotyledons and subsequent leaves one by one with a scalpel until all primordia are visible (Cookson et al. 2010). Measuring the leaf initiation rate or the plastochron – the time interval between the emergence of two successive leaves, i.e. the reciprocal of leaf initiation rate – requires sequential destructive measurements over time increasing the number of replicates to be grown together in a single experiment (Cookson et al. 2010). These tasks are even more tedious in *A. thaliana* than in other plants due to its small size and, as a consequence, quantitative datasets with phenotypic characterization of these variables are often restricted to a limited number of genotypes and/or environmental conditions (Cookson et al. 2005; Skirycz et al. 2010; Vanhaeren et al. 2010).

Even though recent advances in microscopy and image analysis now allow the assessment of early stages of leaf growth from leaf initiation on the meristem until emergence, the throughput of these early measurements is typically low because of the technical constraints associated with destructive measurements and the necessity to grow a high number of plants together for sufficient replicates at each sampling time point (Cookson et al. 2005; Wuyts et al. 2010; Vanhaeren et al. 2010). At the whole plant scale, these methods have shown that subtle differences in shoot apical meristem volumes contribute to subsequent differences in leaf emergence rate and rosette expansion rate (Clark et al. 1995; Vanhaeren et al. 2010). Similarly, at the individual leaf scale, early changes in leaf expansion rate or changes in the initial size of the primordium impact the growth dynamics later on and can alter final leaf size (Granier and Tardieu 2009). This is due to the exponential behaviour of early growth phase in

which absolute leaf expansion rate at a given time depends on the leaf area at that time. It is then easy to misinterpret growth curves, as early developmental variation can cause misleading differences in late expansion (Arvidsson et al. 2011; Zhang et al. 2012).

As an example, at the whole plant scale, differences in rosette area observed at a given date after sowing in a set of genotypes (Figure I-3 (a)) are no longer apparent when the x-scale is expressed in ‘days after leaf 2 emergence’ revealing here that differences in leaf area observed throughout the growing period were due to differences in germination rates and/or early development (Figure I-3 (b)). This illustrates that different conclusions can be drawn if the time scale is normalized by developmental stages (Arvidsson et al. 2011; Zhang et al. 2012). It is important to keep in mind that phenotypic differences reported at a given date after sowing can be due to differences in germination rate.

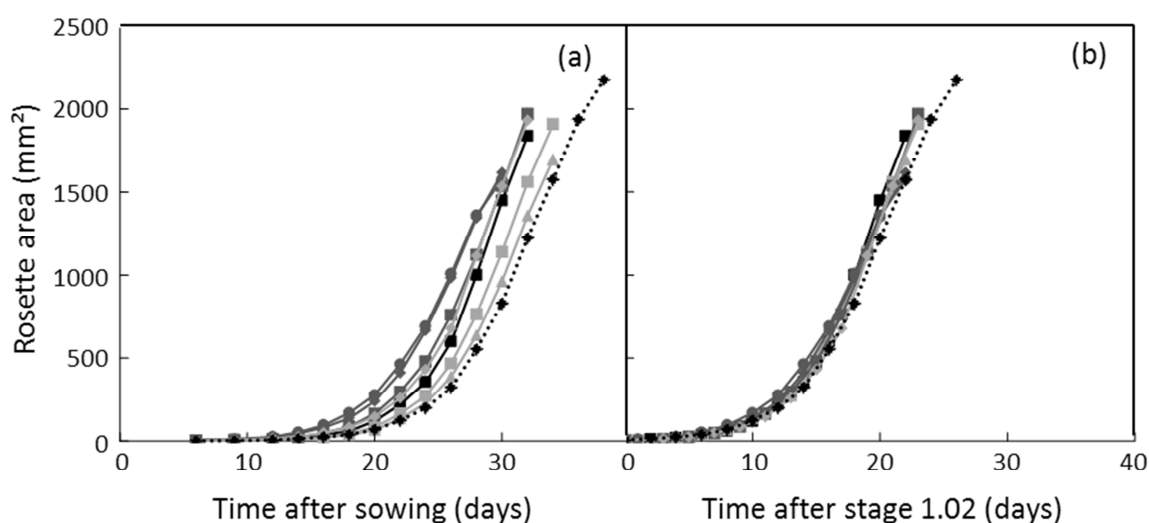


Figure I-3. Dynamics of whole rosette area over time expressed either as days after sowing (a) or days after stage 1.02 (b), i.e. when the second leaf emerged in the center of the rosette. Rosette area was determined on top view images automatically taken in the PHENOPSIS platform (see Figure II-1 (c)). Each point is the mean of 3 plants. From Lièvre et al. 2013.

### 3.3. Comparing growth of successive leaves along the shoot

In the majority of leaf growth phenotyping studies, conclusions on individual leaf growth are drawn from the comparisons of leaves of same rank, i.e. having the same position on the plant shoot (Figure II-1 (b)). However, as developed in this chapter, section 1, it is known that leaves initiated during different phases of the shoot development display different morphological characteristics, such as final leaf area, leaf blade length-to-width ratio, number and depth of serrations present on the leaf margin and pattern of trichomes covering. This phenomenon, called heteroblasty, is observed among the whole plant kingdom and can be



affected by genetics and/or environmental cues (Lawson and Poethig 1995; Kerstetter and Poethig 1998; Sylvester and Parker-Clark 2001; Poethig 2013; Matsoukas 2014). There are also examples showing that when dynamics of expansion are compared among successive leaves of a same plant, it appears that they grow at different rates and that the whole duration of their expansion may also differ (Dosio et al. 2003; Cookson et al. 2007). Those considerations question the relevance of comparing leaves of same rank between two plants that, because of environmental cues or genetic differences, display overall differences in their shoot development, such as in vegetative phase change and/or final number of rosette leaves.

Heteroblasty is often linked with the concept of juvenile and adult vegetative developmental phases in the plant. Although some of the morphological changes within *A. thaliana* shoot seem to occur gradually making the distinction between the different phases subtle, the formation of trichomes on the abaxial face of leaves has been identified as a relevant marker of the onset of the juvenile-to-adult phase transition (Chien and Sussex 1996; Telfer et al. 1997; Kerstetter and Poethig 1998). On *A. thaliana* leaves, trichomes are specialized cells that taking form of branched ‘hairs’. They are differentiated from a single epidermal cell. It is commonly admitted that the first rosette leaf having at least one trichome on its abaxial surface is the first one that was initiated after the onset of the vegetative phase change. Similarly, the first leaf whose abaxial face is fully and homogeneously covered with trichomes is considered as the first leaf initiated after the phase change was completed and the plant has entered the adult vegetative phase. By using this marker, the identity of each leaf of a rosette can easily be determined, that can give access to information concerning the different phases of a plant vegetative development (Figure I-4). These kinds of inputs are not commonly used in *A. thaliana* leaf growth phenotyping experiments, whereas it could be interesting to add this data, measured at the individual leaf scale but analyzed as a series along the shoot.

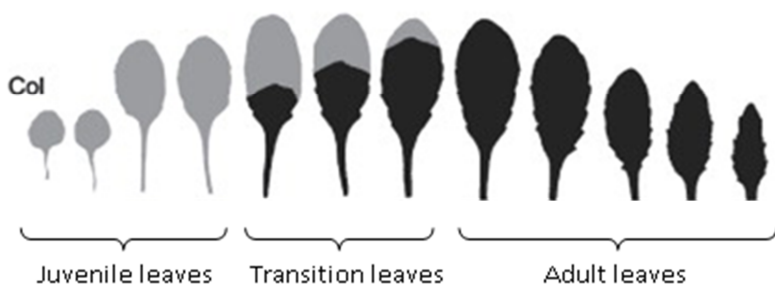


Figure I-4. Patterns of trichomes observed on the abaxial leaf surface of Col-0 plant. The identity of leaves is deduced from these observations. Grey: absence of abaxial trichomes. Black: presence of abaxial trichomes. Adapted from Hunter et al. 2003.

### 4. AIM AND STRATEGY OF THE PhD

As shown in this chapter, analyzing the plant growth phenotype is complex due to the existence of interactions between processes occurring at different scales, and due to a variation along the shoot of growth traits measured on individual leaves, that is linked to the global phase changes occurring at the shoot level. In this context, the aim of my PhD was to propose and test methods for analyzing such a complexity, considering traits at three different scales of the shoot system: tissue (leaf epidermis), organ (leaf), and shoot (leaf series), as well as their relationship to phase changes in shoot ontogeny. This required development of appropriate statistical methods for integrating different types of data – tissular, morphological, dimensional and dynamics. From a practical point of view, the main challenge of my PhD was to elaborate a functional pipeline of analysis methods for phenotyping plant growth at these three scales, and to help taking into account the effect of leaf position within the shoot.

Our first concern was to acquire proper datasets that were necessary for these analyses: tissular data on successive leaves of a plant, dynamics of leaf expansion of successive leaves on a plant, and morphological and dimensional traits giving insights into the leaf status. Despite the availability of an automated phenotyping platform in the host laboratory and hundreds of genotypes already grown in this platform since 2002, such a dataset was not available at the onset of this work. Consequently, the first part of my work was to acquire datasets at each scale of analyses, with enough replicates for further robust statistical analyses.

To this end, two first experiments were performed on genotypes with contrasted shoot development. Genotypes were selected, on their leaf-growth phenotype contrasts – without any knowledge on the underlying genes controlling these leaf growth differences. We first present a detailed analysis of these genotypes, considering separately the different leaves and the growth variables at the different macroscopic scales, to highlight the contrasts of leaf developmental features (Chapter II).

Subsequently, we built integrated models that allowed a global detection of plant phase changes using all measurements made at different scales, leading to general considerations regarding this matter. Analyzing successions of developmental phases and changes of leaf growth traits within each of these phases allowed us to characterize the effect on shoot development of the selected genotypes mutations (Chapter III).

We then focused on the development of a new pipeline of analysis of tissular data that enables detecting populations of cells in the epidermis. This pipeline combined a semi-

automatic method for segmenting leaf epidermis imprints and gamma mixture model used to analyze the obtained cell area distributions (Chapter IV).

Finally, we used one genotype dataset to show how the tissular data that we obtained can be integrated to the multi-scale analysis method. It finalized the integration of the data from the cell to the shoot scale (Chapter V).

## **CHAPTER II. USE AND LIMITS OF EXISTING PIPELINE OF ANALYSES FOR COMPARING LEAF GROWTH PHENOTYPES IN *A. THALIANA*: A TEST ON 90 T-DNA INSERTION LINES**

### **1. INTRODUCTION**

*Arabidopsis thaliana* (*A. thaliana*) offers important advantages for researches in genetics and molecular biology by its relatively small genome size compared to crops (Meinke et al. 1998; AGI 2000). Maize, for example, has a genome of approximately 2,400 Megabase pairs (Mbp), i.e. around 19 times the size of the *A. thaliana* genome while the wheat genome is 16,000 Mbp, i.e. 128 times larger than the *A. thaliana* one. Since the completion of the *A. thaliana* genome sequence, extensive genetic maps of all 5 chromosomes, together with efficient methods for mutagenesis and plant transformation have delivered a large range of genetic and genomic resources to the plant biology research community (Alonso et al. 2003; O'Malley and Ecker 2010). Different sources of genetic diversity are exploited in *A. thaliana* to identify genes controlling many processes involved in plant growth and functioning (see following listing). In many examples, they have allowed to identify genes controlling the development of plant leaf area.

- Natural lines corresponding to different ecotypes with natural variation have been collected by many groups and have been used for Genome Wide Association Mapping (Atwell et al. 2010).
- Recombinant inbred or near isogenic populations have been created by crosses of different ecotypes to produce genetically structured populations that are intensively used to map quantitative trait loci and/or genes controlling shoot growth and developmental processes (Tisné et al. 2008; Méndez-Vigo et al. 2010).
- Mutated lines have been created from wild-type genetic backgrounds in which specific individual genes or small groups of 'candidate' genes supposed to be involved in processes associated to leaf growth control are altered via inactivation or over-expression (Cookson et al. 2005; Massonnet et al. 2011).
- Mutated lines have been created from a same reference genetic background in which genes or small groups of genes are altered randomly at random positions along the whole genome (Berna et al. 1999).

For our study, we took advantage of the recently distributed collection of T-DNA homozygous *A. thaliana* mutant lines, originating from the SALK institute. This collection was generated by the group of Prof. J.R. Ecker (Alonso et al. 2003). At the onset of our work, a preliminary high-throughput leaf growth screen was performed in Jose Luis Micol group in Alicante (Elche University). The objective was to systematically search for mutant leaf phenotypes in the whole collection by a simple screening, looking for easily identifiable leaf phenotypes. Over 14,000 lines representing 10,846 genes have been analyzed so far leading to the identification of 255 leaf mutants, including genotypes with altered leaf shape, leaf color and/or leaf area compared to the wild-type. Interestingly, only 15 % of the genes thus identified had been previously described in the literature. This genetic material represents a mine of new information concerning the genetic control of leaf growth. During this first screening, plants were grown *in vitro*. Leaf phenotype was scored 18 days after sowing (DAS). Qualitative criteria to identify lines with unambiguous different leaf phenotypes compared to the wild-type were: the size of the rosettes compared to the wild-type Columbia (Col-0) – i.e. bigger/smaller – and the visible differences in leaf color and/or shape (n = 30); see Wilson-Sánchez et al. 2014 and poster by Micol et al. (International Conference on Arabidopsis Research, ICAR 2012) in Appendix VIII.

Our first objective at the onset of this work was to select a few *A. thaliana* genotypes with contrasted leaf development from the experiment performed in Jose Luis Micol group to develop our pipeline of analyses. Two preliminary experiments (thereafter referred as Experiments 1 and 2) were performed using routine phenotyping methods developed in LEPSE (Cookson et al. 2010) on 90 SALK mutants (issued from the 255 leaf mutants identified by Jose Luis Micol group). Different leaf growth variables were measured 18 days after sowing (as done in Jose Luis Micol group experiment) and at flowering (as usually done in LEPSE experiments) to identify genotypes with contrasted leaf production and/or leaf expansion. The main results of these experiments were presented at the ICAR 2012 (see poster by Lièvre et al. in Appendix VII). Afterwards, a subset of 4 contrasted genotypes was selected from Experiments 1 and 2, for a more in depth study (Experiment 3) with more replicates per genotype. Additional leaf growth traits were measured during this experiment, in order to obtain a more complete overview of the plants shoot growth. As it has been reported in angiosperms, *A. thaliana* shoot development involves at least three distinct post-embryonic phases: a reproductively incompetent phase (i.e. juvenile vegetative phase), a reproductively competent (i.e. adult vegetative phase) and a reproductive phase (Poethig 2003). The juvenile-to-adult transition (vegetative phase change) involves changes in several

leaf traits such as patterns of abaxial trichomes production, number of serrations, length-to-width ratio of the leaf blade and size of the petiole and leaf blade (Röbbelen 1957; Telfer et al. 1997; Steynen et al. 2001; Méndez-Vigo et al. 2010). The presence of trichomes on the abaxial face of the leaf is the clearest marker of the onset of the vegetative phase change, so we added this trait to the growth measurements performed in Experiment 3. To complete the overview of *A. thaliana* shoot growth, individual leaf expansion curves were also measured.

This chapter presents the multi-scale phenotyping analyses performed on the 90 SALK T-DNA lines and their wild-type, the selection of 4 contrasted genotypes and the more detailed phenotypical analysis of these 4 genotypes which constitute the baseline dataset used in Chapter III for the development of our pipeline of analysis methods.

## **2. MATERIAL AND METHODS**

### **2.1. Biological material**

The 90 T-DNA insertion lines used in the Experiments 1 and 2 are listed in Appendix I. Seeds used were kind gifts from José Luis Micol group. All of them are SALK mutants issued from the Columbia 0 (Col-0) genetic background. They are homozygous lines as checked by José Luis Micol group.

### **2.2. Growth conditions**

The 3 experiments were performed in a growth chamber equipped with the PHENOPSIS automaton (Granier et al. 2006), under controlled air temperature, air humidity and incident light. For the 2 first experiments, 14 plants of each line (SALK lines and Col-0) were grown together. Experiment 3 was designed to build solid analysis models from the generated data. For this purpose, the number of replicates per genotype was increased up to 60 plants. Seeds were sown in pots filled with a mixture (1:1) of a loamy soil and organic compost. Micrometeorological conditions were kept constant during the whole duration of the experiments and homogeneous within the growth chamber (Granier et al. 2006). Light was on during 16 h per day and provided by a bank of cool-white fluorescent tubes and HQi lamps. It was measured at the plant level using a photosynthetic photon flux density (PPFD) sensor (LI-190SB, Li-Cor, Lincoln, NE, USA). Mean micro-meteorological conditions were constant over the whole duration of experiments and are indicated in Table II-1. Soil water content was determined before sowing to estimate the amount of dry soil in each pot. Subsequent changes

in pot weight were due to changes in soil water content (the plant weight was considered as negligible as it was never higher than 1 g). This allowed the computation and daily automatic adjustment of soil water content to 0.35 g water g<sup>-1</sup> dry soil by the PHENOPSIS automaton (Granier et al. 2006) with a modified one-tenth-strength Hoagland solution (Hoagland and Arnon 1950), from sowing to the end of the experiments.

|  | Experiment 1 | Experiment 2 | Experiment 3 |
|--|--------------|--------------|--------------|
| Day length (h)   | 16           | 16           | 16           |
| Incident PPFD ( $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$ )  | 176          | 175          | 178          |
| Air temperature (°C)   | 21.9         | 22.2         | 20.4         |
| Air humidity (%)   | 61.4         | 59.6         | 60.6         |
| Soil water content (g H <sub>2</sub> O g <sup>-1</sup> dry soil) | 0.35         | 0.35         | 0.35         |

Table II-1. Meteorological conditions applied during the experiments.

### 2.3. Growth measurements

During the three experiments, daily zenithal pictures were taken by the PHENOPSIS automaton (Figure II-1 (b)) during rosette growth.

For the first two experiments, 7 plants of each genotype were harvested 18 days after sowing whereas the remaining 7 others were harvested at a common stage of plant development, stage 6.00, i.e. first flower open (as defined in Boyes et al. 2001). All leaves that formed the rosette had stopped their expansion at this stage.

At harvesting, rosette fresh weight (without root system or flowering axes) was measured (mg). Then, successive true leaves of the rosette (cotyledons were excluded) were excised without their petiole, stuck on a sheet of paper and scanned as shown in Figure II-1 (a). Individual leaf areas (mm<sup>2</sup>) were automatically measured with image analysis software ImageJ (Rasband , National Institutes of Health, USA). Final rosette area (mm<sup>2</sup>) was computed as the sum of individual leaf areas. The number of rosette leaves was estimated by counting the number of true leaves produced after the production of the cotyledons until the emergence of the primary flowering axis. In Experiments 1 and 2, duration of the vegetative phase was determined as the time elapsed between the sowing date and the emergence of the inflorescence, i.e. the bolting stage. This stage was determined by naked-eye on daily zenithal images taken by the PHENOPSIS automaton. In Experiment 3, in order to overcome possible differences in germination time, duration of the vegetative phase was determined as the time elapsed between the apparition of the first true pair of leaves and the apparition of the last

rosette leaf. Mean leaf apparition rate ( $\text{leaf day}^{-1}$ ) was computed as the final number of rosette leaves divided by the duration of vegetative phase.

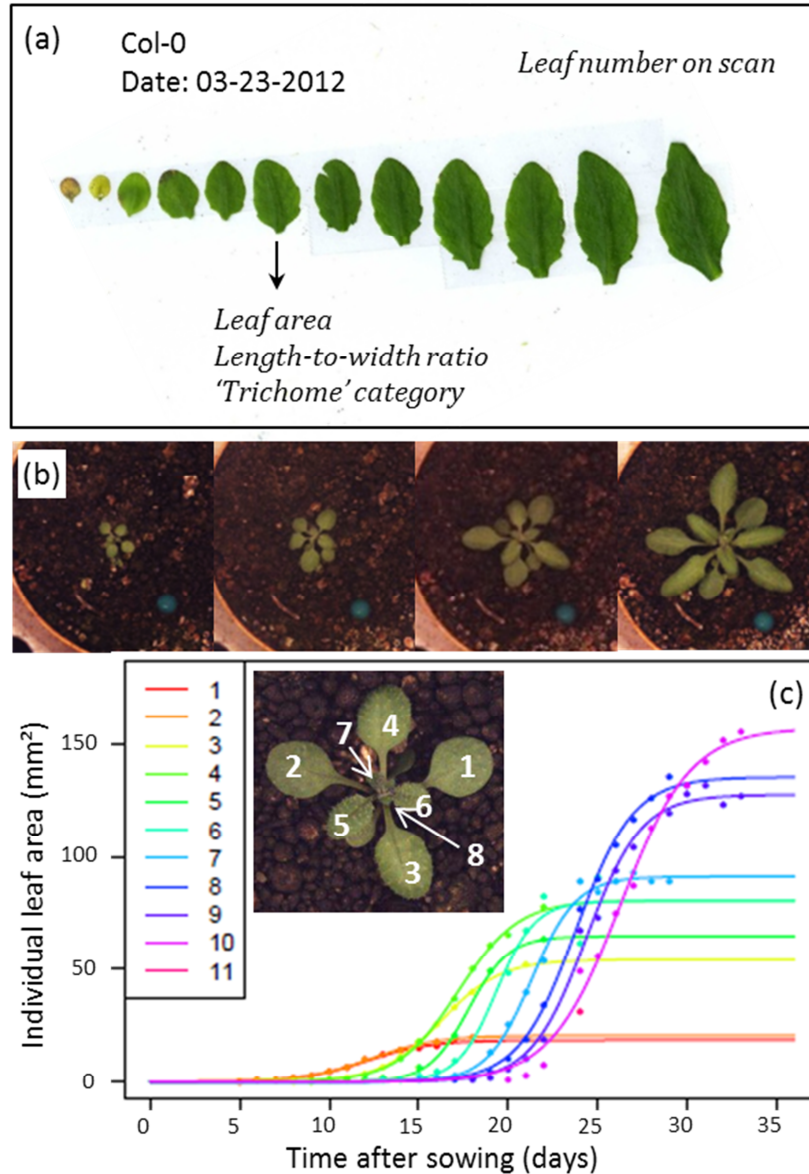


Figure II-1. Different types of images taken during the experiments are shown with associated leaf growth traits. (a) Scan of a dissected rosette after harvest, with different growth traits measured (destructive measurements). (b) Zenithal images taken by the PHENOPSIS automaton. (c) Changes in individual leaf areas can be measured on zenithal images (non-destructive measurements). When changes in individual leaf areas are plotted over time, sigmoids can be fitted on these data. Inset: close-up of a zenithal image of a rosette with respective leaf ranks (indicated per order of apparition).



Additional measurements were performed on the plants grown during Experiment 3. In addition to individual leaf area, maximal leaf blade length (mm) and width (mm) were automatically measured on the leaf scans using the image analysis software ImageJ 1.43C and blade length-to-width ratio was computed. In addition leaf abaxial surface was observed under a binocular (X10) for each successive leaves just before they were scanned to identify those with or without trichomes. Rosette leaves were divided into three categories depending on the level of covering of their abaxial surface by trichomes. Based on the literature (Telfer et al. 1997; Bollman et al. 2003), leaves without abaxial trichomes were classified as juvenile leaves, leaves with trichomes (at least one) that did not entirely span their abaxial surface as transition leaves (trichomes were not distributed homogeneously over the lamina but mainly present at the base of the leaf), and leaves with trichomes that span their entire abaxial surface (trichomes distributed homogeneously and at least one within 2 mm of the leaf margin of the distal tip) as adult leaves. To limit possible stochastic effects, the subsequent leaves on the shoot had to meet the same criterion or to correspond to the next leaf category.

Projected individual leaf areas were measured using image analysis software ImageJ on the daily zenithal pictures of the rosette taken by the PHENOPSIS automaton by drawing the visible part of leaf blade. Petioles were not included in the measurement. In case of overlap between leaves making a leaf contour too uncertain to be drawn, or when a leaf had revolute margins (phenomenon of leaf curling) at a point where about 50 % or more of the leaf area was hidden, the leaf was not measured. The measured areas were used to estimate individual leaf expansion curves. To do this estimation we chose a nonlinear logistic regression model as the curves seemed to have a sigmoidal shape with a certain symmetry at the inflection point (Torres and Frutos 1989). The parametric sigmoid function used was

$$Y = \frac{A}{1 + \exp\{-(t - M) / B\}}$$

where  $A$  is the value of the sigmoid curve plateau – i.e. the estimated final value of the leaf area;  $B$  is a characteristic growth time parameter (inverse of growth rate); and  $M$  is the time corresponding to the inflection point of the sigmoid curve (time of maximum growth). Figure II-1 (c) shows a collection of estimated logistic functions for successive leaves from leaf areas measured during the plant growth. However, because of overlapping and leaf curling phenomena, the estimated final leaf area was frequently under-estimated in comparison to the measurement made when the plants were harvested and the leaves separated and stuck on a sheet of paper for scan measurements. This bias increased with leaf rank (because the last

leaves emerged were the most revolute) and with leaf expansion (increase of the overlap between leaves). To our knowledge, there is no model enabling to this bias. Our objective was not to study precisely leaf expansion kinetics but rather to compare expansion curves; we therefore identified a method to compute a corrected dynamical variable that would enable a relevant analysis of leaf expansion kinetics from a comparative point of view. We assumed that the biased estimation of the final leaf surface within the logistic growth model does not affect substantially the estimation of the characteristic growth duration or of the time of maximum growth. It allowed us to compute each leaf maximum absolute leaf expansion rate ( $LER_{max}$ ,  $mm^2 day^{-1}$ ) using the characteristic growth duration parameter  $B$  estimated using leaf areas measured on zenithal rosette pictures and the corresponding final leaf area obtained by dissecting the rosette after harvest (unbiased final leaf area,  $LA$ ):

$$LER_{max} = LA/4B$$

#### 2.4. Statistical analyses

All statistical analyses were done using R software (R Core Team 2014). The normality of each variable distribution was evaluated for each genotype using Shapiro-Wilk test. Homogeneity of variances between genotypes was tested for each variable using Levene tests. Differences between Columbia 0 and mutant lines computed means were tested for each variable using Student t-tests. Linear regression between leaf growth variables were estimated on the whole data. Pearson correlation coefficients were computed and their significance was tested using the associated R function.

### 3. RESULTS

#### 3.1. Lines with contrasted leaf phenotypes could not be identified robustly when leaf growth variables were compared 18 days after sowing

When the set of 91 genotypes grown together was classified according to total rosette area or number of leaves 18 days after sowing (Figure II-2 (a) and (b)), the ranking did not reflect what was observed at flowering when rosette development was completed (Figure II-2 (c) and (d)). Interestingly, in this dataset a negative non-linear correlation was found between rosette area measured 18 days after sowing and the duration of leaf production estimated as the number of days between the sowing date and bolting (Figure II-2 (e)). The latter stage is defined as the emergence of the inflorescence in the center of the rosette, just after the last rosette leaf has emerged (Boyes et al. 2001). This illustrates that rosettes expanding rapidly

early in development have a tendency to stop producing leaves and transit to the reproductive phase earlier than those that expand more slowly in the beginning of their development (Figure II-2 (e)). They therefore do not necessarily have a larger final leaf area (Figure II-2 (d)).

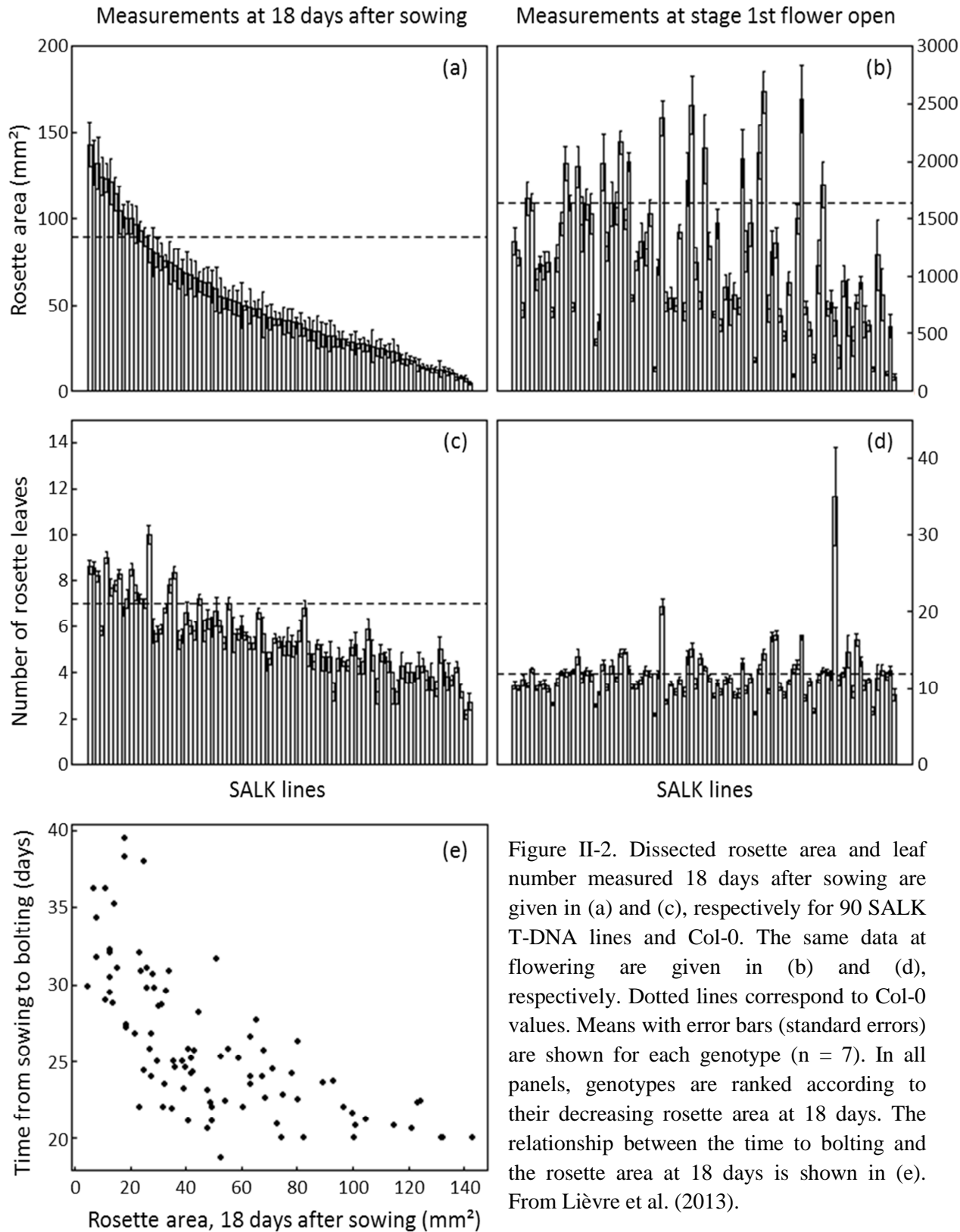


Figure II-2. Dissected rosette area and leaf number measured 18 days after sowing are given in (a) and (c), respectively for 90 SALK T-DNA lines and Col-0. The same data at flowering are given in (b) and (d), respectively. Dotted lines correspond to Col-0 values. Means with error bars (standard errors) are shown for each genotype ( $n = 7$ ). In all panels, genotypes are ranked according to their decreasing rosette area at 18 days. The relationship between the time to bolting and the rosette area at 18 days is shown in (e). From Lièvre et al. (2013).

### **3.2. Many lines were affected on the different growth variables measured at the opening of the first flower when compared to the wild-type**

Fresh weight and rosette area were affected in the same way for the majority of the lines. Fresh weight of the rosette, measured at flowering, significantly differed from that of the wild-type (Col-0) in 66 lines. It was increased in only 7 of them and smaller in the remaining 59 (Figure II-3 (a)). The results obtained for final rosette area were almost the same: it was significantly larger in the same 7 genotypes and smaller in the remaining 58 lines for which significant differences from Col-0 rosette area was found (Figure II-3 (b)). Two T-DNA lines had significant differences in comparison to Col-0 fresh weight value but not in rosette area data (SALK\_145203 and SALK\_122867). On the contrary, only 1 line exhibited a significantly smaller rosette area without having a decrease in rosette fresh weight (SALK\_024759). Figure II-4 (a) shows the strong correlation found between these two variables.

Final leaf 6 area was affected in a fewer lines than the above-described traits: 52 lines exhibited a significant difference in comparison with Col-0. Three genotypes displayed a larger leaf 6 area, whereas it was reduced for the other 49 (Figure II-3 (c)).

The final number of rosette leaves was the less affected among the analyzed traits. Significant differences with Col-0 mean number of rosette leaves (12) were found for only 44 T-DNA lines, 12 of which had an increased number of leaves. Among these 12 genotypes, 8 also exhibited larger rosette area, whereas 29 of the 32 lines with reduced number of leaves also had decreased final rosette area. It illustrates the positive correlation existing between these two traits (Figure II-4 (c) and (d)).

Three T-DNA lines presented a compensation phenomenon between significant alterations in their leaf 6 area and number of rosette leaves, leading to a final rosette area showing no significant difference compared to Col-0 one. Two of them (SALK\_086630 and SALK\_129352) displayed smaller leaf 6 area coupled with an increased number of rosette leaves, whereas the other one (SALK\_018664) had on the contrary larger individual leaf area and a lower final number of leaves.

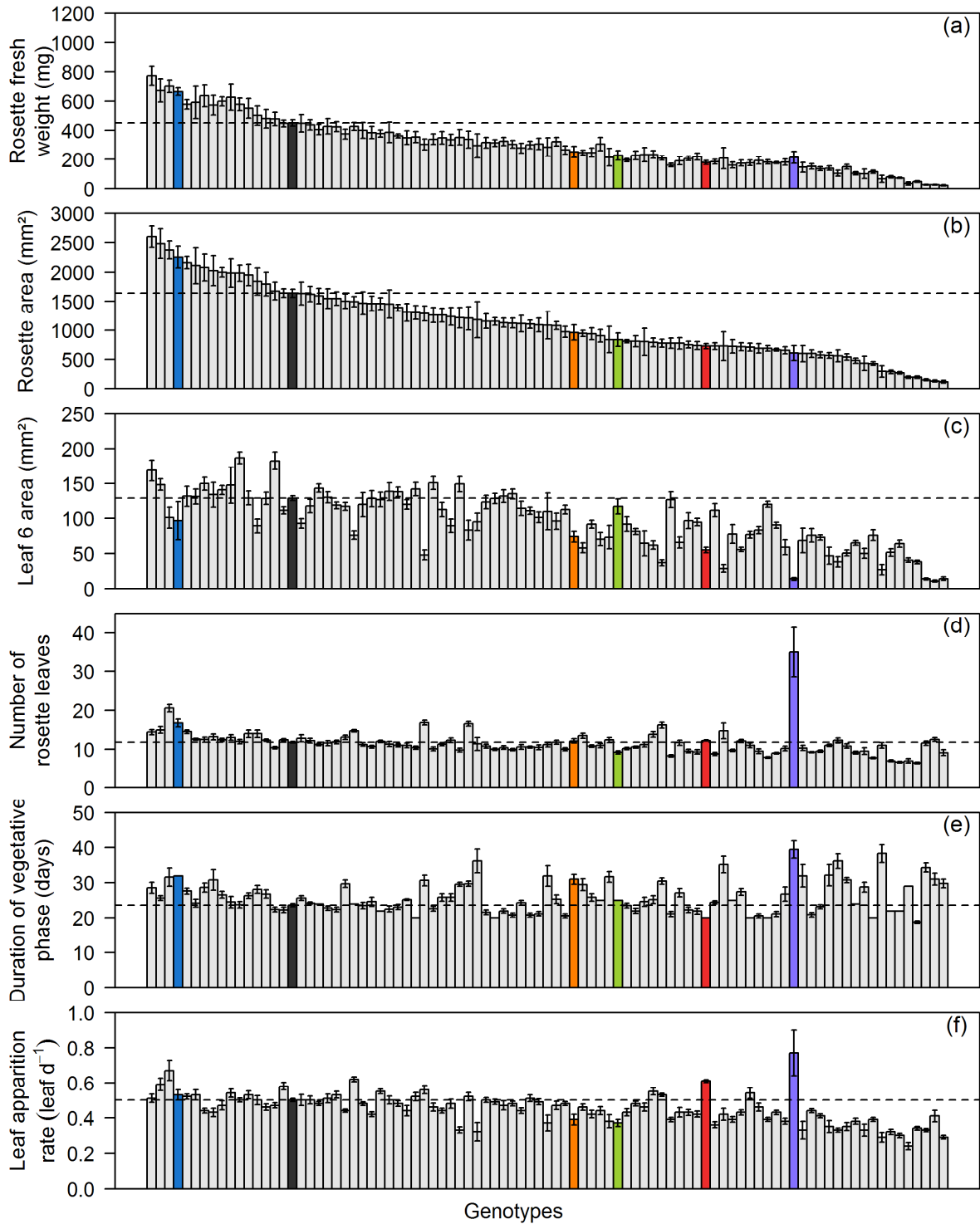


Figure II-3. Fresh weight of the rosette (mg; (a)), dissected rosette area (mm²; (b)), leaf 6 area (mm²; (c)), number of rosette leaves (d), duration of vegetative phase (days after sowing; (e)) and mean leaf apporation rate (leaf day<sup>-1</sup>; (f)) are given for 90 SALK T-DNA lines and Col-0 grown during Experiments 1 and 2. The first 4 traits were measured at flowering whereas the two others are the components of leaf apporation dynamics and were computed over time (see material and methods). Dotted lines and dark gray bars correspond to Col-0 values. Means with standard deviation are shown for each genotype (n = 7). In all panels, genotypes are ranked according to their decreasing rosette area at flowering. Five lines that have an interest for our in depth study are identified by colored bars: SALK\_025730, blue; SALK\_055458, orange; SALK\_048174, green; SALK\_126071, red; SALK\_064915, purple.

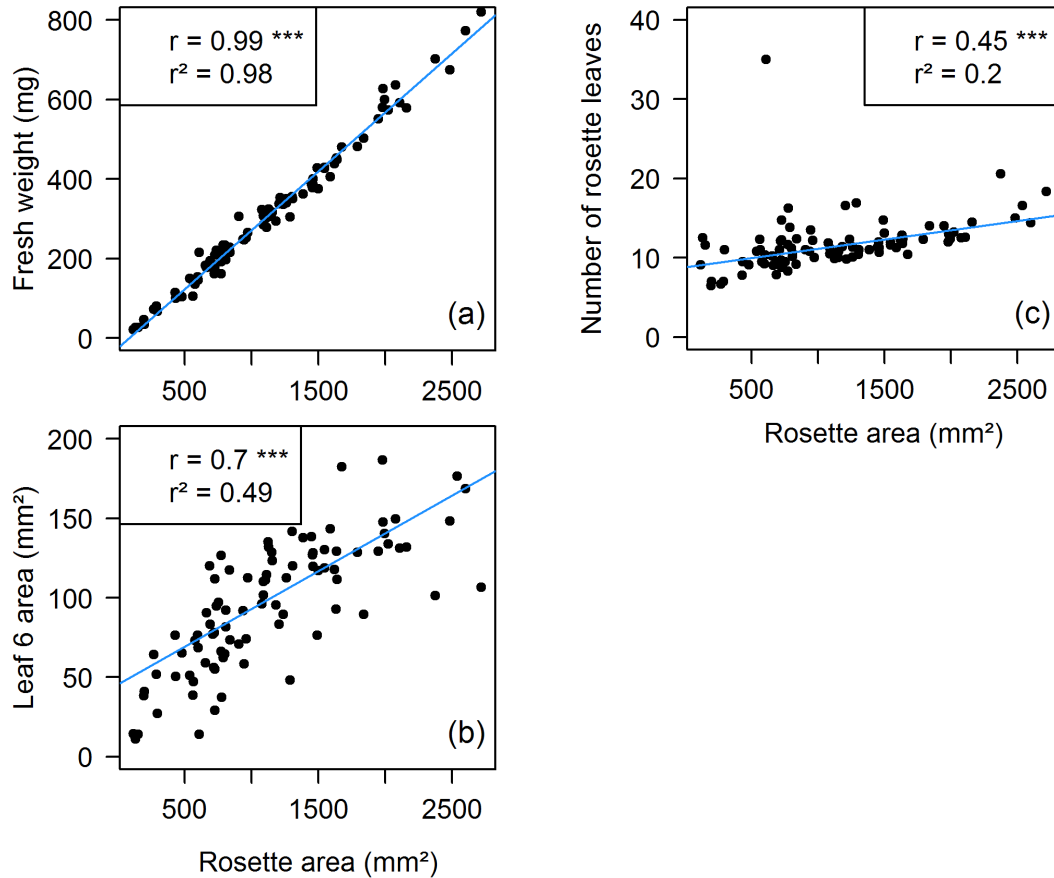


Figure II-4. Linear relationships between rosette area (mm<sup>2</sup>) and (a) fresh weight (mg) of the rosette; (b) leaf 6 area (mm<sup>2</sup>); (c) number of rosette leaves with all T-DNA lines. Correlations were estimated using the values measured at flowering on the plants grown during Experiments 1 and 2. For greater readability, on the graphs each point represents the mean data for a genotype ( $n = 7$ ). The linear relationships estimated between the variables are represented by a blue line.  $r$ , Pearson correlation coefficient.  $r^2$ , coefficient of determination. \*\*\*,  $p$ -value  $< 0.001$ .

Concerning the dynamics of leaf apparition, more lines were affected in duration of vegetative phase than in mean rate of leaf apparition (respectively 57 and 49 with significant differences in comparison to Col-0, Figure II-3 (e) and (f)). The duration of vegetative phase was mostly extended compared to Col-0 (38 lines) whereas mean leaf apparition rate was mainly reduced in altered T-DNA lines compared to Col-0 (42 lines).

Opposite alterations in leaf apparition dynamics led to a compensation phenomenon regarding the final number of rosette leaves, which was no significantly different compared to Col-0 for 18 lines that yet exhibited significant changes in their leaf apparition dynamics. For examples, see SALK\_009798 or SALK\_136507 for extended vegetative phase with lower leaf apparition rate. SALK\_126071 and SALK\_145203 are the only two lines showing the opposite type of compensation.

