Universität Potsdam

IMPRS Primary Metabolism and Plant Growth

AG Molekularbiologie - Prof. Dr. Bernd Müller- Röber
AG Plant Development and Stress Adaptations - Dr. Justyna Olas

Elucidating the molecular basis of enhanced growth in the *Arabidopsis thaliana* accession Bur-0

Dissertation

zur Erlangung des akademischen Grades
"doctor rerum naturalium" (Dr. rer. nat.)
in der Wissenschaftsdisziplin " Molekulare Pflanzenphysiologie"

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
Institut für Biochemie und Biologie
der Universität Potsdam

von

Catalina Moreno Curtidor

Ort und Tag der Disputation:
Potsdam, den 05.11.2021

Supervisors:

Prof. Dr. Bernd Müller-Röber (Principal supervisor)
University of Potsdam

Prof. Dr. Salma Balazadeh (Second supervisor)
Leiden University

Dr. Justyna Olas (Mentor)
University of Potsdam

Reviewers:

Prof. Dr. Bernd Müller-Röber
University of Potsdam

Prof. Dr. Michael Lenhard
University of Potsdam

Prof. Dr. Gerco Angenent
Wageningen University

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-52681
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-526814

Declaration

I hereby declare that I have written this thesis on my own and without the use of sources and aids other than those indicated accordingly in the text. This work has not been previously submitted to any examination authority in the same or a similar form.

Potsdam, 24th March 2021

Catalina Moreno Curtidor

Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne Hilfe Dritter und ohne Zuhilfenahme anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe. Die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen sind als solche kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Potsdam, den 24.03.2021

Catalina Moreno Curtidor

Table of contents

Abstract

The life cycle of flowering plants is a dynamic process that involves successful passing through several developmental phases and tremendous progress has been made to reveal cellular and molecular regulatory mechanisms underlying these phases, morphogenesis, and growth. Although several key regulators of plant growth or developmental phase transitions have been identified in Arabidopsis, little is known about factors that become active during embryogenesis, seed development and also during further postembryonic growth. Much less is known about accession-specific factors that determine plant architecture and organ size. Bur-0 has been reported as a natural *Arabidopsis thaliana* accession with exceptionally big seeds and a large rosette; its phenotype makes it an interesting candidate to study growth and developmental aspects in plants, however, the molecular basis underlying this big phenotype remains to be elucidated. Thus, the general aim of this PhD project was to investigate and unravel the molecular mechanisms underlying the big phenotype in Bur-0.

Several natural Arabidopsis accessions and late flowering mutant lines were analysed in this study, including Bur-0. Phenotypes were characterized by determining rosette size, seed size, flowering time, SAM size and growth in different photoperiods, during embryonic and postembryonic development. Our results demonstrate that Bur-0 stands out as an interesting accession with simultaneously larger rosettes, larger SAM, later flowering phenotype and larger seeds, but also larger embryos. Interestingly, inter-accession crosses (F1) resulted in bigger seeds than the parental self-crossed accessions, particularly when Bur-0 was used as the female parental genotype, suggesting parental effects on seed size that might be maternally controlled. Furthermore, developmental stage-based comparisons revealed that the large embryo size of Bur-0 is achieved during late embryogenesis and the large rosette size is achieved during late postembryonic growth. Interestingly, developmental phase progression analyses revealed that from germination onwards, the length of developmental phases during postembryonic growth is delayed in Bur-0, suggesting that in general, the mechanisms that regulate developmental phase progression are shared across developmental phases.

On the other hand, a detailed physiological characterization in different tissues at different developmental stages revealed accession-specific physiological and metabolic traits that underlie accession-specific phenotypes and in particular, more carbon resources during embryonic and postembryonic development were found in Bur-0, suggesting an important role of carbohydrates in determination of the bigger Bur-0 phenotype. Additionally, differences in the cellular organization, nuclei DNA content, as well as ploidy level were analyzed in different tissues/cell types and we found that the large organ size in Bur-0 can be mainly attributed to its larger cells and also to higher cell proliferation in the SAM, but not to a different ploidy level.

Furthermore, RNA-seq analysis of embryos at torpedo and mature stage, as well as SAMs at vegetative and floral transition stage from Bur-0 and Col-0 was conducted to identify accession-specific genetic determinants of plant phenotypes, shared across tissues and developmental stages during embryonic and postembryonic growth. Potential candidate genes were identified and further validation of transcriptome data by expression analyses of candidate genes as well

as known key regulators of organ size and growth during embryonic and postembryonic development confirmed that the high confidence transcriptome datasets generated in this study are reliable for elucidation of molecular mechanisms regulating plant growth and accession-specific phenotypes in Arabidopsis.

Taken together, this PhD project contributes to the plant development research field providing a detailed analysis of mechanisms underlying plant growth and development at different levels of biological organization, focusing on Arabidopsis accessions with remarkable phenotypical differences. For this, the natural accession Bur-0 was an ideal outlier candidate and different mechanisms at organ and tissue level, cell level, metabolism, transcript and gene expression level were identified, providing a better understanding of different factors involved in plant growth regulation and mechanisms underlying different growth patterns in nature.

Zusammenfassung

Der Lebenszyklus blühender Pflanzen ist ein dynamischer Prozess, der das erfolgreiche Durchlaufen mehrerer Entwicklungsphasen impliziert. Es wurden enorme Fortschritte gemacht, um zelluläre und molekulare Regulationsmechanismen zu entschlüsseln, die diesen Phasen, der Morphogenese und dem Wachstum zu Grunde liegen. Obwohl mehrere Schlüsselregulatoren des Pflanzenwachstums oder der Entwicklungsphasenübergänge in Arabidopsis identifiziert wurden, ist nur wenig über Faktoren bekannt, die sowohl während der Embryogenese als auch während der Samenentwicklung und dem weiteren Wachstum aktiv werden. Noch viel weniger ist über akzessionspezifische Faktoren bekannt, die die Pflanzenarchitektur und Organgröße bestimmen. Bur-0 wurde als eine natürliche Arabidopsis-Akzession mit außergewöhnlich großen Samen und großer Blattrosette beschrieben. Ihr Phänotyp macht sie zu einem interessanten Kandidaten für die Untersuchung von Wachstums- und Entwicklungsaspekten in Pflanzen, jedoch muss die molekulare Basis, die diesem großen Phänotyp unterliegt, noch entschlüsselt werden. Daher war das allgemeine Ziel dieser Doktorarbeit, die molekularen Mechanismen, die dem großen Phänotyp in Bur-0 zu Grunde liegen, zu entschlüsseln und zu verstehen.

Mehrere natürliche Arabidopsis-Akzessionen und spät blühende Mutantenlinien wurden in dieser Studie analysiert, so auch Bur-0. Die Phänotypen wurden durch eine detaillierte Analyse der Rosettengröße, der Samengröße, der Blütezeit, der Sprossapikalmeristemgröße und des Wachstums in verschiedenen Photoperioden, während der embryonalen und postembryonalen Entwicklung charakterisiert. Unsere Ergebnisse zeigen, dass Bur-0 als interessanter Akzession mit gleichzeitig größeren Blattrosetten, größerem Sprossapikalmeristem (SAM), späterem Blühphänotyp und größeren Samen, aber auch größeren Embryonen auffällt. Interessanterweise führten Kreuzungen zwischen den Akzessionen (F1) zu größeren Samen als die elterlichen selbstgekreuzten Akzessionen, insbesondere wenn Bur-0 als weiblicher elterlicher Genotyp verwendet wurde, was auf elterliche Effekte auf die Samengröße hindeutet, die möglicherweise mütterlicherseits kontrolliert werden. Darüber hinaus ergaben Vergleiche auf Basis von Entwicklungsstadien, dass die große Embryogröße von Bur-0 während der späten Embryogenese erreicht wird und die große Blattrosette während des späten postembryonalen Wachstums. Interessanterweise ergaben Analysen der Entwicklungsphasenprogression, dass ab der Keimung die Länge der Entwicklungsphasen während des postembryonalen Wachstums bei Bur-0 verzögert ist, was darauf hindeutet, dass im Allgemeinen die Mechanismen, die die Entwicklungsphasenprogression regulieren, über die Entwicklungsphasen hinweg geteilt werden.

Andererseits ergab eine detaillierte physiologische Charakterisierung in verschiedenen Geweben in unterschiedlichen Entwicklungsstadien akzession-spezifische physiologische und metabolische Merkmale, die den akzession-spezifischen Phänotypen zu Grunde liegen. Insbesondere wurden mehr Kohlenstoff-Ressourcen, während der embryonalen und postembryonalen Entwicklung in Bur-0 gefunden, was auf eine wichtige Rolle von Kohlenhydraten bei der Bestimmung des größeren Bur-0-Phänotyps hindeutet. Zusätzlich wurden Unterschiede in der zellulären Organisation, dem DNA-Gehalt der Nuklei sowie dem

Ploidiegrad in verschiedenen Geweben/Zelltypen analysiert und wir fanden heraus, dass die größere Organgröße in Bur-0 hauptsächlich auf die größeren Zellen und auch auf eine höhere Zellproliferation im SAM zurückzuführen ist, aber nicht auf einen anderen Ploidiegrad.

Darüber hinaus wurden RNA-seq-Analysen von Embryonen im Torpedo- und Reifestadium sowie SAMs im vegetativen und Florenübergangsstadium von Bur-0 und Col-0 durchgeführt, um akzession-spezifische genetische Faktoren für Pflanzenphänotypen zu identifizieren, die in allen Geweben und Entwicklungsstadien während des embryonalen und postembryonalen Wachstums auftreten. Potenzielle Kandidatengene wurden identifiziert und eine weitere Validierung der Transkriptomdaten durch Expressionsanalysen neuartiger Kandidatengene sowie bekannter Schlüsselregulatoren für Organgröße und -wachstum während der embryonalen und postembryonalen Entwicklung bestätigte, dass die in dieser Studie generierten Transkriptomdatensätze mit hoher Zuverlässigkeit für die Aufklärung molekularer Mechanismen zur Regulierung des Pflanzenwachstums und akzessionspezifischer Phänotypen in Arabidopsis geeignet sind.

Insgesamt trägt diese Doktorarbeit zur Forschung im Bereich der Pflanzenentwicklung bei, indem sie eine detaillierte Analyse der Mechanismen liefert, die dem Wachstum und der Entwicklung auf verschiedenen Ebenen der biologischen Organisation zu Grunde liegen, wobei der Schwerpunkt auf Arabidopsis-Akzessionen mit bemerkenswerten phänotypischen Unterschieden liegt. Dafür war die natürliche Akzession Bur-0 ein idealer Ausreißerkandidat und es wurden verschiedene Mechanismen auf Organ- und Gewebeebene, Zellebene, Stoffwechsel, Transkript- und Genexpressionsniveau identifiziert, was ein besseres Verständnis der verschiedenen Faktoren, die an der Regulierung des Pflanzenwachstums beteiligt sind, und der Mechanismen, die den verschiedenen Wachstumsmustern in der Natur zu Grunde liegen, ermöglicht.

Abbreviations

°C	Degree Celsius
µg	Microgram
µL	Microliter
µm	Micrometer
µM	Micromolar
ANOVA	Analysis of variance
Alst-1	Alston-1, *Arabidopsis thaliana* accession
Ang-0	Angleur-0, *Arabidopsis thaliana* accession
Araport	The Arabidopsis Information Portal
bp	Base pair
Bur-0	Burren-0, *Arabidopsis thaliana* accession
cDNA	Complementary DNA
Cen-0	Caen-0, *Arabidopsis thaliana* accession
cm	Centimeter
Col-0	Columbia-0, *Arabidopsis thaliana* accession
DAPI	4′,6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide
DEPC	Diethyl dicarbonate
DTT	Dithiothreitol
EDTA	Ethylene diamine tetra-acetic acid
eFP Browser	Electronic Fluorescent Pictographic Browser
EtOH	Ethanol
et al.	*et alia* (and others)
F1	First generation after crossing two parents
G2-phase	Gap two-phase
h	Hours
H_2O	Dihydrogen oxide
HCl	Hydrogen chloride
kb	Kilobase
L1	Layer one
L*er*-1	Landsberg *erecta*-1, *Arabidopsis thaliana* accession

LiCl	Lithium chloride
Lip-0	Lipowiec-0, *Arabidopsis thaliana* accession
M	Molar
mg	Milligram
$MgCl_2$	Magnesium chloride
min	Minute
mL	Milliliter
mM	Millimolar
MOPS	3-Morpholinopropane-1-sulfonic acid. Buffer
M-phase	Mitosis-phase
NaCl	Sodium chloride
$NaHCO_3$	Sodium hydrogen carbonate
Na_2CO_3	Sodium carbonate
NaOH	Sodium hydroxide
NCBI	National Center for Biotechnology Information
ng	Nanogram
PANTHER	Protein Analysis Through Evolutionary Relationships
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pH	Log of the hydrogen ion concentration in an aqueous solution
qRT-PCR	Quantitative real time-polymerase chain reaction
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT	Room temperature
Sap-0	Slapy-0, *Arabidopsis thaliana* accession
SDS	Sodium dodecyl sulfate
Sei-0	Seis am Schlern-0, *Arabidopsis thaliana* accession
S-phase	Synthesis-phase
TAIR	The Arabidopsis Information Resource
tRNA	Transfer ribonucleic acid
Ts-1	Tossa de Mar-1, *Arabidopsis thaliana* accession
Tris	Tris (Hydroxymethyl) aminomethane
Tukey–HSD	Tukey-Honestly Significant Difference
Ws-2	Wassilewskija-2, *Arabidopsis thaliana* accession

List of Figures

List of Supplementary Tables

List of Supplementary Figures

1. INTRODUCTION

1.1. Plant growth and development

The life cycle of flowering plants is a dynamic process that involves successful passing through several developmental phases such as embryogenesis, germination, vegetative and reproductive growth. In general, the development of a flowering plant begins with division of a fertilized egg to form an embryo with a polarized organization: the apical part will form the shoot, the basal part, the root, and the middle part, the stem (Alberts *et al*., 2002). Initially, cell division occurs throughout the body of the embryo, however, as the embryo grows, addition of new cells becomes restricted to small clusters of undifferentiated cells known as meristems, thus morphogenesis of a developing plant also depends on orderly cell divisions followed by strictly oriented cell expansions (ten Hove *et al*., 2015).

The initial specification and establishment of apical meristems (shoot and root tips) and the three fundamental tissues (epidermis, vasculature and ground tissue) occur in the early embryo and their maintenance and renewal continues throughout the life of a plant, enabling plants to grow by sequentially adding new organs, such as leaves branches, flowers and roots to build complex postembryonic structures (Alberts *et al*., 2002; ten Hove *et al*., 2015). The mature plant is typically made of many copies of a small set of standardized modules. The positions and times at which those modules are generated are strongly influenced by the environment, but because plants cannot move, they adapt their growth and development to the respective environment (Alberts *et al*., 2002).

Environmental cues, especially light, can cause the expression of genes that switch the shoot apical meristem (SAM) from a leaf-forming to a flower-forming mode. Depending on the type of the organ initiated by the SAM, the plant postembryonic development has been divided into vegetative and reproductive growth. During the vegetative phase, the SAM produces leaves (juvenile and adult), while during reproductive phase, flowers are produced instead (Clark, 1997).

Tremendous progress has been made to reveal cellular and molecular regulatory mechanisms underlying plant developmental phases, morphogenesis, and growth. Most of what is currently known is based on studies using the model organism *Arabidopsis thaliana* because it has a small genome, is relatively easy to genetically manipulate and availability of whole-genome information of a large collection of naturally inbred lines (accessions) that are products of natural selection under diverse ecological conditions provide a powerful resource for determining how genetic variation translates into phenotypic variation.

1.1.1. Embryo and seed development

Embryogenesis in higher plants can be broadly divided into three overlapping phases. The first phase is morphogenesis, during which the polar axis of the plant body is defined with the specification of the shoot and root apices and the embryonic fundamental tissues are formed.

The second phase is embryo maturation, characterized by the accumulation of storage reserves and growth. During the third phase, the embryo prepares for desiccation and enters a period of developmental arrest (West & Harada, 1993).

Embryos are a fascinating study object because within a period that is limited by the time until the seed desiccates (in seed plants), sufficient cells need to be generated during embryogenesis to build a new body and these cell divisions need to be ordered in a precise way to achieve a species-specific morphology. Furthermore, the complex architecture of an adult plant is the result of iterations of the same elementary developmental processes that first occur during early embryogenesis: organ initiation, growth, and pattern formation (ten Hove *et al*., 2015).

In flowering plants in general, after the egg is fertilized, the diploid embryo starts to develop and polarity becomes apparent at the first cell division, which is asymmetric and produce a smaller apical cell and a larger basal cell. The apical cell will generate most of the resulting embryo, whereas the basal cell will give rise to a nutritive structure (the suspensor) and a portion of the root meristem (Long, 2006). The suspensor attaches the embryo to the adjacent nutritive tissue and provides a pathway for the transport of nutrients (Alberts *et al*., 2002).

The shoot apical meristem (SAM) and the root apical meristem (RAM) are formed during early embryogenesis and they contain niches with pluripotent stem cells that constantly divide and give rise to two types of daughter cells: the cells that stay in the center remain stem cells, whereas daughter cells that are displaced to the periphery of the meristem enter a developmental pathway that leads to differentiation for organ formation (Nakajima & Benfey, 2002; Weigel & Jürgens, 2002; ten Hove *et al*., 2015). At the same time, it is possible to distinguish the first tissue precursors: epidermal cells, forming the outermost layer of the embryo, ground tissue cells, occupying most of the interior, and vascular tissue cells forming the central core, whereas endosperm tissues (the second product of double fertilization) surround the embryo and provide it with nutrients as it develops (Alberts *et al*., 2002).

Great progress has been made in the last decade to understand the cellular and genetic regulatory mechanisms that direct organ initiation, growth, and pattern formation as well as particular steps during early embryogenesis. Important transcriptional, epigenetic, and hormonal regulators have been identified and genetic regulatory mechanisms that direct zygote development, apical and basal cell fate determination, shoot and root domains determination, establishment of epidermis, vasculature, and ground tissue, as well as establishment of the RAM and SAM stem cell niches (stem and organizer cells) in the early embryo have been widely reported (reviewed in Palovaara *et al*., 2016). Most of the studies have been made on Arabidopsis embryos, an excellent model for studying such processes, but also a challenge due to its very small size and encapsulation in a small seed.

Briefly, during early embryogenesis following zygote divisions, the apical and basal domains are established and patterning in the embryo is dependent on auxin and cytokinin control. The basal domain is formed first and asymetric divisions and specification of the hyprophysis (the precursos of the quiscent center (QC)) involves auxin-dependent factors like *TARGET OF MONOPTEROS7* (*TMO7*), the transcription factor *NO TRANSMITTING TRACT* (*NTT*), its two

paralogs *WIP DOMAIN PROTEIN4* (*WIP4*) and *WIP5*, which are trigerred by auxin through *MONOPTEROS* (*MP*) and *PLETHORA* (*PLT*) (not regulated directly by MP), all identified as critical regulators of auxin-dependent root formation (Crawford *et al*., 2015).

Furthermore, ectopic cytokinin signaling in the basal derivative of the hypophysis interferes with the stereotypical cell division pattern of the root pole and auxin dampens this signaling *via* two negative cytokinin regulators, *ARABIDOPSIS RESPONSE REGULATOR7* (*ARR7*) and *ARR15* (Müller & Sheen, 2008). Moreover, auxin/cytokinin cross talk is crucial for vascular and ground tissue formation, growth and cell-specific division patterns. The RAM and the fundamental tissues are also established and *ARABIDOPSIS THALIANA MERISTEM LAYER1* (*ATML1*) acts as master transcriptional regulator of epidermal cell fate in the plant embryo (Abe *et al*., 2003).

The apical domain is established later, *TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS1* (*TAA1*) and *YUC1/4* initiate auxin production at the proembryo apex and *PIN1* is polarized in the inner proembryonic cells to mediate basal auxin transport (Robert *et al*., 2013). Expression of several HD-ZIP III family members is spatially restricted to the upper cells and further asymetric divisions are regulated by auxin signaling and *WOX* transcription factors. Later on, the SAM is established and its organization is regulated during early embryogenesis by *WUSCHEL* (*WUS*), *SHOOT MERISTEMLESS* (*STM*), *CLAVATA3* (*CLV3*) and HD-ZIP III family members (Prigge *et al*., 2005; ten Hove *et al*., 2015).

Afterwards, the rudiment of the shoot begins to produce the cotyledons leaves, one in the case of monocots and two in the case of dicots (Alberts *et al*., 2002; ten Hove *et al*., 2015). Later on, during the final phase embryo growth is exclusively characterized by events of cellular expansion and subsequent cell differentiation without cell divisions, development usually halts, and the embryo is stabilized by dehydration inside the seed, which can remain dormant for a very long time. Upon rehydration, the seeds germinate and embryonic development resumes (Alberts *et al*., 2002; Locascio *et al*., 2014).

On the other hand, seed development has been widely studied in the last years and in general, it is determined by the coordinated and highly controlled communication and growth of all seed components (embryo, endosperm, and seed coat) (Rewiewd in Nowack *et al*., 2010). The diploid embryo embodies the structure of the future adult plant and has all the necessary elements for plant development after germination, the triploid endosperm constitutes the reservoir for all the nutrients that the embryo will use during development until the new plant becomes autotrophic and the seed coat (derived from diploid maternal integuments of the ovule) protects the vital part of the seed from mechanical injury, predators and drying out (Locascio *et al*, 2014).

Seed development can be broadly categorized in two main phases: (I) Morphogenesis, which covers all the processes involving formation and structural development of all the components of the mature seed (embryo, endosperm, and integuments) and phase (II) Maturation, where the seed loses up to 95% of its water content (desiccation), nutrients are stored in the endosperm (Monocots) or in the cotyledons (Eudicots), cell cycle activities are stopped, RNA and protein

synthesis decrease and during late maturation the seed is metabolically quiescent (state of dormancy) (Raz *et al*., 2001; Santos-Mendoza *et al*., 2008; Locascio *et al*., 2014).

Seed development has been well described in *Arabidopsis thaliana* as a model plant for the Eudicots and *Zea mays* L. (maize) for the Monocots. Even though Monocots and Eudicots share most of the seed structures (Figure 1), the processes that lead to seed development and maturation are remarkably different between the two models in the later stage of endosperm development (Locascio *et al*, 2014). While in Arabidopsis the endosperm is absorbed at the end of the maturation phase to provide space for the embryo to grow, in maize the endosperm persists and covers other important roles on embryo development and seed organization (reviewed in Locascio *et al*., 2014).

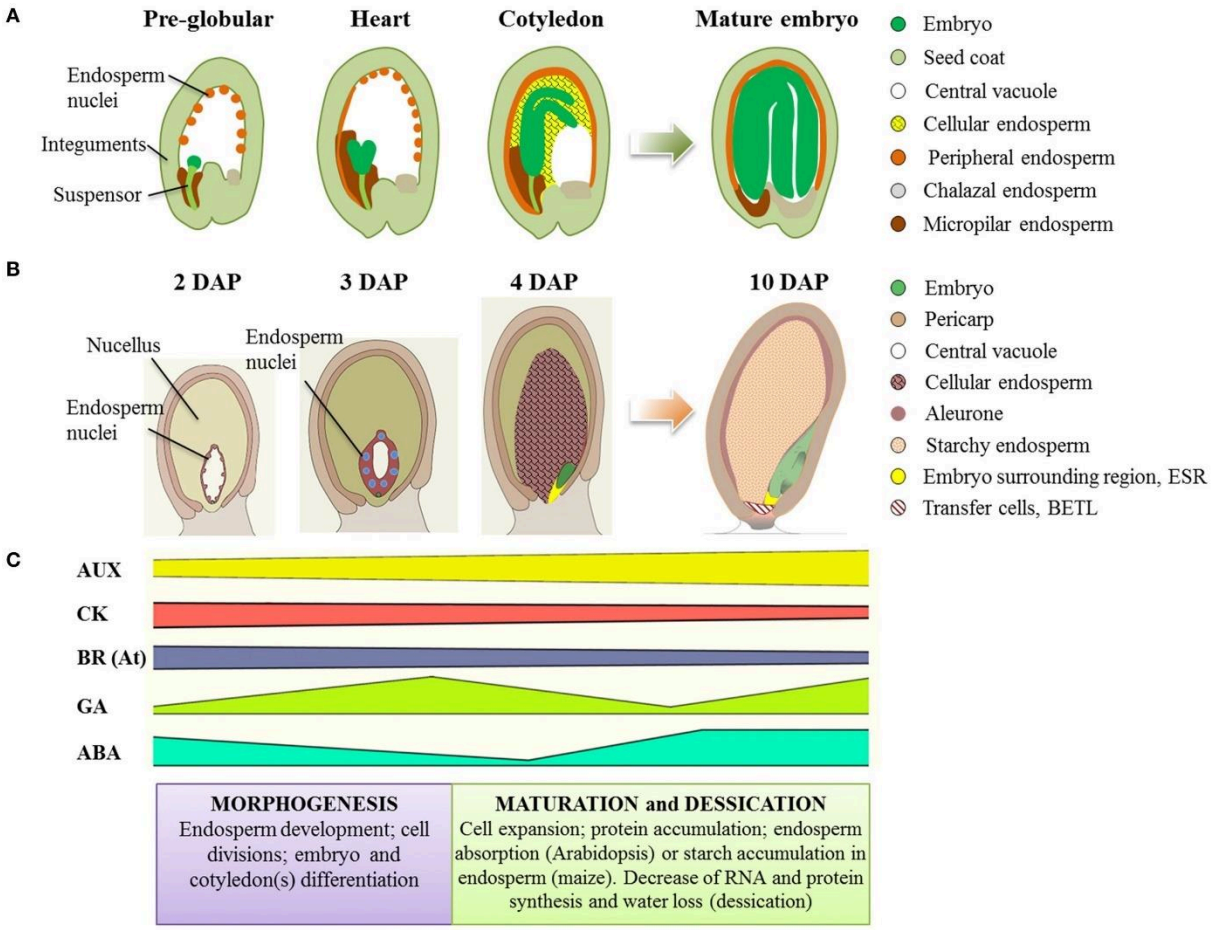


Figure 1. Seed development in Arabidopsis and Maize. (A) Schematic representation of seed development in Arabidopsis. Embryo development stages are indicated. (B) Schematic representation of seed development in Maize. Stages indicate days after pollination (DAP). (C) Schematic trend of hormone accumulation during seed development (Locascio *et al.*, 2014).

In Arabidopsis, regulation of seed development has been well studied and is marked by an initial period of active endosperm proliferation followed by embryo growth (reviewed in Sun *et al*., 2010), as well as maturation processes ocurring in the seed components (embryo, endosperm, seed coat) (reviewed in Santos-Mendoza *et al*., 2008). Briefly, soon after fertilization, the endosperm nuclei undergo successive mitotic divisions without cell wall formation, generating the multinucleate endosperm. This phase is followed by cellularization of the endosperm and the definition of three regions: the micropylar, the peripheral and chalazal

endosperm (Sørensen *et al*., 2002). The cellularized endosperm acts as nourishing tissue that is consumed by the embryo and reduced until a single peripheral endospermic cell layer in the mature seed. The embryo goes through a period of cellular expansion and differentiation, grows and fills the seed volume and the main storage products (lipids and proteins) accumulate in the cotyledons (Santos-Mendoza *et al*., 2008; Sun *et al*., 2010). Starch and hexoses accumulate only transiently during seed development, while sucrose and some oligosaccharides gradually accumulate towards late maturation phase, being sucrose the most abundant soluble saccharide in the dry seed (Baud *et al*., 2002).

In the past two decades regulation of seed development and seed size has been widely studied through functional analyses and characterizations of mutants, transcriptomes, and QTLs mapping. The role of different transcriptional, epigenetic, hormonal, peptide and sugar signaling regulators as well as gene networks regulating growth and interdependent relationship between the three seed compartments (embryo, endosperm and integuments) have been deciphered in Arabidopsis and crop species like rice, maize and soybean (reviewed in Savadi, 2018).

Briefly, in the developing seed tissues cell cycles are regulated by cyclin-dependent kinases complexes (CDK/CYC) and their cyclin-dependent kinase inhibitors (CKIs), which are main cell cycle regulators (De Veylder *et al*., 2007; Dante *et al*., 2014). The differentiation of seed tissues at different stages also involves the ordered elimination of cells by the programmed cell death (PCD) process (Domínguez & Cejudo, 2014). Furthermore, cell fate is regulated by *LEAFY COTYLEDON1* (*LEC1*), *LEC2*, and *FUSCA3* (*FUS3*), which are principally expressed in the embryo and endosperm with a determinant function for both embryo development and for initiation and maintenance of the maturation phase (Lotan *et al*., 1998; Gazzarrini *et al*., 2004).

Moreover, a complex network of transcriptions factors (TFs) also regulates seed development and many of them have been well characterized (reviewed in Agarwal *et al*., 2011). For example, TFs like *APETALA2* (*AP2*), *AUXIN RESPONSE FACTOR2* (*ARF2*), *LARGE IN CHINESE* (*DA1*) have been reported as negative regulators of seed development, whereas others like *HAIKU1* (*IKU1*), *IKU2*, *MINISEEDS3* (*MINI3*), *AGAMOUS LIKE62* (*AGL62*) *TRANSPARENT TESTA GLABRA2* (*TTG2*), *SHORT HYPOCOTYL UNDER BLUE1* (*SHB1*) and *KLUH* (*KLU*) have been reported as positive regulators of seed development (Savadi, 2018). Some of these key regulators play an important role in specific seed tissues at particular stages, but others have a determinant function in different seed tissues as well as in different stages of seed development.

Additionally, seed development pathways are in general further regulated by the interaction of transcriptional regulators, phytohormones, peptide and sugar signaling regulators. The phytohormones cytokinins, brassinosteroids, and especially auxins are considered important signaling molecules in seed development (Sun *et al*., 2010), whereas abscisic acid (ABA) and Gibberellins (GAs) play an important role in the progression of seed maturation (Seo *et al*., 2006). Moreover, small peptides and sugar-mediated signaling are known to regulate seed development. For example, it has been reported that *AP2* together with the *Arabidopsis thaliana*

SUCROSE TRANSPORTER 5 (*AtSUC5*) is involved in the control of sucrose ratio and seed mass (Baud *et al*., 2005).

Another relevant mechanism controlling gene expression during seed development is exerted by microRNAs (miRNAs). It has been observed that miRNAs are expressed from early to later stages during seed development. More specifically, they seem to be implicated in the control of embryogenesis and embryo patterning, also affecting germination processes (Nodine & Bartel, 2010; Seefried *et al*., 2014). Additional key regulators and gene networks controlling seed development as well as the interdependent relationship between the three seed compartments are summarized in Figure 2, in the model reported by Savadi (2018).

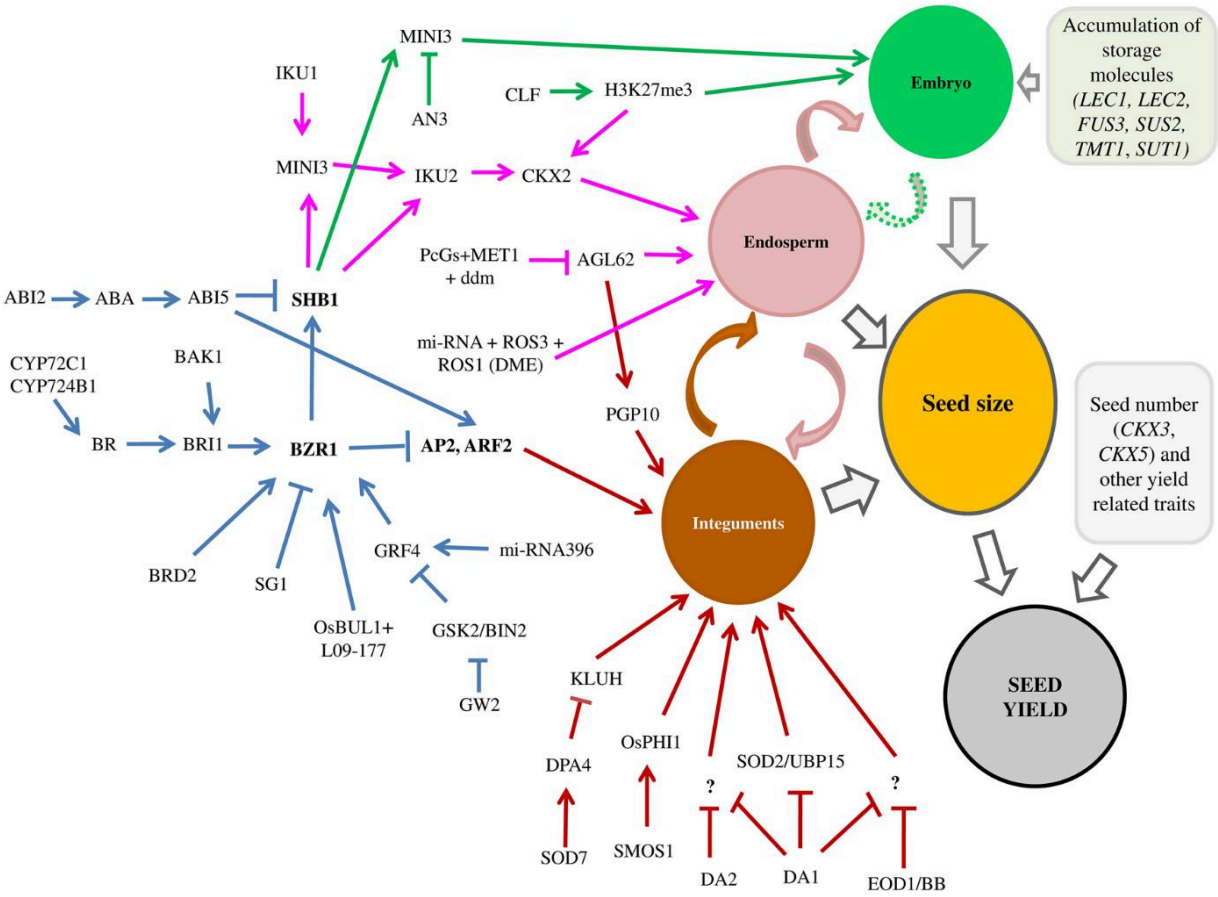


Figure 2. Gene networks regulating seed size and interdependent relationship between the three seed compartments. Seed size is governed by the coordinated growth of integuments, endosperm, and embryo. Embryogenesis-specific pathways (green color lines); endosperm proliferation-specific pathways (pink color lines), integument elongation-specific pathways (maroon color) (Savadi, 2018).

Moreover, epigenetic control of seed development and seed size has been also reported. At some genes there are parent-of-origin differences in the expression of the maternal and paternal alleles and this is referred to as imprinting (Waters *et al*., 2013). Regulation of genomic imprinting is complex and may involve DNA methylation, histone modifications and non-coding RNAs. Among the different seed tissues only endosperm development was known to be under epigenetic control, nevertheless, in maize the *MEE1* gene has been reported to be imprinted in the embryo (Jahnke & Scholten, 2009) and more recent genome wide approaches involving RNA-seq analysis in Arabidopsis (Nodine & Bartel, 2012), rice (Luo *et al*., 2011),

and maize (Waters *et al*., 2013) have also reported the presence of several potentially imprinted genes in embryos.

1.1.2. Postembryonic development

As soon as the seed coat ruptures during germination, a dramatic enlargement of nonmeristematic cells occurs, driving first the emergence of a root to establish an immediate foothold in the soil, and then of a shoot. This is accompanied by rapid and continual cell divisions in the apical meristems. The rapidly growing root and shoot test the environment and the root increases the plant's capacity for taking up water and minerals from the soil, while the shoot increases its capacity for photosynthesis (Alberts *et al*., 2002). After germination, plants undergo several developmental transitions (Figure 3) and the SAM starts a highly coordinated cell division program that continues throughout plants life cycle (Bäurle & Dean, 2006; Nakajima & Benfey, 2002).

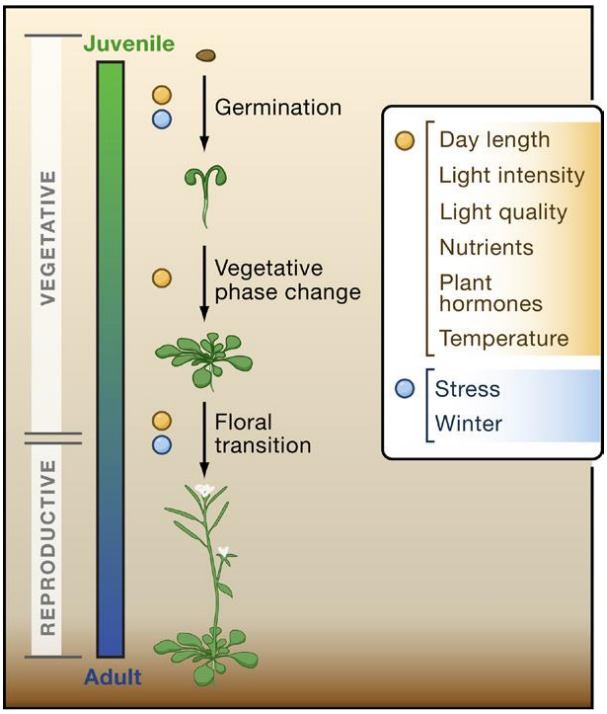


Figure 3. Developmental transitions during the plant life cycle. Germination is the transition between embryonic and postembryonic development. In the vegetative phase, the seedling progresses from the juvenile state into the adult state (vegetative phase change). The third major transition is the floral transition from the adult vegetative state to the reproductive state. Whereas the progression from the juvenile to the adult vegetative state is gradual, the floral transition is usually abrupt. All developmental transitions are regulated by environmental signals such as available nutrients, day length, light intensity, light quality, and ambient temperature as well as endogenous signals transmitted by plant hormones. Cold temperature and stress affect germination and the floral transition (Bäurle & Dean, 2006).

From germination onwards, the course of plant development is strongly influenced by signals from the environment. The shoot has to push its way rapidly up through the soil and must open its cotyledons and begin photosynthesis only after it has reached the light. The timing of this transition from rapid subterranean sprouting to growth in presence of light cannot be genetically programmed, because the depth at which the seed is buried is unpredictable. The developmental

switch is controlled instead by light, which, among other effects, acts on the seedling by inhibiting production of brassinosteroids (Alberts *et al*., 2002).

Shoot growth in higher plants is dependent on the activity of the shoot apical meristem (SAM), which contains a small, dome-shaped population of undifferentiated cells used for the formation of new lateral organs and for stem cells renewing (Scofield *et al*., 2014). The SAM has a highly organized structure subdivided into three domains: the central zone (CZ) of pluripotent stem cells, the peripheral zone primordia that contributes to the production of lateral organs, and the rib zone (RZ). Cells in the peripheral zone and the RZ are rich in cytoplasm and divide rapidly, the CZ is maintained by an underlying organizing center (OC) and below the OC is the RZ, which is responsible for the elongation of the stem (Yruela, 2015).

After germination, the seedling passes through a juvenile vegetative phase, where it is not competent to flower (Bäurle & Dean, 2006). In flowering plants like *Arabidopsis thaliana* during the vegetative phase the shoot meristems produce leaves on their flanks in regular patterns called phyllotaxy and within a plant, leaf shape and size can vary depending on developmental stage and growth conditions (Yruela, 2015). This is followed by the transition to the adult vegetative phase, where the seedling can respond to floral inductive signals (Bäurle & Dean, 2006).

The change to the subsequent generative phase is called floral transition, which is regulated by multiple flowering pathways that are controlled by environmental and endogenous factors and with the transition to flowering, the plant enters the reproductive phase (Bäurle & Dean, 2006; Yruela, 2015). Great progress has been made to understand the molecular basis of the juvenile to-adult vegetative phase change and the transition to flowering and parallels between the two transitions have begun to emerge. Both are significantly modified by similar environmental conditions and gibberellins, both involve transmissible signals originating from outside the shoot apical meristem and both transitions are regulated by repressive pathways that prevent the transition from occurring precociously (reviewed in Bäurle & Dean, 2006).

In general, the transition from vegetative to reproductive growth is mainly controlled by day length, which is perceived in leaves and induces a systemic signal, called florigen that moves through the phloem to the shoot apex. Vernalization (exposure to cold) is also an important factor regulating the floral transition (Yruela, 2015). Additionally, the effects of sugars on floral transition have been well studied in Arabidopsis and it has been reported that increased leaf carbohydrate export and starch mobilization are required for flowering, thus phloem carbohydrates might have a critical function (Corbesier *et al*., 1998) and sugars regulate floral transition by positively and negatively regulating the expression of floral identity genes (Ohto *et al*., 2001).

After floral induction, the primary meristem changes its activity from the production of leaf primordia to the production of floral primordia or floral meristems. This switch from vegetative to reproductive growth is important for the right timing of the floral initiation, which is essential for the optimal production of fruits and seeds in flowering plants and ensure reproductive processes (Meyerowitz, 1997; Yruela, 2015).

1.1.3. Key growth regulators shared throughout plant development

Significant advances have been made to understand the molecular mechanisms underlying plant development, as well as the switches between developmental phases, so-called transitions, particularly in the model plant *Arabidopsis thaliana.* With some exceptions, most genes expressed in the embryo are also expressed during postembryonic growth, which raises the question of the genetic and molecular relationships that exist between the various developmental and metabolic phases. In recent years, it has become evident that the genetic networks regulating different phases of the plant´s life cycle share some common factors. Interestingly, approximately 8000 diverse mRNAs, including 554 transcription factor (TF) mRNAs have been identified by microarray analysis as common factors shared by seeds, flower structures and whole rosettes (Le *et al*., 2010).

Given that the SAM largely determines the general architecture of a plant, meristem establishment and maintenance provide the capacity to accomplish the complex species-specific characteristics of the adult plant (Clark, 2001; Spencer *et al*., 2007). In Arabidopsis, several molecular pathways have been identified that regulate the SAM organization, and many of the genes involved are expressed during embryogenesis (ten Hove, *et al*., 2015). Among the key regulatory genes that become active during embryogenesis and play an essential role in postembryonic development in Arabidopsis are the SAM maintenance and establishment genes *SHOOT MERISTEMLESS* (*STM*), *CLAVATA 3* (*CLV3*) and *WUSCHEL* (*WUS*) (Nakajima & Benfey, 2002).

Significant changes in the morphology and size of organs occur when SAM organization or growth is disrupted (Hu *et al*., 2003). In Arabidopsis, *stm-1* mutant does not form the SAM during embryogenesis, and adventitious meristems produce only single leaves postembryonically. The weak *stm* mutant allele *stm-2* also has no SAM, but the adventitious meristems that form postembryonically can form at times shoots with fewer organs than usual (Clark *et al*., 1996; Meyerowitz, 1997). *Wus* mutants fail to properly organize a shoot meristem in the embryo, and during postembryonic development defective shoot meristems are initiated repetitively but terminated prematurely in aberrant flat structures (Laux *et al*., 1996). In contrast, *clv3* mutants have a strong phenotype with more cells in their embryonic SAMs, and enlarged SAMs are produced during postembryonic growth (Clark *et al*., 1995).

Another example of factors essential for plant growth is TREHALOSE-6-PHOSPHATE SYNTHASE (TPS1), which likely regulates the metabolic status of the plant through trehalose-6-phosphate (T6P) biosynthesis. Interestingly, the *tps1* mutation does not develop mature seeds (Eastmond *et al*., 2002). Furthermore, heterozygous mutant *tps1-2* plants were transformed with an inducible *TPS1* (*ind-TPS1*) construct and the *ind-TPS1* transgenic *tps1*/*tps1* plants showed that the root meristematic region is dramatically decreased, leaf growth is reduced, and floral transition is absent (Dijken *et al*., 2004). Transgenic tobacco (*Nicotiana tabacum*) plants expressing *Escherichia coli* homologs of TPS and trehalose-6-phosphate phosphatase show a positive correlation between trehalose-6-phosphate levels and photosynthetic activity, suggesting a regulatory role for trehalose-6-phosphate in plant carbohydrate metabolism (Paul *et al*., 2001). Moreover Fichtner *et al*. (2020) found that TPS1 protein is localized predominatly

in the phloem-loading zone, guard cells in leaves, root vasculature and the shoot apical meristem and they also reported an important role of TPS1 in regulation of sucrose metabolism and sucrose signaling in Arabidopsis plants. TREHALOSE-6-PHOSPHATE SYNTHASE1 (TPS1) catalyzes the synthesis of the sucrose-signaling metabolite trehalose 6-phosphate (Tre6P) in *Arabidopsis thaliana* and Trep6P has been implicated as a regulatory factor in several developmental transitions like embryogenesis, flowering and shoot branching, providing information of sucrose status and plant´s capacity to supply sucrose for the new grow that will follow the transition (reviewed in Fichtner & Lunn, 2021).

1.2. Molecular mechanisms underlying organ size regulation in plants

1.2.1. Organ size regulated by growth-promoting and growth-repressor factors

Final organ size is remarkably constant within a given species, suggesting that a species-specific size checkpoint terminates organ growth in a coordinated and timely manner (Disch *et al*., 2006). In plants, organ growth is controlled by genetic and environmental factors and several genetic determinants of final organ size, as well as pathways that may link organ growth with environmental conditions have been identified (Figure 4) and characterized in *Arabidopsis thaliana* (Krizek, 2009). Briefly, growth-promoting and growth-restricting factors like the growth promotor *AUXIN-REGULATED GENE INVOLVED IN ORGAN SIZE* (*ARGOS*) (Hu *et al*., 2003), *AINTEGUMENTA* (*ANT*) (Mizukami & Fischer, 2000), the non-cell autonomous growth promoter *KLUH* (*KLU*) (Anastasiou *et al*., 2007), the cellular biomass-stress resistance regulator *Arabidopsis TOR gene* (*AtTOR*) (Deprost *et al*., 2007), the plant growth repressor *AUXIN RESPONSE FACTOR 2* (*ARF2*) (Okushima *et al*., 2005), the E3 ubiquitin ligase growth respressor *BIG BROTHER* (*BB*) (Disch *et al*., 2006), and the organ size repressor *LARGE IN CHINESE* (*DA1*) (Li *et al*., 2008) have been identified, which mainly affect organ size without affecting organ shape and whenever determined, their loss- or gain-of-function phenotype showed opposite effects on organ size (Krizek, 2009).

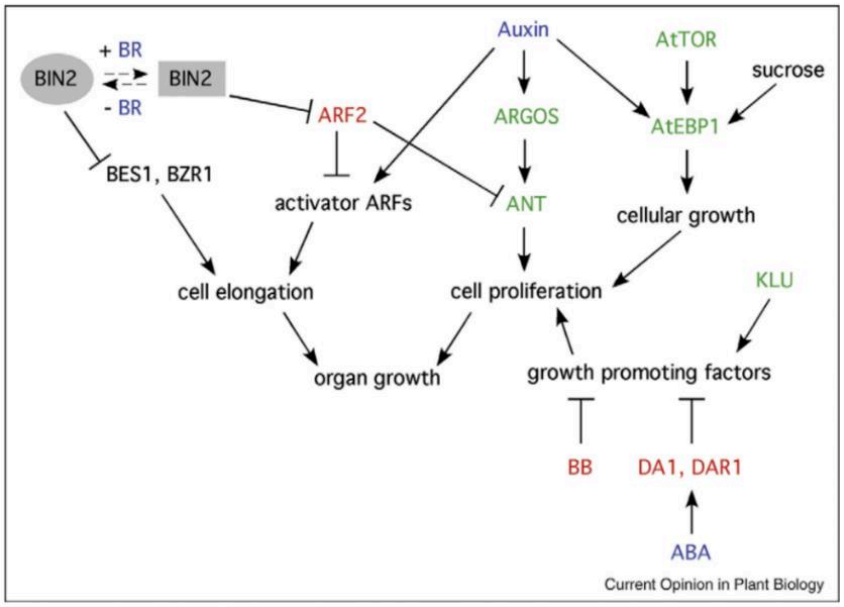


Figure 4. Pathways controlling final organ size. Plant hormones are shown in blue, proteins promoting growth are shown in green, and proteins restricting growth are shown in red (Krizek, 2009).

1.2.2. Organ size regulation and ploidy level

Cell and organ enlargement associated with ploidy level has been observed in many organisms. Ploidy describes the number of genome copies contained in a single nucleus. When more than two copies are present, the cell and nucleus can be described as polyploid (Robinson *et al*., 2018). A change in ploidy directly changes two parameters: (1) the bulk amount of chromatin in the nucleus and (2) the copy number of each gene. Two forms of polyploidy are often considered: allopolyploidy, which originates from interspecies hybrids, and autopolyploidy, which originates from intraspecies genome duplication events (whole-genome multiplication, WGM). Advances in this research field have been made in Arabidopsis polyploidy, but also in other agronomic plant species (reviewed in del Pozo & Ramirez-Parra, 2015).

The effects of an increase in ploidy are numerous and include changes in gene expression, nuclear size, cell size, and the size of organs and organisms (Yu *et al*., 2010; del Pozo & Ramirez-Parra, 2015; Robinson *et al*., 2018). WGM events are often associated with increases in plant size, cell size, and increased plant vigor (Otto, 2017). Natural plant polyploid species often exhibit improved growth vigour and adaptation to adverse environments, conferring evolutionary advantages that can also be used in crop breeding programs (del Pozo & Ramirez-Parra, 2015).

Another mechanism that increases ploidy in plant cells is endoreduplication, which occurs in individual cells during differentiation and allows cells to increase their ploidy above the organism's "base" ploidy level, sometimes dramatically. Endoreduplication occurs *via* the endocycle, an alternative to the mitotic cell cycle (Robinson *et al*., 2018). This cycle includes gap (G) and DNA synthesis (S) phases, but omits mitosis (M), causing multiple copies of the genome to be retained in a single nucleus. All diploid (2C) cells replicate their DNA to 4C during S phase of the cell cycle. Cells in the mitotic cycle divide into two 2C (diploid) daughter cells, whereas cells that endoreduplicate remain 4C and may enter consecutive endocycles to become 8C, 16C, 32C, and so on (Robinson *et al*., 2018). Endoreduplication and WGM both create polyploid cells, but these cells are probably not identical in their cytology. Because WGM increases ploidy in the zygote and in all descendant somatic cells, WGM polyploid cells must undergo mitotic divisions with extra genome copies. In contrast, endoreduplicated cells arise during terminal differentiation and very rarely divide (Roeder *et al*., 2010).

Endopolyploidy is present in most tissues in most angiosperms and in Arabidopsis, extensive endoreduplication has been found in several somatic tissues, which is developmentally regulated according to the age of the tissues, their position within the plant and is strongly associated with increased cell size (Galbraith *et al*., 1991). Cell size has important physiological implications, determining both the surface area/volume ratio and the ratio of cytoplasm/DNA, thereby likely impacting nutrient uptake rates, protein concentrations, and transcription frequencies. Cell size and growth rates vary strongly within a plant according to tissue and developmental stage and growth and size are evidently regulated in coordination with the cell cycle (Willis *et al*., 2016).

1.2.3. Organ size regulated by cell number or cell size

Plant organ growth occurs by an initial proliferative phase in which cell numbers increase while their size remains fairly constant, followed by dramatic cell size increases that cease when the set size of the organ is reached (Li *et al.*, 2008). Coordination of organ growth with cell division, cell expansion and cell differentiation is essential for organ-size determination in plants, as well as maintenance of meristematic competence in developing organs (Mizukami & Fischer, 2000; Mizukami, 2001). The size of a multicellular organism, its organs and tissues depend on the number and size of constituent cells. Cell number, in turn, depends on the rate of cell division, the number of dividing cells, and the duration of the cell proliferation phase during development, while the size of non-dividing cells is influenced by cell growth and cell expansion. In plants, cell number generally seems to make a larger contribution than cell size to the size of comparable organs (Mizukami, 2001; Dante *et al*., 2014).

In Arabidopsis some genetic determinants involved in the regulation of cell expansion during organ growth have been identified including *EXPANSIN10* (*EXP10*) (Cho & Cosgrove, 2000), *ARGOS-LIKE* (*ARL*) (Hu *et al*., 2006), *REGULATORY PARTICLE AAA-ATPASE 2a* (*RPT2a*) (Sonoda *et al*., 2009) and *ORGAN SIZE RELATED2* (*OSR2*) which affect the duration or the rate of cell expansion, and regulate the final size of cells and organs (Qin *et al*., 2014). Nevertheless, the consistent size of a given cell type shows that cells have mechanisms to measure their own size and adjust their growth rate, or rate of cell division, to maintain uniformity. However, how cells coordinate growth and division to achieve a particular cell size remains a fundamental question in biology and the understanding of this basic property of cells is limited, in part, by the lack of quantitative data on cellular growth and size kinetics over multiple generations, especially in higher eukaryotes (Ginzberg *et al*., 2015; Willis *et al*., 2016).

1.3. Organ size regulation and metabolic status

Plants can display photosynthesis, respiration, and fermentation at the same time in different tissues through a complex regulatory system that involves sugar signaling and integrates different metabolic, developmental, and environmental signals to control metabolic mode and activity (Rolland *et al*., 2002). The growth condition has a large impact on the values of metabolic traits, on the connectivity between metabolic traits and influences the connectivity between metabolism and growth (Sulpice *et al*., 2013).

Sugars such as sucrose, glucose, and fructose have an essential function in plant intermediary and respiratory metabolism and are the substrate for synthesizing complex carbohydrates such as starch and cellulose. Moreover, sugars provide the building blocks for amino acid and fatty acid biosynthesis and essentially all other compounds present in plants (Smeekens, 2000). The immediate products of photosynthetic carbon assimilation in the light are partitioned between sucrose (immediately available for growth) and starch (the most abundant carbohydrate reserve in plants). Starch accumulates in the leaf through the day, at night it is degraded to produce sucrose and is almost but not totally remobilized by the end of the night in plants growing with an adequate level of nutrients and favorable temperature (Lu *et al*., 2005; Smith & Stitt, 2007).

Photosynthesis is active primarily in mature leaf mesophyll cells, and photosynthate is transported, primarily as sucrose, to meristems and developing organs like young leaves, roots, flowers, fruits, and seeds. Light and sugars regulate organ growth by a coordinated modulation of gene expression and enzyme activities in both carbohydrate-exporting (source) and carbohydrate-importing (sink) tissues where hexoses as well as sucrose have been recognized as important signal molecules in source-sink regulation (Roitsch, 1999; Rolland *et al*., 2002).

Several studies have reported a central role of sugars in interactions that integrate light perception, stress responses, and hormone signaling (reviewed in Smeekens, 2000), and coordinate carbon and nitrogen metabolism (reviewed in Coruzzi & Bush, 2001; Stitt & Krapp, 1999). A genetic approach to dissect the complex mechanisms that underlie sugar sensing and signaling in plants has been implemented in the last years using Arabidopsis as a model plant, and a large collection of sugar signaling mutants has been isolated, based on either sugar-regulated gene expression or sugar-insensitive or sugar-hypersensitive phenotypes during germination and seedling development (Sheen *et al*., 1999; Gibson *et al*., 2001; Rook *et al*., 2001).

In general, low sugar status enhances photosynthesis, reserve mobilization, and export, while high sugar status promotes growth and carbohydrate storage, therefore sugars have a key role as long distance messengers of whole-organism carbohydrate status as well as substrates for both cellular metabolism, local carbohydrate sensing systems and for sugar-regulated gene expression in vascular plants (reviewed in Koch, 1996). This ensures optimal synthesis, use and distribution of resources and energy among tissues and organs and allows the adaptation of carbon metabolism to changing environmental conditions, which involves changes in cell division rates and global alterations in metabolic activities (Grossman & Takahashi, 2001).

The effect of carbon allocation on organ and whole plant architecture is illustrated most dramatically by carbohydrate storage and the associated cell expansion in reserve organs such as roots, fruits, seeds, and tubers, however sugars can also act as morphogens, providing positional information to the cell cycle machinery and different developmental programs (Rolland *et al*., 2002). For example, gradients of sugars have been reported to correlate spatially with mitotic activity in *Vicia faba* embryos (Borisjuk *et al*., 1998) and Pien *et al*., (2001) reported spatially regulated carbohydrate metabolism within the meristem in tomato and suggested involvement of carbohydrate metabolism in organogenesis.

Different mechanisms regulating the link between carbon balance and growth processes have been reported in recent years (reviewed in Smith & Stitt, 2007). Analyses of transcriptional and translational changes during carbon starvation and re-supply suggest that growth may be affected at several levels by carbon availability (Price *et al*., 2004). During starvation there are general decreases in the levels of transcripts necessary for cell division, cell cycle, DNA synthesis and repair and biosynthesis of proteins, as well as a general decrease of transcripts encoding proteins for cell wall synthesis and proteins that are believed to modify the cell wall during expansion growth. On the other hand, recovery from starvation involves very rapid increases in expression of many biosynthesis and growth-related genes that are down regulated during starvation (Osuna *et al*., 2007).

Furthermore, the role of sugars in promoting cell division has been known for many years and the best understood link between carbon availability and growth is in control of the G1/S transition in the cell cycle, which is also influenced by hormonal status, since sugars and hormones act interactively (Riou-Khamlichi *et al.*, 1999). Menges & Murray (2002) reported differences in expression timing and activity of cell cycle-related genes and cell cycle progression in cell cultures where sucrose was removed and resupplied. The effect of sucrose levels in the G1 to S transition has been further investigated and Menges *et al*. (2006) reported that in cultured Arabidopsis cells, the D-cyclin protein CYCD3;1 dominantly drives the G1/S transition and in sucrose-depleted cells the decline in CYCD3;1 level leads to G1 arrest. Expression of the *CYC3;1* gene is strongly up regulated by sucrose and because the protein turns over rapidly, the presence of sucrose is essential to maintain protein levels and hence the G1/S transition rate (Planchais *et al*., 2004; Menges *et al*., 2006).

1.4. Natural variation studies on plant architecture and organ morphology

Several *Arabidopsis thaliana* natural accessions (1001 genome project) have been collected from wild populations growing throughout the world and tremendous variation has been reported in their physiological, morphological and life history traits, providing a powerful resource for determining how genetic variation translates into phenotypic variation (1001 Genomes, 2018). Few systems share the key advantage of *A. thaliana* for complementary forward genetics approaches: the ready availability of whole-genome information of a large collection of naturally inbred lines (accessions) that are products of natural selection under diverse ecological conditions, all available from the stock centers (The 1001 Genomes Consortium, 2016).

Several studies aiming to identify and characterize genetic determinants responsible for organ size regulation have been focused on single key genes. Although great progress has been made in identifying specific genes contributing to trait variation, the high level of gene interactions underlying quantitative traits makes it unlikely that single genes studies would provide mechanistic models for such traits, or that such studies would have greater predictive power than quantitative genetic models (Roff, 2007). Identification of genomic regions that control quantitative traits has been possible using quantitative trait loci (QTLs) and genome wide association (GWAS) mapping, both powerful tools to identify genetic basis underlying natural variation and traits of agronomical importance in wild and crop plants, although it should be noted that all mapping studies are biased in the sense that they can only detect alleles that explain a sufficient fraction of the variation in the mapping population (Atwell *et al*., 2010; Bartoli & Roux, 2017).

Developmental traits related to plant architecture and organ morphology have been studied in *A. thaliana*, where natural variation for seed size (Alonso-Blanco *et al*.,1999), leaf morphology (Pérez-Pérez *et al*., 2002), and flower size (Juenger *et al*., 2005) have been described. In addition, the analysis of natural variation has also identified molecular mechanisms involved in growth regulation and many mapping studies have been performed for plant growth and size resulting in the identification of many quantitative trait loci (QTLs) (reviewed in Alonso-Blanco *et al*., 2009). Nevertheless, few of the underlying identified genes have been cloned and

thus, elucidation of additional players explaining a large part of the plant size variation observed in nature seem to be scarce (Bac-Molenaar *et al*., 2015).

1.4.1. Quantitative trait locus (QTL) approach

With the development of quantitative trait locus (QTL) mapping the identification of genomic regions controlling a quantitative trait has become feasible (Kearsey, 1998; Miles & Wayne, 2008) and the analysis of multiple traits in the same experimental mapping population enables the detection of loci with pleiotropic effects on various characters (Alonso-Blanco *et al*., 1999).

QTL studies require very large sample sizes, they can only map differences that are captured between the initial parental strains and some loci will remain undetected because the parental strains are unlikely to contain segregating alleles of large effect at every locus contributing to variation. Furthermore, identifying the actual loci that affect a quantitative trait involves demonstrating causality, the quest for individual genes within a QTL is frequently assisted by the identification of a priori candidate genes using classical reverse genetics or bioinformatics and a functional relationship between the candidate gene and the QTL must then be further demonstrated (Miles & Wayne, 2008).

Once a QTL has been identified, molecular techniques can be employed to narrow the QTL down to candidate genes, however the number of times that individual genes have been identified following a QTL mapping study remains small relative to the effort invested in QTL analysis. One reason for this discrepancy is that many QTL map to regions of the genome of perhaps 20 centimorgans (cM) in length, and these regions often contain multiple loci that influence the same trait (Kearsey & Farquhar, 1998; Miles & Wayne, 2008). Even so, QTL analysis has allowed identification of loci (and alleles at the same loci) of interest, as well as specific segregating alleles with relevant implications in medicine and agriculture studies and the availability of high-throughput technologies, genomic and proteomic data will enable further extensions and adaptations of QTL analysis (Kearsey & Farquhar, 1998; Miles & Wayne, 2008).

1.4.2. Genome wide association analysis (GWAS) approach

Genome-wide association studies (GWAS) are becoming increasingly popular in genetic research, and they are an excellent complement to QTL mapping. Whereas QTL contain many linked genes, which are then challenging to separate, GWAS can identify many unlinked individual genes or even nucleotides (Miles & Wayne, 2008). Rather than looking for marker–trait associations in a population with known relationships (such as the members of a pedigree, or the offspring of an experimental cross), GWAS approach look for associations in the general population of "unrelated" individuals, thus phenotypically similar individuals may share alleles inherited identical by descent, alleles that will be surrounded by short ancestral marker haplotypes that can be identified in genome-wide scans (Aranzana *et al*., 2005).

Association mapping has two main advantages over traditional linkage mapping methods. First, the fact that no pedigrees or crosses are required often makes it easier to collect data. Second,

because the extent of haplotype sharing between unrelated individuals reflects the action of recombination over very large numbers of generations, association mapping has several orders of magnitude higher resolution than linkage mapping. Nevertheless, GWAS studies also require very large sample sizes and although common alleles with major effects from as few as 96 accessions have been identified (Atwell *et al*., 2010), a much larger sample is required for most traits (The 1001 Genomes Consortium, 2016) and these studies are riddled with large expected numbers of false positives (Miles & Wayne, 2008).

GWAS studies can be followed up by functional validation studies of the causative genes and furthermore, by analysis of transcriptional and/or post-transcriptional regulation, interactions with proteins, identification of the downstream signaling pathways and so on, thereby providing additional candidate genes for breeding programs (Bartoli & Roux, 2017). Although GWAS remains limited to organisms with genomic resources, it is a powerful genetic approach to identify individual genes or even nucleotides that contribute to a particular phenotype, to study natural variation and to identify traits of agricultural importance (Miles & Wayne, 2008; Atwell, *et al*., 2010).

1.5. The *Arabidopsis thaliana* accession Burren 0 (Bur-0)

Burren (Bur-0) is an Arabidopsis natural accession originally collected more than 50 years ago (1959) in the Burren region of County Clare and County Galway in western Ireland (Tabib *et al*., 2016), a zone internationally recognized for its karstic geology and an exceptional flora that includes arctic-alpine and northern plant communities often growing in great abundance (Feeser & O'Connell, 2019). The Burren region extends from sea level to heights over 300 m and has a temperate climate with mean daily temperatures around 14°C in July and August (warmest months) and less than 5°C in January and February (coolest months) (EFNCP, 2020).

Although Bur-0 is not one of the most commonly used Arabidopsis accessions, it belongs to the Arabidopsis 1001 genome project collection (Alonso-Blanco *et al*., 2016) and its big phenotype makes it an interesting candidate to study growth and developmental aspects in plants. In general, Bur-0 has been reported in the literature as a particularly large rosette, big seed, and late flowering phenotype independent of vernalization (Werner *et al*., 2005; Herridge *et al*., 2011; Camargo *et al*., 2014). In addition, Bur-0 has been reported to have high nitrogen use efficiency, to be poorly tolerant to a high nitrogen supply (Chardon *et al*., 2010) and to be relatively insensitive to low carbon conditions (Sulpice *et al*., 2013), however, few studies have reported molecular determinants of the Bur-0 phenotype.

One example is the mapping analysis reported by Werner *et al*. (2005), who studied natural variation in the photoperiod pathway by identifying late-flowering accessions with limited *FLC* activity and little response to vernalization. They found that Bur-0 has a strong loss-of-function allele at *FLC*, an alternatively spliced *FLC* allele that behaves as a null allele that does not respond to *FRI*, thus its recessive late-flowering phenotype may be independent of *FLC*.

