

Understanding the genetic regulation of acidity in *Citrus* fruit

Elliott Jordan Eden Atkins

A thesis presented for the degree of Doctor of Philosophy

University of East Anglia, UK

John Innes Centre, UK

June 2022

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Contents

Contents.....	1
Abstract.....	6
List of Tables	7
List of Figures	8
List of Supplementary Tables.....	12
List of Supplementary Figures.....	13
Gene Glossary	15
Acknowledgements.....	17
Chapter 1: General Introduction.....	19
1.1 The genus <i>Citrus</i>	20
1.1.1 Ancestry and diversification of <i>Citrus</i>	20
1.1.2 The economic importance of <i>Citrus</i>	21
1.2 Quality traits of <i>Citrus</i>	24
1.2.1 Determinants of fruit flavour	24
1.2.2 Organic acids in <i>Citrus</i>	24
1.2.3 Other PhPH5 structural homologs	26
1.3 Regulation of acidification and the flavonoid pathway	28
1.3.1 MYB-bHLH-WDR complexes.....	28
1.3.2 The ‘acidless’ phenotype in <i>Citrus</i>	29
1.4 The hypotheses	31
Chapter 2: General Materials and Methods	33
2.1 Materials	34
2.1.1 <i>Citrus</i> fruit	34
2.1.2 Transgenic <i>N. tabacum</i> plants.....	34
2.1.3 <i>N. benthamiana</i> plants.....	35
2.1.4 Bacterial strains.....	35

2.1.5 Antibiotics	35
2.1.6 Plasmids	36
2.1.7 Media recipes.....	36
2.1.8 Reference genomes	36
2.2 Methods.....	38
2.2.1 Primer design	38
2.2.2 Polymerase chain reaction.....	38
2.2.3 PCR and agarose gel purification	38
2.2.4 Plant tissue homogenisation.....	40
2.2.5 Plasmid DNA extraction	40
2.2.6 Quantification of nucleic acids.....	40
2.2.7 DNA sequencing.....	40
2.2.8 cDNA synthesis.....	40
2.2.9 Real-time quantitative PCR analysis	41
2.2.10 Preparation of competent <i>E. coli</i> cells	42
2.2.11 <i>E. coli</i> transformation.....	42
2.2.12 Preparation of electrocompetent <i>Agrobacterium</i> cells.....	43
2.2.13 <i>Agrobacterium</i> transformation.....	43
2.2.14 Restriction digestion and ligation	43
2.2.15 Overexpression plasmid construction	44
2.2.16 TOPO cloning.....	45
Chapter 3: Metabolic and transcriptomic analyses of acidless <i>C. sinensis</i> varieties	46
3.1 Introduction	47
3.1.1 The roles of AtTT2 and AtMYB5 in proanthocyanidin regulation.....	47
3.2 Materials and Methods.....	50
3.2.1 pH and carboxylic acid analyses.....	50
3.2.2 Cellular localisation of proanthocyanidins in <i>C. sinensis</i> seeds	50
3.2.3 Extraction and quantification of proanthocyanidins	51

3.2.4	Identification of <i>C. sinensis</i> homologs	53
3.2.5	Multiple sequence alignment	53
3.2.6	Phylogenetic analyses	53
3.2.7	Gene structural analyses.....	54
3.2.8	<i>Citrus</i> genomic DNA extraction.....	55
3.2.9	<i>Citrus</i> RNA extraction.....	55
3.2.10	RNA sequencing	56
3.3	Results.....	57
3.3.1	Fruit sugar, organic acid, and pH analyses.....	57
3.3.2	Proanthocyanidin quantification and localisation	59
3.3.3	Genome-wide identification of <i>C. sinensis</i> R2R3MYB transcription factors	66
3.3.4	Genome-wide identification of <i>C. sinensis</i> MATE transporters	71
3.3.5	Genotypic analyses of acidic and acidless <i>C. sinensis</i> varieties	75
3.3.6	Transcriptomic analyses of acidic and acidless <i>C. sinensis</i> varieties.....	77
3.4	Discussion.....	89
3.4.1	Potential roles within the CsR2R3MYB and CsMATE family	89
3.4.2	Genotypic characterisation of <i>nicole</i> ^{soro}	89
3.4.3	The <i>nicole</i> mutant retains PAs but loses acidity	90
3.4.4	Candidate targets of Nicole	92
Chapter 4: Characterisation of the regulatory function of the Noemi-Nicole MBW complex		94
4.1	Introduction	95
4.1.1	Introduction	95
4.1.2	The function of other MYB5-homologs	95
4.2	Materials and Methods.....	97
4.2.1	<i>Nicotiana</i> plant DNA and RNA extraction	97
4.2.2	Qualitative proanthocyanidin analyses in <i>Nicotiana</i>	97
4.2.3	Extraction and quantification of proanthocyanidins	98

4.2.4	Dual-luciferase reporter plasmid construction.....	98
4.2.5	Agroinfiltration of <i>N. benthamiana</i> leaves	99
4.2.6	Dual-luciferase reporter assays.....	100
4.3	Results.....	101
4.3.1	Genotyping and gene expression in transformed <i>N. tabacum</i> lines	101
4.3.2	Quantification of proanthocyanidins in <i>N. tabacum</i>	105
4.3.3	Dual-luciferase reporter gene assays.....	113
4.4	Discussion.....	119
4.4.1	Proanthocyanidin biosynthesis and associated genes are induced by both Nicole and Iris MBW complexes when ectopically expressed in <i>N. tabacum</i>	119
4.4.2	Nicole and Iris-driven MBW complexes directly transactivate different gene promoters in <i>C. sinensis</i>	121
Chapter 5: Generation and efficacy analysis of Nicole-targeting multi-sgRNA CRISPR-Cas9 constructs.....		123
5.1	Introduction	124
5.1.1	A gene-editing approach in <i>Citrus</i>	124
5.1.2	CRISPR-Cas9 genome-editing in plants	124
5.1.3	<i>Agrobacterium rhizogenes</i> -mediated transformation of <i>C. limon</i>	126
5.2	Materials and Methods.....	128
5.2.1	Identification of potential gRNA targets within <i>Nicole</i>	128
5.2.2	CRISPR/Cas9 plasmid construction	128
5.2.3	Hairy root transformation of <i>C. limon</i>	133
5.2.4	Hairy root DNA extraction.....	134
5.2.5	Identification of CRISPR-Cas9-edited alleles.....	134
5.2.6	Analyses of CRISPR-Cas9-edited alleles	134
5.3	Results.....	135
5.3.1	Identification and selection of gRNAs.....	135
5.3.2	Construction of multi-sgRNA CRISPR-Cas9 plasmids	135

5.3.3	<i>A. rhizogenes</i> transformation efficiency in <i>C. limon</i>	136
5.3.4	Cas9 nuclease activity derived gene edits	143
5.3.5	Guide RNA efficacy and protein functionality analysis	150
5.4	Discussion.....	156
5.4.1	Multi-species compatibility of <i>Nicole</i> -targeting gRNAs	156
5.4.2	Maximising MYB loss of function potential in gRNA design	156
5.4.3	<i>A. rhizogenes</i> transformation efficiency in <i>C. limon</i>	158
5.4.4	CRISPR-Cas9 gRNA efficacy analyses via <i>A. rhizogenes</i> -mediated transformation in <i>Citrus</i>	160
5.4.5	Functional analyses of CRISPR-Cas9 mutagenised <i>Nicole</i> alleles.....	161
5.4.6	Concluding remarks	163
Chapter 6:	Discussion.....	164
6.1	Final conclusions	165
6.1.1	The first reported divergence from the pleiotropic phenotype of acidless mutants	165
6.1.2	<i>Nicole</i> and <i>Noemi</i> associate in an MBW complex and is necessary for citrate bioaccumulation	165
6.1.3	<i>Nicole</i> has dissociated from the typical PA-role of MYB5-like homologs.....	166
6.1.4	Specialisation of <i>Nicole</i>	168
6.1.5	Concluding remarks	170
References	172
Appendices	191

Abstract

Citrus is the most economically important genus in the family Rutaceae. Fruit acidity is also a major determinant of fruit and juice taste, but its regulation has not been elucidated. Improvements through breeding, while difficult, could be better informed by understanding the genetic regulation of fruit acidity in *Citrus* at the molecular level. In this study, I used RNA sequencing transcriptomic data, phylogenetic analyses, and metabolic characterisation of acidless *C. sinensis* varieties to propose a model for the regulation of hyperacidification in *Citrus* fruits. By analogy the mechanism in *Petunia*, the *AtMYB5* and *PhPH4* homolog *Nicole* forms a complex with bHLH transcription factor Noemi, to govern fruit acidity and citrate accumulation via transcriptional regulation of *PH*-like genes.

Phylogenetic analyses of the *C. sinensis* genome and genotypic analyses of acidless varieties Sorocaba, Verde R1 and Verde R2 led to the discovery of *nicole^{soro}*, a mutant allele of *Nicole* containing an LTR-retrotransposon insertion which has disrupted functionality. These varieties provide the first example of disassociation of the pleiotropic link between acidity and proanthocyanidin (PA) biosynthesis in an acidless *Citrus* mutant. *Nicole* is not essential for PA biosynthesis in *Citrus* due to the accumulation of PAs in mutant seeds, and uniquely has lost the ability to activate the expression of PA transporter *CsTT12*, typically activated by *AtMYB5* homologs. Conversely, *Nicole* can directly activate P_{3A} - and P_{3B} -ATPases, *CsPH5* and *CsPH1*, to facilitate hyperacidification and citrate accumulation by avoiding the counteraction of H^+ /PA-antiporter, *CsTT12*.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

List of Tables

Table 2.1.1 Bacterial strains used within the thesis and their respective antibiotic selection and purpose.	35
Table 2.1.2 Antibiotics used within the thesis and their respective solvent and working concentrations.	36
Table 2.2.1 Typical PCR composition per 50 µl reaction.....	39
Table 2.2.2 Typical PCR thermocycler parameters.	39
Table 2.2.3 RT PCR composition per 20 µl reaction.....	41
Table 2.2.4 RT qPCR thermocycler parameters.	41
Table 2.2.5 Restriction digestion composition per 50 µl reaction.....	44
Table 2.2.6 Ligation composition per 20 µl reaction.	44
Table 3.2.1 Proanthocyanidin HPLC analyses separation gradient of acetonitrile versus 0.1 % (w/v) formic acid in H ₂ O.....	52
Table 5.2.1 Golden Gate digestion-ligation composition per 15 µl reaction.....	132
Table 5.2.2 Golden Gate Level 1 and Level 2 digestion-ligation thermocycler parameters.	132
Table 5.3.1 Highly specific gRNA target sequences present in CsNicole.	137
Table 5.3.2 Transformation efficiency of <i>C. limon</i> using <i>A. rhizogenes</i> strains K599, ATCC15834 and MSU440.	141
Table 5.3.3 Mutation rates of each type attributed to either pEA13, pEA15 or overall.	150
Table 5.3.4 Guide RNA efficacy.....	155

List of Figures

Figure 1.1.1 Global annual production of sweet oranges in 2019.....	22
Figure 1.4.1 Proposed regulation of anthocyanin, proanthocyanidins and hyperacidification by transcriptional MBW regulator complexes in <i>Citrus</i>	31
Figure 3.1.1 Schematic of the flavonoid pathway.	48
Figure 3.1.2 Simplified schematic of proton homeostasis and PA transport into the vacuole of <i>Arabidopsis</i> endothelium cells.....	49
Figure 3.3.1 Acidless Sorocaba, VerdeR1 and VerdeR2 <i>C. sinensis</i> fruit immediately before sampling.....	57
Figure 3.3.2 Juice pH of five <i>C. sinensis</i> varieties: Navel, Valencia, Sorocaba, Verde R1 and Verde R2.....	58
Figure 3.3.3 Organic acid concentration in the juice of five <i>C. sinensis</i> varieties.	60
Figure 3.3.4 Soluble sugar concentration in the juice of five <i>C. sinensis</i> varieties.	61
Figure 3.3.5 <i>C. sinensis</i> seeds before and after ≥ 10 min staining with 0.3 % DMACA reagent.	62
Figure 3.3.6 Soluble proanthocyanidin concentration in the juice of five <i>C. sinensis</i> varieties.	63
Figure 3.3.7 Free catechin and epicatechin monomer concentration in the juice of five <i>C. sinensis</i> varieties.	64
Figure 3.3.8 Cellular localisation of proanthocyanidins in sectioned Valencia and Vaniglia <i>C. sinensis</i> seeds.....	65
Figure 3.3.9 Amino acid conservation of R2 and R3 MYB repeats across all 68 putative R2R3MYBs identified in the <i>C. sinensis</i> genome.....	66
Figure 3.3.10 Maximum-likelihood amino acid phylogeny of 68 putative CsR2R3MYBs, 125 Arabidopsis R2R3MYBs and 21 characterised R2R3MYBs from other plant species.....	68
Figure 3.3.11 ClustalO amino acid alignment of <i>C. sinensis</i> R2R3MYBs structurally homologous to AtMYB123 and AtMYB5.....	69
Figure 3.3.12 Maximum-likelihood amino acid phylogeny and gene structure of 68 putative CsR2R3MYB.....	70
Figure 3.3.13 Maximum-likelihood amino acid phylogeny of 35 putative CsMATEs, 56 Arabidopsis MATEs and 20 characterised MATEs from other plant species.	73
Figure 3.3.14 Maximum-likelihood amino acid phylogeny and gene structure of 35 putative CsMATEs.....	74

Figure 3.3.15 Genetic characterisation of the mutant nicole ^{soro} allele.....	76
Figure 3.3.16 Differentially expressed gene counts in Vaniglia, relative to Navel.	78
Figure 3.3.17 Differentially expressed gene counts in Sorocaba, relative to Navel.	79
Figure 3.3.18 Differentially expressed gene counts in Verde R1, relative to Navel.	80
Figure 3.3.19 Differentially expressed gene counts in Verde R2, relative to Navel.	81
Figure 3.3.20 Significantly downregulated gene counts in Vaniglia, Sorocaba, Verde R1 and Verde R2 <i>C. sinensis</i> in comparison to Navel.	82
Figure 3.3.21 RNAseq FPKM values in Navel, Vaniglia, Sorocaba, Verde R1 and Verde R2 <i>C. sinensis</i> of the 15 most significantly downregulated genes.	85
Figure 3.3.22 RNAseq FPKM values in Navel, Vaniglia, Sorocaba, Verde R1 and Verde R2 <i>C. sinensis</i> of 16 genes of interest.....	86
Figure 3.3.23 RNAseq FPKM values in Navel, Vaniglia, Sorocaba, Verde R1 and Verde R2 <i>C. sinensis</i> of 17 genes involved in citrate metabolism or transport.....	87
Figure 3.3.24 Gene expression in Navel, Vaniglia, Sorocaba, Verde R1 and Verde R2 <i>C. sinensis</i> of 10 genes of interest via RT-qPCR.....	88
Figure 4.3.1 PCR amplification of Noemi T-DNA from gDNA extracted from overexpression <i>N. tabacum</i> transformed lines.....	102
Figure 4.3.2 PCR amplification of Nicole T-DNA from gDNA extracted from overexpression <i>N. tabacum</i> transformed lines.....	103
Figure 4.3.3 PCR amplification of Iris T-DNA from gDNA extracted from overexpression <i>N. tabacum</i> transformed lines.....	103
Figure 4.3.4 RT-qPCR analyses of T-DNA and various candidate genes of interest in overexpression <i>N. tabacum</i> lines.....	104
Figure 4.3.5 Scatter plots of Nicole and Iris expression against NtLDOX expression quantified in RT-qPCR analyses of overexpression <i>N. tabacum</i> lines.	106
Figure 4.3.6 Scatter plots of Nicole and Iris expression against NtANR expression quantified in RT-qPCR analyses of overexpression <i>N. tabacum</i> lines..	107
Figure 4.3.7 Scatter plots of Nicole and Iris expression against NtPH1 expression quantified in RT-qPCR analyses of overexpression <i>N. tabacum</i> lines.	108
Figure 4.3.8 Qualitative phenotypic changes in overexpression <i>N. tabacum</i> lines.....	110
Figure 4.3.9 Proanthocyanidin quantification of via colorimetric DMACA assays.	111
Figure 4.3.10 HPLC quantification of free catechin and epicatechin monomer concentration of soluble PA extracts.	112

Figure 4.3.11 Dual-luciferase pNoemi reporter activation by co-expression of various combinations of <i>C. sinensis</i> MYB and bHLH-encoding genes.	114
Figure 4.3.12 Dual-luciferase pNicole reporter activation by co-expression of various combinations of <i>C. sinensis</i> MYB and bHLH-encoding genes.	114
Figure 4.3.13 Dual-luciferase pCsPH1 reporter activation by co-expression of various combinations of <i>C. sinensis</i> MYB and bHLH-encoding genes.	115
Figure 4.3.14 Dual-luciferase pCsPH5 reporter activation by co-expression of various combinations of <i>C. sinensis</i> MYB and bHLH-encoding genes.	115
Figure 4.3.15 Dual-luciferase pCsLDOX reporter activation by co-expression of various combinations of <i>C. sinensis</i> MYB and bHLH-encoding genes.	116
Figure 4.3.16 Dual-luciferase pCsANR reporter activation by co-expression of various combinations of <i>C. sinensis</i> MYB and bHLH-encoding genes.	117
Figure 4.3.17 Dual-luciferase pCsTT12 reporter activation by co-expression of various combinations of <i>C. sinensis</i> MYB and bHLH-encoding genes.	118
Figure 5.2.1 Simplified schematic view of construction of Level 1 (L1) plasmids using Golden Gate cloning technology.	130
Figure 5.2.2 Simplified schematic view of construction of Level 2 (L2) plasmids using Golden Gate cloning technology.	131
Figure 5.3.1 Annotated exon 1 nucleotide sequences of <i>C. sinensis</i> , <i>C. limon</i> and <i>F. hindsii</i> Nicole alleles.	138
Figure 5.3.2 Multi-sgRNA CRISPR-Cas9 plasmid map of pEA13.	139
Figure 5.3.3 Negative control plasmid map of pEA16 containing no CRISPR-Cas9 genetic elements.	139
Figure 5.3.4 <i>Dra</i> III restriction digestion of L2 CRISPR-Cas9 plasmids.	140
Figure 5.3.5 Typical hairy root development \geq 8 weeks after <i>A. rhizogenes</i> transformation treatment of <i>C. limon</i> epicotyls, with and without 50 mg L ⁻¹ kanamycin selection.	142
Figure 5.3.6 PCR amplification of <i>CINicole</i> exon 1 from hairy root DNA extracted from <i>C. limon</i> transformed by ATCC15834 <i>A. rhizogenes</i> cells harbouring pEA13 or no recombinant plasmid.	144
Figure 5.3.7 PCR amplification of <i>CINicole</i> exon 1 from hairy root DNA extracted from <i>C. limon</i> transformed by ATCC15834 <i>A. rhizogenes</i> cells harbouring pEA15.	145
Figure 5.3.8 TOPO clone sequencing of <i>Nicole</i> exon 1 alleles, derived from pEA13-transformed <i>C. limon</i> , aligned to the <i>CINicole.1</i> allele.	146

Figure 5.3.9 TOPO clone sequencing of <i>Nicole</i> exon 1 alleles, derived from pEA13-transformed <i>C. limon</i> , aligned to <i>CINicole.2</i> allele.....	147
Figure 5.3.10 TOPO clone sequencing of <i>Nicole</i> exon 1 alleles, derived from pEA15-transformed <i>C. limon</i> , aligned to <i>CINicole.1</i> allele.....	148
Figure 5.3.11 TOPO clone sequencing of <i>Nicole</i> exon 1 alleles, derived from pEA15-transformed <i>C. limon</i> , aligned to <i>CINicole.2</i> allele.....	149
Figure 5.3.12 Amino acid sequence translated from TOPO clone sequencing of <i>Nicole</i> exon 1 alleles, derived from pEA13-transformed <i>C. limon</i> , aligned to <i>CINicole.2</i> allele.	152
Figure 5.3.13 Amino acid sequence translated from TOPO clone sequencing of <i>Nicole</i> exon 1 alleles, derived from pEA13-transformed <i>C. limon</i> , aligned to <i>CINicole.1</i> allele.	152
Figure 5.3.14 Amino acid sequence translated from TOPO clone sequencing of <i>Nicole</i> exon 1 alleles, derived from pEA15-transformed <i>C. limon</i> , aligned to <i>CINicole.1</i> allele.	153
Figure 5.3.15 Amino acid sequence translated from TOPO clone sequencing of <i>Nicole</i> exon 1 alleles, derived from pEA15-transformed <i>C. limon</i> , aligned to <i>CINicole.2</i>	154
Figure 6.1.1 Simplified schematic of vacuolar hyperacidification in <i>Citrus</i> fruit due to the specialisation of R2R3MYB transcription factor, <i>Nicole</i>	169

List of Supplementary Tables

Supplementary Table 1 Plasmids used in this thesis, their description, selection, and construction method.	191
Supplementary Table 2 Media recipes used in Chapter 5.	192
Supplementary Table 3 List of gene accession IDs from the <i>C. sinensis</i> v1.1 Phytozome genome, nomenclature based on phylogenetic analyses, BLAST analyses and naming from previous publications.....	193
Supplementary Table 4 List of primers used in this thesis, sequence, target and purpose..	196
Supplementary Table 5 Sequence of nicole ^{soro}	205

List of Supplementary Figures

Supplementary Figure 1 Absorbance (OD _{640 nm}) of various concentrations of 15 µl (+)-catechin standard reacted with 85 µl 0.3% DMACA over time.....	198
Supplementary Figure 2 (+)-Catechin standard curves of absorbance (OD _{640 nm}) after varying lengths of time reacting with 0.3% DMACA reagent.	199
Supplementary Figure 3 Neighbour-joining phylogenetic tree constructed from a ClustalO MSA of all 125 R2R3MYBs from <i>Arabidopsis</i> and <i>Citrus</i> proteins containing at least one PFAM MYB domain (PF00249) which share greatest homology to AtMYB5 and AtMYB123.	200
Supplementary Figure 4 PCR amplification of <i>Noemi</i> from gDNA extracted from <i>C. sinensis</i> juice.....	201
Supplementary Figure 5 PCR amplification of <i>Iris</i> from gDNA extracted from <i>C. sinensis</i> juice.	201
Supplementary Figure 6 PCR amplification of <i>CsTTG1</i> from gDNA extracted from <i>C. sinensis</i> juice.....	202
Supplementary Figure 7 PCR amplification of <i>CsPH3</i> from gDNA extracted from <i>C. sinensis</i> juice.....	202
Supplementary Figure 8 PCR amplification of <i>CsPH1</i> from gDNA extracted from <i>C. sinensis</i> juice.....	203
Supplementary Figure 9 PCR amplification of <i>CsTT12</i> from gDNA extracted from <i>C. sinensis</i> juice.....	203
Supplementary Figure 10 PCR amplification of <i>CsPH5</i> from gDNA extracted from <i>C. sinensis</i> juice.....	204
Supplementary Figure 11 Pearson correlation heatmap of all <i>C. sinensis</i> RNAseq biological replicates.....	206
Supplementary Figure 12 Biological process gene ontology enrichment in significantly downregulated genes in Vaniglia relative to Navel.	207
Supplementary Figure 13 Biological process gene ontology enrichment in significantly downregulated genes in Sorocaba relative to Navel.	208
Supplementary Figure 14 Biological process gene ontology enrichment in significantly downregulated genes in Verde R1 relative to Navel.	209
Supplementary Figure 15 Biological process gene ontology enrichment in significantly downregulated genes in Verde R2 relative to Navel.	210
Supplementary Figure 16 Primer efficiency analyses for RT-qPCR.....	211

Supplementary Figure 17 PCR amplification of <i>NPTII</i> T-DNA from gDNA extracted from overexpression <i>N. tabacum</i> transformed lines.....	212
Supplementary Figure 18 Primer efficiency analyses for RT-qPCR.....	213
Supplementary Figure 19 Bleached leaf discs from overexpression <i>N. tabacum</i> lines, prior to DMACA staining.	214
Supplementary Figure 20 Absorbance (OD _{640 nm}) of various concentrations of 4 µl (+)-catechin standard reacted with 96 µl 0.3% DMACA over time.	215
Supplementary Figure 21 (+)-Catechin standard curves of absorbance (OD _{640 nm}) after varying lengths of time reacting with 0.3% DMACA reagent.	216
Supplementary Figure 22 Proanthocyanidin quantification of via colorimetric DMACA assays.	217

Gene Glossary

Gene	Protein class	Species	References (if applicable)
CsANR	Anthocyanidin reductase	<i>Citrus sinensis</i>	N/A
Noemi	Basic helix-loop-helix transcription factor	<i>Citrus sinensis</i>	(Butelli et al., 2019)
CsDFR	Dihydroflavonol-4-reductase	<i>Citrus sinensis</i>	N/A
CsLDOX	Leucoanthocyanidin dioxygenase	<i>Citrus sinensis</i>	N/A
CsLAR	Leucoanthocyanidin reductase	<i>Citrus sinensis</i>	N/A
CsTT12	Multi-drug and toxic compound extrusion transporter	<i>Citrus sinensis</i>	N/A
Iris	Myeloblastosis transcription factor	<i>Citrus sinensis</i>	N/A
Marys	Myeloblastosis transcription factor	<i>Citrus sinensis</i>	N/A
Nicole	Myeloblastosis transcription factor	<i>Citrus sinensis</i>	N/A
Ruby	Myeloblastosis transcription factor	<i>Citrus sinensis</i>	(Butelli et al., 2012)
CsPH5	P _{3A} -ATPase	<i>Citrus sinensis</i>	(Shi et al., 2015)
CsPH1	P _{3B} -ATPase	<i>Citrus sinensis</i>	N/A
CsTTG1	WD-repeat protein	<i>Citrus sinensis</i>	N/A
CsPH3	WRKY transcription factor	<i>Citrus sinensis</i>	N/A
PhAN1	Basic helix-loop-helix transcription factor	<i>Petunia hybrida</i>	(Spelt, 2002)
PhAN2	Myeloblastosis transcription factor	<i>Petunia hybrida</i>	(Quattrocchio et al., 1999)
PhPH4	Myeloblastosis transcription factor	<i>Petunia hybrida</i>	(Quattrocchio et al., 2006)
PhPH5	P _{3A} -ATPase	<i>Petunia hybrida</i>	(Verweij et al., 2008)
PhPH1	P _{3B} -ATPase	<i>Petunia hybrida</i>	(Faraco et al., 2014)
PhAN11	WD-repeat protein	<i>Petunia hybrida</i>	(de Vetten et al., 1997)
PhPH3	WRKY transcription factor	<i>Petunia hybrida</i>	(Verweij et al., 2016)
AtBAN	Anthocyanidin reductase	<i>Arabidopsis thaliana</i>	(Baudry et al., 2004)
AtTT8 / AtbHLH42	Basic helix-loop-helix transcription factor	<i>Arabidopsis thaliana</i>	(Nesi et al., 2000)
AtDFR	Dihydroflavonol-4-reductase	<i>Arabidopsis thaliana</i>	(Shirley et al., 1992)
AtLDOX	Leucoanthocyanidin dioxygenase	<i>Arabidopsis thaliana</i>	(Abrahams et al., 2003)

AtTT12	Multi-drug and toxic compound extrusion transporter	<i>Arabidopsis thaliana</i>	(Marinova et al., 2007)
AtMYB5	Myeloblastosis transcription factor	<i>Arabidopsis thaliana</i>	(Li et al., 2009)
AtPAP1	Myeloblastosis transcription factor	<i>Arabidopsis thaliana</i>	(Borevitz et al., 2000)
AtTT2 / AtMYB123	Myeloblastosis transcription factor	<i>Arabidopsis thaliana</i>	(Nesi et al., 2001)
AtAHA10	P _{3A} -ATPase	<i>Arabidopsis thaliana</i>	(Baxter et al., 2005)
AtTTG1	WD-repeat protein	<i>Arabidopsis thaliana</i>	(Baudry et al., 2004)
AtTTG2	WRKY transcription factor	<i>Arabidopsis thaliana</i>	(Gonzalez et al., 2016)
VvANR	Anthocyanidin reductase	<i>Vitis vinifera</i>	(Gagné et al., 2009)
VvMYC1	Basic helix-loop-helix transcription factor	<i>Vitis vinifera</i>	(Hichri et al., 2010)
VvLAR	Leucoanthocyanidin reductase	<i>Vitis vinifera</i>	(Gagné et al., 2009)
VvAM1	Multi-drug and toxic compound extrusion transporter	<i>Vitis vinifera</i>	(Gomez et al., 2009)
VvAM3	Multi-drug and toxic compound extrusion transporter	<i>Vitis vinifera</i>	(Gomez et al., 2009)
VvMYB5a	Myeloblastosis transcription factor	<i>Vitis vinifera</i>	(Amato et al., 2019)
VvMYB5b	Myeloblastosis transcription factor	<i>Vitis vinifera</i>	(Amato et al., 2019)
VvMYBA1	Myeloblastosis transcription factor	<i>Vitis vinifera</i>	(Kobayashi et al., 2004)
VvMYBA2	Myeloblastosis transcription factor	<i>Vitis vinifera</i>	(Kobayashi et al., 2004)
VvMYBPA1	Myeloblastosis transcription factor	<i>Vitis vinifera</i>	(Bogs et al., 2007)
VvMYBPA2	Myeloblastosis transcription factor	<i>Vitis vinifera</i>	(Terrier et al., 2009)
VvMYBPAR	Myeloblastosis transcription factor	<i>Vitis vinifera</i>	(Koyama et al., 2014)
VvPH5	P _{3A} -ATPase	<i>Vitis vinifera</i>	(Amato et al., 2019)
VvPH1	P _{3B} -ATPase	<i>Vitis vinifera</i>	(Amato et al., 2019)
VvWD1	WD-repeat protein	<i>Vitis vinifera</i>	(Matus et al., 2010)
VvWRKY26	WRKY transcription factor	<i>Vitis vinifera</i>	(Amato et al., 2019)

Acknowledgements

With my thesis now complete, I have taken a moment to realise how much of an impact both the people who have helped me, and the pandemic have had on my personal and PhD life. The first lockdown fell during what students commonly call the most important year for data collection, and I was hit with instant dread and suddenly felt overwhelmed by the task ahead. Many people have supported me throughout the entirety of my PhD, and more so during those lockdown periods we like to forget. I would like to acknowledge and thank them now.

Firstly, my supervisory team Cathie and Eugenio. Before I started, I was “technically” an Ecologist, who happened to find out he loved molecular plant biology in the final year of his undergraduate degree. I managed to get a year’s experience working in the plant transformation team next door at TSL. A special mention goes to Jodie for encouraging me to start this PhD. Of course, I see now that while tissue culture skills are invaluable, they did not make up for my lack of experience in molecular biology (I had one PCR under my belt). Despite this, Cathie and Eugenio were happy to take me on, and I am incredibly grateful for everything I have been able to learn. Since now I can officially remove my Ecologist title, and be known as a molecular plant biologist instead! I would like to thank Eugenio for all his practical tips and tricks, and seemingly unending knowledge of troubleshooting and *Citrus*. I would like to thank Cathie also for always encouraging and supporting me when I needed it most. I think she could see the pandemic was not a great experience for me and she put her full support behind me, like when we applied for an extension period.

I am also incredibly grateful that Cathie encouraged taking on students throughout my PhD, without which I would never have met my dear friend Chiara. Chiara I would like to thank you for bringing your tireless positive outlook to the lab every single day you were here. Despite the terrible time you had with Norwich accommodation, your top priority was to make as many people smile as possible. You were the first real friend I made at JIC, and while I was sad to see you leave, I am thrilled you have since started your own PhD journey. I would like to thank members of my lab, past and present: Jie, Fabio, Ronan. They all gave me countless hours of practical advice and it has been a pleasure to work alongside you. Special thanks go to Jie for being a temporary therapist when the write-up period began!...

I would like to acknowledge the BioImaging and Metabolic services teams for all their hard work during my last year. Elaine was so approachable and willing to dedicate as much time as necessary to understand exactly what I needed from the microscopy work. Thank you to Lionel and Paul for letting me queue jump where possible and for handing back great data.

I want to thank, and apologise to, my best friends Harry and Chris. I have neglected you a lot and we have seen less of each other than ever before since I moved to Norwich. Yet when we meet it feels like we're still at sixth form. Thank you for being so patient. The same goes for my family. The past four years have really shown me how supportive you all are. You have all encouraged me so much and never fail to mention oranges whenever I see you. You have also been just as supportive to my partner, Shannon, and treat her just like your own. Particularly Mum and Dad, who I know would do anything for both of us.

Finally, I want to thank my partner, Shannon. We met at the start of our PhDs, and I feel so lucky to have gone through the whole journey with you. You have had such a difficult time over the past year but have still made me your priority, especially during my write-up. I am eternally grateful to have met you and for your love and support. I am so proud of you and everything you have achieved, and I am so excited to have two Doctors of Philosophy in the family!

Chapter 1:
General Introduction

1.1 The genus *Citrus*

1.1.1 Ancestry and diversification of *Citrus*

The most economically important genus in the family Rutaceae is *Citrus*, a genus comprised of flowering trees and shrubs. There is a considerable proliferation of interspecific hybrids within the genus, amongst which are some of the most widely recognised and commercially important fruits, such as oranges, lemons, and limes. *Citrus* fruits are often characterised by their high acidity and vitamin C content, which is a recognised health-promoting constituent of the juice. Final flavour is also predominantly determined by the ratio between acidity and soluble sugar content, which in turn dictates the necessity for additional sugars in products such as juices. This thesis presents hypotheses regarding the mechanisms that regulate fruit acidity in *Citrus* and how it has been investigated, including the identification of the first acidless mutants lacking the pleiotropic link between proanthocyanidins (PAs) and low acidity, transcriptomic analyses, functional characterisation of a myeloblastosis (MYB) transcription factor that regulates acidity in *Citrus* and the generation of mutant alleles via CRISPR/Cas9 editing.

There are two classification systems considered for *Citrus* taxonomy, presented by Swingle and Reece (1967) and Tanaka (1977). There is ambiguity between the two systems as to the acceptance of hybrids and variant species as true species. In short, Swingle and Reece (1967) proposed 6 genera within Rutaceae, including *Fortunella*, *Microcitrus* and *Citrus*. *Citrus* was divided into two subgenera, *Citrus* with 10 species, and *Papeda* with 6, disregarding many hybrid cultivars as true *Citrus* species. Tanaka (1977), however, considered a more comprehensive view of *Citrus*, comprising 173 species which include the variants or hybrids as species themselves. When morphological and biochemical analyses were considered, it was recognised that there are in fact only a small number of true, or progenitor, species in *Citrus* (Barrett and Rhodes, 1976, Scora, 1975). Initially, there were 3 ancestral species, *C. maxima* (pummelos), *C. reticulata* (mandarins) and *C. medica* (citrons), accepted as progenitors, from which many interspecific hybrids are derived.

Recently, however, dissecting genome composition of many *Citrus* species identified 10 true species (Wu et al., 2018). Wu et al. (2018) suggest that the *Citrus* clade also includes genera

Microcitrus, *Eremocitrus* and *Fortunella*, from which some members are considered within the 10 true *Citrus* species enumerated. Furthermore, analysis of allelic proportions of five of these progenitors in 50 *Citrus* accessions elucidated further the ancestry of major commercial varieties. These data support the view that the most popular varieties are interspecific hybrids derived from citrons, mandarins and pummelos, but also the more recent theory that species from *Fortunella* were also involved in hybridisations, producing limes and calamondin (Wu et al., 2018, Carbonell-Caballero et al., 2015, Swingle and Reece, 1967). Additionally, mandarins were divided into three types, the first of which is a pure species, while sequential introgressions with pummelo gave rise to two more types, termed early and late-admixture mandarins.

Propagation of *Citrus* plants is achieved either by grafting or from germination of seed. These interspecific hybrids reproduce apomictically, a trait which characterises most of the important commercial *Citrus* fruits that are cultivated. This process is defined as the polyembryony of maternal nucellar tissue, leading to development of seeds asexually. Consequently, there is preservation of genetic makeup between generations, regardless of propagation method, and it is difficult to develop new accessions with desirable traits using conventional breeding techniques. The proliferation of the *Citrus* genus likely occurred during an expansion in southeast Asia during the late Miocene epoch (Wu et al., 2018), because of random somatic mutations. Domestication of *Citrus* would have also facilitated diversification with the selection and propagation of any emerging variants with preferable traits.

1.1.2 The economic importance of *Citrus*

Citrus comprises some of the most economically valuable and important fruit tree crops globally. The fruits are grown within the *Citrus*-belt, 35° north and south of the equator (FAO, 2021, Berk, 2016), and utilised as both fresh and processed products, such as juices (Figure 1.1.1). Cultivation of oranges is dominated by China, Brazil and India as 47% of global production was attributed to these three countries alone in 2019 (FAO, 2021). Total *Citrus* production during 2019 in all 116 *Citrus*-cultivating countries was estimated at over 228 million tonnes, with oranges, clementines, mandarins, tangerines and satsumas being the leading varieties.

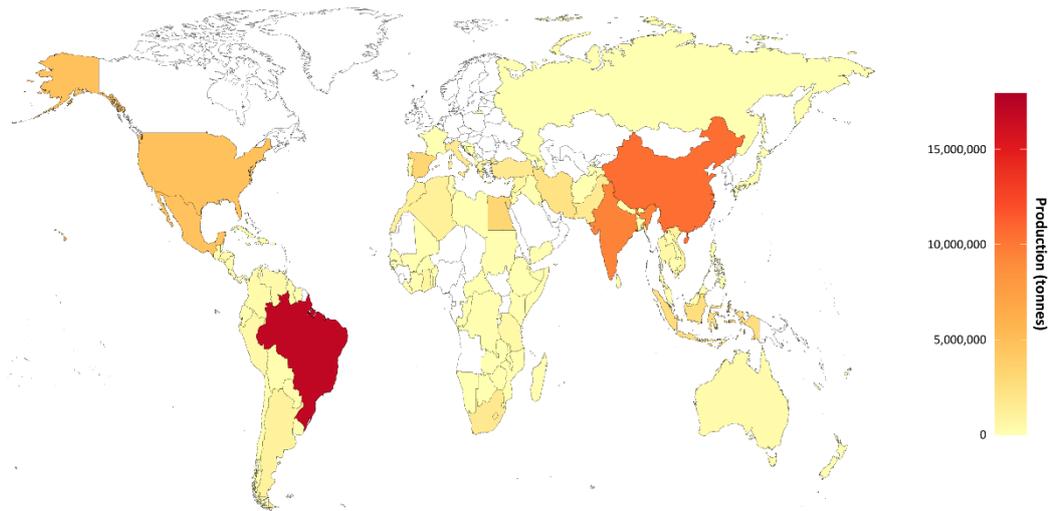


Figure 1.1.1 Global annual production of sweet oranges in 2019. Plotted using data from FAO (2021). Extracted from: <https://www.fao.org/faostat/en/#home>. Date of access: 14/10/2021.

Citrus fruits are widely appreciated as nutrient-rich, healthy foods which provide natural carbohydrates, negligible fat, and health-promoting benefits (Economos and Clay, 1999). They are frequently recognised as an excellent dietary source of ascorbic acid (vitamin C), which is responsible for 65-100 % of the antioxidative capability of *Citrus* juices (Gardner et al., 2000). Grapefruit, sweet orange and tangerine contain 79 mg, 70 mg and 26 mg of vitamin C per whole fresh fruit, respectively (Economos and Clay, 1999), and more than an adult male’s recommended daily intake of vitamin C can be sourced from a single glass of orange juice (Berk, 2016). Consequently, *Citrus* fruit consumption can even provide disease prevention or treatment, for example, for scurvy. This was first discovered during clinical trials in 1747 (Lind, 1757).

In addition, the fruits also offer a source of fibre and a wide range of additional health-benefiting compounds, including other vitamins, minerals and flavonoids (Duarte et al., 2016, Economos and Clay, 1999). These secondary metabolites found in *Citrus* have the potential to exhibit anti-inflammatory, anti-oxidative and even anti-cancer bioactivity, in addition to reducing the risk of cardiovascular disease (Favela-Hernández et al., 2016, Lv et

al., 2015). In summary, *Citrus* fruits provide an abundant supply of nutritional compounds and have disease-preventative potential, whether consumed fresh or as processed juices.

1.2 Quality traits of Citrus

1.2.1 Determinants of fruit flavour

Citrus fruits contain many compounds, including vitamins, minerals, and sugars, that contribute to the final flavour. However, a major determinant of fruit taste in *Citrus* is acidity. While a positive trait for some species, providing an iconic and desirable sour taste in *C. limon* (lemon), it is a negative trait in varieties for which a sweeter taste is sought. Ultimately, the sweetness or bitterness of *Citrus* fruits is attributed to relative levels of total soluble sugars (TSS) and acidity. This ratio is particularly important in juice production, as it determines the necessity for additional sugars to sweeten the taste. As a result, highly acidic products can have negative impacts on dental health, particularly in children, and may contribute to overnutrition, owing to added sugars. Simply selling fruits can be hindered by the TSS:acid ratio as a minimum level must be exceeded (Fang et al., 1997). As an increase of fruit TSS to sweeten products could have further negative implications on human health, manipulating the other components of the ratio has become a priority (Etienne et al., 2013). Improvements through breeding, while difficult and impossible in some species, could be better informed by understanding the genetic regulation of fruit acidity in *Citrus* at the molecular level. This was the focus and objective of my PhD project.

1.2.2 Organic acids in Citrus

The main constituent of organic acids in *Citrus* fruit is citric acid (Albertini et al., 2006, Chen et al., 2013, Guo et al., 2016, Li et al., 2017, Lin et al., 2015, Zhou et al., 2018, Yamaki, 1989), which varies in vacuolar content during the three phases of ripening. Chronologically, fruits undergo a cell division stage (I), followed by expansion (II), and finally, ripening (III). Generally, citric acid accumulates through stages I and II, peaking in the latter, and then declines during phase III (Albertini et al., 2006, Cercos et al., 2006, Moufida and Marzouk, 2003, Moon and Mizutani, 2002). In lemons and limes however, citric acid remains constant following its peak in stage II (Albertini et al., 2006). In contrast, TSS increase, and consist predominantly of sucrose, glucose, and fructose.

In plants the vacuole serves as a reservoir for many compounds, such as sugars, organic acids and anthocyanins transported into the vacuolar-lumen by an electrochemical gradient across the tonoplast. This is governed in most cases by vacuolar pyrophosphatase (V-PPase) and vacuolar ATPase (V-ATPase) proton pumps, translocating cytosolic H⁺ ions into the vacuole and generating an electrochemical H⁺ gradient (Drozdowicz and Rea, 2001, Gaxiola et al., 2007, Schumacher and Krebs, 2010, Shen et al., 2013). Generally, the difference in vacuolar (pH_{vac} ~6) and cytosolic (pH ~7) acidity is relatively small as a result. However, some specialised cells in the plants become hyperacidified, such as *C. limon* fruit juice cells (pH_{vac} ~2) (Müller and Taiz, 2002, Echeverria et al., 1992, Müller et al., 1996) and epidermal cells in *Petunia* petals (pH_{vac} ≤ 5) (Quattrocchio et al., 2006).

The current understanding of vacuolar citrate accumulation in fruit cells is referred to as the 'acid trap' mechanism (Martinoia et al., 2007, Etienne et al., 2013). Citrate accumulation is determined by the combination of both H⁺ transport across the tonoplast and cytosolic citrate content. As citrate is a weak acid, the acid can easily dissociate and exist as either a monoanion, dianion or trianion. The dominant form in the vacuolar lumen is the protonated species, in contrast to the trianion form in the cytosol. As trianion citrate³⁻ species are transported into the tonoplast via inward-rectifying channels (Rentsch and Martinoia, 1991), they drive a large influx of protons in parallel. Vacuolar pH decreases while providing the driving force for increased uptake of citrate³⁻. The citrate³⁻ anions are instantly protonated to citrateH²⁻, thereby maintaining the electrochemical gradient ($\Delta\psi$) and sustaining transport from the cytosol. An effective acid trapping mechanism is dependent on the pH of the vacuole and electrochemical gradient across the endomembrane. Citrate homeostasis is also dependent on a H⁺/citrate symporter (CsCit1), facilitating citrateH²⁻ efflux, and enzymes directly involved in citrate metabolism, such as citrate synthase (CS), and degradation-related genes such aconitase and glutamine synthase (Chen et al., 2013, Shimada et al., 2006).

Recently, another type of proton pump has been identified in *Citrus*, and data suggests it is highly associated with citric acid levels. *PH5* encodes a P_{3A}-ATPase proton pump in *Petunia hybrida* is essential for the hyperacidification of epidermal cells in petals (Verweij et al., 2008). The cells also accumulate vacuolar anthocyanin pigments whose colour is pH-dependent. Consequently, *ph5* mutants exhibit variations in colour. Various other members

of the P_{3A}-ATPase pump family have been identified in plants, all localising to the plasma membrane (DeWitt et al., 1996, DeWitt and Sussman, 1995, Lefebvre et al., 2005, Lefebvre et al., 2004). However, PhPH5 is tonoplast-bound and drives the hyperacidification of the vacuole (Verweij et al., 2008). The structural homolog of PhPH5 in *Citrus* has since been studied and variation in the activity of this tonoplast-localised P-ATPase has been suggested to be responsible for variations in citrate between species. PhPH5-like genes were characterised by Shi et al. (2015) and transcriptomic analyses suggested down regulation of PhPH5-like genes was associated with low citrate in acidless varieties. Similar results have been reported since in *Citrus* (Aprile et al., 2011, Strazzer et al., 2019, Shi et al., 2021, Shi et al., 2019). In particular, the overexpression of *CsPH5* in tomato, which lacks a putative *PH5* homolog, increased citric acid levels (Shi et al., 2019).

Comparative analyses of organic acids during fruit maturation of acidic and acidless *Citrus* varieties have provided insight on the mechanistic determination of acidity and citrate homeostasis. Several acidless *Citrus* varieties of sweet orange, lime and lemon exhibit reduced accumulation of citric acid, measured as titratable acidity (Huang et al., 2016, Albertini et al., 2006). This was likely a consequence of limited citrate vacuolar uptake, rather than minimal synthesis, due to abnormal proton pump activity. This included activity of *CsPH5* (Aprile et al., 2011), and therefore lower protonation activity and higher luminal pH. Furthermore, differences in citrate levels in acidic and acidless *Citrus* fruits have been shown multiple times to be unrelated to variations in CS expression levels (Guo et al., 2016, Chen et al., 2013, Hussain et al., 2017, Sadka et al., 2001, Lin et al., 2015, Lu et al., 2016, Yu et al., 2012). With all these points in consideration, the evidence suggests that proton pump activity, particularly by P_{3A}-ATPase *CsPH5*, is a key determinant of citrate levels.

1.2.3 Other PhPH5 structural homologs

A widespread search for P-ATPases within the plant kingdom was conducted to investigate the evolution of the protein family. Functional *PhPH5* homologs were identified in many plant species and their phylogenetic relationship examined (Li et al., 2016). The results indicated that P-ATPase-driven acidification of the vacuole first appeared in gymnosperms before flowers, indicating influence on petal colouration, as seen in *Petunia* (Quattrocchio et al., 2006, Verweij et al., 2008), could not be the only process associated with hyperacidification. One obvious suggestion is to facilitate the bioaccumulation of PAs, which

can be found in some gymnosperms (Niemann, 1988, Stafford and Lester, 1986), because the bioaccumulation of PAs require the H⁺ gradient powered by AHA10 in *Arabidopsis*, the putative structural homolog of PhPH5 (Marinova et al., 2007). In addition to *Petunia* and *Citrus*, P_{3A}-ATPases involved in hyperacidification have been reported in grapevine, apple, and soybean (Amato et al., 2019, Sundaramoorthy et al., 2020, Ma et al., 2019). However, the transcriptional regulation of P-ATPase H⁺ pumps has been well studied with regards to hyperacidification in *Petunia*, which has provided the basis upon which my hypothesis was built.

1.3 Regulation of acidification and the flavonoid pathway

1.3.1 MYB-bHLH-WDR complexes

MBW complexes are highly conserved transcriptional regulatory complexes found in plants. They regulate many processes, such as vacuolar acidity, anthocyanin biosynthesis and PA biosynthesis. The complex is comprised of two classes of transcription factor proteins and a co-regulator protein: MYB, basic helix-loop-helix (bHLH) and WD-repeat (WDR). Within an MBW complex, the respective MYB factor generally provides specificity for target genes conferred by the ability to bind DNA, while bHLH and WDR proteins can be relatively promiscuous, regulating many different processes with other MYB proteins. WRKY transcription factors can also interact with MBW complexes to alter target specificity, enhance expression or even be necessary for expression (Amato et al., 2019, Gonzalez et al., 2016, Verweij et al., 2016). However, recently it has been claimed that binding of the WDR protein to the MBW complex works in competition with its binding to the WRKY protein, at least in the MBW complex conferring trichome initiation in *Arabidopsis* (Lloyd et al., 2017).

In *Arabidopsis thaliana*, anthocyanins, PAs, and mucilage synthesis are governed by regulatory MYB factors AtPAP1, AtTT2/MYB123 and AtMYB5, respectively, which establish independent complexes with bHLH AtbHLH42 (referred to as AtTT8) and WDR AtTTG1 (Baudry et al., 2004, Borevitz et al., 2000, Nesi et al., 2001). A similar model has been proposed in *Petunia*. Vacuolar hyperacidification in epidermal *P. hybrida* petal cells is an important process in determining petal colouration. The cells also accumulate vacuolar anthocyanin pigments whose exact colour is pH-dependent. The regulation of these two processes is determined by MBW complexes that share common subunits, differing only in the MYB factor (Verweij et al., 2008, Quattrocchio et al., 2006, Spelt, 2002). Both complexes comprise bHLH protein PhAN1, WDR protein PhAN11, and MYB factors PhAN2 (anthocyanin-regulating) or PhPH4 (acidification-regulating) structurally homologous to AtPAP1 and AtMYB5, respectively. The MBW complex together with WRKY factor PhPH3 governs the expression of *PhPH5* and *PhPH1* in *Petunia*, which encode tonoplast-bound P_{3A}^- and P_{3B}^- ATPases, respectively (Quattrocchio et al., 2006, Faraco et al., 2014, Verweij et al., 2008). These act as a dimeric proton pump which hyperacidifies vacuoles in petal epidermal cells. Historically, P_{3B}^- -ATPase encoding genes were thought to be absent from the plant kingdom and only found in bacteria (Kühlbrandt, 2004).

In grapevine (*Vitis vinifera*), VvMYC1 (bHLH) and VvWD1 (WDR) also form a multitude of complexes with various MYB factors (Matus et al., 2010). Anthocyanin and PA pathways are controlled by complexes comprising MYB constituents VvMYBA1 & VvMYBA2, and VvMYBPA1 & VvMYBPA2, respectively (Bogs et al., 2007, Kobayashi et al., 2004, Terrier et al., 2009). Recently, Amato et al. (2019) identified genes encoding proteins homologous to PhPH4 and AtMYB5 in grapevine (*VvMYB5a* and *VvMYB5b*) that control the expression of *PhPH1* and *PhPH5* homologs *VvPH1* and *VvPH5*, mediating vacuolar acidification. In addition, the WRKY factor VvWRKY26 is recruited by VvMYB5a, boosting *VvPH5* expression 8-10 fold in dual-luciferase assays.

The conservation of this hyperacidification mechanism between *Petunia* and grapevine is demonstrated in *Petunia ph1*, *ph5*, *ph3* and *ph4* mutants, as the expression of the respective *V. vinifera* homologs can restore pH, petal colouration and target gene expression (Amato et al., 2019, Amato et al., 2016, Li et al., 2016). However, the expression of *A. thaliana PhPH3* homolog *AtTTG2*, which regulates trichome development, seed coat mucilage and PA synthesis, also complements *Petunia ph3* mutants by restoring expression of *PH5* (Gonzalez et al., 2016, Verweij et al., 2016). Similarly, a MYB transcription factor has recently been characterized in *Litchi chinensis* named *LcMYB5* due to its structural homology to *AtMYB5* and *VvMYB5a* & *b*. *LcMYB5* expression correlates with the expression of key genes in the anthocyanin synthesis pathway. A significant decrease in pH was also observed when *LcMYB5* was expressed in *Petunia* and tobacco (Lai et al., 2019).

1.3.2 The 'acidless' phenotype in *Citrus*

Butelli et al. (2012, 2017) have elucidated the genetic regulation of anthocyanin biosynthesis in *Citrus*, identifying *Ruby*. *Ruby* is homologous to *Petunia PhAN2* and *Arabidopsis AtPAP1* and similarly governs the accumulation of anthocyanins, encoding a constituent of an MBW complex. Butelli et al. (2019) present examples, however, where a lack of anthocyanins is observed despite expressing functional *Ruby* alleles, such as the Corsican citron (*C. medica*). Relative to the WT Poncire commun, this variety is characterised by a lack of anthocyanins and PAs, and a higher fruit pH (pH_{vac} 5.42 in comparison to pH_{vac} 2.45) (Butelli et al., 2019). These traits define the 'acidless' phenotype in *Citrus* and is attributed to deletions or

retrotransposon insertions in the *Citrus* gene *Noemi*, encoding the bHLH component that forms an MBW complex with Ruby (Butelli et al., 2019, Reuther et al., 1967). This phenotype is also observed in acidless accessions of *C. limettiodes*, *C. limetta*, and *C. sinensis* (Butelli et al., 2019).

1.4 The hypotheses

Considering *Noemi* is a structural homolog of *PhAN1* and *AtTT8*, and their function is conserved across anthocyanin, PA and ‘acidification’ regulatory MBW complexes, I hypothesised that these processes are regulated by three separate MBW complexes respectively, each comprising bHLH *Noemi*, a WDR protein and a distinct MYB transcription factor in *Citrus* (Figure 1.4.1).

While *Citrus* has retained *PH*-like genes (Li et al., 2016, Shi et al., 2015), independent evolutionary losses have been observed frequently, meaning many plant species lack them. Furthermore, the putative *PhPH4* homologs in *Arabidopsis* (*AtMYB5*) and rice (*OsMYB4*), are involved in mucilage synthesis, seed coat development and chilling tolerance, rather than vacuolar hyperacidification (Li et al., 2009, Soltesz et al., 2012, Vannini et al., 2004). However, the involvement of four *AtMYB5* homologs in hyperacidification has been demonstrated in grapevine, *Petunia*, and *Litchi* (Amato et al., 2019, Lai et al., 2019, Quattrocchio et al., 2006).

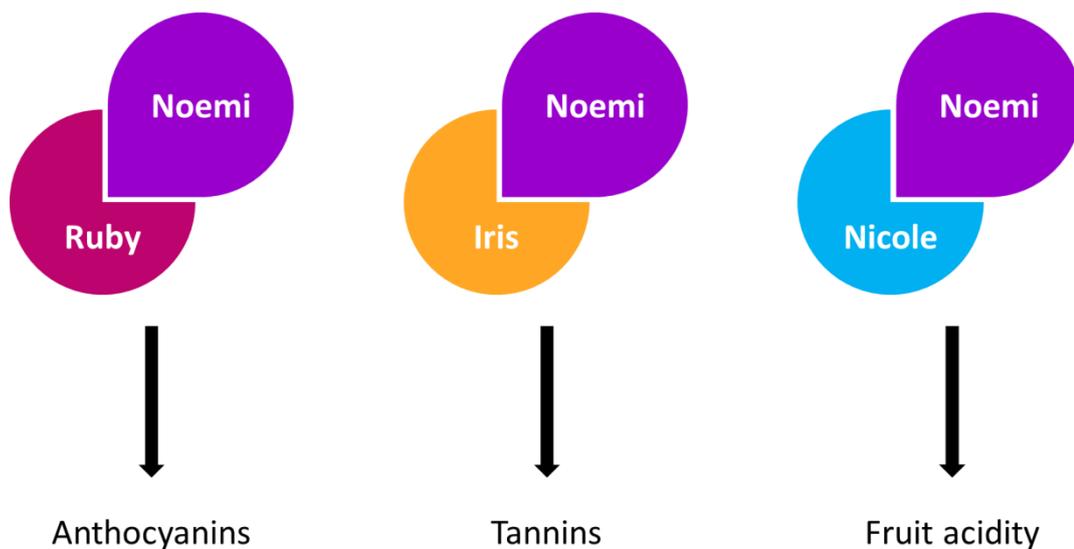


Figure 1.4.1 Proposed regulation of anthocyanin, proanthocyanidins and hyperacidification by transcriptional MBW regulator complexes in *Citrus*. Each MBW complex involves a unique MYB protein (*Ruby*, *Iris*, or *Nicole*) in association with bHLH protein *Noemi*.

By analogy to hyperacidification in *Petunia*, I propose the *PhPH4* homolog in *Citrus*, provisionally named *Nicole*, mediates fruit acidity with *Noemi* via transcriptional regulation of *PH*-like genes, such as *CsPH1* and *CsPH5*. Previous transcriptome analyses of acidic and acidless varieties of sweet orange and mandarin have revealed a strong correlation between acid accumulation and gene expression of *Nicole*, *Noemi*, and a *PH5*-like homolog (Huang et al., 2016, Li et al., 2015). Recent qPCR analysis suggests a reduction of *PH1* and *PH5* expression is correlated with lower *Noemi* (Strazzer et al., 2019), validating previous work of (Butelli et al., 2019).

Nicole has recently been considered to be involved in the PA biosynthetic pathway, showing complementation of the *Arabidopsis tt2* phenotype (Zhang et al., 2020). However, only a weak restoration of function was observed, similar to that of *AtMYB5* (redundant PA regulator) overexpression in *tt2* mutants. Despite this, a *Citrus AtTT2* homolog was not also tested in complementation assays. By analogy to PA regulation by *AtTT2* and *VvMYBPA1 & 2* (Bogs et al., 2007, Nesi et al., 2001, Terrier et al., 2009) a structurally homologous gene, provisionally named *Iris*, has been identified as the candidate gene for PA regulation with *Noemi* in *Citrus*.

Until now, characterised acidless varieties have only been attributed to mutations in *noemi*. Here I present the first example of natural *nicole* mutants in *Citrus*: Lima Sorocaba, Lima Verde R1 and Lima Verde R2. Preliminary analyses indicated these fruits are acidless, yet produce PAs in their seeds. This is the first evidence of acidless *Citrus* mutants that do not conform to the pleiotropic phenotype typically observed. To reveal genes associated with *Nicole*, I have performed differential expression analyses on these 3 *nicole* mutants, a *noemi* mutant Vaniglia, and wild type Navel. In addition, genome-wide identification analyses of the *Citrus* R2R3MYB transcription factor and multi-drug and toxic compound extrusion (MATE) transporter families were performed. Consequently, the putative *AtMYB5* homolog in *C. sinensis* was found and a mutant allele isolated and characterised functionally. Finally, a library of CRISPR/Cas9-edited *nicole* alleles were generated which will be invaluable for characterising gene and protein structure of *Nicole* in the future.

Chapter 2:
General Materials and Methods

2.1 Materials

2.1.1 Citrus fruit

Wild-type *C. sinensis* varieties Navel and Valencia are acidic sweet orange variants used as comparative controls in various experiments within this thesis. Dr Concetta Licciardello (CREA) provided RNA from one ripe Navel biological replicate. An additional two ripe Navel fruits and three ripe Valencia fruits were sourced from Tesco and Waitrose, respectively, for juice sampling and subsequent analyses. As juice was not available for the first Navel biological replicate, Valencia was used as the primary wild-type control in metabolic analyses where only two biological Navel replicates were available. Valencia fruit was also used to obtain seeds due to the lack of seed production in Navel oranges.

Four acidless sweet orange varieties, Lima Verde 1, Lima Verde 2, Lima Sorocaba and Vaniglia were compared with the wild-type varieties in transcriptomic and metabolic analyses. Dr Concetta Licciardello provided RNA samples from three ripe Vaniglia fruit. Three fruits of each of the remaining acidless varieties were sourced from the Sao Paulo region in Brazil. Sorocaba was obtained from a local market, whereas Verde 1 and 2 originate from large *Citrus* farms in Santa Cruz do Rio Pardo and Mogi Guaçu, respectively.

Degree of ripeness in terms of days post-anthesis are not available for the fruit used in this thesis. However, the fruit were obtained during the respective maturation season for each variety. Carboxylic acid and soluble sugar levels show typical variations throughout sweet orange fruit development and so this was measured as a proxy for ripeness.

Primofiori *C. limon* fruit were obtained from a local supermarket for seed collection and germination for subsequent *A. rhizogenes*-mediated transformation of epicotyl tissue.

2.1.2 Transgenic *N. tabacum* plants

Seeds of transgenic *35S:Noemi/Nicole/Iris N. tabacum* were generated and provided by Dr. Eugenio Butelli. *N. tabacum* transformed with *35S:Noemi* were crossed with both *35S:Nicole*

and 35S:*Iris N. tabacum* to generate lines expressing both the genes encoding bHLH protein Noemi and either of the MYB transcription factors.

2.1.3 *N. benthamiana* plants

Lab Strain and Northern Territory varieties of *N. benthamiana* were used, depending on plant availability, for overexpression, dual-luciferase and VIGS assays via agroinfiltration. Plants were grown under a 16/8 h light/dark cycle at 24 °C and 20 °C during the day and night, respectively.

2.1.4 Bacterial strains

E. coli, *A. tumefaciens* and *A. rhizogenes* strains used are listed in Table 2.1.1.

2.1.5 Antibiotics

All antibiotics used and their respective working concentrations are detailed Table 2.1.2.

Table 2.1.1 Bacterial strains used within the thesis and their respective antibiotic selection and purpose.

Species	Strain	Antibiotic selection	Purpose
<i>E. coli</i>	DH5 α		General plasmid cloning
<i>E. coli</i>	ccdB Survival™ 2		Cloning of ccdB plasmids
<i>A. tumefaciens</i>	GV3101	Rifampicin, Gentamicin	Agroinfiltration
<i>A. rhizogenes</i>	K599	Streptomycin	<i>C. limon</i> transformation
<i>A. rhizogenes</i>	ATCC15834		<i>C. limon</i> transformation
<i>A. rhizogenes</i>	MSU440		<i>C. limon</i> transformation

Table 2.1.2 Antibiotics used within the thesis and their respective solvent and working concentrations.