### 3.3. Selection of 4 lines with contrasted leaf growth phenotypes for a more in depth phenotypical study

One specific line, SALK\_064915 (purple bars on Figure II-3), particularly differs from the others, with the strongest alterations in leaf growth variables, of which the highest increases in both duration of vegetative phase and mean leaf apparition rate, even if the later was not significant. These changes in leaf apparition dynamics resulted in a mean final number of rosette leaves largely higher than the one scored in Col-0 (respectively 35 and 12). This line also presents the most drastic decrease in leaf 6 area, with a final leaf 6 area that reaches only 11 % of Col-0 leaf 6 area. Despite its high number of leaves and because its low individual leaf area this mutant final rosette area reaches 37 % of Col-0 final rosette area. Consequently, SALK\_064915 plants had extremely shrunken and compact rosettes with high numbers of senescent leaves on the lower metamers (Figure II-5). The dissection of the rosette at flowering and the identification of leaf ranks on the rosette were very laborious and, subsequently, the line unsuitable for the rest of our study.

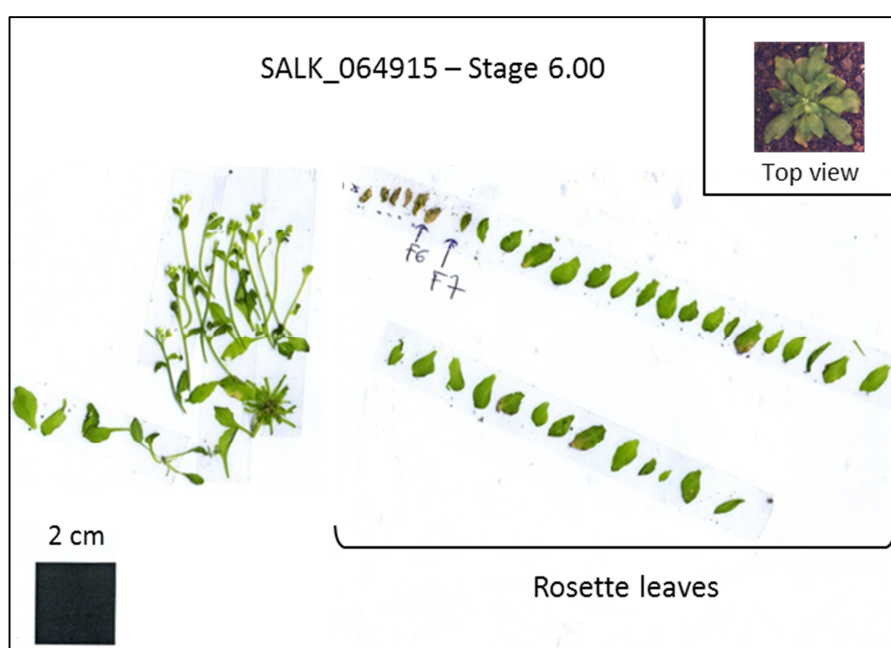


Figure II-5. Scan of the dissected rosette of a SALK\_064915 plant grown during Experiment 1 and harvested at flowering. Insert: zenithal picture of the same plant taken by the PHENOPSIS automaton 37 days after the apparition of the first two true leaves.

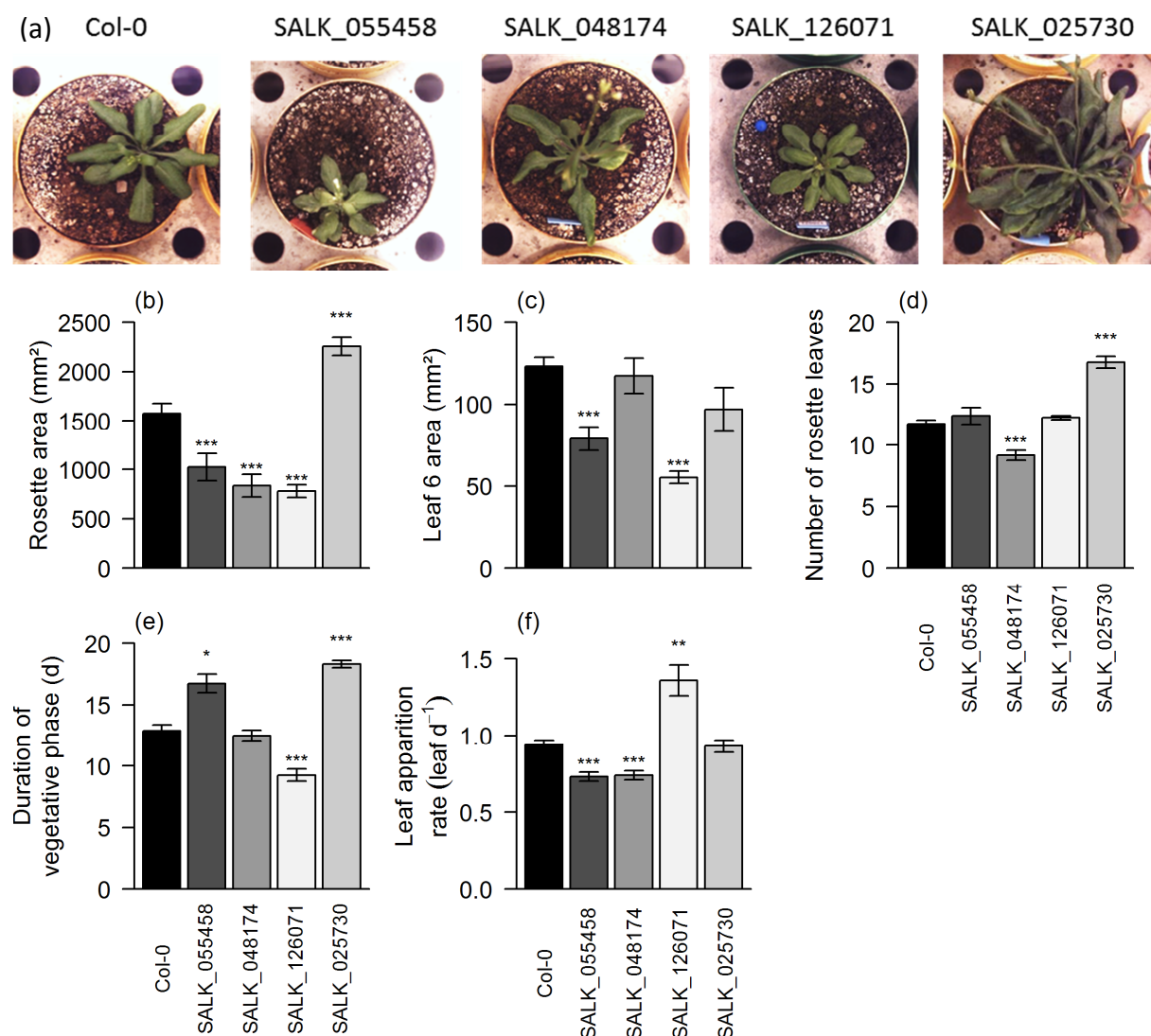


Figure II-6. (a) Zenithal pictures of a characteristic rosette of each line taken by the PHENOPSIS automaton just before flowering. (b) to (f) Results of leaf growth measurements obtained from Experiments 1 and 2 for Col-0 and the 4 T-DNA insertion lines selected as showing contrasted leaf growth phenotype in comparison with their Col-0 wild-type. The variables shown are: (b) rosette area (mm<sup>2</sup>), (c) leaf 6 area (mm<sup>2</sup>), (d) number of rosette leaves, (e) time to bolting (days after sowing) and (f) mean leaf apparition rate (leaf day<sup>-1</sup>). The first 4 traits were measured at flowering whereas the two others are the components of leaf apparition dynamics and were computed over time (see material and methods). Means with standard errors are shown for each genotype (n = 7). Stars indicate when the T-DNA lines value is significantly different compared to the one scored in Col-0 (Student t-tests): \*, 0.01 < p-value < 0.05; \*\*, 0.001 < p-value < 0.01; \*\*\*, p-value < 0.001.

Four genotypes were selected from Experiments 1 and 2: SALK\_025730, SALK\_048174, SALK\_055458 and SALK\_126071 (Figure II-6 (a)). They were chosen because they all exhibited strongly contrasted leaf growth phenotype in comparison to the Col-0 wild-type when measured at flowering. We selected them trying to focus on T-DNA insertions that do not affect every measured trait in order to minimize the number of interactions between leaf-



growth processes that could be affected by the mutations. Differences described in the following paragraphs are all significant according to the Student t-tests that were performed on the dataset.

SALK\_025730 was one of the rare lines with a larger final dissected rosette area than Col-0 one (Figure II-6 (b)). This increase of final rosette area was explained by a larger number of rosette leaves and no significantly altered individual leaf areas (respectively Figure II-6 (b) and (d)). This larger number of rosette leaves was due to a longer duration of leaf apparition (Figure II-6 (f)).

In the three other lines, rosette areas were significantly smaller than Col-0 (Figure II-6 (b)). For SALK\_048174, it was due to a shorter number of rosette leaves, triggered by a lower mean rate of leaf apparition (respectively Figure II-6 (c) and (f)).

The two other lines had no significant differences concerning the number of rosette leaves but exhibited smaller final leaf 6 areas when compared to Col-0 (Figure II-6 (d)). They were also altered in the leaf apparition dynamics, but for each one the alterations of the two underlying variables, duration and rate, compensated each other: SALK\_055458 plants had a longer duration of leaf apparition than Col-0 but a lower mean leaf apparition rate, whereas SALK\_126071 exhibited the opposite phenotype – i.e. a shorter duration of leaf apparition and a higher mean leaf apparition rate (Figure II-6 (e) and (f)).

### **3.4. In depth phenotyping analysis of the 4 contrasted lines**

#### ***3.4.1. Consistency of leaf growth variables measured on the 4 contrasted lines during the different experiments***

The SALK\_025730 line was first identified as a genotype of interest and selected for more in depth phenotyping analyses (see section above). However, it was excluded of further analyses because of its complex rosette architecture. Its successive leaves largely overlapped and were very revolute leaves. Additional measurements, particularly estimation of individual leaf expansion kinetics from zenithal images, could thus not be done accurately on this line. Therefore, this line was grown and imaged in Experiment 3 but finally not used for further in depth phenotyping analyses. Only 3 lines were kept for the complete in-depth analysis.

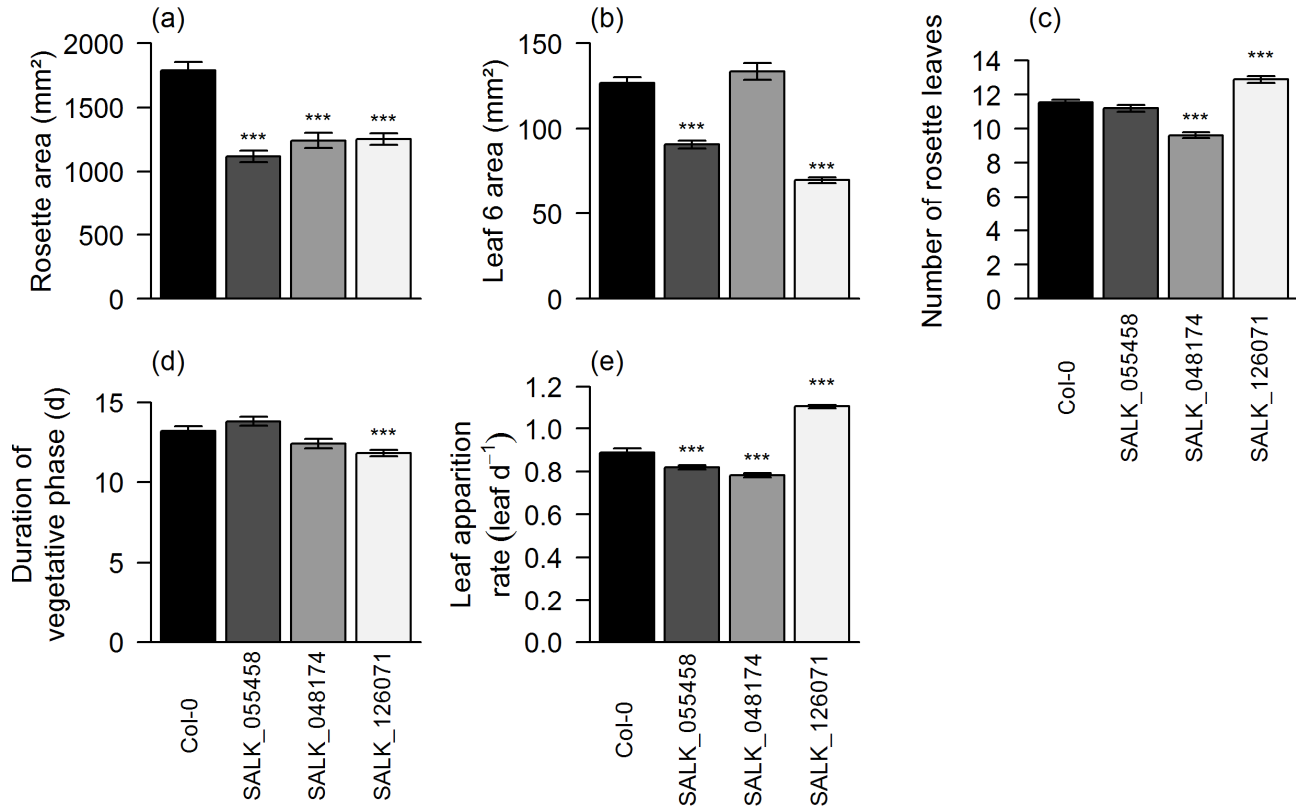


Figure II-7. Histograms of leaf growth measurements obtained from Experiment 3 for Col-0 and the 3 T-DNA insertion lines selected. The variables shown are: (a) dissected rosette area (mm<sup>2</sup>), (b) leaf 6 area (mm<sup>2</sup>), (c) number of rosette leaves, (d) duration of the vegetative phase (days after apparition of the first two leaves) and (e) mean leaf apperition rate (leaf.day<sup>-1</sup>), all measured at flowering. Means  $\pm$  SE (n  $\sim$  50). Stars indicate when the T-DNA lines value is significantly different when compared to Col-0 (Student t-tests): \*, 0.01 < p-value < 0.05; \*\*, 0.001 < p-value < 0.01; \*\*\*, p-value < 0.001.

| Genotype    | Rosette area (mm <sup>2</sup> ) |     | Leaf 6 area (mm <sup>2</sup> ) |     | Number of rosette leaves |      | Vegetative phase (days) |      | Leaf apperition rate (leaf day <sup>-1</sup> ) |     |
|-------------|---------------------------------|-----|--------------------------------|-----|--------------------------|------|-------------------------|------|--|-----|
|             | Mean                            |     | Mean                           |     | Mean                     |      | Mean                    |      | Mean   |     |
| Col-0       | 1790                            |     | 127                            |     | 11.5                     |      | 13.2                    |      | 0.89   |     |
| SALK_055458 | 1118                            | *** | 91                             | *** | 11.2                     | n.s. | 13.8                    | n.s. | 0.82   | *** |
| SALK_048174 | 1242                            | *** | 134                            | *** | 9.6                      | ***  | 12.4                    | n.s. | 0.78   | *** |
| SALK_126071 | 1252                            | *** | 70                             | *** | 12.9                     | ***  | 11.8                    | ***  | 1.1  | *** |

Table II-2. Results of leaf growth measurements obtained from Experiment 3 for Col-0 and the 3 T-DNA insertion lines selected as showing contrasted leaf growth phenotype in comparison with their wild-type Col-0. The variables shown are: (a) dissected rosette area (mm<sup>2</sup>), (b) leaf 6 area (mm<sup>2</sup>), (c) number of rosette leaves, (d) duration of the vegetative phase (days after apparition of the first two leaves) and (e) mean leaf apperition rate (leaf.day<sup>-1</sup>). Values are means computed for each genotype (n  $\sim$  50). Stars indicate when the T-DNA lines value is significantly different compared to Col-0 (Student t-tests): \*, 0.01 < p-value < 0.05; \*\*, 0.001 < p-value < 0.01; \*\*\*, p-value < 0.001.

Regarding the other genotypes selected, the results obtained in Experiment 3 were mostly consistent with the ones made during Experiments 1 and 2 (Figure II-6 and Figure II-7). For the 3 T-DNA lines final rosette areas were still significantly smaller than Col-0 (p-values < 0.0005; Table II-2 and Figure II-7 (a)).

SALK\_048174 plants still displayed a lower mean leaf apparition rate, leading to a lower number of rosette leaves and consequently to a smaller rosette area (Figure II-7).

The data also confirmed that the final leaf 6 areas of SALK\_055458 and SALK\_126071 were significantly smaller than Col-0 ones (Figure II-7 (b)). However, for these two lines, differences were found concerning the dynamics of leaf apparition. Indeed, the slight increase in the duration of vegetative phase observed for SALK\_055458 with respect to Col-0 was not significant in the second experiment (Figure II-7 (d)). Despite this difference, there was still no significant difference for the final number of rosette leaves between those 2 genotypes, indicating that the lower leaf apparition rate of the mutant line was still compensated (Figure II-7 (c)).

On the contrary, the increase in SALK\_126071 leaf apparition rate was higher than previously observed, and, instead of simply compensating the reduced duration of vegetative phase, led to a significantly higher number of rosette leaves in these T-DNA line plants than in Col-0 plants (p-value < 0.0005; Figure II-7 (c) and (e)).

#### ***3.4.2. Analysis of vegetative phase changes gives further insights into the differences between the numbers of rosette leaves.***

The presence of trichomes on the abaxial face of rosette leaves is used as a marker of the developmental phase during which a leaf as emerged (juvenile vegetative phase, juvenile-to-adult transition and adult vegetative phase). By considering the number of leaves attributed to each of those phases as a time indicator, it is possible determinate the onset and the duration of each phase.

The analysis of the segmentations of the shoot development in developmental phases considering abaxial trichomes as markers of vegetative phase changes highlighted new differences between Col-0 and the 3 T-DNA insertion lines grown during Experiment 3. Significant differences were found in the number of leaves forming the juvenile and transition phases for the three mutant lines, whereas SALK\_055458 and SALK\_048174 exhibited the same mean number of leaves initiated in the adult phase as Col-0 plants (Table II-3).

SALK\_055458 showed a longer series of juvenile leaves, tested as significant even if it was only by about 1 leaf, and a shorter series of transition leaves, by about 1 leaf too (Table II-3). These opposed and quantitatively equivalent changes in developmental phases are consistent with the fact that there was no significant difference between SALK\_055458 and Col-0 final numbers of rosette leaves. However it gave the information that this line was slightly affected in both the onset and the duration of the vegetative phase change.

| Genotype    | Juvenile phase<br>length (leaves) |     | Transition phase<br>length (leaves) |     | Adult phase<br>length (leaves) |      |
|-------------|-----------------------------------|-----|-------------------------------------|-----|--------------------------------|------|
| Col-0       | 7.2                               |     | 3.2                                 |     | 1.1                            |      |
| SALK_055458 | 7.9                               | *** | 1.9                                 | *** | 1.1                            | n.s. |
| SALK_048174 | 6.8                               | *   | 1.5                                 | *** | 1.3                            | n.s. |
| SALK_126071 | 11.2                              | *** | 1.5                                 | *** | 0.3                            | ***  |

Table II-3. Developmental phases lengths measured by the number of leaves in each phase, for Col-0 and the 3 T-DNA insertion lines. As explained in the Material and Methods section, the segmentation was based on the absence (juvenile phase), partial presence (transition phase) or homogeneous presence (adult phase) of abaxial trichomes on the successive leaves. Values corresponds to means computed for each genotype (n = 50). Stars indicate the T-DNA lines for which the value is significantly different compared to Col-0 (Student t-tests): \*, 0.01 < p-value < 0.05; \*\*, 0.001 < p-value < 0.01; \*\*\*, p-value < 0.001.

The analysis of these developmental phases also showed that the lower number of rosette leaves exhibited by SALK\_048174 plants was mainly due to a shorter transition phase (by almost 2 leaves; Table II-3). The number of leaves initiated during the juvenile phase was also lower than Col-0 one, but about less than 1 leaf and the significance level of the difference was low. Therefore, the onset of vegetative phase change in SALK\_048174 plants occurred almost at the same time as it did in Col-0 plants, but did not take as long to be completed.

In SALK\_126071, there were a larger number of leaves initiated during the juvenile phase that was responsible of the increase in final number of rosette leaves. Both the juvenile-to-adult transition and the adult phase were shorter than in Col-0 plants (the majority of the plants did not even have an adult phase), but the resulting difference in number of leaves was not enough to compensate the one triggered by the previously mentioned increase of leaves initiated during the juvenile phase (respectively about 2 and 4 leaves; Table II-3).

### ***3.4.3. Variation in final areas between successive rosette leaves of a same genotype; consistency with results observed for a leaf at a given rank***

Differences between individual leaf area of the 90 mutant lines and their wild-type Col-0

grown during Experiments 1 and 2 were assessed from the comparison of leaf 6 area mean values. However it was uncertain that these differences observed at one specific leaf rank were conserved at the other ranks. In order to test this, areas of each successive rosette leaves obtained from Experiment 3 were compared rank by rank and sorted by order of leaf apparition on the rosette. The first observation that could be made was that final leaf area tended to increase for each new leaf produced after the second rank. However, the last produced leaf had the tendency to exhibit a smaller leaf area than the preceding one (Figure II-8 (a)).

The differences observed when comparing final leaf 6 areas of Col-0 and mutant lines were conserved for most leaf ranks. For the SALK\_048174 line, leaf areas had no significant differences compared to Col-0 ones, except for the twelfth (and last) emerged leaf. For the SALK\_055458 and SALK\_126071 lines, plants exhibited smaller leaf areas for every rosette leaf (Figure II-8), indicating that the effect of the mutations on leaf expansion concerned all leaf ranks.

#### ***3.4.4. Variation of $LER_{max}$ and length-to-width ratio between successive rosette leaves; relationship with the final area of the considered leaf***

Maximal leaf expansion rates ( $LER_{max}$ ) was affected similarly to individual leaf areas. The observations made for these variables were rigorously the same, whether it was regarding the variation of the variable between the successive leaves of a same genotype or the consistency of the effect of a mutation among the different leaf ranks of a given T-DNA line (Figure II-8 (b)). Figure II-9 (a) illustrates the strong correlation existing between the leaf final area and the  $LER_{max}$  of the same leaf. This correlation was expected as the final area of each leaf was used to compute the corresponding  $LER_{max}$ , yet, the Pearson coefficient of correlation ( $r = 0.97$ ) showed how tightly these 2 leaf growth variables are related. In fact, the coefficient of determination showed that 94 % of the variability of  $LER_{max}$  can be explained by the linear relationship between this variable and individual leaf area.

Individual leaf blade length-to-width ratio, as individual leaf areas, tended to increase with leaf rank, reflecting an increase in the length of the successive leaves (Figure II-8 (c)). However this increase was not similar for all the genotypes grown during Experiment 3. In SALK\_048174 plants, this trend was quite marked, whereas it was the less pronounced in SALK\_126071 plants, that, in comparison with the other genotypes, had the most elongated leaves at the lowest ranks and the rounder leaves at the highest ranks.

The ranks presenting the highest mean leaf areas were not necessarily the ones with the most elongated leaves (Figure II-8). These two variables were still linked by a linear relationship. Figure II-9 (b) shows the correlation existing between leaf final area and the length-to-width ratio of the same leaf. The coefficient of determination computed indicates that about 64 % of the variability observed among the length-to-width ratio values can be explained by this relationship.

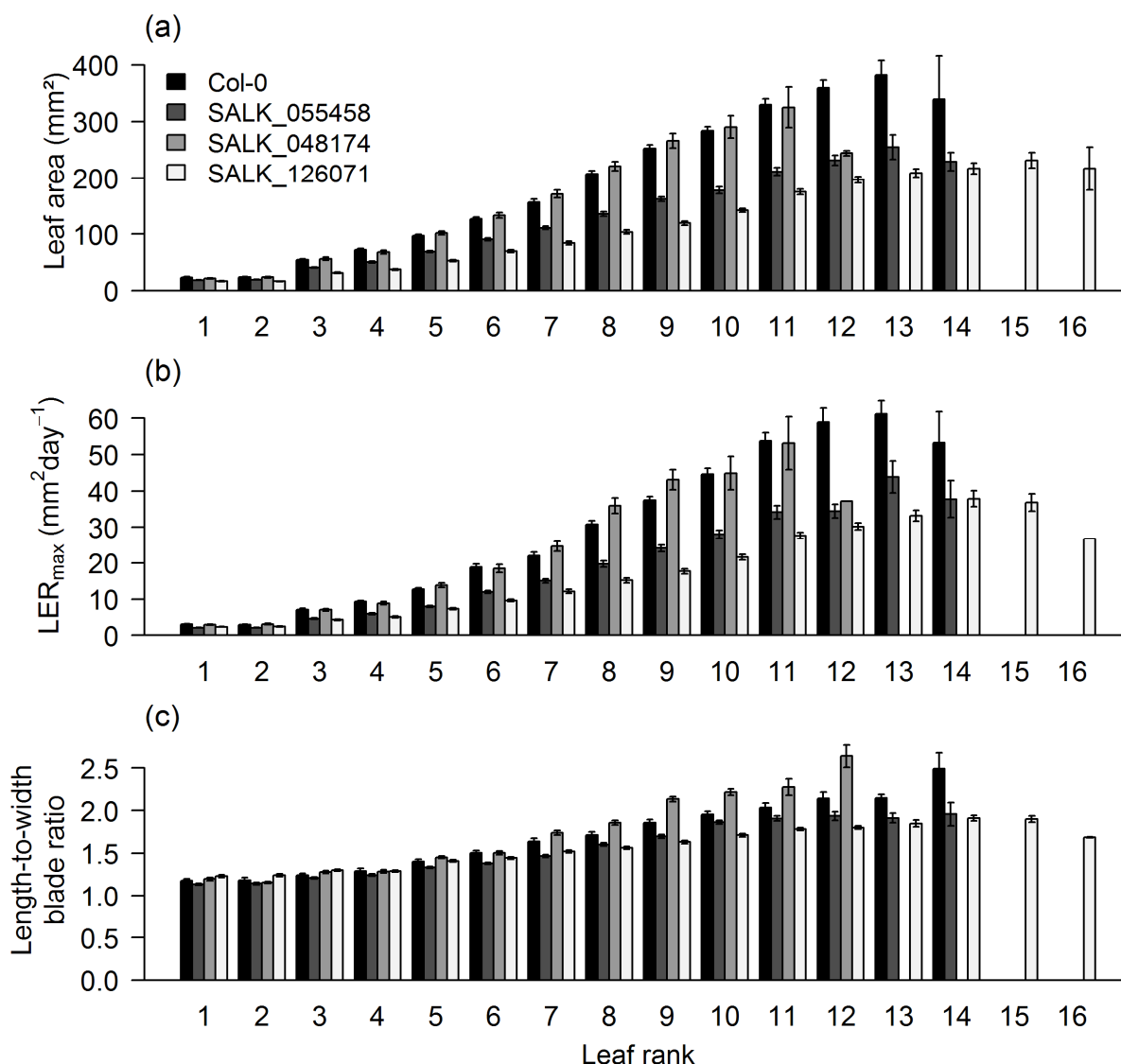


Figure II-8. Individual leaf areas (mm<sup>2</sup>; (a)), maximal leaf expansion rates (LER<sub>max</sub>, mm<sup>2</sup> day<sup>-1</sup>; (b)) and leaf blade length-to-width ratio (c) are represented rank by rank, sorted by leaf apparition order. Bars are the mean value by rank computed for each genotype ( $\pm$  SE).

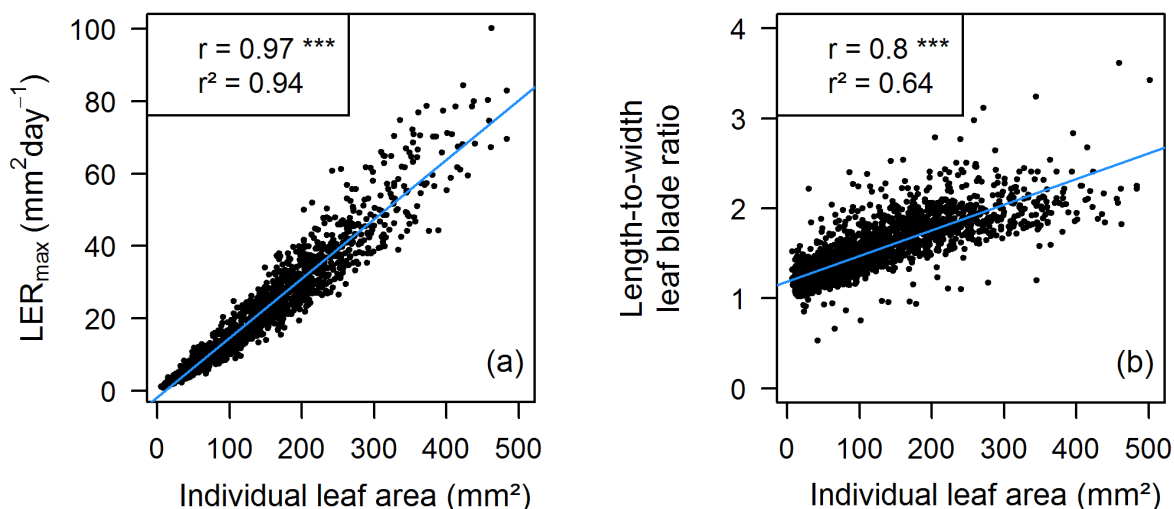


Figure II-9. Linear relationships between individual leaf area (mm<sup>2</sup>) and (a) corresponding maximal leaf expansion rate (LER<sub>max</sub>, mm<sup>2</sup> day<sup>-1</sup>); (b) leaf blade length-to-width ratio. Each point represents an individual leaf ( $n = 2,274$ ). The linear relationship estimated between the variables is represented by a blue line.  $r$  is the Pearson correlation coefficient and  $r^2$  the coefficient of determination. \*\*\*,  $p$ -value  $< 0.001$ .

A noticeable observation for those four leaf growth variables (individual leaf area, LER<sub>max</sub>, characteristic growth duration and leaf blade length-to-width ratio) was the absence of variation between the first two leaves of a given genotype. It was the most striking with the final leaf area and LER<sub>max</sub> variables for which the first two true rosette leaves values were very similar one to another and quite lower than the mean values of the third 3 leaf.

## 4. DISCUSSION

### 4.1. Early stage leaf growth phenotyping does not reflect the final leaf growth phenotype

One of the first outputs from our studies is related to leaf growth phenotyping methods. Due to technical constraints, leaf growth analyses have often been based on measurements of a trait or a few traits of interest at a given date, without taking in account plant developmental stages. The comparisons of leaf growth traits measured at 18 DAS and at the end of the rosette development on 91 *A. thaliana* lines (Experiments 1 and 2) showed that phenotypes identified at a given date are not conserved over time and that a trade-off exists between the initial rate of growth and growth duration. These observations support that conclusions drawn on traits measured at a given date cannot be generalized to the whole growing period and another analysis at another date or stage may lead to different conclusions, whatever the scale of growth analysis.

The trade-off between the initial rate of growth and growth duration is embedded in the evolutionary framework of life-history theory that predicts that the covarying values of individual characters are constrained by the use of a limited amount of resources (Vasseur et al. 2012). In annual plants such as *A. thaliana*, fitness increases with body size, which can be achieved by high relative growth rate or by extending the period of growth (Roff 2007; Metcalf and Mitchell-Olds 2009). Interestingly a similar trade-off has been observed at the individual leaf scale (Cockcroft et al. 2000). Genotypes with high initial relative leaf expansion rate have generally a shorter duration of leaf expansion (Cockcroft et al. 2000; Cookson et al. 2005).

Experiments 1 and 2 had allowed a general characterization of the leaf growth phenotype of 90 T-DNA insertion lines issued from the SALK collection (Alonso et al. 2003). Only a very small proportion of the lines identified in our screen were mutated on genes that had been previously described with a role in leaf development, suggesting that the SALK collection has a strong potential for identifying new genes controlling growth and improving known molecular networks. Most of the grown genotypes had smaller final rosette (and individual leaf) areas than their wild-type Col-0, suggesting that a disruption of various molecular pathways leads to a down-regulation of leaf expansion.

### **4.2. Considerations regarding the results of the more in depth phenotyping analysis**

Experiment 3 was performed growing a larger number of plant replicates per genotype with the measurements of additional leaf growth variables. Comparison of the results obtained for the same genotypes but with different sample sizes highlighted the fact that analyzing datasets with small numbers of independent replicates using statistical hypothesis tests can lead to groundless conclusions. The matter of sample size determination is quite documented (Lwanga and Lemeshow 1991; Lenth 2001; Noordzij et al. 2011; Beck 2013) and even if examples are often taken in medical research, this problematic has to be emphasized to conduct meaningful studies in plant phenotyping studies.

We extended the study of the effect of T-DNA insertions on individual leaf growth to each successive rosette leaf, instead of focusing on a given leaf rank. Comparing growth traits of individual leaves rank by rank when plants have the same number of leaves is not problematic but a large variation in leaf number is often reported among *A. thaliana* genotypes. In the population of recombinant inbred lines used in (Tisné et al. 2008), for example, the number of rosette leaves varied from 7 to 30 leaves among the different lines.



In our case, the comparison of Col-0 and SALK\_055458 showed that the T-DNA insertion had the same effect on each rosette leaf. However, this type of comparison can become tricky when plants with significantly different numbers of leaves are studied. In this case, conclusions drawn from the analysis of the first appeared leaves can differ from those drawn when leaves of higher ranks are compared. That could be explained by differences in the developmental phase of the plant when the compared leaves were initiated, since leaves initiated during different plant developmental phases can present different morphological and physiological differences (Poethig 1990, 2003; Matsoukas 2014). One solution to be able to perform an analysis of the effect of a mutation or a treatment on individual leaf growth that would be representative of the effect observed among the shoot, would be to synchronize flowering time among plants as presented in Lièvre et al. 2013; see Chapter I, Figure I-2.

The number of successive leaves sharing the same identity – defined by the presence of trichomes on their abaxial face based on Telfer et al. 1997 – is often interpreted as a characteristic of the duration of the corresponding developmental phase. Here, we made the observation that SALK\_126071 plants have on average an extended series of juvenile leaves compared to Col-0, without any reduction of transition leaf number and a less marked decrease in adult leaf numbers. If the number of leaves sharing the same identity really reflected the duration of the developmental phase, the complete vegetative phase should thus be longer in the mutant than in its wild-type. Nevertheless, we showed here that, on the contrary, SALK\_126071 vegetative phase was shorter than Col-0 ones, and that the larger final number of rosette leaves was due to an increased mean apparition rate.

Comparison of shoot development for the analysis of genetic and/or environmental effects can be tricky because of the multiplicity of the variables that could be involved and because of the asynchronicity between plants development, raising the question whether comparison of leaves based on their rank is relevant. Comparing developmental phases (succession and variation of the leaf-growth traits within each phase) could be a solution. However the definition of such phases might be complex. In the literature, observed heteroblasty is often linked to the acquisition of reproductive competence, and vegetative development in *A. thaliana* is frequently divided into three developmental phases (juvenile, transition and adult) using abaxial trichomes patterns as discriminating criterion. However, several studies highlighted that the relationship between the variation of morphological leaf traits and plant phase change are not yet clear and that distinguishing genetically determined ontogenetic changes from those due to plasticity can be difficult (Diggle 2002; Poethig 2010; Huijser and

Schmid 2011). For example, we observed here that the first two true leaves of the rosette were both very similar one to another and quite different from the following leaves, considering their area, shape and maximal expansion rate. Differences between the first postcotyledonary leaves and the subsequent juvenile leaves have been reported in *A. thaliana* (Telfer et al. 1997; Steynen et al. 2001) and other species such as *Pseudopanax crassifolius* (Gould 1993), *Fraxinus pennsylvanica* (Merrill 1986), *Desmodium paniculatum* (Wulff 1985), *Medicago* (Damerval 1983), maize (Bongard-Pierce et al. 1996), wheat and barley (Dornbusch et al. 2010). It could be explained by some degree of regulation by the seed environment on the leaves initiated during embryogenesis (Bongard-Pierce et al. 1996; Jones 1999). The juvenile vegetative phase (in terms of reproductive competence) could thus be divided into two developmental phases, and, consequently, phenotypic studies of the shoot vegetative development might need to be based on four different phases: seedling, juvenile, transition and adult phases.

These considerations show the interest of developing a new method of analysis to decipher the structure of *A. thaliana* rosette into different developmental phases, taking into account the variations of different leaf-growth traits along the shoot.



## CHAPTER III. INTEGRATIVE MODELS FOR ANALYZING *A. THALIANA* ROSETTE DEVELOPMENTAL PHASES

### 1. INTRODUCTION

The genome sequences of model plants such as *A. thaliana* ecotypes and *Brachypodium distachion* are available. In addition, rice, maize, sorghum (*Sorghum vulgare*), barley (*Hordeum vulgare*) and many dicot and monocot crops are either sequenced or soon will be. To harness this genomic information for agricultural application, it has to be carefully and comprehensively linked to plant phenotype. However, phenotypic descriptions of genome-wide knockout collections are often limited because phenotyping technologies have often been reported as the bottleneck of phenomic studies (Furbank and Tester 2011). The annotation ‘no visible phenotype’ occurs frequently and it maybe reveals a lack of capacity for the plant community to analyze the subtle phenotypic effects of various genetic modifications. Shoot growth phenotyping per se is not an obstacle anymore in *A. thaliana* since technologies have been developed and are still improved to measure plant traits on the rosette (Granier et al. 2006, 2014; Arvidsson et al. 2011; Zhang et al. 2012; Fiorani et al. 2012). The challenge is now to integrate the different types of traits that can be acquired at different scales during phenotyping studies or to identify the most pertinent trait(s) to compare genotypes. In leaf growth phenotyping studies for example, one can have to analyze simultaneously dimensional and morphological data on individual leaves (such as leaf areas, leaf blade length-to-width ratio or number of serrations on the leaf margin), leaf growth dynamics data (such as leaf expansion rate, duration of leaf expansion), and data measured at the leaf scale but analyzed as series along the shoot when shoot developmental phases are investigated (e.g. changes in abaxial trichomes covering with the leaf rank).

When measurements are made for each individual leaf, comparisons are usually made by analyzing the quantitative or qualitative characteristics of leaves at a given rank – i.e. leaves that have the same position on the shoot according to the order of their emergence after germination (Willmann and Poethig 2011; Friedli and Walter 2014). As discussed earlier, this framework can be used when plants that are compared exhibit the same number of leaves, but, it raises a problem when they have different leaf numbers as it is often the case when comparing different genotypes or a same genotype grown in various environmental conditions. In this context, the phenotypic analysis of the plant shoot structuring it into

successive developmental phases seems to be the most appropriate framework to compare shoot development of different genotypes.

Such structuring is based on more or less abrupt changes in the shoot characteristics (e.g. Dambreville et al. in press). We will refer to these changes as the plant heteroblasty. The term ‘heteroblastic’ was initially introduced by Goebel (1900). It was used to oppose plant species in which substantial differences were observable between earlier and later formed phytomers, to species with small and gradual changes along the shoot – the ‘homoblastic’ species. Nowadays, the definition of the term ‘heteroblasty’ has been extended, and this term is commonly used to refer to changes in leaf size and/or shape occurring during ontogeny, even if they are subtle. It has to be noted that ‘heteroblasty’ is sometimes used to refer to ontogenetic changes only (Zotz et al. 2011). When it refers to the combination of the observable morphological changes, these changes can be due to several reasons such as ontogeny, physiological aging, increase in the shoot size (Evans 1972; Day et al. 2002; Mencuccini et al. 2007), environmental heterogeneity such as soil humidity, light quality, temperature or nutrients availability (Ray 1987; Lee and Richards 1991; Jones 1995; Bruni et al. 1996; Fisher et al. 2002; Cutri et al. 2013) and vegetative phase change (Poethig 2003).

Gradual changes occurring along the shoot can be related to individual leaf size and shape, internode size, branching patterns but also leaf and/or internode growth (Guédon et al. 2001; Andrieu et al. 2006; Cookson et al. 2007; Willmann and Poethig 2011). These changes in shoot traits are usually analyzed individually during phenotyping studies, partly because of a lack of appropriate statistical methods. There is therefore a real challenge in developing methods that would allow the segmentation of shoot axis development into successive developmental phases based on the integration of several traits reflecting the heteroblastic changes.

During the analysis presented in Chapter II we identified three T-DNA insertion lines (SALK\_055458, SALK\_048174 and SALK\_126071) with contrasted leaf growth development. They were used here to build multivariate segmentation models in order to analyze the structuring of *A. thaliana* rosette. We grew these 3 lines and their Col-0 wild-type with a large number of replicates for each genotype ( $n = 60$ ). The throughput of the experiment could be considered as low when considering the number of genotypes grown together but was high when considering the number of measured plants in total. This high number of replicates was necessary to compute more robust statistical models to get more trustworthy conclusions. Most kinematic data that are necessary to quantify spatial and

temporal changes in leaf growth were obtained here from hand-made measurements that are time-consuming. As explained in Chapter I, section 1, a leaf can be apprehended as a whole growing organ, whom development can sometimes control cell division and expansion processes (Tisné et al. 2008; Tsukaya 2013), making the study of *A. thaliana* shoot development at the macroscopic scale relevant. Because of these considerations and of the high number of plants that we wanted to grow together for statistical relevance, we first focused on the macroscopic scale of shoot development. Because of the shortness of the internodes in *A. thaliana* rosette, we focused on leaf characteristics.

Even if the main objective of the work presented in this chapter was to develop a model for *A. thaliana* shoot growth phenotyping taking into account shoot structuring, the model outputs have been used (1) to question how vegetative phase change and heteroblasty are delimited and explore some of the problems that arise in the explicit or implicit link between them and (2) to characterize the 3 T-DNA insertion lines used to develop this approach.

## **2. MATERIAL & METHODS**

### **2.1. Plant material and growth conditions**

Four *Arabidopsis thaliana* (L.) Heynh genotypes were grown during the experiment: Columbia 0 natural accession (Col-0) and three homozygous T-DNA insertion mutant lines issued from the SALK institute collection created by the group of Prof. J.R. Ecker (Alonso et al. 2003). Seeds used were provided by José Luis Micol group (Universidad Miguel Hernandez, Elche, Alicante, Spain).

