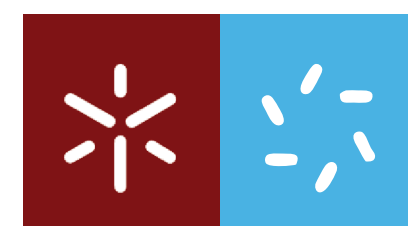




Gonçalo Emanuel Elias da Silva

**Mining and characterization of the candidate genes for distorted segregation in chromosome 4 of tomato.**

**Universidade do Minho**  
Escola de Ciências







**Universidade do Minho**

Escola de Ciências

Gonçalo Emanuel Elias da Silva

**Mining and characterization of the  
candidate genes for distorted segregation  
in chromosome 4 of tomato.**

Tese de Mestrado

Mestrado em Biologia Molecular, Biotecnologia e  
Bioempreendedorismo em Plantas

Trabalho efetuado sob a orientação de

**Professor Doutor Antonio Granell Richart**

**Professora Doutora Teresa Lino-Neto**

## DECLARAÇÃO

**Nome:** Gonçalo Emanuel Elias da Silva

**Endereço electrónico:** goncaloedasilva@gmail.com

**Telefone:** 917183863

**Número do Bilhete de Identidade:** 14399673

**Título tese:** "Mining and characterization of the candidate genes for distorted segregation in chromosome 4 of tomato"

**Orientador(es):** Professor Doutor Antonio Granell Richart e Professora Doutora Teresa Lino-Neto

**Ano de conclusão:** 2017

**Designação do Mestrado:** Mestrado em Biologia Molecular, Biotecnologia e Bioempreendedorismo em Plantas

Gonçalo Emanuel Elias da Silva, cartão do cidadão número 14399673, residente na rua Virgílio Ferreira N.º5, Montemor-o-Novo, Portugal, declaro que, de acordo com a lei em vigor, não é permitida a reprodução de qualquer parte desta tese.

Universidade do Minho, \_\_\_/\_\_\_/\_\_\_\_\_

Assinatura: \_\_\_\_\_



## **Agradecimentos**

Este último ano, apesar de ter sido um dos mais stressantes e cansativos da minha vida foi sem duvida o mais compensador. Sinto que cresci profissionalmente, mas mais do que isso foi uma experiencia de vida extremamente enriquecedora. Nada teria sido possível sem o apoio de diversas pessoas e a quem tenho de agradecer.

Ao excelentíssimo reitor da Universidade do Minho e à administração do Instituto de Biología Molecular y Celular de Plantas por me permitirem realizar esta tese.

Ao Doutor Antonio Granell Richart, quero agradecer toda a confiança, apoio e oportunidades que me concedeu durante este último ano.

Ao Doutor Antonio Monforte Gilabert, obrigado por toda a orientação, disponibilidade e ensinamentos que me transmitiu.

À Doutora Clara Pons Puig, o mais profundo obrigado por toda a dedicação, paciência e apoio prestado durante a realização do projeto e elaboração da tese. Não poderia ter pedido uma melhor supervisora e tudo o que fez por mim nunca será esquecido.

A todos no laboratório, Maria José, Carlos, Soledad, Pablo, Sara, Joan, Sílvia, Diego, Eugenio, e de uma maneira geral a todos no IBMCP e na UPV que contribuíram direta ou indiretamente para este trabalho, obrigado por toda a entreaajuda durante esta experiencia.

Ao Doutor Rafael Fernández Muñoz e a toda a equipa do Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora" em Algarrobo-Costa, Málaga pela colaboração neste projecto.

Espero ter tirado o maior partido desta experiência e sei que tive a sorte de aprender com os melhores, pessoas que admiro profissional e pessoalmente.

À Professora Doutora Teresa Lino-Neto e Professora Doutora Ana Cunha, pela ajuda, conselhos e atenção que me deram.

À minha família, em especial aos meus pais, obrigado por serem sempre pacientes e pelo apoio incondicional mesmo com a distancia que nos separava.

Ao meu irmão por me ajudar a ver o lado positivo mesmo com tudo parecia mais negro.

Ao Alberto, Sónia, Paula e toda a família por tudo o que fizeram por mim e por me fazerem sentir em casa. Sinto que ganhei mais uma família.

E aos meus amigos, obrigado pelos bons momentos, embora que com o passar do tempo sejam cada vez mais escassos sei que estarão lá sempre que precisar.

Esta tese foi realizada no Instituto de Biología Molecular y Celular de Plantas (IBMCP), em Valência, Espanha e contou com fundos financeiros do Consejo Superior de Investigaciones Científicas (CSIC).

A mobilidade foi financiada pelo programa Erasmus Placement.

## **“Mining and characterization of the candidate genes for distorted segregation in chromosome 4 of tomato”**

### **Abstract**

*Solanum lycopersicum* L. genetic variability was drastically diminished by successive genetic bottlenecks induced by the domestication process. The wild species of tomato, *S. pimpinellifolium*, is a small red-fruited plant native to Peru and is assumed as an ancestor species of the domesticated *S. lycopersicum*. One of the strategies to induce genetic variability to cultivated tomato is the development of introgression lines (ILs) containing a single segment of a donor wild genome in the genetic background of an elite tomato cultivar. In 2014, a genomic library of ILs that incorporates variability from the *S. pimpinellifolium* accession TO-937 in the genetic background of *S. lycopersicum* cultivar “MoneyMaker” was developed. During the development of the IL collection, a region on the distal portion of chromosome 4 showed a segregation distortion (SD) favouring TO-937 alleles in detriment of “MoneyMaker” alleles. Recently, the SD region was mapped to a 39Kb region of chromosome 4 containing seven gene annotations. The preliminary studies to assert gametic, post gametic and/or zygotic indicated that the SD was most probably caused by post-gametic or zygotic selection and it was a sex-independent phenomenon. The present study aims to characterize the genes included in the SD region and to propose a possible mechanism for the SD. Expression profile analysis by qRT-PCR and sequencing of genomic and transcriptomic sequences indicated a strong expression in the reproductive tissues of the two Heat-Shock Protein (HSP) genes contained in the SD region. Haplotyping of reciprocal and self-pollinating crosses between the SD haplotypes and “MoneyMaker” gave new insights about the gametic and zygotic character of the SD. The analysis of natural sequence variations of the SD region revealed this region diverged in wild tomato accessions. Additionally, a reverse genetic approach was initiated to assess if the HSPs are the cause of the SD using the GoldenBraid 3.0 standard assembly to create *Agrobacterium*-mediated transformation vectors, two CRISPR/Cas9 expression cassettes for the silencing of the HSP genes, and 3 expression cassettes.

## **“Mining and characterization of the candidate genes for distorted segregation in chromosome 4 of tomato”**

### **Resumo**

A variabilidade genética de *Solanum lycopersicum* L. foi drasticamente diminuída por sucessivos efeitos de gargalo genéticos induzidos pelo processo de domesticação. A espécie de tomate selvagem, *S. pimpinellifolium*, é uma pequena planta de frutos vermelhos nativas do Peru e é assumida como um ancestral do *S. lycopersicum* domesticado. Uma das estratégias para induzir variabilidade genética ao tomate cultivado é o desenvolvimento de linhas de introgressão (ILs - introgression lines) que contêm um único segmento do genoma selvagem no genoma de um cultivar de elite. Em 2014, foi desenvolvida uma biblioteca genômica de ILs que incorpora variabilidade da acessão TO-937 de *S. pimpinellifolium* no genoma do cultivar "Moneymaker" de *S. lycopersicum*. Durante o desenvolvimento da IL, uma região na porção distal do cromossoma 4 revelou uma segregação distorcida (SD) favorecendo os alelos TO-937 em detrimento dos alelos "Moneymaker". Recentemente, a região SD foi mapeada para uma região de 39Kb do cromossoma 4 contendo sete genes. Os estudos preliminares para afirmar o carácter gamético, pós-gamético e/ou zigótico indicaram que a SD provavelmente é causada pela seleção pós-gamética ou zigótica e é um fenómeno independente do sexo. O presente estudo tem como objetivo a caracterização dos genes incluídos na região SD e um possível mecanismo de SD. A análise de perfil de expressão por qRT-PCR e sequenciação do genoma e transcriptoma indicou uma elevada expressão em tecidos reprodutores de dois genes contidos na região genômica da SD que codificam *Heat-Shock Proteins* (HSP). Haplotipagem de cruzamentos recíprocos e auto-cruzamentos entre os haplótipos de SD e "Moneymaker" revelou novas pistas acerca do carácter gamético e zigótico da SD. A análise da variação natural da região SD revelou uma significativa diversidade em acessões de tomate selvagem. Além disso, usando o sistema de clonagem GoldenBraid 3.0 para criar vetores de transformação mediada por *Agrobacterium*, duas cassetes de expressão CRISPR / Cas9 para o silenciamento dos genes HSP e 3 cassetes de expressão para foram desenvolvidas para futura aplicação.

## **Table of contents**

<b>OBJECTIVES:</b>	<b>1</b>
<b>LITERATURE REVIEW</b>	<b>2</b>
<b>1. Botanical description of Solanum tomato species.</b>	<b>3</b>
1.1 Solanum lycopersicum var. Heinz and Solanum lycopersicum var. "MoneyMaker".	4
1.2 Solanum pimpinellifolium accession TO-937.	5
<b>2. Economic importance of the cultivated tomato.</b>	<b>5</b>
<b>3. Origin and domestication of tomato.</b>	<b>6</b>
<b>4. Phenotypic and genetic variability of wild tomato species as a source of variability for cultivar improvement.</b>	<b>7</b>
<b>5. Exploiting genetic variability available for tomato cultivar improvement by using marker assisted selection (MAS) and genetic maps of quantitative trait loci (QTL).</b>	<b>8</b>
<b>6. Single nucleotide polymorphism markers in the tomato genome and high throughput genotyping.</b>	<b>10</b>
6.1 High resolution melting - HRM	11
6.2 Kompetitive Allele Specific PCR genotyping assay - KASP™	13
<b>7. S. pimpinellifolium accession TO-937 library of introgression lines (IL) in the genetic background of tomato cultivar "Moneymaker".</b>	<b>14</b>
<b>8. Segregation Distortion.</b>	<b>15</b>
8.1 Influencing physiological and genetic factors and causes of segregation distortion.	16
8.2 SD in tomato species	17
8.3 Mapping QTL markers linked to distorted segregation.	19
8.4 Segregation distortion in TO-937 derived populations.	20
<b>9. Defining landmarks for stages in tomato flowering and fruit development.</b>	<b>21</b>
<b>10. Genome engineering in crops and the CRISPR/Cas9 system.</b>	<b>24</b>
<b>11. Synthetic Biology and standard assembly.</b>	<b>29</b>
11.1 GoldenBraid 3.0.	31
<b>MATERIALS AND METHODS</b>	<b>37</b>
<b>1. Plant material.</b>	<b>38</b>
<b>2. Seed disinfection and germination.</b>	<b>39</b>
<b>3. DNA extraction, evaluation and quantification.</b>	<b>39</b>
<b>4. HRM – high resolution melting genotyping method.</b>	<b>40</b>
<b>5. KASP primer design and procedure.</b>	<b>41</b>

<b>6. Assessment of SD mechanism.</b>	<b>42</b>
6.1 Self-pollination crosses.	42
6.2 Reciprocal cross MM x IL4-4H.	43
6.2.1 Evaluation of the progeny of the cross MM x IL4-4H.	43
<b>7. Sequencing of the genes within the SD region-</b>	<b>43</b>
<b>8. Analysis of candidate gene expression using qRT-PCR comparative C<sub>t</sub> analysis.</b>	<b>45</b>
8.1 Sample collection for total RNA extraction.	45
8.2 RNA extraction, purification and quantification.	46
8.3 cDNA synthesis.	47
8.4 Primer design for qRT-PCR	47
8.5 Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR) of gene expression between plants carrying different SD haplotypes.	47
8.6 Data analysis by the Comparative Ct ( $\Delta\Delta C_t$ )	48
8.7 cDNA sequencing	49
<b>9. Methods used for genetic constructs based on the standardized assembly system GoldenBraid.</b>	<b>49</b>
9.1 Preparation of electrocompetent DH5 $\alpha$ Escherichia coli cells.	50
9.2 Restriction-ligation assembly reactions.	50
9.3 CRISPR/Cas9 knockout constructs.	51
9.3.1 Selection and design of the CRISPR/Cas9 guide RNAs.	51
9.3.2 GB CRISPR Domestication	51
9.3.3 CRISPR Multipartite Assemblies in $\alpha$ -level Destination Vectors	52
9.3.4 CRISPR Binary Assemblies in $\Omega$ -level Destination Vectors	52
9.3.5 CRISPR Binary Assembly in $\alpha$ -level Destination Vectors	53
9.3.6 Confirmation of the final constructs of CRISPR/Cas9 by sequencing.	53
9.4 Homologous over-expression and reporter cassettes assembly.	54
9.4.1 Sequencing of the promoter region of the Solyc04g081640 gene in <i>S. lycopersicum</i> and <i>S. pimpinellifolium</i> .	54
9.4.2 GB Domestication	54
9.4.3 Multipartite Assemblies in $\alpha$ -level Destination Vectors	55
9.4.4 Binary Assembly in $\Omega$ -level Destination Vectors	55
9.5 Transformation of <i>E. coli</i> electrocompetent cells by electroporation	56
9.6 Plasmid DNA extraction and purification from <i>E. coli</i> cultures	57
<b>RESULTS</b>	<b>58</b>
<b>CHAPTER I – PHENOTYPING AND NATURAL VARIATION OF THE LOCUS IN CHROMOSOME 4 CAUSING A SD DISORDER IN TOMATO.</b>	<b>59</b>
<b>1. Validation of plant material by genotyping.</b>	<b>59</b>
1.1 Genotyping of F <sub>2</sub> B <sub>2</sub> and F <sub>2</sub> B <sub>3</sub> families.	59
<b>2. Effect of gametic and zygotic factors on SD in reciprocal crossing BC<sub>1</sub>F<sub>1</sub> populations.</b>	<b>60</b>
<b>3. Seed counting of fruits from the self-pollination crosses on the three haplotypes.</b>	<b>61</b>
<b>4. Analysis of the natural variation in the SD region.</b>	<b>62</b>
4.1 Validation of the haplotype in the SD region in resequenced wild accessions	62
4.2 Analysis of the SD in the selected wild accessions	65

<b>CHAPTER II – GENOMIC AND TRANSCRIPTOMIC ANALYSIS OF THE CANDIDATE GENES.</b>	<b>67</b>
1. <b>Candidate genes in the SD region and functional in silico analysis.</b>	<b>67</b>
2. <b>Expression analysis of the candidate genes by qRT-PCR and Comparative Ct (<math>\Delta\Delta C_t</math>) analysis.</b>	<b>70</b>
3. <b>Candidate gene sequencing, structural prediction and SNP effects.</b>	<b>73</b>
<b>CHAPTER III – DEVELOPMENT OF BIOTECHNOLOGY TOOLS FOR SD FUNCTIONAL ANALYSIS</b>	<b>77</b>
1. <b>Design and construction of novel expression cassettes for <i>Solanum lycopersicum</i> using GoldenBraid 3.0 toolbox.</b>	<b>77</b>
1.1 CRISPR/Cas9 cassettes for the silencing of candidate genes.	77
1.2 Homologous over-expression of <i>Solanum pimpinellifolium</i> Solyc04g081640 gene in <i>Solanum lycopersicum</i> genetic background.	80
1.3 A reporter construct for the evaluation and spatial expression pattern driven by the <i>Solanum pimpinellifolium</i> and <i>Solanum lycopersicum</i> Solyc04g081640 gene promoter.	81
<b>DISCUSSION</b>	<b>84</b>
<b>CHAPTER I - PHENOTYPING AND NATURAL VARIATION OF THE LOCUS IN CHROMOSOME 4 CAUSING A SD DISORDER IN TOMATO</b>	<b>85</b>
1. <b>Effect of gametic and zygotic factors on SD in reciprocal crossing populations and on seed set of self-pollination crosses from the three SD haplotypes.</b>	<b>85</b>
2. <b>Analysis of the natural variation in the SD region.</b>	<b>86</b>
<b>CHAPTER II – GENOMIC AND TRANSCRIPTOMIC ANALYSIS OF THE CANDIDATE GENES.</b>	<b>88</b>
1. <b>Expression analysis of the candidate genes by qRT-PCR</b>	<b>88</b>
2. <b>Candidate gene sequencing, structural prediction and SNP effects.</b>	<b>89</b>
<b>CHAPTER III – DEVELOPMENT OF BIOTECHNOLOGY TOOLS FOR SD FUNCTIONAL ANALYSIS.</b>	<b>93</b>
1. <b>CRISPR/Cas9 cassettes for the silencing of candidate genes.</b>	<b>93</b>
<b>Conclusions</b>	<b>95</b>
<b>Future work</b>	<b>97</b>
<b>References</b>	<b>99</b>
<b>Supplementary Material</b>	<b>117</b>

## Tables

<b>Table 1</b> – Tomato plant accessions from the <i>S. pimpinellifolium</i> , <i>S. lycopersicum</i> var. <i>cerasiforme</i> and “Moneymaker” used in this study.....	38
<b>Table 2</b> – List of primers used for HRM genotyping. ....	41
<b>Table 3</b> – Cycling method for the KASP analysis used in the ABI 7500 Real Time PCR System.....	42
<b>Table 4</b> – Recycling method for the KASP analysis used in the ABI 7500 Real Time PCR System.....	42
<b>Table 5</b> – Primer pairs used in the sequence of the cDNA of the gene Solyc04g081640.1.....	49
<b>Table 6</b> – List of primer pairs for the construction of the Heterodimers for the gRNAs of the CRISPR tools. ....	51
<b>Table 7</b> – List of primers used in the sequencing of the fragments from the CRISPR/Cas9 knockout constructs pDGB3_alpha1:Tnos:NptII:Pnos-SF:U6-26:SL4g30.1:sgRNA:35s:hCas9:Tnos and pDGB3_alpha1:Tnos:NptII:Pnos-SF:U6-26:tRNA-gRNA40.1.1:tRNA-gRNA40.1.2:35s:hCas9:Tnos. ....	54
<b>Table 8</b> – Primers for the domestication of the new GB parts. ....	55
<b>Table 9</b> – Chi-square test for distorted segregation in the progeny of the reciprocal crosses.....	61
<b>Table 10</b> – Results of the seed counting from self-pollination crosses tomato fruits. ....	62
<b>Table 11</b> – Annotations available in the Sol Genomics Network for the candidate genes in the Heinz reference genome.....	69
<b>Table 12</b> – List of variants detected in <i>S. lycopersicum</i> SP4-4 against the 150 and 360 tomato resequence project references.....	73

## Figures

<b>Figure 1</b> – Phylogenetic representation of the 17 tomato species in the Solanum genus with the separation of the three sections of the genus, Lycopersicon, Juglandifolia, and Lycopersicoides. ....	4
<b>Figure 2</b> – Geographical distribution of the Solanum genus tomato species on the pacific coast of South America and Galapagos islands.....	6
<b>Figure 3</b> – Example of a HRM analysis result.....	12
<b>Figure 4</b> – Allelic discrimination plot of a KASP analysis.....	14
<b>Figure 5</b> – Marker-assisted selection (MAS) pipeline applied by Barrantes (2014) to develop a <i>S. pimpinellifolium</i> accession TO-937 library of introgression lines (IL) in the genetic background of <i>S. lycopersicum</i> cv. “Moneymaker”.....	15
<b>Figure 6</b> – Schematic representation of the distal portion of chromosome 4 of tomato. ....	20
<b>Figure 7</b> – Chronological representation of the 10 landmarks of floral development for tomato species.....	23
<b>Figure 8</b> – Chronological representation of the 10 landmarks of fruit development for tomato species..	24
<b>Figure 9</b> – Conformational structure change and activation upon binding of the guide RNA to the Cas9 protein.....	27
<b>Figure 10</b> – Schematic view of the GB syntax and description of each one of the 10 elements and respective 5’ and 3’ fusion overhangs. ....	32
<b>Figure 11</b> – Schematic representation of the GoldenBraid system. ....	34
<b>Figure 12</b> – Schematic of the GB grammar for the construction of CRISPR target TUs for plants. ....	35
<b>Figure 13</b> – Genomic SD region defined by Barrantes, et al. (2014) and narrowed by Fakhret (2016). ..	60
<b>Figure 14</b> – Neighbor-joining tree representing the proximity of other tomato accessions to the <i>S. pimpinellifolium</i> TO937 based on the total number of equal SNPs. ....	64
<b>Figure 15</b> – Physical map of the KASP markers. ....	65
<b>Figure 16</b> – Schematic representation of the SD genomic region mapped by Fakhret (2016). ....	69
<b>Figure 17</b> – Plotting of the results of the qRT-PCR of the candidate gene Solyc04g081630 encoding to a HSP90.....	72
<b>Figure 18</b> – Plotting of the results of the qRT-PCR of the candidate gene Solyc04g081640 encoding to a HSP90.....	72
<b>Figure 19</b> – Predicted protein sequence alignment of the <i>S. pimpinellifolium</i> and <i>S. lycopersicum</i> sequences of the genes Solyc04g081630, Solyc04g081650, Solyc04g081660.....	74
<b>Figure 20</b> – Alignment of the Solyc04g081640 gene sequences from the <i>S. lycopersicum</i> Heinz reference genome, <i>S. lycopersicum</i> SP 4-4 and the <i>S. pimpinellifolium</i> and <i>S. lycopersicum</i> var. <i>cerasiforme</i> accessions.....	75



**Figure 21** – Graphical representation of the nucleotide sequences of the genomic annotation (a.), sequenced genomic DNA (b.) and sequenced cDNA (c.) of the Solyc04g081640 gene in in the SP4-4....76

**Figure 22** – Schematic representation of the CRISPR/Cas9 silencing plasmids.....79

**Figure 23** – Assembly of the CRISPR/Cas9 construct for Solyc04g081630.. .....79

**Figure 24** – Assembly of the CRISPR/Cas9 construct for Solyc04g081640.. .....80

**Figure 25** – Assembly of the plasmid  
pDGB3\_omega1::35s:SPSolyc04g081640.1CDS:Tnos::Pnos:NptII:Tnos .....81

**Figure 26** – Schematic representation of the reporter plasmids. ....83

## Supplementary materials

<b>Annex 1</b> – Sequences and physical properties of the primers designed for qRT-PCR, sequencing and other amplification reactions. ....	118
<b>Annex 2</b> – Sequences and physical properties of the primer triplets for the KASP markers.....	122
<b>Annex 3</b> – $C_T$ analysis raw data calculations and plotted graphs of the qRT-PCR experiment. ....	123
<b>Annex 4</b> – List of GB elements used and created with the GB 3.0 toolkit for the assembly of the CRISPR/Cas9 cassettes and expression cassettes. ....	131
<b>Annex 5</b> – Results of the HRM genotyping. ....	136
Annex 6 – Results of the KASP analysis performed on <i>S. pimpinellifolium</i> and <i>S. lycopersicum</i> var. <i>cerasiforme</i> samples.....	137
<b>Annex 7</b> – Results of the in silico analysis of the candidate genes and respective orthologs in <i>Arabidopsis thaliana</i> . ....	138
<b>Annex 8</b> – Alignment of the Solyc04g081640 gene proximal promoter region from the <i>S. lycopersicum</i> var. “Moneymaker” (MM) and <i>S. lycopersicum</i> IL 4-4 PP.....	146

## Abbreviations list

♂ – male

♀ – female

A – adenine

acc. – accession

aka – also known as

ANT – anthers

BC – backcross

C – cytosine

CAPS – Cleaved Amplified Polymorphic Sequence

Cas – CRISPR-associated

CDS – coding sequence

COMAV – Centro de Conservación y Mejora de la Agrodiversidad Valenciana

CRISPR – clustered regularly interspaced short palindromic repeats

crRNA – CRISPR RNA

DSB – double strand breaks

DNA – deoxyribonucleic acid

ER – endoplasmic reticulum

FRET – Förster/fluorescence resonance energy transfer

G – guanine

GB – GoldenBraid

GFP – green fluorescent protein

GMO – genetically modified organisms

GO – Gene Ontology

gRNA – guide RNA

HDR – homology-directed repair

HRM – high-resolution melting

HSP – Heat-shock protein

IBMCP – Instituto de Biología Molecular y Celular de Plantas

IL – introgression line

InDel – insertion/deletion

ITAG – International Tomato Annotation Group

KASP – Kompetitive Allele Specific PCR genotyping assay

MAS – Marker Assisted Selection

MoClo – Golden Gate Modular Cloning

NGS – next-generation sequencing

NHEJ – non-homologous end-joining

NIL – Near Isogenic Line

OVA – ovaries

PCR – polymerase chain reaction  
qRT-PCR – quantitative real-time polymerase chain reaction  
QTL – Quantitative Trait Loci  
RFLP – Restriction Fragment Length Polymorphism  
RIL – Recombinant Introgression Line  
RNA – ribonucleic acid  
RNAi – RNA interference  
rpm – rotations per minute  
SC – self-compatible  
SD – Segregation Distortion  
SDL – Segregation Distortion Loci  
SGN – Sol Genomics Network  
sgRNA – single guide RNA  
SI – self-incompatible  
SL – *Solanum lycopersicum*  
SNP – single nucleotide polymorphism  
SolCAP – Solanaceae Coordinated Agricultural Project  
SP – *Solanum pimpinellifolium*  
SSR – Single Sequence Repeat  
T – thymine  
TALEN – transcription activator-like effector nuclease  
TGRC – Tomato Genetic Resource Center  
T<sub>m</sub> – melting temperature  
tracrRNA – auxiliary trans-activating crRNA  
TRD – Transmission Ratio Distortion  
TSS – transcription starting site  
TU – transcription unit  
var. – variety  
VPE – Vacuolar Processing Enzymes  
ZFN – zinc-finger nuclease

## **Objectives:**

- Identify and characterize the gene(s) in the SD region responsible for the SD phenomenon;
- Investigate natural variability of the SD region in wild tomato accessions;
- Continue the study of the effect of gametic and zygotic factors on SD;
- Create CRISPR/Cas9 and expression analysis tools to study gene and promoter function.

# **Literature Review**

## 1. Botanical description of *Solanum* tomato species.

The tomato plant is a Magnoliopsida (dicotyledon) plant from the Solanaceae family, also known as nightshades. The Solanaceae family comprises 98 genera with some species having high economic importance, most notably, the potato (*Solanum tuberosum*), tomato (*Solanum lycopersicum*), pepper (*Capsicum annuum*), eggplant (*Solanum melongena*), and tobacco (*Nicotiana tabacum*). The most represented genus is in fact *Solanum*, accounting for almost half of all species in the Solanaceae family and being one of the most represented genus in all angiosperms (Peralta *et al.*, 2008).

The *Solanum* genus is separated in 3 sections of tomato plants: *Lycopersicon*, *Lycopersicoides* and *Juglandifolia* (Figure 1). The *Solanum* section *Lycopersicon* contains 13 species, 12 of those are wild species of tomato and only one, *S. lycopersicum*, exists as domesticated or feral plant. The wild species from the *Lycopersicon* section are: *S. arcanum*, *S. cheesmaniae*, *S. chilense*, *S. chmielewskii*, *S. corneliomulleri*, *S. galapagense*, *S. habrochaites*, *S. huaylasense*, *S. neorickii*, *S. pennellii*, *S. peruvianum* and *S. pimpinellifolium*. Another two sections of the *Solanum* genus are the *Lycopersicoides*, that contains two species: *S. lycopersicoides* and *S. sitiens*, and *Juglandifolia*, comprised by the *S. juglandifolium* and *S. ochranthum* species (Knapp and Peralta, 2016).

The tomato wild species are native to the west coast of South America (Figure 2). The eleven wild species from the *Lycopersicon* section inhabit dry and rocky open regions of the Andean mountains and valleys or not so dense forests from sea-level to 4000m. Two exceptions are the two orange/yellow-fruited species native to the Galapagos islands, *S. cheesmaniae* and *S. galapagense*. They occupy dry, open, rocky slopes from seashore to 1600m. Similarly, the *Lycopersicoides* section plants, *S. lycopersicoides* and *S. sitiens*, inhabit continental slope regions but in high altitude regions between 1250m to 3500m. The *Juglandifolia* section species, *S. juglandifolium* and *S. ochranthum*, are mostly found in high altitude montane cloud forests (Knapp and Peralta, 2016).

*S. pimpinellifolium* species is the closest wild relative of domesticated tomato *S. lycopersicum*, also known as currant tomatoes, varies from the domesticated *S. lycopersicum* varieties mainly on fruit size and metabolite composition. *S. pimpinellifolium* fruits are small, almost perfectly round, red or orange in colour when ripe and ranging from 0.5 to 1 cm. It is also SC-autogamous or facultative allogamous. Flowers are similar in size to the cultivated tomatoes. Is

usually found in arid sandy soils near water sources or on the periphery of farming fields in the regions of Ecuador and Peru from sea level to 500 m in altitude. (Viquez Zamora, 2015)

Domesticated *Solanum lycopersicum* tomato plants are cultivated worldwide, usually in moderate humid places. Immature fruits are in general green and ripe fruits are typically red and vary in size, shape and metabolic composition depending on the variety. It is a SC-autogamous or facultative allogamous plant. (Viquez Zamora, 2015)

From a scientific perspective, tomato became a model species for studying fleshy fruited plants (Meissner *et al.*, 1997), the main reasons are the large quantity of genomic data available (Mueller *et al.*, 2005; 100 Tomato Genome Sequencing Consortium, 2014) and the ample collection of accessions in germplasm banks (Bai and Lindhout, 2007).

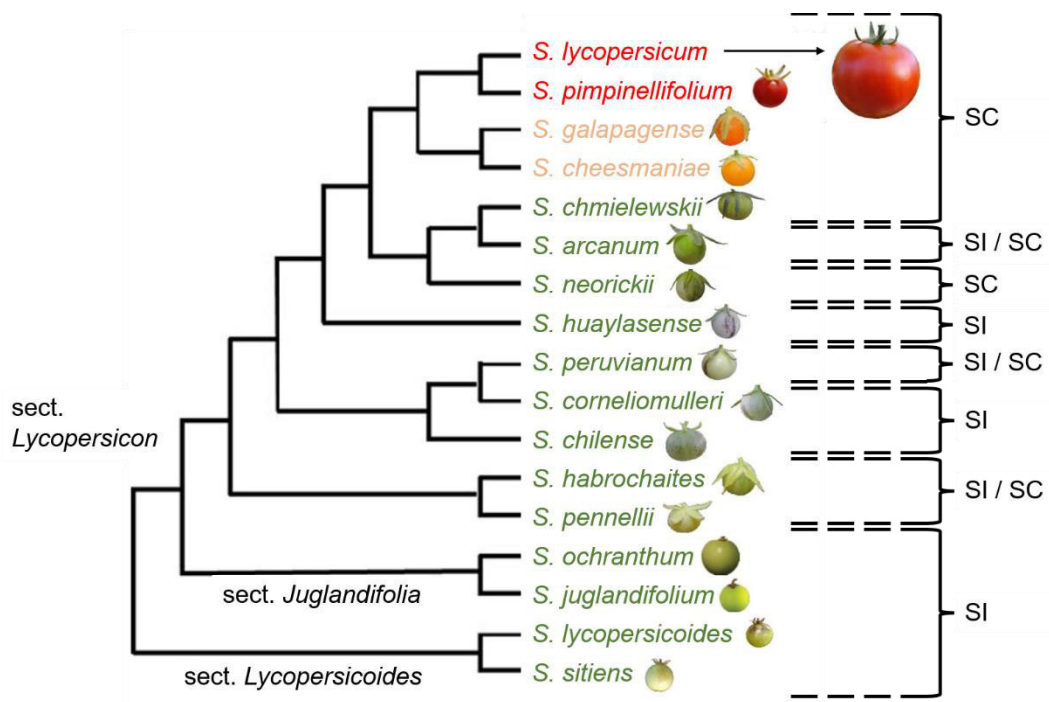


Figure 1 – Phylogenetic representation of the 17 tomato species in the *Solanum* genus with the separation of the three sections of the genus, *Lycopersicon*, *Juglandifolia*, and *Lycopersicoides*. On the left of the tree there is the indication of the self-compatibility (SC) and self-incompatibility (SI). Adapted from Bedinger *et al.*, 2011. Tomato fruit photos from Li and Chetelat, 2015.

### 1.1 *Solanum lycopersicum* var. Heinz and *Solanum lycopersicum* var. “MoneyMaker”.

*S. lycopersicum* var. Heinz 1706-BG was the cultivar used in the tomato genome sequencing project (Andolfo *et al.*, 2013). Like other *S. lycopersicum* accessions is self-compatible (SC) and has a diploid number of 24 chromosomes ( $2n=2x=24$ ) with the approximate genome size



of 950Mb. This genome is predicted to have 33,810 coding genes as well as 1,406 non-coding genes (Tomato Genome Consortium, 2012; 100 Tomato Genome Sequencing Consortium, 2014). The gene annotations were made by the International Tomato Annotation Group (ITAG) using *ab initio* gene prediction methods and evidence-based sequence data. *S. lycopersicum* var. “MoneyMaker” is another cultivar similar to the Heinz variety, both genomic and phenotypically, that was used as recurrent parent in the development of the introgression line (IL) collection by Barrantes (2014).

## **1.2 *Solanum pimpinellifolium* accession TO-937.**

*S. pimpinellifolium* accession TO-937 is a red-fruited wild variety of tomato from Peru and, as such, is close to the region where tomato was first domesticated. This wild accession differs from others because it has type IV glandular trichomes on its leaves and stems that contain acylsucroses not observed in cultivated tomatoes or in other wild accessions. This trait confers resistance to pests like the two-spotted spider mite *Tetranychus urticae* (Alba *et al.*, 2009). Adding to the pest resistance, TO-937 accession has some traits valuable for domesticated tomato related to fruit quality, such as, high content in ascorbic acid, sugars, organic acids, carotenoids and volatile compounds (Rambla *et al.*, 2014; Lima-Silva *et al.*, 2012). For these reasons, it was chosen as the donor genome in the development of the genomic libraries of ILs by Barrantes (2014) that will be discussed later.

## **2. Economic importance of the cultivated tomato.**

Commercial varieties of the *Solanum lycopersicum* L. are one of the most important agricultural productions worldwide, having a production of more than 170 million metric tons in 2014 from more than 5 million hectares. European Union production rounds 17 million metric tons from 264 thousand hectares of cultivated area. Spain produces circa 5 million tons and Portugal 1.4 million, from 55 thousand and 18 thousand hectares of cultivated area, respectively. The world biggest producers are China, India, USA and Southeast Asia countries (FAOSTAT: <http://www.fao.org/faostat/en/>).

Tomato varieties cultivated in different parts of the world change according to cultural, gastronomical and historical factors but also for their adaptation to climate and resistance to pests.

More than 1000 varieties are stored in the Tomato Genetic Resource Center (TGRC: <http://tgrc.ucdavis.edu/index.aspx>) germplasm bank for scientific studies.

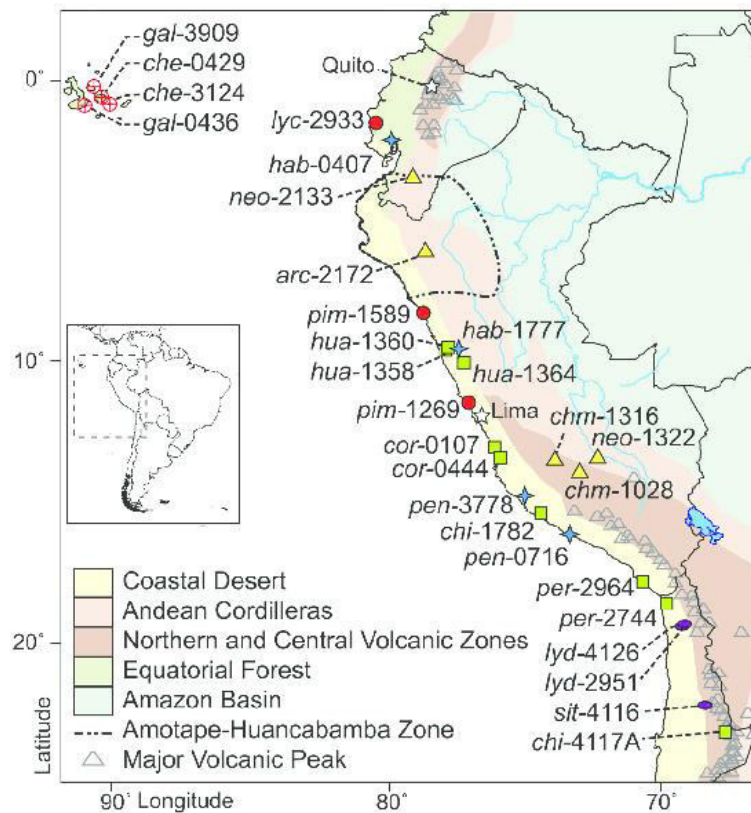


Figure 2 – Geographical distribution of the *Solanum* genus tomato species on the Pacific coast of South America and Galapagos islands. The 4 digit numbers indicate the TGRC code of that species accession. gal – *S. galapagense*; che – *S. cheesmaniae*; lyc – *S. lycopersicum*; hab – *S. habrochaites*; neo – *S. neorickii*; arc – *S. arcanum*; pim – *S. pimpinellifolium*; hua – *S. huaylasense*; cor – *S. corneliomulleri*; pen – *S. pennellii*; chi – *S. chilense*; chm – *S. chmielewskii*; neo – *S. neorickii*; per – *S. peruvianum*; lyd – *S. lycopersicoides*; sit – *S. sitiens*. Image from Pease *et al.*, 2016.

### 3. Origin and domestication of tomato.

Tomato (*Solanum lycopersicum*) is originally from the Andean region that is now part of Peru, Ecuador, Bolivia, Colombia and Chile but the time of domestication remains to this day a debated subject. Some researchers propose a domestication in the region of origin, other hypothesis suggests that it was domesticated in Central America where is now Mexico. (Peralta *et al.*, 2008). These hypotheses were formulated based on botanical, linguistic and historical aspects (Bauchet and Causse, 2012; Peralta *et al.*, 2008).

Among the tomato wild species, *S. pimpinellifolium* is the only one to present red fruits and it was assumed to be the ancestral of modern tomato. However, *S. pimpinellifolium* is only present in South America, not in Mexico. The most accepted explanation is that the domesticated *S. lycopersicum* came from *S. lycopersicum* var. *cerasiforme* which in turn originated from the *S. pimpinellifolium*. Contrary to *S. pimpinellifolium*, *S. lycopersicum* var. *cerasiforme* is present in both Mexico and Andean region (Bauchet and Causse, 2012). Therefore, likely the first

domestication occurred directly from *S. pimpinellifolium* to *S. lycopersicum* var. *cerasiforme* and later to *S. lycopersicum*. This 2-step domestication correlates to the geographical area where the plants are found and is supported by the decrease of population effective size during *S. lycopersicum* var. *cerasiforme* domestication. (Bai and Lindhout, 2007)

Tomato arrived in Europe after the discovery of the Americas in the 15<sup>th</sup> century and recent evidence indicates that it started to be studied, cultivated and consumed within the 16<sup>th</sup> century (<http://traditom.eu/en/history/>) possibly later in Northern Europe (Bai and Lindhout, 2007).

Traditional farmers probably selected varieties that were adapted to their climatic conditions and preferences as suggested by the phenotypic and genotypic variability present in the traditional European gene pool (<https://traditom.upv.es/> and unpublished info from the IBMCP lab). During the the last century, the number of tomato varieties has been increased due to intensive breeding efforts. Current varieties have mostly expanded the already existing variability in size, shape, colour, genes for disease resistance, abiotic stress tolerance, and other valuable traits.

The latest milestone occurred with the introduction of genomics (Mueller *et al.*, 2005), enabling a faster and more precise development of new varieties and the introduction of novel ones. On a genomic level, due to domestication by inbreeding, the genetic diversity of cultivated tomato had an unceasing decrease caused by successive genetic bottlenecks (Bauchet and Causse, 2012; van der Knaap *et al.*, 2002). Comparatively to wild tomato species, domesticated species are estimated to have less genetic diversity but the develop of heirloom tomato cultivars selected and adapted to certain environmental factors should be considered a source for genetic variability and valuable traits (Knapp and Peralta, 2016)

#### **4. Phenotypic and genetic variability of wild tomato species as a source of variability for cultivar improvement.**

Although geographically dispersed from the Galapagos island to the northwest part of South America, *S. cheesmaniae*, *S. galapagense*, *S. lycopersicum* and *S. pimpinellifolium* form a morphological and genetical close group (Figure 1) being the only species with red or orange fruits, hinting for a possible incomplete lineage between this species (Koenig *et al.*, 2013). In fact, all these species are self-compatible (SC), i.e. autogamous (Bedinger *et al.*, 2011), even though *S. pimpinellifolium* is a facultative allogamous if the conditions are appropriate (Rick *et al.*, 1978).

Alternatively, and concerning the mating systems of the tomato clade, the sections *Juglandifolia* and *Lycopersicoides* are self-incompatible (SI). The SI species have floral structures that promote outcrossing. Some species, most notably, *S. habrochaites* and *S. pennellii* (Figure 1), present both SC and SI mating system (Bedinger *et al.*, 2011). The determinant factor for one or other crossing system seems to be the geographical positioning of the plant that changes the phenotype of the flowers (Rick *et al.*, 1978).

The understanding of mating barriers between the wild accessions and cultivated tomato, as well as the genetic variability, is important for the current and future use of wild accessions for the improvement of tomato cultivars (Bedinger *et al.*, 2011). The incompatibility of the mating systems of two tomato species can affect the development of hybrid populations that have greater variability, especially if the aim is to confer variability to cultivars.

Different wild species can be used as a source for genetic and phenotypic diversity for different traits mainly caused by the adaptation to their native environment. The *S. cheesmaniae* and *S. galapagense* have tolerance to high concentrations of salt. *S. neorickii* is resistant to bacterial diseases and *S. peruvianum*, *S. huaylasense*, *S. arcanum* and *S. corneliomuelleri* also to virus and fungi. *S. pennellii* and *S. chilense* are resistant to drought. *S. habrochaites* is more suited to colder climates. Alternatively, *S. pimpinellifolium* is tolerant to heat and insect induced stresses. Other than abiotic resistance traits, *S. pimpinellifolium* is more suited to improve quality traits of the cultivated tomato (Barrantes, 2014).

## **5. Exploiting genetic variability available for tomato cultivar improvement by using marker assisted selection (MAS) and genetic maps of quantitative trait loci (QTL).**

Traditionally, the main agricultural traits used to improve in tomato and other crops have been yield, quality, tolerance or resistance to stresses. These traits however result from interactions of several genes and the environment. The genes involved in complex traits are called quantitative trait locus (QTL).

DNA-based molecular markers can be associated to specific traits or chromosomes and help breeders select molecular markers instead of phenotypic markers. Marker Assisted Selection (MAS) changed plant breeding from selection of phenotypes (using morphophysiological markers) to the selection of genes based on marker genotypes. Therefore, QTLs can more easily be selected by MAS (Bai and Lindhout, 2007). The application of MAS allows for the substitution of field trials

by molecular testing to monitor the presence of desirable marker/ gene/ trait at the seedling level and, by doing so, reducing the time and cost for developing new cultivars. Also, it made possible to choose traits that do not have apparent phenotypical manifestations, or are not easy to screen, like resistance to disease.

To improve tomato cultivars and overcome the reduced genetic variability of domesticated tomato the implementation of genetic maps and marker assisted selection (MAS) of QTLs has revealed to be crucial. An extensive list of markers is available to the scientific community at the Sol Genomics Network (SGN, <http://solgenomics.net/>), including Restriction Fragment Length Polymorphism (RFLP), Cleaved Amplified Polymorphic Sequence (CAPS), microsatellites (SSR – Single Sequence Repeat), and single nucleotide polymorphism (SNP) markers.

Before incorporating genetic variability to tomato cultivars, the first step is to develop genetic maps in order to study the underlying QTL / genetic variability. Mapping population generation starts by selecting two genetically divergent parent genotypes, the recurrent parent or recipient and the donor parent. Due to the lack of genetic variability among domesticated cultivars, the most common way is to use inter-specific crosses between wild and cultivated tomatoes to create hybrid populations (Knapp and Peralta, 2016; Grandillo, 2014). The most used species are *S. pennellii*, *S. habrochaites* and especially *S. pimpinellifolium* since they are a good source for genes/traits of interest, are easily crossable, and do not present problems in hybrid crossing like F<sub>1</sub> sterility (Barrantes, 2014). The first high-density genetic map of tomato species was developed using restriction fragment length polymorphism (RFLP) markers of a segregating interspecific F<sub>2</sub> population from a *S. lycopersicum* x *S. pennellii* cross (Tanksley *et al.*, 1992).

As said before, the incompatibility of the matting systems of tomato species (Bedinger *et al.*, 2011) can create a major setback on creating mapping populations. But in the cases where the two parents are compatible it is possible to create segregating recombinant populations. Such examples of segregating populations to study and map QTLs are the F<sub>2</sub> populations, backcross (BC), introgression lines (ILs), recombinant introgression lines (RILs) and near isogenic lines (NIL).

After crossing two distinct homozygote parents (P), the first generation of the cross (F<sub>1</sub>) will consist of full heterozygous individuals. In this heterozygotic population, during meiosis, recombination events can occur among chromosomes, creating chimeric chromosomes (Roeder *et al.*, 1997). For this reason, the first type of crosses used to develop QTL maps were F<sub>2</sub> and BC populations. F<sub>2</sub> segregating populations are generated by self-crossing the F<sub>1</sub> population and BC population by crossing the F<sub>1</sub> population with one of the parents, either the recurrent or donor

parent. But these types of interspecific cross population have some drawbacks, like the presence of multiple segregating QTLs in the whole genome that mask the effect of each other (Grandillo *et al.*, 1999) and therefore are not as well suited to map QTLs for crop improvement.

On the other hand, each Introgression line (IL) in a population of ILs, contains only a small portion of the donor genome in the genetic background of the recurrent parent and all together cover the whole genome of the donor in a small number of lines (traditionally around 50-60 for tomato), and are much more suited for QTL mapping. Since the genetic background is almost all similar to the recurrent parent, the phenotypical changes can be pinpointed to the small genetic segment of the donor parent and can be studied as traits with Mendelian (or quasi-Mendelian) inheritance (Rousseaux *et al.* 2005; Schauer *et al.*, 2006).

In tomato, cultivated varieties are logically used as the recurrent parent and the wild species as the donor. Several ILs libraries have been developed for wild relatives of cultivated tomatoes such as, for *S. habrochaites* (Monforte and Tanksley, 2000), and *S. pennellii* (Eshed and Zamir, 1995). IL collections can be applied to a multitude of QTL studies, such as, the identification, verification of effects, and study of interactions with environment, genetic background and other QTLs (Barrantes *et al.*, 2014)

## **6. Single nucleotide polymorphism markers in the tomato genome and high throughput genotyping.**

Briefly, a single nucleotide polymorphism (SNP) is a variation at a single nucleotide position in a DNA sequence of a single species. SNPs are the most common polymorphism in the genomes of higher organisms, i.e., animals and plants (Huq *et al.*, 2016). In plants, SNP variants are at high density across the genome, for example in tomato 6.1 SNPs per Kb are observed in the whole genome (Kim *et al.*, 2014).

Adding to the fact that SNPs are the most common variation, the genomic position of a SNP can affect gene expression by variation on upstream and downstream regulatory sequences and gene function, most commonly for creating nonsense and missense mutations in coding sequences, hence affecting the phenotype (Hirakawa *et al.*, 2013). The study of SNPs is then crucial for functional genomics.

Having a way to measure genetic polymorphism, genetic distance and population differentiation is decisive to efficiently manage germplasm resources for crop improvement and association mapping (Hamilton *et al.*, 2012)

With the recent advances in next-generation sequencing (NGS) technologies, obtaining a complete genome sequence is increasingly more precise and cost-effective. High throughput genotyping together with NGS allows for the construction of highly precise high-density genetic maps (Sim *et al.*, 2012).

A NGS re-sequencing analysis done by Hirakawa *et al.* (2013) obtained a whole-genome sequence that not only included SNP data but also insertions/deletions (InDels) just by mapping the obtained reads against the reference genome.

In the studying of tomato, several works regarding genome-wide SNPs were conducted (Hamilton *et al.*, 2012; Kim *et al.*, 2014; Hirakawa *et al.*, 2013). All these works together with the tomato genome sequences available will further enhance the knowledge not only about gene function but more relevant for phenotypic variation for crop trait improvement (100 Tomato Genome Sequencing Consortium, 2014; Tomato Genome Consortium, 2012).

Based on NGS-derived transcriptome sequence data Sim *et al.* (2012) developed the first large SNP genotyping array for tomato, SolCAP array, that was used to create high-density linkage maps for interspecific F<sub>2</sub> populations.

Other than the genome-wide mapping of SNPs, there are important techniques to focus on small portions of the genome or in specific alleles. Two such methods of simple SNP genotyping are the high-resolution melting (HRM) and Kompetitive Allele Specific PCR genotyping assay (KASP) described in detail below.

### **6.1 High resolution melting - HRM**

The High-Resolution Melting (HRM) analysis was first described by Wittwer *et al.* back in 2003, in an attempt to develop a genotyping and mutation screening system without labeled oligonucleotides, overcoming the liabilities of the systems previously available.

HRM can identify PCR products amplified in the presence of a double-stranded DNA (dsDNA) dye using unlabeled primers (Wittwer *et al.*, 2003). It is a cost-effective method, since it only uses a traditional PCR reaction, a common dsDNA saturating dye and melting hardware and software for analysis, and it is simple and fast, because it can be performed in a closed-tube reaction and the analysis accomplished in a very brief time after the completion of the PCR reaction (Liew, M. *et al.*, 2004).

The principle of the method is that different amplicons have different melting temperatures ( $T_m$ ). The differences in the  $T_m$  of the dsDNA are caused by the GC content, length, and sequence of the amplicon (Ririe *et al.*, 1997).

As the temperature increases, the dsDNA denatures to ssDNA and the dye loses affinity to the DNA lowering the fluorescence that is detected by the sensor. This drop is more accentuated in the  $T_m$  (i.e. the temperature at which 50% of the DNA is double stranded and 50% is single stranded) of an amplicon marking the inflex point of the melting curve. (Figure 3a) The relative amplicon signal for each sample is based in the fluorescence change inside the region defined by the pre-melting region, when samples are dsDNA and the emitted fluorescence is 100%, and the post-melting region, when samples are ssDNA and the emitted fluorescence is 0%. The change of the fluorescence in this region is used to calculate the relative signal of each amplicon and to discriminate the change in the melt curve behavior between DNA samples.