Antibiotic	Solvent	Working conc. ($\mu\text{g ml}^{-1}$)
Ampicillin	H ₂ O	50
Chloramphenicol	Ethanol	34
Gentamicin	H ₂ O	50
Kanamycin	H ₂ O	50-100
Rifampicin	DMSO	50
Tetracycline	70% ethanol	10
Cefotaxime	H ₂ O	400

2.1.6 Plasmids

All plasmids used and their respective antibiotic selection, purpose and method of construction are listed in Supplementary Table 1

2.1.7 Media recipes

All media used and their respective recipes are listed in Supplementary Table 2.

2.1.8 Reference genomes

Reference genomes for *C. sinensis* and *Nicotiana* species were used for primer design, phylogenetics and RNA sequencing. Gene accession IDs for all genes referred to in the thesis are listed in Supplementary Table 3.

C. sinensis genome sequences were downloaded from Phytozome (<https://phytozome-next.jgi.doe.gov/>, *C. sinensis* v1.1, Wu et al., 2014a) and the *Citrus* Pan-genome to Breeding

Database (<http://citrus.hzau.edu.cn/index.php>, *C. sinensis* v2.0, Xu et al., 2013). There is some minor variation in sequence and gene annotation between these two reference genomes. As a result, standard PCR and RT qPCR primers were designed to bind to regions identical in both reference genomes. However, RNA sequencing and phylogenetic analyses used the Phytozome sweet orange genome exclusively.

Draft genomes for both *N. tabacum* (*N. tabacum* v1.0, Edwards et al., 2017) and *N. benthamiana* (*N. benthamiana* v1.0.1, Bombarely et al., 2012) were accessed from on the Sol Genomics Network (<https://solgenomics.net/>).

2.2 Methods

2.2.1 Primer design

The Primer3 website (<https://primer3.ut.ee/>) was used to design primers for regular PCR, sequencing, and qPCR. If necessary, basic local alignment search tool (BLAST) was used to check primer specificity to the target sequence with CLC Main Workbench v8.0.1 and the respective genome sequence. Optimal primer design considered a length of 18-28 bp, a GC content of 35-65 %, a T_m of 60 °C and minimal secondary structures.

RT qPCR primers were designed to span exon-exon junctions or were separated by large (> 400 bp) introns, to avoid amplification of residual DNA contamination, and to produce an amplicon length of 75-200 bp. Primers designed for more specific purposes, such as GoldenGate cloning and sgRNA synthesis, may differ from these optimal conditions due to the nature of the experiment. All primers used are listed in Supplementary Table 4.

2.2.2 Polymerase chain reaction

G-Storm thermal cyclers (Kapa Biosystems) were used for all PCRs. For standard applications the reaction comprised Taq DNA polymerase (Qiagen) and the respective buffer. Phusion (NEB) and the corresponding Phusion HF Buffer were used for higher fidelity applications, such as sequencing. Where bacterial colonies or cultures were used as DNA template the denaturation step was extended to release plasmid DNA from bacterial cells. Typical reaction composition and thermocycler parameters are detailed in Table 2.2.1 and Table 2.2.2.

2.2.3 PCR and agarose gel purification

PCR amplicons, restriction digestion products and other nucleic acid samples requiring downstream application, such as sequencing, were purified or extracted from TBE buffer agarose gels (0.5 – 2.0 % w/v) using the QiaQuick PCR Purification Kit (Qiagen) and QiaQuick Gel Extraction Kit (Qiagen), respectively, according to the manufacturer's protocol.

Table 2.2.1 Typical PCR composition per 50 μ l reaction (a: Taq DNA polymerase (Qiagen); b: Phusion (NEB)).

Component	50 μl reaction	Final concentration
DNA template:		
Genomic	~ 100 ng	~ 2 ng μ l ⁻¹
Plasmid	~ 10 ng	~ 0.2 ng μ l ⁻¹
PCR buffer	5 ^a / 10 ^b μ l	1X
25 mM dNTPs	0.4 μ l	200 μ M
10 μ M Forward Primer	2.5 μ l	0.5 μ M
10 μ M Reverse Primer	2.5 μ l	0.5 μ M
Polymerase	0.25 ^a / 0.5 ^b μ l	2.5 ^a / 1 ^b units
H ₂ O	to 50 μ l	

Table 2.2.2 Typical PCR thermocycler parameters (a: Taq DNA polymerase (Qiagen); b: Phusion (NEB); c: extension time was dependent on the respective polymerase manufacturer's specifications for target product size).

Step	Temperature ($^{\circ}$C)	Time
Initial denaturation:		
Colony PCR	95 ^a	5 min
Standard PCR	95 ^a / 98 ^b	30 sec
25 – 35 cycles:		
Denaturation	95 ^a / 98 ^b	30 sec
Annealing	55 - 72	30 sec
Extension	72	15 sec – 2 min ^c
Final extension	72	10 min

2.2.4 Plant tissue homogenisation

Plant tissue homogenisation was achieved by disrupting samples contained in 2 ml Eppendorf's with a 0.5 mm steel bead in a TissueLyser LT (Qiagen) at 50 Hz for 2 min. Samples and the TissueLyser LT adapter were incubated on dry ice for ≥ 30 min prior to disruption.

2.2.5 Plasmid DNA extraction

Plasmid DNA was extracted from 24 h 10 ml bacterial LB liquid cultures. *E. coli* and *Agrobacterium* cultures were incubated at 37 °C and 28 °C, respectively, and agitated at 220 rpm. Cells were collected by centrifugation at 3,300 rpm for 15 min and processed with the QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturer's protocol.

2.2.6 Quantification of nucleic acids

Nucleic acid concentration and purity was measured using NanoDrop One spectrophotometer (Thermo Scientific).

2.2.7 DNA sequencing

DNA samples were prepared using the Mix2Seq Kit (Eurofins) according to the manufacturer's protocol and sent to Eurofins Scientific for overnight sequencing. Raw chromatogram sequence data was assembled and analysed using default settings in CLC Main Workbench v8.0.1.

2.2.8 cDNA synthesis

For qPCR analysis cDNA was synthesised from extracted RNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's protocol.

2.2.9 Real-time quantitative PCR analysis

RT qPCR was used to validate *C. sinensis* RNAseq transcriptomic data and to analyse gene expression in stably and transiently transformed tobacco. The cDNA samples, synthesised from RNA as previously detailed, were diluted 5-fold with dH₂O. Reaction composition using the SYBR Green JumpStart Taq ReadyMix Kit (Sigma) is detailed in Table 2.2.3. RT qPCR experiments were conducted with the X96 Touch Real-Time PCR Detection System (Biorad) using thermocycler parameters detailed in Table 2.2.4.

Table 2.2.3 RT PCR composition per 20 µl reaction.

Component	20 µl reaction
cDNA template	5 µl
10 µM Forward Primer	1 µl
10 µM Reverse Primer	1 µl
SYBR Green Taq ReadyMix	10 µl
H ₂ O	to 20 µl

Table 2.2.4 RT qPCR thermocycler parameters.

Step	Temperature (°C)	Time
Initial denaturation	94	4 min
40 – 50 cycles:		
Denaturation	94	10 sec
Annealing	60	10 sec
Extension	72	15 sec
Final extension	72	1 min

RT qPCR primers were first tested on a 5-fold serial dilution of a cDNA mix of all samples within the respective experiment to confirm primer efficiency equalled $100\% \pm 10\%$. Primer efficiency was calculated with the following equation, where m = gradient of the standard curve:

$$\left(10^{(-1/m)} - 1\right) \times 100$$

Each sample had 3 technical replicates and gene expression was normalised against housekeeping genes Actin and EF1 α for *Citrus*, and Actin for *Nicotiana* RT qPCRs. Normalised expression values (ΔC_q) were calculated with the following equation:

$$2^{(Cq_{HK} - Cq_x)}$$

Where Cq_{HK} and Cq_x represent quantitation cycle of the housekeeping gene(s) and the gene of interest, respectively.

2.2.10 Preparation of competent *E. coli* cells

One ml of a 24 h 10 ml *E. coli* LB liquid culture was used to inoculate a 100 ml LB culture. This was incubated at 37 °C, 220 rpm, overnight. When an OD₆₀₀ of 0.35 was reached the culture was incubated in 50 ml falcon tubes on ice for 15 min. Cells were collected by centrifugation at 1,800 g for 10 min at 4 °C and resuspended in 30 ml of cool 0.1 M CaCl₂. Following a 30 min incubation on ice, the culture was centrifugated again at 1,800 g, 4 °C for 10 min. Finally, the bacterial pellet was resuspended in 2 ml of cool 0.1 M CaCl₂ and 2 ml 30 % glycerol and stored at -80 °C in 50 μ l aliquots.

2.2.11 *E. coli* transformation

A 50 μ l aliquot of competent *E. coli* cells were thawed and incubated on ice for 20 min with ≤ 50 ng of plasmid DNA. After a 50 s incubation at 42 °C, and 2 min incubation on ice, 950 μ l SOC medium was added to the cells. *E. coli* cells were then agitated for 1 h 30 min at 37 °C, 220 rpm. A 100 μ l aliquot of transformed cells were spread on LB agar medium and incubated under selective conditions overnight at 37 °C.

2.2.12 Preparation of electrocompetent *Agrobacterium* cells

A 250 ml LB liquid culture was inoculated with a single *Agrobacterium* colony and grown overnight at 28 °C, 220 rpm. The culture was diluted to OD₆₀₀ 0.6, and 200 ml was cooled on ice for 15 min. By collecting cells by centrifugation at 5,000 rpm for 10 min, 4 °C, the cells were washed sequentially with 200 ml, 100 ml, 40 ml, and 4 ml of 10 % glycerol. Finally, the cells were resuspended in 2 ml 10 % glycerol and stored in 50 µl aliquots at -80 °C.

2.2.13 *Agrobacterium* transformation

Aliquots of 50 µl electro-competent *A. tumefaciens* or *A. rhizogenes* cells were thawed on ice. The cells were transferred with ≤ 200 ng plasmid DNA to a pre-cooled electroporation cuvette. To transfer recombinant plasmid DNA into the cells a high voltage pulse was applied to the cuvette (field strength: 12.5 kV/cm; capacitance: 25 µF; resistance: 200 Ω). *Agrobacterium* cells were resuspended in 1 ml SOC and incubated at 28 °C, 220 rpm. After 3 h, 100 µl of the suspension was aliquoted and spread onto LB agar medium supplemented with appropriate antibiotics. Transformed *A. tumefaciens* and *A. rhizogenes* colonies developed by day 3 and 5, respectively, of incubation at 28 °C.

2.2.14 Restriction digestion and ligation

High fidelity restriction enzymes supplied by New England Biolabs were used in this thesis. The reaction composition is listed in Table 2.2.5 and the thermocycler parameters for incubation and heat inactivation steps were performed according to the manufacturer's restriction enzyme specifications. If heat inactivation was not possible for a restriction enzyme the reaction was inactivated and digested products purified with the QiaQuick PCR Purification Kit (Qiagen) mentioned previously.

Ligation of digested DNA products at an insert to vector molar ratio of 3:1 was achieved with T4 DNA ligase (New England Biolabs), according to the manufacturer's protocol. The ligation reaction composition is detailed in Table 2.2.6. Once complete, 5 µl of the reaction were then transformed into competent *E. coli* cells. Reactions containing no insert were also transformed into *E. coli* to check background levels of self-ligation or incomplete digestion of the recipient plasmid.

Table 2.2.5 Restriction digestion composition per 50 μ l reaction.

Component	50 μl reaction
DNA	1 μ g
rCutSmart Buffer	5 μ l
Restriction enzyme(s)	1 μ l (each)
H ₂ O	to 50 μ l

Table 2.2.6 Ligation composition per 20 μ l reaction (a: Insert DNA required for a 3:1 insert to vector ratio was calculated per reaction based on insert size).

Component	20 μl reaction
Vector DNA	50 ng
Insert DNA	variable ^a
T4 DNA Ligase Buffer	2 μ l
T4 DNA Ligase	1 μ l
H ₂ O	to 20 μ l

2.2.15 Overexpression plasmid construction

The CDS of *C. sinensis* genes encoding R2R3-MYB, bHLH and WRKY proteins were amplified from WT sweet orange cDNA and cloned into the pEAQ-HT-DEST1 vector downstream of a 35S promoter. These plasmids were constructed by Dr Eugenio Butelli using Gateway Cloning technology.

2.2.16 TOPO cloning

TOPO cloning was utilised when isolation of multiple PCR amplicons could not be achieved by excision and extraction from agarose gels due to similarity in bp size. PCR purified samples were processed using the Zero Blunt TOPO cloning Kit (Invitrogen) according to the manufacturer's protocol. This method ligates blunt-end products derived from Phusion-mediated PCRs into the VECTOR. Following *E. coli* transformation, all amplicons within the PCR product sample were sequenced by extracting DNA from multiple colonies. Plasmid DNA extraction was performed on up to 6 colonies per *C. limon* PCR when possible.

Chapter 3:
Metabolic and transcriptomic analyses of acidless
C. sinensis varieties

3.1 Introduction

3.1.1 The roles of AtTT2 and AtMYB5 in proanthocyanidin regulation

PA biosynthesis occurs late in the flavonoid pathway, sharing intermediary metabolites with the anthocyanin pathway before the possibility of branching off from leucoanthocyanidins and anthocyanidins by catalysis of PA-specific structural proteins LAR and ANR (Figure 3.1.1) (Abrahams et al., 2003, Tanner et al., 2003, Pang et al., 2007, Liu et al., 2016b, Xie et al., 2003, Paolocci et al., 2007, Matsui et al., 2016, Li et al., 2019a). These genes are responsible for biosynthesis of stereochemical isomers (+)-catechin and (–)-epicatechin, respectively, which are initiating monomeric precursor flavan-3-ols of PA polymers. While LDOX is not solely committed to the PA pathway, it is essential in species that exclusively bioaccumulate epicatechin-comprised PAs. One example is *Arabidopsis*, which lacks a putative LAR homolog (Abrahams et al., 2003, Tanner et al., 2003).

In *Arabidopsis*, PA bioaccumulation is governed primarily by AtTT2-driven MBW complexes comprising WDR AtTTG1 and a bHLH (AtTT8/AtGL3/AtEGL3) in any tissue producing PAs (Xu et al., 2014, Nesi et al., 2001, Baudry et al., 2004, Nesi et al., 2000). These complexes regulate the expression of key structural PA-related genes *AtDFR*, *AtLDOX*, *AtBAN/ANR*, *AtAHA10* and the MATE transporter *AtTT12*. Transport into and vacuolar pH homeostasis is summarised in Figure 3.1.2. In comparison, the AtMYB5-containing MBW complex functions only in the endothelium regarding PA regulation and can regulate only a subset of genes within the PA biosynthetic pathway (*AtDFR*, *AtLDOX*, *AtTT12*). Furthermore, *tt2 Arabidopsis* mutants produce no PAs, whereas *myb5* mutants only exhibit a reduction in accumulation (Nesi et al., 2001). AtMYB5 is also only able to partially recover the wildtype phenotype in *tt2* mutants (Xu et al., 2014). In fact, AtMYB5 also has a weak role in trichome development but its main function concerns seed coat development (Gonzalez et al., 2009, Li et al., 2009).

With regards to PA regulation a similar story is seen in grapevine. AtTT2-like putative homologs VvMYBPA1 and VvMYBPA2, and VvMYBPAR are the major determinants of PA regulation, whereas distantly related VvMYB5a and VvMYB5b can induce promoters of select genes within the flavonoid pathway only (Bogs et al., 2007, Deluc et al., 2008, Koyama et al., 2014, Terrier et al., 2009, Hichri et al., 2010). *VvMYB5a* and *VvMYB5b* show continual

expression in tissues lacking flavonoids, and in grape skin after veraison when *VvANR* and *VvLAR* expression is low and PA biosynthesis has halted, also suggesting functional redundancy. More recently, Amato et al. (2019) reported the recruitment of a WRKY factor by *VvMYB5a* specifically which is required for the regulation of genes in the vacuolar acidification pathway by *VvMYB5a* and *VvMYB5b* participating in MBW complexes.

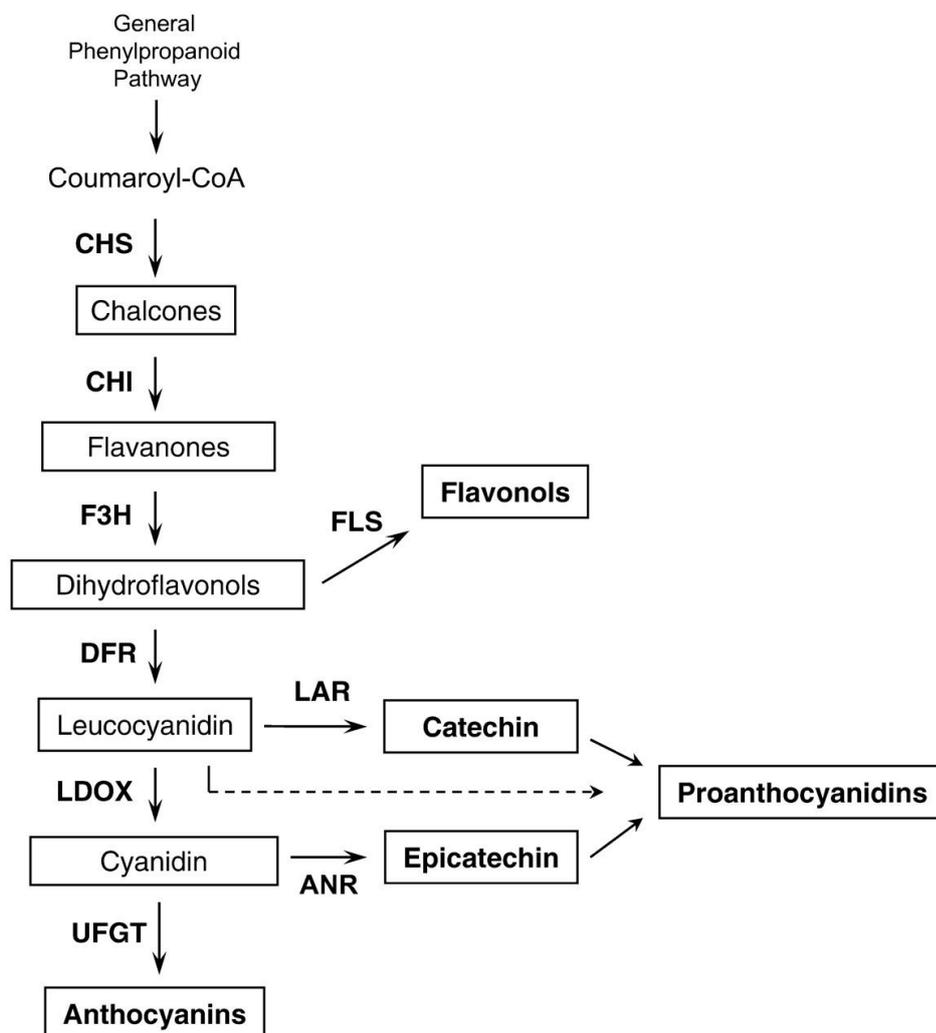


Figure 3.1.1 Schematic of the flavonoid pathway, including the proanthocyanidin sub-branch by catalysis of LAR and ANR to produce catechin and epicatechin, respectively. Figure reused and unmodified from Bogs et al. (2007), made available for unrestricted use, distribution, and reproduction under the terms of the Creative Commons CC BY public license (<https://creativecommons.org/licenses/by/4.0/legalcode>).

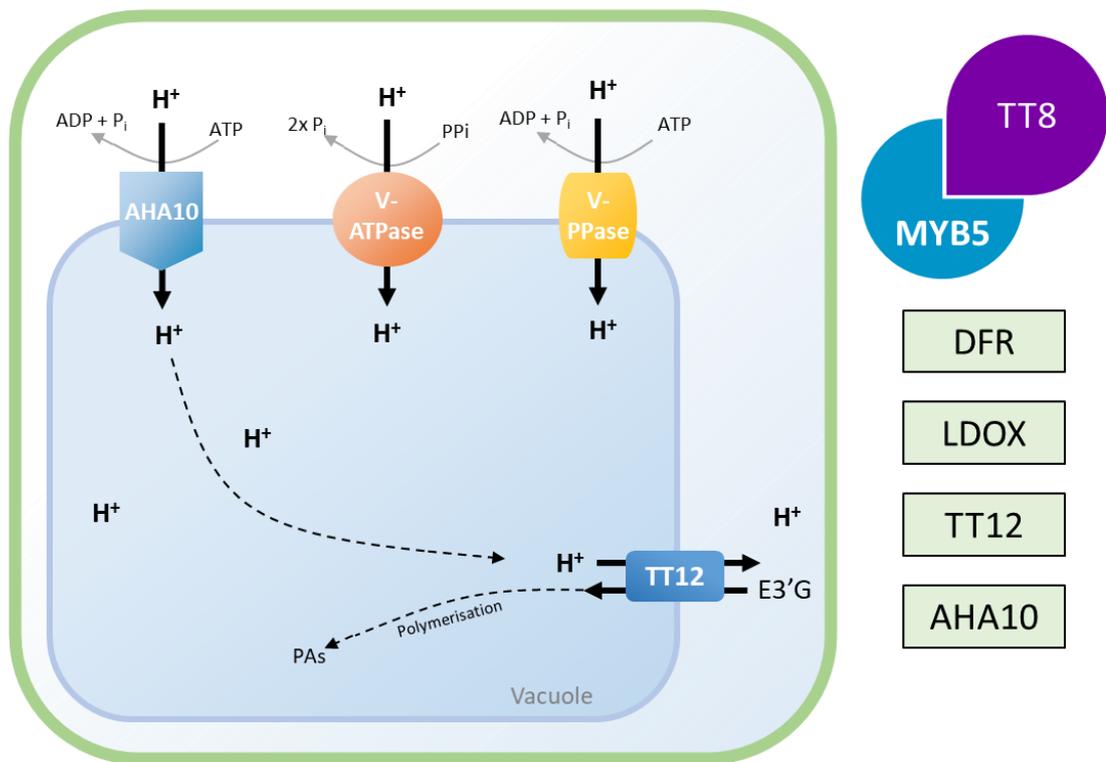


Figure 3.1.2 Simplified schematic of proton homeostasis and PA transport into the vacuole of *Arabidopsis endothelium* cells. E3'G: 3'-glucosylated epicatechin. List of PA-related genes activated by MBW complex comprising AtMYB5 and AtTT8 are in green (right).

3.2 Materials and Methods

3.2.1 pH and carboxylic acid analyses

The pH of fresh juice sampled from Navel, Valencia, Sorocaba, Verde R1 and Verde R2 sweet oranges was measured following centrifugation at 4,000 rpm for 15 min. Organic acids and soluble sugars were quantified by gas chromatography-mass spectrometry (GCMS) as described previously by Lin et al. (2015). A 1 ml aliquot of fresh juice, sampled from Navel, Valencia, Sorocaba, Verde R1 and Verde R2 sweet oranges was added to 1.4 ml methanol and incubated at 70 °C for 15 min. Samples were centrifugated at 14,000 rpm for 10 min and the supernatant stored until analysis at -80 °C. A 10 µl aliquot of 1.2 mg ml⁻¹ ribitol (internal standard) was added to 10 µl sample and vacuum dried. The residue was incubated in 50 µl 20 mg ml⁻¹ methoxyamine in pyridine at 30 °C with agitation for 1 h 30 min. The samples were then incubated at 37 °C with agitation for a further 30 min following the addition of 100 µl N-methyl-N-trimethylsilyltrifluoroacetamide.

Each sample (1 µl) was injected into the GCMS in split mode with a Phenomenex Zebron ZB5-HT inferno 30 m (+5 m guard) x 250 µm x 0.25 µm column and a 10:1 split ratio. Temperature of the injector was 250 °C and the helium carrier gas had a 0.85 ml min⁻¹ flow rate. Column temperature was kept at 60 °C for 1 min, increased at a rate of 10 °C min⁻¹ to 325 °C, then held at this temperature for 10 min. The temperature of the MS transfer line was 290 °C. The mass spectra of m/z 50 – 600 was scanned with a solvent delay of 5 min 54 sec. Source temperature was set to 250 °C. Organic acids and soluble sugars were quantified by calibration against citric acid, malic acid, sucrose, glucose, and fructose standard curves.

3.2.2 Cellular localisation of proanthocyanidins in *C. sinensis* seeds

External seed coats were removed from Valencia, Vaniglia, Sorocaba, Verde R1 and Verde R2 seeds. One end of the seed was sliced with a scalpel to ease infiltration of the staining and fixative solutions into the sample. Seed samples were then stained with ice-cold 4-(dimethylamino)cinnamaldehyde (DMACA) solution (0.3 % (w/v) DMACA, 50 % (v/v) methanol, 3 M HCl) for 20 min, washed with 70 % ethanol 4 times and stored in 70 % (v/v)

ethanol. Seeds were first photographed then Valencia and Vaniglia seeds processed as described by Abeynayake et al. (2011) for characterising cellular localisation of PAs.

Stained samples were fixed by 1 min vacuum-infiltration in 50 mM sodium phosphate buffer (pH 7.4) containing 4 % (w/v) paraformaldehyde and 6 % (w/v) glutaraldehyde and incubated at 4 °C for 2 h. Samples were washed in 50 mM sodium phosphate buffer (pH 7.4) three times for 5 min. The fixed samples were then embedded in LR white resin. First, they were dehydrated in sequential washes (ethanol series: 30 %, 60 %, 70 %, 90% and 100 %) 15 min each. The ethanol was removed, and the sample was immersed in 3:7 LR white resin to 100 % ethanol for 1 h at ambient temperature. Samples were then immersed in 7:3 LR white resin to 100 % ethanol for another hour, followed by 14 h immersion in 100 % LR white resin. Seed samples were then incubated for 14 h at 60 °C under vacuum. Cross sections were taken using a microtome (6 – 10 µm). Images were taken at 40x magnification using differential interference contrast (DIC) microscopy.

3.2.3 Extraction and quantification of proanthocyanidins

PAs were extracted from fresh sweet orange juice sampled from Navel, Valencia, Sorocaba, Verde R1 and Verde R2 fruit. *Citrus* juice was extracted twice with 10X volume (v/w and v/v, respectively) of an extraction solution, containing 70 % (v/v) acetone and 0.5 % (v/v) acetic acid, and room temperature sonication for 30 min. Extractions were centrifugated at 8,000 rpm for 10 min and supernatants containing soluble PAs were pooled.

Soluble PA samples were washed three times with chloroform, and a further three times with hexane, before freeze-drying overnight. PA powder was resuspended in 1 ml extraction solution per 1 g of fresh starting material fresh. Quantification of total soluble PAs was calculated by measuring absorbance at 640 nm following a reaction of 15 µl sample with 85 µl DMACA solution (0.3 % (w/v) DMACA, 50 % (v/v) methanol, 3 M HCl). Absorbance was measured using a CLARIOstar Plus plate reader at approximately 3 min intervals starting 2 min after initiating the reaction, with constant agitation in between. Values were calibrated against a (+)-catechin standard curve.

For high-performance liquid chromatography (HPLC) analyses, soluble PA extracts were diluted 5-fold in H₂O and run on a Shimadzu Nexera UHPLC with Prominence diode array detector (UV/vis absorbance) and a 2020 single quad mass spectrometer. Separation was on a 50 × 2.1 mm, 2.6 μm particle size Kinetex EVO C18 column (Phenomenex) using the gradient of acetonitrile versus 0.1 % (w/v) formic acid in H₂O detailed in Table 3.2.1, run at 0.6 ml min⁻¹ at 40 °C. Detection was by UV absorbance collecting full spectra from 200-600 nm at 12.5 Hz with a time constant of 0.08 sec, and by positive mode electrospray MS. The mass spec collected full spectra from *m/z* 100-900 in 0.1 sec and also monitored *m/z* 291⁽⁺⁾ by single-ion-monitoring for 50 msec. Spray chamber conditions were 250 °C desorbation line, 200 °C heat block, 1.5 L min⁻¹ nebulizer gas, and 15 L min⁻¹ drying gas. Injection volume was 5 μl. Flavan-3-ol monomers were quantified at 279 nm by calibration against (+)-catechin and (-)-epicatechin standard curves. Following acid-catalysis of soluble PA extracts, flavan-3-ol terminal subunits from PA polymers are released in addition to extension subunits. Terminal subunit concentration was determined by the subtraction of free flavan-3-ol monomer concentration in uncleaved PA samples from the total flavan-3-ol monomer concentration in the acid-catalysed samples.

Table 3.2.1 Proanthocyanidin HPLC analyses separation gradient of acetonitrile versus 0.1 % (w/v) formic acid in H₂O.

Time (min:sec)	Acetonitrile (%)
0:00	2
2:00	10
4:00	35
6:00	95
7:00	95
7:10	2
9:10	2

3.2.4 Identification of *C. sinensis* homologs

Genes encoding proteins structurally homologous to those encoded by genes of interest from other species, such as *A. thaliana*, *P. hybrida* and *V. vinifera*, were identified via BLAST alignment of amino acid sequences with the Phytozome *C. sinensis* genome. Phylogenetic analyses also informed the identification for R2R3MYB transcription factors and MATE transporters. Identified genes were compared with those previously identified in literature. All genes referred to in this thesis, their Phytozome accession ID, related publications (or source), and my designation are included in Supplementary Table 3.

3.2.5 Multiple sequence alignment

Alignments of a few sequences which did not require phylogenetic analysis were performed using default settings in CLC Main Workbench v8.0.1. In comparison, multiple sequence alignments (MSA) were generated using default Clustal Omega (ClustalO) method settings in the online tool available from EMBL-EBI (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) when phylogenetic analysis was required downstream.

3.2.6 Phylogenetic analyses

Genome-wide analyses of MYB and MATE protein families in the *C. sinensis* v1.1 Phytozome reference genome were performed by protein sequence structure analyses, MSA and subsequent phylogenetic tree construction with homologous *A. thaliana* gene families. Other characterised proteins from different species were also included to discern the likely metabolic function of *C. sinensis* proteins. Initially, candidate genes encoding MYB TFs and MATEs in *C. sinensis* were identified by reference genome AA sequence analysis of PFAM domains PF00249 (MYB-like DNA binding domain) and PF01554 (MATE), respectively.

Candidate sweet orange MYB-like TFs containing two predicted MYB domains were selected and designated as R2R3-CsMYBs. An MSA, using MEGA as previously described, of homologous R2 and R3 MYB domains was generated to investigate AA conservation across the R2R3-MYB proteins in *C. sinensis* and validate results against previous findings that used a different reference genome (Hou et al., 2014).

Full-length R2R3-CsMYB AA sequences were also subjected to an MSA alone and with R2R3-MYBs from other plants species. These included 125 R2R3-AtMYBs and other characterised R2R3-MYB proteins from *V. vinifera* (n=5), *P. hybrida* (n=2), *L. chinensis* (n=2), *Medicago truncatula* (n=3), *Malus domestica* (n=4), *Prunus persica* (n=3), *Glycine max* (n=1) and *Fragaria ananassa* (n=1). The maximum-likelihood (ML) substitution model that best described each alignment was determined using PhyML (Guindon et al., 2010). ML phylogenetic trees were constructed using the best substitution model, for the respective alignment (stated in corresponding figure captions), with RAxML (1,000 bootstrap replicates) (Stamatakis, 2014).

Candidate *C. sinensis* MATEs were filtered according to typical plant MATE protein features: ≥ 400 AAs and 8-12 transmembrane domains. MATE transmembrane domain prediction was achieved using TMHMM v2.0 (<https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>). Proteins that adhered to these constraints were designated as CsMATEs. The full-length AA sequences of CsMATEs were aligned alone and with 56 AtMATE (AtDTX) and other functionally characterised MATE proteins from *V. vinifera* (n=4), *Brassica rapa* (n=1), *B. oleracea* (n=1), *Medicago truncatula* (n=2), *Malus domestica* (n=2), *N. tabacum* (n=3), *Solanum lycopersicum* (n=1), *Eucalyptus camaldulensis* (n=1), *Oryza sativa* (n=2), *Glycine max* (n=1), *Triticum aestivum* (n=1) and *Zea mays* (n=1) as previously described. The ML substitution model that best described each alignment was determined using PhyML (Guindon et al., 2010). ML phylogenetic trees were constructed using the best substitution model, for the respective alignment (stated in corresponding figure captions), with RAxML (1,000 bootstrap replicates) (Stamatakis, 2014).

3.2.7 Gene structural analyses

Gene structures of the R2R3-MYB and MATE gene families were mapped to the corresponding phylogenetic tree using the Gene Structure Display Server (Hu et al., 2015).

3.2.8 Citrus genomic DNA extraction

Fruit DNA was extracted from juice sampled from fresh sweet oranges. A 500 µl aliquot of juice was incubated on ice for 20 minutes with 3.5 ml TES extraction buffer (Tris 0.2 M pH 8, EDTA 1 mM, SDS 1 %). DNA was isolated by centrifugation at 8,000 rpm for 10 minutes following sequential 4 ml phenol (pH 7.9) and 4 ml chloroform washes. DNA in the aqueous phase was precipitated with 0.10X volume 3 M sodium acetate (pH 5.2) and 2X volume 100 % ethanol on ice for 20 minutes before 14,000 rpm centrifugation for 5 minutes. The pellet was washed with 70 % ethanol centrifugated at 14,000 rpm for 5 minutes. After removing the supernatant and drying at room temperature, RNA contamination was eliminated by resuspending the pellet in 50 µl RNase H₂O solution (0.1 mg ml⁻¹) and incubating at 37 °C for 30 minutes.

3.2.9 Citrus RNA extraction

Citrus fruit RNA was extracted from 3 ml juice from two Navel, and 3 Lima Sorocaba, Lima Verde R1 and Lima Verde R2 fruits. The Lima fruits were relatively ripe, with wholly orange skin. The exact degree of ripening in terms of days post anthesis was unknown, however. All centrifugation steps took place at 4 °C. Juice was incubated for 5 minutes at 50 °C with 3 ml extraction buffer (Tris 0.2 M, NaCl 0.2 M, EDTA 50 mM, SDS 2 %) and 60 µl β-mercaptoethanol. Separation of aqueous and organic phases was achieved by centrifugation at 4,000 rpm for 30 minutes. The aqueous phase was then repeatedly washed with 1X volume chloroform, and centrifugated at 4,000 rpm for 15 minutes, until the intermediate phase became clear. RNA in the aqueous phase was precipitated with 0.5X volume lithium chloride (6 M) overnight at 4 °C. RNA was pelleted and then washed with 70 % ethanol by centrifugation at 14,000 rpm for 40 and 20 minutes, respectively. After removing the supernatant, the pellet was dried and resuspended in 50 µl DEPC treated dH₂O. All 50 µl RNA were purified with the RNeasy Plant Mini Kit (Qiagen) as detailed in the manufacturer's instructions. Briefly, RNA samples were mixed with binding buffer and ethanol and applied to a RNeasy spin column to bind RNA to the membrane. Following two washes with an ethanol-based washing buffer the RNA was eluted with H₂O.

3.2.10 RNA sequencing

Citrus RNA samples were sent to Novogene for low input PE150 Illumina sequencing. Novogene conducted the initial analyses, providing FPKM values for all genes in all samples and significance values when compared to expression in WT Navel fruit juice. R (v4.0.5) was used to generate lists of differentially expressed genes in the acidless varieties and GO enrichment analyses using R package topGO (Alexa and Rahnenführer, 2009).

3.3 Results

3.3.1 Fruit sugar, organic acid, and pH analyses

Organic acid content and pH measurements were performed on fresh juice sampled from acidic and acidless *C. sinensis* varieties. Qualitatively, the sweet orange fruit sampled were ripe (Figure 3.3.1). All acidless varieties were wholly orange with no green regions on the skin. Fruit at different degrees of ripeness were not available, however. Since soluble sugar levels show typical variation throughout sweet orange fruit development, this was measured as a proxy for ripeness. As mentioned in the General Materials and Methods section (2.1.1), it must be noted that only two biological replicates were available for metabolic analyses of Navel juice, as biological replicate Navel A comprised an RNA sample only, which was provided early during my PhD by Dr Concetta Licciardello.

Fruit juice pH of all varieties tested were significantly different to Navel (P -value < 0.05; Figure 3.3.2). Valencia sweet oranges, also considered WT, were marginally more acidic, averaging a pH of 3.71 compared to 4.09 in Navel. Verde R1 and Verde R2 juice pHs, however, were considerably higher at 5.78 and 5.89, respectively. Sorocaba fruit juice pH was almost neutral with a pH of 6.42.

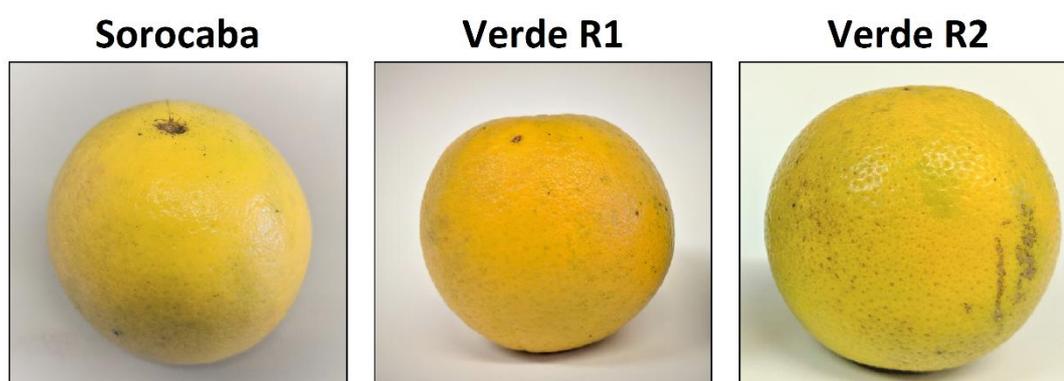


Figure 3.3.1 Acidless Sorocaba, VerdeR1 and VerdeR2 *C. sinensis* fruit immediately before sampling.

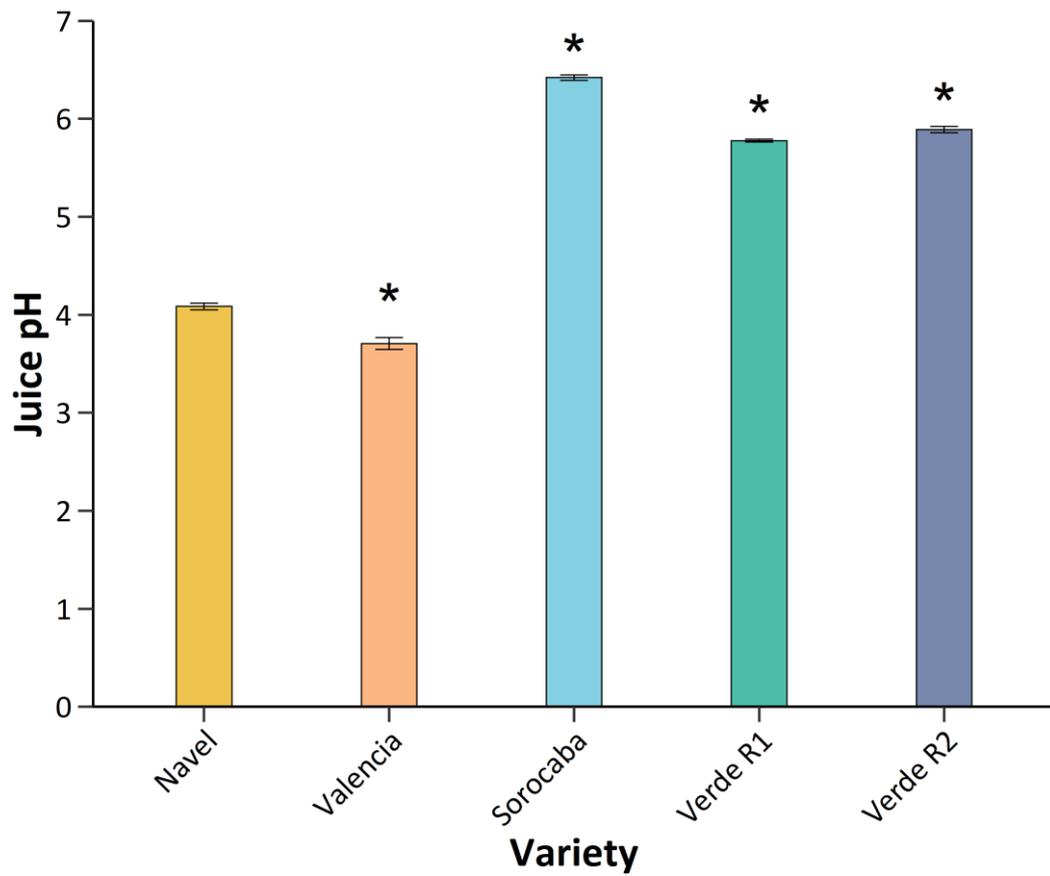


Figure 3.3.2 Juice pH of five *C. sinensis* varieties: Navel, Valencia, Sorocaba, Verde R1 and Verde R2. Asterisks indicate significant difference relative to the Navel control. Values and error bars presented represent the mean of 3 (2 for Navel) biological reps \pm se.

Juice concentrations of organic acids (citric and malic acid) were quantified by a GCMS. Citric acid was the most dominant organic acid in the WT varieties (Figure 3.3.3a). Conversely, citric acid content of all three acidless varieties were considerably lower relative to both WT varieties. Navel and Valencia citric acid concentration in juice averaged at 11.06 and 15.09 mg ml⁻¹, respectively. In comparison, Sorocaba, Verde R1 and Verde R2 had concentrations ranging only 0.28 - 0.67 mg ml⁻¹. Despite this great reduction, compared to Navel, the difference was found to be only almost significant (Sorocaba: *P*-value = 0.068; Verde R1: *P*-value = 0.071; Verde R2: *P*-value = 0.069). Valencia citric acid concentration was not significantly different to that in Navel (*P*-value = 0.204).

Juice malic acid content of two acidless varieties, Sorocaba, and Verde R2, was significantly reduced, compared to Navel (*P*-value < 0.05; Figure 3.3.3b). Notably, malic acid levels in Sorocaba, Verde R1 and Verde R2 were not significantly different from the other WT variety, Valencia. Concerning the composition of organic acids in fruit juice, the concentration of malic acid was approximately half of citric acid levels in Navel (5.62 mg ml⁻¹). Conversely, in all acidless varieties the malic acid content was up to 6.5-fold greater than citric acid.

Soluble sugar content was also quantified via GCMS analyses of fruit juice (Figure 3.3.4). Total soluble sugars levels were relatively similar amongst all sweet orange varieties (Navel = 95.4 mg ml⁻¹; Valencia = 80.8 mg ml⁻¹; Sorocaba = 68.3 mg ml⁻¹; Verde R1 = 86.7 mg ml⁻¹; Verde R2 = 63.3 mg ml⁻¹). By comparison, differences in soluble sugar composition were observed. In Navel and Valencia sucrose accounted for around 50 % of all sugars, whereas glucose and fructose levels were greater than sucrose in every acidless variety. Sucrose content of Sorocaba, Verde R1 and Verde R2 juice was approximately half of that quantified in Navel. Despite this, there were no significant differences in concentration of any soluble sugar in juice of any variety, relative to Navel.

3.3.2 Proanthocyanidin quantification and localisation

Noemi Citrus mutants typically exhibit a pleiotropic phenotype consisting of a lack of anthocyanins, PAs, and reduced acidity. 4-(dimethylamino)cinnamaldehyde (DMACA; Treutter, 1989) staining of *Citrus* seed is a relatively fast method to confirm the presence of PAs in the seed coat. Figure 3.3.5 displays seeds from the acidless Vaniglia and three Lima

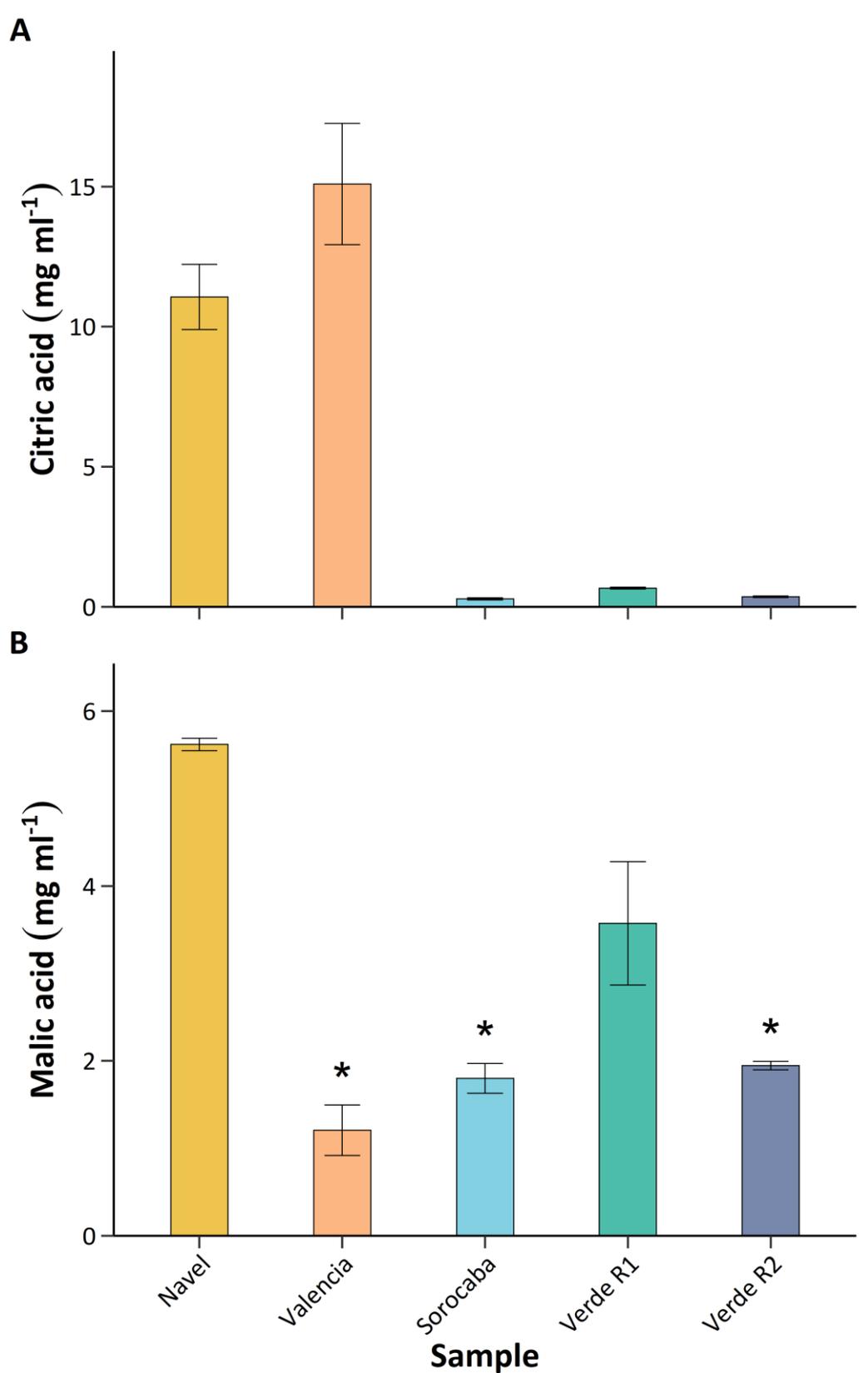


Figure 3.3.3 Organic acid concentration (A: citric acid, B: malic acid) in the juice of five *C. sinensis* varieties: Navel, Valencia, Sorocaba, Verde R1 and Verde R2. Asterisks indicate significant difference relative to the Navel control. Values and error bars presented represent the mean of 3 (2 for Navel) biological reps \pm se.

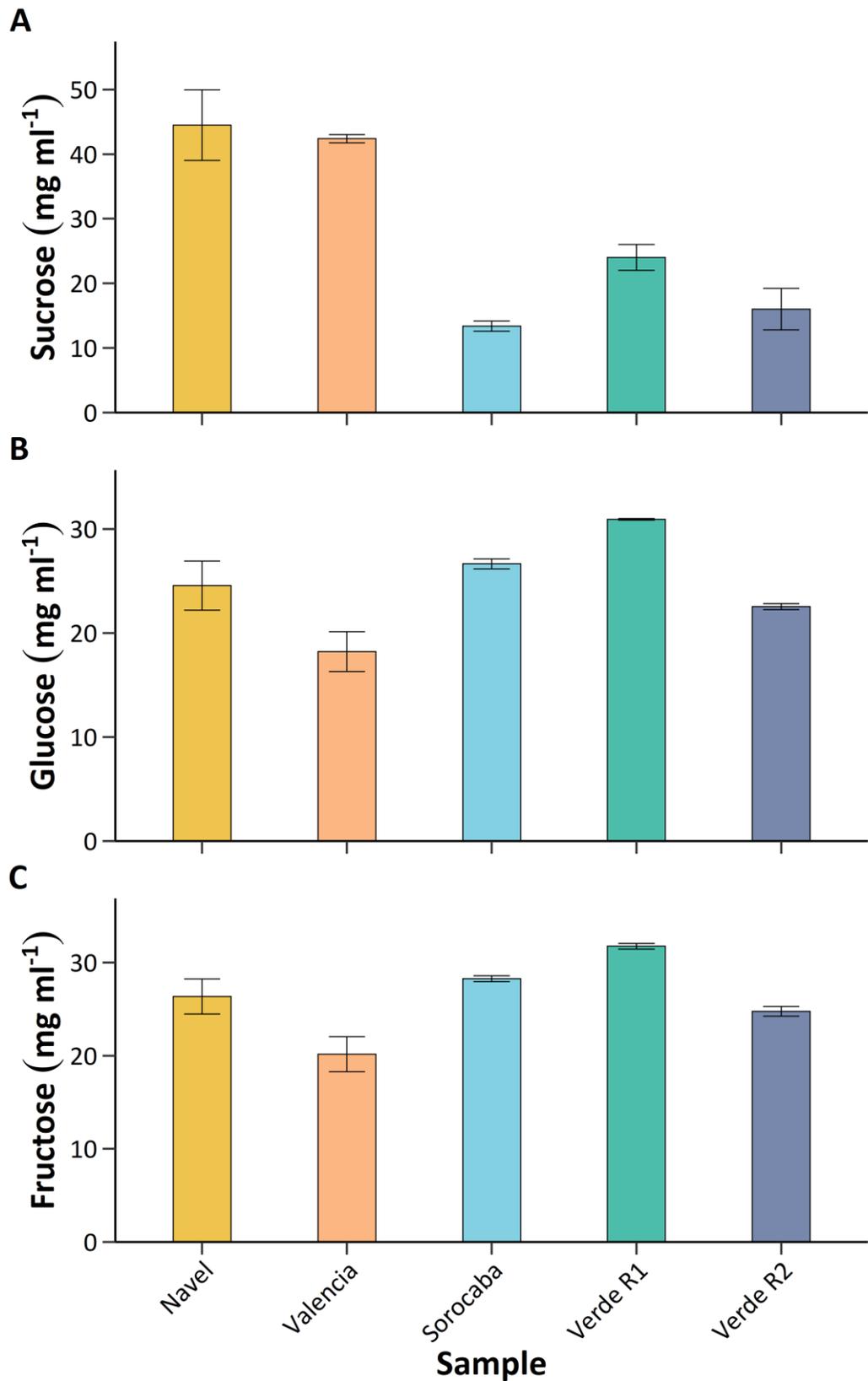


Figure 3.3.4 Soluble sugar concentration (A: sucrose, B: glucose, C: fructose) in the juice of five *C. sinensis* varieties: Navel, Valencia, Sorocaba, Verde R1 and Verde R2. Asterisks indicate significant difference relative to the Navel control. Values and error bars presented represent the mean of 3 (2 for Navel) biological reps \pm se.

varieties before and after DMACA staining. It was evident that, unlike Vaniglia, the Lima varieties all produced condensed tannins in the seed coat.

Despite producing PAs in seeds, PAs were not detectable in juice extracts from any variety (Figure 3.3.6) via colorimetric DMACA staining assays. Absorbance at 640 nm of (+)-catechin standards over time was plotted to assess the optimal time point to calculate a standard curve (Supplementary Figure 1). Absorbance at ~11 min offered the greatest goodness-of-fit R^2 value and lowest P -value for the standard curve linear regression model (Supplementary Figure 2). The lowest detectable concentration of PAs, as catechin equivalent, was 0.1 mM. PAs were only observed in the Valencia+ samples to which catechin had been added prior to extraction. A total recovery of 114 % was calculated.

HPLC quantification of catechin and epicatechin levels in *Citrus* juice were also performed (Figure 3.3.7). Similarly, free catechin and epicatechin monomers were not detected in any PA extract from sweet orange juice, except for catechins in Valencia+ (0.48 mg ml^{-1}). Total catechin recovery calculated from HPLC-based quantification was 64.5 %, in contrast to that calculated from colorimetric DMACA assay quantification.

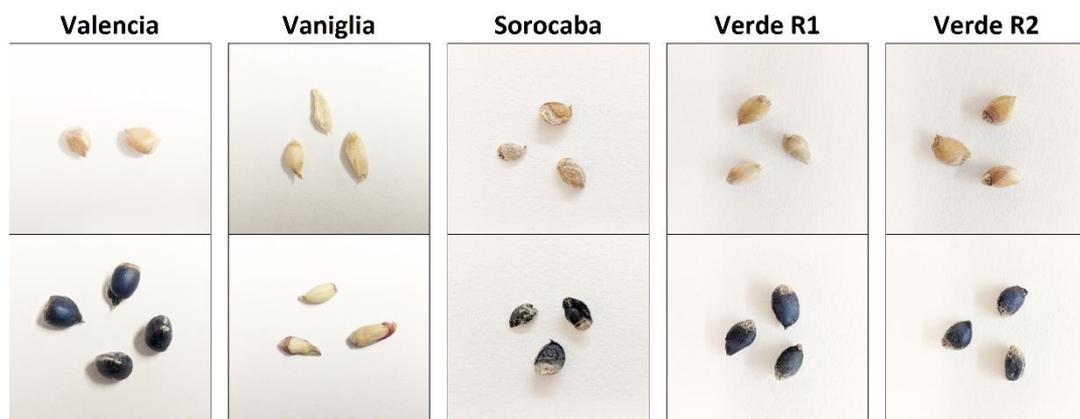


Figure 3.3.5 *C. sinensis* seeds before (top row) and after (bottom row) ≥ 10 min staining with 0.3 % DMACA reagent and 100 % ethanol washes from five varieties: Valencia, Vaniglia, Sorocaba, Verde R1 and Verde R2.

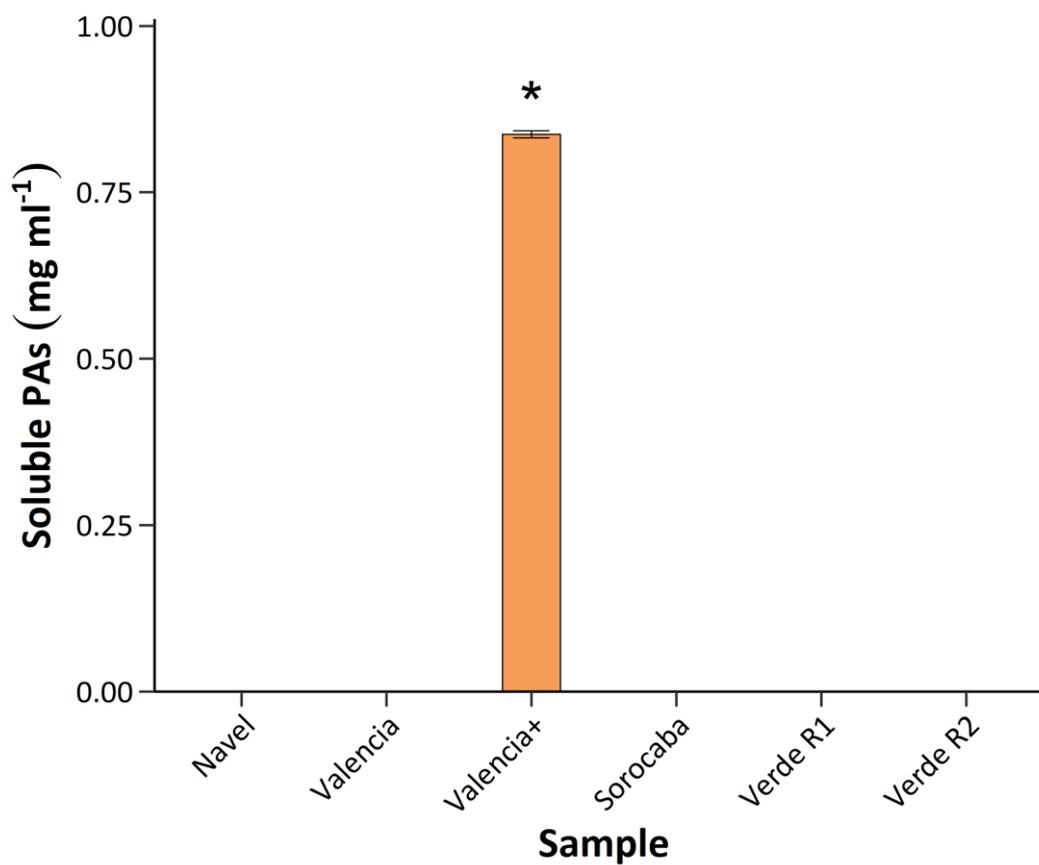


Figure 3.3.6 Soluble proanthocyanidin concentration in the juice of five *C. sinensis* varieties: Navel, Valencia, Sorocaba, Verde R1 and Verde R2. Catechin was added to Valencia+ samples prior to extraction to calculate recovery percentage. Asterisks indicate significant difference relative to the Navel control. Values and error bars presented represent the mean of 3 (2 for Navel) biological reps \pm se.

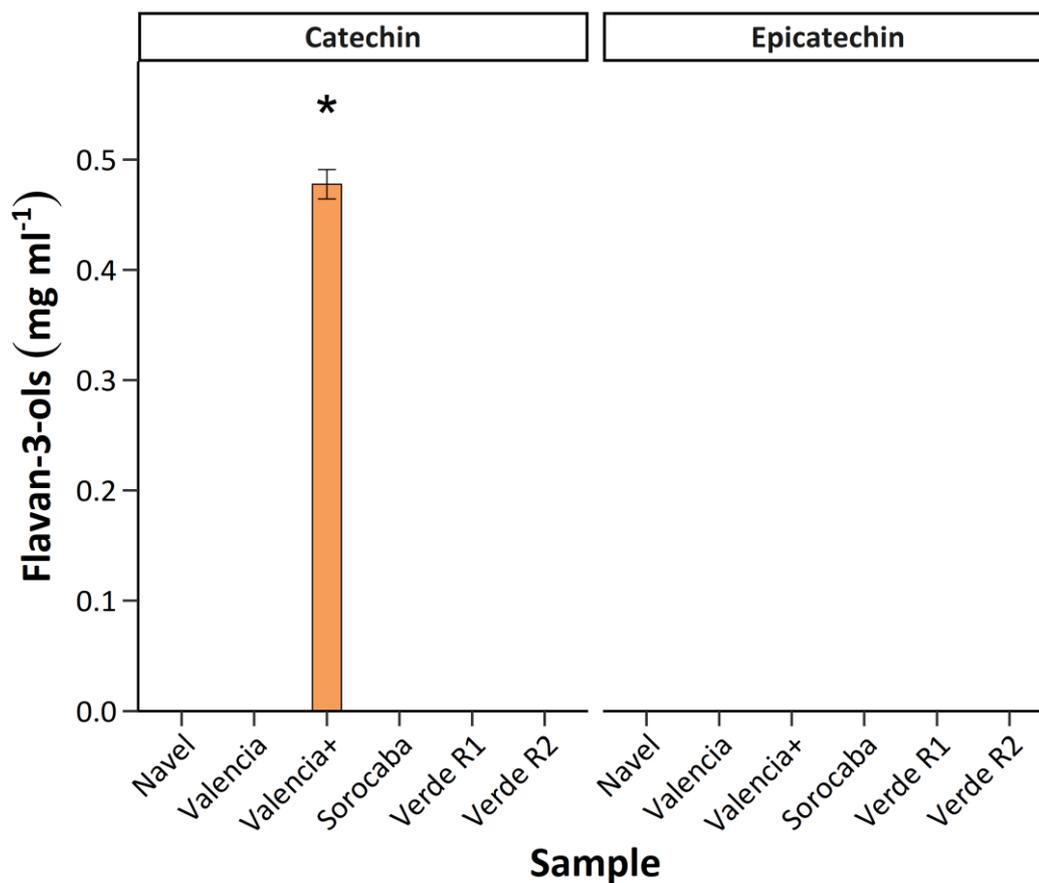


Figure 3.3.7 Free catechin and epicatechin monomer concentration in the juice of five *C. sinensis* varieties: Navel, Valencia, Sorocaba, Verde R1 and Verde R2. Catechin was added to Valencia+ samples prior to extraction to calculate recovery percentage. Asterisks indicate significant difference relative to the Navel control for the respective metabolite. Values and error bars presented represent the mean of 3 (2 for Navel) biological reps \pm se.