Sixty plants of each line were grown together. Seeds were sown in pots filled with a mixture (1:1) of a loamy soil and organic compost. Plants were grown in a growth chamber equipped with the PHENOPSIS automaton under controlled air temperature, air humidity and incident light. Micrometeorological conditions were kept constant during the whole duration of the experiment and homogeneous within the growth chamber (Granier et al. 2006). Light was on during 16 h per day and provided by a bank of cool-white fluorescent tubes and HQi lamps. It was measured at the plant level using a photosynthetic photon flux density (PPFD) sensor (LI-190SB, Li-Cor, Lincoln, NE, USA). Mean micro-meteorological conditions are indicated in Table III-1. Soil water content was determined before sowing to estimate the amount of dry soil in each pot. Subsequent changes in pot weight were due to changes in soil water content (the plant weight was considered as negligible as it was never higher than 1 g).

This allowed the computation and daily automatic adjustment of soil water content to 0.35 g water g<sup>-1</sup> dry soil by the PHENOPSIS automaton (Granier et al. 2006) with a modified one-tenth-strength Hoagland solution (Hoagland and Arnon 1950), from sowing to the end of the experiments.

|  |      |
|--|------|
| Day length (h)   | 16   |
| Incident PPFD ( $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$ )  | 178  |
| Air temperature (°C)   | 20.4 |
| Air humidity (%)   | 60.6 |
| Soil water content (g H <sub>2</sub> O g <sup>-1</sup> dry soil) | 0.36 |

Table III-1. Mean meteorological conditions applied during plant growth.

## 2.2. Growth measurements – Image analyses and variables selection

During the experiment, zenithal images of the plants were taken on a daily basis by the PHENOPSIS automaton to follow rosette growth. Plants were harvested when the first silique was formed in order to ensure that each individual rosette leaf had completed its expansion (as previously shown in the group in previous experiments in the same growing conditions). The successive leaves of the rosette were excised without their petiole, stuck on a sheet of paper in the order of emergence and scanned.

### 2.2.1. Leaf morphological characterization

Individual leaf area (mm<sup>2</sup>), maximal leaf blade length (mm) and width (mm) were automatically measured on the leaf scans with an image analysis software (ImageJ 1.43C, Wayne Rasband, National Institutes of Health, USA) and leaf blade length-to-width ratio was computed.

Leaves were divided into three categories depending on the level of covering of their abaxial surface by trichomes: leaves without abaxial trichomes, leaves with trichomes (at least one) that did not entirely span their abaxial surface, and leaves with trichomes that span their entire abaxial surface (trichomes distributed homogeneously and at least one within 2 mm of the leaf margin of the distal tip). To limit spurious measurement fluctuations, the subsequent leaves on the shoot had to meet the same criterion or to correspond to the next leaf category. In this study, we did not assume that the leaf covering with trichomes defined *a priori* a developmental phase (the three categories corresponding respectively to juvenile, transition and adult leaves according to Telfer et al. 1997 and Bollman et al. 2003). This morphological

variable was used in conjunction with leaf shape, dimension and growth variables to define developmental phases with less *a priori*.

### 2.2.2. Characterization of leaf expansion kinetics

Projected individual leaf areas were measured with the image analysis software ImageJ 1.43C on the daily zenithal pictures of the rosette taken by the PHENOPSIS automaton by drawing the visible part of leaf blade. Petioles were not included in the measurement. In case of overlap between leaves making a leaf contour too uncertain to be drawn, or when a leaf had revolute margins (phenomenon of leaf curling) at a point where about 50 % or more of the leaf area was hidden the leaf was not measured. The measured areas were used to estimate individual leaf expansion curves (Torres and Frutos 1989). The logistic function used to model leaf expansion is given by

$$Y = \frac{A}{1 + \exp\{-(t - M) / B\}}$$

where  $A$  is the upper asymptote, i.e. the estimated final leaf surface,  $B$  is the characteristic growth duration (inverse of growth rate), and  $M$  is the time corresponding to the inflection point (time of maximum growth).

The growth duration between  $\alpha A$  and  $(1 - \alpha)A$  with  $0 < \alpha < 0.5$  given by

$$t_{1-\alpha} - t_{\alpha} = 2B \log\left(\frac{1-\alpha}{\alpha}\right)$$

only depends on  $\alpha$  and  $B$ . Therefore, parameter  $B$  corresponds to the growth duration between  $\alpha A$  and  $(1 - \alpha)A$  for  $\alpha = 1 / \{1 + \exp(0.5)\} \approx 0.38$ .

Classically the leaf expansion curves are summarized by computing the maximal leaf expansion rate rate ( $\text{LER}_{\max}$ ):

$$\text{LER}_{\max} = A / 4B$$

However, in our case, because of leaf overlap and leaf curling phenomena, the estimated final leaf area  $A$  was frequently underestimated, making the estimated  $\text{LER}_{\max}$  biased. Moreover, even by using the unbiased final leaf area obtained by rosette dissection to compute  $\text{LER}_{\max}$ , the strong correlation between these two variables makes the use of both redundant. We assumed that the biased estimation of the final leaf area within the logistic growth model because of leaf overlap does not affect substantially the estimation of the characteristic growth duration. Consequently, we used this estimated parameter to



characterize individual leaf expansion kinetics.

### **2.2.3. Characterization of the shoot vegetative development**

The number of rosette leaves was defined as the number of true leaves produced between the apparition of the two cotyledons and the emergence of the primary flowering axis (cotyledons were excluded). Duration of the vegetative phase was determined as the time elapsed between the emergence of the first pair of true leaves and the emergence of the last rosette leaf. Mean leaf apparition rate (leaf day<sup>-1</sup>) was computed as the final number of rosette leaves divided by the duration of the vegetative phase.

## **2.3. Integrative multi-scale models for analyzing *A. thaliana* rosette developmental phases**

### **2.3.1. Building of multivariate sequences for each *A. thaliana* rosette**

Data take the form of multivariate sequences indexed by the leaf rank. The leaf description at each successive rank combines a morphological variable (ordinal categorical variable with ordered categories no trichome, partial covering with trichomes, full covering with trichomes), a shape descriptor (length-to-width blade ratio), the final leaf area and a dynamic variable (characteristic growth duration extracted from the estimation of a logistic growth model on the basis of leaf expansion follow-up data).

Because of leaf overlap, the larger the leaf is, the more its surface is underestimated. We thus chose to incorporate in the multivariate sequences the final leaf area obtained by dissecting the rosette.

### **2.3.2. Segmentation models**

A key question was whether the rosette growth pattern takes the form of a trend (i.e. gradual change in the mean level for quantitative variables such as the leaf area) or of a succession of developmental phases separated by abrupt changes (for instance juvenile leaves without abaxial trichomes followed by transition leaves partially covered with abaxial trichomes and then adult leaves fully covered with abaxial trichomes). After some exploratory analyses, we chose to build two-scale segmentation models on the basis of multivariate sequences that combine developmental phases and linear trends within phases for quantitative variables. These two-scale models are semi-Markov switching models (SMSMs) that generalize hidden semi-Markov chains by incorporating regression models as observation models. They are formally defined in Appendix II. In our context, the succession and duration of developmental phases (coarse scale) are represented by a non-observable semi-Markov

chain while the different leaf descriptors within a developmental phase (fine scale), are represented by observation models attached to each state of the semi-Markov chain. Hence, each state of the semi-Markov chain represents a developmental phase. A semi-Markov chain is defined by three subsets of parameters:

- Initial probabilities to model which is the first phase occurring in the sequence measured,
- Transition probabilities to model the succession of developmental phases,
- Occupancy distributions attached to non-absorbing states (a state is said to be absorbing if, after entering this state, it is impossible to leave it) to model the phase length in number of successive leaves. We used, as possible parametric state occupancy distributions binomial distributions  $B(d, n, p)$ , Poisson distributions  $P(d, \lambda)$  and negative binomial distributions  $NB(d, r, p)$  with an additional shift parameter  $d \geq 1$ .

A SMSM adds observations models to the non-observable semi-Markov chain:

- In the case of the abaxial trichome variable, observation models are simple categorical observation distributions and the probabilities of observing each category within a developmental phase were estimated directly. In the case of the three quantitative variables (length-to-width blade ratio, final leaf area, characteristic growth duration), we chose to model trends within developmental phases using simple linear regression models because of the short length of phases (between 2 and 11 successive leaves).

‘Left-right’ SMSM composed of successive transient states followed by a final absorbing state were estimated on the basis of each dataset. A state is said to be transient if after leaving this state, it is impossible to return to it. In a ‘left-right’ model, the states are thus ordered and each state can be visited at most once. Each estimated model was used to compute the most probable state sequence for each observed sequence (Guédon 2003). The restored state sequence can be viewed as the optimal segmentation of the corresponding observed sequence into sub-sequences, each corresponding to a given developmental phase.

It is generally assumed, when using SMSMs, that the sequence length is independent of the process that is supposed to have generated the sequence (Guédon 2003). This assumption entails that the time spent in the last visited state is ‘censored’ or truncated, i.e. the last vegetative developmental phase was randomly truncated by the transition to the reproductive

phase. This is indeed unrealistic. In this study, it was chosen to assume instead that the end of an observed sequence systematically coincides with the transition from the current state to an extra absorbing ‘end’ state modeling implicitly the reproductive phase (no observation model is associated with this end state). Hence, at the end of an observed sequence, the process systematically jumps to the absorbing end state. In this way, it is possible to model explicitly the duration of the last vegetative developmental phase.

### ***2.3.3. Selecting the number of states of the underlying semi-Markov chain***

Combining results from the literature (Poethig 1990; Telfer et al. 1997) and exploratory analyses, it was rapidly obvious that at least three transient states were necessary (corresponding roughly to the first two seedling leaves, juvenile leaves and adult leaves). We then evaluated the possible number of states from 3 onward (the absorbing ‘end’ state being not counted).

### ***2.3.4. Illustration with the model built for the Col-0 dataset***

We built a five-state model (four transient states corresponding to four successive vegetative phases and an absorbing ‘end’ state corresponding to the reproductive phase) on the basis of the Col-0 multivariate sequences; see Figure III-1. The four vegetative phases will be referred to as the seedling phase, the juvenile phase, the transition phase and the adult phase with reference to the literature; see justifications below. As shown in the transition graph (Figure III-1 (c)), the succession was not completely deterministic since the adult phase can be skipped with probability 0.27 – and the transition phase with probability 0.02. Phase length distributions (expressed in number of successive leaves) associated with each developmental phase are also shown, above the transition graphs. The mode of each distribution represents the most probable number of leaves that would compose the corresponding phase. Coupled with the probabilities of the phases as a function of the leaf rank (Figure III-1 (b)), it allows the deduction of the most represented segmentation (or segmentations if two phases strongly overlap) of a Col-0 plant (Figure III-1 (a)). The seedling phase was in a large majority composed of the first two leaves of the shoot. It was discriminated by the three quantitative variables: final leaf area, length-to-width ratio of the leaf blade and characteristic growth duration parameter. In opposition to the following phases (i.e. the juvenile phase), there was almost no trend in this phase for these traits (Figure III-1 (d) and Table III-5). The juvenile phase had 5 leaves in average and mostly began with the third emerged leaf. The transition phase was constituted from 1 to 5 leaves, with an average of 3 leaves. For the plants that displayed an adult phase, it was mainly composed of 1 leaf (9

sequences had 2 leaves assigned to this phase, and only 1 sequence had 3 leaves in this phase). The discriminations between the juvenile, transition and adult phases followed exactly the organization of the trichome variable: none of the leaves assigned to the juvenile phase exhibited abaxial trichomes, every leaves assigned to the transition phase had a partial covering of their abaxial face with trichomes, and those attached to the adult phase had their abaxial face fully covered with trichomes. The most structuring quantitative variable was the final leaf area, with changes in the slopes estimated in the successive phases (Figure III-1 (d), Table III-5). On the contrary, the less structuring quantitative variable was the leaf-to-width ratio that, except for the first two emerged leaves, conserved the same slope along the shoot (Table III-5).

The graph showing the probability for a leaf at rank  $n$  to be assigned to a given phase indicates that the segmentation of Col-0 plants into developmental phases was quite synchronous among the 49 individuals (Figure III-1 (b)). The main overlaps observed between two successive phases were at rank 8 (leaves mostly assigned to the transition phase, and in smaller proportion to the juvenile phase), and at rank 11 (leaves assigned in the same proportions to the transition and the adult phases).

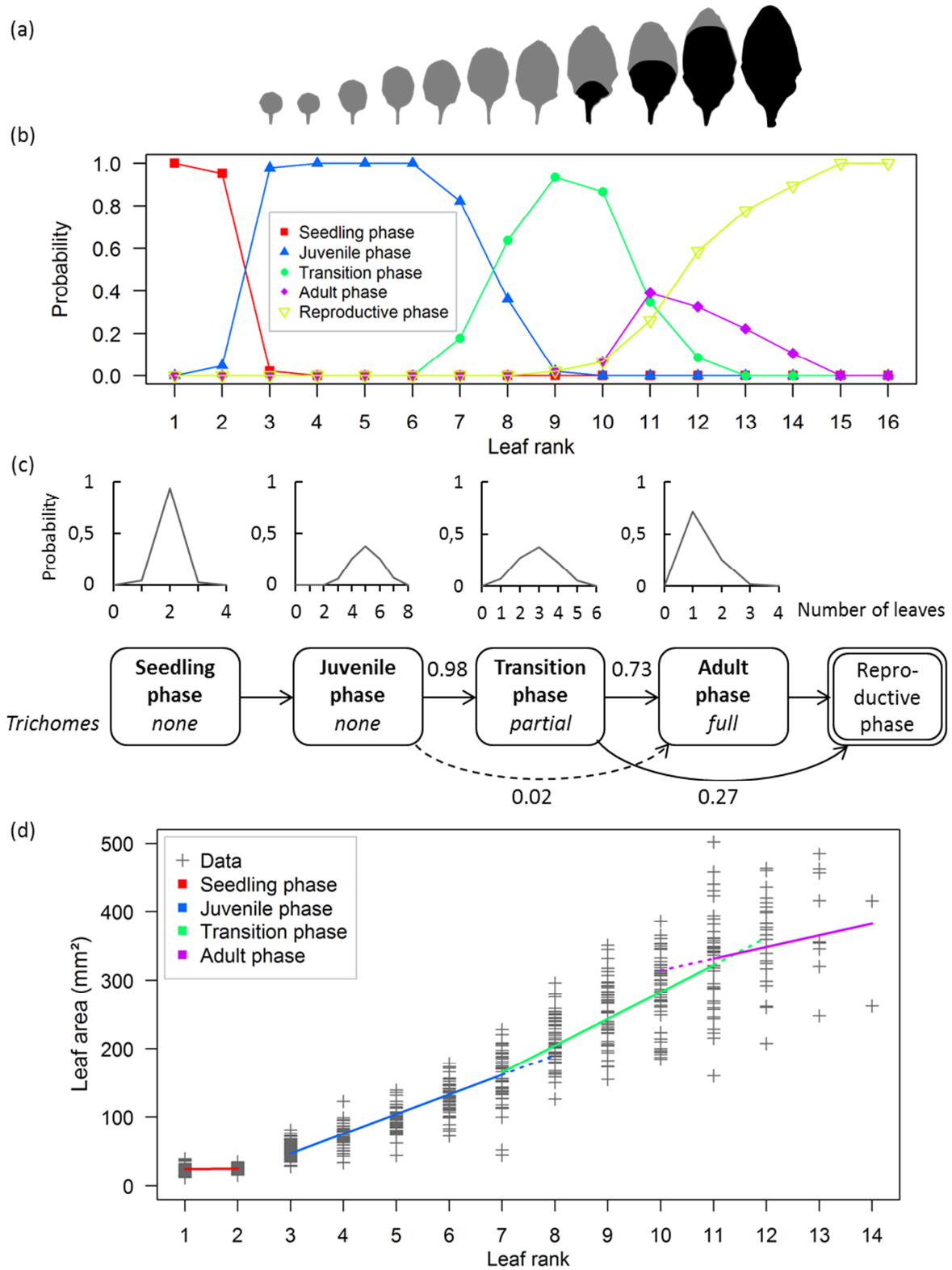


Figure III-1. Representations of the segmentation model main outputs obtained for Col-0 dataset. (a) Representation of a typical Col-0 plant, dissected view to show each leaf rank. Leaves are shown in order of production, from left to right, and shaded to indicate no abaxial trichomes (light gray) and abaxial trichomes (black) (b) Probabilities of the developmental phases as a function of the leaf rank. (c) Graph of transition between phases and phase length distributions. (d) Final leaf area values as a function of the leaf rank. Trends within each phase are shown in color. Plain lines: most probable phase (i.e. majority of the leaves at the rank have been assigned to the phase). Dotted lines: alternative phase (with respect to the most probable phase); shown if probability > 0.04.

### 3. RESULTS

#### 3.1. Validation of the assumption of segmentation in developmental phases

|             | Minimum probability | % of individuals above the given thresholds |     |      | Number of possible segmentations |      |
|-------------|---------------------|---|-----|------|----------------------------------|------|
|             |                     | 0.75  | 0.9 | 0.95 | Interval                         | Mean |
| Col-0       | 0.64                | 90  | 59  | 35   | 3 → 32                           | 12.7 |
| SALK_048174 | 0.53                | 70  | 27  | 14   | 6 → 18                           | 11.7 |
| SALK_055458 | 0.42                | 80  | 63  | 33   | 6 → 32                           | 16.5 |
| SALK_126071 | 0.46                | 68  | 28  | 13   | 6 → 18                           | 15.3 |

Table III-2. Segmentation uncertainty: minimum posterior probability of the optimal segmentation, proportions of individuals whose posterior probability of the optimal segmentation is above given thresholds (0.75, 0.9 and 0.95), number of possible segmentations (lower and higher values, mean).

The posterior probabilities of the optimal segmentations (i.e. weight of the optimal segmentation among all the possible segmentations of a given observed sequence) were most often high: 90 % above 0.75 and 59 % above 0.9 for Col-0 to be related to an average of about 12 possible segmentations. The results were similar for SALK\_055458 while the segmentation ambiguity was a bit higher for SALK\_0481784 and SALK\_126071; see Table III-2. This strengthened the validity of the segmentation assumption. It also allows us to use the segmentations to accurately interpret the different model outputs.

#### 3.2. Comparison of the segmentations into developmental phases between Col-0 and each T-DNA insertion line

##### 3.2.1. The succession of states is almost deterministic

‘Left-right’ five-state SMSMs (four transient states modeling successive vegetative phases and an absorbing end state modeling the reproductive phase) were built for Col-0, SALK\_055458 and SALK\_048174 datasets. These models were then used to segment the multivariate sequences into successive developmental phases. The fact that states could not be skipped in most cases was the result of the iterative estimation procedure. The estimated transition distribution for each transient state  $i$  was thus almost degenerate i.e.,  $p_{i\ i+1} \approx 1$  and  $p_{ij} = 0$  for  $j \neq i+1$ . The only exception was state 2 with  $p_{2\text{end}} = 0.27$  for Col-0, 0.22 for SALK\_055458 and 0.08 for SALK\_0481784 (Table III-3).

Since the adult phase was systematically absent for SALK\_126071, we built a four-state SMSM for this genotype. In this case, the estimated transition probability matrix (i.e. every transition distributions corresponding to the transient states) was degenerate.

|             | Probability of skipping a phase |             |
|-------------|---------------------------------|-------------|
|             | Transition phase                | Adult phase |
| Col-0       | 0.02                            | 0.27        |
| SALK_055458 | 0                               | 0.22        |
| SALK_048174 | 0                               | 0.08        |
| SALK_126071 | 0                               | -           |

Table III-3. Probabilities of skipping a phase (i.e.  $1 - p_{i i+1}$ ) in Col-0, SALK\_055458, SALK\_048174 and SALK\_126071. Probabilities for skipping seedling and juvenile are equal to 0, so only those for transition and adult phases are shown.

### 3.2.2. *The categorical trichome variable is strongly structuring for the whole shoot development*

For the four genotypes, the seedling and juvenile phases were the only ones composed only of leaves without abaxial trichomes, indicating that this variable had no role in the definition of these two developmental phases but could be used to discriminate the juvenile phase from the transition one. For the three lines for which an adult phase was estimated, it consisted only of leaves with their abaxial face fully covered with trichomes. However, differences were observable between the genotypes for the transition phase. In Col-0 and SALK\_048174, the estimated observation distributions for the trichome variable for this developmental phase showed that it was almost only composed of leaves with abaxial surface partially covered with trichomes (Table III-4). Thus, the discrimination between the juvenile, transition and adult phases fitted almost perfectly the trichomes categories, indicating a strong structuring role of the variable in the segmentation of the multivariate sequences.

|             | No   | Partial | Full |
|-------------|------|---------|------|
| Col-0       | 0    | 1       | 0    |
| SALK_055458 | 0.1  | 0.87    | 0.03 |
| SALK_048174 | 0.04 | 0.96    | 0    |
| SALK_126071 | 0.25 | 0.66    | 0.09 |

Table III-4. Estimated observation distribution for the categorical trichome variable (3 ordered categories: no trichome, partial covering and full covering with trichomes) for the transition phase.

In SALK\_055458 the estimated observation distributions for the trichome variable showed that the ‘partial covering’ category was still largely the majority one. However, the probability to observe leaves belonging to different trichome categories in the transition phase was higher than in Col-0 and SALK\_048174 (Table III-4).

In contrast, in SALK\_126071 the transition phase was more often composed of a mixture of leaves belonging to the different trichome categories (Table III-4). In this case, the proportion of leaves fully covered with abaxial trichomes in the transition phase can be explained by the fact that adult leaves were too rare to enable the definition of a phase for them and consequently some adult leaves were modeled within the transition phase.

For the three SALK genotypes, the definition of the transition phase could thus not be strictly made on the ‘partial covering with abaxial trichomes’ category. Coupled with the fact that this variable cannot discriminate the seedling phase from the juvenile one, this indicates that the quantitative variables had also a structuring role in the segmentation of the multivariate sequences into successive developmental phases.

### ***3.2.3. The structuring role of the quantitative variables depends on the developmental phases and on the genotypes***

In extreme scenarios the variation along the shoot of the different leaf growth variables can either be described as a trend – i.e. a gradual change of variables along the shoot – or as a succession of stationary phases separated by breakpoints. The models we built segmented the rosette in three or four distinct developmental phases, depending on the genotype. However, we assumed that the different quantitative leaf traits (leaf blade length-to-width ratio, final leaf area and characteristic growth duration) would not be stationary within each phase and therefore we chose to model trends within developmental phases using simple linear regression models (see Appendix III for graphic representations). The observation of marked changes of slopes between two successive phases was used as an indicator of the structuring role of the variable. The observation of a marked shift differentiating the values of two successive phases at a given point was also used as an indicator of the structuring role of the variable.

We have seen previously that trichomes could not be used to differentiate seedling leaves from juvenile leaves. In contrast, every quantitative leaf growth variable exhibited a marked change of slope between these two successive phases, except leaf blade length-to-width ratio in SALK\_126071 (Table III-5). In most cases, the slope estimated for the seedling phase was



not significantly different from zero, reflecting the similarity of the two first epicotylar leaves development.

Discrimination between the juvenile, transition and adult phases was strongly influenced by the categorical trichome variable. However, some of the quantitative traits had also a structuring role for some of these phases, depending on the genotypes. The leaf blade length-to-width ratio was the less structuring variable for Col-0, SALK\_048174 and SALK\_126071, with none or only slight changes in slopes between the 3 last phases (Table III-5 (a)). In SALK\_055458, the variable kept the same slope between the juvenile and the transition phases, but became stationary in the adult phase. It could thus participate to the differentiation of the transition and the adult phases. Despite the minor structuring role of this variable, for Col-0 and SALK\_055458 a slight shift could be noted between the juvenile and the transition (around, respectively 9 and 12; see Appendix III Figures SIII-1 (a) and (b)) phases that was interpretable since there was no change in slopes. In a similar way, for SALK\_048174, a slight shift (around 9; see Appendix III Figure SIII-1 (c)) separated the leaves assigned to the transition phase and those assigned to the adult phase. That indicates a more or less marked change in the leaf blade shape between these two developmental phases.

The analysis of final leaf area along the shoot showed that this variable had a strong structuring role for Col-0, SALK\_048174 and SALK\_126071 (Table III-5 (b)). It discriminates the four phases for Col-0 with a marked change in slope between two successive phases. The same observations were made regarding SALK\_126071 juvenile and transition phases. For SALK\_048174, slopes for juvenile and transition phases were similar but differentiated by a marked shift (around 30 mm<sup>2</sup>; Table III-7 and Appendix III Figure SIII-2 (c)). On the contrary, for SALK\_055458 plants, final leaf areas did not allowed to discriminate leaves belonging to the juvenile, transition and adult phases, which exhibited similar slopes (Table III-5 (b)), without marked shifts between these developmental phases.

The characteristic growth duration was the only trait that was decreasing along the shoot – i.e. the latter in the rosette development a leaf is initiated, the shorter its expansion is. This variable was also structuring for the genotypes in which four developmental phases were identified (Col-0, SALK\_055458 and SALK\_048174). The juvenile and transition phases were differentiated by a change in slope since the variable was modeled by a decreasing trend during the juvenile phase and was stationary during the transition phase (Table III-5 (c)). The transition and adult phases had slopes that did not significantly differed from 0 (Table III-5 (c)) but shifts that differentiated them (respectively about 3, 4 and 3 hours; Table III-7 and

Appendix III Figures SIII-3 (a), (b) and (c)). Regarding SALK\_126071, comparison of the estimated slopes showed that the characteristic growth duration differentiated the seedling and juvenile phases, but barely the juvenile phase from the transition one (Table III-5 (c) and Appendix III Figure SIII-3 (d)).

The fact that the final leaf area increases along the shoot indicates that the corresponding reduction of the expansion duration was overcome by a stronger increase of the expansion rate along the shoot.

| <b>(a) Length-to-width ratio</b>          | Seedling phase |      | Juvenile phase | Transition phase |      | Adult phase |      |
|---|----------------|------|----------------|------------------|------|-------------|------|
| Col-0                                     | -2.4           | n.s. | 8.5            | 8.6              |      | 9.3         |      |
| SALK_055458                               | 1.8            | n.s. | 6.8            | 6.5              |      | 1.7         | n.s. |
| SALK_048174                               | 1.2            | n.s. | 9.9            | 12.2             |      | 12.4        |      |
| SALK_126071                               | 3.0            |      | 6.0            | 1.8              |      | -           |      |
| <b>(b) Final leaf area</b>                | Seedling phase |      | Juvenile phase | Transition phase |      | Adult phase |      |
| Col-0                                     | -0.6           | n.s. | 28.7           | 39.2             |      | 16.8        |      |
| SALK_055458                               | 0.6            | n.s. | 19.2           | 22.2             |      | 16.8        |      |
| SALK_048174                               | 10.9           |      | 26.7           | 31.3             |      | 16          | n.s. |
| SALK_126071                               | 2.1            |      | 16.3           | 8.8              |      | -           |      |
| <b>(c) Characteristic growth duration</b> | Seedling phase |      | Juvenile phase | Transition phase |      | Adult phase |      |
| Col-0                                     | -0.1           | n.s. | -1.2           | -0.7             | n.s. | 0.5         | n.s. |
| SALK_055458                               | 1.2            | n.s. | 2.2            | -0.1             | n.s. | -0.4        | n.s. |
| SALK_048174                               | 2.3            |      | -1.0           | 0.1              | n.s. | 0.6         | n.s. |
| SALK_126071                               | -1.2           | n.s. | -0.7           | -0.8             |      | -           |      |

Table III-5. Slopes estimated for each quantitative variable and for each successive developmental phase. When the estimated slope was not significantly different from 0, 'n.s.' was indicated near to the value.

#### 3.2.4. Developmental phases are markedly synchronous between the individuals of a given genotype

The 'left-right' SMSMs were built based on the hypothesis that the developmental phases were asynchronous between the individuals for a genotype. Thus, the model built for a given genotype enabled to segment individually each multivariate sequences corresponding to a rosette of this genotype. As a byproduct of the model building, we were able to compute the probability of the developmental phases as a function of the leaf rank, allowing the evaluation

of the degree of synchronism between the individuals of a genotype. It has to be noted that this evaluation takes into account both within- and between-individual variation – i.e. the variations between the different possible segmentations of an individual and the variations between the different individuals. Figure III-2 shows the probability graphs drawn for each genotype. In the four lines, the overlap between the seedling and juvenile phases was inexistent or very minor. It was also slight between the juvenile and transition phases, with only one leaf rank where these two phases were equally probable. It shows the synchronicity of all developmental phases between the individuals for SALK\_126071 (as the succession was only composed of these three developmental phases; Figure III-2 (d)), and of the seedling and juvenile phases for the three other genotypes. For Col-0 and SALK\_048174 the occurrence of the adult phase was also quite synchronous among the individuals (Figure III-2 (a) and (c)). A larger overlap was observed between the transition and adult phases for SALK\_055458, reflecting a lower synchronicity between the individuals for these two phases (Figure III-2 (b)). Overall, the developmental phases were markedly synchronous between the individuals of a given genotype.

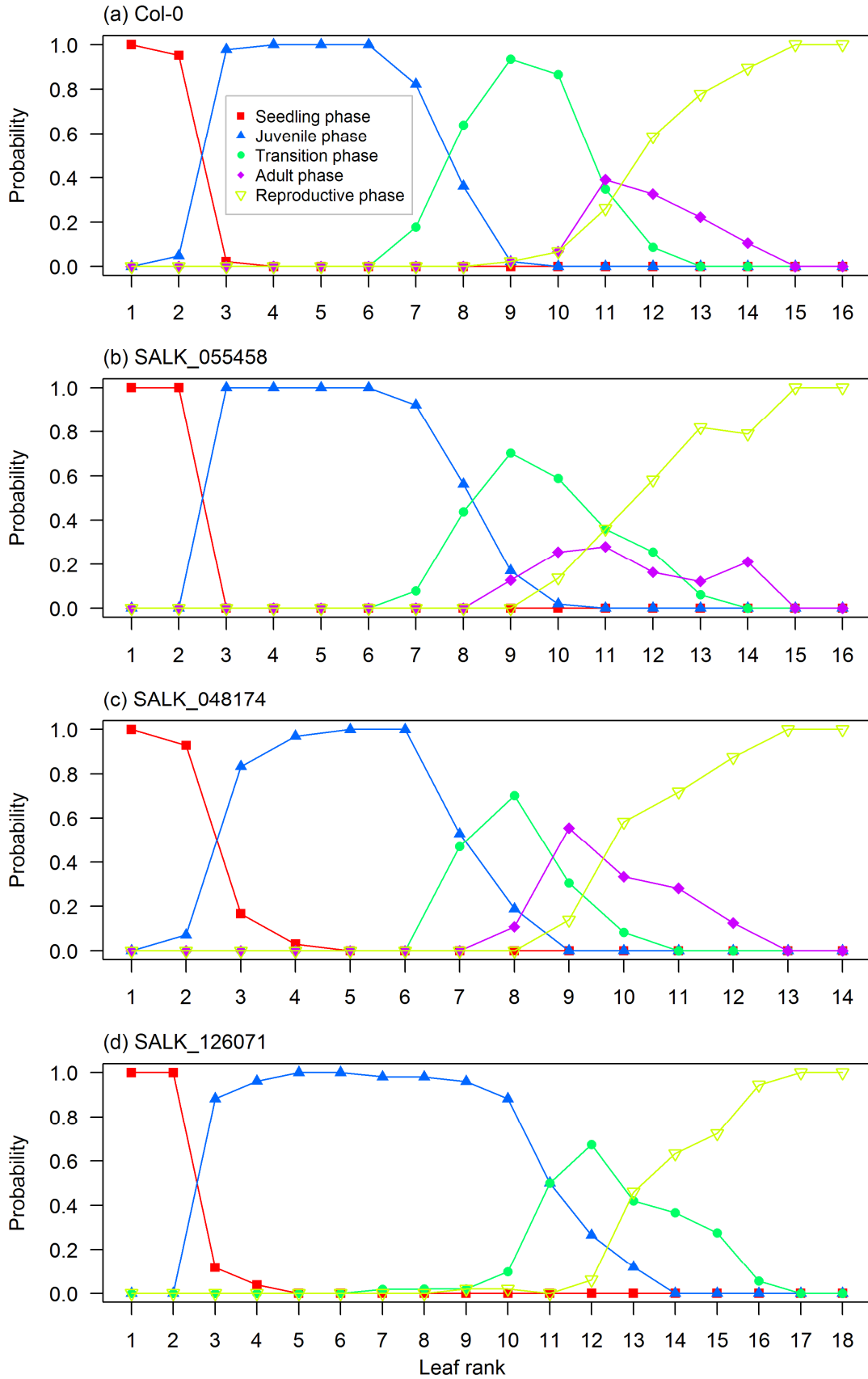


Figure III-2. Probability of the developmental phases as a function of the leaf rank for each genotype: (a) Col-0, (b) SALK\_048174, (c) SALK\_055458 and (d) SALK\_126071. Red: seedling phase. Blue: juvenile phase. Green: transition phase. Purple: adult phase. Lemon green: reproductive phase.

### 3.2.5. *The length of the juvenile and transition phases varies between the genotypes*

The average length of each developmental phase, expressed in number of successive leaves, is given in Table III-6. It showed that the seedling phase was mostly composed of the first two leaves, even if for a small proportion of the plants, it had 3 leaves because of missing data due to senescent leaves at the first or second leaf rank. The length of the adult phase was also similar (a bit more than 1 leaf in average) for the three concerned genotypes. However, differences were found between the juvenile and transition phases estimated for each line.

Regarding these outputs, SALK\_055458 had the phase lengths closest to Col-0 ones. The mean length for each developmental phase was almost exactly the same for this mutant and its wild-type (Table III-6), resulting in similar final number of rosette leaves for both genotypes (Figure III-3 (c)). SALK\_048174 estimated juvenile phase was similar to the one found in Col-0. However, the transition phase defined by SALK\_048174 model was in average about 1 leaf shorter than in Col-0 plants. In SALK\_126071, the juvenile phase was composed of 8 leaves on average, making it far larger than the one identified in Col-0 (Figure III-2, Table III-6). As for the transition phase, it seemed to be not impacted by the T-DNA insertion, the number of leaves assigned to it being less than 1 leaf lower than in Col-0 dataset.

|             | Mean length of phase in number of leaves $\pm$ standard deviation |                |                  |               |
|-------------|---|----------------|------------------|---------------|
|             | Seedling phase  | Juvenile phase | Transition phase | Adult phase   |
| Col-0       | 2.0 $\pm$ 0.2   | 5.0 $\pm$ 1.0  | 2.9 $\pm$ 1.0    | 1.3 $\pm$ 0.5 |
| SALK_055458 | 2.0 $\pm$ 0.0   | 5.5 $\pm$ 1.0  | 2.4 $\pm$ 1.0    | 1.2 $\pm$ 0.4 |
| SALK_048174 | 2.1 $\pm$ 0.5   | 4.4 $\pm$ 1.1  | 1.5 $\pm$ 0.6    | 1.4 $\pm$ 0.5 |
| SALK_126071 | 2.2 $\pm$ 0.5   | 8.2 $\pm$ 1.4  | 2.2 $\pm$ 0.9    | -             |

Table III-6. Mean length of each developmental phase and associated standard deviation for each genotype. Phase length is expressed in number of successive leaves.

### 3.3. **Effect of the mutations on the values of leaf blade length-to-width ratio, final leaf area and characteristic growth duration**

In addition to altering the organization of plant shoot development, the mutations of the SALK lines could also affect the leaf growth traits themselves. The asynchronous segmentations of the plants into successive developmental phases allowed us to compare leaves that had developed during the same developmental phase. That overcame the problem of comparison of leaves rank by rank that can be irrelevant when developmental phases are asynchronous.

| (a) Length-to-width ratio (x100)   |       | Seedling phase |        | Juvenile phase |        | Transition phase |        | Adult phase |        |
|------------------------------------|-------|----------------|--------|----------------|--------|------------------|--------|-------------|--------|
|                                    |       | Lower          | Higher | Lower          | Higher | Lower            | Higher | Lower       | Higher |
| Col-0                              | Rank  | 1              | 2      | 3              | 8      | 7                | 12     | 10          | 14     |
|                                    | Value | 117            | 115    | 123            | 166    | 167              | 209    | 192         | 230    |
| SALK_055458                        | Rank  | 1              | 2      | 3              | 9      | 7                | 13     | 9           | 14     |
|                                    | Value | 113            | 115    | 119            | 160    | 159              | 198    | 192         | 200    |
| SALK_048174                        | Rank  | 1              | 2      | 3              | 8      | 7                | 10     | 8           | 12     |
|                                    | Value | 118            | 120    | 124            | 173    | 178              | 214    | 199         | 249    |
| SALK_126071                        | Rank  | 1              | 3      | 3              | 13     | 10               | 16     |             |        |
|                                    | Value | 123            | 129    | 127            | 187    | 179              | 190    |             |        |
| (b) Final leaf area                |       | Seedling phase |        | Juvenile phase |        | Transition phase |        | Adult phase |        |
|                                    |       | Lower          | Higher | Lower          | Higher | Lower            | Higher | Lower       | Higher |
| Col-0                              | Rank  | 1              | 2      | 3              | 8      | 7                | 12     | 10          | 14     |
|                                    | Value | 24             | 25     | 46             | 190    | 166              | 362    | 315         | 382    |
| SALK_055458                        | Rank  | 1              | 2      | 3              | 9      | 7                | 13     | 9           | 14     |
|                                    | Value | 19             | 19     | 36             | 151    | 115              | 248    | 179         | 263    |
| SALK_048174                        | Rank  | 1              | 2      | 3              | 8      | 7                | 10     | 8           | 12     |
|                                    | Value | 19             | 30     | 50             | 184    | 187              | 281    | 263         | 327    |
| SALK_126071                        | Rank  | 1              | 3      | 3              | 13     | 10               | 16     |             |        |
|                                    | Value | 16             | 20     | 24             | 187    | 182              | 235    |             |        |
| (c) Characteristic growth duration |       | Seedling phase |        | Juvenile phase |        | Transition phase |        | Adult phase |        |
|                                    |       | Lower          | Higher | Lower          | Higher | Lower            | Higher | Lower       | Higher |
| Col-0                              | Rank  | 1              | 2      | 3              | 8      | 7                | 12     | 10          | 14     |
|                                    | Value | 49.0           | 48.9   | 47.8           | 41.6   | 41.6             | 38.2   | 35.6        | 37.6   |
| SALK_055458                        | Rank  | 1              | 2      | 3              | 9      | 7                | 9      | 9           | 14     |
|                                    | Value | 54.8           | 55.9   | 55.6           | 42.6   | 42.2             | 41.7   | 38.1        | 35.9   |
| SALK_048174                        | Rank  | 1              | 2      | 3              | 8      | 7                | 10     | 8           | 12     |
|                                    | Value | 45.2           | 47     | 47.8           | 42.8   | 40.4             | 40.6   | 37.4        | 39.6   |
| SALK_126071                        | Rank  | 1              | 3      | 3              | 13     | 10               | 16     |             |        |
|                                    | Value | 42.8           | 40.3   | 46.0           | 39.4   | 39.5             | 34.9   |             |        |

Table III-7. The lower and the higher ranks of the leaves modeled in each developmental phase and the corresponding values predicted by the linear models are shown for each genotype. (a) Leaf blade length-to-width ratio (x100); (b) Final leaf area (mm<sup>2</sup>); (c) Characteristic growth duration (hours).

SALK\_055458 leaves exhibited a lower blade length-to-width ratio than Col-0 leaves, indicating a tendency to be more rounded, except for the ones of the seedling phase (Table III-7 (a)). Their characteristic growth duration increased during the seedling and the juvenile phases. This trait was also decreased in transition and adult leaves – except for the last formed

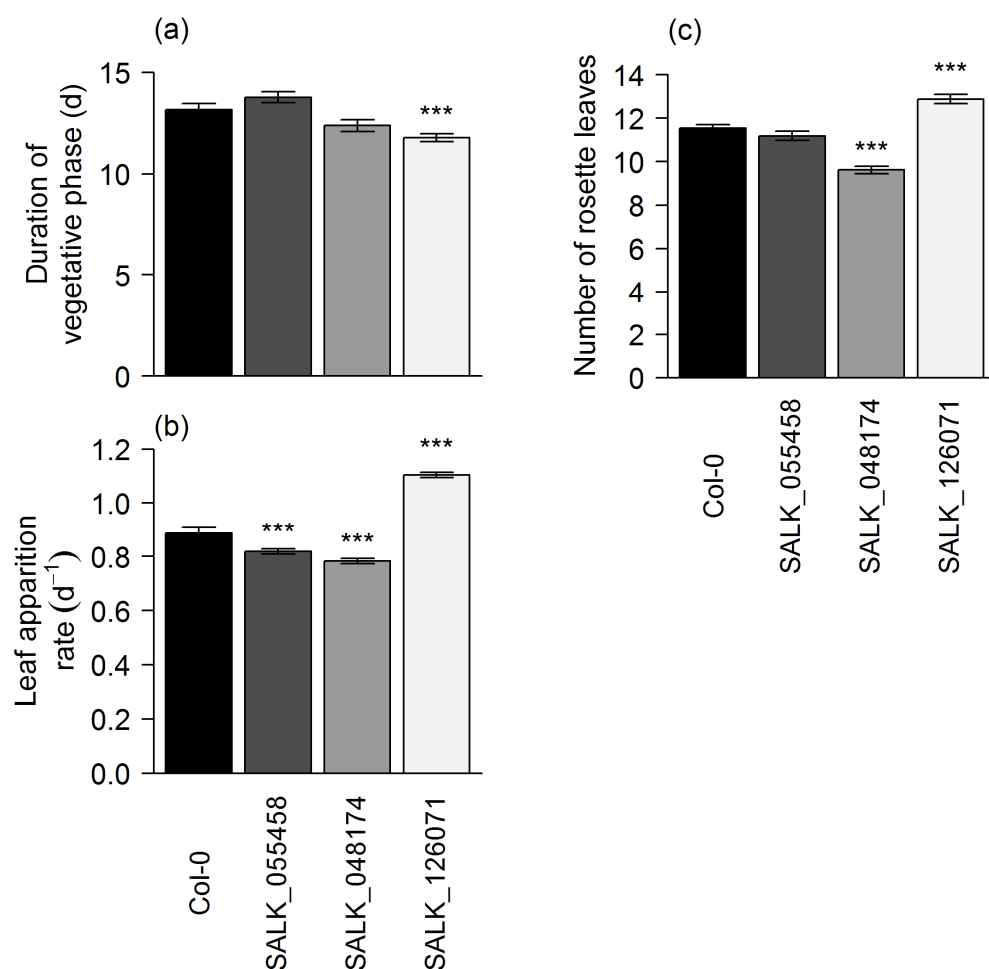


Figure III-3. (a) Duration of the vegetative phase (days after leaf 2 emergence), (b) mean leaf apparition rate (leaf day<sup>-1</sup>) and (c) final number of rosette leaves of Col-0 and the 3 T-DNA lines. Means with error bars are shown for each genotype ( $n \approx 50$ ). Stars indicate when the T-DNA lines value is significantly different in comparison to Col-0 (Student t-tests): \*,  $0.01 < p\text{-value} < 0.05$ ; \*\*,  $0.001 < p\text{-value} < 0.01$ ; \*\*\*,  $p\text{-value} < 0.001$ .

leaves – but less markedly (Table III-7 (c)). Despite this higher duration of expansion within each phase, SALK\_055458 leaves were smaller than wild-type ones (Table III-7 (b)), implying a decrease in individual leaves expansion rate.

