UNDERSTANDING THE GENETIC AND MORPHOLOGICAL BASIS OF BUSHY ROOT AND BIFURICATE, TWO MUTATIONS AFFECTING PLANT ARCHITECTURE IN SOLANUM LYPERSICUM L.

SCHOOL OF WATER, ENERGY and ENVIRONMENT

PhD in ENVIRONMENT and AGRIFOOD
Academic Year: 2013 - 2017

Supervisors:

Prof Andrew Thompson
Dr Fady Mohareb

March, 2017
Understanding the genetic and morphological basis of bushy root and bifurcate, two mutations affecting plant architecture in Solanum lycopersicum L.

Supervisors:

Prof Andrew Thompson
Dr Fady Mohareb

March, 2017
ABSTRACT

The classical ethyl methanesulfonate (EMS) tomato mutant *bushy root* (*brt*) was studied using a homozygous near isogenic line (*brt*NIL) in the Micro-Tom (MT) genetic background. The mutation has a pleiotropic phenotype comprising slow seedling development, which may be a consequence of a maternally-inherited small seed phenotype, and a more compact, smaller but not bushier, root phenotype. The number of lateral roots, total root length and taproot size are all smaller in *brt*NIL than the WT. The *BRT* locus was mapped to a 137 kbp region containing 9 candidate genes on chr 12; an InDel in the promoter region of *Solyc12g014590* – containing two highly conserved pirin domains (Pirin_C and Pirin), was detected. Different expression patterns were confirmed by transcriptomic results, supporting *Solyc12g014590* as the gene responsible for the *brt* phenotype.

A naturally occurring recessive mutant named *bifuricate* (*bif*) shows an increase in inflorescence (truss) branching in comparison to the wild type (WT) control line, LAM183. In addition, the number of flowers per truss was 235% higher in *bif* plants than WT. Low temperature is known to increase truss branching, and so a four day low temperature treatment was applied and it was demonstrated that flowering increased significantly more in *bif* than in LAM183. The *BIF* locus was mapped to a 2.01 Mbp interval of chromosome 12 containing 53 genes. All coding region polymorphisms in the interval were surveyed, and two genes *Solyc12g019420* (a BTB/TAZ transcription factor) and *Solyc12g019460* (a MAP kinase) contained one stop codon predicted to disrupt gene function; both genes are excellent candidates for inflorescence branching control based on literature evidence. A newly developed introgression browser was used to demonstrate that the origin of the *bif* mutant haplotype is *Solanum galapagense*.

Keywords: branching; BTB-POZ BTB-TAZ; bushiness; EMS-mutant; low temperature; MAP kinase; root architecture; root branching; seed area; *Solanum galapagense*; transcriptomics; truss branching; Ulp1.
Mãe, muito obrigado por acreditar em mim.

Sinto sua falta todos os dias.
ACKNOWLEDGEMENTS

Firstly, I would like to express my profound gratitude to Prof Andrew Thompson, for all the advice and help during my journey. Thank you, Prof Thompson. Your support meant more than you will ever know, and I will never be able to express how grateful I am for this opportunity.

I would also like to thank Dr Fady Mohareb, as my second supervisor. Your support and kind words (especially during some panicking episodes in these last few months) were invaluable.

I am constantly complaining about the weather, in Cranfield, but I had a great time on the last 42 months. Surely, this would not be possible without the amazing people that I have known and learnt so much from.

In our research group, I would like to specially thank Zoltan, who started his journey in Cranfield with me, and taught me about molecular biology and so much more (e.g. good words in English and French) and Pedro, for helping me through one of the hardest moments of my life. Guys, in the last few months, we have shared much more than language, lab work, and weekend rotas. I am very grateful for your friendship and for your support. I would also like to thanks Tom, for always being there for me – especially when I had endless problems with bioinformatics. My heartfelt thanks to Jane, Jake, Melody and Emmanuel; I am not sure if I would be able to present this thesis if not by your support and love.

I am very grateful for the support provided by the staff during my time here. I would like to specially thank Zoe, Angel, Carol (and Dominick), Maria Carmen, Emma, S., Elsa, Sharon, Debbie, Adrienne and Senita.

Senita, I am not only thankful for your help with my IT problems, I am very grateful for your friendship and support – especially in these last final crazy weeks. You are the best and I am sure you will love Brazil.

During my time in Cranfield, I met special people. Some have become dear friends and will be with me for the rest of my life.
Alicia Rodriguez Sixtos Higuera Simoes..., I am so glad that we started our PhD at the same day/time. Your friendship is special and means the world for me. Alicinha, thanks for your unconditionally love and friendship through the brightest and darkest time. I would also like to thank you for loving my Baileys and brigadeiros, you are a hero.

Daniela, Pieter, Celine and Jolien, Thanks for the amazing time and trips. And, as promised, I would like to especially thank Daniela e Pieter for bringing me Kebab during my long nights in the glasshouse, and helping me when I was drowning tomatoes seedlings. You are the best and will be always be part of me.

I would also like to thank the Portuguese gang Joana: Francisco, Andre, Anish and Pablo; my friends from the water department; The Spanish gang, in special Esther, Fernando, Maria (ballena), Antonio, Juani, Francisco and Angela. Also would like to thank the support and love from the Mexicans that I met Cranfield, in particular, Citlali, Omar, Octavio, Victoria, Kazuo, Ali, Francis, Juan Carlos, Georgina, Silvia and Ricardo. Thanks guys!

I would also like to thank my fellow Brazilians in Cranfield, Tiago Matos, Ingrid, Guilherme, Vinicius, Renan and in special, Thiago Castro. Obrigado.

Thiago, thanks for the long chats over the phone, your everlasting support and your warm friendship.

Back in Brazil, I want to thank my friends and family; who did everything to support my decisions, no matter how crazy they were. Thank you all very much.

Specially, Silvia and Joyce who are much more than friends; are my sisters and endured me for hours on Skype and here in England! Silvia, thanks for the helping with my endless genotyping experiments. Joyce, thanks for sharing with me your experience and beers in Chester, I will always root for your success. You are amazing, and I am grateful for having you both in my life.

Tati and Glenda, your personality and advice were essential during the last years. Thank you for every laugh and for making every moment special.
Esther and Leonardo, Thanks for your support and help with the project. It was
great to work with you both.

Judia, Sonia, Gaby and Mayara thanks for your support. I love you girls.

I am very grateful for siblings; Regiane, Renato, Junior, Fernanda and Thiago. I
am especially grateful for my loving and caring sister-in-law Sandra and my cute
nephew Bruno, who contributed so much in my PhD journey. Junior, thanks for
your love.

Thiago, you are a part of me. Thank you for everything, including your
astonishing math skills and your support. I love you very much.

It was a long and very arduous journey; I would not be able to conclude the PhD
without the support of my family. In special, My dad, dinda, Andressa, Amanda,
Conceicao, Josiane, Mariana, Carolina, Iaura, Durvalino, Percilia, Lindomar,
Laine, Wagner, Regiane and Jefferson.

I would like to thank and name every member of my family and friends who
helped me through my time here. But, I would require 50 extra pages – so I
hope you know how much you have contributed for my growth as a human
being and as a researcher. What I am today is the result of my contact with you.
So, thanks for your help, comfort and support.

Lastly, I would like to thank The Queen, Elizabeth II. Thank you for letting me in
the country. It is very much appreciated.
# TABLE OF CONTENTS

ABSTRACT ......................................................................................................................... i  
ACKNOWLEDGEMENTS ....................................................................................................... iii  
LIST OF FIGURES ................................................................................................................ x  
LIST OF TABLES .................................................................................................................. xi  
LIST OF EQUATIONS .......................................................................................................... xii  
LIST OF ABBREVIATIONS ................................................................................................. xiii  

## 1 GENERAL INTRODUCTION ......................................................................................... 18  
1.1 Plant plasticity ............................................................................................................. 19  
   1.1.1 Sink and source association ............................................................................... 20  
1.2 Root architecture and root classification ..................................................................... 23  
   1.2.1 Classification systems in roots .......................................................................... 23  
   1.2.2 The molecular control of the root architecture .................................................. 25  
1.3 Shoot development and architecture .......................................................................... 29  
   1.3.1 The molecular basis of the shoot control ............................................................ 31  
   1.3.2 Shoot architecture in plant domestication ............................................................ 38  
1.4 Plant domestication in tomato: wild species as germplasm source ......................... 39  
1.5 Technologies for genomics ........................................................................................... 41  
   1.5.1 Sequencing tools and approaches ..................................................................... 41  
   1.5.2 Genome sequencing and genotyping techniques ............................................... 43  
   1.5.3 CRISPR-Cas9 .................................................................................................... 46  
   1.5.4 Bioinformatics' tools for data analysis ............................................................... 47  
   1.5.5 Reference genomes ......................................................................................... 50  
1.6 Aim ............................................................................................................................... 51  
   1.6.1 Objectives ........................................................................................................ 51  

## 2 FINE MAPPING AND PHENOTYPIC CHARACTERISATION OF THE  
TOMATO BUSHY ROOT MUTATION ON CHROMOSOME 12 ........................................ 53  
2.1 INTRODUCTION ......................................................................................................... 55  
2.2 MATERIALS AND METHODS ..................................................................................... 59  
   2.2.1 Plant material .................................................................................................... 59  
   2.2.2 Seed extraction ................................................................................................. 59  
   2.2.3 Plant growth ...................................................................................................... 60  
   2.2.4 Root phenotyping ............................................................................................. 60  
   2.2.5 Seed phenotyping ............................................................................................. 61  
   2.2.6 NGS genomic data generation and analysis .................................................... 61  
   2.2.7 DNA Extraction for genotyping ....................................................................... 62  
   2.2.8 Genotyping by polymerase chain reaction (PCR)-based markers .................... 62  
   2.2.9 Statistical analysis ............................................................................................. 64  
   2.2.10 RNA-seq data generation and analysis ............................................................. 65
2.3 RESULTS ............................................................................................................. 66
  2.3.1 Root development and characterisation ......................................................... 66
  2.3.2 Bushiness ...................................................................................................... 66
  2.3.3 Seed area and maternal inheritance .............................................................. 68
  2.3.4 Resequencing of brtNIL and MT and delimiting the introgression ... 71
  2.3.5 Recombinants selection and mapping interval ............................................. 74
  2.3.6 SNP data from resequenced lines in fine-mapping ..................................... 76
  2.3.7 Transcriptomics ............................................................................................ 77
  2.3.8 Candidate gene ............................................................................................ 77
2.4 DISCUSSION ...................................................................................................... 78
2.5 ACKNOWLEDGEMENTS .................................................................................. 83

3 IDENTIFICATION AND CHARACTERISATION OF BIFURICATE, A NOVEL LOCUS ON CHROMOSOME 12 CONTROLLING TRUSS BRANCHING AND FLOWER NUMBER IN TOMATO ........................................ 84
  3.1 INTRODUCTION ............................................................................................... 86
  3.2 MATERIAL AND METHODS .......................................................................... 89
    3.2.1 Plant material .............................................................................................. 89
    3.2.2 Plant growth ............................................................................................... 89
    3.2.3 Cold Experiment ........................................................................................ 90
    3.2.4 DNA extraction, NGS genomic data generation and variant calling . 91
    3.2.5 SolCaps genotyping ..................................................................................... 93
    3.2.6 Linkage test ................................................................................................ 95
    3.2.7 DNA extraction for genotyping ................................................................. 95
    3.2.8 Genotyping by polymerase chain reaction (PCR)-based markers... 96
    3.2.9 Statistical analysis ....................................................................................... 97
    3.2.10 Similarity map analysis ............................................................................ 97
  3.3 RESULTS .......................................................................................................... 98
    3.3.1 The bifuricate phenotype: initial characterization and inheritance ... 98
    3.3.2 Truss development and characterisation in the contrasting inbred lines...... 98
    3.3.3 Environmental interactions – Cold effect .............................................. 101
    3.3.4 The BIFURICATE gene: Map-based cloning .......................................... 104
    3.3.5 Resequencing of bif and LAM183 inbred lines ....................................... 104
    3.3.6 High resolution gene mapping ............................................................... 108
    3.3.7 Candidate gene analyses .......................................................................... 109
    3.3.8 Germplasm origin of the genomic region encompassing the BIF locus .................................................................................................................. 109
  3.4 DISCUSSION .................................................................................................... 117
  3.5 ACKNOWLEDGEMENTS ................................................................................. 122

4 GENERAL DISCUSSION and FUTURE WORK ................................................. 123
REFERENCES ......................................................................................................................... 125

APPENDICES .......................................................................................................................... 170
Appendix A : .......................................................................................................................... 170
Method for DNA extraction with Chelex 100 in 96-well format ...................................... 170
Appendix B Transcriptomics data ..................................................................................... 172
Appendix C ............................................................................................................................. 173
Appendix D ............................................................................................................................. 174
LIST OF FIGURES

Figure 1: Examples of different root systems and architecture .......................... 24
Figure 2: Phosphate homeostasis regulated by miR399 and PHO2 ...................... 27
Figure 3: Hormonal control of the NO₃⁻ assimilation pathway and transport ....... 30
Figure 4: Inflorescence development in tomato ............................................ 33
Figure 5: Network of genes involved with meristem fate in tomato inflorescences .......................................................... 36
Figure 6: Genome editing using CRISPR-Cas9 ............................................. 48
Figure 7: Seeds pedigree ............................................................................. 59
Figure 8: Root development and taproot profile in brtNIL and MT ................. 67
Figure 9: Root system architecture .............................................................. 68
Figure 10: Seed area profile ................................................................. 69
Figure 11: Genome-wide SNPs uniquely discovered in bushy root (brtNIL) and MT ................................................................. 72
Figure 12: Genome-wide unique InDels discovered in bushy root (brtNIL) and MT ................................................................................. 73
Figure 13: SNP discovery in bushy root (brtNIL) and Micro-Tom (MT) across chromosome 12 ........................................................................... 74
Figure 14: Seeds pedigree ................................................................. 90
Figure 15: Transfer points ............................................................................ 92
Figure 16: Characterization of the first and second truss of LAM183 (bif*) and bifuricate (bif) inbred line plants ....................................................... 99
Figure 17: Floral production in the F₂ population ............................................. 100
Figure 18: Cold effects on both parental lines ............................................. 103
Figure 19: Genome-wide SNPs uniquely discovered in bifuricate and LAM183 .................................................................................. 106
Figure 20: Genome-wide unique InDels discovered in bifuricate and LAM183 .................................................................................. 107
Figure 21: SNP discovery in bifurcate (bif) and LAM183 across chromosome 12 .................................................................................. 108
Figure 22: Similarity tree based on the SNPs in the defined mapping interval (2.01 Mbp) on chromosome 12 ...................................................... 110
**Figure 23:** SNP similarities map in *bifuricate*, LAM183, Heinz 1706 and other genomes.

**Figure 24:** High contrast SNP similarities map in *bifuricate*, LAM183, Heinz 1706 and 28 other lines.

**Figure 25:** SNP heat map in the mapping interval.

**Figure 26:** SNP heat map in the *bifuricate* (*bif*) gene mapping interval using the lines that are more similar to *bif* locus.

**Figure 27:** Origin of the *Solanum galapagense* accessions.
LIST OF TABLES

Table 1: Classical Mendelian mutations affecting tomato's root system. ........ 56
Table 2: SNPs used to develop KASP markers on chromosome 12 .......... 63
Table 3: Phenotype characterisation and comparison between Micro-Tom (MT) and bushy root (brtNIL) parental lines. ................................................................. 70
Table 4: Recombinants genotyping and representation.............................. 75
Table 5: Polymorphisms at the mapping interval.. ...................................... 76
Table 6: Gene expression profile in bushy root (brtNIL) and Micro-Tom (MT) . 78
Table 7: Candidate genes for the brtNIL phenotype. ................................ 79
Table 8: SolCap markers used for genotyping........................................... 93
Table 9: SNPs used to develop the KASP markers on chromosome 12 ....... 96
Table 10: Phenotype characterisation and comparison between LAM183 and bituricate (bif) parental lines............................................................... 100
Table 11: Genotyping of recombinants for fine mapping of bif.................. 111
Table 12: Candidate genes for the bif phenotype ..................................... 114
LIST OF EQUATIONS

Equation 1: Mathematical expression used to calculate bushiness.................. 61
LIST OF ABBREVIATIONS

ABA-insensitive4  abi4
ABERRANT SPIKELET AND PANICLE1  ASP1
Abscisic acid  ABA
albescent  alb
Allele-specific PCR  AS-PCR
Aluminium  Al
ANANTHA  AN
APETALA1  AP1
Apical Meristem  AM
Auxin indole-3-acetic  AUX/IAA
Auxin indole-3-butyric  AUX/IBA
Beta carotene  B
bifurcate  bif
BLIND  bl
BRANCHED1  BRC1
Branching Meristem  BM
BRANCHING SILKLESS1  BD1
Brassinosteroids  BRs
Bric-A-Brac/Tramtrack/Broad complex  BTB
Bulk Segregant Analysis  BSA
Burrows-Wheeler Aligner  BWA
bushy root  brt
bushy root- Near Isogenic Line  brtNIL
Charge-Couple Device  CCD
Chromosome  Chr
<table>
<thead>
<tr>
<th>Term</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleaved Amplified Polymorphic Sequence</td>
<td>CAPS</td>
</tr>
<tr>
<td>CLEVATA</td>
<td>CLE</td>
</tr>
<tr>
<td>Clustered Regularly Interspaced Short Palindromic Repeats</td>
<td>CRISPR</td>
</tr>
<tr>
<td>Combinatorial Fluorescence Energy Transfer</td>
<td>CFET</td>
</tr>
<tr>
<td>Common Primer Extension</td>
<td>CPE</td>
</tr>
<tr>
<td>COMPOUND INFLORESCENCE</td>
<td>s</td>
</tr>
<tr>
<td>C-TERMINALLY ENCODED PEPTIDE</td>
<td>CEP</td>
</tr>
<tr>
<td>CUPSHAPED COTYLEDONS</td>
<td>CUC</td>
</tr>
<tr>
<td>CyclinD2</td>
<td>CYCD2</td>
</tr>
<tr>
<td>Cytokinin</td>
<td>CK</td>
</tr>
<tr>
<td>Days After Germination</td>
<td>DAG</td>
</tr>
<tr>
<td>Days After Sowing</td>
<td>DAS</td>
</tr>
<tr>
<td>decreased apical dominance</td>
<td>dad</td>
</tr>
<tr>
<td>Denaturing high-performance liquid chromatography</td>
<td>DHPLC</td>
</tr>
<tr>
<td>DENSE AND ERECT PANICLE1</td>
<td>DEP1</td>
</tr>
<tr>
<td>Deoxynucleotide Triphosphates</td>
<td>dNTPs</td>
</tr>
<tr>
<td>Dicotyledonous</td>
<td>dicot</td>
</tr>
<tr>
<td>dwarf</td>
<td>d</td>
</tr>
<tr>
<td>dwarf root</td>
<td>drt</td>
</tr>
<tr>
<td>Ethyl Methanesulfonate</td>
<td>EMS</td>
</tr>
<tr>
<td>FALSIFLORA</td>
<td>FA</td>
</tr>
<tr>
<td>FAR-RED ELONGATED HYPOCOTYL3</td>
<td>FHY3</td>
</tr>
<tr>
<td>Floral Meristem</td>
<td>FM</td>
</tr>
<tr>
<td>Fold Change</td>
<td>FC</td>
</tr>
<tr>
<td>Food and Agriculture Organization of the United</td>
<td>FAO</td>
</tr>
</tbody>
</table>

xiv
<table>
<thead>
<tr>
<th>Term</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NATIONALS</td>
<td></td>
</tr>
<tr>
<td>FRIZZY PANICLE</td>
<td>FZP</td>
</tr>
<tr>
<td>Fruit weight</td>
<td>FW</td>
</tr>
<tr>
<td>Genetic Modified Organisms</td>
<td>GMOs</td>
</tr>
<tr>
<td>Genome Analysis Tool Kit</td>
<td>GATK</td>
</tr>
<tr>
<td>Genotyping-By-Sequencing</td>
<td>GBS</td>
</tr>
<tr>
<td>Gibberellic acid</td>
<td>GA</td>
</tr>
<tr>
<td>GOBLET</td>
<td>GOB</td>
</tr>
<tr>
<td>High Affinity Transport System</td>
<td>HATS</td>
</tr>
<tr>
<td>High-Throughput</td>
<td>HT</td>
</tr>
<tr>
<td>Homologous Recombination</td>
<td>HR</td>
</tr>
<tr>
<td>HT-Next Generation Sequencing</td>
<td>HT-NGS</td>
</tr>
<tr>
<td>INDETERMINATE FLORAL APEX1</td>
<td>IFA1</td>
</tr>
<tr>
<td>INDETERMINATE SPIKELE1</td>
<td>IDS1</td>
</tr>
<tr>
<td>Inflorescence Meristem</td>
<td>IM</td>
</tr>
<tr>
<td>Inorganic phosphate</td>
<td>P$_i$</td>
</tr>
<tr>
<td>Insertion and Deletions</td>
<td>InDel</td>
</tr>
<tr>
<td>Iron</td>
<td>Fe</td>
</tr>
<tr>
<td>ISOPENTENYLTRANSFERASE</td>
<td>IPT</td>
</tr>
<tr>
<td>JOINTLESS</td>
<td>j</td>
</tr>
<tr>
<td>Kompetitive Allele Specific PCR</td>
<td>KASP</td>
</tr>
<tr>
<td>LANCELOTA</td>
<td>LA</td>
</tr>
<tr>
<td>LATERAL ORGAN BOUNDARIES1</td>
<td>LOB1</td>
</tr>
<tr>
<td>LATERAL SUPPRESSOR</td>
<td>Ls</td>
</tr>
<tr>
<td>LEAFY</td>
<td>LFY</td>
</tr>
<tr>
<td>LEAFY HULL STERILE1</td>
<td>LHS1</td>
</tr>
<tr>
<td>Term</td>
<td>Abbreviation</td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>MALDI- time of flight mass spectrometry</td>
<td>MALDI-TOF MS</td>
</tr>
<tr>
<td>Matrix-assisted laser desorption/ionization</td>
<td>MALDI</td>
</tr>
<tr>
<td>MicroRNA</td>
<td>miR</td>
</tr>
<tr>
<td>Micro-Tom</td>
<td>MT</td>
</tr>
<tr>
<td><em>miniature</em></td>
<td>mnt</td>
</tr>
<tr>
<td>mitogen-activated protein kinase</td>
<td>MAP kinase</td>
</tr>
<tr>
<td>Molecular inversion probe</td>
<td>MIP</td>
</tr>
<tr>
<td>Monocotyledons</td>
<td>monocot</td>
</tr>
<tr>
<td><em>more axillary branching</em></td>
<td><em>max1</em></td>
</tr>
<tr>
<td>Mutator-like element</td>
<td>MULE</td>
</tr>
<tr>
<td>Next generation sequencing</td>
<td>NGS</td>
</tr>
<tr>
<td>Phosphite</td>
<td>PO$_3^-$</td>
</tr>
<tr>
<td>NITRATE-REGULATED1</td>
<td>ANR1</td>
</tr>
<tr>
<td>Nitrogen</td>
<td>N</td>
</tr>
<tr>
<td>Non-homologous end joining</td>
<td>NHEJ</td>
</tr>
<tr>
<td><em>P-glycoprotein</em></td>
<td><em>PGP</em></td>
</tr>
<tr>
<td>Phosphorus</td>
<td>P</td>
</tr>
<tr>
<td>PhytochromeB</td>
<td>phyB</td>
</tr>
<tr>
<td>PIN-FORMED1</td>
<td>PIN1</td>
</tr>
<tr>
<td>PINOID</td>
<td>PID</td>
</tr>
<tr>
<td>Polymerase chain reaction</td>
<td>PCR</td>
</tr>
<tr>
<td><em>ramosus</em></td>
<td><em>rms</em></td>
</tr>
<tr>
<td>Reduced-representation sequencing</td>
<td>RRS</td>
</tr>
<tr>
<td><em>REVERSED GERM ORIENTATION</em></td>
<td><em>RGO1</em></td>
</tr>
<tr>
<td>RNA-sequencing</td>
<td>RNA-seq</td>
</tr>
<tr>
<td>Rolling circle amplification</td>
<td>RCA</td>
</tr>
</tbody>
</table>
self-pruning

Shoot apical meristem

Single Molecule Real Time

single-base extension

Single-nucleotide polymorphism

SISTER OF INDETERMINATE SPIKELET1

Spikelet meristem

Strigolactones

SUPERNUMERARY BRACT

tasselssed4

Tasselssed6

TEOSINTE BRANCHED1

TERMINAL FLOWER1

TILLER ANGLE CONTROL

transcriptional adapter zinc finger

True single molecule sequencing

UNIFLORA

UNUSUAL FLOWER ORGAN

Variant Call Format

Variant Effect Predictor

Whole genome resequencing

Wild type

Zero-mode wave-guides

sp

SAM

SMRT

SBE

SNP

SID1

SM

SL

SNB

ts4

Ts6

TB1

TFL1

TAC1

TAZ

tSMS

UF

UFO

VCF

VEP

WGR

WT

ZMWs
1 GENERAL INTRODUCTION

Tomato is a crop with high agricultural value, and the tomato world production has increased about 40% during the last ten years. Tomato cultivation takes place under protection in glasshouses or plastic tunnels and in open fields (Schwarz et al. 2013; FAOSTAT 2014). The cultivation method has a big impact on production efficiency and energy use (Muigai et al. 2003; Kirda et al. 2004; Hatirli et al. 2006; Ozkan et al. 2011). One reason for this improvement in world production is the results of research and breeding, which led to increased fruit quality and yield. According to the Food and Agriculture Organization of the United Nations (FAO), in 2013 the world production was around 163 million tonnes (FAOSTAT 2014). China, India and USA are the biggest tomatoes producers. Brazil is the eighth biggest tomato producer (fresh market) worldwide, with an average production of 4.3 million tonnes per year (FAOSTAT 2014). Around 42% of the world’s production comes from countries with favourable conditions of solar radiation, air temperature and precipitation (Jędrszczyk et al. 2012). Food demand is sharply increasing; in 2100 the world population is predicted to reach 10 billion (DESA 2014), thus the pressure to increase food production is exponentially rising (Schmautz et al. 2016; Suhl et al. 2016; Bernstad et al. 2017).

