The Quest for a Key Gene Controlling Acquisition of Cell Competence During In Vitro Organogenesis in Tomato

Chanaka Roshan Abeyratne

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation
Abeyratne, Chanaka Roshan, "The Quest for a Key Gene Controlling Acquisition of Cell Competence During In Vitro Organogenesis in Tomato" (2017). Graduate Theses, Dissertations, and Problem Reports. 5019.
https://researchrepository.wvu.edu/etd/5019

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.
THE QUEST FOR A KEY GENE CONTROLLING ACQUISITION OF CELL COMPETENCE DURING *IN VITRO* ORGANOGENESIS IN TOMATO

Chanaka Roshan Abeyratne

Thesis submitted to the Davis College of Agriculture, Natural Resources and Design at West Virginia University in partial fulfillment of the requirements for the degree of

Master of Science

in

Genetics and Developmental Biology

Prof. Vagner A. Benedito, Ph.D., Chair
Prof. Jianbo Yao, Ph.D.
Prof. Teiya Kijimoto, Ph.D.

Division of Plant and Soil Sciences
Morgantown, West Virginia
2017

Keywords: Tomato, MicroTom, *in vitro* organogenesis, Acquisition of organogenic competence, *Rg1* allele, Shoot regeneration, RT-qPCR,

Copyright 2017 Chanaka Roshan Abeyratne
Plants show a remarkable capacity for vegetative propagation through *de novo* organogenesis. Unlike their wild relatives, most commercially important crop species lack this ability, which is largely attributed to their inability to acquire regeneration competence by dedifferentiation of somatic cells. *In vitro* micropropagation relies heavily on overcoming such regeneration recalcitrance for its success in biotechnology and clonal propagation in agriculture. Here, we investigate a key locus (Regeneration1) responsible for acquisition of competence that confers high rates of de novo shoot and root organogenesis in tomato. *In vitro* shoot regeneration assay was performed in several *Solanum pennellii* introgression lines into *S. lycopersicum* cv. Micro-Tom (MT) to help identify 29 candidate *RG1* genes from a genomic region that originally occupied 6.16 Mbps with 293 genes. We used recalcitrant genotype MT as well as the highly regenerant *S. pennellii* and its introgression line MT-Rg3C in a subsequent *in vitro* shoot regeneration assay to characterize the expression of candidate genes using RT-qPCR. Two candidate genes were upregulated in both *S. pennellii* and MT-Rg3C in comparison to MT. Since the dominant recalcitrance trait may be conferred by mutations leading to lack of function in MT, sequence variations of the 29 candidates were compared between the *S. pennellii* and MT genomes. Ten genes were identified with mutations that may cause loss of function. Furthermore, 11 novel proteins with unknown functions were identified with polymorphisms for further investigation. Here we propose at least 10 candidate genes for functional characterization in other highly regenerant genotypes to ultimately end the quest for a key gene controlling organogenesis in tomato.

**Keywords:** Tomato, MicroTom, *In vitro* organogenesis, Organogenic competence, *RGl* allele, Shoot regeneration, RT-qPCR
ACKNOWLEDGEMENTS:

First and foremost I thank my advisor and my mentor Dr. Vagner Benedito for his sage advice, guidance and above most paving the way for me to follow my passion in biology and genetics. He was my lighthouse when navigating the deep sea of knowledge in molecular genetics and research. I like to thank Dr. Benedito for believing in me and for his friendship to which I owe my whole scientific career. I would like to thank Dr. Jianbo Yao and Dr. Teiya Kijimoto for offering me guidance and direction for this research while serving in my graduate committee. I take this opportunity to thank Dr. Yao’s lab for allowing us to use their research facility for RT-qPCR. I take this opportunity to thank Dr. Barton Baker for granting me the opportunity to join the Genetics and Developmental Biology graduate program at WVU.

I thank my collaborators Dr. Lazaro Peres and Dr. Maisa de Siqueira Pinto of Laboratory of Hormonal Control of Plant Development, Department of Biological Sciences, Escola Superior de Agricultura ‘Luiz de Queiroz’ (ESALQ), University of São Paulo (USP) Piracicaba, Brazil for providing the biological material for this research. Especially I thank Dr. Siqueira Pinto, for her friendship and guidance in carrying out this research.

I thank the Evansdale Green house staff at WVU, Sue Myers and Gail Sikorsky for their help and care for the biological material used in this research.

Furthermore, I thank my lab mates especially Prasanthi Koganti and Lucas Maia for their support during the whole of this project and for their undying curiosity and energy as fellow students and researchers. I thank them for being true friends. I thank Dr. Nipuni Palliyaguru for her generous advice, moral support and true friendship. Morgantown can be such a lonely barren place for an international scholar if not for their friendship.

Last but not least I thank my family for being supportive of my work and allowing me to follow my passion. I thank my mother who supports me and encourage me to follow my dreams. I thank my brother for being a pillar of support in both good and bad times and especially putting up with me for this long in life. I thank my wife and my
daughter for being patient with me and supporting my lofty dreams.

This research could have not been possible if not for the advice, support, love and care of above individuals and many more whom are not mentioned in this short space. I thank you all from the bottom of my heart.
TABLE OF CONTENTS:

ABSTRACT: ......................................................................................................................... I

ACKNOWLEDGEMENTS: ..................................................................................................... iii

TABLE OF CONTENTS: ..................................................................................................... v

CHAPTER ONE ......................................................................................................................... 1

INTRODUCTION AND RESEARCH RATIONALE ................................................................. 1

1.1 What drives organogenesis in plants? ........................................................................ 1

1.2 The Regeneration1 (Rg1) allele confers high rates of in vitro organogenesis in tomato .... 3

1.3 Background and Goals .............................................................................................. 5

1.4 Specific Aims of this Research .................................................................................. 6

CHAPTER TWO ......................................................................................................................... 9

LITERATURE REVIEW ........................................................................................................... 9

2.1 The uniqueness of plant organogenesis and regeneration in relation to animals 9

2.2 Adventitious organogenesis and somatic embryogenesis in plants: pluripotency versus totipotency ................................................................. 10

2.3 In vitro organogenesis in plants .............................................................................. 12

2.4 Experimental characterization of phases in in vitro organogenesis ...................... 13

2.5 Acquisition of competence: dedifferentiation or redifferentiation? ....................... 16

2.6 Understanding the molecular mechanisms of shoot regeneration ....................... 18

2.7 Molecular pathways of in vitro organogenesis ...................................................... 19

2.8 The role of plant hormones during in vitro organogenesis in tomato ................. 28

2.9 The possible role of epigenetics on in vitro organogenesis .................................... 30

CHAPTER THREE .................................................................................................................... 32

MATERIAL AND METHODS ................................................................................................. 32

3.1 Plant material and growth conditions ..................................................................... 32

3.2 In vitro tissue culture ............................................................................................... 32

3.3 Expression analysis of Rg1 candidates during in vitro organogenesis using RT-
qPCR .......................................................................................................................... 33

3.4 Statistical layout of the experiment and the statistical analysis ....................... 34

3.5 DNA sequence analysis .................................................................................... 35

CHAPTER FOUR ........................................................................................................... 36

RESULTS AND DISCUSSION .................................................................................. 36

4.1 Narrowing down the Rg1 locus to two sub regions with 29 candidates ............ 36

4.2 Gene expression analysis using RT-qPCR and integration with public RNA-Seq metadata ........................................................................................................... 37

4.3 Expected gene expression profile of RG1 candidate in the genotypes under investigation ............................................................. 38

4.4 Evaluation of Rg1 candidates based on gene sequence polymorphism leading to loss-of-function products ................................................................. 43

CHAPTER FIVE ........................................................................................................... 52

CONCLUSIONS AND PERSPECTIVES ................................................................ 52

5.1 The Rg1 candidate gene shortlist ..................................................................... 52

REFERENCES: ......................................................................................................... 55

APPENDIX 1 .............................................................................................................. 105

Heritability of the high regeneration capacity in tomato and comparisons across related species ......................................................................................................... 105

APPENDIX 2 .............................................................................................................. 107

Micro-Tom (MT) as a model system to study in vitro regeneration ....................... 107

APPENDIX 3 .............................................................................................................. 108

The SnpEff and the subsequent Perl coding scripts ............................................... 108

Figure 2. MT-Rg3C in vitro regeneration phenotype .............................................. 76

Figure 3. Rg1 locus on chromosome 3 of tomato genome ..................................... 77

Figure 4. Stages of plant in vitro organogenesis ..................................................... 78

Figure 5. Molecular pathway of in vitro organogenesis ......................................... 79

Figure 6. Statistical layout for analyzing gene expression for Rg1 candidates ......... 80
Figure 7. Heatmap of 28 candidate gene expression within the Rg1 locus region along with tomato homologs of established markers of in vitro regeneration..................81

Figure 8. Rg1 candidate gene expression over the experimental time-course. ............85

Figure 9. Expected Rg1 expression pattern...............................................................89

Table 1. Functional annotations of Rg1 candidate gene shortlist.................................90

Table 2. RT-qPCR primers used for gene expression analysis....................................91

Table 3. Rg1 candidate gene short list and polymorphisms seen in comparison to tomato reference genome..............................................................93
CHAPTER ONE

INTRODUCTION AND RESEARCH RATIONALE

1.1 What drives organogenesis in plants?

Plants have a remarkable capacity to form new organs in ectopic locations throughout their lifecycle (Ikeuchi et al., 2016; Huijser & Schmid, 2011). Being largely sessile organisms, this ability provides them with an edge to survive and propagate amidst environmental stresses (Gaillochet & Lohmann, 2015; Kareem et al., 2016b). However, across species and often across genotypes within the same species, plants vary in their capacity to carry out de novo adventitious organogenesis. This ability to regenerate organs in ectopic locations is a heritable trait mainly attributed to either the presence of competent pluripotent cells in mature tissues or the ability of differentiated cells to acquire regeneration competence (Atta et al., 2009; Christianson & Warnick, 1983).

Research on plant in vitro micropropagation have been formative in experimentally identifying a distinct phase responsible for acquisition of competence of explant cells. This unique phase is decoupled, yet shared by hormone-mediated organ (shoot and root) induction, which later produces cells committed to an organ fate (Christianson & Warnick, 1983). However, it is noteworthy that the more committed a cell is to a programmed fate, less competent it is for induction towards a different regeneration fate (Lombardi-Crestana et al., 2012). Although a theoretical framework for acquisition of competence that requires a reversal of cell fate programming to achieve a less specialized, the early meristematic character in plants is a logical system for studying de novo organogenesis since the exact mechanism that takes place and the key molecular regulators involved in this early phase are yet to be characterized (Sugimoto et al., 2011; Motte et al., 2014; Gaillochet & Lohmann, 2015; de Siqueira Pinto et al., 2017).

Beyond providing a fundamental understanding of plant development, the identification and characterization of key molecular components and associated pathways for acquisition of competence by explant cells, will immensely benefit applied research in biotechnology (Pino et al., 2010; De Faria et al., 2002). Modern biotechnological tools, such as the CRISPR/Cas9 gene editing system, rely heavily on our ability to carry out in
vitro organogenesis (Bradshaw, 2016). Thus, it is clear that the commercial production of plants with economically desired traits acquired by manipulation of molecular pathways as well as maintaining clonally important transgenic material for research, all depend on the high success rate of micropropagation that, in turn largely depend on the capacity of explant cells to acquire organogenic competence (Gerszberg et al., 2015; Lombardi-Crestana et al., 2012; Atta et al., 2009).

Just as hormones, such as auxin and cytokinin, are involved in de novo organogenesis in intact plant tissue, artificial growth media containing different ratios of these hormones supplemented with a carbon source and macro/micronutrients are used in micropropagation. These systems can induce in vitro callogenesis, shoot organogenesis, or rhizogenesis. Although a specific exogenously supplied ratio of growth regulators (auxin:cytokinin), rather than their absolute concentrations, is needed for the cells of a given explant type to initiate organ induction (Skoog & Miller, 1957), experimental evidence is lacking for such a dependence on a specific medium type for acquisition of competence, which makes it experimentally more challenging to manipulate this early, ephemeral phase. Since the acquisition of cell competence precedes all other phases of in vitro organogenesis, such as hormone-mediated induction to fate determination and the subsequent morphogenesis (organ development) by growth and differentiation (Christianson & Warnick, 1983a), understanding this process at the molecular level is imperative in order to harness its potential to be used productively in biotechnology. To this end, genotypes that exhibit in vitro and ex vitro phenotypes with enhanced efficiency or aberrations in acquisition of cell competence, provide a useful resource to diametrically study molecular and genetic perturbations to find culprit genes involved during de novo organogenesis (Lombardi-Crestana et al., 2012; Kareem et al., 2015; Atta et al., 2009; de Siqueira Pinto et al., 2017).
1.2 The Regeneration1 (Rg1) allele confers high rates of *in vitro* organogenesis in tomato

Over two decades ago, Koornneef *et al.* (1993) identified two distinct quantitative trait loci (QTLs) responsible for high rates of *in vitro* organogenesis in tomato, coined *Regeneration1* (Rg1) and 2 (Rg2). Both Rg1 and Rg2 are additive (i.e., non-epistatic) dominant alleles displaying Mendelian segregation. Both loci had implied function during acquisition of regeneration competence by explant tissue. They were originally introgressed from the tomato wild species, *S. peruvianum* into *S. lycopersicum* cv. MsK8 using classical breeding techniques (Koornneef *et al.*, 1987). In the current study, we focus our efforts on identifying the gene responsible for the high organogenesis trait in the Rg1 locus.

The Rg1 allele was phenotypically characterized by inducing high shoot regeneration capacity on both, root explants and callus (Koornneef *et al.*, 1993). Using traditional breeding with backcrossing and selfing progeny lines (Lima *et al.*, 2004), further introgressed Rg1 from *S. lycopersicum* cv. MsK8 into the tomato model, cv. Micro-Tom (MT) to create a near isogenic line (NIL), MT-Rg1 (*Figure 1A*). Indeed, MT-Rg1 shows significant increase in *in vitro* organogenesis of both shoot and root compared to the wild-type MT, when 8-day old cotyledon explants are induced on media containing appropriate phytohormones (*Figure 1D-G*). Direct incubation of Rg1 cotyledon explants on shoot induction medium (SIM, which is supplemented with the cytokinin BAP) showed a robust 60% shoot regeneration rate, which is significantly higher than the MT mean regeneration rates ranging around 30-40% (Pino *et al.*, 2010). Similar rates were observed for root regeneration from cotyledons using RIM (Lombardi-Crestana *et al.*, 2012). The shoot regeneration capacity of MT-Rg1 root explants is equivalent to that reported in (Koornneef *et al.*, 1993).

*Solanum pennellii* cv. LA716 is another desert dwelling wild species of tomato from Peru that was documented to show high *in vitro* regeneration capacity (Arikita *et al.*, 2012). A study using 46 introgression lines (ILs) of *S. pennellii* into *S. lycopersicum* cv. M82 identified four key ILs (3-2, 7-1, 8-3 and 10-2) displaying high shoot and root regeneration rates of cotyledon explants (Arikita *et al.*, 2013). These ILs also showed
higher in vitro shoot and root regeneration capacity (Figure 2A-I), with 60-80% of explants regenerating shoots (Figures 2J-L). In order to be consistent with nomenclature used in a previous study exploring tomato fruit color for *S. pennellii* ILs, bins carrying critical regeneration loci identified using these IL population were named as RG3C, RG7H, RG8F, and RG10F (Liu *et al.*, 2003). These loci were then further introgressed into the MT background and the high in vitro regeneration capacity of these MT-ILs was confirmed using the same in vitro regeneration assay for shoot and root using cotyledon explants (Figure 2J) (Siqueira Pinto *et al.*, 2017). Importantly, the RG3C bin includes the *RgL* locus previously identified by morphological markers (Koornneef *et al.*, 1993). (cf. Appendix I for a review of heritability of the high regeneration trait on other tomato wild species and introgression lines). Notwithstanding, this phenomenon has not yet been fully characterized at the molecular genetics level, which is underscored by the experimental gap in the literature reviewed. The ability to improve both shoot and root organogenesis programs implies that the *RgL* allele may play a fundamental role in a developmental process shared by both shoot and root organogenesis, such as acquisition of cell competence.

Based on information from the ‘150 Tomato Genome Resequencing Project’ ([www.tomatogenome.net](http://www.tomatogenome.net)) (Kobayashi *et al.*, 2014; Causse *et al.*, 2013), the original genomic region of *RgL* first introgressed into the cv. MsK8 and mapped by (Koornneef *et al.*, 1993) contains a genomic region of 6.16 Mb with 293 genes laying between the morphological genetic markers, *CrtR-b* (*Solyc03g007960* – white flower, *wf*) and *PSY1* (*Solyc03g031860* - yellow fruit, *r*). These 293 genes (ITAG2.3 release) are genuine candidates for conferring the high regeneration phenotype. Furthermore, a separate study (Azevedo, 2012) used cleaved amplified polymorphic sequence (CAPS) markers to further map *RgL* in a *S. peruvianum* x MT mapping population. The study narrowed down the candidate *RgL* to a 4.7-Mb region containing 139 genes between the markers *CrtbR-b* and P5 (*Solyc03g025320*) on the short arm of chromosome 3. Furthermore, our efforts described here were able to narrow down *RgL* genomic region to 29 candidate genes contained within 2.18 Mb by comparing results from three independently published experiments (Figure 3; Arikita *et al.*, 2013; Chitwood *et al.*, 2013; de Siqueira Pinto *et al.*, 2017).
1.3 Background and Goals

*In vitro* assays, such as shoot regeneration from cotyledon or root explants, can be used as a reproducible phenotypic test to differentiate between organogenic recalcitrance and high regeneration capacity (Koornneef et al., 1993) (Figure 1D,F). The organogenic capacity of explants of a certain genotype is largely attributed to the pre-existence of regeneration competent cells within its tissues, or the ability to acquire regeneration competent cells in the form of organogenic callus (Christianson & Warnick, 1983). Competent cells within a mass of callus can be induced to a committed cellular fate toward organogenesis, by plant hormones that are either endogenously synthesized or exogenously supplied (Skoog & Miller, 1957; Ikeuchi et al., 2016; Hill & Schaller, 2013).

We hypothesize that the *Rg1* culprit gene codes for a key player in the molecular pathway producing competent cells. The *Rg1 in vitro* phenotype shows increased shoot and root regeneration from both shoot and root explants (Lombardi-Crestana et al., 2012; Siqueira Pinto et al., 2017). Since the shoot and root organogenic pathways share the same early phase when producing regeneration competent cells, *Rg1* may act either by increasing the number of cells acquiring competence or increasing the efficiency of existing cells to acquire competence. The analogous lines MT-*Rg1* and MT-*RG3C* both contain natural variants of the *Rg1* allele introgressed from the tomato wild species *S. peruvianum* and *S. pennellii*, respectively. Indeed, tomato natural wild species in general display higher organogenic capacity in comparison to their domesticated counterparts (Peres et al., 2001). This trait is considered as an adaptation to harsh ecological niches occupied by the wild species in order to fend off and recover from various biotic and abiotic stresses (Arikita et al., 2012). Associated *ex vitro* phenotypes, such as prolific branching, and extensive investment of resources in above- and below-ground biomass used for foraging resources are associated with highly organogenic wild species and are often absent in domesticated species (Pino et al., 2010; Lombardi-Crestana et al., 2012). All these *ex vitro* phenotypes rely heavily on the availability and the activity of a meristematic tissue competent for organogenic induction. Indeed, MT-*Rg1* and MT-*RG3C* both show these *ex vitro* phenotypes [highly branched stem (i.e., high branching index), short stature (Figure 1A) with thick petioles (Figure 1C)], and bifurcated
tricotyledons (Figure 1B)] in comparison to the recalcitrant MT genotype. This observation supports the hypothesis that the elusive \( Rg1 \) allele introgressed from tomato wild species is involved in higher efficiency in the phase of acquisition of competence. Experiments with \textit{Arabidopsis thaliana} identified that pericycle cells in the root, or pericycle-like cells in aerial tissues, serve as founder cells for organogenic callus formation during the initial phase of acquisition of competence (Atta \textit{et al.}, 2009; Kareem \textit{et al.}, 2015). Therefore, we expect the \( Rg1 \) culprit gene to show either aberrations of gene expression or loss of gene product functionality within explant pericycle or pericycle-like cells during induced organogenesis. Changes to molecular pathways that gives rise to organogenic competent cells could result from differential transcriptional regulation of key molecular regulators (e.g., transcription factors, as well as chromatin remodelers, such as histone modifiers, DNA (de)methylases) or functional aberrations of gene products due to lack-of-function mutations that result in truncated peptides, non-functional domains, or misfolded proteins.

The long-term goal of this research is to unveil the genetic pathways involved in plant organogenesis. By taking advantage of unique genetic resources available in the tomato cv. Micro-Tom system, as well as preliminary information available on \( Rg1 \) genetic mapping, phenotypical characterization of introgression lines, and a comprehensive whole-genome sequence dataset, we are in an advantageous position to answer our research question: \textit{Which gene in the \( Rg1 \) locus is responsible for inducing high organogenic response in vitro?} In other words, what is the identity of the long-sought \( Rg1 \) gene?

1.4 Specific Aims of this Research

1.4.1 Identify and shortlist candidate genes for \( Rg1 \) based on their differential expression profile

As a first step, we characterize the differential expression of the 29 \( Rg1 \) candidate genes between highly regenerant \textit{S. pennellii} (accession LA716), the recalcitrant cv. MT model, and the introgression line MT-\textit{RG3C}. It is well established that the \( Rg1 \) locus is involved in making cotyledon cells more competent to regeneration by hormone-induced
mechanisms (Koornneef et al., 1993; Lombardi-Crestana et al., 2012). Although there are no genetic markers so far identified for acquisition of cell competence during organogenesis of any plant species, bona fide markers for shoot determination and development are available in Arabidopsis thaliana (Gallois et al., 2002; Cary et al., 2002; Atta et al., 2009). WUSCHEL (WUS) and SHOOT MERISTEMLESS (STM) are two such marker genes for which homologs have been characterized in tomato and have been used in shoot regeneration studies (Lombardi-Crestana et al., 2012; Siqueira Pinto et al., 2017). Therefore, expression data from these marker genes, along with the 29 candidate genes of the Rgl locus, in a time-course of incubating cotyledon explants on SIM from contrasting genotypes (MT), MT-RG3C and S. pennellii (LA716) will be collected to reveal their expression profile during in vitro shoot organogenesis. Since the Rgl locus is demarcated between Solyc03g007960 and Solyc03g025320 (encompassing the 29 candidate genes and containing the key molecular regulator of acquisition of organogenic competence (i.e., cell dedifferentiation), we will confine our focus within this specific genomic region. Since the Rgl allele is dominant with Mendelian segregation, we expect the highly regenerant tomato variants S. pennellii and MT to show a diametric expression pattern while MT-RG3C could be intermediate or similar to S. pennellii in expression of Rgl.