Contrary to SALK\_055458, SALK\_048174 leaves were slightly more elongated than Col-0 ones, except for the seedling leaves, as indicated by the measurements of leaf blade length-to-width ratio. Final leaf areas values were similar to Col-0 values except for adult leaves that were smaller in the mutant plants than in the wild-type. In a similar way, the mutation did not affect the characteristic growth duration of the leaves, except for the adult ones, that had a slightly extended duration of expansion. This indicates that leaves assigned to the adult phase must have a lower expansion rate in the mutant than in the wild-type.

In SALK\_126071 plants, length-to-width ratio was slightly higher than in Col-0 for the

first-emerged leaves but the increase between the lower and the higher nodes was less marked, explaining that the last emerged leaves exhibited more rounded blades in the mutant than in the wild-type. As in SALK\_055458, plants exhibited smaller leaves than Col-0 plants at every leaf rank. The seedling leaves were the only ones for which the characteristic growth duration was clearly affected, with a shortened duration of expansion. It also had a tendency to be reduced within the juvenile and transition phases but it was not as pronounced.

### **3.4. Analysis of global plant shoot development: effects of the mutations on the duration of the vegetative phase, mean leaf apparition rate and final number of leaves**

SALK\_055458 had a significantly lower mean leaf apparition rate compared to Col-0 and the two genotypes showed no significant differences regarding the duration of their vegetative phase (Figure III-3 (a) and (b)). However, the mutant line exhibited a similar final number of rosette leaves (Figure III-3 (c)). This could be explained by the fact that even if the slight lengthening of the vegetative phase in the mutant was not tested as being significant, it was likely still enough to compensate the decrease of the mean leaf apparition rate.

As SALK\_055458 plants, SALK\_048174 ones exhibited a significantly lower mean leaf apparition rate (Figure III-3 (b)). Since it was not compensated by a lengthening of the vegetative phase, the mutant exhibited an average final number of rosette leaves significantly lower than Col-0 (Figure III-3 (a)).

Comparison of vegetative phase durations showed that SALK\_126071 plants took less time than Col-0 ones to reach the vegetative-to-reproductive transition (Figure III-3 (a)). However, the mutant plants also exhibited a largely higher mean leaf apparition rate, leading to an increase of the final number of rosette leaves (Figure III-3 (b) and (c)).

## **4. DISCUSSION**

A developmental phase corresponds to a period of time during which an organism or an organ displays specific morphological and physiological traits. The determination of these phases generally relies on the observation of a discontinuity in the development, but the transition between two successive developmental phases can be subtle. When growth and development are considered together, pre-defined developmental stages are generally used as stone-marks for the collection of growth data (e.g. Kalu and Fick 1981; Borreani et al. 2003). A recent study on mango (*Mangifera indica* L.) highlighted a strong coordination between



growth and development of shoot growth unit – i.e. the vegetative axis and its attached leaves – and the inflorescence (Dambreville et al. in press). Here, we used two morphological traits (abaxial trichomes patterns, leaf blade length-to-width ratio) combined with a dimensional trait (final leaf area) and a dynamic trait (the characteristic growth duration) to identify successive developmental phases in *A. thaliana* rosette. In this aim, a pipeline of methods combining semi-automatic image analyses and SMSMs was developed. The different parameters of SMSMs – transitions probabilities between the different phases, phase length distributions, and observation models for each trait within each developmental phase – allowed us to quantitatively analyze how *A. thaliana* rosette structure was affected by the mutations in the different genotypes.

### **4.1. The successive developmental phases structuring *A. thaliana* rosette are determined by a combination of leaf growth traits**

Vegetative development in plants is generally divided into two developmental phases: juvenile and adult phases (Poethig 2003). However, heteroblastic variation often follows more complex patterns. Our approach enabled the identification and characterization of a first developmental phase that was strongly conserved between the four studied genotypes despite their difference in whole shoot development previously reported in Chapter II. This phase is composed of the first two postcotyledonary leaves that share the same characteristics regarding the various measured leaf growth traits. It is discriminated from the following phase by a change in slope for the three quantitative variables. The identification of a seedling phase is consistent with similar observations made in previous studies in *A. thaliana* (Telfer et al. 1997; Steynen et al. 2001) and in a variety of other species such as *Desmodium paniculatum* (Wulff 1985), *Fraxinus pennsylvanica* (Merrill 1986), *Pseudopanax crassifolius* (Gould 1993), *Eucalyptus* (Boland et al. 2006) or wheat and barley (Dornbusch et al. 2010). We chose to refer to this developmental phase as the ‘seedling phase’ in accordance with some previous publications (Gould 1993; Boland et al. 2006). As discussed in the previous chapter, this phase could correspond to the preformation of leaves during embryogenesis, these leaves being thus subjected to the seed environment (Bongard-Pierce et al. 1996; Jones 1999).

As previously stated, two main developmental phases constitute the vegetative development – the juvenile and the adult phases. However studies of heteroblasty highlighted a transition phase, in several species, including species in which the heteroblastic changes along the shoot are subtle such as *A. thaliana*. This phase corresponds to the gradual change from juvenile leaves characteristics to adult leaves characteristics (Bongard-Pierce et al. 1996;

Kerstetter and Poethig 1998; Jones 1999). In addition to the seedling phase, these three phases – juvenile, transition and adult phases – were identified using SMSMs in Col-0, SALK\_055458 and SALK\_048174 phases, in accordance with the literature.

Commonly, in *A. thaliana*, the presence or absence of abaxial trichomes is the only marker used to define developmental phases. Here we showed that even if this variable is quite determinant for structuring the plant shoot in successive developmental phases, depending on the studied genotypes it may not be sufficient to define an accurate structuring. Indeed, this trait cannot be used to discriminate the seedling phase from the juvenile phase. In Col-0 the juvenile, transition and adult phases correspond exactly to the three categories defined for this trait – i.e. no covering, partial covering and full covering of the leaf abaxial surface with trichomes. Length-to-width ratio is the less useful variable to decipher the structuring of plant shoots, whereas the combined analysis of final leaf area and characteristic growth duration complete the information given by the abaxial trichome observation. Nevertheless, the matching between these developmental phases and the trichome variable categories were not as perfect for SALK\_055458 and SALK\_048174 indicating a role of at least some of the quantitative traits used to build SMSMs. As a consequence, the mean phase lengths estimated using only the abaxial trichome categories (see Chapter II section 3.4.2) and the ones estimated within SMSMs show some differences. These differences, even if they are slight, led to different conclusions concerning the effect of T-DNA insertions in the rosette structuring.

SALK\_126071 was a particular case in our study, since only three developmental phases were identified using SMSM. The seedling phase was still identifiable. In the three other genotypes the discrimination between the juvenile, transition and adult phases strongly matched with the three abaxial trichome categories. However, only a very few SALK\_126071 plants presented leaves with abaxial surfaces fully covered with trichomes. We therefore considered that the missing phase was the adult one. Since only a few leaves had abaxial surfaces fully covered with trichomes, an adult phase might be estimated if the experiment was conducted with a higher number of individuals. This genotype also stood out from the others by the fact that the discrimination between the juvenile and the transition phases cannot be done solely using the abaxial trichome variable either, since 25 % of the leaves assigned to the transition phase using the estimate SMSM do not have any trichome on their abaxial face.

The analysis of molecular actors that have been recently identified as regulating the vegetative phase change – the microRNAs miR156 and miR157 and their direct targets the

*SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (SPL) family of transcription factors – showed that they are differentially expressed in juvenile and adult tissues of a wide range of plants (Wu and Poethig 2006; Wang et al. 2008; Poethig 2013). Analyzing their expression is thus a reliable way to investigate vegetative phase change. Another reliable method to determine the onset of vegetative phase change is to conduct reciprocal transfer experiments, as it enables the determination of the acquisition of the reproductive competency. This method involves regularly transferring plants between inductive and non-inductive growth conditions (for more details, see Mozley and Thomas 1995; Adams et al. 2003; Matsoukas et al. 2013). However it can be laborious to apply these methods in high-throughput analyses. Since the vegetative phase change is marked by coordinated changes, sometimes subtle, in the leaves characters such as leaf shape and size, patterns of trichome covering, vascular patterns, epicuticular wax and phyllotaxy, these changes are sometimes used as markers of the juvenile vegetative to adult vegetative phase. Therefore, vegetative phase change has often been conflated with the concept of heteroblasty, sometimes to such an extent that in some cases morphological changes can be ascribed to the wrong process. For example, in *A. thaliana*, change in the covering of the leaf abaxial surface with trichomes is one of the more marked changes, making it the most commonly used marker of the vegetative phase change in this species. However this variable can be altered independently of the reproductive competency acquisition by some mutations, and is therefore not completely reliable. Thus, depending on the traits and on the genotypes investigated, results on defining the onset of the vegetative phase change and the number of nodes this transition encompasses can differ. This matter has been pointed out in several reviews (Jones 1999; Zotz et al. 2011; Poethig 2013; Matsoukas 2014). It is due to the involvement of other phenomena – such as aging, size increasing (Day et al. 2002; Mencuccini et al. 2007) and environmental heterogeneity – in the establishment of the heteroblasty. Since our pipeline of analysis methods determines developmental phases based on several leaf traits, it reflects the plant heteroblasty in its broadest sense. Associated to reliable measurements of the vegetative phase change onset, it could be used to study the links and possible uncoupling between these two processes.

#### **4.2. Our analysis method gives new insights concerning the effect of the mutations in T-DNA insertions lines.**

Integrating different leaf traits – morphological, dimensional and dynamic – at the shoot scale allows obtaining a more global view of the alternations a plant can undergo. The identification of developmental phases structuring *A. thaliana* rosette using SMSMs enables

comparing changes in leaf growth traits among leaf successions with a biological meaning (i.e. belonging to same developmental phase) rather than comparing the alterations at one arbitrary chosen leaf rank. Our approach also enables to determine possible changes in the importance of the structuring role of the respective leaf growth traits, possibly creating interesting trails for research on heteroblasty control.

Concerning the genotypes analyzed during this study, we showed that SALK\_055458 was the line that had the closest shoot profile compared to the Col-0 wild-type. Nevertheless, we observed in this line a delay in bolting time, an increased phyllochron – i.e. more time elapses between the emergences of two successive leaves than in Col-0 – and smaller individual leaves. SALK\_055458 is a knockout line for At1g14040. This gene encodes a PHO1 (phosphate exporter) homolog protein (PHO1;H3). A recent publication (Khan et al. 2014) has identified *PHO1;H3* as being involved in the crosstalk between zinc deficiency signaling and the regulation of phosphate homeostasis in *A. thaliana*. The phenotypic alterations that were observed in the mutant are therefore likely to be due to some imbalance in these elements homeostasis.

SALK\_048174 T-DNA is located in At4g11540, a gene coding for a Cysteine/Histidine-rich C1 domain family protein. Nothing has been published so far on the function of this protein. It is annotated as having a protein-disulfide reductase activity and being involved in intracellular signal transduction and oxidation-reduction process (The Arabidopsis Information Resource 2014). According to our results, we can assume that it is involved – directly or not – in mechanism(s) controlling leaf initiation during the vegetative development of the rosette, resulting to an extended phyllochron. The analysis of the segmentation model indicates that this decrease in the number of rosette leaves in SALK\_048174 only concerned leaves initiated during the transition phase. The effect of the mutation could thus specifically affect the activity of the shoot apical meristem during the juvenile-to-adult transition. A more precise follow-up of the phenology would be necessary to test this hypothesis.

SALK\_126071 plants were the ones presenting the strongest differences when compared to Col-0 rosette structuring. In this line, the T-DNA is inserted in an exon of At1g04730, a protein coding gene also named *CHROMOSOME TRANSMISSION FIDELITY 18 (CTF18)*. This gene is mainly expressed in the shoot apical meristem during the whole plant development (vegetative and inflorescence apex). Takahashi et al. (2010) demonstrated that the corresponding protein contributes to sister chromatid cohesion in *A. thaliana* cells, in accordance with what was determined in other eukaryotes. Takahashi's study did not report

any effect of the mutation on leaf size. However plants were not grown in the same conditions as those applied in our experiment and measurements were made earlier in the plant development (21-day-old plants). Nevertheless, they showed that plants with aggravated sister chromatid cohesion defect exhibited smaller leaf areas due to restricted cell division. The decrease in individual leaf area that we observed in our experiment could therefore be due to an enhanced effect of the mutation in our growth conditions. It could also be related to the reduction in the duration of the plant vegetative development, as leaves would have less time to expand after their initiation. Here, comparison with Col-0 showed that SALK\_126071 plants had an increased mean leaf apparition rate and an increased final number of rosette leaves. This increase in final number of leaves can be mainly attributed to the lengthening of the juvenile phase alone whereas no adult phase could be estimated. The mutation thus strongly alters the plant vegetative development. That could be directly linked to the activity of CTF18 in the shoot apical meristem. The decreased time necessary for the plants to reach bolting might be a compensation phenomenon. It is interesting to observe that some alterations of SALK\_126071 phenotype – lengthened juvenile phase, increased final number of leaves and increased leaf apparition rate – are similar to the ones observed in plants overexpressing miR156 (Schwab et al. 2005; Wu and Poethig 2006; Wang et al. 2008, 2009; Wu et al. 2009; Shikata et al. 2009). However, in these studies the flowering time was delayed, whereas SALK\_126071 plants rather exhibited shortened vegetative phases. Therefore, even if there might be links between pathways involving these genes, they are most likely indirect.

In conclusion, the establishment of developmental phases based on the analysis of morphological, dimensional and dynamic leaf traits integrated at the shoot scale enabled the determination of the structuring role of the different variables in *A. thaliana* shoot development and a more accurate characterization of the effect of the T-DNA insertions at the macroscopic scale.

## **CHAPTER IV. A NEW PIPELINE OF METHODS COMBINING MACHINE LEARNING APPROACHES FOR IMAGE SEGMENTATION AND GAMMA MIXTURE MODELS FOR ANALYZING CELL DIMENSIONS IN *A. THALIANA* LEAF EPIDERMIS**

### **1. INTRODUCTION**

Growth and development of multi-cellular organisms are characterized by a complex coordination of cell division and cell expansion (Donnelly et al. 1999; Rymen et al. 2010). In plants, these cellular processes have been tracked over time and space over simple leaves (as opposed to compound leaves) and roots. Detailed kinematic analyses have revealed complex and tightly controlled spatial and temporal patterns of cell division and cell expansion and how they are modified in genetic variants or upon environmental changes (Bernstein et al. 1993; Granier and Tardieu 1998; Granier et al. 2000; De Veylder et al. 2001; Beemster et al. 2005; Cookson et al. 2005; Fiorani and Beemster 2006). Studies linking molecular data and kinematically determined cellular growth variables with similar spatial and temporal resolution have given insights into the genetic networks controlling cell proliferation and growth (Kalve et al. 2014).

Techniques and frameworks of analyses have been developed to quantitatively and dynamically assess the cellular processes underlying shoot organ formation and growth (Donnelly et al. 1999; Rymen et al. 2010). Using these methods, cellular behaviors underlying leaf shape and size variations have been assessed in large collection of mutants, populations of recombinant inbred lines and naturally occurring accessions, giving insights into the genetic control of the cellular variables as well as their coordination and role during leaf development (Horiguchi et al. 2006b; Tisné et al. 2008; Massonnet et al. 2011; Pérez-Pérez et al. 2011; Sterken et al. 2012). In most cases, cell density, cell area and cell number are determined on epidermal peels, imprints or paradermal views of cleared leaves. These measurements are generally limited to the upper epidermis mainly because of technical constraints, but also because the epidermis is often considered as the tissue physically limiting whole leaf expansion (Savaldi-Goldstein and Chory 2008; Marcotrigiano 2010). Mutants with strong changes in mesophyll cell density, but no change in leaf size have been described, whereas epidermal cell size or number are often positively correlated with leaf area (González-Bayón et al. 2006; Tisné et al. 2008; Pérez-Pérez et al. 2011). Some genotypes

however present strong changes in epidermal cell number and size without exhibiting any change in leaf area (Tisné et al. 2011; Massonnet et al. 2011). For example, a drastic modification in epidermal cell density is caused by the *erecta* mutation without any effect on leaf size Figure IV-1. Cell densities are higher for both the adaxial (2x) and abaxial epidermis (3x) very early on in leaf development (5 days after initiation). As a consequence, leaf area cannot be used as a proxy of epidermal cellular variables, highlighting the interest of investigating the tissular scale in parallel to macroscopic traits.

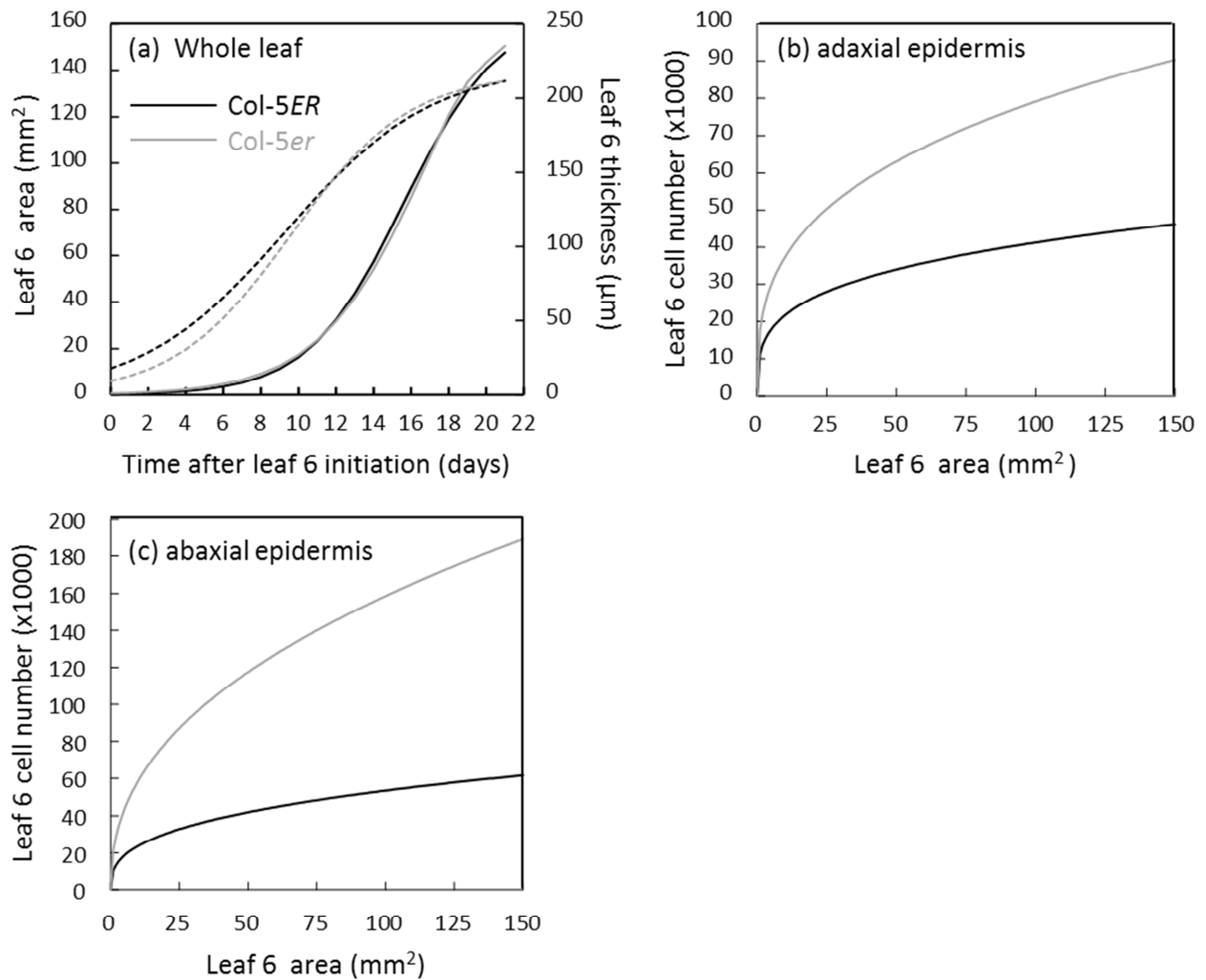


Figure IV-1. Phenotyping cell density in a given tissue does not necessarily reflect the cellular phenotype of other tissues: example of the effects of the *erecta* mutation on cell number in different leaf tissues. Changes over time in leaf surface area (full lines) and thickness (dotted lines) (a), for the 6<sup>th</sup> leaf of *A. thaliana* Col-5 wild-type (Col-5ER, black) and Col-5 harbouring the *erecta* mutation (Col-5er, grey). Cell number increases during leaf blade expansion are shown for the adaxial epidermis (b) and the abaxial epidermis (c) for the two genotypes, revealing a strong effect of the *erecta* mutation on the dynamics of cell division in these tissues. From Lièvre et al. 2013.

Even if phenotyping shoot growth and development at the macroscopic scale is not a real bottleneck anymore in *A. thaliana* due to the range of technologies that have been developed (see Chapter III), there is still a difficulty for now to obtain high-throughput cellular variable datasets. Measurements that are necessary to obtain this type of data are often done manually, and are therefore time-consuming. They are also destructive and usually laborious. This was a first obstacle in the integration of cellular growth related traits in the pipeline of analysis presented in Chapter III. As a consequence, we first developed a protocol enabling the semi-automatic segmentation of leaf epidermis imprints in order to extract cell areas faster and with a limited risk of error due to observer bias. This was made possible by using a new machine learning approach for image segmentation made available during my PhD (Sommer et al. 2011).

Another difficulty in integrating cellular data in our multivariate model comes from the fact that in *A. thaliana*, the epidermis is composed of a mosaic of cells differing substantially in dimensions and presenting a right skewed distribution (see the boxplots in Cookson et al. 2006, see also the Results section in this chapter). In previous leaf growth multi-scale phenotyping studies, cell area in a tissue was commonly summarized by mean (or median) and standard (or mean absolute) deviation (Cookson et al. 2005, 2006; Tisné et al. 2008; Massonnet et al. 2011). However, these statistical descriptors of location and dispersion constitute a poor summary of cell area distributions and are certainly not sufficient for an in depth comparison of genotypes affected in cell cycle or cell expansion for examples. Many studies in different plant species and different organs have reported that the cell dimension mosaic in tissues is at least partly due to the endoreduplication process and then reflects a mosaic in cell DNA content (Boudolf et al. 2004; Vlieghe et al. 2007). Regular cell cycle involves a DNA duplication phase followed by a division phase – i.e. mitosis. Endoreduplication is a cell cycle variant of multicellular eukaryotes in which mitosis is skipped and cells repeatedly replicate their DNA, resulting in cellular polyploidy (Figure IV-2; see De Veylder et al. 2011, for review).

*In vivo* ploidy maps showed that in *A. thaliana* leaf epidermis, this gain in DNA content is positively correlated with pavement cell size (Melaragno et al. 1993; Boudolf et al. 2004). In agreement with these observations, numerous studies highlighted a positive correlation between these two variables (Vlieghe et al. 2005; Dewitte et al. 2007; Tojo et al. 2009; Roeder et al. 2010). We developed a statistical model to analyze cell area distribution in *A*



*thaliana* leaf epidermis as a mixture of overlapping distributions, with the hypothesis that each of these distributions corresponds to a given ploidy level.

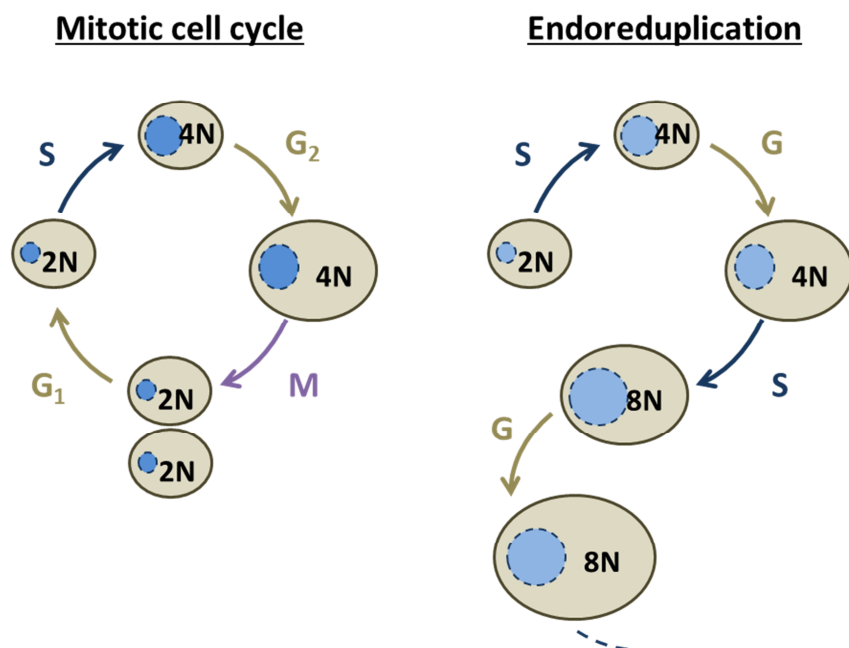


Figure IV-2. Endoreduplication is a modification of the mitotic cell cycle resulting in cellular polyploidy. Ploidy level ( $kN$ ) is indicated in the cells. G<sub>1</sub>, Gap 1 (growth); S, Synthesis (DNA replication); G<sub>2</sub>, Gap 2 (growth); M, Mitosis; G, Gap (growth).

## 2. METHODOLOGICAL DEVELOPMENTS

### 2.1. Plant material and growth conditions

Plants used in this experiment are the Columbia 0 (Col-0) plants grown during Experiment 3 (see Chapter III section 2.1).

Sixty plants of Columbia 0 (Col-0) natural accession of *Arabidopsis thaliana* (L.) Heynh were grown during the experiment. Seeds used were provided by José Luis Micol group (Universidad Miguel Hernandez, Elche, Alicante, Spain).

Seeds were sown in pots filled with a mixture (1:1) of a loamy soil and organic compost. Plants were grown in a growth chamber equipped with the PHENOPSIS automaton under controlled air temperature, air humidity and incident light. Micrometeorological conditions were kept constant during the whole duration of the experiment and homogeneous within the growth chamber (Granier et al. 2006). Light was on during 16 h per day and provided by a bank of cool-white fluorescent tubes and HQi lamps. It was measured at the plant level using a photosynthetic photon flux density (PPFD) sensor (LI-190SB, Li-Cor, Lincoln, NE, USA).

Mean micro-meteorological conditions are indicated in Table IV-1. Soil water content was determined before sowing to estimate the amount of dry soil in each pot. Subsequent changes in pot weight were due to changes in soil water content (the plant weight was considered as negligible as it was never higher than 1 g). This allowed the computation and daily automatic adjustment of soil water content to 0.35 g water g<sup>-1</sup> dry soil by the PHENOPSIS automaton (Granier et al. 2006) with a modified one-tenth-strength Hoagland solution (Hoagland and Arnon 1950), from sowing to the end of the experiments.

|  |      |
|--|------|
| Day length (h)   | 16   |
| Incident PPFD ( $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$ )  | 178  |
| Air temperature (°C)   | 20.4 |
| Air humidity (%)   | 60.6 |
| Soil water content (g H <sub>2</sub> O g <sup>-1</sup> dry soil) | 0.36 |

Table IV-1. Mean meteorological conditions applied during plant growth.

## 2.2. Image acquisition

Plants were harvested at a common stage of plant development, stage 6.00, i.e. first flower open (as defined in Boyes et al. 2001). The successive leaves of the rosette were excised without their petiole, stuck on a sheet of paper and scanned as shown in Figure II-1 (Chapter II). After the leaves were scanned, translucent varnish was spread on the adaxial face of each leaf. The dry varnish made an imprint of the upper epidermis of this leaf. Films were observed with a microscope coupled with the image analysis software Optimas (6.1, Optimas Corporation) and images of two different zones of the leaf were taken for cell area measurement (Figure IV-3).

## 2.3. Image segmentation using ilastik v.0.5.12

To extract the area of each cell in the images, it is first necessary to identify and delimit cells. This process, called segmentation, consists of assigning each pixel of an image to the cell it belongs. Segmentation can either be made manually, automatically or semi-automatically. Automatic segmentation methods are composed of two main steps: 1) classification of the pixels in biologically relevant classes; 2) grouping of the pixels assigned to a given class in one or several connected components. Automatic segmentation methods may include a preliminary training phase in order to build classification rules on the basis of a training set. This training phase may require some manual interventions for labelling

homogeneous regions in tissues. Semi-automatic segmentation methods typically add a post-processing step of manual correction to automatic segmentation methods.

While a manual segmentation method may be less prone to errors because of the observer expertise, it is time consuming (more than 800 images to segment,  $\sim 20$  minutes/image) and is affected by the observer's prior knowledge – i.e. segmentation can change from one observer to the other (Chatelain et al. 2013). An automatic segmentation method is less time consuming than the manual method and is less or not subject to observer's bias but entails errors in pixel classification. Semi-automatic segmentation methods combine the advantages of manual and automatic methods.

Since in our application context it is essential to have as few errors as possible to avoid the side effects of merging or splitting true cells, we chose to design a semi-automatic pipeline of methods for segmentation in order to fasten the process and limit the observer's bias while keeping a final manual correction step to ensure the best possible segmentation. This semi-automatic pipeline of methods for segmentation is based on the ilastik software (Sommer et al. 2011). This software offers a machine-learning algorithm (random forest) that can learn discriminating pixels using a rather extensive set of feature such as pixel color in several spaces (RGB, HSV...), local derivatives (edges) or local entropy (texture).

In our pipeline, the first automatic segmentation that labels the image pixels into three classes (cytoplasm, cell wall and stomates) relies on a random forest classifier implemented in the ilastik software. Since this is a supervised learning method, the images to be segmented have to share common characteristics. In particular, the different classes defined by the observer have to be homogeneous in subsets of images used for the training the classifier.

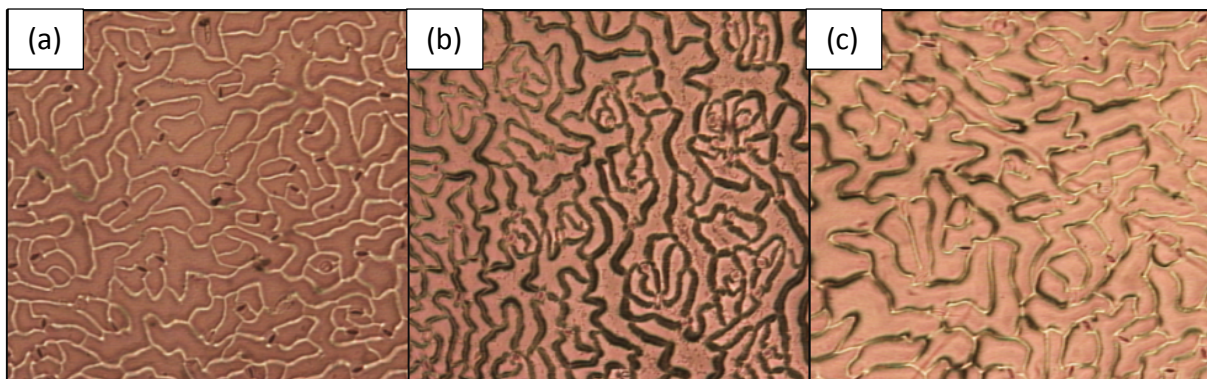


Figure IV-3. Images of leaf upper epidermis imprints belonging to each category: (a) cell walls brighter than the background, (b) cell walls darker than the background, and (c) both situations observable in the same image.

Thus, the first step of our protocol was to establish categories of homogeneous epidermis images on the basis of similarities in the aspect of the background and in the characteristics of cell walls. Epidermis images were grouped into 3 categories: images in which cell walls were brighter than the background, images in which cell walls were darker than the background, and images in which both situations were observable (Figure IV-3).

Once the images were grouped into the three categories, training sets were built independently for each category and epidermis components were manually labelled for the three categories of images (Figure IV-4 (b)). The pixel metrics that were chosen to drive the classifiers relied on color, edge, orientation of the intensity and texture in different neighborhood sizes. These metrics are parameters of the classifiers and are linked to the properties of the epidermis components we want to separate (sharpness of the walls, uniformity of the intensity of cytoplasm pixels ...). The metrics learnt on the basis of the manually labeled pixels were used to assign the rest of the image pixels to the most probable epidermis component class (Figure IV-4 (c)). For each category of images, the training set contained around ten images. The trained random forest classifier was then applied to label the epidermis components in the rest of the images of the category.

Three types of cell components were identified in the images: cell walls, cytoplasm and stomates. However the quality of the images did not enable an accurate delimitation of guard cells. We therefore chose to merge stomates and cell walls together, and only measure areas of pavement cells. Consequently, in the remainder of this section, 'cell walls' will refer to cell walls plus stomates. A morphological smoothing algorithm (binary closing of the cytoplasm) was applied as a first automatic correction step to reassign pixels of cytoplasm that were initially assigned to cell walls by the automatic classification. This filter removed small isolated regions of cell walls. Ultimately, a unique label was attached to each different connected components of the cytoplasm. Each of this connected component thus corresponded to a cell (Figure IV-4 (d)). However, a second correction step had to be made manually, using the image visualization software ImageJ and a Wacom graphic tablet (Figure IV-4 (e)). Maximizing the quality of the ilastik classification with a carefully designed training phase was crucial to reduce the time spent in this step (6 minutes on average per image in our conditions). Finally, from these segmented images in which each pixel was assigned to its proper cell, the areas were computed by counting the number of pixels in the region of the cell and multiplying this number by the image resolution (Figure IV-4 (f) and (h)). During this step, censored and uncensored cells were automatically discriminated,

because they should be considered separately in the statistical analysis (Figure IV-4 (g)).

In order to validate this semi-automatic method of segmentation, we compared the results obtained using this method and those obtained using a manual segmentation for a subset of images (see Appendix IV).

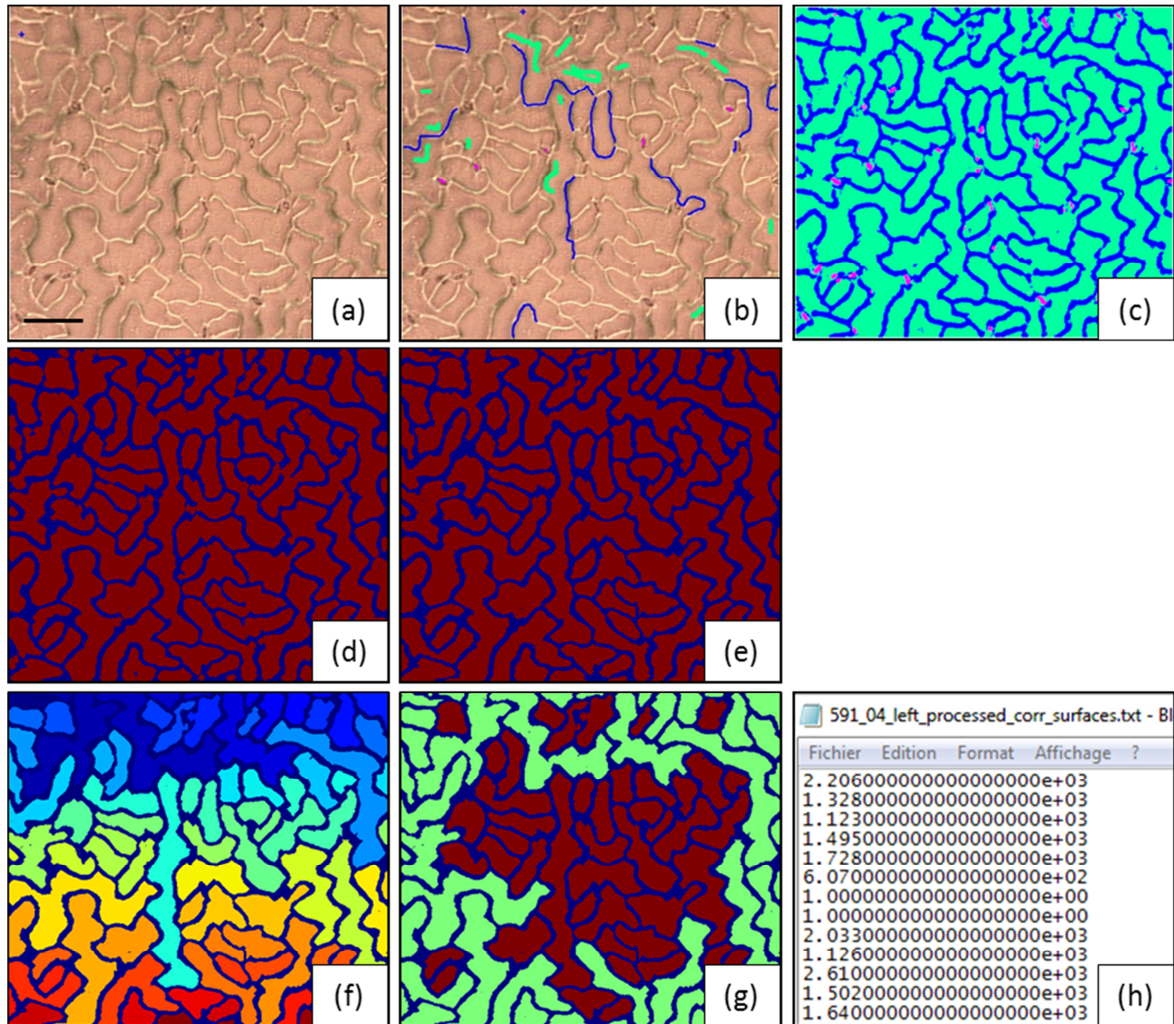


Figure IV-4. The successive steps of the semi-automatic protocol of leaf epidermis image segmentation are illustrated in the different panels. (a) Captured image. (b) Captured image plus manual labelling. Green: cytoplasm, blue: cell walls, pink: stomates. (c) Result of ilastik segmentation. (d) Classification after automatic correction – i.e. merging of cell walls and stomates, and morphological filtering. (e) Classification after manual correction. (f) Identification of individual cells. (g) Discrimination of censored and uncensored cells. (h) Output file after cell area computation. Scale bar: 100  $\mu\text{m}$ .

#### 2.4. Gamma mixture models incorporating a scaling rule between components

The empirical distribution of cell area is roughly unimodal but highly right-skewed; see the histogram in Figure IV-5. If we assume that the high dispersion is mainly explained by endoreduplication with different ploidy levels (e.g. 2C, 4C, 8C and 16C), this entails that the cell areas cannot be partitioned into disjoint classes, each corresponding to a given ploidy level. The empirical distribution of cell area is thus a mixture of overlapping distributions, each corresponding to a given ploidy level. Consequently, some area values can correspond to cells of different ploidies (Melaragno et al. 1993). No well-separated humps corresponding to the different ploidy levels are apparent in the empirical distribution (see the histogram in Figure IV-5) which means that a mixture model is not easily identifiable on this dataset (i.e. different mixture models could fit equally well the data).

We thus chose to integrate a scaling rule in the mixture model in order to overcome this

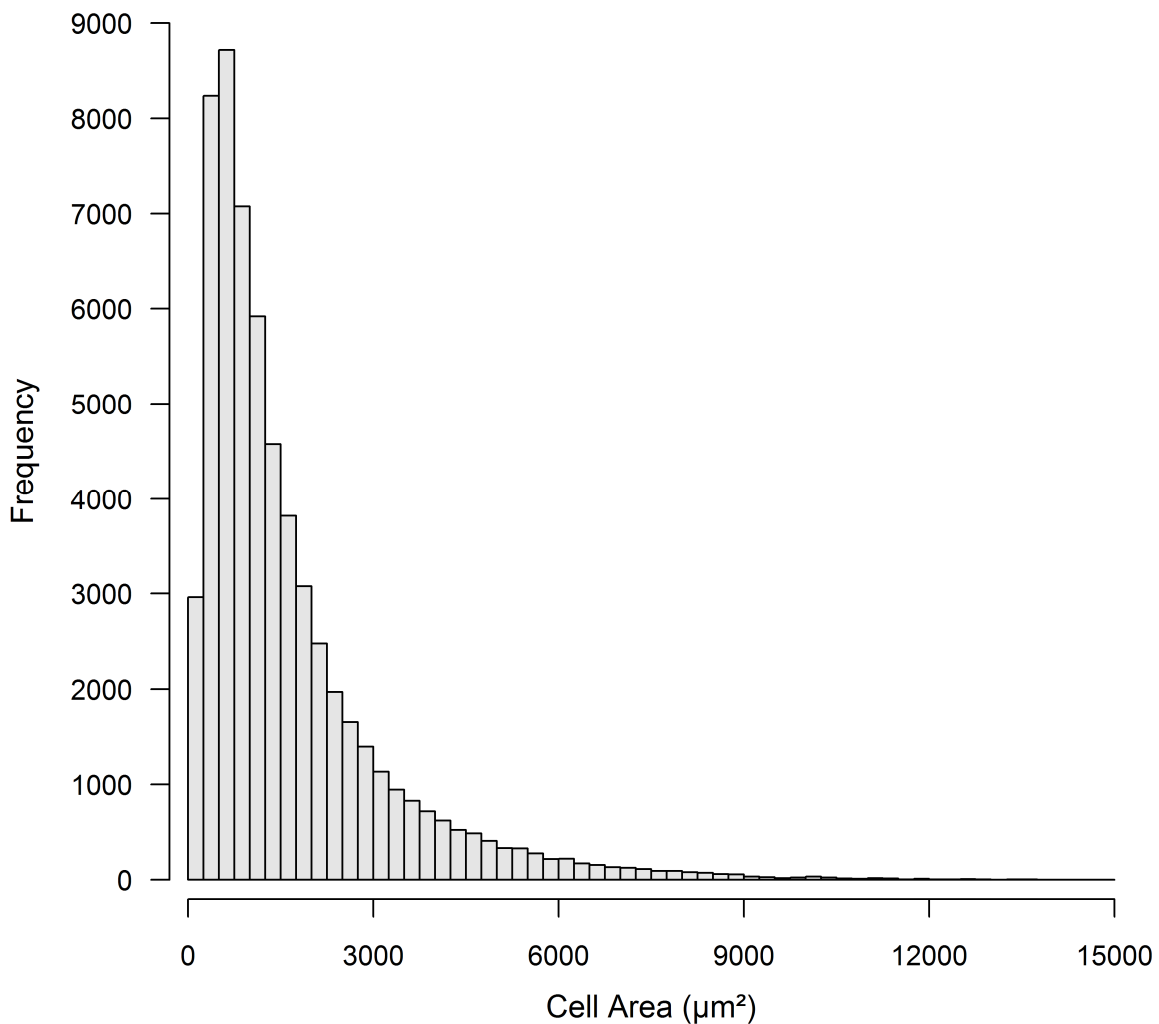


Figure IV-5. Frequency distribution of cell area ( $n=60,286$ ). Cell areas were measured on pictures of leaf epidermis imprints using the semi-automatic segmentation methods presented in section 2.3 of this chapter.



identifiability problem and to test possible assumptions regarding the area of cells with replication of the nuclear DNA. Two scaling rules are standard in probability. This can be illustrated with the example of Gaussian distributions. Let the area of unreplicated cells  $X$  be distributed according to a Gaussian distribution with mean and variance parameters  $\mu$  and  $\sigma^2$ . If  $X \sim N(\mu, \sigma^2)$ , then the two possible assumptions for the cells with replication of the nuclear DNA are:

- $2X \sim N(2\mu, 2\sigma^2)$ ; convolution assumption (sum of independent and identically distributed random variables),
- $2X \sim N(2\mu, 4\sigma^2)$ ; scaling assumption (multiplication of a random variable by a scaling factor).