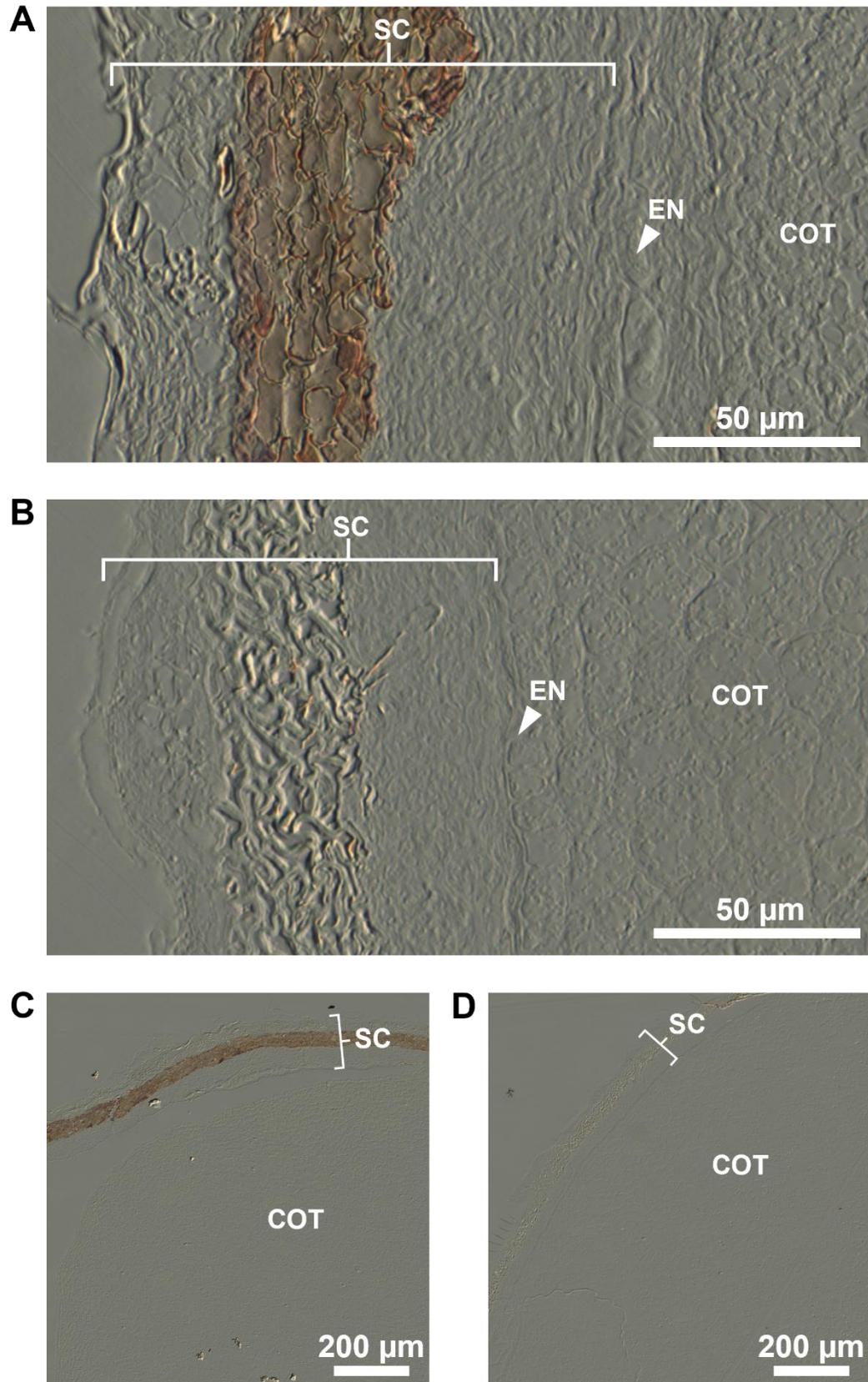


Figure 3.3.8 Cellular localisation of proanthocyanidins (PA) in sectioned Valencia (A, C) and Vaniglia (B, D) *C. sinensis* seeds. PAs in seeds were stained brown by reaction with 0.3 % DMACA reagent prior to LR white resin embedding and sectioning. Images were taken via DIC microscopy at 40x magnification. SC: seed coat, EN: endosperm, COT: cotyledon.

To investigate the cellular localisation of PAs in the coat of sweet orange seeds, seeds were stained with DMACA reagent, sectioned, and visualised at 40x magnification using DIC microscopy (Figure 3.3.8). Staining of PAs was visible exclusively within the seed coat in Navel seeds. Additionally, the images suggest PAs have bioaccumulated within the vacuoles of cells. No evidence of condensed tannins was observed in Vaniglia seeds under magnification in any seed tissues.

3.3.3 Genome-wide identification of *C. sinensis* R2R3MYB transcription factors

In total, 185 loci were identified containing at least one significant PFAM PF00249 domain within the *C. sinensis* genome. Of these, 68 contained two and were provisionally defined CsR2R3MYB1 – 68 in ascending order of their respective accession IDs (Supplementary Table 3). The R2R3MYB amino acid sequences range from 1,697 (CsR2R3MYB1) to 190 (CsR2R3MYB51) in length, averaging at 340. The R2 and R3 MYB domain protein sequences were aligned to analyse amino acid conservation within each domain. Each protein typically contained 48 and 46 residues in the R2 and R3 MYB repeat regions, respectively, and each domain was highly conserved Figure 3.3.9. Of the 51 and 59 amino acids in the R2 and R3

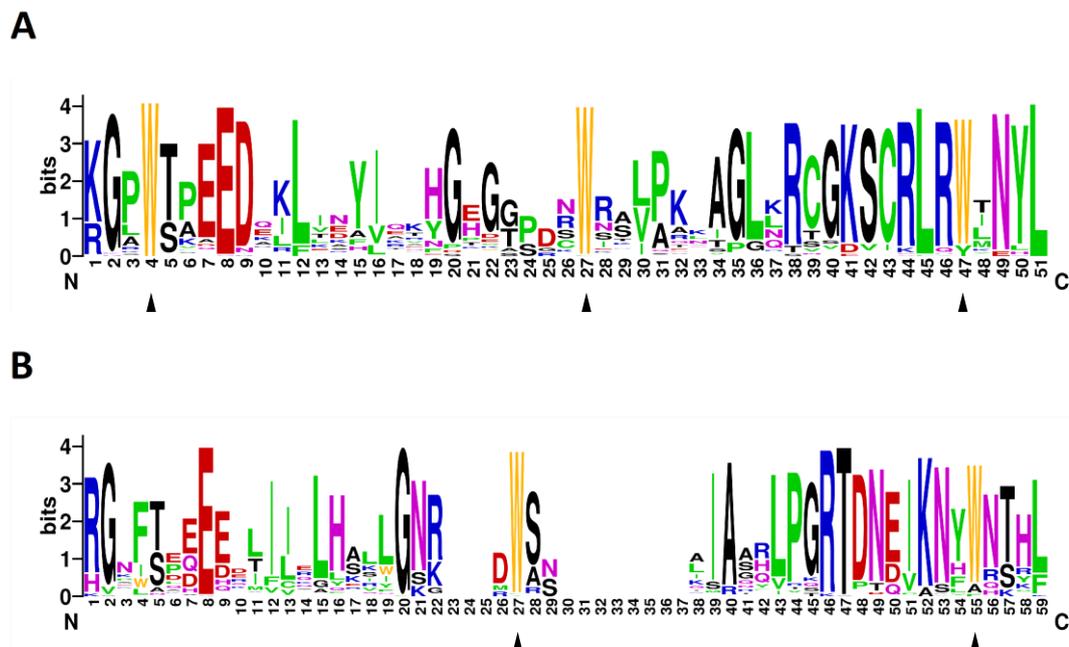


Figure 3.3.9 Amino acid conservation of R2 (A) and R3 (B) MYB repeats across all 68 putative R2R3MYBs identified in the *C. sinensis* genome. Asterisks highlight conserved tryptophan residues, typical of MYB transcription factors.

MYB domain MSAs, 22 and 16 were at least 80 % conserved, respectively. These included 2-3 tryptophan residues in both MYB domain repeats, which are typically characteristic of MYB proteins (Stracke et al., 2001, Ogata et al., 1992).

Multiple sequence alignments and phylogenetic analyses were also performed all 68 CsR2R3MYBs, 125 AtR2R3MYBs and 21 characterised R2R3MYB proteins from other plant species Figure 3.3.10. Potential molecular functions of *Citrus* R2R3MYBs were inferred based on their phylogenetic relationship with characterised proteins from these species. A ML phylogenetic tree (substitution model: Jones-Taylor-Thornton (JTT)+G+F, bootstrap replicates = 1000) was constructed from this alignment.

The R2R3MYBs were divided into 25 subgroups (S1-25) defined by previously established nomenclature and phylogenetic characterisation of the *Arabidopsis* R2R3MYB transcription factor family (Dubos et al., 2010, Kranz et al., 1998, Stracke et al., 2001). Of these subgroups 20 contained at least one CsR2R3MYBs. There were no CsR2R3MYBs in S10, 12, 21, 23, and 25. Conversely, an entire clade of *Citrus* R2R3MYBs only was observed, containing CsR2R3MYBs 2, 22, 23, 24, 26 and 31. S5 was divided into S5-a, containing many positive regulators of the PA pathway in *Arabidopsis*, grapevine, *Medicago*, *Malus* and peach (An et al., 2015, Baudry et al., 2004, Bogs et al., 2007, Gesell et al., 2014, Liu et al., 2014, Terrier et al., 2009, Xu et al., 2014, Zhou et al., 2015a), and S5-b, which includes *AtMYB5* homologs with minor roles in PA regulation that have more recently been shown to regulate hyperacidification in grapevine, *Petunia* and *Litchi* (Lai et al., 2019, Amato et al., 2019, Quattrocchio et al., 2006).

Notably, two CsR2R3MYBs (46 and 35) clustered with PhPH4, VvMYB5a and b, and AtMYB5 in S5-b. CsR2R3MYB46 shares the highest sequence identity with PhPH4, VvMYB5a and b. Therefore, CsR2R3MYB46 was designated as the primary MYB candidate involved in fruit hyperacidification and the encoding gene was provisionally named *Nicole*. Another two CsR2R3MYBs were situated with AtMYB123 homologs such as, VvMYBPA1 and 2, MtMYB14, and PpMYB7, key regulators of PA biosynthesis in S5-a. These were provisionally named Iris (CsR2R3MYB47) and Marys (CsR2R3MYB30). The amino acid sequences of Nicole, Iris and Marys were subjected to an MSA (ClustalO) with AtMYB123, AtMYB5 and homologs from

other species (Figure 3.3.11). It was evident that amino acid conservation was highest around the R2R3MYB domain.

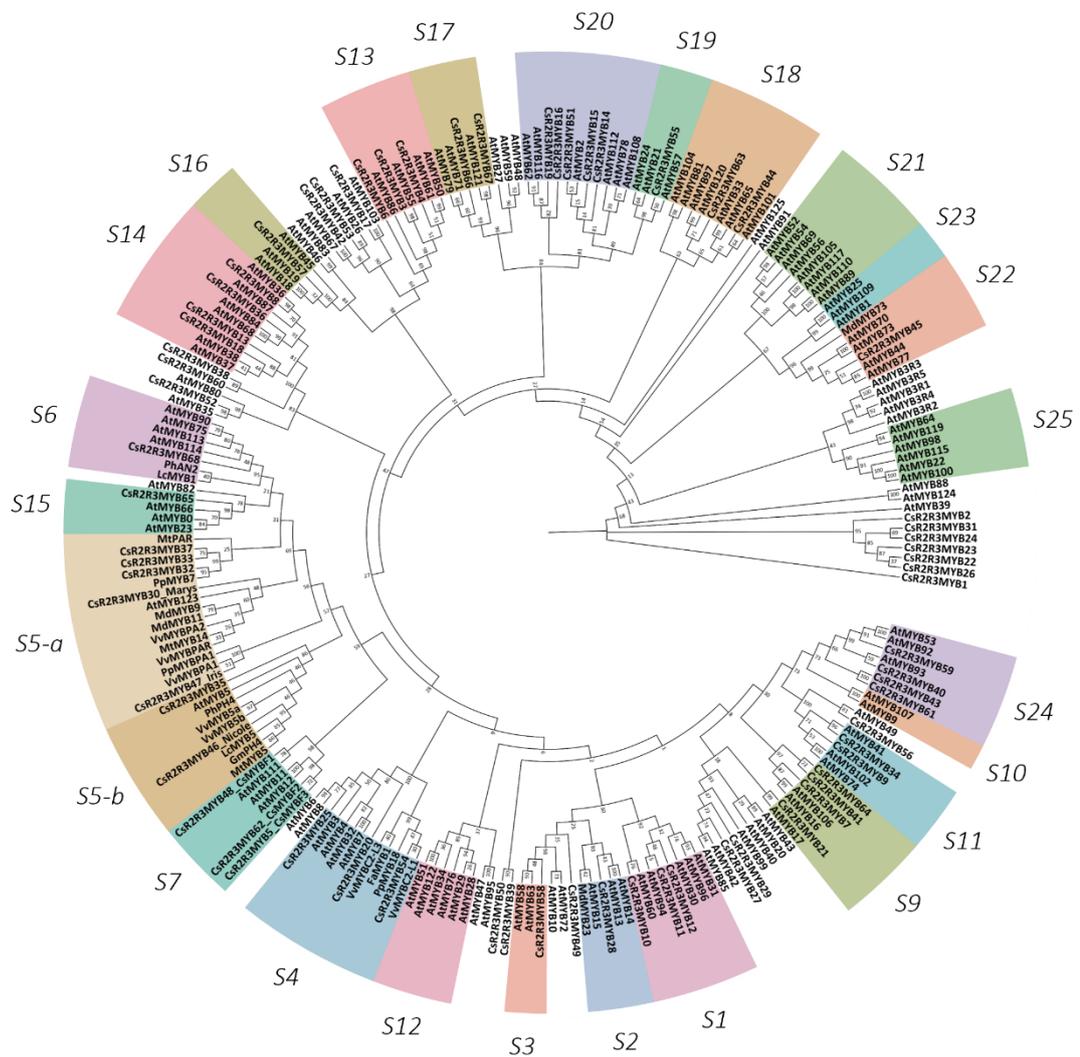


Figure 3.3.10 Maximum-likelihood amino acid phylogeny of 68 putative *CsR2R3MYBs*, 125 *Arabidopsis R2R3MYBs* and 21 characterised *R2R3MYBs* from other plant species. Subgroup (S1-25) designation conforms to the previously established phylogenetic characterisation of the *Arabidopsis R2R3MYB* transcription factor family. Substitution model: JTT+G+F, bootstrap replicates = 1000. Node values are bootstrap value percentages.

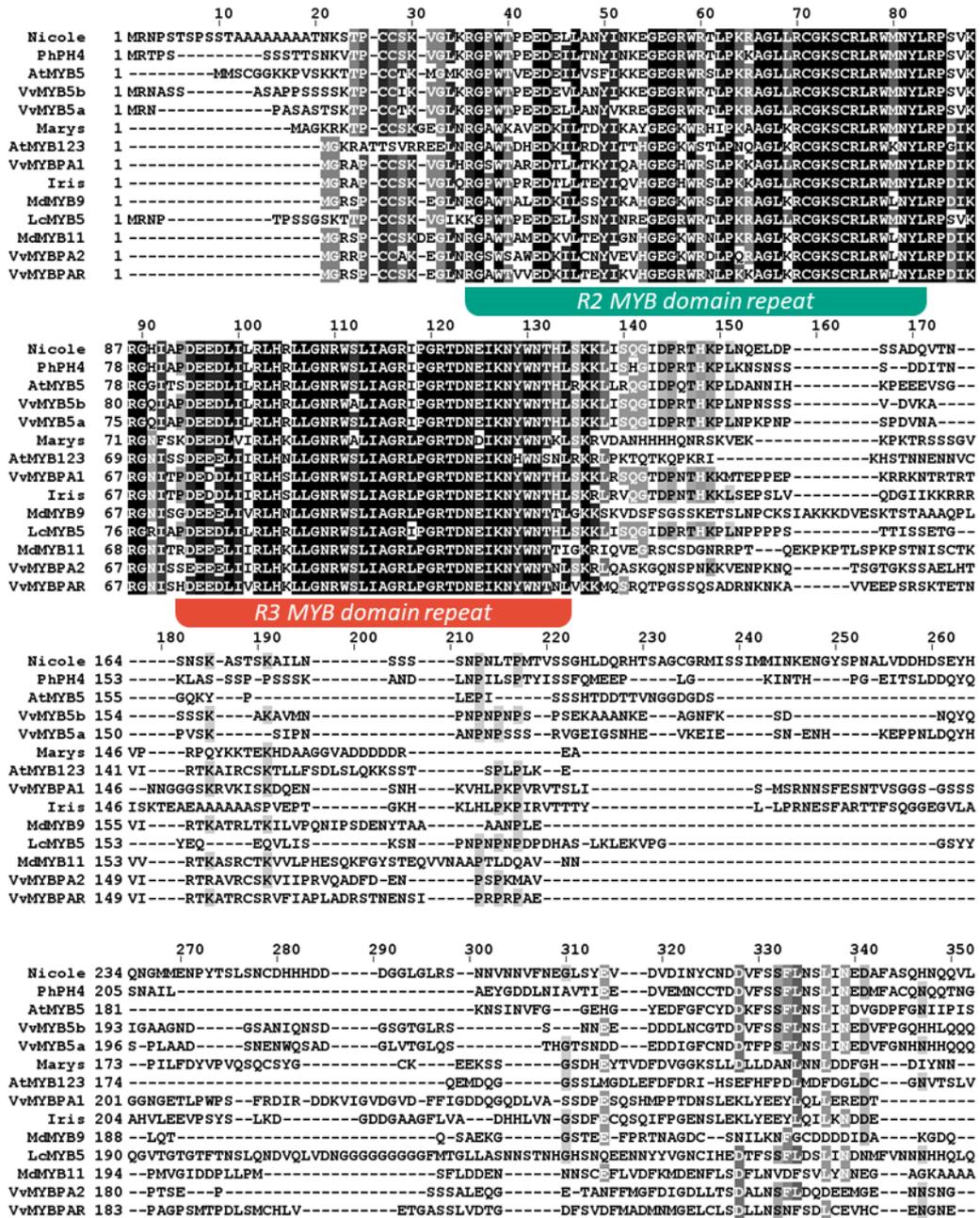


Figure 3.3.11 ClustalO amino acid alignment of *C. sinensis* R2R3MYBs structurally homologous to AtMYB123 and AtMYB5. Sequences were ordered by pairwise identity with Nicole. R2 and R3 MYB domain repeats are indicated by the green and red boxes below the sequence, respectively. Residue conservation is indicated by the black/grey background.

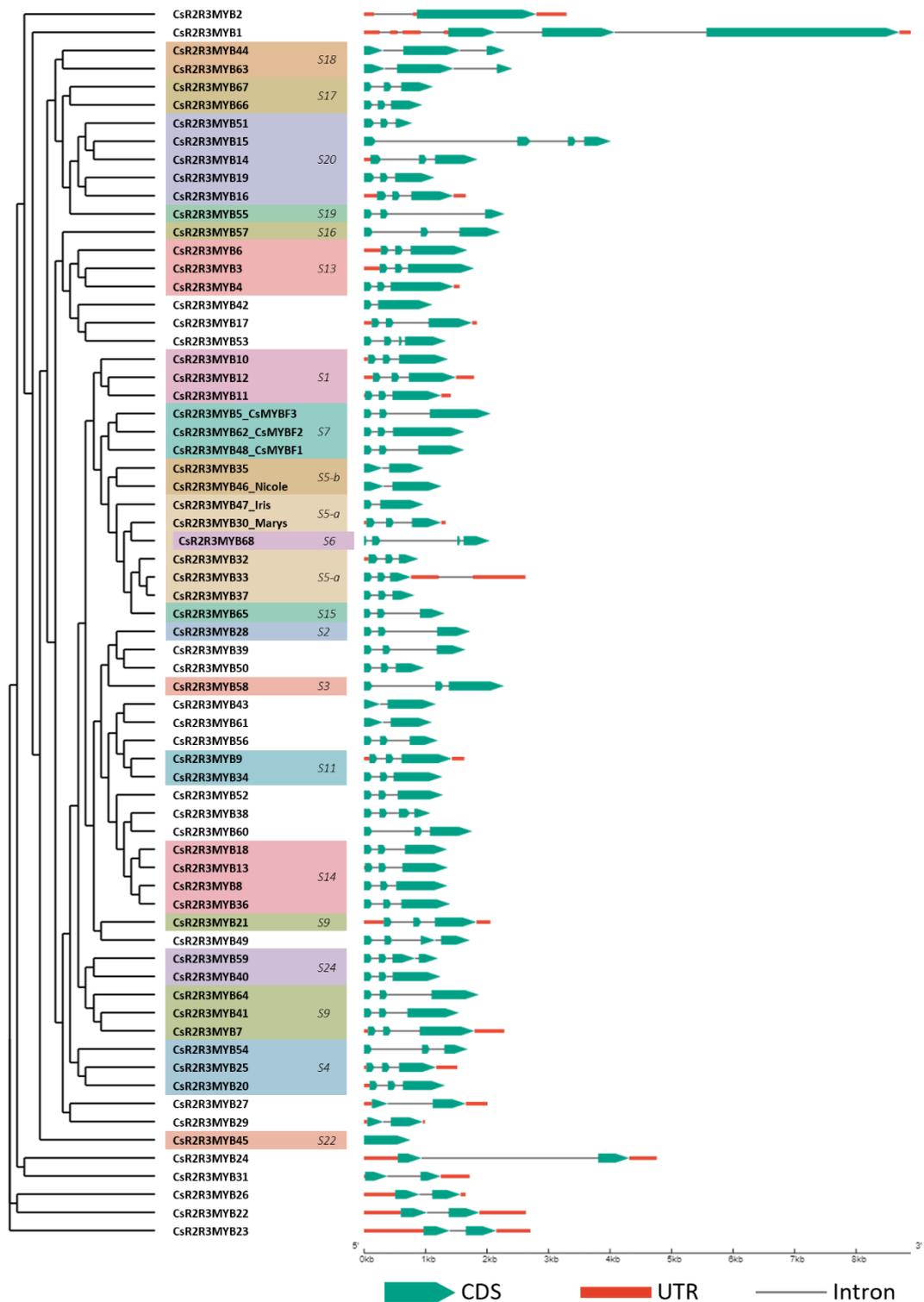


Figure 3.3.12 Maximum-likelihood amino acid phylogeny and gene structure of 68 putative *CsR2R3MYB*. Subgroups (S1-25) were designated according to the phylogenetic relationship between *CsR2R3MYBs* and *Arabidopsis R2R3MYBs* reported earlier. Substitution model: JTT+G+F, bootstrap replicates = 1000. Green arrows, red bars and grey lines represent CDS, UTR and introns, respectively.

Notably, an additional MYB candidate was identified via BLAST searching the *C. sinensis* genome with the AtMYB123 protein as the query sequence. A neighbour-joining (NJ) phylogenetic tree was constructed using a ClustalO MSA of all *Arabidopsis* R2R3MYBs and candidate MYBs from *C. sinensis* with high sequence homology with AtMYB123 (Supplementary Figure 3). This additional candidate identified via BLAST (orange1.1g025602m.g), and provisionally named Miriam, was observed clustered with AtMYB123. However, the predicted Miriam amino acid sequence did not conform to the candidate R2R3MYB requirements, detailed previously, as it does not contain an R2 MYB domain repeat. Furthermore, the gene was cloned and sequenced from Navel orange and *C. clementina* by Dr Butelli, which revealed the sequence was misannotated in the Phytozome genome. The sequenced gene contained a start codon mutation (ATG to ATA). For these reasons, Miriam was not considered as a potential PA regulator for the remainder of this thesis.

To examine the conservation of gene structure within the CsR2R3MYB transcription family I constructed an additional ML phylogenetic tree (substitution model: Jones-Taylor-Thornton (JTT)+G+F, bootstrap replicates = 1000) with all 68 members only alongside their intron-exon sequence (Figure 3.3.12). In general, the CsR2R3MYBs clustered in accordance with their designated subgroups, as defined in the ML phylogeny with the *Arabidopsis* R2R3MYB protein family. Likewise, gene structure within subgroups was relatively conserved, particularly regarding exon number and size. The gene length ranged greatly from 758 – 8,890 bp, with a median of 1,590.5 bp. The median CDS length was 570 bp, commonly comprising 3 exons (~70 % of CsR2R3MYBs) ranging from 42 – 1,940 bp. The most downstream exon was often the largest within each gene. Only two genes contained 1 exon (CsR2R3MYB2 and 45) whereas 6 genes contained the highest number observed, 4. Notably, the subgroups 5, 6 and 7, all of which comprise *Arabidopsis* R2R3MYBs that positively regulate various branches of the flavonoid pathway, were all clustered together.

3.3.4 Genome-wide identification of *C. sinensis* MATE transporters

A total of 52 loci in the *C. sinensis* genome were identified as encoding polypeptides containing PFAM MATE domains (PF01554). These protein sequences were filtered by applying typical plant MATE protein constraints, such as a minimum length of 400 AAs and 8-12 transmembrane domains (Brown et al., 1999, Xu et al., 2019, Zhang et al., 2021, Li et

al., 2002, Ali et al., 2021, Liu et al., 2016d, Santos et al., 2017, Li et al., 2019b). Only 35 of the 52 candidates satisfied these criteria, and were provisionally named CsMATE1 – 35 in ascending order of the respective accession ID.

A multiple sequence alignment of CsMATEs, 20 characterised MATE transporters belonging to other species and 56 *A. thaliana* MATE proteins was performed to infer their endogenous function. These representative MATEs include those reported to elicit Al detoxification via citrate excretion, the transport of PA precursors and acylated-anthocyanins, nicotine transport, iron homeostasis and hypocotyl elongation (Marinova et al., 2007, Gomez et al., 2009, Diener et al., 2001, Li et al., 2002, Wang et al., 2015, Serrano et al., 2013, Durrett et al., 2007, Liu et al., 2009, Maron et al., 2013, Tovkach et al., 2013, Yokosho et al., 2011, Yokosho et al., 2009, Rogers et al., 2009, Sawaki et al., 2013, Wu et al., 2014b, Morita et al., 2009, Zhao et al., 2011, Chai et al., 2009, Thompson et al., 2010, Shoji et al., 2009, Mathews et al., 2003, Zhao and Dixon, 2009, Frank et al., 2011, Pérez-Díaz et al., 2014). Figure 3.3.13 presents the ML phylogenetic relationship between these peptide sequences (substitution model: Le-Gascuel (LG)+G, bootstrap replicates = 1000).

The MATEs were divided into 5 clades (C1-5) defined by the phylogenetic characterisation of the *Arabidopsis* MATE family and the associated nomenclature (Li et al., 2002). Subclades for C1, 2 and 4 were defined in consideration of their large size and protein function of characterised MATEs within. For example, just over 40% of MATEs within the phylogeny were confined to C2 which includes MATEs responsible for transportation of different metabolites. All clades contained at least one CsMATE, including C5, to which only one AtMATE is a member. MATEs clustered within C2-a are all putative AtTT12 homologs shown to transport PA precursors into the vacuole (Chai et al., 2009, Frank et al., 2011, Marinova et al., 2007, Pérez-Díaz et al., 2014, Zhao and Dixon, 2009). Only one *C. sinensis* protein was located within this subclade, CsMATE34, and was provisionally named CsTT12. CsMATE27 and CsMATE31 were found closely related to anthocyanin transporters such as VvAM1 and 3, and MtMATE2 in C2-c (Gomez et al., 2009, Mathews et al., 2003, Thompson et al., 2010, Zhao et al., 2011). C3 contains many citrate exporters required for metal ion (Fe and Al) detoxification and a salicylic acid transporter associated with disease resistance (Durrett et al., 2007, Liu et al., 2009, Maron et al., 2013, Rogers et al., 2009, Sawaki et al., 2013, Serrano et al., 2013, Tovkach et al., 2013, Wu et al., 2014b, Yokosho et al., 2011, Yokosho et al., 2009). There were

two closely related CsMATEs clustered with these citrate exporters (CsMATE10 and 15). No clusters of only CsMATEs were observed distantly related to MATEs from other species.

To examine the conservation of gene structure within the CsMATE family I constructed an additional ML phylogenetic tree (substitution model: LG+G+F, bootstrap replicates = 1000) with all 35 members only alongside their intron-exon sequence (Figure 3.3.14). The CsMATEs

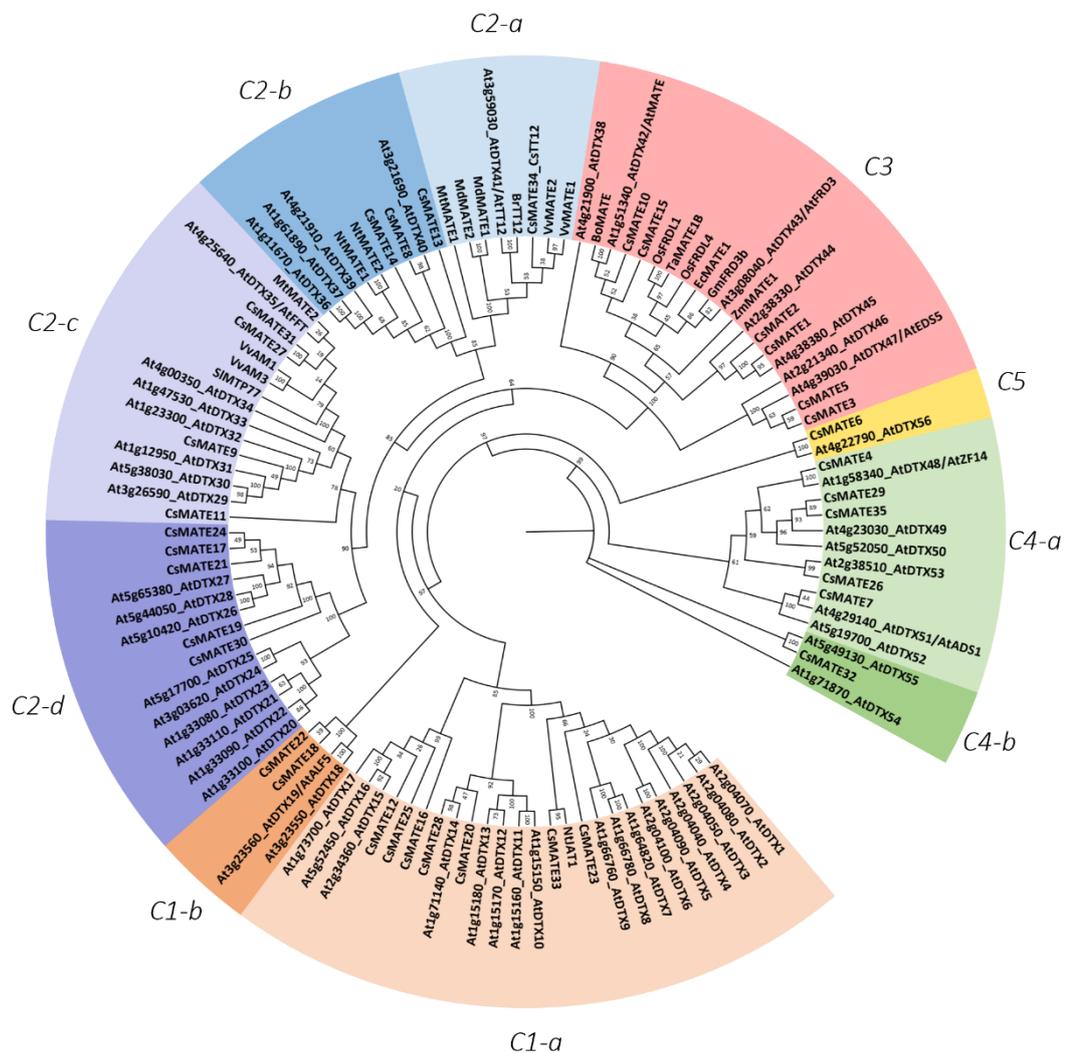


Figure 3.3.13 Maximum-likelihood amino acid phylogeny of 35 putative CsMATEs, 56 Arabidopsis MATEs and 20 characterised MATEs from other plant species. Clade (C1-5) designation conforms to the previously established phylogenetic characterisation of the Arabidopsis R2R3MYB transcription factor family. Substitution model: LG+G, bootstrap replicates = 1000. Node values are bootstrap value percentages.

clustered well, conforming to their clade and subclade designations in the previous phylogeny. Gene structure within clades was also relatively conserved. CsMATE members of C3 all contain 13 or 14 exons, double the median average of 7 for the protein family. The lowest exon number (1) was observed in C4 and C5. These clades also have the longest exons within the CsMATE family of around 1.5 kb. C1 and 2 all have between 5 – 8 exons. CDS length had a median average of 1,500 bp and varied little. Conversely, gene length ranged between 1,541 – 6,290 bp.

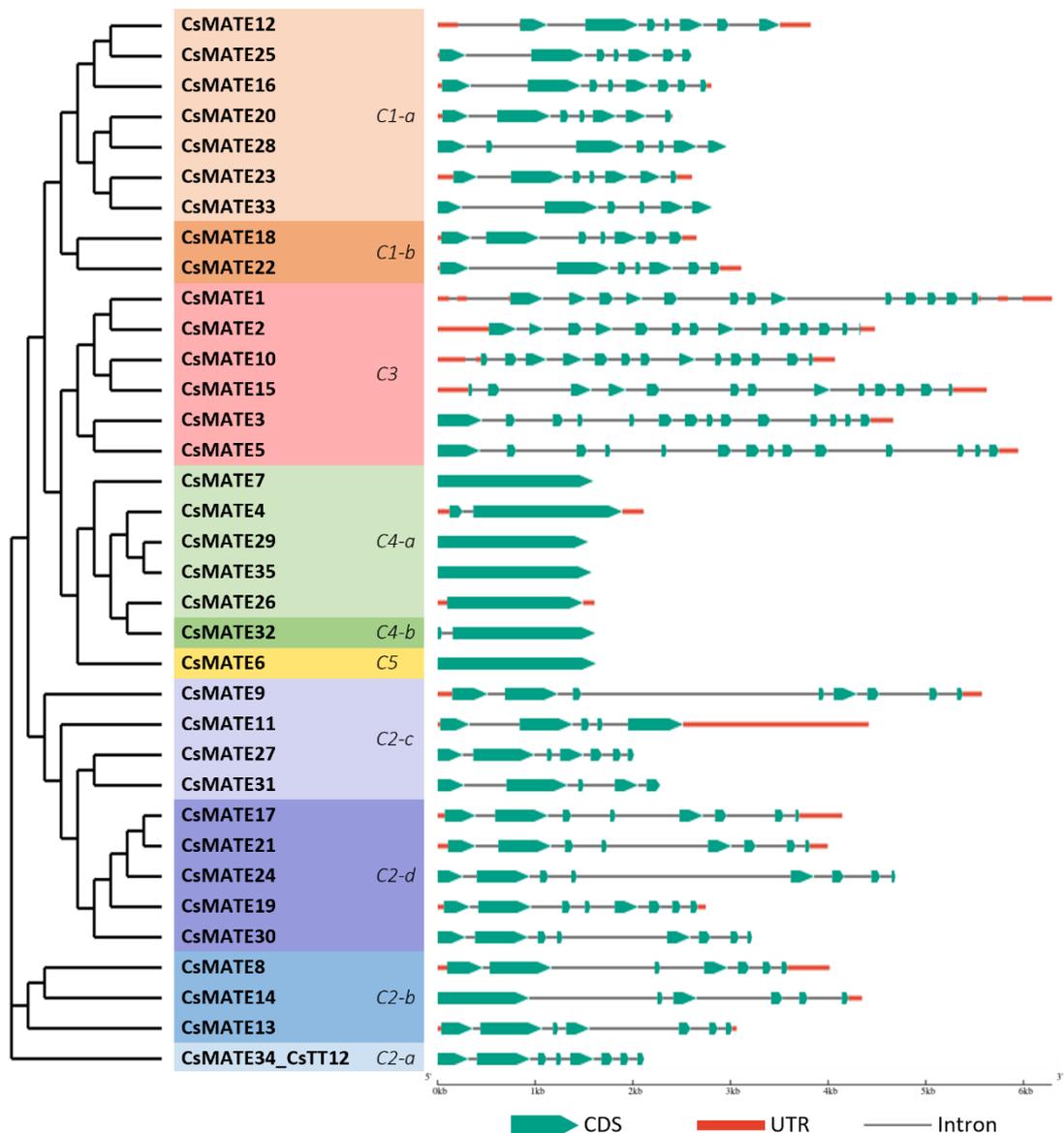


Figure 3.3.14 Maximum-likelihood amino acid phylogeny and gene structure of 35 putative CsMATEs. Clades (C1-5) were designated according to the phylogenetic relationship between CsMATEs and *Arabidopsis* MATEs reported earlier. Substitution model: LG+G+F, bootstrap replicates = 1000. Green arrows, red bars and grey lines represent CDS, UTR and introns, respectively.

3.3.5 Genotypic analyses of acidic and acidless *C. sinensis* varieties

Identification of genes of interest within the Phytozome *C. sinensis* genome was achieved via a combination of BLAST alignments with characterised proteins from other species and the phylogeny data detailed above (for *CsR2R3MYBs* and *CsMATEs* only), and then compared with data from previous publications (Supplementary Table 3). Of note, 8 *PhPH5*-like genes were identified and named *CsPH1-8* by Shi et al. (2015). Only one was strongly expressed in fruit (*CsPH8*). This gene corresponds to the *CsPH5* candidate I identified from the Phytozome genome and will continue to be referred to as *CsPH5* for two reasons. Firstly, to keep consistency with nomenclature of *AtAHA10* homologs in *Petunia* and grapevine (*PhPH5* and *VvPH5*), and secondly, if the *CsPH1-8* name scheme was adopted there may be confusion with the *C. sinensis* homolog of *PhPH1* and *VvPH1*, named *CsPH1*. All other gene names conform to previously published nomenclature.

Several key genes of interest were cloned from the acidless Lima varieties via PCR amplification and sequenced to identify possible mutations. Regarding gene size, all genes amplified from Navel, Sorocaba, Verde R1 and Verde R2 DNA (*Noemi*, *Iris*, *CsTTG1*, *CsPH3*, *CsPH1*, *CsPH5* and *CsTT12*) were identical in size except for *Nicole* (Figure 3.3.15a; Supplementary Figure 4-10). Sequencing of the gDNA PCR amplification of *Nicole*, isolated from acidless mutants Sorocaba, Verde R1 and Verde R2, revealed they shared a mutant allele containing a long terminal repeat (LTR)-retrotransposon insertion in exon 2, which introduced an early stop codon (Figure 3.3.15b; Supplementary Table 5). Identical LTRs 549 bp in size were located at the 5' and 3' ends of the insertion. Open reading frames (ORF) were predicted with the FGENESH online tool and the predicted amino acid sequence was submitted to the NCBI Conserved Domain search tool (Solovyev et al., 2006, Lu et al., 2020). A significant Ty1/copia-type *gag-pol* domain hit was found (PF14223). The LTR-retrotransposon was named *Tcs7x*, and the resulting truncated mutant allele named *nicole^{soro}*. The insertion disrupted the coding sequence of exon 2 at the 3' end of the gene, preventing transcription of the last 195 bp due to the introduction of an early stop codon. The predicted translated protein sequence of *nicole^{soro}* is truncated, containing 326 amino acids in comparison to 375 in the WT allele (Figure 3.3.15c). The remainder of the amino acid sequence, including the R2R3MYB domain, remained intact except for the 13 amino acid C-terminal region encoded by the LTR insertion prior to the stop codon.

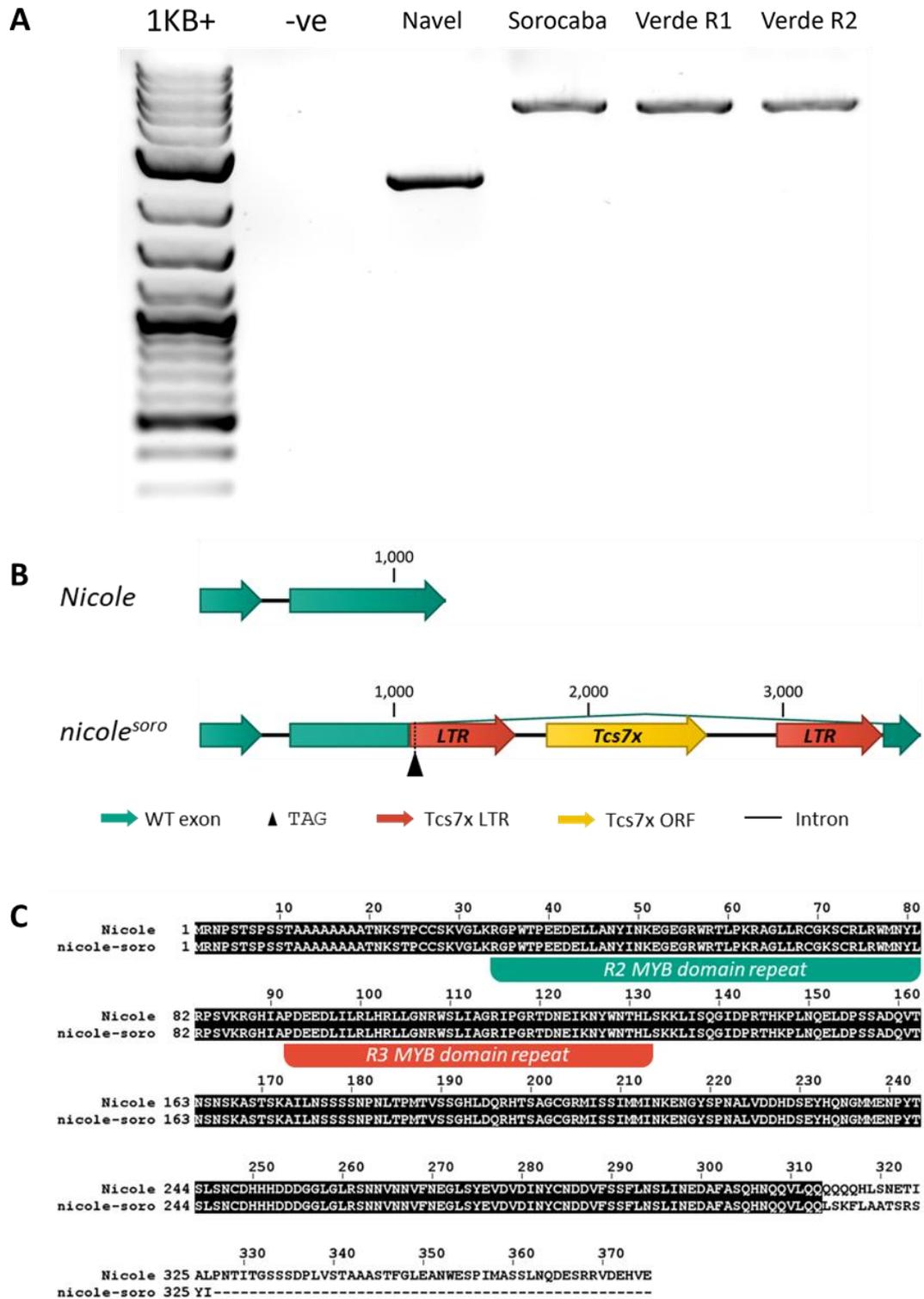


Figure 3.3.15 Genetic characterisation of the mutant *nicole^{SORO}* allele. A: PCR amplification of *Nicole* from gDNA extracted from *C. sinensis* juice. 1KB+: 1KB plus ladder (NEB); -ve: no DNA template. B: Gene structure of *Nicole* and *nicole^{SORO}*. Green arrows and grey lines indicate WT *Nicole* exons and introns, respectively. Red and yellow arrows indicate the LTR and ORF of the LTR-retrotransposon *Tcs7x*, respectively. The black arrow indicates the location of the introduced early stop codon. C: Amino acid sequence of *Nicole* and *nicole^{SORO}*. R2 and R3 MYB domain repeats are indicated by the green and red boxes below the sequence, respectively. Residue conservation is indicated by the black background.

3.3.6 Transcriptomic analyses of acidic and acidless *C. sinensis* varieties

The transcriptome of both *noemi* and *nicole* sweet orange mutants was analysed to identify candidate target genes of *Nicole*. RNA sequencing was performed on Navel (WT), Vaniglia (non-functional *noemi* mutant; Butelli et al., 2019) and the 3 Lima *nicole* mutants. Expression data for primary genes of interest were validated by RT-qPCR. Dr Licciardello from the CREA-OFA institute provided RNA from one Navel and one Vaniglia fruit. Due to the availability of Navel fruits, I extracted RNA from Navel oranges purchased from a local supermarket to increase the number of biological replicates to three. My transcriptomic analysis and RT-qPCR validation was compared between all varieties and Navel, the control sample. Novogene undertook statistical analyses of these data. Genes are considered expressed if the number of fragments per kilobase of exon per million mapped (FPKM) was greater than 1. A strong Pearson correlation between biological replicates within sample groups was observed (Supplementary Figure 11).

I identified 2,300 down regulated genes (P -value < 0.05) in the *noemi* mutant Vaniglia, in comparison to Navel (Figure 3.3.16), of which genes encoding proteins involved in organic acid metabolic biological processes were significantly enriched (Supplementary Figure 12). The *nicole* mutants Sorocaba, Verde R1 and Verde R2 have reduced expression of 2,409, 2,935 and 3,089 genes relative to Navel (Figure 3.3.17, Figure 3.3.18, Figure 3.3.19). Organic acid metabolic processes (GO:0006082) were statistically over-represented in GO enrichment analyses of genes downregulated in all acidless varieties (Supplementary Figure 13-15; GO:0006082 not shown in Supplementary Figure 15 as only the top 20 most significantly enriched GO terms were plotted in the interest of space). A significant overlap between downregulated genes was also apparent. A total of 1,151 genes were common amongst all Lima varieties, 636 of which are also shared by Vaniglia (Figure 3.3.20).

Differential gene expression was explored by plotting expression levels (FPKM) of various genes of interest. Genes in these plots are named in a particular format to provide an indication of possible function if their structural homology had not been manually investigated. The *C. sinensis* Phytozome genome provides additional annotation information, including the following: most significant *Arabidopsis* BLAST hit, GO terms and PFAM domains. In FPKM plots gene accession IDs are followed by my designation (prefixed by “Cs”, with the exception of Noemi, Nicole, Iris, Marys), if available (e.g. orange1.1g037798m.g; Noemi). If a

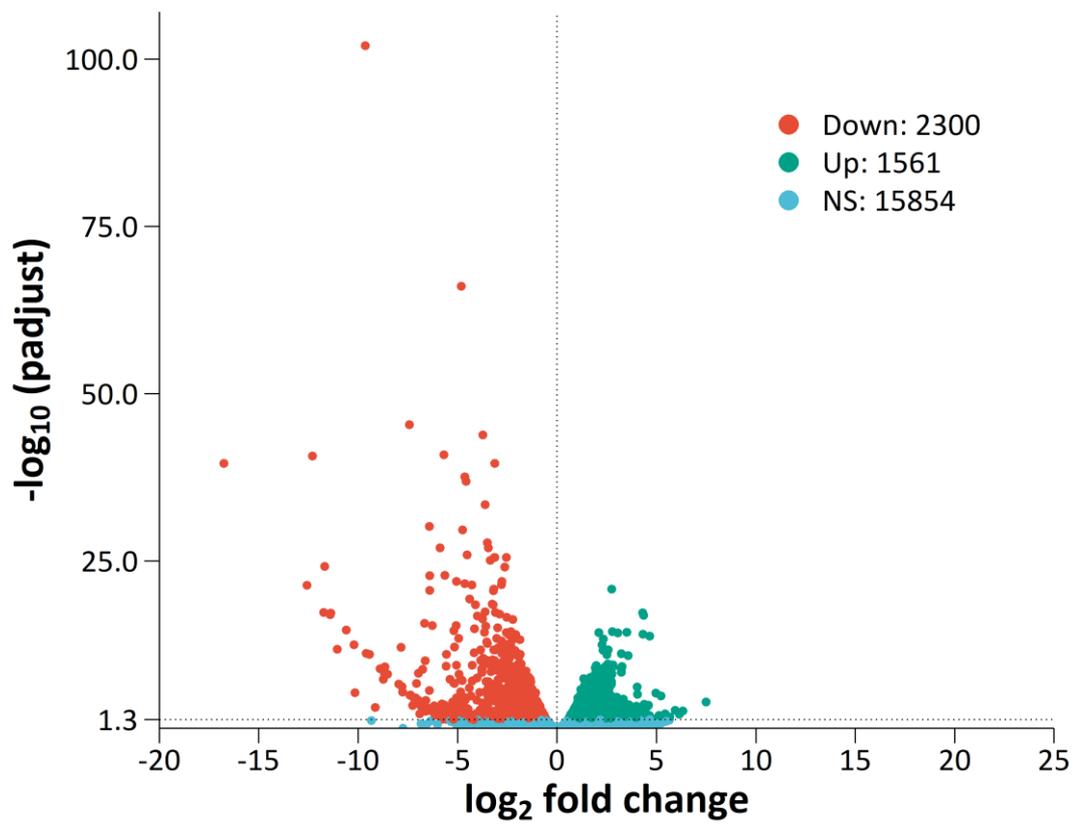


Figure 3.3.16 Differentially expressed gene counts in Vaniglia, relative to Navel.

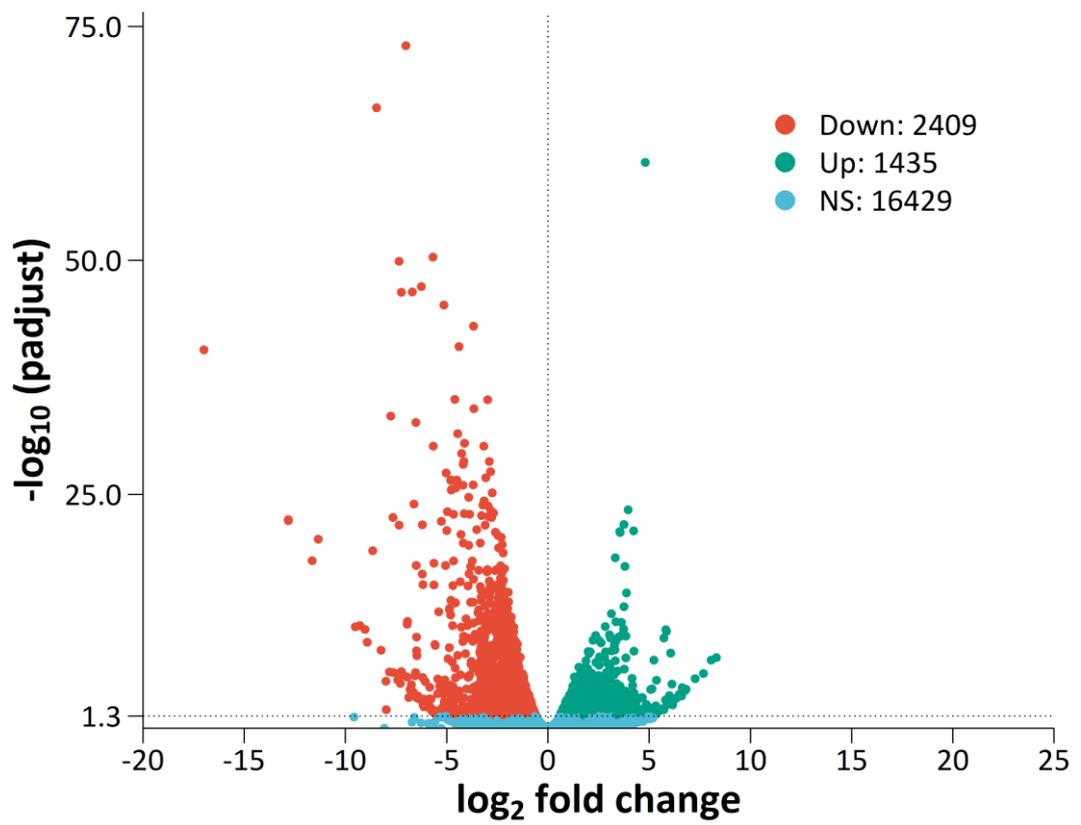


Figure 3.3.17 Differentially expressed gene counts in Sorocaba, relative to Navel.

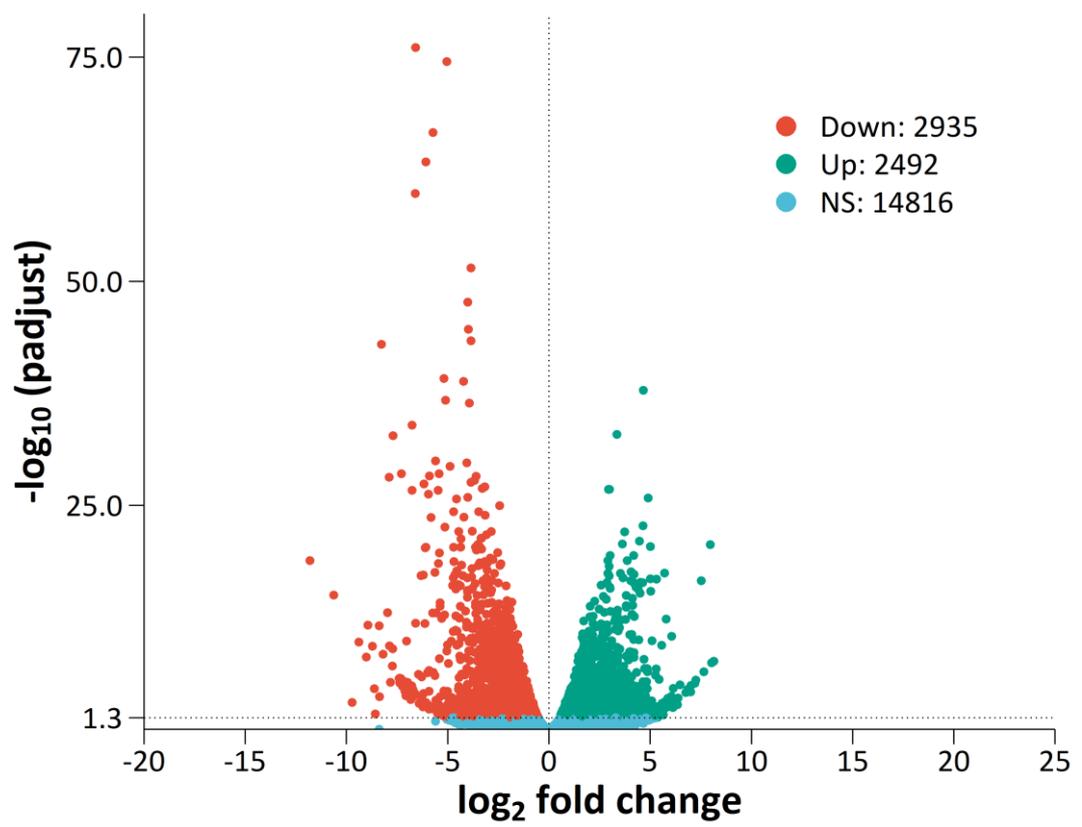


Figure 3.3.18 Differentially expressed gene counts in Verde R1, relative to Navel.

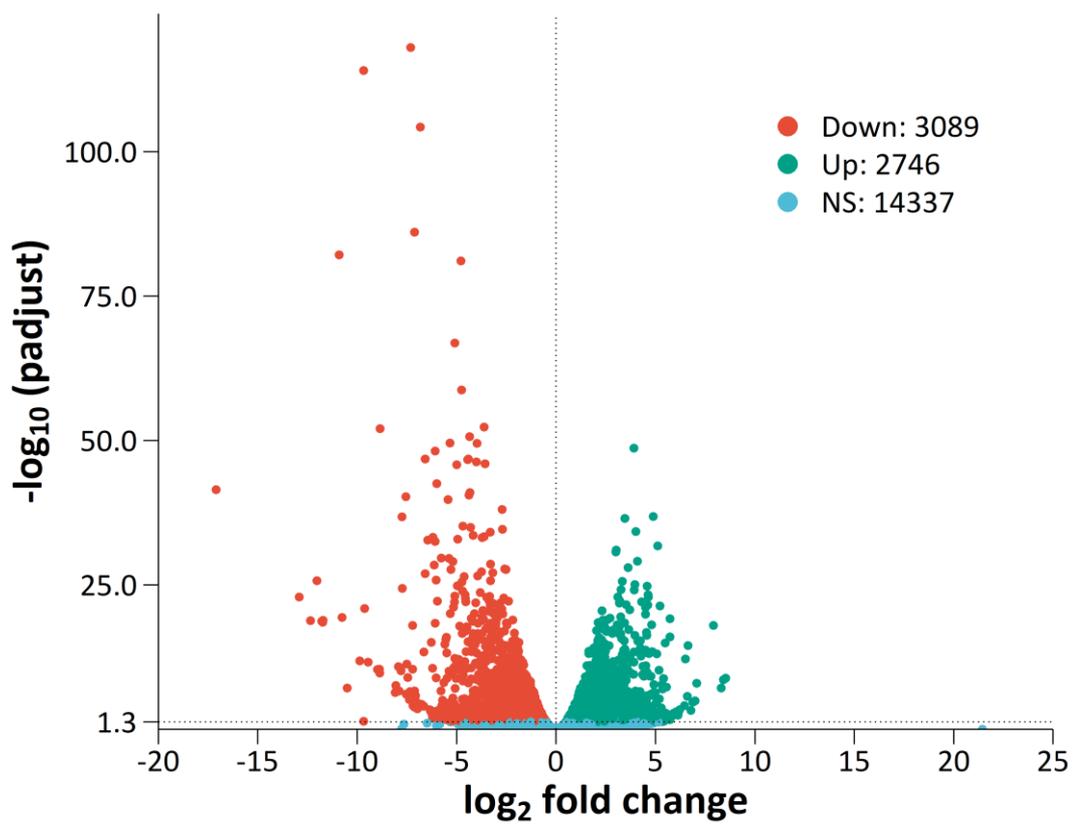


Figure 3.3.19 Differentially expressed gene counts in Verde R2, relative to Navel.

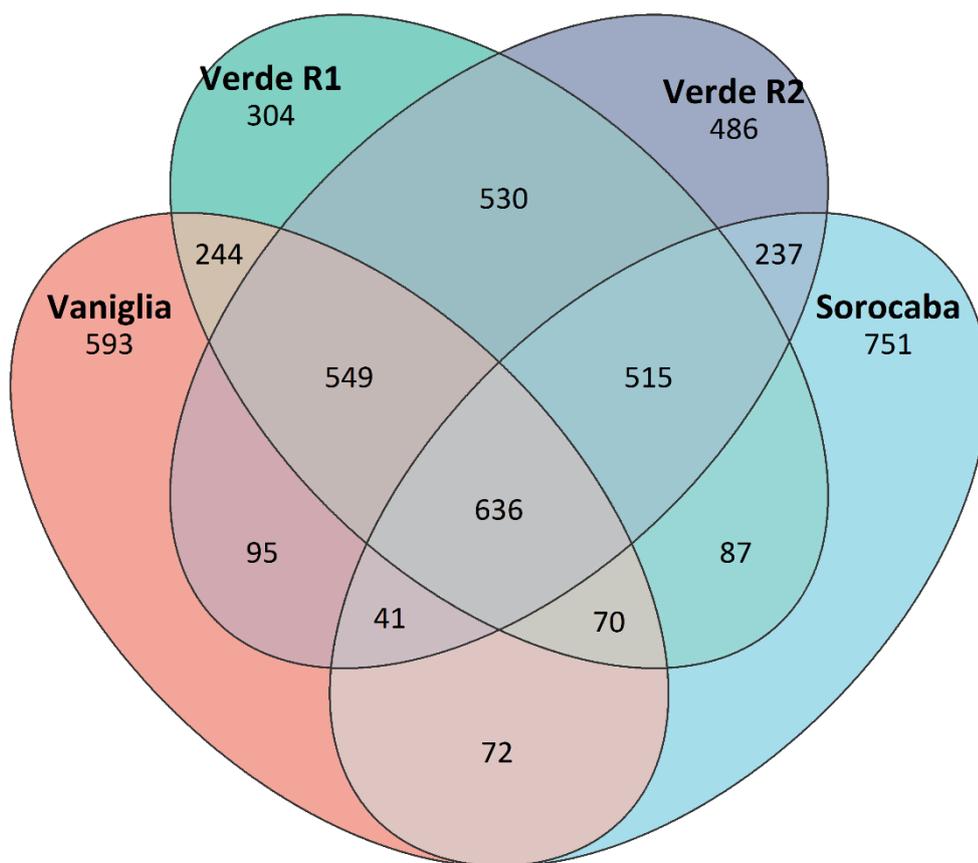


Figure 3.3.20 Significantly downregulated gene counts in Vaniglia, Sorocaba, Verde R1 and Verde R2 *C. sinensis* in comparison to Navel.

gene has not been manually investigated for structural homology and subsequently named, the first available Phytozome annotation field described earlier, if any and in the order stated, follows the Phytozome accession ID (e.g., orange1.1g041277m.g; ABC transporter family).

The top 15 most significantly downregulated genes (average of all acidless mutants), relative to Navel, are presented in (Figure 3.3.21). *CsPH5* was the most downregulated gene and expression was completely abolished in both *noemi* and *nicole* mutants. Another *AtAHA10*-like gene was in this list (orange1.1g037174m.g). The predicted protein sequence appears to be a truncated *CsPH5*-like protein, containing only 301 amino acids with very high structural homology with the C-terminal region of *CsPH5* (94 % identity), in comparison to the 883 amino acids in the *CsPH5* sequence. *CsLDOX* is also in the top 15 most significantly downregulated gene subset, with a lack of expression in all acidless mutants. *CsLDOX* is a structural gene of the flavonoid biosynthetic pathway, catalysing the conversion of leucoanthocyanidins to anthocyanidins which are required for the generation of epicatechin precursors of condensed tannins. The remaining genes are all uncharacterised. A few have significant BLAST hits against the *Arabidopsis* genome, such as orange1.1g043774m.g (*AtRD21*) and orange1.1g028699m.g (*AtLSH1*). *AtRD21* has functions for drought-induced resistance to *Pseudomonas*, and the *N. benthamiana* homolog also has roles in plant immunity (Bozkurt et al., 2011, Liu et al., 2020). *AtLSH1* is a potential regulator of hypocotyl elongation (Lee et al., 2020).

Gene expression of *Noemi*, R2R3MYBs of interest, the WDR-encoding *CsTTG1*, the WRKY-encoding *CsPH3*, probable acidity-related genes and PA pathway genes are presented in Figure 3.3.22. These genes were identified as described previously, primarily using amino acid sequences of genes from *Arabidopsis*, *Petunia*, and grapevine in BLAST alignments with the Phytozome *C. sinensis* genome. Putative homologs were named according to the typical nomenclature of homologous genes in literature. Of note, two anthocyanin reductase-encoding genes were identified and named *CsANR1* and *CsANR2*. *Nicole* is expressed in all varieties except for Vaniglia, the *noemi* mutant. Conversely, while significantly downregulated, *Noemi* was still expressed in the *nicole* mutants, suggestive of partial transcriptional induction by *Nicole*. No expression of *Iris* or *Marys*, the PA-regulating

R2R3MYB candidates, was detected in any sweet orange variety, including the WT control Navel. The WDR and WRKY-encoding genes *CsTTG1* and *CsPH3* were also expressed in all varieties but downregulated approximately 1.5-fold and 2-fold, respectively, in all acidless mutants. Like *CsPH5*, *CsPH1* expression was abolished in Vaniglia, Sorocaba, Verde R1 and Verde R2. *CsDFR* and *CsLDOX* encode members of the flavonoid pathway prior to the PA branch. There was no reliable indication of expression of *CsDFR* in *C. sinensis*. Conversely, *CsLDOX* is highly expressed in Navel, but significantly downregulated in all mutants. Regarding structural genes specific to the PA pathway, no expression of *CsLAR*, *CsANR1*, *CsANR2* was measured in any variety. Furthermore, *CsTT12*, encoding a MATE transporter of PA precursors, was not expressed in any variety, including Navel.