1.4.2 Identify DNA sequence differences that may allude to differences in functional gene products in the Rgl locus

The putative genomic region for Rgl contains 29 genes, some of them coding for gene products with largely unknown functions, while others have potential roles in gene transcription, translation or are involved in secondary active transport and metabolic processes (Table 1). Since Rgl is a dominant allele that confers high regeneration rates to wild tomato species than domesticated varieties of S. lycopersicum, the REGENERATION1 (RG1) gene product could potentially be an inducer of a single or multiple molecular pathways associated with the phase of acquisition of competence.

Here we focus on identifying sequence variants (mainly point mutations and indels) in exonic regions and intron-exon junctions where splice sites are located. We will take into consideration the relative impact of point mutations and indels when evaluating changes
to the gene sequence. With the knowledge that $RgI$ is a dominant allele we expect the recalcitrant genotype MT to contain loss of function sequence variants.
CHAPTER TWO
LITERATURE REVIEW

2.1 The uniqueness of plant organogenesis and regeneration in relation to animals

In both kingdoms, higher order animals and plants produce zygotes during sexual reproduction, which develop into complex multicellular, adult forms mediated through embryogenesis. This feature is common to both groups irrespective of the evolutionary divergence between these lineages at the single cell stage (Cridge et al., 2016; Kuijk et al., 2015; Rensing, 2016). In both, programmed cellular division, growth and planned differentiation result in specialized organs with multicellular, well-defined three-dimensional structures (Ikeuchi et al., 2016). However, higher order animal morphology does not change much throughout adult life. Therefore, animals do not have the capacity to fully regenerate organs if lost, except in a few remarkable cases, such as Urodela amphibians (salamanders), which retain their competence of regenerating complex structures such as limbs, tail, and the spinal cord throughout their lives (Roy & Lévesque, 2006; Looso et al., 2013). Plants, in contrast, can not only replace lost organs but continue to grow and develop new organs such as shoots and roots, continuously changing morphology throughout their lifetime (Goldberg et al., 1994; Drost et al., 2016).

Both, plants and animals use wound healing and regeneration mechanisms to repair or replenish damaged or senesced tissues and organs (Birnbaum & Alvarado, 2008). However, plants inherently express a much higher regeneration capacity of wound healing and de novo organogenesis, than that of higher order animals (Ikeuchi et al., 2016; Kareem et al., 2016). This trait alleviates the disadvantages of a sessile lifestyle which otherwise may have made plants too vulnerable to damage by not being able to move away from harmful environmental stresses, or readily move towards more favorable habitats.

Both, de novo organogenesis and wound healing abilities seen in plants, can be attributed to the superior developmental plasticity of differentiated plant cells (Lup et al., 2016). Driving this plasticity is the capacity of plant cells to acquire competence to form
pluripotent cells capable of producing organogenic primary and secondary meristematic cell niches located throughout their adult structure (Brand et al., 2001). This capability is unsurpassed by higher order animal species, but given that some parallels between development between multicellular organisms in plants and animals can be established (e.g., polarity and asymmetric cell divisions during embryo differentiation, stomata and muscle development programs), it would not be surprising to realize in future that acquisition of competence by animal stem cells could have recruited the same basic genetic mechanisms as plants (Serna, 2009).

### 2.2 Adventitious organogenesis and somatic embryogenesis in plants: pluripotency versus totipotency

Post-embryonic development and *de novo* organogenesis in plants are carried out by primary and secondary meristematic tissues, such as shoot and root apical meristems, procambium, and the vascular cambium (Jones et al., 2013; Goldberg et al., 1994). However, often in response to trauma (Cary et al., 2002), or means of asexual propagation, adventitious organogenesis is observed (Jones et al., 2013). When a plant tissue is cut or harmed, the damage triggers a stress response that results in the neighboring cells close to the wound to proliferate and produce a mass of callus tissue and, at optimum conditions, give rise to founder cells at the base of the callus that ultimately induces the formation of new organ primordia (Kareem et al., 2015, 2016). The organogenic meristematic tissue produced during this process does not directly originate from existing primary or secondary meristems, thus they are considered ectopic or adventitious. The phenomenon of adventitious organogenesis showcases the plasticity of cellular differentiation program, and its capacity to be reversed to early developmental stages that can then be re-differentiated into other non-embryonic cell types. This character of plant cells is known as **pluripotency** (Atta et al., 2009).

*In vitro* micropropagation is a technique that evinces the process of adventitious organogenesis and the capacity of plants to generate pluripotent cells from differentiated cells. During *in vitro* organogenesis, pericycle cells or pericycle-like cells in explants that are differentiated with an arrested cell cycle (Lavenus et al., 2013), acquire competence to divide and become pluripotent, producing a mass of callus that gives rise to an organ
primordium that in turn results in a meristem (Pulianmackal et al., 2014; Atta et al., 2009).

It is important to establish a clear distinction between the processes of adventitious organogenesis and somatic embryogenesis. Somatic embryogenesis results in the development of an entire plant structure, off an embryo generated by a single somatic cell (Fehér, 2015). A single, differentiated somatic cell can be cultured on artificial growth media with appropriate nutrients and growth regulators to produce a somatic embryo (Vasil & Hildebrandt, 1965; Steward et al., 1958). This capability of a somatic plant cell to gain a meristematic character and generate cells with the capacity to differentiate into all types of cell lines, including embryonic tissues, encompasses the character known as totipotency (Sugiyama, 2015; Gaillochet & Lohmann, 2015). Somatic embryogenesis occurs naturally in certain plants, such as in species of the Bryophyllum clade (Crassulaceae family) known as mother-of-thousands. The leaves of these plants can produce individual somatic embryos that arise from individual leaf cells. For these species, in planta somatic embryogenesis serves as the main vegetative reproduction method. Therefore, regardless whether in planta or in vitro, somatic embryogenesis showcases the plasticity of reprogrammed cellular differentiation, i.e. the capacity of differentiated somatic cells to alter or return to initial stages of its developmental fate in order to regain the totipotent, zygotic character.

During this process, the somatic cell is expected to dedifferentiate, become totipotent by activating the cell cycle mechanism, and completely reorganize its transcriptome, leading to a complete overhaul of metabolic and physiological states suited for subsequent embryogenic development. Stress perception, chromatin remodeling and endogenous growth regulator mediated signaling are important components that work in tandem to carry out somatic embryogenesis (Fehér et al., 2003; Loyola-Vargas, 2016). Whether all or some of these same mechanisms and molecular pathways are involved in the early stages of adventitious organogenesis when pericycle or pericycle-like cells acquire competence to form organogenic callus is however still up for debate.
2.3 *In vitro* organogenesis in plants

The process of plant adventitious organogenesis is studied and harnessed for mass propagation using *in vitro* regeneration methods. This technique is more commonly known as tissue culture. Plant hormones produced endogenously by the explant tissues, or growth regulators supplied exogenously via the growth medium, especially auxin and cytokinin, induce organogenesis (Cary *et al.*, 2002). Importantly, explant tissues respond to auxin-cytokinin ratios rather than their absolute concentrations supplied. This was first experimentally demonstrated by Skoog & Miller (1957) in *Nicotiana tabacum* pith explants using indole acetic acid (IAA) as the auxin (Gautheret, 1983) and kinetin as the cytokinin (Miller *et al.*, 1956). A high auxin-to-cytokinin ratio induces root development while a low ratio induces shoot organogenesis in a range of explant types (Skoog & Miller, 1957; Koornneef *et al.*, 1993; Valvekens *et al.*, 1988; Atta *et al.*, 2009; Lombardi-Crestana *et al.*, 2012; Kareem *et al.*, 2015). These media are widely identified as root induction medium (RIM) and shoot induction medium (SIM), respectively. A medium containing a balanced level of both auxin and cytokinin is effective at producing callus, thus called callus induction medium (CIM). Many different renditions that broadly fall into these three categories of growth medium types use both naturally occurring and synthetic growth regulators (auxins: NAA, IBA, 2,4-D; and cytokinins: BAP, iP, zeatin) to successfully induce organogenesis (Plíhalova *et al.*, 2016).

*In vitro* organogenesis can be successfully achieved with both embryonic (e.g., hypocotyl, cotyledon) as well as post-embryonic (e.g., leaf, root, stem) tissues, demonstrating the capacity of a wide range of differentiated tissue types to acquire competence and to carry out *de novo* organogenesis. However, a single *in vitro* regeneration protocol does not fit either explants derived from diverse plant species or all types of explants of a given species. For example, a two-step culture method is the most effective for *Arabidopsis thaliana* (Valvekens *et al.*, 1988). The explant is first incubated on CIM to induce calllogenesis and then transferred to either SIM or RIM to produce shoots or roots, respectively (Valvekens *et al.*, 1988). On the other hand, in tomato, direct incubation on either RIM or SIM without an intervening CIM incubation stage produces sufficient shoot or root organogenesis in regenerant genotypes (de Siqueira Pinto *et al.*, 2017). However, not all explant types and genotypes are easily amenable to these
established regeneration protocols (Arikita et al., 2012; Pino et al., 2010). We know that certain plant genotypes show reduced or complete recalcitrance to organogenesis. The recalcitrance observed in these phenotypes cannot be attributed to a sensitivity of explant cells to the growth regulators in the medium or to homeostasis of endogenous hormones (Boiten et al., 2004; Lombardi-Crestana et al., 2012). Rather, these cells carry an innate recalcitrance that makes them non-responsive to hormonal signaling, which implies the existence of an early stage in organogenesis that prepares them for hormone perception called acquisition of competence. Therefore, identification and characterization of molecular pathways involved in this early stage is imperative for efficient application of micropropagation and understanding of plant development.

2.4 Experimental characterization of phases in in vitro organogenesis

We know that during in vitro organogenesis an explant goes through three distinct phases (Christianson & Warnick, 1983b). (i) Acquisition of competence by explant cells, (ii) Induction of competent cells to commit to an organ (root or shoot) developmental fate, and (iii) Organ determination that will guide cells into differentiation of a fully functional meristem (Figure 4). During the first phase of organogenesis, which is common to both root and shoot, differentiated cells of the explant acquire competence to respond to growth regulators in the media (such as SIM, RIM and CIM), thus acquiring an undifferentiated, pluripotent state. In the second phase, these competent cells are induced by exposure to a specific auxin-to-cytokinin hormonal ratio in the growth media in order to become determined or committed to a specific organ fate of either shoot or root. During the third and final phase (determination), induced cells are reprogrammed to establish organ primordia that will produce the differentiated tissues to develop complete organs via regular morphogenesis that is also observed in shoot and root meristems of intact plants. (Figure 4). These stages were first experimentally described in bindweed (Convolvulus arvensis, Convolvulaceae) (Christianson & Warnick, 1983b) and since then confirmed by various experiments using many established plant model systems (Kareem et al., 2015; Siqueira Pinto et al., 2017).
(i) Acquisition of competence

Although the end result of acquisition of competence is experimentally defined, information regarding the early stages which initiates this process is fragmentary (Motte et al., 2014). Cellular competence can be broadly defined as the “capacity to respond to the inductive effects of a growth medium” (Christianson & Warnick, 1983b; Sugimoto et al., 2010). This requires the differentiated cells to first acquire a developmental state with no distinct commitment to a particular organ fate. However, acquisition of competence is not straightforward to observe for being a transitory, inconspicuous stage that does not necessarily rely on a particular ratio of exogenously supplied hormones in the growth medium, at least for organogenic explants (Koornneef et al., 1993; Siqueira Pinto et al., 2017). The inability to acquire cell competence is generally cited as the main reason for organogenic recalcitrance in explant types (i.e., tissues of the same plant) as well as different genotypes within species and it can explain the reason why some species are more prone to organogenesis than others (Sena et al., 2009; Pulianmackal et al., 2014; Christianson & Warnick, 1983). Therefore, the lack of control over acquisition of competence by explant cells poses a major bottleneck and a rate-limiting step, which girdles efficiency of in vitro organogenesis. This presents a significant hurdle for research in biotechnology and industrial research.

The time required for explant cells to become competent is stable across media but may vary by plant species/genotypes for a given growth condition (e.g., medium composition) (Christianson & Warnick, 1983b; Siqueira Pinto et al., 2017). For example, it has been shown that S. lycopersicum cv. MT takes 48 hours to achieve competence on SIM (Pino et al., 2010), but MT-Rgl1 takes 24 hours on the same medium (Siqueira Pinto et al., 2017), while A. thaliana takes just one day on CIM (Che et al., 2007). Indeed, due to the lack of our understanding of the genetics and mechanisms involved, the time period required for an explant to acquire competence can only be empirically determined. Importantly, the acquisition of competence by explant cells is not dependent on the type of medium upon which the explant is cultured. To this effect, a transfer of explants between culture media such as RIM, CIM or SIM to any order does not change the amount of time it takes for an explant to generate competent cells. However, after
competence is achieved by the explant, the medium upon which the explant is cultured determines the fate of the explant (Koornneef et al., 1993; Siqueira Pinto et al., 2017). Given that acquisition of competence is an early, common stage shared by the subsequent steps of organogenesis fates, an increased capacity to acquire competence should result in higher organogenic capacity to produce both shoot and roots upon induction by the respective medium (Lombardi-Crestana et al., 2012; Koornneef et al., 1993).

(ii) Induction of competent cells to fate commitment

Induction was first introduced as a “stable phenotypic change that persists in the absence of the initiating stimulus” (Binns & Meins, 1979). During organogenesis, competent cells (or callus containing a mass of competent cells) become committed to a specific organ developmental fate by producing a shoot or root meristem, or non-organogenic callus, through incubation on SIM (with low auxin-to-cytokinin ratio), RIM (with high auxin-to-cytokinin ratio) or CIM (with a balanced auxin-to-cytokinin ratio), respectively. Once these cells become committed (induced) for a particular organ fate, they go on to produce specific organ primordia, even after the inductive signal is removed and the explants are transferred to a basal medium devoid of exogenously supplied hormones.

The minimum time required for explant cells to become induced for a particular organ fate can be experimentally observed by progressively transferring explants incubated on an inductive medium to a hormone-free basal medium (BM). However, the minimum time taken by explants on a specific organ-induction medium (RIM or SIM) prior to transfer to a BM, and still trigger the given organogenesis, will reflect the time taken for both phases, acquisition of competence and induction (Koornneef et al., 1993). As discussed earlier, since the minimum time period required for competence acquisition and induction are relatively stable and additive, we can derive the time required for acquisition of competence by subtracting the least time required on inductive medium after pre-incubation (on CIM or BM, for example), from the total minimum incubation time on a certain inductive medium (on RIM or SIM) without pre-incubation. For example, it was observed that MT-\textit{RG3C} requires a minimum of 3 days on SIM for generating shoot. However, preincubation on CIM reduced this to 2 days, thereby
indicating that time for acquisition of competence is 1 day and the minimum time required for shoot induction on SIM is 3 days (Siqueira Pinto et al., 2017). This experimental strategy was first observed in field bindweed (Earle & Torrey, 1965; Christianson & Warnick, 1983), and subsequent experiments have confirmed in various models, including *A. thaliana* and *S. lycopersicum* (Che et al., 2006; Koornneef et al., 1993; Siqueira Pinto et al., 2017).

(iii) Organ determination and differentiation

During the third and final phase of organogenesis, the committed (induced) cells of the early primordium become increasingly reprogrammed and specialized towards establishing either a shoot or root meristem (Christianson & Warnick, 1983; Kareem et al., 2015; Che et al., 2007). This process does not depend on the influence of the specific type of growth hormone ratio and can be carried out even on hormone-free medium (BM). The process is identical to the regular morphogenesis within shoot and root meristems found in intact plants and will not be discussed beyond this point.

Although the determined and differentiated state of a plant cell is stable even through generations of cellular divisions of the founder cells, it cannot be considered a permanent developmental event. Various experiments demonstrated that cell determination can be reversed and that cells can reacquire competence to be again induced to a different organ fate even during the later stages of plant development (Che et al., 2007; Sugimoto et al., 2011; Atta et al., 2009).

2.5 Acquisition of competence: dedifferentiation or redifferentiation?

The process of acquisition of cell competence converts mature differentiated cells into cells that are stem cell-like in nature, which can actively divide and are not committed to a specific organ fate. Dedifferentiation, which is “a process whereby existing transcriptional and translational profiles are erased or altered in order to allow cells to set a new developmental program” (Fehér et al., 2003), is generally accepted as the mechanism by which the explant’s somatic cells become more competent and stem cell-like during acquisition of competence. In theory, once quiescent somatic cells are
dedifferentiated as pluripotent cells, they can actively divide and produce a callus mass containing competent cells that can then be induced by hormonal signaling pathways to become committed (determined) to a specific organ fate as discussed above.

However, the current understanding of dedifferentiation involves a broad range of molecular changes within a cell and is defined to include cellular changes capable of producing somatic embryos (Elhiti et al., 2013; Motte et al., 2014). Whether or not some or all of these molecular mechanisms are employed during acquisition of competence is not yet known. More recent experiments carried out in Arabidopsis via indirect in vitro organogenesis demonstrated that callus, previously defined as an unorganized, undifferentiated mass of cells that can be induced by hormones, does in fact show a certain degree of differentiated organization that is similar to the genetic identity of the early lateral root primordium (LRP), irrespective of the explant type used (Atta et al., 2009; Sugimoto et al., 2010, 2011). This early callus that mimics early LRP transcriptional footprint originates from the active division of selected founder cells within the pericycle layer close to the xylem poles in root explants, and pericycle-like cells associated with the vascular bundle of aerial organs, such as cotyledons, leaves and petals, which lack a native pericycle layer. Since this is observed with both root and shoot explants during shoot and root organogenesis in whichever combination (Liu et al., 2014), it can be thought of as a common developmental pathway employed by plant cells during the early phases of adventitious organogenesis (Atta et al., 2009; Kareem et al., 2016). This model also highlights the structural and functional significance of the pericycle layer as an extended meristem during adventitious organogenesis, and the pericycle-like cells at the xylem poles, as a cellular platform for understanding acquisition of competence.

The pericycle layer as the site of origin for new founder cells has been confirmed with root pericycle-specific markers, along with Arabidopsis GAL4-GFP (J0121) enhancer trap lines designed to reveal this cell type within the developing root (Laplaze et al., 2005) during in vitro regeneration (Sugimoto et al., 2010). Arabidopsis alf4 mutants (Celenza et al., 1995) are incapable of pericycle cell division and therefore, fail to carry out acquisition of competence and specification of founder cells necessary for lateral root
formation irrespective of increased auxin signaling (DiDonato et al., 2004). Experimental evidence that pericycle-like cells are responsible for callus formation using Arabidopsis alf4 mutants can be further confirmed by observing the more pronounced recalcitrance of cotyledon and petals explants during in vitro organogenesis. Che et al., (2007) results also support this observation, since they reported curbed shoot regeneration in Arabidopsis root explants through selective ablation of pericycle cells using cell-specific expression of diphtheria toxin chain A (DTA, an ADP ribosylase that inhibits translation by targeting eEF-2 factor, thereby causing cell death) under the control of a GAL4/UAS inducible promoter (Laplaze et al., 2005).

2.6 Understanding the molecular mechanisms of shoot regeneration

No molecular markers have so far been designated to designate acquisition of competence by explant cells during in vitro organogenesis (Ikeuchi et al., 2016; Motte et al., 2014). Our research is mainly focused on identifying the early genetic pathways involved with acquisition of competence during in vitro organogenesis by using the tomato model. S. pennellii, the recurrent S. lycopersicum cv. MT, and the introgression line MT-RG3C are ideal contrasting genetic systems to study organogenic regeneration in tomato (Appendix 2 describes the rationale for selection of MT as the model system to study hormone induced organogenesis). Moreover, since hormone mediated signaling mechanisms are evolutionarily well conserved in plant species and many key molecular homologs of A. thaliana have already been discovered and characterized in tomato, we can use prevailing evidence in Arabidopsis to build a theoretical framework of the molecular mechanisms involved in tomato organogenesis (Figure 5). By using the molecular markers extensively characterized in Arabidopsis two-step in vitro regeneration protocol (CIM followed by SIM or RIM) (Valvekens et al., 1988), we can study the molecular genetics involved with acquisition of cell competence followed by organ induction and determination during organogenesis. In tomato, direct SIM incubation of root or cotyledon explants is sufficient to produce competent cells that can then be induced and further determined to form organ primordia during in vitro shoot regeneration (Pino et al., 2010; Siqueira Pinto et al., 2017). However, we will discuss the initial molecular mechanisms involved in both shoot and root regeneration using LRP
development pathway. LRP functions as the default adventitious organogenic pathway (Atta et al., 2009) from which shoot regeneration can be induced using factors that influence either endogenous or exogenously supplied auxin:cytokinin ratio.

2.7 Molecular pathways of in vitro organogenesis

2.7.1 Default in vitro organogenesis may be directed towards producing rhizogenesis

The fundamental molecular mechanism that leads to acquisition of competence by explant cells is still obscure. However, in this section we outline possible molecular pathways and genes that might be involved in the acquisition of cell competence in tomato based on the assumption that early stages of in vitro organogenesis follow an LRP initiation-like developmental pathway seen with Arabidopsis. We know that pericycle cells in root explants or the pericycle-like cells close to the vascular bundle in aerial explants become specified as founder cells during early stages of adventitious organogenesis (Atta et al., 2009). These cells may generate a callus mass during in vitro organogenesis (Sugimoto et al., 2010; Chandler, 2011). This pluripotent stage resembles the early LRP initiation program (Atta et al., 2009; Che et al., 2007; Sugimoto et al., 2010), and the resulting early primordia are competent to produce either root or shoot primordia based on the respective growth medium present during induction (Gordon et al., 2009). Below an effort is made to outline possible molecular pathways by amalgamating both, LRP initiation with the current understanding of pathways during early in vitro organogenesis.

Auxin is a key hormone involved in the process of in vitro organogenesis (Dubrovsky et al., 2008; Dharmasiri et al., 2005b). During in vitro incubation, the explant comes in contact with either exogenously supplied auxin when on an auxin-containing medium, or the endogenously produced auxin. This exposure to high auxin levels can cause pericycle (or pericycle-like) cells to accumulate high levels of this hormone in the cytoplasm. The auxin enrichment of these pericycle founder cells is achieved in Arabidopsis by auxin influx secondary transporters, such as AUXIN-RESISTANT1 (AUX1, gene locus: AT2G38120) (Laskowski et al., 2008), or its three functionally redundant AUX1-like transporters (LAX1, AT5G01240; LAX2,
AT2G21050; LAX3, AT1G77690) (Kakani et al., 2009; Marchant et al., 2002). The reduced lateral root initiation seen with aux/lax loss-of-function mutants supports this argument (Swarup et al., 2001). In contrast, expression of PINFORMED1 (PIN1, AT1G73590) auxin efflux transporter in this early stage has been shown to reduce the formation of early LRP-like structures on hypocotyl explants (Pernisová et al., 2009), possibly due to interference with auxin accumulation within the pericycle cells. The mechanism by which pericycle cells enrich selective auxin transportation and how such a transport network is triggered on explant cells and whether the early stages of this molecular pathway is either hormone dependent or independent is not very well understood yet, but certainly deserves more attention in the light of acquisition of competence.