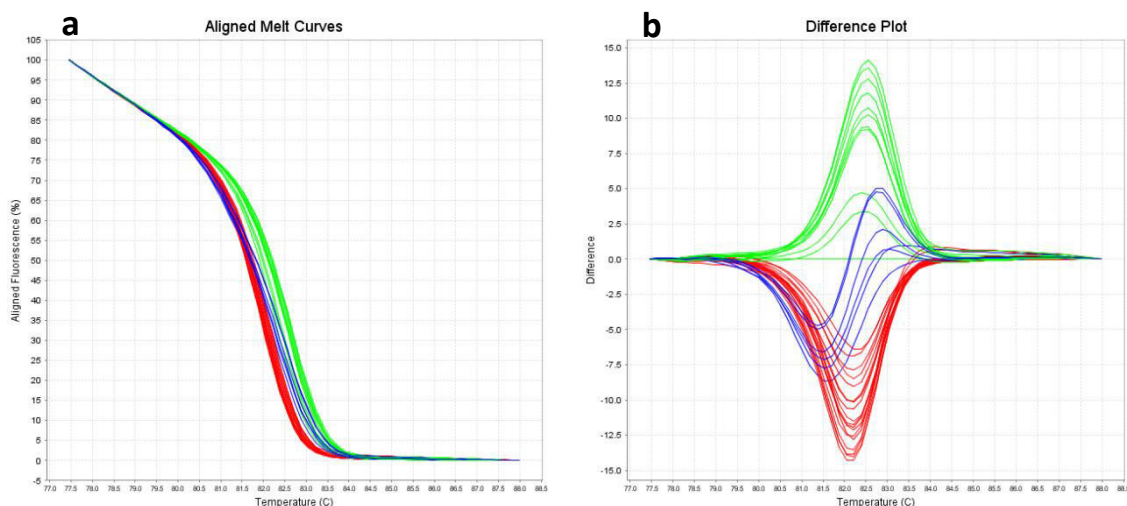


Figure 3 – Example of a HRM analysis result. (a) plotting of the signal shift over the difference in temperature. (b) normalized signal shift over the difference in temperature. Green lines indicate mutant SNP samples, Red lines indicate wild-type SNP samples. Blue lines heterozygote samples.

HRM has high sensitivity for heterozygote samples although in some cases different heterozygote variant may have very similar melting curves to the homozygotes (Wittwer *et al.*, 2009). So, homozygote samples are in fact not identified by a  $T_m$  shift compared to the wild-type but are identified by the melting curve shape (Figure 3b) (Herrmann *et al.*, 2006). The slow increment in temperature allows for the detection of SNPs between samples making HRM a suited technique for SNP genotyping, DNA mapping, and characterization of haplotypes and other DNA sample differentiation analysis. It can even be used to assess methylation of DNA (Wojdacz *et al.*, 2007).



The capabilities of this technique in plant biology, were proved in the identification of species/cultivars (Ganopoulos *et al.*, 2011; Mackay *et al.*, 2008), genome-wide SNP discovery and mapping (Han *et al.*, 2012; Lehmsiek *et al.*, 2008) and screening of GMOs and other modified crops (Lochlainn *et al.*, 2011; Akiyama *et al.*, 2009), just to point some examples.

## **6.2 Kompetitive Allele Specific PCR genotyping assay - KASP™**

The Kompetitive Allele Specific PCR genotyping assay (KASP™ or KASPer) combines the competitive allele-specific PCR with a homogenous, fluorescence-based reporting system. It allows the identification of genetic variation at the nucleotide sequence, namely, single nucleotides polymorphisms (SNPs) and insertions and deletions (InDels) (He *et al.*, 2014).

This technology uses two competitive primers that differ on the 3' end, accordingly to the nucleotide difference between alleles, and a common primer for the amplification of small DNA fragments (<120bp). The reaction mixture uses a FRET cassette (LGCGenomics kit uses FAM and HEX dyes) and a Taq polymerase for identification and amplification, respectively (Thomson, 2014).

The competitive annealing of the two allele-specific primers emits a fluorescence signal for one of the dyes in a homozygote sample. On the other hand, a heterozygote for the alleles generates a mixed signal (<http://www.lgcgroup.com/kasp/>). Figure 4, represents the allelic discrimination plot of a KASP analysis. Results are clustered based on the wavelength of the signal. Red dots indicate the presence of a Wild-type SNP by a high fluorescence of the FAM probe. Blues dots indicate a Mutant SNP by high fluorescence signal of HEX probe. And green dots indicate heterozygote samples by the existence of a mixed signal. Black squares represent non-template controls, therefore not having fluorescence signal.

One of the main advantages of this system for genotyping is the reduction of cost by minimizing the reaction volume and by simplifying the process using 96 or 384-well PCR plaque on a thermocycler equipped with a fluorescence detector. The plaque assay allows the genotyping of few samples with several markers or the opposite, few markers and a large number of samples (He *et al.*, 2014).

It is an excellent tool for genotyping crops and it has been used successful in corn (Semagn *et al.*, 2014), wheat (Neelam *et al.*, 2013), peanut (Khera *et al.*, 2013), and chickpea (Hiremath *et al.*, 2012).

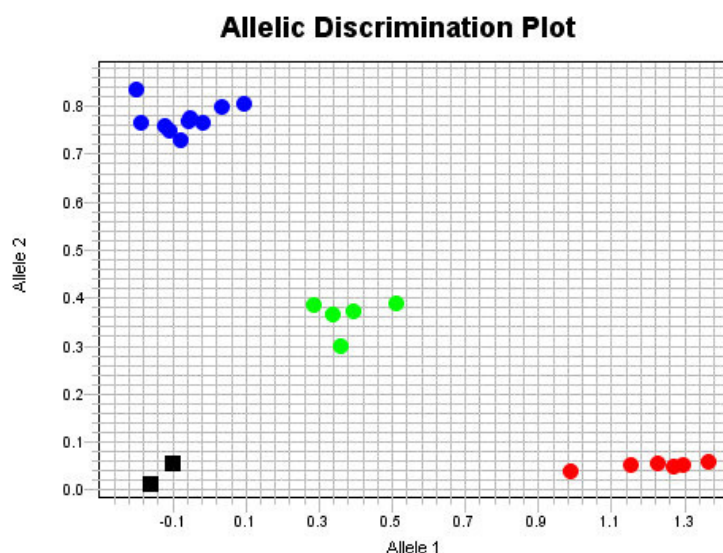


Figure 4 – Allelic discrimination plot of a KASP analysis. Red – Wild-type SNP in homozygosis; Blue – Mutant SNP in homozygosis; Green – heterozygote; Black – non-template controls.

### **7. *S. pimpinellifolium* accession TO-937 library of introgression lines (IL) in the genetic background of tomato cultivar “Moneymaker”.**

TO-937 accession includes resistance to some pests (as spider-mite) and it has potential to increase fruit quality traits of the tomato cultivars. So, an IL collection was developed from this accession in the “Moneymaker” genetic background. The breeding scheme applied by Barrantes *et al.* (2014) for the development of the IL library is depicted in Figure 5. The development of the introgression line started by backcrossing a single F<sub>1</sub> plant from the *S. lycopersicum* (SL) x *S. pimpinellifolium* (SP) cross with the recurrent parent “Moneymaker” (SL). The 100 BC<sub>1</sub> plants that originated from this cross were backcrossed again with SL to produce BC<sub>2</sub> generation. These BC<sub>2</sub> plants were genotyped using a 712 SNPs array and a number of lines selected based on 2 criteria, (1) the entire genome should be represented in those selected lines at least twice with overlapping donor segments, and (2) each selected plant should have the lowest proportion of the donor genome. The BC<sub>2</sub> that followed these criteria were then backcrossed again with SL (recurrent parent) to produce a BC<sub>3</sub> progeny. The BC<sub>3</sub> plants were genotyped with a 96 SNPs array and selected following the same criteria as the BC<sub>2</sub> plants. The selected plants from BC<sub>3</sub> were divided into two groups after genotyping using (HRM). An “elite” group carrying up to 3 introgressions and a “backup” group with more than 3 introgressions. The plants from the elite group were selfed to obtain BC<sub>3</sub>S<sub>1</sub> and those in the backup group backcrossed again to produce BC<sub>4</sub>. The BC<sub>3</sub>S<sub>1</sub> and BC<sub>4</sub> populations were screened by HRM and selfed until plant with single introgressions were obtained.

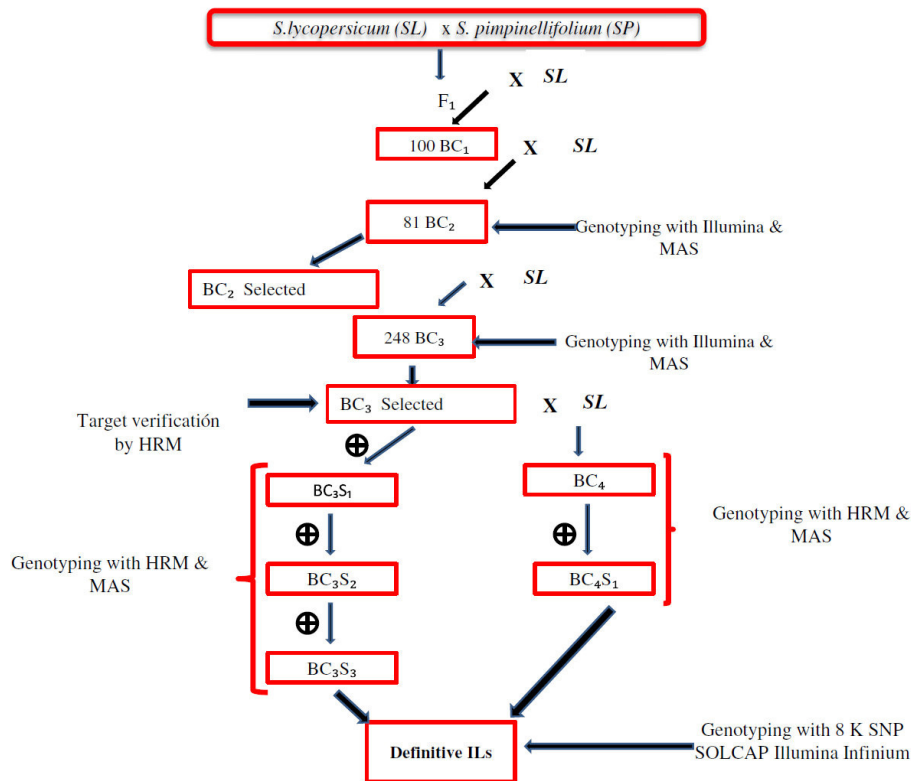


Figure 5 – Marker-assisted selection (MAS) pipeline applied by Barrantes (2014) to develop a *S. pimpinellifolium* accession TO-937 library of introgression lines (IL) in the genetic background of *S. lycopersicum* cv. “Moneymaker”. Image from Barrantes *et al.* (2014).

The definitive set of ILs were genotyped by an 8K Illumina Infinium Array and showed 95.7% of the SL genome and an average number of SP introgressions of 1.9. The whole IL library is composed of a total 53 ILs covering, in total, 94% of the donor SP parent genome. Each IL containing 4.25% of the SP genome.

Most of the ILs were developed in just four generations ( $BC_3S_1$ ) which is lower than other similar works reporting 5 to 10 generations needed. For the complete set five to seven generations where needed.

With this study, Barrantes *et al.* (2014) showed that the implementation of high-throughput genotyping in the intermediate steps of developing ILs reduces significantly the number of generation needed, comparatively to other studies, and increases the isogeny of the genetic background among the ILs, which in turn favours the QTL mapping.

## 8. Segregation Distortion.

Segregation Distortion (SD), or Transmission Ratio Distortion (TRD), describes a natural occurring phenomenon defined by a deviation in the genotype/allele frequency of a locus from the

expected Mendelian ratio (Koide *et al.*, 2012). Since SD is an irregularity to the proper Mendelian segregation and changes the frequency of alleles in a population it is considered one important evolutionary force (Xu, 2008) and a factor for the emergence of reproductive barriers (Hurst and Pomiankowski, 1991).

Genetic elements that alter the common Mendelian segregation and favour their transmission are called 'selfish' genes, and in the case of SD genetic elements are also called segregation distorters (Taylor *et al.*, 2003). The best described examples of SD are the segregation distorter and X-linked meiotic drive in *Drosophila* species (Sandler *et al.*, 1959; Lyttle, 1993) and the t allele system in mice (Silver, 1993).

In plants, the SD phenomenon was first observed in maize (Mangelsdorf and Jones, 1926), and later tomato (Rick 1963), rice (Nakagahra, 1972), wheat (Endo, 1990), coffee (Ky *et al.*, 2000), chickpea (Castro *et al.*, 2011), and cotton (Dai *et al.*, 2016).

Since the SD phenomenon is so widespread along different taxa and the mechanisms by which occurs share many features, one would assume very few ways for the phenomenon to occur exist but, alternatively, many ways of SD occurring exist but we are only capable of detecting the few that share common features (Taylor *et al.*, 2003). Additionally, when a 'selfish' allele is relatively new to a population it has more competition by the wild-type alleles but, when it occupies a large part of the genetic pool of a population it is transmitted at a much high ratio than the wild-type alleles. This is even more evident in the cases where the SD effect is caused by male gametes. If the SD allele has a deleterious effect on male gametes that do not contain that allele, it is transmitted at a substantially higher rate comparatively to others (Taylor *et al.*, 2003).

### **8.1 Influencing physiological and genetic factors and causes of segregation distortion.**

The appearance of SD in a trait or marker can be due to several factors, including genetic and environmental factors or a combination of both (Xu *et al.*, 1997). The genetic factors range from chromosomal structure differences to allelic incompatibility (Rick, 1966). Environmental factors are mostly related to selective pressure on/ expression of alleles caused by the environment (Wang *et al.*, 2005). Either way, the combination of both factors should be taken in account when studying the SD phenomenon.

SD could be due to different mechanisms acting at different stages of the plant life cycle. Shortly, SD could arise from alterations in gametes or by mechanisms after fecundation. Therefore, the causes can be classified as pre-zygotic or post-zygotic, or both.

Pre-zygotic causes indirectly affect the allele ratios observed in the progeny by a direct effect on the proportion of gametes i.e. provoked by the abortion of male gametes (Lin *et al.*, 1992), female gametes (Longley, 1945), or both (Rick, 1966). Pre-zygotic causes are mostly related to male gametes (Xu *et al.*, 1997). Some examples of pre-zygotic factors are the pollen killer genes in *Nicotiana* (Cameron and Moav 1957) and wheat (Loegering and Sear, 1963), gamete eliminator alleles in tomato (Rick, 1966), the gametocidal factor in wheat (Endo, 2007), the pollinic tube capacity of maize (Mangesldorf and Jones, 1926), and female meiotic drive in *Mimulus* (Finseth *et al.* 2015).

Post-zygotic distortion are the ones that occur after fertilization by directly influencing zygotic ratio or zygotic viability. They are controlled mainly by maternal factors and can occur during the maternal-zygotic transition or during embryogenesis after expression of embryo develop related genes (Xu *et al.*, 2013). Causes vary from incompatibility with the endosperm to low fitness or sterile seeds (Barrantes, 2014).

## **8.2 SD in tomato species**

In the tomato clade, the frequency by which phenomena like the SD occurs is related with the phylogenetical distance of the species. The appearance of SD is more frequent in distant species (Anderson *et al.*, 2010).

Just to point some examples, in crosses of *S. arcanum* with *S. lycopersicum*, Fulton *et al.* (1997) found a preference in the inheritance of *S. arcanum* alleles in chromosome 9 and attributed the occurrence to a gamete promoter gene (*Gp*). In a  $F_2$  population from the *S. pennellii* x *S. lycopersicum* cross 80% of the markers favoured wild alleles (de Vicente and Tanksley, 1993). And, 15% was reported in a  $BC_1$  population of *S. habrochaites* x *S. lycopersicum* (Bernacchi and Tanksley, 1997).

One case that has special interest for this thesis is the sex-independent SD mapping to the proximal heterochromatin region of chromosome 4 described by Rick (1966) after analysing several crosses between *S. pimpinellifolium* and *S. lycopersicum*. The underlying gene was proposed to be part of a triallelic system dubbed the *Gamete eliminator* (*Ge*). The allele *Ge*, predominantly found in *S. pimpinellifolium* accessions, is preferably selected over the *Ge* allele,

found in *S. lycopersicum*, in heterozygotic plants ( $Ge^e/Ge^e$ ). The phenomenon was independent of the sex of the gametes and occurred with a penetration around 95%. The third allele,  $Ge^e$ , was considered neutral and did not upset the segregation ratio of the other alleles. Other genes with similar effects were found in other cultivated plant species: the pollen killer of *Nicotiana tabacum* (Cameron and Moav 1957); the *Ki* gene (Loegering and Sear, 1963) and the *Gc2* gene (Finch et al. 1984; Endo 1990) in *T. aestivum vulgare*; and the *S1* and *S2* genes of *Oryza sativa* (Sano et al. 1979; Sano 1983).

From the several examples of SD reported during the mapping of tomato species, most are related to self-incompatibility (SI) or unilateral incompatibility (UI) between the two species being crossed (Barrantes *et al.*, 2014).

It was discussed previously that there is a great diversity of mating systems among tomato species. From autogamous self-compatible (SC) to outcrossing self-incompatible, with some species having facultative mating systems (SC/SI) (Figure 1).

SI is a mechanism of cell-cell communication where the pistil can recognize and reject pollen coming from itself thus preventing inbreeding. UI averts hybridization of closely related species when the male parent is SC and the female parent is SI, however in the reciprocal cross (i.e. ♂ SI x ♀ SC) the pollen rejection does not occur. (Kondo *et al.*, 2002). Therefore, one of the major factors for the achievement or failure of interspecific crosses in the tomato clade is the pollen-style interaction (Bedinger *et al.*, 2011).

Well described mechanisms of SI in *Solanum* involve the genetic control by the S-locus (Bernacchi and Tanksley, 1997; Covey *et al.*, 2010) and S-locus modifiers (Kondo *et al.*, 2002).

Two genes are localized in the S-locus. One that is specifically expressed in the pistil codes for a ribonuclease (S-RNase), and the other in pollen codes for F-Box (SLF) protein. S-RNases control the specificity of pollen rejection in conjunction with other pistil factors by arresting pollinic tube growth in case of compatibility. Pollen is rejected only when the S-haplotype of S-RNase/SFL is identical to either one of the two S-haplotypes in the diploid pistil (Tovar-Méndez *et al.*, 2014).

The best described S-locus modifiers are the HT genes. These modifiers are necessary for SI function but not for its specificity. The HT-A is a stylar expressed asparagine-rich protein and HT-B a protease inhibitor. (Covey *et al.*, 2010; Kondo *et al.*, 2002)

The lost SI in the red/orange fruited species *S. lycopersicum*, *S. pimpinellifolium*, *S. galapagense* and *S. cheesmaniae* (Figure 1) is attributed to mutations in HT genes and/or S-RNase genes (Kondo *et al.*, 2002).

UI is commonly observed in crosses between *S. lycopersicum* and SI tomato species. The molecular mechanism underlying this UI by pollen rejection occur in S-RNase-dependent and S-RNase-independent way. Recently, Tovar-Méndez *et al.* (2017) reported that HT proteins are involved in both S-RNase-dependent and S-RNase-independent mechanisms with crosses between *S. lycopersicum* pollen and *S. pennellii*, *S. habrochaites*, and *S. arcanum* SC accessions. They also stated that an additive effect from multiple pollen–pistil interactions result in the overall compatibility.

### **8.3 Mapping QTL markers linked to distorted segregation.**

SD is often observed during the development of mapping populations. SD is known to affect the estimation of recombination distance between the markers which influences mapping precision and linkage analysis of QTLs (Wang *et al.*, 2005). SD is observed in all types of recombinant populations like F<sub>2</sub>, BC, double haploid (DH), RILs and ILs. The closer the marker to the SD loci bigger is the effect of distortion in the linkage analysis (Xu *et al.*, 1997).

The actual cause of the observed SD in mapping are genes subjected to gametic or zygotic selection. The molecular markers appear to be distorted by their linkage to the segregation distortion loci (SDL) (Xu, 2008). This only presents a setback in QTL mapping because most mathematical models used assume Mendelian segregation. However, ignoring SDL in QTL mapping results in only a slight power loss. Nevertheless, experimental and statistic methods that account for the influence of SDL and benefit the quality of the data have been proposed (Wang *et al.*, 2005; Xu, 2008). Contrary to traditional QTL mapping, where SDLs are only detected by linkage to neutral markers, in these methods a multipoint mapping of SDLs is accomplished before the linkage map is done, eliminating the bias in the statistical analysis. The degree by which the linkage and estimation of genetic distance is affected depends on the degree of dominance of the markers used and the type of cross. Lorieux *et al.* (1995a, 1995b) demonstrated that in backcross populations, when the SD is caused by a single gametophytic factor, the linkage and mapping are unaffected but the order of the linkage groups may be. For example, Lu *et al.* (2002) did not find effects of SD regions in mapping distances when identifying chromosomal regions associated with SD in maize since only one gametophytic factor was present.

Some works accomplished the mapping of both QTLs and SDL. Yamanaka *et al.* (2001) was able to map 4 QTLs for flowering time, 6 for leaflet morphology and 18 SD regions on soybean using 503 markers.

## 8.4 Segregation distortion in TO-937 derived populations.

During the development of the IL library (Barrantes 2014) a non-target introgression, located in the distal region of the q arm of chromosome 4, was observed in 51% of the IL. This non-target introgression favoured the *S. pimpinellifolium* in detriment of the alleles of the recurrent parent “Moneymaker”. Barrantes (2014) fine-mapped the distortion segregation locus to a small 84 Kb region of chromosome 4 (Figure 6).

In 2016, Fakhret made further studies on this region to define the mechanism underlying the SD of chromosome 4, increasing the mapping resolution and to identify if the phenomenon is a gametic and/or zygotic by testing the progeny of reciprocal crosses between heterozygotic plants for the region and homozygotic parents (homozygotic for TO-937 and “Moneymaker” alleles).

Fakhret (2016) was able to narrow down the SD region to 39Kb (Figure 6). This new mapped SD region, contains 7 candidate genes, 5 of those functionally annotated and two with unknown function. A preliminary *in silico* expression analyses suggested 3 genes as the strongest candidates for being responsible for the SD. Those were the unknown protein coding gene Solyc04g081620 and two HSP90 genes, Solyc04g081630 and Solyc04g081640. The tested progeny of the reciprocal crosses implied for a sex independent phenomena with involvement of post-gametic elements for the mechanism of SD.

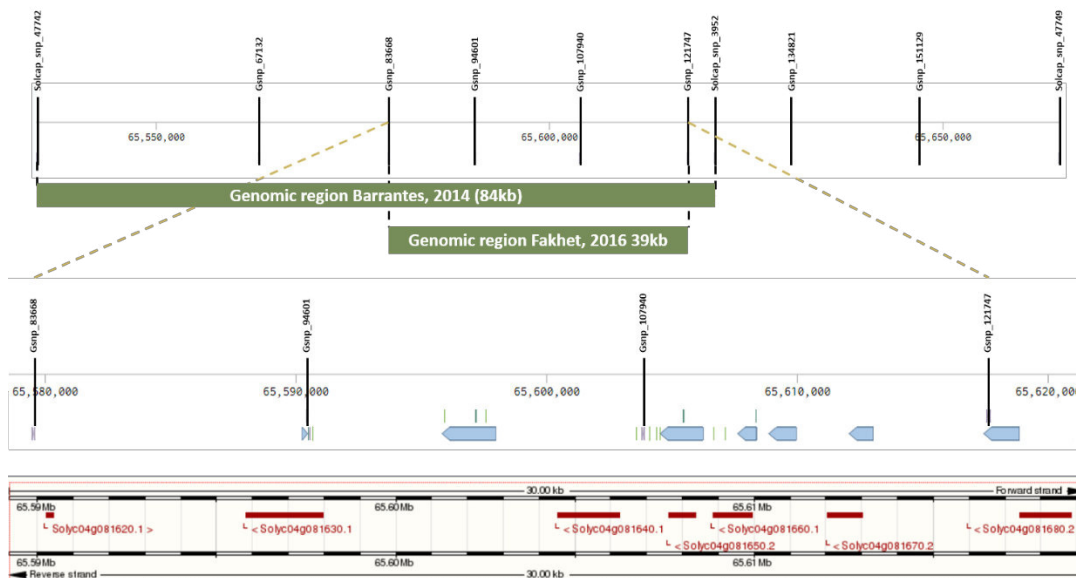


Figure 6 – Schematic representation of the distal portion of chromosome 4 of tomato. Top: genomic regions identified by Barrantes (2014) and Fakhret (2016). 84Kb region flanked by the Solcap\_snp\_3952 and Solcap\_snp\_47742 markers and 39Kb region flanked by the Gsnp\_83668 and Gsnp\_121747. Bottom: location of the 7 candidate genes contained in the SD region identified by Fakhret.



The present work aims, first, to structurally and functionally characterize the candidate genes in the region by evaluating their expression profiles during flower development and fruit set and by sequencing from both the *S. pimpinellifolium* acc. TO-937 and *S. lycopersicum* var. “Moneymaker”. Second, to investigate the natural variability around the SD region in wild accessions showing or not SD. Third, study the effect of gametic and zygotic factors on SD. And fourth, develop tools based on the CRISPR/Cas9 system and the GoldenBraid 3.0 standard assembly toolkit, to assess gene function.

## **9. Defining landmarks for stages in tomato flowering and fruit development.**

Studying the flower/fruit in terms of development, morphology and gene expression patterns helps to identify how structures are preserved or diversified, the sequences involved in the development, the evolution of genetic sequences and, most important, gene function. To largely enhance the knowledge about the genes involved in flower/fruit development relating morphology development and genetics is crucial to have a way to comparatively study several plants from several taxa (Buzgo *et al.*, 2017). The multitude of flower and fruit structures among the plant kingdom creates a setback in the comparative study of floral/fruit development because the morphology and development of the flowers in different species can be remarkably diverse. Also, the same can be said about the genetic factor, although some gene families related to flower morphology are highly conserved, like the MADS-box genes (Ng & Yanofsky, 2001).

In what concerns tomato species, schedules for evaluating floral/fruit development were established (Brukhin *et al.*, 2003) and molecular comparative studies of the Solanaceae family were done (Tanksley *et al.* 2004), but the adoption of common landmarks, i.e., a consonance in the delimiting of the floral/fruit development stages, is yet to be put in place effectively.

Flower and fruit development stages (‘landmarks’) are very useful to study gene function in what concerns the temporal, spatial and tissue-specific expression of genes related to organ development (Brukhin *et al.*, 2003).

The following descriptions of flower and fruit development landmarks were established by Buzgo *et al.* (2004) and adopted for tomato species by Xiao *et al.* (2009).

Flower and fruit development have each 10 distinct landmarks. The last landmark of the flower development is also the first landmark of fruit development, which is called anthesis, and marks the turning point between the two.

The formation of the inflorescence and flower initiation is the first established landmark of flower development. The inflorescence formation begins with the transition of the vegetative meristem into an inflorescence meristem. The flower initiation begins after the division of the inflorescence meristem that generates the first flower bud and continues with the other flower buds.

The next landmark is the sepal initiation (aka, initiation of outermost perianth organs). Is defined by the appearing of the sepal primordia around the flower apex of the second flower, forming a helical pattern of  $144^\circ$  with five petals. This stage finishes as sepals cover the flower meristem 4 days after flower initiation. Next the petals start to grow in what is the petal initiation landmark (aka, initiation of inner perianth organs). Five days after flower initiation the stamen primordia starts to develop which marks the stamen initiation landmark. The sepal and petal continue to lengthen during this period. On the 6<sup>th</sup> day the emergence of the carpel marks the carpel initiation landmark. On the 7<sup>th</sup> day after flower initiation the central column that will form the locular cavities arise. During this the stamen filament starts to develop and the two anther lobes appear.

The next landmark, microsporangia formation, marks the beginning of the development of the male reproductive organs on the 8<sup>th</sup> day after floral bud initiation. The primary sporogenous layers are showing, the tapetal cells are binucleate and the microsporocytes are visible at the 9<sup>th</sup> day. The central column elongates and the carpel fuses at the apex of the ovary. Also, the placental development starts. The next landmark is, in fact, the ovule initiation, when the primordial ovule begins to emerge from the placenta.

From this point on, male and female meiosis occur parallelly. Figure 7 presents a schematic timeline of floral development landmarks.

Male meiosis (8<sup>th</sup> landmark) happens 10 days after floral bud initiation when microsporocytes (pollen mother cells) undergo meiotic division. The haploid nuclei of the tetrads are surrounded by a callose wall (polysaccharide wall). The day after the wall breaks and microspores start to release. On the 13<sup>th</sup> day the tapetum is degrading and the microspore are singly divided but encapsulated in a thick wall. The day after, the single spores show vacuoles and go through an asymmetric mitosis. On the 15<sup>th</sup> day the microspores are bi-cellular and in the next day there is a complete differentiation of the generative and vegetative cells. On the 17<sup>th</sup> day the nucleus of the generative cells shows a crescent shape. The second mitosis of the germ cells is only accomplished after pollination.

The 7<sup>th</sup> landmark, ovule initiation, also marks the beginning of the development of female structures. Nine days after flower bud initiation the style and ovule are proximally equal in length and ovule primordia emerges. One day later the spores are completely visible and, on the eleventh day after floral bud initiation, the megasporocyte undergoes the first meiotic division. The day after, the second meiotic division occurs marking the first stage of megagametogenesis. On the 14<sup>th</sup> day the egg apparatus development is marked by the presence of the megaspore at the chalaza end of the ovule. On the 16<sup>th</sup> day the female gamete is almost completely developed. Everything described since the 7<sup>th</sup> landmark comprises the development during the 9<sup>th</sup> landmark, female meiosis.

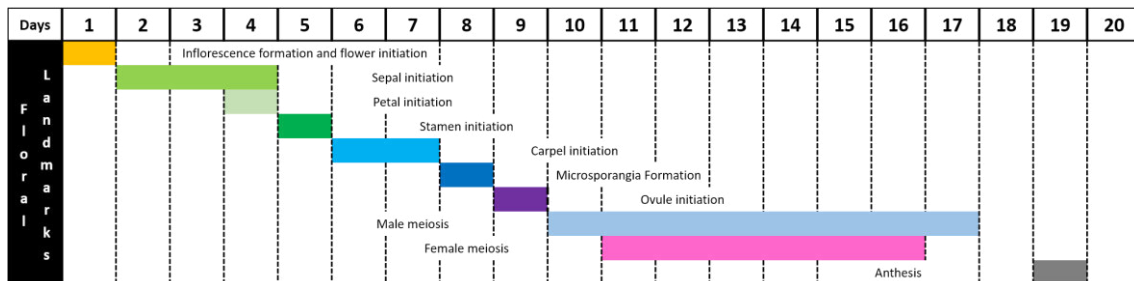


Figure 7 – Chronological representation of the 10 landmarks of floral development for tomato species. Day one (1) is the day when the floral bud splits from the inflorescence meristem and day twenty (20) the last day when anthesis normally occurs.

The tenth and last landmark of flower development is the anthesis. anthesis is both the final landmark of floral development and the first landmark of fruit development.

At the time of this stage, the dehiscence of the anther enables the release of the mature pollen. The pollen that lands on the receptive stigma starts germination and, 6 hours after, reaches the base of the stigma and two hours later the ovules. Ten to twelve hours after pollination the fertilization of the ovule is complete. Fertilization is the second landmark of the fruit development. The senescence of floral organs (petal, stamens and style) evidences a successful fertilization and occurs, normally, 48 hours after anthesis. After fertilization, the cell division and expansion starts. The first of five landmarks of embryo formation, and third in fruit develop landmarks is the 4-16 cell stage embryo. This stage is characterized by a rapid cell division and cell elongation. The growth of the pericarp during these next stages is marked by overlapping of cell division and cell elongation. The 4<sup>th</sup> landmark is the Globular stage embryo, it occurs between the 6<sup>th</sup> and 10<sup>th</sup> day post anthesis (DPA). The 5<sup>th</sup> landmark is the Heart stage embryo. It takes only one day and normally happens between the 10<sup>th</sup> and 12<sup>th</sup> DPA. The Torpedo stage embryo is the 6<sup>th</sup> landmark and is characterized by cotyledon elongation and the start of exponential fruit growth. This stage takes

only one day to occur but occurs between the 13<sup>th</sup> and 16<sup>th</sup> DPA. The last stage of embryo formation, and the 7<sup>th</sup> landmark of fruit develop is called Coiled stage embryo. At this stage, the cotyledon curl inside the seed as they elongate, hence the name. At the end of this stage the embryo is fully mature although the seed is not viable.

The next period, between the 20<sup>th</sup> and 28<sup>th</sup> DPA, is not considered a landmark but is during this time that the seed mature. The 8<sup>th</sup> landmark is seed germination, the fruit is at the end of the mature green stage. The fruit responds to ethylene stimuli when seeds are viable. On the 33<sup>rd</sup> DPA, the fruit enters on what is called the breaker. This is the 9<sup>th</sup> landmark, fruit ripening, the change in pigmentation, from green to red, is clearly visible and all seed are viable. The last and 10<sup>th</sup> stage is the ripe fruit, when the fruit presents itself fully red. Figure 8 depicts the timeline of fruit development, as opposed to flower development is much harder to clearly delimitate when one landmark ends and the other begins.

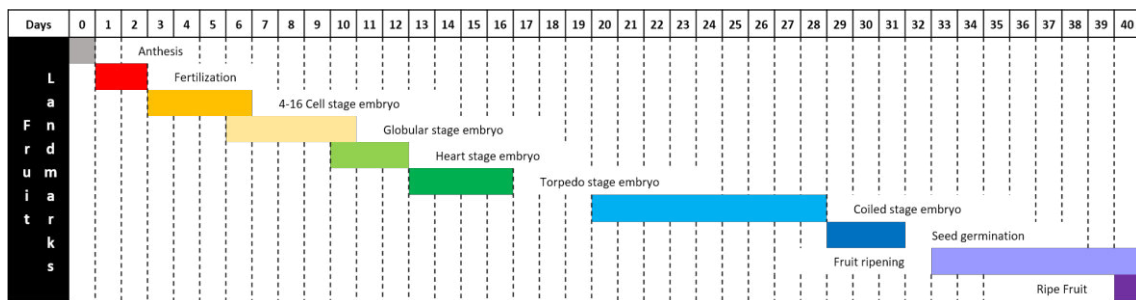


Figure 8 – Chronological representation of the 10 landmarks of fruit development for tomato species. Day zero (0) is the day of anthesis. The days marked are days post anthesis (DPA).

## 10. Genome engineering in crops and the CRISPR/Cas9 system.

The Council for Agricultural Science and Technology (CAST) announced this year (2017) that current agricultural yields will fail to supply food, feed, and fuel before we reach the year of 2050. Being crops the main food and feed resource around the globe, there is a demand for modern technologies that can rapidly improve crop development (Baenziger *et al.*, 2017).

Genome engineering is the process of creating targeted modifications to: the genome by altering DNA coding sequences, outputs by tinkering with the transcripts, and changing the genomic context by shifting epigenetic marks (Hsu *et al.*, 2014).

For years, genome engineering in plants relied on expensive, time-consuming, and unspecific methods, like the application of chemical treatments or radiation, to produce mutants in random genes and, after, screening for desirable traits (Zhang *et al.*, 2014).

For the fields of life science, biotechnology and crop development, the application of technologies that can change specific DNA sequences is a crucial leap forward from previous methods. The deletion, insertion, or modification of the genome of single cells or organism creates an opportunity to access the function of genes and regulatory elements and consequentially helps the quick improvement of crop varieties (Hsu *et al.*, 2014).

In recent years, some important technologies that satisfy the need for a more specific and quicker ways to edit DNA emerged. These are the site-specific nucleases like zinc-finger nucleases (ZNFs), transcription activator-like (TAL) effector nucleases (TALENs). (Wood *et al.*, 2011)

These two technologies can create double strand breaks (DSBs) on specific sequences and, as such, were a significant improvement for site-specific genome editing. (Wood *et al.*, 2011; Ma *et al.*, 2017)

The creation of DSBs in a precise location of the genome activates the cell DNA repairing pathways to act in the desired location. The cell then activates one of two major pathways for DNA damage repair. In eukaryotic cells, DSBs are most commonly repaired by non-homologous end-joining (NHEJ) that occasionally leaves reparation scars on the DNA sequence (Pan *et al.*, 2016). These scars are insertion/deletion mutations (InDel). The NHEJ pathway can be used to produce gene knockouts since InDels occurring in a coding sequence can change the reading-frame and introduce frameshift mutations and early stop codons. (Shalem *et al.*, 2015) Additionally, multiple DSBs can originate large genomic deletions eliminating several genes and regulatory elements. (Cong *et al.*, 2013)

The other major pathway is the homology-directed repair (HDR). Contrary to NHEJ, HDR is mostly active in dividing cells and its efficiency can vary depending on the cell type but it has an important advantage to NHEJ in what concerns genome engineering. The repairing of a DSB by HDR allows for the modification of the targeted sequence by providing the cell with an exogenous DNA template for repair. (Ran *et al.*, 2013) The cell machinery usually uses the sister chromatid to repair the DSB but by introducing a desired sequence template small modifications can be accomplished. (Hsu *et al.*, 2014)

ZFNs and TALENs use customizable modular DNA-binding proteins combined with *FoII* endonuclease catalytic domains to produce the DSBs on the targeted sequence (Wood *et al.*, 2011; Ran *et al.*, 2013) and researchers take advantage of the cell DNA repairing pathways to introducing desired changes. These two technologies rely on the ability of personalized proteins to bind to specific DNA targets and necessarily require an extensive screening process. (Hsu *et al.*, 2014)

However, a more recent technology that relies on Watson-Crick base pairing with the target DNA appeared. This technology is the CRISPR/Cas9 system. CRISPR stands for clustered regularly interspaced short palindromic repeats and is a microbial adaptive immune system based on RNA-guides that is widespread in many bacterial and archaea species (Ran *et al.*, 2013). Until now, three CRISPR systems have been identified, namely type I, II, and III. Each one of these types comprises a cluster of CRISPR-associated (Cas) genes, noncoding RNAs, and an array of repetitive elements (Ran *et al.*, 2013). The function of these repetitive elements is to space short variable sequences of exogenous DNA known as protospacers (Makarova *et al.*, 2011). This protospacers contain matching sequences to phages and plasmids and provide acquired immunity against an ensuing exposure to the same sequences that those sources of exogenous DNA have. (Sapranauskas *et al.*, 2011) The group of protospacers is designated as CRISPR RNA (crRNA) array (Ran *et al.*, 2013). A crRNA unit has a 20-nucleotide guide sequence (guide RNA) that pairs to the target DNA (Ran *et al.*, 2013). In the engineered version of the system is up to the user to defined the 20-nt sequence suited for the desired application. The parameters to take in account when designing the guide RNA are discussed later in this work since they can vary depending on how the system will be applied.

The most well studied CRISPR system is the type II (Ran *et al.*, 2013; Sapranauskas *et al.*, 2011). The type II system has a Cas9 nuclease, the crRNA array that encodes for the guide RNAs and an auxiliary trans-activating crRNA (tracrRNA) for the processing of the crRNA array. For practical application of the engineered CRISPR/Cas9 system in genome editing the crRNA 3' end was fused to the tracrRNA 5' end to produce a single guide RNA (sgRNA) (Jinek *et al.*, 2012). The type II system uses just a single Cas9 protein and can be applied with the sgRNA so it was the best candidate to use in genome editing (Pan *et al.*, 2016).

Cas9 nucleases are characterized by their six functional domains: HNH, RuvC, REC I, REC II, Bridge Helix, and PAM-Interacting (Jinek *et al.*, 2014; Nishimasu *et al.*, 2014). Although, protein length is highly variable from species to species the domains of Cas9 proteins maintain a similar architecture (Jinek *et al.*, 2012). The variation is caused mainly by variable conservation of the REC domains (Hsu *et al.*, 2014; Jinek *et al.*, 2014). The HNH, RuvC domains are nuclease domains, therefore, responsible for cleavage of the DNA (Jinek *et al.*, 2014; Nishimasu *et al.*, 2014). REC I and REC II domains seem to be involved in the bidding to the guide RNA, and the Bridge Helix for initiating cleavage activity after target DNA bidding (Nishimasu *et al.*, 2014). Finally, PAM-Interacting is responsible for DNA bidding and gives the PAM specificity of the Cas9. The Cas9 protein remains

inactive until the binding of the crRNA:tracrRNA guide RNA (Jinek *et al.*, 2014) or the sgRNA in the engineered CRISPR system (Jinek *et al.*, 2012; Nishimasu *et al.*, 2014). The binding promotes a conformational change and activation of the Cas9 (Figure 9).

Upon activation by conformational change the complex looks for the target DNA sequence by binding first to the protospacer adjacent motif (PAM).

The PAM flanks the 3' end of the DNA target site but it is not included in the guide RNA sequence. In the adaptive immunity of bacteria and archaea the reconnaissance of PAM sequences is useful to discriminate endogenous from exogenous DNA (Shah *et al.*, 2013).

Several PAM sequences exist depending on the CRISPR of the host species. The canonical form of the PAM sequence is the 5'-NGG-3' (Nishimasu *et al.*, 2014). It is the PAM sequence of the *Streptococcus pyogenes* Cas9 and it was the one considered when designing the guide RNAs in this work.

It is assumed that the RNA-DNA heteroduplex begins at the PAM site and continues to the 5' end of the RNA to check for sequence complementarity (Sternberg *et al.*, 2014). The binding of the PAM and a complementarity of the target DNA to the guide RNA activates the nuclease activity of the HNH and RuvC domains that cleaved the DNA at about 3bp upstream of the PAM (Nishimasu *et al.*, 2014). One important aspect to take in account when designing guide RNA is that the complementarity does not have to be total for the Cas9 to cleave a similar target sequence. What this means is that a small fraction of off-target mutations can occur (Wu *et al.*, 2014).

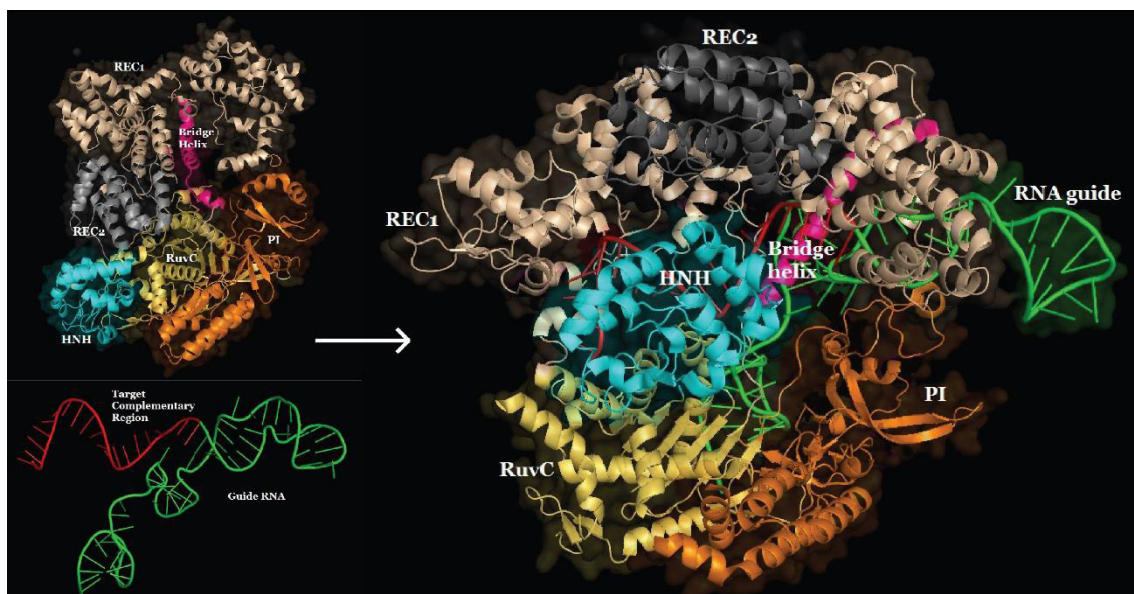


Figure 9 – Conformational structure change and activation upon binding of the guide RNA to the Cas9 protein. Crystal imaged from Anders *et al.*, 2014.

In what concerns the engineering application of the CRISPR/Cas9 system it is possible to change the PAM specificity by replacing the PAM-interacting domain by an ortholog enabling the targeting of other genomic regions (Nishimasu *et al.*, 2014). This also means that other structural domains could possibly be changed by orthologs to optimize Cas9 for other parameters such as DNA binding, cleavage or even protein size and aminoacidic sequence (Hsu *et al.*, 2014).

After understanding how the CRISPR/Cas9 system functions it is clear the advantages that this technology has versus the ZFNs and TALENs (Ran *et al.*, 2013). For starters, ZFNs and TALENs are difficult to engineer and more time-consuming since it requires the design and assembly of whole new proteins. Second, the efficiency of Cas9 is greater than that of the DNA-binding proteins hence reducing the screening time for the desired mutation. And third, since the customization of the guide RNA is simpler and it relies on the conjugation of the Cas9 nuclease to the sgRNA it is possible to target several genomic loci at once by providing one Cas9 gene and multiple sgRNAs, in what is called a multiplex approach (Ma *et al.*, 2017; Li *et al.*, 2017; Cong *et al.*, 2013).

In plant research, CRISPR/Cas9-based genomic engineering was proven to be efficient in several species like the plant model *Arabidopsis thaliana* (Jiang *et al.*, 2013; Mao *et al.*, 2013), important crops such as, tobacco (Jiang *et al.*, 2013), rice (Jiang *et al.*, 2013; Zhang *et al.*, 2014), tomato (Brooks *et al.*, 2014; Li *et al.*, 2017), and soybean (Sun *et al.*, 2015), and in other species for example *Citrus sinensis* (Jia *et al.*, 2014), *Marchantia polymorpha* (Sugano *et al.*, 2014), and *Populus* (Fan *et al.*, 2015).

For application in plants, CRISPR/Cas9-based genomic engineering can be done by incorporating the Cas9 coding sequence and the desired sgRNA in a T-DNA cassette for stable transformation (Vázquez-Vilar *et al.*, 2016) or transient expression (Ron *et al.*, 2014). This opens new possibilities to overcome the regulation of transgenic plants in Europe. Since European regulation focuses on the process by which the plant is transformed conventional mutagenesis and genetic engineering could possibly be classified under different definitions, especially in the case when using CRISPR/Cas9 only a few nucleotides are excised from the genome (Bortesi *et al.*, 2015).

One of the many possible applications of the CRISPR/Cas9 system in plant biology and biotechnology is the loss-of-function (i.e. knock-out) to determine gene function. Until the advent of the CRISPR/Cas9 system the RNA interference (RNAi) was the preferred method for assessing gene function. Even though RNAi is a useful method it has some important drawbacks compared to the CRISPR/Cas9 system. First, by interfering with RNA transcripts what occurs is a knock-down



of expression and not a complete silencing of the gene. Second, RNAi, being less complex, has a lot more off-target silencing that affects the desired outcome (Xu *et al.*, 2006). As explained before, the use of CRISPR/Cas9 really creates a knock-out by targeting genomic loci (Shalem *et al.*, 2015).

Other than loss- or gain-of-function experiments, the specific binding of the Cas9 to a specific target can be explored for other gene characterization procedures in plant biology. For example, by inactivating the catalytic domains of the Cas9, what is called a dead Cas9 (dCas9), it is possible to create a RNA-guided homing device. (Hsu *et al.*, 2014) The dCas9 can then be associated to transcriptional repressors/activators to modulate gene expression (Piatek *et al.*, 2015; Gilbert *et al.*, 2013), or to fluorescence protein to study chromosome structure and to visualize genomic loci (Anton *et al.*, 2014).

Recently, in the field of crop development, Rodríguez-Leal *et al.* (2017), demonstrated that CRISPR/Cas9-based genome editing of promoters in tomato can generate *cis*-regulatory alleles that provide a continuum of variation. This process is an acceleration process to the conventional domestication of quantitative trait loci (QTL).

The list of possible applications of the CRISPR/Cas9 system is already quite extensive and compasses several fields of life science but the efficiency, simplicity and standardization of this system will bring more innovative applications and accelerate the processes of gene discovery and characterization.

## **11. Synthetic Biology and standard assembly.**

The definition of what synthetic biology is remains a debated subject since it concerns whole gene systems, instead of focusing in specific genes or pathways, making this an interdisciplinary field of study, and because it includes a wide variety of methodologies for the manipulation of living systems (Andrianantoandro *et al.*, 2006; Balmes & Voytas, 2015). It is common to make analogies between Synthetic biology and mechanical, electrical, and computer engineering (Andrianantoandro *et al.*, 2006; Khalil & Collins, 2010). This happens because the engineering philosophy of modularization, rationalization, and modelling (Khalil & Collins, 2010) is present in the biological counterpart. So, one widely accepted definition is one from Endy (2005), stating that «Synthetic biology aims to use modular, well-characterised biological parts to predictably construct novel genetic devices and complex cell-based systems following engineering principles». In short, the objective of synthetic biology is to create or modify living beings, conferring them characteristics that do not exist naturally (Haseloff & Ajioka, 2009). Synthetic biology, not

exclusive to plant synthetic biology, presents the possibility of changing and enhancing complex biological systems either to produce primary products like biomass and food but also in secondary applications such as pharmaceuticals, fuels and polymers (Haseloff & Ajioka, 2009 and Patron *et al.*, 2015).

The field of plant synthetic biology shows great potential not only in the production of primary metabolites but also in what concerns secondary metabolites. Additionally, a lot of the restraints presented with this type of work in animal cells do not concern plant studies. Especially, the ethical restraints but also the easy and less expensive production of plant tissues (Baltes & Voytas, 2015).

As previously stated, one of the analogies between synthetic plant biology and engineering is the modularization of components (Khalil & Collins, 2010). In the engineering case, the standardization of components (e.g. mechanical parts, electronic circuits, etc.) allowed a significant increase in the speed of innovation and decrease in production costs. This standardization of parts, interexchange and reusability of those parts is what defines the basis of synthetic biology (Patron *et al.*, 2015). About standardization, having a simpler system eases not only the adoption of the technology by users, but also helps in the automation and efficiency of the assembly (Sarrion-Perdigones *et al.*, 2011). The first assembly standard with the simplistic principle of idempotency was the BioBrick assembly standard 10 (Shetty *et al.*, 2008, Knight, 2003). Idempotency means that the resulting product of an assembly has the same key structure elements as the parts that created it and so can be used as an input in another assembly, i.e. reused (Patron *et al.*, 2015, Sarrion-Perdigones *et al.*, 2011). Although BioBrick allows for idempotency it has the drawback of only being capable of binary assemblies which slows the developments process of new standard parts.

Several assembly methods enable the multipart assembly but lack in the capability of making idempotent modules and, for that reason, most of them are not adopted by biotechnologists (Sarrion-Perdigones *et al.*, 2013). A good example is the Gibson assembly (Gibson *et al.*, 2009). On the other hand, there is the Gateway cloning system (Karimi *et al.*, 2002), that allows for multipartite assemblies but, since it involves recombination of specific sequences that delimit the module leave scars in the assembly.

As seen so far, a good assembly standard for plant synthetic biology must have several features, all at the same time. Shortly, and defined by Sarrion-Perdigones *et al.* (2013), an ideal system is not only fast and efficient but has to follow the principle of idempotency, allowing the

reusability of new parts enabling unlimited growth of the library of parts. Also, it is crucial that the assembly method does not leave scars in the assembly, or at least is scar-benign. And, most importantly, it needs to be defined by a strict set of rules so it can be used by as many biotechnologist, since this is what defines what a standard is.