Another example was reported by Sureshkumar *et al*. (2009), who determined how copy number variation in tandemly repeated, short DNA sequences can underlie phenotypic

variability. These authors found an "*irregularly impaired leaves*" (*iil*) phenotype in Bur-0 under short days at temperatures above 27°C. The *iil* phenotype occurs due to a dramatic expansion of a TTC/GAA trinucleotide repeat (from 23 copies in Col-0 and Pf-0 to more than 400 copies in Bur-0). When the triplet is present in a homozygous state, causes a temperature-dependent reduction in *IIL1* transcript levels and severely impairs growth in Bur-0 (Sureshkumar *et al*., 2009). Later on, Tabib *et al*. (2016) collected and analized several *A. thaliana* wild Irish populations to assess whether the GAA/TTC repeat expansion was persistent, and whether it was unique to the surroundings of the Burren region. The authors demonstrated that the *IIL1* repeat expansion is a cryptic genetic variation that is revealed only under specific environmental conditions and the *IIL1* triplet repeat expansion, which causes severe growth impairments at high ambient temperatures, has been maintained in the Irish populations for over 50 years.

1.6. Aim and objectives

Even though the Arabidopsis natural accession Bur-0 has been often highlighted in natural variation studies due to its exceptionally large phenotype, the molecular basis underlying this interesting phenotype remain to be elucidated. Although several key regulators of plant growth or developmental phase transitions have been identified in Arabidopsis, little is known about factors that become active during embryogenesis, seed development and also during further postembryonic growth. Much less is known about accession-specific factors that determine plant architecture and organ size. In order to understand the mechanisms driving organ size variation in natural populations, identification of genes responsible for this variation is crucial. Moreover, a comprehensive understanding of gene regulatory networks operating at different stages of development requires a wide transcriptome coverage in different tissues/cell types of the developing organs as well as studies within Arabidopsis accessions with remarkable phenotypical differences. Thus, the general aim of this PhD project was to investigate and understand the molecular mechanisms underlying the big phenotype in the Arabidopsis accession Bur-0. To do so, the following main objectives were defined:

- Determine whether rosette size, seed size, flowering time and SAM size are generally correlated traits in different *A. thaliana* natural accessions including Bur-0 and mutant lines and identify possible marker traits for accession-specific phenotypes by a detailed phenotypical characterization during embryonic and postembryonic development as well as in different photoperiods,

- Investigate the extent to which the physiological status might contribute to the big phenotype observed in Bur-0, by a detailed physiological characterization in different tissues/stages during embryonic and postembryonic development, comparing Bur-0 to other accessions,

- Determine whether the enlarged organs observed in Bur-0 are determined by differences in cell size/number, ploidy level and/or expression of cell cycle regulators in different tissues/cell types during embryonic and postembryonic development, comparing Bur-0 to Col-0,

- Identify accession-specific genetic determinants of plant phenotypes, shared across tissues and developmental stages during embryonic and postembryonic growth by RNA-seq analysis of embryos at torpedo and mature stage, as well as SAMs at vegetative and floral transition stage from Bur-0 and Col-0. Moreover, identification of potential candidate genes and further validation of transcriptome data by expression analyses of candidate genes as well as known key regulators of organ size and growth during embryonic and postembryonic development.

2. MATERIAL AND METHODS

2.1. Growth conditions

Seeds were sown on soil and stratified at 4°C in dark for two days in 6 cm diameter pots filled with 3:1 soil: vermiculite substrate. Afterwards, pots were transferred to controlled growth chambers at the University of Potsdam, unless otherwise indicated. Plants were grown in long day (LD) (16 h light/8 h dark) (Percival AR-36L2, CLF PlantClimatics) and in short day (SD) photoperiods (8 h light/16 h dark) at 22°C with photosynthetically active radiation of 160 µmol m^{-2} s^{-1} at the plant level (Fitotron® SGC 120, Weiss Technik UK Ltd). Stratification treatment was not performed in SD. The age of the plants was recorded as days after germination (DAG), considering one DAG as the fourth day after starting the experiment in the growth chamber. After 4 DAG plants were thinned to three plants per pot until 14 DAG; afterwards, plants were thinned to one individual per pot. Growth conditions used for phenotyping experiments using Phytotyping^{4D} imaging system are described in *Section 2.4.1*.

2.2. Seed stocks

Seeds from the *Arabidopsis thaliana* accessions Col-0, L*er*-1, Ws-2, Bur-0 and from previously described Arabidopsis late flowering mutant lines *tsf-1*(Michaels *et al*., 2005), *ft-10* (Yoo *et al*., 2005), *soc1-6* (Alonso *et al*., 2003) and *fd-3* (Abe *et al*., 2005) were provided by Dr. Justyna Olas (University of Potsdam, in-house collection). The mutant lines were in Col-0 background. Additional Arabidopsis accessions Lip-0, Sei-0, Ts-1, Cen-0, Sap-0, Alst-1 and Ang-0 were provided by Dr. Corina Fusari (Max Planck Institute of Molecular Plant Physiology, in-house collection).

2.3. Embryo stage determination

Progression of embryo development was analyzed over time (days after pollination, DAP) for the Arabidopsis accessions Col-0 and Bur-0. Plants were hand pollinated and properly developed siliques were harvested at 5, 6, 7, 8, 10 and 12 DAP ($n \geq 4$). Ovules were extracted and cleared with chloral hydrate for Nomarski imaging as described by Figueiredo *et al*. (2016) and cleared ovules were mounted on glass slides to identify embryo stages using a Differential Interference Contrast (DIC) Microscope Zeiss Axioimager M2 (Carl Zeiss Microscopy, Germany).

2.4. Phenotype analysis

2.4.1. Analysis of rosette size, hyponasty, relative expansion rate (RER) and leaf initiation rate (LIR)

Rosette area of various Arabidopsis accessions and flowering time mutant lines grown in LD and SD photoperiods was analyzed from 4 to 14 DAG, by measuring the rosette surface area using *ImageJ* software (Schneider *et al*., 2012) (n=30). Additionally, 3D area, hyponasty and

relative expansion rate (RER) of the Arabidopsis accessions Col-0, L*er*-1, Ws-2 and Bur-0 were analyzed by Dr. Federico Apelt from the MPI of Molecular Plant Physiology, using the light-field camera imaging system Phytotyping^{4D} (*n*=8) as described by Apelt *et al.* (2015). Plants were grown in a growth chamber (model E-36L; Percival Scientific Inc) under LD, SD and neutral day (ND; 12h light/12 dark) photoperiods, with photosynthetically active radiation 160 µmol photons m^{-2} sec^{-1} at the plant level. Plants were imaged from 14 to 28 days after sowing (DAS). Afterwards, high-resolution images provided by Dr. Federico Apelt were used to identify the appearance of newly visible leaves and calculate the leaf initiation rate (LIR) per ecotype/photoperiod by counting the leaves produced every three-days.

2.4.2. Flowering time analysis

Flowering time was determined based on days to bolting (DTB = day on which the first flower buds are visible, and the main stem has approx. 0.5 cm high) for plants grown in LD and SD photoperiods (*n*=30). Additionally, flowering time was determined based on total leaf number (TLN) as previously described by Olas *et al*. (2019), for plants grown in LD photoperiod (*n*≥10).

2.4.3. Vegetative phase change analysis

The vegetative phase change was analyzed by identifying trichomes on the abaxial leaf surface (lower side of the leaf). Juvenile leaves were defined as rosette leaves without abaxial trichomes and adults leaves as rosette leaves with at least one trichome on the adaxial side of the leaf, respectively (Telfer *et al*., 1997). Juvenile leaf number (JLN) was determined for the Arabidopsis accessions Col-0, L*er*-1, Ws-2 and Bur-0, grown in LD photoperiod (*n*=15).

2.4.4. Seed analyses

2.4.4.1. Seed germination

Seed germination was analyzed in LD photoperiod for the Arabidopsis accessions Col-0, L*er*-1, Ws-2 and Bur-0. Seeds were sterilized according to the protocol provided by Dr. Justyna Olas. Briefly, seeds were immersed in a sterilization solution containing 5 mL 70% EtOH with Triton X-100, mixed by inverting for 10 min, placed on a sterile filter paper and dried under a laminar flow hood. Sterile seeds were sowed on Petri plates with Murashige-Skoog (MS) medium in three replicates, each containing 100 seeds and stratified at 4°C in the dark for two days. Seeds were considered germinated when the radicle emerged and reached 2 mm in length. Germination parameters like final germination percentage (FGP), mean time to germination (MTG) and germination index (GI) were determined as described by Moreno *et al*. (2018).

2.4.4.2.Seed yield

Seed yield was determined for the Arabidopsis accessions Col-0, L*er*-1, Ws-2 and Bur-0 grown in LD photoperiod. Single plants were bag packed when siliques started to turn brown to yellow and watering was stopped. Plants were left to dry out and paper bags were collected. Seed were

released pressing the bags gently, passed through a mesh, collected into 1.5 mL Eppendorf® tubes (Eppendorf AG, Germany) and stored at 4°C until use. Total seed weight produced per plant, seed weight (as weight of 100 seeds) and total seed number produced per plant were determined as yield parameters (n = 6).

2.4.4.3.Seed size

Dry seeds from the Arabidopsis accessions Col-0, L*er*-1, Ws-2 and Bur-0 were imaged under a stereoscope equipped with a digital camera in three replicates ($n \geq 20$ seeds per replicate). Seed area was analyzed using *ImageJ* software (Schneider *et al*., 2012).

2.5. *Physiological analysis*

2.5.1. Shoot biomass

Shoots from the Arabidopsis accessions Col-0, L*er*-1, Ws-2 and Bur-0 were harvested from plants grown in LD and SD photoperiod over time, from early vegetative stage (3 DAG) until late reproductive stage (3 days after floral transition in LD, 10 days after floral transition in SD respectively). Samples were harvested in time intervals of 3 days in LD and 10 days in SD photoperiod.

Shoots were harvested and fresh samples were individually weighted and packed in aluminum foil (previously weighted). Afterwards, packed samples were dried at 55°C for one week and dry shoot weight was estimated by subtracting the weight of the aluminum foil from the total weight of the packed sample. Biomass based relative growth rate (RGR) was calculated as described by Bai-Han *et al*. (2018):

$$RGR = \frac{(\ln w2 - \ln w1)}{(t2 - t1)}$$

where w1 was the biomass at harvest time 1 (t1), w2 at t2, respectively, and (t2 – t1) was the time frame analyzed (n=10).

2.5.2. Metabolite content

Metabolite content (protein, soluble sugars and starch) was measured from dissected embryos at late torpedo and mature stage, dry seeds, and whole rosettes of Col-0 and Bur-0 plants grown in LD photoperiod. Briefly, rosette samples were harvested at times points corresponding to vegetative stage (4 DAG for both accessions) and floral transition stage per accession (10 DAG for Col-0 and 21 DAG for Bur-0), as well as over time between 4 and 14 DAG. Rosette and dry seed samples were harvested into screw cap microtubes, in three biological replicates (at least 10 rosettes each and 100 mg of seeds each, respectively) and kept in liquid nitrogen. Frozen rosette and seed tissues were grinded into fine powder in a bead beater with sterile nuclease-free beads and kept at -80°C until use. Before use, aliquots were transferred to new screw cap microtubes.

Embryos at late torpedo and mature stages were individually hand-dissected under a stereomicroscope and collected in RNAlater (Thermo Fisher Scientific, USA). After hand pollination, siliques at 8 and 10 DAP were harvested, placed in a petri dish and opened under a dissection microscope by peeling off the valves using micro-dissection forceps. Exposed ovules were carefully removed and each ovule was gently squeezed to push the embryo out. Each embryo was collected with a syringe needle and rapidly transferred to 100 µL of RNAlater (Thermo Fisher Scientific, USA). Each biological replicate contained approximately 250 dissected embryos (n=3). Samples were kept at 4°C until use. Before use, RNAlater was carefully removed by pipetting and embryos were smashed with a pestle.

Metabolites were measured in collaboration with Dr. Maria Grazia Annunziata from the MPI of Molecular Plant Physiology. Briefly, ethanol extractions were made from powdered seeds, rosettes and smashed embryos. After centrifugation steps, supernatants were used for enzymatic detection of soluble sugars (glucose, fructose and sucrose) according to Stitt *et al*. (1989) and pellets were used for starch enzymatic detection (Hendriks *et al*., 2003) and total protein estimation using the Bradford (1976) method with bovine serum albumin as standard. Spectrophotometric/enzymatic assays were performed in 96-well micro plates and absorbance was determined using a micro plate reader. Two technical replicates were measured per biological replicate ($n \geq 3$).

2.6. Inter-accession crosses

Cross pollination between Bur-0 and L*er*-1, Col-0, Ws-2 was performed, respectively. After hand cross pollination seeds from developed siliques were harvested, sowed and plants were grown in the same conditions as parental plants under LD photoperiod. The offspring´s seed phenotype was analyzed on F1.

2.7. Histological and morphological analyses

Plant tissue used for histological analyses included shoot apices, siliques, leaves, inflorescences and stems at different stages of development and from different Arabidopsis accessions. Samples were harvested and immersed in 10 mL of FAA-fixative solution (5 mL EtOH, 0.5 mL acetic acid, 1 mL formaldehyde and 3.5 mL of DEPC water), freshly prepared before use in 50 mL falcon tubes. Shoot apices were first collected and placed on a petri dish to cut the oldest leaves and dissect the oldest leaf primordia using scalpel blades and micro-dissection forceps prior immersion in the FAA-fixative solution, while every other plant tissue was directly harvested in the FAA-fixative solution without intermediate dissection steps.

Afterwards, samples were transferred to embedding cassettes and placed into a tissue procesor machine Leica ASP200S (Leica, Germany) for fixation overnight. On the next day, samples were transferred to a heated paraffin embedding station HistoCore Arcadia (Leica, Germany), embedded into wax blocks, and sectioned (8 µm thickness) using a rotary microtome Leica RM2255 (Leica, Germany). Tissue sections were placed on Menzel polysine-coated glass slides (Thermo Scientific Gerhard Menzel GmbH, Germany) and incubated overnight at 42°C on a heating block. Afterwards, the slides were stored at 4°C until use.

2.7.1. Staining

Before staining, tissue sections were dewaxed and re-hydrated by dipping and incubating the slides in the following solutions: Histoclear (2 times, 10 min each), 100 % EtOH (2 times, 2 min each), 95% EtOH (1 min), 90% EtOH (1 min), 80% EtOH (1 min), 60% EtOH + 0.75% NaCl (1 min), 30% EtOH + 0.75% NaCl (1 min), 0.75% NaCl (1 min), PBS 1x (1 min). Afterwards, samples were stained as follows:

2.7.1.1. Toluidine blue – morphological staining

After dewaxing, slides were left to dry in a heating block at 42°C, then toluidine blue/sodium borate solution 0.01% was applied on top making sure that tissue sections were fully covered, and then incubated for 5-7 min in a heating block at 42°C. Slides were gently washed with water, incubated for 2 min in 80% EtOH and dried at RT. Next, a few drops of glycerol were aplied, samples were covered with cover slides and imaged with a Nikon Eclipse E600 bright field-DIC Microscope (Nikon, Japan). Toluidine blue staining was done for longitudinal tissue sections of shoot apices and siliques. Image analysis was done using *ImageJ* software (Schneider *et al*., 2012).

2.7.1.2.Calcofluor white – fluorescent cell wall staining

After dewaxing , excess PBS 1x was gently removed and 100 µL of Calcofluor white solution (100 µL Calcofluor white (1 mg/mL) added to 1 mL of water) was applied on top, slides were covered with cover slides and incubated in the dark at RT for at least 15 min. Next, cover slides were carefully removed, slides were shortly dipped in PBS 1x and covered again with clean cover slides. Calcofluor white staining was done for longitudinal tissue sections of shoot apices and siliques. Samples were imaged with a fluorescent-DIC Microscope Zeiss Axioimager M2 (Carl Zeiss Microscopy, Germany), using the default settings for Calcofluor white imaging. Image analysis was done using *ImageJ* software (Schneider *et al*., 2012).

2.7.1.3. DAPI – fluorescent nuclei staining

After dewaxing , excess of PBS 1x was gently removed and 100 µL DAPI solution (958 µL of 0.1 M Phosphate buffer (pH 7), 40 µL of 25 mM EDTA, 1 µL of Triton X-100, 1 µL of DAPI (1 mg/mL)) was applied on top, slides were covered with cover slides and incubated in the dark at RT for at least 30 min. Staining solution was not removed before imaging. DAPI nuclei staining was done for longitudinal tissue sections of shoot apices and siliques, but also for fresh harvested pollen grains. Briefly, pollen grains were stained as follows: flowers were fresh harvested, petals, sepals were removed, and anthers were incubated in 1.5 mL Eppendorf ® tubes (Eppendorf AG, Germany) with 300 µL of DAPI solution for at least 30 min at RT in the dark. After incubation, pollen grains were collected by pipetting, mounted on glass slides and covered with cover slides. All DAPI stained samples were imaged with a a fluorescent-DIC Microscope Zeiss Axioimager M2 (Carl Zeiss Microscopy, Germany), using the default settings for DAPI imaging and image analysis was done using *ImageJ* software (Schneider *et al*., 2012).

2.8. Ploidy level analysis by fluorescence activated cell sorting (FACS)

Plant tissue used for ploidy level analysis by fluorescence activated cell sorting (FACS) included shoot apices, leaves, and dissected embryos from the Arabidopsis accessions Col-0 and Bur-0 grown in LD photoperiod. Ploidy level was measured in collaboration with Dr. Frank Machin from the MPI of Molecular Plant Physiology. Nuclear isolation was done according to the protocol provided by Dr. Frank Machin as follows: the desired volume of working solution was prepared fresh before use, considering that 1 mL contains 5 µL RNAse A (10 mg/mL), 5 µL of SYTO13 red and 990 µL of Galbraiths Buffer (45 mM $MgCl_2$, 20 mM MOPS, 30 mM sodium citrate, 0.1% (volume/volume) Triton X-100, pH adjusted to 7.0 with 1 M NaOH).

Embryos at mature stage were individually hand-dissected under a stereomicroscope and collected in RNAlater (Thermo Fisher Scientific, USA). After hand pollination, siliques at 10 DAP were harvested, placed in a petri dish and opened under a dissection microscope by peeling off the valves using micro-dissection forceps. Exposed ovules were carefully removed and each ovule was gently squeezed to push the embryo out. Each embryo was collected with a syringe needle, rapidly transferred to a screw cap microtube with 100 µL of RNAlater (Thermo Fisher Scientific, USA) and samples were kept at 4°C until use. Shoot apices at vegetative (4 DAG for both accessions) and floral transition (10 DAG for Col-0 and 21 DAG for Bur-0) stages were collected and placed on a petri dish in order to remove the oldest leaves and dissect the oldest leaf primordia using scalpel blades and micro-dissection forceps. Rosette leaves from plants at 12 DAG were fresh harvested before use.

Leaf and shoot apex tissue samples were placed on small Petri dishes (5 cm) containing 1 mL of working solution and chopped with a sharp, flat-edge razor blade as fine as possible. For embryo samples, RNAlater was carefully removed by pipetting, 1 mL of working solution was added into the screw cap microtube and embryos were smashed with a pestle. Each tissue/working solution mixture was collected by pipetting and filtered through a 20 µm CellTrics® Filter (Partec, USA), suspended nuclei were collected in 2 mL Eppendorf tubes and kept on ice. Samples were transferred to glass tubes and loaded on a Flow Cytometry Cell Sorter FACS Aria II BD™ (Becton Dickinson, USA). Base ploidy level was determined by presence of 2n and 4n peaks per accession/tissue, respectively.

2.9. Transcript analyses

2.9.1. RNA isolation and cDNA synthesis

Embryo and SAM tissue at different developmental stages was harvested from Col-0 and Bur-0 plants grown in LD photoperiod. Embryos at late torpedo and green mature stages were individually hand-dissected under a stereomicroscope and collected in RNAlater (Thermo Fisher Scientific, USA). After hand pollination, siliques at 8 and 10 DAP were harvested, placed in a petri dish and opened under a dissection microscope by peeling off the valves using micro-dissection forceps. Exposed ovules were carefully removed and each ovule was gently squeezed to push the embryo out. Each embryo was collected with a syringe needle and rapidly transferred to 100 µL of RNAlater (Thermo Fisher Scientific, USA). Each biological replicate contained approximately 250 dissected embryos (n=3). Samples were kept at 4°C until use.

Before use, RNAlater was carefully removed by pipetting and embryos were smashed with a pestle.

SAM tissue was harvested from plants at vegetative (4 DAG for both accessions) and at floral transition (10 DAG for Col-0 and 21 DAG for Bur-0) stages. First, shoot apices were harvested, the youngest possible leaf primordia were removed and the SAM was manually excised using scalpel blades under a dissection microscope. Dissected SAM tissue was immediately transferred to a screw cap microtube and kept in liquid nitrogen. Samples were harvested in three biological replicates, with ≥ 25 dissected SAMs each. Frozen SAM tissue was grinded into fine powder in a bead beater with sterile nuclease-free beads and kept at -80°C until use.

Total RNA from homogenized material was isolated using mirVana™ miRNA Isolation Kit (Thermo Fisher Scientific, USA). Briefly, 500 µL of Lysis/Binding buffer was added per sample, tubes were vortexed vigorously to obtain a homogeneous lysate, then 50 µL of miRNA Homogenate Additive was added to the lysate, samples were mixed by vortexing and incubated on ice for 10 min. For the organic extraction 500 µL of acid-phenol:chloroform was added, mixed by vortexing and samples were centrifugated at RT for 10 min at maximum speed.

After centrifugation, the aqueous phase (upper layer) was transferred to new tubes (volume removed was noted) and 100% EtOH was added in a volume corresponding to 1.25 volume of the removed aqueous phase. The final RNA isolation and washing steps were done according to the manufacturer specifications and for embryo and SAM tissue samples, total RNA was collected in a final elution volume between 30-50 µL nuclease-free water. Total RNA quality and concentration were determined using a NanoDrop 2000® Spectrophotometer (Thermo Scientific, USA) and RNA integrity was verified by gel electrophoresis.

DNA digestion and cDNA synthesis were performed according to the protocol provided by Dr. Justyna Olas. Digestion was done using Turbo DNA-free DNase I kit (Ambion/Life Technologies, Germany). Total RNA was diluted to a final concentration of 5 µg in 20.5 µL sterile distilled water, then 2.5 µL of 10x TURBO DNase™ Buffer, 1 µL of TURBO DNase™ Enzyme and 1 µL sterile distilled water were added. Samples were incubated at 37°C for 1 h, after incubation 2.5 µL of inactivation reagent was added. Samples were incubated at RT for 2 min, mixed by inverting and centrifuged at 14000 rpm for 10 min at 22°C.

After centrifugation, the supernatant was transferred to a clean safe lock Eppendorf tube. cDNA synthesis was performed using SuperScriptTMIII Reverse Transcriptase Kit (Invitrogen/Life Technologies, Germany). Briefly, to digested RNA 2 µL of oligo-$(dT)_{18}$ and 2 µL of 10 mM dNTP were added. Afterwards, samples were incubated 65°C for 5 min. After incubation, samples were kept on ice for 1 min and shortly centrifuged at 4°C. After centrifugation, 8 µL of 5x Buffer, 2 µL of 0.1 M DTT, and 2 µL of SuperScript™ III Reverse Transcriptase enzyme were added. Samples were gently mixed by pipetting and incubated at 50°C for 50 min, followed by 55°C for 10 min and 70°C for 15 min. After incubation, samples were kept on ice for 10 min and stored at -80°C until use. The cDNA quality was analyzed by qRT-PCR using primers for the 3´ and 5´ regions of GAPDH (primers sequences provided in Supplementary Table S1).

2.9.2. Transcriptome analysis RNA-sequencing (RNA-seq)

For RNA-seq analysis, total RNA was isolated from Col-0 and Bur-0 embryo and SAM tissue at different developmental stages as described in *Section 2.9.1* Library preparation and sequencing to generate paired-end (2 × 150 bp) reads was performed by BGI Tech Solutions Co., Ltd, Hong Kong. RNA-seq data processing and statistical analysis was performed by Dr. Federico Apelt and Dr. Saurabh Gupta from the MPI of Molecular Plant Physiology. STAR (version 2.5.2b; Dobin *et al*., 2013) was used to align the reads to the TAIR10 annotation of the genome of *Arabidopsis thaliana* and the expression was quantified per gene using HTSeq (version 0.9.1; Anders *et al*., 2015). Pearson correlation was calculated on DESeq2 normalized counts (variance stabilizing transformation) and distance measure was used to cluster the samples using pheatmap in R (Kolde, 2015) and for PCA analysis using the plotPCA function from the DESeq2 package (Love *et al*., 2014).

Differentially expressed genes (DEGs) data sets were generated for significantly changed genes (log2 fold change >= 1 & FDR < 0.05), up and down regulated, identified at different stages, tissues and accessions during embryonic and postembryonic development with DESeq2. The DEGs were clustered into 30 k-means clusters using pheatmap. The Venn diagrams were generated using Bioinformatics & Evolutionary Genomics http://bioinformatics.psb.ugent.be/webtools/Venn/. and modified manually. Analysis of Gene ontology (GO) was performed using PANTHER16.0– Gene list analysis http://pantherdb.org/.

2.9.3. Quantitative real-time PCR (qRT-PCR) analysis

qRT-PCR was performed according to the protocol provided by Dr. Justyna Olas. A CFX Connect Real-Time PCR system (Bio-Rad, California, USA) for 96 well PCR plates was used. qRT-PCR was performed using SYBR Green PCR Master Mix (Applied Biosystems/Life Technologies) in 10 µL final reaction volume per well (6 µL of SYBR Green mixed with cDNA and 4 µL of primers at 0.5 µM working concentration) in three biological replicates, with three technical replicates each.

Oligonucleotides used for the qRT-PCR measurement were commercially synthesized (Eurofins Genomics GmbH, Germany) or provided by Dr. Justyna Olas. The primers sequences used for qRT-PCR are described in the Supplementary Table S1. Primers were used in final 0.5 µM concentration. Relative quantification of gene expression in different tissues and accessions was performed using the $2^{-\triangle Cq}$ method as described by (Wang *et al*., 2018). Expression level for each target gene was normalized to the expression level of the reference gene *TUBULIN 2* (*TUB2,* AT5G62690) per accession and tissue, respectively.

2.9.4. RNA in situ hybridization

RNA *in situ* hybridization was performed according to the protocol provided by Dr. Justyna Olas. In brief, cloning, probe synthesis and *in situ* hybridization steps were carried out as follows:

2.9.4.1.Cloning for RNA in situ hybridization

2.9.4.1.1. Oligonucleotides and PCR

Oligonucleotides used for PCR reactions were commercially synthesized (Eurofins Genomics GmbH, Germany) and a list of the primers is presented in Supplementary Table S1. PCR reactions were performed in the thermocyclers Eppendorf Mastercycler (Eppendorf AG, Germany) or peQSTAR Thermocycler 96 HPL Gradient (PEQLAB Biotechnologie GmbH, Germany), using *Pfu* High-Fidelity DNA Polymerase (#EP0572) (Thermo Scientific, USA) and *Taq* DNA Polymerase (#M0267L) (New England BioLabs GmbH, Germany). The general thermal profile for the PCR reactions is presented below and extension time and annealing temperature were adjusted according to primers features and amplicon length.

Steps	**Temperature**	**Time**
Step 1: Initial denaturation	98°C	40 sec
Step 2: 40 cycles	98°C	20 sec
	Annealing T°C primers	30 sec
	72°C	1 min /kb
Step 3: Final extension	72°C	10 min
Step 4: Hold	4°C	

Aliquots of the PCR product were used to verify the fragment size/specificity by gel electrophoresis on 1% agarose gel Agarose Basic (#A8963,0500) (AppliChem GmbH, Germany) and total PCR product was purified using a Promega Wizard Kit (Promega, USA).

2.9.4.1.2. A-tailing

A-tailing was performed by mixing 7 µL of purified PCR product with 1 µL of dATP (25 mM), 1 µL of *Taq* DNA polymerase and 1 µL of ThermoPol Reaction Buffer *Taq* (#B9004S) (New England BioLabs GmbH, Germany). Samples were incubated for 30 min at 72°C. Next, ligation using pGEM®-T Easy Vector (Promega, USA) was performed by mixing 3 µL of PCR A-tailing product with 5 µL of 2x Rapid Ligation buffer, 1 µL pGEM®-T Easy Vector and 1 µL of T4 DNA ligase. Samples were incubated at 4°C overnight or at RT for 1 h.

2.9.4.1.3. Transformation of Escherichia coli (DH5α)

Transformation into *E. coli* (*DH5α* competent strain) was done by adding 4 µL of A-tailed plasmid DNA to 50 µL of *E. coli* (DH5α) competent cells. Samples were gently mixed by pipetting and incubated on ice for 30 min, then in a water bath for 45 sec at 42°C and on ice for 2 min. Next, 450 µL of Luria-Bertani (LB) liquid media was added and samples were incubated at 37°C for 1 h. Afterwards the samples were centrifuged at 5000 rpm for 5 min, the supernatant was discarded, the pellet was resuspended in 150 µL of LB liquid media and spread on LB agar plates containing ampicillin (100 mg/mL H_2O), 4 µL isopropyl thiogalactoside (IPTG) and 40 µL 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal). Bacteria cultures were incubated at 37°C overnight.

2.9.4.1.4. Plasmid preparation

Next day, white bacteria colonies were selected for plasmid preparation, according to the following in-house protocol. Each colony was transferred using sterile toothpicks from the LB agar plate to glass tubes containing 5 mL of LB liquid media. The colonies were incubated at 37°C, shaking at 220 rpm overnight. After incubation, 1.5 mL of the liquid culture was transferred to a 1.5 mL Eppendorf tube and centrifuged at maximum speed for 30 sec. The remaining original culture was stored at 4°C. After centrifugation, the supernatants were discarded and the cell pellet was re-suspended in 200 µL of Solution I (50 glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0). Next, 200 µL of Solution II (0.2 N NaOH, 1% SDS) was added to the bacterial suspension and samples were mixed by inversion. Finally, 150 µL of Solution III (3 M Potassium acetate, pH 4.8 adjusted with glacial acetic acid) was added.

Samples were mixed by inversion several times and centrifugated at maximum speed for 15 min. The supernatant was transferred to a new tube and 400 µL of Isopropanol was added. Samples were mixed by inversion and centrifugated at maximum speed for 15 min. The supernatant was removed by pipetting. The pellet was mixed with 500 µL of 70% EtOH by vortexing, then samples were centrifugated at maximum speed for 15 min. Next, the supernatant was discarded, and samples were left to dry at 56°C to remove completely any EtOH traces. Dry pellets were re-dissolved in 30 µL in RNase-free water. Samples were stored at 4°C until use. The list of constructs generated during this study is presented in Supplementary Table S2.

Plasmid DNA was digested with the restriction enzyme *Eco*RI-HF (New England BioLabs GmbH, Germany) by mixing 3 µL of plasmid DNA, 2 µL of 10x digest buffer, 14.7 µL of sterile H_2O and 0.3 µL of *Eco*RI-HF restriction enzyme for a total reaction volume of 20 µL. Samples were incubated at 37°C for 30 min. The size of the product was verified by gel electrophoresis. Finally, samples with correct fragment size were purified with a PCR clean up kit (Promega, USA) and the final concentration was verified using a NanoDrop 2000® Spectrophotometer (Thermo Scientific, USA).

2.9.4.1.5. Probe synthesis

Probes were synthesized using a DIG RNA Labelling Kit (Roche, Germany) and probe synthesis was carried out by mixing 200 ng of cDNA template with 2.0 µL of 10x transcription buffer, 1.0 µL of RNase inhibitor, 2.0 µL of 10x NTP mix, and 2.0 µL of SP7 or T7 enzyme, respectively. Samples were mixed and incubated for 2 h at 37°C in a water bath. Next, 0.25 µL of RNAse free DNAse was added, samples were mixed by vortexing, shortly spin down and incubated for 15 min at 37°C in a water bath.

After incubation, samples were placed on ice and 1µL of EDTA 0.5 M was added, then samples were mixed and 2.5 µL of 4 M LiCl and 75.0 µL of 100% EtOH were added. Samples were mixed again and incubated at - 80°C for 1 h. Afterwards samples were centrifugated at 4°C at maximum speed for 30 min. After centrifugation, the supernatant was discarded, 200 µL of cold 80% EtOH was added to the pellet and samples were centrifugated again at 4°C for 10 min at 14000 rpm.

The supernatant was removed and the pellet was left to dry at RT. The pellet was re-dissolved in 100 µL of DEPC- water and mixed with 100 µL of carbonate buffer (80 mM $NaHCO_3$, 120 mM Na_2CO_3). Probe fragmentation was done by incubating the samples at 60°C in a water bath during the time calculated with the following formula:

$$\text{Time} = (\text{Li} - \text{Lf})/(\text{K} * \text{Li} * \text{Lf}),$$

where Li is the initial length of the probe in kb, Lf is the final length of the probe in kb and K=0.11 kb/min. After incubation, 20 µL of 10x neutralization buffer (10% acetic acid) was added and samples were mixed. Next, 1 µL of glycogen (20 mg/mL), 1 µL of 1 M $MgCl_2$ and 600 µL of 100% EtOH were added, samples were mixed again and incubated overnight at -20°C.

After incubation, samples were centrifugated at 4°C at maximum speed for 30 min, the supernatant was discarded and 200 µL of cold 80% EtOH was added to the pellet. Next, samples were centrifugated again at 4°C, for 10 min at 14000 rpm. The supernatant was removed, the pellet was left to dry at RT. Afterwards, the pellet was re-dissolved in 50 µL of DEPC-water. Next, 5 µL of RNA was transferred to a new tube to check probes fragmentation *via* gel electrophoresis, the remaining RNA was diluted in 450 µL hybridization buffer (deionized formamide, 50% dextrane sulphate, 10× in situ salts, 50×denhardt's, 50 mg/mL tRNA, 50 mM EDTA) and stored at -20°C until use.

2.9.4.1.6. RNA in situ hybridization

Slides containing histological sections of selected plant tissues were dewaxed and re-hydrated as described in *Section 2.6.1*. Then, slides were incubated in pre-warmed proteinase K solution (200 µL of proteinase K dilution buffer and 10.4 µL proteinase K (Roche, Germany)) for 30 min at 37°C, in a water bath. Next, the slides were incubated in 1x PBS for 5 min at RT and dehydrated in the following solutions during 30 sec each: 0.75% NaCl, 30% EtOH + 0.75% NaCl, 60% EtOH + 0.75% NaCl, 80% EtOH, 90% EtOH, 95% EtOH and two times in 100% EtOH. Afterwards, the slides were left to dry at RT for 1 h.

Hybridization was performed by mixing 1-2 µL of the probe with 150 µL of hybridization buffer per slide. The probes were activated by incubating them for 2 min in a water bath at 80°C. Afterwards, the probes were kept on ice, applied on each slide and covered with a cover slide avoiding bubbles. Next, samples were incubated in a humidified box with soaking solution (2x SSC (3 M NaCl, 0.3 M Na-citrate, pH 7.0) in 50% formamide) at 55°C overnight. The list of probes tested in this study is presented in the Supplementary Table S3.

After incubation, the cover slides were gently removed and slides were washed as follows: shortly in 2x SSC, four times in 0.2x SSC at 55°C for 30 min each, once in 0.2x SSC at 37°C for 5 min and two times in 0.2x SSC at RT for 5 min each. Then, slides were incubated in 1x PBS for 5 min at RT and afterwards, samples were incubated in blocking reagent solution (1% blocking reagent (Roche, Germany) in 1x TBS buffer (50 mM Tris pH 9.5 and 0.9% NaCl) and 0.1% Triton X-100 at RT for1 h. Afterwards, immunological detection was carried out by applying 150 µL of Anti-DIG antibody (Roche, Germany) diluted in blocking reagent (dilution

1:1250). Next, samples were covered with cover slides and incubated in a box humidified with DEPC-water at RT for 90 min.

Next, the cover slides were gently removed and slides were washed with blocking reagent solution four times, 30 min each at RT. Afterwards, the slides were washed with TNM-50 buffer (100 mM Tris pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$) two times, for 5 min each at RT. The colorimetric reaction was done by applying 250 µL per slide of NBT/BCIP solution (1:50 dilution of NBT/BCIP (Roche, Germany) in PVA-TNM-50 (10% polyvinyl alcohol (PVA) in TNM-50 buffer). Samples were incubated overnight in a box humidified with DEPC-water at RT in the dark. Samples were imaged with a Nikon Eclipse E600 bright field-DIC Microscope (Nikon, Japan) or with a DIC Microscope Olympus BX-61 (Olympus, Germany).

2.10. Quantification and statistical analyses

Statistical analyses were performed with Microsoft® Excel® version 2012 and R software (R Core Team, 2020). Significant differences between two samples were determined using two-way Student's *t*-test or for multiple comparisons using two-way ANOVA. Means were compared and grouped using Tukey–HSD test. Linear regressions and correlations coefficients were obtained using Microsoft® Excel®.

3. RESULTS

3.1. Phenotypical characterization during embryonic and postembryonic growth

3.1.1. Phenotypical characterization during postembryonic growth

Considering that the Arabidopsis accession Bur-0 has been reported as large rosette, large seed and late flowering time accession, we wanted to determine whether those plant traits are generally correlated in different Arabidopsis accessions and if so, identify factors that can determine accession-specific phenotypes. The rosette, seed and flowering time phenotypes of eleven accessions from different geographical origins were analyzed for plants grown in LD photoperiod including Bur-0, and linear regressions were performed to estimate correlations between plant traits.

Plant growth was further characterized in different photoperiods, focusing on the early, intermediate, and late flowering accessions Ws-2, L*er*-1, Col-0 and Bur-0. Rosette size analyses were initially conducted in LD and SD photoperiods from 2 to 12 DAG. For later time points, a more detailed analysis was performed in collaboration with Dr. Federico Apelt (Max Planck Institute of Molecular Plant Physiology) using a high-resolution phenotyping system (Phytotyping^{4D}) for accurate monitoring of spatio-temporal plant growth behavior under LD, SD and neutral day (ND) photoperiods. The 3D area, hyponasty and relative expansion rate (RER) were determined and images provided by Dr. Federico Apelt were used to identify the appearance of new leaves and calculate the leaf initiation rate (LIR) per accession/photoperiod.

Given that the accessions analyzed in this study have different flowering time phenotypes, we wanted to determine whether the accessions were being compared at the same stage of development. For this purpose, we were interested in a marker trait that allowed a precise differentiation of vegetative and reproductive stages, therefore a detailed morphological analysis of the SAM (shoot apical meristem) was conducted over time to identify the time point for floral transition initiation in each accession. The floral transition was further confirmed at the molecular level by expression analysis of the floral marker gene *APETALA 1 (AP1)* by RNA *in situ* hybridization for the accessions Ws-2, L*er*-1, Col-0 and Bur-0 grown in LD and SD photoperiods.

In order to determine correlations between SAM size and adult plant traits, linear regression analyses were performed between the traits: flowering time, seed area and age-dependent or developmental stage-dependent SAM size in eleven Arabidopsis accessions grown in LD. Further correlations between age-dependent or developmental stage-dependent SAM size and rosette area were analyzed for the accessions Ws-2, L*er*-1, Col-0 and Bur-0 grown in LD. Finally, we wanted to determine whether intermediate developmental stages were also reached at different time points during postembryonic growth, therefore the detailed phenotypical characterization was concluded with a developmental phase progression analysis, including germination, juvenile phase, and flowering time (as total leaf number, TLN), for the accessions Ws-2, L*er*-1, Col-0 and Bur-0 grown in LD photoperiod.

3.1.1.1. Rosette area, seed area and flowering time phenotypes

We analyzed the rosette, seed and flowering time phenotype in eleven accessions from different geographical origins: Ws-2, Lip-0, Sei-0, L*er*-1, Ts-1, Cen-0, Col-0, Sap-0, Alst-1, Ang-0 and Bur-0, reported as early, intermediate, and late flowering time accessions, respectively. Plants were grown in LD photoperiod, rosette area was analyzed over time (from two to 14 DAG), flowering time was determined as days to bolting (DTB = day on which the first flower buds are visible and the main stem has approx. 0.5 cm high) and average seed area was measured per accession.

We found that the average rosette area is significantly different ($p< 0.05$) among all accessions for each of the time points analyzed (Supplementary Table S4). At the last time point analyzed (14 DAG), the Arabidopsis accessions Sei-0, Lip-0, Ts-1 and Sap-0 have the biggest rosettes, Col-0, Cen-0, Ang-0 and Bur-0 have intermediate rosette area, while Ws-2, L*er*-1, and Alst-1 have the smallest rosettes (Figure 5A, B). Flowering time is significantly different among all accessions ($p< 0.05$), Bur-0 has the latest flowering time (45.7 days) and Ws-2 has the earliest flowering time (21.1 days) (Figure 5C). Seed area is also significantly different among the analyzed accessions ($p< 0.05$) and most accessions have an average seed area between 0.80 and 0.12 mm^2, except for Bur-0, which has significantly bigger seeds with an average area of 0.19 mm^2 (Figure 5E, F). No trend was observed among the plant traits analyzed, i.e., accessions with bigger rosettes at 14 DAG did not have a later flowering time, nor bigger seeds and not all accessions with small rosettes at 14 DAG have an early flowering phenotype, neither smaller seeds.

Although rosette area could not be estimated at later time points than 14 DAG using the same method (due to leaf overlapping, which prevented inclusion of all leaves and thus accurate estimation of total rosette area), we observed high variation in the rosette phenotype during late postembryonic growth. Some accessions having a small rosette phenotype at 14 DAG are still among the smallest at bolting time (for example L*er*-1). In contrast, some accessions having a big rosette phenotype at 14 DAG are among the smallest accessions at bolting time (for example Lip-0), while an ecotype like Bur-0 has intermediate rosette area at 14 DAG, but at bolting time has the biggest rosettes among the analyzed accessions (Figure 5D).

Linear regressions including the plant traits rosette area at 14 DAG, flowering time and seed area revealed no correlations among traits ($r < 0.8$ and $p > 0.05$) and linear regressions are broken by outliers (Figure 5G). The regression analysis was repeated grouping the phenotypical data sets according to the flowering phenotype (early, intermediate, late) and for early flowering accessions there are positive, but non-significant correlations between rosette area, seed area and flowering time. No further correlations were found with this approach and linear regressions are also broken by outliers (Supplementary Figure S1), indicating that rosette area, seed area and flowering time are not generally correlated traits in Arabidopsis. Our results suggest that rosette area, seed area and flowering time are independent traits influenced by other factors and none of them is an optimal predictor trait for the adult plant phenotype. Nevertheless, this analysis revealed how the accession Bur-0 stands out from the data distribution as outlier, with particularly larger seeds and later flowering time.

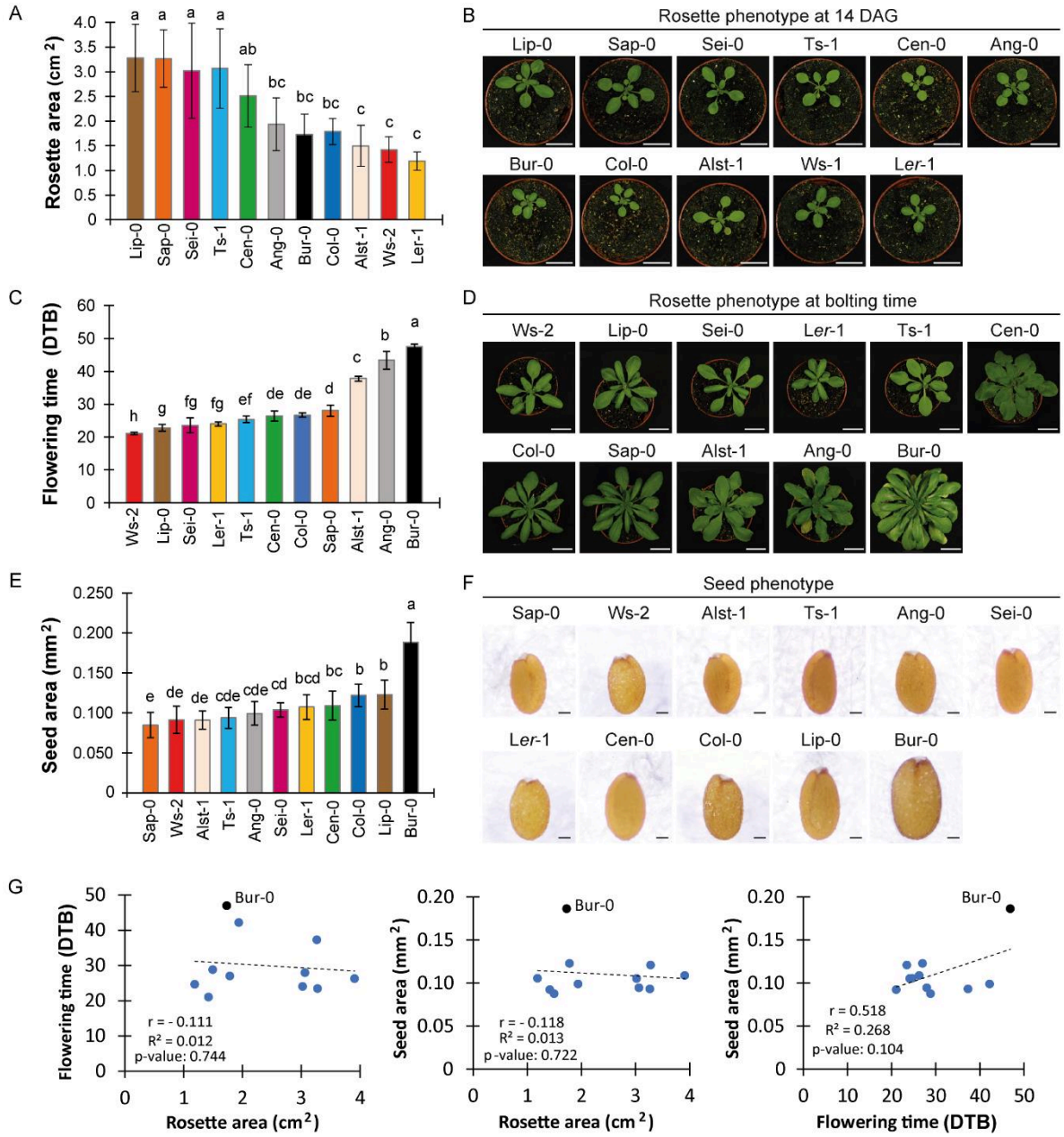


Figure 5. Rosette area, seed area, flowering time, and correlations between plant traits. Eleven *Arabidopsis thaliana* accessions were analyzed and plants were grown in long day photoperiod. (A) Rosette area at 14 days after germination (DAG). *n*=10. (B) Rosette phenotype at 14 DAG. Scale bar = 2 cm. (C) Flowering time as days to bolting (DTB), *n*=10. (D) Rosette phenotype at bolting time per accession. Scale bar = 2 cm. (E-F) Seed area and seed phenotype, *n*=18. Scale bar = 0.1 mm. Error bars indicate ± SD. (G) Linear regressions between flowering time, rosette area at 14 DAG and seed area. Statistical significance was tested with ANOVA and means were compared using Tukey–HSD test. Different letters indicate significant differences at α 0.05.

3.1.1.2. Growth and flowering time phenotypes in different photoperiods

In order to investigate if Bur-0 achieves its phenotype independently of the photoperiod, further analyses were performed focusing on the early, intermediate and late flowering accessions Ws-2, L*er*-1, Col-0 and Bur-0. The rosette size was measured over time (from two to 12 DAG) and flowering time was determined as days to bolting (DTB = day on which the first flower buds are visible and the main stem has approx. 0.5 cm high) for plants grown under LD and SD

photoperiods. Within the time frame analyzed (2 to 12 DAG) the rosette phenotype is smaller in SD than in LD for all accessions (Figure 6A, B) and the rosette area is significantly different among the analyzed accessions at each of the time points and photoperiods analyzed ($p< 0.05$) (Figure 6C, D). In LD photoperiod Ws-2 and Bur-0 have similar and bigger rosette size than Col-0 and L*er*-1, while in SD photoperiod Bur-0 rosettes are larger than the other accessions and Ws-2 rosette area is rather similar to Col-0 or L*er*-1. Moreover, the flowering time analysis demonstrated that Bur-0 flowers much later than the other accessions, both in LD and SD photoperiods (Figure 6E, F) and interestingly, a similar trend as observed in LD is followed in SD for the other accessions, where Ws-2 flowers first, followed by L*er*-1 and then Col-0. Thus, our results indicate that Bur-0 has a late flowering time phenotype in LD as well as in SD photoperiod.

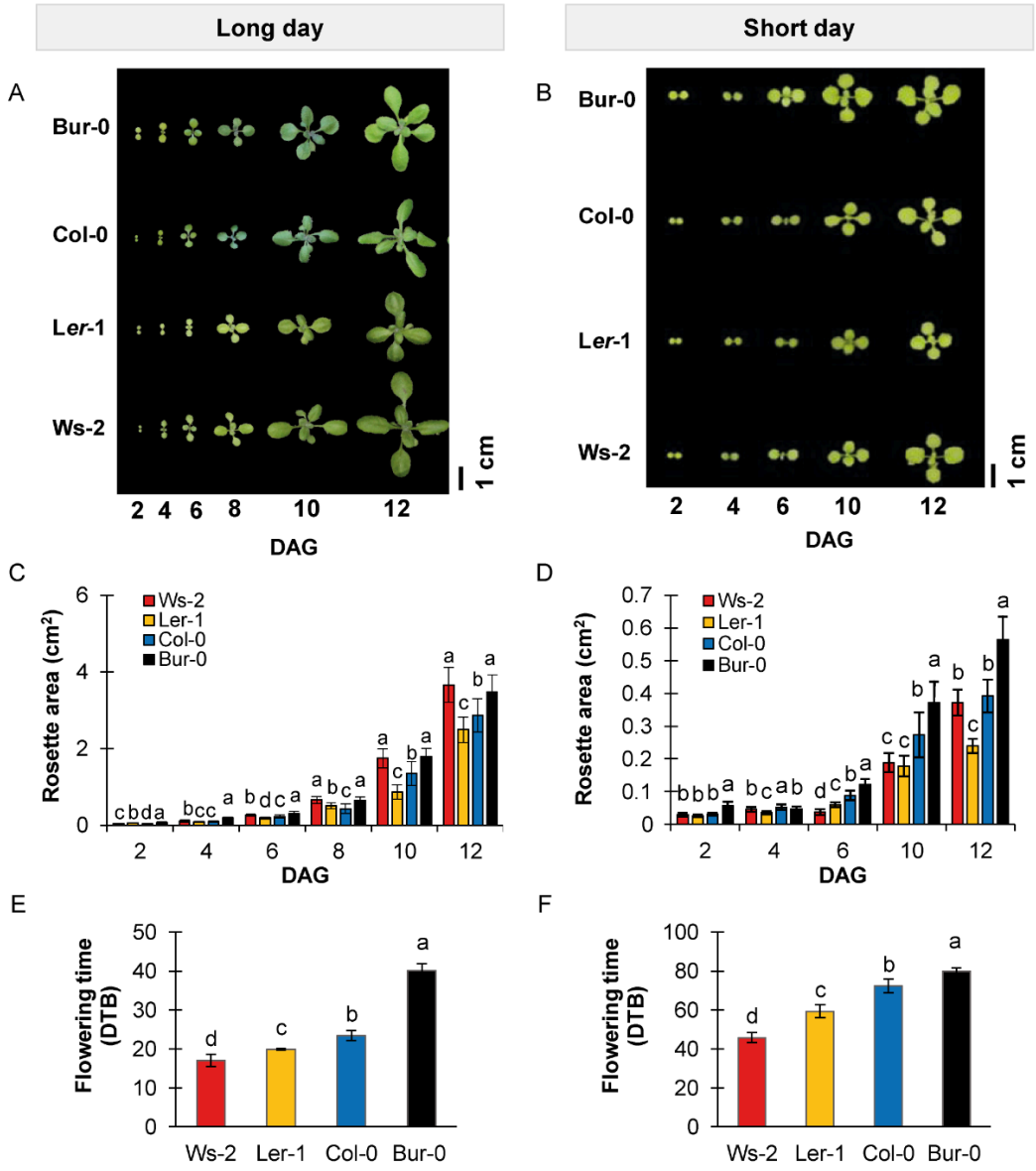


Figure 6. Rosette size and flowering time phenotypes analyzed in LD and SD photoperiods. The Arabidopsis accessions Bur-0, Col-0, L*er*-1 and Ws-2 were analyzed and time is given in days after germination (DAG). Rosette phenotype from two to 12 DAG in (A) LD and (B) SD photoperiods. Rosette area measured from two to 12 DAG in (C) LD and (D) SD photoperiods. *n*=30. Flowering time as days to bolting (DTB) in (E) LD and (F) SD photoperiods. *n*=15. Error bars indicate ± SD. Statistical significance was tested with ANOVA and means were compared with Tukey–HSD test. Different letters indicate significant differences at α 0.05.

Given that our previous results showed that a large rosette phenotype in Bur-0 was evident at bolting time in LD photoperiod (*Section 3.1.1.1.*), a more detailed growth analysis at later time points was performed under LD, SD and ND photoperiods in collaboration with Dr. Federico Apelt from the MPI of Molecular Plant Physiology, using an established high resolution 3D phenotyping system (Phytotyping^{4D}). The Arabidopsis accessions Ws-2, L*er*-1, Col-0 and Bur-0 were imaged during at least one week, starting from 14 days after sowing (DAS) and the results revealed that the total plant 3D surface area over time is bigger in Bur-0 under all photoperiods analyzed, followed by Ws-2, Col-0 and L*er*-1 (Figure 7A). These results confirmed that in LD, SD as well as in ND photoperiod, the big rosette phenotype in Bur-0 is achieved during late postembryonic growth.

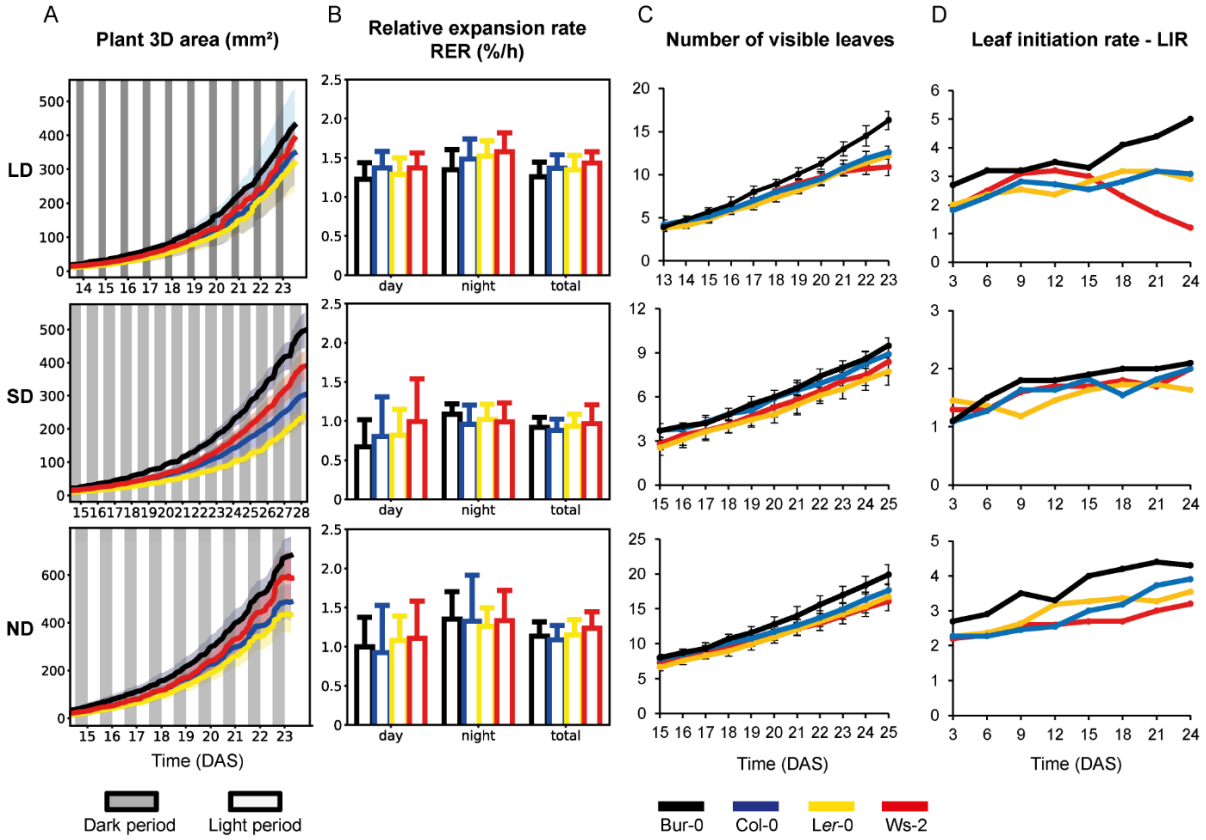


Figure 7. **Growth analysis during late postembryonic development under different photoperiods**. The Arabidopsis accessions Bur-0, Col-0, L*er*-1 and Ws-2 were analyzed using a high-resolution 3D phenotyping system and plants were grown in long day (LD), short day (SD), and neutral day (ND) photoperiods. (A) Total plant 3D surface area. Lines and shadows with the same color represent mean and standard deviation for each ecotype, respectively. (B) Relative expansion rate (RER). (C) Number of visible leaves. (D) Leaf initiation rate. $n \geq 7$. Error bars indicate ± SD. Time is given in days after sowing (DAS).

Moreover, the relative expansion rate (RER) is larger during the night for all accessions under all photoperiods and interestingly, the total RER is similar for all accessions per photoperiod, respectively, indicating similar expansion growth for all accessions (Figure 7B). These results suggest that the bigger plant 3D surface area observed in Bur-0 does not result from a higher RER. In order to determine if the bigger rosette area in Bur-0 might result from a faster leaf initiation rate (LIR), the high-resolution images from the Phytotyping^{4D} analysis provided by Dr. Federico Apelt were used to identify the appearance of newly visible leaves, counting the

leaves produced every three-day (Figure 7C), and calculate the leaf initiation rate (LIR) in each ecotype/photoperiod. We found that the LIR is similar for all accessions grown in SD photoperiod, while in LD and ND photoperiod Bur-0 LIR is higher (Figure 7D), indicating that Bur-0 produces leaves faster than the other accessions during late postembryonic growth. However, it should be noted that in LD photoperiod, at 20 DAS, Ws-2, Col-0 and L*er*-1 are at bolting stage (plants produced flowers but not rosette leaves), thus a plateau in LIR is observed for those accessions much earlier, starting at 18 DAS (Figure 7D).

In addition, the Phytotyping4D analysis provided information regarding the diurnal leaf movement (hyponasty), revealing that in LD conditions Bur-0 is more synchronized with Col-0 pattern, in SD conditions shows larger angles during the day and in the night and is more synchronized with Ws-2 pattern, while in ND conditions is more synchronized with L*er*-1. Overall, L*er*-1 has intermediate leaf movement pattern between Bur-0 and Col-0 and also has a slightly delayed leaf downwards movement towards the end of the night (Supplementary Figure S2).

3.1.1.3. SAM morphological characterization and floral transition analysis

A detailed morphological analysis of the SAM (shoot apical meristem) was done to identify the precise time point of floral transition in each accession. We performed toluidine blue staining on longitudinal sections of shoot apices from the Arabidopsis accessions Ws-2, Lip-0, Sei-0, L*er*-1, Ts-1, Cen-0, Col-0, Sap-0, Alst-1, Ang-0 and Bur-0 grown in LD conditions and harvested at 4, 6, 8, 10, 12, 14, 16, 18, 21 and 23 days after germination (DAG).

Histologically stained sections revealed morphological changes in the SAM from vegetative (flattened, narrow SAM) to floral transition (wider, rounded SAM) and reproductive stages (elongated SAM with floral primordia visible) (Figure 8). In addition, SAM width was measured as a parameter for SAM size estimation and our results showed that the maximal widening of the SAM takes place towards the same time point when the reproductive transition is morphologically visible; afterwards the SAM slightly shrinks (Supplementary Figure S3). The size changes of the SAM over time (maximum widening peak followed by SAM shrinkage) revealed that the floral transition stage is initiated at earlier, intermediate and later time points among the eleven accessions analyzed.

We identified the floral transition stage at 6 DAG for the accession Ws-2, 8 DAG for Sei-1, 10 DAG for L*er*-1 and Col-0, 12 DAG for Lip-0, 14 DAG for Ts-1, 16 DAG for Cen-0, 18 DAG for Alst-1 and Sap-0 and 21 DAG for Ang-0 and Bur-0 (Figure 8). The floral transition stage and time point was also morphologically identified for the early, intermediate, and late flowering accessions Ws-2, L*er*-1, Col-0 and Bur-0 plants grown in SD photoperiod. Toluidine blue staining on SAM longitudinal sections over time revealed that in SD the floral transition also takes place at different time points per accession and it is first initiated for Ws-2 between 20-25 DAG, followed by L*er*-1 between 30-35 DAG, then for Col-0 between 40-45 DAG and lastly for Bur-0 towards 45-50 DAG (Supplementary Figure S4).

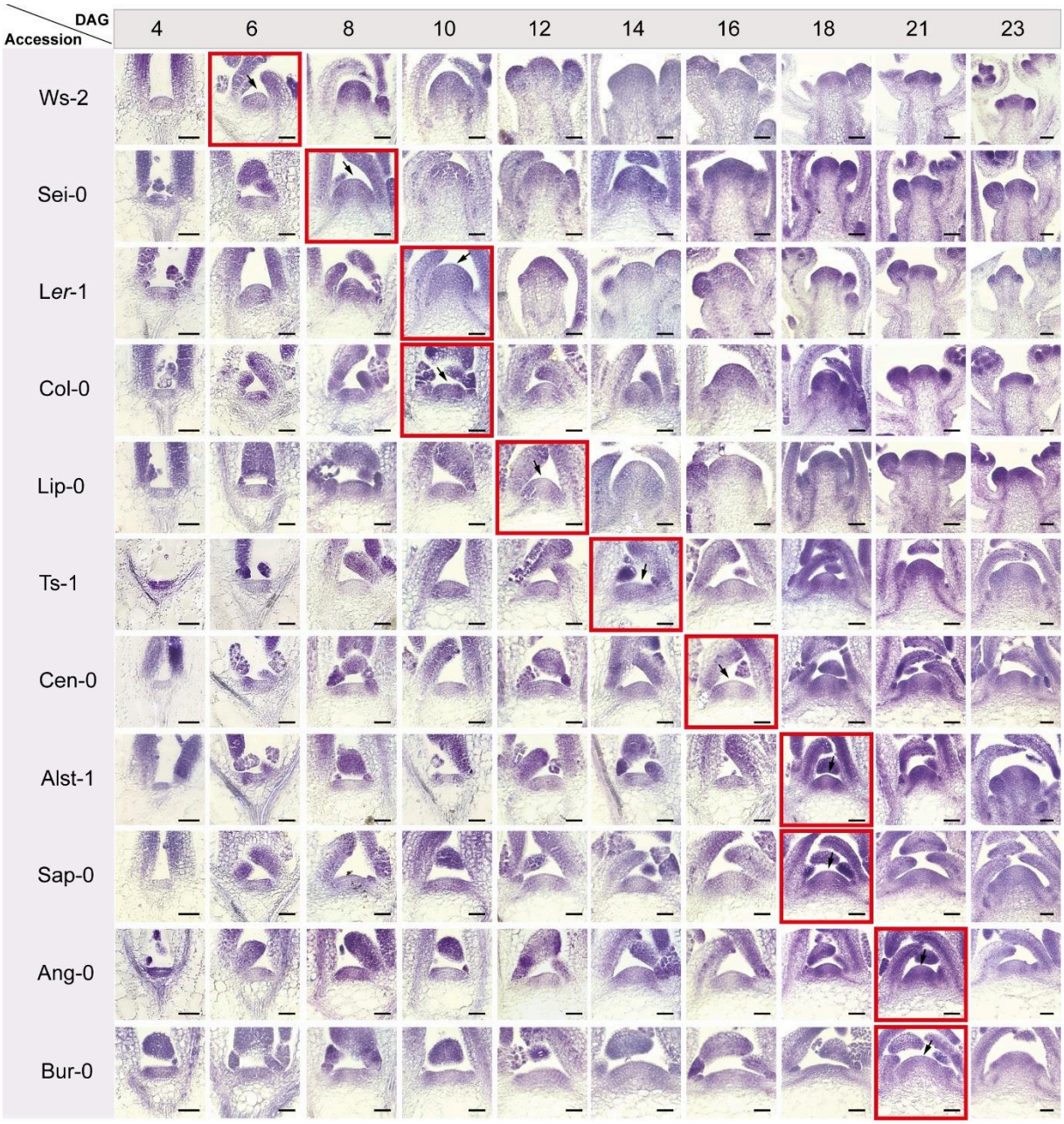


Figure 8. Morphological analysis of the shoot apical meristem (SAM). Eleven *Arabidopsis thaliana* accessions were analyzed and plants were grown in long day photoperiod. Longitudinal sections of shoot apices at different time points were stained with toluidine blue. Arrows indicate the SAM and red frames indicate the time point when the floral transition is initiated in each accession. Scale bar = 100µm. Time is given in days after germination (DAG).