It must be noted that some plant species show increased capacity to acquire competence, thus higher rates of in vitro organogenesis, even in absence of exogenously supplied growth regulators. This implies a genetic potential for higher endogenous production or an increased sensitivity to endogenous levels of auxin (IAA), in some species (Zhao et al., 2001). A key enzyme of the tryptophan-dependent biosynthesis of auxin is coded by YUCCA genes (Zhao et al., 2001). In tomato, ToFZY (Solyc06g065630) codes for a putative ortholog of the YUCCA flavin monooxygenase (Expósito-Rodríguez et al., 2007, 2011). Much of the early endogenous auxin related pathways related to acquisition of competence have not been clearly established yet. Whether the acquisition of cell competence is a function of increased cytoplasmic levels of auxin or increased sensitivity to auxin-mediated signaling, or both, is therefore an open question as well.

In the absence (or under low levels) of auxin, AUXIN RESPONSE FACTORS (ARFs) are constitutively expressed and remain bound to AUX/IAA transcriptional repressors, thus actively repressing the auxin response pathways. When adequate levels of auxin are perceived by a receptor complex composed by Skp1, Cullin (CUL), and the F-box TRANSPORT INHIBITOR RESISTANCE 1 (TIR1) auxin receptor (Dharmasiri et al., 2005a), the so-called SCF^{TIR1} complex, AUX/IAA repressor proteins are ubiquitinated and further degraded by the proteasome, thus releasing ARFs to bind to
target promoters containing *AUXIN RESPONSE ELEMENTS (AREs)* for transcriptional activation of auxin response pathway. Consistently, overexpression of *TIR1* in Arabidopsis was shown to increase shoot regeneration rates while *tir1* loss-of-function mutants showed decreased rates in relation to the wild type (Qiao et al., 2012). The *TIR1* homolog in tomato (*SITIR1*) that mediates auxin signaling pathways have been functionally characterized (Ren et al., 2011; Ren & Wang, 2016).

In Arabidopsis, accumulation of auxin in the pericycle cells leads to degradation of AUX/IAA28 transcriptional repressor, inducing the expression of *GATA23* (AT5G26930), a zinc-finger transcription factor that specifies founder cells within the pericycle (De Rybel et al., 2010). These founder cells then start dividing, with the auxin-mediated degradation of SLR/IAA14 (AT4G14550) proteins (Fukaki et al., 2002) and release of ARF7 and ARF19. Indeed, NPH4/ARF7 (AT5G20730) and ARF19/ARF11 (AT1G19220) function by inducing the direct transcriptional activation of *LATERAL ORGAN BOUNDARIES-DOMAIN16/ASYMMETRIC LEAVES2-LIKE18 (LBD16/ASL18, AT2G42430)*, which initiates the first asymmetric anticlinal division of founder cells (Goh et al., 2012). Auxin response is required for the initial division and specification of founder cells among candidate pericycle cells (Dubrovsky et al., 2008; Benková et al., 2003). Auxin signaling thus causes pericycle or pericycle-like cells to function as an extended meristem in mature tissues of the plant that contain cells arrested in the G1 phase with potential to divide upon appropriate auxin signaling. After the first anticlinal divisions, the receptor-like kinase *ARABIDOPSIS CRINKLY4 (ACR4/C4, AT3G59420)* is expressed in small daughter cells of the asymmetrically divided pericycle cells induced by the SLR/IAA14-ARF7/ARF19 module (De Smet et al., 2008). The expression of *ACR4* suppresses the proliferation of the nearby pericycle cells (De Smet et al., 2008).

We know that auxin-sensitive modules of gene networks, such as *BODENLOS (BDL/IAA12, AT1G04550)*, the B3 transcription factor *MONOPTEROS (MP, AT1G19850)* and *SHORT HYPOCOTYL 2 (SHY2/IAA3, AT1G04240)* are crucial auxin responsive modules during embryonic root development (Hamann et al., 2002) and as per recent studies also implied in lateral root development in *Arabidopsis* (De Smet et al., 2010). Furthermore, the *HOMEODOMAIN-LEUCINE ZIPPER (HD-ZIP)*
transcription factor HOMEBOX PROTEIN 5 (HB5) has been shown to transcriptionally repress and control the spatial expression of BDL during early stages of organogenesis (De Smet et al., 2013). Another protein, TOPLESS (TPL) has been demonstrated to physically interact with IAA12/BDL to repress auxin signaling (Szemenyei et al., 2008). Eleven TOPLESS gene members were identified in tomato where SlTPL1 (Solyc03g117360) showed most correspondence in spatial and temporal expression as well as sequence similarity (Hao et al., 2014). Given the bimodal nature of auxin response modules (SLR/IAA14 and IAA12/BDL), it is not clear which module is more significant for auxin response during the initial phase of in vitro regeneration. The comparative importance each of these pathways have in the subsequent steps of cellular proliferation of pericycle cells as well as the repression of proliferation in non-specified pericycle cells in the vicinity of organogenic foci that will develop into an organ is not clear as well (De Smet, 2012).

Cytokinins (CK) have also been implicated during acquisition of cell competence in Arabidopsis, with increased CK signaling inhibiting LRP initiation (Li et al., 2006). Cytokinin signaling have been found to selectively block pericycle division of founder cells at the G2-M phase transition checkpoint (Li et al., 2006). Repression of CK signaling is therefore important to initiate asymmetric division of the founder cell. Furthermore, there is evidence that an initial auxin signaling causes organogenic callus to express AHK4 cytokinin receptors in founder cells, which henceforth become primed for an increased CK sensitivity later on in the organogenic process (Gordon et al., 2009).

However, the mechanism by which the tissue defines which, out of the many pericycle cells that accumulated auxin, will become specified as founder cells have not been molecularly revealed yet. Whether this is stochastically determined or guided by a genetic poise of selected cells still needs to be answered. Although experimentally identified as a separate phase, no molecular marker or event have been established yet to signify the end of competence acquisition and the beginning of induction stage. Therefore, it is not possible to define the exact physiological boundary between these phases yet.

To this end, it can be observed that after the initial asymmetric division, the
founder cells and their selected progeny go through several anticlinal and periclinal divisions to produce an emerging mass of callus (Peret et al., 2009). Since callus initiation follows an LRP-like developmental pathway (Atta et al., 2009; Sugimoto et al., 2010), most of the genes previously thought to be lateral root developmental markers are in fact expressed in calli or early LRP-like primordia. These include the homeobox transcription factor \textit{WUSCHEL-RELATED HOMEBOX 5} (\textit{WOX5}, AT3G11260) expressed in pericycle cells (Ditengou et al., 2008); the GRAS transcription factor \textit{SHORT–ROOT} (\textit{SHR}, AT4G37650) expressed in the stele (Lucas et al., 2011); another GRAS transcription factor, \textit{SCARECROW} (\textit{SCR}, AT3G54220), which marks root quiescent cells and endodermis (Roy & Lévesque, 2006); as well as the AP2 transcription factor \textit{PLETHORA1} (\textit{PLT1}, AT3G20840), the auxin efflux carrier \textit{PIN-FORMED} (\textit{PIN1}, AT1G73590), the CLAVATA3/ESR-related 19 receptor ligand (\textit{CLE19}, AT3G24225), and \textit{ROOT CLAVATA HOMOLOG} (\textit{RCH}), which are expressed in the root apical meristem. However, they are different in the initial location and polarity of expression within the primordium structures (Kareem et al., 2015). For example, during regeneration, \textit{WOX5} is expressed within the sub-epidermal layer of early regenerating organ primordia, while \textit{SHR} and \textit{SCR} are both expressed throughout the callus (Atta et al., 2009; Sugimoto et al., 2010). As cell divisions occur and the root marker genes mentioned above are expressed, founder cells lose their pericycle identity, indicating a major shift in cellular program described as the re-differentiation of callus toward a pluripotent state (Atta et al., 2009). The expression of these root markers are associated with \textit{PIN1} expression (Tian et al., 2014; Della Rovere et al., 2016; Motte et al., 2014). In tomato, the PIN1 homolog has been identified as \textit{SlPIN1} (Solyc03g118740: Nakayama et al., 2012), being a good candidate as a regeneration marker in our model system. All in all, tomato has 10 PIN and 5 AUX/LAX genes (Pattison & Catalá, 2012), and how these genes are involved in \textit{in vitro} organogenesis or LRP-mediated root regeneration has not been established.

Spatio-temporal regulation of auxin efflux transporters is essential to establish local auxin maxima in target cells within the developing organogenic primordium (Benková et al., 2003; Gordon et al., 2007). If \textit{PIN1} expression takes place slightly earlier during the stage of founder specification, it may interfere with auxin influx and
establishment of the auxin maxima within founder cells. Supporting this hypothesis, Pernisová et al. (2009) showed that inhibition of PIN expression increased organogenesis. In Arabidopsis, PLETHORA (PLT) 3,5,7 are functionally redundant AP2 transcription factors identified as key regulators of PIN1-mediated rhizogenesis (Hofhuis et al., 2013). A similar effect by these three genes on PIN1 patterning leading to the development of an auxin foci at the tip of organogenic callus and thus, organogenesis was also observed (Kareem et al., 2015). Expression of PLT 3,5,7 genes is up-regulated in actively dividing cells of the explant within the first 5 hours of incubation on CIM and is responsible for downstream activation of PLT 1 and 2. In turn, PLT 1,2 expression is responsible for maintaining the stem cell niche in roots (Kareem et al., 2015). PLT 3,5,7 expression is at first restricted to inner organogenic callus cells and later moves to sub epidermal layers of the proliferating callus. Interestingly, the Arabidopsis plt 3,5,7 triple mutant is completely recalcitrant to shoot regeneration from leaf, cotyledon, hypocotyl explants, which shows the important roles they play early on in the acquisition of competence during organ regeneration. Auxin maxima established at the tip of the organogenic primordia (Benková et al., 2003) may reveal the last functional state that can be shared by both shoot and root primordia development during in vitro organogenesis. Therefore, this stage can be considered a significant developmental milestone that should be better investigated to identify molecular marker candidates responsible for divergence of shoot organogenesis program from the LRP initiation pathway when incubate on SIM.

The LRP initiation pathway will progressively develop lateral root primordia amidst continued auxin signaling. However, the molecular components responsible for this progression amidst sustained auxin signaling have not been yet characterized. We know that root development pathway can be perturbed by increased CK signaling during SIM incubation at this stage of development in order to induce shoot organogenesis (Atta et al., 2009). Therefore, from here on we describe the events that follow SIM incubation and putative molecular markers involved in shoot induction.

2.7.2 Induction of shoot organogenesis with SIM incubation

Cytokinin (CK) is the main hormone responsible for assigning shoot identity to the competent organogenic callus (Motte et al., 2014). Callus shows an early LRP-like
differentiation footprint, as discussed above, and when subjected to high CK levels (e.g., when incubated on SIM), it transitions into a shoot development program leading to the formation of a shoot meristem (Gordon *et al.*, 2009) (Figure 5). During adventitious shoot meristem development, signaling in the early primordium depends on hormone transport and homeostasis, which involves CK biosynthesis, as well as its reversible inactivation/activation and degradation (Zalabák *et al.*, 2013). Understanding metabolism and homeostasis mechanisms of this hormone in the plant cell is of paramount importance for enhancing organogenesis rates in recalcitrant explants. We know that H⁺-ATPase-coupled secondary transporter systems are involved in uptake and accumulation of cytokinin in cells (Cedzich *et al.*, 2008), such as PURINE PERMEASES (PUPs) that transport CK bases and nucleoside precursors, as well as EQUILIBRATIVE NUCLEOSIDE TRANSPORTERS (ENTs) (cf. Motte *et al.*, 2014).

There are two main CK biosynthetic pathways identified to date based on which pathway contributes the isoprenoid moiety to adenine (Frébort *et al.*, 2011). The first, and rate-limiting step of endogenous CK biosynthesis is carried out by phosphate-isopentenyl transferases (IPTs) that add an isoprenoid chain to a nucleoside-phosphate (ATP, ADP or AMP) (Kakimoto, 2001; Takei *et al.*, 2001). The exogenous supply of CK by the SIM can be foregone when IPT is overexpressed, since it fast tracks CK biosynthesis. CK has been found in plants in a variety of structural forms but mainly as cis- and trans-zeatin (Sakakibara, 2006). Cytokinin can also be produced by reinstating inactive molecules in the cell pool, by the decarboxylase activity of LONELY GUY (LOG: (Kurakawa *et al.*, 2007). On the other hand, CYTOKININ OXIDASE (CKX) contributes to its degradation, with its high expression being a major contributor to recalcitrance of shoot organogenesis in many cases (Sriskandarajah *et al.*, 2006). Cytokinins can also be momentarily inactivated for storage by the reversible glycosylation at N7 and N9 positions of the molecule (Hou *et al.*, 2004). While this form is resilient against CKX, it becomes readily available to attain the desired cytokinin levels in the cell, when needed (Frébort *et al.*, 2011).

In *Arabidopsis*, CKs are perceived by ARABIDOPSIS HISTIDINE KINASE (AHK) two-component receptors at the plasma membrane (Hwang & Sheen, 2001).
Cytokinin-bound AHKs initiate a phosphorelay cascade by phosphorylating ARABIDOPSIS HISTIDINE PHOSPHOTRANSFER PROTEINS (AHP). AHPs serve as vehicles to transfer phosphate groups to ARABIDOPSIS RESPONSE REGULATORs (ARRs), thereby activating them in the nucleus. Type-B ARRs activate cytokinin response genes while type-A ARRs function in a CK negative feedback loop. Type-C ARRs are not induced by CK but rather are phosphatases that act by blocking the signaling. The paramount importance of CK signaling in in vitro regeneration has been demonstrated using its signaling cascade components. Higuchi et al. (2004) showed that loss-of-function mutations in AHK receptor genes impaired the formation of organogenic green calli/primordia as well as both, root and shoot development, thus hinting at the initial role of CK on gain of cell competence stage of regeneration. Overexpression of type-B ARRs increased shoot regeneration even in absence of exogenous CK (Hwang & Sheen, 2001) whereas mutations in type-B ARRs reduced shoot regeneration capacity. On the other hand, overexpression of type-A ARRs reduced shoot regeneration rates in root explants while their loss-of-function mutations led to great increases of the capacity to regenerate shoots (Buechel et al., 2010). Kiba et al., (2004) showed that the callus was less sensitive to cytokinin and failed to form green foci when type-C ARRs were overexpressed. Furthermore, Hwang & Sheen (2001) showed that CYTOKININ-INDEPENDENT KINASEs, which mimic AHK kinase domains, were able to regenerate shoot in vitro in the absence of CK.

PLT 3, 5 and 7 redundantly play a dual role both early on during CIM incubation as well as later during SIM induced shoot organogenesis (Kareem et al., 2015). PLT 3,5,7 could be the primary target of CK signaling in that the induced high expression of PLT 3,5,7 by CK leads to shoot organogenesis even in CK-free medium (Kareem et al., 2015). However, PLT 3,5,7 could not fully replace auxin in the same manner, implying that they are not direct targets of auxin signaling. This result could, however be due to the aberrant polarization of auxin signaling, since constitutive rather than native gene expression was used in the study (Kareem et al., 2015). PLT 3,5,7 expression is required for both, initiating LRP-like competent primordium cells as well as subsequently establishing the shoot meristem identity (Kareem et al., 2015). However, this dual role of PLT 3, 5 and 7 in both stages of organogenesis makes them unsuitable specific markers for
characterizing either acquisition of competence or induction phases. When transferred to SIM, PLT 3,5,7 combined expression is initially limited to shoot progenitor cells of organ primordia but subsequently they move towards the sub-epidermal layer of the primordium as it goes through further induction. When explants are incubated directly on SIM supplemented just with CK, without a pre-incubation on CIM, these genes are similarly induced.

PLT 3,5,7 transcriptionally regulate the functionally redundant NAC-domain CUP-SHAPED COTYLEDON (CUC1 and CUC2 AT5G53950) transcription factors. In turn, CUC genes are transcriptional regulators responsible for demarcating the functional space of the shoot meristem (Kareem et al., 2015). During in vitro organogenesis, CUC2 expressing cells start to express PIN1, further facilitating the formation of the auxin maxima at the tip of the organogenic primordium (Gordon et al., 2007). SIM incubation also increases the expression of the partially redundant AP2 transcription factors ENHANCER OF SHOOT REGENERATION (ESR1)/DORNROSCHEN (DRN AT1G12980) and ESR2/DRN-LIKE (DRNL AT1G24590). These are also responsible for increasing PIN1 AT1G73590) and the Ser/Thr kinase of PIN1, PINOID (PID AT2G34650) expression, which together with CUC2, induce the PIN1 polarity shift where auxin is transported to the incipient cells of the auxiliary shoot meristem during shoot primordium development (Gordon et al., 2007; Kleine-Vehn et al., 2009; Friml, 2004; Zhang et al., 2010; Motte et al., 2014). The PIN1 polarity shift is assumed to be the key step which marks the end of shoot induction when the organogenic primordium becomes determined with cells committed to produce the shoot meristem (Zhao et al., 2002).

Simultaneously with PIN1 polarity shift, expression of a whole host of well-established shoot meristem development genes is observed within the newly established organogenic primordium, such as the homeodomain transcription factor WUSCHEL (WUS: Mayer et al., 1998), the KNOX homeobox transcription factor SHOOT MERISTEMLESS (STM: Barton & Poethig, 1993), and the receptor-ligand system CLAVATA 1-3 (CLV1-3: Cary et al., 2002). In tomato, functional homologs for these genes are well characterized: the WUS homolog is coded by Solyc02g083950, the STM
homolog is *Let6/Solyc02g081120*, the *CUC2* homolog is *GOB/Solyc07g062840*, *SLCLV1* is *Solyc04g081590*, *SLCLV2* is *Solyc04g056640*, and *SLCLV3* is *Solyc11g071380* (Xu et al., 2015). STM is involved in SAM initiation and maintaining central zone cells in an undifferentiated state (Barton & Poethig, 1993). WUS functions to maintain the meristematic nature of the shoot primordia while also inducing CLV genes, which together establish the pattern of distinct cell layers of the shoot meristem (Schoof et al., 2000). *CLV3* starts to express at the apex of the shoot primordia, thus restricting WUS expression, which establishes the WUS-CLV negative feedback loop that maintains the shoot meristem (Chatfield et al., 2013). STM and WUS are both required by SAM to maintain stem cells as shown by the high recalcitrance to regenerate shoots in any one of these single mutants (Gallois et al., 2002). Together, these genes are considered *bona fide* markers of shoot determination, which indicates the cellular commitment to the shoot development fate.

2.8 The role of plant hormones during *in vitro* organogenesis in tomato

Most of what we know about hormone interactions on plant organogenesis comes from work on the Arabidopsis model. Below, we review our current understanding on the topic in our tomato model system from studies with the *Rg1* locus from several wild species: *Rg1* from *Solanum peruvianum*, *RG3C* from *S. pennellii*, and *Rg2* from *S. chilense*.

In the cv. Micro-Tom (MT)-introgressed line MT-*Rg1*, higher concentrations in the culture medium of synthetic auxins, such as 2,4-D, or longer incubation periods on auxin-rich media produce negative effects on organogenesis (Lombardi-Crestana et al., 2012). This auxin hypersensitivity-induced recalcitrance was also reported for shoot organogenesis in the cv. MT (Lima et al., 2009). Indeed, in *Arabidopsis*, optimal incubation periods are critical for obtaining high regeneration rates (Valvekens et al., 1988, Che et al., 2007) and prolonged incubations on CIM define the callus fate towards root regeneration. Lombardi-Crestana et al. (2012) reported that the double mutant MT-*Rg1,dgt* (*diageotropica*, which has reduced sensitivity to auxin) showed high organogenesis rates comparable to MT-*Rg1*, whereas MT-*dgt* alone shows high organogenic recalcitrance for both root and shoot. When the first hypothesis was tested by using the auxin sensitive *DR5::GUS* reporter, increased auxin signaling was not
observed (Lombardi-Crestana et al., 2012). Auxin, however, as described above, is also involved in founder cell specification, founder cell proliferation in a callus mass as well as in gaining organogenic competence (cf. Motte et al., 2014).

Boiten et al. (2004) showed that the root regeneration process explained by Koornneef et al. (1993) is independent of exogenous cytokinin (zeatin) in the medium. Furthermore, the endogenous cytokinin levels in MT-Rg1 explants changed only during the latter stages of organ primordia development (fourth week of incubation) (Lombardi-Crestana et al., 2012). This suggests that the culprit gene REGENERATION 1 plays a specific function in the acquisition of competence, possibly during the very early stages of regeneration, upstream of cytokinin signaling.

In an experiment characterizing phenotypes of double mutants, MT-Rg1 was able to reverse the low root and shoot regeneration rate of MT-procera (Lombardi-Crestana et al., 2012). MT-procera is a tomato mutant that constitutively responds to gibberellic acid (GA), even in the absence of this hormone. This is caused by a point mutation in LeGAI (Solyc11g011260) that changes a conserved hydrophobic residue (valine) to a negatively charged amino acid residue (glutamate) in the VHVID domain of this DELLA protein. LeGAI is the only DELLA gene identified in the tomato genome and the mutation changes DELLA DNA-binding domain of the protein (Bassel et al., 2008). DELLA proteins are identified as one of the most important transcriptional repressors of the GA signaling cascade (Locascio et al., 2013). Indeed, GA binds to and inactivates the DELLA repressor activity, thus enabling the transcription of GA-MYB, which functions as a downstream transcription factor that carries out GA signaling. GDI1 and GDI2 are F-box domain GA receptors. They were first identified in Oryza sativa (Sasaki et al., 2003) and subsequently characterized in Arabidopsis as SLEEPY1 (SLY1) (McGinnis et al., 2003) and SNEEZY (SNE/SLY2), which are two functionally redundant homologs (Ariizumi et al., 2011). The GIDx-GA-DELLA complex is targeted to the SCF<sup>GID</sup> complex composed of Skp1, CUL1 and a specific E3 ligase (Kipreos & Pagano, 2000; Shabek & Zheng, 2014) that targets DELLA for degradation by the 26S proteasome complex (Livne et al., 2015). The relationship between GA signaling and acquisition of competence during organogenesis has not been extensively studied. However, we know
that the constitutive GA signaling is associated with an inability to maintain the pluripotent stem cell niche at the SAM in *Arabidopsis* (Jasinski et al., 2005). In tomato, the expression of GA responsive *GAMYB-LIKE1* (*GAMYBL1*) is upregulated in MT-

*procura*, which confirms the increased GA signaling that explains the mutant phenotype. However, in the MT-*Rg1/procera* double mutant, the *GAMYBL1* expression is reduced, while shoot and root *in vitro* regeneration were both comparable to MT-*Rg1* (Lombardi-

Crestana *et al.*, 2012). This relationship hints that *RG1* confers high *in vitro* regeneration and acquisition of competence independently of GA signaling.