One DNA assembly technology that satisfies almost all the before mentioned characteristics is the Golden Gate technology (Engler *et al.*, 2008 and 2009). This technology is such a good candidate for the development of a standard because it can perform multipartite assemblies with high efficiency and leave no scars in the assembly. Instead of using recombination sites like Gateway (Karimi *et al.*, 2002), Golden Gate uses type IIS restriction enzymes to produce 4 nucleotide sticky-end overhangs that can be user defined. Opposite to type II restriction enzymes, type IIS recognize non-palindromic sequences and cut outside the recognition sequences at a defined distance (Pingoud *et al.*, 2001). This feature allows for the recognition sequence to be maintained and the endonuclease activity to be done in an adjacent sequence that can be tailored. The definition of the tailored overhangs also helps in the assembly of multiple parts in a determined order and orientation in a single step of assembly using just one endonuclease (Patron *et al.*, 2015). What Golden Gate lacks for it to be a standard assembly method is the ability for the assembly products to be reused in another assembly (Sarrion-Perdigones *et al.*, 2011).

To overcome this handicap, two strategies that make use of the Golden Gate assembly system were developed, namely, the Golden Gate Modular Cloning (MoClo) (Weber *et al.*, 2011) and GoldenBraid (GB) (Sarrion-Perdigones *et al.*, 2011 and 2013). Although they use the same assembly system and many parts can be interchangeable, the two systems are not fully compatible (Patron *et al.*, 2015). Since both standards share the same syntax and grammar structure those are to be presented in the GB perspective as this was the standard used in this project.

### **11.1 GoldenBraid 3.0.**

GoldenBraid (GB3.0) is standard module DNA assembly technology for Plant Synthetic biology that is founded in the characterization of modules, exchangeability of parts, and easy and cost-effective assembly. GB3.0 allows the creation of shuttle vectors, used in *Agrobacterium*-mediated transformation, containing single transcription units or modules comprising multigenic assemblies (Sarrion-Perdigones *et al.*, 2011 and 2013). The first version of GB (GB1.0) was an attempt to convert the single-use Golden Gate multipartite assembly in the construction of standard parts for use in synthetic biology to a reusable system, by allowing the binary assembly of

multipartite constructs in a double loop cloning strategy (Sarrion-Perdigones *et al.*, 2011). The second, GB2.0, was developed in consonance with the MoClo strategy for the implementation of a common standard enabling the exchangeability of parts. Also, it saw the introduction of several changes to the first version that facilitated the adoption of this tool in the field of Plant Synthetic Biology (Sarrion-Perdigones *et al.*, 2013).

GoldenBraid uses *BsaI* and *BsmBI* type IIS endonucleases, from the Golden Gate system, in the assembly of TU and genetic modules. These enzymes recognize 6bp sequences and leave tailored 4 nucleotides long 5' overhangs (Weber *et al.*, 2011). These tailored overhangs are obtained by putting the recognition and digestion sites in opposite orientation, flanking the desired insert, meaning that the recognition sites disappear after assembly (Sarrion-Perdigones *et al.*, 2011). The standardization of the overhangs enables the multipartite assembly of each GBpart on the correct position (Figure 10). Several contiguous GBparts can be assembled in a GB Superpart (GBSpert) and more practically assembled, for example, a complete CDS (B3-B4-B5), a Promoter (A1-A2-A3-B1-B2) or a Terminator (B6-C1) (Figure 10) (Sarrion-Perdigones *et al.*, 2011 and 2014). This feature aids in the efficiency of the assembly method and in the faster development of GBSperts library.

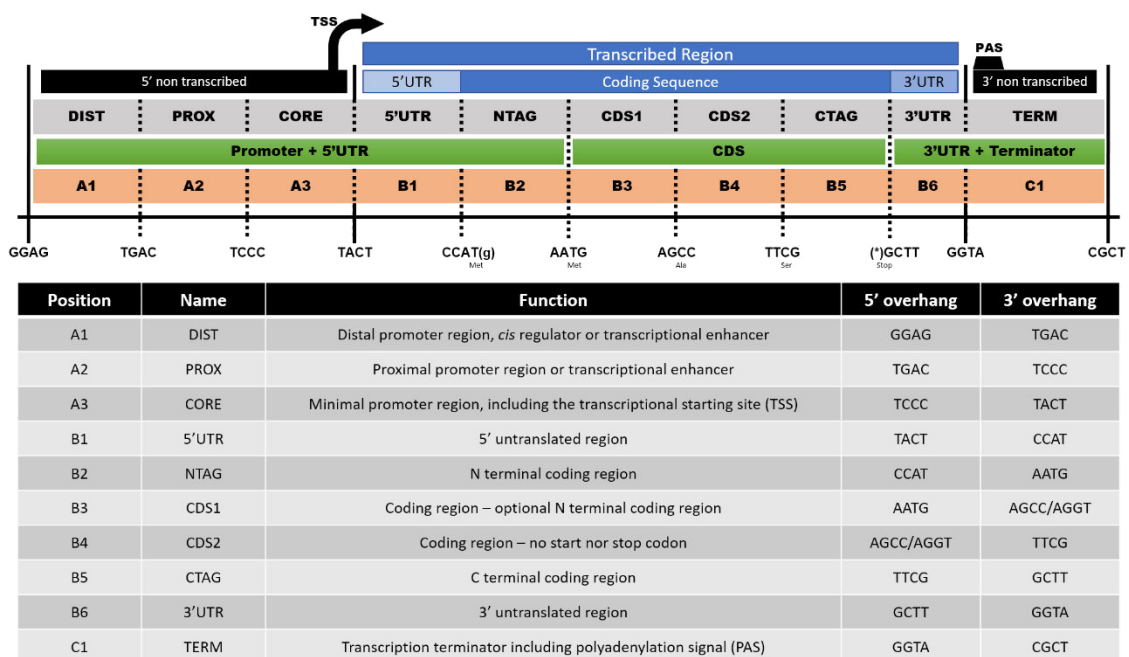


Figure 10 – Schematic view of the GB syntax and description of each one of the 10 elements and respective 5' and 3' fusion overhangs. Adapted from Patron *et al.*, 2015 and Sarrion-Perdigones *et al.*, 2013.

Both the GBparts and GBSperts are stored in a pUPD vector. To get these parts in this vector, and more important to get whatever DNA sequence in the standard, the DNA sequence

suffers a process called domestication. GB Domestication begins by amplifying the DNA sequence, called GBpatch (e.g. CDS, promoter, etc.), with GB-adapted primers. These primers include, in average, 20 nucleotides of the targeted sequence to amplify, and a tail containing the desired 4 nucleotide overhang immediately after the specific part and the recognition sequence for the *BsmBI* endonuclease. This process also removes internal recognition sequences from the DNA fragments, amplifying more than one fragments of the target sequence. In the digestion-ligation reaction with *BsmBI* the pUPD backbone will incorporate the DNA part(s) amplified and lose the recognition site(s) for *BsmBI*. The backbone of this newly formed GBpart or GBSpart contains, however, the recognition site for the *BsaI* endonuclease (Sarrion-Perdigones *et al.*, 2013; Vazquez-Vilar *et al.*, 2015).

Having the GBparts and GBSparts assembled in a pUPD allows for the creation of transcription units (TU) by multipartite assembly in another type of vector, the GB destination vectors (pDGB). The pDGB vectors are, at the same time, binary and shuttle vectors that have a GBcassette. The GBcassette contains the LacZ gene flanked by the recognition sites for both type IIS endonucleases and a watermark, i.e., a restriction site for a type II endonuclease for the identification by fragment length analysis of the plasmid. There are two levels of pDGB vectors, the  $\alpha$ -level and the  $\Omega$ -level. They differ in: (1) the order and orientation of the restriction sites for *BsaI* and *BsmBI*. *BsaI* is used in  $\alpha$ -level assemblies and *BsmBI* in  $\Omega$ -level assemblies. (2) the antibiotic resistance marker of the  $\alpha$ -level vectors is kanamycin and the  $\Omega$ -level is spectinomycin, which allows for counterselection of the plasmids (Sarrion-Perdigones *et al.*, 2013; Vazquez-Vilar *et al.*, 2015).

Simplifying, to assemble a TU in a  $\alpha$ -level pDGB backbone the restriction-ligation reaction is performed with *BsaI* (enzyme in) and to remove the resulting assembled TU the reaction is done with *BsmBI* (enzyme out). In the  $\Omega$ -level the order of the endonucleases is switched. There are eight basic pDGB plasmids, four of them form a basic set for endless cloning (Sarrion-Perdigones *et al.*, 2013), namely, pDGB $\alpha$ 1, pDGB $\alpha$ 2, pDGB $\Omega$ 1, and pDGB $\Omega$ 2. And the remaining 4 (pDGB $\alpha$ 1R, pDGB $\alpha$ 2R, pDGB $\Omega$ 1R, and pDGB $\Omega$ 2R) are used for the assemble in the reverse orientation.

The GB3.0 has, in total, 24 different plasmids. Sixteen of those, 8 pDGB1, based in the pGreenII (Hellens *et al.*, 2000) backbone and another 8 pDGB2, based in the pCAMBIA backbone, were available in the GB2.0. The GB3.0 introduced other 8, the pDGB3, also based in the pCAMBIA backbone (Sarrion-Perdigones *et al.*, 2013; Vazquez-Vilar *et al.*, 2015). The TU assembled in one

level of the destination vectors can them be assembled on the other level by binary assembly. The double-loop iterative strategy that gives the name to the GoldenBraid system is represented in Figure 11A. Basically, two TU in an  $\alpha$ -level destination vector can be assembled in an  $\Omega$ -level vector if they have complementary *BsmBI* sticky ends (Sarrion-Perdigones *et al.*, 2013 and 2014). The same process can be performed with a *BsaI* reaction to pass from the  $\Omega$ -level to the  $\alpha$ -level. It is this feature that allows for the continuous enhancement of the complexity of the cassettes developed using the GB system (Figure 11B).

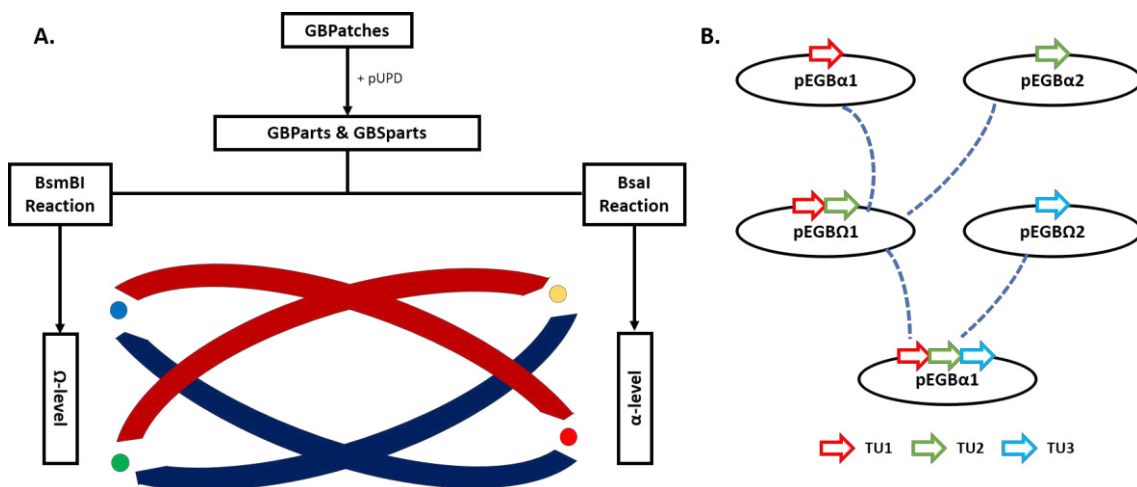


Figure 11 – A. Schematic representation of the GoldenBraid system. After domestication in a pUPD vector the GBparts and GBSparts can be assembled in the  $\Omega$ -level or  $\alpha$ -level destination plasmids. B. In turn those can be binarily assembled in alternating levels forming more complex assemblies.  $\alpha$ -level TU can be assembled in  $\Omega$ -level destination vectors. And  $\Omega$ -level TU or modules can be combined in  $\alpha$ -level destination vectors.

All that was stated before helps explained the process by which the GB system is used to create new genetic sequences for use in Plant synthetic biology. The iterative assembly and the possibility of generating new GBparts (domestication), contributes for the studies of gene function and expression. For the purpose of this project both study approaches were used. GB standard was used to construct expression cassettes in the study of promoter activity and gene function for sequences not available in the GB database in combination with GBparts already domesticated.

All that was stated before helps explained the process by which the GB system is used to create new genetic sequences for use in Plant synthetic biology. The iterative assembly and the possibility of generating new GBparts (domestication), contributes for the studies of gene function and expression. For the purpose of this project both study approaches were used. GB standard was used to construct expression cassettes in the study of promoter activity and gene function for sequences not available in the GB database in combination with GBparts already domesticated.

The GoldenBraid system is splendidly suited to be applied in the assembly of CRISPR/Cas9 cassettes targeting one or several genomic targets. In 2016, Vazquez-Vilar *et al.*, (2016) reported the creation of a toolbox of GB-adapted gRNA-Cas9 elements for construction of binary vectors. These tools allow for the CRISPR-based genome engineering in plants using *Agrobacterium*-mediated transformation.

The construction of the guide RNA (gRNA) for this tool has a specific GB syntax depending if the plant tissue to modify is from the Liliopsida class (formerly, Monocotyledon) or from the Magnoliopsida class (formerly, Dicotyledon). Figure 12 highlights the different elements in the creation of a gRNAs for CRISPR/Cas9 cassettes.

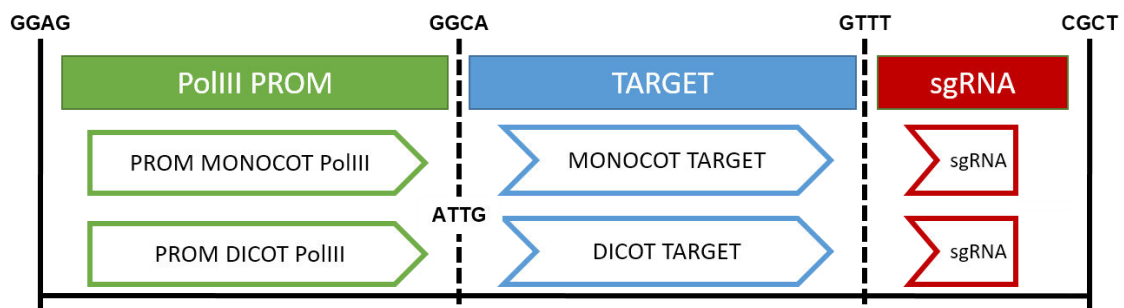


Figure 12 – Schematic of the GB grammar for the construction of CRISPR target TUs for plants. MONOCOT is used in plants from the Liliopsida and DICOTS for the Magnoliopsida plants. The fusion site between the Polymerase III promoter and the target is different in each one of the classes. GGCA for Liliopsida and ATTG for Magnoliopsida. Adapted from GoldenBraid 3.0 website (<https://gbcloing.upv.es/search/>)

The targets are comprised of sequences with about 20 nucleotide sequences with 4 nucleotide overhangs on both sides for the ligation with the rest of the elements of the gRNA. The RNA polymerase III elements available in the GB database are the *Oryza sativa* U3 RNA PoIII promoter for monocotyledon targets and the *Arabidopsis thaliana* U6-26 and U6-1 RNA PoIII promoters for dicotyledons. For the gRNA multipartite assembly in a pDGB $\alpha$  destination vector, the digestion-ligation reaction with *Bsa*I involves one of the promoters, the target DNA sequence with the proper syntax overhangs and a scaffold. The scaffold (sgRNA) is the combination of the sequences of bacterial crRNA and tracrRNA minus the target sequence. The final construct of this assembly is the TU for the sgRNA used in the CRISPR/Cas9 system. The iterative assembly using the GB system adds the TU for the Cas9 protein in the same construct. This pipeline of assembly can be used in a multitude of target constructs, reducing the time and effort to produce sgRNAs. The GB-gRNA/Cas9 toolbox eases the cloning of T-DNA CRISPR/Cas9 tools and provides the system traceability and exchangeability (Vazquez-Vilar *et al.*, 2016). The CRISPR/Cas9 cassettes

for the knockout of the candidate genes were constructed using this simple, time effective, and less expensive technology.

The GB website (<https://gbcloning.upv.es>) has a collection of web tools to assist in the GB cloning process, providing information about the GB entries and correspondent sequence, as well as helping in the definition of the protocols to follow in the assemble of the constructs.



## **MATERIALS AND METHODS**

## 1. Plant material.

All the lines and accessions used in this study are listed in Table 1.

Seeds of the F<sub>2</sub>B<sub>2</sub> and F<sub>2</sub>B<sub>3</sub> populations derived from the crossing of the IL SP4-4, containing the SD region, with the recurrent parent *S. lycopersicum* var. "MoneyMaker". "MoneyMaker" and F<sub>2</sub> plants were used as genotyping controls, in expression analyses and the determination of the pre- or post-zygotic nature of the SD.

Wild accessions of *S. pimpinellifolium* and *S. lycopersicum* var. *cerasiforme* (Table 1) were obtained from Tomato Genetic Resource Center (TGRC) (Dept. of Plant Sciences, UC-Davis). In collaboration with Doctor Rafael Fernandez (Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora", Malaga, Spain), wild accessions were crossed with *S. lycopersicum* var. "MoneyMaker". The F<sub>1</sub> of each successful cross was self-crossed to generate F<sub>2</sub> segregating populations.

Table 1 – Tomato plant accessions from the *S. pimpinellifolium*, *S. lycopersicum* var. *cerasiforme* and "MoneyMaker" used in this study.

<b>TS code/ name</b>	<b>TGRC code</b>	<b>Species or Cross</b>	<b>Origin or Donor</b>
<b>TS-15</b>	<b>LA 2093</b>	<i>S. pimpinellifolium</i>	El Oro, Ecuador
<b>TS-16</b>	<b>LA 1246</b>	<i>S. pimpinellifolium</i>	Loja, Ecuador
<b>TS-437</b>	<b>LA 1578</b>	<i>S. pimpinellifolium</i>	La Libertad, Peru
<b>TS-182</b>	<b>LA 2183</b>	<i>S. pimpinellifolium</i>	Amazonas, Peru
<b>TS-413</b>	<b>LA 1242</b>	<i>S. pimpinellifolium</i>	Guayas, Ecuador
<b>TS-420</b>	<b>LA 2184</b>	<i>S. pimpinellifolium</i>	Amazonas, Peru
<b>TS-265</b>	<b>LA 0400</b>	<i>S. pimpinellifolium</i>	Piura, Peru
<b>TS-50</b>	<b>LA 0417</b>	<i>S. pimpinellifolium</i>	Guayas, Ecuador
<b>TS-14</b>	<b>LA 1547</b>	<i>S. pimpinellifolium</i>	Carchi, Ecuador
<b>TS-96</b>	<b>LA 1456</b>	<i>S. lycopersicum</i> var. <i>cerasiforme</i>	Vera Cruz, Mexico
<b>TS-124</b>	<b>LA 1245</b>	<i>S. pimpinellifolium</i>	El Oro, Ecuador
<b>TS-301</b>	<b>LA 2688</b>	<i>S. lycopersicum</i> var. <i>cerasiforme</i>	Madre de Dios, Peru
<b>F<sub>2</sub>B<sub>2</sub></b>	-	IL SP4-4 x <i>S. lycopersicum</i> var. "MoneyMaker"	IBMCP, Valencia, Spain
<b>F<sub>2</sub>B<sub>3</sub></b>	-	IL SP4-4 x <i>S. lycopersicum</i> var. "MoneyMaker"	IBMCP, Valencia, Spain
<b>MM</b>	<b>LA2706</b>	<i>S. lycopersicum</i> var. "MoneyMaker"	Henri Laterrot

## **2. Seed disinfection and germination.**

Disinfection of seeds was performed by using a combined chemical and thermal treatment. Seeds were immersed in a solution of 10% (w/v) trisodium phosphate dodecahydrate ( $\text{Na}_3\text{PO}_4 \cdot 12\text{H}_2\text{O}$ ) for 3 hours, with constant agitation, and after, washed with distilled water at least 5 times. Next, seeds were immersed in a 30% (v/v) solution of commercial bleach for one hour with constant agitation. After the bleach treatment, seeds were washed with distilled water at least 9 times. Following the chemical treatment, seeds were air dried on a filter paper for 24 hours. To ensure seeds are completely dry, seed were incubated for another 24 hours inside a paper envelope in a box with silica gel. Thermal treatment was performed at 78-80°C for 24 hours.

For germination, the disinfected seeds were placed in Petri dishes over a paper filter on watered cotton wool and then covered with another moist paper filter. Germination induction was done by incubating seeds, in darkness, at 37°C for 24 hours. Seeds were transferred to a growth chamber in light at 28°C, until the emergence of the radicle and cotyledons. Seedlings were planted in soil into 10 x 6 inserts and grown in a greenhouse with the temperature regime set to 25°C during the day and 17°C at night at the IBMCP facilities. Plants reaching around 20 cm, having a good fitness, and starting the inflorescence development were transferred to COMAV greenhouses with a temperature regime between 22°C and 28°C during the day and natural day/night cycles.

## **3. DNA extraction, evaluation and quantification.**

Genomic DNA was extracted by using the Cetyl-Trimethyl-Ammonium-Bromide (CTAB) method (Doyle and Doyle, 1991) with some specific modifications. One to two centimetres leaflets were collected and maintained in ice while harvesting the rest of the samples and then frozen in liquid nitrogen. Frozen samples were grinded in the Geno/Grinder 2000 (SPEX Sample Prep, New Jersey, USA) with two stainless steel balls ( $\varnothing$  3 mm), at 1500 strokes per minute for 15 seconds. Grinded tissue was homogenized with 300  $\mu\text{L}$  of CTAB buffer (2% CTAB (m/v), 1.4 M NaCl, 100mM Tris-HCl, 20 mM EDTA). Nucleic acids were separated from other cell components by adding 300  $\mu\text{L}$  Chloroform: Isoamyl alcohol (24:1 v/v) and centrifuged at 3000 rpm for 10 minutes. After centrifugation, the aqueous phase was transferred to a new tube. Nucleic acids were precipitated with one volume of cold Isopropanol (-20°C) followed by a centrifugation at 4000 rpm for 10 minutes. Nucleic acids were washed with 400  $\mu\text{L}$  of Washing Buffer (76% Ethanol and 10 Mm

Ammonium Acetate) and then centrifuged at 4000 rpm for 10 minutes. Supernatant was discarded and the DNA air dried for 10 minutes. To remove RNA, nucleic acids were resuspended in 30  $\mu$ L of TE buffer (Tris 10 Mm, EDTA 1 mM, pH 8) and 0.2  $\mu$ L of RNase at 10 mg/mL and incubated for an hour at 37°C with constant shaking. The quality of DNA was evaluated by electrophoresis in a 0.8% agarose/TAE buffer (1x, Tris - acetic acid - EDTA) gel with 0.1  $\mu$ L/mL of ethidium bromide (EtBr). DNA concentration (ng/ $\mu$ L) was quantified using a Nanodrop spectrophotometer.

#### **4. HRM – high resolution melting genotyping method.**

The PCR amplification and determination of the corresponding melting curves were performed in a 7500 Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). Three SolCAP markers flanking the SD region and three Gsnp SNP markers inside the region identified by Barrantes (2014) were used for HRM genotyping (Table 2). Four controls were used. A blank control for the amplification, containing ddH<sub>2</sub>O instead of gDNA, and DNA of three previously confirmed haplotypes for the SD region: *S. pimpinelifolium* haplotype (PP), *S. lycopersicum* var. “Moneymaker” haplotype (MM), and heterozygous haplotype (H).

The reaction was performed in a final volume of 10  $\mu$ L, containing 2  $\mu$ L of gDNA (5 ng/ $\mu$ L), 0.3  $\mu$ L of Primer F and Primer R at 10  $\mu$ M, 5  $\mu$ L of AccuMelt™ HRM SuperMix (Quanta Bio, Inc.), 0.16  $\mu$ L of Mg<sup>2+</sup> and 2.24 $\mu$ L of ddH<sub>2</sub>O. AccuMelt™ HRM SuperMix relies on the STYO®9 green-fluorescence dye. The HRM conditions were: denaturation or holding stage at 95°C for 10 min; 40 cycles of amplification at 95°C for 15s followed by 60°C for a minute; melting stage starts at 95°C for 10s then 60°C for 1 min and increases back to 95°C by a ramp rate of 1% and 200 data points per 1°C. The discriminant allele analysis is performed using the Applied Biosystems High Resolution Melting Software (HRM v2.0). Raw fluorescence for each amplicon is plotted versus temperature (Fig 1a) to define the melting temperature ( $T_m$ ) (i.e. the temperature at which 50% of the DNA is double stranded and 50% is single stranded).

The discriminant allele  $T_m$  (melting temperature) analysis was performed using the Applied Biosystems High Resolution Melting Software (HRM v2.0). Briefly, emitted fluorescence for each sample was plotted against the temperature (melt curve) and the amplicon  $T_m$  and relative amplicon signal were calculated. The  $T_m$  and relative signal for each amplicon and DNA sample together with the melting curves obtained for the controls were used to discriminate and assign the genotype to tested DNA samples.

Table 2 – List of primers used for HRM genotyping.

Primer	Sequence (5'→3')	T <sub>m</sub> (°C)	Physical position 5'
solcap_snp_47742 F	CACTCCAAAACATCGTCAGCTC	56.2°C	65534784
solcap_snp_47742 R	ACTGTTCATTGCTGTTGGC	57.1°C	65534908
solcap_snp_3952 F	TCTTTCTTCGTGCAACCTTCTG	55.4°C	65621010
solcap_snp_3952 R	CGATATAAAGTTGAAGGAGATCGGTC	55°C	65621127
solcap_snp_47749 F	TTCTGGGTGCTGTCGGATG	57.4°C	65664775
solcap_snp_47749 R	TGGCAAGGCCACTCATGA	57°C	65664906
Gsnp_67132 F	ACACCCCTAGGAACAAAGATG	54.4°C	65562965
Gsnp_67132 R	CGACCTTGCCCAATATCAT	52.4°C	65563114
Gsnp_94601 F	CATGTGCTGCTTTGCTTGAT	54.3°C	65590442
Gsnp_94601 R	CAATGGCAGCAACTGGATTA	53.3°C	65590556
Gsnp_121747 F	CAAGCTTTGTCTCCTGTAAGCA	51.3°C	65617535
Gsnp_121747 R	GGGAGGGGAAAGGCTTTAG	55.2°C	65617684

## 5. KASP primer design and procedure.

The primers for KASP genotyping were designed against the genomic sequence SL 2.50 available in the SOL genomic network site (<https://solgenomics.net/>) using the Web tool Benchling. Eleven KASP primer pairs were designed within the SD region of chromosome 4. The full list of primers is available in Annex 2, as well as the genomic coordinates for the targets.

KASP genotyping analysis was performed in an ABI 7500 Real Time PCR System using the LGC genomics KASP™ genotyping chemistry.

The reactions were performed in a 96-well plate containing a volume of 10 µL per well. The reaction mixture consisted in 2 µL of gDNA (10 ng/µL), 5 µL of 2X KASP Master mix (containing: FRET reporting system, the FAM and HEX labelled cassettes; DNA polymerase; ROX passive reference dye; dNTPs; MgCl<sub>2</sub>; and an optimised buffer), 1 µL of the Forward Common primer (1 µM), 1 µL of the Reverse FAM primer (wild-type - MM) (1 µM), 1 µL of the Reverse HEX primer (mutant - PP) (1 µM). The reaction was performed following manufacturer guidelines (Table 3).

Table 3 – Cycling method for the KASP analysis used in the ABI 7500 Real Time PCR System. The temperatures in brackets refer to a special case where the primer  $T_m$  was lower.

Step	Procedure	Temperature (°C)	Time	Num. of cycles
1	Pre-PCR read	30.0	1 min	1
2	Activation	94.0	15 min	1
3	Denaturing	94.0	20 sec	10
	Annealing/Elongation	61.0 (59.0) ( $\Delta T$ : -0.6 °C/cycle)	1 min	
4	Denaturing	94.0	20 sec	26
	Annealing/Elongation	55.0 (53.0)	1 min	
5	Post-PCR Read	30.0	1min	1

Fluorescence signals were clustered around a specific wavelength and shown in a dispersion plot by the ABI 7500 Real Time PCR System software. A signal consistent with the FAM probe indicates a wild-type SNP and a signal with similar wavelength of the HEX probe indicates mutant SNP. A mixed signal indicates a heterozygote.

In cases where the results did not cluster, the plaque was used for an additional two step cycling and the results plotted again (Table 4).

Table 4 – Recycling method for the KASP analysis used in the ABI 7500 Real Time PCR System.

Step	Procedure	Temperature (°C)	Time	Num. of cycles
1	Denaturing	94.0	20 sec	3
	Annealing/Elongation	55.0	1 min	
2	Post-PCR Read	30.0	1min	1

## 6. Assessment of SD mechanism.

### 6.1 Self-pollination crosses.

Self-pollination crosses were performed in *S. lycopersicum* var. “Moneymaker” (MM) plants and *S. lycopersicum* SP4-4 (PP and H) plants haplotyped by HRM. To do the crosses, one or two days before anthesis, female flowers were emasculated (removing the stamen anthers and filament), the petals were also removed. At anthesis, and early in the morning, the anthers of male flower with the same haplotype were collected and dried for 20 to 30 min in petri plates with silica gel. Fecundation was performed using a scalpel. Pollen was collected and deposited in the stigma

of the emasculated plants. The fruits were let to develop until the breaker stage of maturation. In that time, the fruits were collected and the seeds removed, cleaned and counted.

## **6.2 Reciprocal cross ♀ MM x ♂ IL4-4H.**

The reciprocal cross was performed using as a female *S. lycopersicum* var. “Moneymaker” (MM) and *S. lycopersicum* SP4-4 H as a male. The crossing and seed collecting procedure was the same applied to the self-pollination crosses. The results of this experiment were added to the ones previously obtained by Fakhret (2016).

### **6.2.1 Evaluation of the progeny of the cross ♀ MM x ♂ IL4-4H.**

Seventy-five seeds derived from the reciprocal cross ♀ MM x ♂ IL4-4H were put to germinate by the method described before. From those nearly half of them (36) germinated and were transplanted to 10 x 6 soil cups in the IBMCP greenhouse for development. 18 plants were genotyped by HRM using the markers Solcap\_47742, Gsnp\_83668, and Solcap\_3952 (Barrantes, 2014; Fakhret, 2016).

## **7. Sequencing of the genes within the SD region-**

The genomic sequence of the genes in the SD region was obtained using DNA from the *S. lycopersicum* SP4-4 that have the PP haplotype in homozygosis for the region and from DNA from *S. lycopersicum* var. “Moneymaker”. Two independent DNA extractions from two different plants with confirmed haplotype by HRM were performed.

The genomic regions were sequenced by using overlapping PCR fragments. Sequencing primers were designed against the “Tomato WGS Chromosomes database (SL2.5, ITAG2.4)”. In all cases, primer design conditions were: amplicon length around 800bp that overlap with adjacent amplicons of the same gene by ~200bp, melting temperature ( $T_m$ ) from 56°C to 62°C, a GC content around 50%, limited self-complementarity, absence of secondary structures, and terminal G or C when possible. To ensure single and specific amplicon, primers were checked for nucleotide variants present in 150 and 360 *Solanum* resequenced accessions (100 Tomato Genome Sequencing Consortium, 2014; Lin *et al.*, 2014) and *in Silico* PCR was done using the tool from

the Sol Genomics Network ([https://solgenomics.net/tools/in\\_silico\\_pcr](https://solgenomics.net/tools/in_silico_pcr)). The full list of primers is available on Annex 1.

PCR amplifications were performed in a final volume of 25  $\mu\text{L}$  using a Techne TC-512 thermocycler with heated lid. PCR reactions consisted in mixture of 5  $\mu\text{L}$  of gDNA at 5  $\text{ng}/\mu\text{L}$ ; 1  $\mu\text{L}$  of each primer (Forward and Reverse) at 10  $\mu\text{M}$ ; 0,38  $\mu\text{L}$  of TAQ DNA polymerase at 5  $\text{U}/\mu\text{L}$  (Biotools, B & M Labs, S.A.); 2,5  $\mu\text{L}$  of TAQ Buffer at 10x (Biotools, B & M Labs, S.A.); 0,5  $\mu\text{L}$  of dNTPs at 2,5  $\text{mM}$ ; and 1  $\mu\text{L}$  of  $\text{Mg}^{2+}$  at 50  $\text{mM}$ ; completed with 13,62  $\mu\text{L}$  of ddH<sub>2</sub>O to the final volume of 20  $\mu\text{L}$ . The PCR condition were 94°C for 5 min for the denaturing step, followed by 30 cycles of amplification (annealing and extension) with 94°C for 30s, 58°C for 30s and, 72°C for 30s. Completed with a final extension of 72°C for 5 min.

PCR amplifications and fragment sizes were confirmed by 1% agarose gel, stained with EtBr (1  $\mu\text{L}/\text{mL}$ ), using the Gene Ruler 100 bp Plus DNA Ladder as marker.

Sequencing was performed on PCR products. Before, sequencing any excess primer was removed treating directly PCR reaction with Affymetrix® ExoSAP-IT® PCR Product Clean-up (Affymetrix, Inc., Cleveland, Ohio, USA) as follows. For every 5  $\mu\text{L}$  of PCR product, 2  $\mu\text{L}$  of ExoSAP-IT® were added and incubated, in the thermocycler, at 37°C for 15 min to degrade remaining primers and nucleotides. ExoSAP-IT® reagent was inactivated at 80°C for 15 min.

DNA sequencing was performed by the “Secuenciación de ADN y análisis de la expresión génica” service in IBMCP using Sanger sequencing method.

Sequencing files were curated in the MEGA software (v7.0.18, release #7160630-x84) with the Trace tool. The assembly of the resulting sequences for each gene was done using the Heinz reference genome of *S. lycopersicum* (build SL2.50) to avoid mismatches in the order of the fragments.

The analysis of the sequences obtained from the genome of the *S. lycopersicum* var. “MoneyMaker” and the *S. lycopersicum* IL4-4 PP was done using the build SL 2.5 and respective annotations, ITAG 2.4. Consensus sequences for each gene, in the PP haplotype and in MM haplotype, were aligned against each other and against Heinz genomic sequence to check for polymorphisms and for prediction of gene structures.

Each genomic sequence obtained from the *S. lycopersicum* var. “MoneyMaker” and *S. lycopersicum* SP 4-4 genomes was aligned with the Heinz annotation for the prediction of the respective gene sequence, mRNA sequence, coding sequence, and, finally the protein sequence, in this order. Also, when possible, some gene structures like 5' and 3' untranslated sequences of



the mRNA (5'UTR and 3'UTR) were assigned. The resulting predicted protein was checked for aminoacidic substitutions and protein models bioinformatically generated to check for differences in the catalytic domains and folding.

Domain finding and protein structure was analysed *in silico* using PROSITE (<http://prosite.expasy.org/>) and SWISS-MODEL (<https://swissmodel.expasy.org/interactive>).

For visualization and manual annotation of the PP sequences the Unipro UGENE software (v1.26.3 64-bit version) was used (Okonechnikov *et al.*, 2012).

Arabidopsis gene homologs were identified using the tblastx BLAST program (<https://blast.ncbi.nlm.nih.gov>) by using the predicted CDS sequence for each gene. Sequence similarities with an *E*-value threshold  $\leq 0.001$  were considered as significant as well as sequences with high identity and a coverage of  $>80\%$  to the query sequence.

The *in silico* expression analysis for the tomato sequences was done using TomExpress ([http://gbf.toulouse.inra.fr/tomexpress17b/www/new\\_query.php](http://gbf.toulouse.inra.fr/tomexpress17b/www/new_query.php)) and Tomato Expression Atlas (<http://tea.solgenomics.net/>).

Both consensus sequences from MM and PP were subjected to this analysis.

The expression of the *A. thaliana* genes that resulted from tblastx search were analysed by the BAR tool - Arabidopsis eFP Browser (<http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>).

The Gene Ontology (GO) of the *A. thaliana* and *Solanum* genes was checked using the UniProt website (<http://www.uniprot.org/>).

## **8. Analysis of candidate gene expression using qRT-PCR comparative $C_T$ analysis.**

### **8.1 Sample collection for total RNA extraction.**

To assert gene expression differences throughout the flower and fruit development, samples of ovaries and anthers were collected, as well as fecundated ovaries, at defined stages as defined by Buzgo *et al.* (2004) and Xiao *et al.* (2009). The landmarks chosen as the reference time point was the anthesis (ODPA – Days Post anthesis). Tissues collected after anthesis correspond to self-pollinated flowers.

In all cases, an average of 3 flowers/fruit were collected and pooled for total RNA extraction. For each landmark, 2 to 3 repetitions of the collection were done to eliminate differences in gene expression caused by the collection method or by environmental conditions. All the replicates were

collected and corresponded to either *S. lycopersicum* var. “Moneymaker” plants, as control, and *S. lycopersicum* SP4-4 PP and H haplotypes.

Samples from ovaries and stamens were collected in two time intervals before the reference time (0 DPA). The first one, Bud Stage (BS), when flowers were early in development and still closed. Second one, the day before anthesis (-1 DBA – Days Before anthesis), i.e., day of emasculation. On the day of anthesis (0 DPA) both ovaries and stamens were collected.

Only ovaries were collected after fecundation on the 4<sup>th</sup>, 7<sup>th</sup>, 10<sup>th</sup>, and 13<sup>th</sup> day after anthesis (4 DPA, 7 DPA, 10 DPA, and 13 DPA).

## **8.2 RNA extraction, purification and quantification.**

RNA extraction from ovaries and anthers was performed using the RNeasy® Plant Mini Kit (Qiagen®, cat. nos. 74904) according to manufacture instructions. Briefly, flowers were dissected and the ovaries and anthers were flash frozen in liquid nitrogen. The tissue was grinded and homogenised in a mortar with liquid nitrogen in presence of 450 µL of RLT buffer, containing dithiothreitol (DTT) and allowed to thaw. Extracts were transferred to a 1.5 mL microcentrifuge tube and vortexed vigorously for a few seconds. Cell debris was removed by filtering lysate in a QIAshredder spin column and centrifuging for two minutes at 14,000 rpm. Next, the supernatant of the flow-through was transferred, carefully, to let the cell debris pellet undisturbed on the bottom of the collection tube to a new 2 mL microcentrifuge tube. Nucleic acids bound to the resin in RLT buffer by adding 0.5 volume of isopropanol to the lysate. The mixture was transferred to a RNeasy® mini spin column attached to a 2 ml collection tube and centrifuged for 15s at 8,500 x g. The flow-through was discarded. Nucleic acids were washed twice with 700 µL of RW1 buffer and centrifugation. An additional washing step was performed with 500 µL of RPE. The column was transferred to a new 2 mL collection tube and centrifuged at 14,000 rpm for a minute to completely dry it. RNA was eluted in a clean 1.5 mL microcentrifuge tube with 50 µL of RNase free water. To increase RNA concentration, the elution step was repeated with the flow through.

RNA quality was checked in a 1% agarose gel using GeneRuler 100bp Plus DNA Ladder as marker.

Genomic DNA traces were eliminated using Ambion® DNA-free™ Kit DNase Treatment and Removal Reagents. 2 µg of RNA were mixed with 2 µL of 10x DNase I Buffer and 0.5 µL of rDNase I (2 U/µL), and water performing a total volume of 20 µL. The mixture was incubated for half an hour at 37°C. After incubation, to eliminate the DNase, 4.5 µL of DNase Inactivation Reagent were

added and gently mixed with the sample. After 2 min of incubation at RT the microtubes were centrifuged for 2 min at 10,000 x g. The reagent is a viscous white solution that, after centrifuge, forms a pellet in the bottom of the tube. The eluted RNA (about 20 µL) was then transferred to a new tube, quantified in the NanoDrop and stored in -80°C until further use.

### **8.3 cDNA synthesis.**

cDNA was synthesized using Prime script 1st strand cDNA Synthesis kit (Takara Bio USA, Inc.). For cDNA synthesis, RNA was concentrated, by vacuum centrifuge, in a Speedvac concentrator for 12 min (-1 µl/min) to less than 5 µg of RNA in a total volume of 8 µl. The total RNA was mixed with 1 µl of 10mM dNTP mixture and 1 µl of 50µM oligo dT Primer. This mixture is incubated for 5 min at 65°C. Afterwards a mixture of 5x PrimeScript Buffer (4 µl), RNase Inhibitor (0.5 µl, 20 U), PrimeScript RTase (1 µl, 200 U) and ddH<sub>2</sub>O to a total volume of 10 µl was prepared. After mixing gently, the sample is incubated for 50 min at 42°C. Finally, for 15 min the mixture is heated at 75°C to inactivate the RTase.

### **8.4 Primer design for qRT-PCR**

For detection of the target sequences using Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR) several pairs of primers were designed, at least one pair corresponding to each one of the target genes. Primers were designed following the same basic stringent properties applied to the sequencing primers. The preferable length of 20-24 bp and an amplicon target length between 80 bp and 200 bp was designed when possible. All primers were screened for possible mismatches/ off targets by search against the “Tomato Genome cDNA database (ITAG 2.40)”. Primer specificity was analyzed by *in silico* PCR using the *in silico* PCR tool from the Sol Genomics Network ([https://solgenomics.net/tools/in\\_silico\\_pcr](https://solgenomics.net/tools/in_silico_pcr)). The full list of qRT-PCR primers is in Annex 1.

### **8.5 Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR) of gene expression between plants carrying different SD haplotypes.**

The qRT-PCR amplification was performed in a ABI 7500 Real Time PCR System. The reaction mixture consists of 10 µL of SYBR® Premix Ex Taq™ (Tli RNaseH Plus, ROX Reference Dye II) (RR420A), 1 µL of cDNA template (50 ng/µL), 1.2 µL of each primer, forward and reverse, and

6.6  $\mu\text{L}$  of ddH<sub>2</sub>O. The run method used consisted on a holding stage at 50°C for 2 min, followed by a second holding stage at 95°C for 10 min, to activate the reagents and denature the cDNA template. And finally, 40 cycles of 15s at 95°C and 60°C for 1 min for amplification. Clathrin adaptor complexes medium subunit gene (CAC), was used as an endogenous control as in González-Aguilera *et al.* (2016).

### 8.6 Data analysis by the Comparative Ct ( $\Delta\Delta\text{Ct}$ )

RNA expression data analysis was done with the DataAssist™ v3.01 Software (<https://www.thermofisher.com>) and extra punctual calculation preformed with Microsoft Office Excel software. Ovary samples from the *S. lycopersicum* MM plants collected at anthesis (MM 0 DPA) were used as the reference sample and the CAC target as the endogenous control. A maximum C<sub>T</sub> value was set (40.0), this value corresponds to the number of cycles defined in the qRT-PCR reaction. Outlier values were rejected in the average C<sub>T</sub> calculations.

The comparative C<sub>T</sub> (aka,  $\Delta\Delta\text{C}_T$  method) (Pfaffl, 2001; Livak and Schmittgen, 2001) was used to calculate the fold-changes of the target genes between the samples.

Two experimental replicates of each sample, composed by pooled samples (i.e. Organ/time), were used to calculate the average C<sub>T</sub> and the respective standard deviations for the target and for the endogenous control.

The  $\Delta\text{C}_T$  is the relation between the average C<sub>T</sub> of the target sample minus the average C<sub>T</sub> of the endogenous control (Formula 1), in this case the constitutively expressed gene CAC, for every sample.

$$\Delta\text{C}_T = \text{C}_{T \text{ target}} - \text{C}_{T \text{ endogenous control}}$$

Formula 1 – Calculation of the  $\Delta\text{C}_T$  value for each sample average.

The standard deviation (*s*) of the  $\Delta\text{C}_T$  value, like the  $\Delta\text{C}_T$  value itself, is calculated using the standard deviations of the target and endogenous control for the same sample by the Formula 2.

$$s = (s_1^2 + s_2^2)^{1/2} \Leftrightarrow s = \sqrt{(s_1^2 + s_2^2)}$$

Formula 2 – Calculation of the standard deviation of the  $\Delta\text{C}_T$  value. *s*<sub>1</sub> and *s*<sub>2</sub> are the standard deviation of the target and reference, respectively.

Finally, to calculate the  $\Delta\Delta\text{C}_T$  value, the  $\Delta\text{C}_T$  of the reference sample (MM ovaries ODPA) was subtracted to the  $\Delta\text{C}_T$  of other sample for a given target (Formula 3).

$$\Delta\Delta\text{C}_T = \Delta\text{C}_{T \text{ sample}} - \Delta\text{C}_{T \text{ reference sample}}$$

Formula 3 – Calculation of the  $\Delta\Delta\text{C}_T$  value for each sample.

As the  $\Delta\Delta C_T$  value is calculated subtracting the  $\Delta C_{T \text{ reference}}$  value, which is constant, the standard deviation of the  $\Delta\Delta C_T$  value is the same as the  $\Delta C_{T \text{ target}}$ .

The final comparison of the  $\Delta\Delta C_T$  value between samples was done by expressing values as a fold-change. The fold change was calculated by incorporating the standard deviation of the  $\Delta\Delta C_T$  value by the expression  $2^{\Delta\Delta C_T}$  by  $\Delta\Delta C_T + s$  and  $\Delta\Delta C_T - s$ .

Finally, for each target the fold-change was converted to a  $\log_2$  range and plotted for each biological group. All sample groups, after normalization, were represented as a fold difference to the reference group. The plotting was done using the  $\log_2$  range that incorporates the standard deviation.

### 8.7 cDNA sequencing

Primers detailed in the Table 5 were used to amplify fragments covering the full annotated cDNA of the candidate gene Solyc04g081640.1. Primer design, PCR conditions and sequencing curation were the same as the one applied to the genomic sequences. 10 ng/ $\mu$ L of treated PCR products were sequenced in the “Secuenciación de ADN y análisis de la expresión génica” service in IBMCP using Sanger sequencing method.

Table 5 – Primer pairs used in the sequence of the cDNA of the gene Solyc04g081640.1.

Primer	Sequence (5'→3')	Length (bp)	Tm (°C)	GC (%)
qRT_Sl4g0818640E1F	TCGACTGTTATGGGGTTGTAA	22	57.38	40.91
40_1_R	TTCGGATATGTTTGTGTTGTC	22	59.9	40.9
qRT_Sl4g0818640E1F	TCGACTGTTATGGGGTTGTAA	22	57.38	40.91
MM cDNA seq R	CCTCTCAAATCGGAGGAAGTTG	22	58.73	50.00

### 9. Methods used for genetic constructs based on the standardized assembly system GoldenBraid.

The following descriptions concern the methods used for the development of the CRISPR/Cas9 knockout constructs, *pDGB3\_alpha1::Tnos:NptII:Pnos-SF:U6-26:gRNA30.1:sgRNA::35s:hcas9:Tnos* and *pDGB3\_alpha1::Tnos:NptII:Pnos-SF:U6-26:tRNA-gRNA40.1.1:tRNA-gRNA40.1.2::35s:hCas9:Tnos*, the develop of the Homologous over-expression construct *pDGB\_omega1::35s:SPSolyc04g081640.1CDS:Tnos::Pnos:NptII:Tnos*, and reporter

cassettes *pDGB3\_omega1::SLSolyc04g081640.1Prom:DsRed:Tnos::Pnos:NptII:Tnos* and *pDGB3\_omega1::SPSolyc04g081640.1Prom:DsRed:Tnos::Pnos:NptII:Tnos*.

All GB elements, intermediate constructs and final constructs are described in Annex 4.

### **9.1 Preparation of electrocompetent DH5 $\alpha$ *Escherichia coli* cells.**

Electrocompetent *E. coli* cells stored in glycerol stock were plaque spread onto a LB plate (1% tryptone, 0.5% yeast extract, 11% sodium chloride, and 1.5% agar) and grow overnight at 37°C. A single colony was picked and inoculated in 15 mL of LB (1% tryptone, 0.5% yeast extract, 1% sodium chloride) and grown on a shaker at 37°C for 24h. The 15mL starter culture was transferred to 1 L of LB medium and grown at 37°C for 3 hours in the shaker. When the OD<sub>600</sub> reached 0.4 the culture was chilled on ice for approximately 30 minutes. Simultaneously, the 4 centrifuge bottles were also on ice. After chilling, 250 mL was poured in each centrifuge bottle and centrifuged at 4,000 x g for 15 minutes at 4°C. The resulting pellet was resuspended in 250 mL of cold 1mM HEPES buffer/10% glycerol, and centrifuged again at 4,000 x g for 15 minutes at 4°C. The supernatant was discarded and the pellet resuspended in half the volume of 1 mM HEPES buffer/10% glycerol. The content of the 4 bottles was then combine in just two for a final volume of 500 mL, the mixture centrifuged again at 4,000 x g for 15 minutes at 4°C. This time, the supernatant was discarded has before but the final pellet was resuspended in 2 mL of ice cold 10% glycerol and aliquoted (40 $\mu$ L) to 1.5 mL microcentrifuge tubes and frozen in liquid nitrogen. The tubes were then stored at -80°C until further use.

### **9.2 Restriction-ligation assembly reactions.**

Assembly reactions were performed using Bsal (NEB, R0535S or R0535L) or BsmBI (Fermentas, ER0451) type IIS endonucleases and T4 DNA ligase (Promega, M180B) in 25 cycle digestion/ligation reactions of 2 minutes at 37°C, then 5 minutes at 16°C. The GB elements added vary in accordance to the assembly but the restriction-ligation cycle was always the same.

### 9.3 CRISPR/Cas9 knockout constructs.

#### 9.3.1 Selection and design of the CRISPR/Cas9 guide RNAs.

The selection and design of the gRNAs (guide RNA) for the CRISPR/Cas9 was done using Benchling CRISPR guide RNA design tool ([benchling.com/crispr](http://benchling.com/crispr)). The guide RNAs were selected based on three different features: (1) target location, to have a successful knockout of gene function target RNA must produce a deletion that creates a different reading frame or interrupts sequence of the gene, preferably closest to the beginning of the coding sequence, i.e., the first exon; (2) specificity, to access gene function based on the knockout of the gene, CRISPR/Cas9 must cut only the coding sequence of the targeted gene; and (3) efficiency, a well-designed target must be before a PAM sequence (5'-NGG-3', canonic form of the Protospacer Adjacent Motif) for the action of Cas9 to take place in the selected target. The last two were selected based on the good on-target efficiency score and off-target effects scores as given by the tool.

After designing the gRNAs, two oligos were designed in order to produce a heterodimer that contains overhangs compatible to the GB domestication vector and to ensure the correct orientation of the target in the final construction.

The heterodimers were obtained by mixing 5µL of each primer (Table 6) at the concentration of 1µM and let them anneal for at least 30 minutes at room temperature.

Table 6 – List of primer pairs for the construction of the Heterodimers for the gRNAs of the CRISPR tools. SL4g30 primers have BsaI compatible overhangs and SL4g40.1/.2 have BsmBI compatible overhangs.

Primer	Sequence (5'→3')	Length (bp)	T <sub>m</sub> (°C)	GC (%)	Heterodimer
SL4g30 F	ATTGGCAGGTTGTAGTTCAAAG	23	57.8	39.1	ATTGGCAGGTTGTAGTTCAAAG CGTCCAACATCAAGTTTCCAAA
SL4g30 R	AAACCTTTTGAATCAACCTGC	23	58.0	39.1	
SL4g40.1 F	GTGCACATTCACAAAATAATTGAAG	25	56.5	32	GTGCACATTCACAAAATAATTGAAG TGTAAGTGTTTATTAACCTCCAAA
SL4g40.1 R	AAACCTCAATTATTTGTGAATGT	25	54.7	24	
SL4g40.2 F	GTGCAAGAACCACAGAGAAAATTG	25	61.2	44	GTGCAAGAACCACAGAGAAAATTG TTATTGGGTGCTCTTTTAACCCAAA
SL4g40.2 R	AAACCAATTTCTCTGTGGTCTT	25	59.6	36	

#### 9.3.2 GB CRISPR Domestication

After the selection of the targets, the CRISPR domesticator tool was used to create domesticated sgRNA sequences flanked by standard GB overhangs that assemble with the remaining GB standard CRISPR elements (<https://gbcloning.upv.es/do/crispr/>).