These two assumptions only differ in the way the dispersion parameter is affected: in the convolution case, the variance is multiplied by the scaling factor while in the scaling case, the standard deviation is multiplied by the scaling factor.

The family of Gaussian distributions is closed under convolution and scaling (in the probabilistic sense) and the parameters of the Gaussian distribution for the different ploidy levels are tied. In this case, the estimation of a mixture model consists of estimating the mean and variance parameters of the Gaussian distribution  $N(\mu, \sigma^2)$  for unreplicated cells on the basis of all the data assuming different ploidy levels for each measured cell area.

We investigated the use of Gaussian distributions but the results were rather unsatisfactory on our data. We then look for another family of parametric distributions closed under convolution and scaling and more appropriate for our application. We identified the family of gamma distributions as an interesting candidate. Two properties of interest of the gamma distributions are the fact that these distributions are defined on positive real numbers (and not on all the real numbers such as Gaussian distributions) and are not constrained to be symmetric but are right-skewed, which is a more realistic assumption for a dimension resulting from a growth process.

The probability density function of the gamma distribution of shape parameter  $\alpha$  and scale parameter  $\beta$  is given by

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \exp\left(-\frac{x}{\beta}\right),$$

with

$$\alpha = \frac{\{E(X)\}^2}{\text{Var}(X)}, \quad \beta = \frac{\text{Var}(X)}{E(X)} \quad (1)$$

Let the area of unreplicated cells  $X$  be distributed according to a gamma distribution of shape parameter  $\alpha$  and scale parameter  $\beta$ . If  $X \sim Ga(\alpha, \beta)$ , then the two possible assumptions for the cells with replication of the nuclear DNA are:

- $2X \sim Ga(2\alpha, \beta)$ ; convolution assumption,
- $2X \sim Ga(\alpha, 2\beta)$ ; scaling assumption.

It should be noted that the scaling rule only affects one parameter (fixed scale parameter and convolution only affecting the shape parameter or fixed shape parameter and scaling, in a probabilistic sense, only affecting the scale parameter).

We also investigated tying parameter assumptions intermediate between independent distributions and full tying of the two parameters. The idea was to let unconstrained the distribution means but to tie the dispersion to the mean for each distribution. More precisely, if  $X \sim Ga(\alpha, \beta)$ , then:

- $2X \sim Ga(\lambda\alpha, \beta)$ ; convolution assumption (constant variance to mean ratio i.e. coefficient of dispersion; see (1)),
- $2X \sim Ga(\alpha, \lambda\beta)$ ; scaling assumption (constant standard deviation to mean ratio i.e. coefficient of variation; see (1)).

The probability density function of the gamma mixture model with tied parameters is given by

$$f(x; \pi_0, \dots, \pi_{M-1}, \alpha, \beta) = \sum_{n=0}^{M-1} \frac{\pi_n x^{\alpha_n-1}}{\Gamma(\alpha_n)\beta_n^{\alpha_n}} \exp\left(-\frac{x}{\beta_n}\right),$$

where the component weights  $\pi_0, \dots, \pi_{M-1}$  sum to one and  $\alpha_n = 2^n \alpha$  and  $\beta_n = \beta$  (convolution assumption) or  $\alpha_n = \alpha$  and  $\beta_n = 2^n \beta$  (scaling assumption). A mixture model is



thus a linear combination of elementary distributions termed components. The weight distribution is defined on the consecutive integers  $0, 1, \dots, M - 1$  corresponding, according to our hypothesis, to the possible numbers of replications (category  $n$  thus corresponds to cell ploidy  $2^{n+1}$ ) The mean number of endocycles per category is then 0, 1, 2 or 3 for 2C, 4C, 8C or 16C respectively.

For each category  $n$ , the weight distribution was calculated and it reflected the frequency of cells in the distribution. As a consequence it was used to estimate the endoreduplication factor (EF), based on the proposition of Cookson et al. (2006), i.e.

$$EF = (0 \times \pi_{2C}) + (1 \times \pi_{4C}) + (2 \times \pi_{8C}) + (3 \times \pi_{16C})$$

This value gives an estimation of the mean number of endocycles per 100 cells, so if all the cells were in 2C, this value would be 0, and if all the cells were 16C, the EF would be 300.

A gamma mixture model can be estimated by an application of the Expectation-Maximization (EM) algorithm, the tying constraints being directly incorporated in the M-step of the algorithm; see McLachlan and Peel 2000. For the selection of the number of mixture components and the comparison between mixture models with different tying constraints between parameters, we applied the Bayesian Information Criterion (BIC). For a gamma mixture model with tied parameters, this penalized likelihood criterion is given by

$$BIC(M) = 2 \sum_{i=1}^R \log f(x_i; \hat{\pi}_0, \dots, \hat{\pi}_{M-1}, \hat{\alpha}, \hat{\beta}) - (M + 1) \log R,$$

where  $M + 1$  is the number of free parameters ( $M - 1$  free weights and 2 gamma distribution parameters) and  $R$  the sample size.

### 3. RESULTS

#### 3.1. Identification of the scaling rule and selection of the number of components of the mixture model

We first estimated gamma mixture models without tying constraints between component parameters (i.e. a common scale parameter corresponding to the convolution assumption). BIC favors 3- and 4-independent-component mixture models. We thus investigated in details these two models. The 3-component model was close to the scaling assumption (shape

parameter common to the components) and the ratio of means (and standard deviation) for two consecutive components was close to 2; see Table IV-2. The mean area of unreplicated cells ( $890 \mu\text{m}^2$ ) was realistic assuming that it is underestimated because only uncensored cells fully included in the image were taken into account in the estimation (examples of cell areas distribution in Cookson et al. 2006; Fujikura et al. 2007; Elsner et al. 2012); see a discussion of this issue in the concluding remarks.

| Estimated ploidy according to the model categories |                  |      |       |      |       |      |        |          |  |
|--|------------------|------|-------|------|-------|------|--------|----------|--|
|  |                  | 2C   |       | 4C   |       | 8C   |        | BIC      |  |
| Independent parameters                             | $\pi_n$          | 0.49 |       | 0.29 |       | 0.22 |        |          |  |
|  | $\alpha$ $\beta$ | 2.32 | 383.8 | 1.86 | 897.7 | 2.23 | 1436.2 | -1016870 |  |
|  | $\mu$ $\sigma^2$ | 890  | 584   | 1670 | 1224  | 3208 | 2147   |          |  |
| Tied parameters                                    | $\pi_n$          | 0.49 |       | 0.27 |       | 0.24 |        |          |  |
|  | $\alpha$ $\beta$ | 2.42 | 341.8 | 2.42 | 683.5 | 2.42 | 1367.1 | -1016780 |  |
|  | $\mu$ $\sigma^2$ | 827  | 532   | 1653 | 1063  | 3306 | 2126   |          |  |

Table IV-2. 3-component mixture model under the scaling assumption: Estimated shape and scale parameters under different tying assumptions (independent parameters, tied shape and scale parameters). Component weights ( $\pi_n$ ) are indicated on the first row, shape ( $\alpha$ ) and scale ( $\beta$ ) parameters on the second row, mean ( $\mu$ ) and standard deviation ( $\sigma^2$ ) in  $\mu\text{m}^2$  on the third row for each tying parameter assumption.

| Estimated ploidy distribution |      |      |      |
|-------------------------------|------|------|------|
| Leaf rank                     | 2C   | 4C   | 8C   |
| 1                             | 0.32 | 0.2  | 0.48 |
| 2                             | 0.33 | 0.16 | 0.51 |
| 3                             | 0.26 | 0.32 | 0.42 |
| 4                             | 0.23 | 0.4  | 0.37 |
| 5                             | 0.32 | 0.39 | 0.29 |
| 6                             | 0.35 | 0.43 | 0.22 |
| 7                             | 0.43 | 0.33 | 0.24 |
| 8                             | 0.46 | 0.35 | 0.19 |
| 9                             | 0.53 | 0.3  | 0.17 |
| 10                            | 0.6  | 0.25 | 0.15 |
| 11                            | 0.68 | 0.17 | 0.15 |
| 12                            | 0.78 | 0.08 | 0.14 |
| 13-14                         | 0.9  | 0    | 0.1  |

Table IV-3. 3-component mixture model with tied scale and shape parameters under the scaling assumption: estimated ploidy distribution and mean number of endocycles for each leaf rank. The shape and scale parameters are common to the different models estimated for each leaf rank and were previously estimated on the pooled dataset.

We then re-estimated the component weights for each leaf rank sub-sample (Table IV-3) on the basis of the components previously estimated on the pooled sample (Table IV-2), parameters being tied according to the scaling assumption (shape parameter common to the components and ratio of scale parameters for two consecutive components equal to 2). Higher weights were estimated for 2C and for 8C than for 4C for the lowest leaf ranks suggesting that the first leaves had a low proportion of cells in 4C. When compared to the literature on the level of ploidy measured by flow cytometry in leaves of *A. thaliana* this was biologically unrealistic (Melaragno et al. 1993; Vlieghe et al. 2005; Ferjani et al. 2007). In addition, *A. thaliana* mature leaves have cells with ploidy levels reaching 16C and this was not the case here with the 3-component model which considered that implicitly the weight for 16C was nil, on the contrary of what was found in the literature (e.g Galbraith et al. 1991; Melaragno et al. 1993; Boudolf et al. 2004; Fujikura et al. 2007). For these reasons, we thus discarded this 3-component model.

|                        |                      | Estimated ploidy according to the model categories |     |      |     |      |      |      |      |          |  |
|------------------------|----------------------|--|-----|------|-----|------|------|------|------|----------|--|
|                        |                      | 2C   |     | 4C   |     | 8C   |      | 16C  |      | BIC      |  |
| Independent parameters | $\pi_n$              | 0.56   |     | 0.28 |     | 0.12 |      | 0.04 |      |          |  |
|                        | $\alpha \quad \beta$ | 2.8  | 270 | 6.69 | 274 | 9.3  | 385  | 9.06 | 694  | -1016830 |  |
|                        | $\mu \quad \sigma^2$ | 756  | 452 | 1833 | 709 | 3582 | 1175 | 6287 | 2089 |          |  |
| Tied parameters        | $\pi_n$              | 0.61   |     | 0.22 |     | 0.14 |      | 0.03 |      |          |  |
|                        | $\alpha \quad \beta$ | 2.2  | 402 | 4.39 | 402 | 8.79 | 402  | 17.6 | 402  | -1016970 |  |
|                        | $\mu \quad \sigma^2$ | 883  | 596 | 1767 | 843 | 3533 | 1192 | 7066 | 1686 |          |  |

Table IV-4. 4-component mixture model under the convolution assumption: estimated shape and scale parameters under different tying assumption (independent parameters and tied shape and scale parameters). Component weights ( $\pi_n$ ) are indicated on the first row, shape ( $\alpha$ ) and scale ( $\beta$ ) parameters on the second row, mean ( $\mu$ ) and standard deviation ( $\sigma^2$ ) in  $\mu\text{m}^2$  on the third row for each tying parameter assumption.

For parameter interpretation, it is important to consider the weight of components. In particular, the scale and shape parameters estimated for 8C and moreover for 16C are far less reliable than the scale and shape parameters estimated for 2C and 4C because of the lower number of values corresponding to these categories, of their larger dispersion and of the absence of visible hump.

The 4-component model was close to the convolution assumption (scale parameter common to the components) and the ratio of means for two consecutive components was

close to 2 (and the ratio of standard deviation close to  $\sqrt{2}$ ); see Table IV-4. The mean area of unreplicated cells ( $756 \mu\text{m}^2$ ) was realistic assuming that it is underestimated because only uncensored cells were taken into account in the estimation and that the proportion of censored cells was rather high (roughly 2 uncensored cells for a censored cell).

| Leaf rank | Estimated ploidy distribution |      |      |      | Estimated |
|-----------|-------------------------------|------|------|------|-----------|
|           | 2C                            | 4C   | 8C   | 16C  | EF        |
| 1         | 0.43                          | 0.27 | 0.22 | 0.08 | 96        |
| 2         | 0.44                          | 0.23 | 0.26 | 0.07 | 97        |
| 3         | 0.39                          | 0.33 | 0.22 | 0.06 | 95        |
| 4         | 0.39                          | 0.36 | 0.20 | 0.05 | 92        |
| 5         | 0.47                          | 0.32 | 0.17 | 0.04 | 79        |
| 6         | 0.51                          | 0.32 | 0.14 | 0.03 | 69        |
| 7         | 0.57                          | 0.26 | 0.14 | 0.03 | 65        |
| 8         | 0.60                          | 0.26 | 0.11 | 0.03 | 57        |
| 9         | 0.66                          | 0.21 | 0.11 | 0.02 | 48        |
| 10        | 0.72                          | 0.16 | 0.10 | 0.02 | 41        |
| 11        | 0.79                          | 0.10 | 0.09 | 0.02 | 34        |
| 12        | 0.88                          | 0.02 | 0.08 | 0.02 | 24        |
| 13-14     | 0.93                          | 0    | 0.06 | 0.01 | 15        |

Table IV-5. 4-component mixture model with tied scale and shape parameters under the convolution assumption: estimated ploidy distribution (frequencies) and estimated endoreduplication factor (EF) for each leaf rank. The shape and scale parameters are common to the different models estimated for each leaf rank and were previously estimated on the pooled dataset.

We then re-estimated the component weights for each leaf rank sub-sample (Table IV-5) on the basis of the components previously estimated on the pooled sample (Table IV-4), parameters being tied according to the convolution assumption (scale parameter common to the components and ratio of shape parameters for two consecutive components equal to 2). The weight distributions (with low weight estimated for 16C whatever the leaf rank considering that implicitly the weight for 32C was nil) are far more biologically realistic than the weight distributions estimated for the 3-component mixture model. Ploidy levels of 32C can sometimes be detected in studies using flow cytometric analyses (Galbraith et al. 1991; Cookson et al. 2006; Ferjani et al. 2007), but they could correspond to trichomes nuclei, since these particular differentiated cells are known to reach such ploidy level (Hülkamp et al. 1994). As trichomes were not included in our measurements, the absence of fifth component

corresponding to the 32C ploidy level is not surprising. We thus focus on this 4-component model relying on the convolution assumption. Figure IV-6 represents the frequency distribution of cell areas with the estimated mixture law and the underlying gamma components.

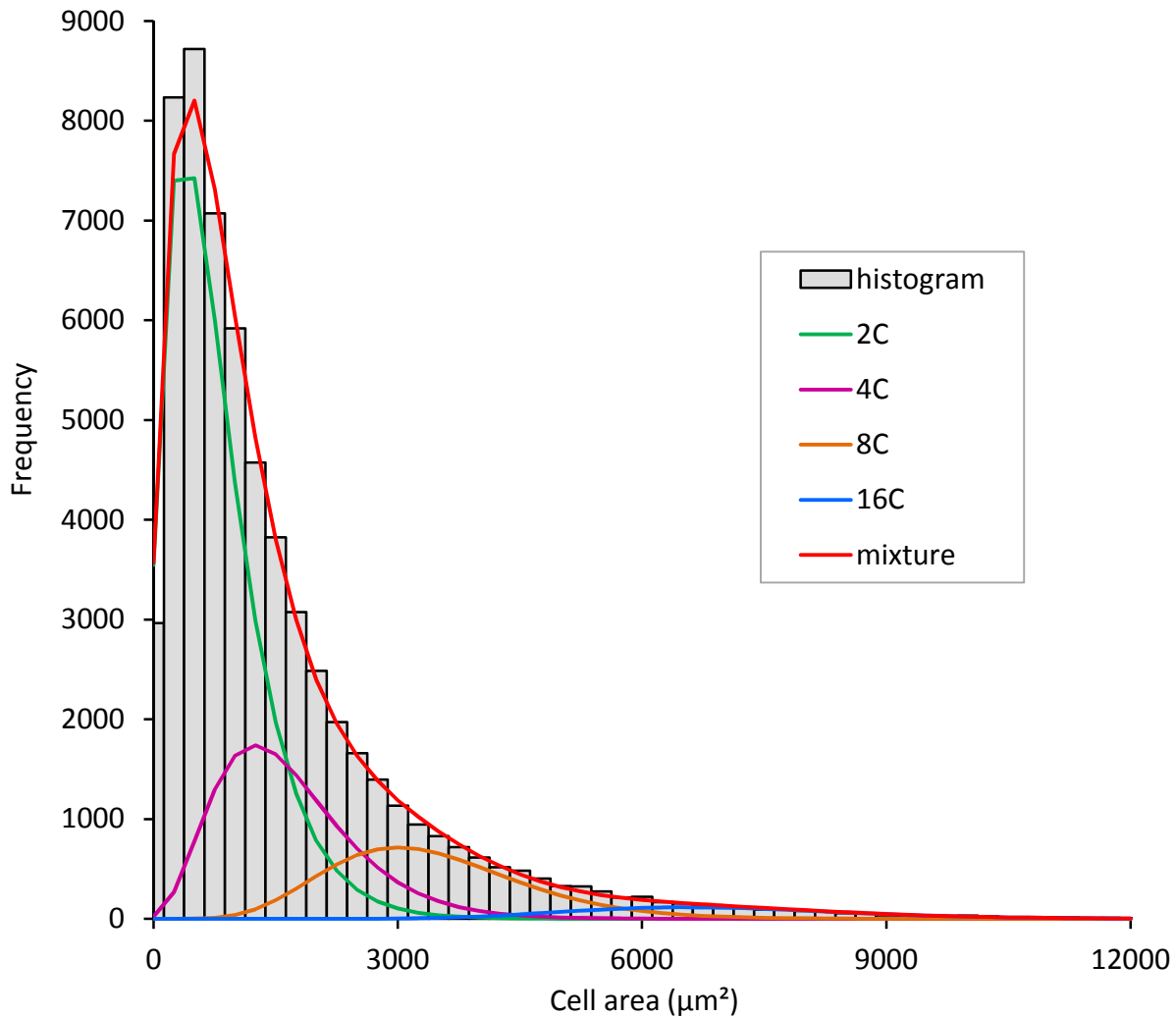


Figure IV-6. Frequency distribution of cell areas of adaxial leaf epidermal cells, excluding trichomes and guard cells ( $n = 60,286$ ) and representation of the 4-component mixture model with tied scale and shape parameters under the convolution assumption. To facilitate readability, values superior to 12,000  $\mu\text{m}^2$  do not appear on the graph (27 cells).

### 3.2. Estimation of the endoreduplication factor in the epidermis using the 4-component mixture model

The distribution estimated using the 4-component mixture model showed that for most leaves along the shoot axis, the proportion of cells in the first category was higher than the proportion of cells in others with a tendency to decrease from a category to the next one

(Table IV-5). According to our hypothesis that each category corresponds to a level of ploidy, this would reflect that in each leaf there was a higher proportion of cells with low level of ploidy which is consistent with results reported in the literature using flow cytometry measurements to estimate ploidy level (Vlieghe et al. 2005; Ng et al. 2012). Comparing the distribution along the shoot axis, this model indicates that the proportion of cells in the first category had a tendency to increase with leaf rank whereas for the second category it tended to increase with the first ranks and then decreased after the 6<sup>th</sup> leaf. For the third and fourth categories, the proportion of cells had a tendency to decrease with leaf rank. As a consequence of these shoot structuring, the last leaves formed on the rosette had almost all cells in a unique category.

From these data, the average number of endocycles per 100 leaf pavement cells could be estimated and it decreased along the shoot (Table IV-5). This is in accordance with preliminary results obtained in our team from flow cytometric analysis performed on mature rosette leaves of Col-0 plants that showed the same tendency in plants grown in short day conditions and thus exhibiting higher number of rosette leaves (LEPSE, data not shown). According to our results, this decrease in the estimated average number of endocycles along the shoot reflects a lower proportion of large pavement cells in the leaves at the upper nodes of the rosette.

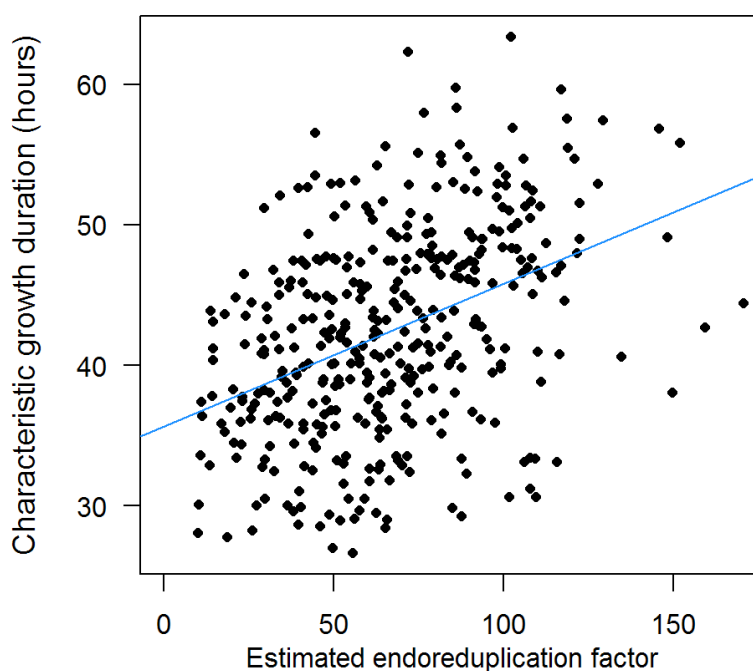


Figure IV-7. Linear relationship between the estimated endoreduplication factor and the characteristic growth duration (hours). Each point represents an individual leaf ( $n = 416$ ). The blue line represents the linear relationship estimated between the variables.

Pearson correlation coefficient  $r = 0.40$ ;  $p$ -value = 0.

By investigating the relationships between the estimated mean number of endocycles and the estimated characteristic growth duration for each leaf (see Chapter III section 2.2.2 for computation method), a positive correlation was identified between the estimated mean ploidy level and the leaf expansion duration (Figure IV-7;  $r = 0.40$ ,  $p\text{-value} = 0$ ). This could reflect that cells with low level of ploidy are those who have less time to undergo through successive numbers of endocycles in leaves rapidly reaching their final size.

## **4. DISCUSSION**

### **4.1. Semi-automatic method for measuring cell areas in *A. thaliana* epidermis imprints: an improvement for analyzing cell size diversity**

The prevalence of the cell size diversity has been reported throughout the epidermis of most *A. thaliana* organs such as hypocotyls, sepals and leaves but the function of having a wide range of pavement cell sizes is still unclear. This high diversity in cell size is a problem to compare genotypes or to analyze the effect of environmental conditions or even to compare cell size between different leaves structuring the whole shoot on robust bases. Getting robust descriptors of cell dimensions in a tissue requires the measurements of a high number of cells which is incompatible with current solutions for automated cell measurements that relies either on semi-automated or manual analysis of the imaging data, or on expensive and proprietary software. In this chapter, we proposed and tested a method for the semi- automatic segmentation and the analysis of large series of leaf epidermal imprints images from phenotyping experiment. The method relies on recent advances on machine learning applied to cell biology and involves minimal user interaction. It enables the acquisition of large and reliable cell area datasets necessary to establish the statistical analyses that may follow.

### **4.2. Analyzing cell size diversity in *A. thaliana* epidermis using a 4-component gamma mixture model**

Cell size is controlled by both ploidy-dependent and ploidy-independent pathways (Cookson et al. 2006; Breuer et al. 2010). This could explain the fact that cells with different ploidy levels can reach the same size, and thus the overlap between two successive components in the mixture model, resulting in a cell area distribution that does not exhibit well-separated humps. However, as previously mentioned, several studies highlighted a positive correlation between cell size and ploidy level in *A. thaliana* leaf epidermis. One difficulty to investigate this covariation is that flow cytometric

analyses are destructive and have to be performed on fresh material, which does not enable to make cell area measurements on the same leaves. Ploidy levels and cell areas are thus measured on different cells (e.g. Vlieghe et al. 2005; Cookson et al. 2006; Dewitte et al. 2007; Roeder et al. 2010). A method enables to assess ploidy and to measure areas on the same cells. It consists in staining nuclei in leaf epidermal peels using DAPI (4',6-diamidino-2-phenylindole) which allows the visualization of both cell walls and nuclei using epi-fluorescence microscopy. Cell ploidy is assessed based on the size of the nuclei which reflects the quantity of genetic material. This method enables to directly analyze the correlation between a cell ploidy level and its size. It has been applied in a very few studies on *A. thaliana* leaves (Melaragno et al. 1993; Boudolf et al. 2004; Elsner et al. 2012), but they all confirmed the positive correlation between these two variable. The study published by Melaragno et al. 1993 is one of the rare study presenting average cell area for each ploidy level detected in Col-0 mature rosette leaves.

One have to keep in mind that we presented the data are area-biased because of the limited size of the image with respect to the area of the largest cells (images covering 356,000  $\mu\text{m}^2$  of the leaf with 81 uncensored cells on average whereas some cells can reach areas up to 15,000  $\mu\text{m}^2$ ). The larger the cell is, the higher is the probability of intercepting an image border and thus of being truncated or censored. Only cells fully included in the image are currently used in the gamma mixture model estimation and this indeed affects all the identification and model selection procedure. The next step in this study will thus consist in extending the estimation algorithm along the lines proposed by Chauveau 1995 in order to use both uncensored and censored cell area for the estimation of the gamma mixture model. Because of this consideration and of the absence of well-separated humps corresponding to the different ploidy levels in the empirical distribution the estimated parameter values should be interpreted with caution. However, the fact that the parameters estimated for the components are realistic indicates that the mean numbers of endocycles estimated for each leaf rank are reliable. This is why the decrease in this estimated mean number of endocycles along the shoot can be interpreted, in this context, as a reduction of the amount of large cells, leading to a decrease in the mean cell area. Such decrease of average leaf pavement cell area along *A. thaliana* shoot has been reported in some studies (Cookson et al. 2007; Dewitte et al. 2007; Usami et al. 2009). Since individual final leaf area increases along the shoot, it indicates that, on the contrary, the epidermal cell number increases in the leaves at the upper nodes of the rosette. It supports the thesis that the final leaf area is most driven by cell number than cell area.



In conclusion, the proposed gamma mixture model enabled a more thorough analysis of the cell area measurements than simple location (mean or median) and dispersion (standard or mean absolute deviation) descriptors, by estimating both average cell areas for cells with a given ploidy level, and variation of leaf ploidy level along the shoot. Coupled with the semi-automatic segmentation method that we developed for images of leaf epidermis imprints, it enables to take into account the variability of *A. thaliana* leaf pavement cell area linked to endoreduplication even in high-throughput phenotypic experiments. Using this pipeline of methods (image analysis combined with mixture models) on different genotypes and/or in different environmental conditions could add value to phenotypic studies that would investigate leaf growth traits from the tissular scale to the shoot scale.

## CHAPTER V. CONCLUSIONS AND PERSPECTIVES

### 1. ADDED VALUE OF OUR PIPELINE FOR THE ANALYSIS OF *A. THALIANA* SHOOT DEVELOPMENT

Characterization of processes controlling plant shoot growth and how it is affected by genetic and environment factors posed a difficult challenge due to the large number of genes in the genomes, and the diversity of environmental conditions that influence shoot growth phenotypes. Another difficulty is that plant shoot growth phenotype has to be analyzed dynamically and with a multi-scale approach with sufficient throughput and resolution (see Chapters I & II, Lièvre et al. 2013). My PhD project has led to the development a new pipeline of analysis, combining image analysis techniques and statistical models to integrate at the shoot scale leaf growth and developmental traits that are commonly analyzed separately. The Figure V-1 schematically represents the obtained pipeline in its integrality. This pipeline enables the structuring of *A. thaliana* rosette into successive developmental phases based on the changes in several leaf growth variables along the shoot. The organ scale was investigated with morphological (covering of the leaf abaxial surface with trichomes, leaf blade length-to-width ratio), dimensional (final leaf area) and dynamic (characteristic growth duration) traits (Figure V-1 Panels A and B). Their computation was made using both manual and semi-automatic methods – i.e. ImageJ macro and R scripts developed in our lab. Regarding the analysis of individual leaf expansion, we determined that the characteristic growth duration could bring more additional information than the commonly used maximum leaf expansion rate since this latter variable is strongly correlated to the final leaf area (Chapter II). Testing this pipeline on 3 T-DNA insertion lines and their wild-type (Columbia 0, referred to as Col-0), we demonstrated its usefulness as it enabled to define successive developmental phases integrating the gradual changes along the shoot of different morphological, dimensional and dynamics leaf growth traits. By analyzing this structuring and the changes in the values of each trait within each developmental phase (Figure V-1 Panel D), we were able to gain new insights on the effect of the mutations on plant shoot development in comparison with more classical analyses that were used in Chapter II.

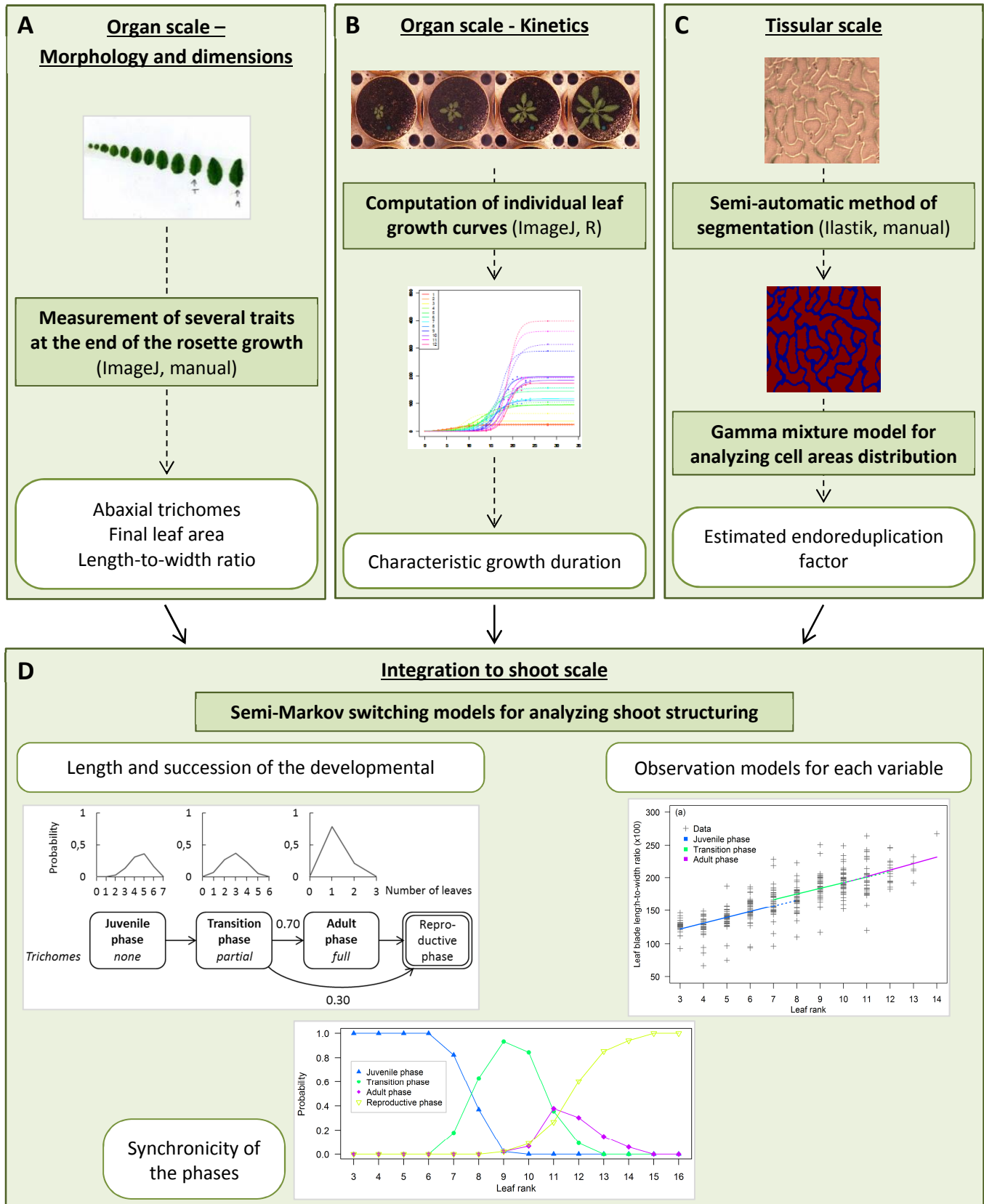


Figure V-1. Schematic representation of the pipeline of analysis methods developed during this PhD, from cell to shoot. The light green panels correspond to: A, the morphological and dimensional measurements at the organ (leaf) scale; B, the kinetics at the organ (leaf) scale; C, the tissular (epidermis) scale; and D, the integration of the different variables at the shoot scale. Image analysis techniques and statistical analyses used for each scale are indicated in the dark green inserts. The outputs obtained for each scale are indicated in the white inserts.

Commonly, leaf growth phenotyping studies mainly focus on one or a few leaf ranks. An originality of this work lies in the fact that developmental and growth traits used to build Col-0, SALK\_055458, SALK\_048174 and SALK\_126071 datasets were measured at different scales for each individual leaf composing the rosette. The traits we investigated have been reported to vary along *A. thaliana* shoot in studies presenting results for different leaf ranks (Telfer et al. 1997; Cookson et al. 2007; Willmann and Poethig 2011). The abaxial trichome variable which is the only one on which different categories of leaves can be defined – i.e. absence, partial covering and full covering of the abaxial leaf surface with trichomes. Changes in the other traits are rather gradual than abrupt and can be very subtle. Therefore, even if they are considered as heteroblastic traits, the delimitation of clear phases in the shoot development based on these criteria can be tricky. Integrating these different traits at the shoot scale to build multivariate segmentation models, we demonstrated that developmental phases can be identified along *A. thaliana* rosette shoot, and that these phases are defined by more than one variable (Figure V-1 Panel D). The most structuring variable was the abaxial trichome variable for the 4 genotypes we investigated, as expected, but, as shown in the Chapter III, some other traits also played a role in the discrimination of the successive phases. Among the quantitative variables, the most structuring ones are the final leaf area and the characteristic growth duration, whereas the less structuring one is the leaf blade length-to-width ratio. This result was not expected since changes in leaf shape have often been reported as markers of heteroblasty (Ashby 1948; Kerstetter and Poethig 1998; Steynen et al. 2001; Willmann and Poethig 2011).

The multi-scale pipeline of analysis methods allowed to determine which leaf growth traits are the most important for the shoot structuring into successive developmental phases, and possibly which ones are the most discriminant between different genotypes and/or environmental conditions. Thus, it might eventually be used to select variables of interest on which to focus during leaf growth phenotyping studies, and consequently optimizes the workload for high-throughput phenotyping.

## **2. ADDED VALUE OF OUR PIPELINE FOR THE ANALYSIS OF CELL AREA DIVERSITY IN *A. THALIANA* LEAF EPIDERMIS**

Our in-depth analysis of cell area distribution in leaf epidermis revealed the complexity of tissular development that results from different processes with complex interactions: cell division, cell expansion and endoreduplication. The semi-automatic method based on a

machine learning procedure that we developed to analyze cell area in *A. thaliana* leaf epidermis enabled fast and reliable segmenting images of leaf epidermal imprints in comparison to manual methods (Figure V-1 Panel C). It allowed increasing the number of cells that could be measured together on the same leaf and then allowed to work on the whole distribution of cell area diversity in the tissue with enough replicates to make robust statistical models. This image segmentation method has a particular interest as it facilitates the integration of cell area measurements in high-throughput phenotypic analyses. The development of a gamma mixture model for analyzing the resulting data is also a great improvement since it gives access to more complete information regarding cell area distribution in the leaf epidermis than classic statistical descriptors of shape and dispersion. Our hypothesis that the pavement cell area distribution observed in *A. thaliana* leaf epidermis is a mixture of distributions corresponding to the different ploidy levels that can be reached in this tissue is supported by the literature as different model parameters followed biological laws (Melaragno et al. 1993; Boudolf et al. 2004). However, this hypothesis could not be fully validated here. Applying these methods to different mutants affected in the endoreduplication process could be useful to validate this hypothesis and to test to what extent cell area distributions can be used (1) to estimate the mean cell area for cell categories corresponding to each ploidy level and (2) to estimate the mean number of endoreduplication cycles per cell (see Chapter IV). This validation is ongoing in LEPSE in the frame of an ANR project (CKI-Stress, 2013-2017) on the role of endoreduplication in leaf growth with the phenotypic analyses of *A. thaliana* mutants affected in cell cycle inhibitors and which present an increase in the endoreduplication factor in their leaves.

### **3. INTEGRATING THE ESTIMATED ENDOREDUPLICATION FACTOR IN OUR PIPELINE TO ANALYSE SHOOT STRUCTURING**

Imaging and statistical methods developed in Chapter IV could be used in the future to complete this pipeline with cellular growth related traits. As an example, we were able to integrate the estimated endoreduplication factor (EF) computed for each leaf of Col-0 plants as a new variable into the Col-0 multivariate sequences built using Experiment 3 data (see Chapter III). In this new setting, the leaf description at each successive rank combines a morphological variable (ordinal categorical variable with ordered categories no trichome, partial covering by trichomes, full covering by trichomes), a shape descriptor (length-to-width blade ratio), a dimensional variable (final leaf area), a dynamical variable (characteristic

growth duration extracted from the estimation of a logistic growth model on the basis of leaf expansion follow-up data) and the newly incorporated tissular variable (estimated EF). We used semi-Markov switching models (SMSMs, defined in Chapter III) to segment multivariate sequences into successive developmental phases. Concerning the observation models of each variable in each developmental phase, we made the same assumption as in Chapter III: categorical observation distributions for the trichome variable and linear regression models for the four quantitative variables: leaf blade length-to-width ratio, final leaf area, characteristic growth duration and estimated EF.

Because of the amount of missing tissular data concerning the first two epicotylar leaves, the first developmental phase modeled without the estimated EF – i.e. the seedling phase – could not be modeled when this variable was integrated. The lack of tissular data for these two first leaves was due to their frequent senescence when the plants were harvested, making the corresponding epidermal imprints unusable. We therefore built a four-state model (three transient states corresponding to three successive vegetative phases and an absorbing ‘end’ state corresponding to the reproductive phase). For these three developmental phases – i.e. juvenile, transition and adult phases – the estimated parameters of the model were very similar to the ones estimated without the estimated EF (Figure V-2). The succession was also not completely deterministic since the adult phase could be skipped with probability 0.3, which is close to the probability found without the tissular data (0.27).

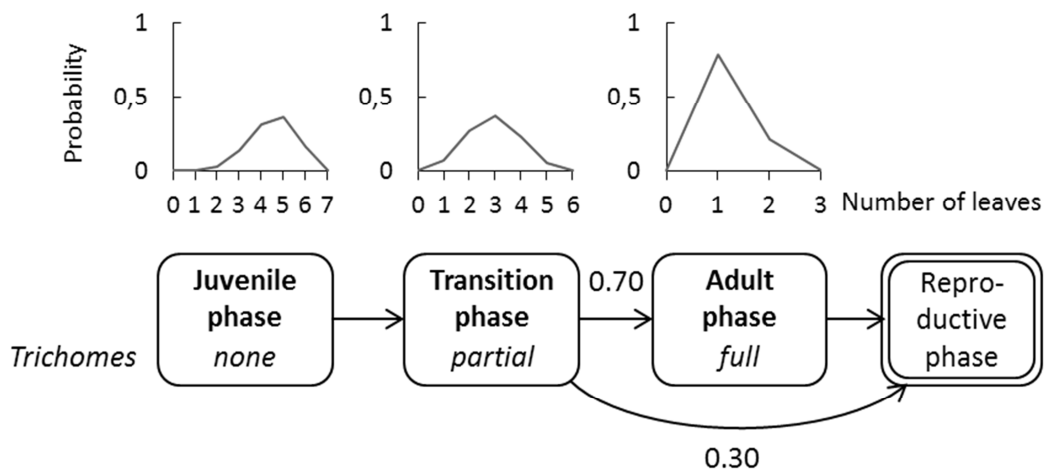


Figure V-2. Graph of transition between phases and phase length distributions estimated using the Col-0 multivariate sequences integrating the estimated endoreduplication factor as an additional variable.

|                      | Mean length of phase in number of leaves $\pm$ standard deviation |                |                  |               |
|----------------------|---|----------------|------------------|---------------|
|                      | Seedling phase  | Juvenile phase | Transition phase | Adult phase   |
| Without estimated EF | $2.0 \pm 0.2$   | $5.0 \pm 1.0$  | $2.9 \pm 1.0$    | $1.3 \pm 0.5$ |
| With estimated EF    | -   | $4.5 \pm 1.0$  | $2.9 \pm 1.0$    | $1.2 \pm 0.2$ |

Table V-1. Mean length of each developmental phase and associated standard deviation for the SMSMs with and without the estimated endoreduplication factor (EF) integrated in the multivariate sequences. Phase length is expressed in number of successive leaves.