The ethylene effect on *in vitro* regeneration depends on the genotype and the explant tissue under consideration (Biddington, 1992). In Arabidopsis, mutants with impaired ethylene sensitivity show decreased levels of shoot regeneration while mutants in which ethylene was constitutively expressed showed higher shoot regeneration rates (Chatfield & Raizada, 2008). In tomato, the *Never Ripe* (*Nr*) is a mutant with impaired ethylene receptor carrying a single amino acid change in the sensor region of the receptor (Wilkinson *et al.*, 1995), which was used to create a double mutant with *Rg1*. The double mutant showed similar rates of shoot and root regeneration to MT-*Rg1*, which indicates that higher propensity to regenerate conferred by the *Rg1* allele is not affected by ethylene insensitivity (Lima *et al.*, 2009).

### 2.9 The possible role of epigenetics on *in vitro* organogenesis

Epigenetic regulatory mechanisms may shed some light to the repression of active genes or activation processes that are associated with organogenic callus formation, which is potentially crucial during the initial phase of competence acquisition. A genome-wide transcriptome analysis in Arabidopsis shoot and root explants identified 1,342 genes that are differentially expressed in callus-forming cells during the early phase of regeneration (Xu *et al.*, 2012). These genes include several transcription factors and epigenetic regulators linked with chromatin reorganization, such as histone modifiers and DNA methylation. Indeed, DNA methylation and histone modification have been implicated in organogenic callus formation in Arabidopsis (Grafi & Avivi, 2004). Studies showed the PRC2 complex, and related histone-3-lysine-27-trimethylation (H3K27me3), induced gene repression that is required for callus formation (He *et al.*, 2012). Furthermore,
histone-3-deacetylases (HDACs) are also necessary for callogenesis (Lee et al., 2016). Increased expression of HDACs, such as AtHDA9, AtHDT4, AtHDT2 and AtHDT1, are associated with increased organogenic callus formation (Lee et al., 2016). Notwithstanding, although an increased expression of histone modifiers is positively correlated with organogenic callus formation, their specific genomic target sites are yet to be discovered. Although similar studies have not yet been carried out in tomato, homologs of these key epigenetic modifiers are present in Solanum genomes, and their functions might be evolutionarily well conserved in the plant lineage. Right now, we can only wonder whether the RG1 activity is involved with epigenetic mechanisms, and the functional studies of the Rg1 locus might reveal the answer. Our current efforts are to take one step in this direction.
CHAPTER THREE
MATERIAL AND METHODS

3.1 Plant material and growth conditions

The MT-RG3C line used was introgressed to the BC6F2 generation, thus being considered a near-isogenic line (NIL). This genotype was obtained from the Laboratory of Hormonal Control of Plant Development, ESALQ-USP (Piracicaba, SP, Brazil), and contains the Rg3C locus from S. pennellii homologous to Rg1 locus from S. peruvianum (Arikita et al., 2013; Siqueira Pinto et al., 2017). Plants were grown in the greenhouse at the Plant and Soil Sciences Division, West Virginia University (USA). The growth conditions were controlled at mean temperature of 25±1°C, 11.5h/13h photoperiod (winter/summer), 250–350 μmol photons m\(^{-2}\) s\(^{-1}\) PAR irradiance, and manual irrigation to field capacity once a day. Seeds were germinated in bulk in 250 mL pots with a 1:1 mixture of commercial potting mix Miracle-Gro (Scotts Miracle-Gro Company) and expanded vermiculite supplemented with 1 g L\(^{-1}\) 10:10:10 NPK and 4 g L\(^{-1}\) dolomite limestone (MgCO\(_3\) + CaCO\(_3\)). Upon appearance of the first true leaf, seedlings of each genotype were individually transplanted to pots containing the soil mix described above, except that NPK supplementation was increased to 8 g L\(^{-1}\).

3.2 In vitro tissue culture

Seeds from S. pennellii, S. lycopersicum cv. Micro-Tom (MT) and the introgressed line MT-Rg3C were surface-sterilized by soaking in 20 mL 30% (v/v) commercial bleach (2.7% NaClO) plus two drops of commercial detergent for 15 min, followed by three rinses with sterile water. The seeds were then germinated on media containing half strength MS salts (Murashige & Skoog, 1962), half strength B5 vitamins (Gamborg et al., 1968), 30 g L\(^{-1}\) sucrose, and 2.3 g L\(^{-1}\) phytagel (Sigma Chemical Co.). Medium pH was adjusted to 5.8 before autoclaving. Approximately 40 seeds were sown in Magenta flasks containing 30 mL of medium. Cultures were sealed with Parafilm sealing and incubated at 25±1°C in the dark for 4 days, followed by 8 days under 16-h photoperiod provided with cool white fluorescent lamp (45 μmol PAR m\(^{-2}\) s\(^{-1}\)).
Cotyledons were excised from 12-day-old (after imbibing) seedlings. The distal and proximal tips were removed, and the cotyledons were divided transversally into two or three pieces (0.3-0.5 cm section). Explants were placed with the abaxial side down, immediately after isolation, onto semi-solid medium composed by MS salts with B5 vitamins, 30 g L$^{-1}$ sucrose, 2.3 g L$^{-1}$ agar (Sigma Aldrich), and 5 µM benzylaminopurine (BAP). Sixteen cotyledonary explants were cultured per Petri dish (90 × 15 mm), with 10 plates per genotype. Plates were sealed with Parafilm and maintained under 16-h photoperiod provided with cool white fluorescent lamp (45 µmol PAR m$^{-2}$ s$^{-1}$ at 25±1°C).

3.3 Expression analysis of Rg1 candidates during in vitro organogenesis using RT-qPCR

Samples for RNA analysis were collected at 4 different time points (days 0, 1, 2, and 6) in three biological replications for each genotype during in vitro shoot organogenesis induction using 12-day old cotyledon explants of *S. pennellii*, *S. lycopersicum* cv. MT and MT-Rg3C genotypes. Samples were snap-frozen in liquid nitrogen and stored at -80°C for RNA extraction.

Incubated explants were macerated using 2.0-mm diameter zirconia beads (2.0-mm diameter, BioSpec Products, #11079124zx) and total RNA was isolated using the mirVana miRNA isolation kit (ThermoScientific, #AM1561) according to the manufacturer's instructions. Extracted RNA was analyzed for quality using 2% agarose gel electrophoresis and quantified in NanoDrop 2000 spectrophotometer (ThermoScientific). The total RNA was treated with TURBO DNA-free kit (ThermoScientific), and quantified again using NanoDrop 2000 before reverse transcription to cDNA with oligo(dT) primer using SUPERSCRIPT III 1ST Strand Synthesis kit (ThermoScientific) according to the manufacturer's instructions.

Quantitative RT-PCR (qPCR) reactions were performed on 96-well qPCR plates in 5 µL total volume using 2X Power SYBR Green PCR Master Mix (Applied BioSystems) fluorescent chemistry, on CFX96 Thermal Cycler System (Bio-Rad Laboratories). Expression data were normalized using the Pfaffl method (Pfaffl, 2001) and the constitutive glyceraldehyde 3-phosphate dehydrogenase (GAPDH, Solyc05g014470) and elongation factor 1-alpha (EF-1α, Solyc06g005060) housekeeping
Expression levels of housekeeping genes, markers for cellular determination (WUSCHEL, Solyc02g083950; TKn2/Let6, Solyc02g081120) and 28 candidate genes within the Rg1 locus were determined in all three genotypes using qPCR primer pairs (Table 2). qPCR primers were designed using primer3 online tool (www.genome.wi.mit.edu/genome_software/other/primer3.html). The primer design parameters and RT-qPCR analysis were guided by Udvardi et al. (2008). qPCR primer pair efficiencies were calculated using LinRegPCR software, v.7.5 (Ramakers et al., 2003).

The expression data were first obtained as quantitation cycle (Cq) values in triplicate technical reps for each of the three biological reps for a given treatment (genotype-collection point combination). Data were analyzed according to (Pfaffl, 2001) using the Equation 1 below, where MT-Day 0 was used as the control for calculating the ratios. Geometric mean of respective Cq values for a given treatment or control was used in respective expressions of the equation.

\[
Pfaffl\ ratio\ (r) = \frac{E^{\Delta C_q}_{GOI}}{E^{\Delta C_q}_{HKG}} = \frac{E^{MT\ day0}_{GOI}}{E^{MT\ day0}_{HKG}} / \frac{E^{Treatment\ GOI}}{E^{Treatment\ HKG}}
\]

Equation 1: (GOI – gene of interest which includes organogenesis markers and Rg1 candidates, HKG – housekeeping genes that are constitutively expressed, Treatment – includes any one of genotype, collection point combination).

3.4 Statistical layout of the experiment and the statistical analysis

Residuals of Pfaffl ratios (r) were tested for normality of distribution before applying the a priori Shapiro-Wilk W test to evaluate goodness of fit (Shapiro & Wilk, 1965). Heteroscedasticity of residuals were also analyzed using the Levene’s test (Levene, 1960). R ratios were normalized using an appropriate transformation protocol (either natural log or square root transformation). The a priori test used a two-way nested analysis of variance (ANOVA) was carried out for each gene of interest in order to test r ratios based on the layout on Figure 6. Genotype (G) and collection point (T), as well
interaction of these main effects (GxT), were evaluated. In the case of significant treatment effects, the mean of relative expression (MRE) for each treatment were compared using Tukey-Kramer, HSD pairwise multiple comparisons test (Tukey, 1949). All data were analyzed using JMP v12 (SAS Institute) on the fit-model platform.

3.5 DNA sequence analysis

The variant call files (VCF) for *S. pennellii* genomic sequence are publicly available at EBI ([http://www.ebi.ac.uk/ena/data/view/PRJEB5235](http://www.ebi.ac.uk/ena/data/view/PRJEB5235)) and were sourced from (Chitwood et al., 2013). VCFs were aligned with tomato reference genome database annotation version SL2.40, using SnpEff software (Cingolani et al., 2012), a variant annotation and effect prediction tool in order to analyze the possible impact of genetic variations (i.e., indels and amino acid changes) within the putative *Rg1* genomic region (2.18 Mb). The original polymorphism list was filtered and processed using “Perl” programming language to identify the medium to high-impact polymorphisms for the candidate genes (The SnpEff and the subsequent Perl coding scripts used are available in the [Appendix 3](#)). The filtered list was then passed through the SIFT online tool at [http://blocks.fhcrc.org/sift/SIFT.html](http://blocks.fhcrc.org/sift/SIFT.html) (Ng & Henikoff, 2003), which evaluates non-synonymous single nucleotide polymorphisms (nsSNPs) and its effect on amino acid substitutions. SIFT uses sequence and/or structure to predict the effect of an amino acid substitutions on protein function in order to derive a score that reflects the relative impact of the substitution (Ng & Henikoff, 2006). Furthermore, NCBI conserved domain analysis for the candidate genes were carried out in order to annotate significant polymorphisms to functional protein motifs and domains, especially in the case of indels.
CHAPTER FOUR
RESULTS AND DISCUSSION

4.1 Narrowing down the $RgI$ locus to two sub regions with 29 candidates

Based on information from the ‘150 Tomato Genome Resequencing Project’ (www.tomatogenome.net); (Kobayashi et al., 2014; Causse et al., 2013), the original genomic region of $RgI$ first introgressed into the cv. MsK8 and mapped by Koornneef et al., (1993) contains 293 genes laying between the morphological genetic markers, CrtR-b (Solyc03g007960 – white flower, $wf$) and PSY1 (Solyc03g031860 - yellow fruit, $r$). These 293 genes (ITAG2.3 release) are genuine candidates for conferring the high regeneration phenotype. Furthermore, a separate study (Azevedo, 2012) used cleaved amplified polymorphic sequence (CAPS) markers to further map $RgI$ in a S. peruvianum x MT mapping population. The study narrowed down the candidate $RgI$ region to a region containing 139 genes between the markers CrtbR-b and P5 (Solyc03g025320) on the short arm of chromosome 3 (ITAG2.3 release).

Independently, (Arikita et al., 2013) was able to characterize a locus (RG3C) in S. pennellii that is orthologous to $RgI$ in S. peruvianum. Chitwood et al. (2013) created 76 S. pennellii introgression (ILs) to cv. M82 background, including those used by Arikita et al. (2013), which is useful to provide a much higher resolution of the genomic region concerned. (Koenig et al., 2013) derived a correlation score system based on phenotypic and transcriptomic data for several vegetative traits, such as leaf complexity, shape and size, as well as cellular traits, such as stomatal density and epidermal cell for the respective genotypes studied. The method facilitated an increased resolution of each IL in order to narrow down the span of each introgressed genomic region. With this approach, IL3-2 and IL3-3 were both identified as non-contiguous introgressions from S. pennellii, as shown in Figure 3. These regions were identified as 3.2a and 3.2b for IL3-2, and 3.3a and 3.3b for IL3-3. With this observation, Chitwood et al. (2013) was able to reallocate unique bins, out of which the bins D-3C.1 and D-3C.2 are significant concerning the $RgI$ locus.

Siqueira Pinto et al. (2017) carried out an in vitro organogenesis assay to analyze
the shoot regeneration capacity of *S. pennellii* introgression lines IL3-2 and IL3-3 using the same cotyledon-based *in vitro* shoot regeneration assay followed by Arikita *et al.* (2013). The genotypes used in this experiment were selected ILs backcrossed into MT. As shown in Figure 2-J, MT-IL3-2 was found to be highly regenerant, while MT-IL3-3 showed the same recalcitrance as the control MT (Siqueira Pinto *et al.*, 2017). Together, these results allow us to assume that the *RgL* allele is located within the distinctive bins D-3C.1 and D3-3C.2 stipulated by Chitwood *et al.* (2013). These two bins together contain 602 genes: D-3C.2 contains Solyc03g025100 to Solyc03g045090 (493 genes) while d-3C.1 contains Solyc03g006910 to Solyc03g008010 (109 genes). Although these experiments individually do not provide significant resolution, combined with the abovementioned regeneration assays (Azevedo, 2012; Siqueira Pinto *et al.*, 2017; Arikita *et al.*, 2013; Koornneef *et al.*, 1987) and bin analyses (Arikita *et al.*, 2013; Chitwood *et al.* 2013), we can narrow down the *RgL* locus to two small regions containing just 29 candidate genes (Table 1; Figure 3). The upper limit of the first region (bin D-3C.1) is demarcated by *CrtR-b* (Solyc03g007960) with its lower limit being Solyc03g008010. The upper limit of the second region (bin D-3C.2) is demarcated by Solyc03g025100, while the lower limit, by the P5 CAPS marker (Solyc03g025320). Altogether, from an initial list of 293 candidate genes involved with the *RgL* phenotype, by using independent genetic analyses and the concept of genetic bins derived from distinct introgression lines, we were able to shortlist just 29 genes, which can now be further explored in more detail.

### 4.2 Gene expression analysis using RT-qPCR and integration with public RNA-Seq metadata

A heatmap for expression of 29 candidate genes and 4 canonical markers of late regeneration during the phase of induction was generated to identify spatio-temporal patterns of expression within the *RgL* locus (Figure 7). Regarding their expression profiles, genes in this genomic region are clustered into main clades. The first cluster contains genes that are constitutively expressed across all organs of the plant, while the other clade consists of genes with more specific expression across different organs and stages of development. With the knowledge that *REGENERATION1* confers increased cell competence in regeneration within the first 24–48 hours on SIM for both root and cotyledon explants (Lombardi-Crestana *et al.*, 2012), we compared this metadata to our
own gene expression results across all candidate genes in the RgI genomic region.

Furthermore, our RT-qPCR time-course dataset revealed based on the significant differential mean relative expression (MRE) within the whole or a fraction of the early time window for competence acquisition, five candidate genes with expression induced during early organogenesis (Solyc03g025110, Solyc03g025190, Solyc03g025250, Solyc03g025260, and Solyc03g025300) (Figure 8). Interestingly, all of the five candidates belong to the second main clade of the RNA-Seq heatmap (Figure 7). The expression patterns and possible molecular mechanisms as they relate to possibly facilitating acquisition of competence are discussed below.

Two of the 28 candidate genes analyzed failed to produce any RT-qPCR amplification (Solyc03g025180 and Solyc03g025240) across any genotype or time point and were excluded from further analysis. The RNA-Seq data revealed that Solyc03g025240 expression does not occur in root, stem or leaf. However, RNA-Seq data for Solyc03g025180 indicate that our qPCR primers were not optimal for this transcript, although it is also likely that this gene was not expressed in the samples used in our experiment. Therefore, it should be noted that the RNA-Seq data used to derive the heatmap corresponds to intact plant organs while we analyzed the expression of cotyledon explants incubated on growth medium.

We identified one gene with constant expression across incubation time and genotypes, Solyc03g025200. This gene is annotated as coding for a protein containing MATE domain (InterPro domain: IPR002528), which is generally involved in transmembrane secondary active transport of diverse molecules, including primary and secondary metabolites.

4.3 Expected gene expression profile of RGI candidate in the genotypes under investigation.

In case the RgI phenotype is due to differential gene expression, we expect the culprit RGI gene to show differential expression within the early time window of in vitro organogenesis during which the acquisition of competence takes place. Any differential expression during day 0 is not indicative of molecular processes that are initiated after
plant material is excised and incubated in SIM (or any other medium such as RIM, CIM). Certainly, \textit{REGENERATION1} alleles from wild tomatoes (e.g., \textit{Rg1} from \textit{S. peruvianum}) is dominant for the regeneration trait, therefore its expression is expected to be activated/upregulated as early as day 1 in \textit{S. pennellii}, which harbors the dominant allele in comparison to MT, which carries the \textit{rg1} recessive allele (Figure 9). This characterization is consistent, whether \textit{RG1} culprit gene function is either directly or indirectly involved in inducing increased competence either as an activator or a repressor of a repressor, respectively. Furthermore, we expect that the \textit{RG1} culprit expression in the heterozygous MT-\textit{RG3C} line to either remain intermediate between the two parental genotypes (MT and \textit{S. pennellii}), or to mimic \textit{S. pennellii} during acquisition of competence (days 1-2).

\textbf{4.3.1 Solyc03g025190 – a putative anthocyanin permease is differentially expressed during acquisition of competence.}

\textit{Solyc03g025190} mean relative expression (MRE) pattern between the three genotypes showed variation across days of incubation on SIM (\(F_{G \times T}=3.6197, p=0.0247\)). While MRE was stable and low in \textit{S. pennellii} across all collection points, MT showed >10-fold increase within just one day and maintained this high level until end of second day (Figure 8). MT-\textit{Rg3C} showed an intermediate MRE during the first day of incubation yet was more similar to MRE levels of MT by the second day on SIM. Variation during the first day of SIM incubation is expected for key genes involved in regeneration competence. RNA-Seq metadata showed the gene to be moderately expressed in leaf as well as portions of the vascular tissue (Matas \textit{et al.}, 2011). This expression pattern is noteworthy, considering the role that pericycle cells play with close proximity to xylem poles in initiating cell division and regeneration labile competent cells as per LRP-mediated shoot regeneration pathway during \textit{in vitro} regeneration.

\textit{Solyc03g025190} is identified as a Multi-antimicrobial transport extrusion (MATE, IPR002528) transmembrane protein possibly controlling the transport of anthocyanins, flavonoids or other metabolites. The 507-amino acid residue gene product contains 12 \(\alpha\)-helical transmembrane regions to anchor it in the cell membrane. Its homolog in \textit{S. pennellii} (\textit{Solanum pennellii} Annotation Release v1.00) contains 507
amino-acid residues and is annotated as a MATE DETOXIFICATION 35-like transporter. In cultivated tomato (S. lycopersicum), it has been named MTP77 permease (Mathews et al., 2003). The effect of anthocyanin accumulation on regeneration has not been investigated.

The Arabidopsis protein ADP1 is a MATE protein localized in endomembrane vesicles that when overexpressed in adp1-D mutants showed signs of altered auxin biosynthesis, yet not intracellular transport, signaling or polar transport (Li et al., 2014). ADP1 was found to reduce precursors [i.e., indole-3-acetonitrile (IAN), indole-3-acetamide (IAM), indole-3-pyruvic acid (IPyA) and indole-3-acetaldehyde (IAAld)] and free auxin (IAA) levels within the cell, thereby implying that it may affect the endogenous auxin biosynthesis. According to our results, auxin biosynthesis should be promoted in S. pennellii due to comparatively low expression of the gene across day 1 and 2. The fact that the competence acquisition is not dependent on exogenous supply of auxin within the first 24-48h is consistent with this hypothesis that wild species have enhanced endogenous biosynthesis of auxin. However, a comparison of auxin signaling using DR5::GUS constructs in MT-Rg1 and MT showed a weaker auxin response signal in MT-Rg1 than MT (Lombardi-Crestana et al., 2012).

We did not observe the characteristic gene expression of RG1 with anthocyanin permease. Nevertheless, what we observed is the opposite where the gene was expressed more in the recalcitrant MT line in comparison to the highly regenerant, S. pennellii. However, as expected MT-RG3C expression remained intermediate (between MT and S. pennellii) on day 1 or was similar to MT on day 2. Therefore, at least at the transcriptional level, the observed expression pattern does not fit the expected pattern of RG1.

4.3.2 Solyc03g025110 and Solyc03g025300 are differentially expressed during acquisition of competence and code for proteins with unknown functions.

Neither of these genes contain known annotated functional protein domains or similar sequences found in other species through BLAST in the NCBI database. Therefore, we are tempted to treat these loci as pseudogenes with conspicuous transcription or, more
Interestingly, that these are genes coding for small peptides with biological activity. Solyc03g025110 showed a pattern in which the expression was very low in *S. pennellii* while in MT it was initially higher, and the expression of MT-*RG3C* was intermediate (*F_{GT}*=9.2393, \(p=0.0003\)) (Figure 8). Solyc03g025110 codes for a very short peptide with 105 amino acid residues. The expression profile of Solyc03g025110 does not fit the gene expression profile expected of *RG1*.