The domestication reaction for the CRISPR/Cas9 final cassette pDGB3\_alpha1:Tnos:NptII:Pnos-SF:U6-26:tRNA-gRNA40.1.1:tRNA-gRNA40.1.2:35s:hCas9:Tnos required two domestication assemblies, one for each heterodimer in the assembly of pUPD2:tRNA-gRNA40.1 and pUPD2:tRNA-gRNA40.2. For these reactions, 2 ng of the gDNA heterodimer target, 75 ng of the GB1208 (for gRNA40.1) or GB1207 (for gRNA40.2) plasmid, 75ng of pUPD2, 5 U of BsmBI, 3 U of T4 ligase and 1  $\mu$ L of ligase buffer, were mixed to a final volume of 10  $\mu$ L.

### 9.3.3 CRISPR Multipartite Assemblies in $\alpha$ -level Destination Vectors

The multipartite assembly in alpha level destination vector was done to produce new CRISPR TU not available in the GB database (pDGB3\_alpha1:U6-26:SL4g30.1:sgRNA and pDGB3\_alpha1:U6-26:tRNA-gRNA40.1:tRNA-gRNA40.2).

Multipartite alpha level reactions were accomplished by mixing all the GB elements to incorporate in a TU in a single tube restriction-ligation reaction. 75 ng each GB part (GB1001, GB0645, pUPD2:tRNA-gRNA40.1, and pUPD2:tRNA-gRNA40.2) were mixed with 75 ng of the alpha level destination vector (pDGB3\_alpha1), 5 U of BsaI, 3 U of T4 ligase and 1  $\mu$ L of ligase buffer, were mixed to a final volume of 10  $\mu$ L.

The exception to the assemble in this level is the construct pDGB3\_alpha1:U6-26:SL4g30.1:sgRNA. In this construct, there was no need to domesticate the heterodimer in a pUPD vector since already presented the BsaI compatible overhangs. For the reaction mix, only 2ng of the heterodimer were added to the reaction. For all the other GB parts (GB1001, GB0645, pDGB3\_alpha1) 75 ng were added.

### 9.3.4 CRISPR Binary Assemblies in $\Omega$ -level Destination Vectors

The binary assembly in omega level destination vectors was used to combine TUs assembled to alpha level vector. For these reactions, 75 ng of the TU in an alpha1 vector (pDGB3\_alpha1:U6-26:SL4g30.1:sgRNA / pDGB3\_alpha1:U6-26:tRNA-gRNA40.1:tRNA-gRNA40.2) and 75 ng of the TU in an alpha2 vector (GB0639) were mixed with 75 ng of an omega level destination vector (pDGB3\_omega2), 5 U of BsmBI, 3 U of T4 ligase and 1  $\mu$ L of ligase buffer, were mixed to a final volume of 10  $\mu$ L. The resulting constructs were pDGB3\_omega2:U6-26:SL4g30.1:sgRNA:35s:hCas9:Tnos and pDGB3\_omega2:U6-26:tRNA-gRNA40.1.1:tRNA-gRNA40.1.2:35s:hCas9:Tnos.



### 9.3.5 CRISPR Binary Assembly in $\alpha$ -level Destination Vectors

To combine modules, two TUs in omega level vectors are combined in an alpha level vector. 75 ng of each omega level vector containing one or more TUs (pDGB3\_omega2:U6-26:SL4g30.1:sgRNA:35s:hCas9:Tnos / pDGB3\_omega2:U6-26:tRNA-gRNA40.1.1:tRNA-gRNA40.1.2:35s:hCas9:Tnos and pDGB3\_omega1R:Tnos:NptII:Pnos-SF) are combined with 75 ng of an alpha level destination vector (pDGB3\_alpha1), 5 U of Bsal, 3 U of T4 ligase and 1  $\mu$ L of ligase buffer, were mixed to a final volume of 10  $\mu$ L.

The resulting constructs are the final CRISPR/Cas9 cassettes pDGB3\_alpha1:Tnos:NptII:Pnos-SF:U6-26:SL4g30.1:sgRNA:35s:hCas9:Tnos and pDGB3\_alpha1:Tnos:NptII:Pnos-SF:U6-26:tRNA-gRNA40.1.1:tRNA-gRNA40.1.2:35s:hCas9:Tnos

### 9.3.6 Confirmation of the final constructs of CRISPR/Cas9 by sequencing.

To confirm that the final GB constructs were correctly assembled and the reading frame of the gRNAs was in order specific fragments of the final vector were amplified and sequenced.

To amplify, 1  $\mu$ L of each primer (F & R) at 10  $\mu$ M (Table 7); 0,38  $\mu$ L of TAQ DNA polymerase at 5 U/ $\mu$ L (Biotools, B & M Labs, S.A.); 2,5  $\mu$ L of TAQ Buffer at 10x (Biotools, B & M Labs, S.A.); 0,5  $\mu$ L of dNTPs at 2,5 mM; and 1  $\mu$ L of Mg<sup>2+</sup> at 50 mM; completed with 13,62  $\mu$ L of ddH<sub>2</sub>O to the final volume of 20  $\mu$ L. The mixture is added to 5  $\mu$ L of plasmid DNA at the concentration of 10 ng/ $\mu$ L to achieve a reaction volume of 25  $\mu$ L. The PCR condition were 94°C for 5 min for the denaturing step, followed by 30 cycles of amplification (annealing and extension) with 94°C for 30s, 58°C for 30s and, 72°C for 30s. Completed with a final 72°C for 5min.

PCR products were run in a 1% agarose gel using the Gene Ruler 100 bp Plus DNA Ladder as marker for confirmation of single amplicon presence.

After confirmation, 10  $\mu$ L of PCR products mixture was added to 4  $\mu$ L of Affymetrix® ExoSAP-IT® PCR Product Cleanup. The mixture was incubated, in the thermocycler, at 37°C for 15 min to degrade remaining primers and nucleotides and then the temperature increased to 80°C for 15min.

After nanodrop quantification, 10 ng/ $\mu$ L of treated PCR products were sent for sequencing in the “Secuenciación de ADN y análisis de la expresión génica” service in IBMCP using Sanger sequencing method.

Table 7 – List of primers used in the sequencing of the fragments from the CRISPR/Cas9 knockout constructs pDGB3\_alpha1:Tnos:NptII:Pnos-SF:U6-26:SL4g30.1:sgRNA:35s:hCas9:Tnos and pDGB3\_alpha1:Tnos:NptII:Pnos-SF:U6-26:tRNA-gRNA40.1.1:tRNA-gRNA40.1.2:35s:hCas9:Tnos.

Primer	Sequence (5'→3')	Length (bp)	T <sub>m</sub> (°C)	GC (%)
pDGB3alpha1_1 LB-Tnos F	GACTGATGGGCTGCCTGTAT	20	59.53	55
pDGB3alpha1_1 LB-Tnos R	AGCGCGCAAACCTAGGATAAA	20	57.98	45
pDGB3alpha1_2 SF-35s F	TGAACAAAGAACAATAGTGGATGAA	25	57.18	32
pDGB3alpha1_2 SF-35s R	AAAGGAGATCAGCTTGGCTCT	21	59.09	47.6
pDGB3alpha1_3 F	AGACGTCAGGTGGCACTTTT	20	59.82	50
pDGB3alpha1_3 R	AAACCTTTTCACGCCCTTTT	20	56.64	40

## 9.4 Homologous over-expression and reporter cassettes assembly.

### 9.4.1 Sequencing of the promoter region of the Solyc04g081640 gene in *S. lycopersicum* and *S. pimpinellifolium*.

The sequencing procedure was the same applied to the confirmation of the final constructs of CRISPR/Cas9 using the primers “SI4g40 prom 1 F”, “SI4g40 prom 1 R”, “SI4g40 prom 2 F”, and “SI4g40 prom 2 R”. The primer pairs used are transversal to both tomato species sequences. Details for the primer pairs are available in Annex 1.

### 9.4.2 GB Domestication

The primers sued for the domestication of the GBpatches assembled in the constructions are available in Table 8.

The amplification was done using 5 µL of genomic DNA (5 ng/µL) from the *S. lycopersicum* varieties as template and by mixing, per reaction, 1 µL of each primer at 10 µM; 0,38 µL of TAQ DNA polymerase at 5 U/µL (Biotools, B & M Labs, S.A.); 2,5 µL of TAQ Buffer at 10x (Biotools, B & M Labs, S.A.); 0,5 µL of dNTPs at 2,5 mM; and 1 µL of Mg<sup>2+</sup> at 50 mM; completed with 13,62 µL of ddH<sub>2</sub>O to the final volume of 20 µL. The mixture is added to the DNA template at the concentration for a reaction volume of 25 µL. The PCR condition were 94°C for 5 min for the denaturing step, followed by 30 cycles of amplification with 94°C for 30s, 58°C for 30s and, 72°C for 1 min. Completed with a final step of 72°C for 5 min.

To confirming the amplification of a single fragment with the expected size the PCR products were run in a 1% agarose gel using the Gene Ruler 100 bp Plus DNA Ladder as marker.

After confirmation, the PCR fragments were ready for assembly in pUPD vectors. This method was used for the assembly of level 0 GBparts: pUPD2\_SPSolyc04g081640.1CDS, pUPD2\_SLSolyc04g081640.1Prom, and pUPD2\_:SPSolyc04g081640.1Prom.

The GB domestication restriction-ligation assembly reactions to insert the GB parts (PCR fragments) in pUPD plasmids were accomplished by mixing 40 ng of the correct patch, 75 ng of pUPD vector, 5 U BsmBI, 3 U T4 Ligase and 1  $\mu$ L Ligase Buffer and water to a final volume of 10  $\mu$ L.

Table 8 – Primers for the domestication of the new GB parts. Top: primer pair for the domestication of the promoter region of the Solyc04g081640.1 gene from *Solanum lycopersicum* var. “Moneymaker” and *Solanum lycopersicum* IL4-4 (PP haplotype); Bottom: primer pair for the domestication of the coding sequence of the Solyc04g081640.1 gene from *Solanum lycopersicum* IL4-4 (PP haplotype).

Primer	Sequence (5'→3')	Length (bp)	Tm (°C)	GC (%)
<b>Sl4g40 Prom dom F</b>	GCGCCGTCTCGCTCGGAGAGAAACATAAGTGTGCCCAA	40	57.8	60.0
<b>Sl4g40 Prom dom R</b>	GCGCCGTCTCGCTCACATTTTAACTTGGGTTTGTATGAGTC	45	56.8	44.4
<b>Solyc04g40 CDS F</b>	GCGCCGTCTCGCTCGAATGTCGACTGTTATGGGGTTT	37	57.3	56.8
<b>Solyc04g40 CDS R</b>	GCGCCGTCTCGCTCAAAGCTTACCTCTCAAATCGGAGGAA	40	56.1	55.0

### 9.4.3 Multipartite Assemblies in $\alpha$ -level Destination Vectors

Multipartite assemblies were used to create the level 1 TUs:

- pDGB3\_alpha1:35s:SPSolyc04g081640.1CDS:Tnos;
- pDGB3\_alpha1:SLSolyc04g081640.1Prom:DsRed:Tnos;
- pDGB3\_alpha1:SPSolyc04g081640.1Prom:DsRed:Tnos.

For the pDGB3\_alpha1:35s:SPSolyc04g081640.1CDS:Tnos the GBparts combined were: GB0030, GB0037 and pUPD2\_SPSolyc04g081640.1CDS.

For the other to the GBparts combined were: pUPD2\_SLSolyc04g081640.1Prom / pUPD2\_:SPSolyc04g081640.1Prom, GB0100, and GB0037.

For these assemblies, 75 ng of a GBpart were mixed with 75 ng of the alpha level destination vector (pDGB3\_alpha1), 5 U of Bsal, 3 U of T4 ligase and 1  $\mu$ L of ligase buffer, were mixed to a final volume of 10  $\mu$ L.

### 9.4.4 Binary Assembly in $\Omega$ -level Destination Vectors

Each of the previously assembled TUs (level 1) were combined with another TU (pDGB3\_alpha2:Pnos:NptII:Tnos) in a omega level destination plasmid (pDGB3\_omega1) to form

a module. For these reactions, 75 ng of the TU in an alpha1 vector and 75 ng of the TU in an alpha2 vector were mixed with 75 ng of an omega level destination vector, 5 U of BsmBI, 3 U of T4 ligase and 1  $\mu$ L of ligase buffer, were mixed to a final volume of 10  $\mu$ L.

At this point the expression constructs cassettes were finished, being these omega level vectors the final constructs for the homologous over-expression (pDGB\_omega1::35s:SPSolyc04g081640.1CDS:Tnos::Pnos:NptII:Tnos), and reporting (pDGB3\_omega1::SLSolyc04g081640.1Prom:DsRed:Tnos::Pnos:NptII:Tnos and pDGB3\_omega1::SPSolyc04g081640.1Prom:DsRed:Tnos::Pnos:NptII:Tnos).

### **9.5 Transformation of *E. coli* electrocompetent cells by electroporation**

After each assembly steps, for the cloning of the constructs, aliquots (40  $\mu$ L) of DH5 $\alpha$  *E. coli* electrocompetent cells were thawed in ice until mixing. All the following steps were performed close to a flame to avoid contamination. After adding 1  $\mu$ L of the plasmid DNA from the assembly to the aliquot, the mixture was gently mixed by pipetting being cautious not to cause air bubbles. The mixture was transferred to an electroporation cuvette (1 mm gap EP-101) and placed on ice. The transfer was done carefully not to introduce air bubbles in the cuvette since the presence of those may cause de cells to disrupt. After turning on the electroporator (BTX™-Harvard Apparatus ECM™ 399 electroporator) and setting the power to 1.5 kV, the cuvette is placed in the electroporator holder and the pulse activated. Immediately, 960  $\mu$ L of SOC medium (2% tryptone, 0.5% yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride, 10 mM magnesium chloride, 10 mM magnesium sulphate, and 20 mM glucose) were added and gently mixed in the cuvette and transferred to a 1.5 mL microcentrifuge tube. To recover the cells the mixture was incubated putting the tube horizontally for 1 hour in a 37°C shaker. After incubation, using a glass spreader, 200  $\mu$ L of cells were put in LB agar plates (1% tryptone, 0.5% yeast extract, 1% sodium chloride, and 1.5% agar) containing the selection antibiotic (Chloramphenicol at 34 $\mu$ g/mL, Kanamycin, Ampicillin, or Spectinomycin at 50  $\mu$ g/mL), 0.5 mM of IPTG, and 20  $\mu$ g/mL of Xgal. The plaques are them put for 24 hours at 37°C. the selection method is the blue/white screening, the colonies transformed with the construction appeared with a white colour and the ones with intact backbone vectors appeared blue. The selected colonies were picked and transferred to sterile LB liquid medium containing the appropriate antibiotic.

## **9.6 Plasmid DNA extraction and purification from *E. coli* cultures**

To use plasmid DNA for the constructs in the GB assembly, the plasmid DNA must be extracted and purified after cloning in *E. coli* cultures, either after cloning a construction, to extract standard GB parts stored in *E. coli* glycate, or to obtain pDGB plasmids.

For the isolation of plasmid DNA, it was used the E.Z.N.A.<sup>®</sup> HP Plasmid Mini Kit I, V(capped) Spin (D6943-02) (Omega Bio-tek, Inc.).

To start, 5ml of cultures were grow, overnight in the dark at 37°C, in a 10 ml culture tube containing sterile LB liquid medium (1% tryptone, 0.5% yeast extract, 1% sodium chloride), supplemented with the selection antibiotic (1 µL per mL LB medium), then centrifuged at 10,000 x g for 1 min at room temperature. The culture medium was discarded and the pellet suspended in 250µL of Solution I containing RNase. After vigorous mixing with the vortex the suspension was transferred to a 1.5 mL microcentrifuge tube. To obtain a clear lysate, 250 µL of Solution II were added and the mix incubated for 5 minutes with gentle mixing. Afterward, 350 µL of Solution II were added and the tubes immediately inverted until a white precipitate formed. The tubes were then centrifuged at 14,000 x g for 10 min to compact the white pellet. The supernatant was then transferred to a 2ml tube equipped with a HiBind<sup>®</sup> DNA Mini Column and centrifuged at 14,000 x g for a minute. The filtrate was discarded and the tube reused. Next, 500 µL of HBC Buffer diluted with isopropanol were added to the column and centrifuged again for 60 seconds at 14,000 x g and, like before, the filtrate was discarded and the tube reused. The same procedure was repeated but this time 700 µL of DNA Wash Buffer, diluted with ethanol, were added and this step done one more time. With the aim of completely dry the column and remove the ethanol that can interfere with the next steps, the column was centrifuged at 14,000 x g for 2 min. The column was then transferred to 1.5 mL DNase/RNase free microcentrifuge tubes and 100 µL of heated Elution Buffer (70°C) added. After a 60 second incubation at room temperature the column is centrifuged at 14,000 x g for a minute, and the eluted plasmid DNA is stored at -20°C for subsequent use.

The plasmid DNA was verified by restriction fragment length analysis. Two µL of plasmid DNA were mixed with 0.5 µL of the restriction enzyme (2.5 U), 2 µL of restriction enzyme buffer and 5.5 µL of ddH<sub>2</sub>O to a total volume of 10µL. The mixture was incubated for an hour at 37°C. Digested plasmid DNA was run in a 1% Agarose gel (TAE 1x) using Lambda DNA/EcoRI plus HindIII Marker as reference marker.

## **RESULTS**

## **CHAPTER I – PHENOTYPING AND NATURAL VARIATION OF THE LOCUS IN CHROMOSOME 4 CAUSING A SD DISORDER IN TOMATO.**

A first set of experiments were designed to:

1) further support the effect of gametic or zygotic factors on SD by using BC<sub>1</sub>F<sub>1</sub> populations as a sort of continuation of Barrantes and Fakhret work, done in self-crossed populations of the F<sub>2</sub> homozygous and heterozygous plants as well as “Moneymaker” homozygotes. In those previous studies and mainly in Fakhret work (2016), the genotyping by HRM of the BC<sub>1</sub>F<sub>1</sub> progeny of reciprocal crosses indicated that both the cross of the female IL4-4 PP and the IL4-4H male and its reciprocal cross (♀ IL4-4 H x ♂ IL4-4 PP) showed a deviation from the expected ratio 1:1. In both crosses, a 3:1 proportion between homozygote PP and heterozygote was observed (p-value < 0.005) (Table 9). Progeny test of the female heterozygote cross with the male homozygote MM indicated that in this case also the PP alleles are favoured to the MM alleles by about the same 3:1 ratio (p-value < 0.05) (Table 9). However, in Fakhret study it was not possible to obtain enough seeds from the ILSP\_4-4MM-x- ILSP\_4-4 H cross to get significant information from this cross.

2) to analyse expression of the candidate genes by qRT-PCR in all three haplotypes, and

3) to perform further characterization of candidate genes by sequencing using only homozygotes from F<sub>2</sub>B<sub>2</sub>, F<sub>2</sub>B<sub>3</sub>, and “Moneymaker”.

### **1. Validation of plant material by genotyping.**

#### **1.1 Genotyping of F<sub>2</sub>B<sub>2</sub> and F<sub>2</sub>B<sub>3</sub> families.**

Plants with *S. pimpinellifolium* (PP) haplotype for the SD region, and heterozygotes (MP) were obtained by screening two F<sub>2</sub> families (F<sub>2</sub>B<sub>3</sub>, F<sub>2</sub>B<sub>2</sub> with the SNP markers Solcap\_snp\_47742 and Solcap\_snp\_3952, previously described as flanking the SD region (Figure 13) (Barrantes, 2014; Fakhret, 2016; see Annex 5 for full results).

In the case of the 19 plants from the F<sub>2</sub>B<sub>2</sub> and F<sub>2</sub>B<sub>3</sub> populations, three out of six B<sub>2</sub> plants had a heterozygous haplotype for the two markers, and for the B<sub>3</sub> population, four out of thirteen. The haplotyping by HRM of the two F<sub>2</sub> populations showed a segregation ratio of nearly 3PP:1H. The genotype of 19 “Moneymaker” was also verified with the same markers.

Eight plants with *S. pimpinellifolium* haplotype (PP), 10 plants with “Moneymaker” haplotype (MM) and 7 plants with heterozygote haplotype (MP) were selected for further experiments.

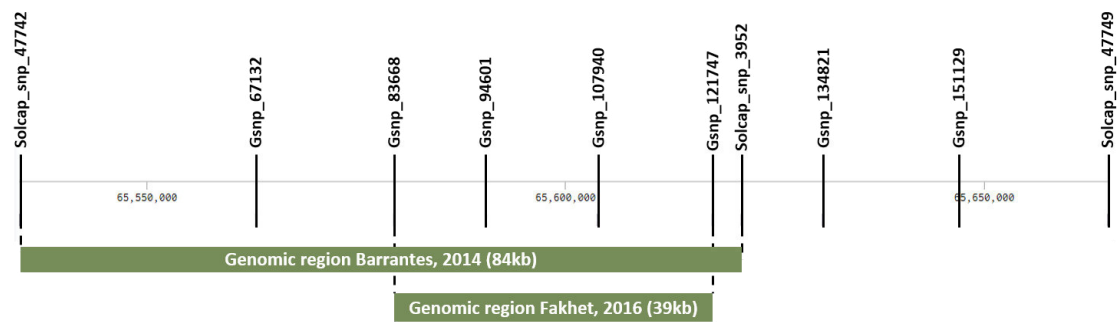


Figure 13 – Genomic SD region defined by Barrantes, et al. (2014) and narrowed by Fakhret (2016). Adapted from Fakhret 2016.

## 2. Effect of gametic and zygotic factors on SD in reciprocal crossing $BC_1F_1$ populations.

To assert if the SD is caused by pre-zygotic factors in male gametes (pollen) or female gametes (embryo-sac), or by post-zygotic factors (zygotic selection), reciprocal  $BCF_1$  crosses were performed using the IL SP4-4 homozygote (PP) and heterozygous (MP) and the recurrent parent “Moneymaker” (MM) (Table 9). The IL SP4-4 in this section are the names assigned to the  $F_2B_2$  and  $F_2B_3$  plants with different haplotype in this region.

Our hypothesis is that, if the SD is caused by pre-zygotic effect of the gene in pollen, the expected SD would only appear in the  $BC_1F_1$  progeny from crosses between heterozygote males and female homozygotes independently of the haplotype of the female gamete ( $\sigma$  IL SP4-4 H x  $\text{♀}$  IL4-4 PP and  $\sigma$  SP 4-4 H x  $\text{♀}$  MM). Correspondingly, if the SD locus acts in the female gamete, the occurrence of SD will only be observed in the  $BC_1F_1$  progeny of the crosses where the pollen comes from homozygotic plants and the maternal gamete from a heterozygote ( $\text{♀}$  IL SP4-4 H x  $\sigma$  MM and  $\text{♀}$  IL SP4-4 H x  $\sigma$  IL SP4-4 PP). However, if the SD is observed in every cross independently of the haplotype of the progenitors then this would indicate that would be caused by post-gametic factors.

Also, it is possible to distinguish from nuclear and cytoplasmic effects on SD. If the maternal gamete contains the SD allele and the SD is observed but it does not occur when the female is a MM homozygote then the SD is probably caused by cytoplasmic factors (Reflinur *et al.*, 2014).

In order to complement Fakhret study, here the  $\text{♀}$  MM x  $\sigma$  SP 4-4 H cross was repeated.

The segregation distortion was studied in seedlings from the  $\text{♀}$  MM x  $\sigma$  SP 4-4 H cross progeny using the markers Solcap\_snp\_47742, Gsnp\_83668, and Solcap\_snp\_3952. Of the 18 plants evaluated, twelve showed the MM haplotype and the remaining 6 the H haplotype.



Previously, the results of this cross gave only 8 seeds resulting in 8 heterozygotic plants. The results obtained here were added to the previous ones, and the significance study is represented in Table 9.

Table 9 – Chi-square test for distorted segregation in the progeny of the reciprocal crosses. Genotype proportions (%) are those corresponding to the genotyping HRM with the markers Solcap\_snp\_47742, Gsnp\_83668, and Solcap\_snp\_3952 for the ♀MM x ♂ILSP\_4-4H cross and the Gsnp\_67132 for the remaining crosses. ILSP\_4-4: homozygote *S. pimpinelifolium*, SP\_4-4 H: heterozygote, MM: homozygote 'Moneymaker', distorted segregation: SD (\*P<0.05, \*\*P<0.005).

Crosses	Expected genotype frequency			Observed genotype frequency			Total	χ <sup>2</sup> Test	P-value	Results
	PP	H	MM	PP	H	MM				
♀ ILSP_4-4PP x ♂ ILSP_4-4H	22 (50%)	22 (50%)	0	34	10	0	44	13,1**	0,0003	<b>SD</b>
♀ MM x ♂ ILSP_4-4H	0	13 (50%)	13 (50%)	0	14	12	26	0,15	0,69	<b>Mendelian</b>
♀ ILSP_4-4 H x ♂ MM	0	10 (50%)	10 (50%)	0	14	5	20	4,1*	0,043	<b>SD</b>
♀ ILSP_4-4H x ♂ ILSP_4-4PP	22 (50%)	22 (50%)	0	31	10	0	41	10,23**	0,0014	<b>SD</b>

The combination of both experiments resulted in 14 plants with heterozygote haplotype and 12 plants with MM haplotype, indicating that there was no distortion in the progeny of ♀ MM x ♂ ILSP\_4-4H cross. From these new results, it can be stated that the progeny of all reciprocal crosses present SD except when the MM homozygote is the female gamete donor.

These results provides evidence of a post-gametic selection and most probably due to female gametophytic or cytoplasmic factor or factors because: (1) independently of the sex of the parents a 3PP:1H progeny was observed in the reciprocal crosses where PP homozygotic and heterozygotic plants were involved; (2) the SD expresses when the heterozygote serves as a female gamete donor and is crossed with MM male gametes; (3) when all female gametes have “Moneymaker” alleles, a Mendelian segregation, rather than SD, is observed.

### 3. Seed counting of fruits from the self-pollination crosses on the three haplotypes.

In the gametophytic maternal effect class of mutants the phenotype is apparent only at the post-fertilization stage (Grossniklaus et al., 1998, Brukhin et al., 2005). The fruits of heterozygous

female gametophyte mutants exhibit reduced seed set because 50% of the female gametophytes are mutant and nonfunctional (Drews 2011). To investigate if SD of the chromosome 4 is due to gametophytic effect, the number of mature seeds produced in fruits from *S. lycopersicum* var. “MoneyMaker” homozygotes, *S. lycopersicum* SP 4-4 homozygotes and heterozygotes self-crosses was analysed (Table 10). Only fully developed seed were counted even if aborted seeds were gathered from the fruits.

The *S. lycopersicum* SP 4-4 homozygotes show the biggest average of 26.4 seeds per fruit, gathered from a collection of 4348 seeds in 165 fruits. Similarly, “MoneyMaker” homozygote fruits have an average 24.1 seeds per fruit (2627 seeds/ 109 fruits). The smallest average was obtained in the heterozygotic plants of *S. lycopersicum* SP 4-4. From 70 fruits, 820 seed were counted, averaging 11.7 seeds per fruit which is much lower than the homozygote haplotype. Thus, compared to the homozygotes, fruits from heterozygote plants showed a reduction in fruit set of more than 50%.

Table 10 – Results of the seed counting from self-pollination crosses tomato fruits.

	Total num. fruits	Total num. seeds	Average
♂x♀ <i>S. lycopersicum</i> MM	109	2627	24,1
♂x♀ <i>S. lycopersicum</i> SP 4-4 H	70	820	11,7
♂x♀ <i>S. lycopersicum</i> SP 4-4 PP	165	4348	26,4

#### 4. Analysis of the natural variation in the SD region.

##### 4.1 Validation of the haplotype in the SD region in resequenced wild accessions

To study natural variability in the SD region and to identify natural recombinants in this region, data for the 150 and 360 resequencing tomato projects (100 Tomato Genome Sequencing Consortium, 2014; Lin *et al.*, 2014) was used. A phylogenetic tree was built (Monforte, unpublished results) using all SNPs identified within the SD region as defined by Barrantes (Barrantes 2014). The phylogenetic tree, developed with *S. lycopersicum* and wild species accession showed that *S. pimpinellifolium* exhibited a large variability in this region. Wild accessions were separated in two main branches A and B.

Branch A includes the TO-937 accession while in branch B (Figure 14 branch B) there are some accessions like LA1589 (TS-19) and LA0722 (TS-412) whose progeny did not have distortion

(Barrantes 2014), which suggests that maybe Branch A could also include other SD-prone accessions.

Up to 5 clusters of accessions could be identified in branch A and in order to study the allelic variability in the SD region, 9 accessions from the most representative clusters from branch A were selected (Figure 14). The selected accessions for group 1 were LA2093, LA1246, LA1578. For group 2, LA2183, LA1242, LA2184. Groups 3, 4 and 10 are represented by LA0400, LA0417, and LA1547, respectively.

Based on the proximity to the TO-937 accession and resequencing data, the seven accessions from groups 1, 2 and 3 were expected to have a *S. pimpinellifolium* haplotype in all the extension of the studied region. Groups 4 and 10, however, were expected to have different haplotypes than the other groups, caused by some SNPs, although they would be closer to the *S. pimpinellifolium* TO-937 haplotype.

Although group 11 is not so tightly clustered as the others, three accessions with special interest were chosen because they represent probable natural recombinants. One *S. pimpinellifolium* accession, LA1245, and two *S. lycopersicum* var. *cerasiforme* accessions, LA1456 and LA2688. Based on the resequencing data the *S. pimpinellifolium* accession LA1245 has many heterozygotic SNPs in the intergenic region between the Heat Shock Protein gene and the Endoplasmic reticulum chaperone gene, Solyc04g081630 and Solyc04g081640, respectively. The two *S. lycopersicum* var. *cerasiforme* accessions could be recombinants in the same intergenic region and for that reason were also chosen.

The haplotype of the selected accessions was confirmed by KASP analysis interrogating the 11 SNPs selected for this region (Figure 15, Annex 6; see annex 2 for primers). Four SNP markers target 4 of the 7 genes included in the 39 Kb SD region (Fakhet 2016). SNP markers 1, 4, 9, and 11 target a variation inside the uncharacterized gene Solyc04g081620.1, the Heat Shocking Protein gene (Solyc04g081630.1), the Endoplasmic reticulum chaperone gene (Solyc04g081640.1), and the Cyclin gene (Solyc04g081650.2), respectively. The remaining SNPs are located in intergenic regions.

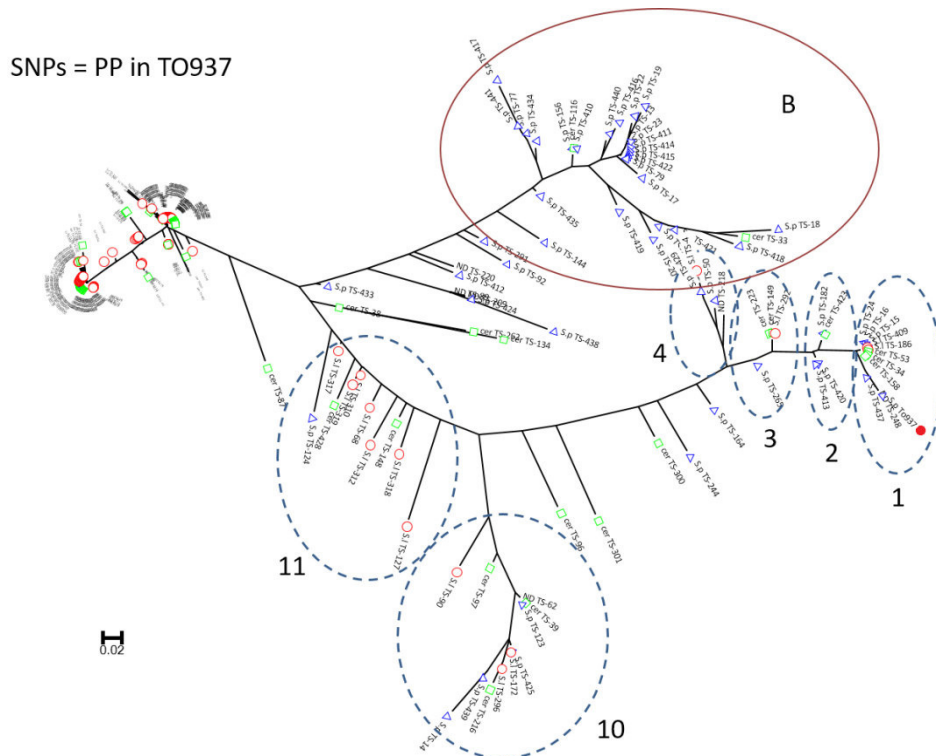


Figure 14 – Neighbor-joining tree representing the proximity of other tomato accessions to the *S. pimpinellifolium* T0937 based on the total number of equal SNPs. 1, 2, 3, 4, 10 and 11 represent the groups in branch A of the *S. pimpinellifolium* and *S. lycopersicum* var. *cerasiforme* accessions. Red dot shows the position of the *S. pimpinellifolium* T0937 accession on the tree. Group B accessions were not used.

Of the 12 accessions analysed only one, LA0400, corresponded exactly to the expected haplotype deduced from the resequenced data. In contrast, LA0417, showed a wild-type (MM) haplotype for the region when it was only expected one wild-type SNP. The accessions from group 1 and two of group 2, LA2183 and LA1242, showed a lot of SNPs in heterozygosis which contrasts with the complete mutant (PP) haplotype that was expected. Accession LA 2184, the remaining accession of group 2, presented an unexpected result by having wild-type (MM) SNP in the first two and last two markers of the studied region and 3 heterozygotic SNPs covering the Heat Shock Protein gene and the Endoplasmic homolog gene. The group 10 accession, LA 1547, gave a similar result with the first two and the second to last marker also showed a wild-type SNP but no heterozygote SNPs. Figure 15 shows the physical distribution of the KASP marker along the SD region and the results of the KASP analysis of the resequenced wild accessions.



Figure 15 – Physical map of the KASP markers. Top: schematic representation of the SD region between the markers Gsnp\_83668 and Gsnp\_121747, each KASP marker is marked on the absolute physical position; Bottom: Results of the KASP analysis of the wild accessions populations of the selected accessions of *S. pimpinellifolium* and *S. lycopersicum var. cerasiforme*. Genomic coordinates of the SNP in the SL2.5 assembly are marked on top of each column. The red boxes represent the wild-type nucleotide, blue boxes the mutant nucleotide and green boxes the cases where both are present.

## 4.2 Analysis of the SD in the selected wild accessions

Based on the KASP results obtained, 5 wild accessions representing all groups identified by phylogenetic analysis except for the group of special interest which was not confirmed, were selected to evaluate whether they show or not the SD phenotype by conducting a progeny test in  $F_2$  populations generated by crossing each of them with “MoneyMaker”. Selected accession LA1578 was chosen to represent group 1 since it has only one heterozygote SNP from all the SNPs considered in the region and, of all 3 accessions in this group, strongly confirmed the expected haplotype. LA0400 accession was chosen to represent group 3 because it confirmed the expected haplotype and LA0417 accession was selected for group 4 but for the opposite reason since it gave a “MoneyMaker” haplotype when it was expected mostly a PP haplotype. LA2184 and LA1547 were selected to represent Group 2 and 10 accessions, and for the possibility of being recombinants possibly shortening the SD region to 16Kb.

Progeny tests are currently being performed at the IBMCP and ISHM. If the SD appears in all progenies then maybe the SD locus is located outside the studied region, which is unlikely.

However, if only some or even one progeny presents the SD, then this would be a starting point for further studying of the SD region and would indicate that not all *S. pimpinellifolium* accessions chosen from the clusters near the *S. pimpinellifolium* TO-937 accession are able to affect the segregation ratio when crossed with the recurrent parent “Moneymaker”. The best scenario would be to obtain SD in crosses involving the wild accession LA2184, since this would mean that the SD is between 65,590,666bp to 65,606,665bp on the distal part of chromosome 4. This would narrow down to a 16 Kb region and include only the HSP90 genes, Solyc04g081630 and Solyc04g081640.

## **CHAPTER II – GENOMIC AND TRANSCRIPTOMIC ANALYSIS OF THE CANDIDATE GENES.**

### **1. Candidate genes in the SD region and functional *in silico* analysis.**

The genomic representation on Figure 16 shows the physical distribution of the candidate gene sequences along the genomic region where SD mapped according to Fakhret (2016). The annotation available on the Table 11 shows the description of the seven candidate genes as it is on the ITAG 2.4 of the Sol Genomics Network for the genome build 2.5 of the *S. lycopersicum* Heinz reference genome. The resulting blast hits for *A. thaliana* homologs and respective GO are included in Table 11.

The Solyc04g081620.1 and Solyc04g081680.2 genes were annotated as unknown protein coding genes. The homology search gave no hits in *Arabidopsis thaliana* for homologous gene sequences. *In silico* expression of these genes indicated that, Solyc04g081620.1 gene has a defined spatial and temporal expression in fruits that starts in the 1 cm fruit, increases to the 2 cm and then is absent in the remaining time of fruit development. It is also expressed in roots and leaves at different time periods, especially in leaf primordia (Annex 7a). The *in silico* analysis of expression for the Solyc04g081680.2 gene revealed that it is expressed in almost all tomato tissues with a higher level of expression in fruits after the breaker stage (Annex 7g).

The Solyc04g081630.1 and Solyc04g081640.1 genes have the same GO and homolog in *A. thaliana*. Both are related to a response to several stresses because they encode molecular chaperones linked to stress response in stabilizing and folding other proteins. These genes have a more localized gene expression, as indicated by the *in silico* expression analysis. The Solyc04g081640.1 gene is expressed mainly during the first days of fruit set until 7 DPA (Annex 7c). Although it belongs to the same family, Solyc04g081630.1 has a more widespread expression pattern, is expressed in other organs and more significantly in roots and flowers (Annex 7b). It is important to mention that these two genes are also highly expressed in the embryo on the 4 DPA, a stage of fruit development characterized by cell division, in the *S. pimpinellifolium* LA 1589 accession that does not present SD (Annex 7b and 7c). The expression results in this accession would support the hypothesis of the participation of this gene in SD and a requirement for expression in embryo development for normal segregating populations.

The Solyc04g081650.2 and Solyc04g081660.1 genes code for type B Cyclins, and share the same homolog in *A. thaliana*. Solyc04g081660.1 is the only one for what there is evidence in

literature for its biological function in tomato. The *in silico* analysis of the cyclin genes showed expression in all plant tissues, as expected for a protein involved in cell division.

The Solyc04g081670.2 has several homologs in *A. thaliana*. All the homologs encode isoforms of Vacuolar Processing Enzymes (VPEs) (Gruis *et al.*, 2002 and 2004; Nakaune *et al.*, 2005). Like the cyclin genes, this VPE gene is expressed in all tissues with no significant differences through development.

Although the *in silico* expression analysis of these last 3 genes showed high expression on the 4 DPA in the LA 1589 accession, the results are more significant in the case of genes encoding for Heat Shock Proteins. Cyclins and VPE genes are expressed at high values throughout all tissues contrasting with the more specific tissue expression shown by Heat Shock Protein 90 genes.

The Blast analysis for *A. thaliana* homologs of the tomato sequences evidenced a close relationship of these sequences in other species where the genes are not characterized.

The predicted mRNA of the candidate genes showed homology with other gene sequences of species of the Solanaceae family (Table 11). Unfortunately, most of them are also predicted sequences of the reference genomes of other species. All tomato sequences presented show similar sequences in the *Nicotiana* genus but phylogenetically are more distant than other *Solanum* species.



Table 11 – Annotations available in the Sol Genomics Network for the candidate genes in the Heinz reference genome. The homologs of *A. thaliana* and the respective GO resulted from blast analysis is described in columns 6 and 7. The gene length and respective coordinates are according to the SL2.5 assembly. \*inferred from direct assay (Zhang *et al.*, 2014).

ITAG 2.4 locus	Description	Length (bp)	Genomic coordinates SL2.50ch04	Solanaceae family homologs	<i>A. thaliana</i> homologs	Gene Ontology (GO) of the <i>A. thaliana</i> homologs
<b>Solyc04g081620.1</b>	Unknown Protein (AHRD V1)	213	65590248..65590460		-	-
<b>Solyc04g081630.1</b>	Heat shock protein 90 (AHRD V1 ***: Q9MB32_ORYSA); contains Interpro domain(s) IPR015566 Molecular chaperone, heat shock protein, endoplasmic	2 146	65595823..65597968			<b>Biological process:</b> response to stress protein folding protein secretion regulation of meristem growth regulation of meristem structural organization response to cadmium ion response to cold response to endoplasmic reticulum stress response to salt stress response to water deprivation <b>Molecular function:</b> ATPase activity ATP binding unfolded protein binding
<b>Solyc04g081640.1</b>	Endoplasmic homolog (AHRD V1 *-: ENPL_HORVU); contains Interpro domain(s) IPR015566 Molecular chaperone, heat shock protein, endoplasmic	1 722	65604519..65606240	<i>S. pennellii</i> (XM_015217347.1); <i>S. tuberosum</i> (XM_015306456.1)	At4g24190	<b>Biological process:</b> cell cycle cell division <b>Cellular component:</b> nucleus
<b>Solyc04g081650.2 (UNIGENE ID: SGN-U587173)</b>	Cyclin B (AHRD V1 ***: B6V744_POPTO); contains Interpro domain(s) IPR004367 Cyclin, C-terminal	752	65607611..65608362	<i>S. pennellii</i> (XM_015216347.1); <i>S. tuberosum</i> (XM_006342070.2)	At1g76310	<b>Biological process:</b> involved in regulation of mitotic cell cycle*
<b>Solyc04g081660.1 (SYMBOL: SLCYCB2_3)</b>	Cyclin B (AHRD V1 ***: B6V744_POPTO)	1 097	65608852..65609948		At1g76310	<b>Molecular function:</b> cysteine-type endopeptidase activity <b>Biological process:</b> proteolysis involved in cellular protein catabolic process vacuolar protein processing
<b>Solyc04g081670.2</b>	Vacuolar processing enzyme-3 (AHRD V1 **:- Q852T0_TOBAC); contains Interpro domain(s) IPR001096 Peptidase C13, legumain	969	65612054..65613022	<i>Capsicum annuum</i> (NM_001324664.1); <i>S. pennellii</i> (XM_015205510.1); <i>S. tuberosum</i> (XM_006342447.2)	At1g62710, At2g25940, At3g20210, At4g32940	
<b>Solyc04g081680.2</b>	Unknown Protein (AHRD V1)	1 423	65617423..65618845		-	-



Figure 16 – Schematic representation of the SD genomic region mapped by Fakhret (2016) (Physical Coordinates Chr4: 65.579.486..65.617.68).

## **2. Expression analysis of the candidate genes by qRT-PCR and Comparative $C_T$ ( $\Delta\Delta C_T$ ) analysis.**

The aim of this experiment was to define if any of the candidate genes show a differential pattern of expression in reproductive tissues that could be associated to the materials showing or not SD. For that we measured and compared the expression levels of candidate genes in pre- and post-anthesis reproductive structures obtained from “MoneyMaker” (MM), SP 4-4PP and SP 4-4H genotypes in several flower and fruit development stages.

The time course sampling covers ovaries and anthers from bud stage (BS), ovaries and anthers one day before anthesis (-1DBA), ovaries and anthers on the day of anthesis (0 DPA), and only ovaries on the 4<sup>th</sup>, 7<sup>th</sup>, 10<sup>th</sup> and 13<sup>th</sup> day post anthesis (4 DPA, 7 DPA, 10 DPA, 13 DPA).

The landmark anthesis was chosen as the reference time landmark (0 DPA) since is where the flower and fruit development overlap. The day of anthesis is when fecundation can occur but the female tissues were collected without fertilization. The sample collected from ovaries of “MoneyMaker” on the day of anthesis was chosen as the reference sample so the fold differences could be show, simultaneously, through time and between haplotypes after fertilization.

To determine expression levels before anthesis, anthers and ovaries were collected in two different times, an earlier sample was picked when the flower is still a closed bud, approximately 8 to 4 days before anthesis. The closed bud represents a stage were male meiosis has not occurred yet. The results of these samples were not included in the studies of comparative  $C_T$  ( $\Delta\Delta C_T$ ) because the data was not relevant in what concerns gene expression in gamete development.

One to two days before anthesis (-1 DBA) landmark represents the stage of gamete functionality without fertilization. For three stages above, both ovaries and anthers were collected. In the next four stages, since the fecundation already occurred, only ovaries were collected. The first collection of samples representing reproductive structures after fecundation is on the fourth day after anthesis (i.e. 4 DPA), when the embryo is at the cell division stage. The next two collections were on the seventh and tenth days after anthesis (i.e. 7 DPA and 10 DPA). These two stages are marked by rapid fruit growth caused mainly by cell elongation and the beginning of cotyledon formation. Finally, on the thirteenth day after anthesis (i.e. 13 DPA), the fruit enters an exponential phase of growth and the cotyledons start to elongate.

The complete set of plotted graphs is available in Annex 3 as well as the tables with the raw values for the  $C_T$  analysis.

The results of the qRT-PCR indicated that Solyc04g081620, Solyc04g081650, Solyc04g081660, Solyc04g081670, and Solyc04g081680 were not differentially expressed between MM and SP 4-4PP and SP4-4H (Annex 3).

The two other candidate genes, Solyc04g081630 (Figure 17) and Solyc04g081640 (Figure 18), show, different expression patterns in “Moneymaker” plants and in SP 4-4 PP and SP 4-4 H genotypes.

The Solyc04g081630 gene (Figure 17), encoding a Heat Shock Protein 90 (HSP90), showed differences in expression between 3 to 6-fold in SP 4-4 PP and SP 4-4 H ovary samples in respect to MM. In fruit tissues, there is an increase in expression between the anthesis to the 7 DPA with a significant drop to the 10 DPA in the homozygous plants. Expression in heterozygous fruits drops from the 4 DPA to 7 DPA but rises again, surpassing significantly the expression levels of the homozygote, by 13 DPA. In anthers samples before anthesis there is no clear expression nor significant differences from the “Moneymaker”.

Solyc04g081640 gene (Figure 18), encoding also a HSP90 showed the highest expression differences between “Moneymaker” plants and SP 4-4 PP and SP 4-4 H. The expression profile of Solyc04g081640 in SP 4-4 PP and SP 4-4 H ovaries follows the same trend as Solyc04g081630 (Figure 17) with values between 9 to 12-fold. The anthers samples collected on the day of anthesis, however, present a higher fold-change in SP 4-4 that in “Moneymaker”.

qRT-PCR analysis indicated that the strongest candidates to encode the SD locus based on the differential expression responses are the Solyc04g081630 and Solyc04g081640.

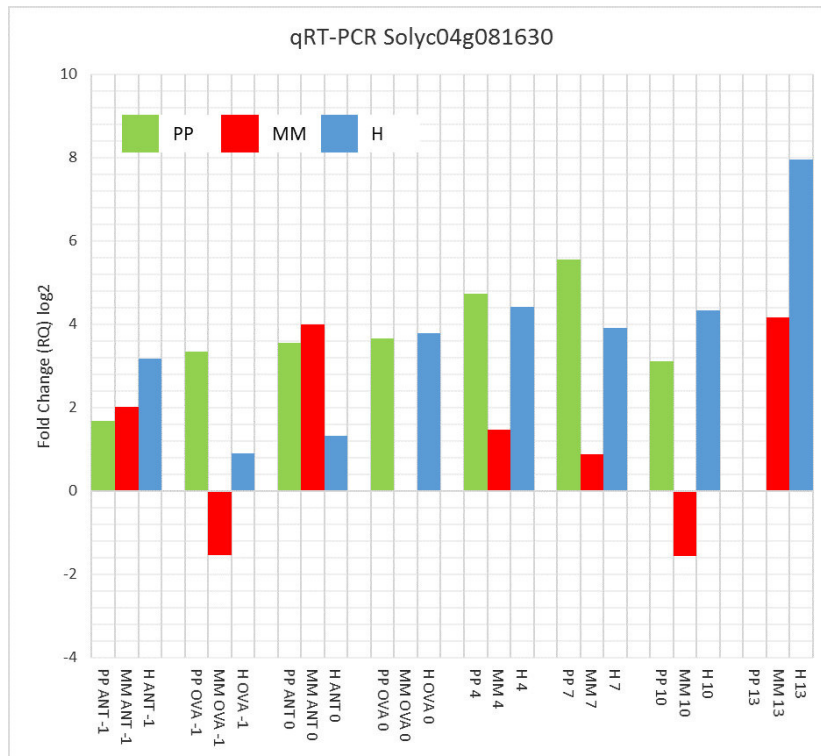


Figure 17 – Plotting of the results of the qRT-PCR of the candidate gene Solyc04g081630 encoding to a HSP90. Green bars represent samples from the *S. lycopersicum* SP 4-4 homozygote. Red bars the *S. lycopersicum* var. “MoneyMaker” control plants and the blue bars the *S. lycopersicum* SP 4-4 heterozygotes.

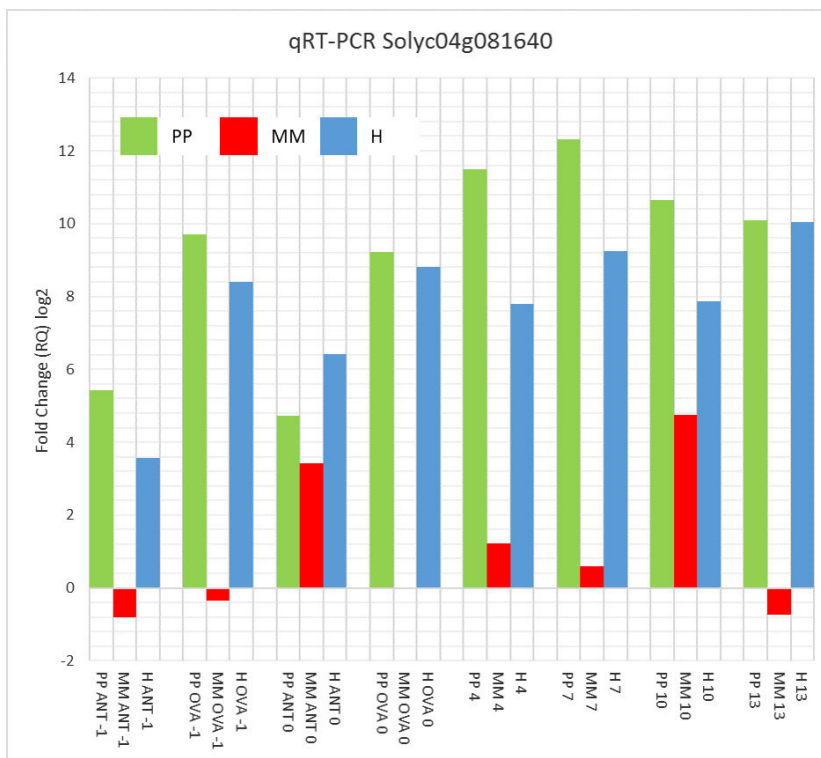


Figure 18 – Plotting of the results of the qRT-PCR of the candidate gene Solyc04g081640 encoding to a HSP90. Green bars represent samples from *S. lycopersicum* SP 4-4 homozygote. Red bars the *S. lycopersicum* var. “MoneyMaker” control plants and the blue bars the *S. lycopersicum* SP 4-4 heterozygotes.