The transcript profiles of genes involved in citrate metabolism and CsMATEs closely related to citrate exporters identified from phylogenetic analyses were also compared (Figure 3.3.23). There were no notable reductions in gene expression observed in the acidless mutants. Genes encoding citrate synthases, *CsCS1* and 2, were not considerably differentially expressed. Further, there were no citric acid degradation-related genes (*CsAco1-3*, *CsIDH1-3*, *CsGAD4* and 5, and *CsGS1-4*) with increased expression in the low citric acid content varieties. The two CsMATEs (10 and 15) were expressed in Navel and slightly down regulated in a few acidless varieties.

RT-qPCR primers were first analysed for efficiency using a cDNA mix of all samples (Supplementary Figure 16). Primers were used for experimental measurements if their efficiency equalled $100\% \pm 10\%$. The RNAseq results were validated via qRT-PCR of various genes of interest (Figure 3.3.24). Expression data are presented as Cq normalised against reference genes *CsEF1 α* and *CsActin*. Statistically significant differences (P -value < 0.05) were observed in all acidless mutants for *Noemi* and *CsPH1* expression. A complete loss of *CsPH5* expression was seen also, however statistical significance was not observed. The expression pattern for most genes was remarkably consistent with the RNAseq results.

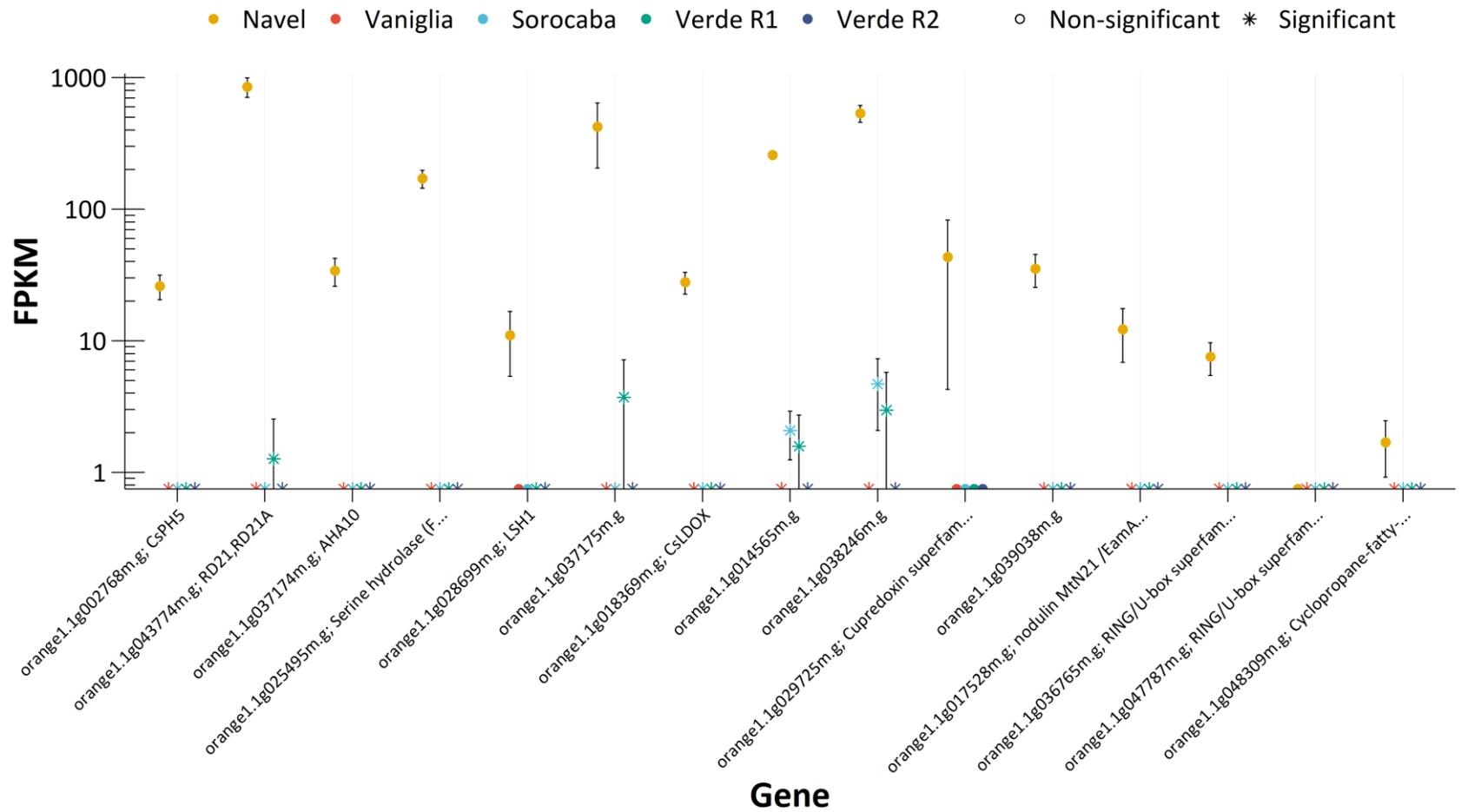


Figure 3.3.21 RNAseq FPKM values in Navel, Vaniglia, Sorocaba, Verde R1 and Verde R2 *C. sinensis* of the 15 most significantly downregulated genes. Asterisks indicate significant difference relative to Navel. Values and error bars presented represent the mean of 3 biological reps \pm se.

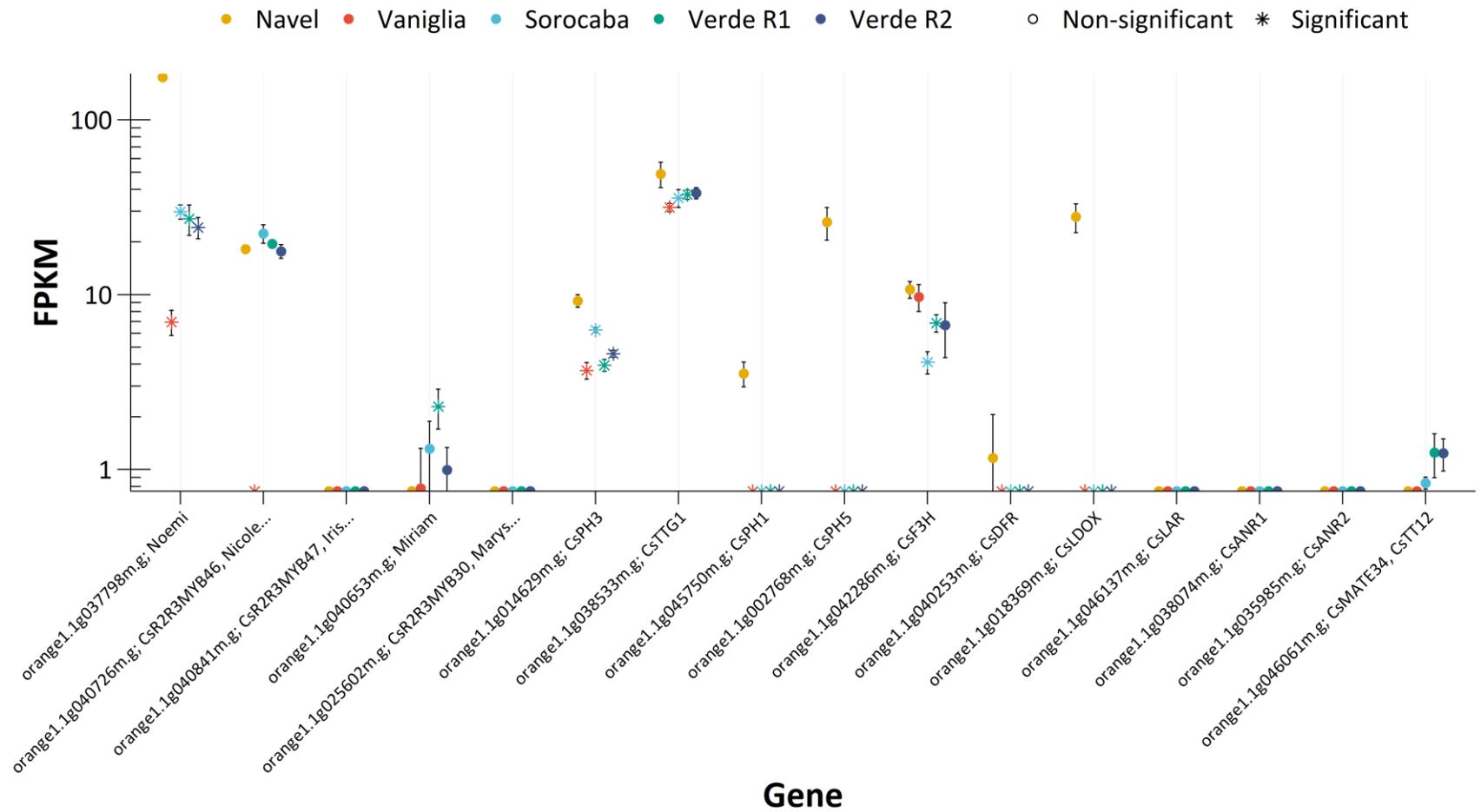


Figure 3.3.22 RNAseq FPKM values in Navel, Vaniglia, Sorocaba, Verde R1 and Verde R2 *C. sinensis* of 16 genes of interest. Asterisks indicate significant difference relative to Navel. Values and error bars presented represent the mean of 3 biological reps \pm se.

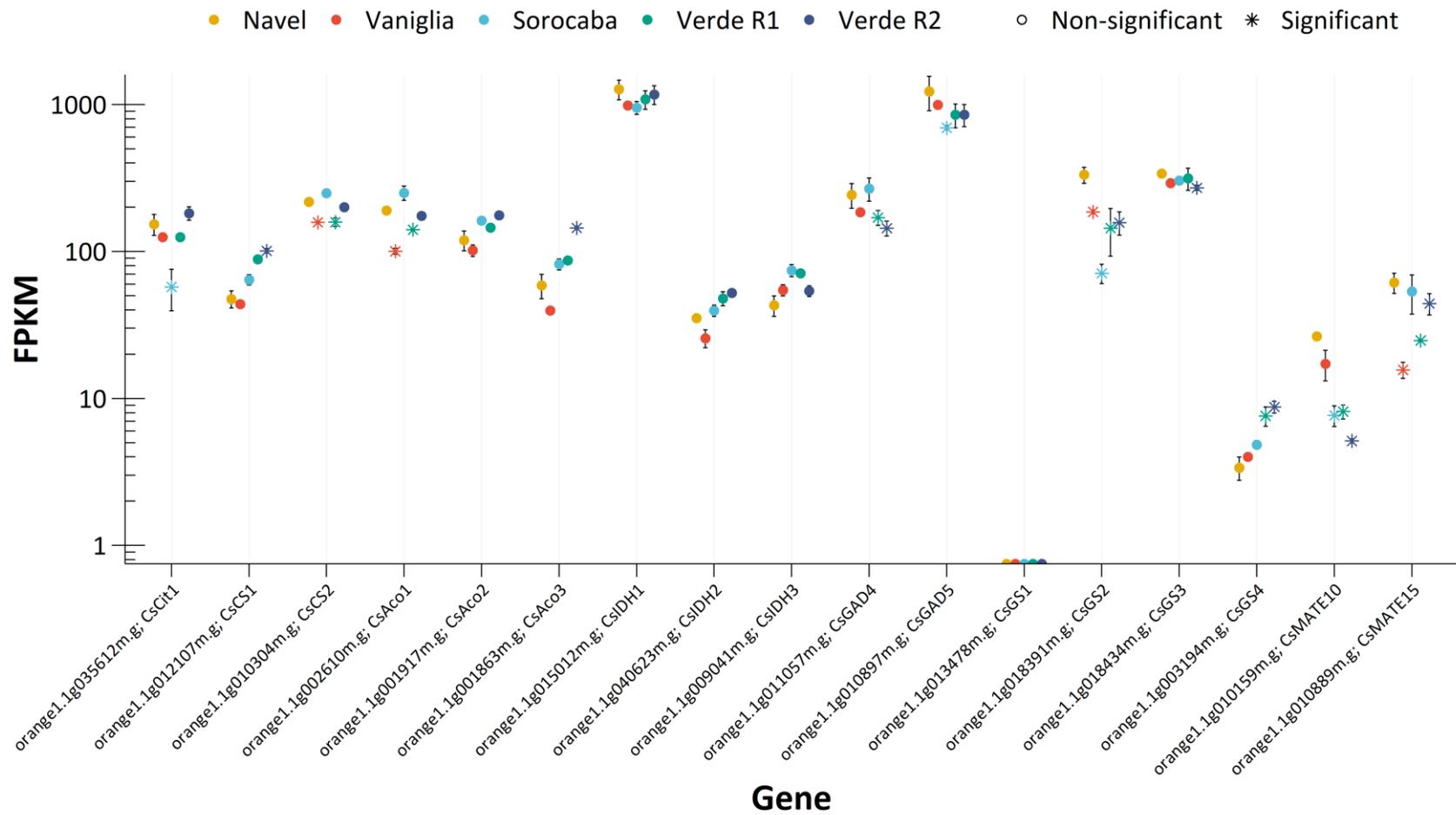


Figure 3.3.23 RNAseq FPKM values in Navel, Vaniglia, Sorocaba, Verde R1 and Verde R2 *C. sinensis* of 17 genes involved in citrate metabolism or transport. Asterisks indicate significant difference relative to Navel. Values and error bars presented represent the mean of 3 biological reps \pm se.

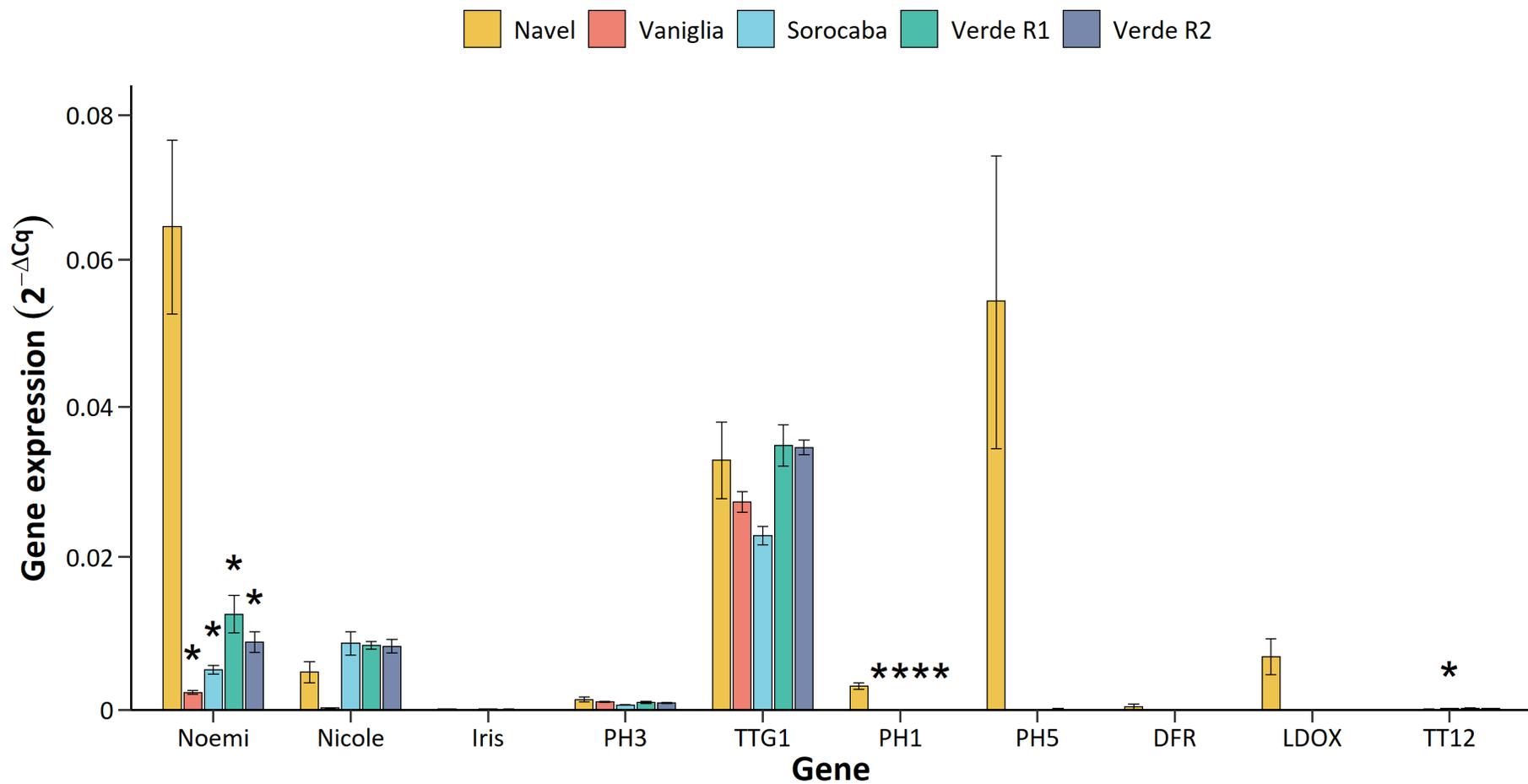


Figure 3.3.24 Gene expression in Navel, Vaniglia, Sorocaba, Verde R1 and Verde R2 *C. sinensis* of 10 genes of interest via RT-qPCR, normalised against housekeeping genes *CsEF1α* and *CsActin*. Asterisks indicate significant difference relative to Navel. Values and error bars presented represent the mean of 3 biological reps \pm se.

3.4 Discussion

3.4.1 Potential roles within the CsR2R3MYB and CsMATE family

Iris is clustered phylogenetically in S5-a, a clade of proteins containing PA-regulators VvMYBPA1 and 2, PpMYBPA1, and AtMYB123 and the VvMYBPA2 putative homolog (Baudry et al., 2004, Terrier et al., 2009, Bogs et al., 2007). VvMYBPA1 and VvMYBPA2 are distantly related within this subgroup. They induce a very similar set of genes in the PA pathway, with only a few specifically regulated by each (Bogs et al., 2007, Terrier et al., 2009). Despite their functional redundancy, a similar phylogenetic profile between VvMYBPA1 and VvMYBPA2 has been reported previously (Terrier et al., 2009).

My phylogenetic analysis confirmed that *Nicole* and *Iris* are good candidate genes regulating acidification and PA biosynthesis respectively in our hypothesised model. Phylogenetic analysis also revealed 35 loci encoding MATE transporters in the sweet orange genome, one of which encodes a protein homologous to the MATE transporter involved in PA and transport, AtTT12 (Marinova et al., 2007).

3.4.2 Genotypic characterisation of *nicole*^{soro}

Until now, only natural *noemi* mutants, such as the sweet orange Vaniglia, have been characterised. Since bHLH proteins can be relatively promiscuous, forming complexes with many MYB transcription factors, any differential expression analysis in Vaniglia is expected to identify many genes involved in independent pathways. Obtaining the Lima varieties has provided an exceptional opportunity to elucidate the function of *Nicole* only, since they are natural *nicole* mutants. I identified an LTR-retrotransposon insertion (*Tcs7x*) in the 3' region of exon 2 of *Nicole*, resulting in the introduction of an early stop codon and subsequently a truncated protein following translation. This mutant allele was named *nicole*^{soro}. *Tcs7x* is likely a Ty1/copia-type LTR-retrotransposon due to the identification of a corresponding PFAM domain within the ORF (PF14223). The only disruption in the *nicole*^{soro} affected the c-terminus, with the R2R3MYB DNA binding domain intact.

However, LTR-retrotransposons can influence gene expression in many ways. In this case, exon-intronic structure has been disrupted, causing premature transcriptional termination, but the location of *Tcs7x* can still transcriptionally silence the gene (Gogvadze and Buzdin, 2009, Sharan et al., 1999). The generation of a C-terminal deletion series facilitated mapping of the activation domain (AD) of AtMYB12, an S7 R2R3MYB transcription factor that regulates flavanols (Stracke et al., 2017). It was shown that a reduction in functionality was first seen after the loss of 46 amino acids, and the AD was in a region close to the C-terminus and highly conserved between members of S7. In contrast, ADs of other R2R3MYBs are located directly at the C-terminus (Goff et al., 1991, Urao et al., 1996). However, while S7 is closely related to S5-b, only functional and transcriptomic analyses can assess the disruption caused by *Tcs7x*. The RNA sequencing data showed no differential expression of *Nicole* in Sorocaba, Verde R1 and Verde R2. The observation of mutually downregulated genes in *nicole* mutants and Vaniglia, which lacks *Nicole* expression, suggests the protein function of *nicole*^{soro} has been disrupted.

3.4.3 The *nicole* mutant retains PAs but loses acidity

Metabolic analyses also suggest a loss of function in *nicole*^{soro}. The acidless mutants exhibit both a fruit juice pH of ~6 and a considerable reduction in citric acid levels. Typically, citric acid accounts for up to 90 % of *Citrus* juice organic acids (Albertini et al., 2006, Chen et al., 2013, Guo et al., 2016, Li et al., 2017, Lin et al., 2015, Zhou et al., 2018). Vacuolar citrate accumulation is attributed to both tonoplastic influx of protons and cytosolic citrate content. Citrate synthase gene expression (*CsCS1* and *CsCS2*) has been shown multiple times to not be responsible for variations in the accumulation of citrate in acidic and acidless *Citrus* fruits (Guo et al., 2016, Chen et al., 2013, Hussain et al., 2017, Sadka et al., 2001, Lin et al., 2015, Lu et al., 2016, Yu et al., 2012). My transcriptomic data are in accordance with these findings as there were no considerable losses in *CsCS1* or *CsCS2* gene expression in any mutant that coincided with the dramatic lack of citric acid concentration quantified in juice. Considering other genes related to citrate metabolism, no other notable changes in expression were observed. A slight but statistically significant decrease in *CsMATE10* and *15* was seen in some acidless mutants. These genes encode MATE transporters structurally homologous to citrate exporters in *Arabidopsis* and *Brassica*, *AtMATE* and *BoMATE* (Wu et al., 2014b, Liu et al., 2009). Their primary role is to enhance Al tolerance and are predominantly expressed in roots. It is interesting to see both expression in Navel and down regulation in a few acidless

varieties. Although, at least in fruit, the reduction in expression may simply reflect the lack of substrate.

Notably the expression of *CsPH1* and *CsPH5* was completely abolished in all acidless mutants, indicating Nicole is a key transcriptional regulator of these genes, particularly because *Noemi* is still expressed in *nicole* mutants. Several studies have proposed CsPH5 is responsible for hyperacidification of the vacuole and facilitates the accumulation of citrate in *Citrus* fruit (Aprile et al., 2011, Strazzer et al., 2019, Shi et al., 2021, Shi et al., 2019). The transcriptomic findings are in accordance with these previous publications, since both a lack of P-ATPase expression (*CsPH1* and 5) and citrate accumulation was observed, and key genes involved in citrate metabolism show no considerable variation in expression. Together, the evidence suggests that the bioaccumulation of citric acid in *Citrus* fruits is predominantly attributed to the activity of proton pumps, particularly *CsPH1* and *CsPH5*. This likely relates to the 'acid trap' mechanism described earlier, whereby the tonoplastic proton gradient, facilitated by at least CsPH5, drives the influx of citrate³⁻ ions (Martinoia et al., 2007, Etienne et al., 2013).

Noemi mutants are characterised by low acidity, and a lack of anthocyanins and PAs. According to our model, we would expect a non-functional MYB transcription factor to affect only the corresponding phenotypic trait associated with the pleiotropic acidless phenotype. DMACA staining qualitatively revealed the presence of tannins in Sorocaba, Verde R1 & R2 seeds, unlike the acidless *noemi* mutant Vaniglia. This was the first example of a mutation without the pleiotropic link between PAs and low acidity in *Citrus* fruits with an acidless phenotype, thereby supporting the model that Nicole regulates hyperacidification, while another MYB transcription factor controls PA biosynthesis. This observation was indicative that another gene is responsible, or that Nicole is not essential, for PA production in *Citrus* fruits.

Furthermore, PAs were localised in cell vacuoles exclusively in the seed coat. This suggests the *AtTT12* homolog, *CsTT12*, is expressed in seeds and is functional, as *AtTT12* is essential for the transport of PA precursor, epicatechin, and subsequent bioaccumulation of PAs in *Arabidopsis* (Marinova et al., 2007). It also suggests *AtBAN/ANR* is expressed to supply the PA precursor substrate for *CsTT12*. *AtTT12* also requires the P_{3A}-ATPase *AtAHA10* to facilitate epicatechin transport by generating a proton gradient (Appelhagen et al., 2015).

Consequently, *CsPH5* is likely expressed in seeds and therefore has likely been transcriptionally activated by a different R2R3MYB transcription factor in the seed coat of *nicole* mutants. In contrast, there were no soluble PAs, or free catechin and epicatechin monomers detected by both colorimetric DMACA assays or HPLC analyses, which directly contradicts the findings of Zhang et al. (2020).

3.4.4 Candidate targets of Nicole

As mentioned, *Noemi* expression was also reduced in Sorocaba, Verde R1 and Verde R2, indicative of a positive feedback loop whereby induction of *Nicole* is *Noemi*-dependant, but *Nicole* also provides non-essential regulation of *Noemi*. The remaining expression of *Noemi* in the Lima varieties was a significant finding as otherwise it would have been impossible to distinguish target genes specific for *Nicole* from those of *Noemi*, if *Noemi* was no longer expressed in *nicole* mutants.

To generate the dramatically low vacuolar pH observed in *Citrus* (hyperacidification), relative to the degree of hyperacidification observed in *Petunia* petals, more genes may be involved in *Citrus*. This may include additional P-ATPases, which provide the principal driving force for hyperacidity in multiple species (Faraco et al., 2014, Amato et al., 2019). Considering their similar expression profiles, and similarity between orange1.1g037174m.g and the C-terminal region of *CsPH5*, one may speculate whether pH differences between *C. limon* and *C. sinensis* may be attributed in part to a second full length, functional *PH5*-like gene located at the orange1.1g037174m.g locus in lemon. It is clear, at least in *C. sinensis*, that superior hyperacidification capability compared to *Petunia* is not attributed to the expression of multiple *PH5*-like genes. Analogous to hyperacidity in *Petunia*, however, these observations suggest that *Nicole* induces *CsPH1* and *CsPH5*, further deviating from the role of PA-regulator as observed in *AtMYB5*. As previously mentioned, *AtAHA10*, the *CsPH5* and *PhPH5* homolog, is inherently a member of the PA pathway, as its vacuolar proton pump activity drives the transport of epicatechins by *AtTT12* (Appelhaagen et al., 2015, Marinova et al., 2007).

The WRKY protein encoded by *PhPH3* is also required for *PH5* expression in *Petunia* (Verweij et al., 2016). *CsPH3*, the homologous transcription factor in sweet orange, is expressed in all sweet orange varieties, although significantly downregulated in the acidless Vaniglia and

Lima fruits. Despite this, *CsPH3* expression is more than 50% of that observed in WT Navel juice. Considering the completely abolished expression of *CsPH5* in the fruit pulp of *nicole* mutants, this suggests that PH3 does not play an essential role in transactivation of the P-ATPase encoded by *CsPH5* but might be a target of Nicole.

Even in wildtype Navel there was no expression of ANR or LAR in juice which directly contradicts recent suggestions that Nicole is involved in PA synthesis (Zhang et al., 2020). Further, there was no detectable expression of PA-related *CsMATE* gene, *CsTT12*, in Navel. There was a slight increase in *CsTT12* expression in the *nicole* mutants, but this was not significant, statistically, and barely surpassed the 1 FPKM threshold for expression. Regardless of the changes in expression, these results indicate that *CsTT12* is not under regulatory control by Nicole and Noemi since the MATE is not expressed in Navel and up regulated just slightly in *nicole* mutants. LDOX and DFR are involved earlier in the flavonoid pathway, shared between anthocyanin and PA branches. *CsLDOX* expression is completely lost in all acidless mutants, indicating it is a likely target of Nicole. This conforms to the activation of *CsLDOX* homologous genes in *Arabidopsis* and grapevine by the respective Nicole homologs. Conversely, DFR is not expressed in any variety, which does not conform to the activation of *CsDFR* homologous genes in *Arabidopsis*. Broadly, my RT-qPCR findings support the RNAseq-derived transcriptomic data.

Chapter 4:
Characterisation of the regulatory function of the
Noemi-Nicole MBW complex

4.1 Introduction

4.1.1 Introduction

Chapter 3 has detailed significant findings in support of the hypothesis that Nicole is responsible for regulating hyperacidification in *Citrus*. Furthermore, there was evidence that Nicole's regulatory function overlaps with the regulation of the PA pathway. However, the potential target genes of Nicole do not include *CsTT12* and *CsDFR*, the homologs of which typically can be transactivated by AtMYB5 homologs, for example in *Arabidopsis* (Xu et al., 2014, Deluc et al., 2008). In addition, I found PA content was undetectable in *C. sinensis* juice. It is known that a WRKY factor (AtTTG2) is essential for the expression of *AtTT12*, which itself is responsible for vacuolar localisation of PA precursors and subsequent bioaccumulation of PAs (Gonzalez et al., 2016, Marinova et al., 2007). Expression of *CsPH3* was observed in all *Citrus* varieties, suggesting *CsTT12* is simply not a target of Nicole. While the transcriptomic data has been foundational for assessing the role of Nicole in *Citrus*, these data have not provided conclusive evidence of the direct activation of candidate target genes by Nicole.

4.1.2 The function of other MYB5-homologs

The regulatory capabilities of various *AtMYB5* homologs have previously been investigated via overexpression analyses in tobacco and complementation of *Arabidopsis* and *Petunia* mutants. While informative, interpretation of these early experiments can be misleading. *VvMYB5a* and *VvMYB5b* were initially thought to primarily regulate the flavonoid pathway as an induction of structural genes within the pathway and PA synthesis was detected in tobacco transformants (Deluc et al., 2006, Deluc et al., 2008). It was later elucidated that these *AtMYB5* homologs in fact play a more minor role, in comparison to the PA-specific regulator *VvMYBPA1*, and also govern vacuolar hyperacidification (Amato et al., 2019). One key indicator that *VvMYB5a* and *VvMYB5b* had other unknown functions was the continual expression in grape tissues which were no longer synthesising PAs. In a similar fashion, I have shown *Nicole* is highly expressed in fruit, despite completely undetectable levels of soluble PAs or the precursor monomers, catechin and epicatechin in juice. Furthermore, the *nicole^{soro}* mutant varieties still produce PAs in the seed coat. In this chapter, I address the direct targets of Nicole to inform our understanding of its role, and whether *nicole^{soro}* is non-functional.

Furthermore, effective complementation of mutants with homologous genes from other species, despite having different endogenous primary functions, have been reported. For example, the *PhPH3* homolog in *Arabidopsis*, *TTG2*, which regulates trichome development, seed coat mucilage and PA synthesis, but not vacuolar hyperacidification, also complements *Petunia ph3* mutants by restoring expression of *PH5* (Gonzalez et al., 2016, Verweij et al., 2016). Of course, homologs that perform similar endogenous roles can also replace one another. *Petunia ph1*, *ph5*, *ph3* and *ph4* mutants which show altered pH, petal colouration and target gene expression can be complemented by expression of the respective *V. vinifera* homologs (Amato et al., 2019, Amato et al., 2016, Li et al., 2016). Consequently, it is important to establish direct induction specifically of *Citrus* promoters by Nicole.

By analysing *N. tabacum* lines overexpressing Nicole, Iris, Noemi, and combinations of either MYB and Noemi, the induction of genes and metabolic changes can be attributed to the introduction of these *Citrus* genes. However, to investigate direct activation of *Citrus* promoters by *Citrus* MBW complexes, dual-luciferase assays were also conducted, revealing a deviation from the typical MYB5-like regulatory targets.

4.2 Materials and Methods

4.2.1 *Nicotiana* plant DNA and RNA extraction

N. tabacum and *N. benthamiana* DNA was extracted from ≤ 100 mg of frozen homogenised leaf tissue with the DNeasy Plant Mini Kit (Qiagen), according to the manufacturer's protocol. Briefly, homogenised tissue was incubated in a lysis buffer and centrifuged to remove cell debris and other precipitates. The lysate was mixed with a binding buffer and ethanol and applied to a DNeasy spin column to bind DNA to the membrane. Following two wash steps the DNA was eluted with H₂O.

N. tabacum RNA was extracted from ≤ 100 mg of frozen homogenised leaf tissue as described by Vennapusa et al. (2020), with some modifications. The tissue was vortexed with 1 ml of RNA extraction buffer (100 mM Tris-HCl (pH 8), 25 mM EDTA, 2.5 M NaCl, 25 mg ml⁻¹ polyvinylpyrrolidone) and incubated for 5 min at room temperature before adding 100 μ l 20 % sodium dodecyl sulphate (SDS). The sample was vortexed vigorously and incubated for at least 2 min at room temperature, but no longer than 10 min. Following a 10 min centrifugation at 4 °C, 14,000 rpm, RNA was extracted from the supernatant by vortexing with 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1). The sample was centrifuged as described previously and the aqueous phase mixed with 1/3 volume chloroform to remove residual phenol. After vortexing vigorously, the sample was centrifuged again.

The aqueous phase was transferred to a new tube. The sample was mixed by inversion with 1/10 volume 3 M sodium acetate (pH 4.8) and 1 volume isopropanol to precipitate RNA. RNA samples were incubated at -20 °C for at least 30 min. The RNA was pelleted by 20 min centrifugation at 4 °C, 14,000 rpm, and washed twice with 1 ml ice-cold 75% ethanol. Ethanol was removed and the pellet dried and subsequently dissolved in RNase-free H₂O.

4.2.2 Qualitative proanthocyanidin analyses in *Nicotiana*

The staining reagent DMACA allows visualisation of PAs in plant tissues. Leaf discs (1 cm) sampled from stably transformed *N. tabacum* plants were bleached overnight in 3:1 (v/v)

ethanol:glacial acetic acid (GAA) solution with gentle agitation. The liquid was then removed, and leaf discs submerged in the bleaching solution for another 30 min. Samples were then stained in ice-cold DMACA solution (0.3 % (w/v) DMACA, 50 % (v/v) methanol, 3 M HCl) with gentle agitation for 30 min before four washes with 70 % ethanol.

4.2.3 Extraction and quantification of proanthocyanidins

PAs were extracted from stable *N. tabacum* transgenic lines and *V. vinifera* fruit skin and flesh following the same method described in Chapter 3.2.9 (Peel and Dixon, 2007). Likewise, quantification of soluble PAs was achieved using the same colorimetric DMACA staining method as described previously but with one minor modification (Pang et al., 2007). Quantification of total soluble PAs was calculated by measuring absorbance at 640 nm following a reaction of 4 µl sample with 96 µl 0.3 % DMACA solution. Values were calibrated against additional independent catechin standard measurements taken with this adapted sample:DMACA reaction ratio.

Additionally, catechin and epicatechin composition of soluble PA extracts were determined via HPLC analysis. The same method described in Chapter 3.2.3 was followed (Downey et al., 2003). However, *N. tabacum* and grape PA extracts were also subjected to a phloroglucinolysis treatment, followed by HPLC analyses to provide information of subunit composition of polymerised soluble PAs (Downey et al., 2003). Exactly 50 µl of PA extract was dried under vacuum and resuspended in 100 µl phloroglucinol buffer (50 mg ml⁻¹ phloroglucinol, 10 mg ml⁻¹ ascorbic acid, 0.1 N HCl solution in methanol). Samples were incubated at 50 °C for 20 min before neutralisation with 100 µl 200 mM sodium acetate (pH 7.5) on ice. Finally, samples were centrifugated at 14,000 rpm, 4 °C, for 15 min. Soluble PA extracts, before and after phloroglucinol treatment, were then analysed via HPLC following the same method as described in Chapter 3.2.3.

4.2.4 Dual-luciferase reporter plasmid construction

Generation of 9 dual-luciferase reporter plasmids was achieved using traditional restriction digestion and subsequent ligation of digested products: a promoter of interest and the recipient vector pGreenII 0800-LUC. Core and proximal promoters typically reside within the 300 bp region upstream of the 5'-UTR, often including transcription factor binding sites

(Smale, 2001, Pedersen et al., 1999, Lemon and Tjian, 2000, Shahmuradov et al., 2005, Porto et al., 2014). DNA-binding site analyses have shown MYB DNA-binding sites are commonly situated 500 bp upstream of the 5'-UTR (Prouse and Campbell, 2012). Further, cloning of 1,000 bp promoter regions have previously been used in transformation experiments, including *Citrus* (Erpen et al., 2018, Porto et al., 2014, Carvalho and Folta, 2017, Li et al., 2012, Hernandez-Garcia et al., 2009). With these factors considered, *Citrus* gene promoters were amplified via Phusion-mediated PCR amplification of the ~ 1,000 bp region upstream of the 5'-UTR to ensure MYB-related cis-regulatory elements were captured for transactivation assays.

An *NcoI* recognition site is the nearest upstream restriction enzyme site to the 5' end of the firefly luciferase CDS and, following *NcoI* digestion, leaves only a single bp prior to the firefly luciferase start codon. Consequently, when amplifying promoters of interest, 3' primers were designed to contain an *NcoI* restriction enzyme site to allow insertion as close as possible to the firefly luciferase CDS downstream. The restriction enzyme site incorporated into 5' primers was dependent on absence of a recognition site within the promoter target sequence and similarity in digestion incubation and inactivation requirements with *NcoI*. Three of the nine reporter plasmids (*CsPH5*, *CsLDOX* and *CsANR* promoters) were constructed by Dr Eugenio Butelli. The negative and positive control reporter plasmids containing a promoter-less (*pEmpty*) and *p35S*-driven firefly luciferase gene, respectively, were provided by Dr Ronan Broad.

4.2.5 Agroinfiltration of *N. benthamiana* leaves

A single colony or a glycerol stock of GV3101 *A. tumefaciens* harbouring the plasmid of interest was used to inoculate a 10 ml LB culture containing selective antibiotics. This culture was grown overnight at 28 °C, 220 rpm. The cells were washed in 15 ml agroinfiltration buffer following centrifugation at 4,000 rpm for 15 min. *A. tumefaciens* cells were collected again by centrifugation and resuspended in 10 ml agroinfiltration buffer and 200 µM acetosyringone.

After room temperature incubation in the dark with gentle agitation, cultures were diluted with agroinfiltration buffer and 200 µM acetosyringone to an OD₆₀₀ of 0.3 – 0.6, depending

on the experiment. For each treatment the underside of 3 leaves from 4-week-old *N. benthamiana* plants were infiltrated with ~ 1 ml of prepared *A. tumefaciens* culture. Leaf samples were harvested for dual-luciferase assays 3-5 days later.

4.2.6 Dual-luciferase reporter assays

To test transcriptional regulation of candidate target genes, respective promoter regions were cloned into the pGreenII 0800-LUC dual-luciferase reporter vector as previously described. Electrocompetent GV3101 *A. tumefaciens* cells were co-transformed with a dual-luciferase reporter plasmid and helper plasmid pSoup. Reporter plasmids were co-infiltrated with various combinations of *C. sinensis* gene overexpression plasmids (provided by Dr Eugenio Butelli) in *N. benthamiana* leaves.

Two 0.5 mm leaf discs, 3-5 days after *Agro*-inoculation were sampled from all 3 biological replicates and placed in 1.5 transparent Eppendorf tubes containing 100 µl 1X PBS. Samples were then processed and measured with the Dual-Glo Luciferase Assay System (Promega) and GloMax 20/20 Luminometer (Promega) according to the manufacturer's protocol. Firefly luminescence corresponds to the level of gene activation attributed to the promoter of interest. In comparison, the Renilla expression is driven by a 35S promoter. Constitutive expression of Renilla thereby provides an internal luminescence control to compare with that of firefly luciferase. The ratio of firefly luciferase:Renilla luminescence in relative luminometer units (RLU) was calculated for each sample.

4.3 Results

4.3.1 Genotyping and gene expression in transformed *N. tabacum* lines

Stable lines of *N. tabacum* transformed with 35S promoter-driven genes encoding Noemi, Nicole and Iris, isolated from *C. sinensis* Valencia cDNA were generated previously by Dr Butelli. Crosses were performed between the p35S:Noemi line with lines p35S:Nicole and p35S:Iris. The subsequent progeny were used for analyses. Five plants from each line were genotyped via PCR amplification of *NPTII*, *Noemi*, *Nicole* and *Iris* T-DNA. Correctly sized amplicons of each gene tested were amplified successfully in every sample (Figure 4.3.1, Figure 4.3.2 and Figure 4.3.3; Supplementary Figure 17).

Putative homologs of PH1, PH5, LDOX, ANR and TT12 in the *N. tabacum* genomes were identified by BLAST with previously characterised genes from *A. thaliana*, *P. hybrida* and *C. sinensis* genomes. Expression of these genes and T-DNA was quantified via RT-qPCR (Figure 4.3.4), following primer efficiency analyses (Supplementary Figure 18). Unfortunately, primers designed for amplification of *NtPH5* and *NtTT12* offered primer efficiencies outside the target threshold (efficiency = 100% ± 10; linear regression *P*-value < 0.05) and were not included in expression analyses. Gene expression in all plant lines was compared to line p35S:NPTII and statistical significance determined via two sample t-tests. *Noemi* expression was detected exclusively in all plant samples containing *Noemi* T-DNA. In p35S:Noemi a marginal level of variation in expression of *Noemi* was observed across all biological replicates. This expression was significantly different to that in the control. In general, expression of the MYB-encoding gene was quite variable in all lines containing the corresponding T-DNA in their genome. Further, the two lines derived from crosses (p35S:Noemi+Nicole and p35S:Noemi+Iris) exhibited lower expression of the respective MYB-encoding gene.

Low levels of *NtPH1* expression were detected in p35S:Nicole, p35S:Noemi+Nicole and p35S:Noemi+Iris in comparison to undetectable levels in the control. This difference was not significant, however. Expression of *NtLDOX* appeared to have been induced in these three lines to a greater degree, particularly in p35S:Noemi+Iris. Likewise, this difference in

comparison to p35S:NPTII was not found to be statistically significant. Finally, NtANR was expressed in p35S:Nicole, p35S:Noemi+Iris and, to a lower extent, p35S:Noemi+Nicole.

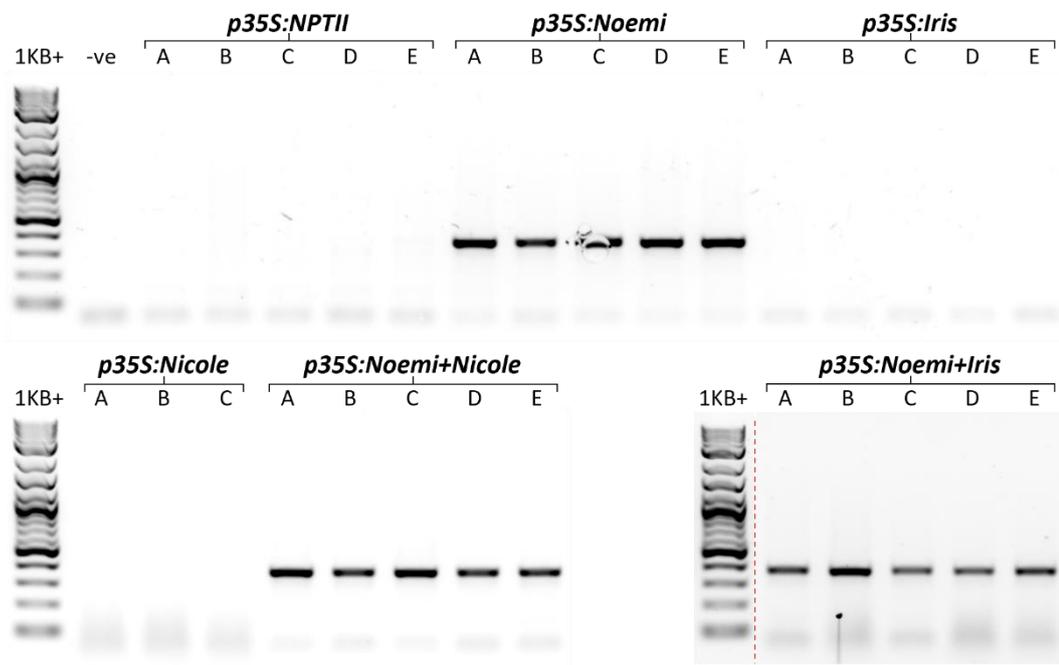


Figure 4.3.1 PCR amplification of Noemi T-DNA from gDNA extracted from overexpression *N. tabacum* transformed lines. 1KB+: 1KB plus ladder (NEB); -ve: no DNA template; red dashed line: indicates gel location where unrelated samples were cropped from the image and remaining areas spliced together.

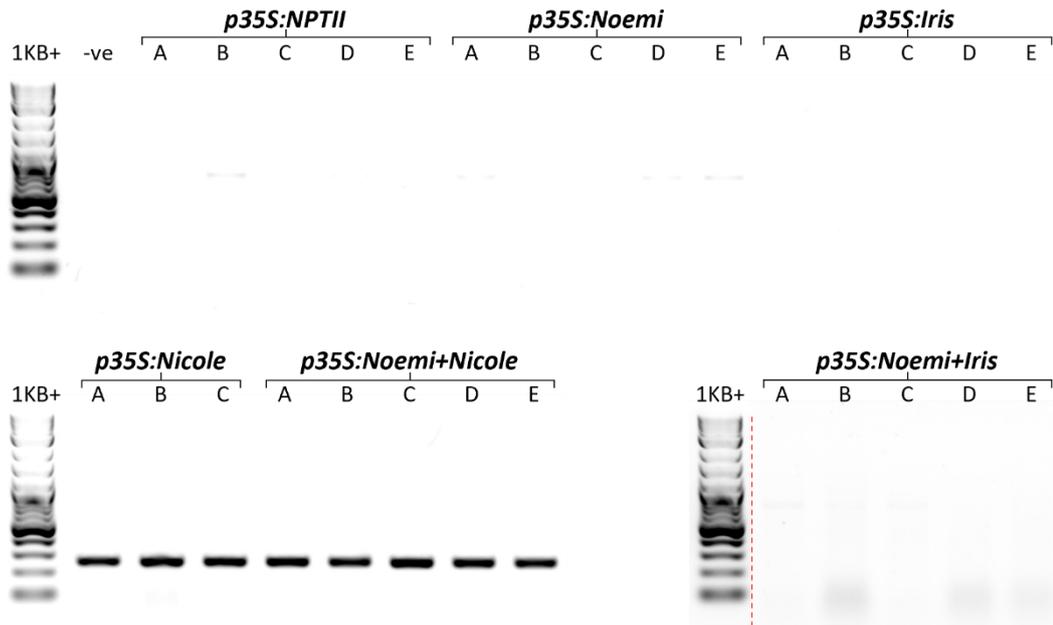


Figure 4.3.2 PCR amplification of Nicole T-DNA from gDNA extracted from overexpression *N. tabacum* transformed lines. 1KB+: 1KB plus ladder (NEB); -ve: no DNA template; red dashed line: indicates gel location where unrelated samples were cropped from the image and remaining areas spliced together.

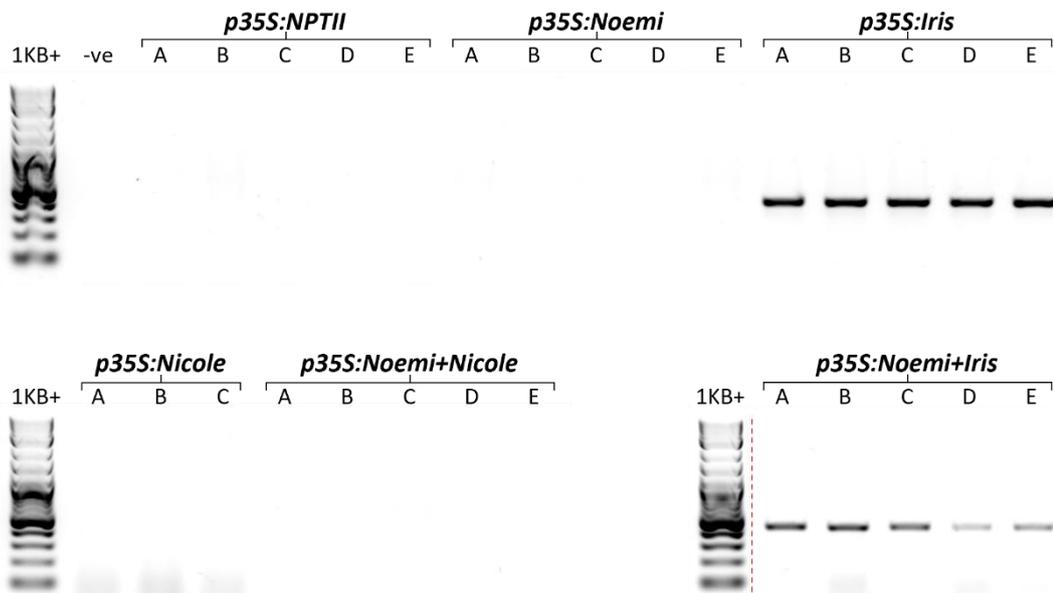


Figure 4.3.3 PCR amplification of Iris T-DNA from gDNA extracted from overexpression *N. tabacum* transformed lines. 1KB+: 1KB plus ladder (NEB); -ve: no DNA template; red dashed line: indicates gel location where unrelated samples were cropped from the image and remaining areas spliced together.

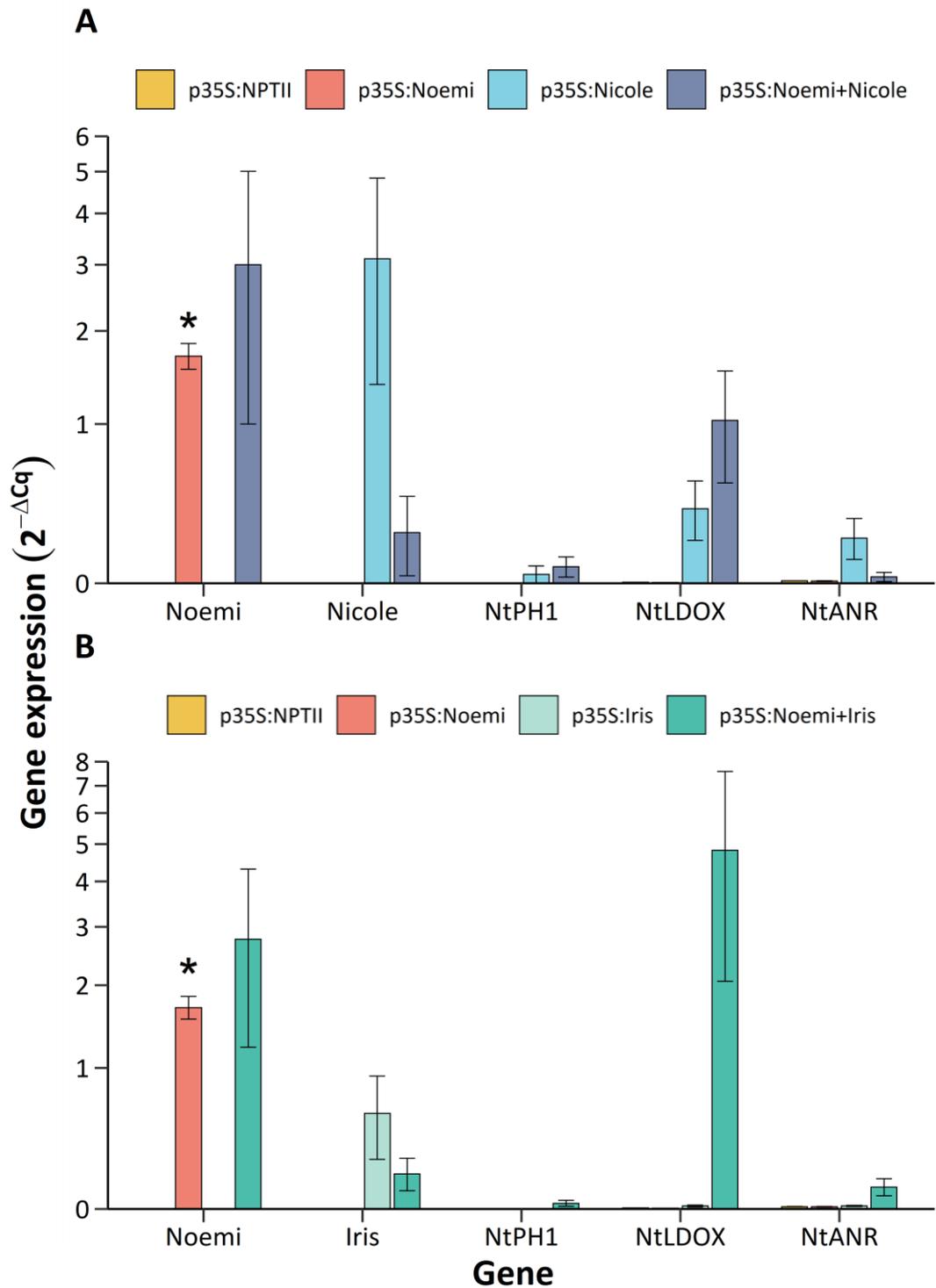


Figure 4.3.4 RT-qPCR analyses of T-DNA and various candidate genes of interest in overexpression *N. tabacum* lines: **A**, Noemi and Nicole-transformed lines; **B**, Noemi and Iris-transformed lines. Values and error bars plotted are means of biological replicates \pm se. Asterisks indicate statistical significance compared to the p35S:NPTII control samples by two sample t-test (P -value < 0.05).

The large within sample variation observed in these expression data of potential target genes is a possible reflection of variation in T-DNA expression. To investigate further a possible correlation between expression of the T-DNA and these candidate target genes a linear regression model was fitted against the data. Between Iris and Nicole expressing lines, the greatest positive correlation between MYB and NtLDOX expression was observed in p35S:Noemi+Iris (p35S:Noemi+Iris: NtLDOX expression = $1.24 + 19.19 * \text{Iris expression}$, $R^2 = 0.44$, $P\text{-value} = 0.22$; p35S:Noemi+Nicole: NtLDOX expression = $0.62 + 1.84 * \text{Nicole expression}$, $R^2 = 0.63$, $P\text{-value} = 0.42$; Figure 4.3.5). These regression models were not statistically significant or strong fits for the data, however.

Similarly, the regression model fitted to NtANR expression was more positive, and statistically significant, when correlated with Iris expression than Nicole (p35S:Noemi+Iris: NtANR expression = $0.03 + 0.45 * \text{Iris expression}$, $R^2 = 0.86$, $P\text{-value} = 0.02$; p35S:Noemi+Nicole: NtANR expression = $0.00 + 0.10 * \text{Nicole expression}$, $R^2 = 0.98$, $P\text{-value} = 0.08$; Figure 4.3.6). The regression model was significant in the p35S:Noemi+Iris sample only. Both models offer high goodness-of-fit R^2 values. A positive correlation was also observed between NtPH1 and MYB expression in p35S:Noemi+Nicole and p35S:Noemi+Iris samples (p35S:Noemi+Iris: NtPH1 expression = $0.01 + 0.07 * \text{Iris expression}$, $R^2 = 0.20$, $P\text{-value} = 0.45$; p35S:Noemi+Nicole: NtPH1 expression = $0.02 + 0.21 * \text{Nicole expression}$, $R^2 = 0.92$, $P\text{-value} = 0.18$; Figure 4.3.7). The correlation between NtPH1 and Nicole expression was stronger than with Iris. In all cases the slope of the line of best fit between candidate target gene and MYB transcription factor was considerably lower, or essentially 0, without the co-expression of Noemi.

4.3.2 Quantification of proanthocyanidins in *N. tabacum*

PA bioaccumulation was investigated via DMACA staining of bleached leaf tissue, colorimetric DMACA assays and HPLC analysis. Leaf discs were first bleached prior to DMACA staining (Supplementary Figure 19). Following staining with DMACA reagent it was evident that PA biosynthesis was only induced in lines co-expressing *Noemi* with either *Nicole* or *Iris* (Figure 4.3.8). Within sample variation was great, with two biological replicates within both p35S:Noemi+Nicole and p35S:Noemi+Iris exhibiting no or very low blue colouration. This corresponded with low expression of the respective MYB-encoding T-DNA. Red pigmentation of flower stamen filaments was observed in all lines containing *Noemi* T-DNA, with and

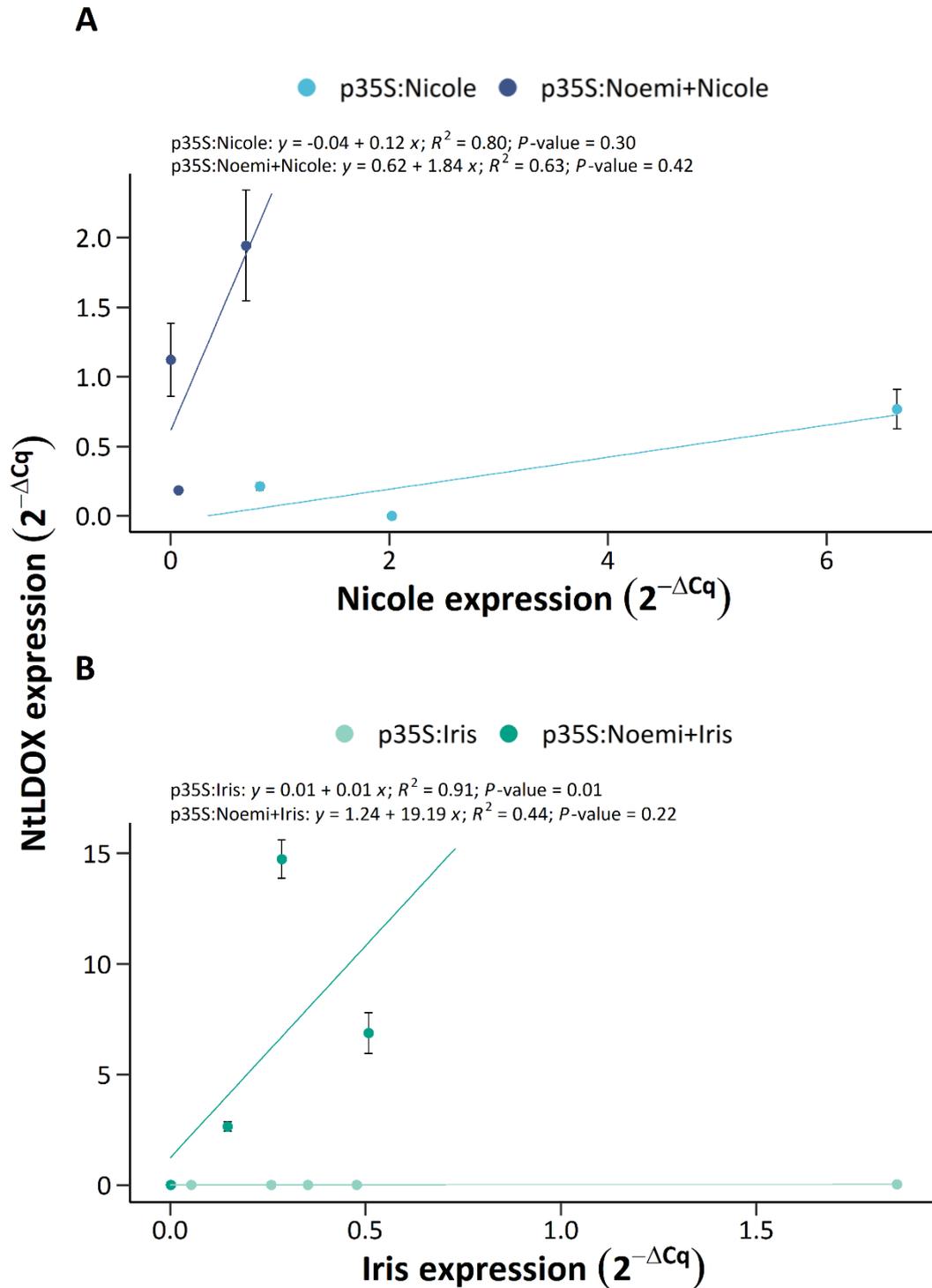


Figure 4.3.5 Scatter plots of Nicole (A) and Iris (B) expression against NtLDOX expression quantified in RT-qPCR analyses of overexpression *N. tabacum* lines. Lines are linear regression lines of best fit. Points and error bars plotted are means of technical replicates \pm se.

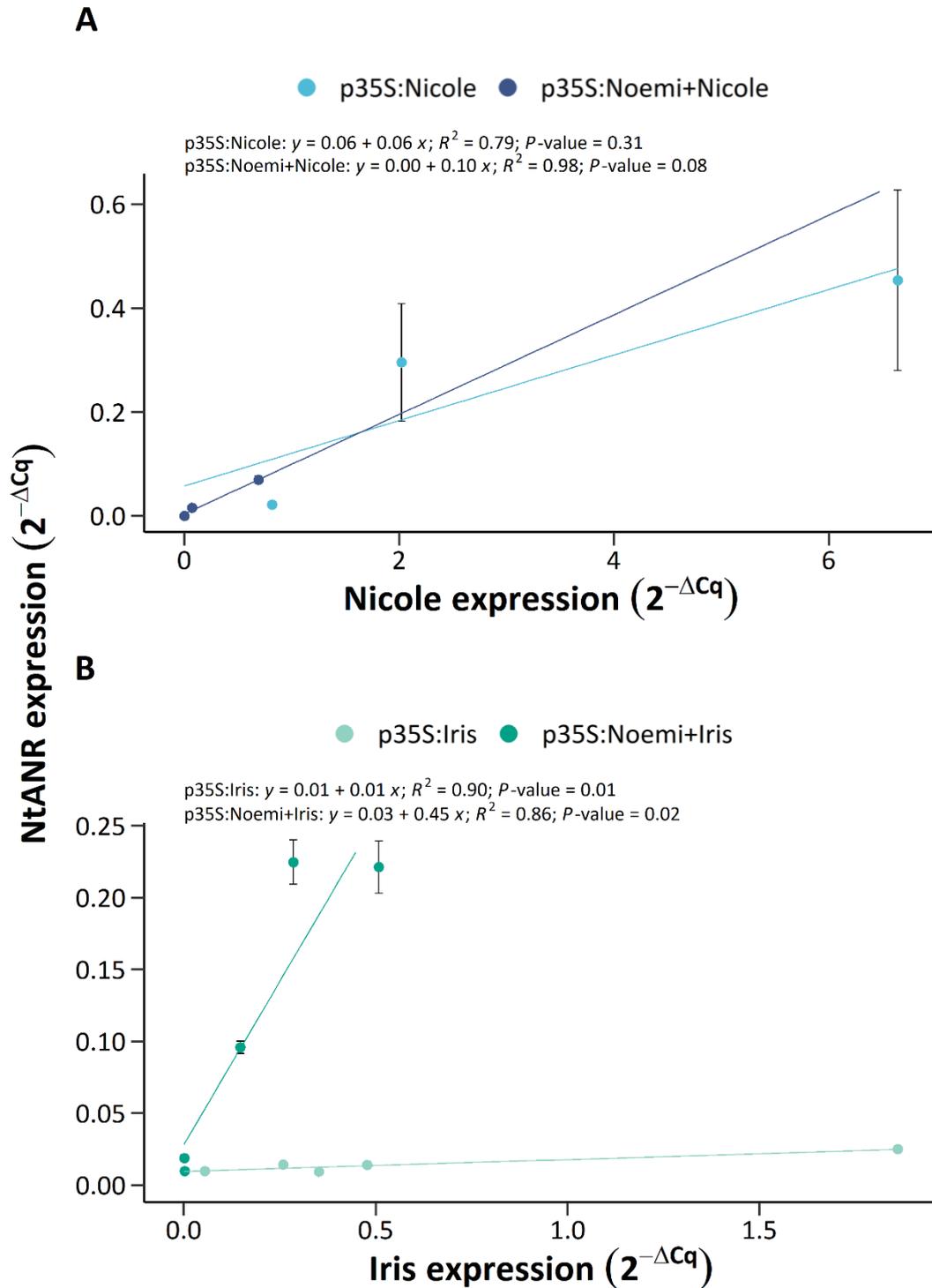


Figure 4.3.6 Scatter plots of Nicole (A) and Iris (B) expression against NtANR expression quantified in RT-qPCR analyses of overexpression *N. tabacum* lines. Lines are linear regression lines of best fit. Points and error bars plotted are means of technical replicates \pm se.