Solyc03g025300 showed variation in MRE among genotypes with modulated expression during incubation on SIM (*F_{GT}=15.1036, \(p<0.0001\)) (Figure 8). However, the predicted 62-aa peptide sequence did not find any match in the BLAST search against the NCBI database. It is interesting to note that the expression of the wild allele was very low or absent and yet showed spurious spikes of expression in the recalcitrant MT on day 0. Although for days 1 and 2, Solyc03g025300 expression in MT-*RG3C* was intermediary between MT and *S. pennellii*, which fits the expected expression pattern for *RG1* in this heterozygous genotype, other two opposing genotypes do not conform to the characteristic expression pattern. Therefore, at least based on the gene expression characterized for these 2 genes coding for short peptides with unknown function, both Solyc03g025110 and Solyc03g025300 can be ruled out of candidacy for the functional *RG1*.

**4.3.3 Solyc03g025250 - A MATE protein that shows differential expression during acquisition of competence**

Solyc03g025250 is another MATE transporter closely related to the JAT2 transporter in *Nicotiana tabacum* as well as the above discussed Solyc03g025190 (MTP77, the putative anthocyanin permease protein - based on unpublished phylogenetic analysis carried out by Adolfo Santos, personal communication). Solyc03g025190 codes for the only MATE protein that has been partially characterized in tomato. It is a putative anthocyanin permease protein with expression activated by the ANT1 MYB transcription factor (Mathews *et al.*, 2003). In tobacco, JAT2 mediates nicotine sequestration to the vacuoles and are exclusively expressed in leaves (Shitan *et al.*, 2014). However, JAT2 is also capable of binding to a wide range of tropane alkaloids (e.g., nicotine, berberine and scopolamine), which are produced when tobacco explants are incubated on various media
supplemented with cytokinin and auxin containing (Khanam et al., 2000; Kim et al., 2008). Indeed, Solanaceae species are avid producers of tropane alkaloids (Jirschitzka et al., 2012), and expression of these secondary metabolites correlates with high root and shoot regeneration rates (Jordan et al., 2006). Additionally, polymorphisms observed in these MATE proteins may indicate adaptations for binding to different ligands or may reflect aberrations leading to loss of function.

Solyc03g025250 showed an expression pattern which spiked >3 fold in S. pennellii on day 1 ($F_{G\times T}=2.8031, p=0.0330$), while MT remained the same until day 2 (Figure 8). The expression in MT-RG3C was intermediate all throughout the period. This expression profile of Solyc03g025110 is in line with expected gene expression profile discussed for a dominant $Rg1$ allele.

4.3.4 Solyc03g025260 - E3 ubiquitin ligase/Retinoblastoma binding protein may be related to related to regulation of the cell cycle in dividing cells during acquisition of competence.

The functional annotation of Solyc03g025260 is related to cell cycle regulation, therefore highly relevant as a potential candidate conferring competence acquisition to explant cells during early organogenesis. The presence of the RING finger domain in the Solyc03g025260 protein is relevant regarding GA physiology. In Arabidopsis, the RING finger domain protein BOTRYTIS SUSCEPTIBLE1 INTERACTOR (BOI) acts similarly to DELLA in suppressing GA signaling (Park et al., 2013). Although the exact mechanism is not clear yet, BOI is believed to target GA-activated transcriptional sites by forming complexes with other important trans-acting proteins. PROCERA (PRO) is the only known DELLA protein coded in the tomato genome, and its loss of function leads to constitutive GA signaling and reduced in vitro regeneration rates of both roots and shoots (Lombardi-Crestana et al., 2012). Therefore, blocking GA response may contribute positively for organogenesis. However, the MT-$Rg1$/procera double mutant showed increased in vitro shoot and root regeneration capacity comparable to MT-$Rg1$, which again confirms a positive role of blocking GA signaling on organogenesis. This mechanism of DELLA-independent GA signal suppression by BOI could explain the observations by which a reduced GAMyb-Like1 expression was also reported, indicating
reduced GA signaling. If \textit{RG1} codes for a BOI-like protein that reduces GA signaling in tomato, we would be able to explain how a reduction in GA signaling can be achieved by the double mutant independent of DELLA function.

Solyc03g025260 MRE varied among genotypes and over time ($F_{G*T}=4.5368$, $p=0.0122$). Although none of the genotypes showed any differential expression on day 0, both MT and MT-\textit{Rg3C} showed approximately three-fold down regulation in MRE on day 1, while \textit{S. pennellii} remained upregulated. \textit{S. pennellii} finally reached low levels achieved by the other genotypes on day 2 of incubation on SIM (Figure 8). In our experiment, \textit{S. pennellii} showed increased MRE of Solyc03g025260 in comparison to other genotypes on day 1. This could lead to increased targeting of cell cycle regulators by a similar mechanism as described above, thereby increasing early cellular proliferation that produces regeneration-competent pericycle or pericycle-like cells amenable for S-phase transition. Although MT and \textit{S. pennellii} genotypes conform to the expected gene expression pattern for \textit{Rg1}, this result is inconsistent with MT-\textit{RG3C} expression pattern of Solyc03g025260 at this time point, which was comparable to MT, where as it is expected to be intermediate or should mimic \textit{S. pennellii}.

\textbf{4.3.5 The majority of candidate genes in the putative \textit{Rg1} region did not show differential expression during acquisition of competence.}

We observed 20 genes with differential MRE on day 0 that failed to show any significant differential expression on days 1 and 2 (Figure 8). Differential MRE on day 0 for these genes can be considered pre-existing and may not reflect the molecular changes that take place during \textit{in vitro} acquisition of competence. However, we noted that Solyc03g025290, which codes for a gene product that has similarity to hydroxyproline-rich glycoprotein family protein, showed differential expression on day 2.

\textbf{4.4 Evaluation of \textit{Rg1} candidates based on gene sequence polymorphism leading to loss-of-function products.}

The putative genomic region for \textit{Rg1} contains 29 annotated gene models, some with largely unknown functions and others with potential roles in gene transcription, translation or involvement in secondary active transport and metabolic processes (Table
We considered the relative impact of point mutations and indels when evaluating changes to the gene sequence, and focused our in-silico sequence analysis on exonic regions and intron-exon junctions where splice sites are located. We did not analyze non-transcribed upstream regions as significant changes to these regions are expected to manifest in gene expression, as analyzed using RT-qPCR.

A 2.18-Mb genomic region within the putative Rg1 region between MT and S. pennellii was analyzed, where a total of 2911 SNPs and indels were identified between S. lycopersicum cv. Heinz 1706 and S. pennellii. We extracted 360 polymorphic sites within the annotated coding regions and splice sites (as per SL2.40 genome annotation) for further analysis. Of these, 178 low-impact synonymous variants were filtered out to highlight 182 missense and other high-impact sequence mutations. The mutations identified included SNPs and indels resulting in non-synonymous amino acid changes, frame-shift as well as gain and loss of stop codons. Eleven out of the 29 genes analyzed were identified with polymorphisms potentially affecting protein function, with 10 genes containing putative deleterious mutations in functional domains (Table 3). Below, we explore the potential impact of these mutants in specific genes.

**Rg1 should carry loss of function mutation in regeneration recalcitrant genotypes.**

Rg1 is a dominant allele (Koornneef et al., 1993; Pino et al., 2010) and confers higher in vitro regeneration rates. Irrespective of whether RG1 functions either directly or indirectly in a single or multiple molecular pathways that are associated acquisition of competence, the gene product should be functional in highly regenerative lines. Indeed, we would then expect that the recessive allele to contain sequence mutation(s) that lead to a loss of function (LoF) of RG1. In this analysis, we expect LoF mutations in regeneration recalcitrant MT genotype in comparison to S. pennellii.

**4.4.1 Solyc03g007960 - \( \beta \)-carotene hydroxylase-2 (CrtR-b2)**

CrtR-b2 is a dominant allele in domesticated tomato species, such as cv. Heinz 1706. It is responsible for the yellow color of petals in the flowers, and catalyzes the reaction of \( \beta \)-carotene conversion to zeaxanthin in the carotenoid biosynthesis pathway. Mutations have a signature recessive white flower phenotype due to substantially (about 80%) lower
xanthophyll synthesis and accumulation in flowers (Galpaz et al., 2006). We identified a deleterious start codon gain variant at the 5’-UTR region of this gene, and three more missense mutations with high probability of affecting protein function (Figure 3). We know that S. pennellii carries yellow flowers (as per flower color of accession collected from Peru, Tomato Genetics Resource Center at UC-Davis: http://tgrc.ucdavis.edu/Data/Acc/AccDetail.aspx?AccessionNum=LA1941). Loss of function mutation in this gene implicates in an obvious phenotype with the recessive white flower phenotype. Therefore, we can assume that mutations in CrtR-b2 sequence mentioned above do not alter the function of the gene product in neither MT nor S. pennellii as they both carry yellow flowers (Table 4).

4.4.2 Solyc03g007970 – putative UDP-N-acetylmuramate-alanine ligase

This gene product contains IPR005758 domain shared by UDP-N-acetylmuramate-alanine ligase proteins. These proteins have not been clearly characterized in plants but in bacteria, they are involved with cytoplasmic peptidoglycan synthesis (Munshi et al., 2013). We identified seven missense mutations (Table 3), but none of them are predicted to significantly alter protein functionality within the conserved domain (Table 4).

4.4.3 Putative novel peptides with unknown functions – identification of many potential high-impact polymorphisms

Several high-impact polymorphisms (missense variations, premature start and codons, in-frame indels, frameshift, and splice variants) were identified in 11 genes that code for small peptides without clear functional annotations: Solyc03g007990, Solyc03g008000, Solyc03g025100, Solyc03g025110, Solyc03g025120, Solyc03g025130, Solyc03g025140, Solyc03g025150, Solyc03g025160, Solyc03g025180 and Solyc03g025300 (Table 3). Since there were no functional domains identified for these proteins, the impact of polymorphisms could not be evaluated. Although some of these genes might code for pseudogenes, while some may represent small/mid-size peptides with biological activity (such as ligands or components of multiprotein complexes). Therefore, the activity of these genes proteins on organogenesis cannot be discarded (Table 4).
4.4.4 Solyc03g008010- PPDE peptidase domain containing protein

Found in viral as well as eukaryotes genomes, this protein contains an IPR008580 domain annotated as putative PPPDE peptidases are thought to play functions by deubiquinrating regulatory proteins (Iyer et al., 2004). Our analysis revealed 4 missense variants, two of which with potential to affect protein function despite being located towards the C-terminus away from the conserved domain. Furthermore, there was a stop gain variant in the middle of the PPDE domain. This means that in S. lycopersicum, the protein sequence is curbed, which may lead to a recessive LoF (Table 3).

4.4.5 Solyc03g025170 - GRAS10 protein sequence changes may not alter protein functionality

Compared to S. lycopersicum cv. Heinz 1706, our analysis revealed six point mutations and one in-frame deletion in S. pennellii, in the coding region of Solyc03g025170 (Table 3), which codes for a GRAS protein (SIGRAS10) (Mayrose et al., 2006). GRAS proteins (IPR005202) are transcription factors mainly involved in hormone signaling and plant development (e.g., SCARECROW, PLETHORA and DELLA), such as root and shoot organogenesis, nodulation, and seed germination (Bolle & Gonzalez, 2015). The GRAS proteins contain leucine heptad repeat I (LHR I), VHIID, leucine heptad repeat II (LHR II), PFYRE and the SAW motifs in their C-terminal end. However, there is considerable sequence variation at the N-terminus, which may confer specific target binding preferences (Tian et al., 2004; Hirsch & Oldroyd, 2009). Out of the six mutations identified, five are located in the GRAS domain but individually identified as tolerated. That said, the combination of individual tolerable mutations might be significant, but unpredictable. The in-frame deletion is located away from the GRAS domain towards the N-terminal end. Therefore, functional changes to SIGRAS10 are not strongly supported by our sequence analysis.

4.4.6 Possible functional changes to MATE transporters

All six MATE transporters within the Rgl1 locus (Solyc03g025190, Solyc03g025200, Solyc03g025210, Solyc03g025230, Solyc03g025240 and Solyc03g025250) are closely related to the JAT2 alkaloid transporter of Nicotiana tabacum (Adolfo Santos, personal
In tobacco, JAT2 mediates nicotine sequestration to the vacuoles and are exclusively expressed in the leaf (Shitan et al., 2014). However, JAT2 is also capable of binding to a wide range of other tropane alkaloids, such as berberine and scopolamine, which are produced when tobacco explants are incubated on various media supplemented with cytokinin and auxin containing (Khanam et al., 2000; Kim et al., 2008). Indeed, Solanaceae species are avid producers of tropane alkaloids (Jirschitzka et al., 2012), and expression of these secondary metabolites correlates with high root and shoot regeneration (Jordan et al., 2006). The polymorphisms observed in these proteins may indicate adaptations for binding to different ligands or may reflect aberrations that lead to LoF.

Solyc03g025190 (MTP77) codes for the only MATE protein that has been partially characterized in tomato. It is a putative anthocyanin permease protein with expression activated by the ANT1 MYB transcription factor (Mathews et al., 2003). Gene ontology and possible gene functions have been discussed in the gene expression analysis section above. The polymorphisms observed in this gene include two missense mutations, one of which is located in the MATE functional domain. However, these changes are tolerated, for protein structure and function as per their SIFT score (Table 3). The other putative MATE sequences showed number of sequence variants that may affect protein structure/function. Solyc03g025200 carries two missense variants with one of them being a Ser to Gly in the MATE domain. Solyc03g025210 also carries two missense variants at position 44 (Leu-Val) and 403(Cys-Arg) both within the MATE domain. Solyc03g025230 shows two missense variants at positions 372 (Arg-Lys) and 411 (Arg-Cys). Solyc03g025240 has one missense variant at 60 (Ala-Thr) and Solyc03g025250 contains one missense variant at 322 (Ile-Val). The impact of these mutations are unknown.

4.4.7 Solyc03g025260 - E3 ubiquitin ligase/Retinoblastoma binding protein

As discussed in gene expression analysis section, Solyc03g025260 is related to cell cycle regulation, and is highly relevant as a potential candidate conferring competence acquisition to explant cells during early organogenesis. The Solyc03g025260 gene contains a single exon encoding a 371-aa peptide in S. lycopersicum. The protein contains
DWNN (IPR014891), CCHC-type Zinc Finger (IPR001878), and RING-type Zinc Finger (IPR013083) domains. This domain architecture and sequence are similar to the retinoblastoma-related (RBR6-like) protein in plants (Gutzat et al., 2012), which is a plant homolog to the mammalian RBBP6 protein (Smit & Sixma, 2014). RBBP6 is a splicing associated protein that plays a key role in induction of apoptosis (Pugh et al., 2006) and regulation of cell cycle in mammals, and apoptosis in Drosophila. It is absent in prokaryotes but is highly conserved across a wide range of eukaryotic species. In humans, RBBP6 has a similar domain structure and was implicated in cellular apoptosis. Overexpression of the human RBBP6 was associated with high proliferation of cervical cancer cell lines by promoting the S-phase through interaction with p53 cell cycle regulation (Moela & Motadi, 2016). However, it should be noted that Solyc03g025260 lacks a protein binding domain (IPR014891) typical of retinoblastoma-like proteins (Pugh et al., 2006), thus it may not play the same role as the human protein.

We identified two disruptive in-frame indels in MT, one of which is a significantly large deletion within the RING-type Zinc Finger domain and a single amino acid residue deletion at the C-terminus. A total of 11 missense variants were identified in MT, one in the DWNN, four in CCHC-type Zinc Finger domain, and rest, except for one towards the C-terminal end of the protein, found within RING-type Zinc Finger domain. Two of these 11 mutations can potentially affect protein function. Given that Rg1 is a dominant allele, these changes indicate a high possibility of a loss of function (LoF) in the MT gene product.

4.4.8 Solyc03g025270 - 36a-like subunit of the Mediator of RNA polymerase II transcriptional machinery

The subunit 36a of Mediator transcriptional co-activator complex (MED36A) of RNA polymerase II relates to more than one functional annotation. Plants use Mediator subunits to regulate transcription of selected regions of the genome (Hemsley et al., 2014). They facilitate the binding of other transcription factors (activation domain) to RNA polymerase II, thereby acting as facilitators and, thus cofactors of the holoenzyme (Conaway & Conaway, 2011). Indeed, Mediator units bind to specific transcription factors and can influence transcription either positively or negatively (Conaway &
The Solyc03g025270 gene product contains an AdoMet-MTase (adenosyl-L-methionine methyltransferase) domain (IPR000692), known to methylate substrates. Within the superfamily of AdoMet-MTases, Solyc03g025270 has the highest sequence similarity to fibrillarin, a component of SnRNP (nucleolar small nuclear ribonucleoprotein) localized to the nucleolus fibrillar region, which plays essential roles in ribosome biogenesis, including pre-rRNA processing and 2′-O-ribose methylation of rRNA and snRNAs (Rakitina et al., 2011). In *Brassica oleracea*, the yeast homologue to NOP1 (Nucleolar Protein 1) rRNA methyltransferase, BoFIB fibrillarin methylates histone H2A (Loza-Muller et al., 2015), thereby combining both of the functions discussed above. Sequence comparisons between both genotypes revealed that Solyc03g025270 contains four high-impact polymorphisms, including a missense mutation, a stop gain, a stop loss and an in-frame indel. The missense mutation was located outside the AdoMet-MTase domain, towards the N-terminus. Complicated set of high impact mutations make it too difficult to specify whether these represent loss of function (LoF) or gain of function (GoF) changes in MT. Nevertheless, this finding is significant, since this protein can potentially coordinate extensive chromatin remodeling necessary for regaining pluripotency potential during the phase of acquisition of cell competence in the early stages of organogenesis.

**4.4.9 Solyc03g025280 - U2 snRNP-associated SURP protein**

Solyc03g025280 gene product includes a conserved RNA-binding domain (IPR012677) with motifs potentially capable of interacting with RNA polymerase II and RNA spliceosome complexes. Although a direct connection of this specific gene with plant organogenesis is not available in the literature, the protein DEAH-Box RNA Helicase (At1g26370) involved in pre-mRNA splicing, which has been demonstrated as important for root initiation in Arabidopsis (Ohtani et al., 2013). We identified seven missense variants, of which at least three SNPs were located within the conserved protein domain, possibly affecting protein function. Furthermore, two mutations, a stop gain and a stop loss, were also identified outside the conserved domains towards the C-terminal end of the protein in *S. pennellii*. Further investigation of Solyc03g025280 sequence aberrations is therefore worthwhile, especially considering early phase of the LRP and its associated...
implication in *in vitro* organogenesis in general.

**4.4.10 Solyc03g025290 - Hydroxyproline-rich glycoprotein**

We identified five missense mutations and one frameshift variation for this gene. Solyc03g025290 shows sequence similarity to hydroxyproline-rich glycoprotein (HRGP) isoform 1, which was characterized in cacao (*Theobroma cacao*) as a cell wall modifier (Verica *et al.*, 2004). HRGPs are a family of proteins comprising extensins, arabinogalactan proteins (AGPs), proline-rich proteins (PRPs) and lectins (Cassab, 1998). Some members, such as extensins are localized in the cell wall and implicated in cell wall extension, thus directly linking them to plant development (Jones *et al.*, 2013). In banana (*Musa spp.* AAA.) HRGPs were identified to play an important role in somatic embryogenesis and regeneration of somatic embryos (Xu *et al.*, 2011). However, the mechanism by which such HRGP enrichment in cell walls affect regeneration has not been characterized. All polymorphisms identified in this protein are located towards the C-terminal end of the main functional domain, with potentially little impact to function.

**4.4.11 Solyc03g025310 - TATEB protein**

This protein is usually targeted to the chloroplasts of plants and contains a Sec-independent protein translocase domain (IPR006312). This domain contains a motif with two arginine residues and was shown to transport folded proteins across the thylakoid membranes of the chloroplast (Berks *et al.*, 2000). We identified 7 missense variants located outside the conserved TATB domain with increased capacity to affect protein structure or function (*Table 3*). Furthermore, a stop gain variant was also found at the Trp22 position of the protein in *S. pennellii*. This could mean that the protein function is lost in *S. pennellii*. However, such pattern does not fit our expectations of LoF in the cultivated tomato, given the dominance of the wild allele.

**4.4.12 Solyc03g025320 - functional changes in a putative A-type BAHD acyltransferase**

Solyc03g025320 codes for a BAHD acyltransferase DCR-like protein, which is a AT-type BAHD acyltransferase containing IPR003480 domain, possibly involved in
transferring acyl groups to a wide range of secondary metabolites, such as flavonoids and other phenolic compounds (Bontpart et al., 2015; Chedgy et al., 2015). They are usually localized in the cytoplasm and contain at least two conserved domains: a HXXXDG domain near the center of the enzyme and a DFGWG motif located near the carboxyl terminus (D’Auria, 2006). Both of these motifs are highly conserved across different types of BAHD proteins and site-directed mutagenesis has shown to largely reduce enzyme activity (Bayer et al., 2004; Suzuki et al., 2001). Both of these motifs are present and conserved in MT as well as S. pennellii. Furthermore, the residue Thr-268, which is functionally significant, is present in both genotypes. In melon, a change of alanine to threonine in this position was sufficient to rescue activity of the naturally non-functional enzyme AAT2 (El-Sharkawy et al., 2005). All other missense mutations were identified as tolerated in terms of structure and function. Importantly, a large in-frame indel in S. pennellii was observed from Glu-224 to Val-233 (residue positions are referent to the Heinz 1706 protein, since it is absent in MT). Motifs that are structurally or functionally significant have not been identified in this region. Although evidence that directly supports AT2-type BAHD acyltransferase activity in plant organogenesis is absent, the tomato SlAT2 (Solyc01g105580) is expressed in glandular trichomes and encodes an acetyl-CoA–dependent acyltransferase enzyme capable of sucrose acetylation in vitro (Schilmiller et al., 2012).
CHAPTER FIVE
CONCLUSIONS AND PERSPECTIVES

5.1 The Rg1 candidate gene shortlist

We were able to confine the original Rg1 locus discovered over two decades ago (Koornneef et al., 1993), containing 293 putative candidates down to 29 candidate genes in two separate sub loci. We capitalized on the preserved genomic synteny across tomato species such as S. pennellii, S. peruvianum and S. lycopersicum within the putative Rg1 genomic region to compare across experiments (Arikita et al., 2012; Chitwood et al., 2013; Siqueira Pinto et al., 2017). Here, moving one step further, by using gene expression characterization and DNA sequence analysis of their coding regions, we were able to eliminate 7 more genes from the 28 candidates: Solyc03g007960, Solyc03g007970, Solyc03g007980, Solyc03g025170, Solyc03g025220, Solyc03g025290 and Solyc03g025310 can be eliminated from the Rg1 candidate list owing to their non-conformity to either gene expression pattern or protein functional characteristics of the dominant Rg1 allele (Table 4). Out of the remaining 21 candidate genes, 11 genes did not have functional annotations. These non-annotated gene sequences are largely sORFs coding for small peptide sequences (sPEP). Recently several sPEPs have been identified in Arabidopsis to be involved in leaf morphogenesis and shoot meristem function (Narita et al., 2004). sPEP have received much attention as functionally important gene products with the advent of innovative molecular techniques, such as ribosome profiling (Ribo-seq), which could shed light to which of these genes are responsible for functional products that are important in acquisition of competence (Andrews & Rothnagel, 2014).