In both cases, the estimated observation distributions for the categorical trichome variable for each developmental phase showed that the discrimination between these phases fits exactly the categories of this variable.

The linear regression models estimated for the quantitative variables common to the two models were almost rigorously alike, except for the one estimated for the final leaf area in the adult phase; see Table V-2 and Appendix V. For the estimated EF, very slight changes of slopes were observed between the successive phases. However, these changes of slopes were still more marked than the ones observed for the length-to-width ratio. Therefore, this variable does not play a strong role in Col-0 shoot structuring, but is still more structuring than the shape descriptor we used.

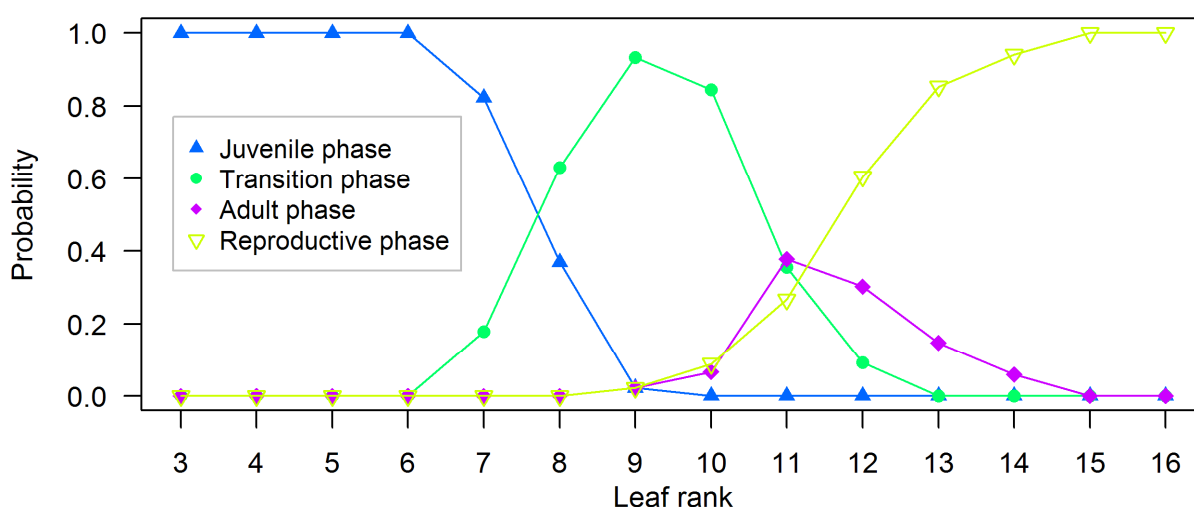


Figure V-3. Probability of the developmental phases as a function of the leaf rank for Col-0 when the estimated endoreduplication factors are integrated to the multivariate sequences. Blue: juvenile phase. Green: transition phase. Purple: adult phase. Lemon green: reproductive phase.

| <b>(a) Length-to-width ratio</b>          | Seedling phase |      | Juvenile phase | Transition phase |      | Adult phase |      |
|---|----------------|------|----------------|------------------|------|-------------|------|
| Without estimated EF                      | -2.4           | n.s. | 8.5            | 8.6              |      | 9.3         |      |
| With estimated EF                         | -              |      | 8.9            | 8.5              |      | 9.8         |      |
| <b>(b) Final leaf area</b>                | Seedling phase |      | Juvenile phase | Transition phase |      | Adult phase |      |
| Without estimated EF                      | -0.6           | n.s. | 28.7           | 39.2             |      | 16.8        |      |
| With estimated EF                         | -              |      | 29.3           | 39.1             |      | 25.0        |      |
| <b>(c) Characteristic growth duration</b> | Seedling phase |      | Juvenile phase | Transition phase |      | Adult phase |      |
| Without estimated EF                      | -0.1           | n.s. | -1.2           | -0.7             | n.s. | 0.5         | n.s. |
| With estimated EF                         | -              |      | -1.3           | -0.6             | n.s. | 0.6         | n.s. |
| <b>(d) Estimated EF</b>                   | Seedling phase |      | Juvenile phase | Transition phase |      | Adult phase |      |
|   | -              |      | -8.6           | -6.2             |      | -8.5        |      |

Table V-2. Slopes of the trends modeled for each quantitative variable within the successive developmental phases. When the slope was not significantly different from 0, ‘n.s.’ was indicated near to the value. EF: endoreduplication factor.

The computation of probability of the developmental phases as a function of the leaf rank showed that the segmentation of Col-0 shoots into successive developmental phases remains markedly synchronous when the estimated EF is integrated in the model (Figure V-3).

These results indicate that incorporating the estimated endoreduplication factor in the multivariate sequences used to build Col-0 segmentation model does not markedly change the structuring of Col-0 shoots in successive developmental phases and the most structuring variables – abaxial trichomes, final leaf area and characteristic growth duration – remain the same than the ones identified as such with the model built without the estimated endoreduplication factors.

This new model illustrates the fact that the analysis approach based on SMSMs presented in Chapter III can be directly applied with supplementary variables incorporated in the multivariate sequences. The constraint comes for the potentially higher occurrence of missing values due to various experimental reasons when incorporating many phenotyping variables.

Adding the estimated endoreduplication factor to our analysis method shows how incorporating new traits can allow the integration of supplementary scales. Considering the tissular variables in shoot development analyses can be of great interest as the control of cell



division and expansion is essential in the establishment of leaf shape and size. Furthermore, Usami et al. 2009 highlighted a regulatory pathway of cell number and area in *A. thaliana* leaves dependent of *miRNA156* and its target *SPL* (*SQUAMOSA PROMOTER BINDING PROTEIN-LIKE*) genes, which are known to affect heteroblasty by their involvement in vegetative phase change. This link between heteroblasty and the regulation of cell number and area further supports the relevance of integrating the tissular scale to analyses investigating the developmental phases that structure shoot development.

#### **4. WHAT NEXT? INCORPORATING NEW VARIABLES IN THE MULTIVARIATE PIPELINE AND USING TARGETED GENOTYPES FOR SPECIFIC QUESTIONS**

It would be interesting to incorporate the estimated endoreduplication factor in other datasets in order to see if the weak structuring effect found in Col-0 is conserved or if this variable can play a stronger role in shoot structuring in other genotypes and/or environmental conditions. Among the genotypes we analyzed, SALK\_126071 (i.e. *ctf18-2*; Takahashi et al. 2010) is known to be affected in the cell cycle and the shoot structuring is markedly disturbed in this mutant. Applying the pipeline of the tissular scale analysis methods to this genotype and building a SMSM using the estimated endoreduplication factor as a supplementary variable might therefore give new insights concerning the link between cell cycle regulation and heteroblasty. One could also consider adding variables such as internode growth variables, presence or not of axillary buds and/or branching data. This could be particularly interesting for applying this pipeline of analysis methods to plant species exhibiting other types of shoot axis. For example, in numerous species, internodes are more elongated than in *A. thaliana* rosette and some studies showed that their length is not homogeneous along the shoot; e.g. in Fournier and Andrieu 2000 (maize); Costa et al. 2012 (*Antirrhinum*); Dambreville et al. in press (mango).

During our study, we used genotypes with mutations that had not been previously described as having an effect on shoot development. Analyzing genotypes affected in specific processes using the methods we developed could help obtaining a more global vision of shoot development. For example, one could consider applying this pipeline to plants with mutations altering the vegetative phase change. This could enable to better understand and/or take into account the links between this process and the heteroblasty. These concepts are of great interest for scientific and economic reasons since they are critically important to understand the regulation of plant development as well as morphological evolution and life-history

variation. The functional significance of heteroblasty – in the sense of ontogenetic changes among the shoot metamers – is extensively discussed in a review of Zotz et al. 2011. The authors advocates that even if demonstrating unambiguous adaptive value to heteroblasty under current ecological conditions may be intricate, this phenomena is likely to be inherited from ancient progenitor in which it provided evolutionary advantage. The regulation of the onset of the vegetative phase change has also been proposed to have ecological functions, since variation in the length of the juvenile vegetative phase can avoid low seed yields to the plants, as it would occur if flowering were to occur while the plants were still small and thus with limited photosynthetic capacity. In a practical perspective, a better understanding of the environmental and genetic factors involved in the regulation of this phase change is critical to improve breeding programs – that are limited by the duration of the juvenile phase – and for crop scheduling, as discussed in Matsoukas 2014.

For this purpose, we performed a fourth experiment in the same growth conditions than those applied during Experiment 3 (see Chapter III), with simple and double mutated lines studied by Patrick Laufs group (Jean-Pierre Bourgin Institut – INRA, Versailles, France) to investigate the role of several miRNAs in the regulation of processes such as cell proliferation in developing organs, timing of organs formation or phyllotaxy. The list of the lines is reported in Appendix VI (Table SVI-1). We used the first steps of our pipeline to constitute multivariate datasets for these lines. Preliminary analyses showed that these lines exhibit marked alterations in their leaf developmental phenotype (see Appendix VI, Figure SVI-1). Analyzing these datasets will therefore bring new insights regarding the links between the different altered processes and the general vegetative shoot development. For instance, three lines were affected in the *miR156/SPL* genes interaction (*MIM156*, *rSPL10* and *rSPL9*), which would enable to directly investigate the crosstalk between vegetative phase change and global rosette heteroblasty. A challenge will be to apply our model to datasets with a limited number of replicates (between 6 and 11 for these lines), which would enable to analyze high-throughput experiments – in the sense of performed on a large number of genotypes/environmental conditions but consequently with a limited number of individuals per genotype/environmental conditions.

Overall, the pipeline of analyzing methods we developed can be seen as a new approach to investigate heteroblasty – in its broadest sense – in *A. thaliana* vegetative shoot, from the tissular scale to the shoot scale. Understanding the function and regulation of heteroblasty in higher plants is of great interest for the reasons above-mentioned, however its study can be

difficult in practical terms. Indeed, truly marked heteroblastic changes are often exhibited by woody species – such eucalyptus (Boland et al. 2006) – that have long life-cycles and mainly uncharacterized genomes (Zotz et al. 2011). Being able to quantify this process in the model plant *A. thaliana* could therefore be very useful to progress in this research field.

## REFERENCES

- Adams SR, Munir M, Valdés VM, Langton FA, Jackson SD. 2003.** Using Flowering Times and Leaf Numbers to Model the Phases of Photoperiod Sensitivity in *Antirrhinum majus* L. *Annals of Botany* **92**: 689–696.
- AGI. 2000.** Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.
- Aguirrezabal L, Bouchier-Combaud S, Radziejwoski A, Dauzat M, Cookson SJ, Granier C. 2006.** Plasticity to soil water deficit in *Arabidopsis thaliana*: dissection of leaf development into underlying growth dynamic and cellular variables reveals invisible phenotypes. *Plant, Cell & Environment* **29**: 2216–2227.
- Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, Gadrinab C, Heller C, Jeske A, Koesema E, Meyers CC, Parker H, Prednis L, Ansari Y, Choy N, Deen H, Geralt M, Hazari N, Hom E, Karnes M, Mulholland C, Ndubaku R, Schmidt I, Guzman P, Aguilar-Henonin L, Schmid M, Weigel D, Carter DE, Marchand T, Risseuw E, Brogden D, Zeko A, Crosby WL, Berry CC, Ecker JR. 2003.** Genome-Wide Insertional Mutagenesis of *Arabidopsis thaliana*. *Science* **301**: 653–657.
- Amasino R. 2010.** Seasonal and developmental timing of flowering. *The Plant Journal* **61**: 1001–1013.
- Andrés F, Coupland G. 2012.** The genetic basis of flowering responses to seasonal cues. *Nature Reviews Genetics* **13**: 627–639.
- Andrieu B, Hillier J, Birch C. 2006.** Onset of Sheath Extension and Duration of Lamina Extension are Major Determinants of the Response of Maize Lamina Length to Plant Density. *Annals of Botany* **98**: 1005–1016.
- Arvidsson S, Pérez-Rodríguez P, Mueller-Roeber B. 2011.** A growth phenotyping pipeline for *Arabidopsis thaliana* integrating image analysis and rosette area modeling for robust quantification of genotype effects. *New Phytologist* **191**: 895–907.
- Ashby E. 1948.** Studies in the Morphogenesis of Leaves. *New Phytologist* **47**: 153–176.
- Atwell S, Huang YS, Vilhjalmsen BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Mulyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JDG, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M. 2010.** Genome-wide association study of 107 phenotypes in a common set of *Arabidopsis thaliana* inbred lines. *Nature* **465**: 627–631.
- Beck TW. 2013.** The Importance of A Priori Sample Size Estimation in Strength and Conditioning Research: *Journal of Strength and Conditioning Research* **27**: 2323–2337.
- Beemster GTS, Veylder LD, Vercruysse S, West G, Rombaut D, Hummelen PV, Galichet A, Gruissem W, Inzé D, Vuylsteke M. 2005.** Genome-Wide Analysis of Gene

## REFERENCES

---

Expression Profiles Associated with Cell Cycle Transitions in Growing Organs of Arabidopsis. *Plant Physiology* **138**: 734–743.

**Berná G, Robles P, Micol JL. 1999.** A Mutational Analysis of Leaf Morphogenesis in Arabidopsis thaliana. *Genetics* **152**: 729–742.

**Berna G, Robles P, Micol JL. 1999.** A mutational analysis of leaf morphogenesis in Arabidopsis thaliana. *Genetics* **152**: 729–742.

**Bernstein N, Silk WK, Läuchli A. 1993.** Growth and development of sorghum leaves under conditions of NaCl stress. *Planta* **191**: 433–439.

**Boland DJ, Brooker MIH, Chippendale GM, Hall N, Hyland BPM, Johnston RD, Kleinig DA, McDonald MW, Turner JD. 2006.** *Forest Trees of Australia*.

**Bollman KM, Aukerman MJ, Park M-Y, Hunter C, Berardini TZ, Poethig RS. 2003.** HASTY, the Arabidopsis ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. *Development* **130**: 1493–1504.

**Bongard-Pierce DK, Evans MMS, Poethig RS. 1996.** Heteroblastic features of leaf anatomy in maize and their genetic regulation. *International Journal of Plant Sciences* **157**: 331–340.

**Borreani G, Roggero PP, Sulas L, Valente ME. 2003.** Quantifying Morphological Stage to Predict the Nutritive Value in Sulla ( L.). *Agronomy Journal* **95**: 1608.

**Boucheron E, Healy JHS, Bajon C, Sauvanet A, Rembur J, Noin M, Sekine M, Khamlichi CR, Murray JAH, Onckelen HV, Chriqui D. 2005.** Ectopic expression of Arabidopsis CYCD2 and CYCD3 in tobacco has distinct effects on the structural organization of the shoot apical meristem. *Journal of Experimental Botany* **56**: 123–134.

**Boudolf V, Vlieghe K, Beemster GTS, Magyar Z, Acosta JAT, Maes S, Schueren EVD, Inzé D, Veylder LD. 2004.** The Plant-Specific Cyclin-Dependent Kinase CDKB1;1 and Transcription Factor E2Fa-DPa Control the Balance of Mitotically Dividing and Endoreduplicating Cells in Arabidopsis. *The Plant Cell Online* **16**: 2683–2692.

**Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Görlach J. 2001.** Growth Stage–Based Phenotypic Analysis of Arabidopsis A Model for High Throughput Functional Genomics in Plants. *The Plant Cell Online* **13**: 1499–1510.

**Breuer C, Ishida T, Sugimoto K. 2010.** Developmental control of endocycles and cell growth in plants. *Current Opinion in Plant Biology* **13**: 654–660.

**Bruni NC, Dengler NG, Young JP. 1996.** Leaf developmental plasticity of Ranunculus flabellaris in response to terrestrial and submerged environments. *Canadian Journal of Botany* **74**: 823–837.

**Chatelain P, Pauly O, Peter L, Ahmadi S-A, Plate A, Bötzel K, Navab N. 2013.** Learning from multiple experts with random forests: application to the segmentation of the midbrain in 3D ultrasound. *Medical image computing and computer-assisted intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* **16**: 230–237.

## REFERENCES

---

- Chauveau D. 1995.** A stochastic EM algorithm for mixtures with censored data. *Journal of Statistical Planning and Inference* **46**: 1–25.
- Chien JC, Sussex IM. 1996.** Differential Regulation of Trichome Formation on the Adaxial and Abaxial Leaf Surfaces by Gibberellins and Photoperiod in *Arabidopsis thaliana* (L.) Heynh. *Plant Physiology* **111**: 1321–1328.
- Clark SE, Running MP, Meyerowitz EM. 1993.** CLAVATA1, a regulator of meristem and flower development in *Arabidopsis*. *Development* **119**: 397–418.
- Clark SE, Running MP, Meyerowitz EM. 1995.** CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. *Development* **121**: 2057–2067.
- Cnops G, Jover-Gil S, Peters JL, Neyt P, Block SD, Robles P, Ponce MR, Gerats T, Micol JL, Lijsebettens MV. 2004.** The rotunda2 mutants identify a role for the LEUNIG gene in vegetative leaf morphogenesis. *Journal of Experimental Botany* **55**: 1529–1539.
- Cockcroft CE, den Boer BGW, Healy JMS, Murray JAH. 2000.** Cyclin D control of growth rate in plants. *Nature* **405**: 575–579.
- Cookson SJ, Chenu K, Granier C. 2007.** Day Length Affects the Dynamics of Leaf Expansion and Cellular Development in *Arabidopsis thaliana* Partially through Floral Transition Timing. *Annals of Botany* **99**: 703–711.
- Cookson SJ, Van Lijsebettens M, Granier C. 2005.** Correlation between leaf growth variables suggest intrinsic and early controls of leaf size in *Arabidopsis thaliana*. *Plant, Cell & Environment* **28**: 1355–1366.
- Cookson SJ, Radziejewski A, Granier C. 2006.** Cell and leaf size plasticity in *Arabidopsis*: what is the role of endoreduplication? *Plant, Cell & Environment* **29**: 1273–1283.
- Cookson SJ, Turc O, Massonnet C, Granier C. 2010.** Phenotyping the Development of Leaf Area in *Arabidopsis thaliana*. In: Hennig L, Köhler C, eds. *Methods in Molecular Biology. Plant Developmental Biology*. Humana Press, 89–103.
- Costa MMR, Yang S, Critchley J, Feng X, Wilson Y, Langlade N, Copsey L, Hudson A. 2012.** The genetic basis for natural variation in heteroblasty in *Antirrhinum*. *New Phytologist* **196**: 1251–1259.
- Cutri L, Nave N, Ami MB, Chayut N, Samach A, Dornelas MC. 2013.** Evolutionary, genetic, environmental and hormonal-induced plasticity in the fate of organs arising from axillary meristems in *Passiflora* spp. *Mechanisms of Development* **130**: 61–69.
- Dambreville A, Lauri PE, Normand F, Guédon Y. in press.** Analysing jointly growth and development of plants using developmental growth stages. *Annals of Botany*.
- Damerval C. 1983.** Étude du développement hétéroblastique chez quelques espèces du genre *Medicago*. *Canadian Journal of Botany* **61**: 2212–2223.

## REFERENCES

---

- Day ME, Greenwood MS, Diaz-Sala C. 2002.** Age- and size-related trends in woody plant shoot development: regulatory pathways and evidence for genetic control. *Tree Physiology* **22**: 507–513.
- Dewitte W, Scofield S, Alcasabas AA, Maughan SC, Menges M, Braun N, Collins C, Nieuwland J, Prinsen E, Sundaresan V, Murray JAH. 2007.** Arabidopsis CYCD3 D-type cyclins link cell proliferation and endocycles and are rate-limiting for cytokinin responses. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 14537–14542.
- Diggle PK. 2002.** A developmental morphologist's perspective on plasticity. *Evolutionary Ecology* **16**: 267–283.
- Donnelly PM, Bonetta D, Tsukaya H, Dengler RE, Dengler NG. 1999.** Cell Cycling and Cell Enlargement in Developing Leaves of Arabidopsis. *Developmental Biology* **215**: 407–419.
- Dornbusch T, Watt J, Baccar R, Fournier C, Andrieu B. 2010.** A comparative analysis of leaf shape of wheat, barley and maize using an empirical shape model. *Annals of Botany*: mcq181.
- Dosio GAA, Rey H, Lecoeur J, Izquierdo NG, Aguirrezábal LAN, Tardieu F, Turc O. 2003.** A whole-plant analysis of the dynamics of expansion of individual leaves of two sunflower hybrids. *Journal of Experimental Botany* **54**: 2541–2552.
- Elsner J, Michalski M, Kwiatkowska D. 2012.** Spatiotemporal variation of leaf epidermal cell growth: a quantitative analysis of Arabidopsis thaliana wild-type and triple cyclinD3 mutant plants. *Annals of Botany* **109**: 897–910.
- Evans GC. 1972.** *The Quantitative Analysis of Plant Growth*. University of California Press.
- Feng G, Qin Z, Yan J, Zhang X, Hu Y. 2011.** Arabidopsis ORGAN SIZE RELATED1 regulates organ growth and final organ size in orchestration with ARGOS and ARL. *New Phytologist* **191**: 635–646.
- Ferjani A, Horiguchi G, Yano S, Tsukaya H. 2007.** Analysis of Leaf Development in fugu Mutants of Arabidopsis Reveals Three Compensation Modes That Modulate Cell Expansion in Determinate Organs. *Plant Physiology* **144**: 988–999.
- Fiorani F, Beemster GTS. 2006.** Quantitative Analyses of Cell Division in Plants. *Plant Molecular Biology* **60**: 963–979.
- Fiorani F, Rascher U, Jahnke S, Schurr U. 2012.** Imaging plants dynamics in heterogenic environments. *Current Opinion in Biotechnology* **23**: 227–235.
- Fisher JB, Posluszny U, Lee DW. 2002.** Shade promotes thorn development in a tropical liana, Artabotrys hexapetalus (Annonaceae). *International Journal of Plant Sciences* **163**: 295–300.
- Fournier C, Andrieu B. 2000.** Dynamics of the Elongation of Internodes in Maize (Zea mays L.): Analysis of Phases of Elongation and their Relationships to Phytomer Development. *Annals of Botany* **86**: 551–563.

## REFERENCES

---

- Franks SJ. 2011.** Plasticity and evolution in drought avoidance and escape in the annual plant *Brassica rapa*. *New Phytologist* **190**: 249–257.
- Friedli M, Walter A. 2014.** Diel growth patterns of young soybean ( *Glycine max* ) leaflets are synchronous throughout different positions on a plant: Synchronous growth patterns of soybean leaflets. *Plant, Cell & Environment*: n/a–n/a.
- Fujikura U, Horiguchi G, Ponce MR, Micol JL, Tsukaya H. 2009.** Coordination of cell proliferation and cell expansion mediated by ribosome-related processes in the leaves of *Arabidopsis thaliana*. *The Plant Journal* **59**: 499–508.
- Fujikura U, Horiguchi G, Tsukaya H. 2007.** Dissection of Enhanced Cell Expansion Processes in Leaves Triggered by a Defect in Cell Proliferation, with Reference to Roles of Endoreduplication. *Plant and Cell Physiology* **48**: 278–286.
- Furbank RT, Tester M. 2011.** Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Science* **16**: 635–644.
- Galbraith DW, Harkins KR, Knapp S. 1991.** Systemic Endopolyploidy in *Arabidopsis thaliana*. *Plant Physiology* **96**: 985–989.
- Ghandilyan A, Barboza L, Tisné S, Granier C, Reymond M, Koornneef M, Schat H, Aarts MGM. 2009.** Genetic analysis identifies quantitative trait loci controlling rosette mineral concentrations in *Arabidopsis thaliana* under drought. *New Phytologist* **184**: 180–192.
- Goebel K. 1900.** *Organography of plants. I. General organography*. New York: Hafner Publishers.
- González-Bayón R, Kinsman EA, Quesada V, Vera A, Robles P, Ponce MR, Pyke KA, Micol JL. 2006.** Mutations in the RETICULATA gene dramatically alter internal architecture but have little effect on overall organ shape in *Arabidopsis* leaves. *Journal of Experimental Botany* **57**: 3019–3031.
- Gonzalez N, Vanhaeren H, Inzé D. 2012.** Leaf size control: complex coordination of cell division and expansion. *Trends in Plant Science* **17**: 332–340.
- Gould KS. 1993.** Leaf Heteroblasty in *Pseudopanax crassifolius*: Functional Significance of Leaf Morphology and Anatomy. *Annals of Botany* **71**: 61–70.
- Granier C, Aguirrezabal L, Chenu K, Cookson SJ, Dauzat M, Hamard P, Thioux J-J, Rolland G, Bouchier-Combaud S, Lebaudy A, Muller B, Simonneau T, Tardieu F. 2006.** PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytologist* **169**: 623–635.
- Granier C, Inze D, Tardieu F. 2000.** Spatial Distribution of Cell Division Rate Can Be Deduced from that of p34cdc2 Kinase Activity in Maize Leaves Grown at Contrasting Temperatures and Soil Water Conditions. *Plant Physiology* **124**: 1393–1402.
- Granier C, Nègre V, Fiorani F. 2014.** Systematic Phenotyping of Plant Development in *Arabidopsis thaliana*. *Phenomics*. CRC Press, 111–141.



## REFERENCES

---

**Granier C, Tardieu F. 1998.** Spatial and temporal analyses of expansion and cell cycle in sunflower leaves. A common pattern of development for all zones of a leaf and different leaves of a plant. *Plant Physiology* **116**: 991–1001.

**Granier C, Tardieu F. 2009.** Multi-scale phenotyping of leaf expansion in response to environmental changes: the whole is more than the sum of parts. *Plant, Cell & Environment* **32**: 1175–1184.

**Guédon Y. 2003.** Estimating hidden semi-Markov chains from discrete sequences. *Journal of Computational and Graphical Statistics* **12**: 604–639.

**Guédon Y. 2005.** Hidden hybrid Markov/semi-Markov chains. *Computational Statistics & Data Analysis* **49**: 663–688.

**Guédon Y. 2007.** Exploring the state sequence space for hidden Markov and semi-Markov chains. *Computational Statistics & Data Analysis* **51**: 2379–2409.

**Guédon Y, Barthélémy D, Caraglio Y, Costes E. 2001.** Pattern analysis in branching and axillary flowering sequences. *Journal of Theoretical Biology* **212**: 481–520.

**Hoagland DR, Arnon DI. 1950.** The water-culture method for growing plants without soil. *Circular. California Agricultural Experiment Station* **347**: 32 pp.

**Horiguchi G, Ferjani A, Fujikura U, Tsukaya H. 2006a.** Coordination of cell proliferation and cell expansion in the control of leaf size in *Arabidopsis thaliana*. *Journal of Plant Research* **119**: 37–42.

**Horiguchi G, Fujikura U, Ferjani A, Ishikawa N, Tsukaya H. 2006b.** Large-scale histological analysis of leaf mutants using two simple leaf observation methods: identification of novel genetic pathways governing the size and shape of leaves. *The Plant Journal* **48**: 638–644.

**Huijser P, Schmid M. 2011.** The control of developmental phase transitions in plants. *Development* **138**: 4117–4129.

**Hülkamp M, Miséra S, Jürgens G. 1994.** Genetic dissection of trichome cell development in *Arabidopsis*. *Cell* **76**: 555–566.

**Hunter C, Sun H, Poethig RS. 2003.** The *Arabidopsis* Heterochronic Gene ZIPPY Is an ARGONAUTE Family Member. *Current Biology* **13**: 1734–1739.

**Itoh Y, Shimizu H. 2012.** Phyllochron dynamics during the course of late shoot development might be affected by reproductive development in rice (*Oryza sativa* L.). *Development Genes and Evolution* **222**: 341–350.

**John PCL, Qi R. 2008.** Cell division and endoreduplication: doubtful engines of vegetative growth. *Trends in Plant Science* **13**: 121–127.

**Jones CS. 1995.** Does Shade Prolong Juvenile Development? A Morphological Analysis of Leaf Shape Changes in *Cucurbita argyrosperma* Subsp. *Sororia* (Cucurbitaceae). *American Journal of Botany* **82**: 346–359.

## REFERENCES

---

**Jones CS. 1999.** An Essay on Juvenility, Phase Change, and Heteroblasty in Seed Plants. *International Journal of Plant Sciences* **160**: S105–S111.

**Jönsson H, Heisler MG, Shapiro BE, Meyerowitz EM, Mjolsness E. 2006.** An auxin-driven polarized transport model for phyllotaxis. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 1633–1638.

**Juenger TE, McKay JK, Hausmann N, Keurentjes JJB, Sen S, Stowe KA, Dawson TE, Simms EL, Richards JH. 2005.** Identification and characterization of QTL underlying whole-plant physiology in *Arabidopsis thaliana*:  $\delta^{13}\text{C}$ , stomatal conductance and transpiration efficiency. *Plant, Cell & Environment* **28**: 697–708.

**Kalu BA, Fick GW. 1981.** Quantifying Morphological Development of Alfalfa for Studies of Herbage Quality1. *Crop Science* **21**: 267.

**Kalve S, De Vos D, Beemster GT. 2014.** Leaf development: a cellular perspective. *Plant Systems Biology* **5**: 362.

**Kawade K, Horiguchi G, Tsukaya H. 2010.** Non-cell-autonomously coordinated organ size regulation in leaf. *Development* **137**: 4221–4227.

**Kazama T, Ichihashi Y, Murata S, Tsukaya H. 2010.** The Mechanism of Cell Cycle Arrest Front Progression Explained by a KLUH/CYP78A5-dependent Mobile Growth Factor in Developing Leaves of *Arabidopsis thaliana*. *Plant and Cell Physiology* **51**: 1046–1054.

**Kerstetter RA, Poethig RS. 1998.** The Specification of Leaf Identity During Shoot Development. *Annual Review of Cell and Developmental Biology* **14**: 373–398.

**Khan GA, Bouraine S, Wege S, Li Y, de Carbonnel M, Berthomieu P, Poirier Y, Rouached H. 2014.** Coordination between zinc and phosphate homeostasis involves the transcription factor PHR1, the phosphate exporter PHO1, and its homologue PHO1;H3 in *Arabidopsis*. *Journal of Experimental Botany* **65**: 871–884.

**Koornneef M, Hanhart CJ, Veen JH van der. 1991.** A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Molecular and General Genetics MGG* **229**: 57–66.

**Koornneef M, Meinke D. 2010.** The development of *Arabidopsis* as a model plant. *The Plant Journal* **61**: 909–921.

**Lawson EJR, Poethig RS. 1995.** Shoot development in plants: time for a change. *Trends in Genetics* **11**: 263–268.

**Lee DW, Richards JH. 1991.** Heteroblastic development in vines. Putz, FE., Mooney, HA., editors. *The Biology of Vines*. Cambridge University Press, 205–243.

**Lenth RV. 2001.** Some Practical Guidelines for Effective Sample Size Determination. *The American Statistician* **55**: 187–193.

**Lièvre M, Wuyts N, Cookson SJ, Bresson J, Dapp M, Vasseur F, Massonnet C, Tisné S, Bettembourg M, Balsera C, Bédiée A, Bouvery F, Dauzat M, Rolland G, Vile D, Granier C. 2013.** Phenotyping the kinematics of leaf development in flowering plants:

## REFERENCES

---

recommendations and pitfalls. *Wiley Interdisciplinary Reviews: Developmental Biology* **2**: 809–821.

**Lwanga SK, Lemeshow S. 1991.** *Sample size determination in health studies: a practical manual*. Geneva : World Health Organization.

**Mandel T, Moreau F, Kutsher Y, Fletcher JC, Carles CC, Williams LE. 2014.** The ERECTA receptor kinase regulates Arabidopsis shoot apical meristem size, phyllotaxy and floral meristem identity. *Development* **141**: 830–841.

**Marcotrigiano M. 2010.** A role for leaf epidermis in the control of leaf size and the rate and extent of mesophyll cell division. *American Journal of Botany* **97**: 224–233.

**Massonnet C, Tisné S, Radziejowski A, Vile D, Veylder LD, Dauzat M, Granier C. 2011.** New Insights into the Control of Endoreduplication: Endoreduplication Could Be Driven by Organ Growth in Arabidopsis Leaves. *Plant Physiology* **157**: 2044–2055.

**Matsoukas IG. 2014.** Attainment of reproductive competence, phase transition, and quantification of juvenility in mutant genetic screens. *Frontiers in Plant Science* **5**.

**Matsoukas IG, Massiah AJ, Thomas B. 2013.** Starch metabolism and antiflorigenic signals modulate the juvenile-to-adult phase transition in Arabidopsis. *Plant, Cell & Environment* **36**: 1802–1811.

**Mauseth JD. 2004.** Giant Shoot Apical Meristems in Cacti Have Ordinary Leaf Primordia but Altered Phyllotaxy and Shoot Diameter. *Annals of Botany* **94**: 145–153.

**McLachlan G, Peel D. 2000.** *Finite Mixture Models*. New York: John Wiley & Sons.

**Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M. 1998.** Arabidopsis thaliana: A Model Plant for Genome Analysis. *Science* **282**: 662–682.

**Melaragno JE, Mehrotra B, Coleman AW. 1993.** Relationship between Endopolyploidy and Cell Size in Epidermal Tissue of Arabidopsis. *The Plant Cell Online* **5**: 1661–1668.

**Mencuccini M, Martínez-Vilalta J, Hamid HA, Korakaki E, Vanderklein D. 2007.** Evidence for age- and size-mediated controls of tree growth from grafting studies. *Tree Physiology* **27**: 463–473.

**Méndez-Vigo B, Andrés MT de, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C. 2010.** Temporal analysis of natural variation for the rate of leaf production and its relationship with flowering initiation in Arabidopsis thaliana. *Journal of Experimental Botany* **61**: 1611–1623.

**Merrill EK. 1986.** Heteroblastic seedlings of green ash. I. Predictability of leaf form and primordial length. *Canadian Journal of Botany* **64**: 2645–2649.

**Metcalf CJE, Mitchell-Olds T. 2009.** Life history in a model system: opening the black box with Arabidopsis thaliana. *Ecology Letters* **12**: 593–600.

## REFERENCES

---

- Mozley D, Thomas B. 1995.** Developmental and photobiological factors affecting photoperiodic induction in *Arabidopsis thaliana* Heynh. *Landsberg erecta*. *Journal of Experimental Botany* **46**: 173–179.
- Nath U, Crawford BCW, Carpenter R, Coen E. 2003.** Genetic Control of Surface Curvature. *Science* **299**: 1404–1407.
- Ng DW-K, Zhang C, Miller M, Shen Z, Briggs SP, Chen ZJ. 2012.** Proteomic divergence in *Arabidopsis* autopolyploids and allopolyploids and their progenitors. *Heredity* **108**: 419–430.
- Noordzij M, Dekker FW, Zoccali C, Jager KJ. 2011.** Sample Size Calculations. *Nephron Clinical Practice* **118**: c319–c323.
- O'Malley RC, Ecker JR. 2010.** Linking genotype to phenotype using the *Arabidopsis* unimutant collection. *The Plant Journal* **61**: 928–940.
- Peaucelle A, Laufs P. 2007.** Phyllotaxy: Beyond the Meristem and Auxin Comes the miRNA. *Plant Signaling & Behavior* **2**: 293–295.
- Pereyra-Irujo GA, Velázquez L, Lechner L, Aguirrezábal LAN. 2008.** Genetic variability for leaf growth rate and duration under water deficit in sunflower: analysis of responses at cell, organ, and plant level. *Journal of Experimental Botany* **59**: 2221–2232.
- Pérez-Pérez JM, Rubio-Díaz S, Dhondt S, Hernández-Romero D, Sánchez-Soriano J, Beemster GTS, Ponce MR, Micol JL. 2011.** Whole organ, venation and epidermal cell morphological variations are correlated in the leaves of *Arabidopsis* mutants. *Plant, Cell & Environment* **34**: 2200–2211.
- Poethig RS. 1990.** Phase Change and the Regulation of Shoot Morphogenesis in Plants. *Science* **250**: 923–930.
- Poethig RS. 2003.** Phase Change and the Regulation of Developmental Timing in Plants. *Science* **301**: 334–336.
- Poethig RS. 2010.** The Past, Present, and Future of Vegetative Phase Change. *Plant Physiology* **154**: 541–544.
- Poethig RS. 2013.** Vegetative phase change and shoot maturation in plants. *Current topics in developmental biology* **105**: 125–152.
- Rasband WS. 1997-2014.** *ImageJ*. Bethesda, Maryland, USA: U. S. National Institutes of Health, <http://imagej.nih.gov/ij/>
- Ray TS. 1987.** Cyclic Heterophylly in *Syngonium* (Araceae). *American Journal of Botany* **74**: 16–26.
- R Core Team. 2014.** *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, <http://www.R-project.org>
- Reinhardt D, Mandel T, Kuhlemeier C. 2000.** Auxin Regulates the Initiation and Radial Position of Plant Lateral Organs. *The Plant Cell Online* **12**: 507–518.

## REFERENCES

---

- De Reuille PB de, Bohn-Courseau I, Ljung K, Morin H, Carraro N, Godin C, Traas J. 2006.** Computer simulations reveal properties of the cell-cell signaling network at the shoot apex in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 1627–1632.
- Röbbelen G. 1957.** Über Heterophyllie bei Arabidopsis thaliana (L.) Heynh. *Berichte der Deutschen Botanischen Gesellschaft* **70**: 39–44.
- Roeder AHK, Chickarmane V, Cunha A, Obara B, Manjunath BS, Meyerowitz EM. 2010.** Variability in the Control of Cell Division Underlies Sepal Epidermal Patterning in Arabidopsis thaliana. *PLoS Biology* **8**.
- Roff DA. 2007.** Contributions of genomics to life-history theory. *Nature Reviews Genetics* **8**: 116–125.
- Rymen B, Coppens F, Dhondt S, Fiorani F, Beemster GTS. 2010.** Kinematic Analysis of Cell Division and Expansion. In: Hennig L, Köhler C, eds. *Methods in Molecular Biology. Plant Developmental Biology*. Humana Press, 203–227.
- Savaldi-Goldstein S, Chory J. 2008.** Growth coordination and the shoot epidermis. *Current opinion in plant biology* **11**: 42–48.
- Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D. 2005.** Specific effects of microRNAs on the plant transcriptome. *Developmental Cell* **8**: 517–527.
- Serrano-Cartagena J, Robles P, Ponce MR, Micol JL. 1999.** Genetic analysis of leaf form mutants from the Arabidopsis Information Service collection. *Molecular & general genetics: MGG* **261**: 725–739.
- Shikata M, Koyama T, Mitsuda N, Ohme-Takagi M. 2009.** Arabidopsis SBP-box genes SPL10, SPL11 and SPL2 control morphological change in association with shoot maturation in the reproductive phase. *Plant & Cell Physiology* **50**: 2133–2145.
- Skirycz A, Bodt SD, Obata T, Clercq ID, Claeys H, Rycke RD, Andriankaja M, Aken OV, Breusegem FV, Fernie AR, Inzé D. 2010.** Developmental Stage Specificity and the Role of Mitochondrial Metabolism in the Response of Arabidopsis Leaves to Prolonged Mild Osmotic Stress. *Plant Physiology* **152**: 226–244.
- Sommer C, Strähle C, Köthe U, Hamprecht F. 2011.** ilastik: Interactive learning and segmentation toolkit. *Eighth IEEE International Symposium on Biomedical Imaging (ISBI)*. 230–233.
- Srikanth A, Schmid M. 2011.** Regulation of flowering time: all roads lead to Rome. *Cellular and Molecular Life Sciences* **68**: 2013–2037.
- Sterken R, Kiekens R, Boruc J, Zhang F, Vercauteren A, Vercauteren I, De Smet L, Dhondt S, Inze D, De Veylder L, Russinova E, Vuylsteke M. 2012.** Combined linkage and association mapping reveals CYCD5;1 as a quantitative trait gene for endoreduplication in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 4678–4683.

## REFERENCES

---

**Steynen QJ, Bolokoski DA, Schultz EA. 2001.** Alteration in flowering time causes accelerated or decelerated progression through Arabidopsis vegetative phases. *Canadian Journal of Botany* **79**: 657–665.

**Sylvester AW, Parker-Clark V. 2001.** Leaf shape and anatomy as indicators of phase change in the grasses: comparison of maize, rice, and bluegrass. *American journal of botany* **88**: 2157–67.

**Takahashi N, Quimbaya M, Schubert V, Lammens T, Vandepoele K, Schubert I, Matsui M, Inzé D, Berx G, De Veylder L. 2010.** The MCM-Binding Protein ETG1 Aids Sister Chromatid Cohesion Required for Postreplicative Homologous Recombination Repair. *PLoS Genet* **6**: e1000817.

**Telfer A, Bollman KM, Poethig RS. 1997.** Phase change and the regulation of trichome distribution in Arabidopsis thaliana. *Development* **124**: 645–654.

**The Arabidopsis Information Resource (TAIR). 2014.**  
<http://www.arabidopsis.org/servlets/TairObject?id=127757&type=locus>.  
[www.arabidopsis.org](http://www.arabidopsis.org).

**Tisné S, Barbier F, Granier C. 2011.** The ERECTA gene controls spatial and temporal patterns of epidermal cell number and size in successive developing leaves of Arabidopsis thaliana. *Annals of Botany* **108**: 159–168.

**Tisné S, Reymond M, Vile D, Fabre J, Dauzat M, Koornneef M, Granier C. 2008.** Combined Genetic and Modeling Approaches Reveal That Epidermal Cell Area and Number in Leaves Are Controlled by Leaf and Plant Developmental Processes in Arabidopsis. *Plant Physiology* **148**: 1117–1127.

**Tisné S, Schmalenbach I, Reymond M, Dauzat M, Pervent M, Vile D, Granier C. 2010.** Keep on growing under drought: genetic and developmental bases of the response of rosette area using a recombinant inbred line population. *Plant, Cell & Environment* **33**: 1875–1887.