The floral transition initiation was further confirmed at the molecular level using the floral marker gene *APETALA1* (*AP1*) for plants grown in LD and SD photoperiod from the accessions Ws-2, L*er*-1, Col-0 and Bur-0. Expression analysis of *AP1* by RNA *in situ* hybridization showed that *AP1* is expressed in floral primordia for all accessions and *AP1* signal is first visible in LD at 8 DAG for Ws-2, at 12 DAG for L*er*-1 and Col-0 and at 26 DAG for Bur-0, while in SD photoperiod *AP1* is only detected in Ws-2 and L*er*-1 samples at 35 DAG and 50 DAG, respectively, indicating that floral formation might take longer than 50 DAG for Col-0 and Bur-0 in SD (Figure 9A, B). Since *AP1* signal is absent at time points when the floral transition initiation was morphologically identified, but first visible in floral primordia at later time points,

these results confirm that the floral transition initiation does take place in each accession towards the time points previously identified in the SAM morphological characterization.

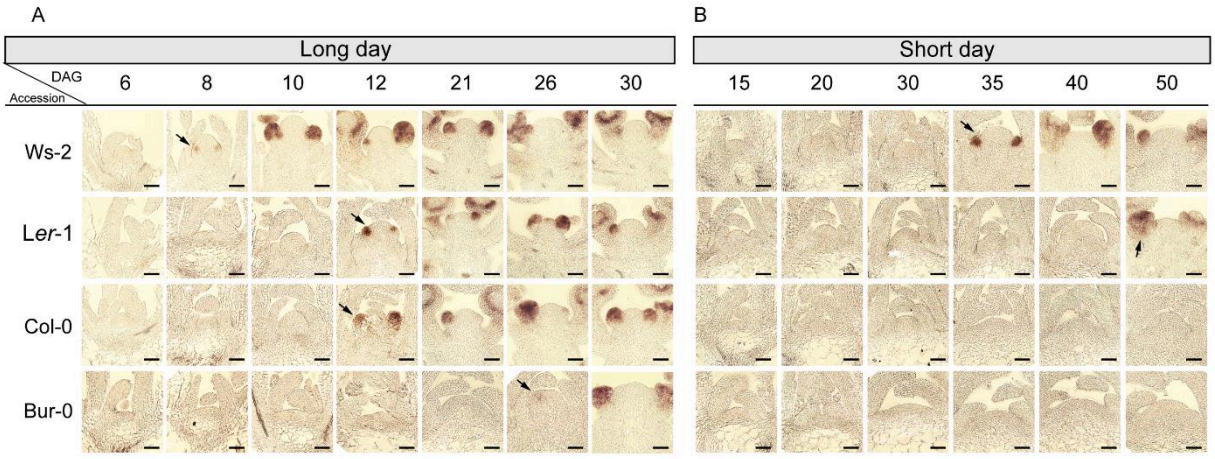


Figure 9. Expression analysis of the floral marker *APETALA1* (*AP1*). RNA *in situ* hybridization using *APETALA1* (*AP1*) as probe on shoot apex longitudinal sections from the *Arabidopsis thaliana* accessions Bur-0, Col-0, L*er*-1 and Ws-2 grown in (A) LD and (B) SD photoperiods. Arrows indicate earliest *AP1*expression signal. Scale bar = 100µm. Time is given in days after germination (DAG).

3.1.1.4. Correlations between SAM size and adult plant traits

SAM size (as the SAM width) was measured over time from four to 14 DAG, at vegetative stage (4DAG for all accessions) and at floral transition stage, respectively, in each ecotype (6 DAG for the ecotype Ws-2, 8 DAG for Sei-1, 10 DAG for L*er*-1 and Col-0, 12 DAG for Lip-0, 14 DAG for Ts-1, 16 DAG for Cen-0, 18 DAG for Alst-1 and Sap-0 and 21 DAG for Ang-0 and Bur-0). The results revealed that changes in SAM size over time are not uniform for all accessions, i.e., a similar size increase at 4, 6 and 8 DAG is observed for all accessions, but afterwards a separation of the data points indicates that at later time points (from 8 to 14 DAG), SAM size continues to increase for some accessions but decreases for others (Figure 10A). On the other hand, SAM size at vegetative and floral transition stages increases uniformly for all accessions (data points grouped), except for three data points that stand out as larger SAMs at floral transition stage (Figure 10B), which correspond to the late flowering accessions Alst-1, Ang-0 and Bur-0.

Linear regression analyses were performed to detect correlations between flowering time (FT), seed area (SA), and age-dependent or stage dependent SAM size. Our results revealed that SAM size at 14 DAG and FT are not correlated traits (Figure 10C), while the correlation between SAM size at 14 DAG and SA is strong and significant ($p< 0.05$) (Figure 10D). On the other hand, no correlations were confirmed between SAM size at vegetative stage and FT (Figure 10E), while SAM size at floral transition stage and FT are strongly and significantly correlated ($p< 0.05$) (Figure 10G). The correlations between SAM size (vegetative or floral transition stage) and SA are moderate and non-significant ($p> 0.05$) (Figure 10F, H).

Our results indicate that early flowering accessions tend to have smaller SAMs (narrower), late flowering accessions tend to have larger SAMs (wider) and interestingly, Bur-0 has the largest

SAM among the analyzed accessions. With this, our results demonstrate that flowering time is a good predictor of SAM size at floral transition stage in the eleven Arabidopsis accessions analyzed.

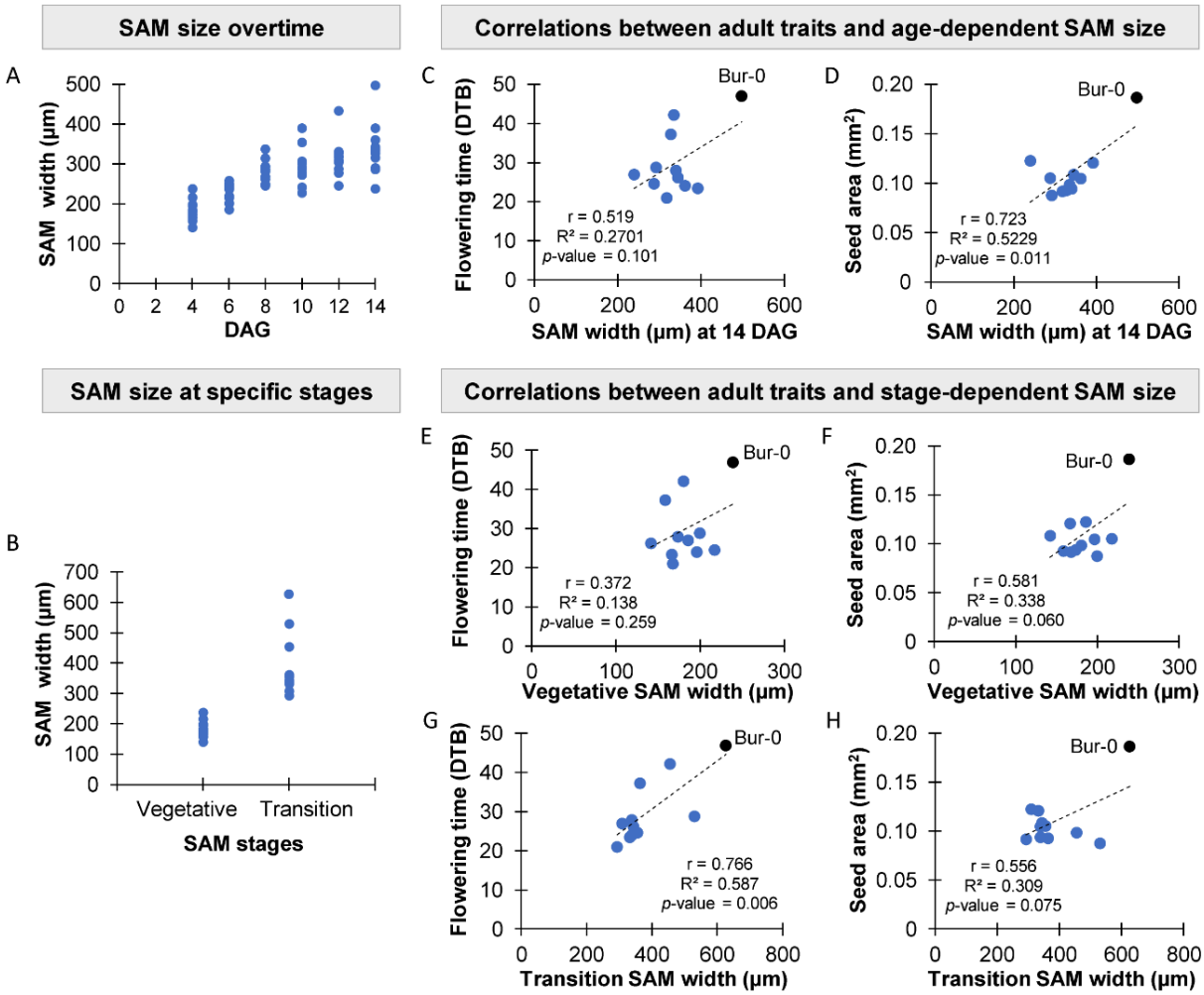


Figure 10. Correlations between SAM size and adult plant traits. Eleven *Arabidopsis thaliana* accessions were analyzed and plants were grown in long day photoperiod. SAM width was considered as SAM size parameter. (A) SAM size from 4 to 14 DAG (B) SAM size at vegetative and floral transition stages. Linear regressions between: (C) flowering time as days to bolting (DTB) and SAM size at 14 DAG, (D) seed area and SAM size at 14 DAG, (E) flowering time as days to bolting (DTB) and SAM size at vegetative stage, (F) seed area and SAM size at vegetative stage, (G) flowering time as days to bolting (DTB) and SAM size at floral transition stage, (H) seed area and SAM size at floral transition stage. Time is given in days after germination (DAG).

3.1.1.5. SAM and rosette size are stage-dependent correlated traits

In order to determine correlations between SAM size and the rosette phenotype, SAM size and rosette area were analyzed once more for Bur-0 (late flowering accession), the early flowering accession (Ws-2) and two intermediate flowering accessions (L*er*-1 and Col-0). In this case, the developmental stages (vegetative and transition) and the chronological age (10 DAG) were considered as reference for comparisons. Linear regressions were performed to determine possible correlations. When the accessions were compared at the same developmental stage, our results revealed that at vegetative stage (morphologically confirmed at 4 DAG for all accessions) SAM size is significantly different among accessions ($p < 0.05$), Bur-0 has bigger

rosettes (although at this stage/time point the rosette composition mainly contains cotyledonary leaves) and the correlation between SAM size and rosette area at vegetative stage is weak and not significant (Figure 11A-D). On the other hand, at floral transition stage (morphologically confirmed at 6 DAG for Ws-2, 10 DAG for Col-0 and L*er*-1 and 21 DAG for Bur-0), Bur-0 has significantly bigger SAM size and rosette area ($p < 0.05$), while SAM size and rosette area are not significantly different between Col-0, L*er*-1 and Ws-2 (Figure 11E-G).

At the floral transition stage, SAM size and rosette area are strongly and significantly correlated traits (Figure 11H). In contrast, when the accessions were compared at the same chronological age (10 DAG), SAM morphological analysis confirmed that all accessions are at different stages of development (Figure 11I), SAM size and rosette area are significantly different among accessions ($p < 0.05$) (Figure 11J, K) and no correlations were confirmed between rosette and SAM size (Figure 11L).

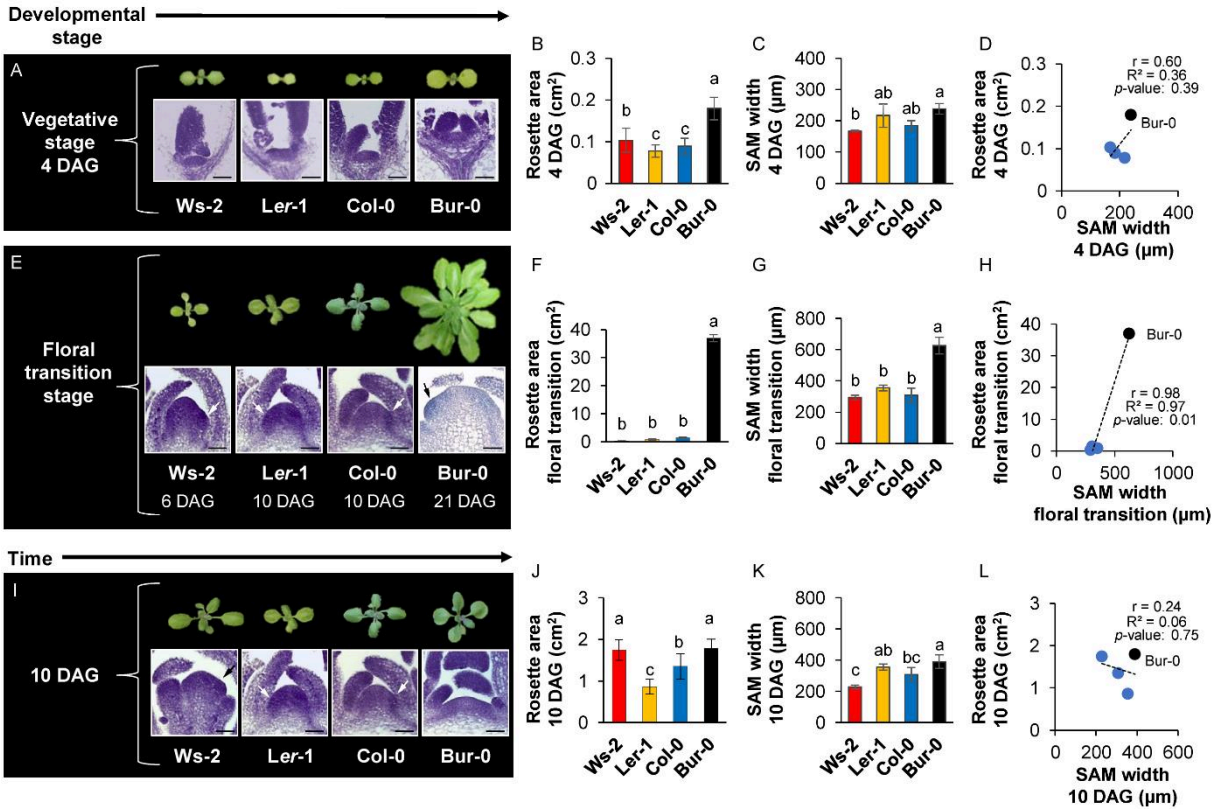


Figure 11. Correlation analyses between SAM size and rosette area. Stage-dependent and age-dependent rosette area and shoot apical meristem (SAM) size (as the SAM width) analyzed in the Arabidopsis accessions Bur-0, Col-0, L*er*-1 and Ws-2, grown in LD photoperiod. Stage-dependent analyses at vegetative stage: (A) rosette and SAM phenotype, (B) rosette area (n=20), (C) SAM width ($n \geq 3$), (D) linear regression between rosette area and SAM size. Stage-dependent analyses at floral transition stage: (E) rosette and SAM phenotype, (F) rosette area (n= 3), (G) SAM width ($n \geq 3$), (H) linear regression between rosette area and SAM size. Age-dependent analysis at 10 DAG: (I) rosette and SAM phenotype, (J) rosette area (n=20), (K) SAM width ($n \geq 3$), (L) linear regression between rosette area and SAM size. Time is given in days after germination (DAG). Scale bar = 100 µm. Arrows indicate emergence of floral primordia. Errors bars indicate ± SD. Statistical significance was tested with ANOVA and means were compared with Tukey–HSD test. Different letters indicate significant differences at α 0.05.

Thus, we could confirm that Bur-0 has the biggest rosette and SAM phenotype among the analyzed accessions, when traits are compared at the same developmental stage. Our results indicate that the flowering time and stage-specific rosette area (at floral transition) are good predictors for SAM size and suggest that SAM size might be an important factor involved in determination of accession-specific adult plant phenotypes.

In order to better understand the causes for the particularly large phenotype in Bur-0, besides the eleven Arabidopsis wild accessions analyzed, a detailed phenotype characterization was done for the late flowering mutant lines *tsf-1, ft-10*, *soc1-6*, and *fd-3*, which are in Col-0 background. Plants were grown in LD photoperiod and rosette area over time, flowering time, seed size and SAM size traits were analyzed and compared to the wild type Col-0. We found that only *fd-3* has significantly bigger rosettes and *soc1-6* significantly bigger seeds than Col-0, but no correlations were confirmed among traits (Supplementary Figure S5).

Furthermore, SAM morphological characterization revealed that floral transition initiation takes place at later time points than the wild type Col-0, SAM size is similar at vegetative stage for the mutant lines and the wild type, while at floral transition stage *soc1-6*, and *fd-3* have larger SAMs than the wild type. In comparison to the wild type, our results indicate that the increased rosette size during late postembryonic growth and wider SAMs observed in the late flowering mutant lines resemble the rosette and SAM phenotype of late flowering natural accessions and none of the late flowering mutant lines have simultaneously larger rosettes, larger SAM, later flowering phenotype and larger seeds than the wild type Col-0.

3.1.1.6. Developmental phase progression during postembryonic growth

In order to better understand how developmental phase progression is regulated during postembryonic growth, the length of intermediate developmental phases like germination and juvenile phase was determined for the late flowering accession Bur-0, the early flowering accession (Ws-2) and two intermediate flowering accessions (L*er*-1 and Col-0), grown in LD photoperiod.

Seed germination parameters like mean time to germination (MTG), germination index (GI) and final germination percentage (FGP) were determined and although all accessions have similar FGP (above 90%), Ws-2 has the highest GI and lowest MGT, i.e., germination is faster and more uniform, germination uniformity is similar for L*er*-1, Col-0, while Bur-0 has the highest MTG (slowest germination speed) (Figure 12A and Supplementary Table S5).

Furthermore, the vegetative phase change was analyzed by counting the juvenile leaves (JLN) in each accession and our results showed that Bur-0 has more juvenile leaves (10) than Col-0 (7), L*er*-1 (6) and Ws-2 (5), indicating an extended juvenile phase in Bur-0, intermediate juvenile phase length in Col-0, followed by L*er*-1 and short juvenile phase length in Ws-2 (Figure 12B). In addition, flowering time based on total leaf number (TLN) confirmed the late flowering phenotype in Bur-0, intermediate flowering phenotype in Col-0 and early flowering phenotype in Ws-2, while the flowering phenotype in L*er*-1 as TLN is rather early and similar to Ws-2 (Figure 12C). Thus, our results revealed different developmental phase progression in

each accession and from germination onwards, the length of developmental phases during postembryonic growth is extended in Bur-0, intermediate in Col-0 and L*er*-1 and shorter in Ws-2 (Figure 12D), suggesting that in general, the mechanisms that regulate developmental phase progression are shared across developmental phases.

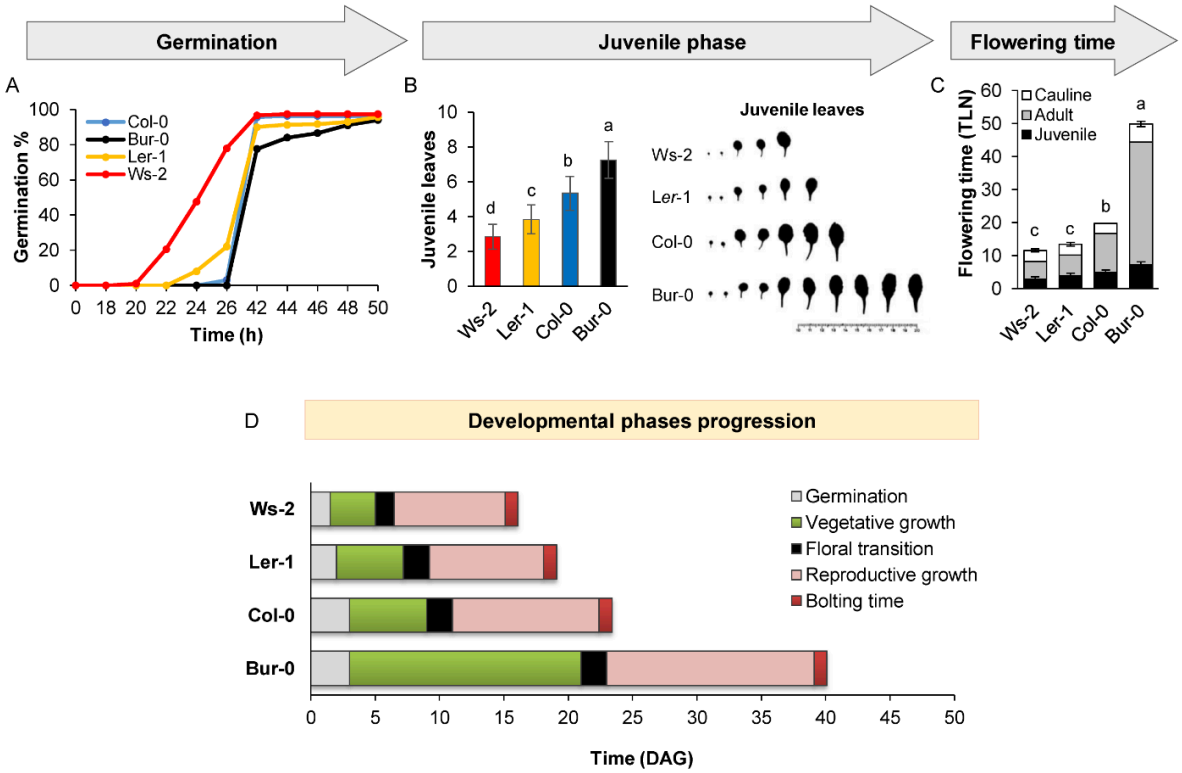


Figure 12. Postembryonic developmental phase progression. The Arabidopsis accessions Bur-0, Col-0, L*er*-1 and Ws-2 were analyzed and plants were grown in LD photoperiod. (A) Germination time, *n*=100. Time is given in hours after transferring the plates to the growth chamber. (B) Juvenile phase, *n*=10. (C) Flowering time (as total leaf number, TLN), *n*=12. (D) Overview of developmental phase progression, time is given as days after germination (DAG). Error bars indicate ± SD. Statistical significance was tested with ANOVA and means were compared with Tukey–HSD test. Different letters indicate significant differences at α 0.05.

3.1.2. Phenotypical characterization during embryonic growth

3.1.2.1. Embryo size analysis

Since our previous results confirmed that Bur-0 has particularly bigger seeds in comparison with other natural accessions, we investigated if the seed size might be attributed to a different embryo size. Longitudinal silique sections were stained with Toluidine blue for morphological identification of embryos at different stages of development in the Arabidopsis accessions Ws-2, Col-0, L*er*-1 and Bur-0 (Figure 13A). Embryo area analysis revealed that all accessions have similar embryo size at heart and torpedo stages (Figure 13B, C), while at late torpedo and mature stages the embryo area is significantly different among all accessions ($p< 0.05$) and Ws-2 has the smallest embryos, Col-0 and L*er*-1 have intermediate embryo area, while Bur-0 has the largest embryos (Figure 13D, E). Thus, our results indicate that seed size can be attributed in large part to the size of its embryo and interestingly, we found that the large embryo size in Bur-0 is achieved only during late embryogenesis.

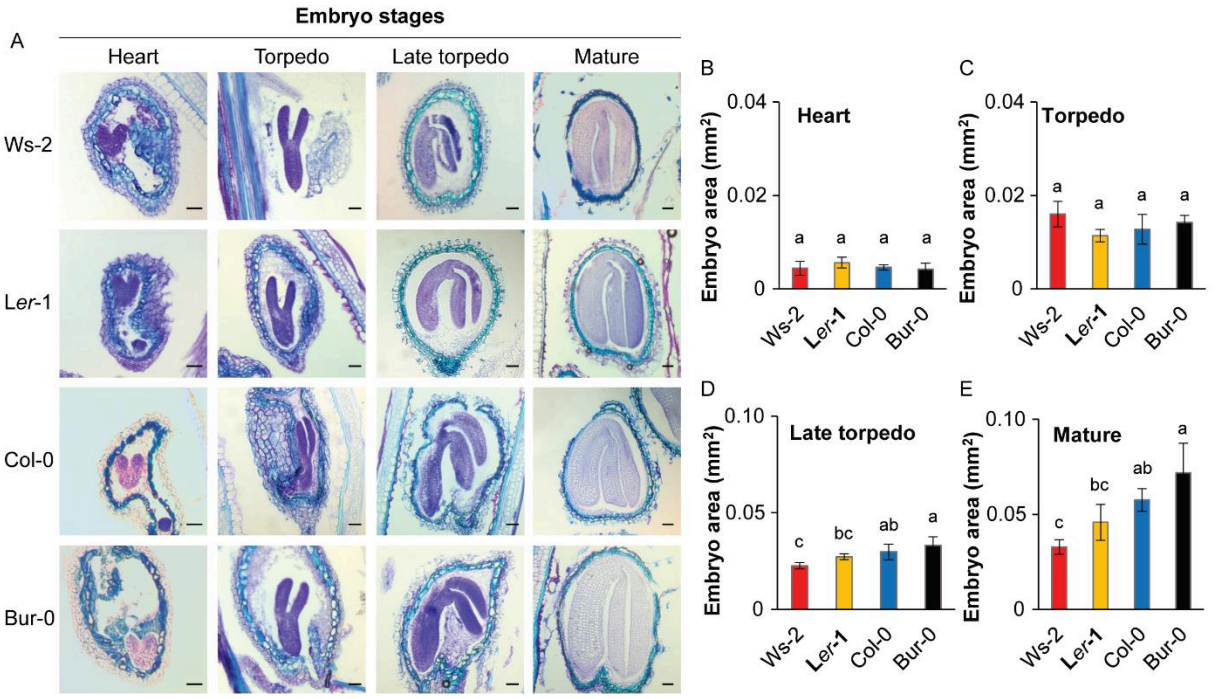


Figure 13. Embryo size analysis at different developmental stages. The Arabidopsis accessions Bur-0, Col-0, L*er*-1 and Ws-2 were analyzed. Longitudinal silique sections were stained with Toluidine blue for morphological identification of embryos at different stages of development. (A) Embryo phenotypes at heart, torpedo, late torpedo and mature stages. Scale bar = 50µm. Embryo area at (B) heart, (C) torpedo, (D) late torpedo and (E) mature stages. Error bars indicate ± SD. *n*=5. Statistical significance was tested with ANOVA and means were compared with Tukey–HSD test. Different letters indicate significant differences at α 0.05.

3.1.2.2. Embryo developmental progression

Embryo development based on days after pollination (DAP) was analysed for the Arabidopsis Bur-0 and Col-0 accessions. Plants were hand pollinated, properly developed siliques were harvested at different time points and ovules were extracted and cleared to identify embryo stages in each accession. We found that globular stages are identified at 5-6 DAP, heart stages at 6-7 DAP, torpedo stages at 7-8 DAP and green mature stages at 10-12 DAP (Figure 14).

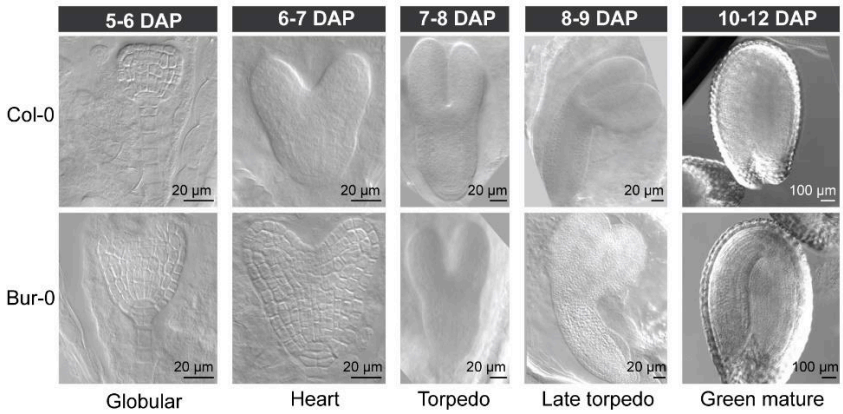


Figure 14. Embryo developmental progression analysis. Plants from the Arabidopsis accessions Col-0 and Bur-0 were hand pollinated, properly developed siliques were harvested ($n \geq 4$) per time point and ovules were extracted and cleared for Nomarski imaging. Embryo stages were identified from globular to mature in each accession. Time is given as days after pollination (DAP).

These results indicate that embryo development is synchronized for both accessions because the same embryo stages were observed around the same days after pollination (DAP). This embryo developmental progression analysis allowed us to determine precise time points to harvest embryo material at particular stages after hand pollination. For further experiments late torpedo embryos were harvested at 8 DAP and green mature embryos at 10 DAP.

3.1.2.3. Seed weigh and seed yield analysis

In order to determine implications of individual seed traits in the total seed yield per plant, seed weight and the total seed yield produced per plant was analyzed for the Arabidopsis accessions Ws-2, Col-0, L*er*-1 and Bur-0. Our results showed that seed weight (as weight of 100 seeds) and the total seed yield produced per plant are significantly different between the accessions analyzed ($p < 0.05$). Interestingly, Bur-0 has the highest seed weight and the lowest total seed yield per plant (Figure 15A, B).

The seed weight/ total seed yield results were used to determine the total seed number produced per plant and interestingly, the accession with the lowest seed weight Ws-2, produces in average 17,000 seeds per plant, while the accession with the highest seed weight Bur-0 produces in average 1,700 seeds per plant, i.e., approximately 10 times fewer seeds that the Ws-2 equivalent. The accessions Col-0 and L*er*-1 produced intermediate seeds number per plant (Figure 15C). Linear regression analysis between seed weight and total seed yield revealed a strong and significant correlation among both traits (Figure 15D).

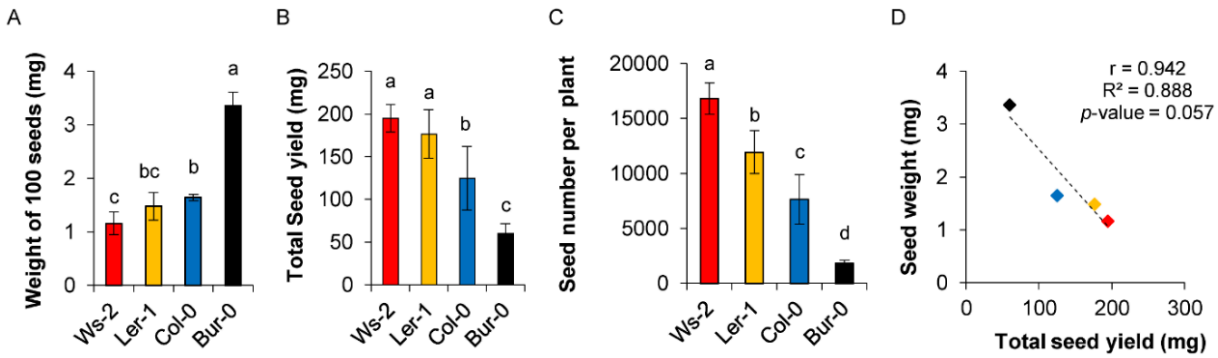


Figure 15. Seed weight and total seed yield produced per plant. The Arabidopsis accessions Bur-0, Col-0, L*er*-1 and Ws-2 were analyzed. (A) Seed weight (as weight of 100 seeds), *n*=5. (B) Total seed yield produced per plant, *n*=6. (C) Total seed number produced per plant, *n*=6. (D) Linear regression between seed weight and total seed. Error bars indicate ± SD. Statistical significance was tested with ANOVA and means were compared with Tukey–HSD test. Different letters indicate significant differences at α 0.05.

Depending on the parameter selected to evaluate seed yield, our results can be interpreted in two ways. One the one hand, if total seeds produced per plant is selected as target trait for seed yield, our results indicate that an early flowering accession with small rosettes like Ws-2 has higher seed yield than a late flowering accession with big rosettes like Bur-0. On the other hand, if the seed size or weight is selected as target trait for seed yield, our results indicate that Bur-0 has a higher seed yield than the other accessions.

3.1.2.4. Parental effects on seed size

In order to investigate parental effects on seed size, inter-accession crosses were performed using Bur-0 (as female and male parental genotype) and L*er*-1, Col-0, Ws-2, respectively. Seed area was measured for the inter-accession crosses and the self-crossed accessions. Interestingly, all inter-accession crosses (F1) result in bigger seeds than the parental self-crossed accessions, particularly when Bur-0 is used as the female parental genotype (Figure 16A, B). These results suggest parental effects on seed size that might be maternally controlled.

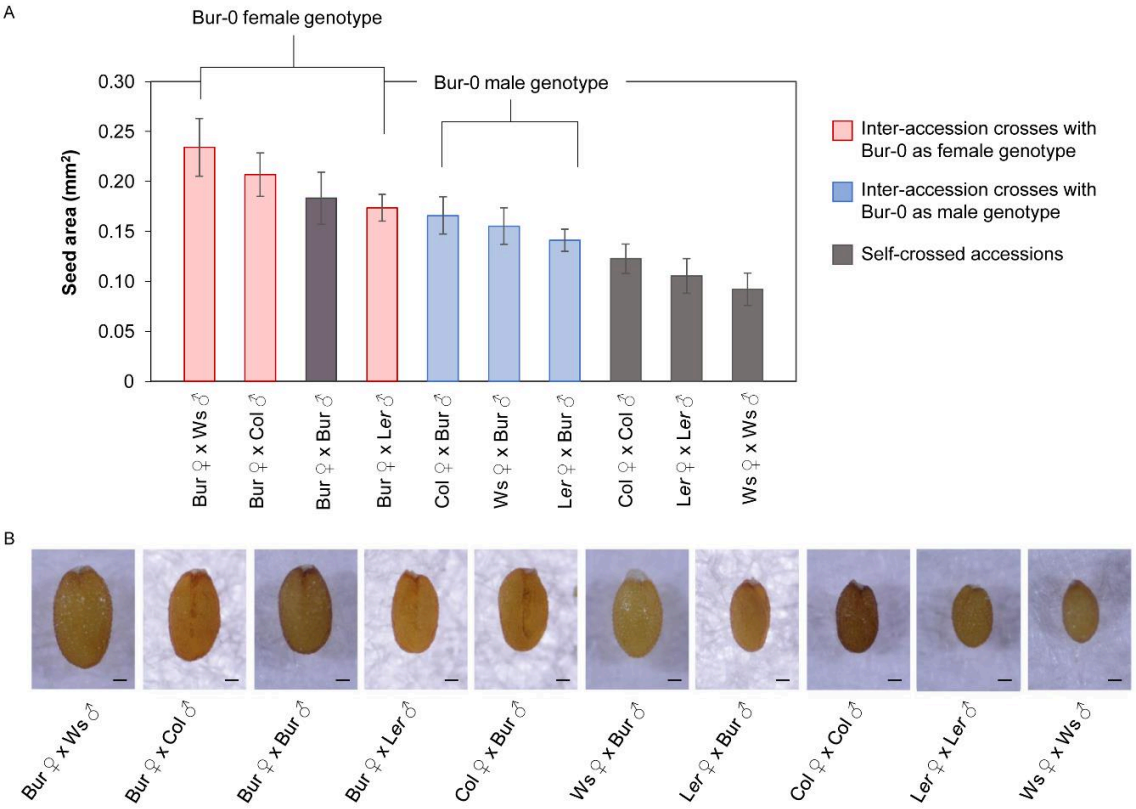


Figure 16. Parental effects on seed size. Inter-accession crosses between the Arabidopsis accessions Bur-0 (as female and male parental genotype), with L*er*-1, Col-0, Ws-2, respectively. (A) Seed area of the F1 and the parental self-crossed accessions ($n \geq 10$). Error bars indicate ± SD. (B) Seed phenotype of the F1 and the parental self-crossed accessions. Scale bar = 0.1 mm.

In summary, our data demonstrate that rosette area, seed area and flowering time are not generally correlated traits in Arabidopsis and suggest that those are independent traits influenced by other factors and none of them is an optimal marker trait (predictor) for the adult plant phenotype. Nevertheless, our analysis revealed how the accession Bur-0 stands out from the data distribution as outlier, with particularly larger seeds and later flowering time, but not with larger rosettes up to 14 DAG. Further growth analysis in different photoperiods and during late postembryonic growth confirmed that the rosette phenotype in Bur-0 is larger than in other accessions, but the larger phenotype is achieved during late postembryonic growth, regardless of the photoperiod. Our results also revealed that the bigger plant 3D surface area observed in Bur-0 does not result from a higher relative expansion rate (RER), but by determining the leaf initiation rate (LIR) per accession we found that Bur-0 produces leaves faster than the other accessions.

On the other hand, we found that early flowering accessions tend to have smaller SAMs (narrower), late flowering accessions tend to have larger SAMs (wider) and interestingly, Bur-0 has the largest SAM among the analyzed accessions. Our results also demonstrate that the stage-dependent SAM size is a good marker trait (predictor) for flowering and rosette size phenotypes in Arabidopsis and suggest that SAM size might be an important factor involved in determination of accession-specific adult plant phenotypes.

Our data demonstrate that among the analyzed natural accessions and mutant lines, no other accession or mutant line stands out as an outlier with larger rosettes, larger SAM, later flowering phenotype and larger seeds as Bur-0. Interestingly, further analysis during postembryonic development revealed that from germination onwards, the length of developmental phases during postembryonic growth is extended in Bur-0, intermediate in Col-0 and L*er*-1 and shorter in Ws-2, suggesting that in general, the mechanisms that regulate developmental phase progression are shared across developmental phases.

The phenotypical characterization during embryonic growth revealed that Ws-2, L*er*-1, Col-0 and Bur-0 have similar embryo size at heart and torpedo stages, while at late torpedo and mature stages the embryo area is significantly different among all accessions and Ws-2 has the smallest embryos, Col-0 and L*er*-1 have intermediate embryo area, while Bur-0 has the largest embryos. Further seed and yield analysis revealed that total seed yield per plant and seed weight are negatively correlated traits, thus Bur-0 produces bigger seeds, but also fewer seeds per plant. These results indicate that final seed size in each accession can be attributed in large part to the size of its mature embryo and interestingly, we found that the large embryo size in Bur-0 is achieved only during late embryogenesis.

Depending on the parameter selected to evaluate seed yield, our results can be interpreted in two ways. One the one hand, if total seeds produced per plant is selected as target trait for seed yield, our results indicate that an early flowering accession with small rosettes like Ws-2 has higher seed yield than a late flowering accession with big rosettes like Bur-0. On the other hand, if the seed size or weight is selected as target trait for seed yield, our results indicate that Bur-0 has a higher seed yield than the other accessions. Finally, inter-accession crosses between Bur-0 (as female and male parental genotype) and L*er*-1, Col-0, Ws-2, respectively, revealed that all inter-accession crosses (F1) result in bigger seeds, particularly when Bur-0 is used as the female parental genotype, suggesting parental effects on seed size that might be maternally controlled.

3.2. Physiological characterization

As part of the phenotype characterization described in *Section 3.1*. we found that during embryonic development Bur-0 produces bigger mature embryos than Col-0, L*er*-1 and Ws-2 and that its large embryo phenotype is achieved during late embryogenesis. In addition, negative correlations between seed weight and total seed yield per plant were identified, thus Bur-0 produces bigger seeds, but also fewer seeds per plant. Moreover, we found that during postembryonic development Bur-0 produces large rosettes, its enlarged phenotype is achieved during late postembryonic growth and it is photoperiod independent.

In order to investigate the extent to which the physiological status might contribute to the adult plant phenotype observed in Bur-0, biomass and metabolite content were analysed in particular tissues/stages where major phenotypical differences were observed during embryonic and postembryonic development, comparing Bur-0 to other Arabidopsis accessions including Col-0, L*er*-1 and Ws-2.

3.2.1. Physiological characterization during embryonic development

3.2.1.1. Published research article:

The results obtained for this subchapter have been published in the following research article:
Moreno Curtidor C, Annunziata MG, Gupta S, Apelt F, Richard SI, Kragler F, Mueller-Roeber B and Olas JJ. (2020). Physiological Profiling of Embryos and Dormant Seeds in Two Arabidopsis Accessions Reveals a Metabolic Switch in Carbon Reserve Accumulation. Frontiers in Plant Science. 11:588433. doi: 10.3389/fpls.2020.588433
Online access: https://doi.org/10.3389/fpls.2020.588433

The published article is presented in the next pages. Figure and page numbering were retained as in the published article. Page numbering of the thesis proceeds along the manuscript. The supplementary information from this publication is presented in Annex A. This research was done in collaboration with members from the Institute of Biochemistry and Biology, University of Potsdam, and members of the Max Planck Institute of Molecular Plant Physiology.

During my PhD project, I contributed to the publication as reported in the Author Contributions section (page 12 of the publication). To be more specific, I contributed in the following way to the publication: I grew the plants, performed the hand-pollination experiments, and analyzed embryo development. I collected seed samples and harvested siliques, extracted ovules and hand-dissected embryo samples. I embedded and sectioned embryo samples. I performed probe synthesis of *SUS* genes, RNA *in situ* hybridization assays, RNA isolation, and cDNA synthesis for qRT-PCR. I did ethanolic extractions and measurements of starch, protein, sucrose, glucose, and fructose content together with Maria Grazia Annunziata from MPI-MP.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 02 December 2020 doi: 10.3389/fpls.2020.588433

Physiological Profiling of Embryos and Dormant Seeds in Two *Arabidopsis* Accessions Reveals a Metabolic Switch in Carbon Reserve Accumulation

OPEN ACCESS

Edited by:
Mingxun Chen, Northwest A and F University, China

Reviewed by:
Marina Gavilanes-Ruiz, National Autonomous University of Mexico, Mexico Enrique Martinez Force, Instituto de la Grasa (IG), Spain

***Correspondence:** Justyna Jadwiga Olas olas@uni-potsdam.de Bernd Mueller-Roeber bmr@uni-potsdam.de

†These authors have contributed equally to this work

Specialty section: *This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science*

Received: 28 July 2020
Accepted: 16 November 2020
Published: 02 December 2020

Citation: *Moreno Curtidor C, Annunziata MG, Gupta S, Apelt F, Richard SI, Kragler F, Mueller-Roeber B and Olas JJ (2020) Physiological Profiling of Embryos and Dormant Seeds in Two Arabidopsis Accessions Reveals a Metabolic Switch in Carbon Reserve Accumulation. Front. Plant Sci. 11:588433. doi: 10.3389/fpls.2020.588433*

Catalina Moreno Curtidor[1,2†], Maria Grazia Annunziata[2†], Saurabh Gupta[2], Federico Apelt[2], Sarah Isabel Richard[1], Friedrich Kragler[2], Bernd Mueller-Roeber[1,2] and Justyna Jadwiga Olas[1*]*

[1]*Department of Molecular Biology, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany,* [2]*Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany*

In flowering plants, sugars act as carbon sources providing energy for developing embryos and seeds. Although most studies focus on carbon metabolism in whole seeds, knowledge about how particular sugars contribute to the developmental transitions during embryogenesis is scarce. To develop a quantitative understanding of how carbon composition changes during embryo development, and to determine how sugar status contributes to final seed or embryo size, we performed metabolic profiling of hand-dissected embryos at late torpedo and mature stages, and dormant seeds, in two *Arabidopsis thaliana* accessions with medium [Columbia-0 (Col-0)] and large [Burren-0 (Bur-0)] seed sizes, respectively. Our results show that, in both accessions, metabolite profiles of embryos largely differ from those of dormant seeds. We found that developmental transitions from torpedo to mature embryos, and further to dormant seeds, are associated with major metabolic switches in carbon reserve accumulation. While glucose, sucrose, and starch predominantly accumulated during seed dormancy, fructose levels were strongly elevated in mature embryos. Interestingly, Bur-0 seeds contain larger mature embryos than Col-0 seeds. Fructose and starch were accumulated to significantly higher levels in mature Bur-0 than Col-0 embryos, suggesting that they contribute to the enlarged mature Bur-0 embryos. Furthermore, we found that Bur-0 embryos accumulated a higher level of sucrose compared to hexose sugars and that changes in sucrose metabolism are mediated by sucrose synthase (SUS), with *SUS* genes acting non-redundantly, and in a tissue-specific manner to utilize sucrose during late embryogenesis.

Keywords: carbon, embryo development, hexoses, metabolites, sucrose, sucrose synthase

INTRODUCTION

In flowering plants, seed development is a highly complex and dynamic process that involves successful progression through several developmental stages leading to the formation of a quiescent seed that germinates later. In this context, seed size is one of the most important agronomic traits affecting seed yield (Kesavan et al., 2013). Therefore, determining the molecular and physiological mechanisms controlling seed development is an important task. Although seeds from different plant species vary greatly in their size, shape, and color, their development largely follows the same principle.

In seed producing plants including *Arabidopsis*, three main phases can be distinguished: embryo morphogenesis, embryo maturation, and seed desiccation (West and Harada, 1993; Harada, 1997). During the first phase, an embryo develops from a fertilized egg cell toward the heart- and torpedo-shaped forms through a series of asymmetric cell divisions. The basic body plan of the embryo with an apical-basal polarity is formed, resulting in the embryo with a morphologically recognizable axis (Capron et al., 2009). In the second phase, embryo maturation occurs, where cell expansion and differentiations replace active cell division (Dante et al., 2014) and storage products, including proteins and oils accumulate (Rolletschek et al., 2005; Baud and Lepiniec, 2010). Lastly, seed desiccation takes place, and the loss of water allows the embryo to enter a quiescent state which further leads to the establishment of a dormant seed (Manfre et al., 2009).

The successful shift between the stages requires the coordinated action of the genetic and molecular programs to support the growth of a developing seed (Le et al., 2010; Radchuk and Borisjuk, 2014). Three genetically distinct compartments exist in a seed: the embryo, the endosperm, and the maternal seed coat (Weber et al., 2005). The embryo and endosperm are derived from the zygotic tissues, while the seed coat develops from maternal integuments (Garcia et al., 2005). Tight interaction of all three elements is required for successful seed development and growth (Nowack et al., 2010). The seed coat protects the developing embryo from external factors to ensure proper development (Radchuk and Borisjuk, 2014), while the endosperm supports embryo growth by delivering nutrients acquired from the mother plant (Melkus et al., 2009). A failure in the development or in the function of the embryo, endosperm or coat will result in defects in the mature seed, or lead to premature embryo abortion (Hehenberger et al., 2012; Figueiredo et al., 2016). Although the embryo leads to the formation of the future adult plant, the developing embryo is highly dependent on the supply of photoassimilates and other nutrients from maternal tissue, particularly photosynthetically active leaves, to sustain cell patterning (Patrick and Offler, 2001). It is well-known that seed maturation is restricted by insufficient carbon supply (Lauxmann et al., 2016). Most of the carbon supplied by the maternal tissue for seed growth is in the form of sucrose (Morley-Smith et al., 2008). Once loaded into the phloem, sucrose is transported to siliques and is imported into developing seeds *via* a set of plasma membrane-localized transporters to provide the energy recourses needed for embryo development and viability (Patrick and Offler, 1995; Tegeder et al., 1999; Baud et al., 2005; Zhang et al., 2007; Chen et al., 2015). In seeds, sucrose is converted to starch, or is broken down by the action of invertase (INV; EC 3.2.1.26) or sucrose synthase (SUS; EC 2.4.1.13) enzymes (Hill et al., 2003; Morley-Smith et al., 2008). While at least 17 INVs are reported in *Arabidopsis* being present in different subcellular localizations (Ruan et al., 2010), only six SUSs are found, acting primarily in non-photosynthetic cells (Fujii et al., 2010). Interestingly, two different phases of sucrose utilization during seed development have been reported (Morley-Smith et al., 2008). During the first phase, when the embryo grows primarily *via* cell division, most of the sucrose in the seed is hydrolyzed to hexoses (glucose and fructose) by the action of INVs (Weschke et al., 2003; Barratt et al., 2009). Hexoses mainly accumulate in the endosperm causing a higher water potential and increased water uptake by the seed. In this phase, a rapid increase in seed volume occurs (Morley- Smith et al., 2008). During the second phase, SUS catalyzes the conversion of sucrose to fructose and uridine diphosphate (UDP)- glucose (Barratt et al., 2009). In this phase, when embryo's cell division ceases and cell expansion increases, sucrose rather than hexoses becomes the major sugar in the seed (Weber et al., 1997b; Borisjuk et al., 2003; Hill et al., 2003; Tomlinson et al., 2004). Although sugars/hexoses have been suggested as a hypothetical signal for seed maturation based on studies performed on legumes (Weber et al., 1998), Hill et al. (2003) showed that most of the generated hexoses in the endosperm do not arrive directly at the embryo (Hill et al., 2003). It thus remains unclear which carbon metabolic signals reach at the developing embryo in *Arabidopsis* to support its growth and development.

Although in the last decades, the molecular mechanisms controlling seed development and, in particular, endosperm cellularization has been well-studied, and many genes regulating seed development have been identified, knowledge about how metabolites contribute to the development of each seed compartment is scarce. This is mainly due to the lack of suitable analytical methods to investigate metabolism occurring in the internal structures of developing seeds. Despite the fact that metabolites provide energy resources for the transition from embryo to seed and that carbohydrate-mediated signaling molecules might direct growth (Wobus and Weber, 1999), no information is available on the sugars, which might contribute to embryo development. To address this, we performed metabolite profiling assays to determine the metabolite content of hand- dissected embryos at late torpedo and mature stages, and dormant seeds, in two *Arabidopsis thaliana* accessions, Col-0 and Bur-0, showing significant differences in seed size. We found that Bur-0 embryos contain much higher carbon reserves compared to Col-0. Our analysis revealed that the sucrose is predominantly degraded *via* SUS pathways in mature embryos, and that *SUS* genes act in non-redundant and rather cell- or tissue-specific manner in sucrose metabolism during late embryogenesis.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Arabidopsis thaliana accessions Columbia-0 (Col-0) and Burren-0 (Bur-0) were used in all experiments. Seeds were obtained

from the in-house collection of the Max Planck Institute of Molecular Plant Physiology. Seeds were sown in 6-cm pots filled with 3:1 soil: vermiculite substrate, stratified at 4°C in the dark for 2 days, and afterward moved to growth chambers (Percival AR-36L2, CLF Plant Climatics GmbH, Wertingen, Germany). Plants were grown in long-day (LD; 16 h light/8 h darkness) condition at 22°C with a photosynthetically active radiation of 160 μmol m^{-2} s^{-1} at the plant level.

Plants were hand-pollinated to analyze the progression of embryo development over time [days after pollination (DAP)]. Embryo developmental stages were determined for both *Arabidopsis* accessions and embryos at late torpedo and mature stages were hand-dissected using an Olympus SZX12 stereomicroscope (Olympus Deutschland GmbH, Hamburg, Germany). Briefly, siliques at 8 and 10 DAP were harvested, placed in a petri dish, and opened under a stereomicroscope by peeling off the valves using micro-dissecting forceps. Exposed ovules were carefully removed, and gently squeezed to release the embryo. Embryos were collected with a syringe needle and rapidly transferred to 100 μl of RNAlater (Thermo Fisher Scientific, Massachusetts, United States). Each biological replicate contained approximately 250 dissected embryos. Samples were kept at 4°C until use. Before use, RNAlater was carefully removed by pipetting.

Seed Parameters, Water Content, and Embryo Size

For determining seed and embryo parameters, plants were harvested at maturity, when siliques were fully ripe. One hundred seeds per accession were weighed (n = 5). For determining seed length, width, and area (n = 20), dried seeds were imaged, and then measured using the *ImageJ* software (NIH, Maryland, United States). Water content was assessed in three biological replicates by determining the fresh weight and subsequent dry weight after 17 h at 105°C (ISTA, 2011). The water content was calculated as the loss in weight as a percentage of the original weight of seeds. Embryo area (n = 20) was measured from images obtained using an Olympus BX-61 microscope (Olympus Europa SE & Co, Hamburg, Germany) and was analyzed using *ImageJ*.

Metabolite Measurements

The total amount of non-structural carbohydrates (starch, sucrose, glucose, and fructose), organic acids (fumarate and malate), total amino acids (AAs), and total protein content was determined in three biological replicates (n = 3; each replicate contained approximately 250 hand-dissected embryos) of Col-0 and Bur-0 embryos at torpedo and mature stages, as well as of dormant seeds. The samples were extracted with boiling 80% (v/v) ethanol and were assayed enzymatically as previously described (Stitt et al., 1989). The supernatants were used for the determination of soluble sugars (Stitt et al., 1989), total AA (Cross et al., 2006), and malate and fumarate (Nunes- Nesi et al., 2007). The pellets were used to determine starch content (Hendriks et al., 2003); protein content was determined using the Bradford method (Bradford, 1976), with bovine serum albumin as standard. The spectrophotometric assays were performed in 96-well microplates, and the absorbance was determined using a Synergy, an ELX-800, or an ELX-808 microplate reader (Bio-Tek, Bad Friedrichshall, Germany). For all assays, two technical replicates were determined per biological replicate. Data analysis was performed as previously described (Annunziata et al., 2017). The total carbon (C) accumulated in metabolites (**Supplementary Table S1**) was calculated as previously described (Lauxmann et al., 2016).

PCA and Heat Maps

The metabolite levels were normalized by z-score (after removing outliers) prior use for principal component analysis (PCA) using the *prcomp* function in the *R* stats package and were plotted using the *ggbiplot* R package. The z-score values were further used for clustering the metabolites and samples *via* hierarchical complete linkage clustering with Euclidean distance using the *pheatmap* R package.

qRT-PCR Analysis

Total RNA from hand-dissected embryos at late torpedo and mature stage of Col-0 and Bur-0 plants was isolated in three biological replicates using *mir*Vana™ miRNA Isolation kit (Thermo Fisher Scientific, Massachusetts, United States). Briefly, embryos were harvested using needles in RNAlater solution (Thermo Fisher Scientific, Massachusetts, United States). Each biological replicate contained approximately 250 hand-dissected embryos. Afterward, RNAlater solution was removed by washing embryos with DEPC-H_20, and pelleted embryos were used for RNA isolation. DNA digestion and cDNA synthesis were performed using Turbo DNA-free DNase I kit (Ambion/Life Technologies, Darmstadt, Germany) and SuperScriptTMIII Reverse Transcriptase Kit (Invitrogen/Life Technologies, Darmstadt, Germany) according to the manufacturer's instructions. The qRT-PCR measurements were carried out using the CFX connect real-time PCR system (Bio-Rad, CA, United States) in a 10-μl total reaction volume in triplicates using SYBR® Green-PCR Master Mix (Applied Biosystems/Life Technologies, Darmstadt, Germany). Expression values were calculated by normalizing the Ct value of the gene of interest to that of the housekeeping gene *TUBULIN 2* (At5g62690); data are presented in graphs as mRNA fold change (Olas et al., 2019). Primer sequences used for the qRT-PCR measurements are listed in **Supplementary Table S2**.

RNA *in situ* Hybridization

For RNA *in situ* hybridization, Col-0 and Bur-0 siliques with embryos were harvested in formaldehyde: acetic acid fixation solution (FAA; 50% EtOH, 5% acetic acid, 3.7% formaldehyde, and 41.3% H_20). The samples were fixed overnight using an automated tissue processor (Leica ASP200S, Leica, Wetzlar, Germany), embedded in wax using an embedding system (HistoCore Arcadia, Leica), and afterward sectioned (8 μm thickness) using a rotary microtome (Leica RM2255; Leica). The slides were stored at 4°C until used for RNA *in situ* hybridization. Probes for *SUCROSE SYNTHASE 1* (*SUS1*;

At5g20830), *SUS3* (At4g02280), and *CYCLINB1;1* (*CYCB1;1*; At4g37490) were generated from cDNAs, and primers used for cloning are listed in **Supplementary Table S2**. RNA *in situ* hybridization was carried out as described (Olas et al., 2019). Briefly, slides were dewaxed by washing in Histoclear II solution and ethanol series. For immunological detection, anti-DIG antibody (Roche, Mannheim, Germany) solution diluted 1:1250 in blocking reagent (Roche) was applied to the slides and incubated at room temperature for 90 min. For the colorimetric detection, the NBT/BCIP stock solution (Roche) diluted 1:50 in 10% polyvinyl alcohol (PVA) in TNM-50 was applied to the slides. The slides were incubated overnight in the dark at room temperature. Sections were imaged with an Olympus BX-61 microscope equipped with a Digital Camera View II, using cellSens Dimension program (Olympus Europa SE & Co, Hamburg, Germany). The figure panels presented in this work were generated using Adobe Photoshop CS5 and Adobe Illustrator CS5.

Statistics

Statistical significance between two ecotypes was calculated using two-tailed, two-sample equal variance Student's *t*-test:
$^{*}p \leq 0.05$; $^{**}p \leq 0.01$; $^{***}p \leq 0.001$.

RESULTS

Bur-0 Accession Has Bigger Seeds and Mature Embryos Than Col-0

Given the crucial role of seed size as an agronomic trait that largely influences seed yield, and the fact that elucidating the mechanisms underlying seed size will help us to improve yield (Kesavan et al., 2013), we decided to investigate how metabolic profiles contribute to embryo development. First, the morphological variations in the seed features in two *A. thaliana* accessions, Col-0 and Bur-0, previously reported as ecotypes with medium and large seed sizes, respectively (Herridge et al., 2011), were analyzed. Consistent with the previous study, Bur-0 seeds were 59% larger at late dry mature stage than Col-0 seeds ($p < 0.001$; **Figures 1A–C**). As changes in seed size are often associated with changes in seed shape, we analyzed seed length, width, and the length-to-width ratio (**Figures 1D,E**). Bur-0 seeds had greater length and width than Col-0 seeds (**Figure 1D**), whereas the ratio of length to width of Bur-0 seeds was not significantly different from that of Col-0 seeds (**Figure 1E**), demonstrating that Bur-0 has enlarged seed size compared with Col-0, while seed shape was similar in the

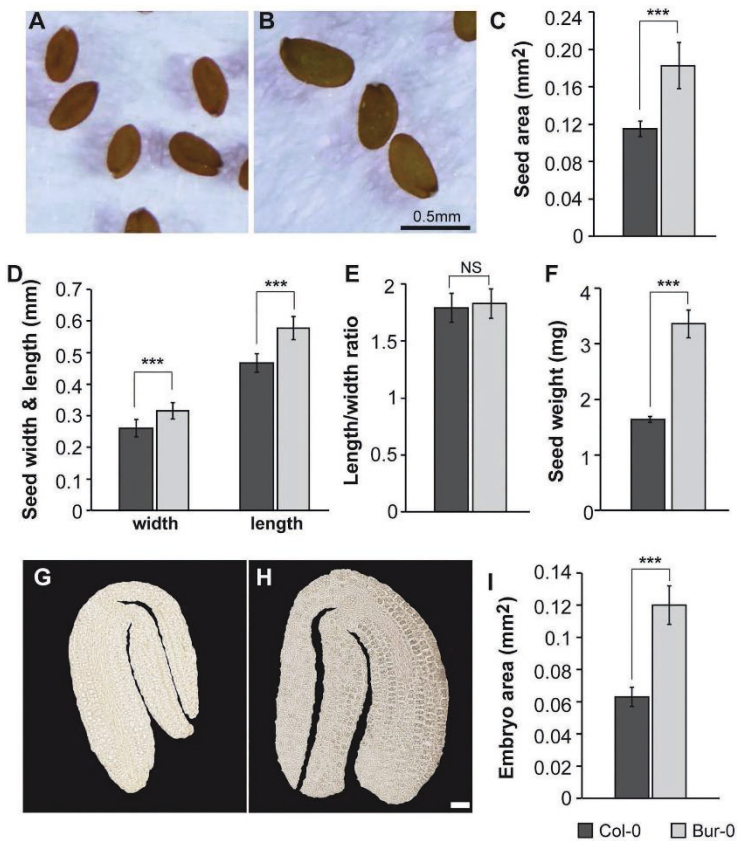


FIGURE 1 | Burren-0 (Bur-0) accession has bigger seeds and mature embryos. **(A,B)** Mature dried seeds of *Arabidopsis thaliana* natural accessions **(A)** Columbia-0 (Col-0) and **(B)** Bur-0. **(C)** Seed area. **(D)** Seed length and width. **(E)** The ratio of length to width. **(F)** Seed weight of 100 mature dried seeds ($n = 5$). **(G,H)** Mature embryos of **(G)** Col-0 and **(H)** Bur-0. Scale, 50 μm. **(I)** Mature embryo area. Error bars indicate s.d. ($n = 20$). Statistically significant difference between accessions was calculated using Student's *t*-test (NS, not significant; $^{***}p < 0.001$).

two accessions. Next, we analyzed the average mass of Bur-0 and Col-0 seeds by weighing batches of 100 mature seeds (**Figure 1F**). In agreement with the observed seed size, mature seeds of Bur-0 plants were on average 48% heavier than Col-0 seeds, although the water content in dry mature seeds was similar in both accessions (**Supplementary Figure S1**).

To test if the increased seed size of Bur-0 might be determined by a change in embryogenesis, we examined the size of mature embryos of Col-0 and Bur-0 plants (**Figures 1G–I**). Interestingly, embryos from mature Bur-0 seeds were about 89% bigger ($p < 0.001$) than those of Col-0 plants.

Thus, to investigate if the enlarged size of Bur-0 mature embryos resulted from changes in cell cycle activity, we analyzed the expression of the mitotic marker gene *CYCLINB1;1* (*CYCB1;1*) in longitudinal sections of early and late torpedo, and mature embryos, by RNA *in situ* hybridization (**Figure 2**). We found that cell division was active in embryos at early torpedo stage (**Figures 2A,D**), while no expression of the cell cycle marker was observed in late torpedo (**Figures 2B,E**) and mature (**Figures 2C,F**) embryos in both accessions, demonstrating that cell division had stopped and that the increased size of Bur-0 embryos during late embryogenesis likely is not triggered by changes in cell division but rather associated with an accumulation of storage products.

Comparison of the Metabolic Profiles in Embryos and Dormant Seeds of Col-0 and Bur-0

As the observed morphological changes during late embryogenesis in Col-0 and Bur-0 embryos might result from differences in storage reserves, we decided to determine the metabolite content in dormant seeds as well as in embryos (without endosperm and coat). For metabolite profiling, we collected embryos at late torpedo and mature stages of Col-0 and Bur-0 plants, so stages in which cell division had stopped (see **Figure 2**). To correlate changes in the metabolite levels with developmental stages during embryogenesis, non-structural carbohydrates (starch, sucrose, fructose, and glucose), organic acids (fumarate and malate), total amino acids (AA), and total protein content were analyzed. Firstly, to ensure that each metabolite was considered equally in the analysis, we performed *z*-score normalization of the metabolite data set. Then, we performed a PCA of all metabolite levels to get an initial overview of the data (**Figure 3A**). Here, the PCA analysis provides information about which samples (three developmental stages: late torpedo, mature embryos, and dormant seeds) are closely related or separated, and which variables (metabolites) contribute to this relationship. The principal component 1 (PC1) and PC2 accounted for 91.6 and 6.5% of the total variation in the data set, respectively. Along the PC1 axis, we identified a clear separation of the embryo samples (torpedo and mature) from the dormant seeds, suggesting that metabolite content of dormant seeds strongly differs from that in embryos, which in part may be due to their reduced water content. This separation was mainly driven by differences in most of the measured metabolites (starch, sucrose, glucose, protein, malate, fumarate, and AA), which were all found to be positive markers of dormant seeds of Col-0 and Bur-0 plants. Along the PC2 axis, mature embryos clearly separated from embryos at late torpedo stage. This separation was driven by fructose and starch (positive markers of mature embryos), indicating that mature embryos display metabolite profiles distinct from those of the torpedo stage, which is more pronounced in Bur-0 than in Col-0.

To further elucidate the differences in the metabolite status of Col-0 and Bur-0 accessions during seed development, we performed hierarchical clustering (**Figure 3B**). The clustered heat map of all metabolites confirmed our previous observation that the metabolic content of dormant seeds is considerably different from that of embryos. Our results clearly demonstrate that the developmental transition from mature embryos to dormant seeds is associated with major metabolic switches in carbon, proteins, and AA accumulation. Moreover, we found that mature embryos accumulate a much higher level of fructose than torpedo stage embryos in both accessions. Importantly, our data revealed that fructose and starch contents were higher throughout late embryogenesis in Bur-0 than Col-0 plants. Moreover, we observed a higher level of total AA content in torpedo Bur-0 embryos than in Col-0 embryos, demonstrating that embryos of both accessions display different metabolic profiles. However, the higher metabolite levels in dormant seeds of both accessions largely masked the metabolic status observed in torpedo and mature embryos in our PC and heat map analyses.

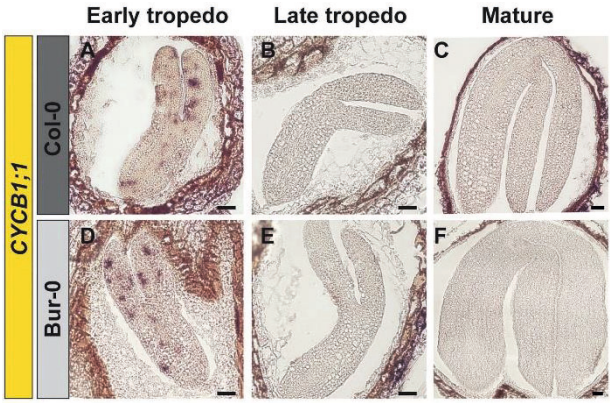


FIGURE 2 | Cell division stops at late torpedo and mature stages of embryogenesis. **(A–F)** RNA *in situ* hybridization on longitudinal sections through **(A,D)** early, **(B,E)** late, and **(C,F)** mature embryos of **(A–C)** Col-0 and **(D–F)** Bur-0 using *CYCLINB1;1* (*CYCB1;1*) as the probe. Scale bars, 100 μm.