Solanaceae plants, e.g. tomato, potato, pepper, eggplant and physalis, are some of the crops with the biggest agricultural value for the fresh or industrial markets (Rigano et al. 2013; Vélez et al. 2016; Gebhardt 2016; Garofalo et al. 2017). Due to a highly conserved genome throughout the Solanaceae family (Fernandez et al. 2009), tomato is used as a genetic model fruit crop to study host-pathogen interactions (Ji, J. Scott, et al. 2007), abiotic stress responses (Bahmani & Maali-Amiri 2015), yield (Marincs et al. 2017) and other key biological processes. This research has the potential to be converted to other fruit crops in tropical (banana, avocado, mango and papaya) and temperate (apple, peach and pear) fruits (Pino-Nunes et al. 2009; Lombardi-Crestana et al. 2012; Quinet et al. 2006; J. Wang et al. 2014; Belović et al. 2016). Tomato is also a good model organism due to other important
characteristics such as the short life cycle (12 weeks, Ichihashi & Sinha 2014), easy growth in open fields and greenhouses, relatively small and diploid genome, self-compatibility, availability of a diverse range of sexually compatible wild species (Emmanuel & Levy 2002), suitability for mutagenesis experiments employing DNA damaging agents such as ethyl methanesulfonate EMS (Just et al. 2013; Sikder et al. 2013), X-ray or fast-neutron (Meissner et al. 1997; Huther et al. 2013; Larbat et al. 2014; Kobayashi et al. 2014). In addition, tomato scientists have access to an extensive collection of single-gene mutants, a frequently updated reference genome, a database with thousands of validated single nucleotide polymorphisms (SNPs) and an active research community connecting a network of more than 50 countries (Fernandez et al. 2009; Shirasawa et al. 2010; The Tomato Genome Consortium 2012; Kim et al. 2014; Kobayashi et al. 2014). These resources allowed the publication of a substantial number of studies in different topics, for instance, the genetic basis of tomato. Breeders and geneticists rely on forward and reverse genetics strategies to uncover genes responsible for a particular phenotype. Forward genetics approaches begin with the description of a phenotype determined by Mendelian inheritance of a single gene. On the contrary, reverse genetics is looking for a phenotype in plants carrying mutation(s) for a defined gene(s).

1.1 Plant plasticity

Different and sometimes hostile environments have contributed to the capacity of plants to adapt their development and fitness (Santos et al. 2015; Mohammed et al. 2016). Phenotypic plasticity – differentiation among cells and tissues developed in response to external influences (Givnish 2002), has been studied in relation to the mechanisms of plant development (López-Bucio et al. 2003; Gandour et al. 2013). Phenotypic plasticity has evolved, but occurs regardless of the plant genotype in response to environmental interactions or under stochastic processes (Van Kleunen et al. 2005). This plasticity consists of changes in plant growth and metabolism (Mohammed et al. 2016). For instance, stem elongation represents an advantage to plants in a dense plant stand, with
competition for light, but early elongation also reduces structural stability and requires diversion of resources away from organs with the fundamental functional roles of photosynthesis (leaves) and water and nutrient capture (roots) (Weinig & Delph 2001; Dechaine et al. 2007).

Plants need to be fully able to acquire light, water and ions from their habitat, thus, plant architecture (see section 1.2.1, page 23) and developmental vigour are essential for the adaptation and exploration of available resources in the surrounding environment and exploitation of available resources. In order to best adapt to the surrounding environment, plants need to regulate their development at different levels, e.g. biomass investment in organs during vegetative growth and reproductive development (Vidyadhar et al. 2015; Poorter 2015). The morphology of plants is due to some important allometric relationships that control size and shape of plant organs (Niklas 2006; Busov et al. 2008). Plant size and architecture are complex traits to study because of coordinated growth between distinct organs. Plant development requires a balance between the fixing and use of carbon containing molecules, in other words, a balance is necessary between the “sink” and the “source” of assimilate in the plant.

1.1.1 Sink and source association

A balance between vegetative shoot growth (source) and the growth of sink organs such as truss and root is required for plant fitness. There must also be a balance between shoot demand for water and ions, and the ability of the root system to provide these resources. Therefore, both root and shoot development are intimately linked (Wissuwa et al. 2005; Vercruyssen et al. 2011; Osorio et al. 2014). Plant architecture is a complex trait that depends on several physiological processes regulating growth and development, for instance, nutrient uptake and carbon partitioning (Barthélémy & Caraglio 2007).

The soil nutrient profile influences plant growth and development. Nutrients as nitrogen and phosphorus are considered major limiting factors to
plant growth (Umehara et al. 2008; Yoneyama et al. 2012; Bianco et al. 2015; Mohammed et al. 2016). Nitrogen (N) is extensively used in fertilizers to ensure high yields, which has caused environmental problems such as eutrophication. The role of N in plant metabolism has been extensively studied and it well characterized (Nagel et al. 2001; Bénard et al. 2009; Yoneyama et al. 2012; Luo et al. 2015) in order to make its use more efficient.

Phosphorus (P) is also crucial for plant development, but is not readily available in many farming sites (Vejchasarn et al. 2016). Thus, P is also used as a fertilizer to increase P availability. It has been reported that lack of regulation and inspection can lead to severe excess of P fertilization, which may contribute to environmental degradation (Cordell et al. 2009; Vejchasarn et al. 2016). Plants with P deficiency are common. In addition, the interaction between P, iron (Fe) and aluminium (Al) makes the ions absorption by the roots harder (Zhou et al. 2009). Over the last decades, different studies were conducted to characterise the plant response to P starvation (Lambers et al. 2015). Plants exposed to environments with low P availability show accumulation of sucrose in the phloem (Hammond & White 2011; Smet et al. 2012). This strategy is commonly adopted to increase carbon resources in the root, which will lead to an enlargement of the root to better explore the surroundings – generating phenotyping plasticity. It has been reported (although, it is not a consensus) that some specific traits in the root system like taproot diameter and length, and root branching are affected by low levels of P (Hammond & White 2011; Niu et al. 2013; Nadira et al. 2014; Li-xiang & Dan 2014).

Underlying phenotype plasticity, i.e. phenotype variation due to external factors, are many mechanistic pathways which rely on gene products whose activity is modulated by the environment to create the variation. Branching of plant roots and shoots is variable and related to genes affecting hormone metabolism, regulatory factors and the cell cycle (Kaul et al., 2000; Chae et al., 2013). All aspects in plant physiology and development are associated with hormone metabolism, for example elongation growth is controlled by auxin
transport and this is affected by several different genes including *P-glycoprotein* (*PGP*) (Higashide et al. 2014); branch outgrowth is affected by strigolactones that act through pathways influenced by genes such as *more axillary branching1* (*max1*) (Martín-Trillo et al. 2011) – see section 1.3.1.2, page 29; also dwarfism and round leaves in plant with deficiency in the production/signalling of the brassinosteroid (BRs) hormone (Shannon & Meeks-Wagner 1991; Busov et al. 2008; Wang & Wu 2015) – see section 1.2.2.2, page 26.

Transcription factors also control plant organ, growth and development – e.g. negative regulation of fruit size affected by the *FW2.2* gene (Cong & Tanksley 2006); reduced branching of lateral root regulated by GRAS and the *LATERAL SUPPRESSOR* (*Ls*) gene (Lozano et al. 2009; Busch et al. 2011); and relation with TCP-domain affected by the *LANCELOTA* (*LA*) gene (Kaul et al., 2000; Busov et al., 2008). The amount of cells is proportional to the plant tissue volume and consequently to the organ size, but the genetic control, as mentioned before, is not final. As an example, the *CyclinD2* (*CYCD2*) gene is reported to regulate the transition between G1 and S (i.e. phases of the cell cycle - (Calegari 2012). Studies on *CYCD2* overexpression led to an increased growth rate, but the plant final size was the same as the WT (Busov et al., 2008).

Studies on root architecture are started by visual characterization and comparison to established models. Research conducted with roots is not straightforward; roots are mostly underground and cannot be easily observed without disturbing the surrounding soil. Nevertheless, root architecture is extensively studied and investigated due to its important functions – uptake of water (Pop et al., 2011) and ions (Jung and McCouch, 2013), anchorage (Tao et al., 2010) and storage (Ubeda-Tomás et al., 2012).


1.2 Root architecture and root classification

Historically, it was arduous to find agreement between plant morphologists regarding root classification and description (Esau, 1977; Zobel and Waisel, 2010; Gratani, 2014). To improve research involving root architecture classification, the international society for root research (ISRR) recommended a standard nomenclature system (Zobel and Waisel, 2010). Currently, four different root categories are commonly used in tomato root classification – taproot, lateral root (LR), shoot-borne root and basal root. The taproot, also called “primary root”, is the first root to develop, from the radicle. LRs, shoot-borne and basal roots are secondary roots. Lateral root is the term adopted to describe any root branching from another. To assist the classification, branching should be considered and described in association with the “original root” (e.g. third-order lateral of the taproot or second-order lateral of the basal root (Zobel and Waisel, 2010; Esau, 1977). Shoot-borne are originated from the shoot, these roots are also considered as adventitious; however, the term “adventitious roots” is only commonly used by plant anatomists when describing monocotyledonous plants (e.g. maize). Basal roots are originated from the hypocotyl (Zobel and Waisel, 2010). The number of lateral, shoot-borne and basal roots is related with the efficiency of water and ion absorption as more roots surface increase the root area/soil volume ratio (Péret et al., 2009a).

However, the mechanisms controlling root initiation are still not clear due to a large number of factors that can influence the morphogenesis. Root initiation has been studied and reviewed in the last two decades (Laplaze et al., 2007; Laskowski et al., 1995; Lavenus et al., 2013; Péret et al., 2009b; Charlton, 1991; Hochholdinger et al., 2004; Guo et al., 2008). Root architecture, i.e. the result of different developmental processes, is ultimately controlled by genetic and environmental factors.

1.2.1 Classification systems in roots

Even with the changes in root architecture caused by wide ecological niche diversity, an overall scheme was proposed by Fitter (1987) to classify the form of the root system by considering: (i) the balance of primary and adventitious
roots: a large number of adventitious roots all connected to the stem base are produced by some species particularly monocotyledons, e.g. sugar-cane; these systems lack a single dominant axis, differently from dicotyledonous plants with a primary root system; (ii) the degree of branching in primary or adventitious roots systems, as described by several models (Tisdall & Oades 1982; Coutts 1983; Fitter et al. 1991); (iii) the plasticity of branching, for example as influenced by nutrient availability (Fitter 1987), (Fig 1).

**Figure 1:** Examples of different root systems and architecture (available at: http://botany.csdl.tamu.edu/FLORA/Wilson/ftp/veg/roots.gif).

Alternative classification systems were suggested by Cannon (1949) based on the root activity and overall morphology. For example, root systems could be classified as very compact and very shallow or very extensive and very deep (Wahid 2000); or classification was based on root origin and development (Cannon 1949), e.g. primary root system type III, which has a long taproot with short laterals, whereas adventitious root system type VII has adventitious roots clustered and originating from an axis of the shoot. The latter classic system is the most used for root anatomists and soil scientists, because it considers different aspects of the plant plasticity, diversity and environment.
1.2.1.1 Root architecture: development and functions

As briefly mentioned before, dicotyledonous (e.g. beetroot) root systems develop from a primary root, thus, dicot root systems have a primary root system. The model of such system (dicot) is constructed based on the number of basal roots in older plants, and length of the laterals (extension covered) relative to the taproot; the root can be classified as taprooted or diffuse (Zobel 2005). On the case of tomato, most of the root system is developed from basal roots (Stofella 1983).

The root system in monocotyledons (monocots) is developed from the short-lived primary root and nodal roots, i.e. adventitious roots, forming a fibrous root system or adventitious system – are called “fasciculate root system” (Chun et al. 2005). In certain cereals, during early stages of the development, the primary root system is the main source of water and ions; until the adventitious root system is formed and established to assume this role (Klepper 1992). Developmental processes, including production of a new main axis, axial growth, radial growth and root senescence and decay (reviewed by Hodge et al. 2009) also impact the root system architecture.

As already mentioned in this study, the root has three main functions: anchorage, water and ion uptake and storage. But, in fact, it was reported that only 10% and 30% of the root length is involved with the assimilation of ions and water, respectively. Different functions between different classes of roots were reported, for example, seminal roots (early roots developed from the radical) have an important role in water absorption, but assimilate less P than the nodal roots – (Mistrik & Mistrikova 1995).

1.2.2 The molecular control of the root architecture

The root system is affected by a wide range of exogenous factors, including biotic factors such as helminths and pathogenic microorganisms, and abiotic factors such as high salinity and the mechanic impedance imposed by soils (Hodge et al. 2009). These factors cause extreme phenotypic differences in
individuals with the same genotype. Different approaches have been used to understand which genes are involved in environmental responses.

Although the phenotypic plasticity in roots is well established, the genetics (and molecular mechanisms) behind it are still obscure. Lately, published works on Arabidopsis is changing this scene by helping to elucidate the mechanisms of lateral root development (Casimiro et al. 2003), inflorescence regulation (Hanano & Goto 2011) and shoot-root interaction (Widman et al. 2014), for example.

1.2.2.1 Responses to nitrogen and phosphorus

Studies carried out in Arabidopsis proved that high concentrations of N in soil reduce LR and taproot length. On the other hand, plants suffering with nitrogen starvation produce longer LRs (Linkohr et al. 2002), but the LR density is decreased, i.e. less LRs are produced. This response is partly controlled by the MADS box transcription factor NITRATE-REGULATED1 (ANR1) as shown by the observation that anr1 mutants have lower LR density in response to nitrate-rich zones (Zhang & Forde 1998). Root branching, and therefore root system architecture, is affected by N transport by the low- and high-affinity transport system (HATS) (Little et al. 2005). The low affinity transport system works when N is abundant, in contrast to HATS, which works when N is limiting (Glass et al. 2002). In addition, the NRT2.1 gene from the NRT2 family of transporter-like proteins that is required for HATS, is reported to be involved with lateral root initiation (Little et al. 2005; Remans et al. 2006).

Experiments where plants were exposed to high levels (10 mM PO$_3^-$) of inorganic phosphate (P$_i$) showed an increase in taproot length, but decrease LR density and suppress LR elongation. Nevertheless, under low/moderate levels (1 mM – 0.5 mM PO$_3^-$), P$_i$ increases the growth of the primary root (Linkohr et al. 2002). Extremely low levels (0.05 mM PO$_3^-$) of P$_i$ cause a severe inhibition to taproot and LR development (López-Bucio et al. 2003). These effects were caused by a stop in meristematic function due to the lack of response to auxin
(López-Bucio et al. 2002); and were not displayed by plants on the control group, treated without P manipulation into the growth medium. Internal P concentration is important to the root architecture (Williamson et al. 2001). \textit{PHO2} and miR-399 (induced under starvation), play a dominant role in P signalling pathway, they regulate the action of each other (Bari et al. 2006) – Figure 2.

\textbf{Figure 2}: Phosphate homeostasis regulated by miR399 and PHO2. Low P concentration in the shoot results in the action of the transcription factor PHR1. PHR1 is reported to be involved with the induction of miR399 (Bari et al. 2006). Transported by the phloem, miR399 acts at the root, inhibiting the action of PHO2; resulting in the upregulation of Pi transporter genes (Pht), increasing the Pi uptake capacity and translocation via the xylem to the shoot (Adapted from: Scheible et al. 2011).

\textbf{1.2.2.2 Phytohormones}

The exogenous factors mentioned above affect the root architecture mainly by changing the plant hormonal profile, thus, its homeostasis. For instance, mutants in the auxin transport mediator \textit{AXR4} show no response to N-rich-zones (Zhang et al. 1999), and the abscisic acid (ABA)-insensitive \textit{abi4} mutant
does not show any reduction in LR density due to increased levels of N (Signora et al. 2001). This suggests that local responses to N are regulated by auxin and ABA would mediate systemic responses (Osmont et al. 2007).

As already described, the transport and signalling of auxin affect the root system architecture directly. Auxin indole-3-butyric acid (IBA) promotes the formation of adventitious roots, and for this reason, its signalling pathway has been extensively studied and discussed in monocotyledonous plants (Laskowski et al. 2006). Auxin indole-3-acetic acid (IAA) induces LR development when used in the growth media (Laskowski et al. 1995). Auxin is involved in all phases of LR development (i.e. initiation, emergence and elongation) (Liu et al. 2013). Acropetal auxin transporters, in the taproot, promote LR formation by stimulating cell division in the pericycle. Basipetal auxin transporters, in the root tip, also affect the LR formation (Casimiro et al. 2003). In addition, studies with mutants and transformed plants confirmed the function of auxin in LR development – mutants with increased auxin biosynthesis had more branched root systems (Boerjan et al. 1995). Accordingly, mutants with lower auxin production or with an impaired auxin signalling present a small root system, with less branching (Hobbie & Estelle 1995).

The well-established antagonism between cytokinin (CK) and auxin is also present in the root. CK regulates root growth/branching negatively – role confirmed by mutants with impaired cytokinin receptors AHK2 and AHK3, which show a quicker root development and more branched roots (Riefler et al. 2006).

Many auxin signalling mutants are also ethylene-insensitive, thus, ethylene is reported to be involved in some auxin responses (Swarup et al. 2002). In maize and Arabidopsis, moderate levels of ethylene are reported to constrain root growth, affecting the quiescent centre and root cap. It was suggested that, in rice, ethylene might regulate LR emergence due to the breakdown of cortical cells (Laskowski et al. 2006). Also in rice, gibberellic acid (GA) is reported to act with ethylene to stimulate adventitious root development.
In general, GA regulates primary root development and it is required for its growth (Kaneko et al. 2003; Gerald et al. 2006).

Brassinosteroids are reported to stimulate the taproot development at low concentrations, but are inhibitory at higher levels (Osmont et al. 2007). BRs will also affect the LR growth, by regulating auxin transport. Studies on BRs (Goda et al. 2002; Müssig et al. 2002; Nakamura et al. 2003; Nemhauser et al. 2004; Kim et al. 2006; Mouchel et al. 2006) suggests an important interaction between auxin and BR hormone signalling; e.g. the function of several genes involved in root development and architecture are regulated by both BR and auxin (Nakamura et al. 2003; Kim et al. 2006). The most reported function of BRs in root is related with its development (Mouchel et al. 2006); mutants with an impaired BR biosynthesis and signalling show a dwarf root phenotype (Noguchi et al. 1999).

In normal conditions, i.e. when the plant is not under stress, exogenous ABA and CK are reported to restrain root development; ABA reduces LR development and primary root growth (Casimiro et al. 2001), cytokinins reduce LR density (Werner et al. 2003). ABA and AUX/IAA interactions have been also suggested for similar alterations, i.e. reduced LR density, in the root system (Beaudoin et al. 2000; De Smet et al. 2003). However, under hydric stress, ABA has been reported to stimulate root growth by preventing the excess of ethylene production (Sharp & LeNoble 2002). ABA was also reported to be responsible for the arrest of shoot growth and development in drying soil (Sharp & LeNoble 2002) – Figure 3.

1.3 Shoot development and architecture

The shoot systems in angiosperms also display enormous phenotypic variation. This is mainly due to alterations to the truss architecture and branching or differences in the “growth habit” (Bell 1992).

Monopodial plants (e.g. Arabidopsis) present a simple architecture. The apical meristem (AM) is indeterminate during the plant growth and
development; all organs are laterally originated (e.g. leaves). Due to the strict flowering control, monopodial plants show distinct vegetative and reproductive phases (Pnueli et al. 2001).

![Diagram of hormonal control of the NO$_3$ assimilation pathway and transport.](image)

**Figure 3:** Hormonal control of the NO$_3$ assimilation pathway and transport. (a) Summary of the concentration and effects of the phytohormones (Auxin, CK = Cytokinin, Ethylene and ABA = Abscisic acid), in different concentrations of NO$_3$. (b) Molecular mechanisms of auxin transport involving the NRT1.1 protein. In low NO$_3$ concentration conditions, the phosphorylated form of NRT1.1 protein acts as an auxin transporter promoting auxin flux out of the root primordium. In higher NO$_3$ conditions, NRT1.1 mediated auxin transport is bypassed which promotes auxin accumulation in the root primordium resulting in lateral root emergence (Adapted from Krouk, 2016).

In sympodial plants (e.g. tomato), the shoot is composed of different axes; the first inflorescence is produced at the termination of the primary vegetative apex (initial segment) after 8 to 12 leaves are produced (depending on the species or cultivar (Pnueli et al. 1998). The axillary bud, below the inflorescence, forms a new vegetative shoot (sympodial segment), which will give continuity to plant growth and development. The shoot architecture of sympodial species alternates after the termination of the initial segment, i.e. the sympodial segment will produce 3 leaves and then terminate with an inflorescence, another sympodial segment will be produced by the axillary bud.
and the process will be repeated indefinitely (Sussex & Kerk 2001; Janssen et al. 2014) – Figure 4.

1.3.1 The molecular basis of the shoot control

Shoot- and root-specific organs start their formation in the postembryonic development (Domagalska & Leyser 2011), and regulation of this process can be exerted at this point. At the same time, secondary meristems are formed, e.g. cambium stripes, and the secondary shoot apical meristem (SAM) activity starts to increase, usually throughout development of leaves and then reduces its growth to form a dormant bud (Nordström et al. 2004). Depending on the source of the dormancy signal, the physiological state of the bud is classified into three different classes: (i) paradormancy, is induced by signals carried from outside the bud, e.g. phytohormones – Cline & Deppong 1999; Anderson et al. 2012); (ii) ecodormancy, is caused by environmental factor on the bud – (Shimizu-Sato & Mori 2001); (iii) endodormancy or true dormancy, is a strong arrest of bud growth and develop induced by internal physiological factors, maintained even in favourable conditions (Bilavcik et al. 2015). The molecular regulation of the bud growth and development is associated with several different mechanisms, outlined below and previously described (Yang & Jiao 2016).

1.3.1.1 Axillary meristem initiation

The axillary meristem is formed at the centre of the boundary zone; this region is essential for the meristem and organ maintenance development (Žádníková & Simon 2014). During the establishment of boundary zones, the transcription factor LATERAL ORGAN BOUNDARIES1 (LOB1) stimulates “phyB (phytochrome B) activation-tagged suppressor1” (BAS1) (Kebrom et al. 2010). BAS1 is reported to negatively regulate brassinosteroid signal transduction (Neff et al. 1999; Bell et al. 2012).
During the initial growth of the leaf primordium, the auxin efflux carrier PIN-FORMED1 (PIN1) is oriented by the leaf primordium (Gendron et al. 2012). After the development of the boundary zone, PIN1 is reoriented towards the SAM (Q. Wang et al. 2014; Y. Wang et al. 2014). The PIN1 synchronisation to SAM happens through the action of the kinase PINOID (PID); which is responsible for the position of PIN1 in the plant (Furutani et al. 2004). An increase of auxin caused by the expression of the transgenic gene *iaaM* in the boundary zone, inhibited axillary meristem development; in contrast, when AUX/IAA protein BODENLOS is used to decrease auxin signalling in this area, the formation of axillary meristem is stimulated (Q. Wang et al. 2014).

Some genes have been reported to affect lateral organ development. For instance, *RPS10B* that encodes the S10e protein (Stirnberg, Liu, et al. 2012) and *FAR-RED ELONGATED HYPOCOTYL3 (FHY3)*, both essential for auxin homeostasis (Stirnberg, Zhao, et al. 2012). In addition, the *CUC* transcription factors, *CUPSHAPED COTYLEDONS 1, 2 and 3*, were reported to have a dominant, and yet redundant, role in meristem formation and development in *Arabidopsis* (Spinelli et al. 2011). In tomato, *GOBLET (GOB)* is an orthologue of the *CUC* genes (Busch et al. 2011). The development of SAM and formation of boundary zones are linked to the expression of these genes (Teichmann & Muhr 2015). *GOB* is down-regulated by BRs; therefore, low BR activity allows the higher expression of *GOB* genes (Bell et al. 2012; Gendron et al. 2012).

### 1.3.1.2 Axillary meristem control

The physiological response of the axillary buds is very dynamic, thus is an important trait for the shoot architecture. The axillary activity is negatively regulated by the shoot apex, i.e. apical dominance (Teichmann & Muhr 2015). The maintenance of apical dominance regulated by an internal signal, i.e. auxin (reported by Thimann & Skoog 1933), from a dominant shoot apex (Snow 1925). The biosynthesis of auxin occurs mainly in young leaves (Ljung et al. 2001), and it is transported downwards in the stem. Apical dominance is interrupted by removing the auxin source. On the other hand, exogenous
application of auxin on the apex can restore apical dominance (Thimann & Skoog 1933).

**Figure 4:** Inflorescence development in tomato. In the vegetative phase, the shoot apical meristem (SAM) initiates a leaf and an axillary meristem. Between the transition, the SAM takes an intermediate, transitional meristem (TM) fate whereas the last vegetative axillary meristem called the sympodial (SYM) takes over shoot growth (Perilli et al. 2012). The TM initiates a new phytomere composed by the inflorescence meristem (IM). TM and IM maturate toward floral meristem (FM) fate and become flowers (F) (Thouet et al. 2012). Each IM initiates another IM in the meantime of maturating to FM (Adapted from Périlleux, Lobet, & Tocquin, 2014).

The negative regulation, i.e. growth inhibition, of the bud is mediated by a signal transported upwards (Thimann & Skoog 1933). Both cytokinins and strigolactones (SL) are suggested to be involved in this negative regulation (Teichmann & Muhr 2015). Studies carried out in plants with different cytokinin levels confirmed the dominant role of CK in the bud control. Bud dormancy was broken when cytokinin was applied to the axillary buds, even with the intact plant apex (Sachs & Thimann 1967). Considering that the bud growth is controlled by auxin; it was suggested that the action of CK in the bud is regulated by auxin (Nordström et al. 2004). Auxin is transported downwards, from plant apex, to the stem reducing the expression of *ISOPENTENYLTRANSFERASE (IPT)* – gene responsible for the biosynthesis of CK (Tanaka et al. 2006).