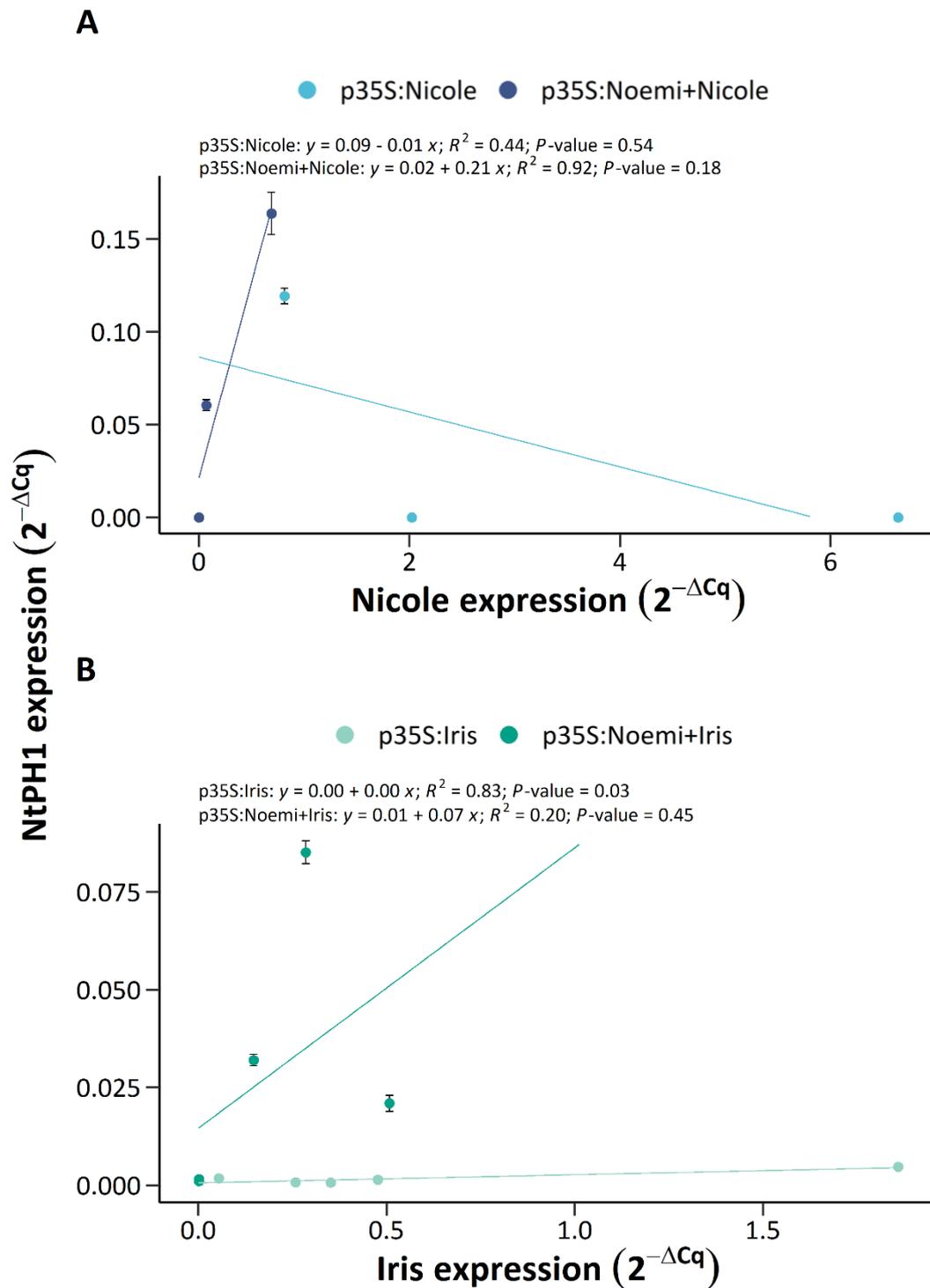


Figure 4.3.7 Scatter plots of Nicole (A) and Iris (B) expression against NtPH1 expression quantified in RT-qPCR analyses of overexpression *N. tabacum* lines. Lines are linear regression lines of best fit. Points and error bars plotted are means of technical replicates \pm se.

without a MYB counterpart. In the co-expression lines petals had a more vibrant colouration compared to all other lines.

Soluble PA extracted from transgenic tobacco lines were quantified via colorimetric DMACA assays, in addition to grape skin and flesh tissues as a positive control (Figure 4.3.9). Following reaction with DMACA reagent absorbance at 640 nm was measured at regular ~ 3 min intervals (Supplementary Figure 20). Catechin standard curves were plotted for each time point and analyses proceeded with the ~ 11 min post-staining dataset, quantifying soluble PAs as catechin equivalents (Supplementary Figure 21). PAs were quantified in both grape tissues, confirming successful application of the extraction protocol and subsequent quantification via colorimetric DMACA assays (skin = 6.09 ± 1.02 mg ml⁻¹; flesh = 0.12 ± 0.07 mg g⁻¹). Soluble PAs were only detected in tobacco lines p35S:Noemi+Nicole (0.45 ± 0.24 mg g⁻¹) and p35S:Noemi+Iris (0.54 ± 0.33 mg g⁻¹). Despite a lack of PA detection in the p35S:NPTII line, statistical significance was not achieved. Similar to the RT-qPCR results, PA content varied greatly within these sample groups. PA content within biological replicates also strongly reflected qualitative leaf staining observations and gene expression of the respective MYB T-DNA (Supplementary Figure 22).

The same soluble PA extracts underwent HPLC analyses to quantify catechin and epicatechin composition. Further, cleavage products derived from phloroglucinolysis treatment of PA extracts were analysed via HPLC to investigate subunit composition (Figure 4.3.10). Within both p35S:Noemi+Nicole and p35S:Noemi+Iris there was approximately the same concentration of free and released terminal subunit monomers (both catechins and epicatechins). Variation within sample groups was high and likely attributed to the variation in T-DNA expression detailed previously. The p35S:Noemi+Nicole samples contained relatively equal levels of catechins and epicatechins, whereas p35S:Noemi+Iris contained a greater proportion of epicatechins.

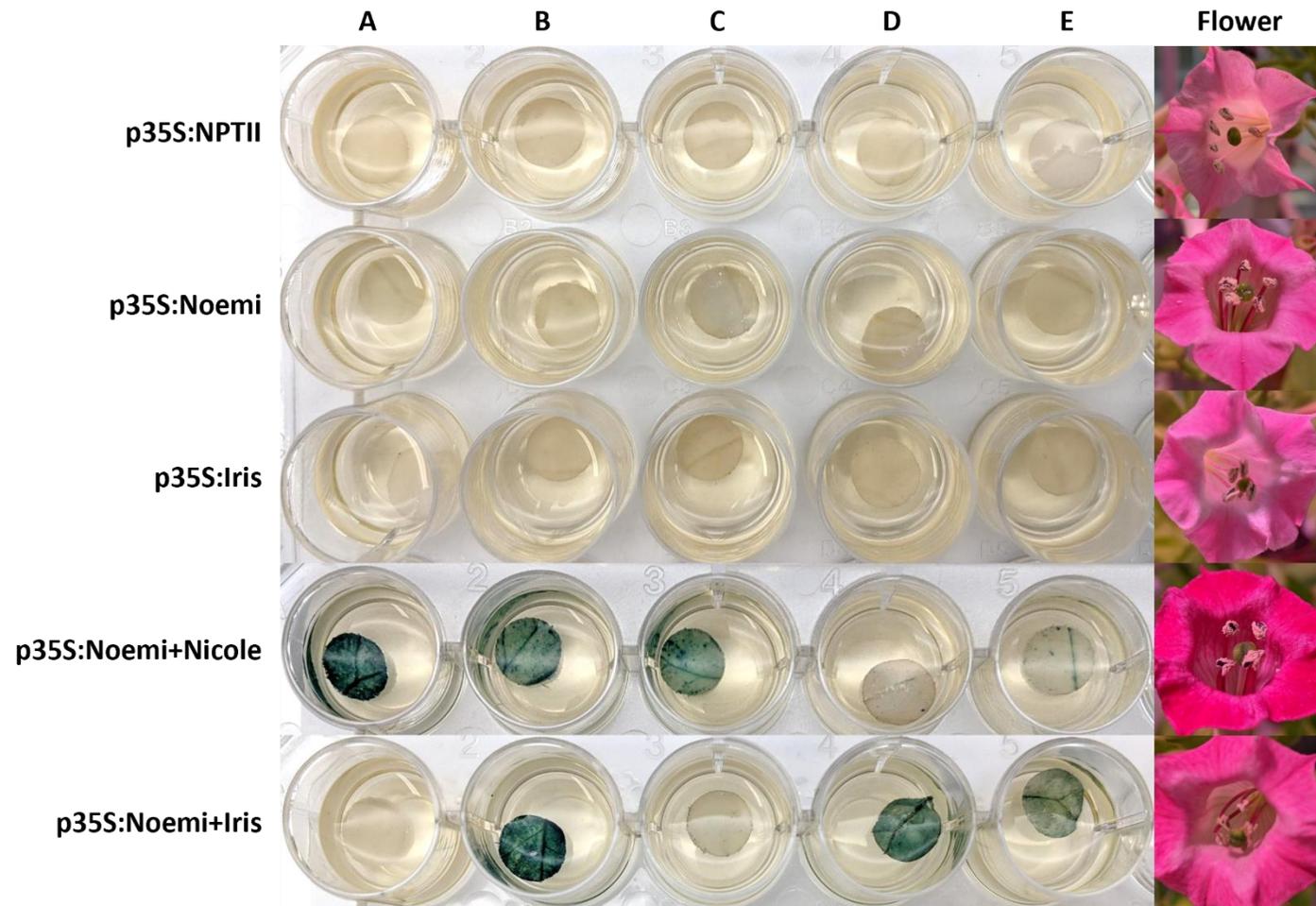


Figure 4.3.8 Qualitative phenotypic changes in overexpression *N. tabacum* lines. Leaf discs from five biological replicates (A-E) per line were bleached and stained with 0.3 % DMACA solution. Flower images are representative of typical flower in each line.

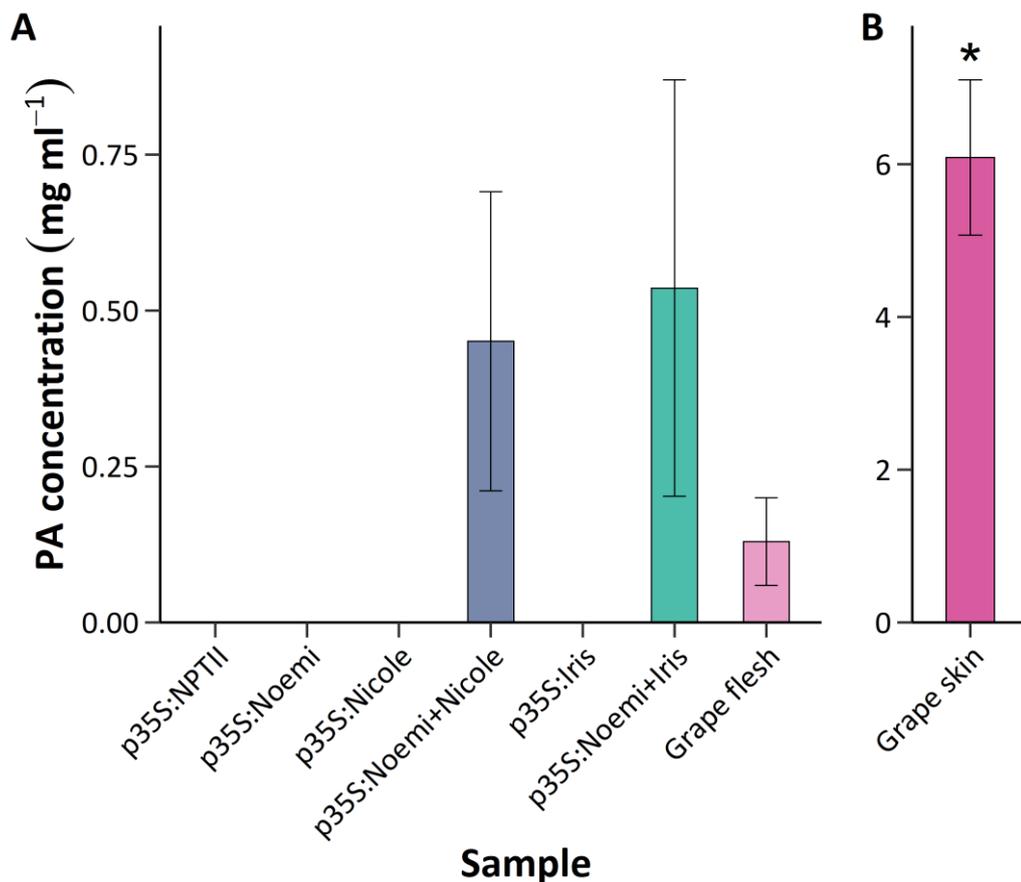


Figure 4.3.9 Proanthocyanidin quantification of via colorimetric DMACA assays. **A**, overexpression *N. tabacum* leaf and grape flesh extracts; **B**, grape skin extracts as a positive control. Values and error bars plotted are means of biological replicates \pm se. Asterisks indicate statistical significance compared to the p35S:NPTII control samples by two sample t-test (P -value < 0.05).

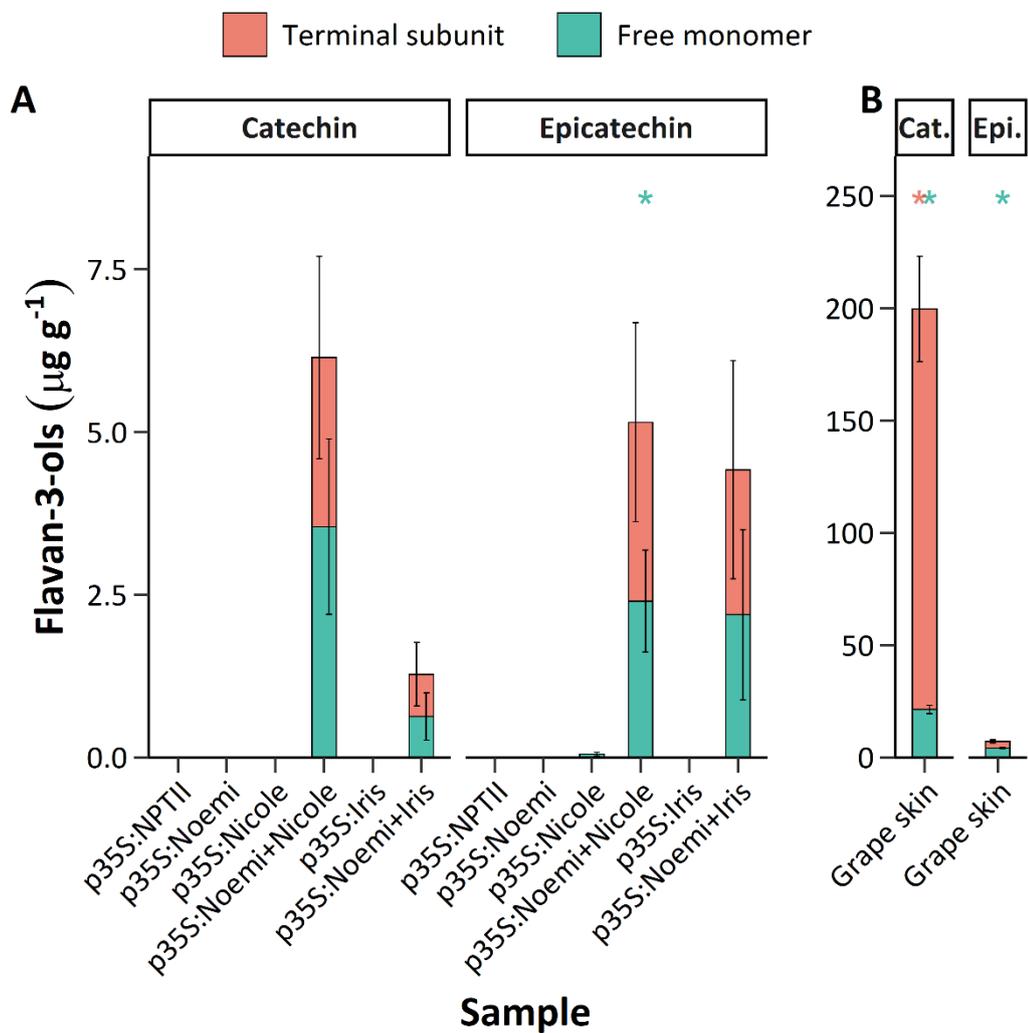


Figure 4.3.10 HPLC quantification of free catechin and epicatechin monomer concentration of soluble PA extracts. **A**, overexpression *N. tabacum* leaf and grape flesh extracts; **B**, grape skin extracts as a positive control. Catechin was added to Valencia+ samples prior to extraction to calculate recovery percentage. Asterisks indicate significant difference relative to the Navel control for the respective metabolite. Values and error bars presented represent the mean of 3 (2 for Navel) biological reps \pm se.

4.3.3 Dual-luciferase reporter gene assays

Regulatory activation of candidate target gene promoters by Nicole, Iris, and Marys was assessed via dual-luciferase reporter assays in *N. benthamiana* leaves. Reporter plasmids were constructed containing a promoter of interest upstream of the firefly luciferase-encoding gene. A firefly luciferase luminescent signal was activated first and measured, followed by quenching of the reaction and initiation of the internal control *Renilla* luciferase reaction and subsequent measurement. Activation of the promoter was quantified by the ratio between firefly luciferase and *Renilla* luminescence. Promoters were cloned via amplification of the ~ 1 kb region upstream of the 5'-UTR of candidate target genes derived from *C. sinensis* Navel or Valenica gDNA. Reporter plasmids were co-infiltrated into *N. benthamiana* leaves with various combinations of *C. sinensis* MYB transcription factors, Noemi and CsPH3. Treatments were compared with the infiltration of the reporter plasmid only, as a negative control via two sample t-tests. A reporter plasmid containing a *p35S*-driven firefly luciferase was also infiltrated as a positive control.

Noemi and *Nicole* promoters were tested for activation by Noemi, Nicole and *nicole*^{so^{ro}} (Figure 4.3.11 and Figure 4.3.12). The *Noemi* promoter required co-expression of Noemi and Nicole for transactivation, whereas *pNicole* was able to be significantly activated by Nicole (*P*-value < 0.05), with and without co-expression of the bHLH transcription factor. Activation of *pNicole* was higher, however, when both Nicole and Noemi were co-infiltrated. While *nicole*^{so^{ro}} was able to activate *pNoemi* and *pNicole* when co-expressed with Noemi, it achieved lower levels of induction compared to Nicole.

Similar patterns of activation of *pCsPH1* and *pCsPH5* were observed (Figure 4.3.13 and Figure 4.3.14). Both promoters were directly and significantly activated by Nicole with Noemi. The participation of *PH3* did not result in a notable increase. The mutant *nicole*^{so^{ro}} was able to significantly induce *pPH1*, albeit to lower levels compared to Nicole, when co-expressed with Noemi. Both Nicole and Iris activated *pCsLDOX* when coexpressed with Noemi, while *nicole*^{so^{ro}} lost this ability (Figure 4.3.15). A 3-fold stronger induction was achieved by Noemi+Iris (4.43 ± 0.26 RLU) in comparison to Noemi+Nicole (1.10 ± 0.19 RLU).

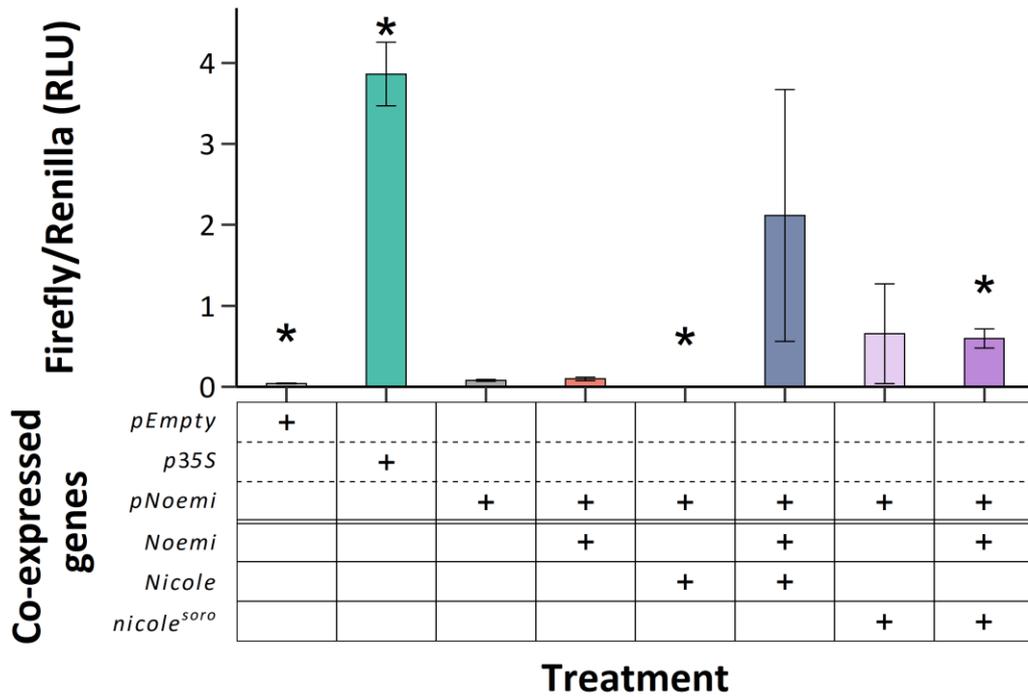


Figure 4.3.11 Dual-luciferase *pNoemi* reporter activation by co-expression of various combinations of *C. sinensis* MYB and bHLH-encoding genes. Values and error bars plotted are means of 6 biological replicates \pm se. Asterisks indicate statistical significance compared to the *pNoemi* reporter plasmid alone by two sample t-test (P -value < 0.05).

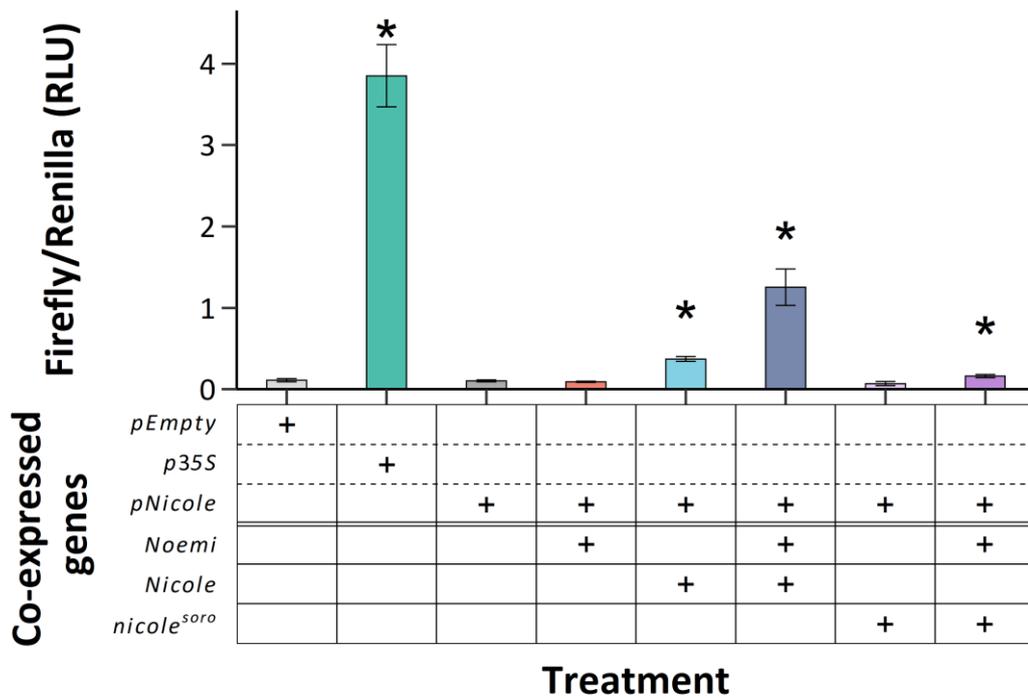


Figure 4.3.12 Dual-luciferase *pNicole* reporter activation by co-expression of various combinations of *C. sinensis* MYB and bHLH-encoding genes. Values and error bars plotted are means of 6 biological replicates \pm se. Asterisks indicate statistical significance compared to the *pNicole* reporter plasmid alone by two sample t-test (P -value < 0.05).

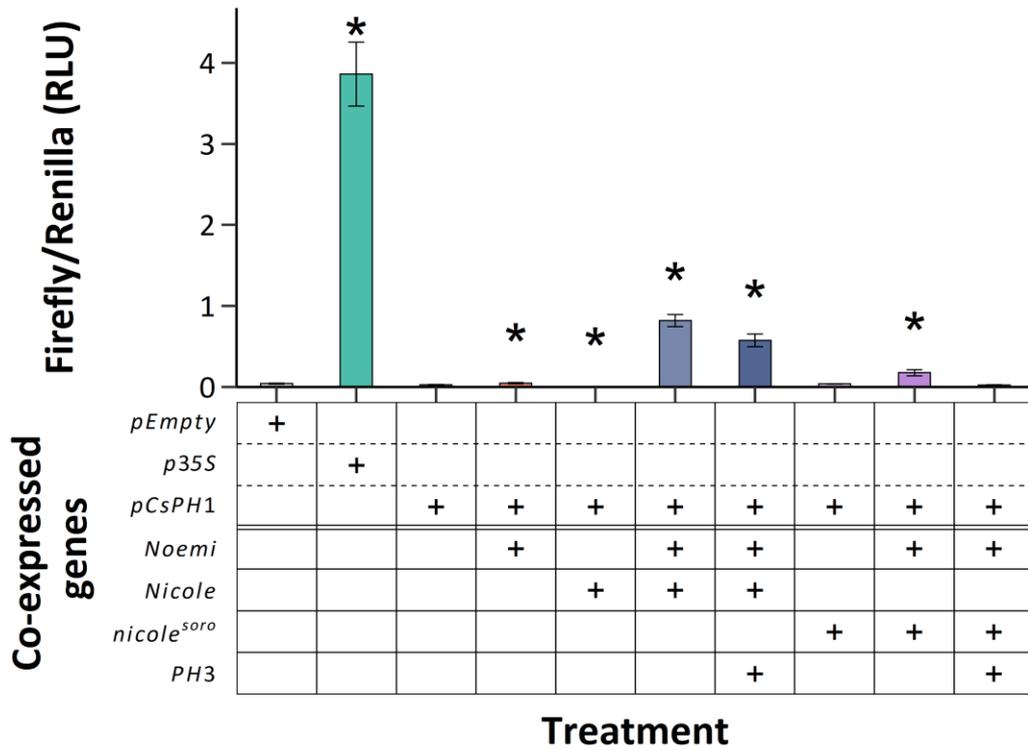


Figure 4.3.13 Dual-luciferase pCsPH1 reporter activation by co-expression of various combinations of *C. sinensis* MYB and bHLH-encoding genes. Values and error bars plotted are means of 6 biological replicates \pm se. Asterisks indicate statistical significance compared to the pCsPH1 reporter plasmid alone by two sample t-test (P -value < 0.05).

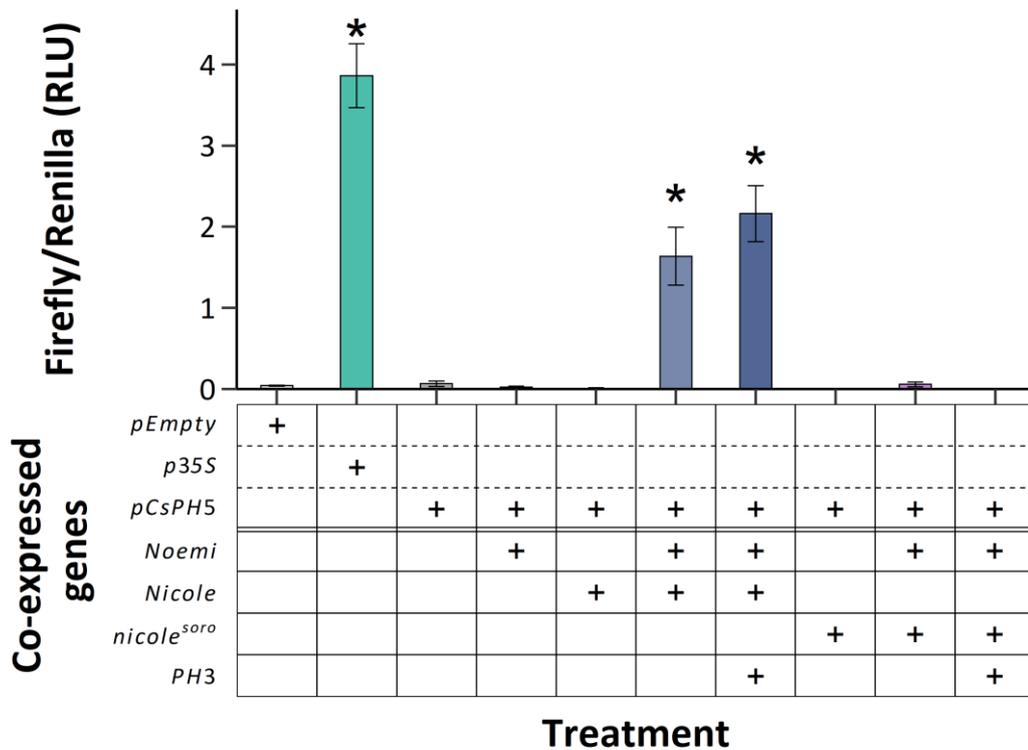


Figure 4.3.14 Dual-luciferase pCsPH5 reporter activation by co-expression of various combinations of *C. sinensis* MYB and bHLH-encoding genes. Values and error bars plotted are means of 6 biological replicates \pm se. Asterisks indicate statistical significance compared to the pCsPH5 reporter plasmid alone by two sample t-test (P -value < 0.05).

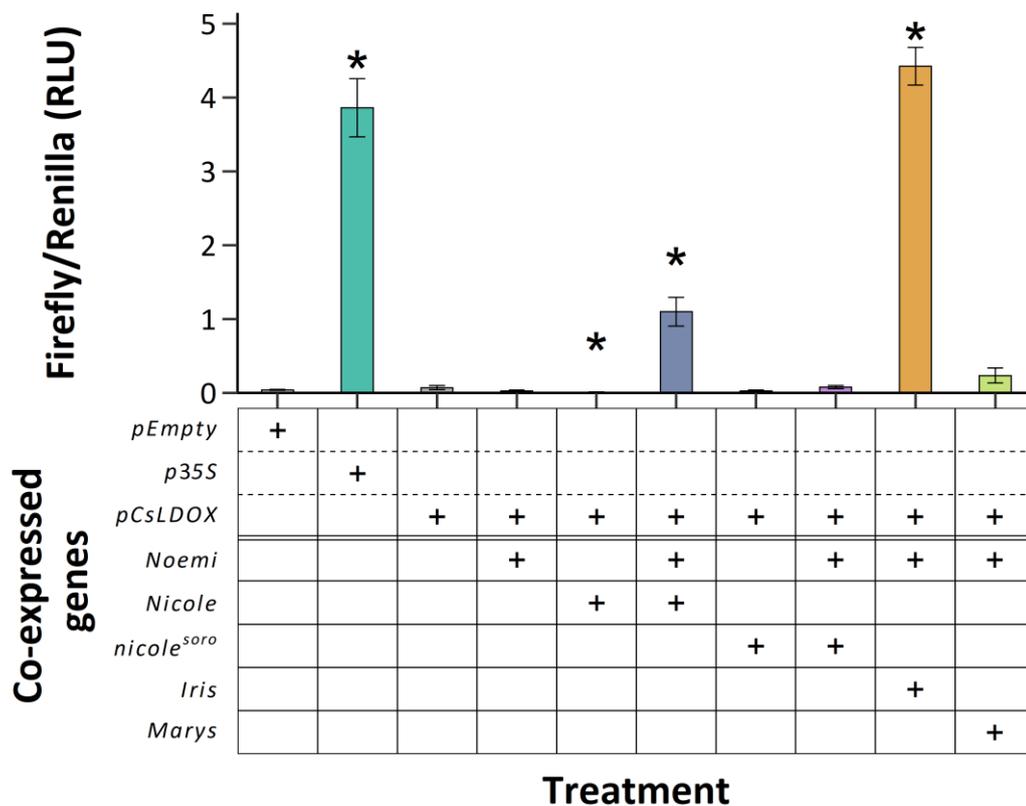


Figure 4.3.15 Dual-luciferase *pCsLDOX* reporter activation by co-expression of various combinations of *C. sinensis* MYB and bHLH-encoding genes. Values and error bars plotted are means of 6 biological replicates \pm se. Asterisks indicate statistical significance compared to the *pCsLDOX* reporter plasmid alone by two sample *t*-test (P -value < 0.05). Induction of *pCsLDOX* by *Iris*-*Noemi* and *Marys*-*Noemi* were performed by Dr E Butelli.

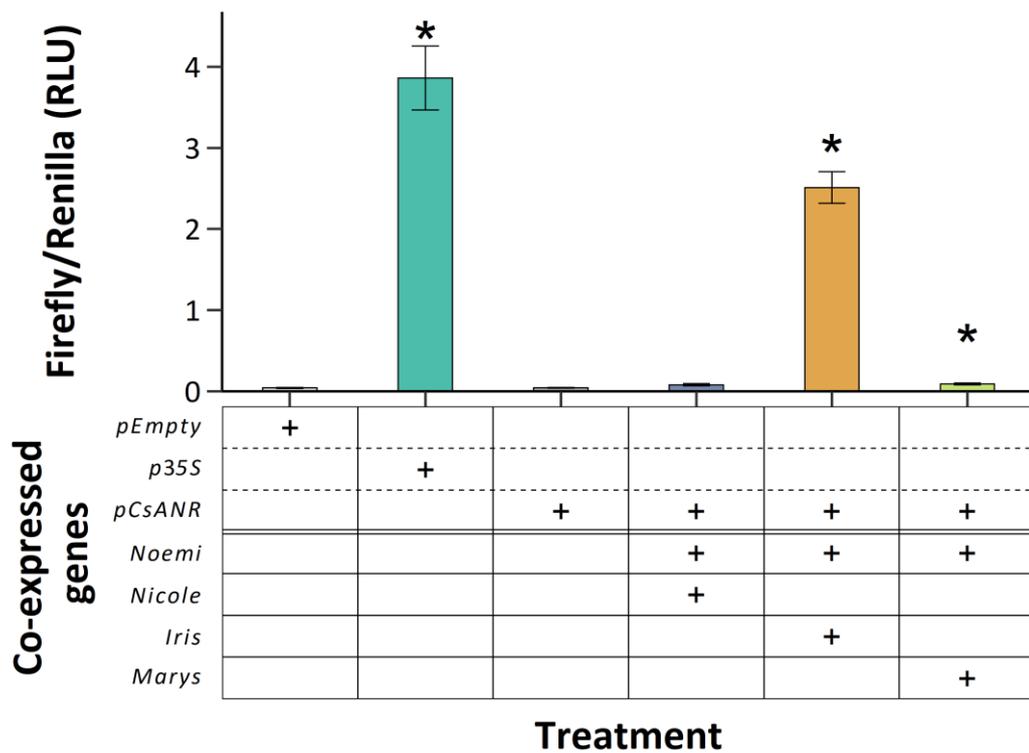


Figure 4.3.16 Dual-luciferase *pCsANR* reporter activation by co-expression of various combinations of *C. sinensis* MYB and bHLH-encoding genes. Values and error bars plotted are means of 6 biological replicates \pm se. Asterisks indicate statistical significance compared to the *pCsANR* reporter plasmid alone by two sample t-test (P -value < 0.05). This experiment was performed by Dr E Butelli.

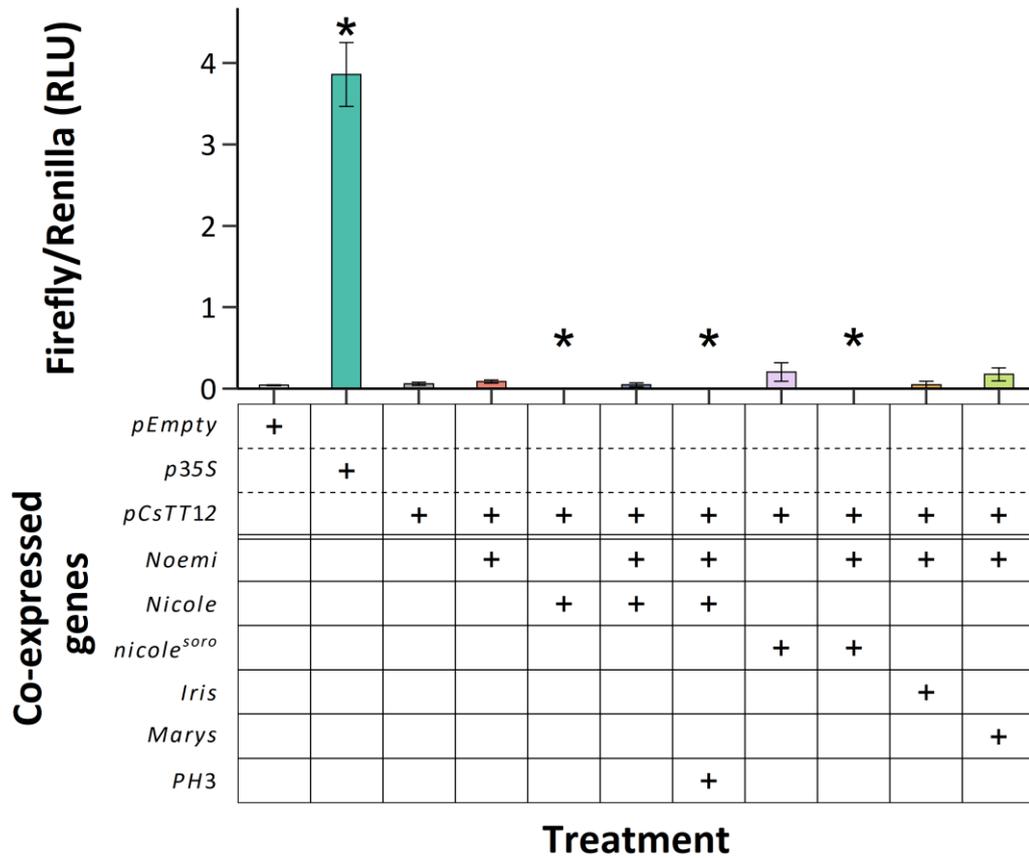


Figure 4.3.17 Dual-luciferase *pCsTT12* reporter activation by co-expression of various combinations of *C. sinensis* MYB and bHLH-encoding genes. Values and error bars plotted are means of 6 biological replicates \pm se. Asterisks indicate statistical significance compared to the *pCsTT12* reporter plasmid alone by two sample t-test (P -value < 0.05). This experiment was performed by Dr E Butelli.

Assays testing the transactivation of *CsANR* and *CsTT12* promoters revealed Nicole was not able to induce the expression of the firefly luciferase, even with Noemi (Figure 4.3.16 and Figure 4.3.17). Further, *pCsTT12* was not activated by any transcription factor combination tested, including PA-regulating candidates Iris and Marys. Notably, *pCsANR* was directly activated by co-expressing Iris and Noemi, but not by Nicole or Marys in combination with Noemi.

4.4 Discussion

4.4.1 Proanthocyanidin biosynthesis and associated genes are induced by both Nicole and Iris MBW complexes when ectopically expressed in *N. tabacum*

As it was not possible, due to time constraints, to assess the function of the MBW components of interest within *C. sinensis* in-vivo, for example by rescuing the *nicole* mutant acidless phenotype with Nicole via stable transformation, *N. tabacum* lines transformed with p35S-driven Noemi, Nicole and Iris cDNA were developed (produced by Dr Butelli). Crosses were performed to produce lines expressing both Noemi and a MYB transcription factor, and all lines were confirmed to contain the corresponding T-DNA via amplification from gDNA. Despite this, a large degree of variation in expression of these introduced genes was observed. This was mirrored in subsequent analyses of expression data of various candidate target genes and PA quantification. As a result, linear regression models were fitted to the expression data to determine correlations between candidate target genes and MYB transcription factors Nicole and Noemi.

My data suggest there is a correlation between Nicole and at least three candidate target genes, *NtPH1*, *NtLDOX* and *NtANR*, which are expressed at marginal or undetectable levels in the negative control line. Iris expression was more strongly correlated with just *NtLDOX* and *NtANR*. Unfortunately, the findings were compromised by a wide variation of MYB expression within sample groups. Although statistical significance was not observed in these data, the induction of typically non-expressed genes in tobacco was notable, nevertheless. It was clear that the expression of these genes was considerably greater or perhaps dependant on co-expression of Noemi and either MYB transcription factor, providing evidence that they form MBW complexes. The expression of these MYBs is lower in p35S:Noemi+Nicole and p35S:Noemi+Iris, compared to the lines expressing each MYB alone. This possibly curtailed the increase in expression of target genes observed when Noemi was expressed together with either MYB.

In parallel, an induction of soluble PA bioaccumulation was observed qualitatively and quantitatively in both p35S:Noemi+Nicole and p35S:Noemi+Iris tobacco. Application of the PA extraction method was effective as grape skin PA concentrations quantified were

comparable to previous reports (Seddon and Downey, 2008). Comparing PA induction by Nicole and Iris was not informative in this case due to the large variation of T-DNA expression. What was notable, however, was the complete lack of PAs in the control line and in all lines expressing only one *Citrus* gene.

The induction of PA bioaccumulation by Nicole is consistent with previous reports in tobacco plants constitutively expressing *AtMYB5* and homologs, such as grapevine *VvMYB5a* and *VvMYB5b*, in tandem with the induction of structural PA-related genes of the flavonoid pathway (Deluc et al., 2006, Deluc et al., 2008). Furthermore, Nicole can partially complement seed PA content in *A. thaliana tt2* mutants and increase soluble PA content in Col-0 (Zhang et al., 2020). Observations such as these resulted in the initial view that the function of *VvMYB5a* and *VvMYB5b* was primarily to regulate flavonoid synthesis. However, both genes are expressed in grapevine tissues that lack flavonoids and the *AtTT2* homolog, *VvMYBPA1*, transactivates PA-related genes more strongly than either *VvMYB5a* or *VvMYB5b* (Cavallini et al., 2014, Hichri et al., 2010).

Unfortunately, *NtPH5* and *NtTT12* RT-qPCR expression data were not measured in tobacco lines due to low primer efficiencies. Not only should these primers be redesigned, but the expression of other structural genes in the acidity/PA pathway, such as *NtF3H*, *NtLAR* and *NtDFR* via RT-qPCR should be prioritised to understand more fully the respective regulatory capabilities of Nicole and Iris in the PA pathway. Despite the lack of data, it is highly likely that *NtTT12* gene expression was induced in both p35S:Noemi+Nicole and p35S:Noemi+Iris tobacco lines as *TT12* is essential for soluble PA bioaccumulation in *Arabidopsis* and soluble tannins were induced in the overexpression tobacco lines (Marinova et al., 2007). In parallel, induction of *CsPH5* is anticipated by analogy to PA bioaccumulation in *Arabidopsis*, whereby vacuolar loading by *TT12* is governed by the proton gradient generated by tonoplast-bound proton pump *AHA10* (putative *CsPH5/PhPH5* homolog in *A. thaliana*: *TT13*) (Appelhagen et al., 2015, Baxter et al., 2005). Finally, the composition of soluble tannins in transgenic tobacco samples may also infer the degrees to which *NtANR* and *NtLAR* are induced, as they catalyse the production of epicatechin and catechin, respectively.

As noted previously, comparisons between Nicole and Iris regarding concentrations of PAs induced is not informative due to the variation in T-DNA expression. However, comparisons

of the PA composition are. It was evident that p35S:Noemi+Iris preferentially induced the synthesis of epicatechins. For both catechins and epicatechins, half of all flavan-3-ol monomers were acting as terminal subunits of polymerised PAs in both PA-induced tobacco lines.

4.4.2 Nicole and Iris-driven MBW complexes directly transactivate different gene promoters in *C. sinensis*

Analogous to *Petunia* and grapevine, I have confirmed that the AtMYB5 and AtbHLH42 putative homologs Nicole and Noemi, respectively, associate as a transcriptional regulatory MBW complex, consistent with previous reports concerning other species (Amato et al., 2019, Quattrocchio et al., 2006, Strazzer et al., 2019, Zhang et al., 2020). The transactivation of *pNoemi*, *pCsPH1*, *pCsPH5* and *pCsLDOX* by Nicole was also dependant on the co-expression of Noemi. Nicole was able to self-induce its own promoter weakly but this activation was stronger when participating in an MBW complex with Noemi. As previously reported in *AtMYB5*-like and *AtTT2*-like homologs, *pCsANR* was only a target of *TT2*-like Iris, when expressed in combination with Noemi (Xu et al., 2014). Interestingly, the *CsTT12* promoter was not a target of Nicole or Iris alone, or when either transcription factor participated in an MBW complex with *Noemi*.

In other plant systems the recruitment of WRKY proteins typify, and in some cases is essential for, transcriptional regulation of target genes by MBW complexes. This could explain the lack of induction of *pCsTT12* by Iris-Noemi. For example, the WRKY *AtTTG2* directly targets and is essential for *AtTT12* expression (Gonzalez et al., 2016). In turn, *TT12* facilitates the polymerisation and bioaccumulation of soluble PAs via tonoplastic transportation of PA precursors (Marinova et al., 2007). *Petunia* PH3, also encoding a WRKY factor, is key for transcription of *PhPH5* and *PhPH1*, and *VvWRKY26* in grapevine boosts *VvPH5* expression up to 10-fold following recruitment by *VvMYB5a* (Faraco et al., 2014, Amato et al., 2019, Verweij et al., 2016). Consequently, the *C. sinensis* putative WRKY homolog of *AtTTG2* and *Petunia* PH3, *CsPH3*, was co-expressed with Nicole and Noemi to establish whether an analagous relationship is present in *Citrus*.

The inclusion of CsPH3 did not induce the activation of any promoter which was not already activated by Noemi-Nicole, such as *pCsTT12*, in contrast to the mechanism reported in *Arabidopsis* (Gonzalez et al., 2016). The only other notable observation was a slightly stronger activation of *pCsPH5* by Noemi-Nicole when co-expressed with CsPH3. The relative unimportance of *CsPH3*, at least regarding the acidity and PA pathway genes tested here was striking, since the respective WRKY homologs are either absolutely essential for, or considerably enhance, target gene expression in other species (Amato et al., 2019, Amato et al., 2016, Gonzalez et al., 2016, Verweij et al., 2016).

In most cases, *nicole^{so^{ro}}* was either unable or could activate only very weakly the promoter targets of its wild-type counterpart. Low but statistically significant levels of induction were seen in *Noemi*, *Nicole* and *PH1* promoters. These data suggest that the *nicole^{so^{ro}}* mutant has completely lost its ability to regulate transcription of *CsPH5* and *CsLDOX*. Similar loss of function in MYB5-like homologs have been documented. The *Petunia ph4-V2153* mutant also contains a transposon insertion and is associated with a loss of *PhPH1* and *PhPH5* expression (Quattrocchio et al., 2006, Verweij et al., 2008).

Chapter 5:
Generation and efficacy analysis of Nicole-
targeting multi-sgRNA CRISPR-Cas9 constructs

5.1 Introduction

5.1.1 A gene-editing approach in *Citrus*

Strong evidence of the regulatory role of Nicole in fruit acidification has been detailed in Chapters 3 and 4, based on transcriptomic and functional analyses of a naturally derived mutant *Nicole* allele in 3 acidless sweet orange varieties. However, the generation of mutant *nicole* alleles in stably transformed acidic sweet orange, using highly specific gene-editing, would complement our findings and allow further attribution of phenotypic changes to Nicole specifically, due to an otherwise identical genetic background to the original parent variety. Similarly, conventional introgression to introduce *nicole* mutants into commercially important *C. sinensis* varieties are not possible due to the sweet orange being an interspecific hybrid propagated normally by grafting and with asexual apomictic seeds.

The clustered regularly interspaced short palindromic repeats (CRISPR) system is a powerful tool that facilitates specific editing of genes. It has been demonstrated to work in a wide variety of plants, including model species *Arabidopsis* and *Nicotiana*, rice and, importantly, *Citrus* (Feng et al., 2013, Li et al., 2013, Nekrasov et al., 2013, Jia and Wang, 2014, Jia et al., 2017, Zhang et al., 2017, Xu et al., 2022). The CRISPR platform has accelerated functional characterisation of single genes in agriculture (Liu et al., 2016c). Due to both the inability to use conventional breeding techniques and notable advantages for functional analysis, a genome-editing approach was taken using CRISPR technology to generate loss-of-function *nicole* alleles in *C. limon*.

5.1.2 CRISPR-Cas9 genome-editing in plants

First described in 1987, the CRISPR immune system is widely distributed amongst prokaryotes (Ishino et al., 1987, Deveau et al., 2010, Horvath and Barrangou, 2010). Together, CRISPR loci and CRISPR-associated (Cas) genes provide defence against invasive bacteriophage genetic elements (Garneau et al., 2010). The type II CRISPR system in *Streptococcus pyogenes* consists of guide sequences that direct nuclease activity by the Cas9 nuclease towards specific genetic targets. This system has been capitalised in plants to perform highly specific gene-editing with relative design ease and efficiency in comparison

to other described editing technologies, such as transcription activator-like effector nucleases (TALENs) and zinc-finger nucleases (ZFNs).

Briefly, plants are transformed with genes encoding a Cas9 nuclease and synthetic gRNA (sgRNA), containing a 20-nt sequence specific to the target gene of interest within the host genome. Target sequences must precede a protospacer adjacent motif (PAM) which is specific for the Cas enzyme used for editing. The commonly used *S. pyogenes* Cas9 homolog requires a 5'-NGG PAM sequence immediately upstream of the 20-nt target sequence within the endogenous target DNA (Jinek et al., 2012). Possible off-targets within the genome can be identified using the BLAST which enables the best gRNA sequences to be selected. Translated Cas9 is directed specifically towards the endogenous DNA target by sgRNA, resulting in a double-stranded break and subsequent initiation of repair pathways. One repair pathway is nonhomologous end joining (NHEJ) which, with no template, can be exploited for the error-prone nature of the repair mechanism to create targeted mutations.

NHEJ can introduce random indels, potentially resulting in early stop codons or frameshifts within the coding sequence (CDS) and subsequent loss of translated protein function (Jinek et al., 2012, Ran et al., 2013). CRISPR-Cas9 gene-editing technology has facilitated the functional characterisation of MYB transcription factors in *Oryza sativa* and *Solanum lycopersicum*, elucidating an involvement in anthocyanin biosynthesis (Yan et al., 2020, Zheng et al., 2021), and identification of MYB transcription factors regulating proanthocyanidin content and trichome development in *Populus tomentosa* and *Gossypium hirsutum*, respectively (Wang et al., 2017, Shangguan et al., 2021). Homozygous mutations have also been reported within the first generation of transgenic plants, attesting the remarkable efficiency of the CRISPR-Cas9 system (Brooks et al., 2014, Zhang et al., 2014). Segregation and removal of the CRISPR-Cas9 T-DNA from transgenic lines may leave a mutant containing only the targeted mutation which may be as small as a single base pair change within the genome. This level of specificity, efficiency, and potential for minimal manipulation of the target genome is advantageous for functional analysis, enabling phenotypic changes due to complete loss of function of a single gene to be assessed, and also in a regulatory sense for future commercialisation.

As a preliminary exploration of the feasibility of editing Nicole to produce low acidity sweet oranges, CRISPR-Cas9 technology was tested for its functionality and efficacy in editing *Nicole* using stable transformation. Multiple gRNA within a single CRISPR-Cas9 construct can facilitate gene-editing at two locations, and simultaneous cleavage at both sites may result in a complete deletion of the region flanked by the two gRNA target sequences. To maximise the generation of knock-out mutations in *Nicole* two multi-sgRNA CRISPR-Cas9 constructs were developed, each containing a unique pair of sgRNAs. In total, four target sequences common to *Nicole* homologs derived from *C. sinensis*, *C. limon* and *Fortunella hindsii* were designed flanking or within the sequence encoding the first repeat of the R2R3MYB DNA binding domain (DBD) of the Nicole protein. Both CRISPR-Cas9 constructs were transformed into *C. limon* via cocultivation with *Agrobacterium rhizogenes*.

5.1.3 *Agrobacterium rhizogenes*-mediated transformation of *C. limon*

The *Fortunella* genus is used as a model for functional gene studies in *Citrus*, particularly when involving genetic transformation (Yang et al., 2007, Zhang et al., 2009, Cao et al., 2015, Yang et al., 2016). Transformation systems in Hongkong kumquat (*F. hindsii*) are well documented, and callus induction rates are amongst the highest in *Citrus* (Deng and Zhang, 1988). More recently, the CRISPR-Cas9 system has been applied successfully to the genus, expanding its potential as a model species (Zhu et al., 2019, Xu et al., 2022). *F. hindsii* benefits from a greatly reduced juvenility period, unlike 5-10 years for most *Citrus* species (Krajewski and Rabe, 1995), as short as 8 months, and is closely related to *Citrus* (Wu et al., 2018, Zhu et al., 2019). Rapid generation of fruit was important due to our interest in phenotypic effects in fruit tissue specifically in prospective *nicole* mutants.

However, due to the lack of Hongkong kumquat plant material and time constraints, an alternative approach was taken. Since *A. tumefaciens*-mediated transformation is a large undertaking (even in *Fortunella*), the selected gRNAs were first analysed for efficacy by taking advantage of hairy root transformations. Hairy root transformations using *A. rhizogenes* have been used for rapid generation of transgenic material by initiating proliferation of roots from transfected explants, particularly when characterising gene function in roots (Kereszt et al., 2007, Cao et al., 2011, Aarrouf et al., 2012). Fast transgenic root development also expedites genotypic analysis compared to conventional *A. tumefaciens* transformations. This is advantageous for determining CRISPR gRNA efficacies prior to proceeding with stable

transformation of commercial *C. sinensis* varieties or *F. hindsii*, which may take many years to produce fruit for characterisation. Here, *C. limon* epicotyls were transformed for gRNA efficacy analysis to better inform any *A. tumefaciens*-mediated transformation of *F. hindsii* by identifying the most effective gRNA combination for phenotypic analysis of fruit in the future. *C. limon* was transformed as an alternative to Hongkong kumquat due to the ample availability of fruit and, therefore, seeds in local supermarkets.

The efficiency of gRNAs in CRISPR-mediated gene editing has been tested via *A. rhizogenes*-mediated transformation in *Populus*, soybean and pea (Bruegmann et al., 2019, Di et al., 2019, Cheng et al., 2021, Li et al., 2023), but is yet to be published in *Citrus*. Further, hairy root transformations have been described in only a few *Citrus* species and relatives: *C. aurantium*, *C. aurantifolia*, *Poncirus trifoliata* (Perez-Molphe-Balch and Ochoa-Alejo, 1998, Chavez-Vela et al., 2003, Xiao et al., 2014). *A. rhizogenes* strains tested previously include MSU440, K599 and A4. K599 was observed to be weak in terms of virulence when co-cultivated with trifoliolate orange, compared to MSU440 (Xiao et al., 2014). These two strains harbouring multi-sgRNA CRISPR-Cas9 plasmids were used to transform *C. limon*, marking the first documentation of *C. limon* susceptibility to any *A. rhizogenes* strain.

5.2 Materials and Methods

5.2.1 Identification of potential gRNA targets within *Nicole*

Nicole homologs derived from *C. limon* and *F. hindsii* were first amplified by PCR, sequenced, and aligned as previously described. Candidate gRNA target sites common to *Nicole* homologs in *C. sinensis*, *C. limon* and *F. hindsii* were identified using CRISPRdirect (<https://crispr.dbcls.jp/>). CRISPRdirect also analyses specificity of the PAM sequence (NGG) plus the 20, 12 and 8 nucleotides adjacent upstream to the target sequence of interest. This was checked against the Phytozome *C. sinensis* v1.1 reference genome. Guide RNAs, including the PAM sequence, were considered as highly specific if a single target site in the gene-of-interest was identified for the 20 and 12 nucleotide sequences upstream of the PAM.

Guide RNAs were incorporated into prospective multi-sgRNA CRISPR-Cas9 plasmids in pairs, gRNA location within the endogenous gene was considered. While a point mutation at a single gRNA location may introduce a frame shift or early stop codons, altering translated amino acid sequence, a pair of gRNAs may also work in parallel to cleave a large fragment from the target. This may increase the frequency of recovering mutant KO alleles.

5.2.2 CRISPR/Cas9 plasmid construction

CRISPR-Cas9 plasmids were constructed by recombination of *Cas9*, sgRNA and *NPTII* level 1 (L1) cassettes into a host level 2 (L2) vector using Golden Gate cloning technology. Golden Gate cloning technology functions initially by amplifying genetic elements of interest, known as cassettes, and inserting them into L1 host vectors. Multiple L1 host vectors exist, each defining the final location within a L2 host vector. Ultimately, a L2 plasmid was constructed by recombining all required L1 cassettes into a L2 host vector.

First, target sequences were amplified by PCR with primers containing 5' extensions of a *BsaI* restriction enzyme recognition site and the necessary overhangs to facilitate ligation within a L1 host vector post *BsaI* digestion. NOS promoter-driven *NPTII* with OCS terminator and double CaMV 35S promoter-driven *Cas9* with NOS terminator sequences were amplified

from pICSL002203 via Phusion-mediated PCR. Overhangs on the forward and reverse primers used to amplify p2x355-Cas9-tNOS were designed to facilitate inverse ligation and orientation into the L1 host vector and, subsequently, the L2 host vector, in comparison to the *NPTII* and sgRNA direction of transcription, similar to Zhang et al. (2017). Synthetic gRNAs were synthesised by Phusion-mediated PCR using the sgRNA scaffold vector pICSL70001 as template DNA, mutually exclusive forward primers containing the new respective *Nicole* gRNA target sequence of interest (EA-009, EA-025, EA-027 and EA-028; Supplementary Table 4), and a universal reverse primer (EA-007; Supplementary Table 4).

L1 constructs were then generated by recombination of PCR products into L1 host vectors (Figure 5.2.1). The pNOS-*NPTII*-tOCS and inverse p2x355-Cas9-tNOS PCR products were cloned into L1 host vectors pICH47732 (position 1), pICH47742 (position 2), respectively, via one-tube *BsaI* digestion-ligation reactions. Together, sgRNA and pICSL90001, the U6 promoter sequence donor, were cloned into L1 host vectors pICH47751 (position 3) or pICH47761 (position 4) via a one-tube *BsaI* digestion-ligation reaction. During the reaction, the *lacZ* gene and both *BsaI* recognition sites in the respective L1 host vector are replaced by the desired genetic elements, allowing blue-white screening of positive DH5 α *E. coli* colonies (Messing et al., 1977, R  ther, 1980). The digestion-ligation reactions contained an insert to vector molar ratio of 3:1 (Table 5.2.1).

L2 plasmids were then constructed by the digestion and recombination of 4 L1 cassettes (*NPTII*, *Cas9*, and an sgRNA pair) and end-linker plasmid pICH47180 into the L2 host vector pAGM4723 (Figure 5.2.2). This was achieved via a one-tube *BpiI* digestion-ligation reaction containing an insert to vector molar ratio of 3:1 (Table 5.2.1). During the reaction, the cytotoxic *ccdB* gene and both *BpiI* recognition sites in pAGM4723 were replaced by the L1 cassettes which allowed selection for positive DH5 α (a *ccdB*-sensitive strain) *E. coli* colonies (Bernard and Couturier, 1992, Bernard et al., 1994). An additional L2 plasmid was constructed by cloning the pNOS-*NPTII*-tOCS cassette only and end-linker plasmid pICH41722 to perform as a negative control during plant transformations. L1 and 2 digestion-ligation reaction thermocycler parameters were identical and are detailed in (Table 5.2.2). L2 insertions were checked via *DraIII* restriction digestion and sequencing. Plasmid maps of all constructed CRISPR-Cas9 plasmids are presented in the results section of this chapter.