### 3. Candidate gene sequencing, structural prediction and SNP effects.

To further investigate which of the candidate genes may underlie the SD locus, SNPs that could correlated with the SD phenotype were searched. All the genes in the SD region were sequenced in both *S. lycopersicum* var. “MoneyMaker” and the *S. lycopersicum* SP4-4 PP. The sequences obtained corresponded to the gene annotations of ITAG 2.4.

Comparison of the genes in *S. lycopersicum* “MoneyMaker” and SP4-4 PP with the Heinz genome reference indicated that “MoneyMaker” genes were, as expected, equal to Heinz genes with no exception, maintaining the same nucleotidic and, therefore, aminoacidic sequences. On the other hand, the PP haplotype gene sequences contained in total 24 polymorphisms, 14 of those where located within the gene sequences and are previously described variants, including 13 SNPs and 1 insertion. Eight polymorphisms are within coding sequences and the remaining 6 in intron regions (Table 12).

Table 12 – List of variants detected in *S. lycopersicum* SP4-4 against the 150 and 360 tomato resequencing project references.

Gene	Physical position	Gene structure	Variant type	Mutation	Polymorphism (positive strand)
Solyc04g081630	65,595,936	Exon 3	SNP	Missense	C->T
	65,595,979	Exon 3	SNP	Synonymous	T->A
	65,597,172	Intron 1	SNP		C->T
	65,597,566	Intron 1	SNP		C->T
	65,597,592	Intron 1	SNP		A->C
Solyc04g081640	65,605,455	Intron 4	SNP		A->C
	65,606,086	Exon 2	Insertion	Synonymous but frame-shifting	GAAAAAAAAA > GAAAAAAAAA
	65,606,095	Exon 2	SNP	Nonsense	C->T
Solyc04g081650	65,608,352	Exon 1	SNP	Missense	G->A
Solyc04g081660	65,608,871	Exon 5	SNP	Missense	A->G
	65,609,414	Exon 3	SNP	Synonymous	T->C
	65,609,623	Intron 2	SNP		C->T
Solyc04g081680	65,617,632	Exon 2	SNP		A->G
	65,618,311	Intron 1	SNP		C->T

To analyse whether these variants could affect protein function, the amino acid sequence and the secondary and tertiary structure of the resulting predicted protein were analysed for each of them. The analysis indicated that Solyc04g081620 and Solyc04g081670 do not present any polymorphism within their genomic sequence. In the case of Solyc04g081630, Solyc04g081650, Solyc04g081660, despite the presence of some polymorphisms in their coding sequences (Figure

16), all of them are synonymous or missense (Figure 19) and therefore have no likely effect on the protein folding as confirmed by *in silico* protein structure analysis. The two polymorphisms in Solyc04g081680 were both synonymous mutations.

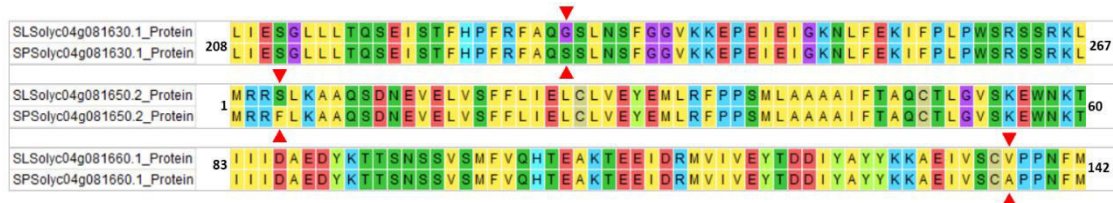


Figure 19 – Predicted protein sequence alignment of the *S. pimpinellifolium* and *S. lycopersicum* sequences of the genes Solyc04g081630, Solyc04g081650, Solyc04g081660. Red arrows signalize the aminoacidic change. Numbers indicate the position of the amino acids on the borders.

According to the ITAG 2.4 annotation (Figure 21a) the gene Solyc04g081640 encodes a Endoplasmic homolog, a protein from the same family of HSP90. Solyc04g081640 gene showed the most substantial differences in the genomic sequence. Considering the coding strand (- strand), the two most significant polymorphisms were located at 146 and 155bp downstream of the transcription starting site. A T insertion was observed in exon 2 at nucleotide 155 (155insT) after a sequence with 8 consecutive thymine (Figure 20 line 2 and Figure 21b). 156insT resulted in a frameshift at the third nucleotide of codon 37 (Figure 20 line 2 and Figure 21b), without aminoacidic substitution, since the codon TTC is change to the TTT both translated to a Phenylalanine. Upstream the 9 thymine repetition, a guanine to adenine substitution (G146A) results in change in the third nucleotide of codon 34 and in the introduction of a premature stop codon (TGA). The predicted truncated protein in *S. pimpinellifolium* would have 34 amino acids of length (Figure 21b) instead of the 159 amino acids predicted for MM.

The presence of these 2 variants was also studied by sequencing and analyses of the *S. pimpinellifolium* and *S. lycopersicum* var. *cerasiforme* accession described earlier in chapter I. Variant analysis indicates that three combinations of these 2 variants exist (Figure 20 lines 3-12). *S. pimpinellifolium* accessions LA0417 and LA1245 and *S. lycopersicum* var. *cerasiforme* accession LA1546 share the combination of “Moneymaker”. *S. pimpinellifolium* accessions LA2093, LA1578, LA1242, LA0400, and LA1547 and *S. lycopersicum* var. *cerasiforme* accession LA2688 share to the TO-937 *S. pimpinellifolium* sequence. One exception is the *S. pimpinellifolium* accession LA2184, that presents the 155insT but not G146A. In summary, according to the genomic sequence of Solyc04081640 most of the *S. pimpinellifolium* accessions analysed showed the substitution or addition that most likely would abolish the function of the gene.





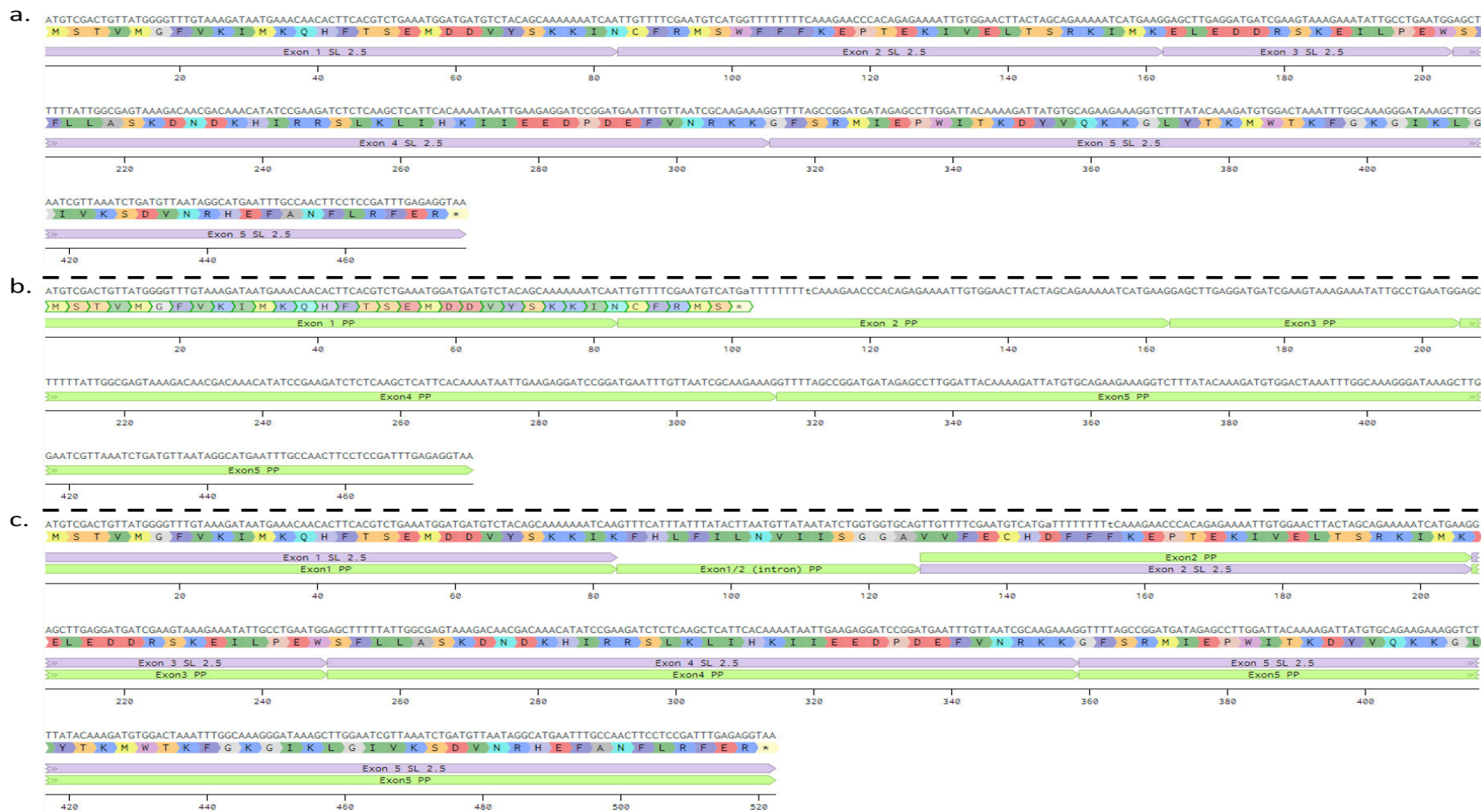


Figure 21 – Graphical representation of the nucleotide sequences of the genomic annotation (a.), sequenced genomic DNA (b.) and sequenced cDNA (c.) of the Solyc04g081640 gene in the SP4-4. Nucleotide sequence (first line), the aminoacidic sequence (second line), genomic annotation in the Heinz reference genome (SL2.5) (purple), genomic annotation obtained by sequencing from the *S. lycopersicum* SP4-4 homozygote samples (green). Graph constructed using Benchling (<https://benchling.com>).



## **CHAPTER III – DEVELOPMENT OF BIOTECHNOLOGY TOOLS FOR SD FUNCTIONAL ANALYSIS**

### **1. Design and construction of novel expression cassettes for *Solanum lycopersicum* using GoldenBraid 3.0 toolbox.**

To further investigate the role of the candidate genes Solyc04g081630 and Solyc04g081640 in regulating SD we generated a suite of genome editing tools. Three tools types were generated to address different questions:

- 1) Would silencing of the HSP 90 genes, Solyc04g0816430 and Solyc04g081640, in the *S. lycopersicum* SP4-4 homozygote (PP), eliminate the SD phenomenon in the reciprocal crosses?
- 2) Would overexpression of the *S. pimpinellifolium* Solyc04g081640 gene in *S. lycopersicum* var. “Moneymaker” skew the segregation of the progeny in the reciprocal cross ♀ MM x ♂ SP 4-4 H?
- 3) Are the tissue location and modulation of expression differences between *S. pimpinellifolium* and *S. lycopersicum* driven by the promoter sequence of the Solyc04g081640 gene determinant for the development of the SD?

All constructs were designed using the GB 3.0 toolbox. The GBparts, used and generated during the development of the final constructs are listed and described, in Annex 4.

The restriction fragment length analysis (Figure 23b and Figure 24b) was used as quality control for all steps along the way to obtain the expected results throughout all 5 different construct pipelines.

#### **1.1 CRISPR/Cas9 cassettes for the silencing of candidate genes.**

To find out whether expression of either Solyc04g0816430 or Solyc04g081640 in the SP 4-4 was linked to the occurrence of SD, two CRISPR/Cas9 cassettes were developed with the GB3.0 toolbox in order to edit the coding sequence of those genes. The technology is working very well in the PBB Lab at the IBMCP and we expect that the modification introduced by NHEJ after the cleavage and endonuclease activity by CRISPR/Cas9 activity will provoke a frame-shift in the coding sequence and create a non-functional protein.

Each cassette targets the Solyc04g081630 and Solyc04g081640 genes. Figure 22a and Figure 22b show the schematic representation of final constructs highlighting the gene structures included in the plasmids.

To knockout the Solyc04g0816430 and Solyc04g081640 genes efficiently, target sites were selected in CRISPR/Cas9-compatible sites close to the beginning of the coding sequence. The targets we selected respecting the limitations of the CRISPR/Cas9 system and GB toolkit. CRISPR/Cas9 20-nt target sequences must precede a canonical PAM sequence and should have no off-target activity.

Based on the selection limitations, 1 sgRNA (form now on mentioned as SL4g30) was designed for gene Solyc04g0816430 (Figure 23a) and 2 sgRNAs (SL4g40.1 and SL4g40.2) for gene Solyc04g081640 (Figure 24a). SL4g30 targets the first exon between 143 and 162bp downstream of the TSS; SL4g40.1 targets the 158 to 177bp and SL4g40.2 targets 483 to 502bp downstream of the TSS located in exon 2 and 4, respectively.

In both vectors, the RNA polymerase III promoter, U6-26, from *A. thaliana* was used to drive expression of the sgRNAs, whereas the expression of Cas9 would be driven by the constitutive promoter, 35S.

Also, both cassettes contain a kanamycin/neomycin resistance TU (NptII) for selection of transformants in plant; confirmation by sequencing revealed the correct assembly of the cassettes for the 3 TUs (sgRNAs, hcas9, and NptII) in each construct (Figure 23c). The sequences were compared with the *in silico* sequences of the constructs. Figure 23a and Figure 24a represent the pipeline of assembly of the constructs *pDGB3\_alpha1::Tnos:NptII:Pnos-SF:U6-26:gRNA30.1:sgRNA::35s:hcas9:Tnos* and *pDGB3\_alpha1::Tnos:NptII:Pnos-SF:U6-26:tRNA-gRNA40.1.1:tRNA-gRNA40.1.2::35s:hCas9:Tnos*, respectively.

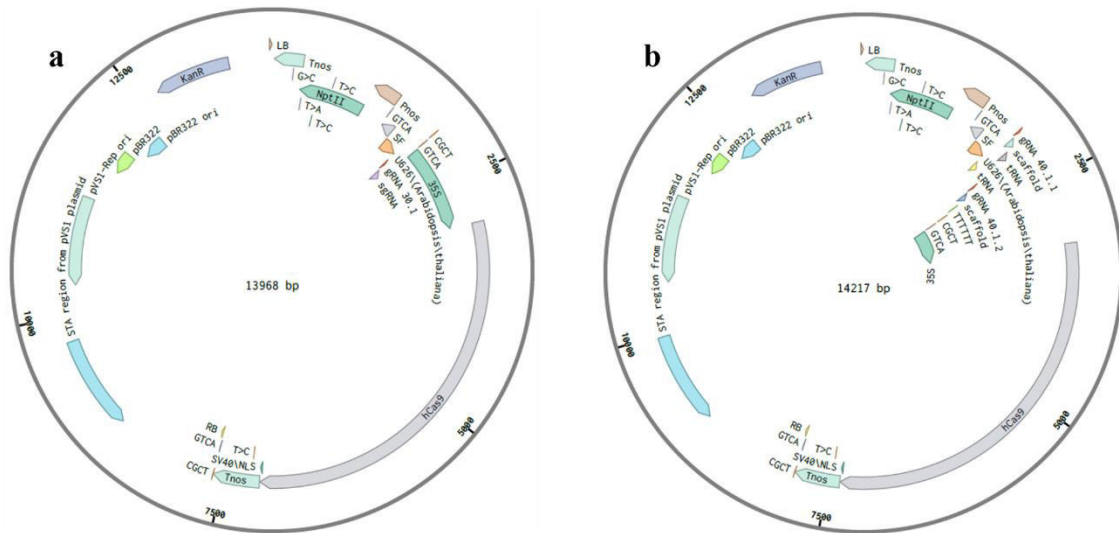


Figure 22 – Schematic representation of the CRISPR/Cas9 silencing plasmids. a. construct for the silencing of the Solyc04g081630 gene. b. construct for the silencing of the Solyc04g081640 gene.

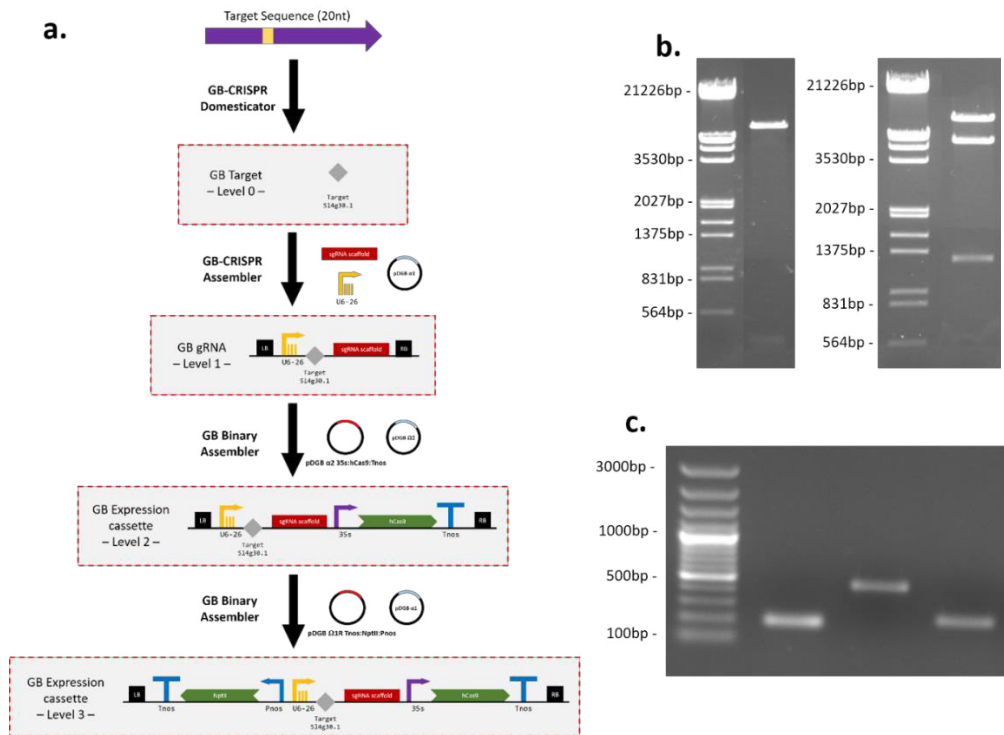


Figure 23 – Assembly of the CRISPR/Cas9 construct for Solyc04g081630. a. Pipeline of assembly using the GB 3.0 toolbox. b. Restriction analysis of the level 1 (left – expected sizes 6345bp and 372bp) and level 2 (right – expected sizes 6674bp, 1236bp, 4620bp, 194bp). c. Amplicons for sequencing of the level 3 cassettes. From left to right: primer pair “pDGB3alpha1\_1 LB-Tnos F” and “pDGB3alpha1\_1 LB-Tnos R” expected size 195bp; primer pair “pDGB3alpha1\_2 SF-35S F” and “pDGB3alpha1\_2 SF-35S R” expected size 478bp; primer pair “pDGB3alpha1\_3 F” and “pDGB3alpha1\_3 R” expected size 236bp.

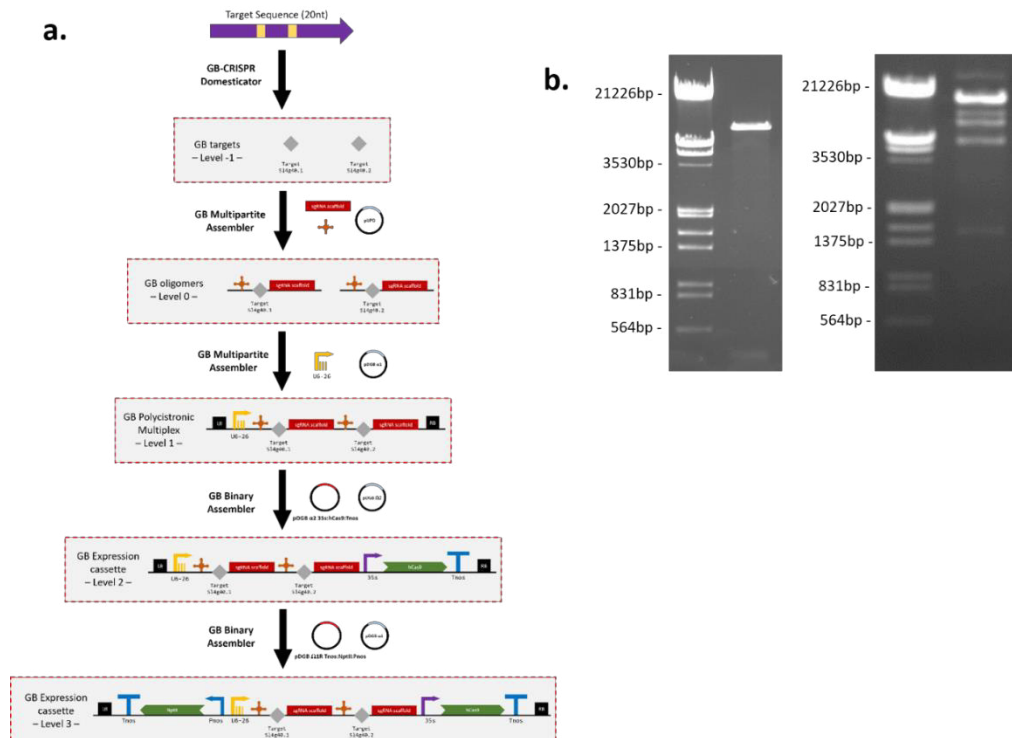


Figure 24 – Assembly of the CRISPR/Cas9 construct for Solyc04g081640. a. Pipeline of assembly using the GB 3.0 toolbox. b. Restriction analysis of the level 1 (left – expected sizes 6345bp and 621bp) and level 2 (right – expected sizes 6674bp, 4620bp, 1485bp, 194bp).

## 1.2 Homologous over-expression of *Solanum pimpinellifolium* Solyc04g081640 gene in *Solanum lycopersicum* genetic background.

In order to explore if the increased expression of *S. pimpinellifolium* Solyc04g081640 gene during reproductive phase is the reason for the SD a homologous over-expression construct was developed.

The GB 3.0 overexpression construct was developed containing the full sequence of the gene under the control of a constitutive promoter (35S) and a Tnos terminator.

The amplification and domestication of the coding sequence of *S. pimpinellifolium* Solyc04g081640.1 was done using primers designed by the GBdomesticator tool against the sequence of the gene from Tomato reference genome SL2.5 (ITAG 2.4). The pipeline of assembly is represented in Figure 25a.

The plasmid designed for the homologous over-expression of *Solanum pimpinellifolium* Solyc04g081640 gene in *Solanum lycopersicum* genetic background is represented in Figure 25b. Like the CRISPR/Cas9 cassettes an antibiotic resistance gene (NptII) was included for transformed tissue selection.

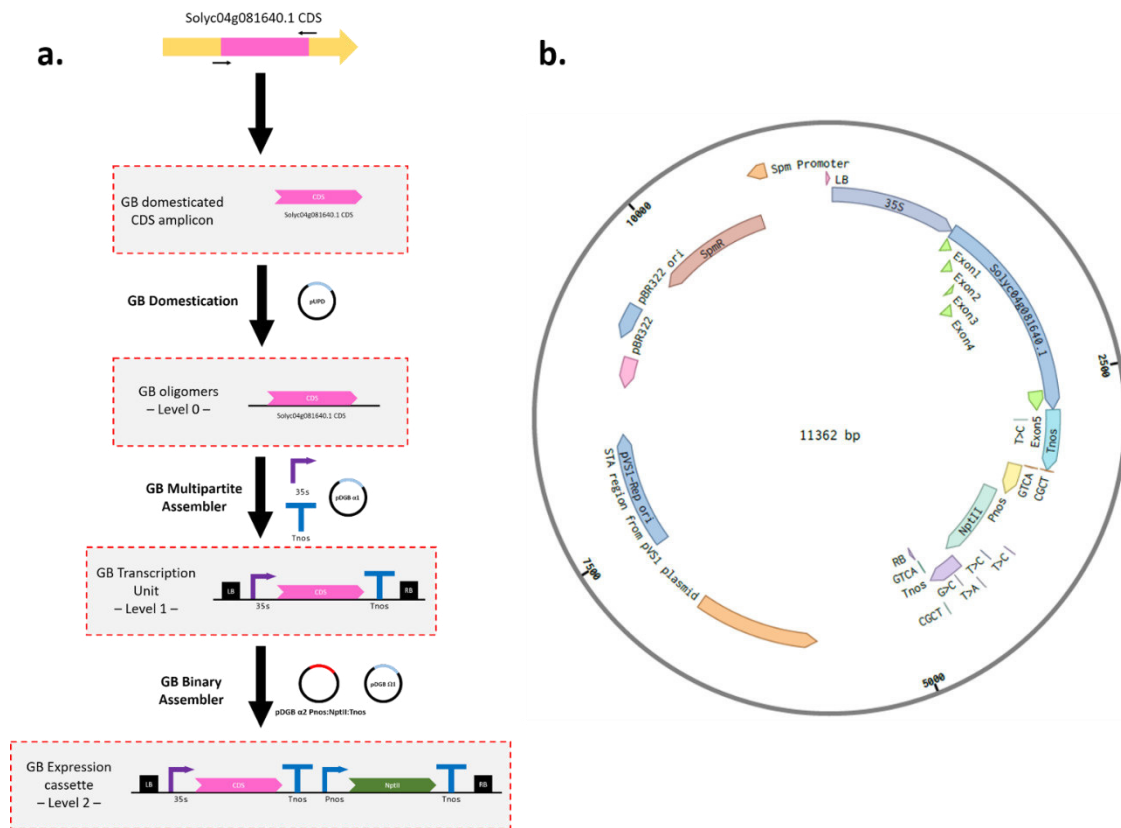


Figure 25 – Assembly of the plasmid pDGB3\_omega1::35s:SPSolyc04g081640.1CDS:Tnos::Pnos:NptII:Tnos a. pipeline of assembly with the GB 3.0 toolbox. b Schematic representation of the plasmid for Homologous over-expression of *Solanum pimpinellifolium* Solyc04g081640 gene.

### 1.3A reporter construct for the evaluation and spatial expression pattern driven by the *Solanum pimpinellifolium* and *Solanum lycopersicum* Solyc04g081640 gene promoter.

The promoter regions corresponding to the gene of interest were amplified by PCR from the SP 4-4 and the “Moneymaker” samples, and verified, by sequencing, that they include 1574 bp and 1552 bp upstream of the corresponding ATG for the gene, respectively. The alignment of the proximal promoter sequences is available in the Annex 8. Although the similarity of the possible promoter region sequences is very high, the proximal sequence from *S. pimpinellifolium* contains a region with 6 CCTT repeated in tandem while it is only repeated 3 times in *S. lycopersicum*. Other polymorphisms found were 6 SNPs and a large InDel.

A search of proximal promoter elements that could modulate, enhancing or repressing, the expression of the gene was done *in silico* using the bioinformatic tools, The Plant Promoter Analysis Navigator (PlantPAN; <http://PlantPAN2.itps.ncku.edu.tw>), TSSPlant (<http://www.softberry.com/berry.phtml?topic=tssplant&group=programs&subgroup=promoter>), and Statistical Motif Analysis in Promoter or Upstream Gene Sequences

(<https://www.arabidopsis.org/tools/bulk/motiffinder/index.jsp>). The *in silico* analysis revealed some possible motifs created by the sequences repetition in the SP 4-4 genome that do not exist in the *S. lycopersicum* var. “MoneyMaker” genome.

To evaluate the effect on gene expression by the different sequences of the proximal promoter regions of the Solyc04g081640 gene from *S. pimpinellifolium* and *S. lycopersicum*, two reporter cassettes to be introduced in each genetic background were constructed.

Both constructs contain a promoter region for modulating the expression of a fluorescent reporter protein (DsRed). The assembly steps required for each final cassette were the same and are schematized in Figure 26a.

The resulting plasmids for *Agrobacterium*-mediated transformation of tomato species to evaluate the activity of the promoter regions of the Solyc04g081640 gene in *S. lycopersicum* and *S. pimpinellifolium* are represented in Figure 26b and Figure 26c, respectively.

The resulting final cassettes are the plasmids *pDGB3\_alpha1::SLSolyc04g081640.1Prom:DsRed:Tnos::Pnos:NptII:Tnos* and *pDGB3\_alpha1::SPSolyc04g081640.1Prom:DsRed:Tnos::Pnos:NptII:Tnos* for the expression of the *S. lycopersicum* and *S. pimpinellifolium* promoters, respectively



## **DISCUSSION**



## **CHAPTER I - PHENOTYPING AND NATURAL VARIATION OF THE LOCUS IN CHROMOSOME 4 CAUSING A SD DISORDER IN TOMATO**

### **1. Effect of gametic and zygotic factors on SD in reciprocal crossing populations and on seed set of self-pollination crosses from the three SD haplotypes.**

Fakhet (2016) analysed reciprocal crosses, between SP4-4 PP, SP4-4 H and “MoneyMaker”, suggesting that the SD distortion of chromosome 4 was due to a sex-independent mechanism and a post-gametic or zygotic selection favoring PP haplotypes in a 3:1 proportion. Even though the gametic factors could not be completely discarded, these results fit, in part, with the proposed mechanism for a single locus mutation model for explaining sex independent the SD observed on chromosome 4 by Rick (1966 and 1971). He hypothesized the existence of a tomato gamete eliminator locus (*Ge*) with incomplete penetrance that has three alleles, *Ge<sup>e</sup>*, *Ge<sup>s</sup>*, and *Ge<sup>r</sup>*. The proposed mechanism implies that on the *Ge<sup>e</sup>/Ge<sup>e</sup>* heterozygote 90% of the *Ge<sup>e</sup>* gametes were eliminated independently of the sex, hence favoring the transmission of the *Ge<sup>e</sup>* allele in a 3:1 proportion in the hybrids. However, this mechanism alone could not explain the 3:1 segregation (PP:H) observed in the F<sub>2</sub> populations (Fakhet 2016), indicating that some other factors must be involved in the SD of chromosome 4 under study. The model proposed by Fakhet 2016 suggests that both gamete elimination and an abortion of 100% of MM zygotes and 50% of MP zygotes could explain the deviation of the F<sub>2</sub> segregation from the expected 9:3:1 ratio.

In the study presented here, we re-analyzed the segregation distortion in the progeny derived from the ♀ MM x ♂ IL 4-4 H cross, adding more individuals to the progeny analyzed by Fakhet (2016). In final experiment, the SD region showed a Mendelian inheritance of 1:1 (MM:H), indicating that the SD is sex-dependent, and giving further support to a post-zygotic mechanism. In addition, our results indicate that the progeny of the SP 4-4 H presented a reduction in seed set about 50% respect to SP 4-4 P and MM (i.e. both homozygotes for the SD locus). This significant reduction in seed set in plants heterozygous for the SD locus suggest a post-zygotic mechanism related with seed abortion.

Given to the fact that cytoplasmic inheritance is always maternal, the appearance of Mendelian inheritance only in the case where the female gamete is the recurrent parent “MoneyMaker” suggests the involvement of female gametophytic or cytoplasmic factors in the SD (Reflinur *et al.*, 2014). Gametophytic mutations are those affecting female gamete development after the meiosis, i.e. during the haploid phase of the life cycle. It can affect the Polygonum-type female gametophyte formation in tomato (megagametogenesis) and it can have a negative

influence on pre- and post-zygotic functioning of the mature gamete during fecundation, through its influence in pollen tube guidance controlled by the embryo-sac (Shimizu and Okada, 2000), fertilization itself (Fu et al., 2000) or seed development controlled by maternal factors (Yadegari et al., 2004). In the gametophytic maternal effect class of mutants, the phenotype is apparent only at the post-fertilization stage (Grossniklaus et al., 1998, Brukhin et al., 2005) Also, it is described that fruits of heterozygous female gametophyte mutants exhibit reduced seed set because 50% of the female gametophytes are mutant and nonfunctional (Yadegari et al., 2004, Drews 2011).

To conclude, Rick's (1966 and 1971) sex-independent hypothesis suited for the most part to the SD of chromosome 4 in tomato but there are clues for an involvement of the female gametophyte associated factors that explain the SD observed.

The hypothesis presented here assumes a single locus involved in the SD phenomena however, the Dobzhansky-Muller model, states that hybrid incompatibilities are caused by the interactions between genes that have functionally diverged in each population and assumes the existence of two or more interacting loci (Koide *et al.*, 2008). This is the case for example of the *Segregation distorter (SD)* of *Drosophila* (Temin *et al.*, 1991). SD systems often involve alleles at a minimum of two closely linked genes, a segregation distorter, and a *cis*-acting target (Koide *et al.*, 2008 and 2012). The current SD locus still harbors several genes, so the previous possibility cannot be discarded

In summary, the currently investigated SD is due to post-zygotic mechanisms, probably by a cytoplasmic factor from *S. pimpinellifolium* present in the female gametophyte that induces seed abortion when "Moneymaker" alleles are present in the zygote. The nature of this factor still needs to be elucidated.

## **2. Analysis of the natural variation in the SD region.**

The implementation of the KASP analysis technique for the haplotyping of the *S. pimpinellifolium* and *S. lycopersicum var. cerasiforme* in the genomic SD region enabled a more accurate discrimination of the SNPs that these accessions present.

Unfortunately, due to time constrains, the progeny of the selected accession cross with the recurrent parent was not phenotyped for SD on time. The most interesting progeny to analyse correspond to the LA2184 accession, with recombinant events within the SD locus that may help to narrow down the region to 16Kb, that includes only the two HSP genes that showed the most significant changes in the expression analysis.

Interestingly, in a RIL of *S. pimpinellifolium* LA2093 in the genetic background of *S. lycopersicum* (NCEBR-1) no SD was reported in chromosome 4 (Kinkade and Foolad, 2013). This could imply that none of the Group 1 accession progeny would present SD, but this hypothesis must be verified.

On an overall perspective of the KASP markers, if the segregation distorter locus is located and characterized, a SD marker can be developed for genotyping of the SD allele shortening the time and cost of identifying wild accessions that when crosses with the elite cultivars present the SD phenomenon.

## **CHAPTER II – GENOMIC AND TRANSCRIPTOMIC ANALYSIS OF THE CANDIDATE GENES.**

### **1. Expression analysis of the candidate genes by qRT-PCR**

The qRT-PCR analysis enabled the relative quantification of the candidate genes expression in ovary/fruit samples at different stages that were obtained from different materials presenting or not SD. The relationship between the spatio-temporal pattern of transcription of the candidate genes also gave hints about how each stage of fruit setting is related to the SD phenomenon.

The two genes that presented significant differential results in the expression analysis were Solyc04g081630 and Solyc04g081640. Solyc04g081630 presented an enhanced expression profile that was similar in the two SP 4-4 samples but contrasted from that of “Moneymaker” samples which, in turn, was consistent with previously published tomato RNAse-based data that were re-analysed here by *in silico* expression analysis (Annex 7).

In “Moneymaker” the anther sample has a high fold increase compared to the corresponding ovary sample on anthesis flowers. The *in silico* expression analysis also shows that this gene is expressed in different plant tissues at different stages of development.

The high expression of the gene in the SP 4-4 tissues did not allow for this gene to be rolled out as a strong candidate like the previous ones.

Focussing on the Solyc04g081640 gene, the *in silico* expression analysis showed a localized and specific expression of this gene in flower and fruit tissues, more specifically in developing seeds, which by itself is a good indicator of the involvement of the gene in the SD phenomenon. Also, the sequencing of the genomic sequence showed a larger number of SNPs and InDels compared to the other candidates and the cDNA sequencing revealed a possible different peptide. But the best indicator is the RNA expression profile. The “Moneymaker” control samples follow the expected pattern of expression thru time consistent with the reported *in silico* expression analysis. The increase of expression in SP 4-4 from the 4<sup>th</sup> to 7<sup>th</sup> DPA, indicates HSP90 genes are expressed during the 4–16 Cell Stage Embryo and Globular Stage Embryo. These stages are characterized by cell division and cell elongation, respectively, and point for an important role in early seed development. Also, since HSP90s have been reported to contribute to stabilization and folding of secreted peptides as those used for cell to cell communication, this could be related to the observed SD by an important role for this protein on pollen tube guidance or embryo-sac rupture upon fecundation (Shimizu and Okada, 2000).

Given that these two genes are our strongest candidates for being responsible for the SD phenomenon they were selected as target genes for the CRISPR/Cas9 knockouts. The knockout of each one of the genes independently in SP4-4 populations could give a deeper insight into the involvement of those genes as the edited genes will be non-functional and therefore would not present SD and support that the higher expression in SP4-4 compared to “Moneymaker” is the cause of the SD.

## **2. Candidate gene sequencing, structural prediction and SNP effects.**

Based on the results for the expression of the candidate genes some illations can be made but most important, the predicted protein structure analysis can give insights about gene function and how they can be related to the SD phenomenon.

The two genes that show the largest differential gene expression changes between the materials of our study encode Heat-shock proteins (HSPs). HSPs act as molecular chaperones involved in protein folding, assembly, translocation and degradation in basal cellular metabolism, and stabilization of other proteins and membranes against heat stress (Wang *et al.*, 2004). These proteins, are molecular chaperones that may or not may be related to thermotolerance and their functions still a subject of active research (Ohama *et al.*, 2017). Although more than one classification of the HSPs exists, one that considers their molecular weight, amino acid sequence homologies and functions was proposed by Gupta *et al.* (2010). In this classification, HSPs are grouped in 5 families: Hsp100, Hsp90, Hsp70, Hsp60, and small heat-shock (sHSP). The nomenclature of the HSPs comes from their molecular weight, i.e., Hsp90 refers to a HSP with a molecular weight in the range of 90 kDa. The high diversity of these proteins derives in part from the sessile nature of plants and from their need for adaptation to heat stress conditions (Al-Wahaibi *et al.*, 2011).

The molecular function of the Hsp90 diverges from the heat stress tolerance more than other HSPs. Hsp90 is abundant in non-heat stressed cells and accounts for almost 1% of total cytosolic proteins (Picard *et al.*, 2002). Hsp90 are expressed in high quantities in non-stressed cells and the increase in stress is just a few fold, so, the observed high fold increase must be explained by other factors. Also, Hsp90 acts more as a stabilizer, regulator (activation/inactivation of other proteins), and molecular chaperon in non-stress conditions in addition to also working in the response to heat and lesion stress.

A homolog gene from *Arabidopsis thaliana* that has high similarity to the two HSP90 genes is the At4g24190. This HSP90-7 gene that encodes an Endoplasmic reticulum chaperone in the *Arabidopsis* genome, and is related to a molecular chaperone role for the processing of cellular secretions. It is required for shoot apical meristem, root apical meristem and floral meristem formation, probably by regulating the folding of CLAVATA proteins (CLVs) (Ishiguro et al., 2002). And, as most of the HSP90 proteins, it is also involved for resistance to ER stresses (Chong et al., 2015). In fact, in the previous works regarding the SD in chromosome 4 of tomato, Fakhret (2016) and Barrantes (2014) reported the homology of the HSP90 to the *A. thaliana* SHEPHERD (SHD) gene involved in pollen tube elongation and meristem proliferation by its action on folding and stabilization of CLV secreted proteins (Ishiguro et al., 2002) hinting for the involvement in fecundation and plant development.

A similar case is presented in the rice hybrid sterility locus *S5*. It encodes three linked ORFs: ORF3+, ORF5+, and ORF4+. ORF3+, a HSP70 homolog, protects from the action of the female gamete killer ORF5+ and the partner ORF4+. ORF3- cannot prevent the action of ORF5+ and ORF5+ and the gametophyte is killed which distorts the segregation by preferential transmission of ORF3+ gametes between *indica* and *japonica* rice hybrids (Yang et al., 2012). One hypothesis is that in these HSP90 a similar mechanism like the *S5* locus could occur where SP4-4 alleles correspond to the ORF3+ alleles and “Moneymaker” alleles to the ORF3-.

A compatible additional hypothesis is that the polymorphisms detected in the Solyc04g081640 coding sequence could also decrease the function of this HSP90 by loss-of-function or by a different ORF being translated. Rutherford and Lindquist (1998) reported that when HSP90 function is impaired in *Drosophila* almost all morphological structures suffer phenotypic variation. These variations were highly influenced by the remaining genetic background. Since HSP90 is related to protein stabilizing and signal transduction, and is involved in several developmental pathways, it controls (“buffers”) the detection of new morphological variants produced by genetic determinants that were previously silent. This existing genomic variation that is not manifested until atypical environmental conditions or following the introduction of novel alleles is what is called cryptic genetic variation (CGV) (Pires et al., 2016).

In *Arabidopsis*, a HSP90 buffering effect on phenotypic variation was also reported (Queitsch et al., 2002) and that the HSP90 buffering effect on genetic variation is common (Sangster et al., 2008). This buffering effect of HSP90 also suggests that it could be related to the evolutionary

process since it lets the accumulation of polymorphisms, i.e. CGV without phenotypical manifestation (Queitsch *et al.*, 2002).

This could mean that although there is an increase in expression, the decrease in functionality of HSP90 may favour the SD by influence in the remaining genetic background or linked genes contained in the SD region, HSPs included, like it was described in *Drosophila* (Phadnis and Orr, 2009).

Recently, Pires *et al.* (2016), showed that the buffer effect in plants can be a sex-dependent mechanism. They hypothesized that in *Arabidopsis* there is a pool of CGV in the paternal regulation of seed development that is buffered by the maternal genome. The paternal variation is only manifested when there is a maternal loss-of-function of the *MEA* gene. Although not directly related to this mechanism, the loss-of function of the HSP90 gene in the SP4-4 could explain the fact that when the female gametophyte has “Moneymaker” alleles there is no SD caused by buffer effect but when the female gametophyte has SP alleles a new segregation appears with incomplete penetrance. This incomplete penetrance on Mendelian traits related of HSP90 buffer effect was also reported in *Dario rerio* (Yeyati *et al.*, 2007). One simplistic explanation is that loss or reduction of HSP90 buffering in ovules containing *S. pimpinellifolium* alleles would mean female gametophytes “prefer” pollen also with the same alleles by some factors contained in the genetic background of “Moneymaker”, explaining why there is no selection when the female gamete is from “Moneymaker” in the reciprocal crosses.

Other functional HSP90 proteins exist in the tomato genome. The impairing of these two would unlikely influence the SD as loss of HSP90 buffering in other loci. Other genes in the *S. lycopersicum* genome characterized as Hsp90 genes are: Solyc06g036290, identified as a Hsp90 gene in chromosome 6. Solyc12g015880 (Hsp90-1) on chromosome 12, that is required in the Mi-1 gene mediated resistance against pathogens and pests, and a Hsp90-2 gene, identified as Solyc07g065840 in chromosome 7. In addition to these three genes, there are 3 genetic sequences related to Hsp90 proteins in what concerns the protein domains. Two of them are chaperone proteins htpG, Solyc05g010670 and Solyc07g047790, and the Solyc10g078930 related to the activation of Heat-shock protein ATPase homolog 1. To conclude, we have shown by both sequence and expression data that the most likely genes related to SD are the two HSP90 and a hypothetical mechanism is described.

Although the other genes in the SD region do not show relevant data for being related to the SD, the GO associated to them can be related to fruit and seed development and fecundation for these reasons some considerations can be done.

There are 2 gene sequences annotated as Cyclin encoding genes, Solyc04g081650 and Solyc04g081660. More precisely, Cyclins of the B subtype. The cyclins from the CYCB1 and CYCB2 classes regulate the progression of the cell cycle from G<sub>2</sub> phase to the mitotic phase (G<sub>2</sub>/M). (Ito, 2014) The GO search of the *Arabidopsis* homolog, At1g76310 gene, that encodes to a Cyclin-B2-4 protein testifies for the relationship of this cyclins in *Solanum* being related to cell division.

Unlike animals, plants have a simpler and specific mechanism for cell cycle-regulated promoter activation of B-type cyclin genes. The Mitosis-specific activator (MSA) motif (Ito, 1998) is a cis-acting element that co-regulates G<sub>2</sub>/M-specific genes activating the phase specific promoters of type B cyclins and other kinesis-like proteins involved in the G<sub>2</sub> and M phases. (Ito, 2014) This means that if the expression data showed an increase of transcription of these genes it would be simultaneous, which is not the case. Additionally, the aminoacidic substitutions do not seem to affect protein folding. There is no apparent alterations of these genes between haplotypes that can be related to the SD mechanisms.

Other gene that did not showed relevant expression profiles in reproductive structures of the SP4-4 is the Solyc04g081670 gene encoding a Vacuolar processing enzyme. VPEs are related to programmed cell death and this protein three-dimensional structure is tightly related to animal caspase 1 (Hatsugai *et al.*, 2015; Hara-Nishimura *et al.*, 2005) The rupture of the vacuole triggered by VPE initiates the proteolytic cascade that provokes programmed cell death (PCD) by cell destruction. (Hatsugai *et al.*, 2015; Hara-Nishimura *et al.*, 2011) The close relation between VPE and the beginning of PCD by modulating other vacuole enzymes is shown by different stimuli that cause apoptosis and development cell death. (Hatsugai *et al.*, 2015) VPE showed to be involved in programmed cell death (PCD) occurred by vacuolar disruption in hybrid seedlings derived from interspecific crosses in the genus *Nicotiana*. (Mino *et al.*, 2007) and PCD is related to embryo-sac disruption upon fecundation (Fu *et al.*, 2000). Since in our case there were not polymorphisms affecting protein sequence or folding nor relevant expression differences for this gene it is unlikely to be the solo cause of the SD.



## CHAPTER III – DEVELOPMENT OF BIOTECHNOLOGY TOOLS FOR SD FUNCTIONAL ANALYSIS.

### 1. CRISPR/Cas9 cassettes for the silencing of candidate genes.

The typical GB 3.0 assembly of a functional cassette of CRISPR/Cas9 to produce a knockout in plants must include a: PolIII promoter, in this case U6-26 Promoter of *A. thaliana*, the desired target sequence; a scaffold that contains the non-targeted part of the nucleotide sequence of the CRISPR/Cas9 complex, a functional Cas9 transcription unit that includes a constitutive promoter, like 35S, a CDS of Cas9, in this case hCas9 (human Cas9), and a terminator (Tnos). Also, to function as a selection gene in plant tissue, a TU for the neomycin phosphotransferase II gene (NptII) is added to the final construct.

Since the process by which the plant tissues will be transformed is the agro-infiltration of tomato leaves with *Agrobacterium tumefaciens*, both constructs contain a Left-border (LB) and a Right-border (RB) making each final cassette a T-DNA in a binary plasmid system.

To do the assembly two different pipelines were used depending of the nature of the designed targets. The construct *pDGB3\_alpha1::Tnos:NptII:Pnos-SF:U6-26:tRNA-gRNA40.1.1:tRNA-gRNA40.1.2::35s:hCas9:Tnos* targeted 2 sequences of the Solyc04g081640.1 gene and the guide RNAs were assembled in a polycistronic TU.

In dicot, RNA polymerase III-dependent U6 promoter requires a G nucleotide at the 5' end of the RNA to start transcription. So, the gRNAs must begin with a guanine nucleotide (G) but, in the case of the target design for the Solyc04g081640.1 gene, it was impossible to design gRNAs that begin with a G and also follow the three characteristics described before. In this case, the gRNAs were designed without the Guanine nucleotide in the 5' position but were assembled with an extra target gRNA in a polycistronic sequence that contains a tRNA before each of the gRNAs to be assembled by GB system (Vazquez-Vilar *et al.*, 2016). The tRNA is eliminated by cell machinery after transcription. In the polycistronic assembly, the step of creating the gRNA for the targets is considered the level -1. Before adding the tRNA and scaffold part to the GB targets forming a level 0 construct.

The construction of single target constructs did not need a domestication step since the heterodimer already had the proper overhangs for the multipartite alpha level assemble. That means it could be directly assembled to the level 1 being the target sequence (heterodimer) the level 0.

The polycistronic construct of the level 1 required the assemble of both GB oligomers with a PolIII U6-26 promoter (GB1001) in an alpha level destination vector (pDGB3\_alpha1). The single target construct (*pDGB3\_alpha1::Tnos:NptII:Pnos-SF:U6-26:gRNA30.1:sgRNA::35s:hcas9:Tnos*) was performed in a similar way but in this case the PolIII U6-26 promoter, the target and the sgRNA (GB0645) were assembled in the same destination vector. After this assemble step both resulting cassettes in the alpha level destination vectors are transcription units of the guide RNAs for the CRISPR/Cas9 tool.

The next step of assembly involved a binary assemble of the guide RNA TU with a human Cas9 encoding TU in an omega level destination vector. Each one of these cassettes, in theory, can express a functional Cas9 with a guide RNA for the knockout of the targeted sequences. For the means of selecting the transformed plants tissues carrying the cassette with a CRISPR/Cas9 cassette, a transcription unit for the neomycin phosphotransferase II gene (NptII), that confers resistance to several antibiotics including, but not limited by, kanamycin and neomycin, was added to the final cassette.

Since the constructs were correctly assembled, and to verify the efficiency of this constructs in the knockout of the targeted genes, they should now be tested in a transient expression experience in *S. lycopersicum* SP 4-4. This experiment will begin by transforming *Agrobacterium tumefaciens*, containing the Helper plasmid (plasmid with the virulence genes of *A. tumefaciens*), and agro-infiltrating tomato leaves. After a few weeks, the genome sequences of the target gene in transformed cells can be sequenced to verify the existence of InDels that produce the knockout.

## Conclusions

- 1) The segregation distortion in the progeny of the ♀ *S. lycopersicum* var. "Moneymaker" x ♂ *S. lycopersicum* SP4-4 H cross revealed an unexpected Mendelian inheritance ratio of 1:1, instead of a 3:1 distorted segregation initially observed, suggesting the involvement of female gametophytic or cytoplasmatic factors in the mechanism of segregation distortion in chromosome 4 of tomato.
- 2) Progeny of self-pollinated plants *S. lycopersicum* SP4-4 heterozygotic for the SD region haplotype has a reduction of seed set of about 50% suggesting SD can be caused by post-zygotic defects.
- 3) More than one factor may contribute to the SD phenomenon. 1. influence of the female gamete haplotype on the post-meiotic gametogenesis or fecundation; 2. a possible cytoplasmic effect; 3. a possible sex-specific pleiotropic effect if the phenomenon is controlled by a single gene.
- 4) KASP analysis successfully discriminated SD haplotypes among *S. lycopersicum* var. "Moneymaker", *S. pimpinellifolium* and *S. lycopersicum* var. *cerasiforme* accessions.
- 5) Different haplotypes in the SD region occur among *S. pimpinellifolium* and *S. lycopersicum* var. *cerasiforme* and cultivated tomato accessions.
- 6) Expression analysis of genes within the SD region revealed an up regulation of expression of two HSP90 genes, Solyc04g081630 and Solyc04g081640, associated to the *S. lycopersicum* SP4-4, either in homo or heterozygosis in the early stages of fruit/embryo development.
- 7) Resequencing of the genomic and transcriptomic sequences of genes within the SD region revealed that Solyc04g081640 produced a transcript that includes the ITAG predicted intron between the first and second exon.
- 8) Solyc04g081640 gene sequences from a number of *S. pimpinellifolium* and *S. lycopersicum* var. *cerasiforme* accession revealed 3 different combination of 2 polymorphisms inside coding sequence.