5.2 Confirming the RGl identity and future directions in characterizing molecular function.

Our results provide at the least 10 promising candidate genes for further functional analysis. None of these candidate genes, except for Solyc03g025190 identified as a potential anthocyanin permease, have yet been characterized in tomato. Therefore, it will be important to investigate MT-Rg1 gene expression of these gene candidates using S. peruvianum and S. pennellii (accession LA716) as positive controls. From thereon in,
we propose the molecular cloning and manipulation of these identified genes and of their regulatory regions in order to confirm the identity of the *RG1* gene. Future experiments may include: i) cloning the full cassette (native promoter plus coding region) for each of the genes from a regenerative genotype into a recalcitrant background genotype, such as MT, thereby complementing gene function; ii) selective knockout of gene expression in *Rg1* candidates identified using the CRISPR/Cas9 system in highly regenerant genotypes, such as MT-*Rg1* and MT-*RG3C* to verify loss of regeneration rates. A useful approach could be using an inducible gene expression construct, such as dexamethasone induced gene expression for tomato (Borghi, 2010) in order to investigate the influence of high expression of *RG1* in the regeneration recalcitrant MT background. Such a system will be useful to distinguish between promoter region and coding region LoF mutations.

Once the gene identity had been established, *RG1* will become an early marker for acquisition of competence during *in vitro* organogenesis. Moreover, revealing the *RG1* identity will provide a functional and temporal landmark for future experiments to study molecular pathways in *in vitro* organogenesis. By further identification of molecular homologs of *RG1* in Arabidopsis, for example, we will be able to study the molecular pathways associated, in a system where most developmental mutants exist and are already functionally characterized (Motte et al., 2014). To this effect, just as it is possible to work with Arabidopsis as a model system, we propose the investigation of cellular and subcellular localization of *RG1* product in MT (Campos et al., 2010). We can confirm the involvement of pericycle or pericycle like cell involvement during acquisition of competence by analyzing the expression domain of *RG1* in tomato tissues. *pRG1::GFP* constructs can be used to identify the exact tissue involved in regeneration. This would provide a robust model that can be compared to that of Arabidopsis LRP initiation pathway proposed for organogenesis (Atta et al., 2009). Subcellular localization of the *RG1* gene product can be investigated using the MT model system containing *pRG1::RG1-GFP*, which may help reveal the molecular function of *RG1*.

Once the RG1 identity has been established, we can sequence the exact genomic region across multiple wild and domesticated species that are both regeneration recalcitrant and highly regenerant. This approach will allow us to compare similarities
and identify key sequence polymorphisms that allude to the \( RG1 \) function. By introgressing \( RG1 \) genomic regions from a multitude of highly regenerant wild species, such as \( S. \) peruvianum (Lombardi-Crestana et al., 2012), \( S. \) pennellii (de Siqueira Pinto et al., 2017), and \( S. \) chilense, we can evaluate the best natural alleles that yield high \textit{in vitro} organogenesis to optimize \textit{in vitro} regeneration rates (Lombardi-Crestana et al., 2012; Pino et al., 2010; Siqueira Pinto et al., 2017). That will encourage identification and introgression of \( RG1 \) in other plant species that are commercially important as well.

Various near-isogenic lines (NILs) of hormonal and developmental mutants are available for MT, allowing developmental mutations to be characterized in a common genetic background (Carvalho et al., 2011). The majority of the abovementioned collection is maintained at the Laboratory of Hormonal Control of Plant Development, at ESALQ/USP, University of São Paulo, in Brazil (http://www.esalq.usp.br/tomato). We can more confidently evaluate the \( RG1 \) function and its involvement in hormone signaling pathways by creating double mutants. Since \( RG1 \) has not been studied in mutants for ABA, ethylene, and salicylic acid signaling pathways, these questions will be addressed in future to evaluate whether \( RG1 \) plays a role in either perception or transduction of these signals during the acquisition of cell competence.

The recent work by (Siqueira Pinto et al., 2017) recognized three other loci in tomato (\( RG7C, RG8C \) and \( RG10C \)) that unequivocally are additively involved in the same phase of acquisition of competence. It will be interesting to investigate how \( RG1 \) (\( RG3C \)) correlate with the additive effects of these genes. Once the identity of \( RG1 \) is known, we will be able to demarcate molecularly the temporal frame of early acquisition of competence, and to build gene networks to unravel an answer to the question that baffled molecular geneticists and plant breeders alike for a longtime: \textit{What is the identity of the gene that confers organogenic competence in plants?}
REFERENCES:


Benková E, Michniewicz M, Sauer M, Teichmann T, Seifertová D, Jürgens G, Friml J.


Lavenus J, Goh T, Roberts I, Guyomarc'h S, Lucas M, De Smet I, Fukaki H, Beeckman T,


Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function.


Shitan N, Minami S, Morita M, Hayashida M, Ito S, Takanashi K, Omote H, Moriyama Y,


Figure 1. MT-Rg1 in vitro and ex vitro phenotypes

(A) Adult plant of the tomato cv. Micro-Tom (MT) model system bearing fruits (left). Introgression line of the MT-Rg1 mutant at similar development stage with yellow fruits (due to the r locus from S. peruvianum) and displaying a highly branched phenotype (right). (B) MT-Rg1 seedlings often show bifurcated tricotyledons and (C) thicker petioles. (D-E) Higher in vitro shoot regeneration phenotype of MT-Rg1. Shoot formation in cotyledonary explants obtained from 12-day old seedlings cultivated on SIM (5.0 µM BAP) for 21 days. (F-G) Higher root organogenesis in MT-Rg1 from cotyledonary explants obtained from 12-day old seedlings on RIM (0.4 µM NAA) for 21 days. Illustrations adapted from Lombardi-Crestana et al. (2012).
Figure 2. MT-Rg3C in vitro regeneration phenotype.

(A-F) Shoot regeneration on 12-day old cotyledon explants from different genotypes. The explants were cultured on SIM with 5.0 μM BAP for 21 days. *S. pennellii* cv. LA716 is the species harboring the critical Rg1 regeneration allele which was introgressed into *S. lycopersicum* cv. M82 background to produce the regenerant M82 introgression lines IL-3-2, IL7-1, IL8-3 and IL9-1, which contains *S. pennellii* genomic introgressions from chromosomes 3, 7, 8 and 9, respectively. (G-I) Root regeneration of 12-day old cotyledon explants on RIM containing 0.4 μM NAA for 21 days. All introgressed lines of cv. M82 showed high root regeneration rates. (J) IL3-2 and IL-3-3 introgression fragments were introgressed into the cv. MT to investigate shoot regeneration capacity. IL3-3 was comparable to MT while IL3-2 showed significantly more shoot regeneration in comparison to other genotypes. (K) In vitro shoot regeneration of 12-day old cotyledon explants from the complete set of the *S. pennellii* introgression lines into cv. M82 background was investigated to identify the most important QTLs for shoot regeneration. Error bars represent mean +/- standard error (n=6, with each repetition composed of a Petri dish containing 15 cotyledons). IL3-2, along with other ILs, was identified as significantly superior for shoot regeneration in comparison to cv. M82. The dashed line represents the mean shoot regeneration percentage for cv. M82. (L) Root regeneration capacity for 12-day old cotyledon explants on RIM was investigated for selected *S. pennellii* introgression lines into *S. lycopersicum* cv. Micro-Tom background. IL3-2, along with other ILs, was identified as significantly superior for root regeneration in comparison to cv. M82. Data used in figure are adapted from Arikita et al., (2013) and de Siqueira Pinto et al., (2017)
Figure 3. Rg1 locus on chromosome 3 of tomato genome

(A) The critical locus for shoot and root regeneration, Rg1, was first mapped to the short arm of tomato chromosome 3 by Koornneef et al., (1993), between the PSY1 and Crtb-R morphological markers that spans a genomic region of around 6.2 Mb and contains 336 annotated genes. (B) By using the P5 CAPS marker Azevedo (2012) narrowed the region down to 136 genes spanning a 4-Mb region. (C) The Rg1 locus can be further mapped within just 29 genes by analyzing the differential regeneration capacity in S. pennellii introgressed lines. Sources: Adapted from Supplementary Figure 4 in Chitwood et al. (2013) and Siqueira Pinto et al., (2017)
Explants become **competent** through redifferentiation (genetic network rewiring) and they enter the **determined** state after induction mediated by hormones. Both of these stages are experimentally identifiable phases the regenerating explant goes through before **differentiating** into a specific organ (shoot or root). Adapted from (Lombardi-Crestana *et al.*, 2012).
Figure 5. Molecular pathway of *in vitro* organogenesis.

Molecular pathway for *in vitro* organogenesis follows a lateral root initiation pathway (LRP) where initial developmental stages are shared by both root and shoot organogenesis, irrespective of the starting explant type. PIN1 polarity shift may mark the achievement of shoot determination, Question marks (?) represent pathways, which have not been identified or characterized yet.
A two-way nested analysis of variance (ANOVA) test was carried out for biological replicates (BRs) nested within each treatment. Each treatment contained three BRs consisting of individual in vitro culture plates containing 12 cotyledon explants of a given genotype.

**Figure 6. Statistical layout for analyzing gene expression for *Rg1* candidates.**

A two-way nested analysis of variance (ANOVA) test was carried out for biological replicates (BRs) nested within each treatment. Each treatment contained three BRs consisting of individual in vitro culture plates containing 12 cotyledon explants of a given genotype.
Figure 7. Heatmap of 28 candidate gene expression within the RgI locus region along with tomato homologs of established markers of *in vitro* regeneration.

The gene expression data for the heat map was based on multiple RNA-seq experiments (Detailed in the data source: http://gbf.toulouse.inra.fr/tomexpress/www/projectsTomExpressA.php). The clustering was based on the level of expression and the spatio-temporal pattern of expression. Two main clusters of gene expression can be identified divided by the two main branches of the tree. Genes belonging to the top main branch are more or less constitutively expressed. The genes in lower branch however, shows patterns of expression that is specific to tissue type and developmental stage.
Figure 8. *Rg1* candidate gene expression over the experimental time-course.

**Figure 8.1** Mean relative expression (MRE) for Solyc03g007960.2.1 (BETA CAROTENE HYDROXYLASE2 protein). The differential gene expression between genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) changed over time on SIM (F Gen*time = 3.7018, p = 0.0241). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits α=0.05). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.2** Mean relative expression (MRE) for Solyc03g007980.1.1. (SEPTUM SITE DETERMINING protein). The differential gene expression between genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) changed over time on SIM (F Gen*time = 10.3869, p = 0.0002). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits α=0.05). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.3.** Mean relative expression (MRE) for Solyc03g007990.1.1 (Unknown protein). The differential gene expression between genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) changed over time on SIM (F Gen*time = 6.5899, p = 0.0019). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits α=0.05). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.4** Mean relative expression (MRE) for Solyc03g008000.1.1 (Unknown protein). The differential gene expression between genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) changed over time on SIM (F Gen*time = 3.2647, p = 0.0353). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits α=0.05). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.5** Mean relative expression (MRE) for Solyc03g008010.1.1 (Unknown protein). The differential gene expression between genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) changed over time on SIM (F Gen*time = 3.2647, p = 0.0353). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits α=0.05). MT Day 0 was used as the control for calculating Pfaffl ratio.
Figure 8.6. Mean relative expression (MRE) for Solyc03g008000.1.1 (PPDE PEPTIDASE DOMAIN CONTAINING protein). The differential gene expression between genotypes S. pennelli (PEN) and S. lycopersicum cv. Micro-Tom (MT) changed over time on SIM (F Gen*time = 3.2647, p = 0.0353). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits α=0.05). MT Day 0 was used as the control for calculating Pfaffl ratio.

Figure 8.7. Mean relative expression (MRE) for Solyc03g025100.1.1 (Unknown protein). The differential gene expression between genotypes S. pennelli (PEN) and S. lycopersicum cv. Micro-Tom (MT) changed over time on SIM (F Gen*time = 4.8919, p = 0.0076). MRE remained stable and low for PEN while MT achieved same low levels after a two days lag (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits α=0.05). MT Day 0 was used as the control for calculating Pfaffl ratio.

Figure 8.8 Mean relative expression (MRE) for Solyc03g025110.1.1 (Unknown protein). The differential gene expression between genotypes S. pennelli (PEN) and S. lycopersicum cv. Micro-Tom (MT) changed over time on SIM (F Gen*time = 9.293, p = 0.0003). MRE remained stable and low for PEN while MT achieved same low levels after a two days lag (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits α=0.05). MT Day 0 was used as the control for calculating Pfaffl ratio.

Figure 8.9. Mean relative expression (MRE) for Solyc03g0025120.1.1 (Unknown protein). The differential gene expression between genotypes S. pennelli (PEN) and S. lycopersicum cv. Micro-Tom (MT) changed over time on SIM (F Gen*time = 4.2419, p = 0.0146). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits α=0.05). MT Day 0 was used as the control for calculating Pfaffl ratio.

Figure 8.10. Mean relative expression (MRE) for Solyc03g0025130.1.1 (Unknown protein). The differential gene expression between genotypes S. pennelli (PEN) and S. lycopersicum cv. Micro-Tom (MT) changed over time on SIM (F Gen*time = 10.2128, p = 0.0003). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits α=0.05). MT Day 0 was used as the control for calculating Pfaffl ratio.

Figure 8.11. Mean relative expression (MRE) for Solyc03g0025140.2.1 (Unknown protein). The differential gene expression was based on genotypes S. pennelli (PEN) and S. lycopersicum cv. Micro-Tom (MT) (F Gen = 16.5854, p <0.0001) as well as time on SIM (F time = 23.0395, p <0.0001). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits α=0.05). MT Day 0 was used as the control for calculating Pfaffl ratio.

Figure 8.12. Mean relative expression (MRE) for Solyc03g0025150.2.1 (Unknown protein). The differential gene expression was based on genotypes S. pennelli (PEN) and S. lycopersicum cv. Micro-Tom (MT) (F Gen = 4.4835, p <0.0273) as well as time on SIM (F time = 24.7671, p <0.0001). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits α=0.05). MT Day 0 was used as the control for calculating Pfaffl ratio.

Figure 8.13. Mean relative expression (MRE) for Solyc03g0025160.2.1 (Unknown protein). The differential gene expression was based on genotypes S. pennelli (PEN) and S. lycopersicum cv. Micro-Tom (MT) (F Gen = 21.6018, p <0.0001) as well as time on SIM (F time = 10.6161, p <0.0001). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits α=0.05). MT Day 0 was used as the control for calculating Pfaffl ratio.
different (error bars represent the upper and lower confidence limits $\alpha=0.05$). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.14. Mean relative expression (MRE) for Solyc03g0025170.1.1 (GRAS protein).** The differential gene expression between genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) changed over time on SIM (F $\text{Gen*time} = 16.9282, p < 0.0001$). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits $\alpha=0.05$). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.15. Mean relative expression (MRE) for Solyc03g025190.2.1 (MATE protein).** The differential gene expression between genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) changed over time on SIM (F $\text{Gen*time} = 3.6197, p < 0.0247$). MRE remained stable and low for PEN while MT achieved showed a sudden spike in MRE by day 1 and maintained it for day 2. RG3C showed intermediate MRE on day 1 but reached levels equal to MT by day 2 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits $\alpha=0.05$). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.16. Mean relative expression (MRE) for Solyc03g025200.2.1 (MATE protein).** Genotypes *S. pennelli* (PEN), RG3C and *S. lycopersicum* cv. Micro-Tom (MT) did not show differential gene expression (F $=0.4602, p <0.8681$). MRE remained stable throughout all the collection points studied (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits $\alpha=0.05$). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.17. Mean relative expression (MRE) for Solyc03g0025210.2.1 (MATE protein).** The gene expression differed between genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) (F $\text{Gen} = 4.7535, p < 0.0220$). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits $\alpha=0.05$). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.18. Mean relative expression (MRE) for Solyc03g0025220.2.1 (MATE protein).** The differential gene expression was based on genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) (F $\text{Gen} = 5.6725, p = 0.0123$) as well as time on SIM (F $\text{time} = 20.4872, p <0.0001$). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits $\alpha=0.05$). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.19 Mean relative expression (MRE) for Solyc03g0025230.2.1 (MATE protein).** The differential gene expression was based on genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) (F $\text{Gen} = 11.2162, p =0.0007$) as well as time on SIM (F $\text{time} = 9.5701, p <0.0015$). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits $\alpha=0.05$). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.20. Mean relative expression (MRE) for Solyc03g025250.2.1 (MATE protein).** MRE changed as explants were incubated on SIM (F $\text{time}=3.5951, p =0.0486$), but the differential expression did not depend on genotype *S. pennelli* (PEN), RG3C and *S. lycopersicum* cv. Micro-Tom (MT) (F $=2.8103, p=0.0867$). (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits $\alpha=0.05$). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.21. Mean relative expression (MRE) for Solyc03g025190260.1.1 (Unknown protein).** The differential gene expression between genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) changed over time on SIM (F $\text{Gen*time} = 4.5368, p <0.0122$). MRE remained stable for PEN on day 0 and 1, but reached low levels that MT and RG3C achieved a day later. (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that
are not significantly different (error bars represent the upper and lower confidence limits $\alpha=0.05$). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.22.** Mean relative expression (MRE) for Solyc03g0025270.2.1 (MEDIATOR OF RNA POL II TRANSCRIPTION SUBUNIT 36-LIKE protein). The differential gene expression was based on genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) (F Gen = 5.5290, p = 0.0134), as well as time on SIM (F time = 9.6470, p < 0.0014). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits $\alpha=0.05$). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.23.** Mean relative expression (MRE) for Solyc03g025280.2.1 (U2 snRNP-associated SURP motif-containing protein-like isoform X2 protein). The differential gene expression between genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) changed over time on SIM (F Gen*time = 6.5899, p = 0.0019). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits $\alpha=0.05$). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.24.** Mean relative expression (MRE) for Solyc03g025290.2.1 (HYDROXYPROLINE RICH GLYCOPROTEIN protein (HRGP)). The differential gene expression between genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) changed over time on SIM (F Gen*time = 4.6926, p = 0.0090). Differential MRE between genotypes limited to day 0 (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits $\alpha=0.05$). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.25.** Mean relative expression (MRE) for Solyc03g025300.1.1 (Unknown protein). The differential gene expression between genotypes *S. pennelli* (PEN) and *S. lycopersicum* cv. Micro-Tom (MT) changed over time on SIM (F Gen*time = 15.1036, p < 0.0001). MRE remained stable and low for PEN while MT achieved same low levels after a two days lag (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits $\alpha=0.05$). MT Day 0 was used as the control for calculating Pfaffl ratio.

**Figure 8.26.** Mean relative expression (MRE) for Solyc03g025310.2.1 (TATEB protein). MRE changed as explants were incubated on SIM (F$_{time}$=20.4149, p<0.0001), but the differential expression did not depend on genotype *S. pennelli* (PEN), RG3C and *S. lycopersicum* cv. Micro-Tom (MT) (F=1.8405, p=0.1874). (Tukey, HSD pairwise comparison). Similar letters denote MRE ratios that are not significantly different (error bars represent the upper and lower confidence limits $\alpha=0.05$). MT Day 0 was used as the control for calculating Pfaffl ratio.
Figure 9. Expected Rg1 expression pattern.

Provided that high shoot and root *in vitro* regeneration is conferred by *Rg1* dominant allele, that shows Mendelian segregation, complete loss of function is expected in a homozygous recessive genotype.
Table 1. Functional annotations of Rg1 candidate gene shortlist.

<table>
<thead>
<tr>
<th>Tomato Gene ID</th>
<th>Functional Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Solyc03g007960 Beta-carotene hydroxylase 2</td>
</tr>
<tr>
<td>2</td>
<td>Solyc03g007970 UDP-N-acetylmuramate-alanine ligase</td>
</tr>
<tr>
<td>3</td>
<td>Solyc03g007980 Septum site-determining protein MinD</td>
</tr>
<tr>
<td>4</td>
<td>Solyc03g007990 Unknown Protein</td>
</tr>
<tr>
<td>5</td>
<td>Solyc03g008000 Unknown Protein</td>
</tr>
<tr>
<td>6</td>
<td>Solyc03g008010 PPPDE peptidase domain-containing protein 1</td>
</tr>
<tr>
<td>7</td>
<td>Solyc03g025100 Unknown Protein</td>
</tr>
<tr>
<td>8</td>
<td>Solyc03g025110 Unknown Protein</td>
</tr>
<tr>
<td>9</td>
<td>Solyc03g025120 Unknown Protein</td>
</tr>
<tr>
<td>10</td>
<td>Solyc03g025130 Unknown Protein</td>
</tr>
<tr>
<td>11</td>
<td>Solyc03g025140 Unknown Protein</td>
</tr>
<tr>
<td>12</td>
<td>Solyc03g025150 Unknown Protein</td>
</tr>
<tr>
<td>13</td>
<td>Solyc03g025160 Unknown Protein</td>
</tr>
<tr>
<td>14</td>
<td>Solyc03g025170 GRAS family transcription factor</td>
</tr>
<tr>
<td>15</td>
<td>Solyc03g025180 Unknown Protein</td>
</tr>
<tr>
<td>16</td>
<td>Solyc03g025190 anthocyanin permease</td>
</tr>
<tr>
<td>17</td>
<td>Solyc03g025200 MATE protein</td>
</tr>
<tr>
<td>18</td>
<td>Solyc03g025210 MATE protein</td>
</tr>
<tr>
<td>19</td>
<td>Solyc03g025220 MATE protein</td>
</tr>
<tr>
<td>20</td>
<td>Solyc03g025230 MATE protein</td>
</tr>
<tr>
<td>21</td>
<td>Solyc03g025240 MATE protein</td>
</tr>
<tr>
<td>22</td>
<td>Solyc03g025250 MATE protein</td>
</tr>
<tr>
<td>23</td>
<td>Solyc03g025260 RBR6-Like protein (E3 ubiquitin ligase like)</td>
</tr>
<tr>
<td>24</td>
<td>Solyc03g025270 mediator of RNA polymerase II transcription subunit 36a-like</td>
</tr>
<tr>
<td>25</td>
<td>Solyc03g025280 RNA-binding protein- Nucleotide-binding, alpha-beta plait</td>
</tr>
<tr>
<td>26</td>
<td>Solyc03g025290 Hydroxyproline-rich glycoprotein family protein (HRGP)</td>
</tr>
<tr>
<td>27</td>
<td>Solyc03g025300 Unknown Protein</td>
</tr>
<tr>
<td>28</td>
<td>Solyc03g025310 TATEB protein</td>
</tr>
<tr>
<td>29</td>
<td>Solyc03g025320 Hydroxycinnamoyl transferase</td>
</tr>
</tbody>
</table>
Table 2. RT-qPCR primers used for gene expression analysis.