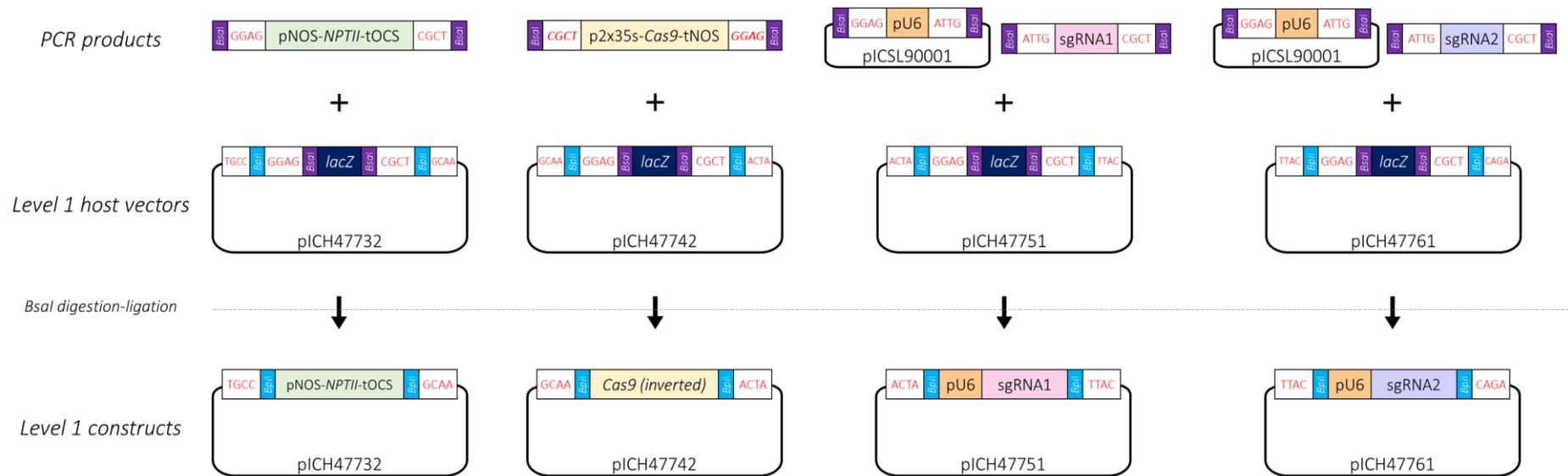


Figure 5.2.1 Simplified schematic view of construction of Level 1 (L1) plasmids using Golden Gate cloning technology. PCR products were first amplified with primers containing 5' extensions of a Bsal recognition site and appropriate overhangs for subsequent digestion-ligation reaction with a L1 host vector. During the Bsal digestion-ligation reaction the selective marker lacZ and Bsal recognition sites were cleaved from the L1 host vector. Once ligation of PCR products and L1 host vector had occurred there were no Bsal recognition sites remaining, allowing a one-tube digestion-ligation reaction. Red text denotes sense-strand overhangs following digestion. The Cas9 cassette was ligated inversely by designing the forward and reverse PCR primer overhang additions appropriately (bold italicised red text). Bsal and BpiI restriction enzyme recognition sites are coloured purple and blue, respectively.

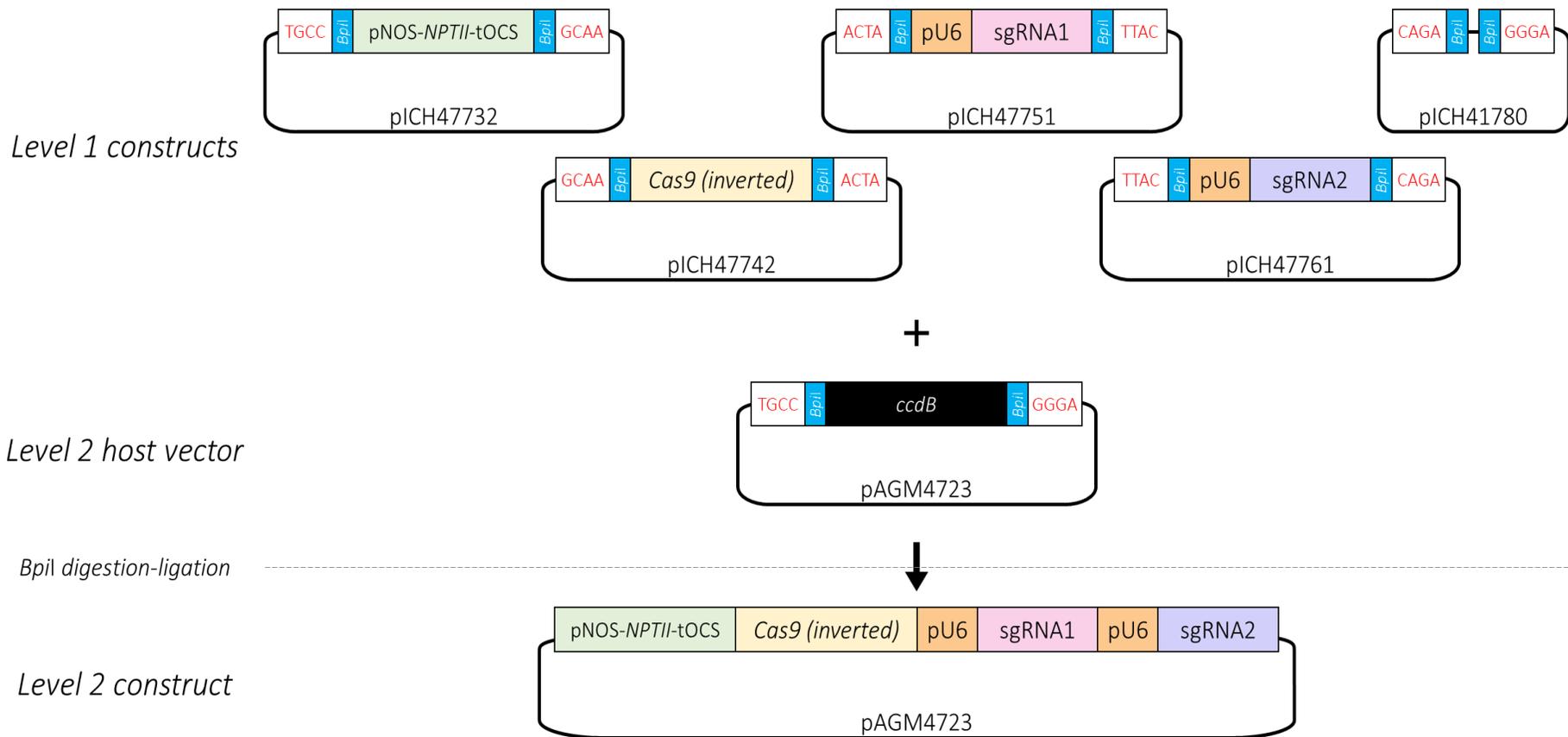


Figure 5.2.2 Simplified schematic view of construction of Level 2 (L2) plasmids using Golden Gate cloning technology. Level 1 (L1) cassettes were recombined into a L2 host vector by performing a one-tube digestion-ligation reaction with all L1 plasmids and a L2 host vector. There were 6 positions within the pAGM4723 cloning site. An end-linker was included when recombining less than 6 L1 cassettes. During the BpiI digestion-ligation reaction the selective marker ccdB and BpiI recognition sites were cleaved from the L2 host vector. Once ligation of L1 cassettes and L2 host vector had occurred there were no BpiI recognition sites remaining, allowing a one-tube digestion-ligation reaction. Red text denotes sense-strand overhangs following digestion. BpiI Restriction enzyme recognition sites are coloured blue.

Table 5.2.1 Golden Gate digestion-ligation composition per 15 μ l reaction (a: Level 1 assembly; b: Level 2 assembly; c: Insert DNA required for a 3:1 insert to vector molar ratio was calculated per reaction per insert size).

Component	15 μl reaction
Vector DNA	100 ng
All insert DNA	variable ^c
10X T4 DNA Ligase Buffer	1.5 μ l
10X Bovine Serum Albumin	1.5 μ l
<i>Bsa</i> I-HF ^a / <i>Bpi</i> I-HF ^b	20 units
T4 DNA Ligase	0.5 μ l
H ₂ O	to 15 μ l

Table 5.2.2 Golden Gate Level 1 and Level 2 digestion-ligation thermocycler parameters.

Step	Temperature ($^{\circ}$C)	Time
27 cycles:		
Digestion	37	3 min
Ligation	16	4 min
Final digestion	37	5 min
Inactivation	80	5 min

5.2.3 Hairy root transformation of *C. limon*

To test gRNA efficacies, hairy root transformations of *C. limon* epicotyls were conducted, as described by Xiao et al. (2014) in trifoliate orange with some alterations, with three *A. rhizogenes* strains. *C. limon* was selected due to the abundance of seeds in fruit in comparison to most commercially available *C. sinensis* varieties. Primofiori *C. limon* fruit were obtained from UK supermarket Tesco for seed collection and germination for subsequent *A. rhizogenes*-mediated transformation of epicotyl tissue. Seeds were sterilised by removing the external seed coat, washed with 30 % bleach for 30 minutes and four washes with sterile dH₂O. Aseptic seeds were soaked in dH₂O overnight and sown by placement on top of solid MS3 medium (Niedz, 2008). To maximise epicotyl tissue growth for transformation, the seeds were incubated in the dark for 4 weeks. Etiolated seedlings were then transferred into 16 h light / 8 h dark conditions for a week prior to transformation (Tan et al., 2009).

Overnight cultures (10 ml) of *A. rhizogenes* strains K599, ATCC15834 and MSU440 harbouring either no recombinant plasmid DNA or a multi-sgRNA CRISPR-Cas9 plasmid were centrifugated at 4,000 rpm for 15 minutes. Pelleted cells were resuspended in 10 ml TY medium containing 100 µM acetosyringone and incubated for 2 h at 28 °C, 220 rpm. Aseptic epicotyls, obliquely cut into ~ 1.5 cm segments, were immersed in *A. rhizogenes* TY cultures adjusted to OD₆₀₀ = 0.7 for 20 min with gentle agitation. As a control, some epicotyls were also immersed in sterile TY containing no *A. rhizogenes* cells.

The explants were blotted dry on sterile filter paper and co-cultivated in the dark on solid co-cultivation media. After 3 days of cocultivation, the epicotyls were soaked in sterile dH₂O containing 400 mg L⁻¹ cefotaxime for 5 min, followed by 5 washes in sterile dH₂O. Hairy root growth was induced by culturing explants on solid root induction media with and without 50 mg L⁻¹ kanamycin selection. Transformation efficiency was determined by calculating the percentage of epicotyls producing roots on selective media. After 8 weeks, root tips were excised and cultured independently on the same medium as a root stock. Up to 100 mg tissue was sampled and flash frozen in liquid N₂.

5.2.4 Hairy root DNA extraction

DNA was extracted from ≤ 100 mg of frozen homogenised root tissue with the DNeasy Plant Mini Kit (Qiagen), according to the manufacturer's protocol. Briefly, homogenised tissue was incubated in a lysis buffer and centrifuged to remove cell debris and other precipitates. The lysate was mixed with a binding buffer and ethanol and applied to a DNeasy spin column to bind DNA to the membrane. Following two wash steps the DNA was eluted with H₂O.

5.2.5 Identification of CRISPR-Cas9-edited alleles

Exon 1 of CRISPR-Cas9 target gene *Nicole* was amplified via Phusion-mediated PCR to identify point mutations via TOPO cloning and sequencing, or large deletions observable via gel electrophoresis. TOPO cloning enabled the isolation of multiple PCR amplicons when this could not be achieved by excision and extraction from agarose gels due to similarity in bp size. PCR purified samples were processed using the Zero Blunt TOPO cloning Kit (Invitrogen) according to the manufacturer's protocol. This method ligates blunt-end products derived from Phusion-mediated PCRs into the pCR™-Blunt II-TOPO vector, disrupting expression of the cytotoxic gene *ccdB* (Bernard and Couturier, 1992, Bernard et al., 1994). Following *E. coli* transformation, all amplicons within the PCR product sample were sequenced by extracting DNA individually from multiple colonies. Plasmid DNA extraction was performed on up to 6 colonies per *C. limon* PCR when possible.

5.2.6 Analyses of CRISPR-Cas9-edited alleles

Sequenced alleles were aligned to the wild-type *Nicole* gene sequence from *C. limon* as previously described, but with manual adjustments. Independent TOPO clones may contain duplicate PCR products. To avoid pseudo-replication, unique alleles sequenced per root are presented only. Translated amino acid sequences were searched using the NCBI conserved domain web tool (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) for Pfam domains to identify loss of predicted R2MYB repeats (E-value = 0.01). Guide RNA efficacy was determined based on percentage of gRNA-specific nucleotide edits and predicted loss of function.

5.3 Results

5.3.1 Identification and selection of gRNAs

The *Nicole* gene sequence was amplified from *C. limon* and *F. hindsii* genomic DNA and sequenced. Two alleles were identified in *C. limon*, differing primarily in the number of GCA repeats in exon 1. All homologs cDNA sequences from these species are extremely high in identity, the lowest pairwise comparison being 95.6% between *CsNicole* and *CINicole*.2. An alignment of exon 1 of each species' *Nicole* alleles was used to cross-reference gRNA targets found within the *Nicole* sequence derived from *C. sinensis* and identify gRNA sequences common to all four alleles. A total of 125 target sequences were found within the genomic sequence of sweet orange *Nicole*, of which 48 were found to be highly specific (Table 5.3.1). Across all *C. limon* and *F. hindsii* *Nicole* alleles 35 of these 48 *C. sinensis* gRNA target sites are present, including any directly upstream 5'-NGG PAM sequence.

My objective was to generate complete loss of function alleles. Consequently, the 23 gRNAs located within exon 1 and in close proximity to the R2MYB repeat encoding region were considered. Ultimately, gRNAs 002, 003, 011 and 019 were selected and paired (Figure 5.3.1). Together, gRNAs 002 and 019 flank the R2MYB repeat encoding region and simultaneous cleavage may result in complete loss of the repeat. This pair of gRNAs, designated as gRNA pair 1, are situated 209 bp apart. Guide RNA pair 2 comprise gRNAs 003 and 011, which are distanced 121 bp apart. Like gRNA_002, gRNA_003 is located shortly upstream of the R2MYB repeat nucleotide sequence. However, gRNA_011 is found in the centre of the R2MYB repeat encoding region. Theoretically, gRNA pair 2 can only facilitate the complete deletion of half of the R2MYB repeat sequence. However, both gRNA within this pair may introduce stop codons that could directly interfere with the transcription of this DBD region, as opposed to only one in gRNA pair 1.

5.3.2 Construction of multi-sgRNA CRISPR-Cas9 plasmids

Synthetic gRNAs were successfully amplified using pICSL90001, containing the sgRNA scaffold sequence, as a DNA template. Sequencing confirmed introduction of gRNA targets 002, 003, 011 and 019 into the scaffold sequence and the presence of flanking *BsaI*

recognition sites. Likewise, pNOS-*NPTII*-tOCS and p2x35S-*Cas9*-tNOS were amplified via PCR but used pICSL002203 as template. L1 plasmids were successfully constructed and L1 cassettes subsequently recombined together to construct two multi-sgRNA CRISPR-Cas9 constructs containing gRNA pair 1 (pEA13; Figure 5.3.2) and 2 (pEA15). In addition, a negative control plasmid containing only pNOS-*NPTII*-tOCS and no CRISPR-Cas9 genetic elements was constructed (pEA16; Figure 5.3.3).

All plasmids contained the correct ligation of L1 cassettes, confirmed via restriction digestion and gel electrophoresis (Figure 5.3.4). There are two *DraIII* recognition sites flanking the pAGM4723 insertion site. As a result, *DraIII* digestion produces bands that accurately represent the L1 cassette insertions and backbone base pair size. All constructed plasmids, pEA13, pEA15 and pEA16 produced identical backbone bands to the pAGM4723 negative control following digestion. A band approximately 1,687 bp was observed in pAGM4723, reflecting the size of the *ccdB* gene sequence that is replaced during the Golden Gate *BpiI* digestion-ligation reaction. The two multi-sgRNA CRISPR-Cas9 constructs pEA13 and pEA15 contain the anticipated insertion of 7,674 bp, whereas only 1,929 bp were introduced in the *NPTII*-only plasmid pEA16. Sequencing of the entire insertion in each plasmid confirmed ligation of the correct sequence and, importantly, gRNA sequences within the sgRNAs. Furthermore, transformation of each *A. rhizogenes* strain was confirmed via colony PCR.

5.3.3 *A. rhizogenes* transformation efficiency in *C. limon*

Very high germination rates of sterilised *C. limon* seeds were observed, providing ample epicotyl tissue. This was exacerbated by 4 weeks of incubation in the dark, greatly expanding epicotyl length. Obliquely cut *C. limon* epicotyls were co-cultivated with *A. rhizogenes* cells harbouring pEA13, pEA15, pEA16 or no recombinant plasmid. This was performed independently in parallel with three *A. rhizogenes* strains: K599, MSU440 and ATCC15834. As a control, some epicotyls were also immersed in TY medium but without *A. rhizogenes* cells. The transformation efficiencies observed are detailed in Table 5.3.2 and the typical explant conditions after 8 weeks are displayed in Figure 5.3.5. Under control conditions, where epicotyls were

Table 5.3.1 Highly specific gRNA target sequences present in *CsNicole*. PAM sequences are flanked by square brackets. Sense or antisense strand gRNA are denoted by + and -, respectively. Position numbers are relative to the start ATG codon. Red sequences are not unanimously shared by all *C. sinensis*, *C. limon* and *F. hindsii* *Nicole* homologs.

Name	Sequence + [PAM]	Strand	Start	End
gRNA_001	AAGAAGATAACAAGACGATG [AGG]	+	-17	6
gRNA_002	GCAACATGGCGTGCTCTTAT [TGG]	-	59	81
gRNA_003	GAGCACGCCATGTTGCAGCA [AGG]	+	66	88
gRNA_004	ACGCCATGTTGCAGCAAGGT [AGG]	+	70	92
gRNA_005	AGCAAGGTAGGGTTAAAGAG [AGG]	+	82	104
gRNA_006	GCAAGGTAGGGTTAAAGAGA [GGG]	+	83	105
gRNA_007	GAGAGGGCCATGGACGCCAG [AGG]	+	99	121
gRNA_008	GCCAGAGGAAGACGAGCTTC [TGG]	+	114	136
gRNA_009	TACATCAATAAAGAAGGCCGA [AGG]	+	142	164
gRNA_010	ATAAAGAAGGCCAAGGCCGG [TGG]	+	149	171
gRNA_011	GGTGGCGAACTCTGCCAAAA [CGG]	+	167	189
gRNA_012	CGAACTTGCCAAAACGGGC [CGG]	+	172	194
gRNA_013	CGGAGCAATCCGGCCCGTTT [TGG]	-	181	203
gRNA_014	CGGGCCGGATTGCTCCGCTG [CGG]	+	187	209
gRNA_015	CTTGCCCGACGGAGCAATC [CGG]	-	191	213
gRNA_016	GCCGGCAACTCTTGCCGCAG [CGG]	-	201	223
gRNA_017	CTGAGACCCCTCCGTTAAACG [AGG]	+	241	263
gRNA_018	ATATGTCCTCGTTTAAACGGA [GGG]	-	247	269
gRNA_019	GATATGTCCTCGTTTAAACG [AGG]	-	248	270
gRNA_020	GGCGATATGTCTCGTTTAA [CGG]	-	251	273
gRNA_021	CTTCGCCTACATCGCCTTCT [CGG]	+	295	317
gRNA_022	TACATCGCCTTCTCGGTAAC [CGG]	+	302	324
gRNA_023	CTTATTACCGGTTACCGAGA [AGG]	-	309	331
gRNA_024	TAATAACCCACGAATACCTT [AGG]	+	336	358
gRNA_025	AATAACCCACGAATACCTTA [GGG]	+	337	359
gRNA_026	TGGAACCCTAAGGTATTCG [TGG]	-	343	365
gRNA_027	TTGGTTTGTACTTCAGAGTT [AGG]	+	391	413
gRNA_028	TTATAGATGGTCTCTGATAG [CGG]	+	456	478
gRNA_029	ATAGATGGTCTCTGATAGCG [GGG]	+	458	480
gRNA_030	GATGGTCTCTGATAGCGGGG [AGG]	+	461	483
gRNA_031	TCTGATAGCGGGGAGGATTC [CGG]	+	468	490
gRNA_032	TGATAGCGGGGAGGATTCGG [GGG]	+	470	492
gRNA_033	GGCTTATCAGCTTCTTACTC [AGG]	-	531	553
gRNA_034	AGTAAGAAGCTGATAAGCCA [AGG]	+	535	557
gRNA_035	TTCAATGGCTTATGAGTTCT [TGG]	-	565	587
gRNA_036	ACTCCAATGACCGTTTCATC [TGG]	+	691	713
gRNA_037	TGACCAGATGAAACGGTCAT [TGG]	-	694	716
gRNA_038	CATTCTACCACAGCCAGCAG [AGG]	-	734	756
gRNA_039	TGATCATCAACTAAAGCGTT [CGG]	-	799	821
gRNA_040	AATTCGATAAACTCGTATAC [GGG]	-	858	880
gRNA_041	CAATTCGATAAACTCGTATA [CGG]	-	859	881
gRNA_042	TGGCTCATCATCGGATCCTT [TGG]	+	1131	1153
gRNA_043	GCCGCTGCAGTCGAAACCAA [AGG]	-	1147	1169
gRNA_044	ACTGCAGCGGCATCAACTTT [TGG]	+	1159	1181
gRNA_045	GAGATTCACAGTTTGCTTCA [AGG]	-	1182	1204
gRNA_046	CTTTGAACCAAGATGAGTCC [AGG]	+	1220	1242
gRNA_047	TGAACCAAGATGAGTCCAGG [AGG]	+	1223	1245
gRNA_048	GAACCAAGATGAGTCCAGGA [GGG]	+	1224	1246

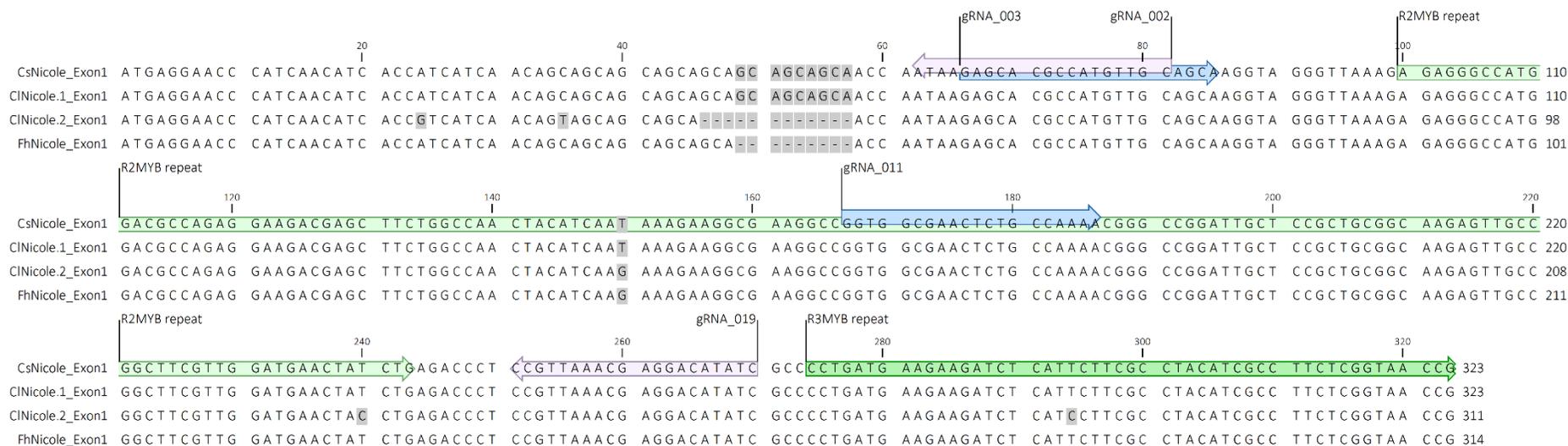


Figure 5.3.1 Annotated exon 1 nucleotide sequences of *C. sinensis*, *C. limon* and *F. hindsii* Nicole alleles. Guide RNA pairs 1 (gRNA_002 and 019) and 2 (gRNA_003 and 011) are coloured pink and blue, respectively. The R2 and R3MYB repeat encoding regions are annotated light and dark green, respectively. Arrows represent strand-sense of annotation. Disrupted arrowhead indicates that the annotation continues past the end of sequence. Different residues are highlighted in grey.

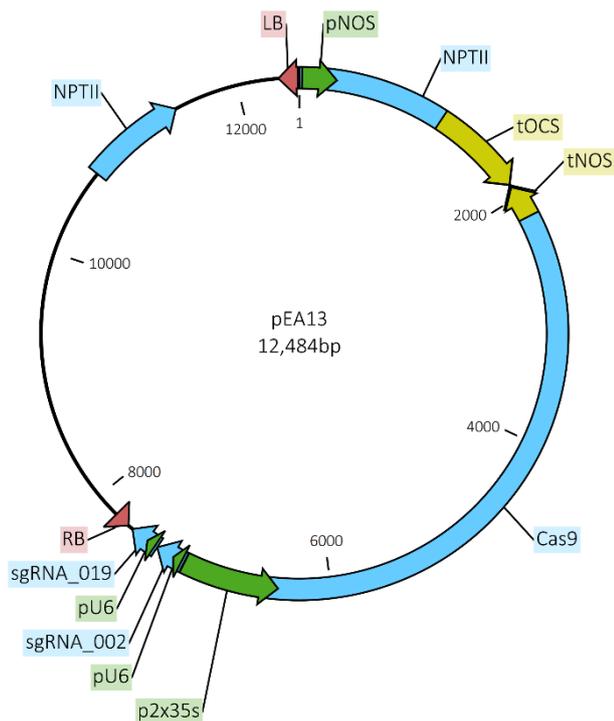


Figure 5.3.2 Multi-sgRNA CRISPR-Cas9 plasmid map of pEA13. Arrows denote the direction of transcription. Borders, promoters, CDS and terminators are annotated red, green, blue, and yellow, respectively. pEA15 is identical except for containing gRNA pair 2 (gRNA_003 and 011) instead of gRNA pair 1 (gRNA_002 and 019).

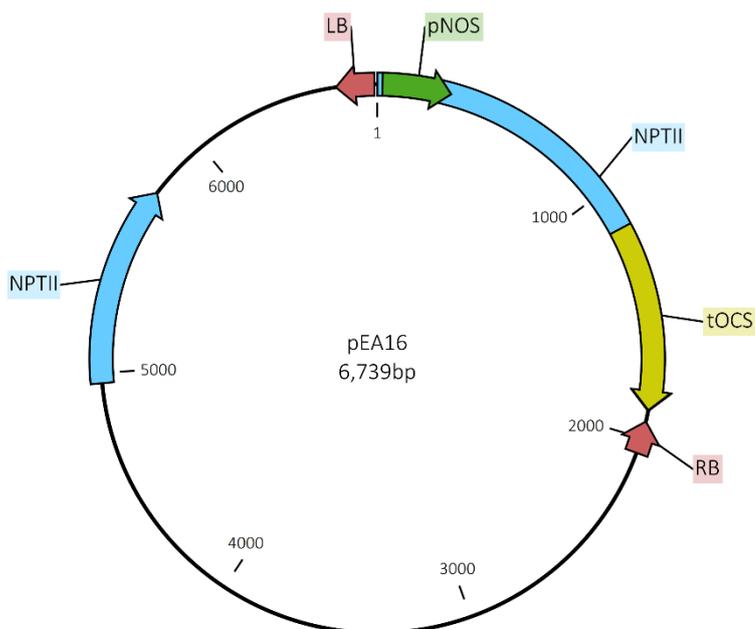


Figure 5.3.3 Negative control plasmid map of pEA16 containing no CRISPR-Cas9 genetic elements. Arrows denote the direction of transcription. Borders, promoters, CDS and terminators are annotated red, green, blue, and yellow, respectively.

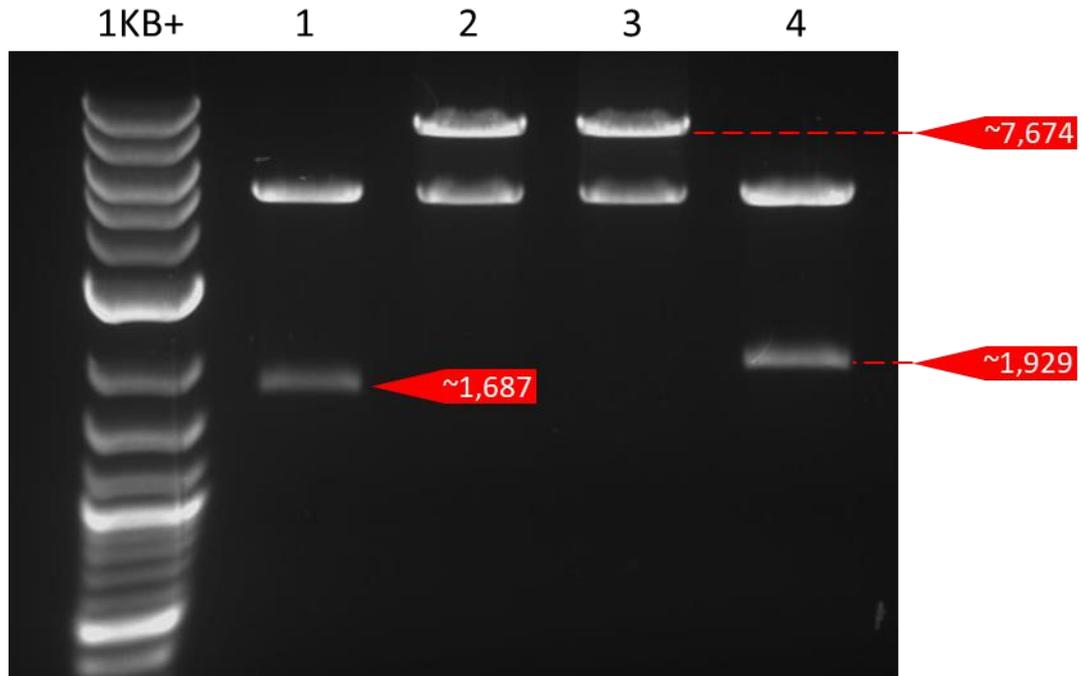


Figure 5.3.4 *DralIII* restriction digestion of L2 CRISPR-Cas9 plasmids. 1KB+: 1KB plus ladder (NEB); 1: pAGM4723; 2: pEA13; 3: pEA15; 4: pEA16. Red arrows indicate approximate band base pair size at the respective position.

immersed in TY medium lacking *A. rhizogenes* cells, there was no development of roots from explants with and without selection. My results also indicated that *C. limon* was not susceptible to K599-mediated transformation as roots did not develop under any conditions tested (data not shown). Where root development was not observed under any treatment explants developed necrosis on wounded epicotyl surfaces.

Conversely, where root induction was seen, it was preceded by callus growth on excised surfaces, usually within 2 – 3 weeks. Root growth typically followed between weeks 6 and 8. Positive results were observed in all MSU440 and ATCC15834 transformations of *C. limon* when harbouring either pEA13, pEA15 or pEA16. Notably, kanamycin selection prevented the induction of roots when transformed with these two strains when lacking a recombinant plasmid. The number of roots per epicotyl varied widely, ranging from one, in many cases, to a large proliferation of > 5. Callus development with no subsequent root growth was also common.

Transformation efficiencies reflected the percentage of epicotyls developing at least one root. Non-transformed ATCC15834 and MSU440 offered transformation efficiencies of 82% and 62%, respectively, when culturing explants without kanamycin. However, higher proportions of root producing epicotyls were observed in MSU440 when harbouring one of the three CRISPR-Cas9 plasmids (44%), compared to ATCC15834 (31%). For both *A. rhizogenes* strains, transformation efficiency was up to 75% lower when transforming *C. limon* with cells containing recombinant DNA on selection, compared to non-transformed cells without selection.

Table 5.3.2 Transformation efficiency of *C. limon* using *A. rhizogenes* strains K599, ATCC15834 and MSU440 after ≥ 8 weeks with and without 50 mg L⁻¹ kanamycin selection.

Strain	Plasmid	Selection	Total epicotyls	With roots	Efficiency (%)
-	-	+	33	0	0
	-	-	100	0	0
ATCC15834	-	+	18	0	0
	-	-	33	27	82
	pEA13	+	50	15	20
	pEA15	+	50	14	28
	pEA16	+	50	18	36
	-	+	12	0	0
MSU440	-	-	37	23	62
	pEA13	+	127	52	40
	pEA15	+	118	45	38
	pEA16	+	130	69	53
	-	+	12	0	0

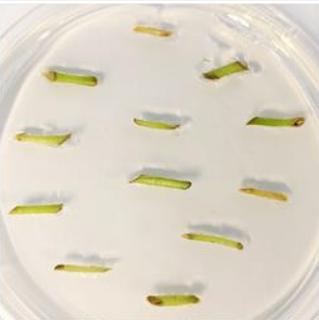
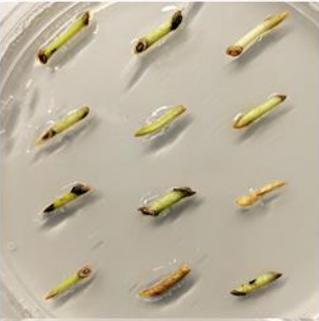
Strain	Plasmid	Selection	Example plate
-	-	-	
MSU440	-	+	
K599	pEA13	+	
MSU440	pEA13	+	
ATCC15834	pEA15	+	

Figure 5.3.5 Typical hairy root development ≥ 8 weeks after *A. rhizogenes* transformation treatment of *C. limon* epicotyls, with and without 50 mg L^{-1} kanamycin selection.

5.3.4 Cas9 nuclease activity derived gene edits

Guide RNA efficacy has been analysed in ATCC15834 transformations of *C. limon*. DNA was extracted from roots where sufficient material was available. Individual roots originating from the same epicotyl were extracted independently. Exon 1 of *CINicole* was then amplified from root and *C. limon* juice DNA and visualised via gel electrophoresis. PCR products were TOPO cloned into pCR™-Blunt II-TOPO vector to isolate all alleles and were sequenced. Due to the nature of TOPO cloning many independent plasmid clones contain copies of the same PCR product. To avoid pseudoreplication, only unique alleles sequenced per root were considered for alignment and gRNA efficacy analysis. Samples are named in the following format: “1a.2”; where “1a” specifies root A from epicotyl 1, and “.2” refers to the TOPO clone number.

Gel electrophoresis allowed visualisation of large-scale edits attributed to Cas9 nuclease activity. This revealed multiple cases of simultaneous cleavage, resulting in large deletions corresponding to the distance between the respective gRNA pair in *CINicole*. For example, bands approximately 244 bp were amplified from pEA13-transformed samples 1c and 2a Figure 5.3.6. Likewise, transformation with pEA15 resulted in mutant alleles approximately 332 bp long in samples 6a and 10a. This provided the first indication of successful CRISPR-Cas9-mediated gene editing of *Nicole*.

Root *CINicole* allele sequencing informed gRNA efficacy analysis further as small-scale mutations were not distinguishable via gel electrophoresis. Unique TOPO cloned sequences were aligned to the respective wild-type *CINicole* allele for analysis. A range of indel mutations, varying in sizes, were observed in exon 1 of *CINicole* as a consequence of transformation with either pEA13 or pEA15.

Sequenced alleles amplified from *C. limon* transformed with pEA13, aligned to *CINicole.1* and *CINicole.2*, are presented in Figure 5.3.8 and Figure 5.3.9, respectively. Gene edits were observed at both gRNA (002 and 019) target sites, often appearing as single nucleotide indels. Larger site-specific deletions occurred, such as in samples 1c.3 and 5a.1. Further, entire deletion of the region flanked by the gRNA sequences was observed in 7a.3. Notably, simultaneous cleavage appears to have taken place in 1c.2, 5a.5 and 6b.3 without the loss of

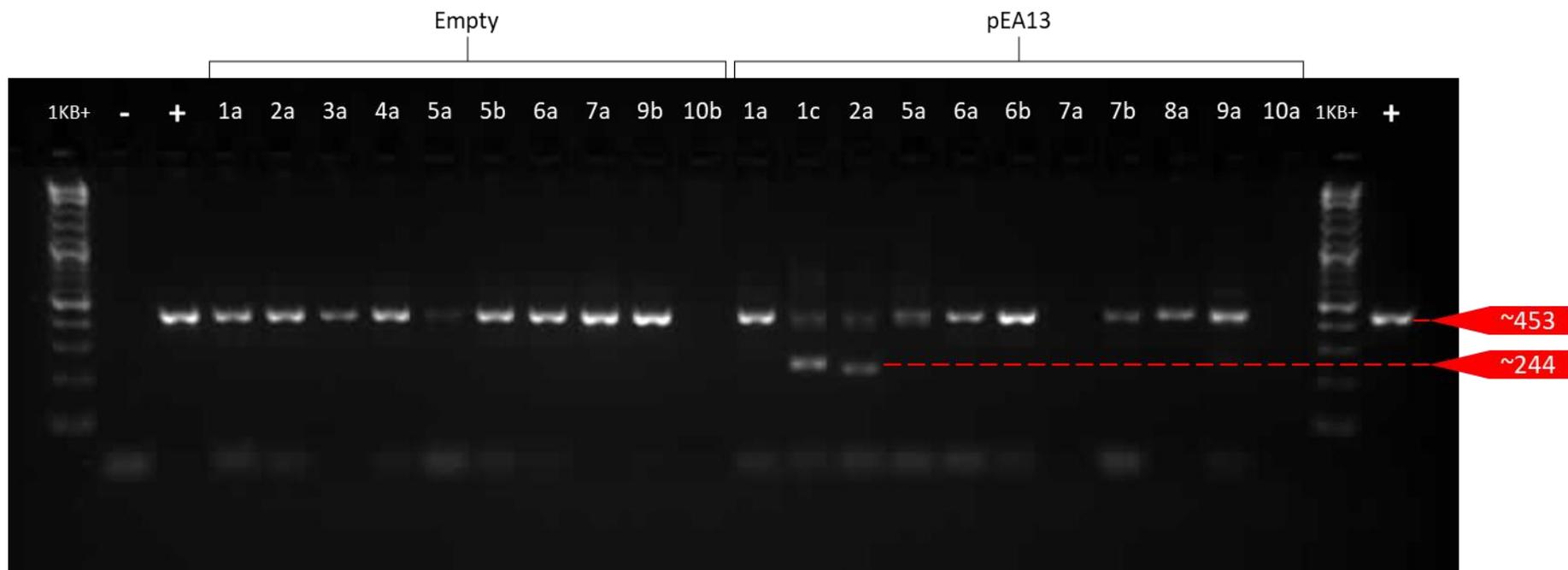


Figure 5.3.6 PCR amplification of *ClNicole* exon 1 from hairy root DNA extracted from *C. limon* transformed by ATCC15834 *A. rhizogenes* cells harbouring pEA13 or no recombinant plasmid (empty). 1KB+: 1KB plus ladder (NEB); - and + denote no DNA and *C. limon* juice DNA template, respectively. Numbers and letters denote individual explants and roots, respectively. Red arrows indicate approximate band base pair size at the respective position.

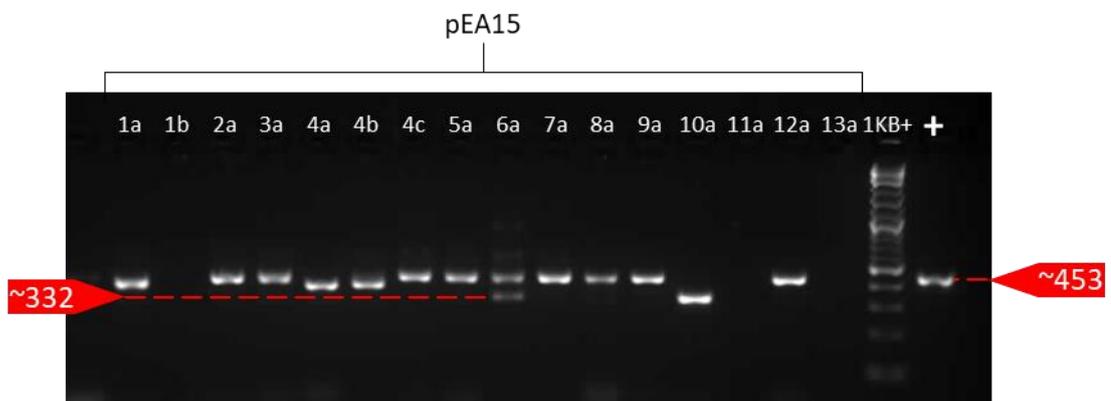
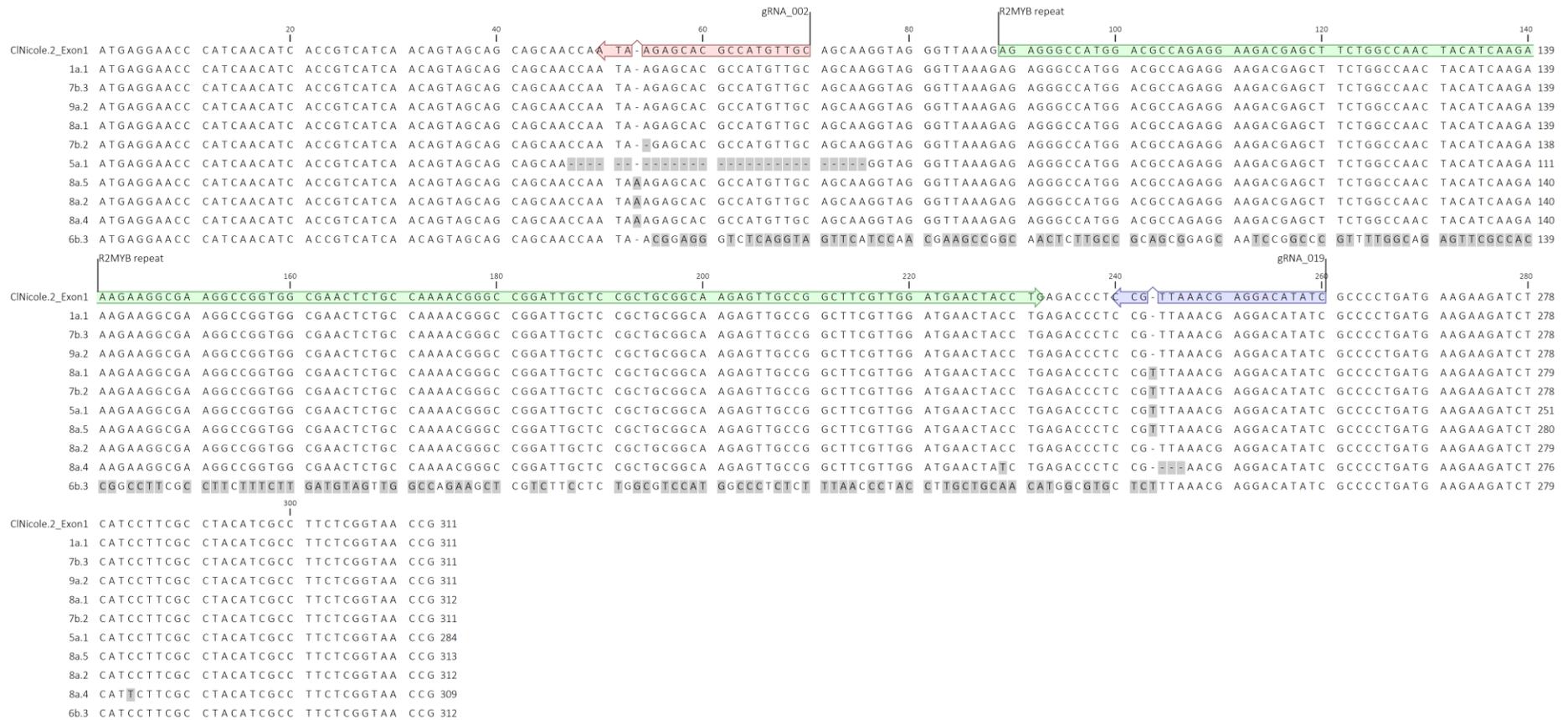


Figure 5.3.7 PCR amplification of *CINicole* exon 1 from hairy root DNA extracted from *C. limon* transformed by ATCC15834 *A. rhizogenes* cells harbouring *pEA15*. 1KB+: 1KB plus ladder (NEB); + denotes *C. limon* juice DNA template. Numbers and letters denote individual explants and roots, respectively. Red arrows indicate approximate band base pair size at the respective position.



Figure 5.3.8 TOPO clone sequencing of Nicole exon 1 alleles, derived from pEA13-transformed *C. limon*, aligned to the CINicole.1 allele. R2MYB repeat encoding sequence, gRNA_002 and gRNA_019 are annotated green, red, and blue, respectively. Arrows represent strand-sense of annotation. Different residues are coloured grey. Sequences were sorted by identity to CINicole.1 allele reference sequence.



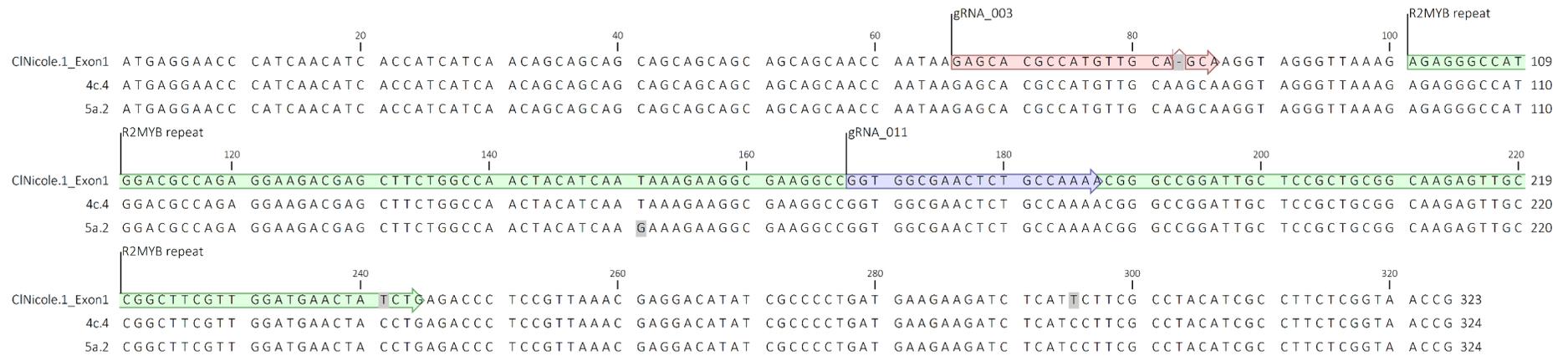


Figure 5.3.10 TOPO clone sequencing of Nicole exon 1 alleles, derived from pEA15-transformed *C. limon*, aligned to CINicole.1 allele. R2MYB repeat encoding sequence, gRNA_003 and gRNA_011 are annotated green, red, and blue, respectively. Arrows represent strand-sense of annotation. Different residues are coloured grey. Sequences were sorted by identity to CINicole.1 allele reference sequence.



the sequence between. The region between both gRNA sites in these samples contains the reverse complement of the original sequence. In comparison, Cas9-activity was dominated by one gRNA, 003, in pEA15 transformations (Figure 5.3.10 and Figure 5.3.11). Similar to pEA13, site-specific indels occurred up to 18 bp in size. In addition, complete deletion of the region flanked by gRNA_003 and gRNA_011 was observed in samples 6a.1 and 10a.2.

In general, different roots originating from the same epicotyl contained different gene-edits. However, within a single root DNA sample, multiple gene-edits were also observed. For example, 4 different mutant alleles were sequenced in pEA13-transformed *C. limon* 8a samples (Figure 5.3.9). The most common type of mutation was 1 bp insertion, accounting for 50% overall, followed by 1 bp deletions at 26.3% (Table 5.3.3).

Table 5.3.3 Mutation rates of each type attributed to either pEA13, pEA15 or overall. Total deletion and reverse complement refer to complete loss of the region flanked by both gRNA target sites within a gRNA pair, and reinsertion of the reverse complement sequence of this region, respectively.

Mutation type	pEA13 (%)	pEA15 (%)	Total (%)
1 bp insertion	46.7	52.2	50.0
≥ 1 bp insertion	0.0	4.3	2.6
1 bp deletion	6.7	4.3	5.3
≥ 1 bp deletion	20.0	30.4	26.3
Total deletion	6.7	8.7	7.9
Reverse complement	20.0	0.0	7.9

5.3.5 Guide RNA efficacy and protein functionality analysis

Exon 1 of sequenced root alleles were translated and aligned to analyse potential loss of function. Amino acid sequences were searched for Pfam domains using the NCBI conserved domain web tool. Mutant alleles were provisionally determined as knock-out mutants if the R2MYB repeat, present in wild-type *CINicole*, was lost. Guide RNA efficacy was evaluated in

terms of number of individual and paired gRNA mutation rates and proportion of subsequent translated protein loss of function.

Translation of exon 1 of pEA13- (Figure 5.3.13 and Figure 5.3.12) and pEA15-transformed (Figure 5.3.14 and Figure 5.3.15) *C. limon* alleles revealed a host of protein mutations, derived from frame shifts, introduced early stop codons, and combinations of both. For example a pEA13-derived allele containing a 17 bp mutation at only one gRNA target site (gRNA_002), in sample 1c.3, resulted in both a frame shift prior to the R2MYB repeat encoding region and, shortly after, an early stop codon. Of course, the complete deletion of the R2MYB repeat nucleotide sequence in sample 7a.3 by gRNA pair 1 would have caused total loss of the R2MYB repeat, despite keeping the following translation in frame. Alleles containing nucleotide edits at gRNA_019 target site only, downstream of the R2MYB repeat encoding sequence, certainly have no impact on translation of the R2MYB repeat region.

Translated mutant alleles from pEA15-transformed *C. limon* also displayed a multitude of mutations leading to failure to identify a significant Pfam R2MYB repeat hit. There were, however, instances where edits at the gRNA_003 target site alone removed up to 33 bp and, at most, 3 amino acids from the start of the R2MYB repeat (samples 1a.2, 3a.2 and 4a.1; Figure 5.3.15). Considering this, and that translation remains in frame, a significant R2MYB repeat was still identified by the NCBI conserved domain web tool.

Overall, 85% of pEA15-transformed *CINicole* alleles containing CRISPR edits were determined as knockout mutants, in comparison to 91% of those derived from pEA13-transformed *C. limon* (Table 5.3.4). In contrast to the knockout efficiency of each CRISPR-Cas9 plasmid the percentage of alleles containing a CRISPR edit, at either gRNA target site within the pair, was higher in pEA15 (91%) than pEA13 (69%). However, considering mutation rates at individual pEA15 gRNA sites, 91% of alleles were edited at the site of gRNA_003 in contrast to 23% at gRNA_011. Guide RNA_019 in pEA13, however, facilitated mutations in 56% of alleles sequenced.

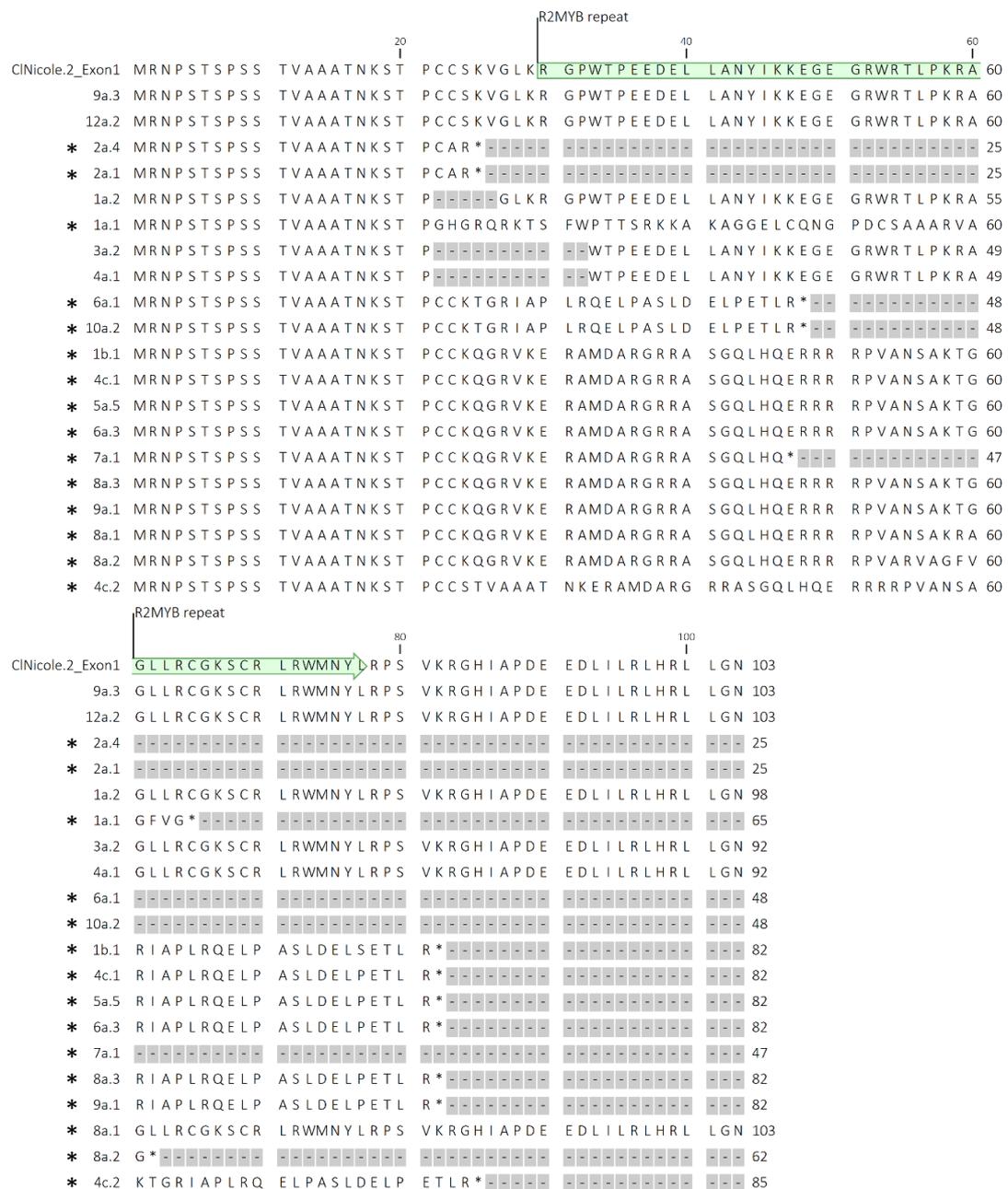


Figure 5.3.15 Amino acid sequence translated from TOPO clone sequencing of Nicole exon 1 alleles, derived from pEA15-transformed *C. limon*, aligned to CINicole.2. R2MYB repeat is annotated green on the CINicole allele 2 reference sequence. Arrow represents direction of translation. Residue gaps are coloured grey. Sequences were sorted by identity to reference sequence. Asterisk indicates loss of Pfam R2MYB repeat.

Table 5.3.4 Guide RNA efficacy in terms of gRNA site or gRNA pair mutation rate and in terms of percentage of recovered edited mutant alleles provisionally determined as knockout mutants.

Guide RNA	Mutation rate (%)	Knockout efficiency (%)
002	63	-
019	56	-
Pair 1 (pEA13)	69	91
003	91	-
011	23	-
Pair 2 (pEA15)	91	85

5.4 Discussion

5.4.1 Multi-species compatibility of *Nicole*-targeting gRNAs

Identification of viable gRNA target sites mutual to *C. sinensis*, *C. limon* and *F. hindsii* was facilitated by gene PCR amplification and sequencing, and published genome sequences (Wu et al., 2014a, Zhu et al., 2019, Guardo et al., 2021). The cDNA sequences of *Nicole* homologs from these three species are very similar, which was mirrored by the number of shared gRNA target sites. This is consistent with two recent phylogenomic analyses studies (Wu et al., 2018, Zhu et al., 2019). Both, in fact, challenge historical *Citrus* taxonomy, placing *Fortunella*, and other genera, within the *Citrus* clade. For example, *Citrus* gene homologs were identified for 97% of *F. hindsii* gene loci. Despite, 13 of the 48 high-specificity *CsNicole* gRNA target sites not being present within *C. limon* and *F. hindsii Nicole* alleles, my results support previous phylogenomic analyses of *Citrus*, and relatives, and demonstrate the potential of multi-species gRNA editing based on sequence compatibility in *Citrus*. Of course, rigorous gRNA specificity analyses should be conducted to identify universal gRNAs, as the analysis here only considered *C. sinensis* off targets. However, for simple gRNA-specific mutation rate analyses using *A. rhizogenes*-mediated transformations, identifying mutual gRNA sites was sufficient to identify efficient gRNAs.

5.4.2 Maximising MYB loss of function potential in gRNA design

Multiple gRNA target sites were tested to compare mutation type, endogenous location within the target, and distance between gRNAs within a pair, on predicted functionality of translated mutant proteins. The gRNA selected here were all within the proximity of the first R2R3MYB domain repeat nucleotide sequence of *Nicole*. R2R3MYB DBDs interact directly with specific DNA sequences and are highly conserved across the plant kingdom. However, MYB DBDs vary in binding preferences, particularly to *cis*-regulatory motifs such as the MYB core and AC-rich elements, as described in chapter 1 (Prouse and Campbell, 2012, Franco-Zorrilla et al., 2014, Kelemen et al., 2015). The 3rd α -helix of both repeats perform as recognition helices, interacting with the core of DNA sequence motifs, while the R3MYB repeat is also wholly responsible for bHLH interaction due to a conserved amino acid motif (Zimmermann et al., 2004). However, the R2 recognition helix is more conserved than the R3 helix 3, at least in the 125 R2R3MYB proteins in *Arabidopsis*, suggesting R3 is more selective

in DNA-binding specificity (Kelemen et al., 2015). Disrupting translation of the MYB DBD will of course affect DNA binding capacity of the mutant MYB protein and subsequent regulatory function. Instances of which have been documented, for example in *Arabidopsis TT2*, *Ipomoea nil MYB1* and *AN2* in tomato (Nesi et al., 2001, Morita et al., 2006, Zhi et al., 2020).

Contrary to the R2R3MYB DBD, the C-terminal domains of R2R3MYB proteins are variable in sequence and length. Particular conserved motifs in non-MYB regions, however, have aided phylogenetical analyses, allowing subgroup clustering of MYBs which typically perform similar biological roles within subgroups (Kranz et al., 1998, Stracke et al., 2001, Jiang et al., 2004, Millard et al., 2019). Millard et al. (2019) hypothesised that the low sequence conservation outside MYB domains, contrary to within, is responsible for the large variety of biological functions within the vast expansion of R2R3MYBs in plants. Notably, these non-MYB regions have been credited for involvement in various functions, such as interactions with other proteins and even transcriptional activation (Stracke et al., 2017, Shin et al., 2007). Further, loss of function alleles with mutations located in the C-terminal regions have been observed in soybean and *Arabidopsis* (Gillman et al., 2011, Zhou et al., 2015b). This suggests non-DBD-targeting gRNA could still enable generation of loss of function alleles. However, regions outside the R2R3MYB DBD are poorly understood structurally and non-MYB regions responsible for the diverse array of molecular functions are poorly defined (Millard et al., 2019).

With these findings in consideration, gRNAs were selected to facilitate gene editing likely to disrupt translation of the R2MYB repeat in *Nicole*. The most upstream gRNAs within gRNA pairs 1 and 2 (gRNA_002 and 003) are situated prior to the R2MYB repeat. This offers the potential to introduce stop codons extremely early in translation of the Nicole protein (~64 bp downstream of ATG), or frame shifts, preventing translation of the MYB domain. The most downstream gRNA within gRNA pairs 1 and 2 (gRNA_019 and 011) are located after and within the R2MYB repeat. The dual gRNA design can facilitate fragment deletion at two loci within *Nicole*. If simultaneous cleavage occurs at gRNA targets in pairs 1 and 2, all and approximately half of the R2MYB repeat encoding sequence could be deleted, respectively. While gRNA pair 2 could not facilitate deletion of the entire R2MYB repeat, independent mutations at gRNA_011 still have the capacity to disrupt translation of the recognition helix. In contrast, gRNA_019-specific edits would not impact translation of the R2 repeat. Despite

this, indels may interfere with the R3MYB repeat sequence shortly upstream and potentially the mRNA splice recognition site at the exon 1-intron boundary.

5.4.3 *A. rhizogenes* transformation efficiency in *C. limon*

Co-cultivation with *C. limon* epicotyls produced a proliferation of transformed hairy roots from transfected tissue. As a result, this system allowed faster genotyping of transformed plant tissue relative to conventional *A. tumefaciens*-mediated transformations. This can be exploited to assess CRISPR gRNA efficacies prior to proceeding with stable transformation of *C. sinensis*, which may take many years to produce fruit for characterisation. As previously mentioned, the Hongkong kumquat is considered as a functional *Citrus* model and reassignment as a species of *Citrus* has been argued (Yang et al., 2007, Zhang et al., 2009, Cao et al., 2015, Yang et al., 2016, Wu et al., 2018, Zhu et al., 2019). *F. hindsii* also benefits from accelerated maturation, which is advantageous for fruit tissue phenotyping. Regardless, time constraints prevented me from starting the stable transformation of *F. hindsii* with *A. tumefaciens* harbouring CRISPR-Cas9 constructs. Since this itself is a large undertaking, this commitment would merit from mutation rate analysis of gRNAs to construct the most optimal multi-sgRNA CRISPR-Cas9 plasmid. Guide RNA candidates are thought to be widely compatible between *C. sinensis*, *F. hindsii* and *C. limon* and the optimised gRNAs and vectors I have identified should be directly transferrable for *C. sinensis* and *F. hindsii* gene editing.

K599 was previously reported to be virulent on *P. trifoliata*, albeit to a lower degree than MSU440 (Xiao et al., 2014). K599 appeared to have no virulence on *C. limon*, in my tests. Findings presented here and previous work suggests K599 is not an effective strain for infecting or transforming *Citrus* and its relatives. In comparison, *C. limon* transformations were a success with MSU440 and ATCC15834, demonstrating comparable efficiencies as in trifoliolate orange (Xiao et al., 2014). It was evident that root development was attributed to transfection with *A. rhizogenes*, because only necrosis was observed in TY control epicotyls either with selection or with no selection. The complete prevention of root growth in empty *A. rhizogenes*-transformed explants without a binary vector when cultured on kanamycin demonstrated that any root tissue on selection is likely transformed and expressing *NPTII*, at least. MSU440 transformations were on average 13% more efficient at inducing root development from explants than ATCC15834. It is also worth noting the striking reduction of

efficiency when transforming with recombinant plasmid DNA compared to without. This was much more severe with ATCC15834. This may simply be attributed to the absence of kanamycin selection on empty-transformed epicotyls. Since MSU440 based hairy root transformations have been studied in *Citrus* and relatives previously, root DNA extractions and subsequent genotyping was performed on ATCC15834 samples.