9) GoldenBraid assembly system was successfully used for development of 5 novel vectors for *Agrobacterium*-mediated transformation of tomato to be applied in: 1. gene knockout using CRISPR/Cas9 system; 2. Homologous over-expression of *S. pimpinellifolium* Solyc04g081640 gene; 3. evaluation and tissue-specific expression of Solyc04g081640 gene promoters of *Solanum pimpinellifolium* and *Solanum lycopersicum*.

## Future work

Several approaches can be used in the continuation of this work to determine exactly what is the gene or genes that provoke the SD.

The sequencing of the totality of the SD region and the improvement of the mapping should give information about other variations outside the sequenced genes and the exact recombination points of the region.

In short term, the application of the homologous over-expression GB 3.0 cassette developed in this work should give insight about the function of the Heat Shock Protein 90 genes. Also, the promoter cassettes of the *S. pimpinellifolium* and *S. lycopersicum* Solyc04g081640 promoter could give clues not only about the tissue specific expression of the genes but also if the differences in promoter sequence are responsible for the increase of expression observed in plants with the *S. pimpinellifolium* haplotype for the SD region.

Since the sample collection times for expression analysis were based in the assumption of previous works that the SD phenomenon was related to post-gametic factors the landmarks of gametogenesis were not considered for analysis. However, given the latest results more gametophyte develop landmarks should be chosen for a similar analysis in the future. This could also be complemented by using other tissues, has complement, to infer in there is also an increase in other plant structures and not only in fruits and flower tissues.

In what concerns protein function, to confirm that the increase in transcription of the Solyc04g081640 gene is accompanied by an increase of the cytoplasmic protein is to use specific antibodies for the predicted aminoacidic sequence and western blot or use one of the tools available in the GB 3.0 toolkit and create fusion proteins with fluorescence (e.g. HSP90-GFP) (Sarrion-Perdigones *et al.*, 2013; Tominaga-Wada *et al.*, 2017), or even a combination of the two methods (Tillmann *et al.*, 2015)

The study of the transcription factors involved (Pattison *et al.*, 2015; Zhang *et al.*, 2016) in the SD phenomenon and gene interaction of the responsible gene(s) with other pathways could also help to understand the mechanism of SD.

In macroscale, histological analysis of the female gametophyte and pollen will help determine if some morphological abnormalities could influence the rate of pollination related with hybrid dysfunction. (Kubo *et al.*, 2017)

Technical difficulties hindered the cytological analysis of pollen tube growth and should be reworked to determine the absence of pollen malformations that could be the reason for skewed segregation observed. Koide *et al.*, 2012

## References

- 100 Tomato Genome Sequencing Consortium et al., 2014. Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *The Plant journal: for cell and molecular biology*, 80(1), p.136–148. Available at: <https://doi.org/10.1111/tpj.12616>.
- Akiyama, H. et al., 2009. A Screening Method for the Detection of the 35S Promoter and the Nopaline Synthase Terminator in Genetically Modified Organisms in a Real-Time Multiplex Polymerase Chain Reaction Using High-Resolution Melting-Curve Analysis. *Biological and Pharmaceutical Bulletin*, 32(11), pp.1824–1829.
- Alba, J.M., Montserrat, M. & Fernández-Muñoz, R., 2009. Resistance to the two-spotted spider mite (*Tetranychus urticae*) by acylsucroses of wild tomato (*Solanum pimpinellifolium*) trichomes studied in a recombinant inbred line population. *Experimental and Applied Acarology*, 47(1), pp.35–47. Available at: <https://doi.org/10.1007/s10493-008-9192-4>.
- Al-Whaibi, M.H., 2011. Plant heat-shock proteins: A mini review. *Journal of King Saud University - Science*, 23(2), pp.139–150. Available at: <http://www.sciencedirect.com/science/article/pii/S101836471000087X>.
- Anders, C. et al., 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, 513(7519), pp.569–573. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4176945/>.
- Andolfo, G. et al., 2013. Overview of tomato (*Solanum lycopersicum*) candidate pathogen recognition genes reveals important *Solanum* R locus dynamics. *New Phytologist*, 197(1), pp.223–237. Available at: <http://dx.doi.org/10.1111/j.1469-8137.2012.04380.x>.
- Andrianantoandro, E. et al., 2006. Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology*, 2(1). Available at: <http://msb.embopress.org/content/2/1/2006.0028.abstract>.
- Anton, T. et al., 2014. Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus*, 5(2), pp.163–172. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4049922/>.
- Baenziger, P.S. et al., 2017. Plant breeding and genetics: a paper in the series on The Need for Agricultural Innovation to Sustainably Feed the World by 2050. *Issue Paper - Council for Agricultural Science and Technology*, (No.57), p.24 pp.
- Bai, Y. & Lindhout, P., 2007. Domestication and Breeding of Tomatoes: What have We Gained and What Can We Gain in the Future? *Annals of Botany*, 100(5), pp.1085–1094. Available at: <http://aob.oxfordjournals.org/content/100/5/1085.abstract>.

- Baltes, N.J. & Voytas, D.F., 2015. Enabling plant synthetic biology through genome engineering. *Trends in Biotechnology*, 33(2), pp.120–131. Available at: <http://www.sciencedirect.com/science/article/pii/S0167779914002376>.
- Barrantes, W., 2014. *Desarrollo de una genoteca de líneas de introgresión entre Solanum lycopersicum y Solanum pimpinellifolium utilizando herramientas genómicas de alto rendimiento y detección de QTL simplicados en calidad de fruto*. Universidad de Costa Rica.
- Barrantes, W. et al., 2014. Highly efficient genomics-assisted development of a library of introgression lines of *Solanum pimpinellifolium*. *Molecular Breeding*, 34(4), pp.1817–1831. Available at: <http://dx.doi.org/10.1007/s11032-014-0141-0>.
- Bauchet, G. & Causse, M., 2012. *Genetic Diversity in Tomato (Solanum lycopersicum) and Its Wild Relatives*, *Genetic Diversity in Plants* P. M. Caliskan, ed., InTech. Available at: <https://www.intechopen.com/books/genetic-diversity-in-plants/genetic-diversity-in-tomato-solanum-lycopersicum-and-its-wild-relatives>.
- Bedinger, P.A. et al., 2011. Interspecific reproductive barriers in the tomato clade: opportunities to decipher mechanisms of reproductive isolation. *Sexual Plant Reproduction*, 24(3), pp.171–187. Available at: <https://doi.org/10.1007/s00497-010-0155-7>.
- Bernacchi, D. & Tanksley, S.D., 1997. An Interspecific Backcross of *Lycopersicon Esculentum* X *L. Hirsutum*: Linkage Analysis and a Qtl Study of Sexual Compatibility Factors and Floral Traits. *Genetics*, 147(2), pp.861–877. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1208205/>.
- Bortesi, L. & Fischer, R., 2015. The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Advances*, 33(1), pp.41–52. Available at: <http://www.sciencedirect.com/science/article/pii/S0734975014001931>.
- Brooks, C. et al., 2014. Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated9 System. *Plant Physiology*, 166(3), p.1292 LP-1297. Available at: <http://www.plantphysiol.org/content/166/3/1292.abstract>.
- Brukhin, V. et al., 2005. The RPN1 Subunit of the 26S Proteasome in Arabidopsis Is Essential for Embryogenesis. *The Plant Cell*, 17(10), pp.2723–2737. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1242268/>.
- Brukhin, V. et al., 2003. Flower development schedule in tomato *Lycopersicon esculentum* cv. sweet cherry. *Sexual Plant Reproduction*, 15(6), pp.311–320. Available at: <https://doi.org/10.1007/s00497-003-0167-7>.
- Buzgo, M. et al., 2017. Towards a comprehensive integration of morphological and genetic studies of floral development. *Trends in Plant Science*, 9(4), pp.164–173. Available at: <http://dx.doi.org/10.1016/j.tplants.2004.02.003>.



- Cameron, D.R. & Moav, R.M., 1957. Inheritance in *Nicotiana Tabacum* Xxvii. Pollen Killer, an Alien Genetic Locus Inducing Abortion of Microspores Not Carrying It. *Genetics*, 42(3), pp.326–335. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1209835/>.
- Carmel-Goren, L. et al., 2003. The SELF-PRUNING gene family in tomato. *Plant Molecular Biology*, 52(6), pp.1215–1222. Available at: <https://doi.org/10.1023/B:PLAN.0000004333.96451.11>.
- Castro, P. et al., 2011. A segregation distortion locus located on linkage group 4 of the chickpea genetic map. *Euphytica*, 179(3), pp.515–523. Available at: <https://doi.org/10.1007/s10681-011-0356-7>.
- Chevalier, E., Loubert-Hudon, A. & Matton, D.P., 2013. ScRALF3, a secreted RALF-like peptide involved in cell–cell communication between the sporophyte and the female gametophyte in a solanaceous species. *The Plant Journal*, 73(6), pp.1019–1033. Available at: <http://dx.doi.org/10.1111/tbj.12096>.
- Chong, L.P. et al., 2015. A highly charged region in the middle domain of plant endoplasmic reticulum (ER)-localized heat-shock protein 90 is required for resistance to tunicamycin or high calcium-induced ER stresses. *Journal of Experimental Botany*, 66(1), pp.113–124. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4265155/>.
- Cong, L. et al., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (New York, N.Y.)*, 339(6121), pp.819–823. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3795411/>.
- Covey, P.A. et al., 2010. Multiple features that distinguish unilateral incongruity and self-incompatibility in the tomato clade. *The Plant Journal*, 64(3), pp.367–378. Available at: <http://dx.doi.org/10.1111/j.1365-3113X.2010.04340.x>.
- Dai, B. et al., 2016. Identification and Characterization of Segregation Distortion Loci on Cotton Chromosome 18. *Frontiers in Plant Science*, 7, p.2037. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5242213/>.
- deVicente, M.C. & Tanksley, S.D., 1993. Qtl Analysis of Transgressive Segregation in an Interspecific Tomato Cross. *Genetics*, 134(2), pp.585–596. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1205500/>.
- Doganlar, S., Frary, A. & Tanksley, S.D., 2000. The genetic basis of seed-weight variation: tomato as a model system. *Theoretical and Applied Genetics*, 100(8), pp.1267–1273. Available at: <https://doi.org/10.1007/s001220051433>.
- Doyle, J., 1991. DNA Protocols for Plants BT - Molecular Techniques in Taxonomy. In G. M. Hewitt, A. W. B. Johnston, & J. P. W. Young, eds. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 283–293. Available at: [https://doi.org/10.1007/978-3-642-83962-7\\_18](https://doi.org/10.1007/978-3-642-83962-7_18).

- Drews, G.N. & Koltunow, A.M.G., 2011. The Female Gametophyte. *The Arabidopsis Book / American Society of Plant Biologists*, 9, p.e0155. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3268550/>.
- Endo, T.R., 2007. The gametocidal chromosome as a tool for chromosome manipulation in wheat. *Chromosome Research*, 15(1), pp.67–75. Available at: <https://doi.org/10.1007/s10577-006-1100-3>.
- Endo, T., 1990. Gametocidal chromosomes and their induction of chromosome mutations in wheat. *遺伝学雑誌*, 65(3), pp.135–152.
- Endy, D., 2005. Foundations for engineering biology. *Nature*, 438(7067), pp.449–453. Available at: <http://dx.doi.org/10.1038/nature04342>.
- Engler, C. et al., 2009. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type II Restriction Enzymes. J. Peccoud, ed. *PLoS ONE*, 4(5), p.e5553. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2677662/>.
- Engler, C., Kandzia, R. & Marillonnet, S., 2008. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE*, 3(11), p.e3647. Available at: <https://doi.org/10.1371/journal.pone.0003647>.
- Engler, C. & Marillonnet, S., 2011. Generation of Families of Construct Variants Using Golden Gate Shuffling BT - cDNA Libraries: Methods and Applications. In C. Lu, J. Browse, & J. G. Wallis, eds. Totowa, NJ: Humana Press, pp. 167–181. Available at: [https://doi.org/10.1007/978-1-61779-065-2\\_11](https://doi.org/10.1007/978-1-61779-065-2_11).
- Engler, C. & Marillonnet, S., 2013. Combinatorial DNA Assembly Using Golden Gate Cloning BT - Synthetic Biology. In K. M. Polizzi & C. Kontoravdi, eds. Totowa, NJ: Humana Press, pp. 141–156. Available at: [https://doi.org/10.1007/978-1-62703-625-2\\_12](https://doi.org/10.1007/978-1-62703-625-2_12).
- Eshed, Y. & Zamir, D., 1995. An Introgression Line Population of *Lycopersicon Pennellii* in the Cultivated Tomato Enables the Identification and Fine Mapping of Yield-Associated Qtl. *Genetics*, 141(3), pp.1147–1162. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1206837/>.
- Fakhet, D., 2016. *High resolution mapping of a locus in chromosome 4 causing distorted segregation in tomato*. Instituto Agronómico Mediterráneo de Zaragoza (IAMZ).
- Fan, D. et al., 2015. Efficient CRISPR/Cas9-mediated Targeted Mutagenesis in *Populus* in the First Generation. *Scientific Reports*, 5, p.12217. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4507398/>.
- Finch, R.A., Miller, T.E. & Bennett, M.D., 1984. “Cuckoo” Aegilops addition chromosome in wheat ensures its transmission by causing chromosome breaks in meiospores lacking it. *Chromosoma*, 90(1), pp.84–88. Available at: <https://doi.org/10.1007/BF00352282>.

- Frary, A. & Doganlar, S., 2003. Comparative Genetics of Crop Plant Domestication and Evolution. *Turkish Journal of Agriculture and Forestry*, 27, pp.59–69. Available at: citeulike-article-id:742395.
- Fu, Y. et al., 2000. Changes in actin organization in the living egg apparatus of *Torenia fournieri* during fertilization. *Sexual Plant Reproduction*, 12(6), pp.315–322. Available at: <https://doi.org/10.1007/s004970000026>.
- Fulton, T.M. et al., 1997. QTL analysis of an advanced backcross of *Lycopersicon peruvianum* to the cultivated tomato and comparisons with QTLs found in other wild species. *Theoretical and Applied Genetics*, 95(5), pp.881–894. Available at: <https://doi.org/10.1007/s001220050639>.
- Ganopoulos, I., Argiriou, A. & Tsaftaris, A., 2011. Microsatellite high resolution melting (SSR-HRM) analysis for authenticity testing of protected designation of origin (PDO) sweet cherry products. *Food Control*, 22(3), pp.532–541. Available at: <http://www.sciencedirect.com/science/article/pii/S0956713510003324>.
- Gibson, D.G. et al., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Meth*, 6(5), pp.343–345. Available at: <http://dx.doi.org/10.1038/nmeth.1318>.
- Gilbert, L.A. et al., 2013. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*, 154(2), pp.442–451. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3770145/>.
- Goldman, I.L., Paran, I. & Zamir, D., 1995. Quantitative trait locus analysis of a recombinant inbred line population derived from a *Lycopersicon esculentum* x *Lycopersicon cheesmanii* cross. *Theoretical and Applied Genetics*, 90(7), pp.925–932. Available at: <https://doi.org/10.1007/BF00222905>.
- González-Aguilera, K.L. et al., 2016. Selection of Reference Genes for Quantitative Real-Time RT-PCR Studies in Tomato Fruit of the Genotype MT-Rg1. *Frontiers in Plant Science*, 7, p.1386. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5021083/>.
- Grandillo, S., Ku, H.M. & Tanksley, S.D., 1999. Identifying the loci responsible for natural variation in fruit size and shape in tomato. *Theoretical and Applied Genetics*, 99(6), pp.978–987. Available at: <https://doi.org/10.1007/s001220051405>.
- Grandillo, S. & Tanksley, S.D., 1996. QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. *Theoretical and Applied Genetics*, 92(8), pp.935–951. Available at: <https://doi.org/10.1007/BF00224033>.
- Grandillo, S., 2014. Introgression Libraries with Wild Relatives of Crops BT - Genomics of Plant Genetic Resources: Volume 2. Crop productivity, food security and nutritional quality. In R. Tuberosa, A. Graner, & E. Frison, eds. Dordrecht: Springer Netherlands, pp. 87–122. Available at: [https://doi.org/10.1007/978-94-007-7575-6\\_4](https://doi.org/10.1007/978-94-007-7575-6_4).

- Grossniklaus, U. et al., 1998. Maternal Control of Embryogenesis by MEDEA, a Polycomb Group Gene in Arabidopsis. *Science*, 280(5362), p.446 LP-450. Available at: <http://science.sciencemag.org/content/280/5362/446.abstract>.
- Gruis, D., Schulze, J. & Jung, R., 2004. Storage Protein Accumulation in the Absence of the Vacuolar Processing Enzyme Family of Cysteine Proteases. *The Plant Cell*, 16(1), pp.270–290. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC301410/>.
- Gruis, D. et al., 2002. Redundant Proteolytic Mechanisms Process Seed Storage Proteins in the Absence of Seed-Type Members of the Vacuolar Processing Enzyme Family of Cysteine Proteases. *The Plant Cell*, 14(11), pp.2863–2882. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC152733/>.
- Gupta, S.C. et al., 2010. Heat shock proteins in toxicology: How close and how far? *Life Sciences*, 86(11), pp.377–384. Available at: <http://www.sciencedirect.com/science/article/pii/S0024320510000044>.
- Hamilton, J.P. et al., 2012. Single Nucleotide Polymorphism Discovery in Cultivated Tomato via Sequencing by Synthesis. *The Plant Genome*, 5, pp.17–29. Available at: <http://dx.doi.org/10.3835/plantgenome2011.12.0033>.
- Han, Y., Khu, D.-M. & Monteros, M.J., 2012. High-resolution melting analysis for SNP genotyping and mapping in tetraploid alfalfa (*Medicago sativa* L.). *Molecular Breeding*, 29(2), pp.489–501. Available at: <https://doi.org/10.1007/s11032-011-9566-x>.
- Hara-Nishimura, I. & Hatsugai, N., 2011. The role of vacuole in plant cell death. *Cell Death and Differentiation*, 18(8), pp.1298–1304. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3172105/>.
- Hara-Nishimura, I. et al., 2005. Vacuolar processing enzyme: an executor of plant cell death. *Current Opinion in Plant Biology*, 8(4), pp.404–408. Available at: <http://www.sciencedirect.com/science/article/pii/S1369526605000774>.
- Haseloff, J. & Ajioka, J., 2009. Synthetic biology: history, challenges and prospects. *Journal of the Royal Society Interface*, 6(Suppl 4), pp.S389–S391. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2843964/>.
- Hatsugai, N. et al., 2015. Vacuolar processing enzyme in plant programmed cell death. *Frontiers in Plant Science*, 6, p.234. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4390986/>.
- He, C., Holme, J. & Anthony, J., 2014. SNP Genotyping: The KASP Assay BT - Crop Breeding: Methods and Protocols. In D. Fleury & R. Whitford, eds. New York, NY: Springer New York, pp. 75–86. Available at: [https://doi.org/10.1007/978-1-4939-0446-4\\_7](https://doi.org/10.1007/978-1-4939-0446-4_7).
- Hellens, R.P. et al., 2000. pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Molecular Biology*, 42(6), pp.819–832. Available at: <https://doi.org/10.1023/A:1006496308160>.

- Herrmann, M.G. et al., 2006. Amplicon DNA Melting Analysis for Mutation Scanning and Genotyping: Cross-Platform Comparison of Instruments and Dyes. *Clinical Chemistry*, 52(3), p.494 LP-503. Available at: <http://clinchem.aaccjnls.org/content/52/3/494.abstract>.
- Hirakawa, H. et al., 2013. Genome-Wide SNP Genotyping to Infer the Effects on Gene Functions in Tomato. *DNA Research*, 20(3), pp.221–233. Available at: <http://dx.doi.org/10.1093/dnares/dst005>.
- Hiremath, P.J. et al., 2012. Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnology Journal*, 10(6), pp.716–732. Available at: <http://dx.doi.org/10.1111/j.1467-7652.2012.00710.x>.
- Hsu, P.D., Lander, E.S. & Zhang, F., 2014. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, 157(6), pp.1262–1278. Available at: <http://www.sciencedirect.com/science/article/pii/S0092867414006047>.
- Huq, MA, Akter, S, Nou, IS, Kim, HT, Jung, YJ, Kang, K., 2016. Identification of functional SNPs in genes and their effects on plant phenotypes. *Journal of Plant Biotechnology*, 43(1), pp.1–11. Available at: <http://www.kspbtjpb.org/journalDOIx.php?id=10.5010/JPB.2016.43.1.1>.
- Hurst, L.D. & Pomiankowski, A., 1991. Causes of Sex Ratio Bias May Account for Unisexual Sterility in Hybrids: A New Explanation of Haldane's Rule and Related Phenomena. *Genetics*, 128(4), pp.841–858. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1204557/>.
- Ishiguro, S. et al., 2002. SHEPHERD is the Arabidopsis GRP94 responsible for the formation of functional CLAVATA proteins. *The EMBO Journal*, 21(5), pp.898–908. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC125899/>.
- Ito, M., 2014. Expression of mitotic cyclins in higher plants: transcriptional and proteolytic regulation. *Plant Biotechnology Reports*, 8(1), pp.9–16. Available at: <https://doi.org/10.1007/s11816-013-0297-9>.
- Ito, M. et al., 1998. A Novel  $\text{cis}$ -Acting Element in Promoters of Plant B-Type Cyclin Genes Activates M Phase-Specific Transcription. *The Plant Cell*, 10(3), p.331 LP-341. Available at: <http://www.plantcell.org/content/10/3/331.abstract>.
- Jia, H. & Wang, N., 2014. Targeted Genome Editing of Sweet Orange Using Cas9/sgRNA. *PLOS ONE*, 9(4), p.e93806. Available at: <https://doi.org/10.1371/journal.pone.0093806>.
- Jiang, W. et al., 2013. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Research*, 41(20), pp.e188–e188. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3814374/>.
- Jinek, M. et al., 2012. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(6096), p.816 LP-821. Available at: <http://science.sciencemag.org/content/337/6096/816.abstract>.

- Jinek, M. et al., 2014. Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science (New York, N.Y.)*, 343(6176), p.1247997. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4184034/>.
- Karimi, M., Inzé, D. & Depicker, A., 2017. GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends in Plant Science*, 7(5), pp.193–195. Available at: [http://dx.doi.org/10.1016/S1360-1385\(02\)02251-3](http://dx.doi.org/10.1016/S1360-1385(02)02251-3).
- Khalil, A.S. & Collins, J.J., 2010. Synthetic Biology: Applications Come of Age. *Nature reviews. Genetics*, 11(5), pp.367–379. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2896386/>.
- Khera, P. et al., 2013. Single Nucleotide Polymorphism–based Genetic Diversity in the Reference Set of Peanut (*Arachis spp.*) by Developing and Applying Cost-Effective Kompetitive Allele Specific Polymerase Chain Reaction Genotyping Assays. *The Plant Genome*, 6. Available at: <http://dx.doi.org/10.3835/plantgenome2013.06.0019>.
- Kim, J.-E. et al., 2014. Genome-Wide SNP Calling Using Next Generation Sequencing Data in Tomato. *Molecules and Cells*, 37(1), pp.36–42. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3907006/>.
- Kinkade, M.P. & Foolad, M.R., 2013. Validation and fine mapping of lyc12.1, a QTL for increased tomato fruit lycopene content. *Theoretical and Applied Genetics*, 126(8), pp.2163–2175. Available at: <https://doi.org/10.1007/s00122-013-2126-5>.
- Knapp, S. & Peralta, I.E., 2016. The Tomato (*Solanum lycopersicum* L., Solanaceae) and Its Botanical Relatives BT - The Tomato Genome. In M. Causse et al., eds. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 7–21. Available at: [https://doi.org/10.1007/978-3-662-53389-5\\_2](https://doi.org/10.1007/978-3-662-53389-5_2).
- Knight, T., 2003. Idempotent Vector Design for Standard Assembly of Biobricks. *MIT Artificial Intelligence Laboratory; MIT Synthetic Biology Working Group*. Available at: <http://hdl.handle.net/1721.1/21168>.
- Koenig, D. et al., 2013. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proceedings of the National Academy of Sciences*, 110(28), pp.E2655–E2662. Available at: <http://www.pnas.org/content/110/28/E2655.abstract>.
- Koide, Y. et al., 2012. Complex genetic nature of sex-independent transmission ratio distortion in Asian rice species: the involvement of unlinked modifiers and sex-specific mechanisms. *Heredity*, 108(3), pp.242–247. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3282388/>.
- Koide, Y. et al., 2008. The Evolution of Sex-Independent Transmission Ratio Distortion Involving Multiple Allelic Interactions at a Single Locus in Rice. *Genetics*, 180(1), pp.409–420. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2535691/>.

- Kondo, K. et al., 2002. Cultivated tomato has defects in both S-RNase and HT genes required for stylar function of self-incompatibility. *The Plant Journal*, 29(5), pp.627–636. Available at: <http://dx.doi.org/10.1046/j.0960-7412.2001.01245.x>.
- Kubo, T. et al., 2017. Two Tightly Linked Genes at the *hsa1* Locus Cause Both F<sub>1</sub> and F<sub>2</sub> Hybrid Sterility in Rice. *Molecular Plant*, 9(2), pp.221–232. Available at: <http://dx.doi.org/10.1016/j.molp.2015.09.014>.
- Ky, C.L. et al., 2000. Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). *Theoretical and Applied Genetics*, 101(4), pp.669–676. Available at: <https://doi.org/10.1007/s001220051529>.
- Lehmensiek, A., Sutherland, M.W. & McNamara, R.B., 2008. The use of high resolution melting (HRM) to map single nucleotide polymorphism markers linked to a covered smut resistance gene in barley. *Theoretical and Applied Genetics*, 117(5), pp.721–728. Available at: <https://doi.org/10.1007/s00122-008-0813-4>.
- Li, R. et al., 2017. Multiplexed CRISPR/Cas9-mediated metabolic engineering of  $\gamma$ -aminobutyric acid levels in *Solanum lycopersicum*. *Plant Biotechnology Journal*, p.n/a-n/a. Available at: <http://dx.doi.org/10.1111/pbi.12781>.
- Li, W. & Chetelat, R.T., 2015. Unilateral incompatibility gene *ui1.1* encodes an S-locus F-box protein expressed in pollen of *Solanum* species. *Proceedings of the National Academy of Sciences*, 112(14), pp.4417–4422. Available at: <http://www.pnas.org/content/112/14/4417.abstract>.
- Liew, M. et al., 2004. Genotyping of Single-Nucleotide Polymorphisms by High-Resolution Melting of Small Amplicons. *Clinical Chemistry*, 50(7), p.1156 LP-1164. Available at: <http://clinchem.aaccjnls.org/content/50/7/1156.abstract>.
- Lima-Silva, V. et al., 2012. Genetic and genome-wide transcriptomic analyses identify co-regulation of oxidative response and hormone transcript abundance with vitamin C content in tomato fruit. *BMC Genomics*, 13, p.187. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3462723/>.
- Lin, S.Y. et al., 1992. Segregation distortion via male gametes in hybrids between Indica and Japonica or wide-compatibility varieties of rice (*Oryza sativa* L). *Theoretical and Applied Genetics*, 84(7), pp.812–818. Available at: <https://doi.org/10.1007/BF00227389>.
- Lin, T. et al., 2014. Genomic analyses provide insights into the history of tomato breeding. *Nat Genet*, 46(11), pp.1220–1226. Available at: <http://dx.doi.org/10.1038/ng.3117>.
- Livak, K.J. & Schmittgen, T.D., 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. *Methods*, 25(4), pp.402–408. Available at: <http://www.sciencedirect.com/science/article/pii/S1046202301912629>.

- Lochlainn, S.Ó. et al., 2011. High Resolution Melt (HRM) analysis is an efficient tool to genotype EMS mutants in complex crop genomes. *Plant Methods*, 7(1), p.43. Available at: <https://doi.org/10.1186/1746-4811-7-43>.
- Loefering, W.Q. & Sears, E.R., 1963. DISTORTED INHERITANCE OF STEM-RUST RESISTANCE OF TIMSTEIN WHEAT CAUSED BY A POLLEN-KILLING GENE. *Canadian Journal of Genetics and Cytology*, 5(1), pp.65–72. Available at: <https://doi.org/10.1139/g63-010>.
- Longley, A.E., 1945. Abnormal Segregation during Megasporogenesis in Maize. *Genetics*, 30(1), pp.100–113. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1209271/>.
- Lorieux, M. et al., 1995a. Maximum-likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross populations. *Theoretical and Applied Genetics*, 90(1), pp.73–80. Available at: <https://doi.org/10.1007/BF00220998>.
- Lorieux, M. et al., 1995b. Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F2 populations. *Theoretical and Applied Genetics*, 90(1), pp.81–89.
- Lu, H., Romero-Severson, J. & Bernardo, R., 2002. Chromosomal regions associated with segregation distortion in maize. *Theoretical and Applied Genetics*, 105(4), pp.622–628. Available at: <https://doi.org/10.1007/s00122-002-0970-9>.
- Lyttle, T.W., 1993. Cheaters sometimes prosper: distortion of mendelian segregation by meiotic drive. *Trends in Genetics*, 9(6), pp.205–210. Available at: <http://www.sciencedirect.com/science/article/pii/0168952593901207>.
- Ma, X. et al., 2017. A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants. *Molecular Plant*, 8(8), pp.1274–1284. Available at: <http://dx.doi.org/10.1016/j.molp.2015.04.007>.
- Mackay, J.F., Wright, C.D. & Bonfiglioli, R.G., 2008. A new approach to varietal identification in plants by microsatellite high resolution melting analysis: application to the verification of grapevine and olive cultivars. *Plant Methods*, 4(1), p.8. Available at: <https://doi.org/10.1186/1746-4811-4-8>.
- Makarova, K.S. et al., 2011. Evolution and classification of the CRISPR–Cas systems. *Nat Rev Micro*, 9(6), pp.467–477. Available at: <http://dx.doi.org/10.1038/nrmicro2577>.
- Mangelsdorf, P.C. & Jones, D.F., 1926. The Expression of Mendelian Factors in the Gametophyte of Maize. *Genetics*, 11(5), pp.423–455. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1200910/>.
- Mao, Y. et al., 2013. Application of the CRISPR–Cas System for Efficient Genome Engineering in Plants. *Molecular Plant*, 6(6), pp.2008–2011. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3916745/>.



- Meissner, R. et al., 1997. A new model system for tomato genetics. *The Plant Journal*, 12(6), pp.1465–1472. Available at: <http://dx.doi.org/10.1046/j.1365-313x.1997.12061465.x>.
- Mino, M. et al., 2007. Cell death in seedlings of the interspecific hybrid of *Nicotiana glauca* and *N. glauca*; possible role of knob-like bodies formed on tonoplast in vacuolar-collapse-mediated cell death. *Plant Cell Reports*, 26(4), pp.407–419. Available at: <https://doi.org/10.1007/s00299-006-0261-z>.
- Monforte, A.J. & Tanksley, S.D., 2000. Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: a tool for gene mapping and gene discovery. *Genome*, v. 43.
- Mueller, L.A. et al., 2005. The Tomato Sequencing Project, the first cornerstone of the International Solanaceae Project (SOL). *Comparative and Functional Genomics*, 6(3), pp.153–158. Available at: <http://dx.doi.org/10.1002/cfg.468>.
- Nakagahara, M., 1972. Genetic Mechanism on the Distorted Segregation of Marker Genes belonging to the Eleventh Linkage Group in Cultivated Rice. *Japanese Journal of Breeding*, 22(4), pp.232–238.
- Nakaune, S. et al., 2005. A Vacuolar Processing Enzyme,  $\delta$ VPE, Is Involved in Seed Coat Formation at the Early Stage of Seed Development. *The Plant Cell*, 17(3), pp.876–887. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1069705/>.
- Neelam, K., Brown-Guedira, G. & Huang, L., 2013. Development and validation of a breeder-friendly KASPar marker for wheat leaf rust resistance locus *Lr21*. *Molecular Breeding*, 31(1), pp.233–237. Available at: <https://doi.org/10.1007/s11032-012-9773-0>.
- Nesbitt, T.C. & Tanksley, S.D., 2002. Comparative Sequencing in the Genus *Lycopersicon*: Implications for the Evolution of Fruit Size in the Domestication of Cultivated Tomatoes. *Genetics*, 162(1), p.365 LP-379. Available at: <http://www.genetics.org/content/162/1/365.abstract>.
- Ng, M. & Yanofsky, M.F., 2001. Function and evolution of the plant MADS-box gene family. *Nat Rev Genet*, 2(3), pp.186–195. Available at: <http://dx.doi.org/10.1038/35056041>.
- Nishimasu, H. et al., 2014. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell*, 156(5), pp.935–949. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4139937/>.
- Ohama, N. et al., 2017. Transcriptional Regulatory Network of Plant Heat Stress Response. *Trends in Plant Science*, 22(1), pp.53–65. Available at: <http://www.sciencedirect.com/science/article/pii/S1360138516301261>.
- Okonechnikov, K., Golosova, O. & Fursov, M., 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, 28(8), pp.1166–1167. Available at: <http://dx.doi.org/10.1093/bioinformatics/bts091>.

- Pan, C. et al., 2016. CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations. *Scientific Reports*, 6, p.24765. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4838866/>.
- Patron, N.J. et al., 2015. Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytologist*, 208(1), pp.13–19. Available at: <http://dx.doi.org/10.1111/nph.13532>.
- Pattison Richard, J. et al., 2015. Comprehensive Tissue-specific Transcriptome Analysis Reveals Distinct Regulatory Programs During Early Tomato Fruit Development. *Plant Physiology*. Available at: <http://www.plantphysiol.org/content/early/2015/06/22/pp.15.00287>.
- Pease, J.B. et al., 2016. Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. *PLOS Biology*, 14(2), p.e1002379. Available at: <https://doi.org/10.1371/journal.pbio.1002379>.
- Peralta, I.E., Knapp, S. & Spooner, D.M., 2008. *Taxonomy of wild tomatoes and their relatives (Solanum sect. Lycopersicoides, sect. Juglandifolia, sect. Lycopersicon; Solanaceae)*, [Ann. Arbor, Mich.]: Amer. Society of Plant Taxonomists.
- Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Research*, 29(9), pp.e45–e45. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC55695/>.
- Phadnis, N. & Orr, H.A., 2009. A single gene causes both male sterility and segregation distortion in Drosophila hybrids. *Science (New York, N.Y.)*, 323(5912), pp.376–379. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2628965/>.
- Piatek, A. et al., 2015. RNA-guided transcriptional regulation in planta via synthetic dCas9-based transcription factors. *Plant Biotechnology Journal*, 13(4), pp.578–589. Available at: <http://dx.doi.org/10.1111/pbi.12284>.
- Picard, D., 2002. Heat-shock protein 90, a chaperone for folding and regulation. *Cellular and Molecular Life Sciences CMLS*, 59(10), pp.1640–1648. Available at: <https://doi.org/10.1007/PL00012491>.
- Pingoud, A. & Jeltsch, A., 2001. Structure and function of type II restriction endonucleases. *Nucleic Acids Research*, 29(18), pp.3705–3727. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC55916/>.
- Pires, N.D. et al., 2016. Quantitative Genetics Identifies Cryptic Genetic Variation Involved in the Paternal Regulation of Seed Development. *PLOS Genetics*, 12(1), p.e1005806. Available at: <https://doi.org/10.1371/journal.pgen.1005806>.
- Queitsch, C., Sangster, T.A. & Lindquist, S., 2002. Hsp90 as a capacitor of phenotypic variation. *Nature*, 417(6889), pp.618–624. Available at: <http://dx.doi.org/10.1038/nature749>.

- Rambla, J.L. et al., 2014. The expanded tomato fruit volatile landscape. *Journal of Experimental Botany*, 65(16), pp.4613–4623. Available at: <http://dx.doi.org/10.1093/jxb/eru128>.
- Ran, F.A. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, 8(11), pp.2281–2308. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3969860/>.
- Ranc, N. et al., 2012. Genome-Wide Association Mapping in Tomato (*Solanum lycopersicum*) Is Possible Using Genome Admixture of *Solanum lycopersicum* var. *cerasiforme*. *G3: Genes/Genomes/Genetics*, 2(8), p.853 LP-864. Available at: <http://www.g3journal.org/content/2/8/853.abstract>.
- Reflinur et al., 2014. Analysis of segregation distortion and its relationship to hybrid barriers in rice. *Rice*, 7, p.3. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4884001/>.
- Rick, C.M., 1963. BARRIERS TO INTERBREEDING IN LYCOPERSICON PERUVIANUM. *Evolution*, 17(2), pp.216–232. Available at: <http://dx.doi.org/10.1111/j.1558-5646.1963.tb03272.x>.
- Rick, C.M., 1966. Abortion of Male and Female Gametes in the Tomato Determined by Allelic Interaction. *Genetics*, 53(1), pp.85–96. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1211010/>.
- Rick, C.M., 1971. The Tomato Ge Locus: Linkage Relations and Geographic Distribution of Alleles. *Genetics*, 67(1), pp.75–85. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1212536/>.
- Rick, C.M., Holle, M. & Thorp, R.W., 1978. Rates of cross-pollination in *Lycopersicon pimpinellifolium*: Impact of genetic variation in floral characters. *Plant Systematics and Evolution*, 129(1), pp.31–44. Available at: <https://doi.org/10.1007/BF00988982>.
- Ririe, K.M., Rasmussen, R.P. & Wittwer, C.T., 1997. Product Differentiation by Analysis of DNA Melting Curves during the Polymerase Chain Reaction. *Analytical Biochemistry*, 245(2), pp.154–160. Available at: <http://www.sciencedirect.com/science/article/pii/S0003269796999169>.
- Rodríguez-Leal, D. et al., 2017. Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. *Cell*, 171(2), p.470–480.e8. Available at: <http://www.sciencedirect.com/science/article/pii/S0092867417309881>.
- Roeder, G.S., 1997. Meiotic chromosomes: it takes two to tango. *Genes & Development*, 11(20), pp.2600–2621. Available at: <http://genesdev.cshlp.org/content/11/20/2600.short>.
- Ron, M. et al., 2014. Hairy Root Transformation Using *Agrobacterium rhizogenes* as a Tool for Exploring Cell Type-Specific Gene Expression and Function Using Tomato as a Model. *Plant Physiology*, 166(2), pp.455–469. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4213079/>.

- Rousseaux, M.C. et al., 2005. QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines. *Theoretical and Applied Genetics*, 111(7), pp.1396–1408. Available at: <https://doi.org/10.1007/s00122-005-0071-7>.
- Rutherford, S.L. & Lindquist, S., 1998. Hsp90 as a capacitor for morphological evolution. *Nature*, 396, pp.336–342. Available at: <http://dx.doi.org/10.1038/24550>.
- Sandler, L., Hiraizumi, Y. & Sandler, I., 1959. Meiotic Drive in Natural Populations of *Drosophila Melanogaster*. I. the Cytogenetic Basis of Segregation-Distortion. *Genetics*, 44(2), pp.233–250. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1209945/>.
- Sangster, T.A. et al., 2008. HSP90-buffered genetic variation is common in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 105(8), pp.2969–2974. Available at: <http://www.pnas.org/content/105/8/2969.abstract>.
- Sano, Y., 1983. A new gene controlling sterility in F1 hybrids of two cultivated rice speciesIts association with photoperiod sensitivity. *Journal of Heredity*, 74(6), pp.435–439. Available at: <http://dx.doi.org/10.1093/oxfordjournals.jhered.a109832>.
- Sano, Y., Chu, Y.-E. & Oka, H.-I., 1979. GENETIC STUDIES OF SPECIATION IN CULTIVATED RICE. *The Japanese journal of genetics*, 54(2), pp.121–132.
- Sapranaukas, R. et al., 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Research*, 39(21), pp.9275–9282. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3241640/>.
- Sarrion-Perdigones, A. et al., 2011. GoldenBraid: An Iterative Cloning System for Standardized Assembly of Reusable Genetic Modules. *PLOS ONE*, 6(7), p.e21622. Available at: <https://doi.org/10.1371/journal.pone.0021622>.
- Sarrion-Perdigones, A. et al., 2014. Design and Construction of Multigenic Constructs for Plant Biotechnology Using the GoldenBraid Cloning Strategy BT - DNA Cloning and Assembly Methods. In S. Valla & R. Lale, eds. Totowa, NJ: Humana Press, pp. 133–151. Available at: [https://doi.org/10.1007/978-1-62703-764-8\\_10](https://doi.org/10.1007/978-1-62703-764-8_10).
- Sarrion-Perdigones, A. et al., 2013. GoldenBraid2.0: A comprehensive DNA assembly framework for Plant Synthetic Biology. *Plant Physiology*. Available at: <http://www.plantphysiol.org/content/early/2013/05/13/pp.113.217661.abstract>.
- Schauer, N. et al., 2006. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat Biotech*, 24(4), pp.447–454. Available at: <http://dx.doi.org/10.1038/nbt1192>.
- Semagn, K. et al., 2014. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Molecular Breeding*, 33(1), pp.1–14. Available at: <https://doi.org/10.1007/s11032-013-9917-x>.

- Shah, S.A. et al., 2013. Protospacer recognition motifs: Mixed identities and functional diversity. *RNA Biology*, 10(5), pp.891–899. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3737346/>.
- Shalem, O., Sanjana, N.E. & Zhang, F., 2015. High-throughput functional genomics using CRISPR-Cas9. *Nature reviews. Genetics*, 16(5), pp.299–311. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4503232/>.
- Shetty, R.P., Endy, D. & Knight, T.F., 2008. Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering*, 2, p.5. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2373286/>.
- Shimizu, K.K. & Okada, K., 2000. Attractive and repulsive interactions between female and male gametophytes in Arabidopsis pollen tube guidance. *Development*, 127(20), p.4511 LP-4518. Available at: <http://dev.biologists.org/content/127/20/4511.abstract>.
- Silver, L.M., 2017. The peculiar journey of a selfish chromosome: mouse haplotypes and meiotic drive. *Trends in Genetics*, 9(7), pp.250–254. Available at: [http://dx.doi.org/10.1016/0168-9525\(93\)90090-5](http://dx.doi.org/10.1016/0168-9525(93)90090-5).
- Sim, S.-C. et al., 2012. Development of a Large SNP Genotyping Array and Generation of High-Density Genetic Maps in Tomato T. Yin, ed. *PLoS ONE*, 7(7), p.e40563. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3393668/>.
- Sternberg, S.H. et al., 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490), pp.62–67. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4106473/>.
- Sugano, S.S. et al., 2014. CRISPR/Cas9-Mediated Targeted Mutagenesis in the Liverwort *Marchantia polymorpha* L. *Plant and Cell Physiology*, 55(3), pp.475–481. Available at: <http://dx.doi.org/10.1093/pcp/pcu014>.
- Sun, X. et al., 2015. Targeted mutagenesis in soybean using the CRISPR-Cas9 system. *Scientific Reports*, 5, p.10342. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4448504/>.
- Tanksley, S.D. et al., 1992. High Density Molecular Linkage Maps of the Tomato and Potato Genomes. *Genetics*, 132(4), pp.1141–1160. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1205235/>.
- Tanksley, S.D. et al., 1996. Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. *Theoretical and Applied Genetics*, 92(2), pp.213–224. Available at: <https://doi.org/10.1007/BF00223378>.
- Tanksley, S.D., 2004. The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *The Plant Cell*, 16(Suppl), pp.S181–S189. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2643388/>.

- Tanksley, S.D. & McCouch, S.R., 1997. Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. *Science*, 277(5329), p.1063 LP-1066. Available at: <http://science.sciencemag.org/content/277/5329/1063.abstract>.
- Taylor, D.R. & Ingvarsson, P.K., 2003. Common Features of Segregation Distortion in Plants and Animals. *Genetica*, 117(1), pp.27–35. Available at: <https://doi.org/10.1023/A:1022308414864>.
- Temin, R.G., 1991. The Independent Distorting Ability of the Enhancer of Segregation Distortion, E(sd), in *Drosophila Melanogaster*. *Genetics*, 128(2), pp.339–356. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1204472/>.
- Thomson, M.J., 2014. High-Throughput SNP Genotyping to Accelerate Crop Improvement. *Plant Breeding and Biotechnology*, 2(3), pp.195–212. Available at: <http://www.plantbreedbio.org/journalDOIx.php?id=10.9787/PBB.2014.2.3.195>.
- Tillmann, B. et al., 2015. Hsp90 Is Involved in the Regulation of Cytosolic Precursor Protein Abundance in Tomato. *Molecular Plant*, 8(2), pp.228–241. Available at: <http://www.sciencedirect.com/science/article/pii/S1674205214000203>.
- Tomato Genome Consortium, 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, 485(7400), p.635–641. Available at: <http://europepmc.org/articles/PMC3378239>.
- Tominaga-Wada, R. et al., 2017. Expression and protein localization analyses of *Arabidopsis GLABRA3 (GL3)* in tomato (*Solanum lycopersicum*) root epidermis. *Plant Biotechnology*, 34(2), pp.115–117.
- Tovar-Méndez, A. et al., 2014. Restoring pistil-side self-incompatibility factors recapitulates an interspecific reproductive barrier between tomato species. *The Plant Journal*, 77(5), pp.727–736. Available at: <http://dx.doi.org/10.1111/tpj.12424>.
- Tovar-Méndez, A., Lu, L. & McClure, B., 2017. HT proteins contribute to S-RNase-independent pollen rejection in *Solanum*. *The Plant Journal*, 89(4), pp.718–729. Available at: <http://dx.doi.org/10.1111/tpj.13416>.
- van der Knaap, E., Lippman, Z.B. & Tanksley, S.D., 2002. Extremely elongated tomato fruit controlled by four quantitative trait loci with epistatic interactions. *Theoretical and Applied Genetics*, 104(2), pp.241–247. Available at: <https://doi.org/10.1007/s00122-001-0776-1>.
- Vázquez-Vilar, M. et al., 2016. A modular toolbox for gRNA–Cas9 genome engineering in plants based on the GoldenBraid standard. *Plant Methods*, 12(1), p.10. Available at: <https://doi.org/10.1186/s13007-016-0101-2>.
- Vazquez-Vilar, M. et al., 2015. Software-Assisted Stacking of Gene Modules Using GoldenBraid 2.0 DNA-Assembly Framework BT - Plant Functional Genomics: Methods and Protocols. In J. M. Alonso & A. N. Stepanova, eds. New York, NY: Springer New York, pp. 399–420. Available at: [https://doi.org/10.1007/978-1-4939-2444-8\\_20](https://doi.org/10.1007/978-1-4939-2444-8_20).

- Viquez Zamora, A.M., 2015. *Exploiting wild relatives of S. lycopersicum for quality traits*. 328, ; Wageningen University. Available at: <http://edepot.wur.nl/353426>.
- Wang, C. et al., 2005. Mapping segregation distortion loci and quantitative trait loci for spikelet sterility in rice (*Oryza sativa* L.). *Genetical Research*, 86(2), pp.97–106. Available at: <https://www.cambridge.org/core/article/mapping-segregation-distortion-loci-and-quantitative-trait-loci-for-spikelet-sterility-in-rice-oryza-sativa-1/368A010AC836E03993AB1277712AEDEF>.
- Wang, W. et al., 2004. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science*, 9(5), pp.244–252. Available at: <http://www.sciencedirect.com/science/article/pii/S1360138504000603>.
- Weber, E. et al., 2011. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE*, 6(2), p.e16765. Available at: <https://doi.org/10.1371/journal.pone.0016765>.
- Wittwer, C.T., 2009. High-resolution DNA melting analysis: advancements and limitations. *Human Mutation*, 30(6), pp.857–859. Available at: <http://dx.doi.org/10.1002/humu.20951>.
- Wittwer, C.T. et al., 2003. High-Resolution Genotyping by Amplicon Melting Analysis Using LCGreen. *Clinical Chemistry*, 49(6), p.853 LP-860. Available at: <http://clinchem.aaccjnls.org/content/49/6/853.abstract>.
- Wojdacz, T.K. & Dobrovic, A., 2007. Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. *Nucleic Acids Research*, 35(6), pp.e41–e41. Available at: <http://dx.doi.org/10.1093/nar/gkm013>.
- Wood, A.J. et al., 2011. Targeted Genome Editing Across Species Using ZFNs and TALENs. *Science (New York, N.Y.)*, 333(6040), p.307. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3489282/>.
- Wu, X. et al., 2014. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotech*, 32(7), pp.670–676. Available at: <http://dx.doi.org/10.1038/nbt.2889>.
- Xiao, H. et al., 2009. Integration of tomato reproductive developmental landmarks and expression profiles, and the effect of SUN on fruit shape. , 21.
- Xiao, H. et al., 2009. Integration of tomato reproductive developmental landmarks and expression profiles, and the effect of SUN on fruit shape. *BMC Plant Biology*, 9, p.49. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2685393/>.
- Xu, P. et al., 2006. Computational Estimation and Experimental Verification of Off-Target Silencing during Posttranscriptional Gene Silencing in Plants. *Plant Physiology*, 142(2), pp.429–440. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1586062/>.

- Xu, S., 2008. Quantitative Trait Locus Mapping Can Benefit From Segregation Distortion. *Genetics*, 180(4), pp.2201–2208. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2600952/>.
- Xu, X. et al., 2013. Gametophytic and zygotic selection leads to segregation distortion through in vivo induction of a maternal haploid in maize. *Journal of Experimental Botany*, 64(4), pp.1083–1096. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3580820/>.
- Xu, Y. et al., 1997. Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). *Molecular and General Genetics MGG*, 253(5), pp.535–545. Available at: <https://doi.org/10.1007/s004380050355>.
- Yadegari, R. & Drews, G.N., 2004. Female Gametophyte Development. *The Plant Cell*, 16(suppl 1), p.S133 LP-S141. Available at: [http://www.plantcell.org/content/16/suppl\\_1/S133.abstract](http://www.plantcell.org/content/16/suppl_1/S133.abstract).
- Yamanaka, N. et al., 2001. An Informative Linkage Map of Soybean Reveals QTLs for Flowering Time, Leaflet Morphology and Regions of Segregation Distortion. *DNA Research*, 8(2), pp.61–72. Available at: <http://dx.doi.org/10.1093/dnares/8.2.61>.
- Yang, J. et al., 2012. A Killer-Protector System Regulates Both Hybrid Sterility and Segregation Distortion in Rice. *Science*, 337(6100), p.1336 LP-1340. Available at: <http://science.sciencemag.org/content/337/6100/1336.abstract>.
- Yates, H.E. et al., 2004. Comparative fine mapping of fruit quality QTLs on chromosome 4 introgressions derived from two wild tomato species. *Euphytica*, 135(3), pp.283–296. Available at: <https://doi.org/10.1023/B:EUPH.0000013314.04488.87>.
- Yeyati, P.L. et al., 2007. Hsp90 Selectively Modulates Phenotype in Vertebrate Development. *PLOS Genetics*, 3(3), p.e43. Available at: <https://doi.org/10.1371/journal.pgen.0030043>.
- Zhang, H. et al., 2014. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnology Journal*, 12(6), pp.797–807. Available at: <http://dx.doi.org/10.1111/pbi.12200>.
- Zhang, S. et al., 2016. Spatiotemporal transcriptome provides insights into early fruit development of tomato (*Solanum lycopersicum*). , 6, p.23173. Available at: <http://dx.doi.org/10.1038/srep23173>.
- Zhang, T. et al., 2014. Genome-Wide Analysis of the Cyclin Gene Family in Tomato. *International Journal of Molecular Sciences*, 15(1), pp.120–140. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3907801/>.