Similar work in peas shows that auxin induces the CK oxidase gene *PsCKX2* (Shimizu-Sato et al. 2008). CK oxidases incapacitate CK action by converting active CK into an inactive form, and as a result, bud dormancy is maintained (Werner et al. 2001).
Root tips are a major site of CK biosynthesis (Miyawaki et al. 2004). However, studies where overexpressing CK mutant roots were grafted to WT scions showed that bud outgrowth was not stimulated (Faiss et al. 1997). These results suggested that cytokinin was not the only hormone involved in bud outgrowth regulation. Mutants showing a different branching pattern in Arabidopsis, more axillary branching (max), (Bennett et al. 2006); pea, ramosus (rms), (Beveridge et al. 1994; Beveridge et al. 1996; Sorefan et al. 2003); petunia, decreased apical dominance (dad), (Simons et al. 2007) and rice, dwarf (d), (Mori et al. 2002; Gao et al. 2009; Jiang et al. 2013) resulted in the identification of the dominant role in branching played by strigolactones. SLs are transported from roots through the xylem (Kohlen et al. 2011) to shoot (Beveridge et al. 1996; Sorefan et al. 2003; Bennett et al. 2006; Beveridge 2006). Biosynthesis of SLs is stimulated by auxin (Sorefan et al. 2003).

The biosynthesis of SL is reported to take place in the root and in the shoot (Auldridge et al. 2006; Umehara et al. 2008; MASHIGUCHI et al. 2009). Three possible repressors of SL signalling were recently identified: DELLA proteins (in Arabidopsis, Nakamura et al. 2013), BES1 gene (in Arabidopsis, Zhou et al. 2013), and the D53 protein (in rice, Zhou et al. 2013). The D53 protein is reported to act by preventing SL degradation (Zhou et al. 2013). The SUPPRESSOR OF MORE AXILLARY GROWTH2 LIKE 7 in Arabidopsis, was suggested to have a similar function as D53 (Cardoso et al. 2014). Although some key precursors of SL are still being characterized and described (Abe et al. 2014), it is known that the SL signalling process involves the hormone binding to a receptor, activating an F-box protein-containing the CSF/E3 ligase complex (Skaar et al. 2013). This process is similar to the signalling process of auxin, jasmonic acid and GA (McSteen & Zhao 2008). The latter regulates the ubiquitination and subsequent degradation of a transcriptional repressor, leading to changes in gene transcription (Hagen & Guilfoyle 2002; Hartweck 2008; Memelink 2009).

After the evidence that apical dominance is regulated by auxin, there is still the transport velocity of auxin in the shoot to be considered: bud
development is initiated before auxin levels in the stem diminished when the apical auxin source was removed (Morris et al. 2005). For this reason, a different regulator could be involved with the bud control (Mason et al. 2014), for example, after the removal of the auxin source, there is escalation of sucrose levels in axillary buds. Additionally, buds broke dormancy after receiving a sucrose treatment (Teichmann & Muhr 2015) and measures of sucrose transport are consistent with the suggestion that sucrose is the acropetal signal regulating bud growth (Phillip 1975). Furthermore, sucrose treatment regulates negatively the branching suppressor \textit{BRANCHED1 (BRC1)} gene (Phillip 1975).

\subsection*{1.3.1.3 The meristematic control on inflorescence architecture}

The shoot architecture of mature plants is determined by the axillary meristem initiation, bud growth and branch patternning (Cline 1997). Changes in theses aspects are responsible for the phenotypic variation observed in different species (Leyser 2005). The genotype is deeply responsible for this phenotypic variation, but, as already mentioned, responses to environment are essential and also have a considerable impact on the plant phenotype (Casal et al. 1986; López-Bucio et al. 2002; Yoneyama et al. 2013). Besides providing diversity to plants, truss architecture is an important trait for the reproductive success in angiosperms; agronomically speaking it is a trait that can strongly influence yield potential.

\textit{Arabidopsis}, tomato (Figure 5), rice and maize are commonly used as models in the literature to describe the molecular mechanisms involved in inflorescence branching (Wang & Li 2008). Rice and maize inflorescence architecture is determinate before the floral meristem (FM) production (Harrop et al. 2016). But, in comparison to rice and tomato, maize shows a different and more complex inflorescence – the architecture development of both male and female in maize inflorescence, tassel and ear, respectively, is different (Lunde & Hake 2005).
Figure 5: Network of genes involved with meristem fate in tomato inflorescences. Interactions between vegetative and reproductive phases are represented by the grey line. FALSIFLORA (FA) is upregulated by a SINGLE FLOWER TRUSS (SFT) signal, both repressing the vegetative growth (veg). TERMINATING FLOWER (TMF) has an antagonistic role to SFT, thus promoting veg by repressing FA. ANANTHA (AN) and FA are required for the maturation of inflorescence meristem (IM) to floral meristem (FM); AN is also reported to determine the FM identity. Different from AN, SINGLE FLOWER TRUSS (SFT) and JOINTLESS (J) are not required for FM identity, but are reported to be involved with the inhibition of the veg in IM. Both act preventing premature maturation of IM toward FM, possibly by repressing FA. COMPOUND INFLORESCENCE (S), on the other hand, accelerates IM maturation (adapted from Périlleux et al., 2014).

Non-orthologue genes have been suggested to influence shoot architecture in a redundant way by controlling the same regulatory pathways. (Kellogg 2007). In Arabidopsis and tomato, the floral meristem (FM) is directly produced by the inflorescence meristem (IM) – see chapter 3, page 80; (Schmitz & Theres 1999). In rice and maize, two intermediate meristems are formed between FM and IM (Tanaka et al. 2013). In grass inflorescences, after the SAM changes fate (i.e. transition between vegetative and reproductive phases), SAM is converted to IM; which will initiate primary, secondary or tertiary branching meristem (pBM, sBM or tBM – depending on the species; e.g. rice produces all types). The BM will induce initiation of the spikelet meristems (SM), which initiate the FMs. Finally, the floral meristem produces florets and
outer organs (pelea). Several genes have been reported to be involved in the initiation and regulation of the intermediate meristems.

The transition between the BM to SM is controlled by BRANCHED SILKLESS1 (BD1) and FRIZZY PANICLE (FZP), in maize and rice, respectively (Chuck et al. 2002; Komatsu et al. 2003). Indeterminate branches are produced in bd1 and fzp mutants (Tanaka et al. 2013), suggesting that these two genes regulate the BM establishment and the SM identity.

Studies with maize have shown that the transition between SM and FM is mediated by INDETERMINATE SPIKELET1 (IDS1) and SISTER OF INDETERMINATE SPIKELET1 (SID1) genes (Chuck et al. 1998; Chuck et al. 2008). Chuck et al. (2008) showed that in ids1 and sid1 double mutants, glumes are uninterruptedly formed, suggesting that the transition between the SM to FM is impaired. It was also suggested that IDS1 and SID1 have redundant functions in the control of inflorescence architecture, and also encode AP2/ERF transcription factors. Regarding the sex determination of the inflorescences, the mutants tasselseed4 (ts4) and Tasselseed6 (Ts6), were used to determine the role of the miR172 microRNA family which inhibits the AP2 domain transcription factors function (Chuck et al. 2007). The INDETERMINATE FLORAL APEX1 (IFA1) and REVERSED GERM ORIENTATION1 (RGO1) genes are reported to be involved with the regulation of the SM identity in maize (Laudencia-Chingcuanco & Hake 2002; Kaplinsky & Freeling 2003). Hence, IDS1, IFA1 and RGO1 have redundant functions controlling the determinacy of the meristems in the reproductive phase of maize (Tanaka et al. 2013).

The SUPERNUMERARY BRACT (SNB) and OsIDS1 genes have orthologous function to IDS1 and SID1 in rice (Lee et al. 2007; Lee & An 2012), suggesting that even with the different inflorescence architectures, the molecular mechanisms controlling the transition from the SM to the FM is conserved between maize and rice. The LEAFY HULL STERILE1 (LHS1) MADS box gene in rice is involved in SM identity – lhs mutants do not show the C terminus of the MADS box protein, resulting in an indeterminate SM as
phenotype (Jeon et al. 2000). Mutations in ABERRANT SPIKELET AND PANICLE1 (ASP1) are reported to impair the auxin signalling, thus affecting inflorescence architecture and spikelet development (Yoshida et al. 2012).

1.3.2 Shoot architecture in plant domestication

Some traits were prioritized during crop domestication, and it is likely that inflorescence architecture is one of those traits (Zhang & Yuang 2014). Different branching profiles in monocot plants are very evident – e.g. panicoids were selected to show a less tillering and reduced axillary development (Teichmann & Muhr 2015).

In cultivated maize crops, ideally, one single female inflorescence (ear) is produced and the apical axis is terminated in a male inflorescence (tassel). In contrast, wild maize species (e.g. teosinte) show multiple axillary branches developed at the same time; each new branch terminates with a tassel (Doebley et al. 1997). One of the most important genes during the maize domestication is TEOSINTE BRANCHED1 (TB1). Doebey (1997) showed that even some minor changes in the expression of TB1 are enough to cause substantial alterations in the shoot phenotype of maize and/or teosinte (Doebley et al. 1997). The DENSE AND ERECT PANICLE1 (DEP1) gene is responsible for the increase in height and number of seeds (grains) per panicle observed in modern rice crops (Huang et al. 2009).

Genes involved with the signalling of SL were also selected during domestication (Teichmann & Muhr 2015). Reduced levels of strigolactones cause high tillering, suggesting that SL play a bigger role than already described in plant architecture (Cardoso et al. 2014).

The limited physical space for farming led to the selection of TILLER ANGLE CONTROL1, TAC1 (Yu et al. 2007). Tillers of the wild rice (Oryza rufipogon) grow horizontally to suppress competition with other plants, but also considerably reduce the density of plants per m² (Galli & Gallavotti 2016). After analysing 152 different accession, Yu et al. (2007) showed that all lines with
mutation on TAC1 show a decreased tiller angle, and concluded that the TAC1 mutation was under heavy selection (Yu et al. 2007).

1.4 Plant domestication in tomato: wild species as germplasm source

Due to morphological trait analysis (e.g. anther anatomy), taxonomists previously considered wild species of tomato as part of genus *Lycopersicon* (Grandillo et al. 2011). However, the latest molecular and phylogenetic results determined that tomato should be classified within the genus *Solanum* with the other important crops, as potato and eggplant (Knapp et al. 2004; Fischer et al. 2011; Grandillo et al. 2011). Regardless of the classification, wild species (WS) have been widely used by breeders and geneticists to use genetic variation into tomato breeding projects.

The in-breeding and selection for agronomic traits in small populations of tomato cultivars resulted in a narrow genetic base between these cultivars (Foolad 2007). On the other hand, wild species show extensive genetic variation. As a result, wild species are used to identify quantitative trait loci (QTLs) carrying polymorphisms related with high value traits for crops such as disease resistance, insect resistance, abiotic stress tolerance, yield, fruit quality and yield, for example (Grandillo et al. 2011).

Research centres in the United States of America started breeding programs using wild tomato species in the beginning of the 20th century, mainly to introduce disease resistance traits into cultivar species (Bai & Lindhout 2007). Afterwards, F1 hybrid tomatoes became very popular and started to dominate the classic cultivars used for fresh market tomatoes. Initially (until the 50's), breeders were selecting and compiling traits for both the industrial processing and fresh markets, but due to different demands, crops started to be bred for different proposes. For fresh market, breeders initially selected for bigger yield, and then post-harvest shelf-life traits were considered very important. Today, the quality and nutritional profiles of fruits are the priority for
consumers. On the other hand, cultivars for industrial processing were bred to carry important qualities for the industry. For instance, tomato varieties were bred for mechanical harvesting by changing the growth habit (sp genes) and the fruit physiology (jointless) (Fridman et al. 2001).

Tomato can be affected by more than two-hundred diseases caused by a large number of pathogens (e.g. viruses and bacteria). All described resistance genes arose from wild species (WS) – i.e. S. chilense, S. peruvianum, S. habrochaites and mainly from S. pimpinellifolium (Foolad et al. 2014). Around 20 resistance genes were introgressed into different cultivars’ genomes, providing resistance against around 42 major diseases. Research conducted in disease resistance represents the biggest use of the WS germplasm (Foolad 2007; Grandillo et al. 2011).

Insect resistance represents a significant advantage for crops and different cultivars. WS are a rich source of resistance against the most relevant insect pests of tomato. This is particularly true for S. habrochaites, which is resistant to at least 18 pest species (Ji, J. W. Scott, et al. 2007). Resistance genes were also found in S. pennelli, S. pimpinellifolium, S. cheesmaniae, S. chmielewshii, S. peruvianum, S. cornelioulleri, S. arcanum and S. chilense (Farrar & Kennedy 1991; Muigai et al. 2003). Several factors (e.g. trichrome characteristics) are related with arthropods resistance, but due to the rich natural diversity of the aggressor organisms, the selection and breeding of this trait has proven to be particularly complicated (Grandillo et al. 2011).

The diversity of environments where WS have been found is also very relevant because it has allowed the evolution of resistance to abiotic stresses such as drought and salinity. Breeding cultivars with such traits is an objective with large economic impact and it has been extensively studied in tomatoes (Cuartero & Fernández-Muñoz 1999; Santa-Cruz et al. 2002; Chaves et al. 2003; Fischer et al. 2011; Sandhu et al. 2014; Rogers & Benfey 2015). As an example, genes related with drought tolerance were identified in S. pimpenellifolium, S. pennelli and S. sitiens; and high tolerance to low
temperatures was demonstrated in *S. habrochaites* (Fischer et al. 2011; Arms et al. 2015).

The market for fresh tomato is very diverse regarding customer demand for traits (e.g. fruit colour, shape, sugar content) (Grandillo et al. 2011). Wild species show a very diverse range of fruit characteristics, for example *S. chmielewskii* present an increase of 40% of the total soluble solids, without significant yield loss. In some regions the red fruit colour is considered as an important trait (Labate et al. 2007). Studies with the *Beta carotene* (*B*) gene on chromosome 6, found in several WS, were essential to understand the fruit colour control in tomato. High expression of *B* increases considerably (15x) the level of pro-vitamin A (β-carotene) in the fruit (Labate et al. 2007; Grandillo et al. 2011).

### 1.5 Technologies for genomics

The analysis of distinctive DNA polymorphisms in individuals and populations is considered the foundation of modern molecular genetics. Such studies allow genotypes to be identified and connected to possible phenotypes. Single nucleotide polymorphisms (SNPs), especially, can be used in phylogenetic studies (McCormack et al. 2013). Improvements and advances in high-throughput (HT) DNA sequencing has allowed more efficient breeding and genetic mapping protocols using SNP markers; these polymorphisms can be directly linked to traits with economic importance, and can be readily fine-mapped to identify candidate genes and causative polymorphisms.

#### 1.5.1 Sequencing tools and approaches

The rapid advance in technology has drastically changed the way complete genome sequencing may be accomplished. Illumina platforms of second- HT-next generation sequencing (HT-NGS) are able to generate billions of bases in a single lane run, which revolutionized the studies of DNA sequences (Schadt
et al. 2010). These second generation of HT-NGS are based on emulsion PCR amplification of DNA fragments. The amplification step was required to amplify the light signal for reliable base detection, but could also result into the addition of wrong bases (Schadt et al. 2010); which was resolved by eliminating this step on the third generation HT-NGS – i.e. based in the sequencing of a single DNA molecule.

Heliscope™ was released as the first single-molecule DNA sequencing platforms. Its sequencer is based on “true single molecule sequencing” (tSMS). In tSMS technology the DNA fragments are hybridized to oligonucleotides after the library preparation (Ozsolak et al. 2010). As a result, up to 28 Gb of data can be generated in a single sequencing run (Fatih Ozsolak & Milos 2011; F Ozsolak & Milos 2011).

Single Molecule Real Time (SMRT™) is a different platform based on real time sequencing in zero-mode wave-guides (ZMWs). The DNA fragment is extended using deoxynucleotide triphosphates (dNTPs) tagged with fluorescent labels, which are detected on a charge-coupled device (CCD) – (Levene et al. 2003; Eid et al. 2009). Pacific Biosciences, responsible for the development of the SMRT™ platform, claims that their sequencing system is capable of generating ~100 Gb/hour, with long reads (>10000 bp) in a single run (Pareek et al. 2011).

Real time single molecule DNA sequencers are currently being produced and improved. VisiGen Biotechnologies has designed a ‘real-time sensor’ for individual nucleotides tagged with specific fluorescent dyes. Although the platform is still under development, it was reported to generate 4 Gb of data per day, with reads 1 kb longer than any current platform (Selvin 2000).

The Nanopore DNA sequencer is different from the platforms previously described, because it is not based on fluorescent nucleotide and detection. The Nanopore DNA sequencer works by passing a DNA fragment through a nanopore while attached to a covalent charged molecule; this process identifies the characteristics (e.g. diameter) of the DNA molecule for base calling (Liang et al. 2017; Leung et al. 2017; Cornelis et al. 2017).
In general, the third generation platforms perform better than the second HT-NGS technologies. The accuracy and volume of generated data is higher, for de novo assembles longer reads can be obtained. Unfortunately, the third generation HT sequencing is still expensive, but with the technology fast developing, it has the potential to be cheaper and more efficient than second-NGS platforms, after proper establishment.

1.5.2 Genome sequencing and genotyping techniques

The traditional genotype identification and detection is accomplished with the use of markers (e.g. Cleaved Amplified Polymorphic Sequence, CAPS) based on allele-specific SNPs. Currently, almost all genotyping methods require PCR to amplify the target region containing the polymorphism, followed by an allelic discrimination step.

The allelic discrimination is based on allele-specific reactions. There are four classical methods: primer extension, hybridization, ligation, and enzymatic cleavage. In the primer extension approach, nucleotides in a primer are incorporated to a DNA template, by a specific enzyme to detect both alleles or with distinct primers for each allele. The common primer extension (CPE) reaction is based in a primer selected/designed to anneal in a sequence juxtaposed to a SNP site, which is amplified by a polymerase enzyme (Sokolov 1990). The primer can be designed to detect a single or multiple SNPs, which increased the commercialization of different CPE systems for SNP genotyping.

Allele detection by fluorescence: systems based in fluorescent signals are commonly used by genotyping platforms, due to its easy implementation and high sensitivity. Besides the usual applications (Kim & Misra 2007), fluorescent methods are employed in a single-base extension, SBE, (Hoogendoorn et al. 1999; Premstaller et al. 2002) and denaturing high-performance liquid chromatography (DHPLC) (Liu et al. 1998; Xiao & Oefner 2001).
Allele detection by chemiluminescence: a method with rapid detection, easy automation and high signal-to-noise ratio of SNP genotyping based in a cascade of enzyme reactions involving inorganic pyrophosphate (Ronaghi et al. 1996). SNP genotyping assays associate the PCR-amplified template DNA with a primer designed based on the SNP site (100). Chemiluminescence is also used for DNA sequencing methods, i.e. Pyrosequencing™ (Kim & Misra 2007).

Allele detection by mass: systems using matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), are the PinPoint assay (Haff & Smirnov 1997; Ross et al. 1998), MassEXTEND™ (Braun et al. 1997; Cashman et al. 2001), SPC-SBE (Kim et al. 2002), and GOOD assay (Sauer et al. 2002). Basically, the allelic discrimination on these platforms takes place by amplifying PCR products as a template. Each extension will have a different mass and correspond to one of the alleles of SNP (Le Hellard et al. 2002; Sanchez et al. 2003).

These platforms are based in SBE detection with primers tagged with fluorescent ddNTPs (Fan et al. 2000; Hirschhorn et al. 2000). Examples of these systems are the SNaPshot® (Kim & Misra 2007), the SNPstream™ assay (Pastinen et al. 1997; Nikiforov et al. 1994; Bell et al. 2002), allele-specific PCR (AS-PCR) (Gibbs et al. 1989; Medintz et al. 2001; Latorra et al. 2003) and Tag-It™ approach (Bortolin et al. 2004).

Hybridization techniques perform the allelic discrimination by detecting the mismatches in the target-probe. The efficiency of the differentiation relies on the length, sequence and SNP site in the probe. Some HT SNP assays were designed to use microarrays, and they have the advantage of enzyme exclusion to perform the allelic distinction – e.g. GeneChip® array (Kennedy et al. 2003; Matsuzaki et al. 2004) and TaqMan® genotyping assay (Holland et al. 1991; Livak 1999).

Ligation approaches uses ligase enzymes to conduct allelic discrimination. Two oligonucleotides are hybridized to a single-stranded DNA and then ligase enzymes assembling them to a single oligonucleotide. Three probes are used by a traditional ligation platforms; 2 allelic-specific and 1
common to merge the adjacent template to the SNP specific probe at the 3’ end due to the higher enzymatic sensitivity (Landegren et al. 1988). Different methods can detect the SNPs in the ligation production – e.g. combinatorial fluorescence energy transfer – CFET (Nilsson et al. 1994; Tong et al. 2001), rolling circle amplification – RCA (Lizardi et al. 1998), molecular inversion probe – MIP (Hardenbol et al. 2003; Hardenbol et al. 2005).

Enzymatic cleavage can also be used for allele discrimination. This method works by the action of enzymes capable of cleaving the DNA, when recognizing specific sequences and structures. For instance, CAPS (Dillon et al. 2006; Foolad 2007; Tezuka et al. 2009). This method does not require any probes. On the other hand, genotyping assays based on restriction enzyme site have limited applicability due to the limited number of SNPs that can be used – e.g. the Invader® assay (Lyamichev et al. 1999).

As described above, many different assays can be used to screen SNPs in different species. These assays need to be robust, rapid and cost-effective (Ragoussis 2009). In my opinion, currently, two methods are particularly efficient for genotyping during genetic mapping, and these have different applications depending on numbers of SNPs and samples.

1.5.2.1 Genotyping-by-sequencing

A recently employed approach used for HT SNP detection and genotyping is genotyping-by-sequencing (GBS) (Andrews et al. 2016; Voss-Fels & Snowdon 2016). Compared the other methods based on NGS sequencing, GBS is significantly cheaper. GBS was successfully used as a breeding tool in important crops – e.g. maize (Elshire et al. 2011) and potato (Uitdewilligen et al. 2013). In association with phenotyping data, the GBS approach can increase the rapidity of mapping and detection of genes regulating important agronomic traits, which can be later introduced into cultivated crops (Edwards et al. 2013).
1.5.2.2 Bulk segregant analysis

The bulk segregant analysis (BSA) technique can dramatically decrease the costs of genotyping, especially when conducted to define an initial mapping position (Salunkhe et al. 2011). The BSA method is based on the sequencing/genotyping of two pools of DNA from individuals with the same phenotype (Venuprasad et al. 2009; Sun et al. 2010). In other words, a segregating population (F2) is divided into two different pools (i.e. bulks), one for each phenotype (Michelmore et al. 1991). These two pools are sequenced by NGS and then screened compared for sequence differences (del Viso et al. 2012).

The number of individuals in each pool is important; enough genetic recombination events represented in both bulks will exclude non-causative SNPs. Furthermore, SNPs linked to a recessive causative mutation (by linkage) will be scored as homozygous in the bulk with the target trait, but will remain heterozygous in WT pools (del Viso et al. 2012). BSA applied to NILs using large pools are very effective (del Viso et al. 2012), especially in the case of mutations caused by ethyl methanesulfonate (EMS) because the exposure to EMS causes a relatively low number of mutations which subsequently need to be resolved (Ansari et al. 2013); this represents an efficient and faster genotyping method.

BSA was successfully used to map diverse traits in different organisms; wheat (Hu et al. 2016), maize (Farooqi et al. 2016), rice (Salunkhe et al. 2011; Vikram et al. 2012; Rani et al. 2013), melon fruit (Chayut et al. 2015), barley (Poulsen et al. 1995), yeast (Dunham 2012), and tomato (Lin et al. 2006; Elsayed et al. 2012; Truong et al. 2015).

1.5.3 CRISPR-Cas9

As an alternative to the traditional breeding, genome editing techniques can be used — e.g. meganucleases; zinc-finger nucleases, (Ito et al. 2015); transcription activator-like effector nucleases, TALEN, (Khlestkina & Shumny
2016) and clustered regularly interspaced short palindromic repeat-associated nuclease protein, CRISPR-CAS (Zsögön et al. 2017). So far, the newest and most efficient system is CRISPR-CAS9 (Shi et al. 2017).

In the CRISPR-Cas9 method, the target locus is broken by the action of Cas9 (CRISPR associated protein 9), and the DNA repair is carried out by non-homologous end joining (NHEJ), or by homologous recombination (HR) (Shi et al. 2017). NHEJ is responsible for some small InDels at the repair junction (Puchta 2017); HR, on the other hand, may occur in a conservative or a non-conservative pathway (Steinert et al. 2016). Although somatic cells mainly use NHEJ pathways (Shi et al. 2017; Puchta 2017), when homologous fragments are near to the target site (i.e. tandem duplications), the induced double-strand break is repaired by HR (Steinert et al. 2016).

Most genome engineering in plants is now performed with the CRISPR-Cas9 system, through the NHEJ pathway. The required nuclease is integrated into the modified plant, and later eliminated from the offspring by genetic segregation to produce plants with no transgenic DNA (Fauser et al. 2014). However, according to the regulations in some countries, these plants are still defined as genetic modified organisms (GMOs), due to the involvement of recombinant DNA technology in their production (Sprink et al. 2016).

CRISPR-Cas9 was used recently as an efficient tool in the genome editing of important crops like rice (Srivastava et al. 2017), maize (Char et al. 2017; Shi et al. 2017) and tomato (Ito et al. 2015; Xu et al. 2016; Soyk et al. 2017) – Figure 6.

1.5.4 Bioinformatics’ tools for data analysis

The analysis and management of genomic data is as important as the chosen HT-SNP genotyping or mapping method. Initially, the raw data is submitted to a quality control tool; e.g. FastQC (Kroll et al. 2014), PRINSEQ (Schmieder & Edwards 2011), MultiQC (Ewels et al. 2016) and Qualimap 2 (Okonechnikov et al. 2015). In some cases, possible contamination may lead to low-quality reads that impact directly the mapping and SNP calling in later steps. For this reason,
filtering and trimming tools – e.g. FASTX-Toolkit (Qi et al. 2015), Trimmomatic (Bolger et al. 2014) and AdapterRemoval 2 (Schubert et al. 2016), are used to prepare the reads in the FASTQ file for the next step.