<table>
<thead>
<tr>
<th>Tomato Gene ID</th>
<th>Functional Annotation</th>
<th>Forward primer sequence 5'-3'</th>
<th>Reverse primer sequence 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solyc03g007960</td>
<td>Beta-carotene hydroxylase 2</td>
<td>ATGGTGTCCTACGATTTTGTGCT</td>
<td>CCTCCGGTTGACCTCTGCTCTGG</td>
</tr>
<tr>
<td>Solyc03g007970</td>
<td>UDP-N-acetylmuramate-alanine ligase</td>
<td>CTTCCCTGATACAGGTGTGGGG</td>
<td>CTGAAGTTATGTCTCCTCTGCTGCTAG</td>
</tr>
<tr>
<td>Solyc03g007980</td>
<td>Septum site-determining protein MinD</td>
<td>ATAGGGGTTTGGGTGGGAAAGGC</td>
<td>GATGATGAAATCCCGGACC</td>
</tr>
<tr>
<td>Solyc03g007990</td>
<td>Unknown Protein</td>
<td>AAGGCTTTGTGGCAAAACTAGG</td>
<td>CCCAGACAGAAAACAAACAGGAG</td>
</tr>
<tr>
<td>Solyc03g008000</td>
<td>Unknown Protein</td>
<td>CGAGTGATGGATGTTCATCTG</td>
<td>TCTTGCAACCCTCCAGTTTC</td>
</tr>
<tr>
<td>Solyc03g008010</td>
<td>PPPDE peptidase domain-containing protein 1</td>
<td>AGCAGGAAACAGCTAAAAACAGAG</td>
<td>CCAGAGTTGCAAGGAGAGTC</td>
</tr>
<tr>
<td>Solyc03g008000</td>
<td>Unknown Protein</td>
<td>AAGCCTCTTGTTGCAAGGTC</td>
<td>CGTAAGGCAAAAGTTCTGTAACTG</td>
</tr>
<tr>
<td>Solyc03g005110</td>
<td>Unknown Protein</td>
<td>GCTGCAAGACACAGTTTTC</td>
<td>ATGGATGCTTCTCTCTGG</td>
</tr>
<tr>
<td>Solyc03g005120</td>
<td>Unknown Protein</td>
<td>CATTTGAGCCCTTTTG</td>
<td>GCTTGATGCAATTAATGAGGTCC</td>
</tr>
<tr>
<td>Solyc03g005130</td>
<td>Unknown Protein</td>
<td>GCAACCCAGTTAAGGTTT</td>
<td>AATGGAGGGCTGAGGTATTTC</td>
</tr>
<tr>
<td>Solyc03g005140</td>
<td>Unknown Protein</td>
<td>AGGAAATTCGGACACGTTATTC</td>
<td>GGCTGCAAGACACAGGTGTTC</td>
</tr>
<tr>
<td>Solyc03g005150</td>
<td>Unknown Protein</td>
<td>TTATGCGCTGTTGAGTTGAC</td>
<td>GAACGAGATCCTCTCTGTCGCGG</td>
</tr>
<tr>
<td>Solyc03g005160</td>
<td>Unknown Protein</td>
<td>CAATGAAGTCAATGCTTTGGCG</td>
<td>GAGCAGTTTCTGAGGTCCAATGCG</td>
</tr>
<tr>
<td>Solyc03g005170</td>
<td>GRAS family transcription factor</td>
<td>TCATGGCTTGCAAATCAGCG</td>
<td>CGAAGCTGCAAATCAATGAC</td>
</tr>
<tr>
<td>Solyc03g005180</td>
<td>Unknown Protein</td>
<td>TGCGTGTTTGTGGCTGCTG</td>
<td>TATCGGCGCCTATCC</td>
</tr>
<tr>
<td>Solyc03g005190</td>
<td>MATE (putative anthocyanin permease)</td>
<td>GAAGCTTTGTTGAGAGGTGATAGCG</td>
<td>GTGTCGTTAACCTCTGTTGTTT</td>
</tr>
<tr>
<td>Solyc03g005200</td>
<td>MATE protein</td>
<td>ATCCCTCTCGGATATCCTCTGT</td>
<td>TCCCCATAGTCCCAGTACTC</td>
</tr>
<tr>
<td>Solyc03g005210</td>
<td>MATE protein</td>
<td>AAGAAGGGTGGACGACACG</td>
<td>CTTGATGCTCAGTCTGATGCTCC</td>
</tr>
<tr>
<td>Solyc03g005220</td>
<td>MATE protein</td>
<td>TGGAATAGAAGGTTGAGACAGCG</td>
<td>CTAATTCTTGGATGCTCCCACATCC</td>
</tr>
<tr>
<td>Solyc03g005230</td>
<td>MATE protein</td>
<td>GATTTGCGACTTGTTGCTG</td>
<td>CAGACGCTGCAACCTTCTTTATCC</td>
</tr>
<tr>
<td>Solyc03g005240</td>
<td>MATE protein</td>
<td>AGGACCTTGCCTGGACATCC</td>
<td>CGAATTGCTCAACCTCTTGG</td>
</tr>
<tr>
<td>Solyc03g005250</td>
<td>MATE protein</td>
<td>TGGAATAGAAGGTTGAGACAGCG</td>
<td>AAGCAGTTTCTGAGGTCTGCG</td>
</tr>
<tr>
<td>Solyc03g005260</td>
<td>RBR6-Like protein (E3 ubiquitin ligase like)</td>
<td>GAGCATGGGTTAGTGATCC</td>
<td>CTCCCCTGAAAGTCTTTGTTG</td>
</tr>
<tr>
<td>Solyc03g005270</td>
<td>mediator of RNA polymerase II</td>
<td>GTCACCTTTGTATCCTCAATCAAAGGC</td>
<td>TFAATTTGCTCCGTTGAGTC</td>
</tr>
<tr>
<td></td>
<td>Gene ID</td>
<td>Description</td>
<td>Forward Sequence</td>
</tr>
<tr>
<td>----</td>
<td>-----------------</td>
<td>--------------------------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>25</td>
<td>Solyc03g025280</td>
<td>transcription subunit 36a-like</td>
<td>ATCAAAAAAAGGAGGCAGGTC</td>
</tr>
<tr>
<td>26</td>
<td>Solyc03g025290</td>
<td>RNA-binding protein- Nucleotide-binding, alpha-beta plaits</td>
<td>ACGRACGTTACTGCTTCTGG</td>
</tr>
<tr>
<td>27</td>
<td>Solyc03g025300</td>
<td>Hydroxyproline-rich glycoprotein family protein (HRGP)</td>
<td>CAATTAACACGTGGCAAAC</td>
</tr>
<tr>
<td>28</td>
<td>Solyc03g025310</td>
<td>TATEB protein</td>
<td>AAGCCAGTCGCAAGAAGC</td>
</tr>
<tr>
<td>29</td>
<td>Solyc02g083950</td>
<td>WUSCHEL</td>
<td>CCATGCATGAAGAGAACATCA</td>
</tr>
<tr>
<td>30</td>
<td>Solyc02g081120</td>
<td>TKn2/Let6</td>
<td>TCCATCGGAATCTCAGAAGC</td>
</tr>
<tr>
<td>31</td>
<td>Solyc05g014470</td>
<td>GAPDH</td>
<td>TTTGTGCGGTAGCAGGTGTC</td>
</tr>
<tr>
<td>32</td>
<td>Solyc04g081490</td>
<td>Beta-tubulin</td>
<td>AAGATGGCATCCACGTGTGT</td>
</tr>
</tbody>
</table>
Table 3. *Rg1* candidate gene short list and polymorphisms seen in comparison to tomato reference genome