In combination with a robust seed sterilisation and germination protocol, transformation of *C. limon* with either *A. rhizogenes* strains MSU440 or ATCC15834 offered a fast and reasonably efficient method of obtaining transformed plant tissue for genotypic studies, and phenotypic if root tissue is of interest specifically. Although, I did not confirm binary vector co-transformation explicitly, the observed induction of root growth strongly suggested successful insertion on exogenous genetic elements. Further, DNA extraction with the DNeasy Plant Mini Kit (Qiagen) was poor, severely constraining subsequent analysis of genetic material. A more effective DNA extraction method in roots should be employed, such as the modified SDS-LiCl method described by Vennapusa et al. (2020). An alternative option for fast genotyping of non-root tissue might be *Xanthomonas citri* subsp. *citri*-facilitated agroinfiltration as described in the first report of CRISPR-Cas9 mutagenesis in *Citrus* (Jia and Wang, 2014). Their edit rates, however, were at best 3.9%, which is incredibly low compared to more recent and conventional transformation studies (Jia et al., 2017, Peng et al., 2017, Zhang et al., 2017, Zhu et al., 2019, Xu et al., 2022).

In the interests of time, the method presented here with suggested improvements provides a relatively simple and quick system to generate transgenic tissue for genotypic analyses in *Citrus*. Based on previous phylogenomic analyses, it is also proposed here that *C. limon* may serve as an effective substitute for *F. hindsii*, with which hairy root transformations have not been yet tested, particularly if plant material is unavailable. This may be more difficult to argue with respect to functional and phenotypic analyses, but when concerning efficacy of mutually shared gRNA target sites, *C. limon* will likely provide a suitable alternative prior to *A. tumefaciens*-mediated transformation of *F. hindsii*. Confirmation could be achieved by hairy root transformation of *F. hindsii* with the same multi-sgRNA CRISPR-Cas9 plasmids presented here.

5.4.4 CRISPR-Cas9 gRNA efficacy analyses via *A. rhizogenes*-mediated transformation in *Citrus*

Mutant *nicole* alleles were sequenced from root DNA extracted from *C. limon* harbouring pEA13 and pEA15. This was achieved by PCR amplification of *Nicole* exon 1 and TOPO cloning. This approach was taken to separate the two *Nicole* alleles in lemon, which differ primarily in the number of GCA repeats spaced in proximity, but upstream, of gRNA_002 and 003 target sites. Of course, the similarity in wildtype and single bp mutant allele nucleotide length also prevents isolation via gel excision and extraction practically. Consequently, to isolate allele amplicons within the PCR samples TOPO cloning was employed. Plasmids extracted from independent *E. coli* colonies therefore contain a single PCR product, some of which may be identical. If possible, up to 6 colonies were selected to capture all allelic diversity within a root sample and unique sequences analysed to avoid pseudoreplication.

My findings demonstrated the successful design and application of *Nicole*-targeting gRNAs, as CRISPR-Cas9 mutations were found at all 4 gRNA target sites. A range of mutation types occurred. The most common of which were single bp insertions, often observed in other CRISPR-Cas9 mutagenesis studies (Ma et al., 2015, Jia et al., 2017, Zhang et al., 2017, Zhu et al., 2019). Roots from which DNA was extracted separately appear to be derived from independent mutation events. However, mutation data also indicate that roots may be chimeric for CRISPR-Cas9 mutagenesis as up to 4 different alleles were amplified. This is a frequent occurrence in *Citrus* transformations (Jia et al., 2017, Zhang et al., 2017). Therefore, it should be noted that the selection of up to 6 TOPO clones per root may actually fail to isolate all alleles within the sample. In some instances, 6 *E. coli* colonies did not develop during TOPO cloning, a suspected reflection of poor root DNA extraction as mentioned previously. To reiterate, an improved DNA extraction protocol is expected to facilitate wider inclusivity of all alleles within a PCR pool.

At the plasmid level, pEA15 performed better than pEA13 with regard to mutation rate, with 91% of unique alleles sequenced containing gene edits at any of the two gRNA target sites. It is clear this efficacy can be attributed to gRNA_003, primarily. The mutation rate reported for each multi-sgRNA CRISPR-Cas9 plasmid is in fact high relative to some previous CRISPR-Cas9 mutagenesis studies in *Citrus* (Jia et al., 2017, Peng et al., 2017, Zhang et al., 2017, Zhu et al., 2019, Xu et al., 2022). These studies, however, involved transformation with *A.*

tumefaciens and analyses of regenerated shoots, rather than *A. rhizogenes* and transformed roots here.

Interestingly, Zhang et al. (2017) assessed the change in gRNA mutation efficiency when transforming *Citrus* with Cas9 driven by the *Arabidopsis* YAO promoter, rather than CaMV 35S promoter, as improvements were seen in *Arabidopsis* (Yan et al., 2015). A mutation rate of 75% was achieved when Cas9 expression was controlled by pYAO, a significant increase compared to under 5% with p35S (Jia and Wang, 2014). Jia and Wang (2014), however, utilised a transient transformation system of *Xanthomonas citri* subsp. *citri*-facilitated agroinfiltration, unlike the stable *Agrobacterium* transformation employed to test pYAO (Zhang et al., 2017). Despite the claimed improvement, recent CRISPR-Cas9 studies in *Citrus* have retained the use of p35S-driven Cas9 and have achieved reported mutation rates comparable to that with pYAO regardless.

It is important to note again that since only unique alleles within a root sample were evaluated, to avoid pseudoreplication, high copy numbers of the wildtype allele will be underrepresented. Evidence for chimerism in transgenic roots was found, containing up to four, and possibly more, mutated *Nicole* sequences. Further, total allelic diversity may not have been assessed as only 6 TOPO clones per root were sequenced. Regardless, evidence of gRNA-guided Cas9 activity has been confirmed in *C. limon* transfected with *A. rhizogenes* cells harbouring both plasmids. It was clear that gRNA_011 was much less effective at editing its target DNA compared to other gRNAs tested. While both gRNA_002 and 003, located upstream of the R2MYB repeat, performed relatively well, gRNA_003 offered close to 100% mutation efficiency. More R2MYB repeat fragment deletions were observed between gRNA sites in pEA13-transformed lemon alleles, but this must be attributed to poor gRNA_011 efficiency in pEA15, and therefore less simultaneous cleavage events. As such, it is hypothesised that a combination of gRNA_003 from pEA015, and gRNA_019 from pEA13 would provide an even more effective combination for generating *Nicole* mutants.

5.4.5 Functional analyses of CRISPR-Cas9 mutagenised *Nicole* alleles

Predicted amino acid translations were generated from all sequenced *Nicole* mutant alleles to assess potential loss of function. Both wildtype *CINicole* alleles contain R2R3MYB repeats,

as Pfam domain hits found with the NCBI conserved domain web tool. Due to time constraints, use of this tool to identify the loss of these R2R3MYB repeats in mutants was used simply as a preliminary effort to predict loss of function. The R2R3MYB domain is essential for interaction with specific DNA sequences (Prouse and Campbell, 2012, Franco-Zorrilla et al., 2014, Kelemen et al., 2015), and any significant disruption in translation of this region is anticipated to eliminate function, as observed in many MYB mutants (Nesi et al., 2001, Morita et al., 2006, Zhi et al., 2020). While poorly understood, involvement of C-terminal regions in many other molecular processes has been reported (Stracke et al., 2017, Shin et al., 2007) and mutations in these regions, distant from the R2R3MYB DBD, have conferred loss of function in some plants (Gillman et al., 2011, Zhou et al., 2015b).

Predicted knockout mutants comprised approximately 90% of all translated mutant alleles. Despite the variation in mutation rate between pEA13 and pEA15-transformed samples, knockout efficiency was comparable. Considering this, and individual gRNA efficacy, nucleotide edits at one gRNA target site upstream of the R2MYB repeat alone seem sufficient. As protein functionality is not explicitly being assessed here via experimentation, a complete fragment loss between two gRNA loci is still anticipated to ensure loss of function most reliably. The R2MYB repeat region within most protein mutants has been completely altered due to single bp insertions prior to the repeat sequence. There are several alleles which have lost a multiple of 3 bp at the start of the R2MYB repeat, losing at most 3 amino acids within the repeat and keeping the remaining translation in frame.

One allele was also observed with an introduced stop codon after the R2MYB repeat (pEA13, 8a.1; Figure 5.3.12). This would be an interesting protein to characterise functionally to further our understanding of the protein-protein interactions of Nicole, particularly as the R3MYB repeat is reported to facilitate bHLH interactions (Zimmermann et al., 2004). Sample 7a.3 from pEA13-transformed lemon contains complete fragment deletion between the two gRNA loci. Naturally, the R2MYB repeat is lost from translation but the remaining downstream sequence remains unchanged due to preservation of the frame of translation. This would offer the opposite analyses, investigating the impact of the loss of R2MYB repeat specifically. Assuming both these mutations would eliminate function completely it would be advantageous from the perspective of attributing any phenotypic changes with the R2R3MYB domain directly since the remaining protein sequence is intact.

Of course, this is a preliminary and predictive analysis. It would be beneficial to transiently express these various *Nicole* mutants in the dual-luciferase reporter assay system presented in chapter 4. This would not only complement the dual-luciferase data itself, in which we hypothesise *Nicole* regulates the expression of candidate target genes, but also complement our genome editing approach by confirming whether knockout mutations have truly been developed. Further, we can build understanding of the roles of individual repeats within the R2R3MYB domain or non-MYB regions. Protein-interaction analyses, such as the yeast-2-hybrid system could likewise elucidate the region of *Nicole* responsible for MBW complex formation.

5.4.6 Concluding remarks

In summary, analyses of gRNA design and efficacy in editing *Nicole* in *Citrus* was performed exploiting hairy root transformations in lemon for the first time with successful results. Cross-compatible gRNAs have been designed to target *Nicole* homologs in *C. sinensis*, *F. hindsii* and *C. limon*, incorporated into multi-sgRNA CRISPR-Cas9 constructs and introduced into the *C. limon* genome. The *A. rhizogenes*-mediated transformation of *C. limon* demonstrated the virulence of three strains, all previously untested in lemon, one previously untested in *Citrus* and relatives. Successful CRISPR-Cas9 mutagenesis was observed via gel electrophoresis and allele sequencing, revealing a suite of mutations in *Nicole*. Finally, mutants have been provisionally assigned as loss of function mutants by assessing loss of the R2MYB repeat.

Further improvements should consider gRNA specificity analyses in all compatible *Citrus* species of interest, a more robust DNA extraction method from roots and a more exhaustive allele isolation approach to ensure adequate representation of all alleles in chimeric root samples. Furthermore, functional characterisation would complement the speculative functional analysis presented.

Chapter 6:
Discussion

6.1 Final conclusions

6.1.1 The first reported divergence from the pleiotropic phenotype of acidless mutants

Noemi mutants have been collectively considered as acidless varieties that exhibit a multi-trait phenotype attributed to the bHLH's inherent ability to associate with many MYB transcription factors. I have reported the first example of a disassociation between fruit pH and condensed tannins in *Citrus*, thereby confirming a MYB transcription factor other than Nicole is responsible, or that Nicole is not essential, for regulating PA synthesis. This directly contradicts recent data (Zhang et al., 2020). The differential expression profile of suspected hyperacidification-related genes associated with *Noemi* concurs with previous reports (Butelli et al., 2019, Huang et al., 2016, Li et al., 2015, Strazzer et al., 2019). However, my novel findings link these associations with Nicole activity, explicitly.

Furthermore, I have presented evidence there is no significant role in the PA pathway played by Nicole due to the synthesis of PAs in the seed coat of non-functional *nicole* mutants. In addition, PAs were undetectable in WT and mutant *C. sinensis* fruits. PAs localise to vacuoles in the seed coat, suggesting CsTT12 is induced, functional, and likely regulated by a different R2R3MYB transcription factor.

6.1.2 Nicole and Noemi associate in an MBW complex and is necessary for citrate bioaccumulation

My results shown that *Noemi* and Nicole cooperate by participating in an MBW complex to induce the expression of *Noemi*, *Nicole*, *CsPH1*, *CsPH5* and *CsLDOX*. Additionally, *nicole*^{soro} functionality was almost completely abolished. This finding complements the RNA sequencing data well, as down regulated genes can now be associated with the loss of function in the *nicole*^{soro} allele. Furthermore, *Noemi* is still expressed in *nicole* mutants and there is no expression of *Iris* or *Marys* in the fruit of any sweet orange variety tested. Consequently, phenotypic and transcriptomic changes can be attributed to *nicole*^{soro} specifically. It is not known whether the loss of function is due to the deletion of the AD or an inability to associate with *Noemi*. Analyses of the library of CRISPR/Cas9-derived mutants

could elucidate the protein domain structure of Nicole via a combination of dual-luciferase reporter assays and yeast-2-hybrid assays.

There was no evidence that Nicole and Noemi regulate the expression of citrate metabolism genes in transcriptomic data. There have been multiple studies recently investigating the relationship between *CsPH5* and citrate levels (Aprile et al., 2011, Strazzer et al., 2019, Shi et al., 2021, Shi et al., 2019). Direct targets of Nicole were confirmed via dual-luciferase reporter assays, including *CsPH1* and *CsPH5*. Co-expression of *CsPH3* was not essential for activation of either promoter but slightly enhanced the induction of *pCsPH5*. In comparison, regulatory functionality was abolished or considerably weakened in the mutant *nicole*^{so^{ro}}. This conforms to findings of previous work that the heterodimeric proton pump activity of *CsPH5* with *CsPH1* is key for citrate bioaccumulation.

6.1.3 Nicole has dissociated from the typical PA-role of MYB5-like homologs

The role of Nicole in regulating PA biosynthesis has been proposed recently (Zhang et al., 2020). Zhang et al. (2020) claim the over expression of *Nicole* in *Arabidopsis tt2* mutants restores seed PA content to that of the wild type. However, the data presented show only a partial restoration to Col-0 levels and the over expression of *Citrus TT2*-like genes, such as *Iris*, was not tested. It is not stated as to the functionality of *Nicole* and *Noemi* in the acidless Succari and Hong Anliu varieties studied since the cDNA was isolated from the control variety Anliu only. However, considering the lack of seed PAs and low *Noemi* expression in their transcriptomic data, the phenotype observed is likely to be attributed to a lack of many MBW complexes formed with *Noemi*, concurrent with that described by Butelli et al. (2019), rather than low *Nicole* expression specifically. Furthermore, the expression profile of *Noemi* and *Nicole* match that of the Vaniglia *noemi* mutant remarkably well. This suggests transcriptomic and phenotypic observations may be associated with low *Noemi* expression, which has been shown to govern a multitude of traits. The complete lack of detectable PAs or precursors in juice, yet high expression of *Nicole* in fruits, indicates regulation of PAs is either a minor role or no role at all.

As previously described, proper function can be restored to *ph Petunia* mutants when complemented by homologous *Arabidopsis* genes, despite having different endogenous

functions. When Noemi was expressed with either Nicole or Iris, genes in the acidity and PA pathways were induced, including *NtPH1*, *NtLDOX* and *NtANR*. I anticipate *NtPH5* and *NtTT12* were also induced, because of the transcriptional requirement of the respective homologs in *Arabidopsis* for the bioaccumulation of PAs, but this has not been unequivocally proven. Iris also seems to show preferential induction of *NtANR* due to the disproportionate level of epicatechin, compared to catechin. Grapevine *VvMYB5a* and *VvMYB5b*, also initiate the production of PAs when expressed in tobacco, despite primarily regulating hyperacidification endogenously (Amato et al., 2019, Deluc et al., 2008). *AtAHA10* and homologs, in non-hyperacidifying species, are inherently PA-related genes.

This likely explains the PA phenotype observed in overexpression tobacco lines. For example, Amato et al. (2019) report grapevine transcriptomics that suggest *VvPH5* facilitates bioaccumulation of metabolites only and hyperacidification is a response only to *VvPH1* induction. These reports, my Lima characterisation and transcriptomics, and phylogenetic analyses support the original model, placing *Nicole* as the MYB candidate regulating hyperacidification in *Citrus* fruits, while another, such as *Iris*, controls the expression of genes in the PA pathway. However, identifying conformity to the acidification mechanism observed in *Petunia* does not explain how pH levels as low as 2 are achieved in *Citrus*. It is not simply a case of increased citrate synthase activity, as activity of this gene appears to vary little between varieties with different levels of acidity (Guo et al., 2016, Chen et al., 2013, Hussain et al., 2017, Sadka et al., 2001, Lin et al., 2015, Lu et al., 2016, Yu et al., 2012).

Undoubtedly, hyperacidification and tannin bioaccumulation are interconnected pathways, likely due the requirement of tonoplastic proton gradients for metabolite transport to the vacuole. In *Citrus*, MYB target genes such as MATE metabolite transporters or additional MWB complex components such as WRKY factors, were considered to potentially play a stronger role in determining the downstream regulatory effects of *Nicole* than previously thought. Perhaps hyperacidity to a vacuolar pH of ~2 not only requires tonoplast-bound P-ATPases, but also a reduction in proton efflux to tip the balance?

6.1.4 Specialisation of Nicole

Concerning WRKY factors, there are instances where MBW coexpression with a WRKY protein enhances target gene expression or is in fact essential. *AtTTG2*, a WRKY factor in *Arabidopsis*, is recruited by an MBW complex comprising *AtTT2*, *AtbHLH42* and *AtTTG1*, and is essential for the expression of *AtTT12* (Gonzalez et al., 2016). A similar mechanism is observed in *V. vinifera* and *Petunia*. The putative *AtTTG2* homologs *VvWRKY26* and *PhPH3* are recruited by *VvMYB5a* and *PhAN11*, respectively, and can enhance, or be necessary for, target gene expression and alter specificity of MBW complexes (Lloyd et al., 2017, Amato et al., 2019, Verweij et al., 2016). It was surprising to observe the homologous WRKY, *CsPH3*, neither enhanced nor enabled the activation of any target gene promoters tested. The data presented here suggest *CsPH3* is not necessary for acidification, at least.

A second possibility concerned MATE transporters, symport proteins that facilitate the influx of metabolites and efflux of protons. For example, *AtTT12* is involved in the bioaccumulation of PAs in *Arabidopsis* by transporting precursor molecules into the tonoplast in exchange for protons (Marinova et al., 2007). The expression of this MATE is primarily controlled by an MBW complex containing the MYB *AtTT2* (Xu et al., 2014). *AtMYB5* can also regulate *AtTT12* expression, however, since complementation of *tt2* mutants partially restores function. *AtTT12* expression is also further reduced in *tt2 myb5* double mutants, in comparison to *tt2* mutants.

However, contrary to typical *AtMYB5* homologs, the RNA sequencing data and promoter activation assays prove that *CsTT12* is not induced by Noemi and Nicole (Xu et al., 2014). Analogous to *AtTT2*, *Iris* was expected to control the expression of *CsTT12*, and the lack of *CsTT12* expression in fruit was initially thought to reflect to the lack of *Iris* or *Marys* expression. However, *Iris* was not able to directly activate the *CsTT12* promoter either. *CsTT12* must be expressed in seeds due to the accumulation of PAs. It is presumed that the lack of promoter induction by *Iris* was attributed to the absence of *CsPH3*. Regardless, that data indicates a divergence from typical regulatory functions of MYB5-homologs observed in *Arabidopsis* and *Petunia*, whereby hyperacidity in *Citrus* fruit is in part attributed to the offset counteraction of *CsPH1* and *CsPH5* heterodimeric proton pump activity by the lack of symport activity of *CsTT12* in fruit.

In conjunction with the results, it is likely that *Nicole* can redundantly promote the expression of a set of PA-related genes. Rather, *Iris* primarily controls the pathway and *Nicole* plays a secondary role, analogous to the differences in target genes between *AtTT2* and *AtMYB5* (Xu et al., 2014). A relatively quick set of experiments to give merit to this hypothesis would include comparing *Iris* complementation of *Arabidopsis tt2* with previous *Nicole* complementation results (Zhang et al., 2020). In addition, functional analysis of *CsTT12* could be assessed by complementation of *Arabidopsis tt12*. Of course, induction of *pCsTT12* should be tested with the co-expression of *Noemi* and *Iris* with *CsPH3*.

Uniquely, *Nicole* has specialised in regulation of acidification not just by the induction of 'new' non-PA genes, such as *CsPH1*, which is anticipated to boost the proton pump activity

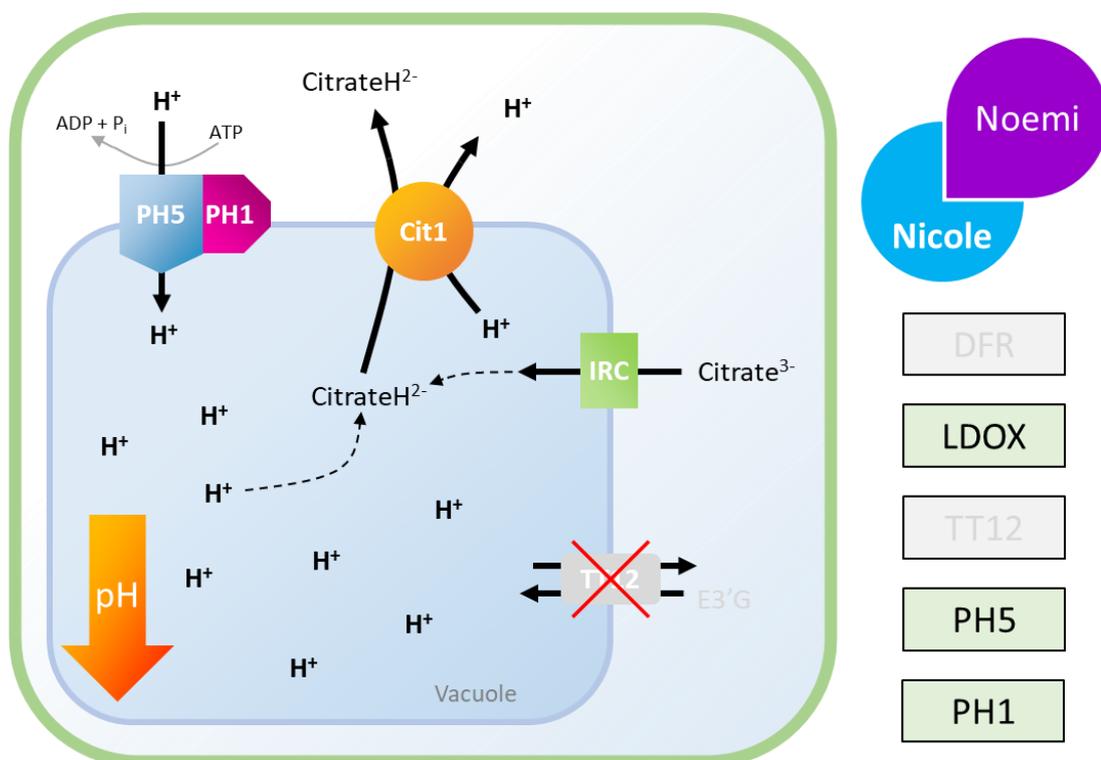


Figure 6.1.1 Simplified schematic of vacuolar hyperacidification in Citrus fruit due to the specialisation of R2R3MYB transcription factor, Nicole. E3'G: 3'-glucosylated epicatechin. List of PA and acidity-related genes activated by MBW complex comprising Nicole and Noemi are in green (right). Target genes for which Nicole has lost the ability to activate are in grey (right).

of CsPH5 beyond PA-transporter requirements, but has also lost the ability to target PA genes (*CsTT12* and *CsDFR*) which may counteract the proton influx. To summarise, AtMYB5 in *Arabidopsis* redundantly promotes the expression of PA-related genes *AtDFR*, *AtLDOX*, *AtTT12* and *AtAHA10*. Nicole is likely able to induce the expression of these genes when ectopically expressed, as evident in overexpression tobacco lines where PAs were synthesised, and PA genes up regulated. Within *Citrus*, however, I propose the following model of fruit acidity regulation (Figure 6.1.1). Nicole is essential for the expression of CsPH1 and CsPH5, and by extension, the bioaccumulation of citrate. Nicole has also lost the ability to induce *CsTT12* expression, thereby minimising proton efflux.

6.1.5 Concluding remarks

The role of Nicole as a participant in MBW-mediated regulation has been examined in relation to hyperacidification and PA synthesis in *Citrus*. In combination with Noemi these MYB transcription factors have been ectopically expressed in *N. tabacum* plants and co-expressed transiently in *N. benthamiana* leaves to elucidate transcriptional regulation of candidate target genes. Metabolite analyses of PAs were conducted due to their induction in tobacco leaves. Nicole and Iris evidently participate in an MBW complex with Noemi due to the resounding lack of transcriptional regulation of all *Citrus* promoters analysed when expressed alone. Both Nicole and Iris, with Noemi, induced the accumulation of PAs in tobacco. Analyses of gene induction, however, suggest Nicole has a more specialised role within *Citrus*. For example, *Nicole* is highly expressed in fruit tissues, where PAs are not synthesised. Further, I observed no induction of key PA genes, including the MATE transporter *CsTT12*, and PAs accumulate in non-functional *nicole* mutants Sorocaba, Verde R1 and Verde R2.

Further gene expression analyses in tobacco overexpression lines of other structural genes in the PA pathway would complement our growing understanding of the primary roles of Nicole and Iris. Complementation and gene expression analyses of mutants *ph4* in *Petunia* and *myb5* and *tt2* of *Arabidopsis* with constitutive expression of Nicole or Iris would be interesting to establish whether Nicole is able to activate PA structural genes of other species. Similarly, repeating p*CsTT12* dual-luciferase reporter assays with the co-expression of *AtMYB5*, *VvMYB5a* and *VvMYB5b*-driven MBW complexes, which have been shown to

induce respective *TT12* homologs, could inform whether *pTT12* is only unresponsive to Nicole.

A dissociation between Nicole and PA-regulation has been shown, while a direct relationship between MYB transcription factor Nicole, P-ATPase proton pumps, citrate accumulation and pH has been presented, offering potential targets for improving taste and fruit quality in *Citrus*.

References

- AARROUF, J., CASTRO-QUEZADA, P., MALLARD, S., CAROMEL, B., LIZZI, Y. & LEFEBVRE, V. 2012. Agrobacterium rhizogenes-dependent production of transformed roots from foliar explants of pepper (*Capsicum annuum*): a new and efficient tool for functional analysis of genes. *Plant cell reports*, 31, 391-401.
- ABEYNAYAKE, S. W., PANTER, S., MOURADOV, A. & SPANGENBERG, G. 2011. A high-resolution method for the localization of proanthocyanidins in plant tissues. *Plant Methods*, 7, 1-6.
- ABRAHAMS, S., LEE, E., WALKER, A. R., TANNER, G. J., LARKIN, P. J. & ASHTON, A. R. 2003. The Arabidopsis TDS4 gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development. *The Plant Journal*, 35, 624-636.
- ALBERTINI, M.-V., CARCOUET, E., PAILLY, O., GAMBOTTI, C., LURO, F. & BERTI, L. 2006. Changes in organic acids and sugars during early stages of development of acidic and acidless citrus fruit. *Journal of Agricultural and Food Chemistry*, 54, 8335-8339.
- ALEXA, A. & RAHNENFÜHRER, J. 2009. Gene set enrichment analysis with topGO. *Bioconductor Improv*, 27, 1-26.
- ALI, E., SAAND, M. A., KHAN, A. R., SHAH, J. M., FENG, S., MING, C. & SUN, P. 2021. Genome-wide identification and expression analysis of detoxification efflux carriers (DTX) genes family under abiotic stresses in flax. *Physiologia Plantarum*, 171, 483-501.
- AMATO, A., CAVALLINI, E., WALKER, A. R., PEZZOTTI, M., BLIEK, M., QUATTROCCHIO, F., KOES, R., RUPERTI, B., BERTINI, E., ZENONI, S. & TORNIELLI, G. B. 2019. The MYB5-driven MBW complex recruits a WRKY factor to enhance the expression of targets involved in vacuolar hyper-acidification and trafficking in grapevine. *The Plant Journal*.
- AMATO, A., CAVALLINI, E., ZENONI, S., FINEZZO, L., BEGHELDO, M., RUPERTI, B. & TORNIELLI, G. B. 2016. A Grapevine TTG2-Like WRKY Transcription Factor Is Involved in Regulating Vacuolar Transport and Flavonoid Biosynthesis. *Front Plant Sci*, 7, 1979.
- AN, X.-H., TIAN, Y., CHEN, K.-Q., LIU, X.-J., LIU, D.-D., XIE, X.-B., CHENG, C.-G., CONG, P.-H. & HAO, Y.-J. 2015. MdMYB9 and MdMYB11 are involved in the regulation of the JA-induced biosynthesis of anthocyanin and proanthocyanidin in apples. *Plant and Cell Physiology*, 56, 650-662.
- APPELHAGEN, I., NORDHOLT, N., SEIDEL, T., SPELT, K., KOES, R., QUATTROCCHIO, F., SAGASSER, M. & WEISSHAAR, B. 2015. TRANSPARENT TESTA 13 is a tonoplast P3A-ATPase required for vacuolar deposition of proanthocyanidins in Arabidopsis thaliana seeds. *The Plant Journal*, 82, 840-849.
- APRILE, A., FEDERICI, C., CLOSE, T. J., DE BELLIS, L., CATTIVELLI, L. & ROOSE, M. L. 2011. Expression of the H⁺-ATPase AHA10 proton pump is associated with citric acid accumulation in lemon juice sac cells. *Functional & integrative genomics*, 11, 551-563.

- BARRETT, H. C. & RHODES, A. M. 1976. A Numerical Taxonomic Study of Affinity Relationships in Cultivated Citrus and Its Close Relatives. *Systematic Botany*, 1, 105-136.
- BAUDRY, A., HEIM, M. A., DUBREUCQ, B., CABOCHE, M., WEISSHAAR, B. & LEPINIEC, L. 2004. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *The Plant Journal*, 39, 366-380.
- BAXTER, I. R., YOUNG, J. C., ARMSTRONG, G., FOSTER, N., BOGENSCHUTZ, N., CORDOVA, T., PEER, W. A., HAZEN, S. P., MURPHY, A. S. & HARPER, J. F. 2005. A plasma membrane H⁺-ATPase is required for the formation of proanthocyanidins in the seed coat endothelium of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 102, 2649-2654.
- BERK, Z. 2016. Chapter 13 - Nutritional and health-promoting aspects of citrus consumption. *Citrus Fruit Processing*. San Diego: Academic Press.
- BERNARD, P. & COUTURIER, M. 1992. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *Journal of molecular biology*, 226, 735-745.
- BERNARD, P., GABARIT, P., BAHASSI, E. M. & COUTURIER, M. 1994. Positive-selection vectors using the F plasmid ccdB killer gene. *Gene*, 148, 71-74.
- BOGS, J., JAFFE, F. W., TAKOS, A. M., WALKER, A. R. & ROBINSON, S. P. 2007. The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development. *Plant Physiology*, 143, 1347-1361.
- BOMBARELY, A., ROSLI, H. G., VREBALOV, J., MOFFETT, P., MUELLER, L. A. & MARTIN, G. B. 2012. A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. *Molecular Plant-Microbe Interactions*, 25, 1523-1530.
- BOREVITZ, J. O., XIA, Y., BLOUNT, J., DIXON, R. A. & LAMB, C. 2000. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *The Plant Cell*, 12, 2383-2393.
- BOZKURT, T. O., SCHORNACK, S., WIN, J., SHINDO, T., ILYAS, M., OLIVA, R., CANO, L. M., JONES, A. M., HUITEMA, E. & VAN DER HOORN, R. A. 2011. *Phytophthora infestans* effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. *Proceedings of the National Academy of Sciences*, 108, 20832-20837.
- BROOKS, C., NEKRASOV, V., LIPPMAN, Z. B. & VAN ECK, J. 2014. Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant physiology*, 166, 1292-1297.
- BROWN, M. H., PAULSEN, I. T. & SKURRAY, R. A. 1999. The multidrug efflux protein NorM is a prototype of a new family of transporters. *Molecular microbiology*, 31, 394-395.
- BRUEGMANN, T., DEECKE, K. & FLADUNG, M. 2019. Evaluating the efficiency of gRNAs in CRISPR/Cas9 mediated genome editing in poplars. *International journal of molecular sciences*, 20, 3623.

- BUTELLI, E., LICCIARDELLO, C., RAMADUGU, C., DURAND-HULAK, M., CELANT, A., RECUPERO, G. R., FROELICHER, Y. & MARTIN, C. 2019. Noemi Controls Production of Flavonoid Pigments and Fruit Acidity and Illustrates the Domestication Routes of Modern Citrus Varieties. *Current Biology*, 29, 158-+.
- BUTELLI, E., LICCIARDELLO, C., ZHANG, Y., LIU, J., MACKAY, S., BAILEY, P., REFORGIATO-RECUPERO, G. & MARTIN, C. 2012. Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell*, 24, 1242-55.
- CAO, D., HOU, W., LIU, W., YAO, W., WU, C., LIU, X. & HAN, T. 2011. Overexpression of TaNHX2 enhances salt tolerance of 'composite' and whole transgenic soybean plants. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 107, 541-552.
- CAO, H., WANG, J., DONG, X., HAN, Y., MA, Q., DING, Y., ZHAO, F., ZHANG, J., CHEN, H. & XU, Q. 2015. Carotenoid accumulation affects redox status, starch metabolism, and flavonoid/anthocyanin accumulation in citrus. *BMC plant biology*, 15, 1-16.
- CARBONELL-CABALLERO, J., ALONSO, R., IBAÑEZ, V., TEROL, J., TALON, M. & DOPAZO, J. 2015. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus Citrus. *Molecular biology and evolution*, 32, 2015-2035.
- CARVALHO, R. F. & FOLTA, K. M. 2017. Assessment of promoters and a selectable marker for development of strawberry intragenic vectors. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 128, 259-271.
- CAVALLINI, E., ZENONI, S., FINEZZO, L., GUZZO, F., ZAMBONI, A., AVESANI, L. & TORNIELLI, G. B. 2014. Functional diversification of grapevine MYB5a and MYB5b in the control of flavonoid biosynthesis in a petunia anthocyanin regulatory mutant. *Plant and Cell Physiology*, 55, 517-534.
- CERCOS, M., SOLER, G., IGLESIAS, D. J., GADEA, J., FORMENT, J. & TALON, M. 2006. Global analysis of gene expression during development and ripening of citrus fruit flesh. A proposed mechanism for citric Acid utilization. *Plant Mol Biol*, 62, 513-27.
- CHAI, Y.-R., LEI, B., HUANG, H.-L., LI, J.-N., YIN, J.-M., TANG, Z.-L., WANG, R. & CHEN, L. 2009. TRANSPARENT TESTA 12 genes from Brassica napus and parental species: cloning, evolution, and differential involvement in yellow seed trait. *Molecular Genetics and Genomics*, 281, 109-123.
- CHAVEZ-VELA, N., CHAVEZ-ORTIZ, L. & PEREZ-MOLPHE BALCH, E. 2003. Genetic transformation of sour orange using Agrobacterium rhizogenes. *Agrociencia*.
- CHEN, M., JIANG, Q., YIN, X.-R., LIN, Q., CHEN, J.-Y., ALLAN, A. C., XU, C.-J. & CHEN, K.-S. 2012. Effect of hot air treatment on organic acid-and sugar-metabolism in Ponkan (Citrus reticulata) fruit. *Scientia Horticulturae*, 147, 118-125.
- CHEN, M., XIE, X., LIN, Q., CHEN, J., GRIERSON, D., YIN, X., SUN, C. & CHEN, K. 2013. Differential expression of organic acid degradation-related genes during fruit development of Navel oranges (Citrus sinensis) in two habitats. *Plant Molecular Biology Reporter*, 31, 1131-1140.

- CHENG, Y., WANG, X., CAO, L., JI, J., LIU, T. & DUAN, K. 2021. Highly efficient *Agrobacterium rhizogenes*-mediated hairy root transformation for gene functional and gene editing analysis in soybean. *Plant Methods*, 17, 1-12.
- DE VETTEN, N., QUATTROCCHIO, F., MOL, J. & KOES, R. 1997. The an11 locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. *Genes & development*, 11, 1422-1434.
- DELUC, L., BARRIEU, F., MARCHIVE, C., LAUVERGEAT, V., DECENDIT, A., RICHARD, T., CARDE, J.-P., MÉRILLON, J.-M. & HAMDY, S. 2006. Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway. *Plant physiology*, 140, 499-511.
- DELUC, L., BOGS, J., WALKER, A. R., FERRIER, T., DECENDIT, A., MERILLON, J. M., ROBINSON, S. P. & BARRIEU, F. 2008. The transcription factor VvMYB5b contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. *Plant Physiology*, 147, 2041-2053.
- DENG, X. & ZHANG, W. 1988. The in vitro induction method of embryonic callus from seedling of *Fortunella hindsii* Swingle. *South China Fruits*, 1988, 7-9.
- DEVEAU, H., GARNEAU, J. E. & MOINEAU, S. 2010. CRISPR/Cas system and its role in phage-bacteria interactions. *Annual review of microbiology*, 64, 475-493.
- DEWITT, N. D., HONG, B., SUSSMAN, M. R. & HARPER, J. F. 1996. Targeting of two Arabidopsis H⁺-ATPase isoforms to the plasma membrane. *Plant Physiology*, 112, 833-844.
- DEWITT, N. D. & SUSSMAN, M. R. 1995. Immunocytological localization of an epitope-tagged plasma membrane proton pump (H⁺-ATPase) in phloem companion cells. *The Plant Cell*, 7, 2053-2067.
- DI, Y.-H., SUN, X.-J., HU, Z., JIANG, Q.-Y., SONG, G.-H., ZHANG, B., ZHAO, S.-S. & ZHANG, H. 2019. Enhancing the CRISPR/Cas9 system based on multiple GmU6 promoters in soybean. *Biochemical and Biophysical Research Communications*, 519, 819-823.
- DIENER, A. C., GAXIOLA, R. A. & FINK, G. R. 2001. Arabidopsis ALF5, a multidrug efflux transporter gene family member, confers resistance to toxins. *The Plant Cell*, 13, 1625-1638.
- DOWNEY, M. O., HARVEY, J. S. & ROBINSON, S. P. 2003. Analysis of tannins in seeds and skins of Shiraz grapes throughout berry development. *Australian Journal of Grape and Wine Research*, 9, 15-27.
- DROZDOWICZ, Y. M. & REA, P. A. 2001. Vacuolar H⁺ pyrophosphatases: from the evolutionary backwaters into the mainstream. *Trends in plant science*, 6, 206-211.
- DUARTE, A., FERNANDES, M. J., BERNARDES, J. P. & MIGUEL, M. G. 2016. Citrus as a component of the Mediterranean diet. *Journal of Spatial and Organizational Dynamics*, 4, 289-304.
- DUBOS, C., STRACKE, R., GROTEWOLD, E., WEISSHAAR, B., MARTIN, C. & LEPINIEC, L. 2010. MYB transcription factors in Arabidopsis. *Trends in plant science*, 15, 573-581.

- DURRETT, T. P., GASSMANN, W. & ROGERS, E. E. 2007. The FRD3-mediated efflux of citrate into the root vasculature is necessary for efficient iron translocation. *Plant physiology*, 144, 197-205.
- ECHEVERRIA, E., BURNS, J. & FELLE, H. 1992. Compartmentation and cellular conditions controlling sucrose breakdown in mature acid lime fruits. *Phytochemistry*, 31, 4091-4095.
- ECONOMOS, C. & CLAY, W. 1999. Nutritional and health benefits of citrus fruits. *Food, nutrition and agriculture*, 11-18.
- EDWARDS, K., FERNANDEZ-POZO, N., DRAKE-STOWE, K., HUMPHRY, M., EVANS, A., BOMBARELY, A., ALLEN, F., HURST, R., WHITE, B. & KERNODLE, S. 2017. A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC genomics*, 18, 1-14.
- ERPEN, L., TAVANO, E., HAKAKAVA, R., DUTT, M., GROSSER, J., PIEDADE, S. M. D. S., MENDES, B. M. J. & MOURÃO FILHO, F. D. A. A. 2018. Isolation, characterization, and evaluation of three *Citrus sinensis*-derived constitutive gene promoters. *Plant cell reports*, 37, 1113-1125.
- ETIENNE, A., GENARD, M., LOBIT, P., MBEGUIE, A. M. D. & BUGAUD, C. 2013. What controls fleshy fruit acidity? A review of malate and citrate accumulation in fruit cells. *J Exp Bot*, 64, 1451-69.
- FANG, D., FEDERICI, C. & ROOSE, M. 1997. Development of molecular markers linked to a gene controlling fruit acidity in citrus. *Genome*, 40, 841-849.
- FAO, F. A. A. O. 2021. FAOSTAT food and agriculture database. <https://www.fao.org/faostat/en/#home>.
- FARACO, M., SPELT, C., BLIEK, M., VERWEIJ, W., HOSHINO, A., ESPEN, L., PRINSI, B., JAARSMA, R., TARHAN, E., DE BOER, A. H., DI SANSEBASTIANO, G. P., KOES, R. & QUATTROCCHIO, F. M. 2014. Hyperacidification of vacuoles by the combined action of two different P-ATPases in the tonoplast determines flower color. *Cell Rep*, 6, 32-43.
- FAVELA-HERNÁNDEZ, J., GONZÁLEZ-SANTIAGO, O., RAMÍREZ-CABRERA, M., ESQUIVEL-FERRIÑO, P. & CAMACHO-CORONA, M. 2016. Chemistry and Pharmacology of *Citrus sinensis*. *Molecules*, 21, 247.
- FENG, Z., ZHANG, B., DING, W., LIU, X., YANG, D.-L., WEI, P., CAO, F., ZHU, S., ZHANG, F. & MAO, Y. 2013. Efficient genome editing in plants using a CRISPR/Cas system. *Cell research*, 23, 1229-1232.
- FRANCO-ZORRILLA, J. M., LÓPEZ-VIDRIERO, I., CARRASCO, J. L., GODOY, M., VERA, P. & SOLANO, R. 2014. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences*, 111, 2367-2372.
- FRANK, S., KECK, M., SAGASSER, M., NIEHAUS, K., WEISSHAAR, B. & STRACKE, R. 2011. Two differentially expressed MATE factor genes from apple complement the *Arabidopsis* transparent testa12 mutant. *Plant Biology*, 13, 42-50.

- GAGNÉ, S., LACAMPAGNE, S., CLAISSE, O. & GÉNY, L. 2009. Leucoanthocyanidin reductase and anthocyanidin reductase gene expression and activity in flowers, young berries and skins of *Vitis vinifera* L. cv. Cabernet-Sauvignon during development. *Plant Physiology and Biochemistry*, 47, 282-290.
- GARDNER, P. T., WHITE, T. A., MCPHAIL, D. B. & DUTHIE, G. G. 2000. The relative contributions of vitamin C, carotenoids and phenolics to the antioxidant potential of fruit juices. *Food Chemistry*, 68, 471-474.
- GARNEAU, J. E., DUPUIS, M.-È., VILLION, M., ROMERO, D. A., BARRANGOU, R., BOYAVAL, P., FREMAUX, C., HORVATH, P., MAGADÁN, A. H. & MOINEAU, S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468, 67-71.
- GAXIOLA, R. A., PALMGREN, M. G. & SCHUMACHER, K. 2007. Plant proton pumps. *FEBS letters*, 581, 2204-2214.
- GESELL, A., YOSHIDA, K., TRAN, L. T. & CONSTABEL, C. P. 2014. Characterization of an apple TT2-type R2R3 MYB transcription factor functionally similar to the poplar proanthocyanidin regulator PtMYB134. *Planta*, 240, 497-511.
- GILLMAN, J. D., TETLOW, A., LEE, J.-D., SHANNON, J. G. & BILYEU, K. 2011. Loss-of-function mutations affecting a specific Glycine max R2R3 MYB transcription factor result in brown hilum and brown seed coats. *BMC plant biology*, 11, 1-12.
- GOFF, S. A., CONE, K. C. & FROMM, M. E. 1991. Identification of functional domains in the maize transcriptional activator C1: comparison of wild-type and dominant inhibitor proteins. *Genes & development*, 5, 298-309.
- GOGVADZE, E. & BUZDIN, A. 2009. Retroelements and their impact on genome evolution and functioning. *Cellular and molecular life sciences*, 66, 3727-3742.
- GOMEZ, C., TERRIER, N., TORREGROSA, L., VIALET, S., FOURNIER-LEVEL, A., VERRIES, C., SOUQUET, J.-M., MAZAURIC, J.-P., KLEIN, M. & CHEYNIER, V. 2009. Grapevine MATE-type proteins act as vacuolar H⁺-dependent acylated anthocyanin transporters. *Plant physiology*, 150, 402-415.
- GONZALEZ, A., BROWN, M., HATLESTAD, G., AKHAVAN, N., SMITH, T., HEMBD, A., MOORE, J., MONTES, D., MOSLEY, T., RESENDEZ, J., NGUYEN, H., WILSON, L., CAMPBELL, A., SUDARSHAN, D. & LLOYD, A. 2016. TTG2 controls the developmental regulation of seed coat tannins in *Arabidopsis* by regulating vacuolar transport steps in the proanthocyanidin pathway. *Dev Biol*, 419, 54-63.
- GONZALEZ, A., MENDENHALL, J., HUO, Y. & LLOYD, A. 2009. TTG1 complex MYBs, MYB5 and TT2, control outer seed coat differentiation. *Developmental biology*, 325, 412-421.
- GUARDO, M. D., MORETTO, M., MOSER, M., CATALANO, C., TROGGIO, M., DENG, Z., CESTARO, A., CARUSO, M., DISTEFANO, G. & MALFA, S. L. 2021. The haplotype-resolved reference genome of lemon (*Citrus limon* L. Burm f.). *Tree Genetics & Genomes*, 17, 1-12.

- GUINDON, S., DUFAYARD, J.-F., LEFORT, V., ANISIMOVA, M., HORDIJK, W. & GASCUEL, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59, 307-321.
- GUO, L.-X., SHI, C.-Y., LIU, X., NING, D.-Y., JING, L.-F., YANG, H. & LIU, Y.-Z. 2016. Citrate accumulation-related gene expression and/or enzyme activity analysis combined with metabolomics provide a novel insight for an orange mutant. *Scientific reports*, 6, 1-12.
- HERNANDEZ-GARCIA, C. M., MARTINELLI, A. P., BOUCHARD, R. A. & FINER, J. J. 2009. A soybean (*Glycine max*) polyubiquitin promoter gives strong constitutive expression in transgenic soybean. *Plant cell reports*, 28, 837-849.
- HICHRI, I., HEPPEL, S. C., PILLET, J., LÉON, C., CZEMMEL, S., DELROT, S., LAUVERGEAT, V. & BOGS, J. 2010. The basic helix-loop-helix transcription factor MYC1 is involved in the regulation of the flavonoid biosynthesis pathway in grapevine. *Molecular plant*, 3, 509-523.
- HORVATH, P. & BARRANGOU, R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science*, 327, 167-170.
- HOU, X.-J., LI, S.-B., LIU, S.-R., HU, C.-G. & ZHANG, J.-Z. 2014. Genome-wide classification and evolutionary and expression analyses of citrus MYB transcription factor families in sweet orange. *PLoS One*, 9, e112375.
- HU, B., JIN, J., GUO, A.-Y., ZHANG, H., LUO, J. & GAO, G. 2015. GSDB 2.0: an upgraded gene feature visualization server. *Bioinformatics*, 31, 1296-1297.
- HUANG, D., ZHAO, Y., CAO, M., QIAO, L. & ZHENG, Z. L. 2016. Integrated Systems Biology Analysis of Transcriptomes Reveals Candidate Genes for Acidity Control in Developing Fruits of Sweet Orange (*Citrus sinensis* L. Osbeck). *Front Plant Sci*, 7, 486.
- HUSSAIN, S. B., SHI, C.-Y., GUO, L.-X., KAMRAN, H. M., SADKA, A. & LIU, Y.-Z. 2017. Recent Advances in the Regulation of Citric Acid Metabolism in Citrus Fruit. *Critical Reviews in Plant Sciences*, 1-16.
- ISHINO, Y., SHINAGAWA, H., MAKINO, K., AMEMURA, M. & NAKATA, A. 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of bacteriology*, 169, 5429-5433.
- JIA, H. & WANG, N. 2014. Targeted genome editing of sweet orange using Cas9/sgRNA. *PLoS one*, 9, e93806.
- JIA, H., ZHANG, Y., ORBOVIC, V., XU, J., WHITE, F. F., JONES, J. B. & WANG, N. 2017. Genome editing of the disease susceptibility gene *CsLOB1* in citrus confers resistance to citrus canker. *Plant Biotechnol J*, 15, 817-823.
- JIANG, C., GU, X. & PETERSON, T. 2004. Identification of conserved gene structures and carboxy-terminal motifs in the Myb gene family of *Arabidopsis* and *Oryza sativa* L. ssp. *indica*. *Genome biology*, 5, 1-11.

- JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *science*, 337, 816-821.
- KELEMEN, Z., SEBASTIAN, A., XU, W., GRAIN, D., SALSAC, F., AVON, A., BERGER, N., TRAN, J., DUBREUCQ, B. & LURIN, C. 2015. Analysis of the DNA-binding activities of the Arabidopsis R2R3-MYB transcription factor family by one-hybrid experiments in yeast. *PLoS one*, 10, e0141044.
- KERESZT, A., LI, D., INDRASUMUNAR, A., NGUYEN, C. D., NONTACHAIYAPOOM, S., KINKEMA, M. & GRESSHOFF, P. M. 2007. Agrobacterium rhizogenes-mediated transformation of soybean to study root biology. *Nature protocols*, 2, 948-952.
- KOBAYASHI, S., GOTO-YAMAMOTO, N. & HIROCHIKA, H. 2004. Retrotransposon-induced mutations in grape skin color. *Science*, 304, 982-982.
- KOYAMA, K., NUMATA, M., NAKAJIMA, I., GOTO-YAMAMOTO, N., MATSUMURA, H. & TANAKA, N. 2014. Functional characterization of a new grapevine MYB transcription factor and regulation of proanthocyanidin biosynthesis in grapes. *Journal of Experimental Botany*, 65, 4433-4449.
- KRAJEWSKI, A. & RABE, E. 1995. Citrus flowering: a critical evaluation. *Journal of Horticultural Science*, 70, 357-374.
- KRANZ, H. D., DENEKAMP, M., GRECO, R., JIN, H., LEYVA, A., MEISSNER, R. C., PETRONI, K., URZAINQUI, A., BEVAN, M. & MARTIN, C. 1998. Towards functional characterisation of the members of the R2R3-MYB gene family from Arabidopsis thaliana. *The Plant Journal*, 16, 263-276.
- KÜHLBRANDT, W. 2004. Biology, structure and mechanism of P-type ATPases. *Nature reviews Molecular cell biology*, 5, 282-295.
- LAI, B., DU, L. N., HU, B., WANG, D., HUANG, X. M., ZHAO, J. T., WANG, H. C. & HU, G. B. 2019. Characterization of a novel litchi R2R3-MYB transcription factor that involves in anthocyanin biosynthesis and tissue acidification. *Bmc Plant Biology*, 19, 13.
- LEE, M., DONG, X., SONG, H., YANG, J. Y., KIM, S. & HUR, Y. 2020. Molecular characterization of Arabidopsis thaliana LSH1 and LSH2 genes. *Genes & genomics*, 42, 1151-1162.
- LEFEBVRE, B., ARANGO, M., OUFATTOLE, M., CROUZET, J., PURNELLE, B. & BOUTRY, M. 2005. Identification of a Nicotiana plumbaginifolia plasma membrane H⁺-ATPase gene expressed in the pollen tube. *Plant molecular biology*, 58, 775-787.
- LEFEBVRE, B., BATOKO, H., DUBY, G. & BOUTRY, M. 2004. Targeting of a Nicotiana plumbaginifolia H⁺-ATPase to the plasma membrane is not by default and requires cytosolic structural determinants. *The Plant Cell*, 16, 1772-1789.
- LEMON, B. & TJIAN, R. 2000. Orchestrated response: a symphony of transcription factors for gene control. *Genes & development*, 14, 2551-2569.
- LI, G., LIU, R., XU, R., VARSHNEY, R. K., DING, H., LI, M., YAN, X., HUANG, S., LI, J. & WANG, D. 2023. Development of an Agrobacterium-mediated CRISPR/Cas9 system in pea (Pisum sativum L.). *The Crop Journal*, 11, 132-139.

- LI, H., TIAN, J., YAO, Y.-Y., ZHANG, J., SONG, T.-T., LI, K.-T. & YAO, Y.-C. 2019a. Identification of leucoanthocyanidin reductase and anthocyanidin reductase genes involved in proanthocyanidin biosynthesis in *Malus crabapple* plants. *Plant Physiology and Biochemistry*, 139, 141-151.
- LI, J.-F., NORVILLE, J. E., AACH, J., MCCORMACK, M., ZHANG, D., BUSH, J., CHURCH, G. M. & SHEEN, J. 2013. Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nature biotechnology*, 31, 688-691.
- LI, L., HE, Z., PANDEY, G. K., TSUCHIYA, T. & LUAN, S. 2002. Functional cloning and characterization of a plant efflux carrier for multidrug and heavy metal detoxification. *Journal of Biological Chemistry*, 277, 5360-5368.
- LI, S.-J., LIU, X.-J., XIE, X.-L., GRIERSON, D., YIN, X.-R. & CHEN, K.-S. 2015. CrMYB73, a PH-like gene, contributes to citric acid accumulation in citrus fruit. *Scientia Horticulturae*, 197, 212-217.
- LI, S.-J., YIN, X.-R., WANG, W.-L., LIU, X.-F., ZHANG, B. & CHEN, K.-S. 2017. Citrus CitNAC62 cooperates with CitWRKY1 to participate in citric acid degradation via up-regulation of CitAco3. *Journal of Experimental Botany*, 68, 3419-3426.
- LI, S. F., MILLIKEN, O. N., PHAM, H., SEYIT, R., NAPOLI, R., PRESTON, J., KOLTUNOW, A. M. & PARISH, R. W. 2009. The *Arabidopsis* MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. *The Plant Cell*, 21, 72-89.
- LI, Y., HE, H. & HE, L.-F. 2019b. Genome-wide analysis of the MATE gene family in potato. *Molecular biology reports*, 46, 403-414.
- LI, Y., PROVENZANO, S., BLIEK, M., SPELT, C., APPELHAGEN, I., MACHADO DE FARIA, L., VERWEIJ, W., SCHUBERT, A., SAGASSER, M., SEIDEL, T., WEISSHAAR, B., KOES, R. & QUATTROCCHIO, F. 2016. Evolution of tonoplast P-ATPase transporters involved in vacuolar acidification. *New Phytol*, 211, 1092-107.
- LI, Z. T., KIM, K.-H., JASINSKI, J. R., CREECH, M. R. & GRAY, D. J. 2012. Large-scale characterization of promoters from grapevine (*Vitis* spp.) using quantitative anthocyanin and GUS assay systems. *Plant science*, 196, 132-142.
- LIN, Q., WANG, C., DONG, W., JIANG, Q., WANG, D., LI, S., CHEN, M., LIU, C., SUN, C. & CHEN, K. 2015. Transcriptome and metabolome analyses of sugar and organic acid metabolism in Ponkan (*Citrus reticulata*) fruit during fruit maturation. *Gene*, 554, 64-74.
- LIND, J. 1757. *A Treatise on the Scurvy*, A. Millar.
- LIU, C., JUN, J. H. & DIXON, R. A. 2014. MYB5 and MYB14 play pivotal roles in seed coat polymer biosynthesis in *Medicago truncatula*. *Plant physiology*, 165, 1424-1439.
- LIU, C., LONG, J., ZHU, K., LIU, L., YANG, W., ZHANG, H., LI, L., XU, Q. & DENG, X. 2016a. Characterization of a citrus R2R3-MYB transcription factor that regulates the flavonol and hydroxycinnamic acid biosynthesis. *Scientific Reports*, 6, 1-16.

- LIU, C., WANG, X., SHULAEV, V. & DIXON, R. A. 2016b. A role for leucoanthocyanidin reductase in the extension of proanthocyanidins. *Nature Plants*, 2, 1-7.
- LIU, D., HU, R., PALLA, K. J., TUSKAN, G. A. & YANG, X. 2016c. Advances and perspectives on the use of CRISPR/Cas9 systems in plant genomics research. *Current Opinion in Plant Biology*, 30, 70-77.
- LIU, J., LI, Y., WANG, W., GAI, J. & LI, Y. 2016d. Genome-wide analysis of MATE transporters and expression patterns of a subgroup of MATE genes in response to aluminum toxicity in soybean. *BMC genomics*, 17, 1-15.
- LIU, J., MAGALHAES, J. V., SHAFF, J. & KOCHIAN, L. V. 2009. Aluminum-activated citrate and malate transporters from the MATE and ALMT families function independently to confer Arabidopsis aluminum tolerance. *The Plant Journal*, 57, 389-399.
- LIU, Y., WANG, K., CHENG, Q., KONG, D., ZHANG, X., WANG, Z., WANG, Q., XIE, Q., YAN, J. & CHU, J. 2020. Cysteine protease RD21A regulated by E3 ligase SINAT4 is required for drought-induced resistance to *Pseudomonas syringae* in Arabidopsis. *Journal of Experimental Botany*, 71, 5562-5576.
- LLOYD, A., BROCKMAN, A., AGUIRRE, L., CAMPBELL, A., BEAN, A., CANTERO, A. & GONZALEZ, A. 2017. Advances in the MYB-bHLH-WD repeat (MBW) pigment regulatory model: addition of a WRKY factor and co-option of an anthocyanin MYB for betalain regulation. *Plant and Cell Physiology*, 58, 1431-1441.
- LU, S., WANG, J., CHITSAZ, F., DERBYSHIRE, M. K., GEER, R. C., GONZALES, N. R., GWADZ, M., HURWITZ, D. I., MARCHLER, G. H. & SONG, J. S. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic acids research*, 48, D265-D268.
- LU, X., CAO, X., LI, F., LI, J., XIONG, J., LONG, G., CAO, S. & XIE, S. 2016. Comparative transcriptome analysis reveals a global insight into molecular processes regulating citrate accumulation in sweet orange (*Citrus sinensis*). *Physiologia Plantarum*, 158, 463-482.
- LV, X., ZHAO, S., NING, Z., ZENG, H., SHU, Y., TAO, O., XIAO, C., LU, C. & LIU, Y. 2015. Citrus fruits as a treasure trove of active natural metabolites that potentially provide benefits for human health. *Chemistry Central Journal*, 9, 68.
- MA, B., LIAO, L., FANG, T., PENG, Q., OGUTU, C., ZHOU, H., MA, F. & HAN, Y. 2019. A Ma10 gene encoding P-type ATPase is involved in fruit organic acid accumulation in apple. *Plant biotechnology journal*, 17, 674-686.
- MA, X., ZHANG, Q., ZHU, Q., LIU, W., CHEN, Y., QIU, R., WANG, B., YANG, Z., LI, H., LIN, Y., XIE, Y., SHEN, R., CHEN, S., WANG, Z., CHEN, Y., GUO, J., CHEN, L., ZHAO, X., DONG, Z. & LIU, Y. G. 2015. A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants. *Mol Plant*, 8, 1274-84.
- MARINOVA, K., POURCEL, L., WEDER, B., SCHWARZ, M., BARRON, D., ROUTABOUL, J.-M., DEBEAUJON, I. & KLEIN, M. 2007. The Arabidopsis MATE transporter TT12 acts as a vacuolar flavonoid/H⁺-antiporter active in proanthocyanidin-accumulating cells of the seed coat. *The Plant Cell*, 19, 2023-2038.