Alterations in Metabolic Profiles of Embryos and Dormant Seeds of Col-0 and Bur-0

Since the PCA and heat map revealed large differences in the metabolite profiles of embryos and dormant seeds, and because fructose and starch mainly contributed to the separation of mature Bur-0 embryos from other samples, we compared the levels of the individual metabolites in Col-0 and Bur-0 (**Figure 4**). A notable difference occurred between embryos and dormant seeds for individual metabolites. In both accessions, we found that

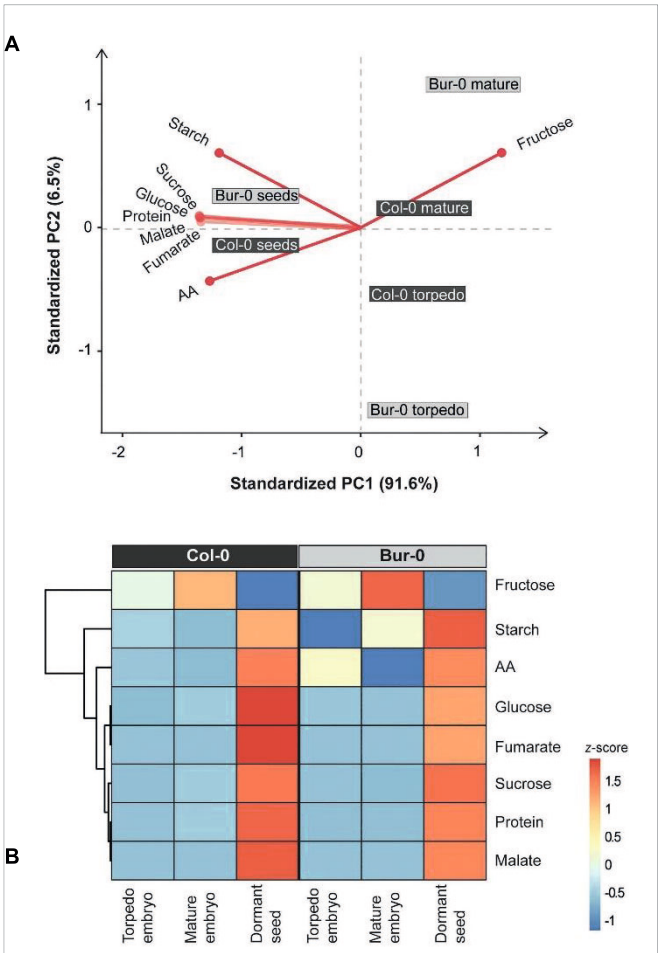


FIGURE 3 | Metabolite content of dormant seeds and embryos at late torpedo and mature stages. **(A)** Principal component analysis (PCA) and **(B)** clustered heat map of the metabolites measured in late torpedo and mature embryos and dormant seeds of Col-0 (dark gray) and Bur-0 (light gray) wild-type plants grown in long-days (LDs; 16 h light/8 h darkness). Metabolites from biological replicates ($n \geq 3$) were averaged and z-score normalized. The percentages of total variance represented by principal component 1 (PC1) and PC2 are shown in parentheses. The loadings of individual metabolites are shown in red.

total protein content, AA, starch, sucrose, and glucose levels were higher in dormant seeds than in torpedo and mature embryos (**Figures 4A**–**E**). Interestingly, all metabolite levels, except fumarate, were not statistically significant different between Bur-0 and Col-0 seeds. Although protein content was very similar across all samples, we found a statistically significant difference between mature embryos (**Figure 4A**). In fact, the protein content in mature Col-0 embryos was about 75% higher than in mature Bur-0 embryos. We noted that AA content generally decreased through embryo development (late torpedo to mature stage; **Figure 4B**), which might result from an increased incorporation of free AA into storage proteins. This observation was more pronounced in Bur-0 embryos where a decrease of 60% was observed from torpedo to mature stage. In agreement with PCA and heat map analyses, we noted that mature Bur-0 embryos accumulated, on average, more starch (26%) and fructose (12%) than mature Col-0 embryos (**Figures 4C,F**). While a greater starch accumulation during late seed development was reported based on metabolic studies in whole seeds (Weschke et al., 2000), we found that the starch level in mature Bur-0 embryos was similar to that seen in dormant seeds (**Figure 4C**). In fact, the starch level in dormant seeds of Bur-0 and Col-0 was only about 40 and 60% higher, respectively, than in corresponding mature embryos suggesting that they already have carbon reserves similar to that of dormant seeds. In addition, no glucose was detected in Col-0 torpedo embryos (**Figure 4E**), indicating that overall carbon reserves might be higher in Bur-0 embryos than in Col-0 embryos during late embryogenesis. Neither in torpedo nor in mature Col-0 and Bur-0 embryos, we detected tricarboxylic acid (TCA) cycle intermediates (fumarate and malate; **Figures 4G,H**), indicating that these organic acids do not serve as alternative carbon sources for developing embryos during late stages of embryogenesis, as previously seen in leaves or flowers (Fahnenstich et al., 2007; Lauxmann et al., 2016). Interestingly, both fumarate and malate were reported to be present in the seed and progressively decrease throughout whole seed development (Fait et al., 2006). The lack of fumarate and malate in embryos during late stages, but their presence in whole seeds containing embryos (Fait et al., 2006), endosperm and coat suggests variation in the activity of the metabolic pathways in the three seed compartments. As our data showed that carbon metabolism is enhanced in Bur-0 embryos, we decided to determine the total carbon accumulated (total carbon was summed from the non-structural carbohydrates and organic acids; **Figure 4I**; **Supplementary Table S1**). While dormant seeds of both accessions had similar total amounts of carbon, mature and in particular late torpedo Bur-0 embryos contained up to 1.2-fold more carbon than Col-0 embryos, confirming that Bur-0 embryos accumulate more carbon during late embryogenesis. Lastly, to access the information about the origin of the different carbon proportions present in the different tissues of the two accessions, we compared the content of hexoses (glucose + fructose) and sucrose (**Figure 4J**). Overall, the hexose levels remained constant in both accessions whereas sucrose progressively increased throughout embryo development until the dormant seed stage. In both accessions, we found a significantly higher level of sucrose compared to hexose sugars in all analyzed tissues except mature Bur-0 embryos. This observation is in agreement with our previous findings showing that very low glucose levels are present in embryos at those stages. Mature Bur-0 embryos contain similar levels of sucrose and hexose, suggesting that the rates of sucrose utilization and sucrose synthesis are similar. Importantly, the hexose-to-sucrose ratio was higher during late embryo development and decreased rapidly upon transition from mature embryo to the desiccate seed (**Figure 4K**), demonstrating that during embryo development a metabolic shift in carbon accumulation occurs. Moreover, we found that the ratio was much higher in late torpedo (11%) and mature (15%) embryos of Bur-0 compared to Col-0 embryos.

In summary, our analysis revealed that dormant seeds and embryos display distinct metabolite profiles. Moreover, we found that Bur-0 embryos in particular at the late torpedo stage contain higher carbon reserves than Col-0 embryos.

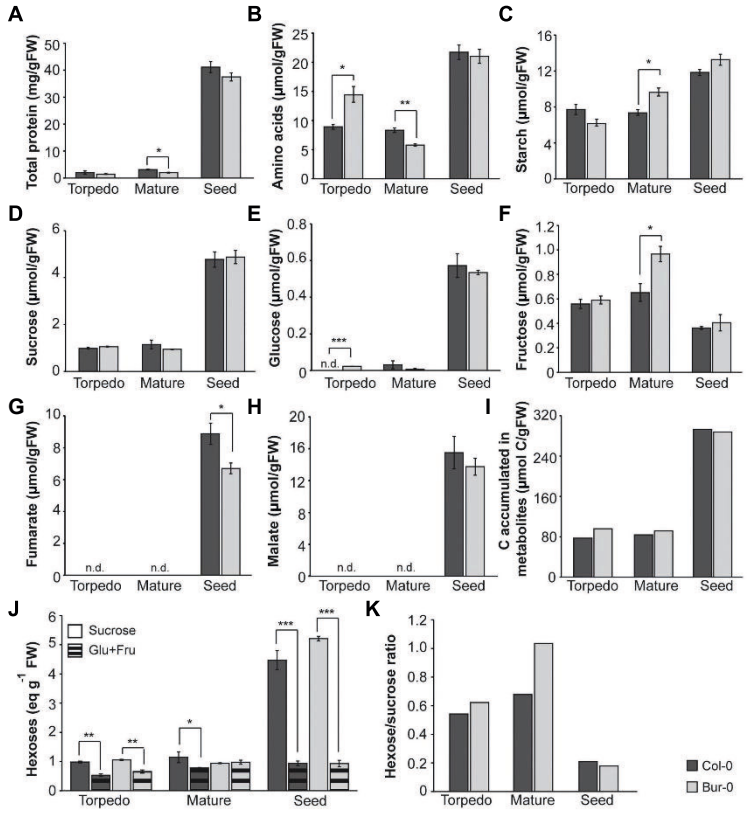


FIGURE 4 | Metabolite content of late torpedo- and mature-stage embryos and dormant seeds of *A. thaliana* accessions Col-0 (dark gray) and Bur-0 (light gray). Plants were grown in LD conditions (16 h light/8 h darkness). **(A)** Total protein content. **(B)** Total amino acids (AAs). **(C)** Starch. **(D)** Sucrose. **(E)** Glucose (Glu). **(F)** Fructose (Fru). **(G)** Fumarate. **(H)** Malate. Note that both fumarate and malate were not detected (n.d.) in embryos of Col-0 and Bur-0 plants. **(I)** Total carbon (C) accumulated in metabolites measured in embryos and seeds. **(J)** Sucrose and hexoses (Glu + Fru). **(K)** Hexose-to-sucrose ratio. Error bars indicate mean ± SEM ($n = 3$). At each time point, statistically significant difference between the two accessions was calculated using Student's *t*-test and is indicated as follows: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

Bur-0 Accumulates More Carbon Resources During the Transition From Late Torpedo to Mature Embryos

Embryos display a carbon status distinct from that of seeds at the late dry mature stage. However, those metabolic differences in embryos are masked by the high metabolite content of dormant seeds. Therefore, we performed a PCA on a *z*-score- normalized data set for metabolites measured only in embryo samples (i.e., excluding dormant seeds; **Figures 5A,B**). Along the PC1, a primary separation of mature Bur-0 from mature Col-0 embryos as well as from late torpedo stage embryos of both accessions (49.4% of total variance) was observed (**Figure 5A**). The separation of mature Bur-0 embryos from the other samples was mainly driven by differences in starch and fructose levels (markers of mature Bur-0 embryos). Moreover, the AA content was responsible for the very close display and separation of Bur-0 and Col-0 torpedo embryos from other mature embryos. Along PC2, a separation of mature Col-0 samples from other embryos (41.1% of total variance) was observed, suggesting that the most distinct metabolic changes between both accessions occurred during the mature stages. Mature Col-0 embryos were marked by sucrose and glucose, along with an influence from total protein and fructose. Generally, high fructose is a marker for mature embryos, whereas high AA marks embryos at torpedo stage.

Next, we generated a heat map to visualize the differences in metabolite levels of late torpedo- and mature-stage embryos (**Figure 5B**). At torpedo stage, there was no difference between fructose level in Bur-0 and Col-0, while we observed higher glucose and sucrose levels in the Bur-0 accession. In contrast, Col-0 torpedo embryos had higher levels of starch, AA, and protein compared to Bur-0. Furthermore, at mature stage, starch and fructose levels were much higher in Bur-0 than Col-0 embryos, while in mature Col-0 embryos a higher content of total protein, sucrose, and glucose was observed. Interestingly, the level of glucose was much higher in mature Col-0 than Bur-0 embryos, while during the torpedo stage the glucose level was higher in Bur-0 embryos.

In summary, the metabolic analysis of Col-0 and Bur-0 embryos revealed that the developmental transition from late torpedo to mature embryos is associated with major metabolic switches in carbon accumulation. Of note, however, Bur-0 embryos accumulate much more carbon reserves than Col-0 embryos. Furthermore, the fact that glucose level rises in Col-0 embryos only at the

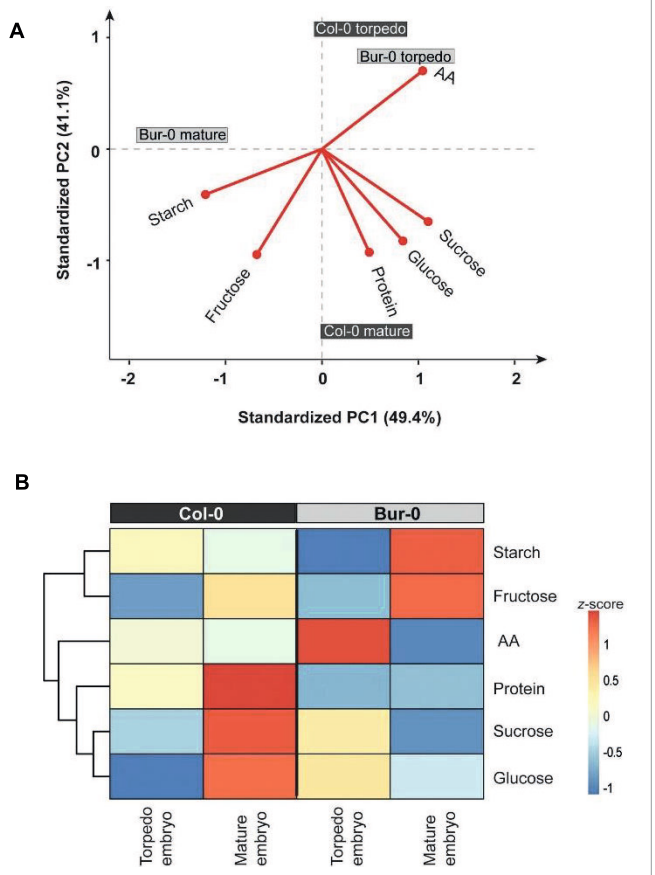


FIGURE 5 | Metabolite content of late torpedo and mature embryos. **(A)** PCA and **(B)** clustered heat map of metabolites measured in torpedo and mature embryos of Col-0 (dark gray) and Buro-0 (light gray) wild-type plants grown in LDs (16 h light/8 h darkness). Metabolites from biological replicates ($n \geq 3$) were averaged and *z*-score normalized. For the PCA, the percentages of total variance represented by PC1 and PC2 are shown in parentheses. The loadings of individual metabolites are shown in red.

mature stage, while in Bur-0 levels were already high at the torpedo stage, suggests differences in carbon metabolism, resulting in a smaller amount of accumulated carbon in Col-0.

Carbon Metabolism Is Enhanced in Bur-0 Embryos

As we detected that a greater proportion of carbon in torpedo and mature embryos is derived from sucrose, we decided to elucidate how sucrose is metabolized in embryos and why Bur-0 embryos accumulate more carbon than Col-0 embryos. INVs and SUSs have been suggested to contribute to sucrose degradation in early and late stages, respectively, of seed development (Tomlinson et al., 2004). Thus, we investigated the expression level of the respective genes involved in sucrose metabolism. We performed quantitative real-time PCR (qRT-PCR) analysis on dissected late torpedo and mature embryos of Col-0 and Bur-0 plants and measured the transcript abundance of all six *SUS* genes (**Figures 6A,B**) and two selected cytosolic *INV* (*CINV*) genes (**Figures 6C,D**) that mediate sucrose breakdown (Winter and Huber, 2000). We found similar expression levels of *SUS* genes in torpedo embryos of Col-0 and Bur-0 (**Figure 6A**). However, we observed a significantly higher expression of *SUS2*, *3*, *4*, and *5* in Bur-0 mature embryos compared to Col-0 (**Figure 6B**), suggesting that all these *SUS* isoforms support embryo development. Importantly, expression of both *CINV1* and *CINV2* was not induced in Bur-0 embryos compared to Col-0, neither in torpedo nor in mature embryos (**Figures 5C,D**), suggesting that during those stages of embryogenesis the *INVs* are not the main factors for sucrose degradation.

As our results suggested SUS-mediated sucrose degradation during the transition from late torpedo- to mature-stage embryos, we performed RNA *in situ* hybridization on longitudinal sections through mature Col-0 and Bur-0 embryos, using a *SUS3*-specific probe (**Figure 6E**). In accordance with the observation that *SUS3* expression was considerably higher in Bur-0 than Col-0 embryos, determined by qRT-PCR (**Figure 6B**), we found that *SUS3* was more strongly induced in mature embryos of Bur-0 than Col-0 plants, particularly in cells of the vasculature (**Figure 6E**). Furthermore, since the qRT-PCR analysis revealed similar *SUS1* and *SUS6* expression levels in mature embryos of both accessions (**Figure 6B**), we analyzed the transcript of *SUS1 via* RNA *in situ* hybridization to validate our results (**Figure 6F**). We found that the expression of *SUS1* in mature embryos of Bur-0 was not visibly different from that of Col-0 plants. Interestingly, the expression domain of *SUS1* was specific for the shoot apical meristem region of mature embryos. Considering that *SUS3* transcript was mainly present in the vasculature of the embryos, while *SUS1* was expressed at the SAM, our results indicate that *SUS* genes not necessarily act redundantly during embryogenesis to degrade the available sucrose but rather function in a tissue-specific manner.

In summary, our data show that degradation of sucrose in mature embryos mainly occurs through SUS pathways, and this metabolic activity appears to be enhanced in Bur-0 embryos.

DISCUSSION

Metabolites, and in particular sugars, play a crucial role during embryo development by providing energy resources for differentiation and growth; therefore, plants in their early developmental stages cannot fully grow without a sufficient and extended supply of carbon (Olas et al., 2020). Although carbohydrate status controls seed formation, and nutrient assimilation pathways are functional in embryos (Gifford and Thorne, 1985; Wobus and Weber, 1999; Neubohn et al., 2000; Olas and Wahl, 2019), little is known about which metabolites or carbon forms are distributed between the internal structures of the seed in *Arabidopsis*. Thus, most of the current knowledge about the importance of metabolites during embryo development is obtained from genetic studies of mutants affected in particular metabolic pathways or metabolic measurements performed on whole seeds, due to the lack of suitable biochemical methods for the investigation of metabolism occurring in the internal

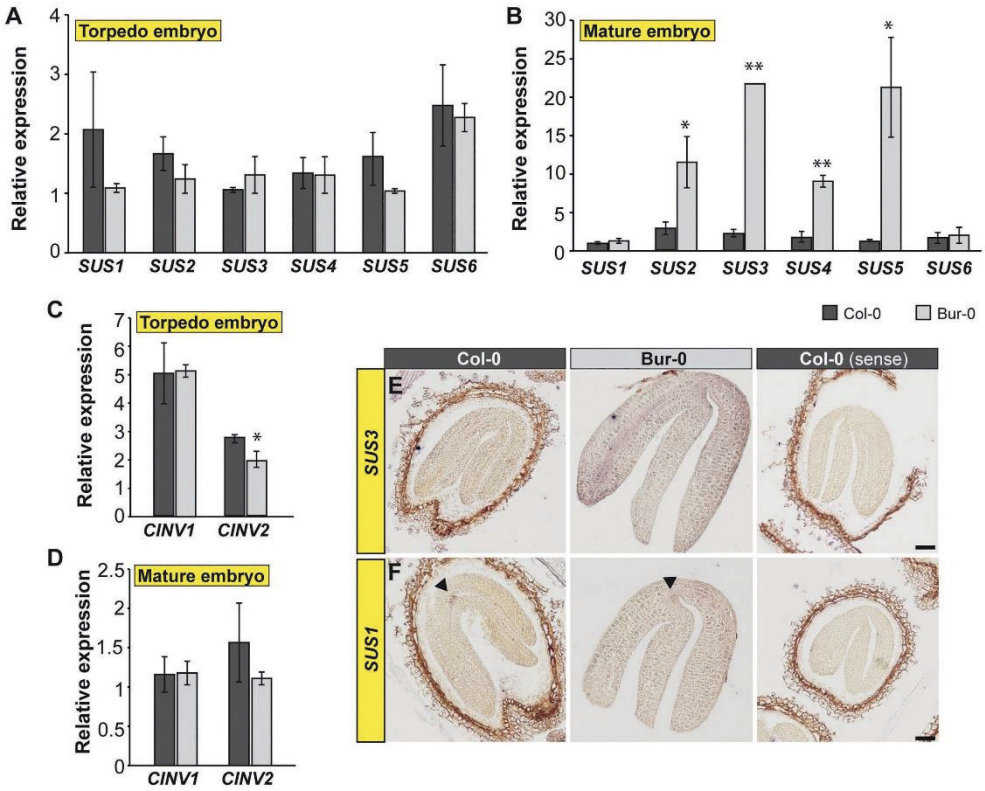


FIGURE 6 | Carbon metabolism is enhanced in mature embryos of Bur-0. **(A–D)** Transcript abundance of **(A,B)** SUCROSE SYNTHASE (SUS) and **(C,D)** cytosolic INVERTASE (CINV) genes in dissected **(A,C)** late torpedo and **(B,D)** mature embryos of Col-0 and Bur-0. Error bars indicate ± s.d. ($n = 3$). At each embryo stage, statistically significant difference between the two accessions was calculated using Student's *t*-test, and is indicated as follows: $^{*}p < 0.05$; $^{**}p < 0.01$. **(E,F)** RNA *in situ* hybridization using specific antisense probes for *SUCROSE SYNTHASE 3* (*SUS3*, **E**) and *SUS1* **(F)** genes on longitudinal sections through mature embryos of Col-0 and Bur-0 plants. Arrows indicate the expression of *SUS1* at the shoot apical meristem. Sense probes were used as control (right side). Scale bars, 50 μm.

seed structures (Heim et al., 1993; Golombek et al., 2001; Eastmond et al., 2002; Tomlinson et al., 2004).

Here, we analyzed morphological variations, metabolite content and transcript abundance of dormant seeds and dissected embryos, lacking endosperm and coat, at late torpedo and mature stages of Col-0 and Bur-0 *A. thaliana* accessions, in order to identify metabolites that report on embryo development during the late embryogenic phase and explain the enlarged seed size previously reported for Bur-0 plants (Herridge et al., 2011).

Variation in Seed Morphological Traits Among Accessions

Information about the magnitude of genetic variation of key seed size traits is of great importance for cultivar development programs that focus on improving seed yield and seedling establishment (Ambika et al., 2014). Small-seed species produce more seeds for a given amount of energy than species with large seeds (Aarssen and Jordan, 2001; Henery and Westoby, 2001; Moles et al., 2005); however, the latter develop seedlings that are often more resistant to abiotic stresses encountered during their establishment (Dong et al., 2016). Our analysis of morphological traits such as seed length, width, length-to- width ratio, area, weight, and embryo size demonstrated that the previously reported large-seed *Arabidopsis* ecotype Bur-0 (Herridge et al., 2011) contains an enlarged embryo compared to that of the medium-seed-size accession Col-0. In particular, mature Bur-0 embryos were significantly bigger (about 89%) than Col-0 embryos. It is well-established that organ growth is determined by cell division occurring at an early stage during organ growth and subsequent cell expansion that is accompanied with rapid synthesis of structural biomass associated with carbon deposition (Olas et al., 2020). As the mature embryo represents the final stage of embryonic development, we hypothesized that the accumulation of structural biomass, protein, and oil, rather than cell division, leads to the enlarged Bur-0 embryos. Indeed, we showed that both late torpedo and mature embryos stopped cell division, suggesting that the increased embryo size is triggered by changes in metabolism associated with an accumulation of storage products.

Developmental Transitions From Late Torpedo to Mature Embryos, and to Dormant Seeds, Are Associated With Metabolic Switches in Carbon Accumulation

The embryo as a non-photosynthetically active tissue is fully dependent on nutrients and carbon supplied by maternal tissues

for proper development (Rolletschek et al., 2005). Previous studies on legume crops demonstrated the relevance of carbon metabolism for seed development (Gifford and Thorne, 1985; Wobus and Weber, 1999; Neubohn et al., 2000). Also for *Arabidopsis*, it is well-established that insufficient carbon supply to developing seeds affects seed maturation (Lauxmann et al., 2016). In fact, a negative relationship exists between seed size and the number of seeds produced by the mother plant due to a limitation of available carbon resources (Harper et al., 1970; Jakobsson and Eriksson, 2003; Li and Li, 2015). Moreover, the importance of carbon metabolism for embryo development was demonstrated by analyzing the *Arabidopsis tps1* null mutant which lacks functional *TREHALOSE-6-PHOSPHATE SYNTHASE 1*. The *tps1* mutant is blocked in the developmental progression of embryos from the torpedo to the mature stage (Eastmond et al., 2002). Overall, we found that Bur-0 embryos accumulated more carbon resources during late embryogenesis than Col-0, and both accessions progressively accumulated carbon content throughout their development. This is in accordance with previous studies showing that once cell division stops, cell expansion increases, and synthesis of storage products starts (Weber et al., 1995; Baud et al., 2002; Hill et al., 2003). Sucrose and starch are the major products of photosynthesis in plants and are considered the most important carbon sources for growth (Stitt and Zeeman, 2012). We found that in both accessions sucrose and starch progressively increased through embryo development. While dormant seeds displayed similar metabolic profiles in both accessions, a notable difference in starch and sucrose levels was observed between them during late torpedo and mature embryo stages. Our finding of a higher sucrose level in Bur-0 late torpedo embryos could explain the increased starch level detected in mature Bur-0 embryos, providing strong evidence for a causal relationship between changes in sucrose catabolism and starch synthesis (Borisjuk et al., 2003). Sucrose imported into the embryo, and then converted to starch would act as a catalyst for the accumulation of more carbon by strengthening the sink status of the seed (da Silva et al., 1997; Andriotis et al., 2010). Furthermore, although fumarate and malate have been suggested to act as alternative storage compounds of fixed carbon in various plant organs, similar to starch and sucrose (Fahnenstich et al., 2007; Araújo et al., 2011), we did neither detect fumarate nor malate in Col-0 and Bur-0 embryos. Undetectable levels of these TCA cycle intermediates in embryos suggest a strongly reduced flux through this pathway. This could potentially be due to a limitation of sufficient oxygen for mitochondrial respiration and the production of ATP and reducing equivalents. Importantly, the lack of fumarate and malate in embryos during late developmental stages, but their presence in whole seeds containing embryos (Fait et al., 2006), endosperm and coat suggests variation in the activity of the metabolic pathways in the three seed compartments.

The route *via* which sucrose is transported to reach the outer seed integument is well-described (Stadler et al., 2005); however, to date little is known about in which form and *via* which pathway carbon from sucrose reaches the developing embryo. Hill et al. (2003) showed that the major pool of hexoses generated in the seed endosperm do not arrive directly at the embryo. Previous studies focusing on developing seeds showed that sucrose utilization in seeds occurs *via* two separate pathways and in two distinct phases (Morley-Smith et al., 2008). The INV pathway operates during early seed development and hydrolyzes sucrose to hexoses (glucose and fructose), which become the major sugars in the seed, while the SUS pathway catalyzes the conversion of sucrose to fructose and UDP-Glc in the late maturation phase (Barratt et al., 2009). UDP-Glc can be used directly, or after transformation into ADP-Glc, for the synthesis of structural and nonstructural carbohydrate macromolecules, respectively (Everard and Loescher, 2017). Here, in contrast to the first phase, sucrose rather than hexoses plays a crucial role for seed growth (Weber et al., 1995; Hill et al., 2003; Morley-Smith et al., 2008). We found that dormant seeds and embryos at late torpedo and mature stages accumulated more sucrose than hexoses in both accessions, except for Bur-0 mature embryos, which contained similar levels of hexoses and sucrose. Those results indicate that sucrose is the major form of carbon in embryos during late embryogenesis. Previous studies performed on oilseeds suggested that the hexose-to-sucrose ratio declines when the transition to storage product accumulation occurs in the embryo (Morley- Smith et al., 2008). In line with these observations, we found that the hexose-to-sucrose ratio was higher in embryos than in dormant seeds. However, hexose-to-sucrose ratio might be even higher during initial embryo development, e.g., in the globular or heart stages, time points not covered in our studies. Importantly, the ratio was in general greater in Bur-0 than Col-0 embryos. The fall in the hexose-to-sucrose ratio in the later stages of seed development has been proposed to be related to the switch from cell division to expansion, and storage product accumulation (Weber et al., 1997a; Borisjuk et al., 1998). Accordingly, one could speculate that the higher ratio of hexose to sucrose in Bur-0 embryos might indicate that cell division is enhanced in Bur-0 embryos compared to Col-0 embryos. However, our analysis revealed that cell division has stopped in embryo stages analyzed in this study, demonstrating that cell division is not enhanced in Bur-0 embryos. Importantly, the small pool of hexoses, the undetected glucose level, and the very high level of fructose in late torpedo and mature embryos suggests that during late *Arabidopsis* embryogenesis the SUS pathway is mainly operative.

A shift in *INV* and *SUS* Expression During Late Embryo Development Mediates the Metabolic Switches in Carbon Accumulation

Previous studies indicated that a shift from INV to SUS activity during seed development mediates sucrose metabolism and triggers the changes in hexose-to-sucrose ratio (Tomlinson et al., 2004). During early embryo growth, which is mainly driven by cell proliferation, most of the sucrose arriving at the developing seed is hydrolyzed by INVs to produce hexoses (glucose and fructose). Thereafter, when cell division fades in the embryo while cell expansion becomes dominant, sucrose rather than hexoses becomes the major sugar in the seed (Weber et al., 1997a; Borisjuk et al., 1998).

In agreement with the previous reports, we found that indeed cell division had stopped in late torpedo and mature embryos. Importantly, we found that several *SUS* genes were induced in mature embryos of Bur-0 plants. We did not observe an upregulation of transcripts of two analyzed cytosolic *INVs*, showing that sucrose in mature embryos is predominantly degraded *via* SUS, as previously reported for whole seeds (Borisjuk et al., 2003; Tomlinson et al., 2004). Tomlinson et al. (2004) suggested that while sucrose utilization for starch synthesis occur *via* SUS enzymes, the INV pathway predominates for oil synthesis. Moreover, recent studies showed that INVs are essential for ovule development through sugar signaling rather than provision of carbon nutrients (Liao et al., 2020). However, here it should be noted that Zuma et al. (2018), using microarray analysis, showed that the cluster containing nine *INV* genes was highly upregulated only during early embryogenesis (i.e., pre-globular, globular, and heart stages), while after the transition to heart stage those genes were downregulated. Similarly, Borisjuk et al. (2003) showed that *INVs* are high during early stages of seed development and that their expression declines throughout seed development. Accordingly, we cannot exclude that INV-mediated sucrose degradation might be the major operative pathway of sucrose utilization during early embryo development, i.e., during stages not covered in our studies.

Importantly, our data suggest non-redundant and rather cell- or tissue-specific functions of *SUS* genes in sucrose metabolism during late embryogenesis. Of note, non-redundant functions of metabolism genes during embryogenesis have been shown; our previous analysis demonstrated that two nitrate assimilation genes, *NIA1* and *NIA2*, complement each other's expression pattern in embryos and act non-redundantly to assimilate nitrate (Olas and Wahl, 2019).

CONCLUSION

In this study, we determine which carbon form is predominant in late *Arabidopsis* embryo stages. By measuring metabolites in dissected late torpedo and mature embryos (without endosperm and seed coat), and in dormant seeds, of Col-0 and Bur-0 *A. thaliana* accessions forming medium and large seeds, respectively (Herridge et al., 2011), we provided evidence that changes in carbohydrate content occur during late embryo development, similarly as reported for whole seeds in legume plants (**Figure 7**; Gifford and Thorne, 1985). More importantly, we showed that the transitions from torpedo to mature embryo, as well as from the mature embryo to desiccation stage, are signified by distinct carbon markers and those changes are pronounced in the accession with the bigger seed size.

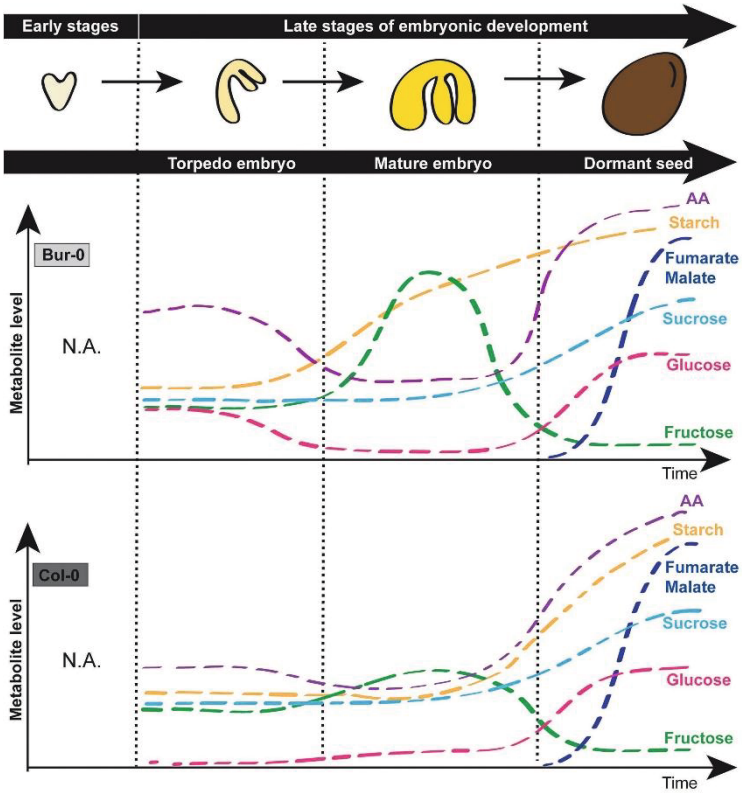


FIGURE 7 | Simplified model describing carbohydrate signatures of embryos during late torpedo and mature stages as well as dormant seeds of *A. thaliana* Col-0 and Bur-0 accessions. Note that Bur-0 embryos accumulate much higher carbon reserves during late embryogenesis than Col-0 embryos. AA, amino acids.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

JO and BM-R conceived and designed the research. CM and MA performed the experiments with the assistance from JO. SR helped with the qRT-PCR assays. JO and MA analyzed the data. SG and FA performed the statistical analyses supervised by FK. BM-R and FK secured the funding. JO generated figures and wrote the paper with contributions from MA and BM-R. All authors contributed to the article and approved the submitted version.

FUNDING

The Deutsche Forschungsgemeinschaft (DFG) provided funding of the Collaborative Research Centre 973 on "Priming and Memory of Organismic Responses to Stress." Additional funding was provided by the Max Planck Society and the International Max Planck Research School "Primary Metabolism and Plant Growth" (IMPRS-PMPG), a joint initiative of the University of Potsdam and the Max Planck Institute of Molecular Plant Physiology and by the European Research Council (ERC), Syg Project 810131 – PLAMORF (FK).

ACKNOWLEDGMENTS

We thank Eike Kamann (University of Potsdam) for generating cDNA constructs used for RNA *in situ* hybridization. We thank Prof. Dr. Dr. h.c. Mark Stitt (Max Planck Institute of Molecular Plant Physiology) for helpful comments on the research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.588433/full#supplementary-material

REFERENCES

Aarssen, L. W., and Jordan, C. (2001). Between-species patterns of covariation in plant size, seed size and fecundity in monocarpic herbs. *Ecoscience* 8, 471–477. doi: 10.1080/11956860.2001.11682677

Ambika, S., Manonmani, V., and Somasundaram, G. (2014). Review on effect of seed size on seedling vigour and seed yield. *Res. J. Seed Sci.* 7, 31–38. doi: 10.3923/rjss.2014.31.38

Andriotis, V. M., Pike, M. J., Kular, B., Rawsthorne, S., and Smith, A. M. (2010). Starch turnover in developing oilseed embryos. *New Phytol.* 187, 791–804. doi: 10.1111/j.1469-8137.2010.03311.x

Annunziata, M. G., Apelt, F., Carillo, P., Krause, U., Feil, R., Mengin, V., et al. (2017). Getting back to nature: a reality check for experiments in controlled environments. *J. Exp. Bot.* 68, 4463–4477. doi: 10.1093/jxb/erx220

Araújo, W. L., Nunes-Nesi, A., and Fernie, A. R. (2011). Fumarate: multiple functions of a simple metabolite. *Phytochemistry* 72, 838–843. doi: 10.1016/j.phytochem.2011.02.028

Barratt, D. P., Derbyshire, P., Findlay, K., Pike, M., Wellner, N., Lunn, J., et al. (2009). Normal growth of *Arabidopsis* requires cytosolic invertase but not sucrose synthase. *Proc. Natl. Acad. Sci. U. S. A.* 106, 13124–13129. doi: 10.1073/pnas.0900689106

Baud, S., Boutin, J.-P., Miquel, M., Lepiniec, L., and Rochat, C. (2002). An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. *Plant Physiol. Biochem.* 40, 151–160. doi: 10.1016/S0981-9428(01)01350-X

Baud, S., and Lepiniec, L. (2010). Physiological and developmental regulation of seed oil production. *Prog. Lipid Res.* 49, 235–249. doi: 10.1016/j.plipres.2010.01.001

Baud, S., Wuilleme, S., Lemoine, R., Kronenberger, J., Caboche, M., Lepiniec, L., et al. (2005). The AtSUC5 sucrose transporter specifically expressed in the endosperm is involved in early seed development in *Arabidopsis*. *Plant J.* 43, 824–836. doi: 10.1111/j.1365-313X.2005.02496.x

Borisjuk, L., Rolletschek, H., Wobus, U., and Weber, H. (2003). Differentiation of legume cotyledons as related to metabolic gradients and assimilate transport into seeds. *J. Exp. Bot.* 54, 503–512. doi: 10.1093/jxb/erg051

Borisjuk, L., Walenta, S., Weber, H., Mueller-Klieser, W., and Wobus, U. (1998). High-resolution histographical mapping of glucose concentrations in developing cotyledons of *Vicia faba* in relation to mitotic activity and storage processes: glucose as a possible developmental trigger. *Plant J.* 15, 583–591. doi: 10.1046/j.1365-313X.1998.00214.x

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

Capron, A., Chatfield, S., Provart, N., and Berleth, T. (2009). Embryogenesis: pattern formation from a single cell. *Arabidopsis Book* 7:e0126. doi: 10.1199/tab.0126

Chen, L. -Q., Lin, I. W., Qu, X. -Q., Sosso, D., McFarlane, H. E., Londoño, A., et al. (2015). A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the *Arabidopsis* embryo. *Plant Cell* 27, 607–619. doi: 10.1105/tpc.114.134585

Cross, J. M., Von Korff, M., Altmann, T., Bartzetko, L., Sulpice, R., Gibon, Y., et al. (2006). Variation of enzyme activities and metabolite levels in 24 *Arabidopsis* accessions growing in carbon-limited conditions. *Plant Physiol.* 142, 1574–1588. doi: 10.1104/pp.106.086629

da Silva, P., Eastmond, P., Hill, L., Smith, A., and Rawsthorne, S. (1997). Starch metabolism in developing embryos of oilseed rape. *Planta* 203, 480–487. doi: 10.1007/s004250050217

Dante, R. A., Larkins, B. A., and Sabelli, P. A. (2014). Cell cycle control and seed development. *Front. Plant Sci.* 5:493. doi: 10.3389/fpls.2014.00493

Dong, R., Jahufer, M., Dong, D., Wang, Y., and Liu, Z. (2016). Characterisation of the morphological variation for seed traits among 537 germplasm accessions of common vetch (*Vicia sativa* L.) using digital image analysis. *New Zeal. J. Agric. Res.* 59, 422–435. doi: 10.1080/00288233.2016.1229682

Eastmond, P. J., Van Dijken, A. J., Spielman, M., Kerr, A., Tissier, A. F., Dickinson, H. G., et al. (2002). Trehalose-6-phosphate synthase 1, which catalyses the first step in trehalose synthesis, is essential for *Arabidopsis* embryo maturation. *Plant J.* 29, 225–235. doi: 10.1046/j.1365-313x.2002.01220.x

Everard, J., and Loescher, W. (2017). "Primary products of photosynthesis, sucrose and other soluble carbohydrates" in *Encyclopedia of applied plant science*. eds. B. Thomas, B. G. Murray and D. J. Murphy, 94–104.

Fahnenstich, H., Saigo, M., Niessen, M., Zanor, M. I., Andreo, C. S., Fernie, A. R., et al. (2007). Alteration of organic acid metabolism in *Arabidopsis* overexpressing the maize C4 NADP-malic enzyme causes accelerated senescence during extended darkness. *Plant Physiol.* 145, 640–652. doi: 10.1104/pp.107.104455

Fait, A., Angelovici, R., Less, H., Ohad, I., Urbanczyk-Wochniak, E., Fernie, A. R., et al. (2006). *Arabidopsis* seed development and germination is associated with temporally distinct metabolic switches. *Plant Physiol.* 142, 839–854. doi: 10.1104/pp.106.086694

Figueiredo, D. D., Batista, R. A., Roszak, P. J., Hennig, L., and Köhler, C. (2016). Auxin production in the endosperm drives seed coat development in *Arabidopsis*. *eLife* 5:e20542. doi: 10.7554/eLife.20542

Fujii, S., Hayashi, T., Mizuno, K. J. P., and Physiology, C. (2010). Sucrose synthase is an integral component of the cellulose synthesis machinery. *Plant Cell Physiol*. 51, 294–301. doi: 10.1093/pcp/pcp190

Garcia, D., Gerald, J. N. F., and Berger, F. (2005). Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in *Arabidopsis*. *Plant Cell* 17, 52–60. doi: 10.1105/ tpc.104.027136

Gifford, R. M., and Thorne, J. H. (1985). Sucrose concentration at the apoplastic interface between seed coat and cotyledons of developing soybean seeds. *Plant Physiol*. 77, 863–868. doi: 10.1104/pp.77.4.863

Golombek, S., Rolletschek, H., Wobus, U., and Weber, H. (2001). Control of storage protein accumulation during legume seed development. *J. Plant Physiol*. 158, 457–464. doi: 10.1078/0176-1617-00357

Harada, J. J. (1997). "Seed maturation and control of germination" in *Cellular and molecular biology of plant seed development*. eds. R. L. Phillips and I. K. Vasil (Springer), 545–592.

Harper, J. L., Lovell, P., and Moore, K. (1970). The shapes and sizes of seeds. *Annu. Rev. Ecol. Evol. Syst.* 1, 327–356. doi: 10.1146/annurev.es.01.110170.001551

Hehenberger, E., Kradolfer, D., and Köhler, C. (2012). Endosperm cellularization defines an important developmental transition for embryo development. *Development* 139, 2031–2039. doi: 10.1242/dev.077057

Heim, U., Weber, H., Bäumlein, H., and Wobus, U. (1993). A sucrose-synthase gene of *Vicia faba* L.: expression pattern in developing seeds in relation to starch synthesis and metabolic regulation. *Planta* 191, 394–401. doi: 10.1007/ BF00195698

Hendriks, J. H., Kolbe, A., Gibon, Y., Stitt, M., and Geigenberger, P. (2003). ADP-glucose pyrophosphorylase is activated by posttranslational redox- modification in response to light and to sugars in leaves of *Arabidopsis* and other plant species. *Plant Physiol*. 133, 838–849. doi: 10.1104/pp.103.024513

Henery, M. L., and Westoby, M. (2001). Seed mass and seed nutrient content as predictors of seed output variation between species. *Oikos* 92, 479–490. doi: 10.1034/j.1600-0706.2001.920309.x

Herridge, R. P., Day, R. C., Baldwin, S., and Macknight, R. C. (2011). Rapid analysis of seed size in *Arabidopsis* for mutant and QTL discovery. *Plant Methods* 7:3. doi: 10.1186/1746-4811-7-3

Hill, L. M., Morley-Smith, E. R., and Rawsthorne, S. (2003). Metabolism of sugars in the endosperm of developing seeds of oilseed rape. *Plant Physiol*. 131, 228–236. doi: 10.1104/pp.010868

ISTA (2011). *International rules for seed testing*. Bassersdorf, Switzerland: International Seed Testing Association.

Jakobsson, A., and Eriksson, O. (2003). A comparative study of seed number, seed size, seedling size and recruitment in grassland plants. *Oikos* 88, 494–502. doi: 10.1034/j.1600-0706.2000.880304.x

Kesavan, M., Song, J. T., and Seo, H. S. (2013). Seed size: a priority trait in cereal crops. *Physiol. Plant.* 147, 113–120. doi: 10.1111/j.1399-3054.2012.01664.x

Lauxmann, M. A., Annunziata, M. G., Brunoud, G., Wahl, V., Koczut, A., Burgos, A., et al. (2016). Reproductive failure in *Arabidopsis thaliana* under transient carbohydrate limitation: flowers and very young siliques are jettisoned and the meristem is maintained to allow successful resumption of reproductive growth. *Plant Cell Environ*. 39, 745–767. doi: 10.1111/pce.12634

Le, B. H., Cheng, C., Bui, A. Q., Wagmaister, J. A., Henry, K. F., Pelletier, J., et al. (2010). Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* 107, 8063–8070. doi: 10.1073/pnas.1003530107

Li, N., and Li, Y. (2015). Maternal control of seed size in plants. *J. Exp. Bot.* 66, 1087–1097. doi: 10.1093/jxb/eru549

Liao, S., Wang, L., Li, J., and Ruan, Y. -L. (2020). Cell wall invertase is essential for ovule development through sugar signaling rather than provision of carbon nutrients. *Plant Physiol*. 183, 1126–1144. doi: 10.1104/pp.20.00400

Manfre, A. J., Lahatte, G. A., Climer, C. R., and Marcotte, W. R. Jr. (2009). Seed dehydration and the establishment of desiccation tolerance during seed maturation is altered in the *Arabidopsis thaliana* mutant *atem6-1*. *Plant Cell Physiol*. 50, 243–253. doi: 10.1093/pcp/pcn185

Melkus, G., Rolletschek, H., Radchuk, R., Fuchs, J., Rutten, T., Wobus, U., et al. (2009). The metabolic role of the legume endosperm: a noninvasive imaging study. *Plant Physiol*. 151, 1139–1154. doi: 10.1104/pp.109.143974

Moles, A. T., Ackerly, D. D., Webb, C. O., Tweddle, J. C., Dickie, J. B., and Westoby, M. (2005). A brief history of seed size. *Science* 307, 576–580. doi: 10.1126/science.1104863

Morley-Smith, E. R., Pike, M. J., Findlay, K., Köckenberger, W., Hill, L. M., Smith, A. M., et al. (2008). The transport of sugars to developing embryos is not via the bulk endosperm in oilseed rape seeds. *Plant Physiol.* 147, 2121–2130. doi: 10.1104/pp.108.124644

Neubohn, B., Gubatz, S., Wobus, U., and Weber, H. (2000). Sugar levels altered by ectopic expression of a yeast-derived invertase affect cellular differentiation of developing cotyledons of *Vicia narbonensis* L. *Planta* 211, 325–334. doi: 10.1007/s004250000305

Nowack, M. K., Ungru, A., Bjerkan, K. N., Grini, P. E., and Schnittger, A. (2010). Reproductive cross-talk: seed development in flowering plants. *Biochem. Soc. Trans.* 38, 604–612. doi: 10.1042/BST0380604

Nunes-Nesi, A., Carrari, F., Gibon, Y., Sulpice, R., Lytovchenko, A., Fisahn, J., et al. (2007). Deficiency of mitochondrial fumarase activity in tomato plants impairs photosynthesis via an effect on stomatal function. *Plant J.* 50, 1093–1106. doi: 10.1111/j.1365-313X.2007.03115.x

Olas, J. J., Fichtner, F., and Apelt, F. (2020). All roads lead to growth: imaging-based and biochemical methods to measure plant growth. *J. Exp. Bot.* 71, 11–21. doi: 10.1093/jxb/erz406

Olas, J. J., Van Dingenen, J., Abel, C., Działo, M. A., Feil, R., Krapp, A., et al. (2019). Nitrate acts at the *Arabidopsis thaliana* shoot apical meristem to regulate flowering time. *New Phytol*. 223, 814–827. doi: 10.1111/nph.15812 Olas, J. J., and Wahl, V. (2019). Tissue-specific *NIA1* and *NIA2* expression in *Arabidopsis thaliana*. *Plant Signal. Behav.* 14:e1656035. doi: 10.1080/15592324. 2019.1656035

Patrick, J., and Offler, C. (1995). Post-sieve element transport of sucrose in developing seeds. *Funct. Plant Biol*. 22, 681–702. doi: 10.1071/PP9950681

Patrick, J. W., and Offler, C. E. (2001). Compartmentation of transport and transfer events in developing seeds. *J. Exp. Bot.* 52, 551–564. doi: 10.1093/jexbot/52.356.551

Radchuk, V., and Borisjuk, L. (2014). Physical, metabolic and developmental functions of the seed coat. *Front. Plant Sci*. 5:510. doi: 10.3389/fpls.2014.00510

Rolletschek, H., Radchuk, R., Klukas, C., Schreiber, F., Wobus, U., and Borisjuk, L. (2005). Evidence of a key role for photosynthetic oxygen release in oil storage in developing soybean seeds. *New Phytol*. 167, 777–786. doi: 10.1111/j. 1469-8137.2005.01473.x

Ruan, Y. -L., Jin, Y., Yang, Y. -J., Li, G. -J., and Boyer, J. S. (2010). Sugar input, metabolism, and signaling mediated by invertase: roles in development, yield potential, and response to drought and heat. *Mol. Plant* 3, 942–955. doi: 10.1093/mp/ssq044

Stadler, R., Lauterbach, C., and Sauer, N. (2005). Cell-to-cell movement of green fluorescent protein reveals post-phloem transport in the outer integument and identifies symplastic domains in *Arabidopsis* seeds and embryos. *Plant Physiol*. 139, 701–712. doi: 10.1104/pp.105.065607

Stitt, M., Lilley, R. M., Gerhardt, R., and Heldt, H. W. (1989). [32] metabolite levels in specific cells and subcellular compartments of plant leaves. *Meth. Enzymol*. 174, 518–552. doi: 10.1016/0076-6879(89)74035-0

Stitt, M., and Zeeman, S. C. (2012). Starch turnover: pathways, regulation and role in growth. *Curr. Opin. Plant Biol*. 15, 282–292. doi: 10.1016/j.pbi.2012.03.016

Tegeder, M., Wang, X. D., Frommer, W. B., Offler, C. E., and Patrick, J. W. (1999). Sucrose transport into developing seeds of *Pisum sativum* L. *Plant J.* 18, 151–161. doi: 10.1046/j.1365-313X.1999.00439.x

Tomlinson, K. L., McHugh, S., Labbe, H., Grainger, J. L., James, L. E., Pomeroy, K. M., et al. (2004). Evidence that the hexose-to-sucrose ratio does not control the switch to storage product accumulation in oilseeds: analysis of tobacco seed development and effects of overexpressing apoplastic invertase. *J. Exp. Bot*. 55, 2291–2303. doi: 10.1093/jxb/erh251

Weber, H., Borisjuk, L., Heim, U., Buchner, P., and Wobus, U. (1995). Seed coat-associated invertases of fava bean control both unloading and storage functions: cloning of cDNAs and cell type-specific expression. *Plant Cell* 7, 1835–1846. doi: 10.1105/tpc.7.11.1835

Weber, H., Borisjuk, L., Heim, U., Sauer, N., and Wobus, U. (1997a). A role for sugar transporters during seed development: molecular characterization of a hexose and a sucrose carrier in fava bean seeds. *Plant Cell* 9, 895–908. doi: 10.1105/tpc.9.6.895

Weber, H., Borisjuk, L., and Wobus, U. (1997b). Sugar import and metabolism during seed development. *Trends Plant Sci*. 2, 169–174. doi: 10.1016/ S1360-1385(97)85222-3

Weber, H., Borisjuk, L., and Wobus, U. (2005). Molecular physiology of legume seed development. *Annu. Rev. Plant Biol.* 56, 253–279. doi: 10.1146/annurev.arplant.56.032604.144201

Weber, H., Heim, U., Golombek, S., Borisjuk, L., and Wobus, U. (1998). Assimilate uptake and the regulation of seed development. *Seed Sci. Res.* 8, 331–346. doi: 10.1017/S0960258500004268

Weschke, W., Panitz, R., Gubatz, S., Wang, Q., Radchuk, R., Weber, H., et al. (2003). The role of invertases and hexose transporters in controlling sugar ratios in maternal and filial tissues of barley caryopses during early development. *Plant J.* 33, 395–411. doi: 10.1046/j.1365-313X.2003.01633.x

Weschke, W., Panitz, R., Sauer, N., Wang, Q., Neubohn, B., Weber, H., et al. (2000). Sucrose transport into barley seeds: molecular characterization of two transporters and implications for seed development and starch accumulation. *Plant J.* 21, 455–467. doi: 10.1046/j.1365-313x.2000. 00695.x

West, M., and Harada, J. J. (1993). Embryogenesis in higher plants: an overview. *Plant Cell* 5, 1361–1369. doi: 10.2307/3869788

Winter, H., and Huber, S. C. (2000). Regulation of sucrose metabolism in higher plants: localization and regulation of activity of key enzymes. *Crit. Rev. Plant Sci.* 35, 253–289. doi: 10.1080/10409230008984165

Wobus, U., and Weber, H. (1999). Sugars as signal molecules in plant seed development. *Biol. Chem.* 380, 937–944. doi: 10.1515/BC.1999.116

Zhang, W. -H., Zhou, Y., Dibley, K. E., Tyerman, S. D., Furbank, R. T., and Patrick, J. W. (2007). Nutrient loading of developing seeds. *Funct. Plant Biol.* 34, 314–331. doi: 10.1071/FP06271

Zuma, B., Dana, M. B., and Wang, D. (2018). Prolonged expression of a putative invertase inhibitor in micropylar endosperm suppressed embryo growth in *Arabidopsis*. *Front. Plant Sci.* 9:61. doi: 10.3389/fpls.2018.00061

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

3.2.2. Physiological characterization during postembryonic development

In order to investigate whether the shoot physiological status might determine shoot phenotypical differences and thus, the large rosette phenotype in Bur-0 plants, biomass and metabolite contents were analysed in rosette samples during postembryonic development. Shoot biomass accumulation (shoot dry weight = SDW) over time as well as the shoot relative growth rate (RGR) of the Arabidopsis accessions Col-0, L*er*-1, Ws-2, and Bur-0 plants grown in LD and SD photoperiods were analysed, within the time frame covered from vegetative until floral transition stage per accession/photoperiod, respectively.

Note that floral transition analyses performed in both photoperiods were presented in *Section 3.1.1.3.* Briefly, we found that in LD the floral transition is initiated for Ws-2 at 6 DAG, L*er*-1 and Col-0 at 10 DAG, Bur-0 at 21 DAG, whereas in SD the floral transition is initiated for Ws-2 between 20-25 DAG, followed by L*er*-1 between 30-35 DAG, then for Col-0 between 40-45 DAG and lastly for Bur-0 towards 45-50 DAG. Additionally, soluble sugars including glucose, fructose and sucrose, starch and total protein levels were measured in rosettes from the Arabidopsis accessions Col-0 and Bur-0 grown in LD photoperiod. Metabolite contents during postembryonic development (from 4 to 14 DAG) and at vegetative and reproductive stages, respectively, were measured in collaboration with Dr. Maria Grazia Annunziata from the MPI of Molecular Plant Physiology.

3.2.2.1. Shoot biomass and relative growth rate (RGR)

Our results showed that shoot biomass (as the shoot dry weight, SDW) follows a similar pattern among accessions, where the SDW increases steadily during early growth, then drastically towards late growth (towards floral transition stage) and this pattern is also observed in both LD (Figure 17A-D) and SD (Figure 17I - L) photoperiod. Nevertheless, we found that the SDW accumulated towards the floral transition differs from accession to accession (end point analyzed 3 days after floral transition in LD and 10 days in SD, respectively). At the respective end point per accession, in LD photoperiod Ws-2 produces 2.4 mg, Ler-1 4.0 mg, Col-0 5.5 mg and Bur-0 110.2 mg of shoot biomass (Figure 17A-D), while in SD photoperiod Ws-2 produces 28.5 mg, Ler-1 75.7 mg, Col-0 240.9 mg and Bur-0 494.5 mg (Figure 17I - L).

These results reveal a photoperiod effect in shoot biomass accumulation, since Bur-0 produces around 46 times more shoot biomass than Ws-2, 27 times more than L*er*-1 and 20 times more than Col-0 in LD, while in SD Bur-0 produces around 17 times more shoot biomass than Ws-2, 6 times more than L*er*-1 and only 2 times more than Col-0. Although differences in shoot biomass accumulation among accessions are less pronounced in SD, Bur-0 plants produce higher shoot biomass and bigger rosettes than the other accessions in both photoperiods (Figure 17Q).

The photoperiod effect in shoot biomass accumulation also revealed interesting responses in each accession. Towards the floral transition (end point) Ws-2 produces 12 times more shoot biomass in SD than in LD, L*er*-1 around 18 times more, Col-0 around 44 times more, while Bur-0 produces only 4.5 times more shoot biomass in SD than in LD. Although the delayed

shoot development induced in SD photoperiod increased the rosette size for all accessions, the photoperiod effect on shoot development and biomass production was particularly interesting in Col-0 (more pronounced) and Bur-0 (less pronounced). End-point values and statistical significance are presented in Supplementary Table S6.

Furthermore, the shoot relative growth rate analysis (biomass based relative growth rate, RGR) from early vegetative to late reproductive stages revealed differences per accession and photoperiod. In LD Ws-2 has a rapid shift from low to high RGR (0.16 to 0.42), in Col-0 and L*er*-1 the RGR increases more steadily (0.20 to 0.39 and 0.11 to 0.25, respectively), while in Bur-0 plants the shoot RGR from early vegetative to late reproductive stages remains uniform (0.29 to 0.31) (Figure 17E-H).

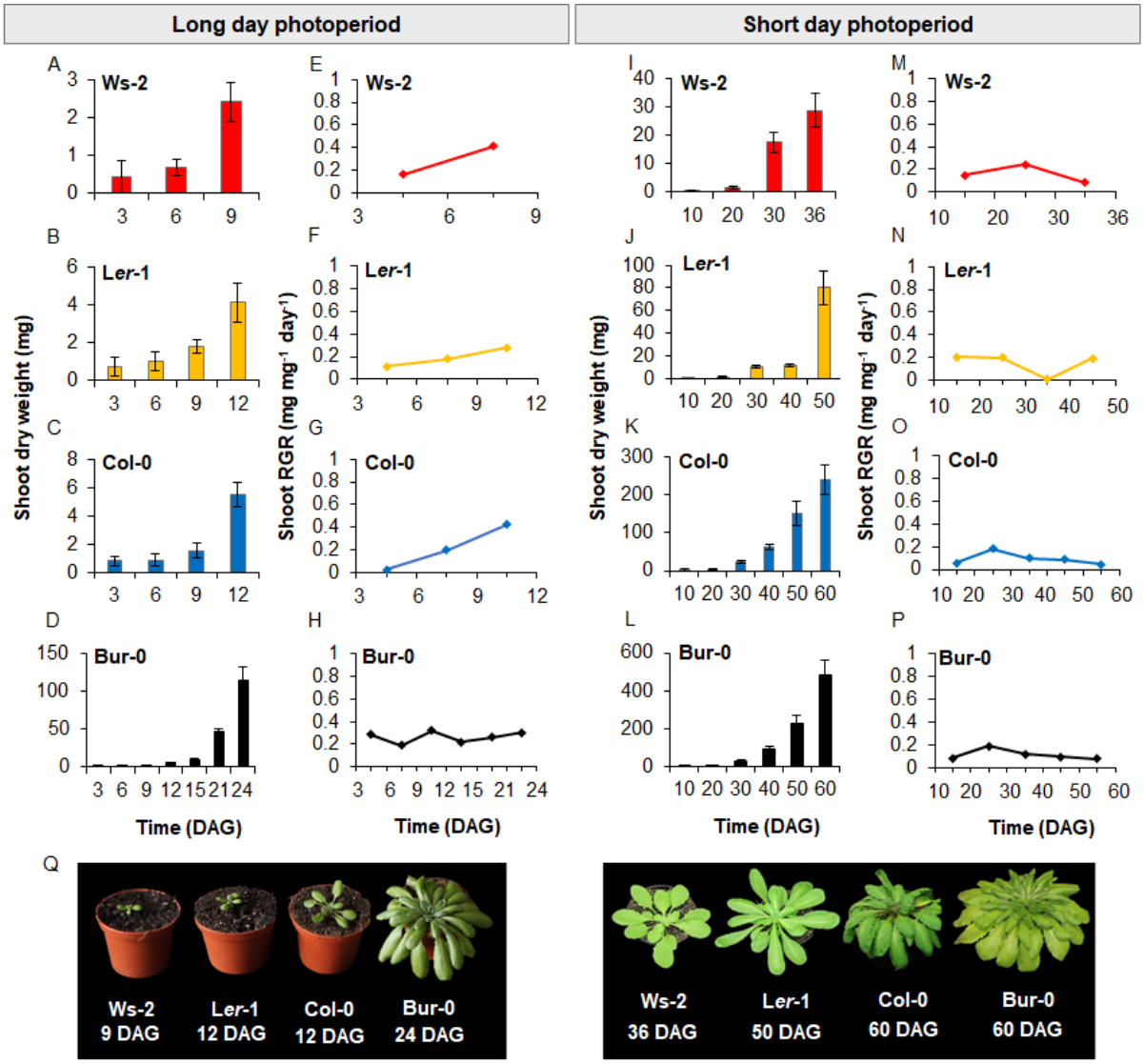


Figure 17. Shoot biomass and dry weight-based relative growth rate (RGR). The Arabidopsis accessions Ws-2, L*er*-1, Col-0 and Bur-0 were analyzed. Shoot biomass is presented as the shoot dry weight and the shoot relative growth rate (RGR) as biomass based relative growth rate. (A-D) Shoot dry weight of plants grown in long days (LD). (E-H) Shoot RGR of plants grown in LD. (I-L) Shoot dry weight of plants grown in short days (SD). (M-P) Shoot RGR of plants grown in SD. (Q) Rosette phenotype of plants grown in LD and SD conditions. Abbreviations: DAG (days after germination). Error bars indicate ± SD (*n*=10). End-point statistical significance is presented in Supplementary Table S6.

In SD the shoot RGR is lower than in LD for all accessions and fewer fluctuations between early and late developmental stages were observed (0.15 to 0.07 in Ws-2, 0.20 to 0.19 in L*er*-1, 0.06 to 0.04 in Col-0 and 0.08 to 0.07 in Bur-0). Interestingly, in SD Col-0 does not follow a similar RGR pattern to L*er*-1 as in LD, but rather resembles the pattern observed for Bur-0 (Figure 17M-P). The highest shoot RGR was recorded for the rapidly growing accession Ws-2 in LD photoperiod and for L*er*-1 in SD photoperiod. End-point values and statistical significance are presented in Supplementary Table S6.

Our results reveal that Bur-0 produces more shoot biomass than the other accessions from early vegetative to late reproductive development, in agreement with the rosette phenotypes observed at the end of this analysis. Instead of a rapid shift from lower to higher shoot RGR from early vegetative to late reproductive development as observed in other accessions, the RGR in Bur-0 remains low and sustained with small fluctuations through postembryonic growth, thus Bur-0 produces more shoot biomass than other accessions at a low RGR.

3.2.2.2.Metabolite contents during postembryonic development

In order to characterize the rosette metabolic status in Bur-0 and Col-0 plants, soluble sugars (sucrose, glucose, fructose), starch and total protein content were measured in rosette samples. Plants were grown in LD photoperiod and metabolite contents were measured over time (4 to 14 DAG) and at different developmental stages during postembryonic growth (vegetative stage: 4 DAG for both accessions; floral transition stage: Col-0 =10 DAG, Bur-0= 21 DAG). Our analysis over time revealed that at 4 DAG Bur-0 has a higher level of total protein content than Col-0, from 6 to 12 DAG the protein levels are similar in both accessions, whereas from 12 to 14 DAG protein levels increase rapidly in Col-0 rosettes (Figure 18A). Starch levels are higher in Bur-0 than in Col-0 plants, but starch content in Col-0 increase steadily over time with small fluctuations, while in Bur-0 starch levels fluctuate drastically over time and starch content is high at 4 DAG, decreases between 4 and 8 DAG and then increases rapidly from 8 DAG onwards (Figure 18B). Bur-0 has overall higher levels of soluble sugars than Col-0, but sucrose and glucose contents fluctuate in a similar pattern for both accessions over time (Figure 18C, D), while fructose levels increase steadily in Col-0 and drastically in Bur-0, particularly from 8 DAG onwards (Figure 18E).

Developmental stage-based comparisons revealed that during vegetative growth the total protein content is significantly higher in Bur-0 rosettes than in Col-0, while during reproductive growth both accessions have similar protein levels (Figure 18F). Interestingly, starch content in Bur-0 rosettes is significantly higher than in Col-0 ($p<0.05$) during vegetative and reproductive growth (Figure 18G). Among the soluble sugars analyzed, we found that sucrose content is significantly higher in Bur-0 rosettes during vegetative growth and during reproductive growth both accessions have similar sucrose levels (Figure 18H). In contrast, glucose and fructose levels are not significantly different between accessions during vegetative growth ($p>0.05$), whereas during reproductive growth glucose and fructose levels are significantly higher in Bur-0 ($p<0.05$).