Figure 6: Genome editing using CRISPR-Cas9. With the help of the guide-RNA (gRNA), the nuclease (Cas9, CRISPR-associate protein) complex binds to the target site and cleaves the double-stranded DNA; activating the double-strand break repair machinery (PAM = protospacer-adjacent motif). In the absence of a homologous repair template, non-homologous end joining can result in InDels disrupting the target sequence (Puchta 2017). Alternatively, precise mutations can be made by providing a homologous repair template (Soyk et al. 2017) (Adapted from: https://www.neb.com/~media/NebUs/Files/Feature%20Articles/Images/FA_Cas9_Fig2_Cas9forGenomeEditing.png).

Most of the important crops count with available reference genomes (Scheben et al. 2017). Thus, the reads are simply aligned to the reference using a read aligner – e.g. BWA-MEM (Li 2013); Bowtie2 (Langmead & Salzberg 2012); Tmap (Zhu et al. 2014) or Novoalign (Hwang et al. 2015). Once the reads are aligned, the steps to detect polymorphisms (i.e. SNPs and InDels) are
conducted by a variant callers (Nielsen et al. 2011) – e.g. the Genome Analysis Tool Kit, GATK (McKenna et al. 2010); SAMtools (H. Li et al. 2009); Freebayes (Garrison & Marth 2012); SOAPsnp (R. Li et al. 2009) and Platypus (Rimmer et al. 2014). Diverse studies were carried out comparing the efficacy of different aligners/variant callers, and they have reported different conclusions (Li 2013; O’Rawe et al. 2013; Pabinger et al. 2014; Zook et al. 2014; Clevenger et al. 2015), but it is clear that the detection method used by each variant caller contributes to these differences (Scheben et al. 2017). The methods used by SOAPsnp, SAMtools and GATK UnifiedGenotyper look for polymorphisms base-by-base. On the other hand, GATK HaplotypeCaller, FreeBayes and Platypus assemble local haplotypes to search for polymorphisms (Scheben et al. 2017). The methods for modelling and estimating sequencing errors are also different; some (e.g. FreeBayes and GATK) use Bayesian methods and others apply the Markov model to estimate error (Garrison & Marth 2012; Li 2013). Thus, due to the different data sets, methods applied and expected outcomes it is virtually impossible to designate a generally superior tool.

For GBS data, there are some common pipelines based on tools such as TASSEL-GBS (Glaubitz et al. 2014), Stacks (Catchen et al. 2011) and UNEAK (Lu et al. 2013). Again, each is used for different types of data sets: TASSEL-GBS is used to with large quantities of low-coverage reads, whereas UNEAK and Stacks are used in de novo methods for species without reference genomes (Scheben et al. 2017).

Once the variant is detected, an annotation step is applied to the data set. The objective is to identify variants in important regions such as exons and gene promoters. The annotation tools Annovar (Wang et al. 2010), SnpEff (Cingolani et al. 2012), Variant Effect Predictor, VEP (McLaren et al. 2010) and VariantAnnotation (Obenchain et al. 2014) require an annotated reference genome or transcript set. There is also a difference in a variant’s predicted impact depending on the tools used for annotation (Scheben et al. 2017), due to the different methodologies applied by the tool. For instance, SnpEff (5kb) and
Annovar (1kb) use different metrics, i.e. distance between gene and variant, to characterise upstream and downstream regions (Hwang et al. 2015).

Data management is also something that needs to be considered when working with genomics. Currently, storage and integration are the biggest challenge faced by bioinformaticians and scientist in the field (Batley & Edwards 2009; Lee et al. 2012). One of the strategies adopted to deal with the vast amount of data is based on data compression using different file formats. For instance, NGS data is encoded in FASTQ files; using SAMtools FASTQ files can be transformed into binary (BAM) files, 3-4 times smaller. Another strategy accomplished with SAMtools (BCFtools) is to store only the detected variants in a variant call format (VCF) file (Danecek et al. 2011); and later compress it using tabix (Li 2011). Unfortunately, it is not uncommon for data to be lost during the analysis pipeline thus, raw FASTQ files are kept for security reasons, which still represents a considerable expenditure with storage hardware. It is being suggested that cloud-based storage systems can offer a solution to that problem (O'Driscoll et al. 2013). Public databases can also be part of the solution of this problem. Broad access to annotation files and other resources can help and serve as a centralised point of information to the research community and the public.

1.5.5 Reference genomes

Making use of second and third generation sequencing and high density SNP arrays, studies relating genotype and phenotype are becoming faster, cheaper and more straightforward. The tomato reference genome (cv. Heinz 1706) is publicly available and extensively used for SNP identification between different Solanum species genomes (Kim et al. 2014). As already described, NGS technology has become more accessible and cost effective, the quality of the outcome improves constantly. This explains why genomic resources are quickly increasing which promotes the whole-genome analyses, transcriptome characterization and detection of variations across various genomes (Sim, Durstewitz, et al. 2012).
The tomato genome reference was published by the Tomato Genome Consortium, in 2012. Since then, it has changed and improved on several occasions. In the beginning of February of 2017, the latest version SL3.0 was announced. The release includes new scaffolding arrangements, updated gene structures and ontology, very important integration of bacterial artificial chromosome (BAC) data, and new RNA-sequencing data (SGN 2017).

1.6 Aim

The aim of this project is to improve the quality and yield of *Solanum lycopersicum* L. by understanding the genetic control of plant architecture. The project will focus on branching of roots and flower trusses.

1.6.1 Objectives

Root architecture

1. To develop methodology for root phenotyping using qualitative imaging, and to use this to phenotype recombinants.
2. To study the environmental interaction of bushy root mutant.
3. To obtain genetic data from the bushy root mutant through deeper sequencing.
4. To develop additional SNP markers within the mapping interval for brt and to use them to obtain a map position for brt.
5. To fine map brt by searching for additional recombination event within the mapping interval.
6. To search for recombination in a F<sub>2</sub> population derived from a cross between Micro-Tom and near isogenic line containing the brt mutation.
7. To identify candidate genes in the mapping interval and test their function using transgenic experiments.
8. To identify the organs where the mutant gene is expressed.
Trusses architecture

1. To characterise the mutant phenotype.

2. To study the penetrance and environmental interaction of bifuricate and FAT.

3. To obtain genetic data from the LAM183 and bifuricate by sequencing both lines.

4. To delimit the introgression on the bifuricate material and determinate if its phenotype is the result of compound inflorescence (s) mutation.

5. To identify single nucleotide polymorphisms (SNP) linked to the bifuricate/flower augmented (FAT) per truss gene for use in mapping.

6. To search for recombination in a F2 population derived from a cross between [(LAM183 for bifuricate) and (Micro-tom for NIL-FAT)].

7. To genetically map bifuricate and NIL-FAT in a segregating population to a small genomic region.

8. To identify the genetic locus responsible for the bifuricate and FAT mutants.

9. To identify and characterise additional genetic sources of truss branching genes.

10. To breed the bifuricate trait into the MT genetic backcross.

11. To identify the organs where the mutant gene is expressed.
2 FINE MAPPING AND PHENOTYPIC CHARACTERISATION OF THE TOMATO BUSHY ROOT MUTATION ON CHROMOSOME 12

Silva Ferreira, D. K.\textsuperscript{1}; Kevei, Z. L.\textsuperscript{1}; Kurowski, T.\textsuperscript{1}; Lazaro E. P. Peres\textsuperscript{2}; Awan, S.Z.\textsuperscript{3}, Sergeant, M.J.\textsuperscript{3}, Mohareb, F.\textsuperscript{1} and Thompson, A. J.\textsuperscript{1}

\textsuperscript{1}Cranfield Soil and AgriFood Institute, College road, Cranfield University, Bedfordshire, MK43 0AL, UK.

\textsuperscript{2}ESALQ, Universidade de São Paulo (USP), Av. Pádua Dias, 11 - São Dimas, Piracicaba - SP, Brazil.

\textsuperscript{3}University of Warwick, Coventry, CV4 7AL, UK.

Correspondence: Andrew Thompson. E-mail: a.j.thompson@cranfield.ac.uk

Highlight

Fine mapping of the classical root mutant \textit{bushy root}, phenotype characterization and detection of insertion in the promoting region of an orthologue of \textit{PR1}, composed by the highly conserved pirin domains.
Abstract

The tomato mutant *bushy root* (*brt*) arose from a cross between tomato accessions “Stock No. 1” and cultivar “Red Cherry” following EMS mutagenesis of pollen (Zobel 1972). Micro-Tom (MT) was used to produce a homozygous near isogenic line (*brt*NIL). The mutation has a pleiotropic phenotype: small seed size was observed, and it was determined that this phenotype is controlled by the maternal tissue; the small seed size is likely to explain the slow early development exhibited by *brt*NIL. The mutant root system was characterised by image analysis and can be described as a compact root system; i.e. Fewer and smaller lateral roots and shorter taproot. The *brt*NIL genome was re-sequenced and assembled in comparison to MT, against the reference *S. lycopersicum* Heinz 1706 (version SL2.50); this shows that the introgression spans 64 Mbp of the 67 Mbp on chr 12. After fine mapping, in the current interval, i.e. 137 kbp, a frameshift insertion was detected in the promoting region of a gene with highly conserved cupin domains. The *Pirin* gene (*Solyc12g014590*) is homologue to *PRN1* in *Arabidopsis* and the transcriptomics data supports *Solyc12g014590* as a strong candidate for the *brt*NIL phenotype.

Key words

EMS-mutant, genome resequencing, root architecture, root mutant, pirin, quercetinase, *Solanum lycopersicum*, seed area, transcriptomics
2.1 INTRODUCTION

Plant architecture and developmental vigour is essential to the exploration of the environment and exploitation of available resources. Root architecture is well known for being plastic and essential for water and ion absorption, anchorage and storage (Nibau et al. 2008). Changes in root function and architecture have resulted in recent enhancements for crop production (Hammer et al. 2009) and plants with improved root systems are of great interest for vegetable production where elite scion genotypes with favourable aboveground traits are grafted as rootstocks that can be bred to address more specialised traits such as resistance to soil borne diseases and scion vigour (Schwarz et al. 2013; Asins et al. 2015). Rootstocks have been selected for diverse conditions, e.g. low nutrients availability (Schwarz et al., 2013), hydric stress (Wang et al., 2012), high salinity (Santa-Cruz et al., 2002) and pest control (Gregory et al., 2013). A lot has been done to understand root system architecture and development, particularly in Arabidopsis and cereal crops, but the knowledge of the underlying genetic mechanisms remains rudimentary in grafted vegetable crops such as tomato.

Recently, making use of second generation sequencing and high density single-nucleotide polymorphism (SNP) arrays, studies relating genotype and phenotype are becoming faster, cheaper and more straightforward. The tomato reference genome (cv. Heinz 1706) is public and extensively used for SNP identification between different genomes (Kim et al. 2014). Array-based genotyping techniques—able to score thousands of markers simultaneously—simplified the use of variants (SNPs) for genetic map construction and genome-wide investigation (Foolad 2007; Sim, van Deynze, et al. 2012). Use of NGS for variant discovery and high-throughput genotyping technologies now provides the opportunity for some classical root mutants (Table 1) to be easily investigated to search for causative genes (Kevei et al. 2015). Study of such mutants will enable a better understanding of the genetic control of root system architecture and development. The tomato mutant collection of the Tomato Genetics Resource Centre established by Dr Charles M. Rick (Davis, California)
(Emmanuel & Levy 2002) includes a small number of 17 classical tomato root mutants.

**Table 1:** Classical Mendelian mutations affecting tomato’s root system.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Phenotype</th>
<th>Chr</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerial roots (aer)</td>
<td>High number of shoot-borne roots produced.</td>
<td></td>
<td>(Rick &amp; Butler 1956)</td>
</tr>
<tr>
<td>Bushy root (brt)</td>
<td>Increases number of lateral roots, radical branches early</td>
<td>Chr 12</td>
<td>(Zobel 1972)</td>
</tr>
<tr>
<td>Bushy root 2 (brt2)</td>
<td>Slow initial development with a kinked root phenotype.</td>
<td>Chr 4</td>
<td>(Voland 1987)</td>
</tr>
<tr>
<td>Diageotropica (dgt)</td>
<td>Low geotropic response, changing the root architecture.</td>
<td>Chr 1</td>
<td>(Zobel 1972)</td>
</tr>
<tr>
<td>Dwarf root (drt)</td>
<td>Reduced hypocotyl and internodes, compact root phenotype.</td>
<td>Chr 2</td>
<td>(Rick &amp; Butler 1956)</td>
</tr>
<tr>
<td>Rosette (ro)</td>
<td>Reduced root branching; single and short taproot.</td>
<td>Chr 2</td>
<td>(Rick &amp; Butler 1956)</td>
</tr>
<tr>
<td>Root suppressed (Rs)</td>
<td>Reduced or non-existent root system.</td>
<td>Chr 4</td>
<td>(Rick &amp; Butler 1956)</td>
</tr>
<tr>
<td>Twisted root (twr)</td>
<td>Abnormal trichomes development. Approximately, 38 days after planting the roots start to twist.</td>
<td>Chr 8</td>
<td>(Rick &amp; Butler 1956)</td>
</tr>
</tbody>
</table>

The bushy root (brt) mutant was recovered in the F2 progeny from a cross between ethyl methanesulfonate (EMS)-treated pollen from cv. California Red Cherry and a plant from Charles Rick’s collection called “Stock No. 1”. The cross was made to provide a preliminary linkage map using classical
phenotype-based genetic markers and to eliminate any accidental self-pollination by including a gene for male sterility in the parental line that acted as pollen recipient (Zobel 1972). Zobel was not able to provide a definitive *brt* phenotypic description due to the complicated genetic background that resulted from this cross. Nevertheless, the mutant was named “*bushy root*” due to early branching of lateral roots compared with the wild type (WT): the lateral root (LR) emergence in *brt* happened between 7 to 8 days after sowing (das) compared with the WT, which branched at 9 das (Zobel 1972). In addition to the early lateral root emergence, the *brt* root system was dwarf which gave the visual impression of being “bushier” than the WT, and the *brt* seedling leaves were reported to develop in a dark green and rugose appearance in early development (Zobel 1972). However, when the taproot had reached approximately 20 cm, the mutant plants grew normally until they became indistinguishable from the WT (Zobel 1972). The dwarf root growth habit and the “bushy” root system of the *brt* mutant are unique characteristics which can be distinguished from other dwarfs in classical linkage tests, including the *dwarf* (*D*) tomato mutant. After histological investigation, an intense starch accumulation was detected in the cortical parenchyma of the basal portion of *brt*, compared with both parents (i.e. cv. California Red Cherry and “Stock No. 1”). In young plants this pattern exhibited in the lower portion on the hypocotyl (Zobel 1973). Linkage results showed that *brt* is linked to *albescent* (*alb*), a classical gene located at the top (short arm) of chromosome 12 (Solanaceae project 2017). The tomato reference sequence (Asamizu et al. 2012) now facilitates the precise mapping of the *brt* gene. Additionally, to the availability of the reference genome of *S. lycopersicu*m cv. Heinz 1706, two large studies have provided systematic resequencing of the genomes of 452 cultivars and wild species of tomato (Aflitos et al. 2014; Lin et al. 2014); several other smaller studies have also added to this list, and the MT (Kevei et al. 2015) genome resequencing data is available (SGN 2017).

Micro-Tom (MT) has been extensively used since it was first established as a model cultivar for genetic studies due to its small size. Micro-Tom (MT) has a strong dwarf phenotype and is the result of a cross between *Florida Basket*
and Ohio 4013-3, initially conceived as a novelty variety (Pons et al. 2014). At least three genes are responsible for the MT phenotype: the self-pruning (sp) gene (Solyc06g074350), which belongs to the CELTS family of regulatory genes (Pnueli et al. 1998; Carmel-Goren et al. 2003; Kobayashi et al. 2014), controls the interchange between vegetative and reproductive cycles in the tomato shoot and is responsible for the determinate growth habit that is suitable for field production with mechanical harvesting (Kaul et al. 2000; Saleem et al. 2013); the dwarf (D) gene (Solyc02g089160) is responsible for short internodes, and rugose, dark-green leaves (Meissner et al. 1997). This gene encodes BR-6-oxidase required for the biosynthesis of brassinolide – the most active brassinosteroid (BR) which was identified from pollen in Brassica napus (Shimada et al. 2001). Thus, MT contains a deficiency in BR biosynthesis – a phytohormone required for cell division and elongation growth (Martí et al. 2006). Finally MT carries the miniature (mnt) mutation (long arm of chromosome 11), that has yet not been well characterized, but it was suggested to be involved with the signalling and metabolism of gibberellin (GA) (Saito et al. 2009) and to contribute to the MT dwarf phenotype.

To facilitate genetic analysis, brt was introgressed into the MT genetic background to create a brt near isogenic line (brtNIL) which was characterized for growth and architecture traits and for its profile of phytohormones in comparison to MT. The results of the brtNIL characterization were somewhat different to Zobel's previous description – the colour of brt leaves was lighter, its early development was slower than the WT (MT) (Pino-Nunes & Peres 2005) and it was suggested that brt might have a low sensitivity to cytokinin which might cause some aspects of the brt phenotype such as smaller leaves and different leaf edge patter, as observed in plants overexpressing cytokinin oxidase (Werner et al. 2001; Bartrina et al. 2011). Auxin and ethylene were also tested, but both the brtNIL and WT showed similar sensitivity to these hormones (Pino-Nunes & Peres 2005; Pino-Nunes & Peres 2009). The brtNIL seeds were characterized as smaller than WT’s seeds, it was suggested that the seeds size was maternally inherited (Pino-Nunes & Peres 2005; Pino-Nunes & Peres 2009). In this paper, we have characterised and analysed the brt phenotype in
the MT background and present a strong candidate gene likely to be responsible for the mutant phenotype.

2.2 MATERIALS AND METHODS

2.2.1 Plant material

Seeds were given accession numbers either prefixed with “DMF” or “WSS” and held in the Cranfield University seed collection. Seeds of MT (accession number WSS1254) and brtNIL (WSS1252) were provided by LEP (University of Sao Paulo - ESALQ); these seeds were sown and self-pollinated to bulk the seed stock (Figure 7).

![Figure 7: Seeds pedigree. Seeds produced in Brazil are represented by the black outline, while seeds produced in Cranfield (UK) are in grey. Genotyping experiments were carried using seeds from seeds package DMF143.](image)

2.2.2 Seed extraction

Red ripe tomato fruits were cut in half and the recovered seeds were submerged in 1 gL⁻¹ brewer’s pectolase (Ritchie, Burton-upon-Trent, UK) in 0.12 M HCl overnight at room temperature. The seeds were then washed thoroughly in tap water, and left to dry at room temperature for one week.
2.2.3 Plant growth

To avoid viral transmission, all seeds were sterilized by immersion for 10 minutes in an aqueous solution of 50% v/v Domestos (Unilever, UK) commercial bleach, equivalent to 2.5% w/v hypochlorite, and then thoroughly washed with tap water to remove residual bleach. After sterilization, seeds were placed in 9 cm Petri dishes containing two layers of Whatman No. 1 filter paper (85 mm diameter) soaked with 3 ml of tap water. Petri dishes were placed in the dark at 25°C for 3 days, and the chitted seeds were sown into 0.54 L round pots, 11 cm diameter x 8.6 cm depth, in a glasshouse with either 70% perlite (Sinclair, LBS Horticulture Ltd, Colne, Lancashire, UK) and 30% horticultural sand (Sporting Surface Supplies Ltd, Surrey, UK.) for phenotyping experiments or with compost (Sinclair multi-purpose Compost, LBS Horticulture Ltd, UK) for seed production. Pots were irrigated according to demand, and were fed twice a week with Hoagland solution (5 mM K$_2$SO$_4$; 1 mM H$_3$PO$_4$; 5 mM Ca(NO$_3$)$_2$; 2 mM MgSO$_4$; 100 µM EDTA Fe-Na; 42.2 µM H$_3$BO$_3$; 9.1 µM MnCl$_2$; 0.76 µM ZnSO$_4$ and 0.32 µM CuSO$_4$. pH = 5.8, adjusted with H$_3$PO$_4$) at half strength before flowering and full strength after flowering.

2.2.4 Root phenotyping

After the plant was removed from the pot and the growth medium was carefully washed away in running tap water, it was placed into a square Petri dish to be photographed. Images were captured using a Nikon Coolpix s8000, 14.2-megapixel digital camera and were analysed by ImageJ (Nakano et al. 2012; Tajima & Kato 2013). Root length was determined by setting a scale, and then drawing over the target roots with a straight or segmented line; length was provided through Image J menu commands: “analyse” > option “measure”). “Bushiness” was the term defined to describe the branching profile in the root system using three parameters obtained from the analysis at ImageJ: tap root length, total root length and number of laterals (Equation 1).
Equation 1: Mathematical expression used to calculate bushiness.

\[
bushiness = \frac{\text{Total root length} \times \text{Number of lateral roots}}{\text{Taproot length}}
\]

2.2.5 Seed phenotyping

The seeds accessions brtNIL (WSS1385, n = 2714 seeds), MT (WSS1414, n = 826) and F1 (MT x brtNIL, n = 719 and brtNIL x MT, n = 788), were scanned at 1800 dpi on a SE A3 USB 600 Pro flatbed scanner (Mustek, Hsin Chu, Taiwan). The seed area was determined from the scanned image using ImageJ (menu “analyse” > option “analyse particles”).

2.2.6 NGS genomic data generation and analysis

Genomic DNA from MT and brtNIL plants was extracted using the DNeasy plant mini kit (Qiagen; Manchester, UK), according to the manufacturer’s instructions. The MT genome was resequenced and made available by Kevei et al. (2015); the brtNIL genome, on the other hand, was resequenced on one lane of Illumina HiSeq 2500 platform generating 132 Million 126 bp paired-end reads. After the quality control performed by FastQC (Schmieder & Edwards 2011), to ensure that the reads used had high quality (on average above 33x); the raw data were mapped to the reference genome SL2.50 and genome annotation ITAG2.4, and analysed on Cranfield University’s GRID, i.e. 32 CPUs 8GB/core, running a Linux environment on the Cranfield University GRID. Firstly, the reads were aligned by the Burrows-Wheeler aligner (BWA, version 0.7.4), using default specifications; secondly, the aligned reads were compressed into a binary (bam) format (Picard tools) and then sorted and indexed by Samtools (version 0.1.19); thirdly, the GATK package (Genome Analysis Tool Kit, Broad Institute, Cambridge, USA, version 3.3.0) was used to realign the Insertions and Deletions (InDels) and variant calling (HaplotypeCaller, using default settings).
This pipeline produced variant call format (VCF) files. The annotation of the VCF files was completed by SnpEff (version 4.0) using ITAG2.40 (associated with tomato genome reference version SL2.50) available on the Sol website (http://solgenomics.net/). Finally, the variants were filtered using GATK’s *variant filtration* tool (Quality Depth < 2, Fisher Strand > 60, Mapping Quality < 40, Haplotype Score > 13 and Mapping Quality RankSum < 12.5) (Kevei et al. 2015). Tom Kurowski filtered unique polymorphisms using a custom Bash unpublished script which excluded polymorphisms shared between the data sets, similar to the mechanism of bedtools (Quinlan & Hall 2010). The filtered VCF file was uploaded to the GenoVerse Genome Browser (Bragin et al. 2012) which can be accessed on: http://elvis.misc.cranfield.ac.uk/GenoverseDemetryus/

### 2.2.7 DNA Extraction for genotyping

Genomic DNA was extracted from young leaves using a protocol based on the use of Chelex-100 (Bio-Rad, Hemel Hempstead, UK) with modifications (Wang et al. 2012; Casquet et al. 2012; Phillips et al. 2012; Walsh 1991; Turan et al. 2015; Pirilea et al. 2016). In brief, ~20-30 mg of leaf tissue from expanding leaves of < 2 cm length was disrupted in an extraction buffer containing 5% w/v Chelex 100 and soda glass balls (Smith Scientific Ltd, Kent, UK) by a mixer mill (VWR, Lutterworth, UK); the samples were subsequently incubated at 100°C for 5 min in a PCR machine with a hot-lid (MJ Research, Canada). Lastly, the samples were centrifuged for 4 minutes at 5,509 × g and stored at 4°C for short-term or at -20°C for long-term (96-well format protocol detailed in Appendix A).

### 2.2.8 Genotyping by polymerase chain reaction (PCR)-based markers

Fine mapping was performed using additional Kompetitive Allele Specific PCR KASP markers designed in the mapping interval. The reactions were performed using 1 µl of purified genomic DNA (~ 0.1 µg) in a 10 µl reaction volume
containing 1x KASP master mix buffer (LGC, Teddington, UK). Using a CFX96 (Bio-Rad, Hemel Hempstead, UK) real-time PCR machine, thermal cycling was initiated at 94°C for 15 minutes, followed by 9 cycles of 94°C for 20 secs, 61 to 55°C for 1 minute (0.6°C drop per cycle), and then 25 cycles of 94°C for 20 sec, 55°C for 1 minute, then the temperature was decreased to 37°C for 1 minute for the final step of fluorescent plate-reading. KASP assays used fluorophores FAM and HEX for distinguishing genotypes; results were analysed in the Allelic Discrimination feature of CFX manager software (BioRad). All KASP assay (primers) were developed by LGC (UK) based on polymorphism and flanking sequence data provided; see Table 2).