**Tojo T, Tsuda K, Yoshizumi T, Ikeda A, Yamaguchi J, Matsui M, Yamazaki K. 2009.** Arabidopsis MBF1s Control Leaf Cell Cycle and its Expansion. *Plant and Cell Physiology* **50**: 254–264.

**Torres M, Frutos G. 1989.** Analysis of germination curves of aged fennel seeds by mathematical models. *Environmental and Experimental Botany* **29**: 409–415.

**Tsukaya H. 2002.** Interpretation of mutants in leaf morphology: Genetic evidence for a compensatory system in leaf morphogenesis that provides a new link between cell and organismal theories. In: Kwang W. Jeon, ed. A Survey of Cell Biology. *International Review of Cytology*. Academic Press, 1–39.

**Tsukaya H. 2013.** Leaf Development. *The Arabidopsis Book / American Society of Plant Biologists* **11**.

**Usami T, Horiguchi G, Yano S, Tsukaya H. 2009.** The more and smaller cells mutants of Arabidopsis thaliana identify novel roles for SQUAMOSA PROMOTER BINDING PROTEIN-LIKE genes in the control of heteroblasty. *Development* **136**: 955–964.

## REFERENCES

---

- Vanhaeren H, Gonzalez N, Inzé D. 2010.** Hide and seek: uncloaking the vegetative shoot apex of *Arabidopsis thaliana*. *The Plant Journal* **63**: 541–548.
- Vasseur F, Violle C, Enquist BJ, Granier C, Vile D. 2012.** A common genetic basis to the origin of the leaf economics spectrum and metabolic scaling allometry. *Ecology Letters* **15**: 1149–1157.
- De Veylder L, Beeckman T, Beemster GTS, Krols L, Terras F, Landrieu I, Schueren EVD, Maes S, Naudts M, Inzé D. 2001.** Functional Analysis of Cyclin-Dependent Kinase Inhibitors of *Arabidopsis*. *The Plant Cell Online* **13**: 1653–1668.
- De Veylder L, Larkin JC, Schnittger A. 2011.** Molecular control and function of endoreplication in development and physiology. *Trends in Plant Science* **16**: 624–634.
- Vlieghe K, Boudolf V, Beemster GTS, Maes S, Magyar Z, Atanassova A, de Almeida Engler J, De Groodt R, Inzé D, De Veylder L. 2005.** The DP-E2F-like Gene DEL1 Controls the Endocycle in *Arabidopsis thaliana*. *Current Biology* **15**: 59–63.
- Vlieghe K, Inzé D, de Veylder L. 2007.** Physiological Relevance and Molecular Control of the Endocycle in Plants. In: Inzé D, ed. *Annual Plant Reviews Volume 32: Cell Cycle Control and Plant Development*. Blackwell Publishing Ltd, 227–248.
- Wang J-W, Czech B, Weigel D. 2009.** miR156-Regulated SPL Transcription Factors Define an Endogenous Flowering Pathway in *Arabidopsis thaliana*. *Cell* **138**: 738–749.
- Wang J-W, Schwab R, Czech B, Mica E, Weigel D. 2008.** Dual Effects of miR156-Targeted SPL Genes and CYP78A5/KLUH on Plastochron Length and Organ Size in *Arabidopsis thaliana*. *The Plant Cell Online* **20**: 1231–1243.
- Willmann MR, Poethig RS. 2011.** The effect of the floral repressor FLC on the timing and progression of vegetative phase change in *Arabidopsis*. *Development* **138**: 677–685.
- Wilson-Sánchez D, Rubio-Díaz S, Muñoz-Viana R, Pérez-Pérez JM, Jover-Gil S, Ponce MR, Micol JL. 2014.** Leaf phenomics: a systematic reverse genetic screen for *Arabidopsis* leaf mutants. *The Plant Journal* **79**: 878–891.
- Wulff RD. 1985.** Effect of Seed Size on Heteroblastic Development in Seedlings of *Desmodium paniculatum*. *American Journal of Botany* **72**: 1684.
- Wu G, Park MY, Conway SR, Wang J-W, Weigel D, Poethig RS. 2009.** The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell* **138**: 750–759.
- Wu G, Poethig RS. 2006.** Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. *Development* **133**: 3539–3547.
- Wuyts N, Palauqui J-C, Conejero G, Verdeil J-L, Granier C, Massonnet C. 2010.** High-contrast three-dimensional imaging of the *Arabidopsis* leaf enables the analysis of cell dimensions in the epidermis and mesophyll. *Plant Methods* **6**: 17.
- Yang L, Conway SR, Poethig RS. 2011.** Vegetative phase change is mediated by a leaf-derived signal that represses the transcription of miR156. *Development* **138**: 245–249.

## REFERENCES

---

**Yang L, Xu M, Koo Y, He J, Poethig RS. 2013.** Sugar promotes vegetative phase change in *Arabidopsis thaliana* by repressing the expression of MIR156A and MIR156C. *eLife* **2**.

**Zhang X, Hause RJ, Borevitz JO. 2012.** Natural Genetic Variation for Growth and Development Revealed by High-Throughput Phenotyping in *Arabidopsis thaliana*. *G3: Genes/Genomes/Genetics* **2**: 29–34.

**Zotz G, Wilhelm K, Becker A. 2011.** Heteroblasty—A Review. *The Botanical Review* **77**: 109–151.



---

## APPENDIX I. LIST OF THE T-DNA INSERTION LINES GROWN IN EXPERIMENTS 1 & 2

90 T-DNA insertion lines were selected for their contrasted leaf growth phenotypes identified in a previous screen performed by José Luis Micol group (Universidad Miguel Hernandez, Elche, Alicante, Spain). In this table, lines are ranked from the one with the highest mean rosette area to the one with the lowest mean rosette area 18 days after sowing. The dotted line indicates the rank of the Col-0 accession (wild-type).

| SALK Line                 | Gene      | SALK Line                 | Gene      | SALK Line                 | Gene      |
|---------------------------|-----------|---------------------------|-----------|---------------------------|-----------|
| SALK_138229C <sup>2</sup> | At1g54840 | SALK_021759C <sup>1</sup> | At4g32200 | SALK_144264C <sup>2</sup> | At3g19510 |
| SALK_080604C <sup>2</sup> | At1g08135 | SALK_034227C <sup>2</sup> | At5g59950 | SALK_070464C <sup>1</sup> | At1g74420 |
| SALK_119457C <sup>2</sup> | At4g30410 | SALK_055996C <sup>2</sup> | At1g32500 | SALK_014243C <sup>1</sup> | At3g52105 |
| SALK_018664C <sup>2</sup> | At4g09000 | SALK_020801C <sup>2</sup> | At2g01450 | SALK_122867C <sup>1</sup> | At1g18900 |
| SALK_032963C <sup>2</sup> | At5g16270 | SALK_148633C <sup>2</sup> | At4g20450 | SALK_015088C <sup>1</sup> | At5g61960 |
| SALK_085503C <sup>2</sup> | At2g33420 | SALK_001496C <sup>2</sup> | At1g62430 | SALK_021618C <sup>1</sup> | At2g47490 |
| SALK_143422C <sup>2</sup> | At2g44650 | SALK_007854C <sup>1</sup> | At4g11120 | SALK_046141C <sup>1</sup> | At2g03190 |
| SALK_054681C <sup>2</sup> | At5g46115 | SALK_015522C <sup>1</sup> | At5g36880 | SALK_064915C <sup>1</sup> | At5g23980 |
| SALK_009736C <sup>2</sup> | At1g17130 | SALK_001004C <sup>2</sup> | At2g30810 | SALK_049200C <sup>1</sup> | At2g04043 |
| SALK_045623C <sup>2</sup> | At5g65050 | SALK_034684C <sup>1</sup> | At2g22420 | SALK_055458C <sup>1</sup> | At1g14040 |
| SALK_021217C <sup>2</sup> | At3g47640 | SALK_020615C <sup>1</sup> | At1g76470 | SALK_048175C <sup>1</sup> | At1g30450 |
| SALK_145203C <sup>1</sup> | At4g08540 | SALK_024759C <sup>1</sup> | At1g48950 | SALK_151603C <sup>1</sup> | At3g62980 |
| SALK_026667C <sup>2</sup> | At5g28250 | SALK_035676C <sup>2</sup> | At1g10050 | SALK_145086C <sup>1</sup> | At1g13740 |
| SALK_126071C <sup>2</sup> | At1g04730 | SALK_056529C <sup>1</sup> | At1g20860 | SALK_075661C <sup>1</sup> | At3g13228 |
| SALK_053198C <sup>1</sup> | At5g24470 | SALK_022117C <sup>2</sup> | At2g01290 | SALK_130499C <sup>2</sup> | At4g31390 |
| SALK_019994C <sup>2</sup> | At1g55370 | SALK_067582C <sup>1</sup> | At2g32540 | SALK_065118C <sup>1</sup> | At1g53140 |
| SALK_019359C <sup>2</sup> | At4g10920 | SALK_148403C <sup>1</sup> | At5g61950 | SALK_063595C <sup>1</sup> | At1g20640 |
| SALK_025598C <sup>2</sup> | At2g46970 | SALK_048174C <sup>1</sup> | At4g11540 | SALK_057785C <sup>1</sup> | At1g22090 |
| SALK_080188C <sup>2</sup> | At5g58970 | SALK_057052C <sup>2</sup> | At5g54770 | SALK_045025C <sup>2</sup> | At3g20550 |
| SALK_003718C <sup>2</sup> | At3g16950 | SALK_072771C <sup>1</sup> | At2g07540 | SALK_079285C <sup>2</sup> | At3g17040 |
| SALK_011867C <sup>1</sup> | At5g51750 | SALK_061494C <sup>1</sup> | At5g35220 | SALK_136507C <sup>2</sup> | At1g78020 |
| SALK_030786C <sup>2</sup> | At3g46790 | SALK_017692C <sup>1</sup> | At2g29670 | SALK_129037C <sup>2</sup> | At4g13590 |
| SALK_011586C <sup>1</sup> | At3g27620 | SALK_142112C <sup>1</sup> | At1g74940 |                           |           |
| SALK_041291C <sup>2</sup> | At2g22680 | SALK_009798C <sup>1</sup> | At2g23090 |                           |           |
| SALK_033455C <sup>1</sup> | At3g28860 | SALK_008561C <sup>1</sup> | At4g20290 |                           |           |
| SALK_129352C <sup>2</sup> | At3g30180 | SALK_040660C <sup>1</sup> | At1g79090 |                           |           |
| SALK_145983C <sup>1</sup> | At4g09340 | SALK_086630C <sup>1</sup> | At3g50400 |                           |           |
| SALK_066708C <sup>1</sup> | At4g24175 | SALK_101771C <sup>2</sup> | At1g18500 |                           |           |
| SALK_075797C <sup>2</sup> | At5g51770 | SALK_003711C <sup>1</sup> | At1g71760 |                           |           |
| SALK_044119C <sup>1</sup> | At2g35040 | SALK_045034C <sup>2</sup> | At5g17660 |                           |           |
| SALK_037549C <sup>1</sup> | At2g23220 | SALK_067017C <sup>1</sup> | At5g52440 |                           |           |
| SALK_004741C <sup>2</sup> | At1g42980 | SALK_019175C <sup>1</sup> | At2g31725 |                           |           |
| SALK_125189C <sup>2</sup> | At1g61310 | SALK_012771C <sup>2</sup> | At4g25410 |                           |           |
| SALK_117972C <sup>1</sup> | At2g44100 | SALK_005153C <sup>1</sup> | At5g13680 |                           |           |

<sup>1</sup> Plants grown in Experiment 1.

<sup>2</sup> Plants grown in Experiment 2.

## APPENDIX II. SUPPLEMENTARY METHODS FOR INTEGRATIVE MODELS FOR ANALYZING *A. THALIANA* ROSETTE DEVELOPMENTAL PHASES

### DEFINITION OF SEMI-MARKOV SWITCHING MODELS AND ASSOCIATED STATISTICAL METHODS

#### *Semi-Markov chains*

Let  $\{S_t\}$  be a semi-Markov chain with finite-state space  $\{0, \dots, J-1\}$ . A  $J$ -state semi-Markov chain  $\{S_t\}$  is defined by the following parameters:

- initial probabilities  $\pi_j = P(S_1 = j)$  with  $\sum_j \pi_j = 1$ ;
- transition probabilities
  - non-absorbing state  $i$ : for each  $j \neq i$ ,  $p_{ij} = P(S_t = j | S_t \neq i, S_{t-1} = i)$  with  $\sum_{j \neq i} p_{ij} = 1$  and  $p_{ii} = 0$  by convention,
  - absorbing state  $i$ :  $p_{ii} = P(S_t = i | S_{t-1} = i) = 1$  and for each  $j \neq i$ ,  $p_{ij} = 0$ .

An explicit occupancy distribution is attached to each non-absorbing state:

$$d_j(u) = P(S_{t+u+1} \neq j, S_{t+u-v} = j, v = 0, \dots, u-2 | S_{t+1} = j, S_t \neq j), \quad u = 1, 2, \dots$$

Since  $t = 1$  is assumed to correspond to a state entering, the following relation is verified:

$$P(S_t \neq j, S_{t-v} = j, v = 1, \dots, t) = d_j(t) \pi_j.$$

We define as possible parametric state occupancy distributions binomial distributions, Poisson distributions and negative binomial distributions with an additional shift parameter  $d$  ( $d \geq 1$ ) which defines the minimum sojourn time in a given state.

The binomial distribution with parameters  $d$ ,  $n$  and  $p$  ( $q = 1 - p$ ),  $B(d, n, p)$  where  $0 \leq p \leq 1$ , is defined by

$$d_j(u) = \binom{n-d}{u-d} p^{u-d} q^{n-u}, \quad u = d, d+1, \dots, n.$$

The Poisson distribution with parameters  $d$  and  $\lambda$ ,  $P(d, \lambda)$ , where  $\lambda$  is a real number ( $\lambda > 0$ ), is defined by

$$d_j(u) = \frac{e^{-\lambda} \lambda^{u-d}}{(u-d)!}, \quad u = d, d+1, \dots$$

The negative binomial distribution with parameters  $d$ ,  $r$  and  $p$ ,  $\text{NB}(d, r, p)$ , where  $r$  is a real number ( $r > 0$ ) and  $0 < p \leq 1$ , is defined by

$$d_j(u) = \binom{u-d+r-1}{r-1} p^r q^{u-d}, \quad u = d, d+1, \dots$$

### *Semi-Markov switching models*

A semi-Markov switching model can be viewed as a pair of stochastic processes  $\{S_t, X_t\}$  where the ‘output’ process  $\{X_t\}$  is related to the ‘state’ process  $\{S_t\}$ , which is a finite-state semi-Markov chain, by a probabilistic function or mapping denoted by  $f$  (hence  $X_t = f(S_t)$ ). Since the mapping  $f$  is such that a given output may be observed in different states, the state process  $\{S_t\}$  is not observable directly but only indirectly through the output process  $\{X_t\}$ . This output process  $\{X_t\}$  is related to the semi-Markov chain  $\{S_t\}$  by the observation (or emission) models. The output process at time  $t$  depends only on the underlying semi-Markov chain at time  $t$ . The extension to the multivariate case is straightforward since, in this latter case, the elementary observed variables at time  $t$  are assumed to be conditionally independent given the state  $S_t = s_t$ .

In the case of the trichome variable, the observation probabilities  $b_{j,1}(y) = P(X_{t,1} = y | S_t = j)$  were directly estimated (categorical observation distribution).

In the case of the quantitative/*interval-scaled* variables (length-to-width ratio, final leaf area, characteristic growth duration), the  $e$ -th output process  $\{X_{t,e}\}$  is related to the state process  $\{S_t\}$ , by a linear trend model

$$X_{t,e} = \alpha_{j,e} + \beta_{j,e} t + \varepsilon_{j,e}, \quad \varepsilon_{j,e} \sim \text{N}(0, \sigma_{j,e}^2).$$

The maximum likelihood estimation of the parameters of a semi-Markov switching model requires an iterative optimization technique, which is an application of the EM algorithm. Once a semi-Markov switching model has been estimated, the most probable state sequence  $\mathbf{s}^*$  with its associated posterior probability  $P(\mathbf{S} = \mathbf{s}^* | \mathbf{X} = \mathbf{x})$  can be computed for each observed sequence  $\mathbf{x}$  using the so-called Viterbi algorithm (Guédon 2003). In our application

## APPENDIX II

---

context, the most probable state sequence can be interpreted as the optimal segmentation of the corresponding observed sequence in successive developmental phases; see (Guédon 2003, 2005, 2007) for the statistical methods for hidden semi-Markov chains that directly apply to semi-Markov switching models.

### APPENDIX III. SUPPLEMENTARY FIGURES FOR INTEGRATIVE MODELS FOR ANALYZING *A. THALIANA* ROSETTE DEVELOPMENTAL PHASES – MACROSCOPIC SCALE

Figure SIII-1 (a) and (b). Leaf blade length-to-width ratio values (x100) are shown as a function of the leaf rank for (a) Col-0 and (b) SALK\_055458. Trends within each phase are shown in color. Plain lines: most probable phase (i.e. majority of the leaves at the rank have been assigned to the phase). Dotted lines: alternative phase (with respect to the most probable phase); shown if probability > 0.04.

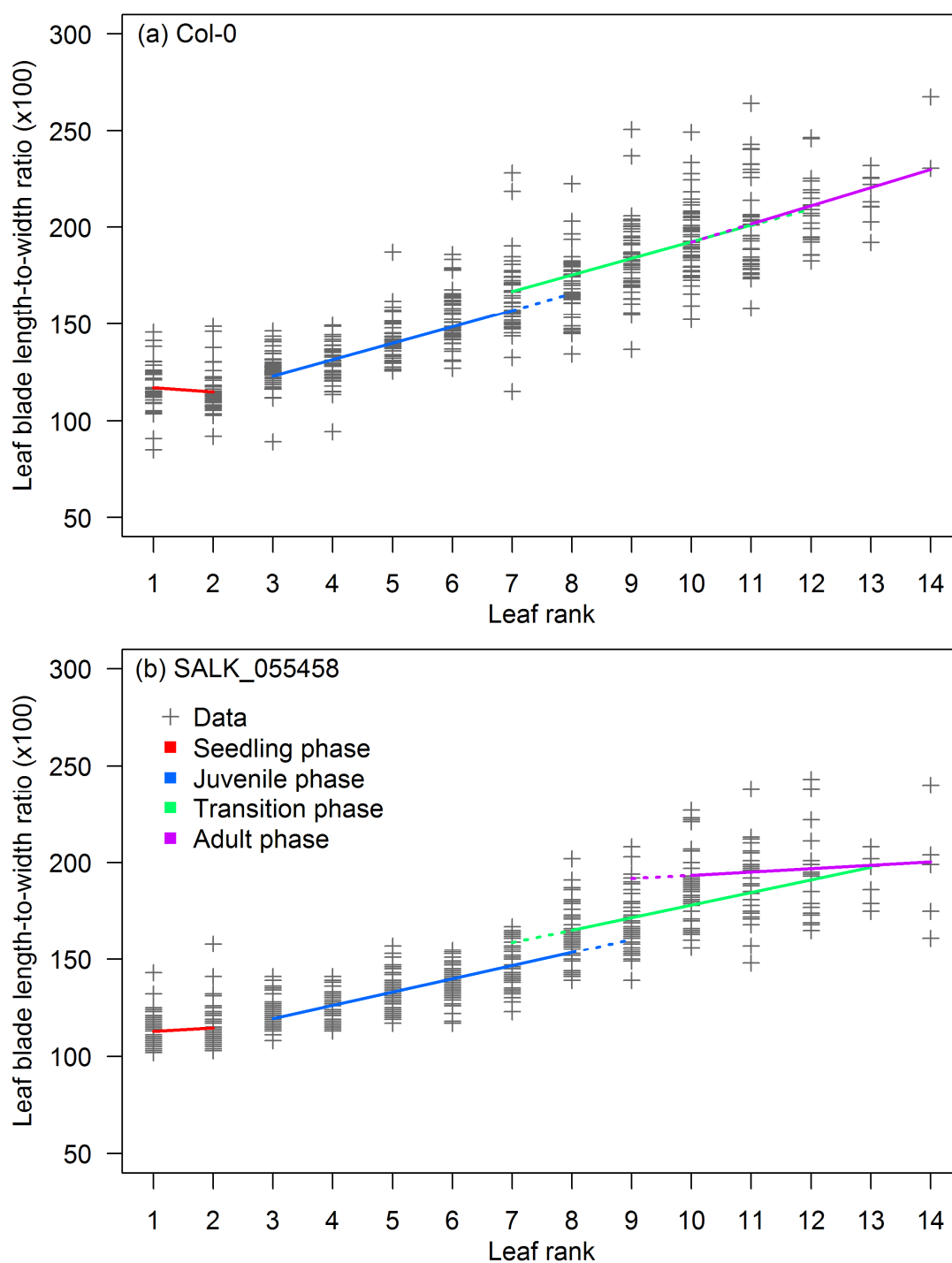


Figure SIII-1 (c) and (d). Leaf blade length-to-width values (x100) are shown as a function of the leaf rank for (c) SALK\_048174 and (d) SALK\_126071.

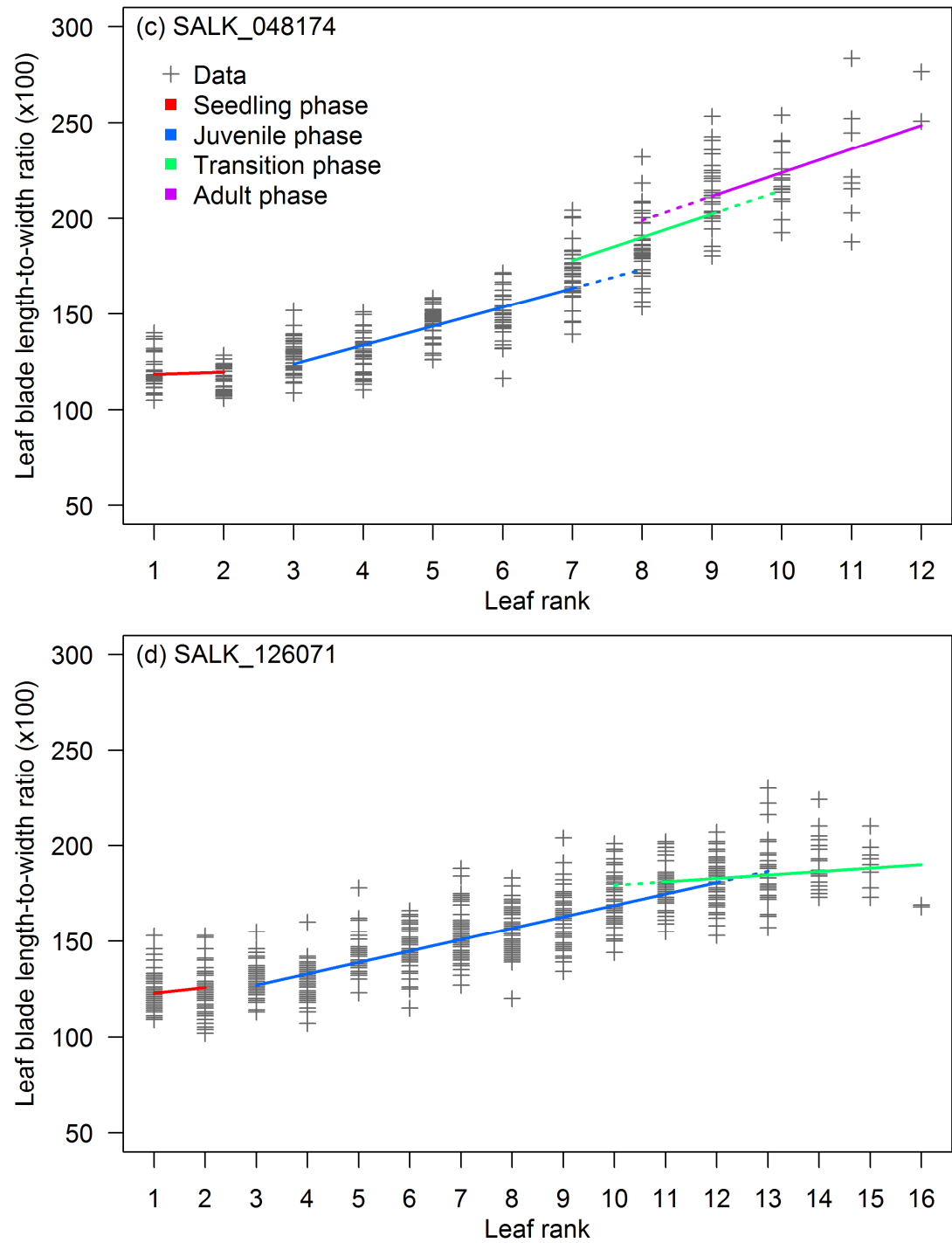


Figure SIII-2 (a) and (b). Final leaf area values ( $\text{mm}^2$ ) are shown as a function of the leaf rank for (a) Col-0 and (b) SALK\_055458. Trends within each phase are shown in color. Plain lines: most probable phase (i.e. majority of the leaves at the rank have been assigned to the phase). Dotted lines: alternative phase (with respect to the most probable phase); shown if probability  $> 0.04$ .

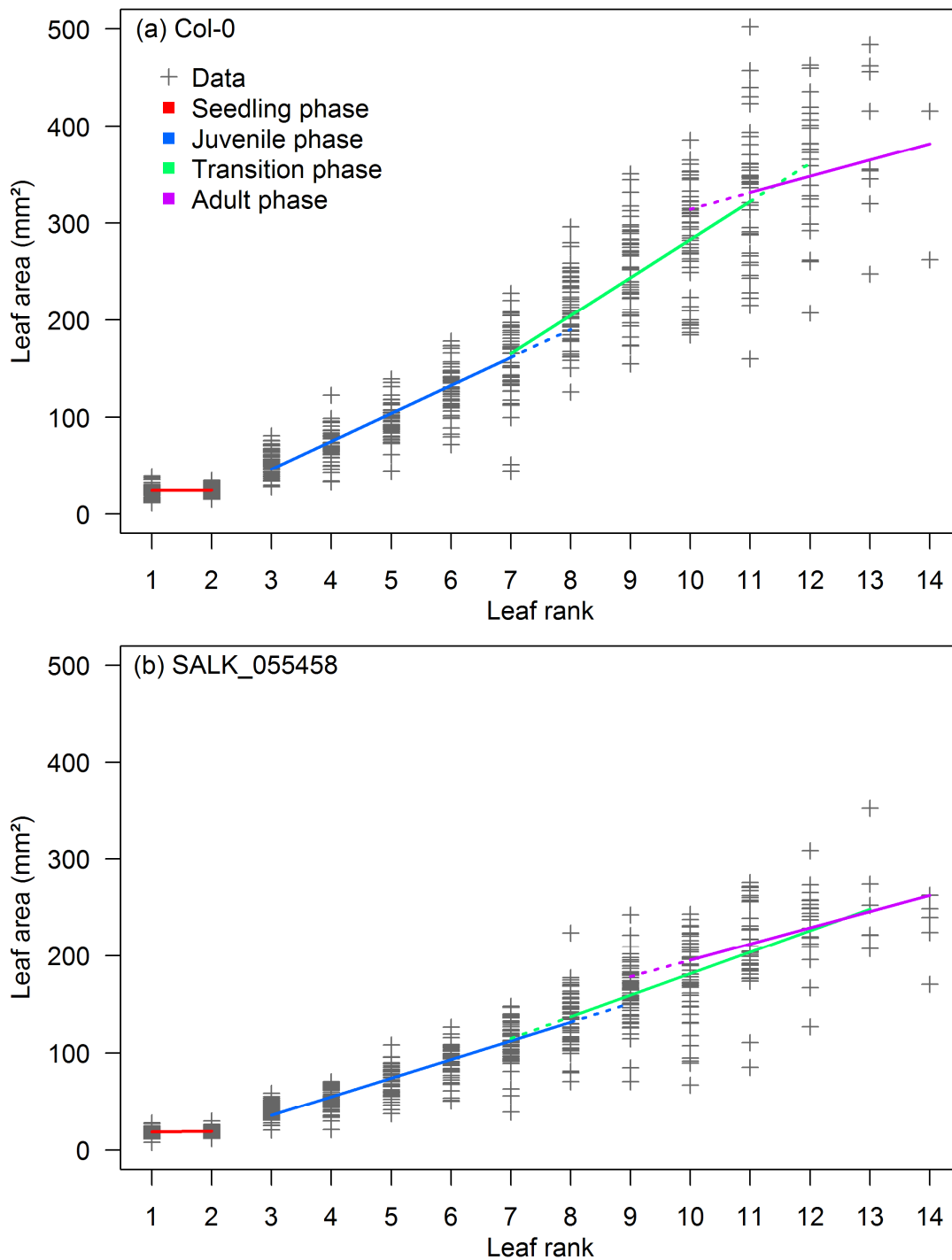




Figure SIII-2 (c) and (d). Final leaf area values (mm<sup>2</sup>) are shown as a function of the leaf rank for (c) SALK\_048174 and (d) SALK\_126071

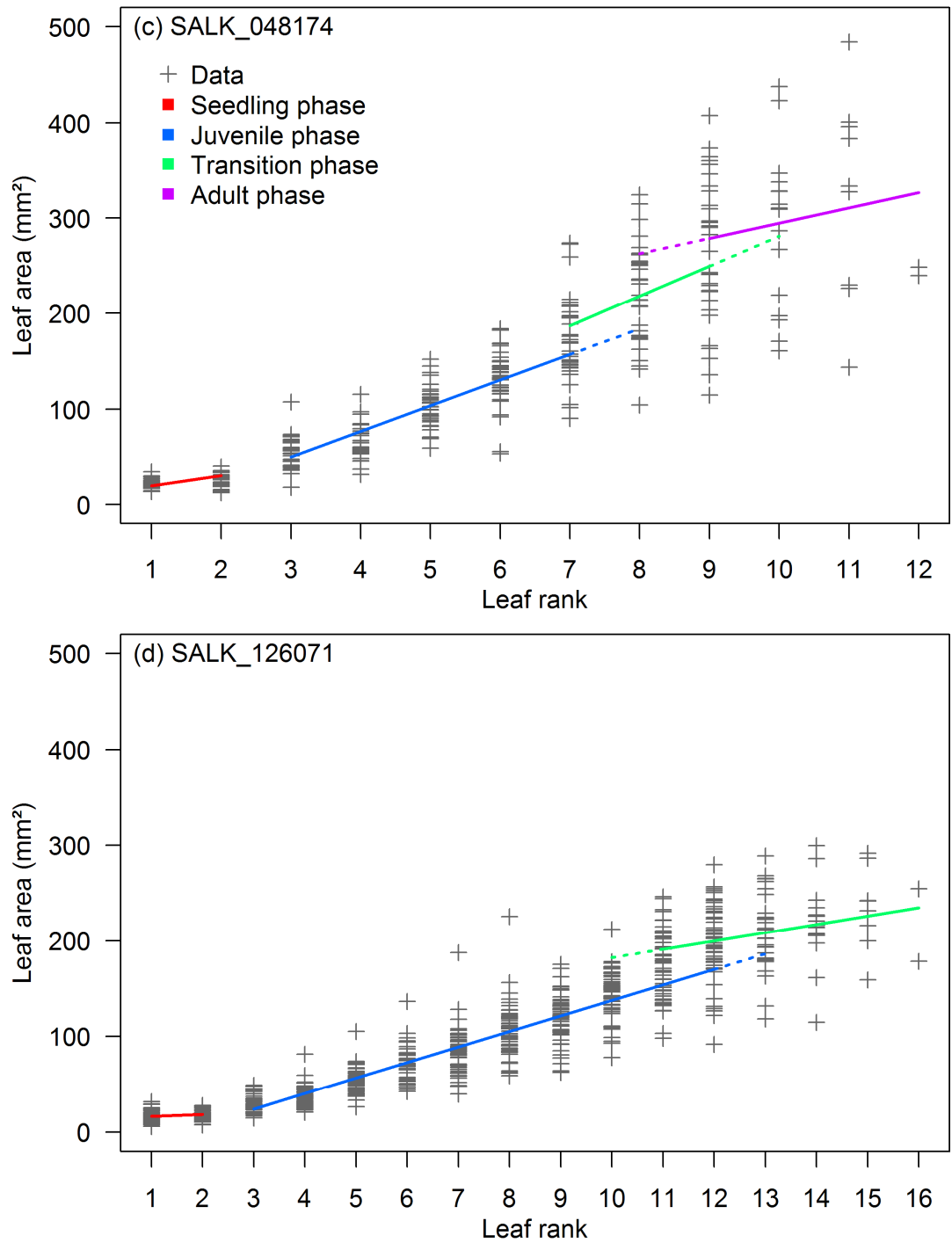


Figure SIII-3 (a) and (b). Characteristic growth duration values (hours) are shown as a function of the leaf rank for (a) Col-0 and (b) SALK\_055458. Trends within each phase are shown in color. Plain lines: most probable phase (i.e. majority of the leaves at the rank have been assigned to the phase). Dotted lines: alternative phase (with respect to the most probable phase); shown if probability > 0.04.

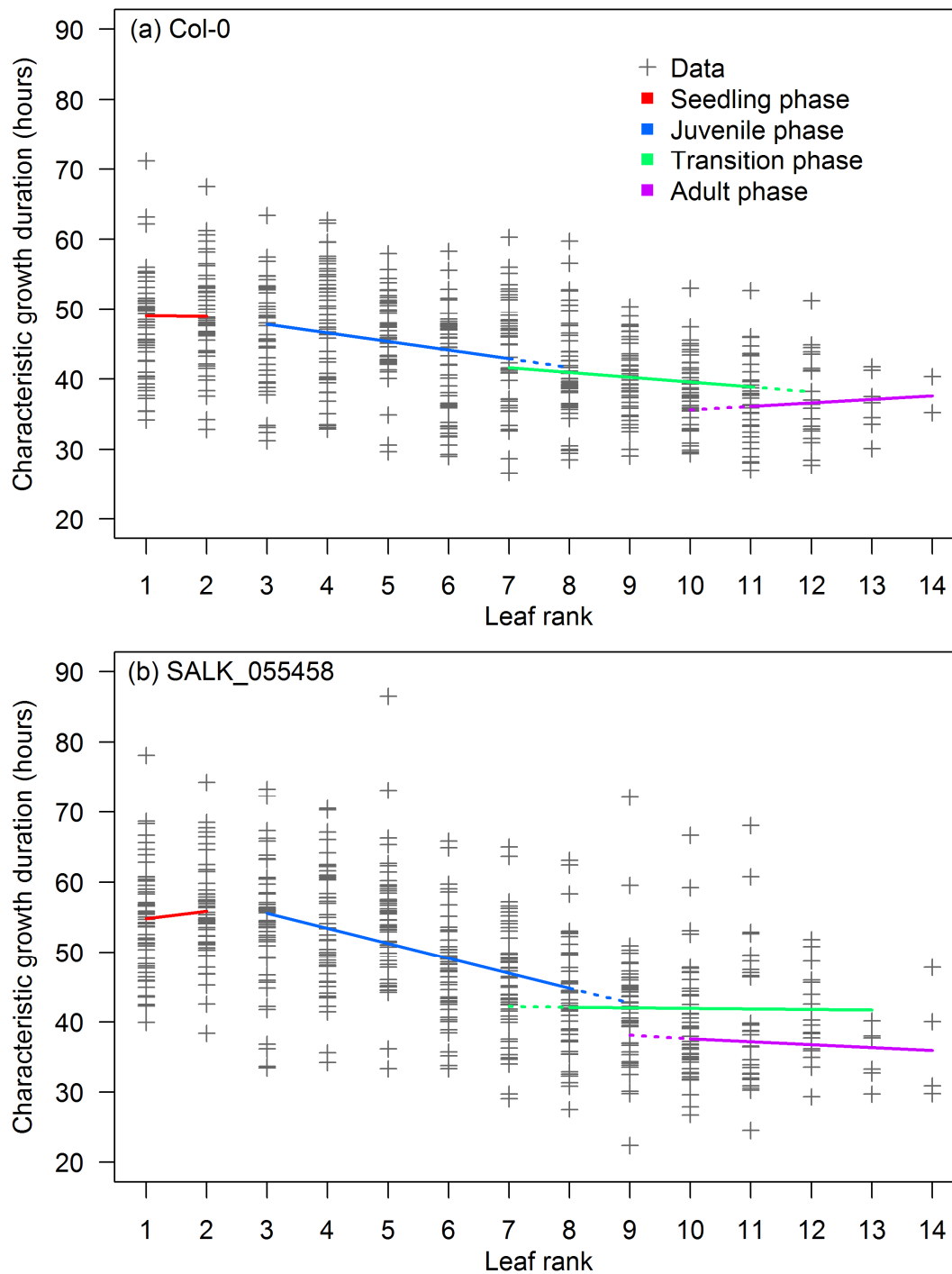
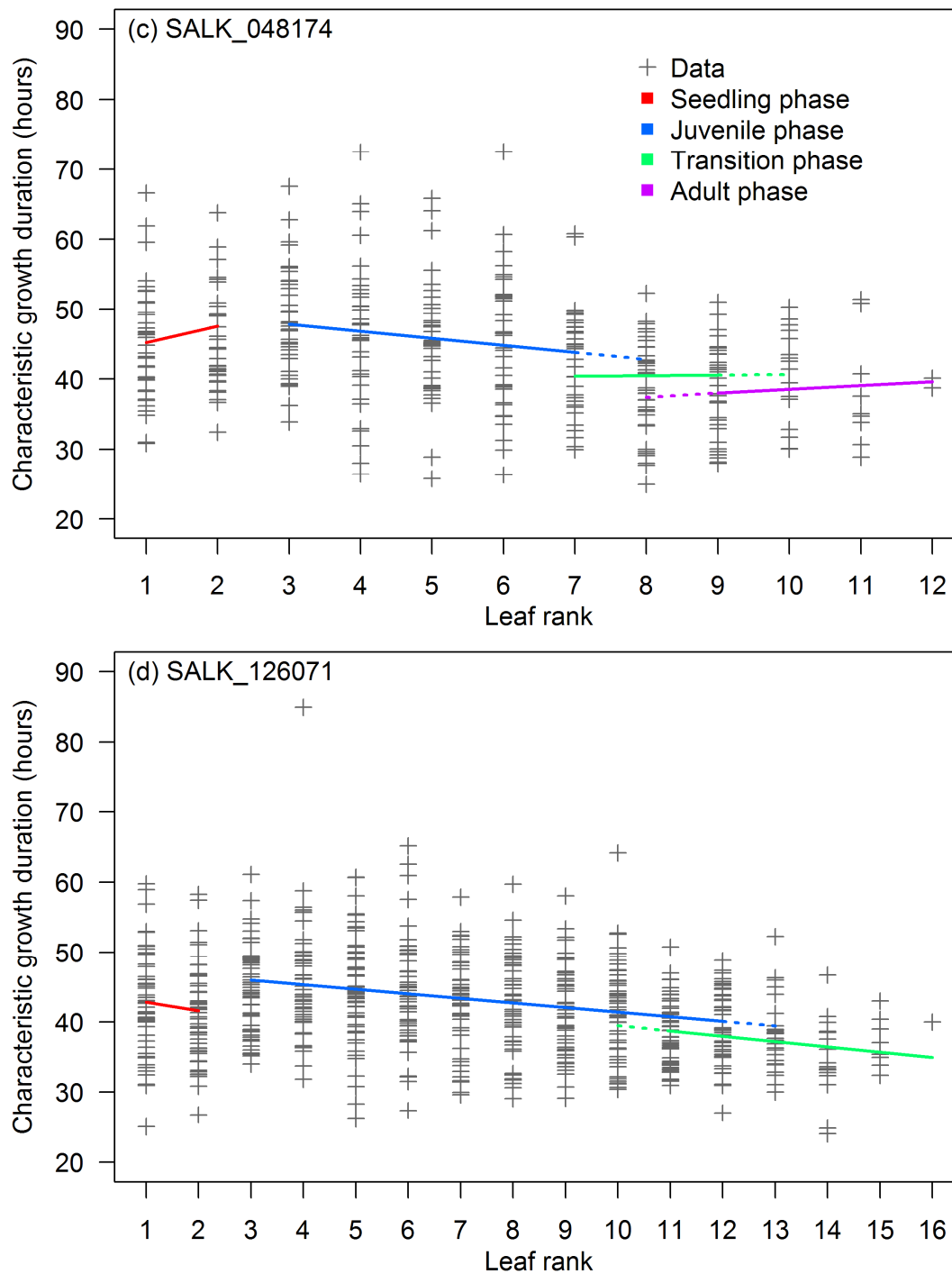


Figure SIII-3 (c) and (d). Characteristic growth duration values (hours) are shown as a function of the leaf rank for (c) SALK\_048174 and (d) SALK\_126071



## APPENDIX IV. COMPARISON BETWEEN MANUAL AND SEMI-AUTOMATIC IMAGE SEGMENTATION METHODS

In order to test the validity of our protocol, we compared a set of 46 images that were segmented both manually (using ImageJ) and semi-automatically (using the previously described protocol); see Figure SV-4. To compare the number of cells found by both segmentations in each image, we computed the distribution of the difference between the number of cells in the semi-automatic segmentation and the number of cells in the manual one, normalized by the number of cells in the semi-automatic segmentation (Figure SV-5).

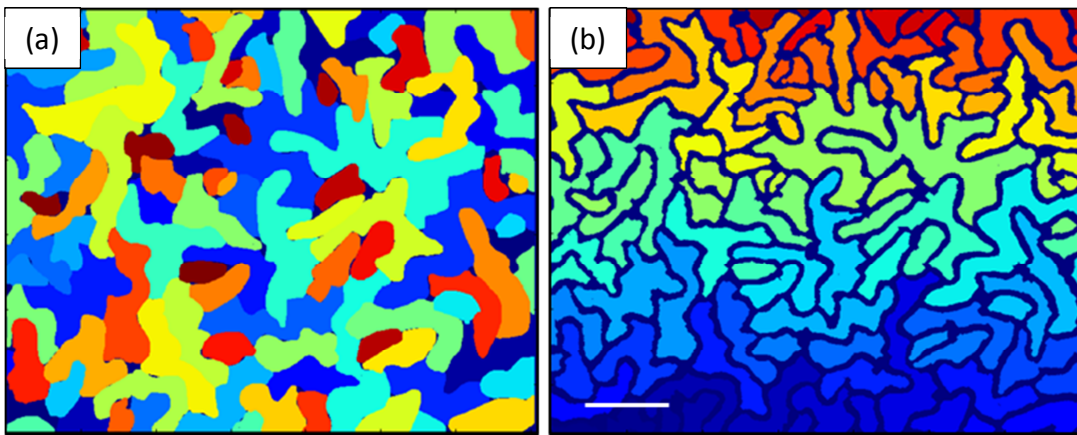


Figure SV-4. Comparison of the two segmentation methods: (a) manual segmentation using ImageJ; (b) semi-automatic segmentation using Ilastik. Scale bar: 100  $\mu\text{m}$ .