Our results reveal that both accessions have a different rosette metabolic status at different stages of development. Vegetative growth starts with higher protein and sucrose levels in Bur-0 compared to Col-0, but towards reproductive growth similar levels of those metabolites are found in both accessions. On the other hand, similar glucose and fructose levels are found in both accessions during vegetative growth, but later on those metabolites are significantly increased in Bur-0 rosettes towards reproductive growth, suggesting that glucose and fructose might play a role of in floral transition related processes for Bur-0. Additionally, the higher starch content found in Bur-0 rosettes during vegetative and reproductive growth indicate that Bur-0 plants accumulate more carbon resources throughout postembryonic development than Col-0.

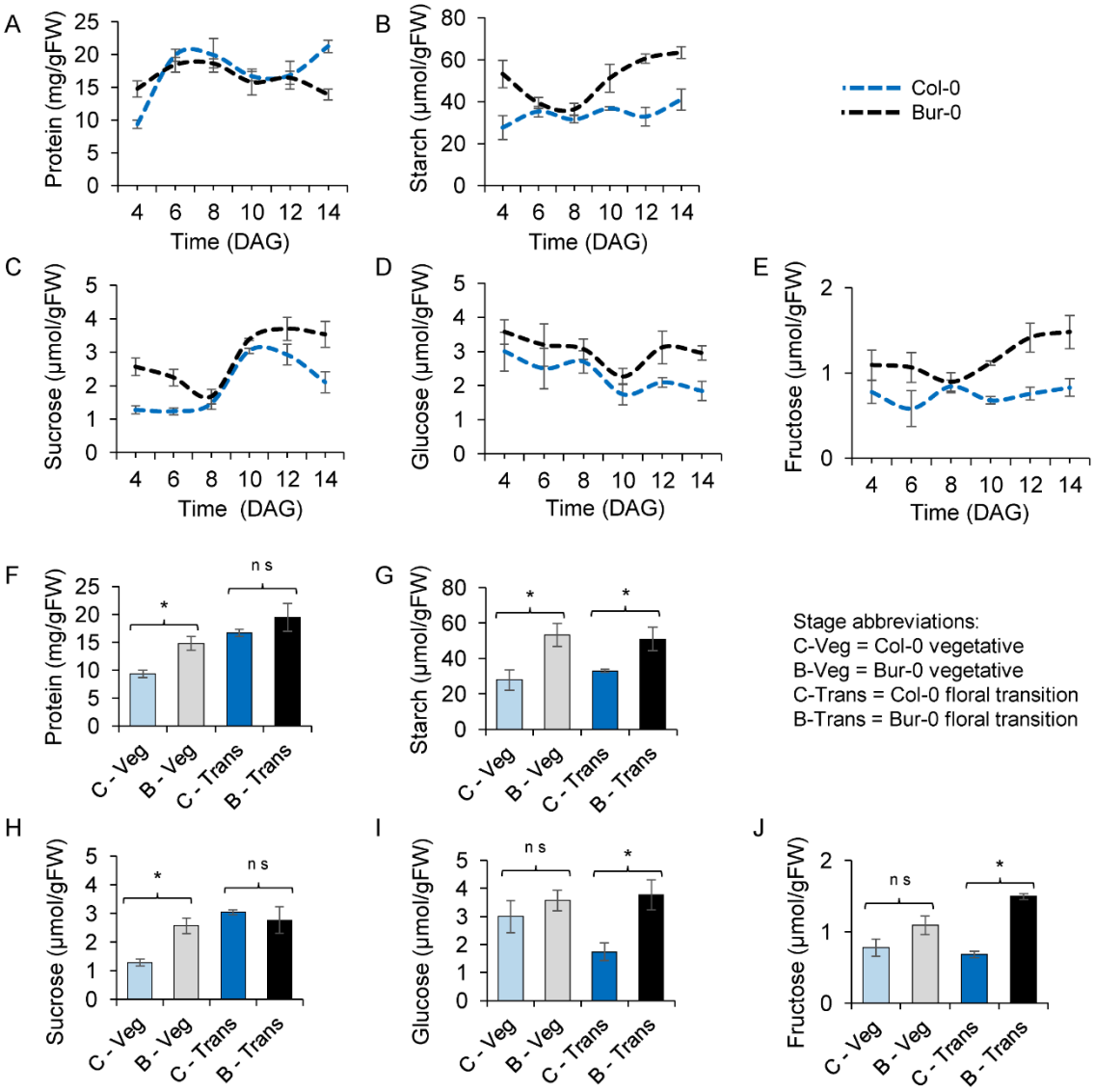


Figure 18. Metabolite content in rosettes of Arabidopsis Col-0 and Bur-0 accessions grown in long day (LD) photoperiod. Metabolite content analyzed overtime from 4 to 14 DAG. (A) Changes in total proteins, (B) starch, (C) sucrose, (D) glucose and (E) fructose levels. Metabolite content analyzed at vegetative and reproductive stages. Levels of (F) total proteins, (G) starch, (C) sucrose, (D) glucose and (E) fructose during vegetative growth (4 DAG for both accessions) and reproductive growth (floral transition at 10 DAG for Col-0 and 21 DAG for Bur-0). Error bars indicate ± SEM. (n = 3). Statistical significance was tested using Student's t-test. *, $p < 0.05$. Abbreviations: ns = not significant. DAG = days after germination.

In summary, our results reveal that Bur-0 produces more shoot biomass than other accessions at a low relative growth rate (RGR). Interestingly, metabolite levels analyzed over time revealed different fluctuation patterns in each accession, in general higher levels of carbohydrates in Bur-0 and similar protein levels in both accessions. On the other hand, metabolite levels analyzed at vegetative and reproductive stages revealed that both accessions have a different rosette metabolic status at different stages of development and in particular, the higher levels of glucose and fructose towards reproductive growth in Bur-0 suggest a role of those metabolites in floral transition related processes. Additionally, the higher starch content in Bur-0 rosettes during vegetative and reproductive growth indicate that Bur-0 plants accumulate more carbon resources throughout embryonic and postembryonic development than Col-0. Thus, accession-specific metabolic traits that underlie the accession-specific phenotypes during embryonic and postembryonic development were identified in this study and our results suggest that the physiological status can determine phenotypical differences among natural Arabidopsis accessions.

3.3. Ploidy level, cell size and cell cycle analyses

As part of the phenotypical and physiological characterization presented in *Sections 3.1* and *3.2*, we compared Bur-0 to other Arabidopsis accessions and found that none of the analyzed natural accessions, or late flowering mutants, have simultaneously a larger rosette, SAM, seed, and larger embryo phenotype as Bur-0. Furthermore, our developmental stage-dependent comparisons revealed that Bur-0 plants produce bigger mature embryos during late embryonic growth, bigger rosettes during late postembryonic growth and we reported that developmental phases progression is overall delayed in Bur-0. Considering that the size of an organ is determined by the size and number of its constituent cells and considering that cell and organ enlargement has been associated with increased ploidy level, we investigated whether the enlarged organs observed in Bur-0 are determined by differences in cell size/number, ploidy level, or expression of cell cycle regulators. Ploidy level by fluorescence activated cell sorting (FACS) was analyzed in collaboration with Dr. Frank Machin from the MPI of Molecular Plant Physiology. This analysis was performed in different organs and cell types of the Arabidopsis accessions Bur-0 and Col-0, grown in LD photoperiod.

3.3.1. Organ size and ploidy level in Col-0 and Bur-0 accessions

Comparison of different organs in the Arabidopsis accessions Col-0 and Bur-0 showed that Bur-0 has bigger seeds, rosettes, and flowers than Col-0 (Figure 19A). Ploidy level was initially analysed using a cell type that does not undergo endoreduplication like pollen grains, and cell size as well as nuclei content were analysed as parameters for ploidy level estimation. Our results showed that Col-0 mature pollen grains are rounded, while Bur-0 pollen grains are ellipsoid and significantly bigger ($p < 0.05$) than in Col-0 (Figure 19B). Interestingly, similar vegetative and sperm nuclei were observed in DAPI-stained pollen grains for both accessions (Figure 19C), indicating similar DNA content and thus, same ploidy level in Col-0 and Bur-0.

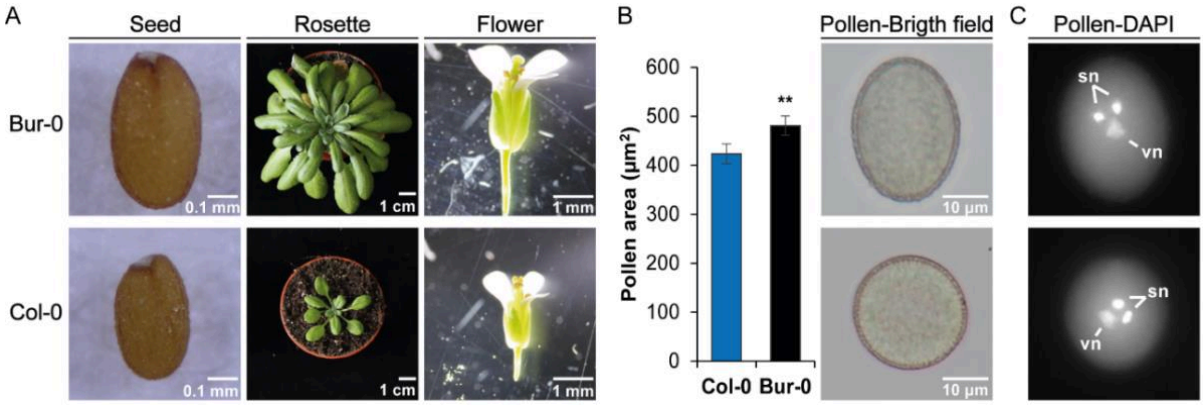


Figure 19. Comparison of organ size and nuclei content in mature pollen grains of the Arabidopsis accessions Bur-0 and Col-0. (A) Seed, rosette and flower phenotype. (B) Pollen area and pollen grain phenotype. Error bars indicate ± SD. ($n = 20$). Statistical significance was tested using Student's *t*-test: **, $p < 0.01$. (C) Nuclei content visualized by DAPI staining (vn = vegetative nucleus, sn = sperm nuclei).

3.3.2. Tissue-specific cell size and ploidy analyses

Our previous results showed that Bur-0 has larger organs as well as bigger pollen grains than Col-0, but both accessions have similar ploidy level, however, ploidy level has been shown to affect the final organ size and it can be different in different cell types/tissues. We, therefore, conducted a more detailed tissue-specific analysis of leaves, mature embryos and shoot apices at vegetative and floral transition stages. Ploidy level was analysed by FACS and visualizing the nuclei content by DAPI-staining, while cell wall fluorescent staining was performed for cell area analysis. Flow cytometry histograms and ploidy distribution charts were provided by Dr. Frank Machin from the MPI of Molecular Plant Physiology.

3.3.2.1. Cell size and ploidy level analysis on leaf tissue

Different cell types were analysed on leaf tissue from Bur-0 and Col-0 plants at 12 DAG, grown in LD photoperiod. Our previous rosette area analysis in LD photoperiod presented in *Sections 3.1.1.1.* and *3.1.1.2.* revealed that both accessions have similar rosettes towards 12-14 DAG, therefore this age was initially selected for tissue-specific analyses on leaves. Ploidy level from rosette leaves was analysed by FACS, nuclei content of guard cells was visualized by DAPI-staining and cell area was measured for epidermis cells. Our results showed that Bur-0 has significantly bigger epidermis cells ($p < 0.05$), thus cell size differences between accessions were found on leaf tissue as well (Figure 20A).

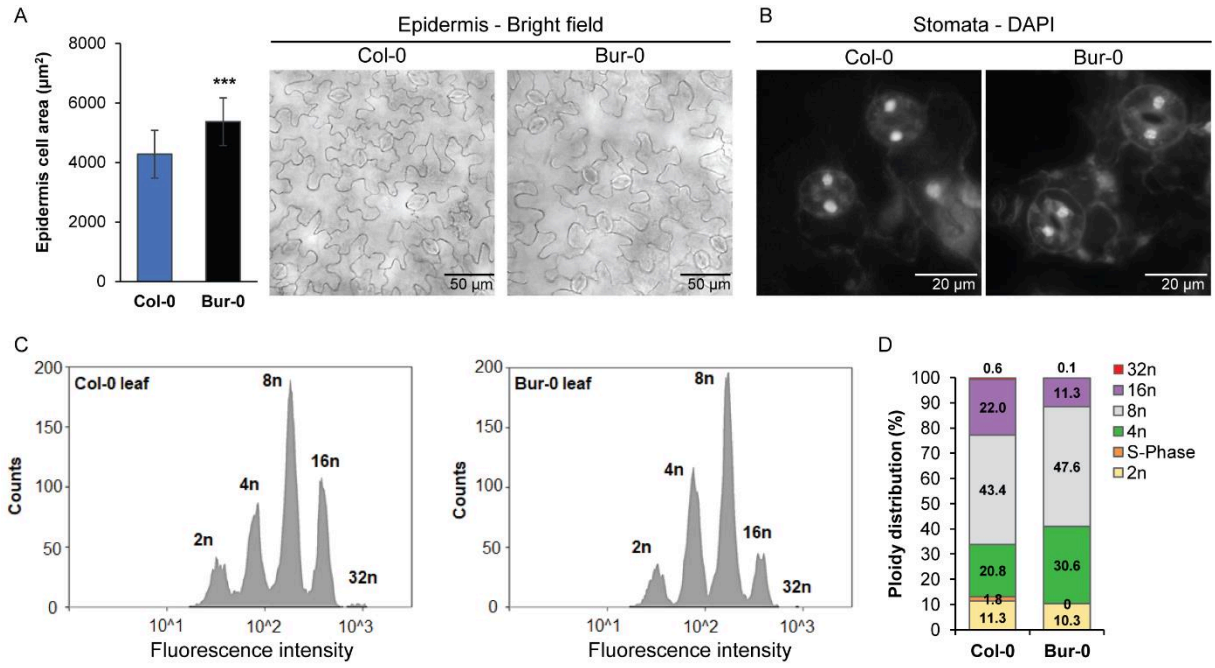


Figure 20. Cell size, nuclei content and ploidy level analyses on leaves. Rosette leaves from the Arabidopsis accessions Bur-0 and Col-0 were analysed from 12-day-old plants grown in long day photoperiod. (A) Epidermis cell area and bright-field images of epidermis cells. Error bars indicate ± SD ($n \geq 10$). Statistical significance was calculated using Student's *t*-test: ***, $p < 0.001$. (B) DAPI-stained nuclei of stomata guard cells. (C) Flow cytometry histograms of leaf nuclei stained with SYTO13 red. (D) Quantification of flow cytometry peaks 2n-32n and ploidy distribution.

Stomatal guard cells are another cell type that does not undergo endoreduplication, and similar nuclei content was observed in Bur-0 and Col-0 DAPI-stained guard cells, indicating similar ploidy level in both accessions (Figure 20B). Furthermore, FACS histograms showed that the peak with the smallest DNA content is 2n for both accessions (Figure 20C), confirming that both accessions have similar ploidy level. The different proportion of 2n-32n nuclei found per accession (Figure 20D) indicate that leaf cells of Bur-0 and Col-0 plants have similar ploidy level, but a different ploidy distribution, suggesting that leaf cells are at different steps along the endoreplication cycling per accession.

3.3.2.2. Cell size and ploidy level analyses on mature embryos

In *Section 3.1.2.1.* we reported that compared to other natural Arabidopsis accessions, Bur-0 has bigger mature embryos. In order to better understand the causes of this large embryo phenotype, cell size, embryo area and ploidy level were analysed using the accessions Bur-0 and Col-0. DAPI-stained mature embryo longitudinal sections revealed that both accessions have similar nuclei content, however, the nuclei organization is different in hypocotyl and cotyledon cells, and the nuclei appear to be more separated from each other in Bur-0 than in Col-0 embryos (Figure 21A). Additionally, longitudinal embryo sections were stained with Calcofluor white for cell wall visualization (Figure 21B), and embryo as well as cell area were analysed.

Our results showed that mature Bur-0 embryos are significantly bigger than those of Col-0 (Figure 21C), and we found that hypocotyl as well as cotyledon cells are significantly bigger in Bur-0 embryos ($p<0.05$) (Figure 21D, E). Ploidy level analyzed by flow cytometry revealed that mature embryo cells in both accessions have the same ploidy level and ploidy distribution (Figure 21F, G). Thus, our results reveal cell size differences between accessions on embryo tissue and indicate that the large mature embryo phenotype in Bur-0 can be attributed to the large area of its constituent cells.

3.3.2.3. Cell size and ploidy level analyses in the SAM

As part of the morphological characterization of the SAM and searching for correlations between SAM size and adult plant traits presented in *Sections 3.1.1.4.* and *3.1.1.5.*, we found that the stage-dependent SAM size is a good marker trait (predictor) for flowering and rosette size phenotypes in Arabidopsis. Our results also suggested that SAM size might be an important factor involved in determination of accession-specific adult plant phenotypes and, interestingly, we found that Bur-0 has a significantly wider SAM compared to other natural accessions during vegetative and floral transition stages. In order to better understand the causes of this large SAM phenotype, L1 cell size, SAM area, cell number and ploidy level were analysed using the accessions Bur-0 and Col-0, grown in LD photoperiod.

Longitudinal sections of vegetative meristems (4 DAG for both accessions) stained with Calcofluor white (Figure 22A) were used for L1 cell size and SAM area analyses. Our results show that cell area of L1 cells is larger in Bur-0 (Figure 22B) and the vegetative SAM is significantly bigger in Bur-0 ($p< 0.05$) (Figure 22C). Moreover, DAPI-stained longitudinal

sections of vegetative meristems revealed that both accessions have similar nuclei content in vegetative SAM cells (Figure 22D). Cell number was estimated by counting the visible DAPI-stained nuclei and we found that Bur-0 has significantly more cells in the SAM at vegetative stage than Col-0 ($p < 0.05$) (Figure 22E). In addition, ploidy level analysis by flow cytometry revealed that shoot apex cells at vegetative stage have the same ploidy level in both accessions (Figure 22F), but different ploidy distribution and more S-phase cells were found in Bur-0 (Figure 22G).

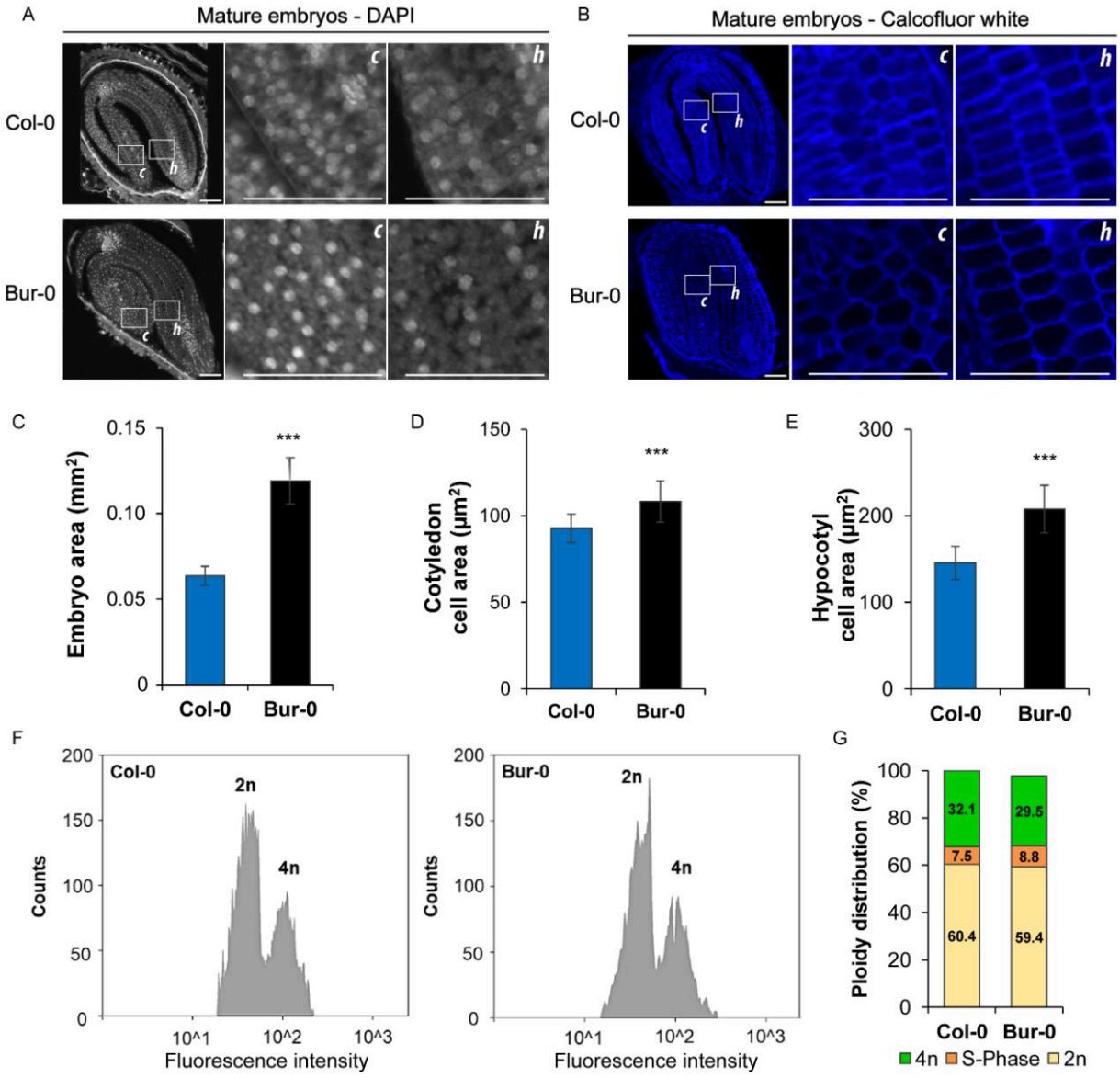


Figure 21. Embryo size, cell size, nuclei content and ploidy level analyses of mature embryos. Embryos at green mature stage of the Arabidopsis accessions Bur-0 and Col-0 were analyzed. (A) Visualization of nuclei content in DAPI-stained hypocotyl and cotyledon cells. (B) Embryo longitudinal sections stained with Calcofluor white. (C) Embryo area. (D) Cotyledon cell area. (E) Hypocotyl cell area. Error bars indicate ± SD ($n \geq 20$). Statistical significance tested with Student's *t*-test: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. (F) Flow cytometry histograms of mature embryo nuclei stained with SYTO13 red. (G) Quantification of flow cytometry peaks 2n-4n and ploidy distribution. c = cotyledon, h = hypocotyl. Scale bar 50 µm.

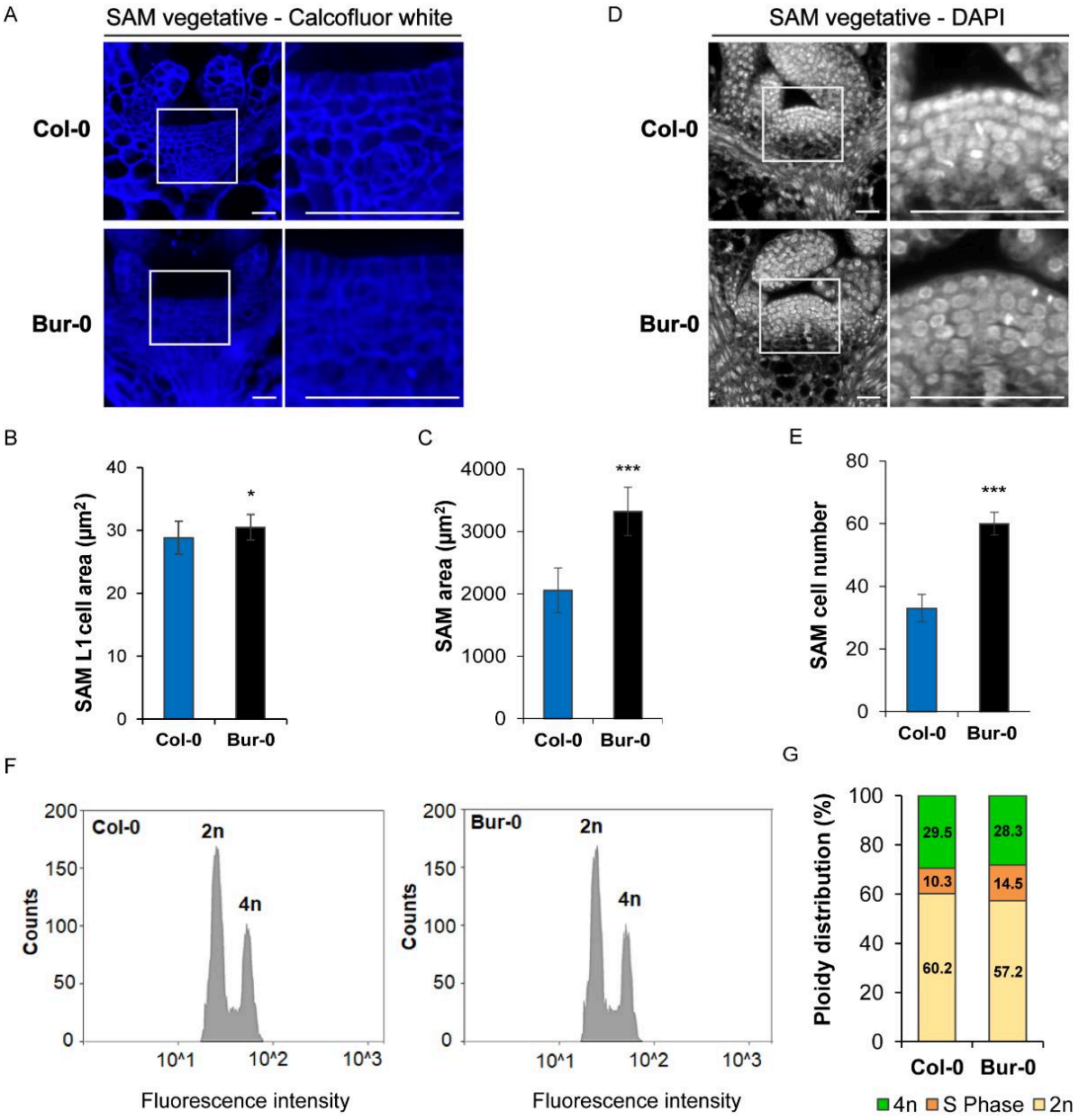


Figure 22. Cell size, cell number, DNA content and ploidy level analyses of the vegetative shoot apical meristem (SAM). The Arabidopsis accessions Bur-0 and Col-0 were analyzed, plants were grown in LD photoperiod and samples at vegetative stage were harvested at 4 days after germination for both accessions. (A) Longitudinal sections of vegetative meristems stained with Calcofluor white. (B) vegetative SAM L1 cell area. (C) Vegetative SAM area. (D) DNA content visualization in DAPI-stained vegetative meristems. (E) Vegetative SAM cell number. Error bars indicate ± SD ($n \geq 5$ SAMs and $n \geq 10$ cells per SAM). Statistical significance tested with Student's *t*-test: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. (F) Flow cytometry histograms of shoot apex nuclei stained with SYTO13 red. (G) Quantification of flow cytometry peaks 2n-4n and ploidy distribution. Scale bar 50 µm.

On the other hand, longitudinal sections of meristems at floral transition stage (10 DAG for Col-0 and 21 DAG for Bur-0) stained with Calcofluor white (Figure 23A) were used for L1 cell size and SAM area analyses. Our results showed that Bur-0 has significantly larger L1 cells (Figure 23B) and significantly bigger reproductive SAMs than Col-0 ($p < 0.05$) (Figure 23C). Additionally, DAPI-stained longitudinal sections of meristems at floral transition stage revealed that both accessions have similar nuclei content in SAM cells (Figure 23D) and Bur-0 has significantly more cells in the SAM at floral transition stage than Col-0 ($p < 0.05$) (Figure 23E).

Moreover, ploidy level analysis by flow cytometry revealed that shoot apex cells at floral transition stage have the same ploidy level for both accessions (Figure 23F), but different ploidy distribution and more 2n, S-phase, but less 4n cells were found in Bur-0 (Figure 23G). Thus, similar ploidy level and cell size differences between accessions were found on SAM tissue as well and differences in ploidy distribution suggest differences in cell cycle progression in the SAM.

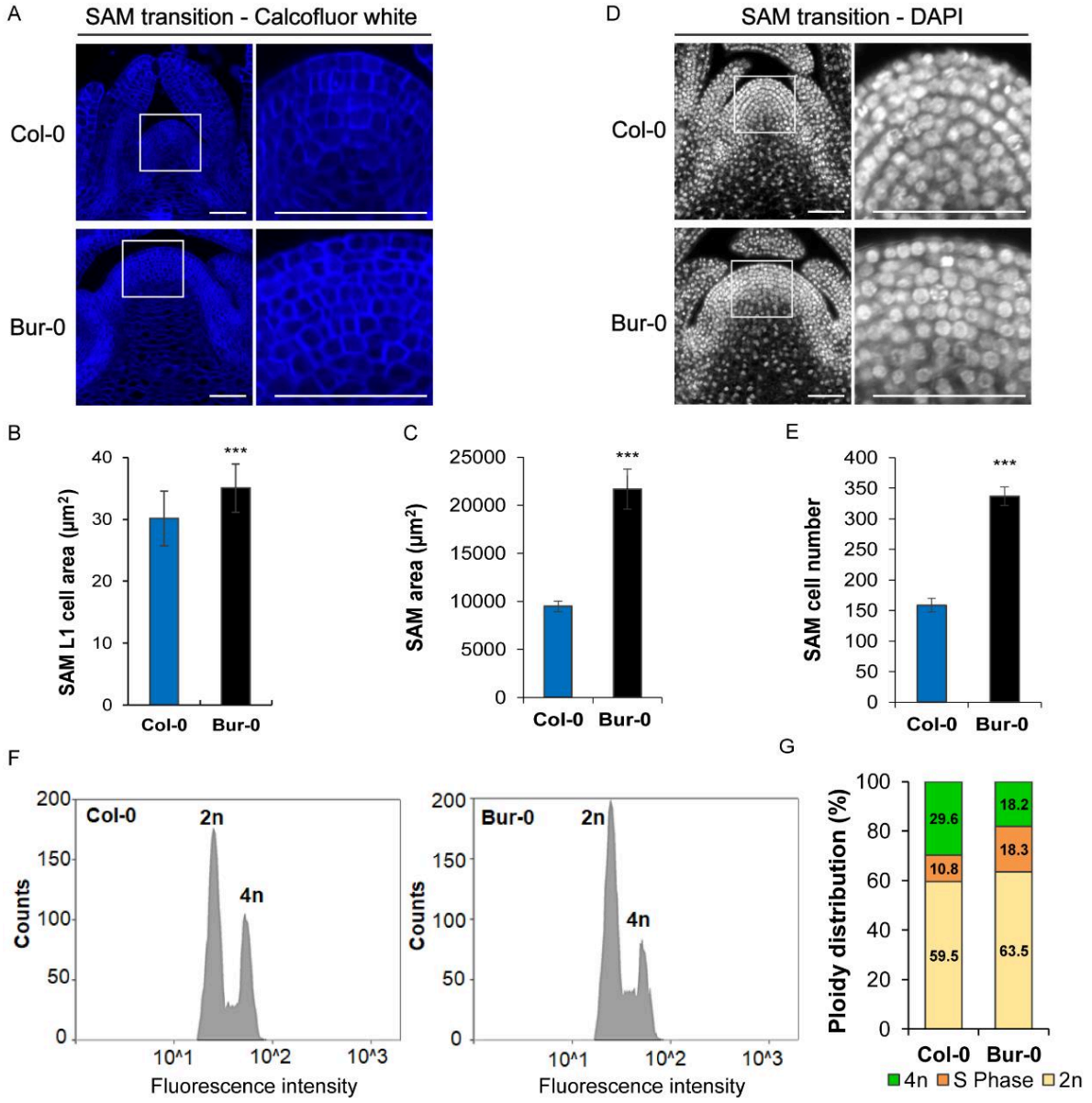
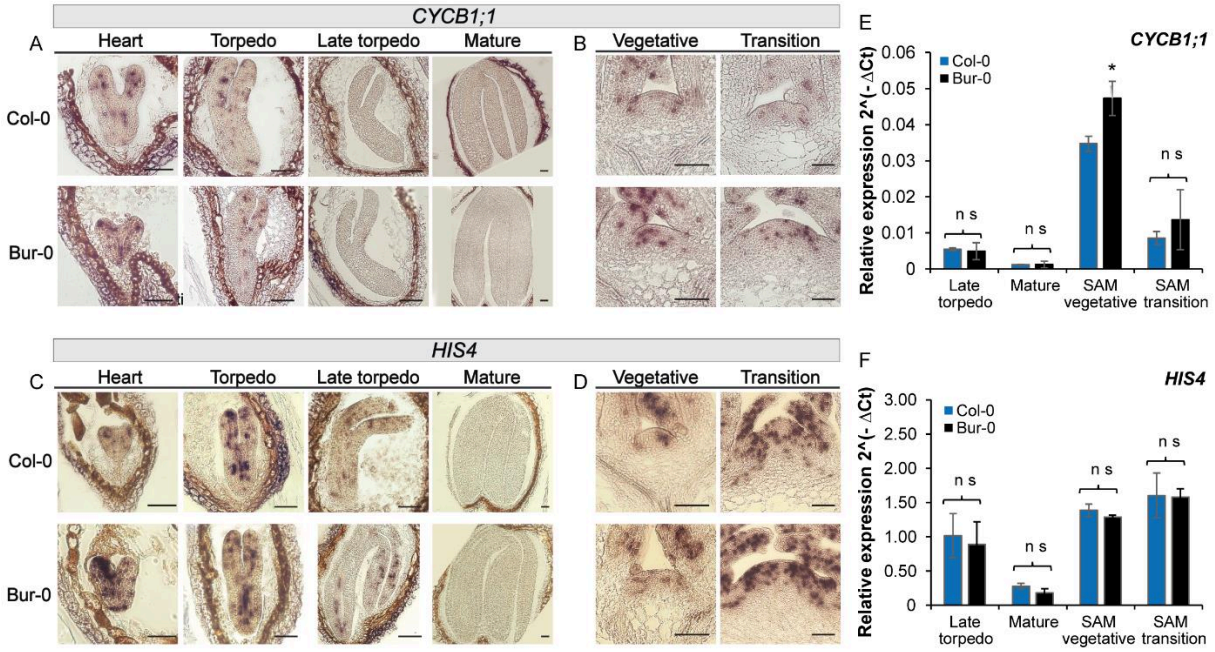


Figure 23. Cell size, cell number, DNA content and ploidy level analyses of the reproductive shoot apical meristem (SAM). The Arabidopsis accessions Bur-0 and Col-0 were analyzed, plants were grown in LD photoperiod and samples at floral transition stage were harvested at 10 and 21 days after germination for Col-0 and Bur-0, respectively. (A) Longitudinal sections of reproductive meristems stained with Calcofluor white. (B) Reproductive SAM L1 cell area. (C) Reproductive SAM area. (D) DNA content visualization in DAPI-stained reproductive meristems. (E) Reproductive SAM cell number. Error bars indicate ± SD ($n \geq 5$ SAMs and $n \geq 10$ cells per SAM). Statistical significance tested with Student's *t*-test: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. (F) Flow cytometry histograms of shoot apex nuclei stained with SYTO13 red. (G) Quantification of flow cytometry peaks 2n-4n and ploidy distribution. Scale bar 50 µm.

3.3.3. Expression analysis of cell cycle regulators

Because our cell size and ploidy analyses revealed tissue-specific differences in cell area and cell number between Bur-0 and Col-0, we investigated whether cell cycle progression in embryos and the SAM is affected in Bur-0 plants compared to Col-0. Expression of two cell cycle markers *CYCLINB1;1* (*CYCB1;1*; G2/M-phase marker) and *HISTONE4* (*HIS4*; S-phase marker) was analyzed by RNA *in situ* hybridization and by qRT-PCR on embryos at different developmental stages and SAMs at vegetative (4 DAG for both accessions) and floral transition stages (10 DAG for Col-0 and 21 DAG for Bur-0).

RNA *in situ* hybridization analysis revealed that both *CYCB1;1* and *HIS4* are expressed (dark dots/positive cells) at heart and torpedo embryo stages and the expression pattern is similar in both accessions. *HIS4* positive cells are less abundant at late torpedo stage, while *CYCB1;1* expression signal is absent in both accessions at the same embryo stage, indicating that cell proliferation decreases during late embryogenesis. In mature embryos none of the cell cycle markers are detected (Figure 24A, C), indicating that cell proliferation is arrested in mature embryos.



[1] Figure 24. Expression analysis of the cell cycle markers *CYCLINB1;1* (*CYCB1;1*) and *HISTONE4* (*HIS4*) on embryos and shoot apical meristems. The Arabidopsis accessions Bur-0 and Col-0 were analyzed and plants were grown in LD photoperiod. (A-D) RNA *in situ* hybridization using *CYCB1;1* and *HIS4* as probes on longitudinal sections of embryos at heart, torpedo, late torpedo and mature stages and vegetative (4 days after germination for both accessions) and shoot apical meristems at floral transition (10 days after germination for Col-0 and 21 for Bur-0). (E, F) Expression levels of *CYCB1;1* and *HIS4* analyzed by qRT-PCR. (n = 3). Error bars indicate ± SD. Statistical significance was tested using Student's t-test: *, $p < 0.05$; n.s., non-significant. Scale bar: 100 µm.

[1] The results corresponding to *CYCB1;1* expression analysis by RNA *in situ* hybridization on embryos at torpedo, late torpedo and mature stages in the Arabidopsis accessions Col-0 and Bur-0 have already been published in the paper presented in *Section 3.2.1.1*

During postembryonic development, positive cells for both cell cycle markers are detected in the SAM and young leaf primordia of Col-0 and Bur-0. However, *CYCB1;1* positive cells in the SAM at vegetative and floral transition stages are more abundant in Bur-0 than in Col-0, indicating higher mitotic activity in Bur-0 SAMs, while *HIS4* expression pattern is similar for both accessions at the same stages (Figure 24B, D), Interestingly, *HIS4* positive cells are more abundant in the SAM at floral transition stage than at vegetative stage, suggesting that cell proliferation increases actively towards the reproductive transition (Figure 24C, D).

Expression analysis by qRT-PCR showed variation of *CYCB1;1* and *HIS4* expression levels according to the developmental stage and tissue analyzed. Expression levels of *CYCB1;1* are low in embryos at late torpedo and mature stage for both accessions. Moreover, *CYCB1;1* transcript is significantly higher ($p < 0.05$) in Bur-0 SAM at vegetative stage, in agreement with the expression pattern of *CYCB1;1* observed by RNA *in situ* hybridization, while at floral transition stage both accessions have similar levels of *CYCB1;1* (Figure 24E). These results indicate higher mitotic activity in the SAM of Bur-0 plants during postembryonic growth. On the other hand, expression levels of *HIS4* are similar for both accessions in all tissues and stages analyzed, however *HIS4* transcript is lower on mature embryos and higher in the SAM at floral transition (Figure 24F), in agreement with the expression pattern of *HIS4* observed by RNA *in situ* hybridization.

In summary, we demonstrated that Bur-0 pollen grains, leaves, mature embryos as well as vegetative and reproductive SAMs have bigger cells than the Col-0 equivalents. However, the nuclei content and ploidy level in somatic and meristematic cells is similar in both accessions, suggesting that the larger organ size in Bur-0 results from its enlarged cells, but not from a different ploidy level. Cell number was determined only for vegetative and reproductive SAMs and we found that Bur-0 SAMs contain more cells than in Col-0, suggesting that the large SAM in Bur-0 results from more cells as well. These results are in agreement with our expression analysis of the cell cycle markers *CYCB1;1* and *HIS4*. Although expression of *HIS4* is similar for both accessions in all tissues and stages analyzed, higher expression of *CYCB1;1* is detected in Bur-0 SAM than in Col-0, particularly during vegetative growth, indicating higher mitotic activity.

Since cell number was not determined on embryo tissue and expression analysis of the cell cycle markers *CYCB1;1* and *HIS4* on embryos revealed that cell proliferation is decreased during late embryogenesis and arrested on mature embryos, we cannot determine from the current analysis if a higher cell proliferation also contributes to the large mature embryo phenotype in Bur-0. Hence, our results provide additional information about tissue-specific cell size/number, ploidy level and cell cycle progression in somatic and meristematic tissues and indicate that the large organ size in Bur-0 can be mainly attributed to its larger cells, and for the vegetative and reproductive SAM, to a higher mitotic activity as well. Our results contribute to a better understanding of the mechanisms that regulate organ size per accession and determine particular phenotypes.

3.4. Transcriptome analysis

In order to identify key regulators shared between developmental phases during embryonic and postembryonic growth, but also to identify accession-specific key regulators that might determine accession-specific phenotypes, RNA-seq analysis was performed for tissues and stages where major phenotypical differences were identified between Bur-0 and Col-0 during embryonic and postembryonic development. Thus, hand dissected late torpedo and green mature embryos as well as SAMs at vegetative (4 DAG for both accessions) and floral transition stages (10 DAG for Col-0 and 21 DAG for Bur-0) from the Arabidopsis accessions Bur-0 and Col-0, grown in LD photoperiod were selected for RNA-seq analysis (Figure 25). Computational analysis of the RNA-seq data (including the graphic presentation shown in figures) was performed by Dr. Federico Apelt and Dr. Saurabh Gupta from the MPI of Molecular Plant Physiology.

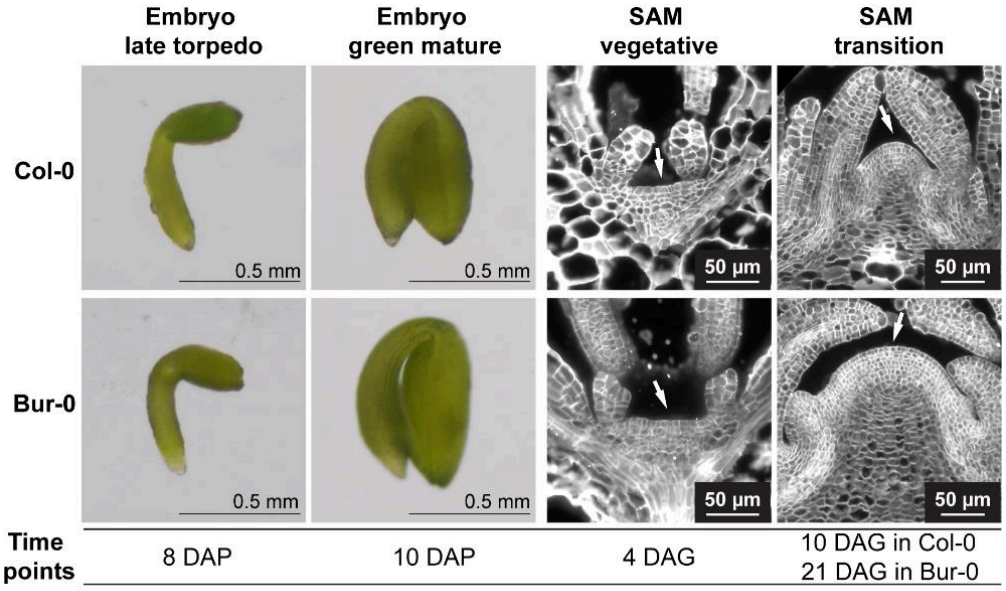


Figure 25. Experimental set up for RNA-seq analysis. Hand dissected embryos at late torpedo and green mature stages as well as shoot apical meristems (SAM) tissue (white arrows) at vegetative and floral transition stages used for RNA-seq analysis from the Arabidopsis accessions Col-0 and Bur-0 grown in LD photoperiod. Time is given as DAP = days after pollination and DAG = days after germination.

3.4.1. Transcriptomes profiles are different according to the tissue, developmental stage and accession

Embryo and SAM transcriptomes at different developmental stages from the Arabidopsis accessions Bur-0 and Col-0 were compared and principal component analysis (PCA) revealed that samples are well separated along the first component into embryo and SAM tissue, which explains 34.5% of the variability and then well separated along the second component (which explains 12.3% of the variability) into vegetative and transition stage for the SAM. Embryo samples are less separated into late torpedo and green mature stages and transcriptome profiles are less separated according to the accession. In addition, all biological replicates were grouped together, showing low variability among samples (Figure 26A). Furthermore, a similar analysis

is presented in a correlation-based heatmap where biological replicates are clustering together. The biggest separation (blue scale, lowest z-score) is observed between the SAM and embryo samples, then based on the developmental stage, and lastly based on the accession (red scale, highest z-score) (Figure 26B).

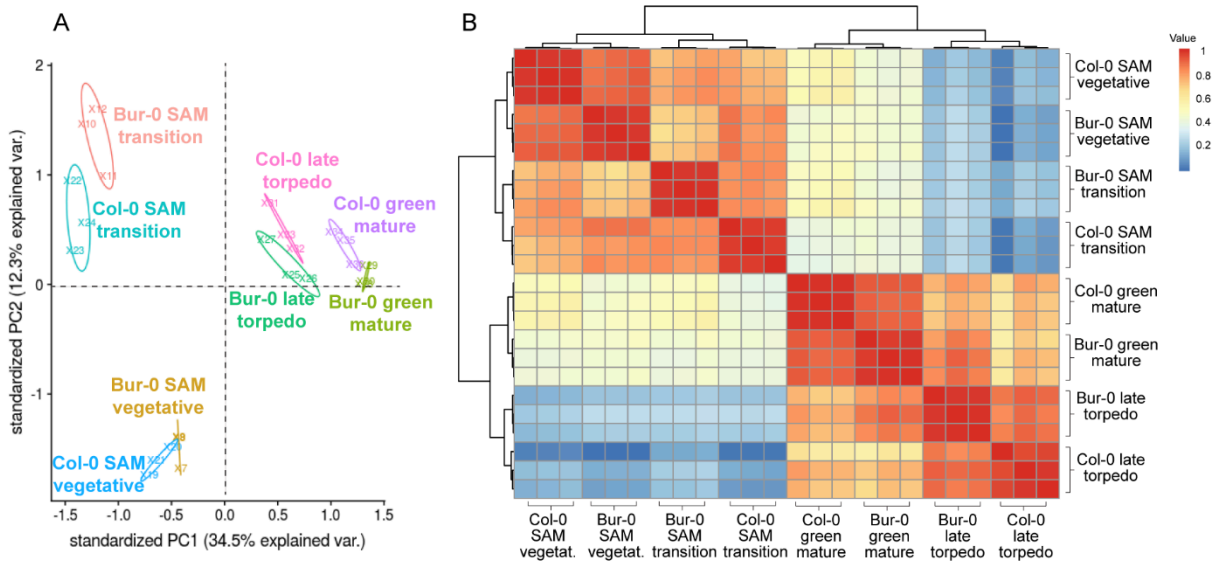


Figure 26. Embryo and shoot apical meristem (SAM) transcriptome profiles are different. (A) Principal component analysis (PCA) of embryo and meristem samples at different developmental stages. (B) Correlation-based heatmap of embryo and meristem samples. Value = z-score. The analysis was performed with three biological replicates (n = 3).

3.4.2. Differentially expressed genes (DEGs)

Genes were grouped based on the implemented cutoff (fold change > 1) into significantly changed and up- and down-regulated genes between the two accessions (Col-0 *vs* Bur-0) at each tissue and developmental stage analyzed. With this approach, at embryo late torpedo stage 2,496 genes were identified as significantly changed, from which 1,497 were up-regulated and 999 down-regulated, at embryo mature stage 2,871 genes were identified as significantly changed, from which 1,401 were up-regulated and 1,470 down-regulated, at SAM vegetative stage 2,457 genes were identified as significantly changed, from which 1,612 were up-regulated and 845 down-regulated and at SAM floral transition stage 4,864 genes were identified as significantly changed, from which 3,545 were up-regulated and 1,320 down-regulated (Figure 27A, B).

Furthermore, an overlap analysis was conducted to identify the number of unique and shared significantly changed genes found in Bur-0 and Col-0 transcriptomes. Our results revealed unique and overlapping groups of genes significantly changed, up- and down-regulated between the two accessions (Col-0 *vs* Bur-0) at each tissue and developmental stage analyzed. Interestingly, stage-specific differentially expressed genes were identified, but also shared between stages. More unique genes were identified for embryo tissue at mature stage (1,534) than at late torpedo stage (1,159) and for SAM tissue more unique genes were identified at floral transition stage (3,166) than at vegetative stage (758), however, between early and late

embryo stages the overlapping genes correspond to 25% of the significantly changed genes and between early and late SAM stages, the overlapping genes correspond to 23% of the significantly changed genes (Figure 27C, D). Additionally, DEGs were grouped into significantly changed and up- and down-regulated genes between tissue stages per accession (embryos at late torpedo *vs* mature stage and SAM at vegetative stage *vs* floral transition stage in Col-0 and Bur-0, respectively) and we identified more genes significantly changed and unique in Bur-0 than in Col-0 (Supplementary Figure S6), however, for further analysis we focus in the DEGs identified between accessions (Col-0 *vs* Bur-0) and not in the DEGs identified per accession.

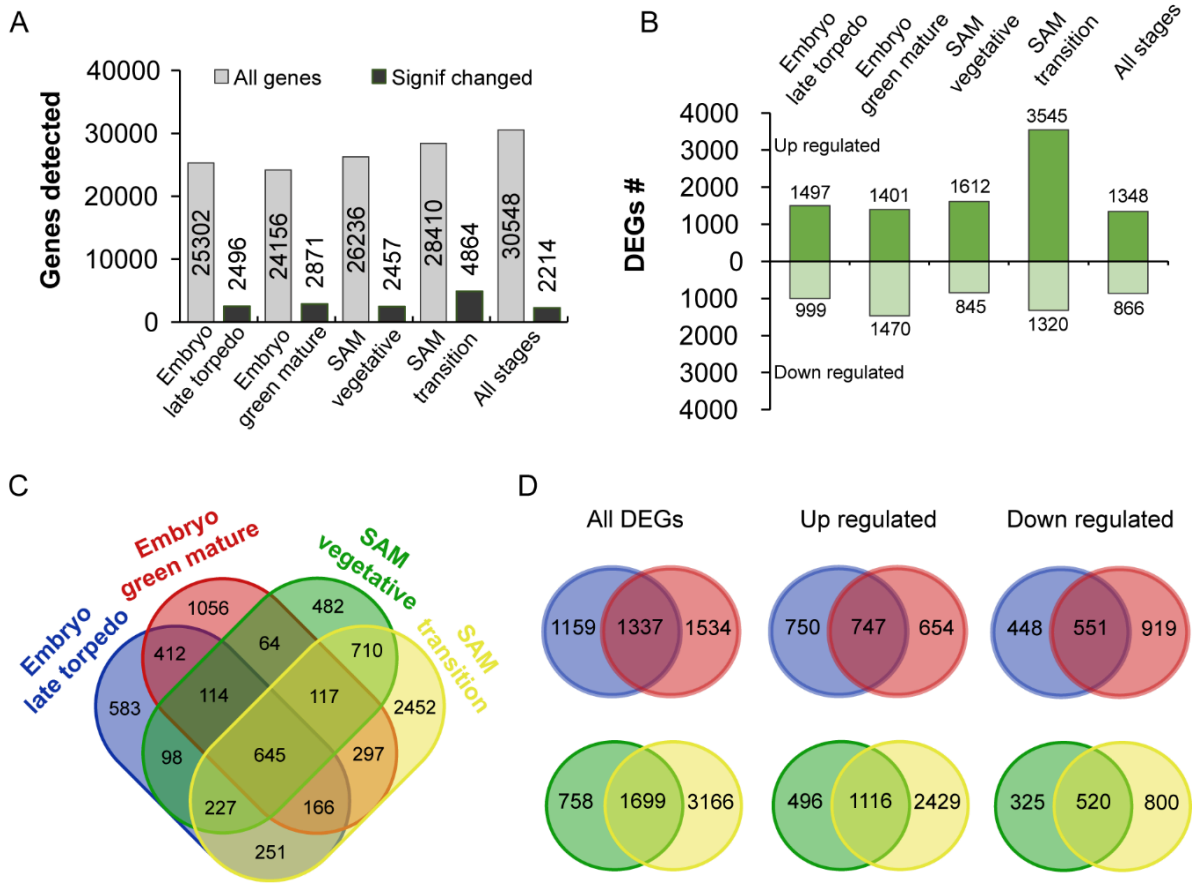


Figure 27. Differentially expressed genes Col-0 *vs.* Bur-0. (A) Total number of genes detected and differentially expressed genes (DEGs) between the Arabidopsis accessions Col-0 and Bur-0, significantly changed per tissue and stage. (B) Differentially expressed genes up and down regulated per tissue and stage. (C) Venn diagrams showing number of differentially expressed genes per tissue, stages and shared between stages. (D) Venn diagrams showing number of differentially expressed genes up and down regulated per tissue, stages and shared between stages. SAM, shoot apical meristem.

3.4.3. Gene Ontology Analysis

We analyzed gene ontology terms (GO) for biological processes using the tool http://www.pantherdb.org/, considering all differentially expressed genes found between the two accessions (Col-0 *vs*. Bur-0). We identified fifteen GO categories enriched, among which ´cellular processes´ (GO:0009987), ´metabolic process´ (GO: 0008152), ´response to stimulus´ (GO: 0050896), ´biological regulation´ (GO: 006507), ´localization´ (GO: 0051179),

´signaling´ (GO: 0023052) and ´growth´ (GO: 0040007) were overrepresented (Figure 28). In addition, we analyzed GO terms for biological process considering DEGs between the two accessions (Col-0 *vs.* Bur-0) at each tissue and developmental stage and we identified up to sixteen GO categories enriched, from which ´cellular processes´ (GO:0009987), ´metabolic process´ (GO: 0008152) and ´response to stimulus´ (GO: 0050896) were also overrepresented as for all DEGs (Supplementary Figure S7).

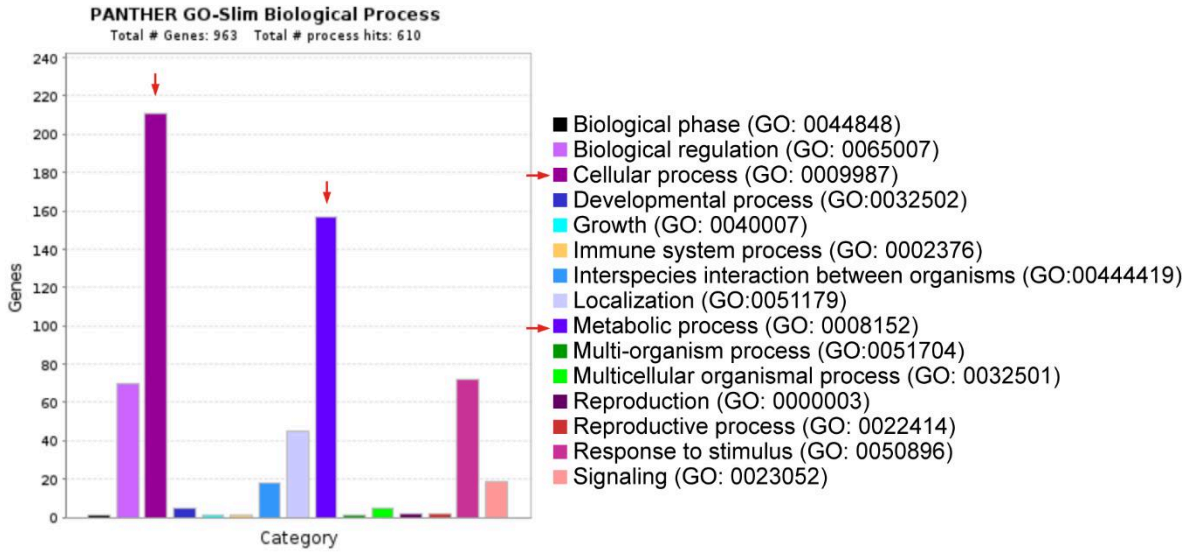


Figure 28. Gene ontology analysis. Gene ontology terms (GO) enriched for biological process considering all differentially expressed genes found between the two accessions (Col-0 *vs.* Bur-0). GO term chart was generated using the tool http://www.pantherdb.org/. Arrows indicate enriched GO categories.

3.4.4. Cluster analysis reveals accession-specific DEGs across tissues and developmental stages

In order to find accession-specific key regulators of plant growth shared between developmental phases, clustering analysis was done considering differentially expressed genes (DEGs) among tissues/stages (embryos at late torpedo and mature stage and SAMs at vegetative and floral transition stage) and the two accessions Bur-0 and Col-0. Hierarchical cluster analysis was performed and our data were grouped into 30 clusters, each of them with a set of genes with higher or lower expression per tissue and stage in Bur-0 and Col-0, respectively (Figure 29A).

Interestingly, two particular clusters contain a set of genes whose expression is high or low, consistently across tissues and developmental stages in Bur-0 (cluster 10 with 41 genes) or in Col-0 (cluster 9 with 44 genes), respectively (Figure 29B, C), revealing candidate genes with potential roles in plant growth and determination of accession-specific phenotypes.

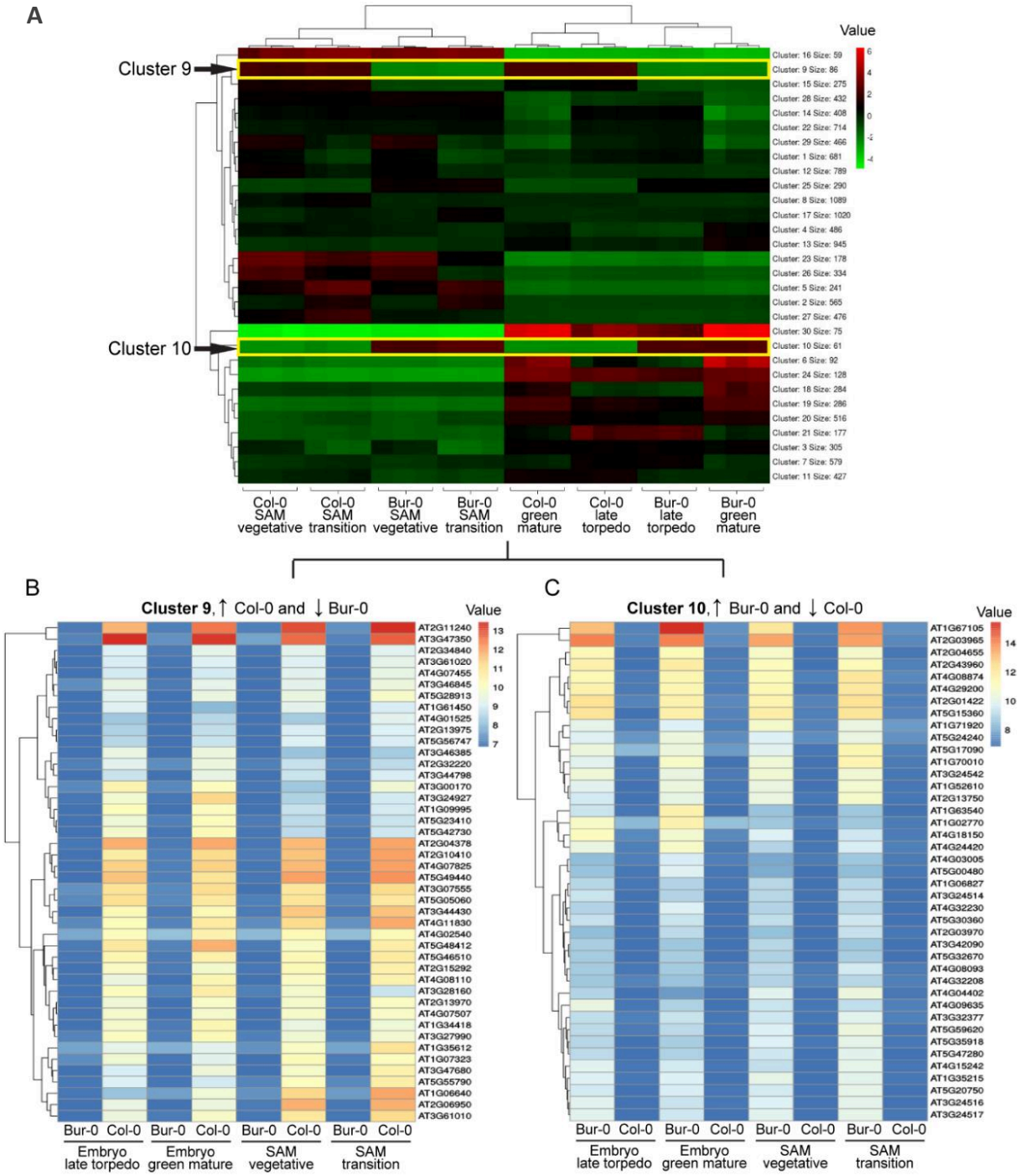


Figure 29. **Hierarchical cluster analysis of differentially expressed genes.** (A) Differentially expressed genes grouped into 30 clusters, each of them with a set of genes with higher or lower expression per tissue and stage in Bur-0 and Col-0, respectively. (B, C) High confidence set of genes from cluster 9 and cluster 10, where expression is high or low in Col-0 or Bur-0, respectively, and consistently across tissues and developmental stages. SAM, shoot apical meristem.

3.4.5. Biological functions of accession-specific genes

The high confidence set of candidate genes whose expression is consistently high or low per accession across tissues and developmental stages (cluster 10 for Bur-0 with 41 genes and cluster 9 for Col-0 with 44 genes) was used to analyze gene ontology terms (GO) for biological process, but due to the low gene number per cluster and also probably because most of them

are unknown/uncharacterized genes, less than 5% of the genes were classified using the tool http://www.pantherdb.org/. Nevertheless, a detailed characterization of the 85 candidate genes was done based on information available in TAIR (https://www.arabidopsis.org/), Araport (https://www.araport.org/), NCBI (https://www.ncbi.nlm.nih.gov/), EnsemblPlants (http://plants.ensembl.org/index.html), ePlant (https://bar.utoronto.ca/eplant/), PANTHER (http://www.pantherdb.org/), Expression Atlas (https://www.ebi.ac.uk/gxa/home), ThaleMine (https://bar.utoronto.ca/thalemine/begin.do), and eFP Browser (http://bar.utoronto.ca/efp2/). Using the information collected from the databases, the candidate genes were classified in the following gene types: protein coding, transposable elements, novel transcribed and undetermined (long noncoding RNA, noncoding RNA, miscellaneous RNA) (Figure 30).

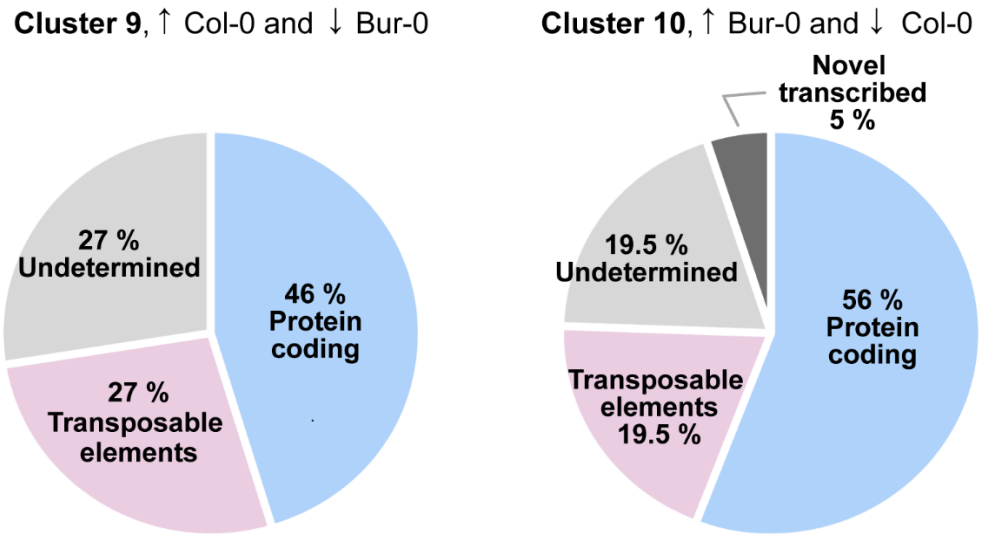


Figure 30. **Biological functions of accession-specific candidate genes.** Gene type classification of differentially expressed genes from cluster 9 and cluster 10, based on database and literature research.

The genes classified under the category protein coding correspond to: ´uncharacterized protein`, ´transmembrane protein`, ´transporter protein`, ´RNA binding / processing`, ´kinase activity`, ´histidine biosynthetic process`, ´response to abscisic acid`, ´response to salt`, ´regulation of flower development`, ´cystatin/monellin superfamily protein`, ´progression of meiosis during early prophase`, ´disease resistance`, ´beta-galactosidase related protein`, ´phosphorylation/signal transduction`, `stress response/heat shock` and `ubiquitin-like superfamily protein`, among others. A detailed list with the candidate genes classification is presented in Supplementary Table S7.

3.4.6. RNA-seq data validation through expression analysis of candidate genes

In order to validate the RNA-seq data, the candidate genes previously classified under the category ´protein coding` were further classified based on the following criteria: Higher expression values consistently across tissues and stages (as shown in heatmaps from Figure 28C, D), higher Log_2 fold change, higher raw read number per tissue/stage and availability of reported information from databases/literature. With this, five candidate genes with higher

expression in Bur-0 (*AT4G08874*, *AT1G71920*, *AT2G03965*, *AT2G43960*, *AT5G15360*) and five with higher expression in Col-0 (*AT3G44430*, *AT4G07825*, *AT5G05060*, *AT4G11830*, *AT2G04378*) were initially selected for expression analysis by qRT-PCR and RNA *in situ* hybridization.