**Table 2:** SNPs used to develop KASP markers on chromosome 12

<table>
<thead>
<tr>
<th>Marker</th>
<th>Position (bp)</th>
<th>SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alfa3.1</td>
<td>9060801</td>
<td>G/A</td>
</tr>
<tr>
<td>Alfa4.2</td>
<td>15142101</td>
<td>T/G</td>
</tr>
<tr>
<td>Alfa13-13 β</td>
<td>17280701</td>
<td>A/G</td>
</tr>
<tr>
<td>Alfa12 β</td>
<td>56204801</td>
<td>A/C</td>
</tr>
<tr>
<td>Alfa1</td>
<td>5897101</td>
<td>G/A</td>
</tr>
<tr>
<td>Beta1</td>
<td>2587701</td>
<td>C/T</td>
</tr>
<tr>
<td>Beta1B</td>
<td>2588327</td>
<td>C/T</td>
</tr>
<tr>
<td>Beta2</td>
<td>3012101</td>
<td>C/T</td>
</tr>
<tr>
<td>Beta3</td>
<td>6464101</td>
<td>T/C</td>
</tr>
<tr>
<td>Beta4</td>
<td>7053001</td>
<td>T/A</td>
</tr>
<tr>
<td>Beta5</td>
<td>7561001</td>
<td>T/A</td>
</tr>
<tr>
<td>Beta6</td>
<td>8083101</td>
<td>T/A</td>
</tr>
<tr>
<td>Beta7</td>
<td>10581601</td>
<td>A/G</td>
</tr>
<tr>
<td>Beta8</td>
<td>12330301</td>
<td>G/C</td>
</tr>
<tr>
<td>Beta9</td>
<td>64145301</td>
<td>T/C</td>
</tr>
<tr>
<td>Gamma1</td>
<td>2793261</td>
<td>G/A</td>
</tr>
<tr>
<td>Gamma2</td>
<td>2899377</td>
<td>A/T</td>
</tr>
<tr>
<td>Gamma</td>
<td>Value</td>
<td>Type</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>Gamma 3</td>
<td>3132238</td>
<td>A/T</td>
</tr>
<tr>
<td>Gamma 4</td>
<td>4193216</td>
<td>T/A</td>
</tr>
<tr>
<td>Gamma 5</td>
<td>4711140</td>
<td>T/A</td>
</tr>
<tr>
<td>Gamma 6</td>
<td>5104880</td>
<td>G/A</td>
</tr>
<tr>
<td>Gamma 7</td>
<td>3024605</td>
<td>A/G</td>
</tr>
<tr>
<td>Gamma 8</td>
<td>3766247</td>
<td>A/T</td>
</tr>
<tr>
<td>Gamma 9</td>
<td>2919476</td>
<td>G/A</td>
</tr>
<tr>
<td>Gamma 10</td>
<td>2559167</td>
<td>G/A</td>
</tr>
<tr>
<td>Gamma 11</td>
<td>7025456</td>
<td>C/T</td>
</tr>
<tr>
<td>Delta 1</td>
<td>4721022</td>
<td>T/G</td>
</tr>
<tr>
<td>Delta 2</td>
<td>5162086</td>
<td>C/A</td>
</tr>
<tr>
<td>Delta 3</td>
<td>5599054</td>
<td>T/C</td>
</tr>
<tr>
<td>Delta 4</td>
<td>5989307</td>
<td>G/T</td>
</tr>
<tr>
<td>Delta 5</td>
<td>6682212</td>
<td>G/T</td>
</tr>
<tr>
<td>Epsilon 1</td>
<td>5266967</td>
<td>G/C</td>
</tr>
<tr>
<td>Epsilon 2</td>
<td>5309547</td>
<td>C/T</td>
</tr>
<tr>
<td>Epsilon 3</td>
<td>5472682</td>
<td>G/C</td>
</tr>
<tr>
<td>Epsilon 4</td>
<td>5518140</td>
<td>A/T</td>
</tr>
<tr>
<td>Epsilon 5</td>
<td>5611274</td>
<td>A/G</td>
</tr>
<tr>
<td>Epsilon 6</td>
<td>5733516</td>
<td>C/A</td>
</tr>
<tr>
<td>Epsilon 7</td>
<td>5811912</td>
<td>T/C</td>
</tr>
<tr>
<td>Zeta 1</td>
<td>5609035</td>
<td>G/A</td>
</tr>
<tr>
<td>Zeta 2</td>
<td>5489626</td>
<td>G/T</td>
</tr>
<tr>
<td>Zeta 3</td>
<td>5498168</td>
<td>A/T</td>
</tr>
<tr>
<td>Zeta 4</td>
<td>5508534</td>
<td>C/T</td>
</tr>
</tbody>
</table>

**2.2.9 Statistical analysis**

Sample standard deviation and standard error were calculated using SigmaPlot. ANOVA was performed to evaluate the effects of the genotypes on the phenotype using SigmaPlot with significant difference claimed if $p < 0.05$ in a Tukey and Duun’s post hoc test.
2.2.10 RNA-seq data generation and analysis

Chitted seeds from *brt*NIL recombinant 4 (DMF163) and MT (WSS1414) were allocated in rockwool blocks and covered by fine vermiculite. Using a randomised design, each rockwool block was placed into a round net basket, i.e. 5.5 cm diameter and 5 cm depth, in the low pressure aeroponics system in the glasshouse. This system works by constantly exposing the rockwool and roots to fine droplets of water (day 0 to 7 after transplanting), or half-strength Hoagland’s nutrient solution (days 8 to 26 after transplanting). 27 days after transplanting the distal part of the root system (i.e. a portion of the taproot and laterals) was sampled from the plants, snap-frozen in liquid nitrogen and stored at −80°C. Total RNA was extracted from the samples using the Spectrum plant total RNA kit (Sigma), according to manufacturer’s instructions. 12 RNA samples, i.e. 3 root- and 3 shoot samples per genotype, were sequenced in 2 lanes, using 100 bp paired-end reads by Illumina HiSeq 2000. The quality control on the RNA-seq data was performed using FastQC (Kroll et al. 2014). Then, the data was trimmed and filtered accordingly using Trimmomatic (Bolger et al. 2014), based on the following parameters: LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15 and MINLEN:40. The gene expression levels were calculated by the RSEM package (Li & Dewey 2011) using the STAR aligner version 2.4.a (Dobin et al. 2013). The expected counts for all transcripts were combined into a single matrix using the rsem-generate-data-matrix command. Differential gene expression analysis was performed in R using the limma package version 3.26.9. Normalisation was performed using the ‘Voom With Quality Weights’ method, i.e. which combine the voom observational-level weights with sample-specific quality weights in a designed experiment, with quantile normalization.
2.3 RESULTS

2.3.1 Root development and characterisation

Different traits were analysed based on the data available in the literature and visual inspection throughout the development of the brtNIL and MT. One of the most evident differences between brtNIL and MT was the taproot length. Initial experiments showed no length data overlap between both genotypes. The experiment was repeated using different (seven) time points during the plant development (Figure 8). In each point, 48 (24/genotype) root systems were analysed. The taproot of brtNIL plants is smaller than WT (MT) in all time points \((p < 0.005)\).

2.3.2 Bushiness

Sixty plants (30 per genotype) were sown (perlite and sand mix), and 25 days after germination their roots were washed and photographed. The results show that brtNIL is 62.2\% less bushy than the WT (MT). The brtNIL mutant presents a more compact (taproot length: brtNIL = 6.7 ± 0.9 and MT = 12.9 ± 0.3) root system with less (number of laterals: brtNIL = 38.6 ± 1.6 and MT = 95.5 ± 3.29) and smaller lateral roots, \(p < 0.005\) (Figure 9).
Figure 8: Root development and taproot profile in brtNIL and MT: (a) Seedlings of Micro-Tom (MT) and bushy root-NIL (brtNIL) after the radical emergence, at the top and bottom panel, respectively. (b) The root system of both parents MT and brtNIL 11 days after sowing. (c) Taproot growth during the development of brtNIL and MT. Several plants were used for each time point (n=24). Means and standard deviations are sown (seeds were sown in filter paper).
2.3.3 Seed area and maternal inheritance

As part of the phenotype characterisation, MT (WSS1414, n= 826), brtNIL (WSS1385, n= 2714) and F₁ (MT x brtNIL, n= 719 and brtNIL x MT, n= 788) seeds were analysed scanned and compared. The variation between both genotypes – brtNIL seeds are smaller (0.043±0.0002) than MT’s (0.58±0.0003); is statistically significant (p < 0.001), but cannot be used as characterization feature, due to overlapping data (Table 3).

It was also determined that this trait is maternally inherited. It is known that brt vegetative phenotype shows a recessive Mendelian segregation (Zobel 1972; Zobel 1973; Voland 1987; Zobel 1991; Pino-Nunes & Peres 2005; Pino-Nunes & Peres 2009). Thus, if the seed size was controlled by the embryo genotype, a bimodal distribution might be expected, and 25% of the seed population will exhibit small seeds and 75% normal-sized seeds. However, this is not the case. The F₁ (MT♀ x brtNIL♂) seed area data shows a sharp peak similar to the MT seed population, differently from the outcome showed by F₁ (brtNIL♀ x MT♂) seed area (Figure 10).
Figure 10: Seed area profile: Seeds from three different genotypes \(brt\)NIL (WSS1385, \(n= 2714\)), MT (WSS1414, \(n= 826\)) and \(F_1\) (MT x \(brt\)NIL, \(n= 719\) and \(brt\)NIL x MT, \(n= 788\)) were analysed compared. Statistical significance is represented by the letters \(p < 0.001\).
Table 3: Phenotype characterisation and comparison between Micro-Tom (MT) and bushy root (brtNIL) parental lines. Statistical differences are represented by letters. Errors are standard errors of the mean; populations used in each experiment per genotype are different and represented in the “N” column, except for the seed area, n = 2714 (brtNIL) and 826 (MT), 719 (F$_1$ MT x brtNIL) and 788 (F$_1$ brtNIL x MT) seeds. DAG = days after germination.

<table>
<thead>
<tr>
<th>Trait</th>
<th>N</th>
<th>MT</th>
<th>brtNIL</th>
<th>$F_1$ MT x brtNIL</th>
<th>$F_1$ brtNIL x MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of lateral roots at 35 DAG</td>
<td>30</td>
<td>95.5±3.3$^a$</td>
<td>38.7±1.7$^b$</td>
<td>n/d</td>
<td>n/d</td>
</tr>
<tr>
<td>Taproot length at 35 DAG (cm)</td>
<td>30</td>
<td>12.3±0.4$^a$</td>
<td>6.8±0.2$^b$</td>
<td>n/d</td>
<td>n/d</td>
</tr>
<tr>
<td>Total root length at 35 (cm) DAG</td>
<td>30</td>
<td>94.4±3.9$^a$</td>
<td>51.9±2.6$^b$</td>
<td>n/d</td>
<td>n/d</td>
</tr>
<tr>
<td>Root area at 34 DAG (cm$^2$), in perlite</td>
<td>10</td>
<td>127.9±6.7$^a$</td>
<td>34.8±1.7$^b$</td>
<td>n/d</td>
<td>n/d</td>
</tr>
<tr>
<td>Root area at 34 DAG (cm$^2$), in compost</td>
<td>10</td>
<td>108.4±5.3$^a$</td>
<td>50.4±3.9$^b$</td>
<td>n/d</td>
<td>n/d</td>
</tr>
<tr>
<td>Root depth at 34 DAG (cm), in perlite</td>
<td>10</td>
<td>43.6±1.9$^a$</td>
<td>19.8±1.9$^b$</td>
<td>n/d</td>
<td>n/d</td>
</tr>
<tr>
<td>Root depth at 34 DAG (cm), in compost</td>
<td>10</td>
<td>25.9±1.1$^a$</td>
<td>10.9±0.6$^b$</td>
<td>n/d</td>
<td>n/d</td>
</tr>
<tr>
<td>Root diameter at 34 DAG (cm), in perlite</td>
<td>10</td>
<td>7.6±0.5$^a$</td>
<td>9.4±0.1$^b$</td>
<td>n/d</td>
<td>n/d</td>
</tr>
<tr>
<td>Root diameter at 34 DAG (cm), in compost</td>
<td>10</td>
<td>6.1±0.1$^a$</td>
<td>7.3±0.9$^b$</td>
<td>n/d</td>
<td>n/d</td>
</tr>
<tr>
<td>Number of mature fruits (per plant)</td>
<td>10</td>
<td>21.1±0.6$^a$</td>
<td>12.8±0.7$^b$</td>
<td>n/d</td>
<td>n/d</td>
</tr>
<tr>
<td>Seed area (mm$^2$ seed$^{-1}$)</td>
<td>-</td>
<td>0.043±0.0002$^a$</td>
<td>0.058±0.0003$^b$</td>
<td>0.060±0.0003$^b$</td>
<td>0.040±0.0004$^a$</td>
</tr>
</tbody>
</table>
2.3.4 Resequencing of brtNIL and MT and delimiting the introgression

Initial genotyping experiments to find recombinants and delimit a target window were carried based on NGS data mapped to version SL2.30 of the reference genome. Results of those experiments placed the brt locus in a window of 54.4 Mbp, on chromosome 12. After the release of version SL2.50, brtNIL and MT were re-sequenced. As a result, from the Illumina sequencing of the genomic DNA, 132 million paired-end 126 bp reads for brtNIL.

To analyse both genomes individually, against the reference or against themselves, shared variants between brtNIL, MT and Heinz 1706 were excluded. Thus, the term “unique” is used to describe polymorphisms that are only present in one of the genotypes. After additional SNP filtration, by excluding polymorphisms with quality values under 210.8 and 112.8 (threshold based on the overall quality score and the validated polymorphism in this project) for SNPs and InDels, respectively, plots were created with 477,123 brtNIL and 147,244 MT unique SNPs (Figure 11) and with 81,437 brtNIL and 66,569 MT unique InDels (Figure 12).

The fact that brtNIL is a NIL of MT, has considerably reduced the amount of unique polymorphism. Very few SNPs were detected on chromosomes 3, 4, 6, 8, 9 and 10. In contrast to chromosomes 1, 2 and 12; where the number of SNPs in brtNIL is always higher than MT – although chromosomes 1 and 2 show peaks of SNPs, i.e. chr1: brtNIL 6,603 and MT 3,725 and chr2: brtNIL 4,871 and MT 1,010, the distribution pattern across the chromosome is the same. Differently than chromosome 12, location of the BRT locus, contains high peaks of SNPs across the whole chromosome. The brtNIL genome also shows more SNPs than MT across chromosome 5 and the top of chromosome 7. Interestingly, the only chromosome covered by more MT SNPs than brtNIL is chromosome 11, probably due to linkage of other loci with mnt.

The distribution of InDels follows the same pattern described above, with the exceptions of chromosomes 1 and 2, i.e. same distribution pattern, but the number of MT InDels overcomes the brtNIL variants – especially on
chromosome 1 where the peak of MT, 5,169 is much higher than brtNIL, 2098. However, Kevei et al. (2015) have also shown the same results, which would suggest that this region on chr1 is probably an introgression in Heinz. An inversion of the distribution pattern of SNPs was detected across chromosome 11, showing peaks of brtNIL 1,926 and MT 1,583.

**Figure 11:** Genome-wide SNPs uniquely discovered in *bushy root* (brtNIL) and MT. Unique = only present in one of the genotypes. Both lines were compared to the genome reference (*S. lycopersicum* cv. Heinz 1706). Polymorphisms with quality lower
than 210.8 (quality index) or shared were excluded from the analysis presented above (unique - SNPs plotted on brtNIL and MT are 477,123 and 147,244, respectively); this threshold was based on validated SNPs on this project.

**Figure 12:** Genome-wide unique InDels discovered in bushy root (brtNIL) and MT. Unique = only present in one of the genotypes. Both lines were compared to the genome reference (*S. lycopersicum* cv. Heinz 1706). Polymorphisms with quality lower than 112.8 (quality index) or shared were excluded from the analysis presented above (InDel plotted on brtNIL and MT are 81,437 and 66,569, respectively).
2.3.5 Recombinants selection and mapping interval

After analysis of the NGS data, the introgression was again delimited (to 92.8% of the chromosome – figure 13) – which means that many SNPs were not discovered during the preliminary analysis (SL2.30). Molecular markers (Alfa, Beta and Gama; Table 4) were used to genotype a population of ~1,000 F₂ plants with the objective of breaking the introgression into smaller bins and define the target region. Genotyping results reduced the introgression to 4.86% (Gamma 4; 4,192,601 - Gamma 11; 7,025,456) of its original size.

Figure 13: SNP discovery in bushy root (brtNIL) and Micro-Tom (MT) across chromosome 12. Both lines were compared to the reference genome (i.e. S. lycopersicum cv. Heinz 1706); only unique SNPs are shown (brtNIL n =477,123 and MT n = 147,244).

A larger population (n = 4,000) of F₂ plants was screened using Gamma 4 and Gamma 11, to detect new recombinants and increase the mapping resolution in the target region. Thirteen recombinants were found and later tested with the other marker sets (Epsilon and Zeta). The current search window (137 kbp – Epsilon 3; 5,472,682 - Delta 3; 5,599,054) contains 9 gene candidates.
Table 4: Recombinants genotyping and representation. Results were shown using representative recombinants and colour code (brt. brt = green; brt.+ = blue and +.+ = black) – Molecular marker are symbolized by the first letter and its positions are represented in bp. The phenotype was confirmed on the next generation (het plants).

<table>
<thead>
<tr>
<th>Marker</th>
<th>Position</th>
<th>G8</th>
<th>G6</th>
<th>D2</th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
<th>Z1</th>
<th>Z2</th>
<th>Z3</th>
<th>E4</th>
<th>D3</th>
<th>Z4</th>
<th>B3</th>
<th>G11</th>
<th>A5</th>
<th>Plant phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014_135</td>
<td>3766247</td>
<td>5104880</td>
<td>5162086</td>
<td>5266967</td>
<td>5309547</td>
<td>5492682</td>
<td>5459626</td>
<td>54098168</td>
<td>5518140</td>
<td>5599054</td>
<td>5609035</td>
<td>6464101</td>
<td>7025456</td>
<td>7561001</td>
<td>brt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014_64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>brt</td>
</tr>
<tr>
<td>2014_148</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>brt</td>
</tr>
<tr>
<td>2015_13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>het</td>
</tr>
<tr>
<td>2015_6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MT</td>
</tr>
<tr>
<td>2015_7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MT</td>
</tr>
<tr>
<td>2015_12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>het</td>
</tr>
<tr>
<td>2015_11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>brt</td>
</tr>
<tr>
<td>2015_4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>het</td>
</tr>
</tbody>
</table>

137 (kbp)

5472682 Mapping interval 5609035
2.3.6 SNP data from resequenced lines in fine-mapping

The VCF files published by other research groups (Aflitos et al. 2014; Lin et al. 2014) containing polymorphisms were used to exclude common variants, i.e. unlikely to be responsible for the *brtNIL* phenotype. The small mapping interval has 69 (64 SNPs and 5 InDels) polymorphisms – 30 (SNPs) and 4 (InDels) which are unique to the *brtNIL* genome, compared the other genomes (includes MT) – Table 5.

**Table 5:** Polymorphisms at the mapping interval. Amino-acid polarity changes are in bold. Base changes are separated by a slash. Annotation = upstream gene variant (UGV), downstream gene variant (DGV), missense variant (MV) and synonymous variant (SV).

<table>
<thead>
<tr>
<th>Polymorphism</th>
<th>Position</th>
<th>Change</th>
<th>Annotation</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP</td>
<td>5481593</td>
<td>T/C</td>
<td>UGV</td>
<td>Solyc12g014510</td>
</tr>
<tr>
<td>SNP</td>
<td>5485750</td>
<td>A/G</td>
<td>UGV</td>
<td>Solyc12g014510</td>
</tr>
<tr>
<td><strong>SNP 5485798</strong></td>
<td>Q11G</td>
<td>MV</td>
<td><strong>Solyc12g014510</strong></td>
<td></td>
</tr>
<tr>
<td>SNP</td>
<td>5485815</td>
<td>L17V</td>
<td>MV</td>
<td>Solyc12g014510</td>
</tr>
<tr>
<td><strong>SNP 5488162</strong></td>
<td>T567R</td>
<td>MV</td>
<td><strong>Solyc12g014510</strong></td>
<td></td>
</tr>
<tr>
<td>SNP</td>
<td>5488258</td>
<td>L599L</td>
<td>SV</td>
<td>Solyc12g014510</td>
</tr>
<tr>
<td>SNP</td>
<td>5488265</td>
<td>C601T</td>
<td>MV</td>
<td>Solyc12g014510</td>
</tr>
<tr>
<td>SNP</td>
<td>5488270</td>
<td>N603T</td>
<td>MV</td>
<td>Solyc12g014510</td>
</tr>
<tr>
<td>SNP</td>
<td>5489045</td>
<td>T/A</td>
<td>DGV</td>
<td>Solyc12g014510</td>
</tr>
<tr>
<td>SNP</td>
<td>5489309</td>
<td>C/T</td>
<td>DGV</td>
<td>Solyc12g014510</td>
</tr>
<tr>
<td>SNP</td>
<td>5489338</td>
<td>G/A</td>
<td>DGV</td>
<td>Solyc12g014510</td>
</tr>
<tr>
<td>SNP</td>
<td>5489357</td>
<td>T/G</td>
<td>DGV</td>
<td>Solyc12g014510</td>
</tr>
<tr>
<td>SNP</td>
<td>5489382</td>
<td>C/A</td>
<td>DGV</td>
<td>Solyc12g014510</td>
</tr>
<tr>
<td>SNP</td>
<td>5489388</td>
<td>G/A</td>
<td>DGV</td>
<td>Solyc12g014510</td>
</tr>
<tr>
<td>SNP</td>
<td>5489390</td>
<td>AC</td>
<td>DGV</td>
<td>Solyc12g014510</td>
</tr>
<tr>
<td>SNP</td>
<td>5486562</td>
<td>C/A</td>
<td>DGV</td>
<td>Solyc12g014520</td>
</tr>
</tbody>
</table>
### 2.3.7 Transcriptomics

Differentially expressed genes were identified using a threshold of 0.25 logFC and a FDR of 0.05 against possible contrasts (Table 6 and Appendix B). In the mapping interval, three genes (Solyc12g014530, Solyc12g014540 and Solyc12g014590) showed a significantly different expression pattern. Further analysis showed that the only gene which showed differential expression between WT and brtNIL in both leaves and roots was the pirin gene (Solyc12g014590) – data not shown.

### 2.3.8 Candidate gene

After the exclusion of common polymorphisms presenting in several genomes, 6 genes – Solyc12g014510, Solyc12g014520, Solyc12g014530,
Solyc12g014550 and Solyc12g014580, Solyc12g014590; were left as possible candidates (Table 6). The excessive number of polymorphisms in a possible Mutator-like element (MULE) – two of the three candidates, and the transcriptomics results suggest that only Solyc12g014590 may be the gene responsible for the brt phenotype (Table 7).

Table 6: Gene expression profile in bushy root (brtNIL) and Micro-Tom (MT)

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Overexpressed genes</th>
<th>Under expressed genes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>brtNIL (root) vs. MT (root)</td>
<td>3786</td>
<td>4440</td>
<td>8226</td>
</tr>
<tr>
<td>brtNIL (leaf) vs. MT (leaf)</td>
<td>117</td>
<td>123</td>
<td>240</td>
</tr>
</tbody>
</table>

2.4 DISCUSSION

As described, the mutant phenotype is a combination of pleiotropic effects caused by the brt gene. The phenotype includes smaller seeds, which is maternally inherited, a phenomenon that has been reviewed by many (Dolan 1984; Koelewijn & Van Damme 2005; Chaudhury & Berger 2001; Wolf & Wade 2009). Maternal effects “are the influences of the maternal genotype or phenotype on the offspring phenotype” (Wolf & Wade 2009); all the offspring, which are exposed to the same conditions on the mother plant, develop a phenotype that is dependent on the maternal genotype. It was suggested that maternal effects can be due to different factors, e.g. fertilization, the number of competing ovules per fruit and cytoplasmic and nuclear effects (Dolan 1984; Roach & Wulff 1987). The phytohormone cytokinin has been reported as a principle regulator in seed development (Bartrina et al. 2011; Swartzberg et al. 2011; Li et al. 2013; Li & Li 2016), so it could be hypothesized that the difference in seed size in brtNIL and MT might be due to the lower cytokinin sensitivity previously described by Pino-Nunes & Peres (2005).
Table 7: Candidate genes for the *brt*NIL phenotype. Unique polymorphisms = only present in *brt*NIL genome compared against the other resequenced lines.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Position</th>
<th>Size (bp)</th>
<th>Description</th>
<th>Unique Polymorphisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solyc12g014510</td>
<td>5485767 - 5488524</td>
<td>2757</td>
<td>Mutator-like transposase</td>
<td>Exon; Introns</td>
</tr>
<tr>
<td>Solyc12g014520</td>
<td>5489565 - 5494483</td>
<td>4918</td>
<td>Ulp1 protease family C-terminal catalytic domain containing protein</td>
<td>Promoter; Exons and Introns</td>
</tr>
<tr>
<td>Solyc12g014530</td>
<td>5498306 - 5503348</td>
<td>5042</td>
<td>L-aspartate oxidase</td>
<td>Intergenic</td>
</tr>
<tr>
<td>Solyc12g014540</td>
<td>5523526 - 5533996</td>
<td>10470</td>
<td>Lysine-specific demethylase 5C</td>
<td>None</td>
</tr>
<tr>
<td>Solyc12g014550</td>
<td>5553309 - 5554724</td>
<td>1415</td>
<td>Syntaxin-52</td>
<td>Intergenic</td>
</tr>
<tr>
<td>Solyc12g014560</td>
<td>5559684 - 5566991</td>
<td>7307</td>
<td>Cell division proteinftsZ</td>
<td>None</td>
</tr>
<tr>
<td>Solyc12g014570</td>
<td>5569788 - 5576277</td>
<td>6489</td>
<td>Glycerophosphoryl phosphodiesterase family protein</td>
<td>None</td>
</tr>
<tr>
<td>Solyc12g014580</td>
<td>5577417 - 5577847</td>
<td>430</td>
<td>Pollen allergen Ole e 6</td>
<td>None</td>
</tr>
<tr>
<td>Solyc12g014590</td>
<td>5578789 - 5581766</td>
<td>2977</td>
<td>Pirin</td>
<td>Promoter</td>
</tr>
</tbody>
</table>
The mutant smaller seeds can be associated with the results obtained from
taproot experiment, *brt*NIL plants show reduced taproot length. The amount of
endosperm available for the embryo is related with the seed size (Li & Li 2016).
Therefore, the small seed size could slow down the initial development of *brt*NIL
in comparison to MT due to limitation of seed reserve.