Chromosomal positions highlighted in grey are available in MT background as well. Domain names are not mentioned if they do not exist or when changes are not contained within a putative domain.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Chromosome Position</th>
<th>Base change</th>
<th>Amino acid change</th>
<th>Variant description</th>
<th>Functional domains affected</th>
<th>SIFT score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solyc03g007960.2.1</td>
<td>2447910</td>
<td>39T&gt;G</td>
<td></td>
<td>5_prime_UTR_premature_start_codon_gain_variant</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2448081</td>
<td>133G&gt;A</td>
<td>Ala45Thr</td>
<td>missense_variant</td>
<td>Fatty acid hydroxylase superfamily</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
</tr>
<tr>
<td></td>
<td>2448153</td>
<td>205A&gt;G</td>
<td>Thr69Ala</td>
<td>missense_variant</td>
<td>Fatty acid hydroxylase superfamily</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
</tr>
<tr>
<td></td>
<td>2448159</td>
<td>211G&gt;C</td>
<td>Glu71Gln</td>
<td>missense_variant</td>
<td>Fatty acid hydroxylase superfamily</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
</tr>
<tr>
<td>Solyc03g007970.1.1</td>
<td>2457468</td>
<td>41A&gt;G</td>
<td>Glu14Gly</td>
<td>missense_variant</td>
<td></td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
</tr>
<tr>
<td></td>
<td>2457615</td>
<td>188G&gt;A</td>
<td>Gly63Asp</td>
<td>missense_variant</td>
<td></td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00.</td>
</tr>
<tr>
<td></td>
<td>2458496</td>
<td>415A&gt;G</td>
<td>Ile139Val</td>
<td>missense_variant</td>
<td>Mur ligase family, Glutamate ligase domain - murC domain</td>
<td>TOLERATED with a score of 1.00</td>
</tr>
<tr>
<td></td>
<td>2458537</td>
<td>456A&gt;T</td>
<td>Leu152Phe</td>
<td>missense_variant</td>
<td>Mur ligase family, Glutamate ligase domain - murC domain</td>
<td>TOLERATED with a score of 0.14.</td>
</tr>
<tr>
<td>Chromosome</td>
<td>Position</td>
<td>SNP</td>
<td>Amino Acid Change</td>
<td>Type</td>
<td>Domain/Function</td>
<td>Annotation</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>-----</td>
<td>-------------------</td>
<td>------</td>
<td>----------------</td>
<td>------------</td>
</tr>
<tr>
<td>Solyc03g007990.1.1</td>
<td>2470518</td>
<td>156A&gt;G</td>
<td>Ter52Ter</td>
<td>splice_region_variant &amp; stop_retained_variant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solyc03g008000.1.1</td>
<td>2471200</td>
<td>494T&gt;G</td>
<td>Ile165Ser</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>Solyc03g008010.2</td>
<td>2472239</td>
<td>666T&gt;G</td>
<td>His222Gln</td>
<td>missense_variant</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
<td></td>
</tr>
<tr>
<td>Solyc03g007990.1.1</td>
<td>2462783</td>
<td>1312A&gt;C</td>
<td>Thr438Pro</td>
<td>missense_variant</td>
<td>Mur ligase family, Glutamate ligase domain - murC domain</td>
<td>TOLERATED with a score of 0.27</td>
</tr>
<tr>
<td>Solyc03g007990.1.1</td>
<td>2463135</td>
<td>1481C&gt;T</td>
<td>Ala494Val</td>
<td>missense_variant</td>
<td>Mur ligase family, Glutamate ligase domain - murC domain</td>
<td>TOLERATED with a score of 0.24</td>
</tr>
<tr>
<td>Solyc03g007990.1.1</td>
<td>2463781</td>
<td>1603G&gt;A</td>
<td>Asp535Asn</td>
<td>missense_variant</td>
<td>Mur ligase family, Glutamate ligase domain - murC domain</td>
<td>TOLERATED with a score of 0.89.</td>
</tr>
<tr>
<td>Solyc03g008000.1.1</td>
<td>2472342</td>
<td>563A&gt;G</td>
<td>Asn188Ser</td>
<td>missense_variant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solyc03g025100.1.1</td>
<td>6955965</td>
<td>120G&gt;C</td>
<td>Arg40Ser</td>
<td></td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>Solyc03g008010.2</td>
<td>2475767</td>
<td>364A&gt;T</td>
<td>Met122Leu</td>
<td>missense_variant</td>
<td>PPPDE putative peptidase domain</td>
<td>TOLERATED with a score of 0.19</td>
</tr>
<tr>
<td>Solyc03g025100.1.1</td>
<td>6957696</td>
<td>117G&gt;A</td>
<td>Trp39*</td>
<td>stop_gained</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>Solyc03g025100.1.1</td>
<td>6957828</td>
<td>249T&gt;A</td>
<td>Ser83Arg</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>Solyc03g025100.1.1</td>
<td>6957839</td>
<td>260T&gt;C</td>
<td>Val87Ala</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>Solyc03g025120.1.1</td>
<td>6964896</td>
<td>203A&gt;G</td>
<td>Tyr68Cys</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>Solyc03g025120.1.1</td>
<td>6964926</td>
<td>173T&gt;A</td>
<td>Leu58*</td>
<td>stop_gained</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>Position</td>
<td>Chromosome</td>
<td>Position</td>
<td>Description</td>
<td>Type</td>
<td>PSI-BLAST Hits</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>----------</td>
<td>-------------</td>
<td>------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>6965028</td>
<td>Solyc03g025130.1.1</td>
<td>71T&gt;G</td>
<td>Val24Gly</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6965034</td>
<td>Solyc03g025130.1.1</td>
<td>65T&gt;A</td>
<td>Leu22His</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6971210</td>
<td>Solyc03g025130.1.1</td>
<td>137T&gt;C</td>
<td>Ile46Thr</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6971222</td>
<td>Solyc03g025130.1.1</td>
<td>125C&gt;G</td>
<td>Ala42Gly</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6971228</td>
<td>Solyc03g025130.1.1</td>
<td>119T&gt;C</td>
<td>Phe40Ser</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6971238</td>
<td>Solyc03g025130.1.1</td>
<td>109C&gt;T</td>
<td>Pro37Ser</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6971273</td>
<td>Solyc03g025130.1.1</td>
<td>74T&gt;C</td>
<td>Leu25Ser</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6971448</td>
<td>Solyc03g025130.1.1</td>
<td>7C&gt;T</td>
<td>Gln3*</td>
<td>stop_gained</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6976178</td>
<td>Solyc03g025140.1.1</td>
<td>166_167insT</td>
<td>Asn56fs</td>
<td>frameshift_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6977883</td>
<td>Solyc03g025150.2.1</td>
<td>3060_3061insGGC</td>
<td>Arg1020_Gly1021insGly</td>
<td>inframe_insertion</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6978048</td>
<td>Solyc03g025150.2.1</td>
<td>2899T&gt;C</td>
<td>Tyr967His</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6978054</td>
<td>Solyc03g025150.2.1</td>
<td>2893T&gt;C</td>
<td>Ser965Pro</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6978157</td>
<td>Solyc03g025150.2.1</td>
<td>2790T&gt;G</td>
<td>His930Gln</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6978237</td>
<td>Solyc03g025150.2.1</td>
<td>2710G&gt;A</td>
<td>Glu904Lys</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6978273</td>
<td>Solyc03g025150.2.1</td>
<td>2674G&gt;C</td>
<td>Ala892Pro</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6978333</td>
<td>Solyc03g025150.2.1</td>
<td>2614T&gt;G</td>
<td>Leu872Val</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6978365</td>
<td>Solyc03g025150.2.1</td>
<td>2582A&gt;C</td>
<td>Asn861Thr</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6978828</td>
<td>Solyc03g025150.2.1</td>
<td>2321T&gt;C</td>
<td>Ile774Thr</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6978879</td>
<td>Solyc03g025150.2.1</td>
<td>2270C&gt;G</td>
<td>Pro757Arg</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6978934</td>
<td>Solyc03g025150.2.1</td>
<td>2215G&gt;A</td>
<td>Gly739Arg</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6978985</td>
<td>Solyc03g025150.2.1</td>
<td>2164C&gt;G</td>
<td>Arg722Gly</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6979039</td>
<td>Solyc03g025150.2.1</td>
<td>2092_2097delGATCTA</td>
<td>Asp698_Leu699del</td>
<td>inframe_deletion</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6979070</td>
<td>Solyc03g025150.2.1</td>
<td>2079C&gt;G</td>
<td>Cys693Trp</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>6979260</td>
<td>Solyc03g025150.2.1</td>
<td>1889C&gt;T</td>
<td>Pro630Leu</td>
<td>missense_variant</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>Gene ID</td>
<td>SNP</td>
<td>Amino Acid Change</td>
<td>Variant Type</td>
<td>Consensus Function</td>
<td>PSI-BLAST Hits</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>--------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>--------------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>Solyc03g025160.2.1</td>
<td>661_663delGGA</td>
<td>Arg221del</td>
<td>inframe_deletion</td>
<td>No PSI-BLAST hits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solyc03g025170.1.1</td>
<td>1649G&gt;C</td>
<td>Cys550Ser</td>
<td>missense_variant</td>
<td>GRAS Domain</td>
<td>TOLERATED with a score of 0.81.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1255C&gt;A</td>
<td>His419Asn</td>
<td>missense_variant</td>
<td>GRAS Domain</td>
<td>TOLERATED with a score of 0.24</td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Position</td>
<td>Mutation</td>
<td>Type</td>
<td>Domain</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>----------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>--------------------------------</td>
<td></td>
</tr>
<tr>
<td>Solyc03g025180.2.1</td>
<td>7008000</td>
<td>-29A&gt;G</td>
<td>missense_variant</td>
<td>5_prime_UTR_premature</td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7008175</td>
<td>47delT</td>
<td>frameshift_variant</td>
<td></td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7008275</td>
<td>145C&gt;T</td>
<td>missense_variant</td>
<td></td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7008284</td>
<td>154G&gt;A</td>
<td>missense_variant</td>
<td></td>
<td>No PSI-BLAST hits</td>
<td></td>
</tr>
<tr>
<td>Solyc03g025190.2.1</td>
<td>7045146</td>
<td>708G&gt;T</td>
<td>missense_variant</td>
<td>MetE Domain</td>
<td>TOLERATED with a score of 0.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7046655</td>
<td>1480G&gt;A</td>
<td>missense_variant</td>
<td></td>
<td>TOLERATED with a score of 0.50</td>
<td></td>
</tr>
<tr>
<td>Solyc03g025200.2.1</td>
<td>7048077</td>
<td>-89C&gt;T</td>
<td>missense_variant</td>
<td>5_prime_UTR_premature</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7048083</td>
<td>-83A&gt;G</td>
<td>missense_variant</td>
<td>5_prime_UTR_premature</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7048172</td>
<td>7C&gt;G</td>
<td>missense_variant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7048446</td>
<td>281A&gt;T</td>
<td>missense_variant</td>
<td>MatE Domain</td>
<td>TOLERATED with a score of 0.10.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7049846</td>
<td>778A&gt;G</td>
<td>missense_variant</td>
<td>MatE Domain</td>
<td>TOLERATED with a score of 0.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7049900</td>
<td>832C&gt;G</td>
<td>missense_variant</td>
<td>MatE Domain</td>
<td>TOLERATED with a score of 0.62</td>
<td></td>
</tr>
<tr>
<td>Gene Accession</td>
<td>Position</td>
<td>Allele</td>
<td>Amino Acid Change</td>
<td>Domain</td>
<td>Functional Impact</td>
<td>P Landscape Score</td>
</tr>
<tr>
<td>---------------</td>
<td>----------</td>
<td>--------</td>
<td>-------------------</td>
<td>--------</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Solyc03g025210.2.1</td>
<td>7050547</td>
<td>1132T&gt;G</td>
<td>Leu378Val</td>
<td>missense_variant</td>
<td>MatE Domain</td>
<td>TOLERATED with a score of 0.11.</td>
</tr>
<tr>
<td>Solyc03g025210.2.1</td>
<td>7050755</td>
<td>1243A&gt;G</td>
<td>Ser415Gly</td>
<td>missense_variant</td>
<td>MatE Domain</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.01</td>
</tr>
<tr>
<td>Solyc03g025210.2.1</td>
<td>7056384</td>
<td>59A&gt;G</td>
<td>Asp20Gly</td>
<td>missense_variant</td>
<td>MatE Domain/Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.30</td>
</tr>
<tr>
<td>Solyc03g025210.2.1</td>
<td>7056420</td>
<td>95C&gt;A</td>
<td>Ala32Glu</td>
<td>missense_variant</td>
<td>MatE Domain/Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.17</td>
</tr>
<tr>
<td>Solyc03g025210.2.1</td>
<td>7056455</td>
<td>130C&gt;G</td>
<td>Leu44Val</td>
<td>missense_variant</td>
<td>MatE Domain/Na+ driven multidrug efflux pump</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
</tr>
<tr>
<td>Solyc03g025210.2.1</td>
<td>7056887</td>
<td>484G&gt;A</td>
<td>Glu162Lys</td>
<td>missense_variant</td>
<td>MatE Domain/Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.18</td>
</tr>
<tr>
<td>Solyc03g025210.2.1</td>
<td>7057255</td>
<td>852G&gt;T</td>
<td>Met284Ile</td>
<td>missense_variant</td>
<td>MatE Domain/Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.45</td>
</tr>
<tr>
<td>Solyc03g025210.2.1</td>
<td>7057755</td>
<td>1136A&gt;C</td>
<td>Tyr379Ser</td>
<td>missense_variant</td>
<td>MatE Domain/Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 1.00</td>
</tr>
<tr>
<td>Solyc03g025210.2.1</td>
<td>7057921</td>
<td>1207T&gt;C</td>
<td>Cys403Arg</td>
<td>missense_variant</td>
<td>MatE Domain/Na+ driven multidrug efflux pump</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00.</td>
</tr>
<tr>
<td>Solyc03g025210.2.1</td>
<td>7058001</td>
<td>1287G&gt;A</td>
<td>Met429Ile</td>
<td>missense_variant</td>
<td>MatE Domain/Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 1.00.</td>
</tr>
<tr>
<td>Solyc03g025210.2.1</td>
<td>7058193</td>
<td>1387C&gt;T</td>
<td>Gln463*</td>
<td>stop_gained</td>
<td>MatE Domain/Na+ driven multidrug efflux pump</td>
<td></td>
</tr>
<tr>
<td>Solyc03g025220.2.1</td>
<td>7063814</td>
<td>712A&gt;G</td>
<td>Asn238Asp</td>
<td>missense_variant</td>
<td>MatE Domain/Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.24</td>
</tr>
<tr>
<td>SNP</td>
<td>Position</td>
<td>Change</td>
<td>Effect</td>
<td>Function</td>
<td>Annotation</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>--------</td>
<td>------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------</td>
<td></td>
</tr>
<tr>
<td>7064324</td>
<td>1042G&gt;A</td>
<td>Ala348Thr</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.70</td>
<td></td>
</tr>
<tr>
<td>7064354</td>
<td>1072A&gt;T</td>
<td>Thr358Ser</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.07</td>
<td></td>
</tr>
<tr>
<td>7064691</td>
<td>1327G&gt;C</td>
<td>Val443Leu</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.42</td>
<td></td>
</tr>
<tr>
<td>Solyc03g025230.2.1</td>
<td>7067224</td>
<td>53C&gt;T</td>
<td>Ala18Val</td>
<td>missense_variant</td>
<td>TOLERATED with a score of 0.78</td>
<td></td>
</tr>
<tr>
<td>7067752</td>
<td>358G&gt;A</td>
<td>Ala120Thr</td>
<td>missense_variant</td>
<td>MatE Domain</td>
<td>TOLERATED with a score of 0.19</td>
<td></td>
</tr>
<tr>
<td>7068915</td>
<td>1042C&gt;A</td>
<td>Leu348Ile</td>
<td>missense_variant</td>
<td>MatE Domain</td>
<td>TOLERATED with a score of 0.38</td>
<td></td>
</tr>
<tr>
<td>7068988</td>
<td>1115G&gt;A</td>
<td>Arg372Lys</td>
<td>missense_variant</td>
<td>MatE Domain</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.05</td>
<td></td>
</tr>
<tr>
<td>7069193</td>
<td>1231C&gt;T</td>
<td>Arg411Cys</td>
<td>missense_variant</td>
<td>MatE Domain</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.02</td>
<td></td>
</tr>
<tr>
<td>7069247</td>
<td>1285A&gt;G</td>
<td>Thr429Ala</td>
<td>missense_variant</td>
<td>MatE Domain</td>
<td>TOLERATED with a score of 0.11</td>
<td></td>
</tr>
<tr>
<td>7069275</td>
<td>1313A&gt;Gj</td>
<td>His438Arg</td>
<td>missense_variant</td>
<td>MatE Domain</td>
<td>TOLERATED with a score of 0.40</td>
<td></td>
</tr>
<tr>
<td>7069394</td>
<td>1318A&gt;T</td>
<td>Thr440Ser</td>
<td>missense_variant</td>
<td>MatE Domain</td>
<td>TOLERATED with a score of 0.05</td>
<td></td>
</tr>
<tr>
<td>7069436</td>
<td>1360G&gt;A</td>
<td>Ala454Thr</td>
<td>missense_variant</td>
<td>MatE Domain</td>
<td>TOLERATED with a score of 1.00</td>
<td></td>
</tr>
<tr>
<td>Solyc03g025240.1.1</td>
<td>7075989</td>
<td>111C&gt;A</td>
<td>Phe37Leu</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.45</td>
</tr>
<tr>
<td>7076030</td>
<td>152G&gt;A</td>
<td>Arg51Lys</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 1.00</td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Position</td>
<td>Mutation</td>
<td>Type</td>
<td>Domain/Function</td>
<td>Impact</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>----------</td>
<td>----------</td>
<td>-----------------</td>
<td>-----------------------------------------------------</td>
<td>---------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Soly03g025250.2.1</td>
<td>49g&gt;A</td>
<td>Val17Ile</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
<td></td>
</tr>
<tr>
<td>7076056</td>
<td>178&gt;C&gt;A</td>
<td>Ala60Thr</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.05</td>
<td></td>
</tr>
<tr>
<td>7076119</td>
<td>241&gt;C&gt;G</td>
<td>Ile18Val</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 1.00</td>
<td></td>
</tr>
<tr>
<td>7076237</td>
<td>359&gt;T&gt;A</td>
<td>Phe120Tyr</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.34.</td>
<td></td>
</tr>
<tr>
<td>7076819</td>
<td>758&gt;C&gt;T</td>
<td>Ser253Leu</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 1.00</td>
<td></td>
</tr>
<tr>
<td>7076848</td>
<td>787&gt;C&gt;A</td>
<td>Val263Ile</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.39.</td>
<td></td>
</tr>
<tr>
<td>7077089</td>
<td>952&gt;C&gt;A</td>
<td>Val318Ile</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.62.</td>
<td></td>
</tr>
<tr>
<td>7077113</td>
<td>976&gt;C&gt;A</td>
<td>Ala326Thr</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.11.</td>
<td></td>
</tr>
<tr>
<td>7077479</td>
<td>1170&gt;C&gt;T</td>
<td>Glu390Asp</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.54.</td>
<td></td>
</tr>
<tr>
<td>7078513</td>
<td>49&gt;C&gt;A</td>
<td>Val17Ile</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
<td></td>
</tr>
<tr>
<td>7079270</td>
<td>488&gt;T&gt;C</td>
<td>Val163Ala</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.38</td>
<td></td>
</tr>
<tr>
<td>7080210</td>
<td>964&gt;C&gt;A</td>
<td>Ile322Val</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
<td></td>
</tr>
<tr>
<td>7080585</td>
<td>1247&gt;C&gt;T</td>
<td>Ser416Phe</td>
<td>missense_variant</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.18</td>
<td></td>
</tr>
<tr>
<td>Gene ID</td>
<td>Position</td>
<td>Chromosome</td>
<td>Variant Type</td>
<td>Nucleotide Change</td>
<td>Protein Effect</td>
<td>Score</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>------------</td>
<td>--------------</td>
<td>------------------</td>
<td>---------------</td>
<td>-------</td>
</tr>
<tr>
<td>Solyc03g025260.1.1</td>
<td>7080888</td>
<td>1372C&gt;A</td>
<td>missense_variant</td>
<td>Leu458Ile</td>
<td>MatE Domain/ Na+ driven multidrug efflux pump</td>
<td>TOLERATED with a score of 0.16.</td>
</tr>
<tr>
<td>Solyc03g025260.1.1</td>
<td>7083313</td>
<td>59G&gt;A</td>
<td>missense_variant</td>
<td>Arg20His</td>
<td>DWNN Domain</td>
<td>TOLERATED with a score of 0.21</td>
</tr>
<tr>
<td>Solyc03g025260.1.1</td>
<td>7083580</td>
<td>326T&gt;A</td>
<td>missense_variant</td>
<td>Val109Glu</td>
<td>RING Zn finger domain</td>
<td>TOLERATED with a score of 0.22</td>
</tr>
<tr>
<td>Solyc03g025260.1.1</td>
<td>7083619</td>
<td>365A&gt;T</td>
<td>missense_variant</td>
<td>Lys122Ile</td>
<td>RING Zn finger domain</td>
<td>TOLERATED with a score of 0.08</td>
</tr>
<tr>
<td>Solyc03g025260.1.1</td>
<td>7083638</td>
<td>385_386ins GTAATA GCA</td>
<td>disruptive_inframe_insertion</td>
<td>Ser128_Lys 129insSerAsnSer</td>
<td>RING Zn finger domain</td>
<td></td>
</tr>
<tr>
<td>Solyc03g025260.1.1</td>
<td>7083721</td>
<td>467C&gt;G</td>
<td>missense_variant</td>
<td>Pro156Arg</td>
<td>RING Zn finger domain</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.03.</td>
</tr>
<tr>
<td>Solyc03g025260.1.1</td>
<td>7083730</td>
<td>476C&gt;T</td>
<td>missense_variant</td>
<td>Thr159Met</td>
<td>RING Zn finger domain</td>
<td>TOLERATED with a score of 0.15</td>
</tr>
<tr>
<td>Solyc03g025260.1.1</td>
<td>7083753</td>
<td>499C&gt;G</td>
<td>missense_variant</td>
<td>Arg167Gly</td>
<td>Zinc knuckle (Zn binding domain)</td>
<td>TOLERATED with a score of 0.83.</td>
</tr>
<tr>
<td>Solyc03g025260.1.1</td>
<td>7083774</td>
<td>520A&gt;G</td>
<td>missense_variant</td>
<td>Asn174Asp</td>
<td>Zinc knuckle (Zn binding domain)</td>
<td>TOLERATED with a score of 0.24.</td>
</tr>
<tr>
<td>Solyc03g025260.1.1</td>
<td>7083960</td>
<td>706A&gt;G</td>
<td>missense_variant</td>
<td>Lys236Glu</td>
<td>RING Zn finger domain</td>
<td>TOLERATED with a score of 0.54</td>
</tr>
<tr>
<td>Solyc03g025260.1.1</td>
<td>7084014</td>
<td>760A&gt;T</td>
<td>missense_variant</td>
<td>Met254Leu</td>
<td>RING Zn finger domain</td>
<td>TOLERATED with a score of 1.00</td>
</tr>
<tr>
<td>Solyc03g025260.1.1</td>
<td>7084068</td>
<td>814G&gt;A</td>
<td>missense_variant</td>
<td>Glu272Lys</td>
<td>RING Zn finger domain</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.05</td>
</tr>
<tr>
<td>Solyc03g025260.1.1</td>
<td>7084280</td>
<td>1029_1031delATC</td>
<td>disruptive_inframe_deletion</td>
<td>Ser344del</td>
<td>RING Zn finger domain</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
</tr>
<tr>
<td>Solyc03g025270.2.1</td>
<td>7085439</td>
<td>741T&gt;A</td>
<td>stop_gained</td>
<td>rp247*</td>
<td>AdoMet_Mtases super family</td>
<td></td>
</tr>
<tr>
<td>Variant ID</td>
<td>Change Type</td>
<td>Change Description</td>
<td>Change Location</td>
<td>Protein Change</td>
<td>Change Type Description</td>
<td>Enzyme Family</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>7086402</td>
<td>690T&gt;C</td>
<td>Ter230Cysext*?</td>
<td></td>
<td></td>
<td>stop_lost</td>
<td>AdoMet_Mtases super family</td>
</tr>
<tr>
<td>7087620</td>
<td>74G&gt;A</td>
<td>Arg25Gln</td>
<td></td>
<td>missense_variant</td>
<td>AdoMet_Mtases super family</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00.</td>
</tr>
<tr>
<td>7087639</td>
<td>49_51delGAG</td>
<td>Gly17del</td>
<td></td>
<td>inframe_deletion</td>
<td>AdoMet_Mtases super family</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00.</td>
</tr>
<tr>
<td>7100908</td>
<td>362T&gt;C</td>
<td>Leu121Ser</td>
<td></td>
<td>missense_variant</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
<td></td>
</tr>
<tr>
<td>7100943</td>
<td>397A&gt;G</td>
<td>Lys133Glu</td>
<td></td>
<td>missense_variant</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.01</td>
<td></td>
</tr>
<tr>
<td>7108755</td>
<td>970C&gt;A</td>
<td>Leu324Met</td>
<td></td>
<td>missense_variant</td>
<td>SWAP Domain Surp module</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.02</td>
</tr>
<tr>
<td>7109875</td>
<td>295G&gt;A</td>
<td>Gly432Glu</td>
<td></td>
<td>missense_variant</td>
<td>VHS, ENTH, and ANTH domain superfamily</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.03</td>
</tr>
<tr>
<td>7110685</td>
<td>1774G&gt;A</td>
<td>Ala592Thr</td>
<td></td>
<td>missense_variant</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.04</td>
<td></td>
</tr>
<tr>
<td>7111088</td>
<td>2177T&gt;C</td>
<td>Ile726Thr</td>
<td></td>
<td>missense_variant</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.05</td>
<td></td>
</tr>
<tr>
<td>7111298</td>
<td>2387C&gt;T</td>
<td>Thr796Met</td>
<td></td>
<td>missense_variant</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.06</td>
<td></td>
</tr>
<tr>
<td>7113022</td>
<td>2659A&gt;T</td>
<td>Lys887*</td>
<td></td>
<td>stop_gained</td>
<td>cwf21 domain</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
</tr>
<tr>
<td>7113041</td>
<td>2678A&gt;C</td>
<td>Ter893Serex1*?</td>
<td></td>
<td>stop_lost</td>
<td>cwf21 domain</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
</tr>
<tr>
<td>Gene Accession</td>
<td>Chromosome</td>
<td>Position</td>
<td>Type</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>------------</td>
<td>-----------</td>
<td>-----------------</td>
<td>--------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solyc03g025290.2.1</td>
<td>7114634</td>
<td>1412_*12d</td>
<td>frameshift_variant&amp;stop_lost</td>
<td>Ter471fs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7114713</td>
<td>1349T&gt;G</td>
<td>missense_variant</td>
<td>Val450Gly</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7114914</td>
<td>1148G&gt;A</td>
<td>missense_variant</td>
<td>Arg383Lys</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7115013</td>
<td>1049A&gt;G</td>
<td>missense_variant</td>
<td>Glu350Gly</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7115049</td>
<td>1013C&gt;G</td>
<td>missense_variant</td>
<td>Ala338Gly</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7115245</td>
<td>817G&gt;T</td>
<td>missense_variant</td>
<td>Val273Leu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solyc03g025300.1.1</td>
<td>7123525</td>
<td>97delA</td>
<td>frameshift_variant</td>
<td>Asn33fs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7123585</td>
<td>39G&gt;C</td>
<td>missense_variant</td>
<td>Arg13Ser</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solyc03g025310.2.1</td>
<td>7131727</td>
<td>65G&gt;A</td>
<td>stop_gained</td>
<td>Trp22*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7131741</td>
<td>79T&gt;A</td>
<td>missense_variant</td>
<td>Tyr27Asn</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7131801</td>
<td>139T&gt;A</td>
<td>missense_variant</td>
<td>Leu47Ile</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7131978</td>
<td>316T&gt;A</td>
<td>missense_variant</td>
<td>Ser106Thr</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7131984</td>
<td>316T&gt;A</td>
<td>missense_variant</td>
<td>Lys108Gln</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromosome</td>
<td>Position</td>
<td>Mutation</td>
<td>Type</td>
<td>Impact on Protein Function</td>
<td>Score</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>----------</td>
<td>-----------------</td>
<td>----------------------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>7134895</td>
<td>629A&gt;C</td>
<td>Tyr210Ser</td>
<td>missense_variant</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7139180</td>
<td>1027T&gt;C</td>
<td>Ser343Pro</td>
<td>missense_variant</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7139188</td>
<td>1035A&gt;T</td>
<td>Gln345His</td>
<td>missense_variant</td>
<td>AFFECT PROTEIN FUNCTION with a score of 0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7141853</td>
<td>790G&gt;A</td>
<td>Ala264Thr</td>
<td>missense_variant</td>
<td>Omega-hydroxypalmitate O-feruloyl transferase</td>
<td>TOLERATED with a score of 1.00</td>
<td></td>
</tr>
<tr>
<td>7141861</td>
<td>782G&gt;A</td>
<td>Gly261Glu</td>
<td>missense_variant</td>
<td>Omega-hydroxypalmitate O-feruloyl transferase</td>
<td>TOLERATED with a score of 0.66</td>
<td></td>
</tr>
<tr>
<td>7141941</td>
<td>669_698del</td>
<td>Glu224_Val233del</td>
<td>missense_variant</td>
<td>Omega-hydroxypalmitate O-feruloyl transferase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7143721</td>
<td>363T&gt;A</td>
<td>His121Gln</td>
<td>missense_variant</td>
<td>Omega-hydroxypalmitate O-feruloyl transferase</td>
<td>TOLERATED with a score of 0.31</td>
<td></td>
</tr>
<tr>
<td>7143729</td>
<td>355G&gt;A</td>
<td>Asp119Asn</td>
<td>missense_variant</td>
<td>Omega-hydroxypalmitate O-feruloyl transferase</td>
<td>TOLERATED with a score of 0.37</td>
<td></td>
</tr>
<tr>
<td>7143764</td>
<td>320A&gt;C</td>
<td>Asp107Ala</td>
<td>missense_variant</td>
<td>TOLERATED with a score of 0.28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX 1

Heritability of the high regeneration capacity in tomato and comparisons across related species

The highly regenerant *Rg1* allele within the short arm of chromosome 3 was first identified in *S. peruvianum* introgression lines in cv. MsK8. However, many wild species of tomato, such as *S. peruvianum* (Koornneef *et al.*, 1993), *S. pennellii* (Arikita *et al.*, 2012), *S. pimpinellifolium* (De Faria *et al.*, 2002) and *S. chilense* (Satoh *et al.*, 2000) are also highly regenerant, while most cultivated varieties of *S. lycopersicum* are largely recalcitrant. This suggests that high regeneration and associated traits are shared by tomato wild genotypes although they may display allelic variation in the same locus.

Interbreeding cultivated tomato species with their wild relatives produce viable progeny (Stevens & Rick, 1986) which attests for the high degree of synteny and genetic homology in this clade (The Tomato Genome Consortium, 2012). For instance, as revealed by bacterial artificial chromosome coupled with fluorescent *in situ* hybridization (BAC-FISH), *S. peruvianum*, *S. pennellii* and other wild *Solanum* species varied from each other by isolated small inversions in chromosomes 5, 6, 7, 9, 10, 11 and 12 (Szinay *et al.*, 2012). This close syntenic relationship allows generating viable hybrid lines segregating for a rich natural allelic variation provided by related wild species (Eshed & Zamir, 1994), such as in the case of *Rg1* allele and associated high regeneration phenotype in MT-*Rg1* (*Figure 1*). Furthermore, as shown by Szinay *et al.* (2012), synteny has been well preserved on chromosome 3, thus enabling us to compare the *Rg1* locus across domesticated and wild species.

(Satoh *et al.*, 2000) identified an *Rg1* allele in *S. chilense* cv. PI128644 located on chromosome 3 by analyzing linkage between high shoot regeneration from root explants with markers based on random amplified polymorphic DNA (RAPD) and RFLP. The allele named *Rg1*-2 showed Mendelian segregation in BC1F2 hybrids between the recalcitrant *S. lycopersicum* cv. KOT, revealing to be a dominant allele in line with what had been demonstrated for *S. peruvianum* (Koornneef *et al.*, 1993). Although this finding does not improve the resolution of the genomic span for *Rg1* allele, it does confirm its importance for shoot regeneration across tomato species.

Large-scale experiments using linkage mapping and phenotypic assays for regeneration have
identified that higher in vitro regeneration capacity is a polygenic (quantitative) trait in tomato, with key quantitative trait loci (QTLs) possibly distributed across several chromosomes (Koornneef et al., 1993; Trujillo-Moya et al., 2011; Arikita et al., 2013). However, the significance of the QTL on chromosome 3 has been reiterated in a recent analysis using BC1 population of tomato wild species S. pennellii cv. PE47 and S. lycopersicum cv. Anl27. The experiment found at least six significant QTLs, including the locus on chromosome 3, associated with high shoot in vitro regeneration in leaf explants (Trujillo-Moya et al., 2011). The physical marker and regeneration phenotype linkage analysis identified that QTLs involved in in vitro regeneration were harbored on chromosomes 1, 3, 4, 7 and 8, concurring with previous findings (Koornneef et al., 1993).

Indeed, Solanum pennellii cv. LA716 is another desert dwelling wild species of tomato from Peru that has been documented to show high in vitro regeneration capacity (Arikita et al., 2012). A study using 46 introgression lines (ILs) of S. pennellii into S. lycopersicum cv. M82 identified four key ILs (3-2, 7-1, 8-3 and 10-2) that are involved in high shoot and root regeneration of cotyledon explants (Arikita et al., 2013). These ILs also showed higher shoot regeneration capacity with 60-80% of explants regenerating shoot (Figures 2A-L). In order to be consistent with nomenclature used in a previous study exploring tomato fruit color for S. pennellii ILs, bins carrying critical regeneration loci identified using these IL population were named as RG3C, RG7H, RG8F, and RG10F (Liu et al., 2003). These loci were then further introgressed into the MT background and the high in vitro regeneration capacity was confirmed using the same in vitro regeneration assay for shoot and root using cotyledon explants (de Siqueira Pinto et al., 2017). Importantly, the RG3C bin includes the Rg1 locus previously identified by morphological markers in (Koornneef et al., 1993).
**APPENDIX 2**

**Micro-Tom (MT) as a model system to study in vitro regeneration**

Regeneration in tomato can be conveniently studied in the dwarf cultivar ‘Micro-Tom’. Solanum lycopersicum cv. Micro-Tom (MT) is a miniature tomato variety first developed for ornamental purposes (Scott & Harbaugh, 1989). It harbors three mutant alleles: the *dwarf* (Bishop *et al.*, 1996), *miniature* (Meissner *et al.*, 1997) and *self-pruning* (Pnueli *et al.*, 1998). The alleles *dwarf* (*d*) and *miniature* (*mnt*) contribute to the small MT size while the dominant allele *self-pruning* (*D*) contributes to its determinate growth. The small size (~15cm in height) conferred by these alleles makes it easy to cultivate MT under limited greenhouse space. Short turn-around time (70-90 days from sowing to fruit ripening), as well as a high amenability to *Agrobacterium tumefaciens* mediated transformation, with >80% transformation efficiency when using cotyledon explants (Pino *et al.*, 2010), makes MT an attractive model for plant molecular genetics, development and physiology (Meissner *et al.*, 1997; Campos *et al.*, 2010; Carvalho *et al.*, 2011). Regarding regeneration rates, MT can be considered recalcitrant, with in vitro regeneration rates of about 60% when using cotyledon as explant (Lombardi-Crestana *et al.*, 2012).

The alleles that MT harbors do not hamper its potential as a model system to study plant development, since they exert minimum pleiotropic influence in comparison to typical tomato varieties (Campos *et al.*, 2010). The MT genome was fully sequenced (Kobayashi *et al.*, 2014b) shortly after the whole genome sequencing of *S. lycopersicum* cv. ‘Heinz 1706’ (International Solanaceae (SOL) Genome Project Consortium, 2012) became publicly available, which has made its well annotated 950 Mb, diploid genome distributed across 12 chromosomes, publicly available for researchers (https://solgenomics.net/). Various near-isogenic lines (NILs) of hormonal and developmental mutants are available for MT, allowing developmental mutations to be characterized in a common genetic background (Carvalho *et al.*, 2011). The majority of the abovementioned collection is maintained at the Laboratory of Hormonal Control of Plant Development, at ESALQ/USP, University of São Paulo, in Brazil (http://www.esalq.usp.br/tomato).
APPENDIX 3
The SnpEff and the subsequent Perl coding scripts
#install snpEff and then run this on your commandline withi the folder that contains your vcf files

java -jar snpEff.jar -v -fi intervals.bed -no-downstream -no-intergenic -no-intron -no-upstream -no-utr SL2.40 RF_074_SZAXP1008753-79.vcf > test.solanum12.ann.vcf

#perl script for filtering out the duplicate lines of SNP information from the vcf files

#!/usr/bin/perl

my $filename = "'/Users/grad/test.solanum11.ann.txt";

my $encoding = ":encoding(UTF-8)";

my $handle = undef;

open($handle, "< $encoding", $filename) or die "$0: cant open $filename for reading: $!";

$match =0; #Just a variable that does not equal the loci of polymorphism

while ($line = <$handle>)

{@fields = split /	/, $line;

$field2 =$fields[1];

if($match != $field2) {

print "$line\n";

}

$match = $field2;

};

close INPUT;

#simple grep function used to filter header information and include information pertinent to gene models

at test.solanum12.ann.vcf |grep -i "Solyc"