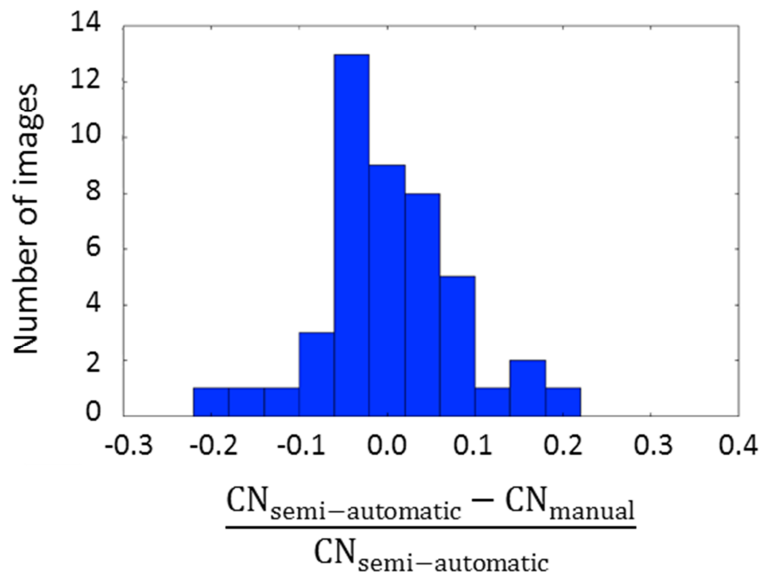


Figure SV-5. Distribution of the normalized differences between numbers of cells identified by manual and semi-automatic segmentations for a given image. CN: number of cells.

## APPENDIX IV

The number of cells tends to be similar between the two segmentation methods (+/- 10 % for 50 cells on average in each image). Individual comparisons by several observers (different perspectives) highlighted that the cases where the semi-automatic method lead to a higher number of identified cells than the manual method were mainly due to cell merging during the manual segmentation.

To quantify the cell area similarity between the two segmentation methods, the overlap between manually and semi-automatically segmented cells was computed. The metric used to quantify this overlap was the dice score ( $D_{C_i, C_j}$ ), i.e. the ratio of twice the area ( $A$ ) of the intersection of the two cells to the sum of the areas of these two cells.

$$D_{C_i, C_j} = \frac{2A(C_i \cap C_j)}{A(C_i) + A(C_j)}$$

where  $C_i$  corresponds to the ImageJ segmentation and  $C_j$  to the ilastik segmentation.

This score is between 0 (no overlap) and 1 (full overlap) (Figure SV-6). The boxplots traced for each pair of segmented images show that the scores are globally comprised between 0.6 and 0.9; see Figure SV-8. It should be noted that cell walls and stomates were not considered as being a part of cells in the semi-automatic segmentation while they were in the manual segmentation; see Figure SV-4 and Figure SV-6 (c). This explains the systematic deviation from 1 of the dice scores. The lowest dice scores reflect either cell merging or splitting (Figure SV-6 (a) and (b)).

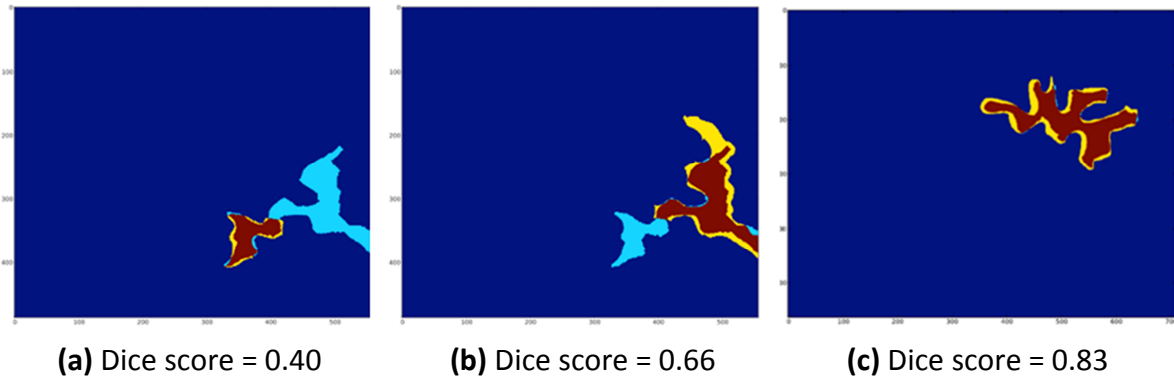


Figure SV-6. The dice score quantifies the overlap between cells found using the manual and semi-automatic segmentation for a given leaf epidermis imprint image. Yellow: cell defined by manual segmentation. Blue: cell defined by semi-automatic segmentation. Brown: overlap. (a) and (b): the manual segmentation identified two separated cells where the semi-automatic segmentation identified a single cell. In (b), the upper yellow area corresponds to a portion of the image that was identified as a separated cell using the semi-automatic method. (c) Good match between the two segmentation methods.

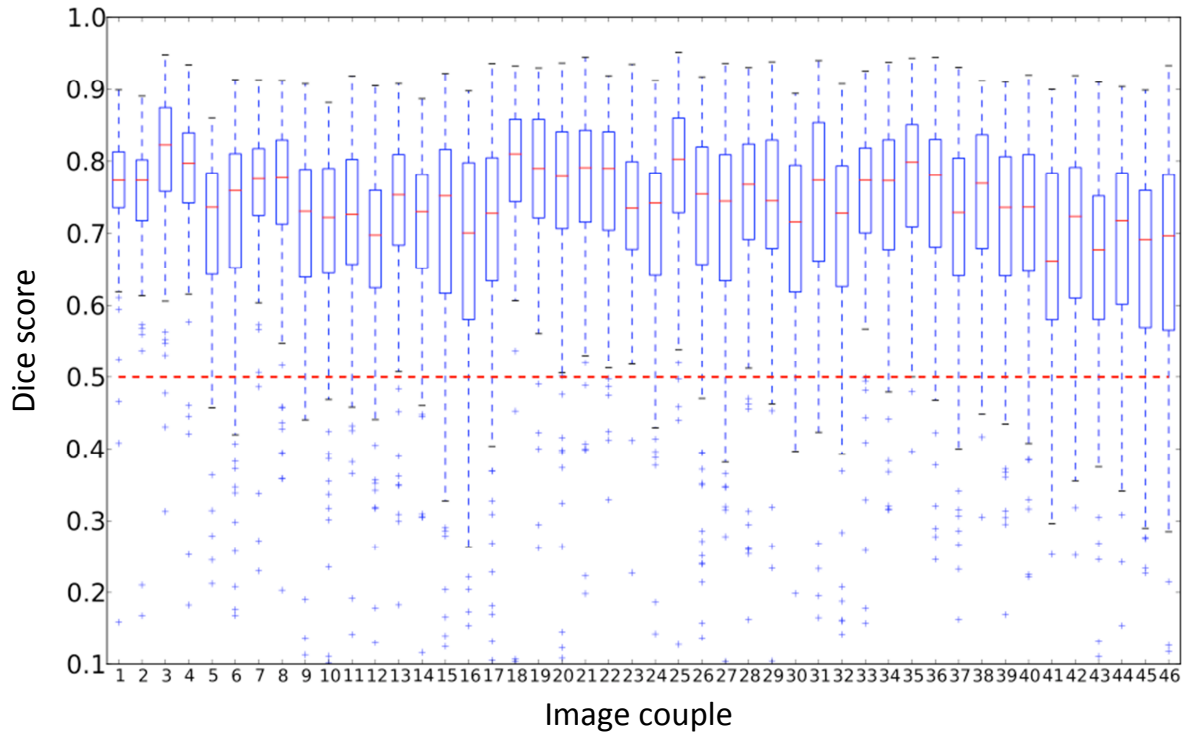


Figure SV-8. Boxplot representing the distribution of dice scores for different images segmented manually using ImageJ and semi-automatically using our protocol.

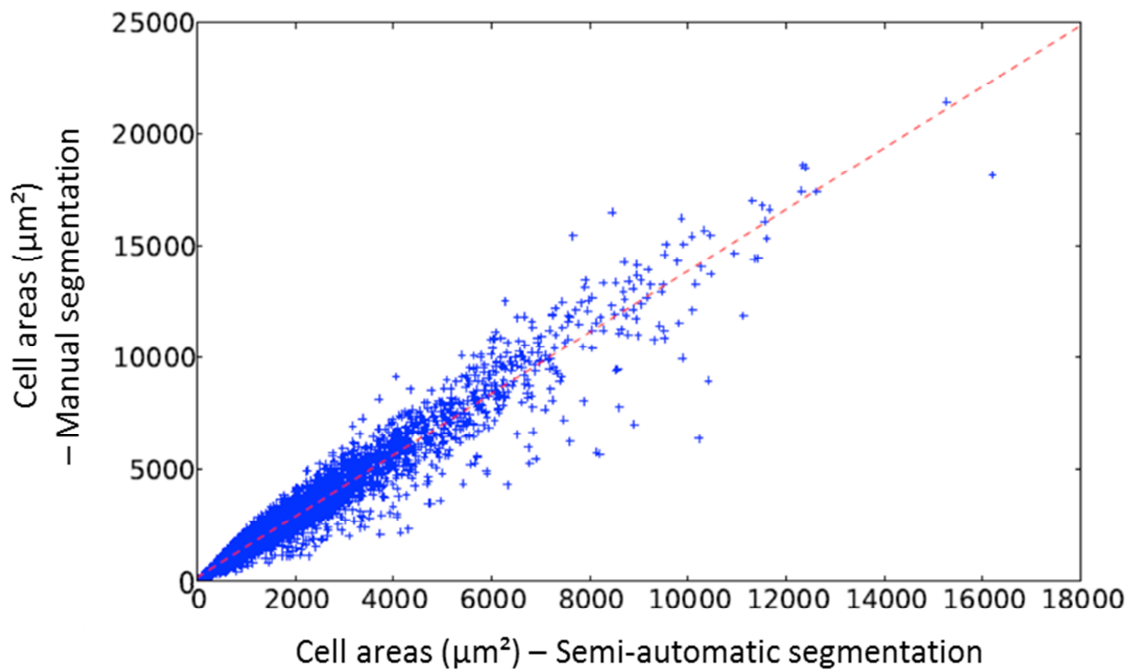


Figure SV-7. Linear relationship between cell areas obtained using the manual and the semi-automatic segmentation methods. Each point represents an image of Col-0 leaf adaxial epidermis imprint ( $n = 6007$ ). The dotted line represents the linear relationship estimated between the two variables. Coefficient of determination  $r^2 = 0.97$ ;  $p$ -value = 0.

## APPENDIX IV

---

The dice score was used to map cells segmented using the two methods by pairing cells with highest pairwise dice score. From this mapping we were able to compare the areas of paired cells. As shown in Figure SV-7, cell areas extracted from the two segmentation methods were strongly correlated ( $r^2 = 0.94$ , p-value = 0), indicating that the methods tend to agree for the estimation of the area. The fact that the slope was different from 1 is due to the inclusion of cell walls and stomates into the cell in the manual method, as explained hereinabove.

The comparison of the two segmentation methods allowed us to validate the semi-automatic segmentation of images. The cells identified using our semi-automatic segmentation protocol were close to the ones segmented manually by an observer. In fact, it can even limit human bias, as we observed that when different perspectives were taken into account, the manual segmentations presented more errors than the ones obtained semi-automatically. In conclusion, our semi-automatic segmentation protocol of leaf epidermis imprints images appears to be reliable, less subject to errors due to human bias and less time consuming than manual segmentation.

## APPENDIX V. SUPPLEMENTARY FIGURES FOR INTEGRATIVE MODELS FOR ANALYZING *A. THALIANA* ROSETTE DEVELOPMENTAL PHASES

Figure SV-1 (a) and (b). (a) Leaf blade length-to-width ratio values ( $\times 100$ ) and (b) final leaf area ( $\text{mm}^2$ ) are shown as a function of the leaf rank for Col-0 dataset incorporating the estimated endoreduplication factor. Trends within each phase are shown in color. Plain lines: most probable phase (i.e. majority of the leaves at the rank have been assigned to the phase). Dotted lines: alternative phase (with respect to the most probable phase); shown if probability  $> 0.04$ .

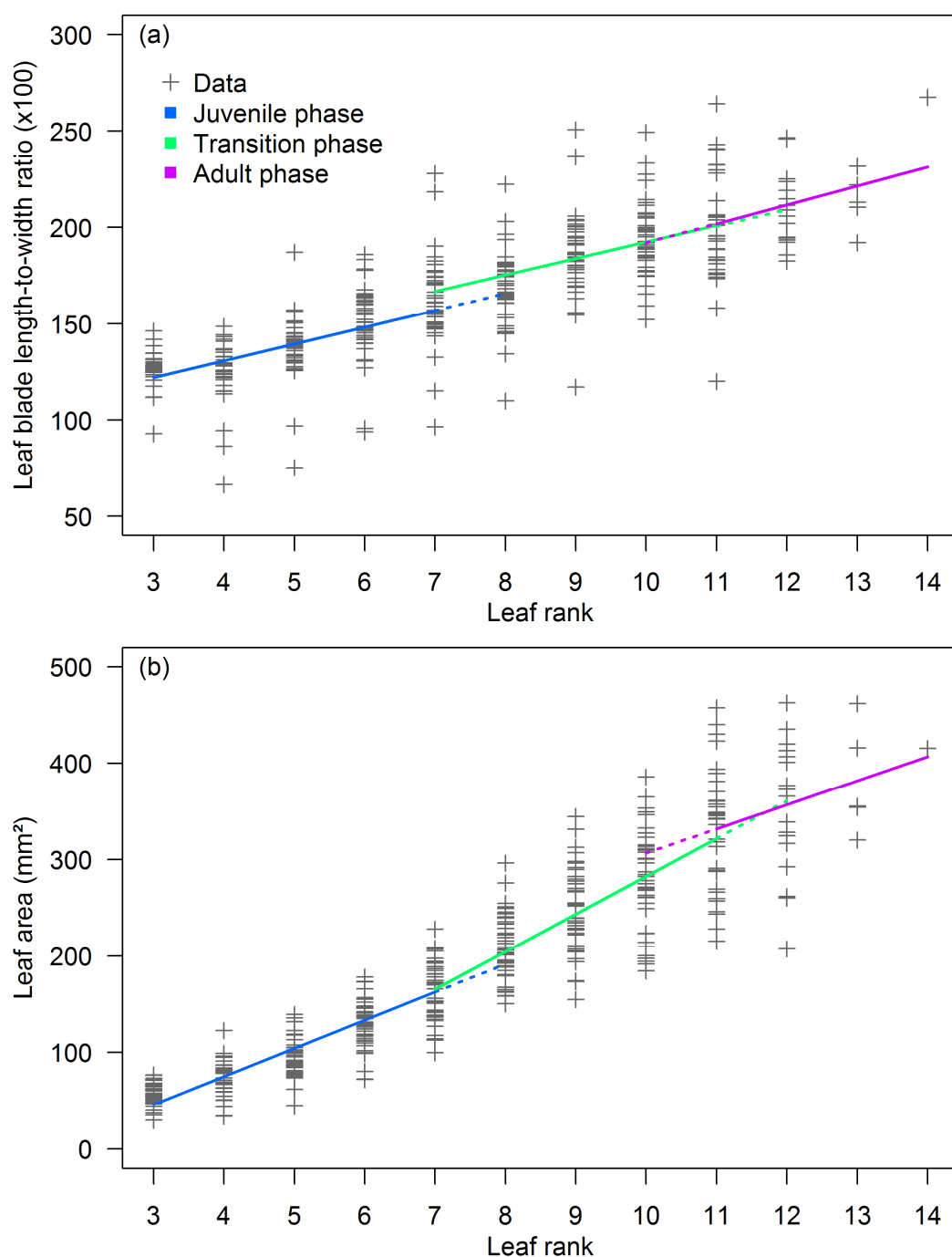
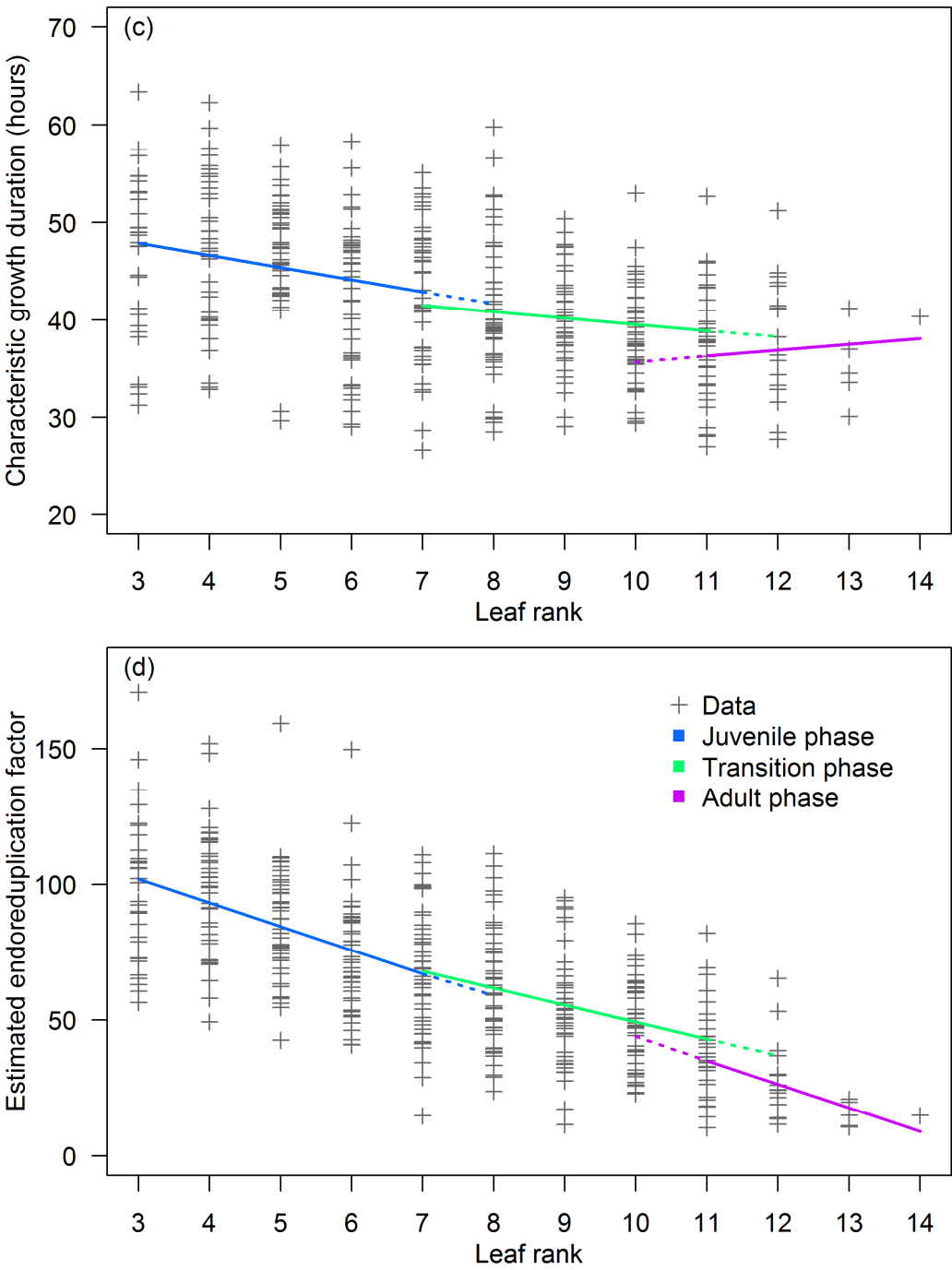




Figure SV-1 (c) and (d). (c) Characteristic growth duration (hours) and (d) estimated endoreduplication factor are shown as a function of the leaf rank for Col-0 dataset incorporating the estimated endoreduplication factor. Trends within each phase are shown in color. Plain lines: most probable phase (i.e. majority of the leaves at the rank have been assigned to the phase). Dotted lines: alternative phase (with respect to the most probable phase); shown if probability > 0.04.



## APPENDIX VI. PRELIMINARY ANALYSES ON MUTANT LINES INVESTIGATING THE CROSSTALK BETWEEN MIRNA-DEPENDENT PATHWAYS AND SHOOT DEVELOPMENT IN *A. THALIANA*

Eight *A. thaliana* mutant lines and their wild-type Columbia 0 were grown together in the same conditions than those applied in Experiment 3 presented in the Chapter III (n = 12). The different leaf traits were computed using the same protocols than the ones described in the Chapter III. Seeds are kind gift from Patrick Laufs unit (Jean-Pierre Bourgin Institut – INRA, Versailles, France).

| Line designation       | Description  |
|------------------------|--|
| Col-0                  | Wild-type Columbia 0   |
| <i>CUC2g-m4</i>        | Line with higher <i>CUP-SHAPED COTYLEDON 2 (CUC2)</i> expression levels as a result of defective <i>miR164</i> -dependent regulation |
| <i>rSPL10</i>          | Line expressing a miR156-resistant of <i>SPL10</i> gene  |
| <i>rSPL10.CUC2g-m4</i> | Line containing both <i>rSPL10</i> and <i>CUC2g-m4</i> mutations   |
| <i>rSPL9</i>           | Line expressing a miR156-resistant of <i>SPL9</i> gene   |
| <i>MIM156</i>          | Line perturbing endogenous miR156 function by overexpressing miR156 mimicry targets ( <i>MIM156</i> )                                |
| <i>kluh (klu)</i>      | Loss-of-function mutant affected in <i>KLUH/CYP78A5</i> gene encoding the cytochrome P450 CYP78A5 monooxygenase                      |
| <i>ago1-27</i>         | Hypomorphic <i>ARGONAUTE 1 (AGO1)</i> mutant   |
| <i>se-1</i>            | Line expressing a weak <i>SERRATE (SE)</i> allele  |

Table SVI-1. List of the mutant lines grown in the fourth experiment mentioned in Chapter V ‘Conclusions and Perspectives’.

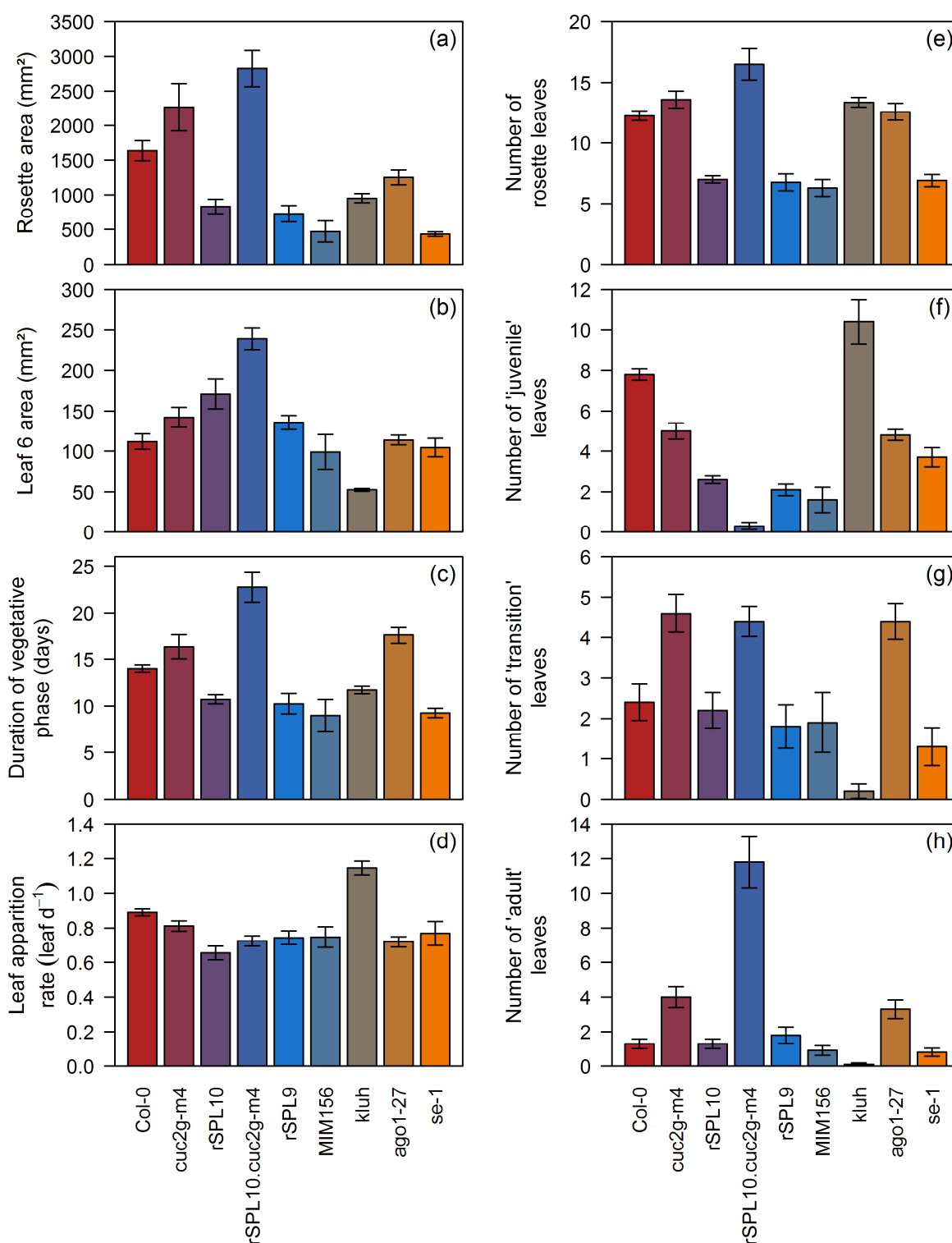


Figure SVI-1. Dissected rosette area (mm<sup>2</sup>; (a)), leaf 6 area (mm<sup>2</sup>; (b)), duration of vegetative phase (days after leaf 2 emergence; (c)), mean leaf apparition rate (leaf day<sup>-1</sup>; (d)), number of rosette leaves (e) and number of leaves without abaxial trichomes ('juvenile' leaves, (f)), with abaxial face partially ('transition' leaves, (g)) and fully ('adult' leaves, (h)) covered by trichomes are given for lines grown during fourth experiment mentioned in Chapter V 'Conclusions and Perspectives'. Means with standard errors are shown for each genotype (between 6 and 12 individual per line).

## Detailed phenotypic analyses of leaf growth in SALK T-DNA lines : A source of variability to reveal new genes controlling shoot development



Maryline Lièvre<sup>1</sup>, Mathilde Bettembourg<sup>1</sup>, Myriam Dauzat<sup>1</sup>, David Wilson-Sanchez<sup>2</sup>,  
José Manuel Pérez-Pérez<sup>2</sup>, María Rosa Ponce<sup>2</sup>, Jose Luis Micol<sup>2</sup>, Christine Granier<sup>1</sup>

<sup>1</sup> LEPSE, UMR 759, INRA-SUPAGRO, France; <sup>2</sup> Universidad Michel Hernández, Spain  
email : lievre@supagro.inra.fr



### Context

Leaf growth control is only partly understood and the circuitry that links the different levels of organization from molecules to molecular networks, cells, tissue, leaf and whole plant remains to be uncovered. Many genes controlling leaf growth have not been characterized as such and have still to be identified.

In this context, we started a large-scale genotype-to-phenotype association in the sequence-indexed Arabidopsis insertion mutant collection [developed by Alonso et al., 2003 Science].

### Plant Material

A first leaf phenotype-based screen was performed on **16400 T-DNA Arabidopsis mutant lines from the SALK institute collection** [Alonso et al., 2003 Science]. From this analysis, **255 lines** were identified with an unequivocal leaf growth phenotype and as homozygous for the T-DNA. Then, **91** were selected for a more in depth phenotypic analysis.

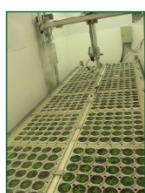
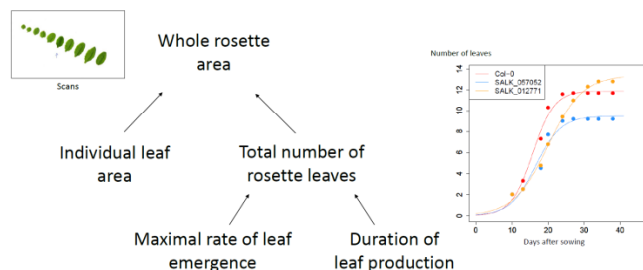
### Experimental Design



**First screen** (Miguel Hernández University of Elche): plants were grown in vitro. Leaf phenotype was scored 18 days after sowing (DAS). Criteria to identify lines with un-ambiguous leaf phenotypes were : bigger/smaller rosettes than the wild-type, different leaf color or different leaf shape (n=30).

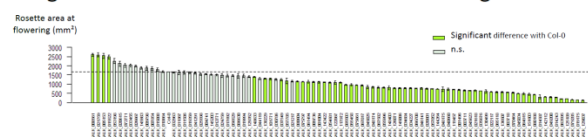
**Second screen** (LEPSE, INRA Montpellier): plants were grown in soil in the PHENOPSIS phenotyping platform [Granier et al., 2006 New Phyt]. Their leaf growth phenotype was scored both at 18 DAS and at flowering (n = 6 to 10).

Quantitative measurements included :

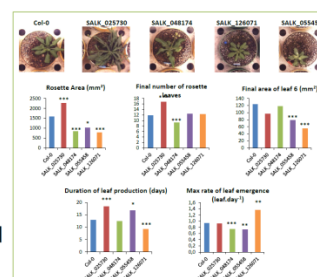


### Results

- ❖ Final rosette area, measured at flowering, significantly differ from that of the wild-type (Col-0) in 63 lines. It was larger in 4 of them and smaller in the remaining 59.



- ❖ Many lines were affected on different growth variables, exhibiting sometimes a compensation phenomenon such as between the rate and duration of leaf production, or the number of leaves and the individual leaf area.

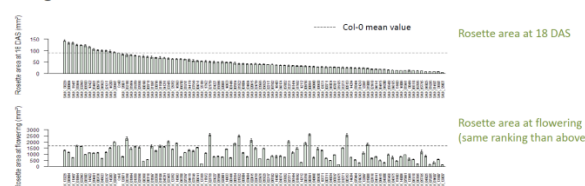


- ❖ Mutated genes of lines identified as different from Col-0 had never been described as involved in shoot development.

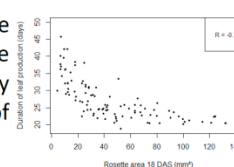
- ❖ Phenotypes were globally well-conserved between the two growth conditions when recorded at 18 DAS.

- ❖ However, they were not conserved over time.

Plants that were bigger than the wild-type at early stages of development could finally have the same size or even smaller rosette areas at the end of their growth.



- ❖ Some of these differences can be explained by the negative correlation between early development and the duration of leaf production phase.



Plants with large rosettes at 18 DAS flowered earlier than others.

### Conclusions

Interestingly, only a very small proportion of the genes identified in our screen had been previously described as required for leaf development suggesting that the SALK collection has a strong potential for identifying new genes controlling growth and improving actual molecular networks. However, results issued from rapid high-throughput screens have to be interpreted with caution because of compensation phenomena during the whole growing period. Our study highlights that it is important to take growth dynamics into account in plant growth phenotyping to avoid misinterpretations.

This work was supported by the Agron-Omics European Project



# Reverse genetics of leaf development

José Luis Micol, David Wilson-Sánchez, Sara Jover-Gil,  
Maryline Lièvre, Christine Granier,  
José Manuel Pérez-Pérez, and María Rosa Ponce.

Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Alicante, Spain.  
jlmicol@umh.es dwilson@umh.es www.agron-omics.eu genetica.umh.es

The interest of the study and eventual manipulation of leaf development lays on the fact that they are the fundamental photosynthetic organ, of vital importance for life in our planet. Although hundreds of mutants with altered leaf development have been isolated using forward genetics, saturation has not yet been reached for the Arabidopsis genome<sup>1</sup>.

The group of Prof. J.R. Ecker, at The Salk Institute, is obtaining a large collection of gene-indexed homozygous T-DNA insertion mutants<sup>2,3</sup>. Aiming to identify novel genes required for leaf growth, we are screening 20,718 of these lines, which correspond to 14,585 genes. So far, we have analyzed the leaf phenotype of 16,400 lines, and identified 518 that show a mutant phenotype with full penetrance and almost constant expressivity (Figure 1 and Table 1). We have genotyped 340 of these genuine leaf mutants, finding that a T-DNA insertion is homozygous at the annotated locus in 252 of them (Figures 2 and 3). Among these 252 genes interrupted, only 40 have previously characterized alleles<sup>4</sup> and, according to TAIR functional annotation, there is no information available on the function of 135 of them. These figures highlight the value of our collection.

We have implemented a public database and a web-based query application that collects the results of our screen and the preliminary characterization of the mutant phenotypes (Figure 4). It will be soon available at <http://agron-omics.umh.es>.

120 genuine leaf mutants with a confirmed homozygous T-DNA insertion have been subjected to detailed time-lapse qualitative and quantitative phenotypic analyses (see poster by Maryline Lièvre *et al.*).

Table 1.- Progress of our screening

|             | T <sub>4</sub> lines |                           | T <sub>5</sub> genuine mutants confirmed |
|-------------|----------------------|---------------------------|--|
|             | studied              | exhibiting anormal leaves |  |
| Completed   | 16,400               | 2,567                     | 518                                      |
| In progress | 2,500                | 450                       | 79*                                      |
| Total       | 18,900               | 3,017                     | 597*                                     |

\*Expected values

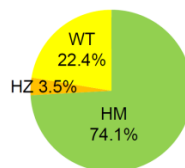


Figure 2.- Genotyping of genuine leaf mutants. Almost three quarters of all lines genotyped are homozygous for the annotated insertion, 3.5% harbour a segregating insertion, and in the remaining 22.4% the insertion is absent from the annotated locus.



Figure 1.- Rosettes of (A) the wild type Col-0 and (B-X) 23 of the leaf mutants identified in our screening. Pictures were taken 21 days after stratification (das). Scale bars: 1 mm.

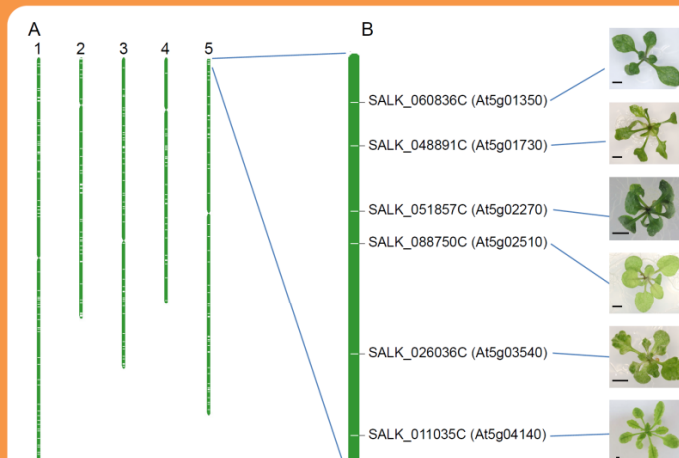
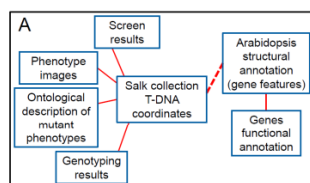


Figure 3.- We performed a phenotype-driven screen on a gene-indexed mutant collection. (A) Position of the 252 homozygous T-DNA insertions confirmed in our laboratory so far, laid out using the TAIR map visualization tool (<http://www.arabidopsis.org/servlets/ViewChromosomes>). Numbers on the top of each bar identify the 5 Arabidopsis chromosomes. (B) Enlargement of a randomly chosen interval showing 6 loci disrupted by T-DNA insertions, with their corresponding Salk codes and rosette phenotypes observed. Scale bars: 2 mm.



GENE QUERY

1 AGI codes 2 At3g52765 At1g55020 At2g27650 At3g47440 At1g74040 At5g21326

3 Only if "Gene keywords" selected: Gene name Gene description GO mol. func. GO cel. comp. GO biol. func.

4 Display 8 records per page

Submit Clear form

PHENOTYPE QUERY

Browse all mutant phenotypes

Search by ontology terms

Only with confirmed annotated T-DNA

Limit results to a physical interval: Chromosome sel. From position (Mb) to position (Mb)

Sort records by Salk code

Submit Reset form

ROSETTE COMPACTNESS Compact Loose

LAMINA COLOR PATTERN Variegated Spotted Blotched Netted

Figure 4.- (A) Simplified database schema. It contains phenotypic information of all T<sub>4</sub> and T<sub>5</sub> lines screened, plate and rosette pictures, and the genotyping results obtained for the lines confirmed as leaf mutants. It also includes the position of the annotated insertions, provided by The Salk Institute Genomic Analysis Laboratory, and the Arabidopsis genome structural and functional annotation, available from TAIR. (B, C) Query page of the database. The user is prompted to choose between a gene or a phenotype query. If the former option is selected (B), three different input types can be chosen: AGI codes, Salk codes, and gene keywords. Other options are also available to refine or format results. If the latter type of query is selected (C), there are several ways to customize it: confirmed mutants can either be browsed or searched by their phenotypes using ontology terms (right hand side of figure C); query can be filtered for mutants with confirmed annotated T-DNA only; finally, it is possible to retrieve only mutants within a physical genomic interval. (D) Example of the results that can be obtained for a given line.

Arabidopsis leaf mutant DB

SALK\_023198C

SCREEN RESULT: Salk code: SALK\_023198C, Position: 1, 377011, UTR: 5' TGA TGA TGA, 3' TGA TGA TGA

GENE INFO: Gene name: SALK\_023198C, Gene description: SALK\_023198C, Gene keywords: SALK\_023198C

ONTOLOGY DESCRIPTION (see additional information):

ENTRY: SALK\_023198C

ATTRIBUTE: Rosette

VALUE: Rosette

Leaf shape: Rosette

Leaf margin: Rosette

Other traits (see ont): Any

## ACKNOWLEDGEMENTS

Research in the laboratory of J.L.M. is supported by grants from the Ministerio de Ciencia e Innovación of Spain [BFU2011-22825 and CSD2007-00057 (TRANSPALANTA)], the Generalitat Valenciana (PROMETEO/2009/112) and the European Commission [LSHG-CT-2006-037704 (AGRON-OMICS)]. D.W.-S. is a predoctoral fellow of the Generalitat Valenciana VALI+d programme.

## REFERENCES

- Berná, G., Robles, P., and Micol, J.L. (1999). *Genetics* **152**, 729-742.
- Alonso, J.M., *et al.* (2003). *Science* **301**, 653-657.
- O'Malley, R.C., and Ecker, J.R. (2010). *Plant J.* **61**, 928-940.
- Lloyd, J. and Meinke, D. (2012). *Plant Physiol.* **158**, 1115-1129.



**Analyse multi-échelles et modélisation de la croissance foliaire chez *Arabidopsis thaliana***  
**– Mise au point et test d'un pipeline d'analyses permettant une analyse intégrée du développement de la cellule à la pousse entière.**

Ce travail est basé sur le constat du manque de méthodes permettant l'analyse intégrée des processus contrôlant le développement végétatif d'*Arabidopsis thaliana* dans les études phénotypiques multi-échelles. Un phénotypage préliminaire de la croissance foliaire de 91 génotypes a permis de sélectionner 3 mutants et des variables d'intérêt pour une étude plus poussée du développement de la pousse. Un pipeline de méthodes d'analyses combinant techniques d'analyse d'images et modèles statistiques a été développé pour intégrer les mesures faites à l'échelle de la feuille et de la pousse. Des modèles multi-phasiques à changements de régime semi-markovien ont été estimés pour chaque génotype permettant une caractérisation plus pertinente des mutants. Ces modèles ont validé l'hypothèse selon laquelle le développement de la rosette peut être découpé en une suite de phases de développement, pouvant varier selon les génotypes. Ils ont aussi mis en évidence le rôle structurant de la variable «trichome abaxial», bien que les phases de développement ne puissent être entièrement expliquées par ce trait. Un 2<sup>nd</sup> pipeline d'analyses combinant une méthode semi-automatique de segmentation d'images de l'épiderme foliaire et l'analyse des surfaces de cellules par un modèle de mélange de lois gamma à paramètres liés par une loi d'échelle a été développé. Ce modèle nous a permis d'estimer la loi du nombre de cycles d'endoréduplication. Nous avons mis en évidence que cette loi dépendait du rang de la feuille. Le cadre d'analyses multi-échelles développé et testé durant cette thèse devrait être assez générique pour être appliqué à d'autres espèces végétales dans diverses conditions environnementales.

---

**Multi-scale analysis and modeling of shoot growth in *Arabidopsis thaliana* –  
Development and testing of a pipeline of analysis methods enabling an integrative  
analysis of the development from cell to shoot scale.**

This study is based on the observation of a lack of methods enabling the integrated analysis of the processes controlling the vegetative development in *Arabidopsis thaliana* during multi-scale phenotypic studies. A preliminary leaf growth phenotyping of 91 genotypes enabled to select 3 mutants and different variables of interest for a more in depth analysis of the shoot development. We developed a pipeline of analysis methods combining image analysis techniques and statistical models to integrate the measurements made at the leaf and shoot scales. Semi-Markov switching models were built for each genotype, allowing a more thorough characterization of the studied mutants. These models validated the hypothesis that the rosette can be structured into successive developmental phases that could change depending on the genotype. They also highlighted the structuring role of the 'abaxial trichomes' variable, although the developmental phases cannot be explained entirely by this trait. We developed a second pipeline of analysis methods combining a semi-automatic method for segmenting leaf epidermis images, and the analysis of the obtained cell areas using a gamma mixture model whose parameters of gamma components are tied by a scaling rule. This model allowed us to estimate the mean number of endocycles. We highlighted that this mean number of endocycles was function of the leaf rank. The multi-scale pipeline of analysis methods that we developed and tested during this PhD should be sufficiently generic to be applied to other plant species in various environmental conditions.