Expression levels of the candidate gene *AT4G08874* (Bur-0 specific) analyzed by qRT-PCR were significantly higher in Bur-0 than in Col-0 ($p < 0.01$) at each tissue and developmental stage analyzed. The expression pattern of *AT4G08874* coincides with the pattern observed from the raw read counts obtained from the transcriptome data, where the transcript was barely detected in Col-0 and highly abundant in Bur-0, particularly in SAM tissue at floral transition stage (Figure 31A). Expression analysis by RNA *in situ* hybridization on longitudinal sections of embryos at torpedo and mature stage and shoot apices at vegetative and floral transition stages showed that *AT4G08874* transcript is detected in Bur-0, but not in Col-0 tissues. In addition, the expression signal in Bur-0 was weak on embryos at torpedo stage and SAM at vegetative stage, but strong on embryos at mature stage and the SAM at floral transition stage (Figure 31B).

On the other hand, expression analysis of the candidate gene *AT3G44430* (Col-0 specific) by qRT-PCR revealed expression levels significantly higher in Col-0 than in Bur-0 ($p < 0.01$) at each tissue and developmental stage analyzed. The expression pattern of *AT3G44430* coincides with the pattern observed from the raw read counts obtained from the transcriptome data, where the transcript was barely detected in Bur-0 and highly abundant in Col-0, particularly in SAM tissue at vegetative and floral transition stage (Figure 31C). Expression analysis by RNA *in situ* hybridization on longitudinal sections of embryos at torpedo and mature stage and shoot apices at vegetative and floral transition stages showed that *AT3G44430* transcript is detected in Col-0, but not in Bur-0 tissues. In addition, the expression signal in Col-0 was weak on embryos at torpedo and mature stages and stronger on the SAM at vegetative stage and floral transition stage (Figure 31D).

Expression analysis of the remaining candidate genes selected for validation was completed by qRT-PCR and raw read counts per gene from RNA-seq data were also plotted for expression pattern comparisons. Our results showed that expression levels of *AT4G07825*, *AT5G05060*, *AT4G11830*, *AT2G04378* in Bur-0 were low or not detected, while expression levels in Col-0 were significantly higher ($p < 0.01$) throughout all tissues and stages, except for *AT4G07825* that was not detected on embryo tissue for any accession (Figure 32A). Likewise, expression levels of *AT1G71920*, *AT2G03965*, *AT2G43960*, *AT5G15360* were low or not detected in Col-0, while expression levels in Bur-0 were significantly higher ($p < 0.01$) throughout all tissues and stages (Figure 32B). Similar expression patterns were observed in the raw read counts per gene from the RNA-seq data set.

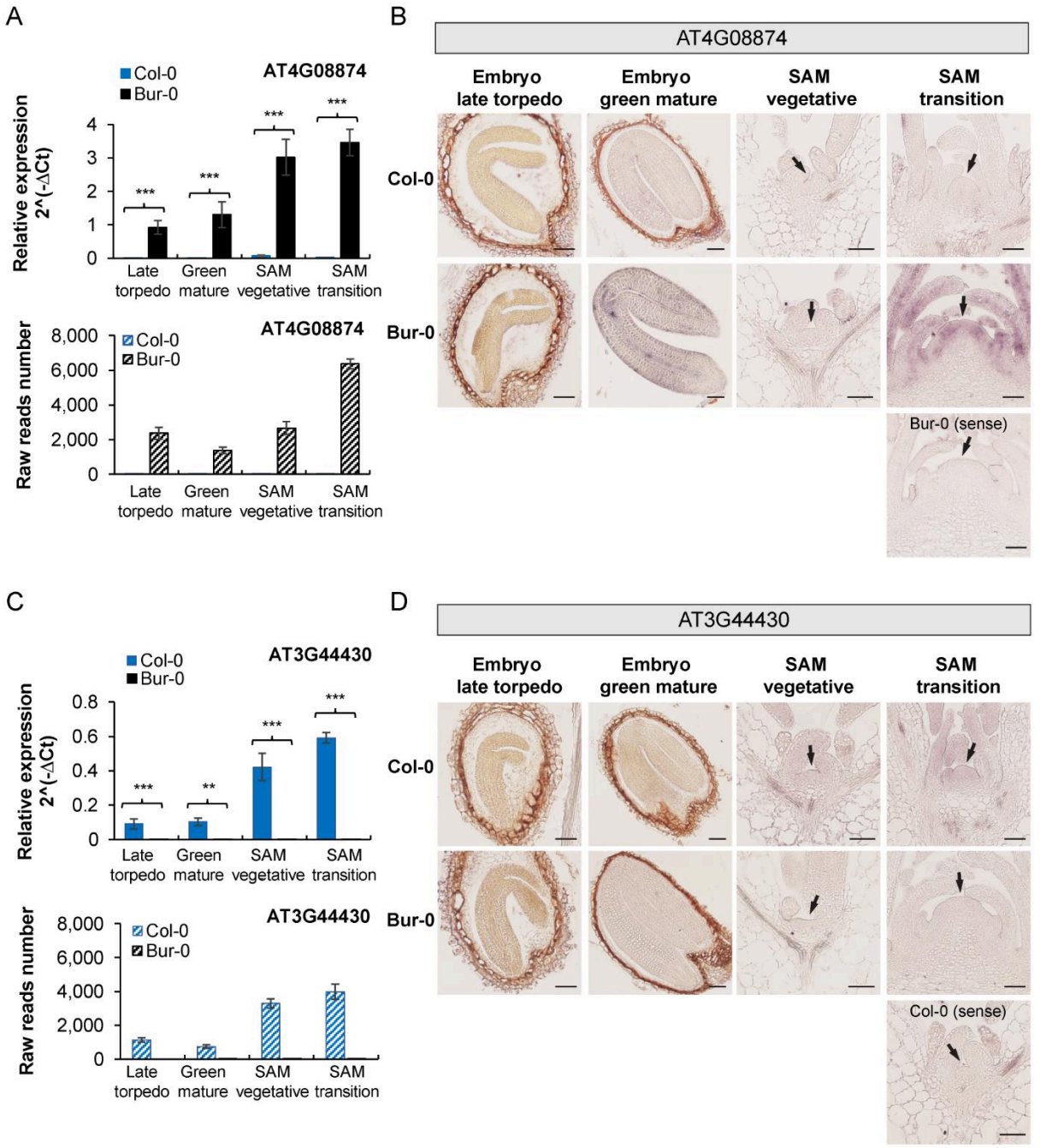


Figure 31. Validation of RNA-seq data by expression analysis of candidate genes by qRT-PCR and RNA *in situ* hybridization. Tissue samples were obtained from the Arabidopsis accessions Bur-0 and Col-0 and plants were grown in long day photoperiod. Expression analysis by qRT-PCR and comparison to the raw reads number from RNA-seq data for the candidate genes (A) *AT4G08874* and (C) *AT3G44430*. Expression analysis by RNA *in situ* hybridization on longitudinal sections of embryos at late torpedo and green mature stages and longitudinal sections of shoot apices at vegetative and floral transition stages, using specific antisense probes for (B) *AT4G08874* and (D) *AT3G44430*, respectively. Sense probes were used as controls. Scale bar = 50µm. Error bars indicate ± SD. Statistical significance tested with Student's *t*-test: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. Arrows indicate the SAM. SAM, shoot apical meristem.

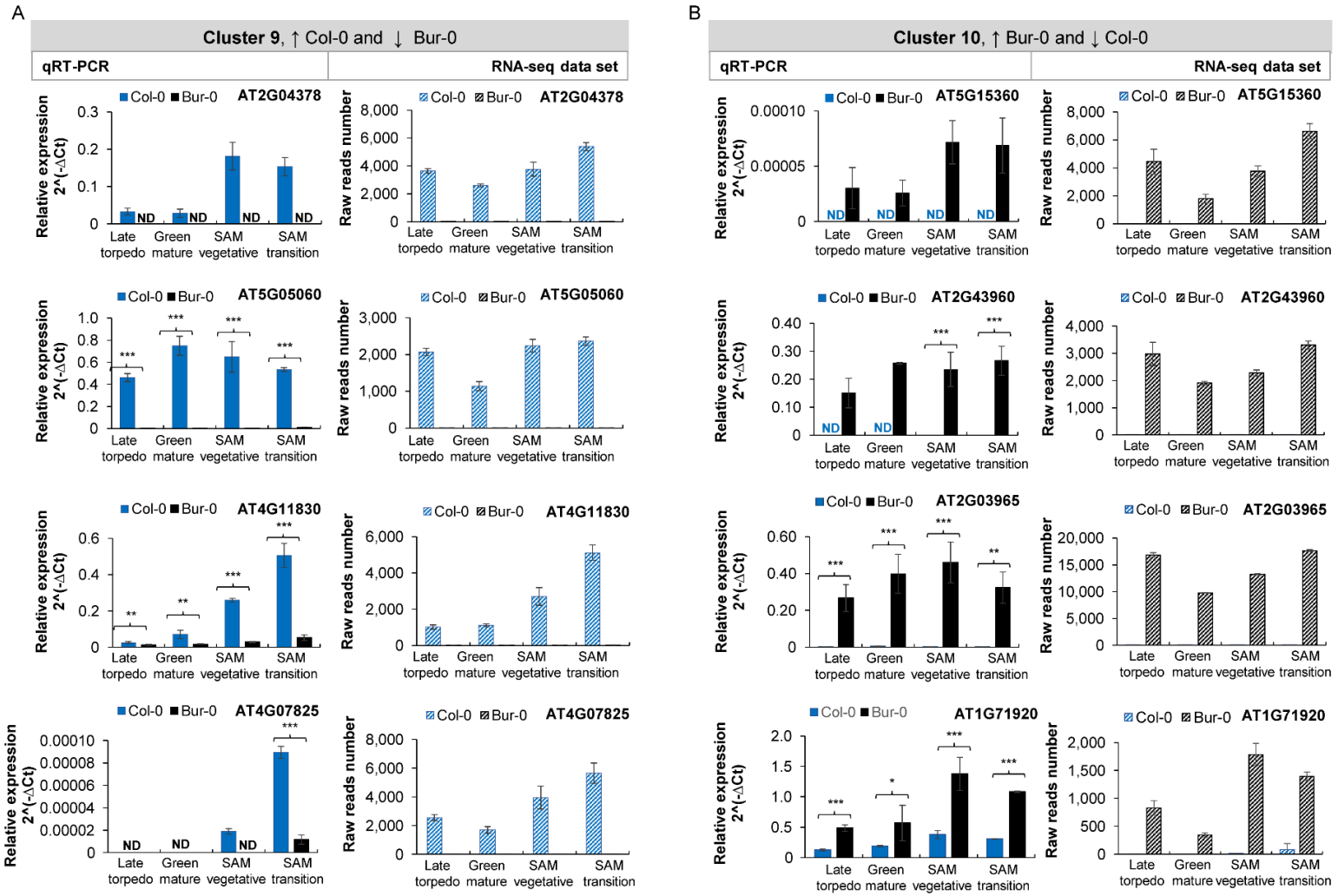


Figure 32. Validation of RNA-seq data by expression analysis of candidate genes by qRT-PCR. Expression analysis of the remaining candidate genes selected for validation and raw read counts per gene from RNA-seq data. (A) Candidate genes from cluster 9 (higher expression in Col-0 than in Bur-0). (B) Candidate genes from cluster 10 (higher expression in Bur-0 than in Col-0). Error bars indicate ± SD. Statistical significance tested with Student's *t*-test: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. ND, not detected. SAM, shoot apical meristem. Tissue from the Arabidopsis accessions Bur-0 and Col-0, grown in long day photoperiod.

Thus, our qRT-PCR and RNA *in situ* hybridization results support our transcriptome data and we could confirm that expression of the candidate genes selected for validation is accession-specific and constant across embryo and SAM tissue at different developmental stages, in agreement with the expression values and raw read counts obtained from our RNA-seq data analysis, indicating that the high confidence gene sets identified from the transcriptome datasets generated in this study are reliable and useful for elucidation of molecular mechanisms regulating plant growth and accession-specific phenotypes in Arabidopsis.

3.4.7. Expression analysis of known regulators of organ size and growth

In order to better understand the molecular basis of the big phenotype in Bur-0, we searched for marker genes reported in the literature that are involved in seed size regulation, plant growth and shoot apical meristem (SAM) maintenance, and we analyzed whether different expression patterns and levels of such known regulators might determine accession-specific phenotypes. Expression analysis of known key regulators was done by RNA *in situ* hybridization on longitudinal sections of embryos and shoot apices at different developmental stages from the Arabidopsis accessions Ws-2, L*er*-1, Col-0 and Bur-0, grown in LD photoperiod. In addition, expression levels of shoot apical meristem (SAM) maintenance genes were analyzed by qRT-PCR on embryo tissue at late torpedo and mature stages and SAM tissue at vegetative and floral transition stages from the Arabidopsis accessions Col-0 and Bur-0, grown in LD photoperiod. Although none of the selected known and previously reported key regulators of seed size, plant growth and shoot apical meristem (SAM) maintenance are found in our high confidence set of DEGs from our RNA-seq data, they are present in the data set corresponding to total genes counted, therefore the raw read counts per gene were also plotted for expression pattern comparisons and further validation of our transcriptome data.

3.4.7.1. Expression analysis of SAM maintenance genes

Our previous results indicated that the SAM might play a crucial role for establishing the big Bur-0 phenotype, a phenomenon that might be determined already during embryogenesis, therefore we analyzed expression of the SAM maintenance genes *CLAVATA3* (*CLV3*), *WUSCHEL* (*WUS*) and *SHOOT MERISTEMLESS* (*STM*), detecting their transcripts during embryonic and postembryonic development by RNA *in situ* hybridization for the Arabidopsis accessions Ws-2, L*er*-1, Col-0 and Bur-0. During embryonic development, we found that *CLV3*, *WUS* and *STM* are expressed in the meristematic zone of embryos and the transcript is detected from heart stage onwards, for all accessions.

Interestingly, we found that *WUS* expression domain is similar for all accessions, while *CLV3* and particularly *STM* expression domains are bigger in Bur-0 than in the other accessions at mature embryo stage, suggesting that the *STM* domain enlargement in Bur-0 takes place during late embryogenesis (Figure 33A, C, E). During postembryonic development, *CLV3, WUS* and *STM* transcripts are detected in the SAM of all accessions from 4 until 21 DAG and *STM* expression signal is visible throughout the SAM, *CLV3* signal is detected in the central zone and *WUS* in the Rib zone (Figure 33B, D, F).

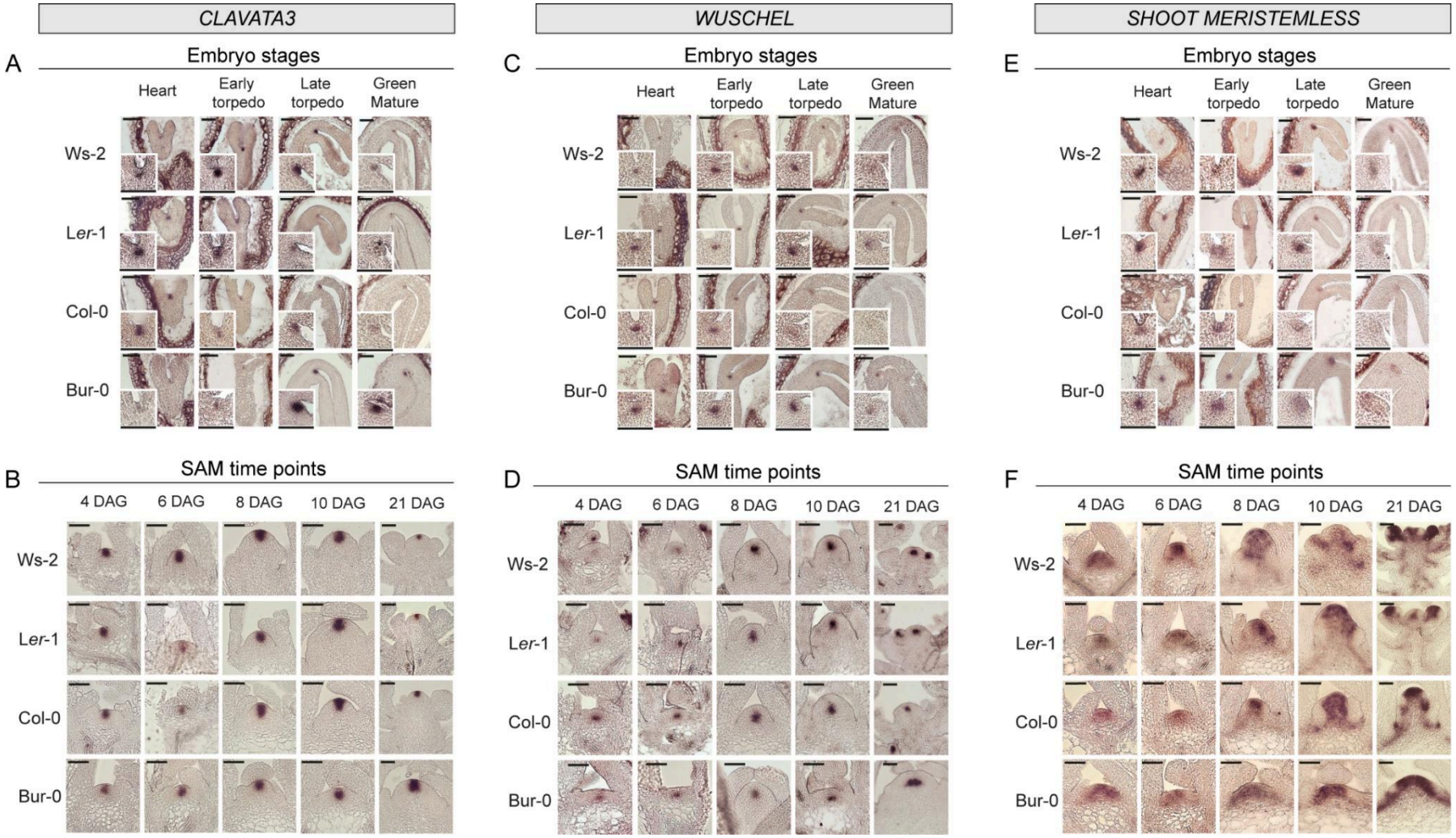


Figure 33. Expression analysis of SAM maintenance genes by RNA *in situ* hybridization. Analysis on longitudinal sections of embryos at different developmental stages and shoot apices overtime, using specific antisense probes for (A, B) *CLAVATA3*, (C, D) *WUSCHEL*, and (E, F) *SHOOT MERISTEMLESS*, respectively. Scale bar = 100µm. SAM, shoot apical meristem; DAG, days after germination. Tissue samples from the Arabidopsis accessions Ws-2, L*er*-1, Col-0 and Bur-0 plants grown in LD.

Moreover, around the time points when floral transition occurs per accession, *STM* signal is absent in the zones where new leaf primordia are initiated and *CLV3* and *WUS* expression domains are bigger. This expression pattern is observed in general for all accessions, confirming the different developmental phase progression revealed and described per accession in the previous sections. Interestingly, particularly bigger *CLV3*, *WUS* and *STM* expression domains are detected in Bur-0 towards the floral transition time point, indicating that expression domains are enlarged during late postembryonic growth. These results suggest a different SAM organization in Bur-0 during late embryonic and postembryonic development.

In order to analyze expression levels of the SAM maintenance genes *CLV3*, *WUS* and *STM* at specific developmental stages where the shift in expression domain size was detected by RNA *in situ* hybridization (late embryogenesis and late postembryonic growth), expression levels of *CLV3*, *WUS* and *STM* were analyzed by qRT-PCR for embryo tissue at late torpedo and mature stages as well as SAM tissue at vegetative and floral transition stages from the Arabidopsis accessions Bur-0 and Col-0. Both accessions have similar *CLV3* expression levels at late torpedo, mature embryo and the SAM at vegetative stage, while significantly lower *CLV3* expression levels are found in Bur-0 SAM at transition stage compared to Col-0 ($p < 0.01$) (Figure 34).

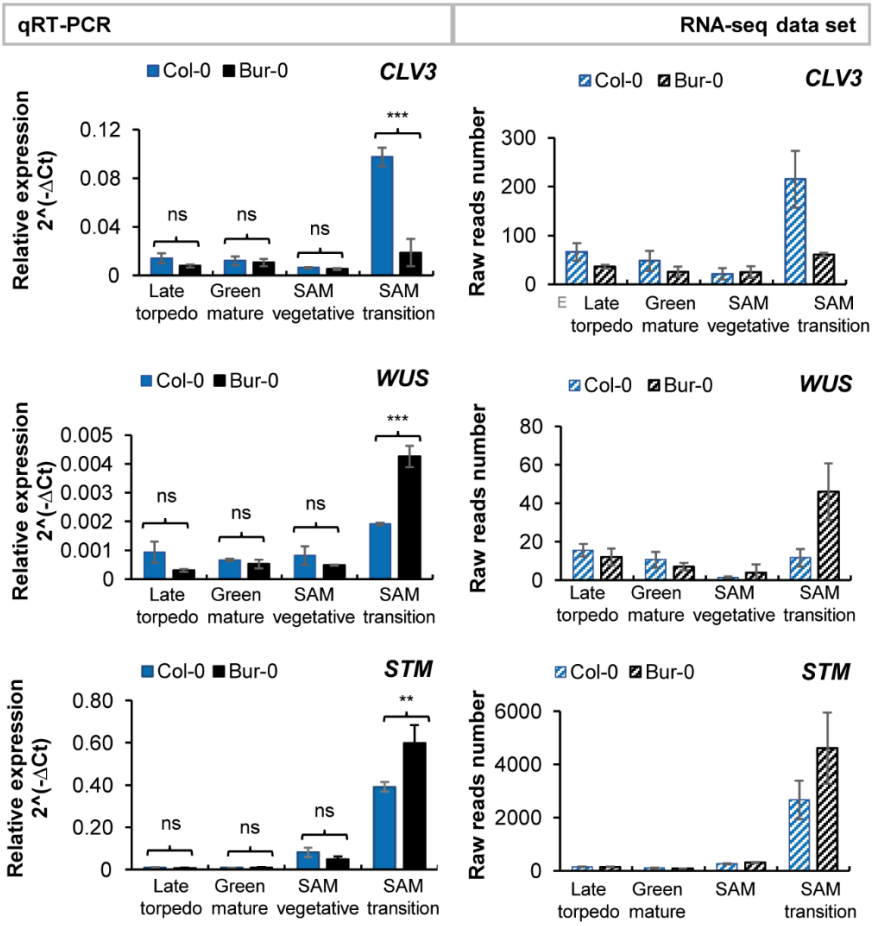


Figure 34. Expression analysis of SAM maintenance genes by qRT-PCR. Expression levels and raw read counts per gene from RNA-seq data of *CLAVATA3* (*CLV3*), *WUSCHEL* (*WUS*), and *SHOOT MERISTEMLESS* (*STM*), respectively. Error bars indicate ± SD. Statistical significance tested with Student's *t*-test: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. ns, not significant; SAM, shoot apical meristem.

On the other hand, expression levels of *WUS* and *STM* are similar for both accessions at late torpedo, mature embryo and the SAM at vegetative stage, while *WUS* and *STM* expression levels are significantly higher in Bur-0 SAM at transition stage compared to Col-0 ($p < 0.05$). A similar expression pattern is observed for *CLV3, WUS* and *STM* raw read counts from our transcriptome data (Figure 34), thus further confirming the reliability of transcriptome datasets generated in this study.

3.4.7.2. Expression analysis of known shoot growth and seed size regulators

Among the genes known from the literature to be involved in shoot growth and seed size regulation, we selected *PIGMENT DEFECTIVE 327* (*PDE327*) because to our knowledge it is the only gene reported to be responsible for a shoot specific phenotype of Bur-0. We also selected the seed size regulator *ARABIDOPSIS G PROTEIN GAMMA SUBUNIT 3* (*AGG3*). Although the function of *AGG3* has been reported and well studied, detailed tissue specific expression and localization of *AGG3* in Arabidopsis has not yet been described.

In order to determine if their expression pattern and localization might determine shoot and seed accession-specific phenotypes, we analysed expression of *PDE327* and *AGG3* by RNA *in situ* hybridization during embryogenic and postembryonic development for the Arabidopsis accessions Ws-2, L*er*-1, Col-0 and Bur-0, grown in LD photoperiod. *PDE327* expression signal is detected on embryos at different developmental stages, in the SAM at vegetative and floral transition stages, as well as in young leaves of all accessions. The strongest signal of *PDE327* is detected on embryo samples during early embryogenesis and on SAM tissue at floral transition stage from the accessions L*er*-1 and Bur-0, indicating that *PDE327* might play a role during the transition from early to late embryogenesis and towards reproductive transition in the SAM, this role might be accession-specific, but is not specific for the Arabidopsis accession Bur-0 (Figure 35A, B).

On the other hand, *AGG3* expression signal on embryo tissue at different developmental stages is weak for all the accessions and stages analysed and it is not detected in the vegetative SAM of any accession. Interestingly, we found that *AGG3* is present in the SAM at floral transition stage, particularly in the region where new floral primordia are formed for all accessions. Induction of *AGG3* during floral transition in the SAM indicates that *AGG3* might be involved in the regulation of the flowering time in *Arabidopsis thaliana* plants, however, this role is not specific for the accession Bur-0 (Figure 35B, C). Although *PDE327* and *AGG3* transcripts are detected on embryo and SAM tissue, their expression patterns and localization are not restricted to the tissue stages where major phenotypical differences have been described in our study for Bur-0 and also do not appear to be involved in determination of accession-specific phenotypes, therefore we did not continue any further analysis with *PDE327* and *AGG3.*

Additional known and previously reported seed size regulators were also selected for expression analysis by RNA *in situ* hybridization, including *APETALA2* (*AP2*), *HAIKU1* (*IKU1*) and *LARGE IN CHINESE 1* (*DA1*). Although *AP2*, *IKU1* and *DA1* transcripts are detected in different organs of the Arabidopsis accession Col-0, only for *AP2* is a positive signal is detected in the SAM at the floral transition stage. Since the expression patterns and localization of the

other known regulators is not restricted to the tissue stages where major phenotypical differences have been described in our study, but rather absent in embryo or SAM at floral transition stage (Supplementary Figure S8), we did not continue further analysis of those genes either.

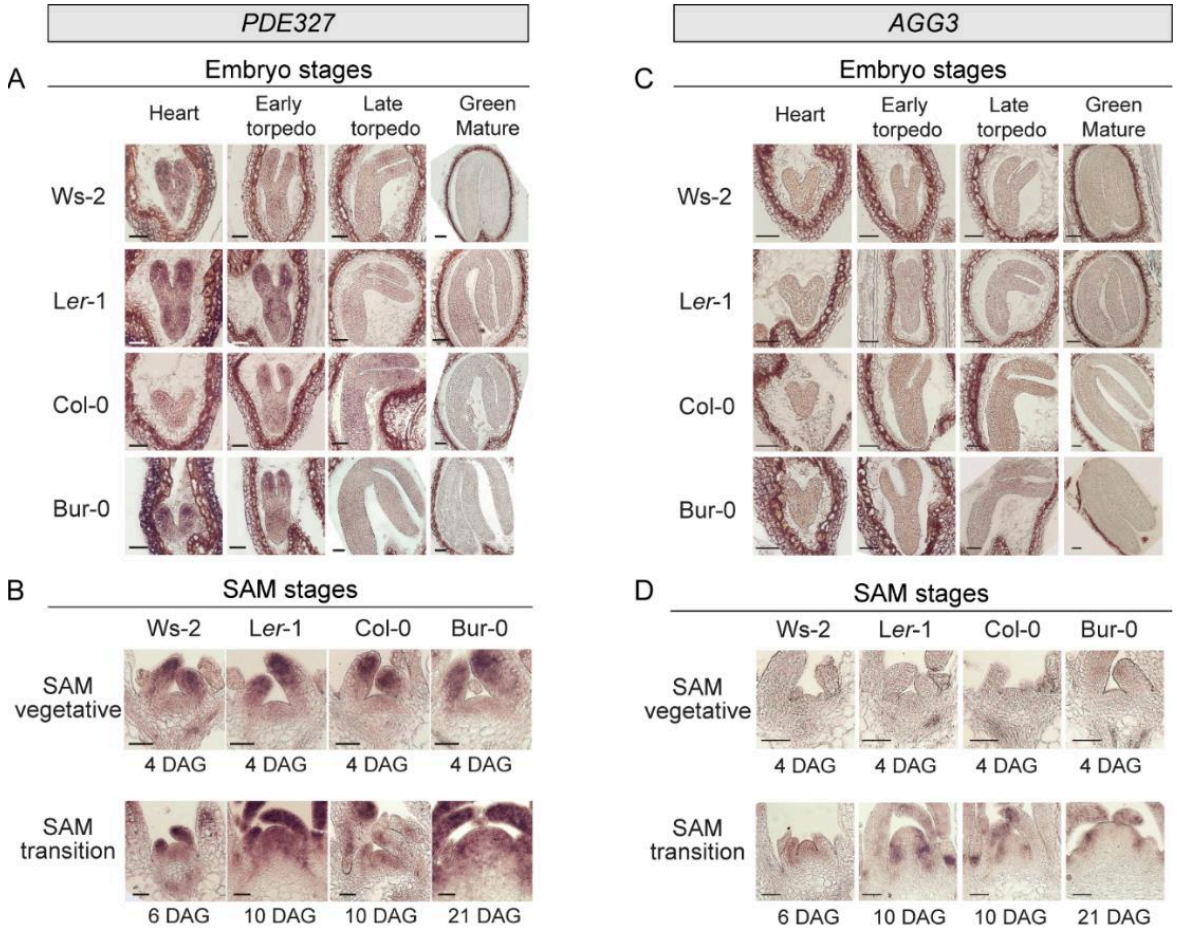


Figure 35. Expression analysis of known regulators of shoot and seed growth. Analysis by RNA *in situ* hybridization on longitudinal sections of embryos and shoot apices at different developmental stages, using specific antisense probes for (A) *PIGMENT DEFECTIVE 327* (*PDE327*) and (B) *ARABIDOPSIS G PROTEIN GAMMA SUBUNIT 3* (*AGG3*), respectively. Scale bar = 100µm. SAM, shoot apical meristem. DAG, days after germination. Tissue samples from the Arabidopsis accessions Ws-2, L*er*-1, Col-0 and Bur-0 plants grown in LD.

In summary, our RNA-seq data revealed different transcriptomes profiles according to the tissue (SAM and embryo), developmental stage (late torpedo and green mature stages for embryos and vegetative and floral transition stages for the SAM) and accession. Furthermore, our analysis identified a high confidence set of candidate genes whose expression is high or low, consistently across tissues and developmental stages in Bur-0 (41 genes) or in Col-0 (44 genes), respectively, revealing candidate genes with potential roles in plant growth and determination of accession-specific phenotypes. Finally, our expression analysis demonstrate that expression of the candidate genes selected for validation is accession-specific and constant across embryo and SAM tissue at different developmental stages, in agreement with the expression values and raw read counts obtained from our RNA-seq data analysis. In addition, expression patterns of known SAM regulators were in agreement with the raw read counts found in our data set, providing further validation of our transcriptome data.

4. DISCUSSION

The general aim of this PhD project was to investigate and understand the molecular mechanisms underlying the big phenotype in the Arabidopsis accession Bur-0. For this purpose, different approaches were considered. First, we wanted to determine whether rosette size, seed size, flowering time and SAM size are generally correlated traits in different *A. thaliana* natural accessions including Bur-0 and mutant lines and identify possible marker traits for accession-specific phenotypes by a detailed phenotypical characterization during embryonic and postembryonic development as well as in different photoperiods. Second, we investigated the extent to which the physiological status might contribute to the big phenotype observed in Bur-0 by a detailed physiological characterization in different tissues/stages during embryonic and postembryonic development, comparing Bur-0 to other accessions. Third, we wanted to investigate whether the enlarged organs observed in Bur-0 are determined by differences in cell size/number, ploidy level and/or expression of cell cycle regulators in different tissues/cell types during embryonic and postembryonic development, comparing Bur-0 to Col-0. Finally, we wanted to identify accession-specific genetic determinants of plant phenotypes, shared across tissues and developmental stages during embryonic and postembryonic growth by RNA-seq analysis of embryos and SAMs at different stages from Bur-0 and Col-0. After identification of potential candidate genes, further validation of transcriptome data by expression analyses of candidate genes as well as known key regulators of organ size and growth were conducted for embryo and SAM tissue at different stages during embryonic and postembryonic development.

4.1. Flowering time, seed size and rosette size are not general correlated traits

In order to determine if the flowering time phenotype correlates with the rosette and seed area phenotype in Arabidopsis, we analyzed those plant traits in eleven accessions from different geographical origins, described as early, intermediate, and late flowering time accessions, grown in LD photoperiod. All traits were significantly different among the analyzed accessions and no trend was observed, i.e., accessions with bigger rosettes during early growth did not have later flowering time, nor bigger seeds and not all accessions found with small rosettes during early growth had early flowering phenotype, neither smaller seeds (Figure 5), indicating high variation among the analyzed traits and accessions.

Interspecific seed size variation has been associated with different habitat characteristics (Westoby *et al*., 1992) and intraspecific variation has been correlated with different fitness components of seedling and adult plants (Krannitz *et al*., 1991). Moreover, Gnan *et al.* (2014) reported that flowering time correlates with seed number per fruit (with late flowering plants producing fewer seeds per fruit than early flowering lines), but not with seed weight. They also suggested that the seed size/number trade-off is enhanced by the limited resources caused by earlier reproduction and thus, flowering time can also alter the seed size/number trade-off in Arabidopsis. In terms of life-history theory, this result makes intuitive sense as early flowering plants should have smaller rosettes and thus reduced resources to invest into reproduction, so, it is likely that the observed modest trade-offs are a consequence of restricted resources and not genetic pleiotropy (Gnan *et al.*, 2014).

Although the seed number produced per fruit was not analyzed in our study, our correlation analysis between the plant traits rosette area at 14 DAG, flowering time and seed area revealed that those traits are not correlated, not when all accessions are compared, neither when the phenotypical data sets are grouped according to the flowering phenotype (early, intermediate, and late). Thus, our data demonstrate that rosette area during early growth, seed area and flowering time are not generally correlated traits in Arabidopsis and suggest that those are independent traits influenced by other factors and none of them is an optimal marker trait (predictor) for the adult plant phenotype. Interestingly, our analysis revealed how the accession Bur-0 stands out from the data distribution as outlier, particularly for the traits seed area and late flowering time, in agreement with the big seed phenotype reported by Herridge *et al*. (2011) and the late flowering time phenotype reported by Werner *et al*. (2005).

Although the average rosette area of Bur-0 was not the largest compared to other accessions during early growth (from two to 14 DAG) (Figure 5A, B), we observed a different rosette size pattern per accession during late growth, at bolting time in LD photoperiod. We were unable to successfully measure the rosette area at bolting time due to overlapping leaves, which prevented an accurate estimation of rosette surface, however, we observed that some accessions having small rosettes during early growth were still smaller during late growth, but others were among the largest accessions during late growth, including Bur-0 (Figure 5D).

Further growth analysis in different photoperiods and during late postembryonic growth confirmed that the rosette phenotype in Bur-0 is larger than in other accessions, but the larger phenotype is achieved during late postembryonic growth, regardless of the photoperiod. Thus, the large rosette phenotype in Bur-0 reported by Camargo *et al*. (2014) was confirmed in our study only during late postembryonic growth. Our results also revealed that the bigger plant 3D surface area observed in Bur-0 does not result from a higher relative expansion rate (RER), but by determining the leaf initiation rate (LIR) per accession we found that Bur-0 produces leaves faster than the other accessions (Figure 11C, D).

The rosette phenotype is a dynamic trait influenced by several factors. Studies on natural variation in rosette size in Arabidopsis have reported that Genome-wide association (GWA) mapping of the rosette temporal growth data resulted in the detection of time-specific quantitative trait loci (QTLs), whereas mapping of model parameters resulted in another set of QTLs related to the whole growth curve, thus the existence of general growth factors with a function in multiple developmental stages has been suggested (Bac-Molenaar *et al*., 2015). Additionally, González *et al*, (2020) reported a Genome-wide association (GWA) mapping analysis on natural variation in Arabidopsis effective leaf surface area (ELSA) (parameter for rosette area) where 710 accessions grown in LD photoperiod were studied and the rosette area of flowered or unflowered plants was measured 45 days after sowing. The authors concluded that rosette's ELSA values do not depend on whether plants have flowered or not but vary along a latitudinal cline: the more northern the origin of a natural accession, the smaller ELSA it has.

4.2. SAM size correlates with adult plant traits

Considering that the accessions analyzed in this study had different flowering phenotypes, it was likely that during postembryonic growth developmental stages were reached at different time points, therefore we were interested in a marker trait that allowed differentiation of vegetative and reproductive stages during postembryonic growth and thus, identification of precise developmental stages. For this purpose, a detailed morphological analysis of the SAM (shoot apical meristem) was done to reveal the precise time point of floral transition per accession.

Floral transition is characterized by an enlargement and mounding-up of the SAM. Mitotic activation, increased SAM size and doming precede the elongation of apical internodes and all these changes occur when floral meristems are first initiated by the SAM (Jacqmard *et al*., 2003). In this study, the SAM size and shape changes over time were morphologically confirmed in the eleven analyzed accessions and the floral transition stage was identified per accession at earlier, intermediate and later time points.

In addition, SAM width was measured as a parameter for SAM size estimation and our results showed that the maximal widening of the SAM takes place towards the same time point when the reproductive transition is morphologically visible; afterwards the SAM slightly shrinks (Supplementary Figure S3). The size changes of the SAM over time (maximum widening peak followed by SAM shrinkage) confirmed that the floral transition stage is initiated at earlier, intermediate, and later time points among the eleven accessions analyzed. Interestingly, these results also revealed a rapid increase in SAM size for early flowering accessions, while in late flowering accessions the SAM size increases slowly and steadily overtime (Supplementary Figure S3).

Our SAM morphological characterization demonstrates that several accessions were at different developmental stages by the time they had the same chronological age (Figure 6), therefore comparisons of phenotypical traits between accessions based on the chronological age of the plants might lead to misinterpretation of results and developmental stage-based comparisons should be considered as well. Our correlation analysis between time-dependent and developmental stage-dependent values for SAM size and adult plant traits revealed that SAM size at floral transition stage correlates with flowering time and rosette area, but not with seed size. No further correlations were confirmed at vegetative stage or for time-dependent comparisons (Figures 7 and 8). Thus, our results indicate that the stage-dependent rosette size and flowering time are good predictors for SAM size phenotypes in Arabidopsis.

In order to better understand the causes for the particularly large phenotype in Bur-0, besides the eleven Arabidopsis wild accessions analyzed, a detailed phenotype characterization was done for the late flowering mutant lines *tsf-1*, *ft-10*, *soc1-6*, *tsf-7* and *fd-3*, which are in Col-0 background. Rosette area during early growth, flowering time, seed size and SAM size traits were analyzed. We found that in comparison to the wild type, the increased rosette size during late postembryonic growth and wider SAMs observed in the late flowering mutant lines resemble the rosette and SAM phenotype of late flowering natural accessions, only *fd-3* has

significantly bigger rosettes and *soc1-6* significantly bigger seeds than Col-0, but no correlations were confirmed among traits and none of the late flowering mutant lines have simultaneously larger rosettes, larger SAM, later flowering phenotype and larger seeds than the wild type Col-0.

Our data demonstrate that among the analyzed natural accessions and mutant lines, no other stands out with larger rosettes, larger SAM, later flowering phenotype and larger seeds as Bur-0. Interestingly, further analysis during postembryonic development revealed that from germination onwards, the length of developmental phases during postembryonic growth is extended in Bur-0, intermediate in Col-0 and L*er*-1 and shorter in Ws-2 (Figure 12). The strong differences in the length of developmental phases between Bur-0 and other natural accessions and the lack of a late flowering natural accession or mutant line with a similar phenotype like Bur-0 made us decide on Col-0 as the reference accession for further comparisons with Bur-0.

4.3. The big Bur-0 phenotype is already determined during embryogenesis

The embryo area analysis revealed that the Arabidopsis accessions Ws-2, L*er*-1, Col-0 and Bur-0 have similar embryo size at heart and torpedo stages (Figure 13B, C), while at late torpedo and mature stages the embryo area is significantly different among all accessions ($p< 0.05$) and Ws-2 has the smallest embryos, Col-0 and L*er*-1 have intermediate embryo area, while Bur-0 has the largest embryos (Figure 13D, E). Thus, our results reveal that embryo size in Bur-0 is considerably enlarged only during late embryogenesis.

Furthermore, embryo development based on days after pollination was analyzed for Bur-0 and Col-0 and we could confirm that development progression is synchronized for both accessions. However, in comparison with the results reported by Le *et al*. (2010) for Ws-0 embryo development, we identified the embryo stages globular, hearth, and torpedo 1-2 days later than the reported time points for Ws-0 and mature green embryos 2-3 earlier. Whether this contrasting results might be attributed to a different developmental phase progression during embryogenesis in different Arabidopsis accessions cannot be determined from our current analysis, but from our observations we can suggest that eight days after pollination (DAP) is an optimal time point to harvest embryos at late torpedo stage and 10-12 DAP to harvest green mature embryos for the Arabidopsis accessions Col-0 and Bur-0.

4.4. Seed weight, seed yield and parental effects on seed size

We analyzed seed weight, total seed weight produced per plant and total seed number produced per plant as seed yield parameters. We found that Bur-0 produces seeds with the highest weight, but a lower total seed yield per plant, while Ws-2 produces seeds with the lowest weight, but a higher seed yield per plant and linear regression analysis revealed a strong and significant correlation between seed weight and total seed yield among the analyzed accessions (Figure 15).

Depending on the parameter selected to evaluate seed yield, our results can be interpreted in two ways. One the one hand, if total seeds produced per plant is selected as target trait for seed

yield, our results indicate that an early flowering accession with small rosettes like Ws-2 has higher seed yield than a late flowering accession with big rosettes like Bur-0. On the other hand, if the seed size or weight is selected as target trait for seed yield, our results indicate that Bur-0 has a higher seed yield than the other accessions. Although the possible reasons why Bur-0 produces such reduced number of seeds per plant cannot be elucidated from our results, technical artefacts because of bagging the plants and possible impairment of the optimal development of the siliques do not apply because in this study siliques were harvested once they had turned completely brown but before they had dropped seeds. Additional analysis of the number of siliques produced per plant or the number of seeds per silique could better explain why Bur-0 produces large seeds, but few seeds per plant.

Given that the seed size and weight are highly influenced by cell size *via* the growth and expansion caused by massive accumulation of storage compounds (proteins, lipids, and/or carbohydrates) and water intake by cotyledon or endosperm cells (Dante *et al*., 2014), our results suggest that the large seed size in Bur-0 can be attributed in large part to the large size of its embryo.

Gnan *et al*. (2014) suggested that the bigger seed size of Bur-0 was due to directional selection and at least some of the variation in seed size within Arabidopsis is due to adaptive processes. Thus, it is possible that the reduction in the trade-off represents a change in allocation pattern due to developmental processes and is not simply a function of more resources due to a later transition to reproduction. In addition, the combination of genetic independence of seed size from seed number suggests that seed size might be a better target for yield and fitness improvement than seed number (Gnan *et al.*, 2014).

Seed size is expected to have a strong effect on fitness and to be under strong stabilizing selection, nevertheless, seed size variation can be also attributed to environmental effects, positional or developmental effects that alter how much of the total resources available for the mother's reproduction is provided to each fruit and seed (resource allocation) (de Jong & Scott, 2007). Gnan *et al.* (2014) conducted a QTL analysis for seed size in *A. thaliana* and for six of the eight seed size QTL identified, the allele conferring the largest seed size was from the Bur-0 accession. At other two QTL, the Bur-0 allele leaded to the second largest seed size.

With respect to seed size, parent-specific gene expression (PSGE) can only evolve when the developing offspring has a strong effect on its own resource acquisition. When seed size is close to the optimum for the maternal parent, there is no internal conflict in the offspring because maternally and paternally derived genes both favor increased seed size. Although the literature generally suggests that the maternal parent controls seed size, a number of observations suggest an additional role for the paternal parent (de Jong & Scott, 2007).

In our study we crossed Bur-0 with L*er*-1, Col-0, Ws-2, respectively, and our results showed that the F1 in all cases, using Bur-0 as parental genotype, either female or male genotype, resulted in bigger seeds (Figure 16). Our results indicate parental effects on seed size and the significant size increase obtained when Bur-0 was used as the female genotype suggests that such effect might be rather maternally controlled.

Scott *et al.* (1998) reported evidence for the parental conflict theory, because they found that Arabidopsis seeds with double the normal dose of paternal genomes produce large endosperms and embryos, while those containing a double dose of maternal genomes have the opposite phenotype, suggesting that in *A. thaliana* larger seeds are produced when the paternal genome is in excess, while an excess of maternal genotype causes reduction or abortion of seeds.

Parental effects on seed size can exist in *A. thaliana*, but natural variation among paternal genotype in seed size has not yet been demonstrated (Scott *et al*., 1998). In addition, molecular characterization of such effect remains scarce and therefore a better understanding of the genetic control of seed size would help evaluate the different evolutionary theories for the maintenance of heritable variation in seed size. From an applied perspective, identification of the contribution of maternal, paternal and developmental factors to seed size and number could be very useful in developing strategies to improve grain yield (Scott *et al.*, 1998).

4.5. Bur-0 has higher carbon resources during embryonic and postembryonic growth

In this work we identified accession-specific physiological and metabolic traits that underlie accession-specific embryo, seed and rosette phenotypes. In particular, more carbon resources during embryonic and postembryonic development were found in Bur-0, suggesting an important role of carbon resources in determination of the bigger Bur-0 phenotype. Since the results and discussion about physiological and metabolic profiling of embryos and dormant seeds have been already published in the paper presented in *Section 3.2.1.1*, only results about postembryonic development will be discussed in this section.

We found that during postembryonic growth, Bur-0 produces more shoot biomass than other accessions at the lowest relative growth rate (RGR). Interestingly, metabolite levels analyzed over time revealed different fluctuation patterns in Col-0 and Bur-0, in general higher levels of carbohydrates in Bur-0 and similar protein levels in both accessions. On the other hand, metabolite levels analyzed at vegetative and reproductive stages revealed that both accessions have a different rosette metabolic status at different stages of development and in particular, the higher levels of glucose and fructose towards reproductive growth in Bur-0 suggest a role of those metabolites in floral transition related processes. Additionally, the higher starch content in Bur-0 rosettes during vegetative and reproductive growth indicate that Bur-0 plants accumulate more carbon resources throughout postembryonic development than Col-0. Thus, accession-specific metabolic traits that underlie the accession-specific phenotypes during embryonic and postembryonic development were identified in this study and our results suggest that the physiological status can determine phenotypical differences among Arabidopsis natural accessions.

In agreement with our results, protein levels have been previously shown to be rather constant in a large set of Arabidopsis accessions (Gibon *et al*., 2009). Moreover, relations between growth rates and the levels of carbon and other central metabolites have been reported in Arabidopsis. For example, Cross *et al*. (2006) reported that the rate of growth is inversely related to the levels of sugars as well as starch levels, and Sulpice *et al*. (2009) also reported that levels of carbohydrates and especially of starch are negatively correlated with growth.

Sulpice *et al.* (2009) reported that many metabolites are negatively correlated to rosette biomass, including starch, protein and several low-molecular-weight metabolites, including sucrose. They also reported that many metabolic traits correlate with each other and therefore some of the correlations with biomass may be secondary. They found some links between metabolites: starch was linked to sucrose, glucose was linked to fructose but not to sucrose or starch, several amino acids were linked, and they reported that starch and, to a lesser extent, total protein integrate metabolic status and also indicates that the regulatory network that determines starch and protein levels contributes to the regulation of biomass. Moreover, they pointed a biological explanation for the negative relation between biomass and metabolites, namely, that large accessions have a modified balance between the carbon supply and growth, which is integrated as a change in starch levels (Sulpice *et al*., 2009).

Nevertheless, Cross *et al*. (2006) also reported that across 24 Arabidopsis accessions, the rosette weight was a high variable parameter that in general was negatively correlated to starch, unrelated to sugars, amino acids, and organic acids, and was positively correlated to several enzyme activities in central metabolism, but some interesting exceptions were identified. A small number of accessions (Bur, Lip-0, Rsch-0, and Bch-1) had high carbohydrate and low amino acid levels, indicating that balance of C and N metabolism might be shifted in these accessions. They also found that for some accessions (Bur, Lip-0, Rsch-0, and Ze-0) the rosette size correlates positively with the levels of carbohydrates or amino acids and this results are in agreement with the results obtained in our study for Bur-0.

Cross *et al*. (2006) indicated that faster-growing accessions often have lower levels of carbohydrates, including lower starch and sugars at the end of the night, than slower-growing accessions, thus they hypothesized that faster-growing accessions are less "conservative" than slower-growing ones and hold less carbohydrate in reserve as starch to cope with unexpected fluctuations of the conditions. This hypothesis could be confirmed with our results, at least from the rosette metabolic status in a faster growing accession like Col-0 and a slower-growing accession like Bur-0. Cross *et al*. (2006) concluded that a larger rosette is frequently accompanied by higher activities of enzymes in central C and N metabolism, and unaltered or slightly decreased levels of central C and N metabolites, indicating that increased growth is driven by increased fluxes due to higher catalytic capacity, rather than increased levels of metabolites. Since we did not measure enzyme activities, we cannot confirm this observation.

On the other hand, different mechanisms regulating the link between carbon balance and growth processes have been reported (reviewed in Smith & Stitt, 2007). Interestingly, pleiotropic metabolic effects of sugar concentration on floral transition, depending on the plant's vegetative growth phase and genetic background have been reported in several late flowering mutant lines (Ohto *et al*., 2001) and studies during carbon starvation and re-supply suggest that growth may be affected at several levels by carbon availability (Price *et al*., 2004). Whether the increased levels of carbohydrates in Bur-0 rosettes also have implications in regulation of developmental phase transitions remains to be elucidated. Analyses of carbon-regulated transcripts combined with measurements of enzyme activities could provide a better understanding of the role that particular metabolites might play in determination of the bigger Bur-0 phenotype during postembryonic development.

4.6. Bur-0 has bigger organs, bigger cells, but similar ploidy level as Col-0

Considering that the size of an organ is determined by the size and number of its constituent cells, we investigated whether differences in cell size, ploidy level, as well as expression of cell cycle regulators might determine accession and tissue-specific phenotypes. We found that Bur-0 pollen grains, leaves, mature embryos as well as vegetative and reproductive SAMs have bigger cells than the Col-0 equivalents, however, the nuclei content and ploidy level in somatic and meristematic cells is similar in both accessions, suggesting that the larger organ size in Bur-0 can result from its enlarged cells, but not from a different ploidy level. The similar ploidy level found in Bur-0 and Col-0 is in agreement with the ploidy level reported for Arabidopsis accessions available in stock centers, which are mostly diploid and the only two natural tetraploid accessions of *A. thaliana* correspond to Stoc and Wa-1 (Schmuths *et al*., 2004).

Additionally, DAPI-stained samples revealed similar nuclei content in both accessions, but the nuclei organization was different and they appeared more separated from each other in Bur-0 than in Col-0 tissues, confirming differences between accessions in terms of cell size. DAPI-stained samples were also used to determine the cell number in vegetative and reproductive SAMs and we found that Bur-0 SAMs contain more cells than in Col-0, suggesting that the large SAM phenotype in Bur-0 can result from more cells as well. These results are in agreement with our expression analysis of the cell cycle markers *CYCB1;1* and *HIS4*.

Although expression of *HIS4* is similar for both accessions in all tissues and stages analyzed, higher expression of *CYCB1;1* is detected in Bur-0 SAM than in Col-0, particularly during vegetative growth, indicating higher cell proliferation. Interestingly, *HIS4* positive cells are more abundant in the SAM at floral transition stage than at vegetative stage in Col-0 and Bur-0, suggesting that cell division increases actively towards the reproductive transition (Figure 24). This results are in agreement with previous publications where increased cell proliferation in the SAM during the floral transition has been reported (Jacqmard *et al.*, 2003).

Since cell number was not determined on embryo tissue and expression analysis of the cell cycle markers *CYCB1;1* and *HIS4* on embryos revealed that cell proliferation is decreased during late embryogenesis and arrested on mature embryos, we cannot determine from the current analysis if a higher cell proliferation also contributes to the large mature embryo phenotype in Bur-0. These results agree with previous publications where it has been described that the final phase of embryo growth is exclusively characterized by events of cellular expansion and subsequent cell differentiation without cell divisions (Alberts *et al*., 2002; Locascio *et al*., 2014).

Larger organs tend to contain more cells than smaller organs, implying that cell division plays a fundamental role in organ size determination during organogenesis (Hu *et al*., 2003). For example, the phenotype found upon overexpression of the plant D-type cyclin *CycD2* in transgenic tobacco resulted in plant growth promotion (Cockcroft *et al*., 2000), but there are cases in which the alteration of cell proliferation is not correlated with changes in organ size and strategies that simply increase expression of cell cycle regulators have not led to increased growth and organ size (Mizukami *et al*., 2000). For example, overexpression of E2Fa and Dpa,

two transcription factors involved in the activation of cell cycle genes, induces extra cell divisions in Arabidopsis plants, but also severely inhibits overall growth of the plant (De Veylder *et al*., 2002) and overexpression of the D-type cyclin *CycD3* in Arabidopsis does not increase organ size, but leads to a disturbed organogenesis, with numerous small, incompletely differentiated cells (Riou-Khamlichi *et al*., 1999).

Besides cell proliferation, recent studies have shown that differences in cell size or cell polar elongation apparently can also contribute to the size difference of plant organ (Kondorosi *et al*., 2000). For example, the mutants *struwwelpeter* (*swp*) with altered organ size (dwarf plants with reduced leaf size and leaf number) shows reduced cell numbers in all aerial organs and in certain cases, this defect is partially compensated by an increase in final cell size (Autran *et al*., 2002). Moreover, Mizukami & Fischer. (2000) reported that cell growth occurred without cell division in *ant-1* petals, resulting in extremely large cells and the loss-of-function *ant-1* mutation uncouples cell proliferation and growth, resulting in organs with fewer cells whose size is larger than normal.

In most tissues, cell proliferation is coordinated with growth such that cells double their size before dividing in two. In general, mutations that block the cell cycle generally do not interfere with cell growth and conversely, mutations affecting metabolism coordinately arrested both cell growth and division (Neufeld *et al*., 1998; Mizukami *et al*., 2000). Although coordination between these two processes has been recognized for decades in a variety of organisms, in the budding yeast *Saccharomyces cerevisiae*, this coordination or 'size control' appears as an inverse correlation between cell size and the rate of cell-cycle progression, but still cell growth and division are not entirely separated processes and dependencies in cell-cycle progression and cell growth within and across cells in a dividing population has been also observed in yeast (Mayhew *et al*., 2017).

Hence, our results provide additional information about tissue-specific cell size/number, ploidy level and cell cycle progression in somatic and meristematic tissues and indicate that the large organ size in Bur-0 can be mainly attributed to its larger cells, and for the vegetative and reproductive SAM, to a higher cell proliferation as well. Our results contribute to a better understanding of the mechanisms that regulate organ size per accession and determine particular phenotypes.

In order to further investigate cell cycle and cell growth regulation in Bur-0 and its role in organ size determination, the constantly dividing SAM meristematic cells would be an ideal cell type for further cell cycle/cell size regulation analyses. Additionally, transcript analysis of additional cell cycle/cell size regulators and/or availability of Bur-0 SAM marker lines that allow analyses on single cells could provide a better understanding of the mechanisms regulating cell size, cell growth and determination of the bigger Bur-0 phenotype, but could also provide a good model for more sophisticated analyses of the interaction of cell proliferation, cell size, organ size and morphogenesis.

4.7. RNA-seq analysis reveals accession-specific developmental regulators, shared across tissues and developmental stages

Our RNA-seq data revealed different transcriptomes profiles according to the tissue (SAM and embryo), developmental stage (late torpedo and green mature stages for embryos and vegetative and floral transition stages for the SAM) and accession. Moreover, differentially expressed genes, significantly changed and up- and down-regulated were identified between the two accessions (Col-0 *vs.* Bur-0) at each tissue and developmental stage analyzed and the overlap analysis revealed unique and shared groups of genes significantly changed, up- and down-regulated between the two accessions (Col-0 *vs.* Bur-0) at each tissue and developmental stage analyzed. Furthermore, the gene ontology terms analysis (GO) for biological process considering all differentially expressed genes found between the two accessions (Col-0 *vs* Bur-0) revealed fifteen GO categories enriched, among which ´cellular processes` (GO:0009987), ´metabolic process` (GO: 0008152), ´response to stimulus` (GO: 0050896), ´biological regulation` (GO: 006507), ´localization` (GO: 0051179), ´signaling` (GO: 0023052) and ´growth` (GO: 0040007) are overrepresented (Figure 28). Interestingly, the two gene ontology terms more enriched correspond to cellular processes and metabolic process, in agreement with differences found between the two accessions in terms of the metabolic status and cell size described in previous sections.

Moreover, clustering analysis considering the DAGs among tissues/stages and the two accessions Bur-0 and Col-0 resulted in 30 clusters, each of them with a set of genes with higher or lower expression per tissue and stage in Bur-0 and Col-0, respectively (Figure 29). Interestingly, two particular clusters contain a set of genes whose expression is high or low, consistently across tissues and developmental stages in Bur-0 (cluster 9 with 41 genes) or in Col-0 (cluster 10 with 44 genes), respectively, revealing candidate genes with potential roles in plant growth and determination of accession-specific phenotypes.

These genes were used to analyze gene ontology terms (GO) for biological process, but due to the low gene number per cluster and also probably because most of them are unknown/uncharacterized genes, less than 5% of the genes were classified. A more detailed characterization of the 85 candidate genes was done based on information available in databases and they were classified in the gene types: Protein coding, transposable elements, novel transcribed and undetermined (Figure 30). The reported functions of some of the protein coding ones are described in the Supplementary Table S7.

4.8. Expression analysis of candidate genes validate RNA-seq data

Five candidate genes with higher expression in Bur-0 (*AT4G08874*, *AT1G71920*, *AT2G03965*, *AT2G43960*, *AT5G15360*), and five with higher expression in Col-0 (*AT3G44430*, *AT4G07825*, *AT5G05060*, *AT4G11830*, *AT2G04378*) were initially selected for expression analysis by qRT-PCR and RNA *in situ* hybridization. Expression levels of the candidate gene *AT4G08874* (Bur-0 specific) analyzed by qRT-PCR were significantly higher in Bur-0 than in Col-0 at each tissue and developmental stage analyzed. The expression pattern of *AT4G08874* coincides with the pattern observed from the raw read counts obtained from the transcriptome data, where the

transcript was barely detected in Col-0 and highly abundant in Bur-0, particularly in SAM tissue at floral transition stage. Expression analysis by RNA *in situ* hybridization showed that *AT4G08874* transcript is detected in Bur-0, but not in Col-0 tissues. In addition, the expression signal in Bur-0 was weak on embryos at late torpedo stage and SAM at vegetative stage, but strong in the SAM at floral transition stage and on embryos at green mature stage.

According to the currently available information in databases, *AT4G08874* encodes a transmembrane protein (Beta-galactosidase related protein) that has high similarity to the protein Q3EA64 or A0A1P8B5I4. Beta galactosidases (BGALs) are glycosyl hydrolases that remove terminal β-D-galactosyl residues from β-D-galactosides. There are 17 predicted BGAL genes in the genomes of both Arabidopsis (*BGAL1-17*) and tomato (*TBG1-17*), all have BGAL activity but their distinct expression profiles and ancient phylogenetic separation indicates that these enzymes fulfil diverse, non-redundant roles in plant biology. The majority of these BGALs are predicted to act during cell wall-related biological processes like remodeling and expansion (Chandrasekar & van der Hoorn, 2016).

The physiological roles of only two Arabidopsis BGALs have been characterized. One is the *mucilage-modified2* (*mum2*) which fails to extrude mucilage from the apoplast upon hydration and is caused by the *bgal6* mutant allele, indicating that BGAL6 alters the hydration properties of mucilage by modifying carbohydrate structures (Dean *et al*., 2007). The second is the *bgal10* mutant with unusual xyloglucan accumulation in cell walls, which is correlated with a reduced silique and sepal length of *bgal10* mutant plants (Sampedro *et al*., 2012). The physiological roles of BGALs have also been studied in other plant species, including plant growth and fruit ripening and softening-related BGALs have been purified from the fruits of muskmelon (*Cucumis melo*) (Ranwala *et al*., 1992), avocado (*Persea americana*) (De Veau *et al*., 1993), coffee (*Coffea arabica*) (Golden *et al*., 1993), apple (Ross *et al*., 1994), mango (*Mangifeera indica*) (Ali *et al*., 1995), and tomato (Carrington & Pressy, 1996).

On the other hand, expression analysis of the candidate gene *AT3G44430* (Col-0 specific) by qRT-PCR revealed significantly higher expression levels in Col-0 than in Bur-0 at each tissue and developmental stage analyzed. The expression pattern of *AT3G44430* coincides with the pattern observed from the raw read counts obtained from the transcriptome data, where the transcript was barely detected in Bur-0 and highly abundant in Col-0, particularly in SAM tissue at vegetative and floral transition stage. Expression analysis by RNA *in situ* hybridization showed that *AT3G44430* transcript is detected in Col-0, but not in Bur-0 tissues and the expression signal in Col-0 was weak on embryos, but stronger in the SAM, particularly at floral transition stage. According to the currently available information in databases, *AT3G44430* encodes a transmembrane protein that is expressed in several plant tissues at different developmental stages, with predicted molecular function and predicted biological function in growth, developmental stages, and plant structure, but no further evidence is available from the literature to ascribe a function, process or component term.

Our results are in agreement with other transcriptome analyses in Arabidopsis were higher expression levels of *AT4G08874* in Bur-0 than in Col-0 have been reported, as well as higher expression levels of *AT3G44430* in Col-0 and very low in Bur-0 (expression levels below the

implemented cutoff) (Gan *et al*., 2011). The expression pattern of *AT4G08874* (Bur-0 specific) and *AT3G44430* (Col-0 specific) reported in our study demonstrates that both are expressed in actively growing organs and transcript accumulation changes in relation to different tissues and stages of development, suggesting developmental specific functions but also an accession-specific role in growth throughout embryonic and postembryonic development.

Expression analysis of the remaining candidate genes selected for validation was completed by qRT-PCR and raw read counts per gene from RNA-seq data were also plotted for expression pattern comparisons. Our results demonstrate similar expression patterns between the qRT-PCR results and the raw read counts per gene from the RNA-seq data set. Thus, our qRT-PCR and RNA *in situ* hybridization results support our transcriptome data. Many of those candidate genes encode proteins that have been ascribed to different functions, such as transporter proteins, RNA Binding / processing, kinase activity, regulation of flower development, progression of meiosis during early prophase, phosphorylation/signal transduction, stress response/heat shock, among others.

4.9. Expression analysis of known key growth regulators validate RNA-seq data

In order to better understand the molecular basis of the big phenotype in Bur-0, we searched for marker genes reported in the literature that are involved in seed size regulation, plant growth and shoot apical meristem (SAM) maintenance and we analyzed whether different expression patterns and levels of such known regulators might determine accession-specific phenotypes. Although none of the selected known and previously reported key regulators are present in our high confidence set of DEGs from our RNA-seq data, they are present in the data set corresponding to total genes counted, therefore the raw read counts per gene were also plotted for expression pattern comparisons and further validation of our transcriptome data.

Because our results indicated that the SAM might play a crucial role for establishing the big Bur-0 phenotype, we analyzed expression of the SAM maintenance genes *CLAVATA3* (*CLV3*), *WUSCHEL* (*WUS*) and *SHOOT MERISTEMLESS* (*STM*). Interestingly, particularly bigger *CLV3*, *WUS* and *STM* expression domains were detected in Bur-0 towards late embryogenesis and the floral transition time point, suggesting a different SAM organization in Bur-0 during late embryonic and postembryonic development (Figure 33D, E, F). Expression levels of *CLV3*, *WUS* and *STM* were analyzed by qRT-PCR for embryo tissue at late torpedo and green mature stages as well as SAM tissue at vegetative and floral transition stages from the Arabidopsis accessions Bur-0 and Col-0. Both accessions have similar *CLV3* expression levels at late torpedo, mature embryo and the SAM at vegetative stage, while significantly lower *CLV3* expression levels were found in Bur-0 SAM at transition stage compared to Col-0 (Figure 34).

On the other hand, expression levels of *WUS* and *STM* are similar for both accessions at late torpedo, mature embryo and the SAM at vegetative stage, while *WUS* and *STM* expression levels are significantly higher in Bur-0 SAM at floral transition stage compared to Col-0. A similar expression pattern was observed for *CLV3, WUS* and *STM* raw read counts from our transcriptome data (Figure 34), thus further confirming the reliability of transcriptome datasets generated in this study.

Meristem maintenance genes are crucial for the establishment of the proper plant body architecture and they are expressed in between plant developmental stages, becoming active already during embryogenesis (Clark, 2001; Sharma & Fletcher, 2002; Scofield *et al*, 2014). In self-regulatory models of the SAM has been proposed that the *CLV-WUS* negative feedback loop has evolved to control the renewal and identity of the stem cells and a positive regulator of *WUS* expression is *STM*, although *STM* may be a more general regulator of meristem cell development, since it is expressed throughout the meristem and not specifically in the stem cell region (Schoof *et al*., 2000).