## **Supplementary Material**

**Annex 1** – Sequences and physical properties of the primers designed for qRT-PCR, sequencing and other amplification reactions.

**Annex 2** – Sequences and physical properties of the primer triplets for the KASP markers.

**Annex 3** –  $C_T$  analysis raw data calculations and plotted graphs of the qRT-PCR experiment.

**Annex 4** – List of GB elements used and created with the GB 3.0 toolkit for the assembly of the CRISPR/Cas9 cassettes and Expression cassettes.

**Annex 5** – Results of the HRM genotyping.

**Annex 6** – Results of the KASP analysis performed on *S. pimpinellifolium* and *S. lycopersicum* var *cerasiforme* samples.

**Annex 7** – Results of the *in silico* analysis of the candidate genes and respective orthologs in *Arabidopsis thaliana*.

**Annex 8** – Alignment of the Solyc04g081640 gene proximal promoter region from the *S. lycopersicum* var. “Moneymaker” (MM) and *S. lycopersicum* IL 4-4 PP.

**Annex 1 – Sequences and physical properties of the primers designed for qRT-PCR, sequencing and other amplification reactions.**

qRT-PCR primer list						
<b>Solyc04g081620 (+ strand)</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	5' Position in Sl chr4
qRT_Sl4g081620F	CTCAGAAGATATGCTGGTGATGT	23	58,4	43,5	111	65,590,274
qRT_Sl4g081620R	GGACGGATAGCACTCAATAAGG	22	60	50		65,590,384
<b>Solyc04g081630 (- strand)</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	5' Position in Sl chr4
qRT_Sl4g0818630E1F	GGTATGGACACACCAGAATCAA	22	59,7	45,5	100	65,598,910
qRT_Sl4g0818630E1R	TGAACTACAACCTGCCACAC	20	57,1	50		65,597,811
<b>Solyc04g081640 (- strand)</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	5' Position in Sl chr4
qRT_Sl4g0818640E1F	TCGACTGTTATGGGGTTTGTA	22	59,4	40,9	156	65,606,237
qRT_Sl4g0818640E2R	CATGATTTTTCTGCTAGTAAGTTCCA	26	60,1	34,6		65,606,038
qRT_SL4g402F	CAACGACAAACATATCCGAAGA	22	56,8	40,9	180	65,605,770
qRT_SL4g402R	ATCCCTTTGCCAAATTTAGTCC	22	56,9	40,9		65,604,589
<b>Solyc04g081650 (- strand)</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	5' Position in Sl chr4
qRT_Sl4g0818650E1F	CGGTCTCTTAAAGCTGCTCAGT	22	60,2	50	98	65,608,356
qRT_Sl4g0818650E2R	GGGAATCGAAGCATTTCACT	22	59,5	40,9		65,608,156
<b>Solyc04g081660 (- strand)</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	5' Position in Sl chr4
qRT_Sl4g0818660E1F	TAGGAGGGCACTAAGCAAAATC	22	59,8	45,5	153	65,609,934
qRT_Sl4g0818660E2R	CTTGTAACGGGACGGTGAAT	20	59,9	50		65,609,699
<b>Solyc04g081670 (- strand)</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	5' Position in Sl chr4
qRT_Sl4g0818670E1F	AACATGTCCTGATATACCAACAAA	25	58,8	32	187	65,612,715
qRT_Sl4g0818670E2R	TCCTTCGTAACATCCTCAACA	22	59,6	40,9		65,612,232
<b>Solyc04g081680 (- strand)</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	5' Position in Sl chr4
qRT_Sl4g0818680E1F	GGGAGAGAGCAGAATCTATTTCG	23	60,7	47,8	92	65,618,594
qRT_Sl4g0818680E1R	CCACCCAATCAAACCAATATC	22	60,3	40,9		65,618,503
<b>Solyc_CAC</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)		
Solyc_CACq-F	CCTCCGTTGTGATGTAAGTGG	21	55,5	47,8		
Solyc_CACq-R	ATTGGTGAAAGTAACATCATCG	23	53,3	39,1		

Sequencing primer list							
<b>Solyc04g081620 (+)</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	Overlap (bp)	5' Position in SI chr4
SI4g081862F	CAATCATGCTGCCTCCTACA	20	59,8	50	618		65,590,027
SI4g081862R	GAGGAAAACAAGGACGTGGA	20	60,1	50			65,590,644
<b>Solyc04g081630 (-)</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	Overlap (bp)	5' Position in SI chr4
SI4g08186301F	CGACACTCCTTGTGTTGTGG	20	60,2	55	679		65,596,305
SI4g08186301R	GGCAAGGCATAGGGACATAC	20	59,4	55			65,595,627
SI4g08186302F	GCCCTTATTGATCCCTAGCC	20	59,9	55	697	323	65,596,679
SI4g08186302R	AGCAACAACCCACTCTCGAT	20	59,7	50			65,595,983
SI4g08186303F	AAATCCAAAGGCGAGCTGAT	20	61,1	45	687	295	65,597,071
SI4g08186303R	GGTCTTCCTCGTGAATTTG	20	59,5	50			65,596,385
SI4g08186304F	CAGCTTTGTAGTTTTGTGGATGA	23	59,3	39,1	740	312	65,597,499
SI4g08186304R	TGGAAAGCCAAAATCCCATA	20	60,3	40			65,597,760
SI4g08186305F	ATCGAGACAAACGCAGGAAT	20	59,7	45	701	214	65,597,986
SI4g08186305R	TCTCCCAAATCTCACTTGC	20	60,2	50			65,597,286
SI4g08186306F	CTATATATTCGGCTCCGCAAC	22	60,9	50	486	120	65,598,279
SI4g08186306R	CAATTGGAGGTTTCCCCTTT	20	60,2	45			65,597,791
<b>Solyc04g081640 (-)</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	Overlap (bp)	5' Position in SI chr4
40_1_F	CGGCATACTGTACGACACATTT	22	59,9	45,5	630		65,606,401
40_1_R	TTCGGATATGTTTGTCTGTTGTC	22	59,9	40,9			65,605,772
40_2_F	AGCCTATGTTTCTCAGCCAAAA	22	60,3	40,9	770	249	65,606,020
40_2_R	TCTATACCCCAAATGGAAACCA	22	60,4	40,9			65,605,251
40_3_F	CAAGTCCTTATTGTCCGATTCC	22	59,8	45,5	680	200	65,605,450
40_3_R	CTGATGGCAGATAACACAATGA	22	58,7	40,9			65,604,771
40_4_F	CTTCGAGGTTAATCCTTAGTTG	24	57,8	41,7	660	252	65,605,022
40_4_R	CCTGAGGTATTTGTAACCAACCA	24	59,5	41,7			65,604,363
<b>Solyc04g081650 (-)</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	Overlap (bp)	5' Position in SI chr4
50_1_F	AACCTTCCACAGGAAAAGTTGA	22	60	40,9	609		65,608,412
50_1_R	CCTAGACAATCCTAAAACCATGC	23	59,1	43,5			65,607,804
50_2_F	GTGCGTCGCTTGATTTTCTAC	21	59,9	47,6	830	182	65,607,985
50_2_R	CACACACCATATTTGAGCCACA	22	61,8	45,5			65,607,156

<b>Solyc04g081660 (-)</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	Overlap (bp)	5' Position in SI chr4
60_1_F	GGAGTGATTAGGCCTCAAATC	22	59,1	45,5	617	244	65,610,107
60_1_R	TGCAGTCTTCTGATTCCTTGAA	22	60	40,9			65,609,491
60_2_F	GATCCCTCCTGTTCTGATTCAC	22	59,9	50	735	275	65,609,734
60_2_R	CCTGCACATGTGTGTGTACATT	22	59,4	45,5			65,609,000
60_3_F	CAGTAGTGGAATACACCGACGA	22	60,1	50	573		65,609,274
60_3_R	TTGTAGTGACCTTTGAATTGG	22	60	40,9			65,608,702
<b>Solyc04g081670 (-)</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	Overlap (bp)	5' Position in SI chr4
70_1_F	TGGAAGGAGAGGAATTCTAGATG	23	58,9	43,5	677	163	65,613,257
70_1_R	TCAAGTATTGTGTGGCCATGA	21	60	42,9			65,612,581
70_2_F	CTGTCTTGATGACAACACGATG	22	59,2	45,5	610	120	65,612,743
70_2_R	CGTCTAGCGTAATCCTGCTT	22	60,1	50			65,612,134
70_3_F	TGTTGAGGATGTTTACGAAGGA	22	59,6	40,9	716		65,612,253
70_3_R	CAAATTGTCCTCAGTCTGTGA	22	60,1	45,5			65,611,538
<b>Solyc04g081680 (-)</b>	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	Overlap (bp)	5' Position in SI chr4
80_1_F	ATTCAATTCAGGGAGTCATTGG	22	60,2	40,9	569	104	65,619,041
80_1_R	AAAGTCTGCACATCCATTTCT	22	60	40,9			65,618,473
80_2_F	TTTCGCAGTAGGTTTCAAGTCA	22	59,9	40,9	481	100	65,618,576
80_2_R	GTCTTCCTCTCTGTGCGTTCT	22	60,1	50			65,618,096
80_3_F	AGACTCAACCTTTTCGCACAGC	22	59,9	45,5	652	210	65,618,195
80_3_R	TCTCCTGTAAGCAGATCACATC	22	57	45,5			65,617,544
80_4_F	AAAGGCATACAGGAGCTCAAAG	22	59,9	45,5	473		65,617,753
80_4_R	CCAACCACCATCCTTCAAAT	20	59,6	45			65,617,281

Other primers used	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)	Amplicon length (bp)	Overlap (bp)	5' Position in SI chr4
cDNA seq R	CAACTTCCTCCGATTTGAGAGG	22	58,7	50	-	-	-
SI4g40 prom 1 F	TCCATTTTCAGACGTGAAGTGTT	22	59.6	40.9	776	132	65.606.182
SI4g40 prom 1 R	ATAGAACACACATGTCCGGTGA	22	60.3	45.5			65.606.936
SI4g40 prom 2 F	CATCTTCCAAATATTCCCCCTA	22	59.2	40.9	908		65.606.826
SI4g40 prom 2 R	AGAAACATAAGTGTGCCCAAG	22	60.4	45.5			65.607.712
SI4g40 Prom dom F	GCGCCGTCTCGCTCGGGAGAGAAACATAAGTGTGCCCAA	40	57.8	60.0	1531	-	65.607.733
SI4g40 Prom dom R	GCGCCGTCTCGCTCACATTTTTAACTTTGGGTTTTGTATGAGTC	45	56.8	44.4		65.606.241	
Solyc04g40 CDS F	GCGCCGTCTCGCTCGAATGTGCGACTGTTATGGGGTTT	37	57.3	56.8	1758	-	65.606.220
Solyc04g40 CDS R	GCGCCGTCTCGCTCAAAGCTTACCTCTCAAATCGGAGGAA	40	56.1	55.0		65.604.540	

**Annex 2 – Sequences and physical properties of the primer triplets for the KASP markers.**

SNP Marker	Primer	Seq 5' -> 3'	Tm (°C)	Chr04 region 5'-3' (SL2.5)
1 T->C	20 A FAM R	GAAGGTGACCAAGTTCATGCTGGTCATATGTGGATGGCAA <b>A</b>	55.77	65.590.486..65.590.505
	20 G HEX R	GAAGGTCGGAGTCAACGGATTGGTCATATGTGGATGGCAA <b>G</b>		
	20 COM F	CCTTCTTCATGTGCTGCTTT	56.1	65.590.435..65.590.454
4 C->T	HSP G FAM R	GAAGGTGACCAAGTTCATGCTCTTTTACCTTTGTACACTT <b>G</b>	49.91	65.597.173..65.597.192
	HSP A HEX R	GAAGGTCGGAGTCAACGGATTCTTTTACCTTTGTACACTT <b>A</b>		
	HSP COM F	CGATGTGAAC TAAGTTTGA	50.42	65.597.150..65.597.168
9 A->C	EndoP T FAM R	GAAGGTGACCAAGTTCATGCTTCTTGCCCTCCCAAAAACT <b>T</b>	55.64	65.605.455..65.605.474
	EndoP G HEX R	GAAGGTCGGAGTCAACGGATTCTTGCCCTCCCAAAAACT <b>G</b>		
	EndoP COM F	TCGGACAATAAGGACTTGAAAG	56.11	65.605.433..65.605.454
11 G->A	Cyc C FAM R	GAAGGTGACCAAGTTCATGCTTATGCATTTATGAGGCGGT <b>C</b>	55.69	65.608.352..65.608.371
	Cyc T HEX R	GAAGGTCGGAGTCAACGGATTTATGCATTTATGAGGCGGT <b>T</b>		
	Cyc COM F	CAGACTGAGCAGCTTTAAGA	55.11	65.608.332..65.608.351
2 A->G	5 A FAM F	GAAGGTGACCAAGTTCATGCTTCCCATTTACCACCTGGGTTTTT <b>A</b>	55.9	65.590.643..65.590.666
	5 G HEX F	GAAGGTCGGAGTCAACGGATTTCCCATTTACCACCTGGGTTTTT <b>G</b>		
	5 COM R	CAACTCTTTCAGCAGAGAATGA	52.58	65.590.691..65.590.712
3 C->T	6 C FAM F	GAAGGTGACCAAGTTCATGCTCCGAAGCTGTTTAGACTGC <b>C</b>	55.4	65.595.917..65.595.936
	6 T HEX F	GAAGGTCGGAGTCAACGGATTCCGAAGCTGTTTAGACTGC <b>T</b>		
	6 COM R	GGAGATTCGACTTTTCATCCG	53.74	65.595.952..65.595.973
5 A->G	8 T FAM R	GAAGGTGACCAAGTTCATGCTCAACGGTAGGTGATACACCATA <b>T</b>	53.8	65.603.580..65.603.602
	8 C HEX R	GAAGGTCGGAGTCAACGGATTCAACGGTAGGTGATACACCATA <b>C</b>		
	8 COM F	CAAATTTAACACGATGTTCCCTC	50.37	65.603.558..65.603.579
6 C->G	9 G FAM R	GAAGGTGACCAAGTTCATGCTTTGTAGGGCAAAGGTTCAA <b>A</b> <b>G</b>	53.5	65.604.102..65.604.123
	9 C HEX R	GAAGGTCGGAGTCAACGGATTTTGTAGGGCAAAGGTTCAA <b>A</b> <b>C</b>		
	9 COM F	GCCTTCCTTAGCAGATTCAATC	53.37	65.604.053..65.604.074
7 G->T	10 G FAM F	GAAGGTGACCAAGTTCATGCTTTGTAACTCAACCAGTCGATT <b>G</b>	52.9	65.604.373..65.604.394
	10 T HEX F	GAAGGTCGGAGTCAACGGATTTTGTAACTCAACCAGTCGATT <b>T</b>		
	10 COM R	GCGTGTAACAACCTGATAAATA	51.47	65.604.460..65.604.482
8 A->T	11 T FAM R	GAAGGTGACCAAGTTCATGCTTCCGATTTGAGAGGTAAGCAT <b>T</b>	53.6	65.604.514..65.604.535
	11 A HEX R	GAAGGTCGGAGTCAACGGATTTCCGATTTGAGAGGTAAGCAT <b>A</b>		
	11 COM F	GCCATGAACGAGTACAAGTAT	52.1	65.604.490..65.604.510
10 A->T	12 T FAM R	GAAGGTGACCAAGTTCATGCTTTGGTAAAAAGAGGTAATTAAC <b>T</b>	48.5	65.606.665..65.606.688
	12 A HEX R	GAAGGTCGGAGTCAACGGATTTTGGTAAAAAGAGGTAATTAAC <b>A</b>		
	12 COM F	CTTAGAATATTCTCCCTTTGTC	48.0	65.606.603..65.606.624

**Annex 3 – C<sub>t</sub> analysis raw data calculations and plotted graphs of the qRT-PCR experiment.**

<b>Results of the average C<sub>t</sub> for each sample</b>																	
<b>Assay</b>	<b>Type</b>	<b>PP ANT -1 (CT)</b>	<b>PP ANT -1 (STDEV)</b>	<b>PP OVA -1 (CT)</b>	<b>PP OVA -1 (STDEV)</b>	<b>PP ANT 0 (CT)</b>	<b>PP ANT 0 (STDEV)</b>	<b>PP OVA 0 (CT)</b>	<b>PP OVA 0 (STDEV)</b>	<b>PP 4 (CT)</b>	<b>PP 4 (STDEV)</b>	<b>PP 7 (CT)</b>	<b>PP 7 (STDEV)</b>	<b>PP 10 (CT)</b>	<b>PP 10 (STDEV)</b>	<b>PP 13 (CT)</b>	<b>PP 13 (STDEV)</b>
Solyc04g081620	Target	31,8605	0,125	31,7676	1,7697	31,747	1,3101	31,6035	1,7993	33,7977	2,4707	33,4536	0,315	39,3096	0,9764	37,3503	3,7473
Solyc04g081630	Target	33,1563	0,3595	31,7375	1,55	31,7377	1,7048	31,1791	0,6759	30,1544	0,7025	35,1434	6,8683	31,5999	1,8835	34,2291	3,699
Solyc04g081640	Target	31,4523	1,0436	27,4261	0,7316	32,6072	0,9492	27,6592	0,6017	25,4426	1,2201	30,4154	6,1707	26,1071	1,7149	26,1974	1,3775
Solyc04g081650	Target	27,5982	0,6246	26,1081	0,221	27,7861	0,5303	26,041	0,8342	25,1726	0,2594	27,7203	4,0929	25,6945	0,5725	26,1465	0,1606
Solyc04g081660	Target	28,063	0,2463	26,8237	0,2029	28,2286	0,4877	26,6011	0,8955	27,456	0,3347	33,5892	9,0663	28,1994	1,0057	27,8678	0,56
Solyc04g081670	Target	27,7787	1,0271	28,3842	0,5033	28,6703	0,7777	27,9927	0,8876	27,8298	1,3869	33,7053	8,902	27,8272	1,8028	27,6088	0,2508
Solyc04g081680	Target	31,8552	0,5758	30,7173	0,1929	31,3826	0,5771	30,3868	0,2915	30,1588	1,2556	33,7904	4,4502	31,862	0,438	31,0374	0,1139
CAC HK	Selected Control	23,5117	1,1342	23,77	1,0152	23,9599	1,2536	23,5103	0,8339	23,5706	1,0998	29,3747	6,64	23,3906	1,1289	22,9316	0,9888
<b>Assay</b>	<b>Type</b>	<b>MM ANT -1 (CT)</b>	<b>MM ANT -1 (STDEV)</b>	<b>MM OVA -1 (CT)</b>	<b>MM OVA -1 (STDEV)</b>	<b>MM ANT 0 (CT)</b>	<b>MM ANT 0 (STDEV)</b>	<b>MM OVA 0 (CT)</b>	<b>MM OVA 0 (STDEV)</b>	<b>MM 4 (CT)</b>	<b>MM 4 (STDEV)</b>	<b>MM 7 (CT)</b>	<b>MM 7 (STDEV)</b>	<b>MM 10 (CT)</b>	<b>MM 10 (STDEV)</b>	<b>MM 13 (CT)</b>	<b>MM 13 (STDEV)</b>
Solyc04g081620	Target	32,3136	1,1452	33,6355	3,3457	31,5944	0,2521	32,9769	1,4486	33,654	1,0984	34,0332	0,8643	33,483	0,73	36,7934	4,5349
Solyc04g081630	Target	33,3215	0,928	36,6021	4,8054	34,7396	5,3916	35,4249	6,4702	35,2754	3,5417	34,6529	2,4911	39,0618	1,3269	30,6602	0,4513
Solyc04g081640	Target	38,1834	2,1405	37,4444	3,1196	37,3646	3,0547	37,4662	2,9675	37,5464	2,8399	36,9766	3,509	34,7835	3,616	37,6215	3,0818
Solyc04g081650	Target	28,5271	0,3028	26,8891	0,6459	29,393	1,2514	27,4473	0,4608	28,3472	0,7441	27,6506	2,3771	28,6061	0,6661	26,7942	0,519
Solyc04g081660	Target	29,3003	0,5846	27,8503	0,8581	32,9	5,0506	28,0448	0,2926	29,9001	0,6426	30,3183	2,458	31,9634	0,6777	29,997	0,5563
Solyc04g081670	Target	29,6831	0,3132	26,7656	0,9413	32,6784	3,9427	27,9377	0,5784	29,6368	1,1083	27,9183	0,129	31,2608	1,7528	28,8151	0,8767
Solyc04g081680	Target	32,9719	0,6384	29,9984	1,3265	33,9093	2,7578	29,9835	0,2137	31,3037	0,41	31,0698	0,0458	31,3833	0,1937	32,1458	0,0516
CAC HK	Selected Control	24,016	1,0746	23,7433	0,8401	27,4125	3,639	24,1035	1,0826	25,4145	1,3504	24,2046	1,4058	26,1819	2,3532	23,5136	1,2561
<b>Assay</b>	<b>Type</b>	<b>H ANT -1 (CT)</b>	<b>H ANT -1 (STDEV)</b>	<b>H OVA -1 (CT)</b>	<b>H OVA -1 (STDEV)</b>	<b>H ANT 0 (CT)</b>	<b>H ANT 0 (STDEV)</b>	<b>H OVA 0 (CT)</b>	<b>H OVA 0 (STDEV)</b>	<b>H 4 (CT)</b>	<b>H 4 (STDEV)</b>	<b>H 7 (CT)</b>	<b>H 7 (STDEV)</b>	<b>H 10 (CT)</b>	<b>H 10 (STDEV)</b>	<b>H 13 (CT)</b>	<b>H 13 (STDEV)</b>
Solyc04g081620	Target	32,3039	1,5332	32,2017	0,3108	34,692	3,3875	34,7558	2,8428	33,9566	0,9295	31,8887	0,8603	33,0179	0,9041	34,8951	0,1382
Solyc04g081630	Target	31,5372	0,7561	34,3397	4,0608	34,6659	4,8511	33,3945	1,0229	30,7876	0,6608	31,4828	0,4133	31,0331	1,071	37,942	2,9105
Solyc04g081640	Target	33,1892	0,3543	28,8885	1,302	31,6005	1,7225	30,4054	1,3748	29,4415	1,6427	28,1912	2,2645	29,5365	0,7582	37,8804	2,4495
Solyc04g081650	Target	27,456	0,668	26,8186	0,6284	28,5484	1,7321	27,8571	1,9336	25,1876	1,2587	26,4415	0,6103	25,7248	3,3252	31,6244	0,7222
Solyc04g081660	Target	27,7513	0,3933	26,9635	0,2652	29,7223	3,1244	29,9904	2,5164	26,7014	0,0689	27,8283	0,1064	30,019	0,6306	40	0
Solyc04g081670	Target	27,8413	0,1351	26,8766	0,1157	31,9438	5,2952	30,8614	3,669	27,1992	0,004	28,453	0,5001	27,7968	0,4935	36,4801	0,7872
Solyc04g081680	Target	31,7032	0,1618	29,883	0,9958	30,8601	0,1686	30,7129	1,1857	29,2183	0,562	31,3228	1,0822	32,3907	0,9313	40	0
CAC HK	Selected Control	23,3975	1,0491	23,9268	1,2216	24,6597	1,745	25,851	1,1257	23,8783	1,437	24,0669	1,1233	24,0434	0,9065	34,5727	1,5674

**Results of the  $\Delta C_T$  for each sample**

Assay	Type	PP ANT -1 (DCT)	PP ANT -1 (STDEV)	PP OVA -1 (DCT)	PP OVA -1 (STDEV)	PP ANT 0 (DCT)	PP ANT 0 (STDEV)	PP OVA 0 (DCT)	PP OVA 0 (STDEV)	PP 4 (DCT)	PP 4 (STDEV)	PP 7 (DCT)	PP 7 (STDEV)	PP 10 (DCT)	PP 10 (STDEV)	PP 13 (DCT)	PP 13 (STDEV)
Solyc04g081620	Target	8,3488	1,1411	7,9976	2,0402	7,7871	1,8133	8,0932	1,9832	10,2271	2,7044	4,079	6,6475	15,919	1,4925	14,4186	3,8756
Solyc04g081630	Target	9,6446	1,1899	7,9675	1,8529	7,7778	2,1161	7,6688	1,0734	6,5839	1,305	5,7687	9,5532	8,2093	2,1959	11,2975	3,8289
Solyc04g081640	Target	7,9405	1,5413	3,6561	1,2514	8,6473	1,5724	4,1489	1,0284	1,872	1,6426	1,0407	9,0646	2,7165	2,0531	3,2658	1,6957
Solyc04g081650	Target	4,0864	1,2948	2,3381	1,039	3,8262	1,3611	2,5306	1,1795	1,602	1,13	-1,6544	7,8001	2,3039	1,2658	3,2148	1,0018
Solyc04g081660	Target	4,5512	1,1607	3,0537	1,0353	4,2686	1,3451	3,0908	1,2237	3,8855	1,1496	4,2145	11,2378	4,8088	1,5119	4,9362	1,1364
Solyc04g081670	Target	4,2669	1,5302	4,6142	1,1331	4,7103	1,4753	4,4824	1,2179	4,2592	1,77	4,3307	11,1056	4,4366	2,1271	4,6771	1,0201
Solyc04g081680	Target	8,3434	1,272	6,9473	1,0334	7,4227	1,38	6,8765	0,8834	6,5882	1,6691	4,4157	7,9934	8,4714	1,2109	8,1057	0,9954
CAC HK	Selected Control	0	1,6041	0	1,4357	0	1,7728	0	1,1794	0	1,5553	0	9,3904	0	1,5965	0	1,3984
Assay	Type	MM ANT -1 (DCT)	MM ANT -1 (STDEV)	MM OVA -1 (DCT)	MM OVA -1 (STDEV)	MM ANT 0 (DCT)	MM ANT 0 (STDEV)	MM OVA 0 (DCT)	MM OVA 0 (STDEV)	MM 4 (DCT)	MM 4 (STDEV)	MM 7 (DCT)	MM 7 (STDEV)	MM 10 (DCT)	MM 10 (STDEV)	MM 13 (DCT)	MM 13 (STDEV)
Solyc04g081620	Target	8,2976	1,5704	9,8921	3,4495	4,1819	3,6477	8,8733	1,8085	8,2395	1,7407	9,8287	1,6502	7,3011	2,4638	13,2797	4,7056
Solyc04g081630	Target	9,3055	1,4199	12,8587	4,8783	7,3271	6,5047	11,3213	6,5602	9,8608	3,7904	10,4483	2,8604	12,8798	2,7015	7,1465	1,3348
Solyc04g081640	Target	14,1674	2,3951	13,7011	3,2308	9,9521	4,7511	13,3626	3,1588	12,1319	3,1446	12,772	3,7801	8,6015	4,3143	14,1079	3,328
Solyc04g081650	Target	4,5111	1,1165	3,1457	1,0597	1,9805	3,8482	3,3438	1,1766	2,9326	1,5418	3,446	2,7617	2,4242	2,4456	3,2806	1,3591
Solyc04g081660	Target	5,2843	1,2234	4,1069	1,2009	5,4875	6,225	3,9413	1,1214	4,4856	1,4955	6,1137	2,8316	5,7815	2,4488	6,4833	1,3738
Solyc04g081670	Target	5,6672	1,1193	3,0223	1,2617	5,2659	5,3653	3,8342	1,2274	4,2223	1,747	3,7137	1,4117	5,0789	2,9342	5,3015	1,5318
Solyc04g081680	Target	8,9559	1,2499	6,255	1,5701	6,4968	4,5659	5,88	1,1035	5,8892	1,4112	6,8652	1,4065	5,2014	2,3612	8,6322	1,2572
CAC HK	Selected Control	0	1,5197	0	1,1881	0	5,1463	0	1,531	0	1,9097	0	1,9881	0	3,3279	0	1,7765
Assay	Type	H ANT -1 (DCT)	H ANT -1 (STDEV)	H OVA -1 (DCT)	H OVA -1 (STDEV)	H ANT 0 (DCT)	H ANT 0 (STDEV)	H OVA 0 (DCT)	H OVA 0 (STDEV)	H 4 (DCT)	H 4 (STDEV)	H 7 (DCT)	H 7 (STDEV)	H 10 (DCT)	H 10 (STDEV)	H 13 (DCT)	H 13 (STDEV)
Solyc04g081620	Target	8,9063	1,8578	8,2749	1,2605	10,0323	3,8106	8,9048	3,0575	10,0783	1,7114	7,8218	1,4149	8,9745	1,2803	0,3224	1,5735
Solyc04g081630	Target	8,1396	1,2932	10,4128	4,2406	10,0062	5,1554	7,5435	1,521	6,9093	1,5817	7,4159	1,1969	6,9897	1,4031	3,3693	3,3057
Solyc04g081640	Target	9,7917	1,1073	4,9617	1,7853	6,9408	2,4519	4,5544	1,7768	5,5632	2,1826	4,1243	2,5278	5,493	1,1818	3,3076	2,9081
Solyc04g081650	Target	4,0585	1,2437	2,8917	1,3737	3,8887	2,4587	2,0062	2,2374	1,3094	1,9103	2,3746	1,2784	1,6813	3,4465	-2,9483	1,7258
Solyc04g081660	Target	4,3538	1,1204	3,0367	1,25	5,0627	3,5787	4,1394	2,7567	2,8231	1,4387	3,7614	1,1284	5,9756	1,1042	5,4273	1,5674
Solyc04g081670	Target	4,4437	1,0578	2,9498	1,227	7,2842	5,5754	5,0104	3,8378	3,3209	1,437	4,3861	1,2296	3,7533	1,0321	1,9074	1,754
Solyc04g081680	Target	8,3057	1,0615	5,9562	1,576	6,2005	1,7531	4,8619	1,6349	5,34	1,543	7,2559	1,5598	8,3473	1,2996	5,4273	1,5674
CAC HK	Selected Control	0	1,4836	0	1,7275	0	2,4678	0	1,592	0	2,0323	0	1,5886	0	1,2819	0	2,2167



**Result of the  $\Delta\Delta C_t$  calculation**

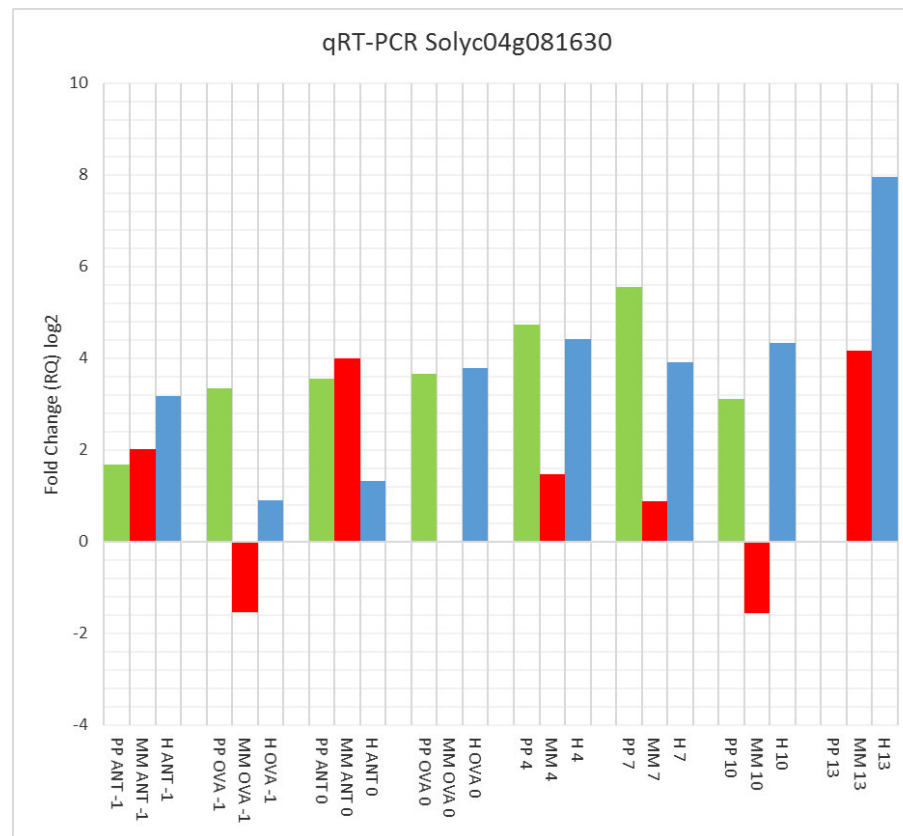
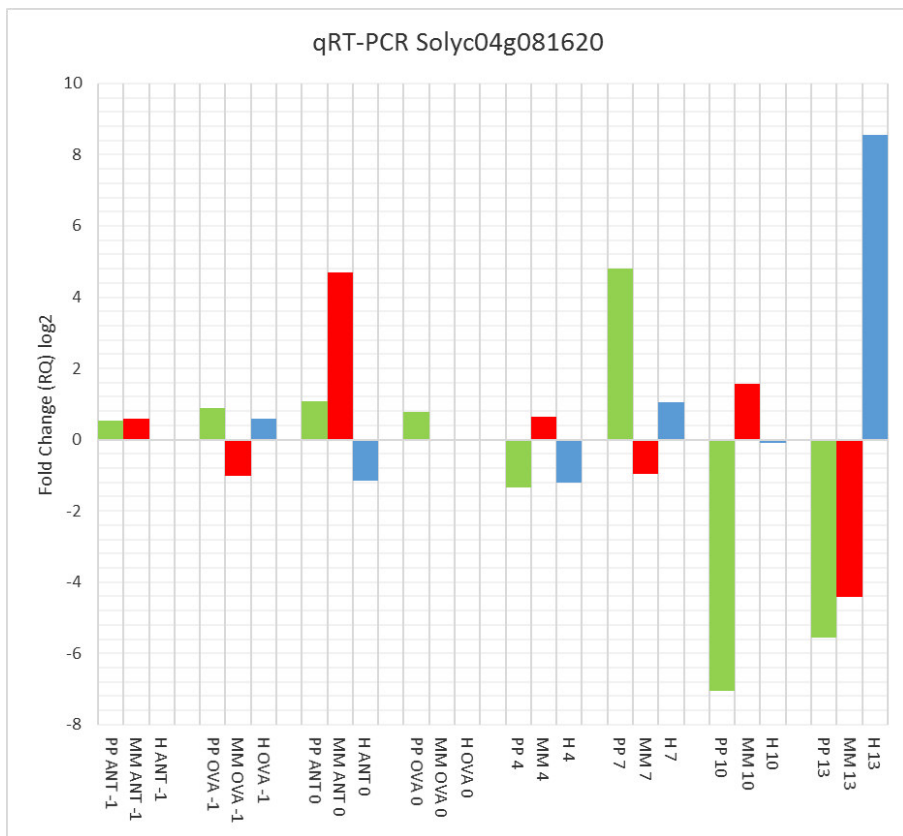
Assay	Type	PP ANT -1 (DDCT)	PP ANT -1 (STDEV)	PP OVA -1 (DDCT)	PP OVA -1 (STDEV)	PP ANT 0 (DDCT)	PP ANT 0 (STDEV)	PP OVA 0 (DDCT)	PP OVA 0 (STDEV)	PP 4 (DDCT)	PP 4 (STDEV)	PP 7 (DDCT)	PP 7 (STDEV)	PP 10 (DDCT)	PP 10 (STDEV)	PP 13 (DDCT)	PP 13 (STDEV)
Solyc04g081620	Target	-0,5245	1,1411	-0,8757	2,0402	-1,0862	1,8133	-0,7801	1,9832	1,3538	2,7044	-4,7943	6,6475	7,0457	1,4925	5,5453	3,8756
Solyc04g081630	Target	-1,6767	1,1899	-3,3538	1,8529	-3,5435	2,1161	-3,6525	1,0734	-4,7374	1,305	-5,5526	9,5532	-3,112	2,1959	-0,0238	3,8289
Solyc04g081640	Target	-5,4221	1,5413	-9,7065	1,2514	-4,7153	1,5724	-9,2137	1,0284	-11,4906	1,6426	-12,3219	9,0646	-10,6461	2,0531	-10,0968	1,6957
Solyc04g081650	Target	0,7426	1,2948	-1,0057	1,039	0,4824	1,3611	-0,8132	1,1795	-1,7418	1,13	-4,9982	7,8001	-1,0399	1,2658	-0,129	1,0018
Solyc04g081660	Target	0,6099	1,1607	-0,8876	1,0353	0,3273	1,3451	-0,8505	1,2237	-0,0558	1,1496	0,2732	11,2378	0,8675	1,5119	0,9949	1,1364
Solyc04g081670	Target	0,4327	1,5302	0,78	1,1331	0,8761	1,4753	0,6482	1,2179	0,425	1,77	0,4965	11,1056	0,6024	2,1271	0,8429	1,0201
Solyc04g081680	Target	2,4634	1,272	1,0673	1,0334	1,5427	1,38	0,9965	0,8834	0,7082	1,6691	-1,4643	7,9934	2,5914	1,2109	2,2257	0,9954
Assay	Type	MM ANT -1 (DDCT)	MM ANT -1 (STDEV)	MM OVA -1 (DDCT)	MM OVA -1 (STDEV)	MM ANT 0 (DDCT)	MM ANT 0 (STDEV)	MM OVA 0 (DDCT)	MM OVA 0 (STDEV)	MM 4 (DDCT)	MM 4 (STDEV)	MM 7 (DDCT)	MM 7 (STDEV)	MM 10 (DDCT)	MM 10 (STDEV)	MM 13 (DDCT)	MM 13 (STDEV)
Solyc04g081620	Target	-0,5757	1,5704	1,0188	3,4495	-4,6914	3,6477	0	1,8085	-0,6338	1,7407	0,9554	1,6502	-1,5722	2,4638	4,4064	4,7056
Solyc04g081630	Target	-2,0158	1,4199	1,5374	4,8783	-3,9942	6,5047	0	6,5602	-1,4605	3,7904	-0,873	2,8604	1,5585	2,7015	-4,1748	1,3348
Solyc04g081640	Target	0,8048	2,3951	0,3385	3,2308	-3,4105	4,7511	0	3,1588	-1,2307	3,1446	-0,5906	3,7801	-4,7611	4,3143	0,7453	3,328
Solyc04g081650	Target	1,1673	1,1165	-0,1981	1,0597	-1,3633	3,8482	0	1,1766	-0,4112	1,5418	0,1022	2,7617	-0,9196	2,4456	-0,0632	1,3591
Solyc04g081660	Target	1,343	1,2234	0,1656	1,2009	1,5462	6,225	0	1,1214	0,5443	1,4955	2,1724	2,8316	1,8402	2,4488	2,542	1,3738
Solyc04g081670	Target	1,833	1,1193	-0,8119	1,2617	1,4317	5,3653	0	1,2274	0,3881	1,747	-0,1205	1,4117	1,2447	2,9342	1,4673	1,5318
Solyc04g081680	Target	3,0759	1,2499	0,375	1,5701	0,6168	4,5659	0	1,1035	0,0092	1,4112	0,9852	1,4065	-0,6786	2,3612	2,7522	1,2572
Assay	Type	H ANT -1 (DDCT)	H ANT -1 (STDEV)	H OVA -1 (DDCT)	H OVA -1 (STDEV)	H ANT 0 (DDCT)	H ANT 0 (STDEV)	H OVA 0 (DDCT)	H OVA 0 (STDEV)	H 4 (DDCT)	H 4 (STDEV)	H 7 (DDCT)	H 7 (STDEV)	H 10 (DDCT)	H 10 (STDEV)	H 13 (DDCT)	H 13 (STDEV)
Solyc04g081620	Target	0,033	1,8578	-0,5984	1,2605	1,159	3,8106	0,0315	3,0575	1,205	1,7114	-1,0515	1,4149	0,1012	1,2803	-8,5509	1,5735
Solyc04g081630	Target	-3,1817	1,2932	-0,9085	4,2406	-1,3151	5,1554	-3,7778	1,521	-4,412	1,5817	-3,9054	1,1969	-4,3316	1,4031	-7,952	3,3057
Solyc04g081640	Target	-3,5709	1,1073	-8,4009	1,7853	-6,4218	2,4519	-8,8082	1,7768	-7,7994	2,1826	-9,2383	2,5278	-7,8696	1,1818	-10,055	2,9081
Solyc04g081650	Target	0,7147	1,2437	-0,4521	1,3737	0,5449	2,4587	-1,3376	2,2374	-2,0344	1,9103	-0,9692	1,2784	-1,6625	3,4465	-6,2921	1,7258
Solyc04g081660	Target	0,4125	1,1204	-0,9046	1,25	1,1214	3,5787	0,1981	2,7567	-1,1182	1,4387	-0,1799	1,1284	2,0343	1,1042	1,486	1,5674
Solyc04g081670	Target	0,6095	1,0578	-0,8844	1,227	3,45	5,5754	1,1762	3,8378	-0,5133	1,437	0,5519	1,2296	-0,0809	1,0321	-1,9268	1,754
Solyc04g081680	Target	2,4257	1,0615	0,0762	1,576	0,3205	1,7531	-1,0181	1,6349	-0,54	1,543	1,3759	1,5598	2,4673	1,2996	-0,4527	1,5674

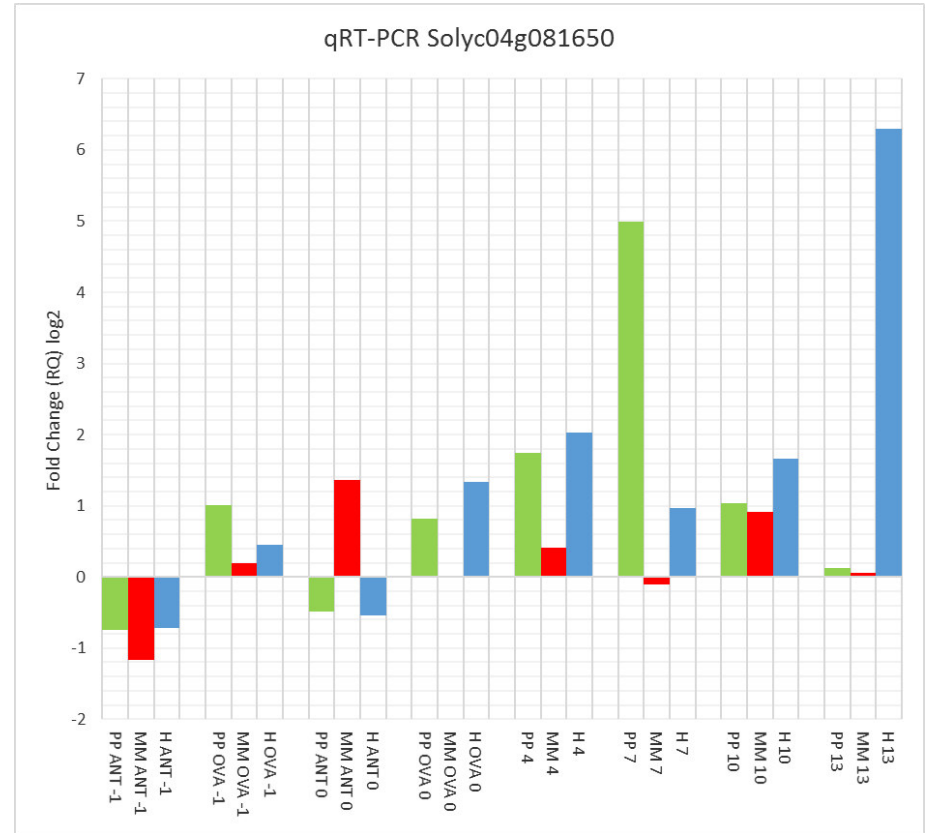
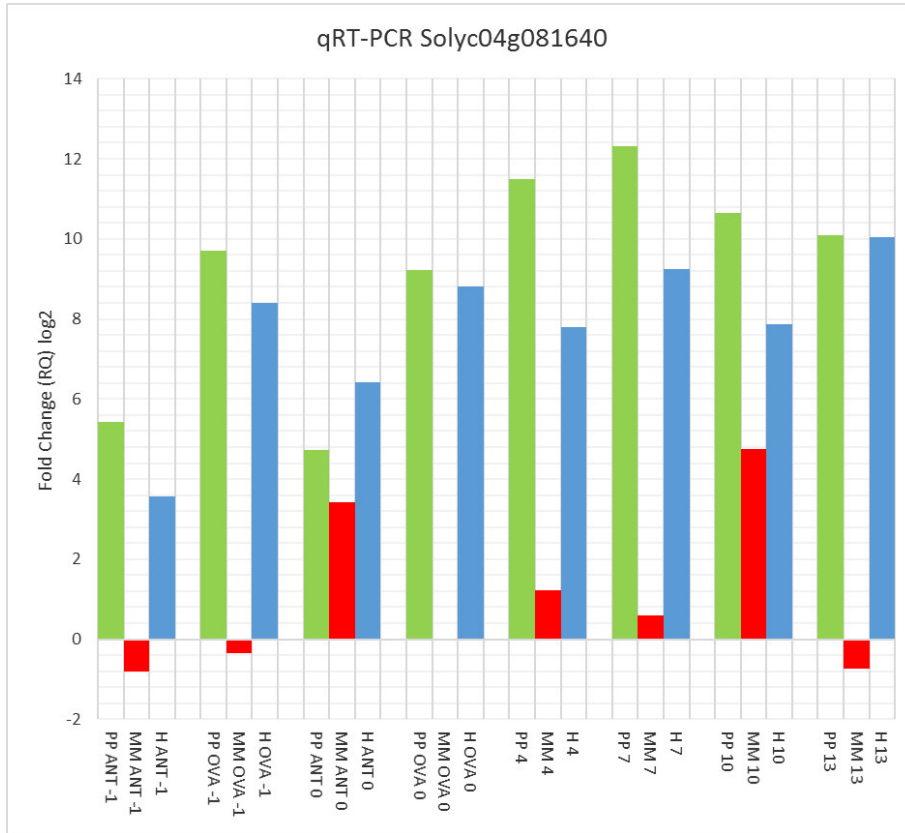
**Result of the  $2^{\Delta\Delta C_t}$  calculation - fold-change**

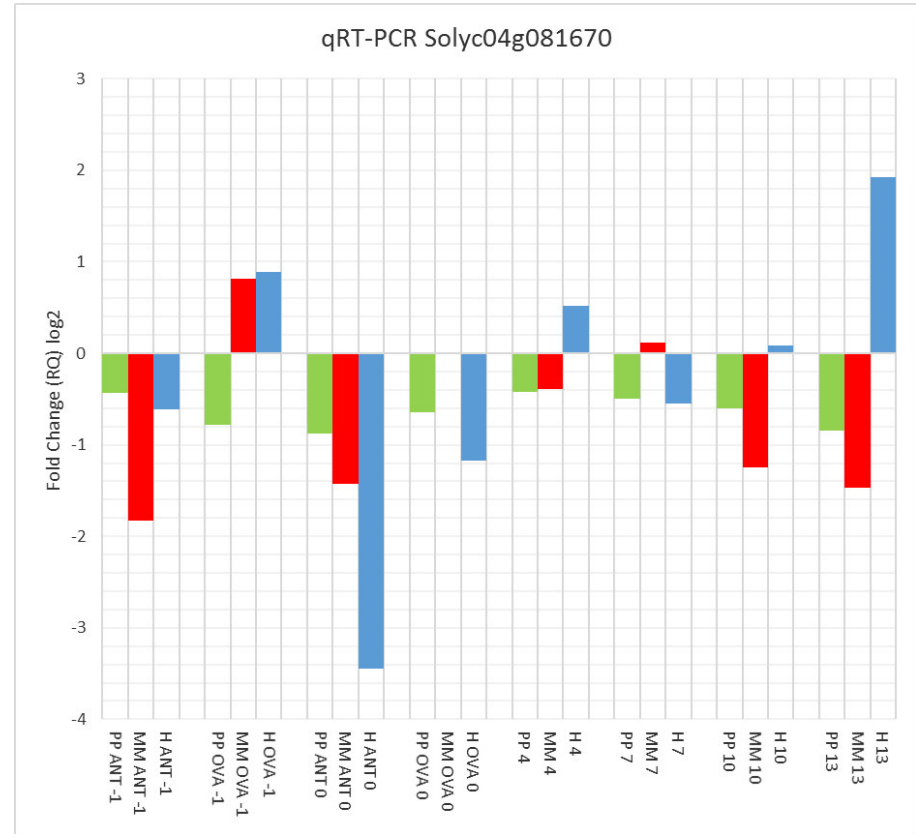
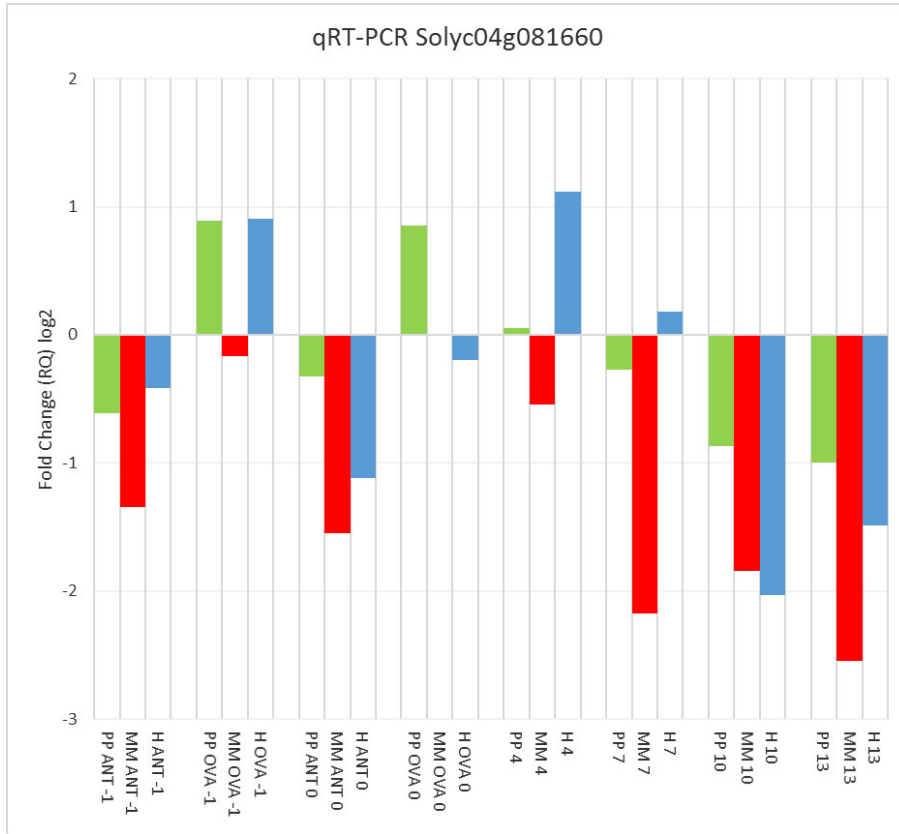
Assay	Type	PP ANT -1	MM ANT -1	H ANT -1	PP OVA -1	MM OVA -1	H OVA -1	PP ANT 0	MM ANT 0	H ANT 0	PP OVA 0	MM OVA 0	H OVA 0	PP 4	MM 4	H 4	PP 7	MM 7	H 7	PP 10	MM 10	H 10	PP 13	MM 13	H 13
Solyc04g081620	Target	1,4385	1,4904	0,9774	1,835	0,4935	1,514	2,1232	25,8386	0,4478	1,7173	1	0,9784	0,3913	1,5517	0,4338	27,7489	0,5157	2,0726	0,0076	2,9737	0,9323	0,0214	0,0472	375,051
Solyc04g081630	Target	3,197	4,044	9,0737	10,2237	0,3445	1,877	11,6602	15,9364	2,4882	12,5753	1	13,7162	26,6753	2,752	21,2888	46,9355	1,8314	14,9848	8,646	0,3395	20,1348	1,0166	18,0609	247,630
Solyc04g081640	Target	42,8757	0,5724	11,8838	835,542	0,7909	338,007	26,2701	10,6334	85,735	593,776	1	448,264	2877,51	2,3469	222,768	5119,92	1,5059	603,956	1602,52	27,117	233,875	1095,09	0,5966	1063,80
Solyc04g081650	Target	0,5976	0,4452	0,6093	2,0078	1,1471	1,3679	0,7158	2,5727	0,6854	1,757	1	2,5273	3,3445	1,3298	4,0965	31,9591	0,9316	1,9577	2,0561	1,8916	3,1655	1,0935	1,0448	78,3627
Solyc04g081660	Target	0,6552	0,3942	0,7513	1,8501	0,8915	1,872	0,797	0,3424	0,4597	1,8031	1	0,8717	1,0394	0,6857	2,1706	0,8275	0,2218	1,1328	0,5481	0,2793	0,2441	0,5018	0,1717	0,357
Solyc04g081670	Target	0,7408	0,2807	0,6554	0,5823	1,7555	1,8459	0,5448	0,3707	0,0915	0,6381	1	0,4425	0,7448	0,7641	1,4273	0,7088	1,0871	0,6821	0,6586	0,422	1,0576	0,5575	0,3617	3,8021
Solyc04g081680	Target	0,1813	0,1186	0,1861	0,4772	0,7711	0,9485	0,3432	0,6521	0,8008	0,5012	1	2,0252	0,6121	0,9936	1,4539	2,7593	0,5051	0,3853	0,1659	1,6006	0,1808	0,2138	0,1484	1,3686

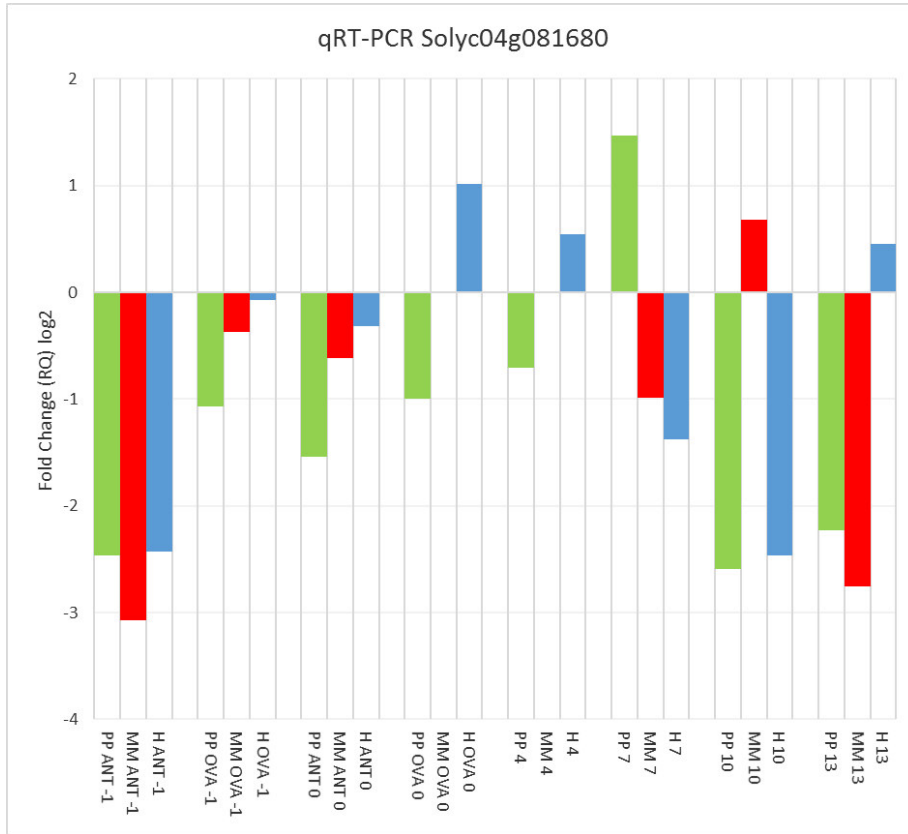
**Result of the  $2^{\Delta\Delta C_t}$  calculation - log<sub>2</sub> fold change**

Assay	Type	PP ANT -1	MM ANT -1	H ANT -1	PP OVA -1	MM OVA -1	H OVA -1	PP ANT 0	MM ANT 0	H ANT 0	PP OVA 0	MM OVA 0	H OVA 0	PP 4	MM 4	H 4	PP 7	MM 7	H 7	PP 10	MM 10	H 10	PP 13	MM 13	H 13
Solyc04g081620	Target	0,5246	0,5757	-0,0330	0,8758	-1,0189	0,5984	1,0862	4,6915	-1,1591	0,7801	0	-0,0315	-1,3537	0,6338	-1,2049	4,7944	-0,9554	1,0514	-7,0398	1,5723	-0,1011	-5,5462	-4,4051	8,5509
Solyc04g081630	Target	1,6767	2,0158	3,1817	3,3538	-1,5374	0,9084	3,5435	3,9943	1,3151	3,6525	0	3,7778	4,7374	1,4605	4,4120	5,5526	0,8729	3,9054	3,1120	-1,5585	4,3316	0,0238	4,1748	7,9520
Solyc04g081640	Target	5,4221	-0,8049	3,5709	9,7066	-0,3384	8,4009	4,7153	3,4105	6,4218	9,2138	0	8,8082	11,4906	1,2308	7,7994	12,3219	0,5906	9,2383	10,6461	4,7611	7,8696	10,0968	-0,7452	10,0550
Solyc04g081650	Target	-0,7427	-1,1675	-0,7148	1,0056	0,1980	0,4520	-0,4824	1,3633	-0,5450	0,8131	0	1,3376	1,7418	0,4112	2,0344	4,9982	-0,1022	0,9692	1,0399	0,9196	1,6624	0,1290	0,0632	6,2921
Solyc04g081660	Target	-0,6100	-1,3430	-0,4125	0,8876	-0,1657	0,9046	-0,3273	-1,5462	-1,1212	0,8505	0	-0,1981	0,0558	-0,5444	1,1181	-0,2732	-2,1727	0,1799	-0,8675	-1,8401	-2,0345	-0,9948	-2,5420	-1,4860
Solyc04g081670	Target	-0,4328	-1,8329	-0,6096	-0,7802	0,8119	0,8843	-0,8762	-1,4317	-3,4501	-0,6481	0	-1,1763	-0,4251	-0,3882	0,5133	-0,4965	0,1205	-0,5519	-0,6025	-1,2447	0,0808	-0,8430	-1,4671	1,9268
Solyc04g081680	Target	-2,4635	-3,0758	-2,4259	-1,0673	-0,3750	-0,0763	-1,5429	-0,6168	-0,3205	-0,9965	0	1,0181	-0,7082	-0,0093	0,5399	1,4643	-0,9854	-1,3759	-2,5916	0,6786	-2,4675	-2,2257	-2,7524	0,4527









## Annex 4 – List of GB elements used and created with the GB 3.0 toolkit for the assembly of the CRISPR/Cas9 cassettes and Expression cassettes.