The available description of the *brt* root phenotype, highly branched root
system, (Zobel 1972; Zobel 1973; Voland 1987; Zobel 1991) is very different
than what was observed on *brt*NIL. The complex genetic background of the
original cross might be responsible for such differences. The wording used was
also challenging, once there was no formal botanical definition of what is
considered “*bushy*” – Zobel (1991) suggests that the increased number of
lateral roots could be responsible for the bushy phenotype in *brt*. However, the
total number of lateral roots in the *brt*NIL was measured (Table 3), and found to
be significantly fewer than in MT (*p* < 0.005), which would make the mutant less
bushy if the definition was simply an increase in lateral roots.

However, many aspects of the root architecture were analysed and our
results based on different models revealed that the most reliable and consistent
way to describe bushiness must consider different parameters of the root
system. Several models were applied to our data and bushiness should be
defined as the “total root length” x “number of lateral roots” / “taproot length”; as
these traits would represent a special and geometrical exploitation of the soil by
the root system. In addition, the data available from the *brt* phenotype was
collected by visual inspection by Zobel. Therefore, the data presented here are
the only quantified and statistically analysed results for this mutant. The *brt*
mutation does not make the root system bushier in the MT genetic background,
but is does make the root system smaller and more compact.

Genotyping experiments confirmed that the *BRT* locus is indeed on the
12th chromosome, as previously reported (Zobel 1972). The introgression in
*brt*NIL represents 92.8% of chr 12, however, through fine-mapping a 137 kbp
mapping interval was defined. From the nine genes located in this interval, six
of them (Solyc12g014510, Solyc12g014520, Solyc12g014530, Solyc12g014550 and Solyc12g014580, Solyc12g014590) contain unique polymorphisms compared to the 452 re-sequenced lines available (Aflitos et al. 2014; Lin et al. 2014).

The Solyc12g014510 gene annotated as a transposon, which would make it a weak candidate for brtNIL. In addition, the excess of unique and shared variants in the exons, introns and promoters of the Solyc12g014510 suggesting that it has a neutral function (Robbins et al. 2011). The lack of polymorphisms in other genes, Solyc12g014540, Solyc12g014560 and Solyc12g014570, excluded them as likely candidates; in addition to the most probably non-causative, intergenic SNPs in other three candidates, i.e. Solyc12g014530, Solyc12g014550 and Solyc12g014580.

Thus, the strongest candidates in the mapping interval are Solyc12g014520 and Solyc12g014590; both genes are differently expressed in MT and brtNIL roots. The Ulp1 gene (Solyc12g014520) mutant of Arabidopsis is reported to have severe changes in phenotype (i.e. reduced plant size and seed production, and slow development, (Murtas et al. 2003; García-Lorenzo et al. 2006; van Leeuwen et al. 2007; Hermkes et al. 2011). In addition to the similar phenotype found in the Arabidopsis Ulp1 mutant, unique non-conservative amino-acid changes were also detected in the brtNIL line. However, the high number of polymorphisms in introns, exons and promoter region of this gene suggests it is not under any selection pressure, and may not encode a functional gene product. Furthermore, Ulp1 homologues are reported to be located next to transposons, which can suggest that Ulp1 is be part of a MULE, a super family of transposons (van Leeuwen et al. 2007).

EMS is usually responsible for very few alterations across the genome (Mohd-Yusoff et al. 2015), so is very unlikely to have been responsible for the large number of polymorphisms observed. These combinations of factors weaken the position of the Ulp1 gene as viable candidate for the BRT locus.

Pirins are conserved bicupins proteins reported to be involved with a series of biological processes – e.g. apoptosis and abscission (Orzaez et al.
2001; Bar-Dror et al. 2011), germination (Lapik & Kaufman 2003; Orozco-Nunnelly et al. 2014), catabolism (Soo et al. 2007) and quercetinase activity (Tranchimand et al. 2010; Widiatningrum et al. 2015) or act as co-transcriptional factors. A single base insertion was detected in the promoter region of the pirin gene (Solyc12g014590) – Table 7. New RNA-seq data available at the new release of version SL3.0 shows expression in the promoter of the pirin in gene (SGN 2017); which suggests that an insertion in the promoter of Solyc12g014590 can, indeed, be responsible for the brtNIL phenotype.

In conclusion, the combination of results shown here suggests that the mutation in the promoter region of the pirin gene is the best candidate in the nine-gene mapping interval to be responsible for the brtNIL compact root system and small seed phenotype. These findings provide a different perspective on the role of the pirin gene in plant development, as it that a possible interaction with cytokinin, due to brtNIL lower CK sensitivity. To confirm Solyc12g014590 as responsible for the brt phenotype, transgenic experiments are required to increase or decrease the expression of the Solyc12g014590 gene, or to knock it out using CRISPR/Cas-9 genome editing.
2.5 ACKNOWLEDGEMENTS

DSF was supported by a Science without Borders PhD scholarship (246783/2012-5), funded by the Brazilian Government and the Brazilian National Council of Research and Development (CNPq). TK was supported by BBSRC grant BB/L011611/1; ZK was supported by BBSRC grant BB/L01954X/1. We are grateful to the Earlham Institute, Norwich for sequencing services. We thank the Tomato Genetics Resource Centre for providing the brt seed accessions to AJT and LEPP. We are also thankful for the previous work carried by Prof Richard Zobel (USA), Dr Lilian Pino-Nunes (Brazil) and Pauline Amiel (France), which helped us to reach the conclusions presented in this paper.
IDENTIFICATION AND CHARACTERISATION OF 
*BIFURICATE*, A NOVEL LOCUS ON CHROMOSOME 12 
CONTROLLING TRUSS BRANCHING AND FLOWER 
NUMBER IN TOMATO

Silva Ferreira, D. K.¹; Kevei, Z. L.¹; Kurowski, T.¹; De Noronha Fonseca, M.E.²,
Boiteux, L.S.², Mohareb, F.¹ and Thompson, A. J.¹

¹Cranfield Soil and AgriFood Institute, College road, Cranfield University, MK43 0AL, 
UK

²National Center for Vegetable Crops Research, CNPH – EMBRAPA Hortaliças, 
Brasília-DF, Brazil

Correspondence: Andrew Thompson. E-mail: a.j.thompson@cranfield.ac.uk

**Highlight**

At a newly discovered locus in tomato, *bifuricate*, a genomic segment introgressed from the wild species *Solanum galapagense*, increased inflorescence branching and flower number and contained two candidate genes, a MAP kinase and a BTB/POZ transcription factor.
Abstract

A naturally occurring recessive mutant named *bituricate* (*bif*) was recovered from a single seed descent program intended to produce high-yield mini-plum tomato breeding lines. Plants with the *bif* phenotype showed an 24-fold increase in inflorescence (truss) branching in comparison to the wild type (WT) control line LAM183, which presented a mean of only 0.16 branches per truss. In addition, the number of flowers per truss was 235% higher in *bif* plants than WT, and a four-day low temperature treatment stimulated branching significantly more in *bif* than in LAM183. The genomes of the LAM183 and *bif* lines were resequenced and single nucleotide polymorphism markers were designed for genetic mapping in an F₂ population. The BIF gene was mapped to a 2.01 Mbp interval of chromosome 12 containing 53 genes. All coding region polymorphisms in the interval were surveyed and five candidate genes displayed altered protein sequences, of which two genes contained stop codons predicted to disrupt gene function. Both of these genes, *Solyc12g019420* (a BTB/TAZ transcription factor) and *Solyc12g019460* (a MAP kinase), are excellent candidates for inflorescence branching control. An introgression browser was used to demonstrate that the origin of this mutant haplotype was *Solanum galapagense*.

Key words
Branching; BTB-POZ BTB-TAZ; genome resequencing; inflorescence architecture; low temperature; MAP kinase; *Solanum galapagense*; *Solanum lycopersicum*. 
3.1 INTRODUCTION

The flowering process and the number of flowers produced by a plant is controlled by a network of approximately 300 genes (Lemmon et al. 2016). An increase in the number of flowers will lead to a greater final fruit yield if supply of assimilate is abundant and reproductive growth is limited by sink strength (Périlleux et al. 2014). Conversely, flower development represents a considerable energy investment (Lemoine et al., 2013), and thus production of more flowers than can be sustained by the assimilate supply is a waste of resources and it may negatively affect the final fruit yield. In situations where assimilate supply is limiting, fruit number is inversely proportional to fruit size, and can be regulated by flower and fruit abscission in response to endogenous and environmental signals (Saglam & Yazgan 1999; Hidayatullah et al. 2008). Growers also manage fruit crops by thinning and pruning trusses to limit fruit number and to achieve larger, more uniform fruits (Cockshull & Ho 1995; Max et al. 2016).

Peduncle length, branching and the number of flowers per unit length of peduncle determines inflorescence architecture and the subsequent number of fruits that can be produced. It also has the potential to influence the uniformity of ripening that begins with proximal fruits and progresses to more distal fruits (Giovannoni 2001). Fruits borne on a highly branched structure will tend to exhibit a higher degree of synchronicity in their ripening, which may be advantageous for commercial fruit production (Ecker 1995; Moxon et al. 2008).

Flower initiation and development in tomato have been well characterized (Lippman et al. 2008), but the genetic mechanisms involved in controlling truss architecture are poorly understood. In the case of *Solanum lycopersicum* L. (tomato), the first inflorescence (truss) is initiated typically after production of 8 to 12 leaves (Lifschitz & Eshel 2006). After the appropriated flowering induction stimulus, the shoot apical meristem (SAM) transitions to the inflorescence meristem (IM), which develops floral meristems (FMs) and ultimately flowers (Lippman et al. 2008; Lozano et al. 2009). Truss architecture is extremely plastic and responsive to environmental factors. Several studies (reviewed by
Gratani 2014) have reported variations in truss architecture in response to external signals (e.g. lower temperatures increase branching and number of flowers – Calvert 1957; Calvert 1959).

In monopodial plants (e.g. *Arabidopsis*) the SAM is indeterminate and the reproductive structures are located on the sides (flanks). Sympodial plants (e.g. tomato), on the other hand, have a determinate SAM and the primary shoot is terminated with an inflorescence; subsequently a new vegetative cycle is initiated, progressing as the new primary shoot – resulting in three new leaves (vegetative nodes) and then again termination of the shoot in a new inflorescence, by the sympodial meristem (Schmitz & Theres 1999; Carmel-Goren et al. 2003; Quinet et al. 2006; Castel et al. 2010; Thouet et al. 2012). This pattern is repeated, forming consecutive sympodial segments which together constitute a sympodial shoot (Samach & Lotan 2007; Kirchoff & Claßen-Bockhoff 2013; Park et al. 2014; Astola et al. 2014).

Although the tomato and *Arabidopsis* model plants have distinct growth and flowering patterns, they share a number of orthologous genes controlling inflorescence architecture. In *Arabidopsis*, four key genes related to meristem identity have been studied in relation to the genetic control of inflorescence architecture: *TERMINAL FLOWER1* (*TFL1*), *APETALA1* (*AP1*), *UNUSUAL FLOWER ORGAN* (*UFO*) and *LEAFY* (*LFY*) (reviewed by: Weigel et al. 1992; Bradley et al. 1997; Kardailsky et al. 1999; Kobayashi et al. 1999; Hanano & Goto 2011; Chandler 2014). *TFL1* is responsible for early flowering after the development of rosette leaves; it delays the transition of IM to FM producing a terminal flower. *AP1* is upregulated in FM before floral stage 1 and it negatively regulates *TFL1* and controls FM initiation. The *UFO* gene is reported to regulate class B functions in the formation of petals and stamens (Samach et al. 1999), growth of floral primordium, and can regulate meristem identity by transforming FM back to IM (Levin & Meyerowitz 1995). *UFO* is also responsible for co-activating the *LFY* gene (Souer et al. 2008), which promotes floral fate by establishing and regulating floral identity (Kobayashi et al. 1999; Yang et al. 1999).
In tomato seven genes are known to be involved in the regulation of inflorescence architecture (Astola et al. 2014): in *FALSIFLORA (FA)*, the tomato orthologue of *LFY*, the IM fails to assume floral identity, remaining in an intermediate stage between vegetative and reproductive states (Allen & Sussex 1996; Molinero-Rosales et al. 1999; Lozano et al. 2009); in *ANANTHA (AN)*, the tomato orthologue of *UFO* (Souer et al. 2008) the IM propagates indefinitely, producing large inflorescences with immature flowers resembling cauliflower arrested inflorescences (Allen & Sussex 1996; Molinero-Rosales et al. 1999; Lozano et al. 2009); in *JOINTLESS (j)* the FM is produced, but, after 3-4 flowers are formed, the IM is converted to a vegetative meristem (VM) (Szymkowiak & Irish 1999; Mao et al. 2000); *BLIND (bl)* regulates the inflorescence architecture by controlling the formation of lateral meristems (Schmitz et al. 2002); the *UNIFLORA (UF)* gene controls the floral transition (Mero & Honma 1982; Dielen et al. 2004; Quinet et al. 2011); the *TERMINAL FLOWER (TMF)* gene upregulates vegetative growth by suppressing *FA* expression (Périlleux et al. 2014) and early inflorescences have only a single flower in *tmf* mutants (Xu et al. 2016). Of the seven known genes that regulate inflorescence architecture, only the *COMPOUND INFLORESCENCE (s)* mutant gene increases the number of flowers by increasing the number of peduncle branching points. The *s* gene encodes a transcription factor related to a gene called *WESCHEL HOMEOBOX* and it was mapped to chromosome 2L (*Solyc02g077390* – 42,337,50 bp, SL2.50 – Lippman et al. 2008; SGN 2017). Lippman et al. (2008) evaluated 6,000 tomato lines, from which 23 showed the highly-branched truss phenotype. Of those, 22 were allelic to *s* and had the same causative mutation (*s-classic* with amino acid substitution G82D), whereas a single line CC5721 showed the highly-branched truss phenotype but contained the wildtype *s* allele. The line CC5721 therefore apparently had a different and unknown genetic cause for the phenotype (Lippman et al. 2008; Krieger et al. 2010). A different mutant allele of the *s* is also reported as *s-MULTIFLORA* (amino acid substitution I94F); it shows an intermediate phenotype between *an* and *s-classic*, having a high number of flowers/branching points and arrested flower development resembling the cauliflower head tissue. In the present paper, we
describe a new tomato locus, *BIFURICATE (BIF)*, which causes a phenotype similar to *s* as well as the effect of a cold treatment on its phenotypic expression.

### 3.2 MATERIAL AND METHODS

#### 3.2.1 Plant material

Seeds of LAM183 and *bifuricate (bif)* inbred lines originated from the tomato breeding program at CNPH, Brasília-DF, Brazil. After one more cycle of self-pollination carried out under glasshouse conditions in the UK (Cranfield University), plants of *bif* and LAM183 that clearly showed high or low branching, respectively, were selected and self-pollinated one more time. It was then confirmed that the contrasting branching phenotypes were inherited in all their progenies. All subsequent phenotyping experiments were conducted using the progeny from these representative stable lines, and they were also used to produce *bif* × LAM183 F₁ and F₂ seeds for genetic mapping (Figure 14).

#### 3.2.2 Plant growth

To avoid potential pathogen transmission, all seeds were sterilized by immersion for 10 minutes in a solution of 50% v/v Domestos (Unilever, UK) commercial bleach, equivalent to 2.5% w/v hypochlorite, and then thoroughly washed with tap water to remove residual bleach. After sterilization, seeds were placed in 9 cm Petri dishes containing two layers of Whatman No. 1 filter paper (85 mm diameter) soaked with 3 ml of tap water. Petri dishes were placed in the dark at 25°C for 3 days, and the chitted seeds were sown into 3 L round pots, 15 cm diameter x 18 cm height, in a glasshouse in peat-based compost (Sinclair multi-purpose compost, LBS Horticulture Ltd, UK). Glasshouse temperature set points were 23/18°C, 14/10 hours, day/night, with light provided by high pressure sodium lamps. Pots were irrigated according to demand, and were fed twice a week with Hoagland solution (5 mM K₂SO₄; 1 mM H₃PO₄; 5
mM Ca(NO$_3$)$_2$; 2 mM MgSO$_4$; 100 µM EDTA Fe-Na; 42.2 µM H$_3$BO$_3$; 9.1 µM MnCl$_2$; 0.76 µM ZnSO$_4$ and 0.32 µM CuSO$_4$. pH:5.8, adjusted using H$_3$PO$_4$) at half concentration before flowering and full concentration after flowering.

**Figure 14**: Seeds pedigree. Seeds produced in Brasilia (Brazil) are represented by the black outline, while seeds produced in Cranfield (UK) are in grey. F$_2$ seeds used on fine mapping were from seeds package DMF269.

### 3.2.3 Cold Experiment

Sixty chitted seeds from each inbred line (*bif*: DMF21 and LAM183: DMF13) were potted in the glasshouse in three randomized blocks (n = 20 plants per genotype per block). After the transfer, the seedlings were left one week in the glasshouse (set point 23°C) for initial establishment. Groups of 24 plants (6
replicates x 2 genotypes x 2 temperatures) were transferred at five different stages of development to two illuminated growth cabinets (208 \( \mu \text{mol m}^{-2} \text{ s}^{-1} \) photosynthetic photon flux) generated by cool white fluorescent bulbs (Panasonic, Bracknell, UK) set to provide a cold treatment (15\(^\circ\)C) and a control treatment (23\(^\circ\)C) (see Figure 15 for treatment regime). After 4 days at the differential temperature, the plants were transferred back to the glasshouse, maintaining the randomised block design. After the last transfer (i.e. week 6), all plants were grown in the glasshouse at a minimum temperature of 23\(^\circ\)C until the third truss was formed and the two first were mature enough to be scored.

3.2.4 DNA extraction, NGS genomic data generation and variant calling

Genomic DNA from LAM183 and \textit{bilf} plants was extracted using the DNeasy plant mini kit (Qiagen; Manchester, UK), according to the manufacturer’s instructions. One lane was used to sequence both genomes, using 126 bp paired-end reads by Illumina HiSeq 2500. After the quality control performed by FastQC (Schmieder & Edwards 2011), to ensure that the reads used had high quality (on average above 33); the raw data were mapped to the reference genome SL2.50 and genome annotation ITAG2.4, and analysed on Cranfield University’s GRID, i.e. 32 CPUs 8GB/core, running a Linux environment on the Cranfield University GRID. Firstly, the reads were aligned by the Burrows-Wheeler aligner (BWA, version 0.7.4), using default specifications; secondly, the aligned reads were compressed into a binary (bam) format (Picard tools) and then sorted and indexed by Samtools (version 0.1.19); thirdly, the GATK package (Genome Analysis Tool Kit, Broad Institute, Cambridge, USA, version 3.3.0) was used to realign the Insertions and Deletions (InDels) and variant calling (HaplotypeCaller, using default settings). This pipeline produced variant call format (VCF) files. The annotation of the VCF files was completed by
Figure 15: Transfer points. Plants were treated at different temperatures for four days – see Materials and Methods.
SnpEff (version 4.0) using ITAG2.40 (associated with tomato genome reference version SL2.50) available on the Sol website (http://solgenomics.net/). Finally, the variants were filtered using GATK’s variant filtration tool (Quality Depth < 2, Fisher Strand > 60, Mapping Quality < 40, Haplotype Score > 13 and Mapping Quality RankSum < 12.5) (Kevei et al. 2015). Unique variants were filtered using a custom bash unpublished script which excluded polymorphisms shared between the data sets, similar to the mechanism of bedtools (Quinlan & Hall 2010). The filtered VCF file was uploaded to the GenoVerse Genome Browser (Bragin et al. 2012) which can be accessed on: [http://elvis.misc.cranfield.ac.uk/GenoverseBifuricate](http://elvis.misc.cranfield.ac.uk/GenoverseBifuricate).

### 3.2.5 SolCaps genotyping

Using the GenoVerse genome browser, 48 SolCap markers (Scott 2010; Sacco et al. 2015; Sim et al. 2015; Kim et al. 2016) observed to be polymorphic between LAM183 and *bifurcate* (2 at the top; 2 at the bottom of each chromosome – avoiding heterochromatin) were manually selected. DNA extraction from leaf tissue of individual plants of the LAM183 × *bifurcate* F2 population (n = 96) and Kompetitive Allele Specific PCR (KASP) genotyping of the 48 SolCap markers (Table 8) was performed by LGC (Teddington, UK).

<table>
<thead>
<tr>
<th>Marker ID</th>
<th>SolCap Reference</th>
<th>Chromosome</th>
<th>Position</th>
<th>SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSF1</td>
<td>solcap_snp_sl_59771</td>
<td>1</td>
<td>6229659</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF2</td>
<td>solcap_snp_sl_16925</td>
<td>1</td>
<td>18519338</td>
<td>T/G</td>
</tr>
<tr>
<td>DSF3</td>
<td>solcap_snp_sl_19068</td>
<td>1</td>
<td>43255975</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF4</td>
<td>solcap_snp_sl_22423</td>
<td>1</td>
<td>59436302</td>
<td>C/G</td>
</tr>
<tr>
<td>DSF5</td>
<td>solcap_snp_sl_17075</td>
<td>1</td>
<td>82722469</td>
<td>A/G</td>
</tr>
<tr>
<td>DSF6</td>
<td>solcap_snp_sl_17448</td>
<td>2</td>
<td>15153844</td>
<td>A/T</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>----------------------</td>
<td>----</td>
<td>------------------</td>
<td>-----</td>
</tr>
<tr>
<td>DSF7</td>
<td>solcap.snp.sl_6255</td>
<td>2</td>
<td>20334143</td>
<td>A/G</td>
</tr>
<tr>
<td>DSF8</td>
<td>solcap.snp.sl_13842</td>
<td>2</td>
<td>29361519</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF9</td>
<td>solcap.snp.sl_67542</td>
<td>2</td>
<td>35723186</td>
<td>A/G</td>
</tr>
<tr>
<td>DSF10</td>
<td>solcap.snp.sl_100592</td>
<td>3</td>
<td>7100090</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF11</td>
<td>solcap.snp.sl_1779</td>
<td>3</td>
<td>50094991</td>
<td>T/G</td>
</tr>
<tr>
<td>DSF12</td>
<td>solcap.snp.sl_21694</td>
<td>3</td>
<td>60059066</td>
<td>A/G</td>
</tr>
<tr>
<td>DSF13</td>
<td>solcap.snp.sl_20714</td>
<td>3</td>
<td>67127672</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF14</td>
<td>solcap.snp.sl_2071114</td>
<td>4</td>
<td>5020084</td>
<td>G/T</td>
</tr>
<tr>
<td>DSF15</td>
<td>solcap.snp.sl_101013</td>
<td>4</td>
<td>54543361</td>
<td>A/T</td>
</tr>
<tr>
<td>DSF16</td>
<td>solcap.snp.sl_69262</td>
<td>4</td>
<td>60067355</td>
<td>C/G</td>
</tr>
<tr>
<td>DSF17</td>
<td>solcap.snp.sl_47590</td>
<td>4</td>
<td>64717474</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF18</td>
<td>solcap.snp.sl_51106</td>
<td>5</td>
<td>8917895</td>
<td>T/G</td>
</tr>
<tr>
<td>DSF19</td>
<td>solcap.snp.sl_51600</td>
<td>5</td>
<td>20576788</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF20</td>
<td>solcap.snp.sl_69404</td>
<td>5</td>
<td>52116995</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF21</td>
<td>solcap.snp.sl_12268</td>
<td>5</td>
<td>63293808</td>
<td>C/G</td>
</tr>
<tr>
<td>DSF22</td>
<td>solcap.snp.sl_34975</td>
<td>6</td>
<td>3502385</td>
<td>A/G</td>
</tr>
<tr>
<td>DSF23</td>
<td>solcap.snp.sl_36705</td>
<td>6</td>
<td>32122683</td>
<td>T/G</td>
</tr>
<tr>
<td>DSF24</td>
<td>solcap.snp.sl_55874</td>
<td>6</td>
<td>37730097</td>
<td>A/G</td>
</tr>
<tr>
<td>DSF25</td>
<td>solcap.snp.sl_54417</td>
<td>6</td>
<td>48367262</td>
<td>A/G</td>
</tr>
<tr>
<td>DSF26</td>
<td>solcap.snp.sl_11180</td>
<td>7</td>
<td>1815826</td>
<td>A/G</td>
</tr>
<tr>
<td>DSF27</td>
<td>solcap.snp.sl_38939</td>
<td>7</td>
<td>57109419</td>
<td>A/G</td>
</tr>
<tr>
<td>DSF28</td>
<td>solcap.snp.sl_55505</td>
<td>7</td>
<td>63665011</td>
<td>A/C</td>
</tr>
<tr>
<td>DSF29</td>
<td>solcap.snp.sl_7305</td>
<td>8</td>
<td>711380</td>
<td>A/C</td>
</tr>
<tr>
<td>DSF30</td>
<td>solcap.snp.sl_56732</td>
<td>8</td>
<td>1916037</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF31</td>
<td>solcap.snp.sl_4374</td>
<td>8</td>
<td>55937442</td>
<td>T/G</td>
</tr>
<tr>
<td>DSF32</td>
<td>solcap.snp.sl_34862</td>
<td>8</td>
<td>63766185</td>
<td>A/G</td>
</tr>
<tr>
<td>DSF33</td>
<td>solcap.snp.sl_28404</td>
<td>9</td>
<td>651775</td>
<td>A/T</td>
</tr>
<tr>
<td>DSF34</td>
<td>solcap.snp.sl_45095</td>
<td>9</td>
<td>6096766</td>
<td>A/G</td>
</tr>
<tr>
<td>DSF35</td>
<td>solcap.snp.sl_29222</td>
<td>9</td>
<td>67465994</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF36</td>
<td>solcap.snp.sl_69743</td>
<td>9</td>
<td>70776666</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF37</td>
<td>solcap.snp.sl_45992</td>
<td>10</td>
<td>49125</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF38</td>
<td>solcap.snp.sl_46021</td>
<td>10</td>
<td>162566</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF39</td>
<td>solcap_snp_sl_16511</td>
<td>10</td>
<td>58189616</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF40</td>
<td>solcap_snp_sl_8834</td>
<td>10</td>
<td>64340314</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF41</td>
<td>solcap_snp_sl_62736</td>
<td>11</td>
<td>5151004</td>
<td>T/G</td>
</tr>
<tr>
<td>DSF42</td>
<td>solcap_snp_sl_732</td>
<td>11</td>
<td>10015478</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF43</td>
<td>solcap_snp_sl_2996</td>
<td>11</td>
<td>30715391</td>
<td>A/G</td>
</tr>
<tr>
<td>DSF44</td>
<td>solcap_snp_sl_53061</td>
<td>11</td>
<td>50649946</td>
<td>A/G</td>
</tr>
<tr>
<td>DSF45</td>
<td>solcap_snp_sl_41168</td>
<td>12</td>
<td>3036369</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF46</td>
<td>solcap_snp_sl_16795</td>
<td>12</td>
<td>10579861</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF47</td>
<td>solcap_snp_sl_59087</td>
<td>12</td>
<td>44105019</td>
<td>T/G</td>
</tr>
<tr>
<td>DSF48</td>
<td>solcap_snp_sl_53957</td>
<td>12</td>
<td>62088020</td>
<td>T/C</td>
</tr>
</tbody>
</table>

### 3.2.6 Linkage test

Linkage between the *BIF* locus and each of the 48 SolCap markers was detected by testing for a statistical difference from a 3:1 segregation ratio for each of the 48 SolCap markers in the sub-population of F2 plants exhibiting the *bif* phenotype. The Chi squared test was applied in Microsoft Excel.