The perturbation of the SAM regulatory network results in phenotypic changes of the plant, for example, the *clv3* mutant (loss of function mutation) develops an enlarged SAM at the time of the transition to flowering (Clark *et al*., 1995; Sharma & Fletcher, 2002). Interestingly, the *clv3* mutant phenotype and the proposed SAM regulatory model resemble our results, where enlarged SAMs, significantly lower levels of *CLV3* and higher levels of *WUS* and *STM* were found in the SAM at floral transition stage of the natural accession Bur-0, but not in the Col-0 equivalent, demostrating that the different SAM organization in Bur-0 plays a key role in determination of its large phenotype.

On the other hand, among the genes known from the literature to be involved in shoot growth and seed size regulation, we selected *PIGMENT DEFECTIVE 327* (*PDE327*) because is the only gene to our knowledge that has been reported responsible for a shoot specific phenotype for Bur-0, and the seed size regulator *ARABIDOPSIS G PROTEIN GAMMA SUBUNIT 3* (*AGG3*) because although the function of *AGG3* has been reported and well-studied, detailed tissue specific expression and localization of *AGG3* in Arabidopsis has not yet been described. Thus, in this study expression of *PDE327* and *AGG3* was analysed by RNA *in situ* hybridization on embryo and shoot apical meristem tissue at different developmental stages, for the Arabidopsis accessions Ws-2, L*er*-1, Col-0 and Bur-0.

Natural variation for shoot growth using quantitative genetic strategies was analyzed by Vlad *et al.* (2010), using a Bur-0 × Col-0 recombinant inbred line set and a fine-mapping strategy. The authors suggested that *PDE327* (*AT4G30720*), which encodes a chloroplast-located protein, might be responsible for phenotypic variation and particularly responsible for a defective growth phenotype (pale and smaller plants at vegetative stage). According to Vlad *et al.* (2010), two copies of *PDE327* are present in the Bur-0 genome, one of which is not functional, and only one (functional) copy is present in the Col-0 genome and this polymorphism can cause natural variation in shoot growth. In our study, the strongest signal of *PDE327* was detected on embryo samples during early embryogenesis, leaf primordia and SAM tissue at floral transition stage for the accessions L*er*-1 and Bur-0 and a similar expression pattern with weaker signal was found in Ws-2 and Col-0 tissues, indicating that *PDE327* might play a role in growth regulation during the transition from early to late embryogenesis and towards the reproductive transition, this role might be accession-specific, but is not specific for Bur-0 (Figure 35A, B).

The *Arabidopsis G Protein GAMMA SUBUNIT 3* (*AGG3*) is involved in the regulation of many agronomically important traits in plants, including seed yield, organ size, abscisic acid (ABA)-

dependent signaling, stress responses and nitrogen use efficiency (Kaur *et al*., 2018). Overexpression of *AGG3* in both Arabidopsis and Camelina resulted in stress resistant and larger plants, which produce bigger seeds compared to control, while the knockout mutants of *AGG3* in Arabidopsis have an opposite phenotype suggesting a direct, positive correlation between the *AGG3* and the observed phenotypes (Roy Choudhury *et al*., 2014). In our study, a weak expression signal of *AGG3* was detected during early embryogenesis and no signal was detected in the SAM at vegetative stage for all the analyzed accessions, suggesting that *AGG3* is absent during late embryogenesis and during early vegetative growth.

Interestingly, we found that *AGG3* transcript is present in the SAM at floral transition stage, particularly in the region where new floral primordia are formed for all accessions. Different factors have been shown to be absent at the vegetative SAM but strongly induced during floral transition, for example *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 3* (*SPL3*, *4* or *5*) (Cardon *et al*., 1997). Thus, induction of *AGG3* during floral transition in the SAM indicates that *AGG3* might be involved in the regulation of the flowering time in *Arabidopsis thaliana* plants, however this role is not specific for the accession Bur-0 (Figure 34B, C).

4.10. Integration of the RNA-seq results with previous research

Growth is a complex trait determined by the interplay between many genes, some of which play a role at a specific moment during development whereas others play a more general role (Bac-Molenaar *et al*., 2015). The functional importance of most growth-related genes is not equal during all developmental stages and in all tissues, and many display specific temporal and spatial expression profiles (Schmid *et al*., 2005). In addition, some genes play an essential role in the overall development of the plant, whereas others are mainly important if the plant has to cope with specific environmental conditions (Geng *et al*., 2013).

These tightly regulated genes form a robust network that enables the plant to complete its life cycle under many different circumstances. A better knowledge of the genetic factors that are involved in growth regulation would help in the understanding of the mechanisms underlying different growth patterns as observed in nature (Bac-Molenaar *et al*., 2015), such dynamic patterns are better understood when growth and its regulation are studied over time, instead of at a single time point (Tessmer *et al*., 2013). Thus, instead of single time point analyses, in this study different stages during embryonic and postembryonic development were analyzed and therefore the high confidence transcriptome datasets generated in this study provide accession, tissue and developmental stage specific gene sets with potential key regulators, useful for elucidation of molecular mechanisms regulating plant growth and development, as well accession-specific phenotypes in Arabidopsis.

On the other hand, elucidation of genetic basis underlying natural variation and traits of agronomical importance in wild and crop plants has been done using different approaches including functional analyses and characterizations of mutants, transcriptomes, and mapping analyses. In Arabidopsis species-wide sequencing studies, one should expect to commonly face new DNA sequences, for which we have no reference and/or no idea of the insertion context, as it is clear that most Arabidopsis accessions have genome sizes 5 to 10% larger than the

reference Col-0 genome (Schmuths *et al*., 2004). In agreement with this, Gan *et al*. (2011) sequenced and accurately assembled the single-copy genomes of 18 Arabidopsis accessions and for each accession there were 497,668–789,187 single-base differences from Col-0, and about 45,000 ambiguous nucleotides.

Gan *et al.* (2011) also found that from 3.07 million SNPs, 45.2% were private to single accessions and they identified 1.20 million indels, and 104,090 imbalanced substitutions, in which a sequence in Col-0 was replaced by a different sequence. Although 57.5% of indels or imbalanced substitutions were shorter than 6 bp, 1.9% were longer than 100 bp, and overall 14.9 Mb of Col-0 sequence was absent in one or more accessions. These results show that the functional consequences of polymorphisms are often difficult to detect and interpret in the absence of gene re-annotation and thus the importance of RNA-seq data for annotation (Gan *et al*., 2011).

Interestingly, beyond quantifying gene expression, the data generated by RNA-seq facilitate the discovery of novel transcripts, identification of alternatively spliced genes and detection of allele-specific expression (Kukurba *et al*., 2015). In this regard our study provides a high confidence dataset of important developmental regulators, which also contains a set of accessions, tissue and stage-specific uncharacterized candidate markers for follow-up gene expression, mutagenesis and functional characterization studies, but also for identification of gene isoforms that could play a role in determination of accession-specific phenotypes.

Moreover, RNA-seq has also revolutionized quantitative trait locus (QTL) analyses because it enables association analyses and the combination of RNA-seq with genetic variation data has enabled the identification of genetic loci correlated with gene expression variation (Kukurba *et al*., 2015). In agreement with this, Qi *et al*. (2021) reported that mapping combined with expression and variant analyses in switchgrass identified a cluster of gene candidates underlying the variation in leaf wax between upland and lowland switchgrass ecotypes. Moreover, Derakhshani *et al*. (2021) reported that the combination of QTL mapping and RNA-seq profiling revealed candidate genes associated with cadmium tolerance in barley.

Alternative resources for the genetic dissection of complex traits include the Multiparent Advanced Generation Inter-Cross (MAGIC) lines, which consist of a set of 527 recombinant inbred lines (RILs) descended from a heterogeneous stock of 19 intermated accessions of *Arabidopsis thaliana* (Kover *et al*., 2009). Interestingly, Bur-0 is one of the founding parents of the MAGIC lines and it has been included in some mapping analyses.

Simon *et al*. (2008) reported an integrated set of 15 new large *Arabidopsis thaliana* recombinant inbred line (RIL) populations optimized for QTL mapping, having Columbia as a common parent crossed to distant accessions like Bur-0. Five of the populations were validated by investigating three traits: flowering time, rosette size, and seed production as an estimate of fitness. The correlation coefficients between the three analyzed traits in the Col-0 x Bur-0 RIL population revealed positive and significant correlations between flowering time and rosette diameter, while a significant negative correlation was obtained between flowering time and total seed weight (Simon *et al*., 2008).

In addition, Simon *et al*. (2008) reported that flowering time variation is explained by six QTL in the Col-0 x Bur-0 RIL population, five of which have allelic effects in the same direction (the Col-0 allele accelerates flowering). Rosette diameter is explained by one to seven QTL, some of them colocalize with flowering time ones and their effects are always in the same direction: the earliest plants are the smallest. Moreover, in the Col-0 x Bur-0 RIL, some of the total seed weight QTL colocalize with QTL for flowering time or flowering time/rosette diameter (Simon *et al*., 2008). Furthermore, Gnan *et al.* (2014) conducted a QTL analysis for seed size in *A. thaliana* and for six of the eight seed size QTL identified, the allele conferring the largest seed size was from the accession Bur-0. At other two QTL, the Bur-0 allele led to the second largest seed size.

A comprehensive understanding of gene regulatory networks operating at different stages of development requires a wide transcriptome coverage in different tissues/cell types of the developing organs as well as studies within Arabidopsis accessions with remarkable phenotypical differences. The expanding potential of RNA-seq to associate phenotypic variations with genetic variation offers an enhanced understanding of gene regulation (Kukurba *et al*., 2015) and once QTL have been mapped, the next step is to identify the genes responsible for these QTL (Simon *et al*., 2008). Here, our RNA-seq data provide a valuable resource for future research directions to see if any of the DEGs identified in this study colocalize with the previously reported QTLs, as a complementary approach to identify target genes responsible for particular accession-specific agronomically important traits like yield, big seed size or large rosettes that can have future application for crop improvement.

4.11. Conclusion

This study demonstrates that no other Arabidopsis accession has simultaneously larger rosettes, larger SAM, later flowering phenotype and larger seeds as observed in Bur-0. In addition, our results show that rosette area, seed area and flowering time are independent traits influenced by different factors and stage-specific rosette size and the flowering time are good predictors of the SAM size at floral transition stage. Moreover, developmental stage-based comparisons revealed that Bur-0 has a large embryo phenotype that is achieved during late embryogenesis and a large rosette phenotype that is achieved during late postembryonic growth. Our detailed characterization provides a valuable resource of precisely defined phenotypical traits to be used as guideline for further characterization studies in Arabidopsis.

Furthermore, accession-specific physiological and metabolic traits that underlie accession-specific phenotypes were identified in this study and in particular, more carbon resources during embryonic and postembryonic development were found in Bur-0, suggesting an important role of carbon resources in determination of the bigger Bur-0 phenotype. The current results provide a baseline for further research that could reveal the role of increased levels of carbohydrates in growth and developmental phase transitions regulation.

Moreover, we found that the large organ size in Bur-0 can be mainly attributed to its larger cells, and for the vegetative and reproductive SAM, to a higher cell proliferation as well, but

not to a different ploidy level. Since our results revealed larger cells in enlarged mature embryos in Bur-0, a tissue and developmental stage where cell proliferation is arrested, our results suggest that a larger cell size might contribute to the bigger organ size in Bur-0 in a greater extent than cell proliferation. These results provide a baseline for future research to elucidate molecular mechanisms regulating cell size, cell growth, but could also provide a good model for more sophisticated analyses of the interaction of cell proliferation, cell size, cell growth, organogenesis and morphogenesis.

Additionally, RNA-seq analysis revealed different transcriptomes profiles according to the tissue (SAM and embryo), developmental stage (late torpedo and green mature stages for embryos and vegetative and floral transition stages for the SAM) and accession. Accession-specific developmental regulators were identified, shared across embryo and SAM tissue at different developmental stages. Expression analyses of candidate genes selected for validation of RNA-seq data, as well as known SAM regulators demonstrate that expression patterns are in agreement with the expression values and raw read counts obtained from ourtranscriptome data. These results demonstrate that the high confidence transcriptome datasets generated in this study are reliable and useful for elucidation of molecular mechanisms regulating plant growth and accession-specific phenotypes in Arabidopsis.

Moreover, our RNA-seq datasets provides a set of accession, tissue and stage-specific uncharacterized candidate markers for follow-up gene expression, mutagenesis and functional characterization studies, but also for identification of gene isoforms that could play a role in determination of accession-specific phenotypes. In addition, our RNA-seq resource provides valuable guidelines for future research directions to elucidate the molecular mechanisms underlying interesting plant traits, but also for identification of target genes responsible for particular accession-specific agronomically important traits like yield, big seed size or large rosettes that can have future application for crops improvement.

Taken together, this PhD project contributes to the plant development research field providing a detailed analysis of mechanisms underlying plant growth and development at different levels of biological organization, focusing on Arabidopsis accessions with remarkable phenotypical differences. For this, the natural accession Bur-0 was an ideal candidate and different mechanisms at organ and tissue level, cell level, metabolism, transcript and gene expression level were identified, providing a better understanding of different factors involved in plant growth regulation and mechanisms underlying different growth patterns in nature.

4.12. Outlook

- In this study Bur-0 was crossed with L*er*-1, Col-0, Ws-2, respectively, and our results showed that the F1 in all cases, using Bur-0 as parental genotype, either female or male genotype, resulted in bigger seeds. This results indicates parental effects on seed size and the significant size increase obtained when Bur-0 was used as the female genotype suggests that such effect might be rather maternally controlled. Parental effects on seed size can exist in *A. thaliana*, but natural variation among paternal genotype in seed size has not yet been demonstrated (Scott *et al*., 1998). In addition, molecular characterization of such effect remains scarce and therefore a better understanding of the genetic control of seed size would help evaluate the different evolutionary theories for the maintenance of heritable variation in seed size. From an applied perspective, identification of the contribution of maternal, paternal and developmental factors to seed size and seed number could be very useful in developing strategies to improve grain yield.

- In this study, a positive relation between the rosette size, rosette biomass and the levels of carbohydrates was identified in Bur-0. Whether the increased levels of carbohydrates in Bur-0 rosettes also have implications in regulation of growth and developmental phase transitions remains to be elucidated. Analyses of carbon-regulated transcripts combined with measurements of enzyme activities could provide a better understanding of the role that particular metabolites might play in determination of the Bur-0 phenotype during postembryonic development.

- We found that the large organ size in Bur-0 can be mainly attributed to its larger cells, and for the vegetative and reproductive SAM, to a higher cell proliferation as well, but not to a different ploidy level. These results provide a baseline for future research and further transcript analysis of cell cycle/cell size regulators and/or availability of Bur-0 SAM marker lines that allow analyses on single cells could provide a better understanding of the mechanisms regulating cell size, cell growth and determination of the bigger Bur-0 organ size, but could also provide a good model for more sophisticated analyses of the interaction of cell proliferation, cell size, cell growth, organogenesis and morphogenesis.

- The transcriptome analysis conducted in this study provides a high confidence dataset of growth and developmental regulators, which also contains a set of accession, tissue and stage-specific uncharacterized candidate markers for follow-up gene expression, mutagenesis and functional characterization studies, but also for identification of gene isoforms that could play a role in determination of accession-specific phenotypes. In addition, our RNA-seq resource provides valuable guidelines for future research directions to elucidate the molecular mechanisms underlying interesting plant traits, but also for identification of target genes responsible for particular accession-specific agronomically important traits like yield, big seed size or large rosettes that can have future application for crops improvement.

5. Acknowledgements

I want to thank Prof. Dr. Bernd Müller-Röber and Dr. Justyna Olas for giving me the opportunity to work on this interesting PhD project under their supervision. I also want to express my gratitude to the whole AG Müller-Röber for their assistance, support and good work environment. I want to thank my PhD advisory committee, Dr. John Lunn, Dr. Duarte Figueiredo and Dr. Friedrich Kragler for their supportive guidance and scientific advice.

Additionally, I want to thank the cooperation partners from the MPI of Molecular Plant Physiology for their involvement and support in different analyses during this PhD project:

- Dr. Maria Grazia Annunziata for metabolites analysis,
- Dr. Frank Machin for ploidy analysis by fluorescence activated cell sorting (FACS),
- Dr. Federico Apelt for plant growth analysis using the phenotyping system Phytotyping^{4D},
- Dr. Federico Apelt and Dr. Saurabh Gupta for RNA-seq data analysis,

Furthermore, I want to express my gratitude to the members of AG-Grebe for the friendly work environment in Haus 20 and special thanks to Dr. Michael Sauer for his help and advice regarding microscopy and imaging experiments. I also want to thank Dr. Duarte Figueiredo for his help and advice regarding embryo tissue processing for different experiments.

Special thanks to Magdalena Jarosz (ERASMUS internship WS/2018) and Pawel Kanejski, (ERASMUS internship WS/2019) for their enthusiasm and help during phenotypical characterization experiments with natural accessions and mutant lines, but also for their engagement with daily work during their internships in AG-Müller-Röber.

I also want to thank my IMPRS colleagues for the friendly environment and helpful scientific discussions, as well as the coordinator of the IMPRS program Dr. Ina Talke for her assistance with different administrative procedures. In addition, I want to thank the team from the International Office and the Potsdam Graduate School (PoGS) for their advice regarding different administrative procedures, as well as the Commission of Equal Opportunity for additional counter-financed support for completion of my PhD.

Finally, I want to express my gratitude to my family for their wholehearted support, understanding, encouragement and for standing by me through every step of my life. Thanks to my friends for their support and for cheering me up when I needed it most.

6. References

Abe, M., Katsumata, H., Komeda, Y., & Takahashi, T. (2003). Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in Arabidopsis. Development. 130: 635-43. DOI: 10.1242/dev.002

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., & Araki, T. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science. 309 (5737): 1052-1056. DOI: 10.1126/science.1115983.

Agarwal, P., Kapoor, S., & Tyagi, A. K. (2011). Transcription factors regulating the progression of monocot and dicot seed development. Bioessays. 33: 189-202. https://doi.org/10.1002/bies.201000107.

Alberts, B., Johnson, A., Lewis, J., *et al*. (2002). Molecular Biology of the Cell. 4th edition. New York: Garland Science. Plant Development. Available from: https://www.ncbi.nlm.nih.gov/books/NBK26922/.

Ali, Z. M., Armugam, S., Lazan, H. (1995). β-Galactosidase and its significance in ripening mango fruit. Phytochemistry. 38: 1109-1114. https://doi.org/10.1016/0031-9422(94)00804-3.

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., . . . Ecker, J. R. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science. 301 (5633):653-7. doi: 10.1126/science.1086391. Erratum in: Science. 2003 Sep; 301(5641): 1849. PMID: 12893945.

Alonso-Blanco, C., Aarts, M., Bentsink, L., Keurentjes, J. J., Reymond, M., Vreugdenhil, D., & Koornneef, M. (2009). What has natural variation taught us about plant development, physiology, and adaptation? The Plant Cell. 21 (7): 1877–1896. https://doi.org/10.1105/tpc.109.068114.

Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K. M., *et al*. (2016). 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. Cell. 166: 481–491. doi: 10.1016/j.cell.2016.05.063.

Alonso-Blanco, C., Blankestijn-de Vries, H., Hanhart, C., & Koornneef, M. (1999). Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences. 96 (8): 4710-4717. DOI: 10.1073/pnas.96.8.4710.

Anastasiou, E., Kenz , S., Gerstung, M., MacLean, D., Timmer, J., Fleck, C., & Lenhard, M. (2007). Control of plant organ size by KLUH/CYP78A5-dependent intercellular signaling. Developmental Cell. 13 (6): 843-56. doi: 10.1016/j.devcel.2007.10.001.

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 31: 166-169. DOI: 10.1093/bioinformatics/btu638.

Apelt F, Breuer D, Nikoloski Z, Stitt M, & Kragler F. (2015). Phytotyping^{4D}: a light-field imaging system. The Plant Journal. 82:693-706. doi:10.1111/tpj.12833.

Apelt, F., Breuer, D., Olas, J. J., Annunziata, M. G., Flis, A., Nikoloski, Z., . . . Stitt, M. (2017). Circadian, carbon, and light control of expansion growth and leaf movement. Plant Physiology. 174 (3) :1949-1968. DOI: 10.1104/pp.17.00503.

Aranzana, M. J., Kim , S., Zhao, K., Bakker, E., Horton , M., Jakob, A., *et al*. (2005). Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. Plos Genetics. 1 (5): e60. https://doi.org/10.1371/journal.pgen.0010060.

Atwell, S., Huang, Y., & Vilhjálmsson, B. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature. 465: 627-631. DOI: https://doi.org/10.1038/nature08800.

Autran, D., Jonak, C., Belcram, K., Beemster, G. T., Kronenberger, J., Grandjean, O., . . . Traas, J. (2002). Cell numbers and leaf development in Arabidopsis: a functional analysis of the *STRUWWELPETER* gene. The EMBO Journal. 21: 6036-6049. https://doi.org/10.1093/emboj/cdf614.

Bac-Molenaar, J. A., Vreugdenhil, D., Granier, C., & Keurentjes, J. J. (2015). Genome-wide association mapping of growth dynamics detects time-specific and general quantitative trait loci. Journal of experimental botany. 66 (18): 5567–5580. https://doi.org/10.1093/jxb/erv176.

Bai, S., Chen, L., Yund, M., & Sung, Z. (2000). Mechanisms of plant embryo development. Current Topics in Developmental Biology. 50: 61-88. https://doi.org/10.1016/S0070-2153(00)50004-0.

Bai-Han, P., Yong-Hong, X., Feng, L., Ye-Ai, Z., & Zheng-Miao, D. (2018). Responses to sedimentation in ramet populations of the clonal plant *Carex brevicuspis*. Frontiers in Plant Science. 9: 512. DOI=10.3389/fpls.2018.00512.

Bartoli, C., & Roux, F. (2017). Genome-wide association studies in plant pathosystems: toward an ecological genomics approach. Frontiers in Plant Science. 8: 763. DOI=10.3389/fpls.2017.00763.

Baud , S., Wuillème, S., Lemoine, R., Kronenberger, J., Caboche, M., Lepiniec, L., & Rochat, C. (2005). The *AtSUC5* sucrose transporter specifically expressed in the endosperm is involved in early seed development in Arabidopsis. The Plant Journal. 43 (6): 824-36. doi: 10.1111/j.1365-313X.2005.02496.x.

Baud, S., Boutin, J. P., Miquel, M., Lepiniec, L., & Rochat, C. (2002). An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. Plant Physiology and Biochemistry. 40:151–160. https://doi.org/10.1016/S0981-9428(01)01350-X.

Bäurle, I., & Dean, C. (2006). The timing of developmental transitions in plants. Cell. 125 (4): 655-64. doi: 10.1016/j.cell.2006.05.005. PMID: 16713560.

Borisjuk, L., Walenta, S., Weber, H., Mueller-Klieser, W., & Wobus, U. (1998). High-resolution histographical mapping of glucose concentrations in developing cotyledons of *Vicia faba* in relation to mitotic activity and storage processes: Glucose as a possible

developmental trigger. The Plant Journal. 15: 583-591. https://doi.org/10.1046/j.1365-313X.1998.00214.x.

Camargo , A., Papadopoulou, D., Spyropoulou , Z., Vlachonasios , K., Doonan , J., & Gay, A. P. (2014). Objective definition of rosette shape variation using a combined computer vision and data mining approach. Plos One. 9 (5): e96889. https://doi.org/10.1371/journal.pone.0.

Cardon, G. H., Höhmann, S., Nettesheim , K., Saedler , H., & Huijser, P. (1997). Functional analysis of the *Arabidopsis thaliana* SBP-box gene *SPL3*: a novel gene involved in the floral transition. Plant Journal for Cell and Molecular Biology. 12 (2): 367-77. doi: 10.1046/j.1365-313x.1997.12020367.x. PMID: 9301089.

Carrington, C., & Pressy, R. (1996). β-Galactosidase II activity in relation to changes in cell wall galactosyl composition during tomato ripening. Journal of the American Society for Horticultural Science. 121: 132–136. https://doi.org/10.21273/JASHS.121.1.132.

Chandrasekar, B., & van der Hoorn, R. A. (2016). Beta galactosidases in Arabidopsis and tomato - a mini review. Biochemical Society Transactions. 44 (1): 150-8. doi: 10.1042/BST20150217. PMID: 26862200.

Chardon, F., Barthélémy, J., Daniel-Vedele, F., & Masclaux-Daubresse, C. (2010). Natural variation of nitrate uptake and nitrogen use efficiency in *Arabidopsis thaliana* cultivated with limiting and ample nitrogen supply. Journal of Experimental Botany. 61 (9): 2293-302. DOI: 10.1093/jxb/erq059. PMID: 20237091.

Cho, H.-T., & Cosgrove, D. J. (2000). Altered expression of expansin modulates leaf growth and pedicel abscission in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences. 97 (17): 9783-9788. doi: 10.1073/pnas.160276997.

Clark, S. E. (1997). Organ formation at the vegetative shoot meristem. The Plant Cell. 9 (7): 1067–1076.

Clark, S. E. (2001). Meristems: start your signaling. Current Opinion in Plant Biology. 4 (1): 28-32. https://doi.org/10.1016/S1369-5266(00)00131-X.

Clark, S. E., Jacobsen, S. E., Levin, J. Z., & Meyerowitz, E. M. (1996). The *CLAVATA* and *SHOOT MERISTEMLESS* loci competitively regulate meristem activity in Arabidopsis. Development. 122: 1567-1575.

Clark, S. E., Running, M. P., & Meyerowitz, E. M. (1995). *CLAVATA3* is a specific regulator of shoot and floral meristem development affecting the same processes as *CLAVATA1*. Development. 121: 2057-2067.

Cockcroft, C. E., den Boer, B. G. W., Healy, J. M., & Murray, J. A. H. (2000). Cyclin D control of growth rate in plants. Nature. 405: 575–579. https://doi.org/10.1038/35014621.

Corbesier, L., Lejeune, P., & Bernier, G. (1998). The role of carbohydrates in the induction of flowering in *Arabidopsis thaliana*: comparison between the wild type and a starchless mutant. Planta. 206: 131-137. https://doi.org/10.1007/s004250050383.

Coruzzi, G., & Bush, D. R. (2001). Nitrogen and carbon nutrient and metabolite signaling in plants. Plant Physiology. 125 (1): 61-64. DOI: 10.1104/pp.125.1.61.

Crawford, B. C. W., Sewell, J., Golembeski, G., Roshan, C., Long, J. A., & Yanofsky, M. F. (2015). Genetic control of distal stem cell fate within root and embryonic meristems. Science. 347: 655-59. DOI: 10.1126/science.aaa0196.

Cross, J. M., Von Korff, M., Altmann, T., Bartzenko, L., Sulpice, R., Gibon , Y., . . . Stitt, M. (2006). Variation of enzyme activities and metabolite levels in 24 Arabidopsis accessions growing in carbon-limited conditions. Plant Physiology. 142 (4): 1574-1588. DOI: 10.1104/pp.106.086629 .

Dante, R. A., Larkins, B. A., & Sabelli, P. A. (2014). Cell cycle control and seed development. Frontiers in Plant Science. 5: 493. doi: 10.3389/fpls.2014.00493.

de Jong, T. J., & Scott, R. J. (2007). Parental conflict does not necessarily lead to the evolution of imprinting. Trends in Plant Science. 12 (10): 439-43. doi: 10.1016/j.tplants.2007.07.003. Epub 2007 Sep 12. PMID: 17855155.

De Veau, E. J. I., Gross, K. C., Huber, D. J., & Watada, A. E. (1993). Degradation and solubilization of pectin by β-galactosidases purified from avocado mesocarp. Physiologia Plantarum. 87: 279-285. https://doi.org/10.1111/j.1399-3054.1993.tb01731.x.

De Veylder, L., Beeckman, T., & Inzé, D. (2007). The ins and outs of the plant cell cycle. Nature Reviews Molecular Cell Biology. 8:655-665. https://doi.org/10.1038/nrm2227.

De Veylder, L., Beeckman, T., Beemster, G. T., de Almeida Engler, J., Ormenese, S., Maes, S., . . . Inzé, D. (2002). Control of proliferation, endoreduplication and differentiation by the Arabidopsis E2Fa-DPa transcription factor. The EMBO Journal. 21: 1360-1368. https://doi.org/10.1093/emboj/21.6.1360.

de Vries, S. C., & Weijers, D. (2017). Plant embryogenesis. Current Biology Magazine. 27: R853-R909. https://www.cell.com/current-biology/pdf/S0960-9822(17)30562-6.pdf.

Dean, G. H., Zheng, H., Tewari, J., Huang, J., Young, D. S., Hwang, Y. T., . . . Haughn, G. W. (2007). The Arabidopsis *MUM2* gene encodes a β-Galactosidase required for the production of seed coat mucilage with correct hydration properties. The Plant Cell. 19 (12): 4007-4021. DOI: 10.1105/tpc.107.050609.

del Pozo, J. C., & Ramirez-Parra, E. (2015). Whole genome duplications in plants: an overview from Arabidopsis. Journal of Experimental Botany. 66: 6991-7003. https://doi.org/10.1093/jxb/erv432.

Deprost, D., Yao, L., Sormani, R., Moreau, M., Leterreux, G., Nicolai, M., . . . Meyer , C. (2007). he Arabidopsis TOR kinase links plant growth, yield, stress resistance and mRNA translation. EMBO Reports. 8 (9): 864-870. https://doi.org/10.1038/sj.embor.7401043.

Derakhshani B, Jafary H, Maleki Zanjani B, Hasanpur K, Mishina K, Tanaka T, *et al*. (2020). Combined QTL mapping and RNA-Seq profiling reveals candidate genes associated with cadmium tolerance in barley. PLoS ONE 15(4): e0230820. https://doi.org/10.1371/journal.pone.0230820

Dijken, A. J. H., Schluepmann, H., & Smeekens, S. C. M. (2004). Arabidopsis Trehalose-6-phosphate synthase 1 is essential for normal vegetative growth and transition to flowering. Plant Physiology. 135 (2): 969-977. http://doi.org/10.1104/pp.104.039743.

Disch, S., Anastasiou, E., Vijay, K., Sharma, K., Laux, T., Fletcher, J. C., & Lenhard, M. (2006). The E3 Ubiquitin Ligase BIG BROTHER controls Arabidopsis organ size in a dosage-dependent manner. Current Biology. 16: 272–279. DOI:https://doi.org/10.1016/j.cub.2005.12.026.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. . Bioinformatics: 29 (1): 15-21. DOI: 10.1093/bioinformatics/bts635.

Domínguez , F., & Cejudo, F. J. (2014). Programmed cell death (PCD): an essential process of cereal seed development and germination. Frontiers in plant science. 5: 366. https://doi.org/10.3389/fpls.2014.00366.

Eastmond, P. J., van Dijken, A. J., Spielman, M., Kerr, A., Tissier, A., Dickinson, H. G., . . . Graham, I. A. (2002). Trehalose-6-phosphate synthase 1, which catalyses the first step in trehalose synthesis, is essential for Arabidopsis embryo maturation. Plant Journal. 29: 225-235. https://doi.org/10.1046/j.1365-313x.2002.01220.x.

EFNCP. (2020). European forum on nature conservation and pastoralism. The Burren. www.efncp.org. http://www.efncp.org/hnv-showcases/ireland-the-burren/the-burren/facts/ (Accessesd on November 2020).

Feeser, I., & O'Connell, M. (2019). Fresh insights into long-term changes in flora, vegetation, land use and soil erosion in the karstic environment of the Burren, western Ireland. Journal of Ecology, 97: 1083-1100. https://doi.org/10.1111/j.1365-2745.2009.01533.x.

Fichtner, F., & Lunn, J. E. (2021). The role of Trehalose 6-Phosphate (Tre6P) in plant metabolism and development. Annual Review of Plant Biology. 72: 1. https://doi.org/10.1146/annurev-arplant-050718-095929.

Fichtner, F., Olas, J. J., Feil, R., Watanabe, M., Krause, U., Hoefgen, R., . . . Lunn, J. E. (2020). Functional features of TREHALOSE-6-PHOSPHATE SYNTHASE1, an essential enzyme in Arabidopsis. The Plant Cell. 32 (6): 1949-1972. DOI: 10.1105/tpc.19.00837.

Figueiredo, D., Batista, R., Roszak, P., Hennig, L., & Köhler, C. (2016). Auxin production in the endosperm drives seed coat development in Arabidopsis. eLife. 5: e20542. DOI: 10.7554/eLife.20542.

Galbraith, D. W., Harkins, K. R., & Knapp, S. (1991). Systemic endopolyplody in *Arabidopsis thaliana*. Plant physiology. 96 (3): 985-989. https://doi.org/10.1104/pp.96.3.985.

Gan, X., Stegle, O., Behr , J., & et al. (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. Nature. 477 (7365): 419-423. DOI: 10.1038/nature10414.

Gazzarrini, S., Tsuchiya, Y., Lumba , S., Okamoto, M., & McCourt, P. (2004). The transcription factor FUSCA3 controls developmental timing in Arabidopsis through the hormones gibberellin and abscisic acid. Developmental Cell. 7 (3): 373-85. doi: 10.1016/j.devcel.2004.06.017.

Geng, Y., Wu, R., Wee, C., Xie, F., Wei, X., Chan, P., . . . Dinneny, J. R. (2013). A spatio-temporal understanding of growth regulation during the salt stress response in Arabidopsis. The Plant Cell. 25 (6): 2132-2154. DOI: 10.1105/tpc.113.112896.

Gibon, Y., Pyl, E. T., Sulpice, R., Lunn, J. E., Höhne, M., Günther, M., & Stitt , M. (2009). Adjustment of growth, starch turnover, protein content and central metabolism to a decrease of the carbon supply when Arabidopsis is grown in very short photoperiods. Plant Cell and Environment. 32 (7): 859-74. doi: 10.1111/j.1365-3040.2009.01965.x. Epub 2009 Feb 19. PMID: 19236606.

Gibson, S. I., Laby, R. J., & Kim, D. (2001). The *sugar-insensitive1* (*sis1*) mutant of Arabidopsis is allelic to *ctr1*. Biochemical and Biophysical Research Communications. 280: 196-203. https://doi.org/10.1006/bbrc.2000.4062.

Ginzberg, M. B., Kafri, R., & Kirschner, M. (2015). On being the right (cell) size. Science. 348 (6236): 1245075. DOI: 10.1126/science.1245075.

Gnan , S., Priest , A., & Kover, P. X. (2014). The genetic basis of natural variation in seed size and seed number and their trade-off using *Arabidopsis thaliana* MAGIC lines. Genetics. 198 (4): 1751-8. doi: 10.1534/genetics.114.170746. Epub 2014 Oct 13. PMID: 25313128; PMCID: PMC4256784.

Golden, K. D., John, M. A., & Kean, E. A. (1993). β-Galactosidase from *Coffea arabica* and its role in fruit ripening. Phytochemistry. 34: 355-360. https://doi.org/10.1016/0031-9422(93)80008-G.

González, R., Butković, A., Rivarez, M. P. S., & *et. al*. (2020). Natural variation in *Arabidopsis thaliana* rosette area unveils new genes involved in plant development. Scientific Reports. 10: 17600. https://doi.org/10.1038/s41598-020-74723-4.

Grant, J. J., Chini, A., Basu, D., & Loake, G. J. (2013). Targeted activation tagging of the Arabidopsis *NBS-LRR* gene, *ADR1*, conveys resistance to virulent pathogens. Molecular Plant-Microbe Interactions. 16 (8): 669-80. doi: 10.1094/MPMI.2003.16.8.669. PMID: 12906111.

Grossman, A., & Takahashi, H. (2001). Macronutrient utilization by photosynthetic eukaryotes and the fabric of interactions. Annual Review of Plant Physiology and Plant Molecular Biology. 52: 163–210. https://doi.org/10.1146/annurev.arplant.52.1.163.

Hanada, K., Sawada, Y., Kuromori, T., Klausnitzer, R., Saito, K., Toyoda, T., . . . Hirai, M. Y. (2011). Functional compensation of primary and secondary metabolites by duplicate genes in *Arabidopsis thaliana*. Molecular biology and Evolution. 28 (1): 377-382. https://doi.org/10.1093/molbev/msq204.

Herridge, R. P., Day, R. C., Baldwin , S., & Macknight, R. C. (2011). Rapid analysis of seed size in Arabidopsis for mutant and QTL discovery. Plant Methods. 7: 3. DOI: 10.1186/1746-4811-7-3.

Hirsch, J., Lefort, V., Vankersschaver, M., Boualem, A., Lucas, A., Thermes, C., . . . Crespi, M. (2006). Characterization of 43 non-protein-coding mRNA genes in Arabidopsis, including the *MIR162a*-derived transcripts. Plant Physiology. 140 (4): 1192-1204, https://doi.org/10.1104/pp.105.073817.

Hu, Y., Poh, H. M., & Chua, N. H. (2006). The Arabidopsis *ARGOS-LIKE* gene regulates cell expansion during organ growth. Plant Journal. 47 (1): 1-9. doi: 10.1111/j.1365-313X.2006.02750.x.

Hu, Y., Xie, Q., & Chua, N. H. (2003). The Arabidopsis auxin-inducible gene *ARGOS* controls lateral organ size. The Plant cell. 15 (9): 1951-1961. https://doi.org/10.1105/tpc.013557.

Jacqmard, A., Gadisseur, I., & Bernier, G. (2003). Cell division and morphological changes in the shoot apex of *Arabidopsis thaliana* during floral transition. Annals of Botany. 91 (5): 571-576. https://doi.org/10.1093/aob/mcg053.

Jahnke, S., & Scholten, S. (2009). Epigenetic resetting of a gene imprinted in plant embryos. Current Biology 19: 1677-1681. DOI:https://doi.org/10.1016/j.cub.2009.08.053.

Juenger, T., Pérez-Pérez, J. M., Bernal, S., & Micol, J. L. (2005). Quantitative trait loci mapping of floral and leaf morphology traits in *Arabidopsis thaliana*: evidence for modular genetic architecture. Evolution and Development. 7 (3): 259-71. doi: 10.1111/j.1525-142X.2005.05028.x. PMID: 15876198.

Kaur, J., Roy Choudhury , S., Vijayakumar, A., Hovis, L., Rhodes, Z., Polzin, R., . . . Pandey, S. (2018). Arabidopsis type III Gγ Protein AGG3 is a positive regulator of yield and stress responses in the model monocot *Setaria viridis*. Frontiers in Plant Science. 9:109. https://doi.org/10.3389/fpls.2018.00109.

Kearsey, M. (1998). The principles of QTL analysis (a minimal mathematics approach). Journal of Experimental Botany. 49: 1619-1623.

Kearsey, M., & Farquhar, A. (1998). QTL analysis in plants; where are we now?. Heredity. 80: 137-142. DOI https://doi.org/10.1046/j.1365-2540.1998.00500.x .

Kim, D.-Y., Scalf, M., Smith, L. M., & Vierstra, R. D. (2013). Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in Arabidopsis. The Plant Cell. 25 (5): 1523-1540; DOI: 10.1105/tpc.112.108613.

Klopffleisch, K., Phan, N., Augustin, K., Bayne, R. S., Booker, K. S., Botella, J. R., . . . Jones, A. M. (2011). Arabidopsis G-protein interactome reveals connections to cell wall carbohydrates and morphogenesis. Molecular Systems Biology. 7: 532. https://doi.org/10.1038/msb.2011.66.

Kover, P. X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I. M., Purugganan, M. D., Durrant, C., & Mott, R. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. PLoS genetics. 5(7). e1000551. https://doi.org/10.1371/journal.pgen.1000551.

Koch, K. E. (1996). Carbohydrate modulated gene expression in plants. Annual Review of Plant Physiology and Plant Molecular Biology. 47: 509-540. https://doi.org/10.1146/annurev.arplant.47.1.509.

Kolde, R. (2015). pheatmap: Pretty Heatmaps. R package version 1.0.8.

Kondorosi, E., Roudier, F., & Gendreau, E. (2000). Plant cell-size control: Growing by ploidy? Current Opinion in Plant Biology. 3: 488-492. https://doi.org/10.1016/S1369-5266(00)00118-7.

Kover, P. X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I. M., Purugganan, M. D., Durrant, C., & Mott, R. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. PLoS Genetics. 5(7). e1000551. https://doi.org/10.1371/journal.pgen.1000551.

Krannitz, P. G., Aarssen, L. W., & Dow, J. M. (1991). The effect of genetically based differences in seed size on seedling survival in *Arabidopsis thaliana* (Brassicaceae). American Journal of Botany. 78: 446-450. https://doi.org/10.1002/j.1537-2197.1991.tb15207.x.

Krizek, B. A. (2009). Making bigger plants: key regulators of final organ size. Current Opinion in Plant Biology. 12 (1): 17-22. doi: 10.1016/j.pbi.2008.09.006.

Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. Cold Spring Harbor protocols. (11): 951–969. https://doi.org/10.1101/pdb.top084970

Laux , T., Mayer, K. F. X., Berger, J., & Jürgens, G. (1996). The *WUSCHEL* gene is required for shoot and floral meristem integrity in Arabidopsis. Development. 122: 87-96. https://dev.biologists.org/content/122/1/87.long.

Le, B. H., Cheng, C., Bu, A. Q., Wagmaister, J. A., Henry, K. F., Pelletier, J., & Goldberg, R. B. (2010). Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proceedings of the National Academy of Sciences of the United States of America. 107 (18): 8063–8070. http://doi.org/10.1073/pnas.1003530107.

Li, Y., Zheng, L., Corke, F., Smith, C., & Bevan, M. W. (2008). Control of final seed and organsize by the *DA1* gene family in *Arabidopsis thaliana*. Genes and Development. 22: 1331-1336. doi: 10.1101/gad.463608.

Lippman, Z., Gendrel, A., Black, M., & *et al*. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature. 430: 471-476. https://doi.org/10.1038/nature02651.

Locascio, A., Roig-Villanova, I., Bernardi, J., & Varotto, S. (2014). Current perspectives on the hormonal control of seed development in Arabidopsis and maize: a focus on auxin. Frontiers in Plant Science. 5: 412. DOI: https://doi.org/10.3389/fpls.2014.00412.

Long, J. A. (2006). Plant development: new models and approaches bring progress. Development. 133: 4609-4612; doi: 10.1242/dev.02676.

Lotan, T., Ohto, M., Yee, K. M., West, M. A., Lo, R., Kwong, R., . . . Harada, J. J. (1998). Arabidopsis *LEAFY COTYLEDON1* is sufficient to induce embryo development in vegetative cells. Cell. 93: 1195-1205. DOI:https://doi.org/10.1016/S0092-8674(00)81463-4.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 15 (12): 550. doi: 10.1186/s13059-014-0550-8.

Lu, Y., Gehan, J. P., & Sharkey, T. D. (2005). Daylength and circadian effects on starch degradation and maltose metabolism. Plant Physiology. 138 (4): 2280-2291. DOI: 10.1104/pp.105.061903.

Luo, M., Taylor, J. M., Spriggs , A., Zhang, H., Wu, X., & Russell, S. (2011). A Genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. Plos Genetics. 7 (6): e1002125. https://doi.org/10.1371/journal.pgen.1002125.

Mayer, K., Schüller, C., Wambutt, R., & *et al*. (1999). Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. Nature. 402: 769-777. https://doi.org/10.1038/47134.

Mayhew, M. B., Iversen, E. S., & Hartemink, A. J. (2017). Characterization of dependencies between growth and division in budding yeast. Journal of the Royal Society Interface. 14: 20160993. https://doi.org/10.1098/rsif.2016.0993.

Menges, M., & Murray, J. A. H. (2002). Synchronous Arabidopsis suspension cultures for analysis of cell-cycle gene activity. The Plant Journal. 30: 203-212. https://doi.org/10.1046/j.1365-313X.2002.01274.x.

Menges, M., Samland, A. K., Planchais, S., & Murray, J. A. H. (2006). The D-Type cyclin *CYCD3;1* is limiting for the G1-to-S-phase transition in Arabidopsis. The Plant Cell. 18 (4): 893-906. DOI: 10.1105/tpc.105.039636.

Meyerowitz, E. M. (1997). Genetic control of cell division patterns in developing plants. Cell. 88: 299-308.

Michaels, S. D., Himelblau, E., Kim, S. Y., Schomburg, F. M., & Amasino, R. M. (2005). Integration of flowering signals in winter-annual Arabidopsis. Plant Physiology. 137 (1): 149-156. doi: 10.1104/pp.104.052811.

Miles, C., & Wayne, M. (2008). Quantitative trait locus (QTL) analysis. Nature Education. 1 (1): 208.

Mizukami, Y., & Fischer, R. L. (2000). Plant organ size control: *AINTEGUMENTA* regulates growth and cell numbers during organogenesis. Proceedings of the National Academy of Sciences. 97 (2): 942-947; DOI: 10.1073/pnas.97.2.942.

Mizukami, Y. (2001). A matter of size: Developmental control of organ size in plants. Current Opinion in Plant Biology. 4: 533-539.

Moreno, C., Seal, C. E., & Papenbrock, J. (2018). Seed priming improves germination in saline conditions for *Chenopodium quinoa* and *Amaranthus caudatus*. Journal of Agronomy and Crop Science. 204: 40-48. https://doi.org/10.1111/jac.12242.

Müller, B., & Sheen, J. (2008). Cytokinin and auxin interplay in root stem-cell specification during early embryogenesis. Nature. 453: 1094-97. doi: 10.1038/nature06943.

Nakajima, K., & Benfey, P. N. (2002). Signaling in and out. The Plant Cell. 14: S265-S276. DOI: 10.1105/tpc.010471.

Neufeld, T. P., de la Cruz, A. F. A., Johnston, L. A., & Edgar, B. A. (1998). Coordination of growth and cell division in the Drosophila wing. Cell. 93: 1183-1193. https://doi.org/10.1016/S0092-8674(00)81462-2.

Nodine, M. D., & Bartel, D. P. (2010). MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes and Development. 24 (23): 2678-92. doi: 10.1101/gad.1986710.

Nodine, M. D., & Bartel, D. P. (2012). Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. Nature. 482. 94-97. https://doi.org/10.1038/nature10756.

Nowack, M. K., Ungru, A., Bjerkan, K. N., Grini, P. E., & Schnittger, A. (2010). Reproductive cross-talk: seed development in flowering plants. Biochemical Society Transactions. 38 (2): 604-612. doi: https://doi.org/10.1042/BST0380604.

Ohto, M., Onai, K., Furukawa, Y., Aoki, E., Araki, T., & Nakamura, K. (2001). Effects of sugar on vegetative development and floral transition in Arabidopsis. Plant physiology. 127 (1): 252-261. https://doi.org/10.1104/pp.127.1.252.

Okushima, Y., Mitina, I., Quach, H. L., & Theologis, A. (2005). AUXIN RESPONSE FACTOR 2 (ARF2): a pleiotropic developmental regulator. The Plant Journal. 43: 29-46. https://doi.org/10.1111/j.1365-313X.2005.02426.x.

Olas, J. J., Van Dingenen, J., Abel, C., Dzialo, M. A., Feil, R., Krapp, A., . . . Wahl, V. (2019). Nitrate acts at the *Arabidopsis thaliana* shoot apical meristem to regulate flowering time. New Phytologist. 223: 814-827. https://doi.org/10.1111/nph.15812.

Orzechowska, M., Gurdek, S., Siwinska, D., & Piekarska-Stachowiak , A. (2016). Cytogenetic characterization of the *Arabidopsis thaliana* natural tetraploid ecotype Warschau stability during *in vitro* regeneration. Plant Cell, Tissue and Organ Culture. 126: 553-560. https://doi.org/10.1007/s11240-016-1006-5.

Osuna, D., Usadel, B., Morcuende, R., Gibon, Y., Bläsing, O., Höhne, M., . . . Stitt, M. (2007). Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings. The Plant Journal. 49: 463-491. https://doi.org/10.1111/j.1365-313X.2006.02979.x.

Otto, S. P. (2017). The evolutionary consequences of polyploidy. Cell. 131: 452–462. DOI:https://doi.org/10.1016/j.cell.2007.10.022.

Palovaara, J., de Zeeuw, T., & Weijers, D. (2016). Tissue and organ initiation in the plant embryo: A first time for everything. Annual Review of Cell and Developmental Biology. 32: 47-75. https://doi.org/10.1146/annurev-cellbio-111315-124929.

Paul, M., Pellny, T., & Goddijn, O. (2001). Enhancing photosynthesis with sugar signals. Trends in Plant Science. 6: 197-200. DOI:https://doi.org/10.1016/S1360-1385(01)01920-3.

Pérez-Pérez, J. M., Serrano-Cartagena, J., & Micol, J. L. (2002). Genetic analysis of natural variations in the architecture of *Arabidopsis thaliana* vegetative leaves. Genetics. 162 (2): 893-915. https://www.genetics.org/content/162/2/893.

Pien, S., Wyrzykowska, J., & Fleming, A. J. (2001). Novel marker genes for early leaf development indicate spatial regulation of carbohydrate metabolism within the apical meristem. The Plant Journal. 25: 663-674. https://doi.org/10.1046/j.1365-313x.2001.01002.x.

Planchais, S., Samland, A. K., & Murray, J. A. H. (2004). Differential stability of Arabidopsis D-type cyclins: CYCD3;1 is a highly unstable protein degraded by a proteasome-dependent mechanism. The Plant Journal. 38 (4): 616-25. doi: 10.1111/j.0960-7412.2004.02071.x. PMID: 15125768.

Price, J., Laxmi, A., Martil, S. K., & Jang, J. C. (2004). Global transcription profiling reveals multiple sugar signal transduction mechanisms in Arabidopsis. The Plant Cell. 16 (8): 2128-2150; DOI: 10.1105/tpc.104.022616.

Prigge, M. J., Otsuga, D., Alonso, J. M., Ecker, J. R., Drews, G. N., & Clark, S. E. (2005). Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in Arabidopsis development. Plant Cell. 17: 61-76. DOI: https://doi.org/10.1105/tpc.104.026161.

Qi, P., Pendergast, T.H., Johnson, A. *et al*. (2021). Quantitative trait locus mapping combined with variant and transcriptome analyses identifies a cluster of gene candidates underlying the variation in leaf wax between upland and lowland switchgrass ecotypes. Theoretical and Applied Genetics. 134, 1957–1975 (2021). https://doi.org/10.1007/s00122-021-03798-y

Qin, Z., Zhang, X., Feng, G., & Hu, Y. (2014). The Arabidopsis *ORGAN SIZE RELATED 2* is involved in regulation of cell expansion during organ growth. BMC Plant Biology: 14: 349. https://doi.org/10.1186/s12870-014-0349-5.

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical. Vienna, Austria. URL https://www.R-project.org/.

Rangwala, S. H., & Richards, E. J. (2010). The structure, organization and radiation of *Sadhu* non-long terminal repeat retroelements in Arabidopsis species. Mobile DNA: 1: 10. https://doi.org/10.1186/1759-8753-1-10.

Ranwala, A. P., Suematsu, C., & Masuda, H. (1992). The role of β-Galactosidases in the modification of cell wall components during muskmelon fruit ripening. Plant Physiology. 100 (3): 1318-1325; DOI: 10.1104/pp.100.3.1318.

Raz, V., Bergervoet, J., & Koornneef, M. (2001). Sequential steps for developmental arrest in Arabidopsis seeds. Development. 128: 243-252. https://dev.biologists.org/content/128/2/243.

Reinhardt, D., Mandel , T., & Kuhlemeier, C. (2000). Auxin regulates the initiation and radial position of plant lateral organs. The Plant Cell. 12 (4): 507-518. DOI: 10.1105/tpc.12.4.507.

Riou-Khamlichi, C., Huntley, R., Jacqmard, A., & Murray, J. A. H. (1999). Cytokinin activation of Arabidopsis cell division through a D-type cyclin. Science. 283: 1541-1544. DOI: 10.1126/science.283.5407.1541.

Robert, H. S., Grones, P., Stepanova, A. N., Robles, L. M., Lokerse, A. S., Alonso, J. M., . . . Friml, J. (2013). Local auxin sources orient the apical-basal axis in Arabidopsis embryos. Current Biology. 23: 2506–12. DOI: 10.1016/j.cub.2013.09.039.

Robinson, D. O., Coate, J. E., Singh, A., Hong, L., Bush, M., Doyle, J. J., . . . Roeder, A. H. K. (2018). Ploidy and size at multiple scales in the Arabidopsis sepal. The Plant Cell. 30 (10): 2308-2329. DOI: 10.1105/tpc.18.00344.

Roeder, A. H., Chickarmane, V., Chickarmane, A., Obara, B., Manjunath, B. S., & Meyerowitz, E. M. (2010). Variability in the control of cell division underlies sepal epidermal patterning in *Arabidopsis thaliana*. Plos Biology. 8: e1000367. https://doi.org/10.1371/journal.pbio.1000367.

Roff, D. A. (2007). A centennial celebration for quantitative genetics. Evolution. 61: 1017-1032.

Roitsch, T. (1999). Source-sink regulation by sugar and stress. Current Opinion in Plant Biology. 2: 198-206. https://doi.org/10.1016/S1369-5266(99)80036-3.

Rolland, F., Moore, B., & Sheen, J. (2002). Sugar sensing and signaling in plants. The Plant Cell. 14 (suppl 1): S185-S205; DOI: 10.1105/tpc.010455.

Rook, F., Corke, F., Card, R., Munz, G., Smith, C., & Bevan, M. W. (2001). Impaired sucrose-induction mutants reveal the modulation of sugar-induced starch biosynthetic gene expression by abscisic acid signalling. The Plant Journal. 26: 421-433. https://doi.org/10.1046/j.1365-313X.2001.2641043.x.

Ross, G. S., Wegrzyn, T., MacRae, E. A., & Redgwell, R. J. (1994). Apple β-Galactosidase (activity against cell wall polysaccharides and characterization of a related cDNA clone). Plant Physiology. 106 (2): 521-528. DOI: 10.1104/pp.106.2.521.

Roy Choudhury, S., Riesselman, A. J., & Pandey, S. (2014). Constitutive or seed-specific overexpression of Arabidopsis G-protein γ subunit 3 (AGG3) results in increased seed and oil production and improved stress tolerance in *Camelina sativa*. Plant Biotechnology Journal. 12 (1): 49-59. doi: 10.1111/pbi.12115. Epub 2013 Sep 17. PMID: 24102738.

Sampedro, J., Gianzo, C., Iglesias, N., Guitián, E., Revilla, G., & Zarra, I. (2012). *AtBGAL10* is the main xyloglucan β-Galactosidase in Arabidopsis, and its absence results in unusual xyloglucan subunits and growth defects. Plant Physiology. 158 (3): 1146-1157. DOI: 10.1104/pp.111.192195.

Santos-Mendoza, M., Dubreucq, B., Baud, S., Parcy, F., Caboche, M., & Lepiniec, L. (2008). Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. The Plant Journal. 54: 608-620. https://doi.org/10.1111/j.1365-313X.2008.03461.x.

Savadi, S. (2018). Molecular regulation of seed development and strategies for engineering seed size in crop plants. Plant Growth Regulation. 84: 401-422. https://doi.org/10.1007/s10725-017-0355-3.

Schmid, M., Davison, T., Henz, S., & *et al*. (2005). A gene expression map of *Arabidopsis thaliana* development. Nature Genetics. 37: 501–506. https://doi.org/10.1038/ng1543.

Schmuths, H., Meister, A., Horres, R., & Bachmann, K. (2004). Genome size variation among accessions of *Arabidopsis thaliana*. Annals of Botany. 93 (3): 317-321. https://doi.org/10.1093/aob/mch037.

Schneider, C., Rasband, W., & Eliceiri, K. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature Methods. 9: 671–675. https://doi.org/10.1038/nmeth.2089.

Schoof, H., Lenhard, M., Haecker, A., Mayer, K. F. X., Jürgens, G., & Laux, T. (2000). The stem cell population of Arabidopsis shoot meristems is maintained by a regulatory loop between the *CLAVATA* and *WUSCHEL* genes. Cell. 100 (6): P635-644. DOI:https://doi.org/10.1016/S0092-8674(00)80700-X.

Scofield, S., Dewitte, W., & Murray, J. A. H. (2014). *STM* sustains stem cell function in the Arabidopsis shoot apical meristem and controls *KNOX* gene expression independently of the transcriptional repressor AS1. Plant Signaling and Behavior. 9: e28934. http://doi.org/10.4161/psb.28934.

Scott, R. J., Spielman, M., Bailey, J., & Dickinson, H. G. (1998). Parent-of-origin effects on seed development in *Arabidopsis thaliana*. Development. 125 (17): 3329-41. PMID: 9693137.

Seefried, W. F., Willmann, M. R., Clausen, R. L., & Jenik, P. D. (2014). Global regulation of embryonic patterning in Arabidopsis by microRNAs. Plant Physiology. 165 (2): 670-687. DOI: 10.1104/pp.114.240846.

Seo , M., Hanada, A., Kuwahara, A., Endo, A., Okamoto, M., Yamauchi , Y., . . . Nambara , E. (2006). Regulation of hormone metabolism in Arabidopsis seeds: phytochrome regulation of abscisic acid metabolism and abscisic acid regulation of gibberellin metabolism. The Plant Journal. 48 (3): 354-66. doi: 10.1111/j.1365-313X.2006.02881.x.

Sharma, V. K., & Fletcher, J. C. (2002). Maintenance of shoot and floral meristem cell proliferation and fate. Plant Physiology. 129 (1): 31-39. DOI: 10.1104/pp.010987.

Sheen, J., Zhou, L., & Jang, J-C. (1999). Sugars as signaling molecules. Current Opinion in Plant Biology. 2: 410-418. https://doi.org/10.1016/S1369-5266(99)00014-X.

Simon, M., Loudet, O., Durand, S., Bérard, A., Brunel, D., Sennesal, F. X., Durand-Tardif, M., Pelletier, G., & Camilleri, C. (2008). Quantitative trait loci mapping in five new large recombinant inbred line populations of *Arabidopsis thaliana* genotyped with consensus single-nucleotide polymorphism markers. Genetics. 178(4), 2253–2264. https://doi.org/10.1534/genetics.107.083899.

Smeekens, S. (2000). Sugar-induced signal transduction in plants. Annual Review of Plant Physiology and Plant Molecular Biology. 51: 49-81. https://doi.org/10.1146/annurev.arplant.51.1.49.

Smith, A. M., & Stitt, M. (2007). Coordination of carbon supply and plant growth. Plant, Cell and Environment. 30: 1126-1149. https://doi.org/10.1111/j.1365-3040.2007.01708.x.

Sonoda, Y., Sako, K., Maki, Y., Yamazaki , N., Yamamoto, H., Ikeda, A., & Yamaguchi, J. (2009). Regulation of leaf organ size by the Arabidopsis RPT2a 19S proteasome subunit. Plant Journal. 60 (1): 68–78. doi: 10.1111/j.1365-313X.2009.03932.x.

Sørensen, M. B., Mayer, U., Lukowitz, W., Robert, H., Chambrier, P., Jürgens, G., . . . Berger, F. (2002). Cellularisation in the endosperm of *Arabidopsis thaliana* is coupled to mitosis

and shares multiple components with cytokinesis. Development. 129: 5567-5576. doi: 10.1242/dev.00152.

Spencer, M. W. W., Casson, S. A., & Lindsey, K. (2007). Transcriptional profiling of the Arabidopsis Embryo. Plant Physiology. 143 (2): 924-940. DOI: 10.1104/pp.106.087668.

Stitt, M., & Krapp, A. (1999). The interaction between elevated carbon dioxide and nitrogen nutrition: The physiological and molecular background. Plant Cell and Environment. 22: 583-621. https://doi.org/10.1046/j.1365-3040.1999.00386.x.

Sulpice, R., Nikoloski, Z., Tschoep, H., Antonio, C., Kleessen, S., Larhlimi, A., . . . Stitt, M. (2013). Impact of the carbon and nitrogen supply on relationships and connectivity between metabolism and biomass in a broad panel of Arabidopsis accessions. Plant Physiology. 162 (1): 347-363. DOI: 10.1104/pp.112.210104.

Sulpice, R., Pyl, E-T., Ishihara, H., Trenkamp, S., Steinfath, M., Witucka-Wall, H., . . . Stitt, M. (2009). Starch as a major integrator in the regulation of plant growth. Proceedings of the National Academy of Sciences. 106 (25): 10348-10353. DOI: 10.1073/pnas.0903478106.

Sun, X., Shantharaj, D., Kang, X., & Ni, M. (2010). Transcriptional and hormonal signaling control of Arabidopsis seed development. Current Opinion in Plant Biology. 13 (5): 611-20. DOI: https://doi.org/10.1016/j.pbi.2010.08.009.

Sureshkumar, S., Todesco, M., Schneeberger, K., Harilal, R., Balasubramanian, S., & Weigel, D. (2009). A genetic defect caused by a triplet repeat expansion in *Arabidopsis thaliana*. Science. 323: 1060–1063. doi: 10.1126/science.1164014.

Tabib, A., Vishwanathan, S., Seleznev , A., Mc Keon, P. C., Downing, T., Dent , C., . . . Balasubramanian, S. (2016). A polynucleotide repeat expansion causing temperature-sensitivity persists in wild Irish accessions of *Arabidopsis thaliana*. Frontiers in Plant Science. 7:1311. DOI=10.3389/fpls.2016.01311.

Telfer, A., Bollman, K. M., & Poething, R. S. (1997). Phase change and the regulation of trichome distribution in *Arabidopsis thaliana*. Development. 124: 645-654. https://dev.biologists.org/content/124/3/645.long.

ten Hove, C., Lu, K., & Weijers, D. (2015). Building a plant: cell fate specification in the early Arabidopsis embryo. Development. 142: 420-430. doi: 10.1242/dev.111500.

Tessmer, O. L., Jiao, Y., Cruz, J. A., Kramer, D. M., & Chen, J. (2013). Functional approach to high-throughput plant growth analysis. BMC Systems Biology. 7 (Suppl 6): S17. https://doi.org/10.1186/1752-0509-7-S6-S17.

The 1001 Genomes Consortium. (2016). 1,135 Genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. Cell. 166 (2): 481-491. DOI: https://doi.org/10.1016/j.cell.2016.05.063.

Thieme, C., Rojas-Triana, M., Stecyk, E., Schudoma, C., Zhang, W., Yang, L., . . . Kragler, F. (2015). Endogenous Arabidopsis messenger RNAs transported to distant tissues. Nature Plants. 1: 15025. https://doi.org/10.1038/nplants.2015.25.

Tsukaya, H., & Beemster, G. T. S. (2006). Genetics, cell cycle and cell expansion in organogenesis in plants. Journal of Plant Research. 119: 1-4. https://doi.org/10.1007/s10265-005-0254-y.

Vlad , D., Rappaport, F., Simon, M., & Loudet , O. (2010). Gene transposition causing natural variation for growth in *Arabidopsis thaliana*. Plos Genetics. 6 (5): e1000945. https://doi.org/10.1371/journal.pgen.1000945.

Wang, J., Abbas, M., Wen, Y., Niu, D., Wang, L., Sun, Y., & Li, Y. (2018). Selection and validation of reference genes for quantitative gene expression analyses in black locust (*Robinia pseudoacacia* L.) using real-time quantitative PCR. Plos One. 13 (3): e0193076. https://doi.org/10.1371/journal.pone.0193076.

Waters, A. J., Bilinski, P., Eichten, S. R., Vaughn, M. W., Ross-Ibarra, J., Gehring, M., & Springer, N. M. (2013). Comprehensive analysis of imprinting in maize reveals allelic variation for imprinting and limited conservation with other species. Proceedings of the National Academy of Sciences. 110 (48): 19639-19644. DOI: 10.1073/pnas.1309182110.

Weigel, D., & Jürgens, G. (2002). Stem cells that make stems. Nature. 415: 751-754. DOI: https://doi.org/10.1038/415751a.

Werner, J. D., Borevitz, J. O., Uhlenhaut, N. H., & Ecker, J. R. (2005). *FRIGIDA*-independent variation in flowering time of natural *Arabidopsis thaliana* accessions. Genetics. 170 (3): 1197-1207. https://doi.org/10.1534/genetics.104.036533.

West, M. A. L., & Harada, J. J. (1993). Embryogenesis in higher plants: An overview. The Plant Cell. 5: 1361-1369. DOI: https://doi.org/10.1105/tpc.5.10.1361.

Westoby, M., Jurado, E., & Leishman, M. (1992). Comparative evolutionary ecology of seed size. Trends in Ecology and Evolution. 7: 368–372. https://doi.org/10.1016/0169-5347(92)90006-W.

Willis, L., Refahi, Y., Wightman, R., Landrein, B., Teles, J., Huang, K. C., . . . Jönsson, H. (2016). Cell size and growth regulation in Arabidopsis. Proceedings of the National Academy of Sciences: 113 (51): E8238-E8246; DOI: 10.1073/pnas.1616768113.

Xin, Q., Shen, Y., Li, X., Lu, W., Wang, X., Han, X., . . . Cheng, Z. (2016). *MS5* mediates early meiotic progression and its natural variants may have applications for hybrid production in *Brassica napus*. The Plant Cell. 28 (6): 1263-1278; DOI: 10.1105/tpc.15.01018.