### List of GB elements used and assembled in the construction of the Knockout vectors

Level -1 GB			
GBdatabase ID	Name	Vector	Description
GB1207	tRNA-gRNA position [n]	pVD1	tRNA and scaffold for the assembly of GBoligomers for the last position (positon [n]) of a polycistronic tRNA-gRNA
GB1208	tRNA-gRNA position [D1_n-1]	pVD1	tRNA and scaffold for the assembly of GBoligomers for the first position (positon [D1_n-1]) of a polycistronic tRNA-gRNA regulated by the U6-26 or U6-1 promoter
Level 0 GB			
GBdatabase ID	Name	Category	Description
GB1001	pAtU6-26	A1-A2-A3-B1-B2c	<i>Arabidopsis thaliana</i> U6-26 RNA polIII promoter
GB0645	psgRNA	B6b-C1	Short guide RNA which is the combination of the bacterial crRNA and tracrRNA into a single guide transcript. It does not contain any target site.
pDGB3_alpha1	pDGB3_alpha1	Vector	Backbone vector used in GB assembly system
pDGB3_omega2	pDGB3_omega2	Vector	Backbone vector used in GB assembly system
pUPD2	pUPD2	Vector	Backbone vector used in GB assembly system
	pUPD2:tRNA-gRNA40.1	Domestication construct	Domestication of a guide RNA targeting the Exon 2 of the Solyc04g081640.1 gene of <i>Solanum lycopersicum</i> (CRISPR-tools)

pUPD2:tRNA-gRNA40.2

Domestication  
construct

Domestication of a guide RNA targeting the Exon 4 of the Solyc04g081640.1 gene of *Solanum lycopersicum* (CRISPR-tools)

### Level 1 GB

GBdatabase ID	Name	Category	Description
GB0639	pEGB:35s:hCas9:Tnos	TU	Transcription unit of the human Cas9. Includes a 35S Promoter, a coding sequence of the human Cas9 (hCas9), and a Tnos Terminator.
	pDGB3_omega1:U6-26:SL4g30.1:sgRNA:35s:hCas9:Tnos	TU	Transcription unit for the guideRNA targeting the Solyc04g081630.1 gene of <i>Solanum lycopersicum</i> (CRISPR-

### Level >1 GB

GBdatabase ID	Name	Category	Description
GB1181	pDGB3_omega1R:Tnos:NptII:Pnos-SF	TU	Transcription unit of the neomycin phosphotransferase II gene (NptII) for Kanamycin/Neomycin resistance. Includes a Pnos Promoter, a coding sequence of the NptII gene, and a Tnos Terminator.
	pDGB3_omega2:U6-26:SL4g30.1:sgRNA:35s:hCas9:Tnos	Module	Module containing 2 TU. (1) Transcription unit for the guideRNA targeting the Solyc04g081630.1 gene of <i>Solanum lycopersicum</i> (CRISPR-tools); (2) Transcription unit of the human Cas9.
	pDGB3_omega2:U6-26:tRNA-gRNA40.1.1:tRNA-gRNA40.1.2:35s:hCas9:Tnos	Module	Module containing 2 TU. (1) Transcription unit for the two guideRNAs targeting the Solyc04g081640.1 gene of <i>Solanum lycopersicum</i> (CRISPR-tools); (2) Transcription unit of the human Cas9.
	pDGB3_alpha1:Tnos:NptII:Pnos-SF:U6-26:SL4g30.1:sgRNA:35s:hCas9:Tnos	Module	Module containing 3 TU. (1) Transcription unit of the neomycin phosphotransferase II gene (NptII) for Kanamycin/Neomycin resistance; (2) Transcription unit for the guideRNA targeting the Solyc04g081630.1 gene of <i>Solanum lycopersicum</i> (CRISPR-tools); (3) Transcription unit of the human Cas9.
	pDGB3_alpha1:Tnos:NptII:Pnos-SF:U6-26:tRNA-gRNA40.1.1:tRNA-gRNA40.1.2:35s:hCas9:Tnos	Module	Module containing 3 TU. (1) Transcription unit of the neomycin phosphotransferase II gene (NptII) for Kanamycin/Neomycin resistance; (2) Transcription unit for the two guideRNAs targeting the Solyc04g081640.1 gene of <i>Solanum lycopersicum</i> (CRISPR-tools); (3) Transcription unit of the human Cas9.



### List of GB elements used and assembled in the construction of the Expression vectors

Level 0 GB			
GBdatabase ID	Name	Category	Description
GB0030	pP35S (PROM+5UTR)	A1-A2-A3-B1-B2	CaMV 35S promoter
GB0100	pDsRed	B3-B4-B5	Red fluorescent protein from <i>Discosoma sp.</i>
GB0037	pTnos (3UTR+Term)	B6-C1	<i>Agrobacterium tumefaciens</i> terminator
pDGB3_alpha1	pDGB3_alpha1	Vector	Backbone vector used in GB assembly system
pDGB3_omega1	pDGB3_omega1	Vector	Backbone vector used in GB assembly system
pUPD2	pUPD2	Vector	Backbone vector used in GB assembly system
	pUPD2_SPSolyc04g081640.1CDS	B3-B4-B5	Domestication of the CDS of the gene Solyc04g081640.1 from <i>Solanum lycopersicum</i> IL4-4 (PP haplotype).
	pUPD2_SLSolyc04g081640.1Prom	A1-A2-A3-B1-B2	Domestication of the Promoter region of the gene Solyc04g081640.1 from <i>Solanum lycopersicum</i> var. "MoneyMaker".
	pUPD2_:SPSolyc04g081640.1Prom	A1-A2-A3-B1-B2	Domestication of the Promoter region of the gene Solyc04g081640.1 from <i>Solanum lycopersicum</i> IL4-4 (PP haplotype).

## Level 1 GB

GBdatabase ID	Name	Category	Description
GB0184	pDGB3_alpha2:Pnos:NptII:Tnos	TU	Transcription unit of the neomycin phosphotransferase II gene (NptII) for Kanamycin/Neomycin resistance. Includes a Pnos Promoter, a coding sequence of the NptII gene, and a Tnos Terminator.
	pDGB3_alpha1:35s:SPSolyc04g081640.1CDS:Tnos	TU	Transcription unit of the gene Solyc04g081640.1 from <i>Solanum lycopersicum</i> IL4-4 (PP haplotype). Contains a 35S Promoter, CDS of the Solyc04g081640.1 gene and a Tnos terminator.
	pDGB3_alpha1:SLSolyc04g081640.1Prom:DsRed:Tnos	TU	Transcription unit of the Promoter region of the gene Solyc04g081640.1 from <i>Solanum lycopersicum</i> var. "MoneyMaker". Contains Promoter region of the gene Solyc04g081640.1, the CDS of the DsRed protein, and a Tnos terminator.
	pDGB3_alpha1:SPSolyc04g081640.1Prom:DsRed:Tnos	TU	Transcription unit of the Promoter region of the gene Solyc04g081640.1 from <i>Solanum lycopersicum</i> IL4-4 (PP haplotype). Contains Promoter region of the gene Solyc04g081640.1, the CDS of the DsRed protein, and a Tnos terminator.

**Level >1 GB parts**

<b>GBdatabase ID</b>	<b>Name</b>	<b>Category</b>	<b>Description</b>
	pDGB_omega1::35s:SPSolyc04g081640.1CDS:Tnos::Pnos:NptII:Tnos	Module	Module containing 2 TU. (1) Transcription unit of the gene Solyc04g081640.1 from <i>Solanum lycopersicum</i> IL4-4 (PP haplotype). (2) Transcription unit of the neomycin phosphotransferase II gene (NptII) for Kanamycin/Neomycin resistance.
	pDGB3_omega1::SLSolyc04g081640.1Prom:DsRed:Tnos::Pnos:NptII:Tnos	Module	Module containing 2 TU. (1) Transcription unit of the Promoter region of the gene Solyc04g081640.1 from <i>Solanum lycopersicum</i> var. "Moneymaker". (2) Transcription unit of the neomycin phosphotransferase II gene (NptII) for Kanamycin/Neomycin resistance.
	pDGB3_omega1::SPSolyc04g081640.1Prom:DsRed:Tnos::Pnos:NptII:Tnos	Module	Module containing 2 TU. (1) Transcription unit of the Promoter region of the gene Solyc04g081640.1 from <i>Solanum lycopersicum</i> IL4-4 (PP haplotype). (2) Transcription unit of the neomycin phosphotransferase II gene (NptII) for Kanamycin/Neomycin resistance.

**Annex 5 – Results of the HRM genotyping.**

0 – Moneymaker (MM) haplotype

1 – Pimpinellifolium (PP) haplotype

2 – Heterozygote (H) haplotype

<b>IL4-4 Sample</b>		<b>B3-1</b>	<b>B3-2</b>	<b>B3-3</b>	<b>B3-4</b>	<b>B3-5</b>	<b>B3-6</b>	<b>B3-7</b>	<b>B3-8</b>	<b>B3-9</b>	<b>B3-10</b>	<b>B3-11</b>	<b>B3-12</b>	<b>B3-13</b>	<b>B2-1</b>	<b>B2-2</b>	<b>B2-3</b>	<b>B2-5</b>	<b>B2-6</b>	<b>B2-9</b>
<b>Marker</b>	<b>Solcap_snp_47742</b>	2	1	2	1	2	1	1	1	1	1	2	1	1	2	1	1	2	1	2
	<b>Solcap_snp_3952</b>	2	1	2	1	2	1	1	1	1	1	2	1	1	2	1	1	2	1	2

<b>MM Sample</b>		<b>MM 1</b>	<b>MM 2</b>	<b>MM 3</b>	<b>MM 4</b>	<b>MM 5</b>	<b>MM 6</b>	<b>MM 7</b>	<b>MM 8</b>	<b>MM 9</b>	<b>MM 10</b>	<b>MM 11</b>	<b>MM 12</b>	<b>MM 13</b>	<b>MM 14</b>	<b>MM 15</b>	<b>MM 16</b>	<b>MM 17</b>	<b>MM 18</b>	<b>MM 19</b>
<b>Marker</b>	<b>Solcap_snp_47742</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<b>Solcap_snp_3952</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

**Annex 6 – Results of the KASP analysis performed on *S. pimpinellifolium* and *S. lycopersicum* var *cerasiforme* samples.**

0 – Moneymaker (MM) / Wild type SNP

1 – Pimpinellifolium (PP) / Mutant SNP

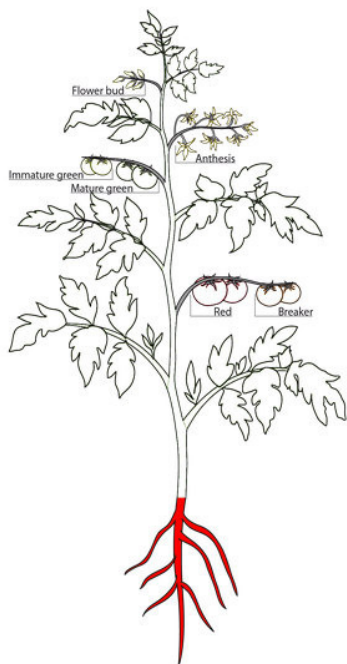
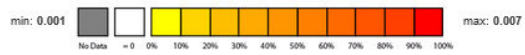
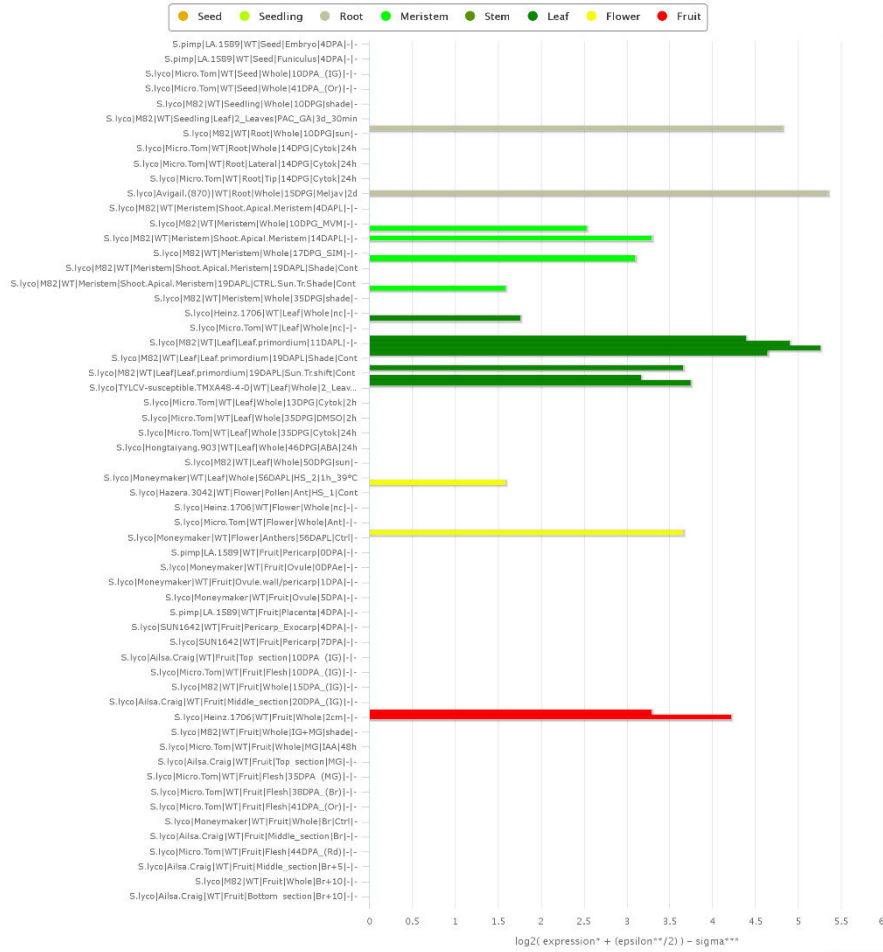
2 – Heterozygote (H) SNP

TGRC	Botanical variety	Category	Cluster-T0937	TS accession	SNP coordinates on chr4 (SL2.5)										
					1	2	3	4	5	6	7	8	9	10	11
LA2093	<i>S. pimpinellifolium</i>	Wild species	1	TS-15	2	2	1	1	2	2	2	2	2	1	2
LA1246	<i>S. pimpinellifolium</i>	Wild species	1	TS-16	1	1	2	2	1	1	2	1	2	1	1
LA1578	<i>S. pimpinellifolium</i>	Wild species	1	TS-437	2	2	2	2	2	1	2	2	2	2	2
LA2183	<i>S. pimpinellifolium</i>	Wild species	2	TS-182	2	2	2	2	2	1	2	2	2	2	2
LA1242	<i>S. pimpinellifolium</i>	Wild species	2	TS-413	1	2	2	2	1	2	2	2	2	2	2
LA2184	<i>S. pimpinellifolium</i>	Wild species	2	TS-420	0	0	2	2	1	1	2	1	2	0	0
LA0400	<i>S. pimpinellifolium</i>	Wild species	3	TS-265	2	2	2	2	2	2	2	2	2	2	2
LA0417	<i>S. pimpinellifolium</i>	Wild species	4	TS-50	0	0	0	0	0	0	0	0	0	0	0
LA1547	<i>S. pimpinellifolium</i>	Wild species	10	TS-14	0	0	2	2	2	2	2	2	2	0	2
LA1456	<i>S. lycopersicum</i> var <i>cerasiforme</i>	Wild species	9	TS-96	0	0	0	0	0	0	0	0	0	0	0
LA1245	<i>S. pimpinellifolium</i>	Wild species	11	TS-124	0	0	2	0	0	0	1	0	2	0	0
LA2688	<i>S. lycopersicum</i> var <i>cerasiforme</i>	Wild species	8	TS-301	2	2	2	2	2	2	2	2	2	2	2

# Annex 7 – Results of the *in silico* analysis of the candidate genes and respective orthologs in *Arabidopsis thaliana*.

## a) Solyc04g081620

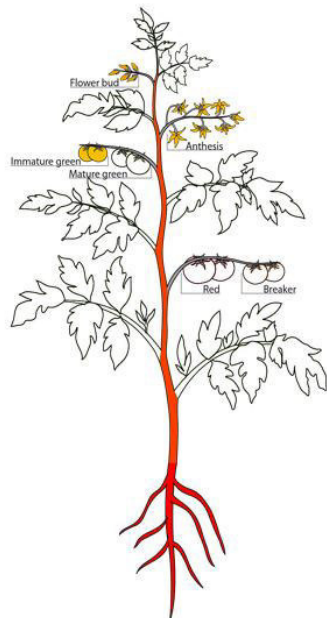
Expression of "Solyc04g081620" in different organs and global conditions in *S. lycopersicum*




Seedling
Fruit set
Fruit tissues
Fruit development
Fruit parts
Hormone treatment: Fruit
Hormone treatment: Vegetative tissues
Biotic interactions

b) Solyc04g081630

Expression of "Solyc04g081630" in different organs and global conditions in *S. lycopersicum*




**Seedling**




**Fruit set**


**Fruit tissues**



**Fruit development**




**Fruit parts**



**Hormone treatment: Fruit**

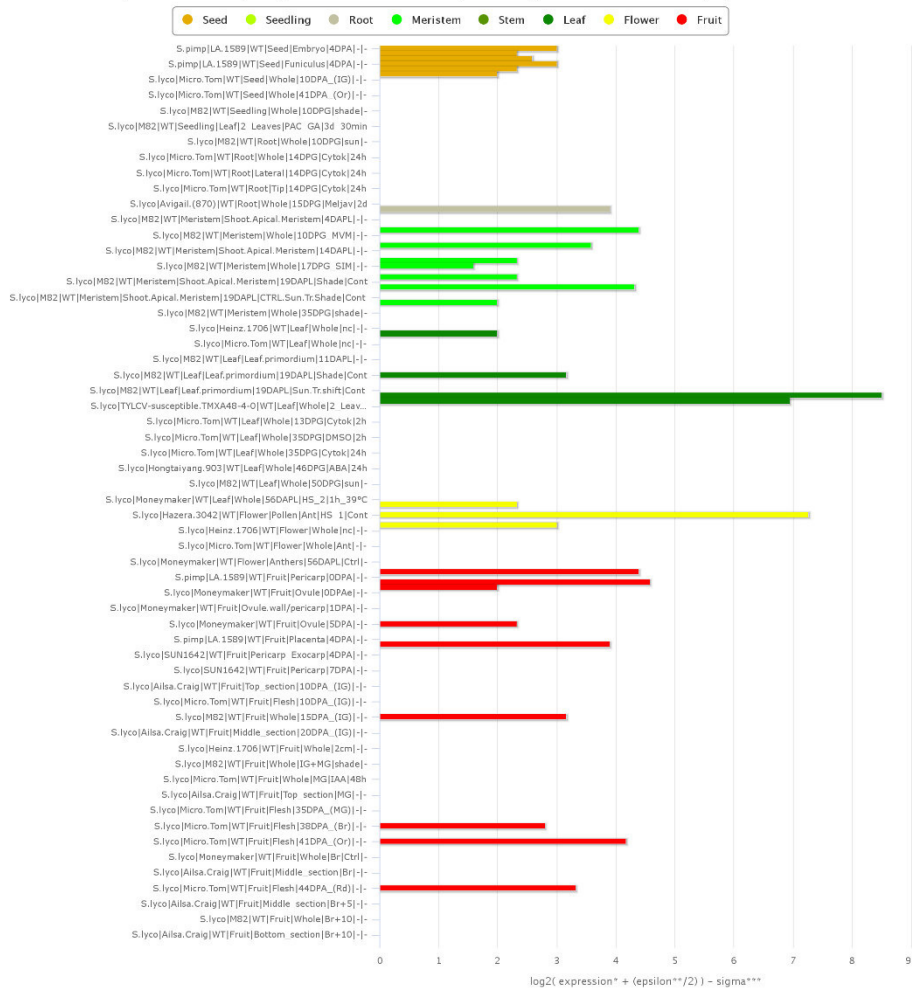
**Hormone treatment: Vegetative tissues**



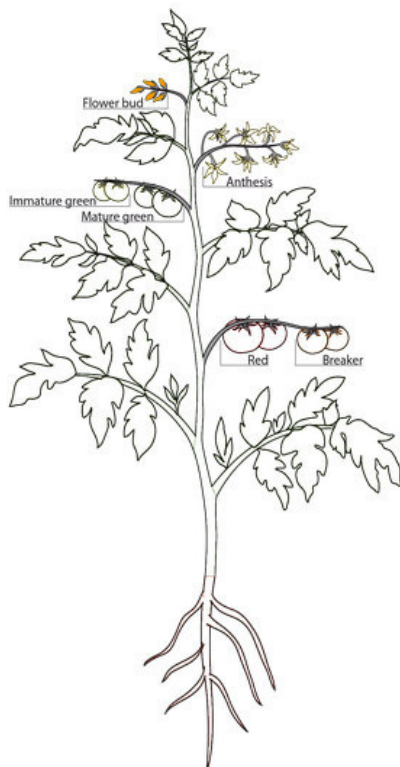
**Biotic interactions**

c) Solyc04g081640

Expression of "Solyc04g081640" in different organs and global conditions in *S. lycopersicum*



d)

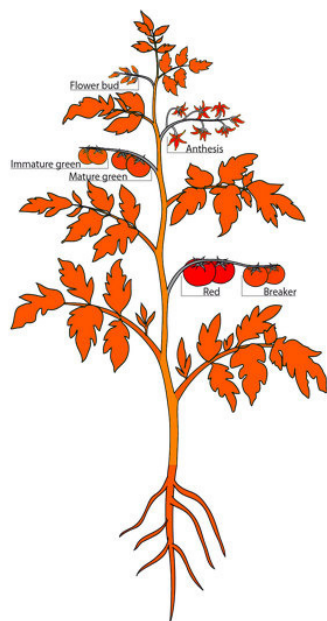
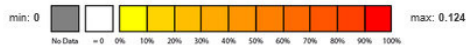


Seedling
Fruit set
Fruit tissues
Fruit development
Fruit parts
Hormone treatment: Fruit
Hormone treatment: Vegetative tissues
Biotic interactions



e) Solyc04g081650

Expression of "Solyc04g081650" in different organs and global conditions in *S. lycopersicum*



**Seedling**

**Fruit set**

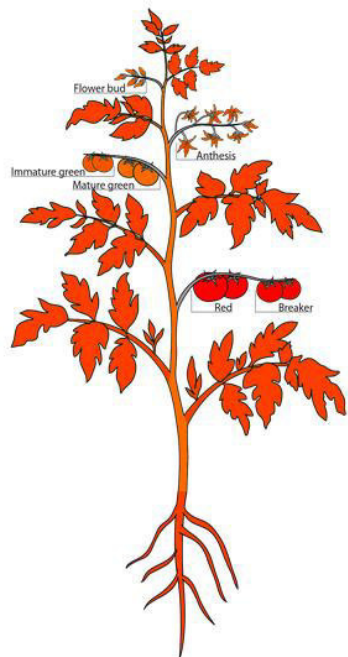
**Fruit tissues**

**Fruit development**

**Fruit parts**

f) Solyc04g081660

Expression of "Solyc04g081660" in different organs and global conditions in *S. lycopersicum*



**Seedling**

**Fruit set**

**Fruit tissues**

**Fruit development**

**Fruit parts**

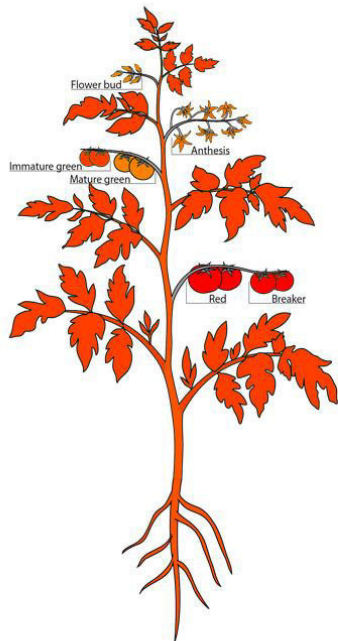
**Hormone treatment: Fruit**

**Hormone treatment: Vegetative tissues**

**Biotic interactions**

g) Solyc04g081670

Expression of "Solyc04g081670" in different organs and global conditions in *S. lycopersicum*



**Seedling**

**Fruit set**

**Fruit tissues**

**Fruit development**

**Fruit parts**

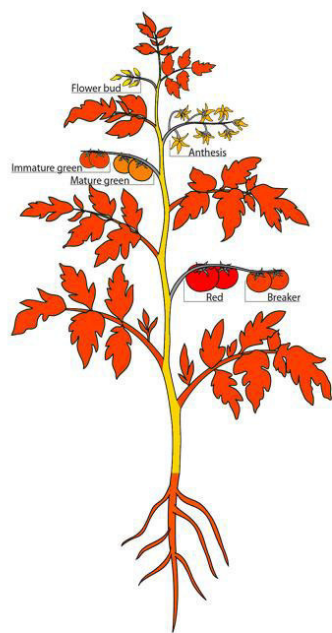
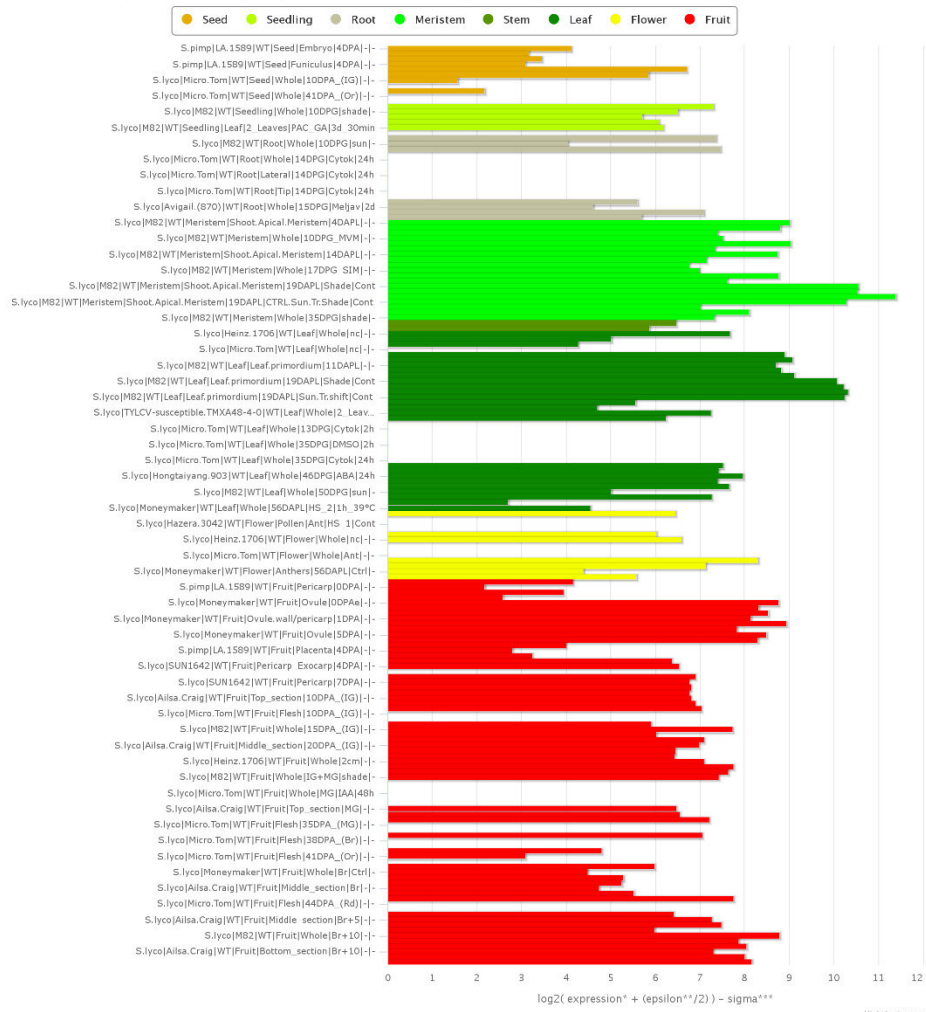
**Hormone treatment: Fruit**

**Hormone treatment: Vegetative tissues**

**Biotic interactions**

## h) Solyc04g081680

Expression of "Solyc04g081680" in different organs and global conditions in *S. lycopersicon*



**Seedling**

**Fruit set**

**Fruit tissues**

**Fruit development**

**Fruit parts**

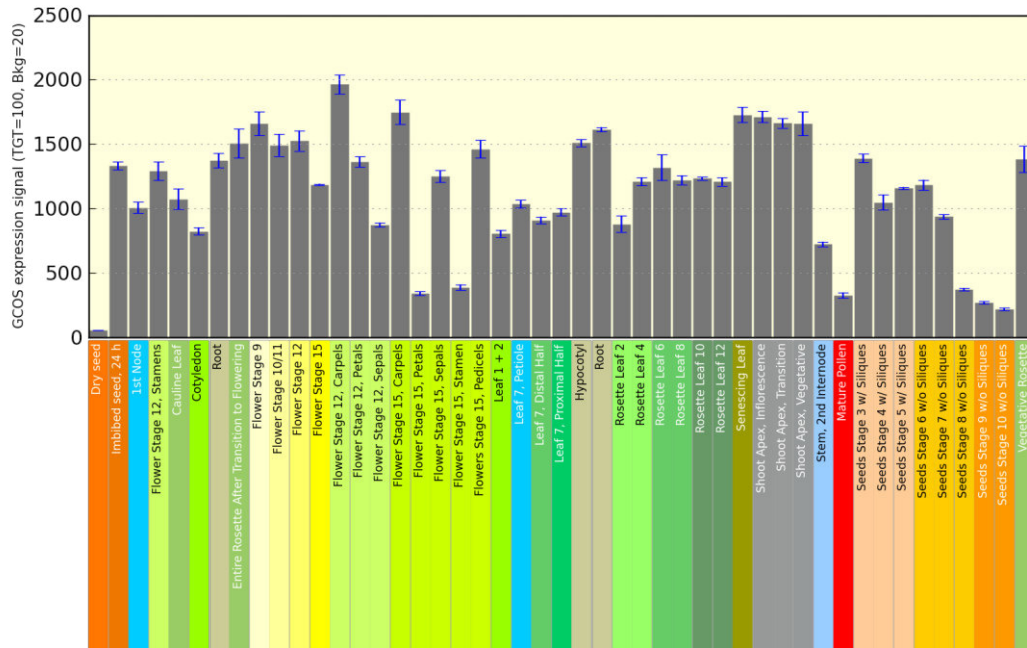
**Hormone treatment: Fruit**

**Hormone treatment: Vegetative tissues**

**Biotic interactions**

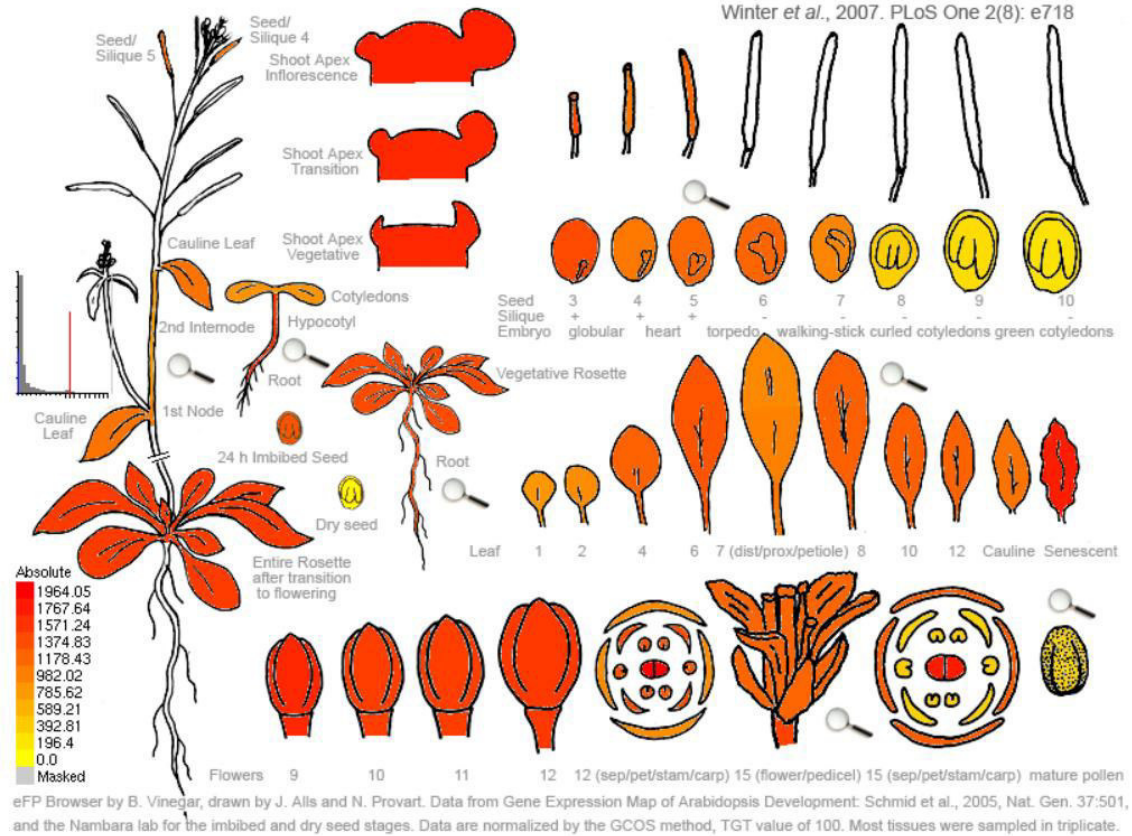


i) At4g24190 (Solyc04g081630 and Solyc04g081640 *Arabidopsis thaliana* homolog)



At4g24190 254166\_at *AtHsp90-7*

Arabidopsis eFP Browser at bar.utoronto.ca  
Winter et al., 2007. PLoS One 2(8): e718



**Annex 8 – Alignment of the Solyc04g081640 gene proximal promoter region from the *S. lycopersicum* var. “Moneymaker” (MM) and *S. lycopersicum* IL 4-4 PP.**

```
#####
# Program: needle
# Rundate: Mon 3 Jul 2017 19:22:30
# Commandline: needle
# -auto
# -stdout
# -asequence emboss_needle-I20170703-192229-0336-44259408-oy.aupfile
# -bsequence emboss_needle-I20170703-192229-0336-44259408-oy.bupfile
# -datafile EDNAFULL
# -gapopen 10.0
# -gapextend 0.5
# -endopen 10.0
# -endextend 0.5
# -aformat3 pair
# -snuceotide1
# -snuceotide2
# Align_format: pair
# Report_file: stdout
#####

#=====#
# Aligned_sequences: 2
# 1: MM_SL4g40Prom(-strand_same_as_the_cds)
# 2: PP_SL4g40Prom_(-strand_same_as_the_cds)
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1574
# Identity: 1546/1574 (98.2%)
# Similarity: 1546/1574 (98.2%)
# Gaps: 22/1574 ( 1.4%)
# Score: 7676.0
#=====#

MM_SL4g40Prom 1 AGAAACATAAGTGTGCCCAAGCTCTGGGGGAAGGGGAGGGTGTGAGT 50
|
PP_SL4g40Prom 1 AGAAACATAAGTGTGCCCAAGCTCTGGGGGAAGGGGAGGGTGTGAGT 50

MM_SL4g40Prom 51 TCTTTTGCTACTGAGAAAACCTCGTTTATGGTTTTTCATGTCATGAACAT 100
|
PP_SL4g40Prom 51 TCTTTTGCTACTGAGAAAACCTCGTTTATGGTTTTTCATGTCATGAACAT 100

MM_SL4g40Prom 101 TGATGTTGTTGACATTATTGGAATGTGAAATCTTGCTTTTCATACTTTTT 150
|
PP_SL4g40Prom 101 TGATGTTGTTGACATTATTGGAATGTGAAATCTTGCTTTTCATACTTTTT 150

MM_SL4g40Prom 151 CTTGAATACATTTACAAGTTCAGTTTCTTCTCAGCAGACATTAAC TCAA 200
|
PP_SL4g40Prom 151 CTTGAATACATTTACAAGTTCAGTTTCTTCTCAGCAGACATTAAC TCAA 200

MM_SL4g40Prom 201 ACTCGGTCTATTTTTAAGTTCAACCCAAC TATTATATTAGGATAAATATG 250
|
PP_SL4g40Prom 201 ACTCGGTCTATTTTTAAGTTCAACCCAAC TATTATATTAGGATAAATATG 250

MM_SL4g40Prom 251 TAAAACCCAAAATGACTCCTGAGAACTTATCAAAATTTGAGCAAAATAG 300
|
PP_SL4g40Prom 251 TAAAACCCAAAATGACTCCTGAGAACTTATCAAAATTTGAGCAAAATAG 300

MM_SL4g40Prom 301 ATACACTAACCAAAATATAATATAACAAAACAAAATAGTCTCCGCTTTTT 350
|
PP_SL4g40Prom 301 ATACACTAACCAAAATATAATATAACAAAACAAAATAGTCTCCGCTTTTT 350

MM_SL4g40Prom 351 GCTTTTCTATAATCTTGGAGCTTCCAATTTTCGAATATGGATATGATTC 400
|
PP_SL4g40Prom 351 GCTTTTCTATAATCTTGGAGCTTCCAATTTTCGAATATGGATATGATTC 400

MM_SL4g40Prom 401 TATATTTATTTTCTATCTATAATTTATATAATAAAGTCTTGGAATGATG 450
|
PP_SL4g40Prom 401 TATATTTATTTTCTATCTATAA TCTATAATAAAGTCTTGGAATGATG 450

MM_SL4g40Prom 451 TGTGTGTTTTATTTCTTCTATTTTTTGTGTGTTTCTTGTTTACTAGAC 500
|
PP_SL4g40Prom 451 TGTGTGTTTTATTTCTTCTATTTTTTGTGTGTTTCTTGTTTACTAGAC 500

MM_SL4g40Prom 501 TTCACCAATCTAATTTTTTTTTATTTTTTTGTGTCATTATCGTAAGTATGTA 550
|
PP_SL4g40Prom 501 TTCACCAATCTAATTTTTTTTTATTTTTTTGTGTCATTATCGTAAGTATGTA 550

MM_SL4g40Prom 551 TTGTTATGTGGCTCAAATATGGTGTGACTTTTATTGGAAGTGTAAAGTT 600
|
PP_SL4g40Prom 551 TTGTTATGTGGCTCAAATATGGTGTGACTTTTATTGGAAGTGTAAAGTT 600
```

MM_SL4g40Prom	601	AAAAA-----ATTATCGCTCCAATTGGAATTGATTACCGATGGAAGT	643
PP_SL4g40Prom	601	AAAAATAACCGATAATAGCTCCAACCTGGAATTGATTACCGATGGAAGT	650
MM_SL4g40Prom	644	TTCAAAGTAATCGATTTCGAAACATTACAATGTGCTTCTCATTGGTGT	693
PP_SL4g40Prom	651	TTCAAAGTAATCGATTTCGAAACATTACAATGTGCTTCTCATTGGTGT	700
MM_SL4g40Prom	694	CTTGATGGTCATTACAAAATATTACCACCTTCACACGATTACATAAAAT	743
PP_SL4g40Prom	701	CTTGATGGTCATTACAAAATATTACCACCTTCACACGATTACATAAAAT	750
MM_SL4g40Prom	744	TTCAATTGTTCTAACGAATCTATAGAAATTCACATAGAACACACATGTCC	793
PP_SL4g40Prom	751	TTCAATTGTTCTAACGAATCTATAGAAATTCACATAGAACACACATGTCC	800
MM_SL4g40Prom	794	GGTGATTGATATATTTATATGAGTGAACAGATGTGCATGAGAAATAGCCT	843
PP_SL4g40Prom	801	GGTGATTGATATATTTATATGAGTGAACAGATGTGCATGAGAAATAGCCT	850
MM_SL4g40Prom	844	AATCAAAATAAATTGATTTTGTAACTATTAATATATAAATTATAGGGG	893
PP_SL4g40Prom	851	AATCAAAATAAATTGATTTTGTAACTATTAATATATAAATTATAGGGG	900
MM_SL4g40Prom	894	AATATTTGGAAGATGATTTTGTAGGTTGTTTGTAGACTTTTAGGATTA	943
PP_SL4g40Prom	901	AATATTTGGAAGATGATTTTGTAGGTTGTTTGTAGACTTTTAGGATTA	950
MM_SL4g40Prom	944	AAATCCAACAATAAATATGAAAAAAAGTGTAGTAGCTACGAGAAAAATA	993
PP_SL4g40Prom	951	AAATCCAACAATAAATATGAAAAAAAGTGTAGTAGCTACGAGAAAAATA	1000
MM_SL4g40Prom	994	CAATGCTAACAAAGTTGGTCATATTTACCAGAAGAATTATTATTCAGAA	1043
PP_SL4g40Prom	1001	CAATGCTAACAAAGTTGGTCATATTTACCAGAAGAATTATTATTCAGAA	1050
MM_SL4g40Prom	1044	TGTTGGTAAAAAAGAGGTAATTAACGGGAAAATTGATAGTGCAAAAT	1093
PP_SL4g40Prom	1051	TGTTGGTAAAAAAGAGGTAATTAACGGGAAAATTGATAGCGCAAAAT	1100
MM_SL4g40Prom	1094	AAAATAGACGCTACCCGACAAAGGGAGAATATCTAAGGGAAAATTCAT	1143
PP_SL4g40Prom	1101	AAAATAGACGCTACCCGACAAAGGGAGAATATCTAAGGGAAAATTCAT	1150
MM_SL4g40Prom	1144	ATATGGCAAACCTAATTGATTAATAAACATTATATGGGTATAGTTTACCT	1193
PP_SL4g40Prom	1151	ATATGGCAAACCTAATTGATTAATAAACATTATATGGGTATAGTTTACCT	1200
MM_SL4g40Prom	1194	AATTACATTCTATGCGTATAGTTTGGTTATTAATAAGACGCTACCCGAC	1243
PP_SL4g40Prom	1201	AATTACATTCTATGCGTATAGTTTGGTTATTAATAAGACGCTACCCGAC	1250
MM_SL4g40Prom	1244	AAAGGAGAATATTTCTAACGGGCATTTAAGTAATATATTCAAAGTTAGCA	1293
PP_SL4g40Prom	1251	AAAGGAGAATATTTCTAACGGGCATTTAAGTAATATATTCAAAGTTAGCA	1300
MM_SL4g40Prom	1294	ATAAAAAATAAATAAATAAATAAATAGCGATTGTTTCTTCCCGGCATACTGT	1343
PP_SL4g40Prom	1301	ATAAAAAATAAATAAATAAATAAATAGCGATTGTTTCTTCCCGGCATACTGT	1350
MM_SL4g40Prom	1344	ACGACACATTTTGTGCTCCTC-----CCCTTCCCTTCCCT	1378
PP_SL4g40Prom	1351	ACGACACATTTTGTGCTCCTC-----CCCTTCCCTTCCCTTCCCTTCCCT	1400
MM_SL4g40Prom	1379	TCCCCAACTCAAAAAACAGAGGAAGAACACACCAACCATTTGGTTGGAA	1428
PP_SL4g40Prom	1401	TCCCCAACTCAAAAAACAGAGGAAGAACACACCAACCATTTGGTTGGAA	1450
MM_SL4g40Prom	1429	AAACACATTACACGGCCAAACATTTTTTCTTCTTACCCGACTCATACAA	1478
PP_SL4g40Prom	1451	AAACACATTACACGGCCAAACATTTTTTCTTCTTACCCGACTCATACAA	1500
MM_SL4g40Prom	1479	AACCCAAAGTTTAAAAATGTCGACTGTTATGGGGTTGTAAAGATAATGAA	1528
PP_SL4g40Prom	1501	AACCCAAAGTTTAAAAATGTCGACTGTTATGGGGTTGTAAAGATAATGAA	1550
MM_SL4g40Prom	1529	ACAACACTTCACGTCTGAAATGGA	1552
PP_SL4g40Prom	1551	ACAACACTTCACGTCTGAAATGGA	1574

#-----  
#-----