### 3.2.7 DNA extraction for genotyping

Genomic DNA was extracted from young leaves using a protocol based on the use of Chelex-100 (Bio-Rad, Hemel Hempstead, UK) with modifications (Wang et al. 2012; Casquet et al. 2012; Phillips et al. 2012; Walsh 1991; Turan et al. 2015; Pirilea et al. 2016). In brief, ~20-30 mg of leaf tissue from expanding leaves of < 2 cm length was disrupted in an extraction buffer containing 5% w/v Chelex 100 and 3 mm soda glass balls (Smith Scientific Ltd, Kent, UK) by a mixer mill (VWR, Lutterworth, UK); the samples were subsequently incubated at 100°C for 5 min in a PCR machine with a hot-lid (MJ Research, Canada). Lastly, the samples were centrifuged for 4 minutes at 5,509 x g and stored at 4°C for short-term or at -20°C for long-term (96-well format protocol detailed in Appendix A).
3.2.8 Genotyping by polymerase chain reaction (PCR)-based markers

Fine mapping was performed using additional polymorphic KASP markers designed in the mapping interval. The reactions were performed using 1 µl of purified genomic DNA (~ 0.16 µg) in a 10 µl reaction volume containing 1x KASP master mix buffer (LGC, Teddington, UK). Using a CFX96 (Bio-Rad, Hemel Hempstead, UK) real-time PCR machine, thermal cycling was initiated at 94°C for 15 minutes, followed by 9 cycles of 94°C for 20 secs, 61 to 55°C for 1 minute (0.6°C drop per cycle), and then 25 cycles of 94°C for 20 sec, 55°C for 1 minute, then the temperature was decreased to 37°C for 1 minute for the final step of fluorescent plate-reading. KASP assays used fluorophores FAM and HEX for distinguishing genotypes; results were analysed in the Allelic Discrimination feature of CFX manager (BioRad). All KASP assays (primers) were developed by LGC (UK) based on polymorphisms and flanking sequence data provided (Table 9).

Table 9: SNPs used to develop the KASP markers on chromosome 12

<table>
<thead>
<tr>
<th>Marker ID</th>
<th>SolCap Reference</th>
<th>Position</th>
<th>SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSF50</td>
<td>solcap_snp_sl_41220</td>
<td>3240286</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF51</td>
<td>solcap_snp_sl_32654</td>
<td>4777800</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF52</td>
<td>solcap_snp_sl_63506</td>
<td>5833035</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF53</td>
<td>solcap_snp_sl_20409</td>
<td>7479839</td>
<td>C/G</td>
</tr>
<tr>
<td>DSF54</td>
<td>solcap_snp_sl_40598</td>
<td>8948057</td>
<td>A/G</td>
</tr>
<tr>
<td>DSF55</td>
<td>solcap_snp_sl_52407</td>
<td>45654100</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF56</td>
<td>-</td>
<td>51569050</td>
<td>G/A</td>
</tr>
<tr>
<td>DSF57</td>
<td>-</td>
<td>55725286</td>
<td>G/A</td>
</tr>
<tr>
<td>DSF58</td>
<td>solcap_snp_sl_42961</td>
<td>59225471</td>
<td>A/T</td>
</tr>
<tr>
<td>DSF59</td>
<td>solcap_snp_sl_53990</td>
<td>61861142</td>
<td>T/C</td>
</tr>
<tr>
<td>DSF60</td>
<td>solcap_snp_sl_53957</td>
<td>62088020</td>
<td>T/C</td>
</tr>
</tbody>
</table>
DSF61  solcap_snp_sl_16796  11159684  A/C
DSF62  solcap_snp_sl_53090  25928732  T/C
DSF63  solcap_snp_sl_38520  38590425  A/G
DSF64  solcap_snp_sl_52402  45339818  A/C
DSF65  solcap_snp_sl_18995  47510753  A/T
DSF66  solcap_snp_sl_52417  48203620  A/T
DSF67  -  50481042  C/G
DSF68  -  8566567  A/T
DSF69  solcap_snp_sl_40598  8948057  A/G
DSF70  solcap_snp_sl_16794  9973851  A/G
DSF71  solcap_snp_sl_16795  10579861  T/C
DSF72  -  10389589  T/A

3.2.9 Statistical analysis

Sample standard deviation and standard error were calculated using SigmaPlot. ANOVA was performed to evaluate the effects of the genotypes and treatments on the phenotype using SigmaPlot with significant difference claimed if p < 0.05 in a Tukey and Duun’s post hoc test. Data was transformed prior to ANOVA to ensure the validity of the normality assumption: for flower number a log(\(x\)) transformation was used and for branching point number data, containing zero values, a log(\(x + 1\)) transformation was used. Data was back-transformed prior to plotting.

3.2.10 Similarity map analysis

Firstly, an Introgression Browser (iBroswer) script was used to extract homozygous SNPs from the VCF files of the accessions available (Aflitos et al. 2014; Lin et al. 2014). Secondly, a custom Bash script used outcome file and FastTreeMP to generate the distance matrices and Newick trees in General Feature Format (GFF). Thirdly, the GFF files were organized into evenly-sized segments using another custom Bash. Lastly, the final result was plotted using
3.3 RESULTS

3.3.1 The bifurcate phenotype: initial characterization and inheritance

Multiple tomato germplasm lines and hybrids were crossed in a single seed descent (SSD) breeding programme at CNPH aiming to obtain small elongated fruits (“miniplum” or “grape” type) combining high-yield and high brix value. One inbred line was selected within the segregating populations due to its high number of branching points and flowers and it was named bifurcate due to the increased truss branching. LAM183 was an alternative inbred line developed from the same breeding programme with similar fruit morphology, high brix value, and general growth habit, but lacking the increase in truss branching. LAM183 and bifurcate lines are therefore phenotypically related, but the precise pedigree and genetic differences between them were unknown at the initiation of this study.

3.3.2 Truss development and characterisation in the contrasting inbred lines

The most obvious phenotypic difference between bif and LAM183 is the significantly higher number of flowers produced on bif trusses (Figure 16). Considering the mean of the first two trusses, bif produced 39.8 ± 1.6 flowers per truss, which was 3.3-fold higher than LAM183, which produced a mean of 12.0 ± 0.3 flowers per truss (Figure 17). The number of branching points within the truss was also affected – bifurcate trusses show a mean of 4.1 ± 1.8 branching points per truss in contrast with LAM183 which shows 0.16 ± 0.37, representing a 25.6-fold difference. Both traits are statistically significantly different between the two inbred lines (phenotype summary, Table 10).
Figure 16: Characterization of the first and second truss of LAM183 (bif<sup>a</sup>) and bifuricate (bif) inbred line plants. Phenotype was scored and numbers of flowers and branch points were recorded at 52 days after germination (DAG) in an F<sub>2</sub> population of 96 plants (A). Images of representative trusses 30 DAG are shown (B) and a schematic diagram illustrates the mean number of flowers (yellow circle) and branching points (red dots) on the first truss (C).
Figure 17: Floral production in the F\textsubscript{2} population. Phenotype was scored and numbers of flowers were recorded at 52 days after germination in an F\textsubscript{2} population of 96 plants in which 71 were scored as the wild type inbred line LAM183 and 25 as \textit{bifuricate} (\textit{bif}). The letters represent the statistical difference between both groups.

Table 10: Phenotype characterisation and comparison between LAM183 and \textit{bifuricate} (\textit{bif}) parental lines. Significant differences are represented by letters. Errors are standard errors of the mean; different population sizes were used: n = 15 (plant height), n = 20 (seeds per fruits); n = 30 (taproot length) and n = 48 (number of flowers and branching points/truss). DAG, days after germination.

<table>
<thead>
<tr>
<th>Trait</th>
<th>LAM183</th>
<th>\textit{bif}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1\textsuperscript{st} truss</td>
<td>2\textsuperscript{nd} truss</td>
</tr>
<tr>
<td>Flowers per truss</td>
<td>12.85 ± 0.42\textsuperscript{a}</td>
<td>11.08 ± 0.38\textsuperscript{b}</td>
</tr>
<tr>
<td>Branch points per truss</td>
<td>0.12 ± 0.04\textsuperscript{a}</td>
<td>0.20 ± 0.05\textsuperscript{a}</td>
</tr>
<tr>
<td>Leaves before first truss</td>
<td>6.66 ± 0.21\textsuperscript{a}</td>
<td></td>
</tr>
<tr>
<td>Plant height at 61 days (cm)</td>
<td>87.7 ± 3.7\textsuperscript{a}</td>
<td></td>
</tr>
<tr>
<td>Taproot length at 13 DAG (cm)</td>
<td>5.4±0.1\textsuperscript{a}</td>
<td></td>
</tr>
<tr>
<td>Seeds per fruit</td>
<td>46.1 ± 1.1\textsuperscript{a}</td>
<td></td>
</tr>
<tr>
<td>Seed area (mm\textsuperscript{2} seed\textsuperscript{-1})</td>
<td>5.8 ± 0.04\textsuperscript{a}</td>
<td></td>
</tr>
</tbody>
</table>
LAM183 plants were observed to exhibit branching in some trusses, and occasionally unbranched trusses were found in the first truss of \textit{bif} plants, although scoring of plants was unambiguous when looking at multiple trusses in older plants. Thus, incomplete penetrance and a genotype × environment interaction was apparently involved with the phenotypic expression of this trait. Environmental factors such as nutrient supply, irradiance and temperature are likely to be important regulators of the programmed development of sink strength.

3.3.3 Environmental interactions – Cold effect

The effects of different temperature treatments in tomato flowering is a well-established phenomenon (Calvert 1957; Calvert 1959). When LAM183 was grown in Brazil, rare truss branching was observed, but this was not consistent with lower temperature in the UK leading to occasional branching in LAM183. Therefore, an experiment was conducted to test if low temperature could induce truss branching, and if there was an interaction between genotype and temperature. LAM183 and \textit{bif} plants were transferred from a glasshouse at 23°C to growth cabinets either at 15°C (cold) or at 23°C (control) for four day periods at weekly intervals over five weeks (Figure 15) during initiation and development of inflorescences, and then the subsequent truss development was recorded (Figure 18). There were more flowers and branching points in \textit{bif} than in LAM183 (\(p < 0.001\); Appendix C) at both the first and second trusses, as expected.

The cold treatment significantly increased the number of flowers produced on the first truss of \textit{bif} (cold = 38.86 ± 2.83; control = 28.80 ± 1.78) regardless of when the plants were exposed to the lower temperature (Appendix C). The same effect was not seen in the first truss of LAM183 (cold = 12.36 ± 0.52; control = 11.85 ± 0.18), and the response of the two genotypes to cold was significantly different (\(p = 0.025\) for the genotype × treatment interaction; Appendix C).
On the second truss there was a significant interaction between treatment and transfer point ($p < 0.001$), because the cold treatment had opposite effects depending on whether the transfer was early or late in truss development: the cold treatment significantly reduced the number of flowers in *bif* up to and including the 3rd transfer (means of the first three transfers: cold = 31.26 ± 1.89; control = 39.61 ± 0.82); after this threshold, the exposure to lower temperatures increased the number of flowers produced (means of the last two transfers: cold = 43.13 ± 5.43; control = 28.27 ± 2.75). This effect of the later transfers was similar to that observed for the first truss at all five transfer points (Figure 18).

The impact of cold treatment on flower number in the second truss was much smaller in LAM183 compared to the *bif* line, but there was no statistical evidence of a different pattern of behaviour between genotypes based on an absence of a significant genotype × treatment interaction ($p = 0.838$; Appendix C) or a genotype × treatment × transfer point interaction ($p = 0.130$; Appendix C).

As expected, the number of branching points followed a similar pattern exhibited by the number of flowers. On the first truss, the *bif* plants showed a significant increase in branching points in the cold treatment (cold = 3.55 ± 0.22; control = 1.93 ± 0.27), compared to a non-significant difference in LAM183 (cold = 0.02 ± 0.02; control = 0), and there was a significant genotype × treatment interaction ($p < 0.004$; Appendix C), but no interaction with transfer point, similar to the observation for flower number.

On the second truss, the *bif* inflorescences were less branched when exposed to lower temperatures up to and including the 3rd transfer (means of the first three transfers: cold = 2.96 ± 0.22; control = 4.07 ± 0.21), and the effect was inverted by the 4th and 5th transfers (cold = 4.14 ± 0.55; control = 2.01 ± 0.23), resulting in a highly significant treatment × transfer point interaction ($p < 0.001$; Appendix C). However, as observed for flower numbers, there was no statistical evidence that the response of the two genotypes to cold was different, indeed LAM183 did have higher number of branching points after the 4th and 5th
transfers to cold treatment (cold = 0.53 ± 0.11; control = 0.26 ± 0.01), but the magnitude of the increase was much less than observed in *bif* line (Figure 18).

**Figure 18:** Cold effects on both parental lines. The numbers of flowers and branch points were recorded at 47 days after germination (DAG) in a population of 60 plants per genotype. Standard error among the replicate is represented by the bars (n = 6) – ANOVA results are shown on Appendix C: In summary, the number of flowers produced on both trusses is statistically different when *bif* plants are exposed to lower temperatures, the same effect is not seen on LAM183; the genotype effect is still present. The number of branching points is also significantly altered. Interestingly, on the second truss, an inverse effect is seen between the first 3 transfers and the last 2, probably caused by the stage of development, i.e. the second truss is formed later, thus, later transfers affect it differently.
3.3.4 The *BIFURICATE* gene: Map-based cloning

A LAM183 × *bif* F₂ population was produced and 96 plants were visually scored for truss branching and flower numbers to establish the inheritance pattern. The number of branching points for LAM183 is not always zero, but the flower number is always significantly lower than the mutant. Therefore, the phenotypic scoring was done using a combination of both traits whereby LAM183 plants have 0 or 1 branching points and max 18 flowers on the first truss. Also, if necessary, each plant could be unambiguously scored by observing multiple trusses in later development.

Twenty-three plants were scored as *bif*; seventy-three plants were scored as LAM183. A Chi-squared test indicated no statistically significant deviation from a 3:1 segregation ratio. Therefore, *bif* is determined by a single recessive gene. In order to genetically map *BIF*, the LAM183 and *bif* lines were resequenced to obtain polymorphic markers, and to provide useful information to assist CNPH-EMBRAPA’s continuing breeding work with this material.

3.3.5 Resequencing of *bif* and LAM183 inbred lines

Illumina sequencing of genomic DNA resulted in 148 million paired-end 126 bp reads for LAM183 and 138 million reads for *bif*. The raw reads were mapped to the tomato genome reference (*S. lycopersicum* cv. Heinz 1706, version SL.2.50), with a 33- and 34-fold coverage for LAM and *bif*, respectively.

Both inbred lines used as parents came from SSD populations with a relatively large genetic basis, so a high degree of polymorphisms was expected at multiple loci. Without any filtering, 96 million and 126 million changes were found between LAM and *bif* (when compared against the reference), respectively. A large number of those polymorphisms could be due to sequencing errors, mis-mapping, or regions with insufficient coverage. Nevertheless, the density of Single Nucleotide Polymorphisms (SNPs) and Insertions and Deletions (InDels) were presented to indicate the regions where LAM183 and *bif* differ genetically.
To analyse both genomes individually, against the reference or against themselves; shared variants between bif, LAM183 and Heinz 1706 were excluded. Thus, the term “unique” is used to describe polymorphisms that are only present in one of the genotypes. After additional SNP filtration, by excluding polymorphisms with quality values under 230.8 (threshold based on the overall quality score and the SolCap validated polymorphism), plots were created with the 625,887 bif and 479,247 LAM183 unique SNPs (Figure 19) and with 77,049 bif and 81,894 LAM183 unique InDels (Figure 20).

The bif genome presents a single relevant peak of SNPs (33,085) at the top of chromosome 1, in contrast with LAM183 that presents more SNPs (11,343) at the bottom of the same chromosome; at the same time, on chromosome 6 LAM183 shows a peak with 45,544 SNPs and bif does not. No relevant information was found on chromosomes 2 (s locus), 3, 9 and 10. Even with repeated SNPs deleted, the same distribution pattern was displayed on chromosomes 4 and 5. On chromosomes 7 and 8 a single small peak of SNPs was detected in each genotype (chromosome 7: bif, 6,651; LAM183, 12,720; chromosome 8: bif, 10,727; LAM183, 6,665).

The most interesting results were found on chromosomes 11 and 12, which are covered by unique SNPs. On chromosome 11, bif is very similar to Heinz 1706 (reference genome) and LAM183 is different. Whereas, on chromosome 12 LAM183 is the closest to Heinz 1706 and bif presents a large number of SNPs across the whole chromosome; suggesting that linkage blocks were preserved during the SSD breeding program. The InDels distribution pattern was consistent with the SNP plot.

Results from the variant calling were used to identify 48 SNP markers for the initial genetic mapping (supplementary table S3). Validated SNPs were used to choose existing SolCap markers; ~4 per chromosome (i.e. 2 at the top, 2 at the bottom). Linkage analyses showed that the bif phenotype was closely linked with two markers on chromosome 12 (DSF46 and DSF47), at the border of the heterochromatin (Figure 21).
Figure 19: Genome-wide SNPs uniquely discovered in *bifurcate* and LAM183. Unique = only present in one of the genotypes. Both lines were compared against the genome reference (S. lycopersicum cv. Heinz 1706); Polymorphisms with quality lower than 230.8 or repeated were excluded from the analysis presented above (unique - SNPs plotted on *bif* and LAM are 625,887 and 479,247, respectively); this threshold was defined based on validated SNPs on this project.
Figure 20: Genome-wide unique InDels discovered in *bifuricate* and LAM183. Unique = only present in one of the genotypes. Both lines were compared against the genome reference (*S. lycopersicum* cv. Heinz 1706). Polymorphisms with quality lower than 230.8 or repeated were excluded from the analysis presented above (InDel plotted on LAM and *bif* are 81,894 and 77,049, respectively).
Figure 21: SNP discovery in *bifuricate* (*bif*) and LAM183 across chromosome 12. Both lines were compared against the reference genome (i.e. *Solanum lycopersicum* cv. Heinz 1706); only unique SNPs are shown (*bif* \( n = 367,568 \) and LAM183 \( n = 38,460 \)). The four dots are representing the SolCap markers using for primarily mapping (markers characterized by the red dots are linked with the *BIF* locus).

3.3.6 High resolution gene mapping

The population of 96 F\(_2\) plants was genotyped with additional markers. Recombinant plants identified with the first batch of markers (i.e. DSF50-DSF60; see supplementary table S3) narrowed the target region from 59.05 Mbps (DSF45: 3,036,369 bp – DSF48: 62,088,020 bp) to 44.08 Mbps (DSF53: 7,479,839 bp and DSF56: 51,569,050 bp). DSF53 and DSF56 were used to screen an F\(_2\) population of 6,000 plants: 600 recombinants were recovered and phenotyped. A new batch of seven markers (i.e. DSF61-DSF67) was used to increase the resolution in the target area on the 600 recombinants. The genotyping of the recombinants reduced the target window to 3.68 Mbps.
(DSF53: 7,479,839 bp – DSF61: 11,159,684 bp). A small subset of the original recombinants (n = 19), were genotyped with another series markers (i.e. DSF68-DSF72); the result narrowed the region containing the BIF gene to 2.01 Mb (DSF68: 8,566,567 – DSF71: 10,579,861) containing 53 candidates (Table 12).

3.3.7 Candidate gene analyses

From this list of 53 genes (appendix D), many were unlikely to be functional: four transposons and 17 apparently pseudogenes (which had no expression recorded in transcriptomics databases and/or presented a single short exon). Another 26 genes presented only synonymous (15) or conservative missense changes in exons (11).

These analyses narrowed down to six remaining candidate genes. Four of which, have SNPs changing the amino-acid polarity (Solyc12g019130, Solyc12g019140, Solyc12g019200 and Solyc12g019320); therefore, considered as moderate effects on the coded proteins. The other two genes in the interval displayed polymorphisms considered as being of potentially high phenotypic impact – a frameshift mutation on the first exon of Solyc12g019420 and a stop codon gained on fifth exon of Solyc12g019460 (Table 13).

3.3.8 Germplasm origin of the genomic region encompassing the BIF locus

The SNPs in the mapping interval of bif were compared to LAM183 and other 63 tomato accessions (Figure 22); this region in bif was found to be very similar to S. galapagense accessions LA1044, LA1401, LA0483, and LA0528. At a chromosomal level, the differences observed between the other wild species (S. habrochaites, S. arcanum, S. pennellii, S. chilense, S. peruvianum, S. huaylasense, S. corneliomulleri, S. chmielewskii, and S. neorickii) were suggesting that there was an introgression from S. galapagense on chromosome 12 of the bif line (Figure 23 and Figure 24).
Figure 22: Similarity tree based on the SNPs in the defined mapping interval (2.01 Mbp) on chromosome 12. Accessions of Solanum wild species (64) were grouped according to how similar they were in comparison to the bifurcate genome. Black stars represent the lines selected to be part of Figure 25 and Figure 26.
**Table 11** Genotyping of recombinants for fine mapping of *bif*. Results were summarised using representative recombinants and a colour code (*bif*. *bif* = purple; *bif*.+ = pink) – DSF marker are symbolized by the first letter and positions are represented in bp.

<table>
<thead>
<tr>
<th>Marker</th>
<th>D51</th>
<th>D53</th>
<th>D68</th>
<th>D69</th>
<th>D70</th>
<th>D72</th>
<th>D71</th>
<th>D61</th>
<th>D62</th>
<th>D63</th>
<th>D58</th>
<th>D59</th>
<th>D60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>4777800</td>
<td>7479839</td>
<td>8566567</td>
<td>8948057</td>
<td>9973851</td>
<td>1039589</td>
<td>10579861</td>
<td>11159684</td>
<td>25928732</td>
<td>38590425</td>
<td>55725286</td>
<td>59225471</td>
<td>62088020</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plant</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015_1255</td>
<td><em>bif</em></td>
</tr>
<tr>
<td>2015_612</td>
<td><em>bif</em></td>
</tr>
<tr>
<td>2016_1674</td>
<td><em>bif</em></td>
</tr>
<tr>
<td>2016_1689</td>
<td><em>bif</em></td>
</tr>
<tr>
<td>2016_67</td>
<td>LAM183</td>
</tr>
<tr>
<td>2016_269</td>
<td>LAM183</td>
</tr>
</tbody>
</table>

2.01 Mbp

8566567 Mapping interval 10579861
Figure 23: SNP similarities map in *bifuricate*, LAM183, Heinz 1706 and other genomes. Each signal covers a 50kb bin. The grey scale represents the differences in the bin, i.e. higher number of different SNPs result in darker tone. The red box highlights *bif* and the 4 accessions of *Solanum galapagense* (LA1044, LA1401, LA0483 and LA0528). This figure can be downloaded for visualization from: http://dx.doi.org/10.17862/cranfield.rd.4721560.
Figure 24: High contrast SNP similarities map in *bifuricate*, LAM183, Heinz 1706 and 28 other lines. Each signal is equivalent to a 50kb bin. The grey scale represents the differences in the bin, i.e. more different SNPs in the site, darker the tone. 23 accessions of wild species (i.e. *S. habrochaites*, *S. arcanum*, *S. pennellii*, *S. chilense*, *S. peruvianum*, *S. huaylasense*, *S. corneliomulleri*, *S. chmielewskii* and *S. neorickii*) were deleted from this analysis to increase the contrast of the lines that are similar to *bif*; the red box highlights *bif* and the other accessions of *S. galapagense*. Selected accessions of *S. lycopersicum* can be seen on Figure 22. This figure can be downloaded for visualization from: http://dx.doi.org/10.17862/cranfield.rd.4721560.
Table 12: Candidate genes for the bif phenotype

<table>
<thead>
<tr>
<th>Gene</th>
<th>Position</th>
<th>Size (bp)</th>
<th>Type of polymorphism in coding region</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solyc12g019130</td>
<td>9244294 - 9245985</td>
<td>1691</td>
<td>Non-conservative change</td>
<td>T381A and G194R</td>
</tr>
<tr>
<td>Solyc12g019140</td>
<td>9293132 - 9295123</td>
<td>1991</td>
<td>Non-conservative change</td>
<td>D206G</td>
</tr>
<tr>
<td>Solyc12g019200</td>
<td>9547294 - 9548500</td>
<td>1206</td>
<td>Non-conservative change</td>
<td>Q29P and G35C</td>
</tr>
<tr>
<td>Solyc12g019320</td>
<td>9971385 - 9976838</td>
<td>5453</td>
<td>Non-conservative change</td>
<td>I249T</td>
</tr>
<tr>
<td>Solyc12g019420</td>
<td>10234086 - 10236575</td>
<td>2489</td>
<td>Frameshift InDel</td>
<td>7^8insT</td>
</tr>
<tr>
<td>Solyc12g019460</td>
<td>10385358 - 10395971</td>
<td>10613</td>
<td>Stop gained</td>
<td>L291*</td>
</tr>
</tbody>
</table>

After removing them from the analysis (Figure 24), it was clear that even though the bif-containing locus is similar to the S. galapagense, some regions (0.1 to 2.5 Mbp and ~63.5 to 66 Mbp) were closer related to S. lycopersicum than to S. galapagense. It was also possible to notice that the S. galapagense accession LA0528 (sequenced and published by Lin et al. 2014) is different from the other three S. galapagense accessions (sequenced and published by Aflitos et al. 2014). Similar results can be obtained, specifically for the mapping interval (Figure 26); using the same lines selected on Figure 22 (for all lines, see figure 25), S. galapagense LA0528 displays variants that are not present in LA0483, LA1401, LA1044 and bif. However, it was verified that all four accessions display the same causative SNP on the fifth exon of Solyc12g019460. The InDel causing a frameshift mutation in Solyc12g019420 was detected by Aflitos et al. (2014) in all three sequenced accessions.
However, we were not able to verify if LA0528 displayed the same InDel as, the VCFs provided by Lin at al. 2014 only contain SNPs.