Yoo, S. K., Chung, K. S., Kim, J., Lee, J. H., Hong, S. M., Yoo, S. J., . . . Ahn, J. H. (2005). *CONSTANS* activates *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* through *FLOWERING LOCUS T* to promote flowering in Arabidopsis. Plant Physiology. 139 (2): 770-778. DOI: 10.1104/pp.105.066928.

Yruela, I. (2015). Plant development regulation: Overview and perspectives. Journal of Plant Physiology. 182: 62–78.

Yu, Z., Haberer, G., Matthes, M., Rattei, T., Mayer, K. F. X., Gierl, A., & Torres-Ruiz, R-A. (2010). Impact of natural genetic variation on the transcriptome of autotetraploid *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the Unitated states of America. 107: 17809-17814. doi: 10.1073/pnas.1000852107.

I. Supplementary tables

Supplementary Table S1. Primer sequences used in this study.

Gene ID	Gene name	Primer sequence 5´ -----> 3´	
Primers used for qRT-PCR			
AT5G62690	***TUBULIN2* (*TUB2*)**	*FW	GAGCCTTACAACGCTACTCTGTCTGTC
		*RV	ACACCAGACATAGTAGCAGAAATCAAG
AT2G28740	***HISTONE4* (*HIS4*)**	*FW	CCTTTAGAAAATGTCAGGTCG
		*RV	GTTTAACACCACCTCTACGAGC
AT4G37490	***CYCLINB1;1* (*CYCB1;1*)**	*FW	TCGGTTCTTGTCGGTTAAGCC
		*RV	CCTGTGGTGGCCAAATTTCTT
AT2G27250	***CLAVATA3* (*CLV3*)**	*FW	CAAGGACTTTCCAACCGCAAG
		*RV	GGTTCACATGATGGTGCAACG
AT1G62360	***SHOOT MERISTEMLESS* (*STM*)**	*FW	TCATGGCTCATCCTCACTACC
		*RV	CCTGTTGGTCCCATAGATGC
AT2G17950	***WUSCHEL* (*WUS*)**	*FW	AACCAAGACCATCATCTCTATCATC
		*RV	CCATCCTCCACCTACGTTGT
AT2G03965	**Not reported**	FW	GAGTGCTCTACATCCCCTGC
		RV	ACCTAACTCTCTCACTGGGC
AT5G15360	**Not reported**	FW	CCCAAGCCAAAGGATGGAGT
		RV	ACTAAAGGCGCAAGCGATCT
AT4G08874	**Not reported**	FW	CTTCCATCGTGGAGAGCACC
		RV	CCCGCAGCGACTAAGAGATT
AT1G71920	**Not reported**	FW	ACCGCATCGCTGAAGTTGTA
		RV	CAACAGATCGTCCTCGCTGA
AT2G43960	**Not reported**	FW	ATCGACGAACTGTCTGCGTT
		RV	TATCCCGTAGAGCCAACCGA
AT2G04378	**Not reported**	FW	TGCTTCTTAGCATCGCCAGA
		RV	CAAGGTGAATATCCGCTCGT
AT4G07825	**Not reported**	FW	CACCATGTTTTCTGACGCCG
		RV	GGTTCGCCGGAAAACCTCTA
AT3G44430	**Not reported**	FW	GCGTCGGGACTTAGCTCTTC
		RV	CCGTTGAAAACCGGTGAAGG
AT5G05060	**Not reported**	FW	CGAGCCTCATGAGACCAACA
		RV	TGTGATTGCAGCAGATGGGT
AT4G11830	**Not reported**	FW	GGACCAAGAGAACCATGGCA
		RV	GGTTTTGAAGCCATCCAGCG
Primers used for cDNA template quality assessment			
AT1G13440	***Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)* 5´**	*FW	TCTCGATCTCAATTTCGCAAAA
		*RV	CGAAACCGTTGATTCCGATTC
AT1G13440	***Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)* 3´**	*FW	TTGGTGACAACAGGTCAAGCA
		*RV	AAACTTGTCGCTCAATGCAATC
Primers used for cloning			
AT4G08874	**Not reported**	FW	ATGATCGATTACTCTCGGAAAACC
		RV	TCAATCAAACCCAAAGAGGGGTAC
AT4G07825	**Not reported**	FW	ATGTATTTTGAAAATCGGTCTAC
		RV	TCAACCTGGGCTGCTGTGATTATG
AT3G44430	**Not reported**	FW	ATGTTTTCTGACGCCGGCGG
		RV	TCAGAATTGGGCCGAAGTAG
AT4G30720	***PIGMENT DEFECTIVE 327* (*PDE327*)**	FW	ATGTCTCTCTCGCAAACGAATTTCC
		RV	TCAGTACTTTACAAGTCCAGCACC
AT5G20635	***ARABIDOPSIS G PROTEIN GAMMA SUBUNIT 3* (*AGG3*)**	FW	ATGTCTGCTCCTTCTGGCGG
		RV	TTAGAAAGCTAAACAACAAGG
AT2G35230	***HAIKU1* (*IKU1*)**	FW	ATGGATAGGCCTAGACAAAATG
		RV	CTAGTAATCATTCCATCTTGGAC
AT1G19270	***LARGE IN CHINESE* (*DA1*)**	FW	ATGGAGGTGAATGATGGTG
		RV	TTAAACCGGGAATCTACCG

Sequences with (*) correspond to primers provided by Dr. Justyna Olas from the in-house stock.

Supplementary Table S2. List of constructs generated for RNA *in situ* hybridization.

Gene ID	Gene name	Construct	Source
AT4G08874	**Not reported**	pGEMTeasy-874	Generated by Eike Kamann
AT4G07825	**Not reported**	pGEMTeasy-825	Generated by Eike Kamann
AT3G44430	**Not reported**	pGEMTeasy-430	Generated by Eike Kamann
AT5G20830	***SUCROSE SYNTHASE 1* (*SUS1*)**	pGEMTeasy-*SUS1*	Generated by Eike Kamann
AT4G02280	***SUCROSE SYNTHASE 3* (*SUS3*)**	pGEMTeasy-*SUS3*	Generated by Eike Kamann
AT2G35230	***HAIKU1* (*IKU1*)**	pGEMTeasy-*IKU1*	Generated by Eike / Catalina
AT1G19270	***LARGE IN CHINESE* (*DA1*)**	pGEMTeasy-*DA1*	Generated by Catalina Moreno
AT5G20635	***ARABIDOPSIS G PROTEIN GAMMA SUBUNIT 3* (*AGG3*)**	pGEMTeasy-*AGG3*	Generated by Catalina Moreno
AT4G30720	***PIGMENT DEFECTIVE 327* (*PDE327*)**	pGEMTeasy-*PDE327*	Generated by Catalina Moreno

Supplementary Table S3. Probes used for RNA *in situ* hybridization.

Gene ID	Gene name	Probes	Source
AT4G08874	**Not reported**	SP6	Synthesized by Catalina Moreno
		T7	
AT4G07825	**Not reported**	SP6	Synthesized by Catalina Moreno
		T7	
AT3G44430	**Not reported**	SP6	Synthesized by Catalina Moreno
		T7	
AT4G30720	***PIGMENT DEFECTIVE 327* (*PDE327*)**	SP6	Synthesized by Catalina Moreno
		T7	
AT5G20635	***ARABIDOPSIS G PROTEIN GAMMA SUBUNIT 3* (*AGG3*)**	SP6	Synthesized by Catalina Moreno
		T7	
AT2G35230	***HAIKU1* (*IKU1*)**	SP6	Synthesized by Catalina Moreno
		T7	
AT1G19270	***LARGE IN CHINESE* (*DA1*)**	SP6	Synthesized by Catalina Moreno
		T7	
AT5G20830	***SUCROSE SYNTHASE 1* (*SUS1*)**	SP6	Synthesized by Catalina Moreno
		T7	
AT4G02280	***SUCROSE SYNTHASE 3* (*SUS3*)**	SP6	Synthesized by Catalina Moreno
		T7	
AT2G28740	***HISTONE4* (*HIS4*)**	SP6	Provided by Dr. Justyna Olas
AT4G37490	***CYCLINB1;1* (*CYCB1;1*)**	SP6	Provided by Dr. Justyna Olas
AT2G27250	***CLAVATA3* (*CLV3*)**	SP6	Provided by Dr. Justyna Olas
AT1G62360	***SHOOT MERISTEMLESS* (*STM*)**	SP6	Provided by Dr. Justyna Olas
AT2G17950	***WUSCHEL* (*WUS*)**	SP6	Provided by Dr. Justyna Olas
AT4G36920	***APETALA2* (*AP2*)**	SP6	Provided by Dr. Justyna Olas

Supplementary Table S4. Average rosette area over time. Eleven *Arabidopsis thaliana* accessions grown in long day photoperiod were analyzed from 4 to 14 days after germination (DAG). *n* = 10 plants. ± SD. Statistical significance was tested using ANOVA (***0.001; **0.01; *0.05) and the means were compared using Tukey–HSD test. Different letters indicate significant differences at α 0.05.

Time point	Accession	Rosette area (cm^2) Average ± SD	Time point	Accession	Rosette area (cm^2) Average ± SD
4 DAG ***	Ws-2	0.072 ± 0.011 cde	**10 DAG** ***	Ws-2	0.578 ± 0.056 c
	Sei-0	0.117 ± 0.021 a		Sei-0	0.892 ± 0.285 a
	L*er*-1	0.072 ± 0.012 cde		L*er*-1	0.481 ± 0.066 cd
	Lip-0	0.099 ± 0.017 ab		Lip-0	0.973 ± 0.210 a
	Col-0	0.075 ± 0.006 cd		Col-0	0.650 ± 0.084 bc
	Ts-1	0.045 ± 0.010 f		Ts-1	0.673 ± 0.151 bc
	Cen-0	0.052 ± 0.008 ef		Cen-0	0.823 ± 0.114 ab
	Alst-1	0.022 ± 0.005 g		Alst-1	0.353 ± 0.065 d
	Sap-0	0.082 ± 0.009 bc		Sap-0	0.847 ± 0.137 ab
	Ang-0	0.054 ± 0.011 def		Ang-0	0.637 ± 0.114 bc
	Bur-0	0.100 ± 0.031 ab		Bur-0	0.850 ± 0.162 ab
6 DAG ***	Ws-2	0.152 ± 0.022 cde	**12 DAG** ***	Ws-2	0.930 ± 0.146 ef
	Sei-0	0.252 ± 0.026 a		Sei-0	1.522 ± 0.484 abc
	L*er*-1	0.139 ± 0.021 ef		L*er*-1	0.750 ± 0.133 f
	Lip-0	0.222 ± 0.034 ab		Lip-0	1.777 ± 0.388 a
	Col-0	0.148 ± 0.008 de		Col-0	1.086 ± 0.162 cdef
	Ts-1	0.106 ± 0.021 f		Ts-1	1.433 ± 0.436 abcd
	Cen-0	0.140 ± 0.030 ef		Cen-0	1.445 ± 0.220 ab
	Alst-1	0.050 ± 0.006 g		Alst-1	0.790 ± 0.182 ef
	Sap-0	0.181 ± 0.026 cd		Sap-0	1.717 ± 0.221 a
	Ang-0	0.125 ± 0.020 ef		Ang-0	1.076 ± 0.264 def
	Bur-0	0.186 ± 0.044 bc		Bur-0	1.197 ± 0.283 bcde
8 DAG ***	Ws-2	0.285 ± 0.042 cd	**14 DAG** ***	Ws-2	1.421 ± 0.256 c
	Sei-0	0.574 ± 0.108 a		Sei-0	3.022 ± 0.961 a
	L*er*-1	0.242 ± 0.042 cd		L*er*-1	1.190 ± 0.182 c
	Lip-0	0.572 ± 0.089 a		Lip-0	3.278 ± 0.680 a
	Col-0	0.317 ± 0.037 bcd		Col-0	1.784 ± 0.262 bc
	Ts-1	0.256 ± 0.049 cd		Ts-1	3.064 ± 0.805 a
	Cen-0	0.332 ± 0.060 bc		Cen-0	2.508 ± 0.632 ab
	Alst-1	0.133 ± 0.025 e		Alst-1	1.497 ± 0.417 c
	Sap-0	0.399 ± 0.055 b		Sap-0	3.267 ± 0.583 a
	Ang-0	0.237 ± 0.053 d		Ang-0	1.939 ± 0.537 bc
	Bur-0	0.389 ± 0.070 b		Bur-0	1.733 ± 0.413 bc

Supplementary Table S5. Germination parameters. Final germination percentage (FGP), germination index (GI) and mean germination rate (MGT) analyzed for the *Arabidopsis thaliana* accessions Col-0, Bur-0, L*er*-1 and Ws-2. Data are the means of three replicates, 100 seeds each. ± SD. Statistical significance was tested using ANOVA (***0.001; **0.01; *0.05) and the means were compared using Tukey–HSD test. Different letters indicate significant differences at α 0.05.

Accession	FGP (%) ns Average ± SD	GI *** Average ± SD	MGT (h) *** Average ± SD
Col-0	96.3 ± 1.1 [a]	48.17 ± 0.5 [b]	45.90 ± 0.10 [a]
Bur-0	94.6 ± 2.3 [a]	45.15 ± 1.4 [b]	46.26 ± 0.05 [a]
L*er*-1	95.6 ± 3.7 [a]	54.50 ± 3.0 [b]	44.80 ± 0.40 [b]
Ws-2	97.3 ± 3.0 [a]	96.33 ± 8.1 [a]	41.10 ± 0.45 [c]

Supplementary Table S6. End point shoot biomass (dry weight based) and shoot relative growth rate (RGR, biomass based). The *Arabidopsis thaliana* accessions Col-0, Bur-0, L*er*-1 and Ws-2 were analyzed in long day (LD) and short day (SD) photoperiods. End point analyzed 3 days after floral transition in LD and 10 days in SD per accession, respectively. *n* = 10. ± SD. Statistical significance was tested using ANOVA (***0.001; **0.01; *0.05) and the means were compared using Tukey–HSD test. Different letters indicate significant differences at α 0.05.

Accession	Shoot biomass (mg) *** Average ± SD	RGR (mg mg^{-1} day^{-1}) ** Average ± SD	Shoot biomass (mg) *** Average ± SD	RGR (mg mg^{-1} day^{-1}) *** Average ± SD
	LD photoperiod		SD photoperiod	
Col-0	5.50 ± 0.51 [b]	0.39 ± 0.09 [ab]	240.92 ± 39.17 [a]	0.04 ± 0.03 [b]
Bur-0	110.24 ± 18.89 [a]	0.31 ± 0.05 [ab]	494.58 ± 79.80 [a]	0.07 ± 0.02 [b]
L*er*-1	4.05 ± 1.22 [b]	0.25 ± 0.11 [b]	75.71 ± 13.97 [b]	0.19 ± 0.01 [a]
Ws-2	2.41 ± 0.57 [b]	0.42 ± 0.10 [a]	28.57 ± 6.90 [c]	0.07 ± 0.03 [b]

Supplementary Table S7. Characterization of biological functions of differentially expressed genes from in cluster 9 and cluster 10. High confidence set of candidate genes whose expression is consistently high or low per accession across tissues and developmental stages (cluster 10 for Bur-0 with 41 genes and cluster 9 for Col-0 with 44 genes). In total, 85 candidate genes are classified as protein coding, transposable elements, novel transcribed and undetermined (including long noncoding RNA, noncoding RNA, miscellaneous RNA), based on information collected from databases as indicated in the main text and literature research.

ATG Code	Gene type classification	Name	Description	Expression Map/predicted localization	Source
Cluster 10, Bur-0 higher and Col-0 lower expression					
AT1G63540	Protein coding	Unknown	Transporter. Subfamily F2K11.10. Hydroxyproline-rich glycoprotein family protein (90% Identity to Q9SH38).	High expression in mature embryos and anthers with pollen. Expressed in several accessions.	Databases.
AT2G04655	Undetermined	Unknown	Long noncoding RNA.	Not available.	Databases.
AT3G24514	Protein coding	Unknown	Hypothetical, uncharacterized protein.	Not available.	Databases.
AT1G52610	Transposable element	Unknown	Mutator-like transposase family. Blast match to GB:AAA21566 mudrA of transposon=MuDR (MuDr-element) (*Zea mays*).	Not available.	Databases.
AT1G70010	Transposable element	Unknown	Not available.	Not available.	Databases.
AT2G03965	Protein coding	Unknown	Uncharacterized protein which has a signal sequence, a peptide usually present at the N-terminus of proteins and which is destined to be either secreted or part of membrane components.	Bur-0 specific.	Databases.
AT1G67105	Undetermined	Unknown	Other RNA. Noncoding RNA.	Not available.	Databases.
AT2G13750	Transposable element	Unknown	CACTA-like transposase family (Ptta/En/Spm). Blast match to At5g29026.1/8-244 CACTA-like transposase family (Ptta/En/Spm) (CACTA-element) (*Arabidopsis thaliana*).	Not available.	Databases.
AT4G29200	Protein coding	DEG3	Beta-galactosidase related protein. Transmembrane. Over-expressed by salt stress Similar protein (100% identity to Q5BPK5, Q9M0G1).	High expression in several accessions. Expressed in mature embryos.	Databases.
AT5G15360	Protein coding	Unknown	Transmembrane protein. Similar protein (90% identity to Q9LXE7).	High expression in several accessions. Expressed in shoot epidermal cells.	Databases.
AT1G06827	Undetermined	Unknown	Long noncoding RNA.	Not available.	Databases.

AT4G08874	Protein coding	Unknown	Transmembrane protein. Beta-galactosidase related protein. Similar protein (100% identity to Q3EA64, A0A1P8B5I4).	Bur-0 specific. Expressed in flower, fruits, leaves and siliques at early development.	Databases.
AT2G01422	Undetermined	Unknown	Other RNA. Noncoding RNA.	Not available.	Databases.
AT5G00480	Novel transcribed	Unknown	Novel transcribed region.	Not available.	Databases.
AT4G24420	Protein coding	Unknown	RNA-binding (RRM/RBD/RNP motifs) family protein (90% identity to Q9STV2).	High expression in several accessions including Bur-0. Expressed in floral bud.	Databases.
AT4G18150	Protein coding	DUF1296	Serine/Threonine-kinase, putative (DUF1296).	High expression in Bur-0. Expressed in siliques and seeds.	Databases.
AT4G03005	Novel transcribed	Unknown	Novel transcribed region.	Not available.	Databases.
AT1G71920	Protein coding	HISN6B ortholog	Histidinol-phosphate aminotransferase 2, chloroplastic. Similar protein (100% identity to A8MRI5).	High expression in Bur-0.	Databases.
AT5G24240	Protein coding	PI4K GAMMA3 MOP9.5	Phosphatidylinositol 4-kinase gamma 3. Encodes PI4Kc3, localizes to the nucleus and has autophosphorylation activity, but no lipid kinase activity. Overexpression mutants display late-flowering phenotype. Similar protein (90% identity to Q9FNF8).	High expression in several accessions. Expressed in siliques and guard cells.	Databases.
AT5G17090	Protein coding	Unknown	Cystatin/monellin superfamily protein. Similar protein (90% identity to Q9LFJ4).	High expression in several accessions. Expressed in flowers and siliques.	Databases.
AT1G02770	Protein coding	UPF0725	These sequences comprise a paralogous family of hypothetical proteins in Brassicaceae. Length heterogeneity within the family is attributable partly to a 21-residue repeat present in from zero to three tandem copies. One member of the family, protein MS5, has been shown to be important for progression of meiosis during early prophase and is essential for pairing of homologues. It is preferentially expressed in reproductive organs.	Expressed in different accessions and high expression in Bur-0. Expressed in flower, mature siliques, anthers, seeds.	Databases and Xin *et al.* (2016).
AT5G30360	Protein coding	Unknown	Hypothetical, uncharacterized protein.	Not available.	Databases.
AT5G32670	Protein coding	Unknown	START domain protein.	Not available.	Databases.
AT5G35918	Transposable element	Unknown	Gypsy-like retrotransposon family (Athila). Blast match to GB:CAA57397 Athila ORF 1 (*Arabidopsis thaliana*).	Not available.	Databases.
AT5G47280	Protein coding	ADR1-like 3; MQL5.14; MQL5_14	Putative disease resistance protein.	Not available.	Databases and Grant *et al.* (2013).

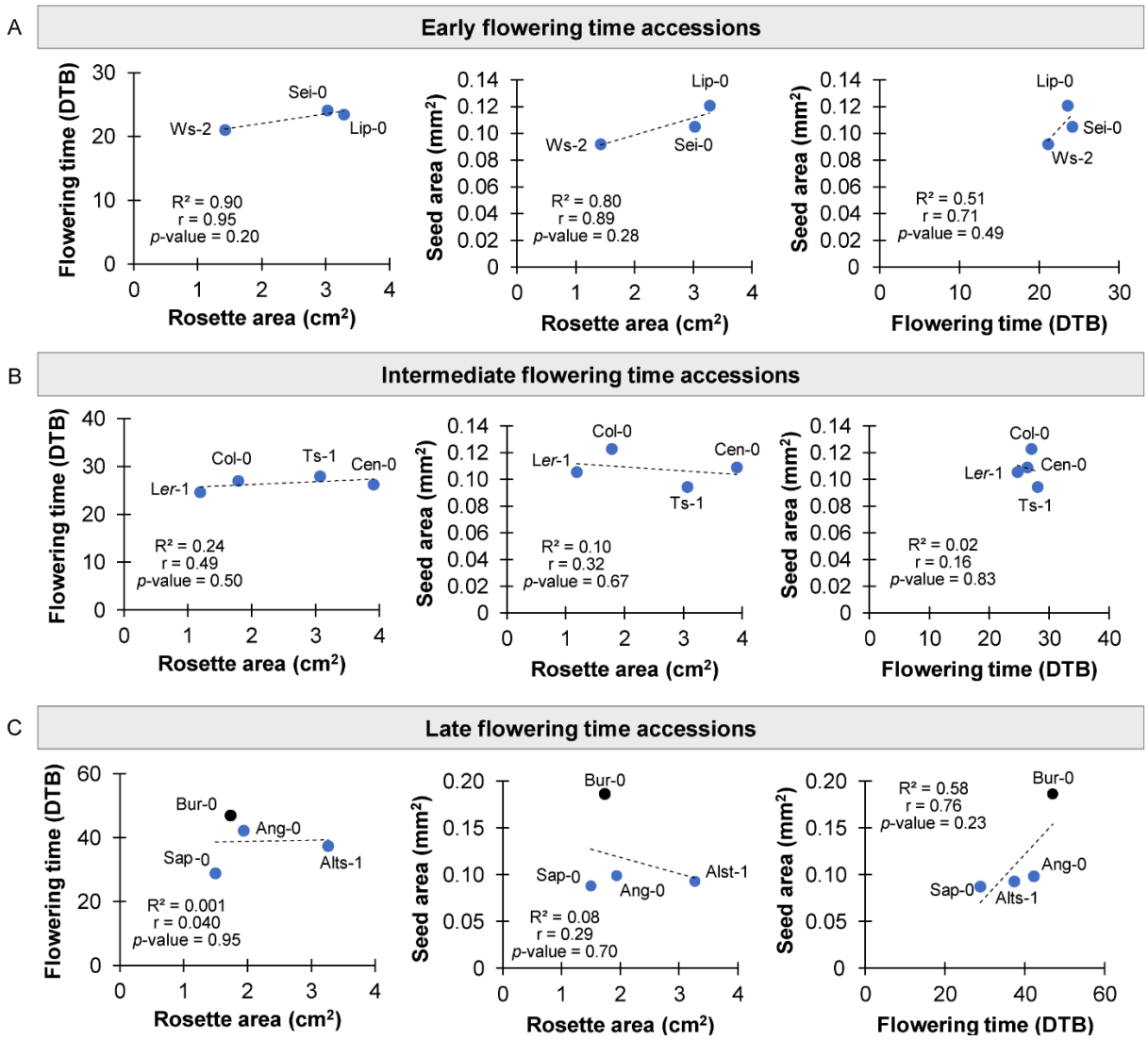
AT5G59620	Transposable element	Unknown	CACTA-like transposase family (Ptta/En/Spm). Blast match to At5g59620.1/14-257 CACTA-like transposase family (Ptta/En/Spm) (CACTA-element) (*Arabidopsis thaliana*).	Not available.	Databases.
AT1G35215	Protein coding	Unknown	Transmembrane protein.	Not available.	Databases.
AT2G03970	Transposable element	Unknown	Pseudogene, similar to SAE1-S9-protein, blast match of 33%25 identity to GP|4760708|dbj|BAA77394.1||AB012866 SAE1-S9-protein (*Brassica rapa*)	Not available.	Databases.
AT2G43960	Protein coding	SWAP	SWAP (Suppressor-of-White-APricot)/surp domain-containing protein RNA processing. Splicing factor.	Not available.	Databases.
AT3G24516	Protein coding	Unknown	Hypothetical, uncharacterized protein.	Not available.	Databases.
AT3G24517	Protein coding	Unknown	Hypothetical, uncharacterized protein.	Not available.	Databases.
AT3G24542	Protein coding	Unknown	Beta-galactosidase related protein	Not available.	Databases.
AT3G32377	Undetermined	Unknown	Miscellaneous RNA.	Not available.	Databases.
AT3G42090	Transposable element	Unknown	Not available.	Not available.	Databases.
AT4G04402	Protein coding	Unknown	Two-component phosphor-relay mediator. HISTIDINE-CONTAINING PHOSPHOTRANSFER PROTEIN 2-RELATED (PTHR28242:SF26).	Not available.	Databases and Lippman *et al*. (2004).
AT4G08093	Undetermined	Unknown	Pseudogene. Miscellaneous RNA.	Not available.	Databases.
AT4G09635	Undetermined	Unknown	Noncoding RNA.	Not available.	Databases.
AT4G15242	Undetermined	Unknown	Noncoding RNA.	Not available.	Databases.
AT4G32208	Protein coding	HSP70	Heat shock protein 70 (Hsp 70) family protein. Stress response.	Not available.	Databases.
AT4G32230	Protein coding	Unknown	Hypothetical, uncharacterized protein.	High expression in several accessions, including Bur-0. Expressed in mature pollen, mature embryos.	Databases and Mayer *et al.* (1999).
AT5G20750	Transposable element	Unknown	Not available.	Not available.	Databases.
Cluster 9, Col-0 higher and Bur-0 lower expression					
AT5G49440	Protein coding	Unknown	Hypothetical, uncharacterized protein.	Not available.	Databases.
AT2G04378	Protein coding	Unknown	Beta galactosidase. Signal peptide.	Not available.	Databases.
AT4G07825	Protein coding	Unknown	Transmembrane protein	Expressed in seeds and inflorescence meristems.	Databases and Mayer *et al.* (1999)
AT2G06950	Transposable element	Unknown	Copia-like retrotransposon family. Blast match to dbj|BAA78426.1| polyprotein (AtRE2-1) (*Arabidopsis thaliana*) (Ty1_Copia-element).	Expressed in shoot apex and mature embryo.	Databases.

AT2G10410	Transposable element	SADHU1-1	NON-CODING RETROTRANSPOSON 1-1, SADHU1-1. Member of Sadhu non-coding retrotransposon family.	Expressed in shoot apical meristem, young leaves, seeds.	Databases and Rangwala & Richards. (2010).
AT5G28913	Transposable element	SADHU4-1	NON-CODING RETROTRANSPOSON 4-1, SADHU4-1. Member of Sadhu non-coding retrotransposon family.	Not available.	
AT4G01525	Transposable element	SADHU5-1	NON-CODING RETROTRANSPOSON 5-1, SADHU5-1. Member of Sadhu non-coding retrotransposon family.	Not available.	
AT3G44430	Protein coding	Unknown	Transmembrane protein.	Expressed in seeds.	Databases.
AT3G61010	Protein coding	Unknown	Ferritin/ribonucleotide reductase-like family protein. Similar to F4JD24.	Expressed in rosette and cauline leaves, floral structures, stem.	Databases.
AT2G11240	Transposable element	Unknown	Non-LTR retrotransposon family (LINE). Blast match to GB:NP_038603 L1 repeat, Tf subfamily, member 23 (LINE-element) (*Mus musculus*).	Not available.	Databases.
AT3G47350	Protein coding	ATHSD2, HSD2,	HYDROXYSTEROID DEHYDROGENASE 2. Transmembrane protein. Encodes a putative hydroxysteroid dehydrogenase (HSD).	Not available.	Databases.
AT3G07555	Undetermined	Unknown	Long noncoding RNA.	Not available.	Databases.
AT5G48412	Undetermined	Unknown	Other RNA. Noncoding RNA	Not available.	Databases.
AT5G05060	Protein coding	Unknown	Cystatin/monellin superfamily protein. Similar protein (90% identity to Q9FF68).	Expressed in shoot apex, inflorescence, seed, siliques, flower buds and during early embryogenesis.	Databases.
AT4G11830	Protein coding	PLD-GAMMA2	PHOSPHOLIPASE D GAMMA 2. Encodes one of three phospholipase D enzymes of the gamma class.	Expressed in guard cells, reproductive organs, inflorescence meristem. Higher expression in Col-0.	Databases.
AT1G06640	Protein coding	Unknown	Encodes a protein whose sequence is similar to a 2-oxoglutarate-dependent dioxygenase. The mRNA is cell-to-cell mobile. Involved in oxidation-reduction process.	Expressed in mature embryo and during silique development.	Databases.
AT1G35612	Transposable element	Unknown	Pseudogene of Ulp1 protease family protein.	Expressed in stem and during silique development.	Databases.
AT5G23410	Undetermined	Unknown	Other RNA. Noncoding RNA	Expressed in embryo torpedo and senescent leaves.	Databases.
AT5G42730	Pseudogene	Unknown	Pseudogene similar to ACT domain-containing protein, similar to F-box family protein.	Expressed in embryo torpedo and senescent leaves.	Databases.

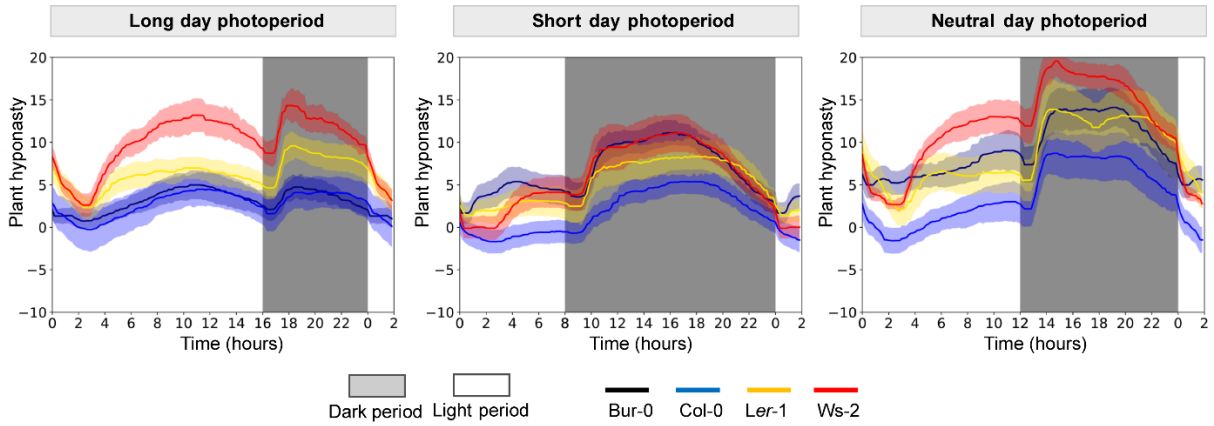
AT5G46510	Protein coding	VICTL	VARIATION IN COMPOUND TRIGGERED ROOT GROWTH RESPONSE-LIKE. Disease resistance protein (TIR-NBS-LRR class) family. VICTL ortholog. Involved in signal transduction, located in cytoplasm, cytosol, has molecular function ADP-binding.	Expressed in rosettes, cauline leaves, SAMs, flowers, imbibed seed.	Databases.
AT4G08110	Transposable element	Unknown	CACTA-like transposase family (Ptta/En/Spm). Blast match to At5g29026.1/8-244 CACTA-like transposase family (Ptta/En/Spm) (CACTA-element) (*Arabidopsis thaliana*).	Expressed in dry seeds, senescence leaves.	Databases.
AT3G47680	Protein coding	Unknown	T23J7.10 Ortholog. MALEYLACETOACETATE ISOMERASE. DNA binding protein, located in chloroplast.	Expressed in dry seeds, senescence leaves and during late embryogenesis.	Databases.
AT5G55790	Protein coding	Unknown	Hypothetical protein. Transmembrane protein. Integral component of membrane, has molecular function. (100% identity to Q9FM60).	Expressed in rosettes, SAMs and early embryo stages. Low expression in Bur-0.	Databases.
AT5G56747	Transposable element	Unknown	Copia-like retrotransposon family. Blast match to GB:CAA72989 open reading frame 1 (Ty1_Copia-element) (*Brassica oleracea*).	Not available.	Databases.
AT1G07323	Undetermined	Unknown	Long noncoding RNA.	Not available.	Databases.
AT1G09995	Protein coding	Unknown	DNA repair DEAD helicase RAD3/XP-D subfamily protein.	Not available.	Databases and Klopffleisch *et al.* (2011).
AT1G34418	Undetermined	Unknown	Noncoding RNA	Not available.	Databases.
AT1G61450	Protein coding	Unknown	CAP-gly domain linker.	High expression in the SAM.	Databases and Hanada *et al.*, (2011).
AT2G13970	Transposable element	Unknown	Mutator-like transposase family. Blast match to GB:AAA21566 mudrA of transposon=MuDR (MuDr-element) (*Zea mays*). Cell-to-cell mobile RNA.	Expressed in collective leaf structure, flower, flower pedicel, guard cell, hypocotyl, petal, plant embryo, root, sepal, shoot apex, stem.	Databases and Thieme *et al.* (2015).
AT2G13975	Transposable element	Unknown	Not available.	Not available.	Databases.
AT2G15292	Undetermined	Unknown	Other RNA. Noncoding RNA.	Not available.	Databases.
AT2G32220	Protein coding	Unknown	Ribosomal L27e protein family. 60S ribosomal protein L27-1. RPL27 A ortholog.	Not available.	Databases.
AT2G34840	Protein coding	F19I3.7; F19I3_7	Coatomer subunit epsilon-2. Vesicle coat protein. Intra-Golgi vesicle mediated transport.	Expressed in rosettes and SAMs.	Databases and Kim *et al*. (2013).
AT3G00170	Undetermined	Unknown	Pseudogene of unknown protein. Miscellaneous RNA	Not available.	Databases.
AT3G24927	Undetermined	Unknown	Pseudogene of unknown protein. Miscellaneous RNA	Not available.	Databases and Hirsch *et al*. (2006).

AT3G27990	Undetermined	Unknown	Other RNA. Noncoding RNA.	Not available.	Databases.
AT3G28160	Transposable element	Unknown	Copia-like retrotransposon family. Blast match to GB:CAA31653 polyprotein (Ty1_Copia-element) (*Arabidopsis thaliana*).	Not available.	Databases.
AT3G44798	Undetermined	Unknown	Other RNA. Noncoding RNA	Not available.	Databases.
AT3G46385	Undetermined	Unknown	Pseudogene of unknown protein. Miscellaneous RNA.	Not available.	Databases.
AT3G46845	Protein coding	Unknown	Hypothetical, uncharacterized protein.	Not available.	Databases.
AT3G61020	Undetermined	Unknown	Pseudogene of unknown protein. Miscellaneous RNA	Expressed in leaves, SAMs, embryos.	Databases.
AT4G02540	Protein coding	Unknown	Cysteine/Histidine-rich C1 domain family protein. CHP-RICH ZINC FINGER PROTEIN-RELATED.	Expressed in leaves and SAMs.	Databases.
AT4G07455	Protein coding	Unknown	Hypothetical, uncharacterized protein.	Not available.	Databases.
AT4G07507	Transposable element	Unknown	Pseudogene, replication protein A1 -related, similar to putative replication protein A1.	Not available.	Databases.

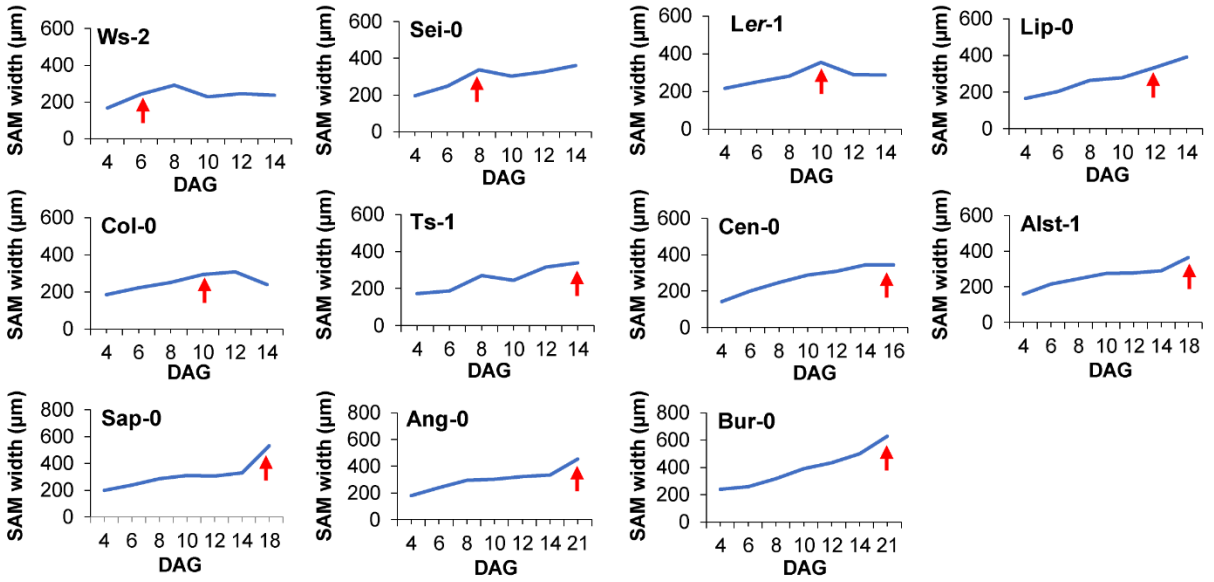
II. Supplementary Figures



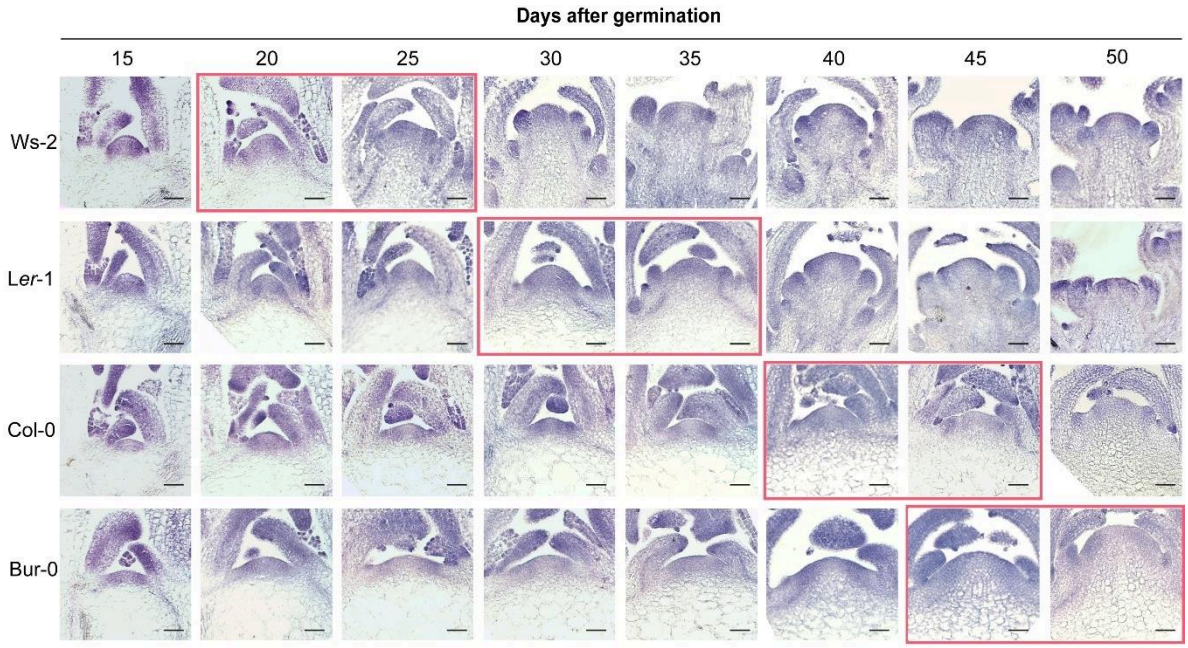
Supplementary Figure S1. Correlations between adult plant traits in early, intermediate and late flowering time accessions. Eleven *Arabidopsis thaliana* accessions were analyzed and plants were grown in long day photoperiod. Linear regressions between the traits: rosette area at 14 days after germination (DAG), flowering time as days to bolting (DTB) and seed area were performed for accessions grouped according to (A) early, (B) intermediate and (C) late flowering time phenotype.



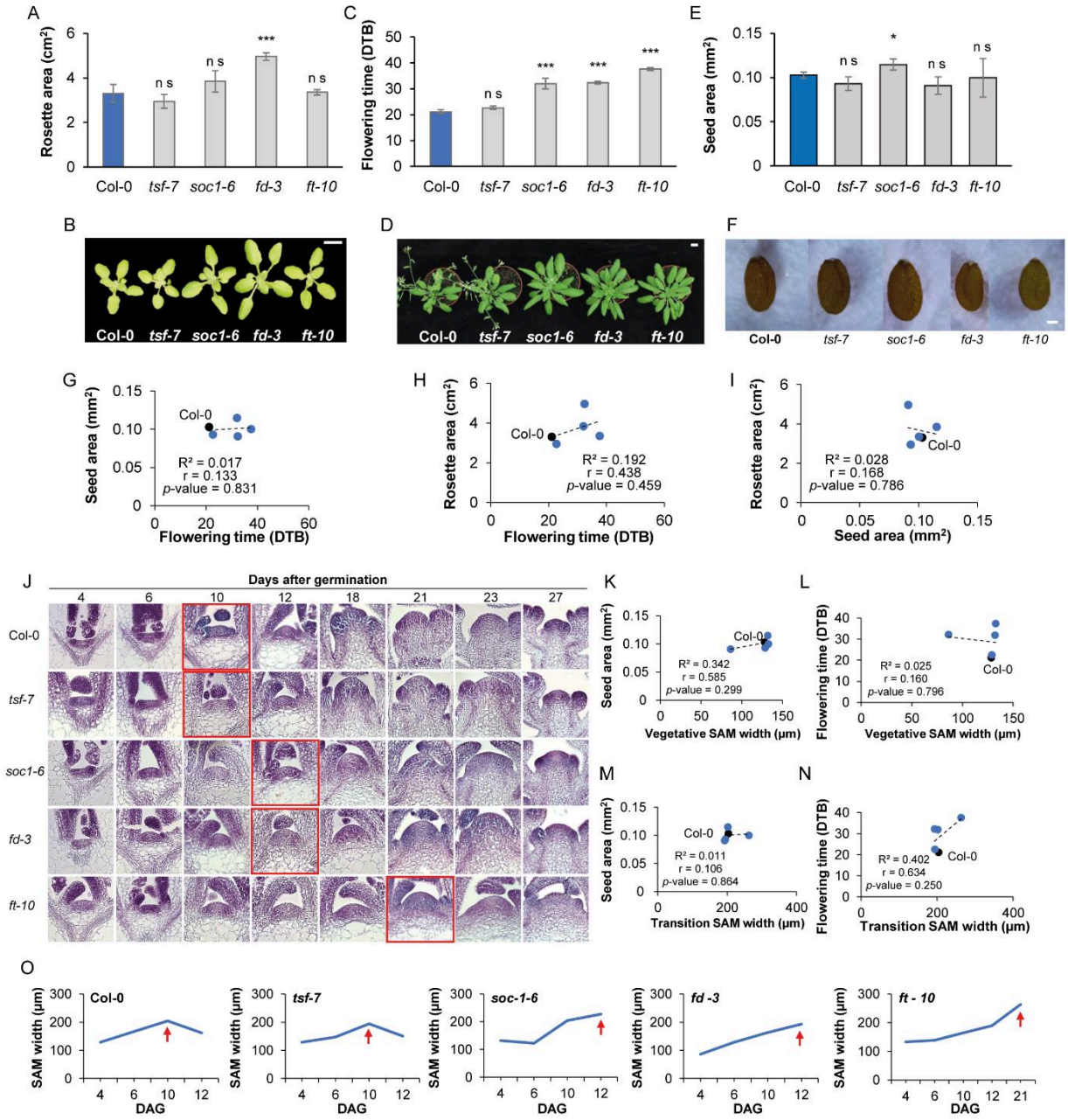
Supplementary Figure S2. Diurnal hyponasty in different photoperiods. The Arabidopsis accessions Bur-0, Col-0, L*er*-1 and Ws-2 were analyzed using a high-resolution 3D phenotyping system and plants were grown in long day, short day, and neutral day photoperiods. $n \geq 7$. Lines and shadows with the same color represent mean and standard deviation for each accession, respectively.



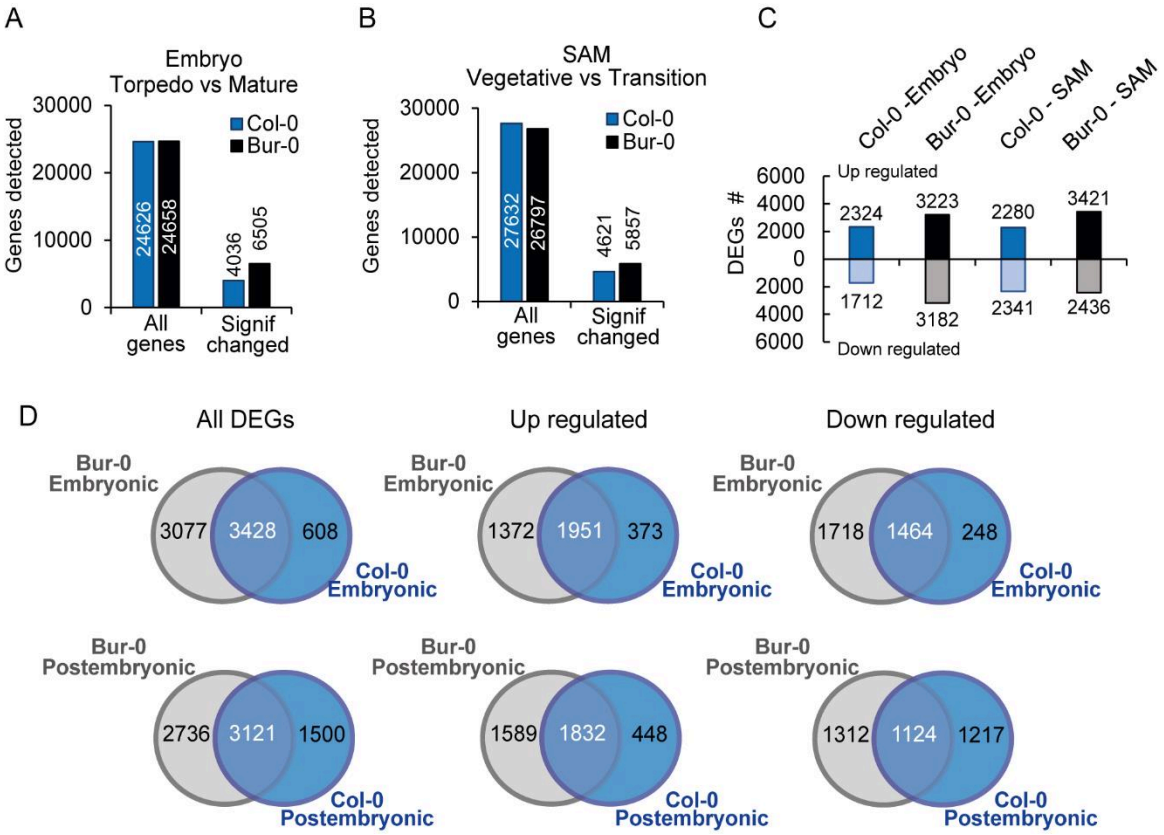
Supplementary Figure S3. Shoot apical meristem (SAM) size changes over time. SAM width was measured as parameter for SAM size in eleven *Arabidopsis thaliana* accessions grown in LD photoperiod. Time is given in days after germination (DAG). ($n \geq 3$). Arrows indicate the time points when the floral transition initiation is morphologically confirmed in each accession as presented in the main text.



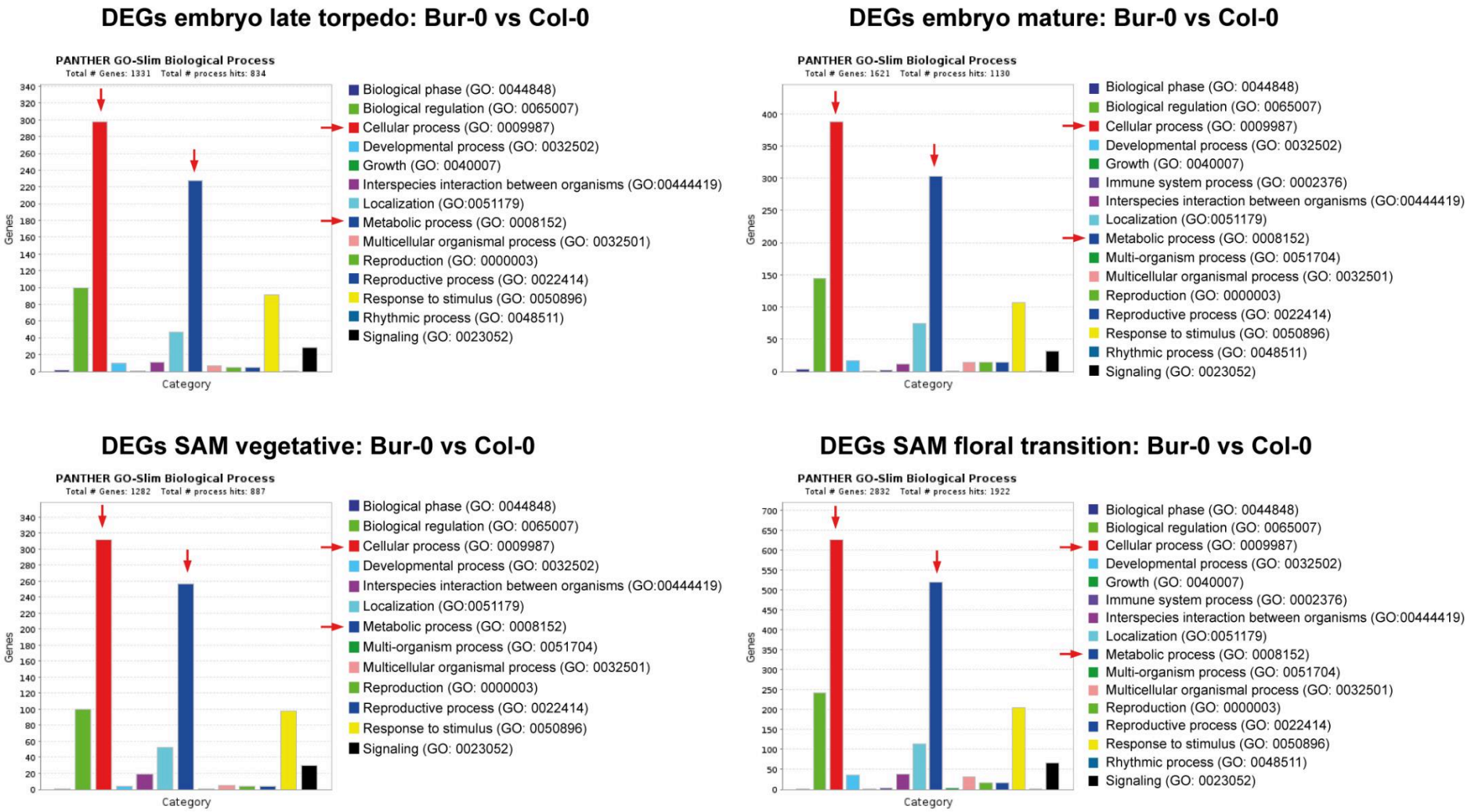
Supplementary Figure S4. Morphological analysis of the shoot apical meristem (SAM) in short day photoperiod (SD). The *Arabidopsis thaliana* accessions Ws-2, L*er*-1, Col-0 and Bur-0 were analyzed and plants were grown in SD. Longitudinal sections of shoot apices at different time points were stained with toluidine blue. Red frames indicate the time point when the floral transition is initiated in each accession. Scale bar = 100µm.



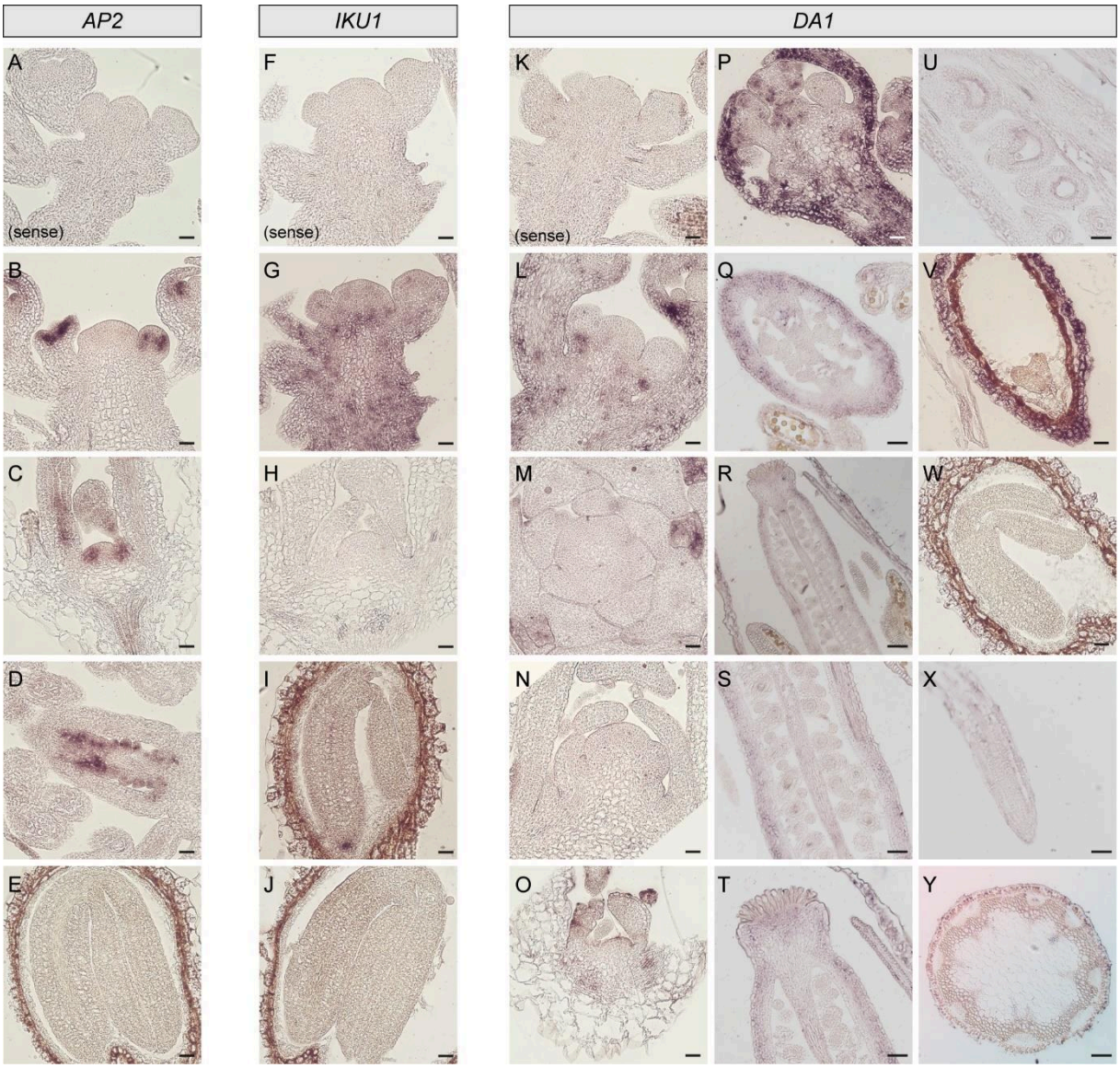
Supplementary Figure S5. Phenotypical characterization of Arabidopsis late flowering mutant lines. The *Arabidopsis thaliana* late flowering mutant lines *tsf-1*, *ft-10*, *soc1-6*, *fd-3* and the wild type Col-0 were analyzed and plants were grown in long day photoperiod. (A) Rosette area at 12 DAG ($n \geq 5$), (B) Rosette phenotype at 12 DAG. Scale bar = 1 cm, (C) Flowering time as days to bolting (DTB) ($n \geq 5$), (D) Rosette phenotype after bolting time. Scale bar = 1 cm, (E) Seed area ($n=60$) and (F) seed phenotype. Scale bar = 0.1 mm. Error bars indicate ± SD. (G-I) Linear regressions between flowering time, rosette area at 12 DAG and seed area. (J) Longitudinal sections of shoot apices at different time points stained with toluidine blue. Red frames indicate the time point when the floral transition is initiated. Scale bar = 100µm. (K, L) Linear regressions between flowering time as days to bolting (DTB), seed area and SAM width at vegetative stage (4 DAG for all mutant lines and the wild type). (M, N) Linear regressions between flowering time as days to bolting (DTB), seed area and SAM width at floral transition stage for each mutant line and the wild type, respectively. (O) SAM width changes from 4 to 12 DAG, ($n \geq 3$). Statistical significance tested with Student's *t*-test: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$, comparing each mutant line to the wild type Col-0. DAG = days after germination.



Supplementary Figure S6. Differentially expressed genes (DEGs) among tissue stages per accession. Differentially expressed genes (DEGs) identified among embryos at late torpedo *vs* mature stage and SAM at vegetative stage *vs* floral transition stage in the Arabidopsis accessions Col-0 and Bur-0, respectively**.** (A) Total number of genes detected and DEGs in Col-0 and Bur-0, significantly changed per embryo stage. (B) Total number of genes detected and DEGs in Col-0 and Bur-0, significantly changed per SAM stage. (C) Differentially expressed genes up and down regulated per tissue and accession. (D) Ven diagrams showing total number of DEGs as well as up and down regulated DEGs per tissue, accession and shared between accessions. SAM= shoot apical meristem.



Supplementary Figure S7. Gene ontology analysis per tissue and developmental stage. Gene ontology terms (GO) enriched for biological process considering differentially expressed genes (DEGs) between the two accessions (Col-0 vs Bur-0) at each embryo and SAM stage analyzed. GO term chart was generated using the tool http://www.pantherdb.org/. Arrows indicate the GO categories more enriched. SAM = Shoot apical meristem.



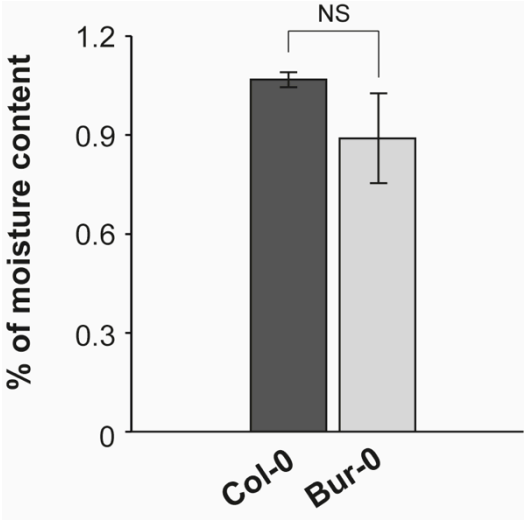
Supplementary Figure 8. Expression analysis of known seed growth regulators by RNA *in situ* hybridization. Expression pattern on different tissues from the Arabidopsis accession Col-0 analyzed by RNA *in situ* hybridization using specific antisense probes for *APETALA 2* (*AP2*) on longitudinal sections of (A, B) Inflorescence meristems, (C) SAM at floral transition stage (10 DAG), (D) gynoecium and ovule primordia, (E) mature embryo. Expression analysis using specific antisense probes for *HAIKU1* (*IKU1*) on longitudinal sections of (F, G) Inflorescence meristems, (H) SAM at floral transition stage (10 DAG), (I) late torpedo embryo and (J) mature embryo. Expression analysis using specific antisense probes for *LARGE IN CHINESE 1 (DA1)* on longitudinal sections of (K, L) Inflorescence meristems, reproductive SAM at 15 DAG (M) transversal section and (N) longitudinal section, (O) SAM at floral transition stage (10 DAG), (P) Flower towards stage 8, (Q) flower towards stage 14, (R) early silique, (S) early ovules, (T) stigma, (U) developed ovules, (V) embryo at heart stage, (W) mature embryo, (X) root tip and (Y) stem transversal section. Sense probes were used as controls. Scale bar: (A-P, V, W, R and Y = 100µm), (Q, S, T, U, X, 50µm). SAM = shoot apical meristem. DAG = days after germination. Plants were grown in long day photoperiod.

ANNEX

A. Supplementary information from the published research article:

Moreno Curtidor C, Annunziata MG, Gupta S, Apelt F, Richard SI, Kragler F, Mueller-Roeber B and Olas JJ. (2020). Physiological Profiling of Embryos and Dormant Seeds in Two Arabidopsis Accessions Reveals a Metabolic Switch in Carbon Reserve Accumulation. Frontiers in Plant Science. 11:588433. doi: 10.3389/fpls.2020.588433
Online access: https://doi.org/10.3389/fpls.2020.588433

Supplementary Figure and Tables from the published research article are kept as presented in the journal and titles or numbering are not modified according to the thesis structure in order to keep the same format as presented in the journal.



Supplementary Figure S1. Water content in dormant seeds. The water content was calculated as the loss in weight as a percentage of the original weight of seeds. Error bars indicate s.d. (n=3). Statistically significant difference between accessions was calculated using Student's t-test (NS; not significant).

Table S1. Total carbon (C) accumulated in torpedo, mature embryos and dormant seeds of Col-0 and Bur-0 plants growing in long photoperiod (LD, 16h light/8h darkness. Values are ± SEM (n=3).

	Stage	Starch (µmol C6 eq. g-1 FW)	Sucrose (µmol/gFW)	Glucose (µmol/gFW)	Fructose (µmol/gFW)	Malate (µmol/gFW)	Fumarate (µmol/gFW)	Amino Acids (µmol/gFW)	C stored in metabolites (µmolC/gFW)
Col-0	Torpedo	6,20	0,98	0,00	0,56	0,00	0,00	8,91	77,28
	Mature	7,10	1,15	0,03	0,65	0,00	0,00	8,34	83,84
	Seed	11,84	4,77	0,57	0,36	15,52	9,10	21,71	293,15
Bur-0	Torpedo	5,73	1,06	0,02	0,59	0,00	0,00	14,05	90,12
	Mature	9,24	0,94	0,01	0,97	0,00	0,00	5,80	88,82
	Seed	13,26	4,87	0,53	0,40	13,75	6,71	21,03	284,37

Table S2. Primer sequences used in this study.

Gene (Atg number)	Oligo name	Sequence (5'→3')
Oligonucleotides used for qRT-PCR		
TUB2	*TUB2_F*	GAGCCTTACAACGCTACTCTGTCTGTC
At5g62690	*TUB2_R*	ACACCAGACATAGTAGCAGAAATCAAG
SUS1	*SUS1_F*	AGTTCACTGCGGATATTTTCG
At5g20830	*SUS1_R*	CCCAACAGTTTCTTTGCTTCCA
SUS2	*SUS2_F*	TGCCATGAATAATGCCGATTTC
At5g49190	*SUS2_R*	TCTTCACTTTGTCGAGCCTCG
SUS3	*SUS3_F*	GACCAGACTGATGAGCATGTCG
At4g02280	*SUS3_R*	TCTTCACTTTGTCGAGCCTCG
SUS4	*SUS4_F*	AAGGAATCGTTCGCAAATGG
At3g43190	*SUS4_R*	TTTCAGCGGCAACATCCTC
SUS5	*SUS5_F*	GCAGTGGTAATTCCTCCGAAC
At5g37180	*SUS5_R*	TCCTCTTACTGCGAACGCTACG
SUS6	*SUS6_F*	CGGAGGCCAGGTTGTTTACAT
At1g73370	*SUS6_R*	AGGCTTGAATCCGAGACCTTGT
CINV1	*CINV1_F*	TTTGACTCTCTCTGAGACACC
At1g35580	*CINV1_R*	ATGACCTCTCTGCCATCTCC
CINV2	*CINV2_F*	ATGCCAGCGAGTTTCAAG
At4g09510	*CINV2_R*	CAACCATAGAACAACCGTCAG
Oligonucleotides used for cloning		
CYCB1;1	*CYCB1;1_c_F*	TCGGTTCTTGTCGGTTAAGCC
At4g37490	*CYCB1;1_c_R*	CCTGTGGTGGCCAAATTTCTT
SUS1	*SUS1_c_F*	ATGGCAAACG CTGAACGTAT G
At5g20830	*SUS1_c_R*	TCAATCATCTTGTGCAAGAGG
SUS3	*SUS3_c_F*	ATGGCAAACCCTAAGCTCAC
At4g02280	*SUS3_c_R*	TCAGTCATCGGCGGTTGAAG