**Figure 25:** SNP heat map in the mapping interval. Each bin is equivalent to 10kb. The grey scale represents the differences in the bin, i.e. more different SNPs in the site, darker tone. The black box highlights the similarity between *bif* and the other *Solanum galapagense* (LA1044, LA1401, LA0483 and LA0528) accession
Figure 26: SNP heat map in the *bifuricate* (*bif*) gene mapping interval using the lines that are more similar to *bif* locus. (complete analysis with all 64 accessions Figure 25); The grey scale represents the differences in the bin when compared to the *bif* line, i.e. more different SNPs in the site, darker tone. Each bin is equivalent to 10 kb.
3.4 DISCUSSION

Although the transition between the vegetative and reproductive stages is well known (Yang et al. 1999; Kobayashi et al. 1999; Quinet et al. 2006; Samach & Lotan 2007; Kwiatkowska 2008), the mechanisms synchronizing the inflorescence development on the initial and the sympodial segments are still poorly understood (Hong et al. 2010; Quinet et al. 2011; Thouet et al. 2012; Poyatos-Pertiñez et al. 2016). Thus, at this moment we were not able to explain the factors responsible for the significant difference on flower production between the first and second truss on LAM183, or why this phenomenon is not shown on bif plants.

And yet, the plant and environment interaction results gave us a better perspective on the mutant characterization. Tomato plants, in general, are very sensitive to the cold (Shah et al. 2016). The cold stress initiates reactions at different levels (i.e. cellular, molecular and physiological) to adapt to the new environment (La Porta et al. 2015). As shown on Figure 18, the interaction between lower temperature, genotype and different exposures is very complex. But, consistently with what was demonstrated by Calvert (1957 and 1959), lower temperatures increase the number of flowers in tomato (Lewis 1953; Calvert 1957; Calvert 1959; Sawhney 1983; Adams et al. 2001). Interestingly, the interaction of the genotype and treatment was also significant.

The BIF initial mapping established linkage with two markers close to the heterochromatin border, on chromosome 12. The phylogenetic tree of the mapping interval suggests that the source of BIF locus is S. galapagense. In the mapping interval specifically, bif is clustered with S. galapagense LA1044 (Figure 22), due to the high level of similarities between both genomes. Interestingly, on the Isla Santa Cruz (origin of the less similar line, LA0528) the vegetation is significantly denser than on the other Galapagos islands, which might have contributed for the superior genetic variations among the S. galapagense accessions (Figure 27). These germplasms (LA0483, 1956; LA1401, 1971; LA1044, 1966 and LA0528, 1957) were collected decades ago.
and no detailed descriptions of the inflorescence architecture were provided to the tomato germplasm collection of TGRC.

The fact that *S. galapagense* introgression does not cover the top or bottom chromosome is very interesting. Quantitative trait loci (QTLs) related to plant height, number of nodes, flower node number, leaf number, and plant fresh weight were described in LA0483 (Goldman et al. 1995; Paran et al. 1997), although they are all located close to the top (TG68, TG50A, TG111 and TG367) and bottom (TG296, TG350 and TG602) of chromosome 12 (Goldman et al. 1995; Paran et al. 1997; SGN 2017).

**Figure 27**: Origin of the *Solanum galapagense* accessions. The locations shown above are based on the description provided by the germplasm donor to the Tomato Genetics Resource Centre (TGRC, 2017).
Due to the proximity to the heterochromatin, approximately 6000 F2 plants were needed to be screened in order to define the map position of \textit{bif} to a 2 Mbp segment, which is a region with an overall low gene density (~1 expressed gene per 62,000 bp). Results from the molecular markers on the current window, narrowed the number of possible candidates to 53 genes. Of those, 48 were discarded as candidates for four reasons: 4 were transposons, 17 artefacts, 15 had synonymous changes and 11 conservative missense changes; consolidating 6 genes as candidates for the \textit{BIF} locus. Although, we cannot exclude \textit{Solyc12g019130}, \textit{Solyc12g019140}, \textit{Solyc12g019200} and \textit{Solyc12g019320} it is not likely that they are responsible for the phenotype, especially when considering the tissue where these genes were reported to be expressed (mainly on roots, seeds and leafs, respectively). In addition, the putative effect caused by those changes on \textit{Solyc12g019130}, \textit{Solyc12g019140}, \textit{Solyc12g019200} and \textit{Solyc12g019320} is moderate when compared to high effect polymorphisms present in the other two candidate genes (\textit{Solyc12g019420} and \textit{Solyc12g019460}).

The frameshift insertion in the first codon of \textit{Solyc12g019420} (7^8insT) changes the next eleven amino-acids until gaining a stop codon, clearly abolishing the protein function and causing loss of BTB domain. The stop codon gained due to a SNP on \textit{Solyc12g019460} (L291*) disrupt the amino-acid sequence of the MAPK, in the PK\_c and STKc domains. In addition to the high quality score of the mentioned polymorphisms (\textit{Solyc12g019420}: 7^8insT; and \textit{Solyc12g019460}: L291*), both were also detected by the tomato genome sequencing consortium (Aflitos et al. 2014), which suggests that they are not sequencing/mapping errors, but \textit{“real variants”}.

Blast results showed that \textit{Solyc12g019460} and MPK6 in \textit{Arabidopsis} contain similar domains. Mitogen-activated protein kinases (MAPKs) are involved in the regulation of many different attributes in eukaryotes (Bush & Krysan 2007), including plant growth and development (Müller et al. 2010; Xu & Zhang 2015). In 2012, it was reported that due to its role on cell proliferation, MPK6 was involved in the inflorescence stem and pedicel elongation in \textit{Arabidopsis}, which would affect its inflorescence architecture (Meng et al.
2013). In the homologous gene in tomato, a mutation in *bif* is creating a stop codon in the middle of the polypeptide sequence, most probably resulting in a dysfunctional protein, implicating MPK6 as a strong candidate for the *BIF* locus (Table 13).

At the same time, a single base insertion (thymine – 10,236,554 bp) caused a frameshift on the transcript of *Solyc12g019420*, leading to the change of Arg to a stop codon. *Solyc12g019420* is a BTB (Broad complex, Tramtrack, and Bric-a-brac)/TAZ (Transcriptional adapter zinc binder) protein domain – also known as BTB/POZ (Pox virus zinc-finger). This class of protein was reported to be involved in a different number of signalling pathways such as cytoskeleton regulation (Kang et al. 2004), transcriptional repression (Ahmad et al. 2003), tetramerization and gating of ion channels (Du & Poovaiah 2004), protein ubiquitination/degradation (Mandadi et al. 2009). Recently published results from the protein-protein interactions of the tomato TERMINATING FLOWERING (TMF) gene indicated its association with three transcription cofactors with BTB domains – SIBOP1 (*Solyc04g04040220*), SIBOP2 (*Solyc10g0794460*) and SIBOP3 (*Solyc10g079750*), demonstrated that the interaction of this three SIBOP are involved with the control of inflorescence architecture in tomato due to its dominant role in oligomerization (Xu et al. 2016) – also making *Solyc12g019420* as a strong candidate for the *BIF* locus.

In conclusion, to verify which of the two strongest candidates are responsible for the *bif* phenotype (i.e. highly branched trusses with more flower and bigger seeds), more recombinants should be generated to separate both mutated genes. But, as they are closely located on chromosome 12, to find recombinants between the two candidate genes would be extremely laborious. Thus, transgenic experiments related to the individual genes seem to be the best alternative for the functional analyses of both genes. Polymorphisms scored within candidate genes with high phenotypical effects are uncommon. In our mapping interval we located two, of them, in genes with highly conserved domains. Thus, the *bif* phenotype might be caused by one of these more likely candidate genes or the combined effect of the truncated proteins of MPK6 and BTB/TAZ. In addition, extensive characterization experiments on diverse *S.*
galapagense germplasm indicated that trusses with 2-3 branches were frequently found in different accessions (Darwin et al. 2003). Unfortunately, the accessions were branching was found are not mentioned, nevertheless supports the data presented in the present study.
3.5 ACKNOWLEDGEMENTS

This project was initiated by a UK-Brazil Sustainable Research Networks Grant from the British Council awarded to AT and LSB. DSF was supported by a Science without Borders PhD scholarship (246783/2012-5), funded by the Brazilian Government and the Brazilian National Council of Research and Development (CNPq). TK was supported by BBSRC grant BB/L011611/1; ZK was supported by BBSRC grant BB/L01954X/1. We thank Ryan Ramsey of Syngenta for highlighting truss branching as a commercially interesting trait prior to the initiation of this project. We are grateful to the Earlham Institute, Norwich for sequencing services.
4 GENERAL DISCUSSION and FUTURE WORK

Breeders are constantly examining traits found in plants that perform better than others in different conditions, e.g. drought resistance, (Chapin et al. 1993; Barbosa et al. 2014). Depending on the studied trait, the research focuses on the above- or belowground part of the plant (Khamzina et al. 2006; Noulèkoun et al. 2017). Even though, the response in different parts of the plant is reported to be different due to several reasons, the whole-plant perspective is required to understand the metabolic control of the plant architecture.

Some research groups are now pursuing genes reported to act on particular traits to understand its global effects on plant architecture. For instance, the CLAVATA (CLE) gene family in Arabidopsis, which have a dominant role in plant growth, development and environmental responses (Endo et al. 2014; Wang et al. 2016). Recently, it was reported the interaction between the CLE peptides and cytokinin (Qiang et al. 2013; Endo et al. 2014; Wang et al. 2016), and thus, the effect caused on the whole plant architecture, e.g. overexpression of CLE14/CLE20, produced a plant with a reduced root phenotype (Meng & Feldman 2010). Another example of a gene controlling the general architecture would be the C-TERMINALLY ENCODED PEPTIDE, CEP (Ohyama et al. 2009). Recent work by Delay et al.( 2013) reported that this family of regulatory peptides is involved in the regulation of different aspects of the root-, e.g. LR development, and shoot architecture, e.g. SAM maintenance (Christina Delay et al. 2013).

The work carried in this PhD project is especially important when considering Solyc12g019460 (SIMAPK1, candidate gene for bif). MAP kinase signaling cascades regulate many stress and defense responses in plants, but they also control many aspects of plant growth and development. Tropotrophic effects are clear in the mutant, experiments to link the brt root architecture to the shoot architecture are still to be done. I would suggest experiments with treatments of the mutant with different concentrations of P and N to verify the impact in the root architecture and how this would reflect the shoot development and architecture.
In addition, grafting experiments would be interesting to connect the shoot and root phenotype. Transcriptomics results showed that the candidate gene \textit{(Solyc12g014590)} is differently expressed in \textit{brt}NIL tissues; the gene is downregulated in \textit{brt}NIL roots, when compared to \textit{brt}NIL leaves, and MT roots and leaves. It would be also interesting to see if the \textit{brt} shoot phenotype, i.e. dwarf with reduced yield, remains the same with WT root system. Lastly, a study of the genetic interaction of \textit{brt} with other root mutants (e.g. \textit{dwarf} and \textit{rosette}) in double and triple mutants would extend our understanding about the root architecture regulation and its molecular basis.

Even though the genetic work on \textit{brt} is slightly more advanced due to the small mapping window, the phenotypic characterisation of the \textit{bifuricate} (\textit{bif}) accomplished; including its reaction to environmental effects (i.e. low temperature exposure). Although both candidate genes \textit{(Solyc12g019420} and \textit{Solyc12g019420}) are expressed in the root and shoot (INRA 2017), grafting experiments would be useful once again to characterize the mutant shoot behaviour in a different rootstock – i.e. whether a vigorous root system increase or reduce the yield, once the sink strength is augmented. I would suggest an experiment to test different levels of N and P again to see how the alterations caused in the root architecture affect the phenotype. As mentioned before, there are similarities between the \textit{bif} and \textit{s} phenotype; to verify if the phenotype (more branching points and flowers) is enhanced by the other locus (epistasis), a cross (F1) between them, followed by the selection and characterization of double mutants on the F2 population would be required.

Genome editing, e.g. CRISPR/Cas-9, and transgenic experiments would reveal in both mutants if \textit{Solyc12g014590} as responsible for the \textit{brt}NIL and \textit{Solyc12g019420} and/or \textit{Solyc12g019420} for \textit{bif} phenotype.
REFERENCES

Abe, S. et al., 2014. Carlactone is converted to carlactonoic acid by MAX1 in Arabidopsis and its methyl ester can directly interact with AtD14 in vitro. Proceedings of the National Academy of Sciences, 111(50), pp.18084–18089.


Bell, P.R., 1992. Green plants: their origin and diversity. *Green plants: their origin and diversity*.


Bénard, C. et al., 2009. Effects of low nitrogen supply on tomato (*Solanum lycopersicum*) fruit yield and quality with special emphasis on sugars, acids, ascorbate, carotenoids, and phenolic compounds. *Journal of Agricultural and Food Chemistry*, 57(10), pp.4112–4123.


Bragin, E., 2012. Interactive genome browser based on html5 and related web technologies. Cranfield University, UK.

Busch, B.L. et al., 2011. Shoot branching and leaf dissection in tomato are regulated by homologous gene modules. The Plant Cell, 23(10), pp.3595–3609.


Cingolani, P. et al., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, 6(2), pp.80–92.


DESA, 2014. Department of economic and social affairs, (June/16).


Ito, Y. et al., 2015. CRISPR/Cas9-mediated mutagenesis of the RIN locus that regulates tomato fruit ripening. Biochemical and Biophysical Research Communications, 467(1), pp.76–82.


Kevei, Z. et al., 2015. Resequencing at ≥ 40-fold depth of the parental genomes of a *Solanum lycopersicum* × *S. pimpinellifolium* recombinant inbred line population and characterization of frame-shift indels that are highly likely to perturb protein function. *G3; Genes|Genomes|Genetics*, 5(5), pp.971–981.


Kim, H.B. et al., 2006. The regulation of *DWARF4* expression is likely a critical mechanism in maintaining the homeostasis of bioactive brassinosteroids in *arabidopsis*. *Plant Physiology*, 140(2), pp.548–557.


Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.


Müller, J. et al., 2010. Arabidopsis MPK6 is involved in cell division plane control during early root development, and localizes to the pre-prophase band,


Poyatos-Pertiñe, S. et al., 2016. Genetic interactions of the unfinished flower development (ufd) mutant support a significant role of the tomato UFD gene in regulating floral organogenesis. *Plant Reproduction*.


Qi, X. et al., 2015. De novo transcriptome sequencing and analysis of *Coccinella septempunctata* L. in non-diapause, diapause and diapause-terminated states to identify diapause-associated genes. *BMC Genomics*, 16(1).


Schwarz, D. et al., 2013. Rootstocks can enhance tomato growth and quality characteristics at low potassium supply. *Scientia Horticulturae*, 149, pp.70–79.


Shirasawa, K. et al., 2010. SNP discovery and linkage map construction in cultivated tomato. *DNA Research*, 17(6), pp.381–391.


Solanaceae project, 2017. International genome sequencing project (Feb). Available at: http://solgenomics.net/solanaceae-project/index.pl.

Soo, P.-C. et al., 2007. Pirin regulates pyruvate catabolism by interacting with the pyruvate dehydrogenase E1 subunit and modulating pyruvate dehydrogenase activity. Journal of Bacteriology, 189(1).

Sorefan, K. et al., 2003. MAX4 and RMS1 are orthologous dioxygenase-like genes that regulate shoot branching in Arabidopsis and pea. Genes Dev, 17.


Soyk, S. et al., 2017. Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato. Nature Genetics, 49(1).


Zsögön, A. et al., 2017. Genome editing as a tool to achieve the crop ideotype and de novo domestication of wild relatives: Case study in tomato. *Plant Science*, 256.
APPENDICES

Appendix A:

Method for DNA extraction with Chelex 100 in 96-well format

A.1 Plate preparation

1. Prepare the extraction buffer (5% w/v Chelex): add 2.5 g of Chelex 100 (Bio-Rad, Hemel Hempstead, UK) into a 50 ml Falcon tube and make up to 50 ml with freshly obtained, deionized (MilliQ) water ( autoclaving not required).

2. Add two 4 mm soda glass balls (Smith Scientific Ltd, Kent, UK) to each well of a 96-well deep-well plate (Starlabs, Milton Keynes, UK).

At room temperature, add a small piece (~20-30 mg) of young leaf tissue (e.g. small developing leaves from the apex which are less than 15 mm in length) to each well, placing it on top of the glass balls. Differences in the amount of plant material across the plate will affect final concentration (NB. Be careful with the labelling and plate orientation).

A.2 Tissue disruption and DNA extraction

1. Add 300 µl of the extraction buffer (Chelex 5% w/v) to each well;

2. Close the plate with a sealing mat (Starlabs) using a seal applicator (Starlabs). Alternatively, you can seal the plate with an adhesive seal (Starlabs), which is easier to use and is disposable.

3. Disrupt the tissue for 30 seconds at a frequency of 30 Hz;

4. Transfer 100 µl of supernatant to a PCR plate (Starlabs) and seal the wells using 8-strip PCR Caps (Starlabs).

5. Incubate the plate at 100°C for 5 min in a PCR machine (MJ Research, Canada) with a heated lid to prevent evaporation.
6. Centrifuge (5430R, Eppendorf, Stevenage, UK) using a swing-out rotor at 3800 rpm for 4 minutes at ambient temperature or 20°C.

Plates can be stored at 4°C for short-term (< 1 month) or frozen at -20°C for longer term. If using after storage, re-spin to bring down condensation and ensure the pellet is secure at the bottom of the tube.
## Appendix B Transcriptomics data

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>logFC</th>
<th>Average Expression</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Differently expressed gene in root tissues</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solyc12g014530</td>
<td>-0.5418533</td>
<td>2.076818</td>
<td>-0.004924884</td>
</tr>
<tr>
<td>Solyc12g014540</td>
<td>-0.6101559</td>
<td>5.381991</td>
<td>-0.004924884</td>
</tr>
<tr>
<td>Solyc12g014590</td>
<td>-0.8864737</td>
<td>3.076808</td>
<td>5.679906e-05</td>
</tr>
</tbody>
</table>

(brtNIL vs. MT)

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>logFC</th>
<th>Average Expression</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differently expressed genes between root and leaf tissues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solyc12g014530</td>
<td>-1.536916</td>
<td>2.076818</td>
<td>2.292134e-06</td>
</tr>
<tr>
<td>Solyc12g014540</td>
<td>0.4795581</td>
<td>5.381991</td>
<td>0.01629394</td>
</tr>
<tr>
<td>Solyc12g014560</td>
<td>-0.8852245</td>
<td>2.643699</td>
<td>3.730434e-05</td>
</tr>
<tr>
<td>Solyc12g014590</td>
<td>3.204554</td>
<td>3.076808</td>
<td>1.822179e-06</td>
</tr>
</tbody>
</table>

(brtNIL vs. brtNIL)

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>logFC</th>
<th>Average Expression</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differently expressed genes between root and leaf tissues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solyc12g014530</td>
<td>-0.7690019</td>
<td>2.076818</td>
<td>0.0005604215</td>
</tr>
<tr>
<td>Solyc12g014540</td>
<td>0.8085703</td>
<td>5.381991</td>
<td>0.0003326477</td>
</tr>
<tr>
<td>Solyc12g014560</td>
<td>-0.6221753</td>
<td>2.643699</td>
<td>0.0005505298</td>
</tr>
<tr>
<td>Solyc12g014580</td>
<td>3.253546</td>
<td>-3.203736</td>
<td>0.004786132</td>
</tr>
<tr>
<td>Solyc12g014590</td>
<td>4.082314</td>
<td>3.076808</td>
<td>1.410241e-07</td>
</tr>
</tbody>
</table>
### Appendix C

#### C.1 Cold transfer experiment: summary of ANOVA analysis

<table>
<thead>
<tr>
<th>Flowers</th>
<th>1st truss</th>
<th>2nd truss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>LSD (5%)</td>
</tr>
<tr>
<td>Genotype (G)</td>
<td>&lt;.001</td>
<td>0.1120</td>
</tr>
<tr>
<td>Treatment (T)</td>
<td>0.004</td>
<td>0.1120</td>
</tr>
<tr>
<td>Transfer point (TP)</td>
<td>0.126</td>
<td>0.1771</td>
</tr>
<tr>
<td>G x T</td>
<td>0.025</td>
<td>0.1584</td>
</tr>
<tr>
<td>G x TP</td>
<td>0.574</td>
<td>0.2505</td>
</tr>
<tr>
<td>T x TP</td>
<td>0.909</td>
<td>0.2505</td>
</tr>
<tr>
<td>G x T x TP</td>
<td>0.611</td>
<td>0.3542</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Branching points</th>
<th>1st truss</th>
<th>2nd truss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>LSD (5%)</td>
</tr>
<tr>
<td>Genotype (G)</td>
<td>&lt;.001</td>
<td>0.1431</td>
</tr>
<tr>
<td>Treatment (T)</td>
<td>0.001</td>
<td>0.1431</td>
</tr>
<tr>
<td>Transfer point (TP)</td>
<td>0.435</td>
<td>0.2263</td>
</tr>
<tr>
<td>G x T</td>
<td>0.004</td>
<td>0.2024</td>
</tr>
<tr>
<td>G x TP</td>
<td>0.646</td>
<td>0.3201</td>
</tr>
<tr>
<td>T x TP</td>
<td>0.929</td>
<td>0.3201</td>
</tr>
<tr>
<td>G x T x TP</td>
<td>0.749</td>
<td>0.4526</td>
</tr>
</tbody>
</table>
## Appendix D

### D.1 Candidates for *BIFURICATE*

<table>
<thead>
<tr>
<th>Gene</th>
<th>Polymorphisms</th>
<th>Apparent artefact</th>
<th>Expression data (TomExpress)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solyc12g017930.1.1</td>
<td>Promoter</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g017950.1.1</td>
<td>Intron</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g017960.1.1</td>
<td>Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g017970.1.1</td>
<td>Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g018970.1.1</td>
<td>Exon</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g018980.1.1</td>
<td>-</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g018990.1.1</td>
<td>Promoter and Intron</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019000.1.1</td>
<td>Promoter and Intron</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019010.1.1</td>
<td>Intron and Exon</td>
<td>NO</td>
<td>Shoot; Flowers; Fruits</td>
</tr>
<tr>
<td>Solyc12g019020.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Flower buds; Mature fruits</td>
</tr>
<tr>
<td>Solyc12g019040.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Flowers; roots</td>
</tr>
<tr>
<td>Solyc12g019050.1.1</td>
<td>Promoter</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019060.1.1</td>
<td>Promoter</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019070.1.1</td>
<td>Promoter and Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019080.1.1</td>
<td>Intron</td>
<td>NO</td>
<td>Flowers</td>
</tr>
</tbody>
</table>

174
<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>NO/YES</th>
<th>Tissue/Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solyc12g019090.1.1</td>
<td>Intron and Exon</td>
<td>NO</td>
<td>Shoot; Fruit</td>
</tr>
<tr>
<td>Solyc12g019100.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Leaf; Fruit</td>
</tr>
<tr>
<td>Solyc12g019110.1.1</td>
<td>Intron</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019120.1.1</td>
<td>Promoter and Exon</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019130.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Root</td>
</tr>
<tr>
<td>Solyc12g019140.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Root</td>
</tr>
<tr>
<td>Solyc12g019150.1.1</td>
<td>Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019160.1.1</td>
<td>Promoter and Intron</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019170.1.1</td>
<td>Prom, Intr and Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019180.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019190.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Seeds; Fruit</td>
</tr>
<tr>
<td>Solyc12g019200.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Mature fruit</td>
</tr>
<tr>
<td>Solyc12g019210.1.1</td>
<td>Prom, Intr and Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019220.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Root</td>
</tr>
<tr>
<td>Solyc12g019230.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Root</td>
</tr>
<tr>
<td>Solyc12g019240.1.1</td>
<td>Intron</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019260.1.1</td>
<td>Promoter and Intron</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
<td>Expression</td>
<td>Tissue(s)</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
<td>--------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Solyc12g019270.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019280.1.1</td>
<td>Prom, Intr and Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019290.1.1</td>
<td>Promoter and Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019300.1.1</td>
<td>Promoter</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019310.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Shoot; Root</td>
</tr>
<tr>
<td>Solyc12g019320.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Root; Leaf; Flower</td>
</tr>
<tr>
<td>Solyc12g019330.1.1</td>
<td>Promoter and Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019340.1.1</td>
<td>Promoter and Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019350.1.1</td>
<td>Promoter and Exon</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019360.1.1</td>
<td>Promoter and Exon</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019370.1.1</td>
<td>Promoter and Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019380.1.1</td>
<td>Promoter and Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019390.1.1</td>
<td>Promoter and Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019400.1.1</td>
<td>Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019410.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Root; LEAF</td>
</tr>
<tr>
<td>Solyc12g019420.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Flower; Fruit</td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
<td>Expression</td>
<td>Location</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------</td>
<td>------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Solyc12g019430.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Seed; Root tip; Fruit</td>
</tr>
<tr>
<td>Solyc12g019440.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019450.1.1</td>
<td>Prom, Intr and Exon</td>
<td>YES</td>
<td>-</td>
</tr>
<tr>
<td>Solyc12g019460.1.1</td>
<td>Prom, Intr and Exon</td>
<td>NO</td>
<td>Shoot and root</td>
</tr>
<tr>
<td>Solyc12g019470.1.1</td>
<td>Promoter and Exon</td>
<td>NO</td>
<td>-</td>
</tr>
</tbody>
</table>