- MARON, L. G., GUIMARÃES, C. T., KIRST, M., ALBERT, P. S., BIRCHLER, J. A., BRADBURY, P. J., BUCKLER, E. S., COLUCCIO, A. E., DANILOVA, T. V. & KUDRNA, D. 2013. Aluminum tolerance in maize is associated with higher MATE1 gene copy number. *Proceedings of the National Academy of Sciences*, 110, 5241-5246.
- MARTINOIA, E., MAESHIMA, M. & NEUHAUS, H. E. 2007. Vacuolar transporters and their essential role in plant metabolism. *Journal of experimental botany*, 58, 83-102.
- MATHEWS, H., CLENDENNEN, S. K., CALDWELL, C. G., LIU, X. L., CONNORS, K., MATHEIS, N., SCHUSTER, D. K., MENASCO, D., WAGONER, W. & LIGHTNER, J. 2003. Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. *The Plant Cell*, 15, 1689-1703.
- MATSUI, K., HISANO, T., YASUI, Y., MORI, M., WALKER, A. R., MORISHITA, T. & KATSU, K. 2016. Isolation and characterization of genes encoding leucoanthocyanidin reductase (FeLAR) and anthocyanidin reductase (FeANR) in buckwheat (*Fagopyrum esculentum*). *Journal of plant physiology*, 205, 41-47.
- MATUS, J. T., POUPIN, M. J., CANON, P., BORDEU, E., ALCALDE, J. A. & ARCE-JOHNSON, P. 2010. Isolation of WDR and bHLH genes related to flavonoid synthesis in grapevine (*Vitis vinifera* L.). *Plant Molecular Biology*, 72, 607-620.
- MESSING, J., GRONENBORN, B., MÜLLER-HILL, B. & HOPSCHEIDER, P. H. 1977. Filamentous coliphage M13 as a cloning vehicle: insertion of a HindII fragment of the lac regulatory region in M13 replicative form in vitro. *Proceedings of the National Academy of Sciences*, 74, 3642-3646.
- MILLARD, P. S., KRAGELUND, B. B. & BUROW, M. 2019. R2R3 MYB transcription factors—Functions outside the DNA-binding domain. *Trends in plant science*, 24, 934-946.
- MOON, D.-G. & MIZUTANI, F. 2002. Changes in soluble solids, acidity, and abscisic acid contents in different portions of fruit during maturation of Satsuma mandarin. *HORTICULTURE ENVIRONMENT and BIOTECHNOLOGY*, 43, 107-112.
- MORITA, M., SHITAN, N., SAWADA, K., VAN MONTAGU, M. C., INZÉ, D., RISCHER, H., GOOSSENS, A., OKSMAN-CALDENTEY, K.-M., MORIYAMA, Y. & YAZAKI, K. 2009. Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in *Nicotiana tabacum*. *Proceedings of the National Academy of Sciences*, 106, 2447-2452.
- MORITA, Y., SAITOH, M., HOSHINO, A., NITASAKA, E. & IIDA, S. 2006. Isolation of cDNAs for R2R3-MYB, bHLH and WDR transcriptional regulators and identification of c and ca mutations conferring white flowers in the Japanese morning glory. *Plant and Cell Physiology*, 47, 457-470.
- MOUFIDA, S. D. & MARZOUK, B. 2003. Biochemical characterization of blood orange, sweet orange, lemon, bergamot and bitter orange. *Phytochemistry*, 62, 1283-1289.
- MÜLLER, M. & TAIZ, L. 2002. Regulation of the lemon-fruit V-ATPase by variable stoichiometry and organic acids. *Journal of Membrane Biology*, 185, 209-220.
- MÜLLER, M. L., IRKENS-KIESECKER, U., RUBINSTEIN, B. & TAIZ, L. 1996. On the Mechanism of Hyperacidification in Lemon: COMPARISON OF THE VACUOLAR H⁺-ATPase

- ACTIVITIES OF FRUITS AND EPICOTYLS (*). *Journal of Biological Chemistry*, 271, 1916-1924.
- NEKRASOV, V., STASKAWICZ, B., WEIGEL, D., JONES, J. D. & KAMOUN, S. 2013. Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nature biotechnology*, 31, 691-693.
- NESI, N., DEBEAUJON, I., JOND, C., PELLETIER, G., CABOCHE, M. & LEPINIEC, L. 2000. The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in *Arabidopsis* siliques. *The Plant Cell*, 12, 1863-1878.
- NESI, N., JOND, C., DEBEAUJON, I., CABOCHE, M. & LEPINIEC, L. 2001. The *Arabidopsis* TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *The Plant Cell*, 13, 2099-2114.
- NIEDZ, R. P. In vitro germination of citrus seed. *Proceedings of the Florida State Horticultural Society*, 2008. 148-151.
- NIEMANN, G. J. 1988. Distribution and evolution of the flavonoids in gymnosperms. *The flavonoids*. Springer.
- OGATA, K., HOJO, H., AIMOTO, S., NAKAI, T., NAKAMURA, H., SARAI, A., ISHII, S. & NISHIMURA, Y. 1992. Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. *Proceedings of the National Academy of Sciences*, 89, 6428-6432.
- PANG, Y., PEEL, G. J., WRIGHT, E., WANG, Z. & DIXON, R. A. 2007. Early steps in proanthocyanidin biosynthesis in the model legume *Medicago truncatula*. *Plant physiology*, 145, 601-615.
- PAOLOCCI, F., ROBBINS, M. P., MADEO, L., ARCIONI, S., MARTENS, S. & DAMIANI, F. 2007. Ectopic expression of a basic helix-loop-helix gene transactivates parallel pathways of proanthocyanidin biosynthesis. structure, expression analysis, and genetic control of leucoanthocyanidin 4-reductase and anthocyanidin reductase genes in *Lotus corniculatus*. *Plant Physiology*, 143, 504-516.
- PEDERSEN, A. G., BALDI, P., CHAUVIN, Y. & BRUNAK, S. 1999. The biology of eukaryotic promoter prediction—a review. *Computers & chemistry*, 23, 191-207.
- PEEL, G. J. & DIXON, R. A. 2007. Detection and quantification of engineered proanthocyanidins in transgenic plants. *Natural Product Communications*, 2, 1934578X0700201008.
- PENG, A., CHEN, S., LEI, T., XU, L., HE, Y., WU, L., YAO, L. & ZOU, X. 2017. Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene CsLOB1 promoter in citrus. *Plant Biotechnol J*.
- PÉREZ-DÍAZ, R., RYNGAJILLO, M., PÉREZ-DÍAZ, J., PEÑA-CORTÉS, H., CASARETTO, J. A., GONZÁLEZ-VILLANUEVA, E. & RUIZ-LARA, S. 2014. VvMATE1 and VvMATE2 encode putative proanthocyanidin transporters expressed during berry development in *Vitis vinifera* L. *Plant cell reports*, 33, 1147-1159.

- PEREZ-MOLPHE-BALCH, E. & OCHOA-ALEJO, N. 1998. Regeneration of transgenic plants of Mexican lime from *Agrobacterium rhizogenes* transformed tissues. *Plant Cell Reports*, 17, 591-596.
- PORTO, M. S., PINHEIRO, M. P. N., BATISTA, V. G. L., DOS SANTOS, R. C., DE ALBUQUERQUE MELO FILHO, P. & DE LIMA, L. M. 2014. Plant promoters: an approach of structure and function. *Molecular biotechnology*, 56, 38-49.
- PROUSE, M. B. & CAMPBELL, M. M. 2012. The interaction between MYB proteins and their target DNA binding sites. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1819, 67-77.
- QUATTROCCHIO, F., VERWEIJ, W., KROON, A., SPELT, C., MOL, J. & KOES, R. 2006. PH4 of *Petunia* is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell*, 18, 1274-91.
- QUATTROCCHIO, F., WING, J., VAN DER WOUDE, K., SOUER, E., DE VETTEN, N., MOL, J. & KOES, R. 1999. Molecular analysis of the anthocyanin2 gene of *petunia* and its role in the evolution of flower color. *The Plant Cell*, 11, 1433-1444.
- RAN, F., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, 8, 2281-2308.
- RENTSCH, D. & MARTINOIA, E. 1991. Citrate transport into barley mesophyll vacuoles—comparison with malate-uptake activity. *Planta*, 184, 532-537.
- REUTHER, W., WEBBER, H. J. & BATCHELOR, L. D. 1967. History, World Distribution, Botany, and Varieties. University of California, Division of Agricultural Sciences.
- ROGERS, E. E., WU, X., STACEY, G. & NGUYEN, H. T. 2009. Two MATE proteins play a role in iron efficiency in soybean. *Journal of Plant Physiology*, 166, 1453-1459.
- RÜTHER, U. 1980. Construction and properties of a new cloning vehicle, allowing direct screening for recombinant plasmids. *Molecular and General Genetics MGG*, 178, 475-477.
- SADKA, A., DAHAN, E., OR, E., ROOSE, M. L., MARSH, K. B. & COHEN, L. 2001. Comparative analysis of mitochondrial citrate synthase gene structure, transcript level and enzymatic activity in acidless and acid-containing Citrus varieties. *Functional Plant Biology*, 28, 383-390.
- SANTOS, A. L. D., CHAVES-SILVA, S., YANG, L., MAIA, L. G. S., CHALFUN-JÚNIOR, A., SINHAROY, S., ZHAO, J. & BENEDITO, V. A. 2017. Global analysis of the MATE gene family of metabolite transporters in tomato. *BMC plant biology*, 17, 1-13.
- SAWAKI, Y., KIHARA-DOI, T., KOBAYASHI, Y., NISHIKUBO, N., KAWAZU, T., KOBAYASHI, Y., KOYAMA, H. & SATO, S. 2013. Characterization of Al-responsive citrate excretion and citrate-transporting MATEs in *Eucalyptus camaldulensis*. *Planta*, 237, 979-989.
- SCHUMACHER, K. & KREBS, M. 2010. The V-ATPase: small cargo, large effects. *Current opinion in plant biology*, 13, 724-730.

- SCORA, R. W. 1975. On the History and Origin of Citrus. *Bulletin of the Torrey Botanical Club*, 102, 369-375.
- SEDDON, T. J. & DOWNEY, M. O. 2008. Comparison of analytical methods for the determination of condensed tannins in grape skin. *Australian Journal of Grape and Wine Research*, 14, 54-61.
- SERRANO, M., WANG, B., ARYAL, B., GARCION, C., ABOU-MANSOUR, E., HECK, S., GEISLER, M., MAUCH, F., NAWRATH, C. & MÉTRAUX, J.-P. 2013. Export of salicylic acid from the chloroplast requires the multidrug and toxin extrusion-like transporter EDS5. *Plant physiology*, 162, 1815-1821.
- SHAHMURADOV, I. A., SOLOVYEV, V. V. & GAMMERMAN, A. 2005. Plant promoter prediction with confidence estimation. *Nucleic acids research*, 33, 1069-1076.
- SHANGGUAN, X., YANG, Q., WU, X. & CAO, J. 2021. Function analysis of a cotton R2R3 MYB transcription factor GhMYB3 in regulating plant trichome development. *Plant Biology*, 23, 1118-1127.
- SHARAN, C., HAMILTON, N. M., PARL, A. K., SINGH, P. K. & CHAUDHURI, G. 1999. Identification and characterization of a transcriptional silencer upstream of the human BRCA2 gene. *Biochemical and biophysical research communications*, 265, 285-290.
- SHEN, J., ZENG, Y., ZHUANG, X., SUN, L., YAO, X., PIMPL, P. & JIANG, L. 2013. Organelle pH in the Arabidopsis endomembrane system. *Molecular plant*, 6, 1419-1437.
- SHI, C.-Y., HUSSAIN, S. B., HAN, H., ALAM, S. M., LIU, D. & LIU, Y.-Z. 2021. Reduced expression of CsPH8, a P-type ATPase gene, is the major factor leading to the low citrate accumulation in citrus leaves. *Plant Physiology and Biochemistry*, 160, 211-217.
- SHI, C.-Y., HUSSAIN, S. B., YANG, H., BAI, Y.-X., KHAN, M. A. & LIU, Y.-Z. 2019. CsPH8, a P-type proton pump gene, plays a key role in the diversity of citric acid accumulation in citrus fruits. *Plant science*, 289, 110288.
- SHI, C.-Y., SONG, R.-Q., HU, X.-M., LIU, X., JIN, L.-F. & LIU, Y.-Z. 2015. Citrus PH5-like H⁺-ATPase genes: identification and transcript analysis to investigate their possible relationship with citrate accumulation in fruits. *Frontiers in plant science*, 6, 135.
- SHIMADA, T., NAKANO, R., SHULAEV, V., SADKA, A. & BLUMWALD, E. 2006. Vacuolar citrate/H⁺ symporter of citrus juice cells. *Planta*, 224, 472-480.
- SHIN, R., BURCH, A. Y., HUPPERT, K. A., TIWARI, S. B., MURPHY, A. S., GUILFOYLE, T. J. & SCHACHTMAN, D. P. 2007. The Arabidopsis transcription factor MYB77 modulates auxin signal transduction. *The Plant Cell*, 19, 2440-2453.
- SHIRLEY, B. W., HANLEY, S. & GOODMAN, H. M. 1992. Effects of ionizing radiation on a plant genome: analysis of two Arabidopsis transparent testa mutations. *The Plant Cell*, 4, 333-347.
- SHOJI, T., INAI, K., YAZAKI, Y., SATO, Y., TAKASE, H., SHITAN, N., YAZAKI, K., GOTO, Y., TOYOOKA, K. & MATSUOKA, K. 2009. Multidrug and toxic compound extrusion-type

- transporters implicated in vacuolar sequestration of nicotine in tobacco roots. *Plant physiology*, 149, 708-718.
- SMALE, S. T. 2001. Core promoters: active contributors to combinatorial gene regulation. *Genes & development*, 15, 2503-2508.
- SOLOVYEV, V., KOSAREV, P., SELEDOV, I. & VOROBYEV, D. 2006. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome biology*, 7, 1-12.
- SOLTESZ, A., VAGUJFALVI, A., RIZZA, F., KEREPESI, I., GALIBA, G., CATTIVELLI, L., CORAGGIO, I. & CROSATTI, C. 2012. The rice Osmyb4 gene enhances tolerance to frost and improves germination under unfavourable conditions in transgenic barley plants. *Journal of Applied Genetics*, 53, 133-143.
- SPELT, C. 2002. ANTHOCYANIN1 of Petunia Controls Pigment Synthesis, Vacuolar pH, and Seed Coat Development by Genetically Distinct Mechanisms. *The Plant Cell Online*, 14, 2121-2135.
- STAFFORD, H. A. & LESTER, H. H. 1986. Proanthocyanidins in needles from six genera of the Taxodiaceae. *American journal of botany*, 73, 1555-1562.
- STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312-1313.
- STRACKE, R., TURGUT-KARA, N. & WEISSHAAR, B. 2017. The AtMYB12 activation domain maps to a short C-terminal region of the transcription factor. *Zeitschrift für Naturforschung C*, 72, 251-257.
- STRACKE, R., WERBER, M. & WEISSHAAR, B. 2001. The R2R3-MYB gene family in Arabidopsis thaliana. *Current opinion in plant biology*, 4, 447-456.
- STRAZZER, P., SPELT, C. E., LI, S. J., BLIEK, M., FEDERICI, C. T., ROOSE, M. L., KOES, R. & QUATTROCCHIO, F. M. 2019. Hyperacidification of Citrus fruits by a vacuolar proton-pumping P-ATPase complex. *Nature Communications*, 10.
- SUNDARAMOORTHY, J., PARK, G. T., LEE, J.-D., KIM, J. H., SEO, H. S. & SONG, J. T. 2020. A P3A-type ATPase and an R2R3-MYB transcription factor are involved in vacuolar acidification and flower coloration in soybean. *Frontiers in Plant Science*, 11, 1880.
- SWINGLE, W. T. & REECE, P. C. 1967. The botany of Citrus and its wild relatives. *The citrus industry*, 190-430.
- TAN, B., LI, D. L., XU, S. X., FAN, G. E., FAN, J. & GUO, W. W. 2009. Highly efficient transformation of the GFP and MAC12.2 genes into precocious trifoliolate orange (*Poncirus trifoliata* L. Raf), a potential model genotype for functional genomics studies in Citrus. *Tree Genetics & Genomes*, 5, 529-537.
- TANAKA, T. 1977. Fundamental discussion of Citrus classification. *Studia Citrologica*, 14, 1-6.
- TANNER, G. J., FRANCKI, K. T., ABRAHAMS, S., WATSON, J. M., LARKIN, P. J. & ASHTON, A. R. 2003. Proanthocyanidin biosynthesis in plants: purification of legume leucoanthocyanidin reductase and molecular cloning of its cDNA. *Journal of Biological Chemistry*, 278, 31647-31656.

- TERRIER, N., TORREGROSA, L., AGEORGES, A., VIALET, S., VERRIES, C., CHEYNIER, V. & ROMIEU, C. 2009. Ectopic Expression of VvMybPA2 Promotes Proanthocyanidin Biosynthesis in Grapevine and Suggests Additional Targets in the Pathway. *Plant Physiology*, 149, 1028-1041.
- THOMPSON, E. P., WILKINS, C., DEMIDCHIK, V., DAVIES, J. M. & GLOVER, B. J. 2010. An Arabidopsis flavonoid transporter is required for anther dehiscence and pollen development. *Journal of experimental botany*, 61, 439-451.
- TOVKACH, A., RYAN, P. R., RICHARDSON, A. E., LEWIS, D. C., RATHJEN, T. M., RAMESH, S., TYERMAN, S. D. & DELHAIZE, E. 2013. Transposon-mediated alteration of TaMATE1B expression in wheat confers constitutive citrate efflux from root apices. *Plant Physiology*, 161, 880-892.
- TREUTTER, D. 1989. Chemical reaction detection of catechins and proanthocyanidins with 4-dimethylaminocinnamaldehyde. *Journal of Chromatography A*, 467, 185-193.
- URAO, T., NOJI, M. A., YAMAGUCHI-SHINOZAKI, K. & SHINOZAKI, K. 1996. A transcriptional activation domain of ATMYB2, a drought-inducible Arabidopsis Myb-related protein. *The Plant Journal*, 10, 1145-1148.
- VANNINI, C., LOCATELLI, F., BRACALE, M., MAGNANI, E., MARSONI, M., OSNATO, M., MATTANA, M., BALDONI, E. & CORAGGIO, I. 2004. Overexpression of the rice Osmby4 gene increases chilling and freezing tolerance of Arabidopsis thaliana plants. *The Plant Journal*, 37, 115-127.
- VENNAPUSA, A. R., SOMAYANDA, I. M., DOHERTY, C. J. & JAGADISH, S. 2020. A universal method for high-quality RNA extraction from plant tissues rich in starch, proteins and fiber. *Scientific reports*, 10, 1-13.
- VERWEIJ, W., SPELT, C., DI SANSEBASTIANO, G. P., VERMEER, J., REALE, L., FERRANTI, F., KOES, R. & QUATTROCCHIO, F. 2008. An H⁺ P-ATPase on the tonoplast determines vacuolar pH and flower colour. *Nat Cell Biol*, 10, 1456-62.
- VERWEIJ, W., SPELT, C. E., BLIEK, M., DE VRIES, M., WIT, N., FARACO, M., KOES, R. & QUATTROCCHIO, F. M. 2016. Functionally Similar WRKY Proteins Regulate Vacuolar Acidification in Petunia and Hair Development in Arabidopsis. *Plant Cell*, 28, 786-803.
- WANG, L., RAN, L., HOU, Y., TIAN, Q., LI, C., LIU, R., FAN, D. & LUO, K. 2017. The transcription factor MYB115 contributes to the regulation of proanthocyanidin biosynthesis and enhances fungal resistance in poplar. *New Phytologist*, 215, 351-367.
- WANG, R., LIU, X., LIANG, S., GE, Q., LI, Y., SHAO, J., QI, Y., AN, L. & YU, F. 2015. A subgroup of MATE transporter genes regulates hypocotyl cell elongation in Arabidopsis. *Journal of experimental botany*, 66, 6327-6343.
- WU, G. A., PROCHNIK, S., JENKINS, J., SALSE, J., HELLSTEN, U., MURAT, F., PERRIER, X., RUIZ, M., SCALABRIN, S. & TEROL, J. 2014a. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature biotechnology*, 32, 656.
- WU, G. A., TEROL, J., IBANEZ, V., LOPEZ-GARCIA, A., PEREZ-ROMAN, E., BORREDA, C., DOMINGO, C., TADEO, F. R., CARBONELL-CABALLERO, J., ALONSO, R., CURK, F., DU,

- D., OLLITRAULT, P., ROOSE, M. L., DOPAZO, J., GMITTER, F. G., ROKHSAR, D. S. & TALON, M. 2018. Genomics of the origin and evolution of Citrus. *Nature*, 554, 311-316.
- WU, X., LI, R., SHI, J., WANG, J., SUN, Q., ZHANG, H., XING, Y., QI, Y., ZHANG, N. & GUO, Y.-D. 2014b. Brassica oleracea MATE encodes a citrate transporter and enhances aluminum tolerance in Arabidopsis thaliana. *Plant and Cell Physiology*, 55, 1426-1436.
- XIAO, X., MA, F., CHEN, C.-L. & GUO, W.-W. 2014. High efficient transformation of auxin reporter gene into trifoliolate orange via Agrobacterium rhizogenes-mediated co-transformation. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 118, 137-146.
- XIE, D.-Y., SHARMA, S. B., PAIVA, N. L., FERREIRA, D. & DIXON, R. A. 2003. Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. *Science*, 299, 396-399.
- XU, L., SHEN, Z.-L., CHEN, W., SI, G.-Y., MENG, Y., GUO, N., SUN, X., CAI, Y.-P., LIN, Y. & GAO, J.-S. 2019. Phylogenetic analysis of upland cotton MATE gene family reveals a conserved subfamily involved in transport of proanthocyanidins. *Molecular biology reports*, 46, 161-175.
- XU, Q., CHEN, L.-L., RUAN, X., CHEN, D., ZHU, A., CHEN, C., BERTRAND, D., JIAO, W.-B., HAO, B.-H. & LYON, M. P. 2013. The draft genome of sweet orange (Citrus sinensis). *Nature genetics*, 45, 59.
- XU, W., GRAIN, D., BOBET, S., LE GOURRIEREC, J., THÉVENIN, J., KELEMEN, Z., LEPINIEC, L. & DUBOS, C. 2014. Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB–b HLH–WDR complexes and their targets in Arabidopsis seed. *New phytologist*, 202, 132-144.
- XU, Y., ZHANG, L., LU, L., LIU, J., YI, H. & WU, J. 2022. An efficient CRISPR/Cas9 system for simultaneous editing two target sites in Fortunella hindsii. *Horticulture Research*.
- YAMAKI, Y. T. 1989. Organic acids in the juice of citrus fruits. *Journal of the Japanese Society for Horticultural Science*, 58, 587-594.
- YAN, L., WEI, S., WU, Y., HU, R., LI, H., YANG, W. & XIE, Q. 2015. High-efficiency genome editing in Arabidopsis using YAO promoter-driven CRISPR/Cas9 system. *Molecular plant*, 8, 1820-1823.
- YAN, S., CHEN, N., HUANG, Z., LI, D., ZHI, J., YU, B., LIU, X., CAO, B. & QIU, Z. 2020. Anthocyanin Fruit encodes an R2R3-MYB transcription factor, SIAN2-like, activating the transcription of SIMYBATV to fine-tune anthocyanin content in tomato fruit. *New Phytologist*, 225, 2048-2063.
- YANG, L., HU, W., XIE, Y. M., LI, Y. & DENG, Z. N. 2016. Factors affecting Agrobacterium-mediated transformation efficiency of kumquat seedling internodal stem segments. *Scientia Horticulturae*, 209, 105-112.
- YANG, L., XU, C. J., HU, G. B. & CHEN, K. S. 2007. Establishment of an Agrobacterium-mediated transformation system for Fortunella crassifolia. *Biologia Plantarum*, 51, 541-545.

- YOKOSHO, K., YAMAJI, N. & MA, J. F. 2011. An Al-inducible MATE gene is involved in external detoxification of Al in rice. *The Plant Journal*, 68, 1061-1069.
- YOKOSHO, K., YAMAJI, N., UENO, D., MITANI, N. & MA, J. F. 2009. OsFRDL1 is a citrate transporter required for efficient translocation of iron in rice. *Plant physiology*, 149, 297-305.
- YU, K. Q., XU, Q., DA, X. L., GUO, F., DING, Y. D. & DENG, X. X. 2012. Transcriptome changes during fruit development and ripening of sweet orange (*Citrus sinensis*). *Bmc Genomics*, 13, 13.
- ZHANG, F., LEBLANC, C., IRISH, V. F. & JACOB, Y. 2017. Rapid and efficient CRISPR/Cas9 gene editing in Citrus using the YAO promoter. *Plant Cell Rep*, 36, 1883-1887.
- ZHANG, H., ZHANG, J., WEI, P., ZHANG, B., GOU, F., FENG, Z., MAO, Y., YANG, L., ZHANG, H. & XU, N. 2014. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant biotechnology journal*, 12, 797-807.
- ZHANG, J. C., TAO, N. G., XU, Q., ZHOU, W. J., CAO, H. B., XU, J. A. & DENG, X. X. 2009. Functional characterization of Citrus PSY gene in Hongkong kumquat (*Fortunella hindsii* Swingle). *Plant Cell Reports*, 28, 1737-1746.
- ZHANG, W., LIAO, L., XU, J., HAN, Y. & LI, L. 2021. Genome-wide identification, characterization and expression analysis of MATE family genes in apple (*Malus domestica* Borkh). *BMC genomics*, 22, 1-14.
- ZHANG, Y., YE, J., LIU, C., XU, Q., LONG, L. & DENG, X. 2020. Citrus PH4–Noemi regulatory complex is involved in proanthocyanidin biosynthesis via a positive feedback loop. *Journal of experimental botany*, 71, 1306-1321.
- ZHAO, J. & DIXON, R. A. 2009. MATE transporters facilitate vacuolar uptake of epicatechin 3'-O-glucoside for proanthocyanidin biosynthesis in *Medicago truncatula* and *Arabidopsis*. *The Plant Cell*, 21, 2323-2340.
- ZHAO, J., HUHMANN, D., SHADLE, G., HE, X.-Z., SUMNER, L. W., TANG, Y. & DIXON, R. A. 2011. MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside malonates in *Medicago truncatula*. *The Plant Cell*, 23, 1536-1555.
- ZHENG, J., WU, H., ZHAO, M., YANG, Z., ZHOU, Z., GUO, Y., LIN, Y. & CHEN, H. 2021. OsMYB3 is a R2R3-MYB gene responsible for anthocyanin biosynthesis in black rice. *Molecular Breeding*, 41, 1-15.
- ZHI, J., LIU, X., LI, D., HUANG, Y., YAN, S., CAO, B. & QIU, Z. 2020. CRISPR/Cas9-mediated SIAN2 mutants reveal various regulatory models of anthocyanin biosynthesis in tomato plant. *Plant Cell Reports*, 39, 799-809.
- ZHOU, H., LIN-WANG, K., LIAO, L., GU, C., LU, Z., ALLAN, A. C. & HAN, Y. 2015a. Peach MYB7 activates transcription of the proanthocyanidin pathway gene encoding leucoanthocyanidin reductase, but not anthocyanidin reductase. *Frontiers in Plant Science*, 6, 908.
- ZHOU, M., SUN, Z., WANG, C., ZHANG, X., TANG, Y., ZHU, X., SHAO, J. & WU, Y. 2015b. Changing a conserved amino acid in R2R3-MYB transcription repressors results in

cytoplasmic accumulation and abolishes their repressive activity in Arabidopsis. *The Plant Journal*, 84, 395-403.

ZHOU, Y., HE, W., ZHENG, W., TAN, Q., XIE, Z., ZHENG, C. & HU, C. 2018. Fruit sugar and organic acid were significantly related to fruit Mg of six citrus cultivars. *Food chemistry*, 259, 278-285.

ZHU, C., ZHENG, X., HUANG, Y., YE, J., CHEN, P., ZHANG, C., ZHAO, F., XIE, Z., ZHANG, S. & WANG, N. 2019. Genome sequencing and CRISPR/Cas9 gene editing of an early flowering Mini-Citrus (*Fortunella hindsii*). *Plant biotechnology journal*, 17, 2199-2210.

ZIMMERMANN, I. M., HEIM, M. A., WEISSHAAR, B. & UHRIG, J. F. 2004. Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. *The Plant Journal*, 40, 22-34.

Appendices

Supplementary Table 1 Plasmids used in this thesis, their description, selection, and construction method. OE: overexpression plasmid; DL: dual-luciferase reporter plasmid; Kan: kanamycin.

Plasmid	Description	Selection	Construction
pEA0013	CRISPR-Cas9 – sgRNA_002+019	Kan+ccdB	GoldenGate
pEA0015	CRISPR-Cas9 – sgRNA_003+011	Kan+ccdB	GoldenGate
pEA0016	CRISPR-Cas9 – empty	Kan+ccdB	GoldenGate
pEA0020	OE – Nicole	Kan	Conventional
pEA0021	OE – Noemi	Kan	Conventional
pEA0022	OE – nicole ^{soro}	Kan	Conventional
pEA0027	DL – pCsPH1	Kan	Conventional
pEA0030	OE – pCsPH5	Kan	Conventional
pEA0031	OE – pCsTT12	Kan	Conventional
pEA0032	OE – pNoemi	Kan	Conventional
pEA0033	OE – pNicole	Kan	Conventional
pEA0037	OE – CsPH3	Kan	Conventional
pEA0038	OE – Iris	Kan	Conventional

Supplementary Table 2 Media recipes used in Chapter 5. MS: Murashige and Skoog.

Media	Component	Concentration
Germination MS		
	Agar	8 g L ⁻¹
	Sucrose	30 mg L ⁻¹
Co-culture MS		
	Agar	8 g L ⁻¹
	Sucrose	40 mg L ⁻¹
	Acetosyringone	100 µM
Root induction MS		
	Agar	8 g L ⁻¹
	Sucrose	30 mg L ⁻¹
	Cefotaxime	400 mg L ⁻¹
	Kanamycin	50 mg L ⁻¹

Supplementary Table 3 List of gene accession IDs from the *C. sinensis* v1.1 Phytozome genome, nomenclature based on phylogenetic analyses, BLAST analyses and naming from previous publications (Wu et al., 2014a).

Accession ID	Phylogeny	My designation	Previous designation	Source
orange1.1g000304m.g	CsR2R3MYB1			
orange1.1g001863m.g		CsAco3	CitAco3	(Chen et al., 2012)
orange1.1g001917m.g		CsAco2	CitAco2	(Chen et al., 2012)
orange1.1g002610m.g		CsAco1	CitAco1	(Chen et al., 2012)
orange1.1g002768m.g		CsPH5	CsPH8	(Shi et al., 2015)
orange1.1g003194m.g		CsGS4	CitGS2	(Chen et al., 2012)
orange1.1g006420m.g	CsR2R3MYB2			
orange1.1g007548m.g	CsMATE1			
orange1.1g008290m.g	CsMATE2			
orange1.1g008695m.g	CsMATE3			
orange1.1g008783m.g	CsMATE4			
orange1.1g008932m.g	CsMATE5			
orange1.1g009041m.g		CsIDH3	CitIDH3	(Chen et al., 2012)
orange1.1g009192m.g	CsMATE6			
orange1.1g009588m.g	CsMATE7			
orange1.1g010025m.g	CsMATE8			
orange1.1g010121m.g	CsMATE9			
orange1.1g010159m.g	CsMATE10			
orange1.1g010173m.g	CsMATE11			
orange1.1g010304m.g		CsCS2	CitCS2	(Chen et al., 2013)
orange1.1g010345m.g	CsMATE12			
orange1.1g010561m.g	CsMATE13			
orange1.1g010798m.g	CsMATE14			
orange1.1g010889m.g	CsMATE15			
orange1.1g010897m.g		CsGAD5	CitGAD4	(Chen et al., 2012)
orange1.1g010912m.g	CsMATE16			
orange1.1g011035m.g	CsMATE17			
orange1.1g011057m.g		CsGAD4	CitGAD4	(Chen et al., 2012)
orange1.1g011062m.g	CsMATE18			
orange1.1g011186m.g	CsMATE19			
orange1.1g011296m.g	CsMATE20			
orange1.1g011371m.g	CsMATE21			
orange1.1g011666m.g	CsMATE22			
orange1.1g011678m.g	CsMATE23			
orange1.1g011857m.g	CsMATE24			
orange1.1g012107m.g		CsCS1	CitCS1	(Chen et al., 2013)
orange1.1g012113m.g	CsMATE25			
orange1.1g012464m.g	CsMATE26			
orange1.1g013300m.g	CsR2R3MYB3			
orange1.1g013478m.g		CsGS1	CitGS2	(Chen et al., 2012)
orange1.1g014135m.g	CsR2R3MYB4			
orange1.1g014629m.g		CsPH3		
orange1.1g014829m.g	CsR2R3MYB5		CsMYBF3	(Liu et al., 2016a)
orange1.1g015012m.g		CsIDH1	CitIDH1	(Chen et al., 2012)
orange1.1g016021m.g	CsR2R3MYB6			
orange1.1g016708m.g	CsR2R3MYB7			
orange1.1g017727m.g	CsR2R3MYB8			
orange1.1g018154m.g	CsR2R3MYB9			
orange1.1g018369m.g		CsLDOX		

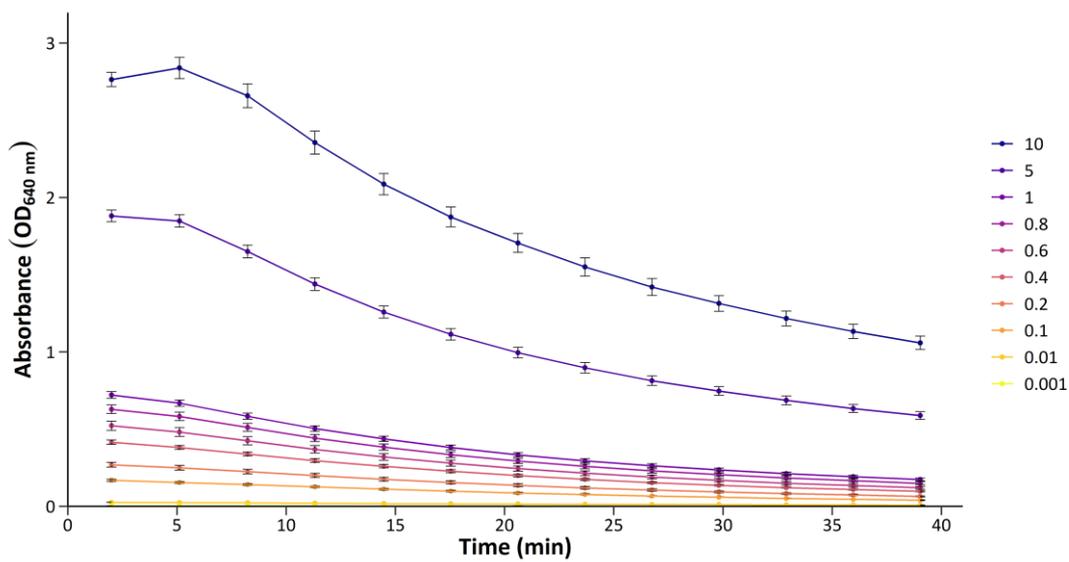
orange1.1g018391m.g	CsGS2	CitGS2	(Chen et al., 2012)
orange1.1g018434m.g	CsGS3	CitGS2	(Chen et al., 2012)
orange1.1g018559m.g	CsR2R3MYB10		
orange1.1g018746m.g	CsR2R3MYB11		
orange1.1g019307m.g	CsR2R3MYB12		
orange1.1g019787m.g	CsR2R3MYB13		
orange1.1g019911m.g	CsR2R3MYB14		
orange1.1g020197m.g	CsR2R3MYB15		
orange1.1g020603m.g	CsR2R3MYB16		
orange1.1g020613m.g	CsR2R3MYB17		
orange1.1g020997m.g	CsR2R3MYB18		
orange1.1g021188m.g	CsR2R3MYB19		
orange1.1g021220m.g	CsR2R3MYB20		
orange1.1g021486m.g	CsR2R3MYB21		
orange1.1g021756m.g	CsR2R3MYB22		
orange1.1g021816m.g	CsR2R3MYB23		
orange1.1g022439m.g	CsR2R3MYB24		
orange1.1g023056m.g	CsR2R3MYB25		
orange1.1g023196m.g	CsR2R3MYB26		
orange1.1g024441m.g	CsR2R3MYB27		
orange1.1g024492m.g	CsR2R3MYB28		
orange1.1g024849m.g	CsR2R3MYB29		
orange1.1g025602m.g	CsR2R3MYB30	Marys	
orange1.1g026855m.g	CsR2R3MYB31		
orange1.1g028843m.g	CsR2R3MYB32		
orange1.1g028922m.g	CsR2R3MYB33		
orange1.1g035514m.g	CsMATE27		
orange1.1g035612m.g	CsCit1	CsCit1	(Shimada et al., 2006)
orange1.1g035629m.g	CsR2R3MYB34		
orange1.1g035985m.g	CsANR2		
orange1.1g036215m.g	CsR2R3MYB35		
orange1.1g036344m.g	CsR2R3MYB36		
orange1.1g036591m.g	CsR2R3MYB37		
orange1.1g037024m.g	CsR2R3MYB38		
orange1.1g037703m.g	CsMATE28		
orange1.1g037798m.g	Noemi		
orange1.1g037956m.g	CsR2R3MYB39		
orange1.1g037998m.g	CsR2R3MYB40		
orange1.1g038074m.g	CsANR1		
orange1.1g038533m.g	CsTTG1		
orange1.1g039016m.g	CsR2R3MYB41		
orange1.1g039070m.g	CsR2R3MYB42		
orange1.1g039198m.g	CsR2R3MYB43		
orange1.1g039708m.g	CsR2R3MYB44		
orange1.1g040253m.g	CsDFR		
orange1.1g040502m.g	CsR2R3MYB45		
orange1.1g040623m.g	CsIDH2	CitIDH2	(Chen et al., 2012)
orange1.1g040653m.g	Miriam		
orange1.1g040726m.g	CsR2R3MYB46	Nicole	CrMYB73 (Li et al., 2015)
orange1.1g040841m.g	CsR2R3MYB47	Iris	
orange1.1g041081m.g	CsR2R3MYB48		CsMYBF1 (Liu et al., 2016a)
orange1.1g041991m.g	CsR2R3MYB49		
orange1.1g042286m.g	CsF3H		
orange1.1g042512m.g	CsMATE29		

orange1.1g042846m.g	CsR2R3MYB50		
orange1.1g042947m.g	CsMATE30		
orange1.1g043161m.g	CsMATE31		
orange1.1g043269m.g	CsR2R3MYB51		
orange1.1g043557m.g	CsR2R3MYB52		
orange1.1g043612m.g	CsR2R3MYB53		
orange1.1g044864m.g	CsR2R3MYB54		
orange1.1g045328m.g	CsMATE32		
orange1.1g045384m.g	CsR2R3MYB55		
orange1.1g045387m.g	CsR2R3MYB56		
orange1.1g045400m.g	CsR2R3MYB57		
orange1.1g045411m.g	CsR2R3MYB58		
orange1.1g045434m.g	CsR2R3MYB59		
orange1.1g045583m.g	CsR2R3MYB60		
orange1.1g045659m.g	CsMATE33		
orange1.1g045750m.g		CsPH1	
orange1.1g046061m.g	CsMATE34	CsTT12	
orange1.1g046075m.g	CsR2R3MYB61		
orange1.1g046093m.g	CsR2R3MYB62	CsMYBF2	(Liu et al., 2016a)
orange1.1g046137m.g		CsLAR	
orange1.1g046419m.g	CsR2R3MYB63		
orange1.1g047101m.g	CsR2R3MYB64		
orange1.1g047269m.g	CsR2R3MYB65		
orange1.1g047839m.g	CsR2R3MYB66		
orange1.1g047854m.g	CsR2R3MYB67		
orange1.1g048224m.g	CsR2R3MYB68		
orange1.1g048788m.g	CsMATE35		

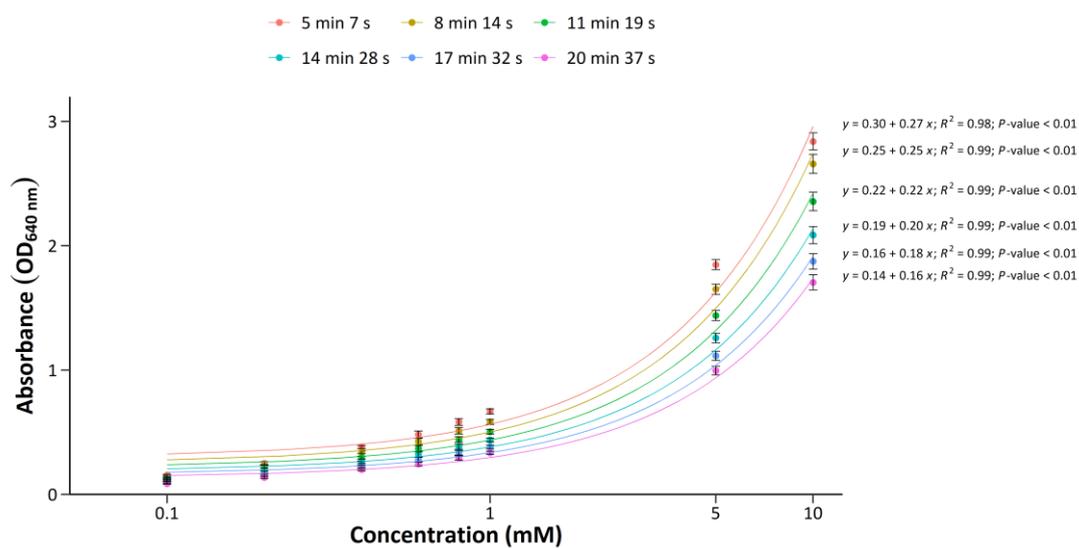
Supplementary Table 4 List of primers used in this thesis, sequence, target and purpose. F: forward; R: reverse.

ID	Sequence	Target	Purpose	Strand
EA-007	TGTGGTCTCAAGCGTAATGCCAACTTTGTAC	pICSL70001-Rev	Amplification	R
EA-009	TGTGGTCTCAATTGGCAACATGGCGTGCTCTTATGTTTTAGA GCTAGAAATAGCAAG	pICSL70001-Nicole-gRNA_002	Amplification	F
EA-025	TGTGGTCTCAATTGGAGCACGCCATGTTGCAGCAGTTTTAGA GCTAGAAATAGCAAG	pICSL70001-Nicole-gRNA_003	Amplification	F
EA-027	TGTGGTCTCAATTGGGTGGCGAACTCTGCCAAAAGTTTTAGA GCTAGAAATAGCAAG	pICSL70001-Nicole-gRNA_011	Amplification	F
EA-028	TGTGGTCTCAATTGGATATGTCCCTCGTTTAAACGGGTTTTAGA GCTAGAAATAGCAAG	pICSL70001-Nicole-gRNA_019	Amplification	F
EA-055	TGTGGTCTCAAGCGGTCAACATGGTGGAGCAC	2x 35S-Cas9-nos	Amplification	F
EA-056	TGTGGTCTCAGGAGTCGATCTAGTAACATAGATGACAC	2x 35S-Cas9-nos	Amplification	R
EA-176	TCTTCCATGGGACTAATTTTAGTCACTAAGGTG	pCsTT12 NcoI -3'	Amplification	R
EA-181	CGAGGTCGACAGTTATCAATTTTGAAGATAGTCT	5'- Sall pCsTT12	Amplification	F
EA-199	TCTTCCATGGTGAACAAATTTTATTGAATGA	pCsPH1 NcoI -3'	Amplification	R
EA-200	CGAGGTCGACGATCAAGAAAAGTCTAATAACACATG	5'- Sall pNoemi	Amplification	F
EA-201	TCTTCCATGGTGAAAGACCCGCTGGTGA	pNoemi NcoI -3'	Amplification	R
EA-202	CGAGGTCGACGTTACCGCATTAGTTTGGGA	5'- Sall pNicole	Amplification	F
EA-203	TCTTCCATGGTCATTTTATAGTAACATTTTATGTA	pNicole NcoI -3'	Amplification	R
EA-204	CGAGCTGCAGGGATTGACCTTTGAGTGC	5'- PstI pCsPH1	Amplification	F
EA-209	CCAAGCAGCATGAAGATCAA	CsActin	qPCR	F
EA-210	ATCTGCTGGAAGGTGCTGAG	CsActin	qPCR	R
EA-213	ACAAGATGGACGCCACCAC	CsEF1a	qPCR	F
EA-214	CAGGGTTGGACCCTTGTACC	CsEF1a	qPCR	R
EA-215	CGGTAACCGATGGTCTCTGA	Nicole	qPCR	F
EA-216	TCCCTTGGCTTATCAGCTTCT	Nicole	qPCR	R
EA-217	CTTCCGGAGTTGGGTACCA	Noemi	qPCR	F
EA-218	TCCTCCGGGACCTTTTCTGT	Noemi	qPCR	R
EA-221	TCGAAACACCTGGTAGTAAGCC	CsPH1	qPCR	F
EA-222	ACATTTGTGCCCATGAAGCAG	CsPH1	qPCR	R
EA-223	CTTTGGAAAAGCAGCACA	CsPH5	qPCR	F
EA-224	ATCCCTGTATGATCGATGCT	CsPH5	qPCR	R
EA-229	ACTCCAAGCGACTATACAGAGG	CsLDOX	qPCR	F
EA-230	TCCAAGCCAAGTGACAACA	CsLDOX	qPCR	R
EA-233	ACCCACCCACCAAGCTAATG	CsTTG1	qPCR	F
EA-234	GAGCGAAAACACTACTGCTC	CsTTG1	qPCR	R
EA-235	AGATGACTGGATGGATGTATTTTGTGT	CsDFR	qPCR	F
EA-236	TGAGACTGGGTGGCATTGAC	CsDFR	qPCR	R
EA-245	GTGGTTCACTTTCTGGTTTGAAG	CsPH3	qPCR	F
EA-246	TTGGTTTGCAGTGTAGGTTGTC	CsPH3	qPCR	R
EA-247	CGGTAGCTCTTGTAAACAAGGC	CsLAR	qPCR	F
EA-248	ACTGTTCCACGTAATAAGAAAGCG	CsLAR	qPCR	R
EA-249	GCTGTCATGCTTTGTTTGGAG	CsTT12	qPCR	F
EA-250	GACAGCCCAACATAAAATGTC	CsTT12	qPCR	R
EA-251	CTCTCCCTTATAGCACTCAAG	CsANR	qPCR	F
EA-252	TGGCTTAATCATGTCAGTCTCTG	CsANR	qPCR	R

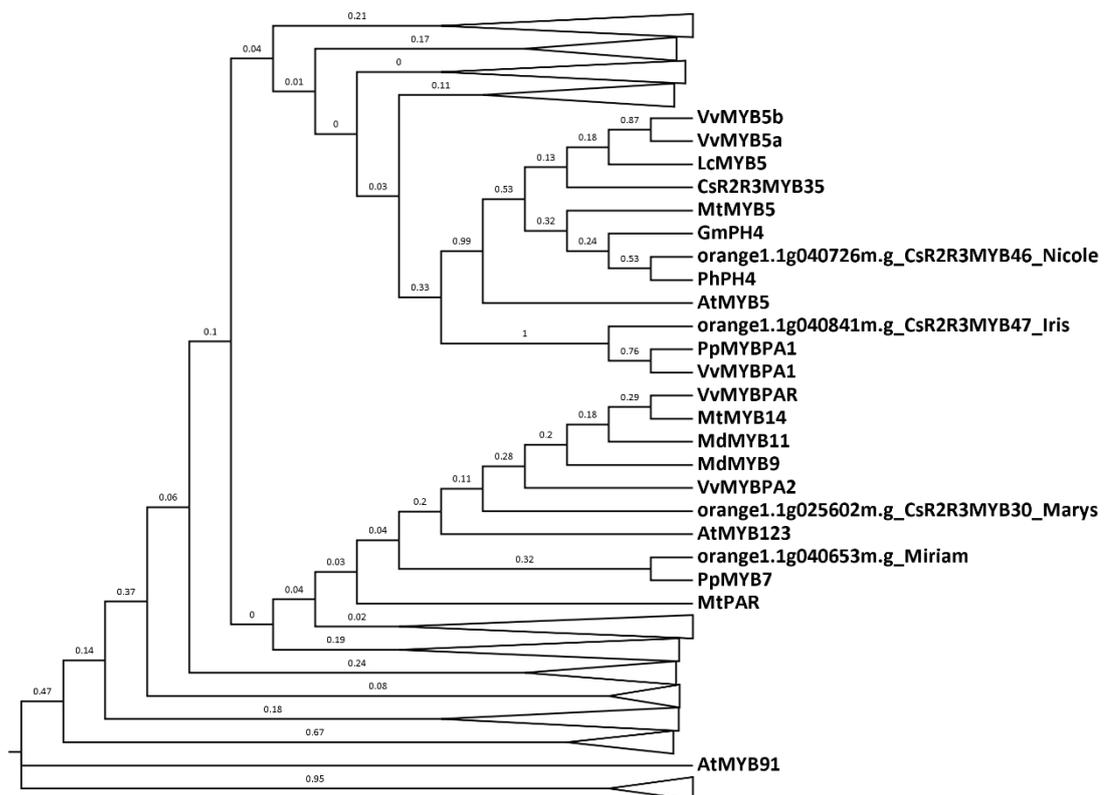
EA-253	AGGCTGGGCTTCTTAGATGTG	Iris	qPCR	F
EA-254	CTCCCTGCGATGAGAGACC	Iris	qPCR	R
EA-255	CAAGATGGATTGCACGCAGG	NPTII cDNA	Genotyping	F
EA-256	GATGTTTCGCTTGGTGGTGC	NPTII cDNA	Genotyping	R
EA-257	GAGAGACGAACCAGCCAACA	Noemi cDNA	Genotyping	F
EA-258	AGCTGGTTTTGGTGGGTGAA	Noemi cDNA	Genotyping	R
EA-259	AGGATTCCGGGGAGAACAGA	Nicole cDNA	Genotyping	F
EA-260	GCCAGCAGAGGTATGACGTT	Nicole cDNA	Genotyping	R
EA-261	CACGGTGAAGGTCATTGGAGA	Iris cDNA	Genotyping	F
EA-262	AATGGGCTTTGGGAGATGGAG	Iris cDNA	Genotyping	R
EA-265	ACTCCGCAGCTCCTAACAAAA	Nicole gDNA full	Amplification	F
EA-266	TTTCATTCCATCCCTCGCGAT	Nicole gDNA full	Amplification	R
EA-267	GAGCGGTCTGATTTGAATGC	Iris gDNA full	Amplification	F
EA-268	TGGGATGCTTGTGAAATGAACTG	Iris gDNA full	Amplification	R
EA-271	TAACGGTCATTTGAGGTTGCGTAG	PH1 gDNA full	Amplification	F
EA-272	TTCTGTCCCTTTTAGCTTCCAGT	PH1 gDNA full	Amplification	R
EA-273	CGGTGTTACACATCTTCTCAATT	TT12 gDNA full	Amplification	F
EA-274	AATTTGGGGTCTCTGCTAATGAT	TT12 gDNA full	Amplification	R
EA-276	GCAGGAGGAAAATGGGAAAGAAAT	PH3 gDNA full	Amplification	R
EA-277	CCAAACCCGAATAGAGTTCAAAG	TTG1 gDNA full	Amplification	F
EA-278	TTCCAAGCAAAGTACATCCAAGC	TTG1 gDNA full	Amplification	R
EA-279	CAGCCCACCAAATCATCATCATC	PH5 gDNA 1/2	Amplification	F
EA-280	TGAATCAAGTAGCATCCGAAGTT	PH5 gDNA 1/2	Amplification	R
EA-281	ACACAGTAACTTCGGATGCTACT	PH5 gDNA 2/2	Amplification	F
EA-282	CTCTATACGCGCTTCTTGACTA	PH5 gDNA 2/2	Amplification	R
EA-283	AGCTGTTACAAAATGGGAAACTGT	PH3 gDNA full	Amplification	F
EA-284	AGTCATTTTCTTGGGCATCTCGA	Noemi gDNA full	Amplification	F
EA-285	AGGCCGGGATTACTTACTTGATC	Noemi gDNA full	Amplification	R
EA-286	CCTCAAGCCTGGTATGGTTGT	NtEF1a	qPCR	F
EA-287	AAACCCACGCTTGAGATCCTT	NtEF1a	qPCR	R
EA-288	AGAAACCCCAAGTACCCTCGTA	NtL25	qPCR	F
EA-289	ACATCTTCTTCACGGCATCCTT	NtL25	qPCR	R
EA-290	TCGTGAACCCAGGGGATCTTTT	NtPH1	qPCR	F
EA-291	AGACACCACACTTGTCCCATATA	NtPH1	qPCR	R
EA-292	TTCTGCTCATTGGTGGTATCCC	NtPH5	qPCR	F
EA-293	TAGTCGATGAGAACCAATGGCC	NtPH5	qPCR	R
EA-294	CCATCTCTATCTGGCTCACAAC	NtTT12	qPCR	F
EA-295	CTGCAGCGCCTAGTCCTAAC	NtTT12	qPCR	R
EA-296	AGCATTCTCACAGAGGGGTG	NtLDOX	qPCR	F
EA-297	TTCTTGAAGAGCTTATGGGCCA	NtLDOX	qPCR	R
EA-298	GATGTTTGTGCGGCCCATATT	NtANR	qPCR	F
EA-299	ATTTGCTAGCTCCGGAACT	NtANR	qPCR	R



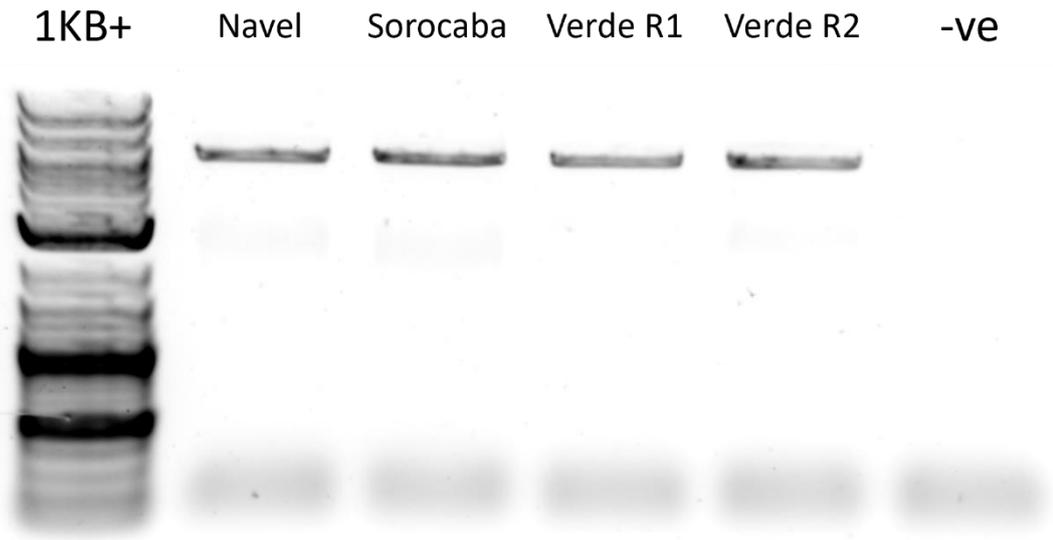
Supplementary Figure 1 Absorbance (OD_{640 nm}) of various concentrations of 15 μ l (+)-catechin standard reacted with 85 μ l 0.3% DMACA over time. Values and error bars presented represent the mean of 3 technical reps \pm se.



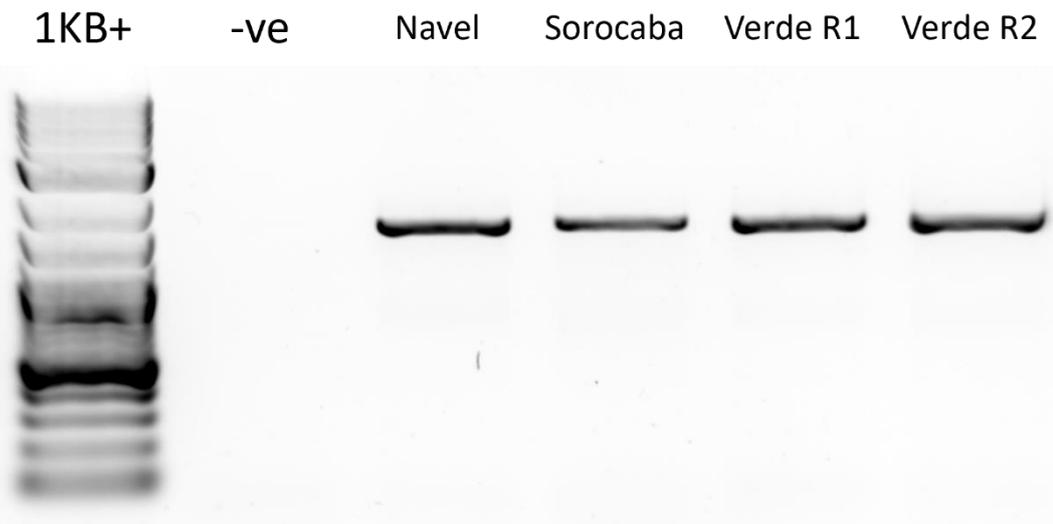
Supplementary Figure 2 (+)-Catechin standard curves of absorbance ($OD_{640\text{ nm}}$) after varying lengths of time reacting with 0.3% DMACA reagent. Lines represent linear regression models. X-axis is log10 transformed. Values and error bars presented represent the mean of 3 technical reps \pm se.



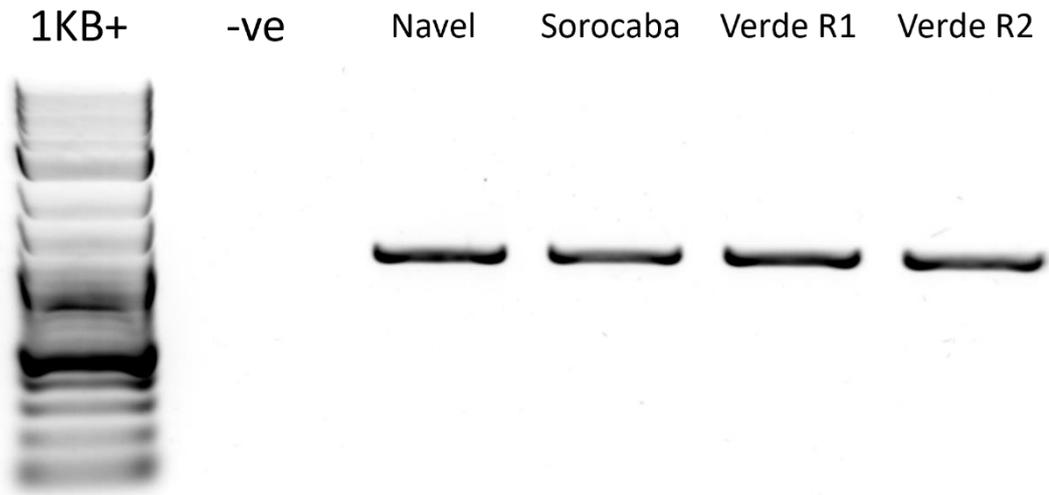
Supplementary Figure 3 Neighbour-joining phylogenetic tree constructed from a ClustalO MSA of all 125 R2R3MYBs from *Arabidopsis* and *Citrus* proteins containing at least one PFAM MYB domain (PF00249) which share greatest homology to AtMYB5 and AtMYB123.



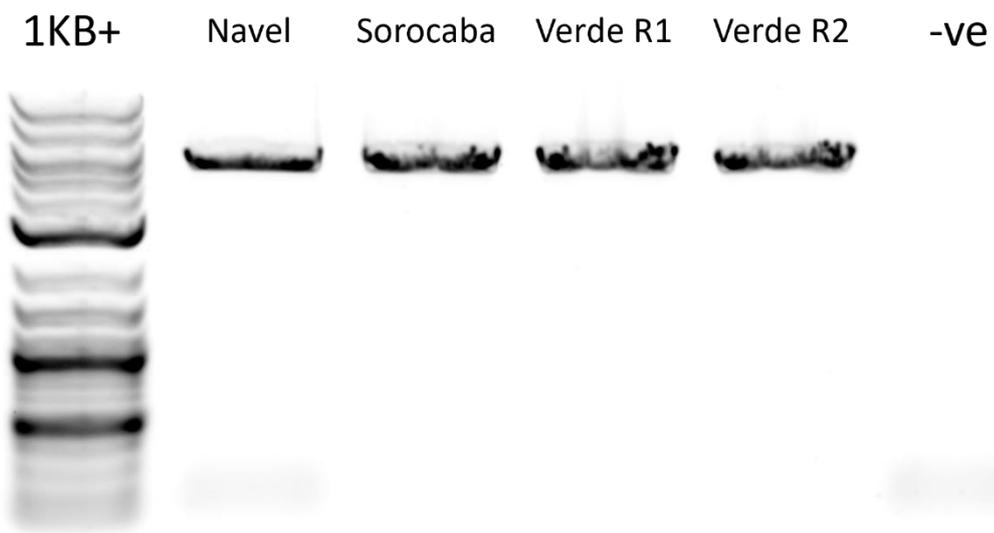
Supplementary Figure 4 PCR amplification of Noemi from gDNA extracted from *C. sinensis* juice. 1KB+: 1KB plus ladder (NEB); -ve: no DNA template.



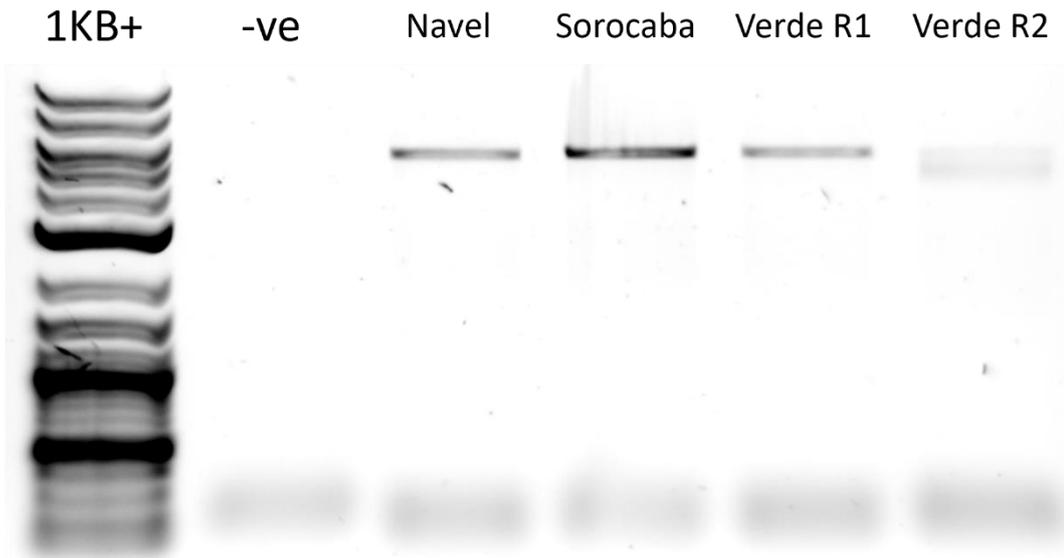
Supplementary Figure 5 PCR amplification of Iris from gDNA extracted from *C. sinensis* juice. 1KB+: 1KB plus ladder (NEB); -ve: no DNA template.



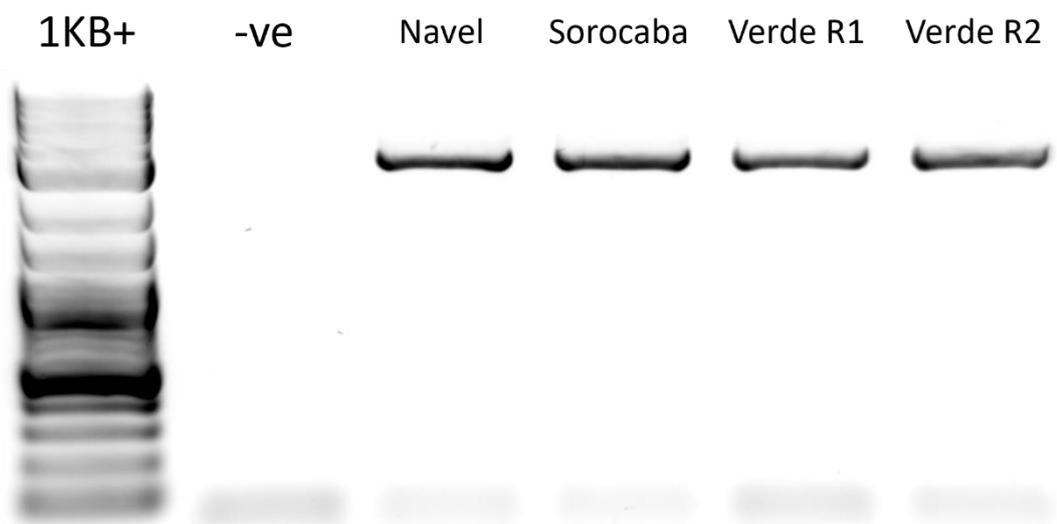
Supplementary Figure 6 PCR amplification of *CsTTG1* from gDNA extracted from *C. sinensis* juice. 1KB+: 1KB plus ladder (NEB); -ve: no DNA template.



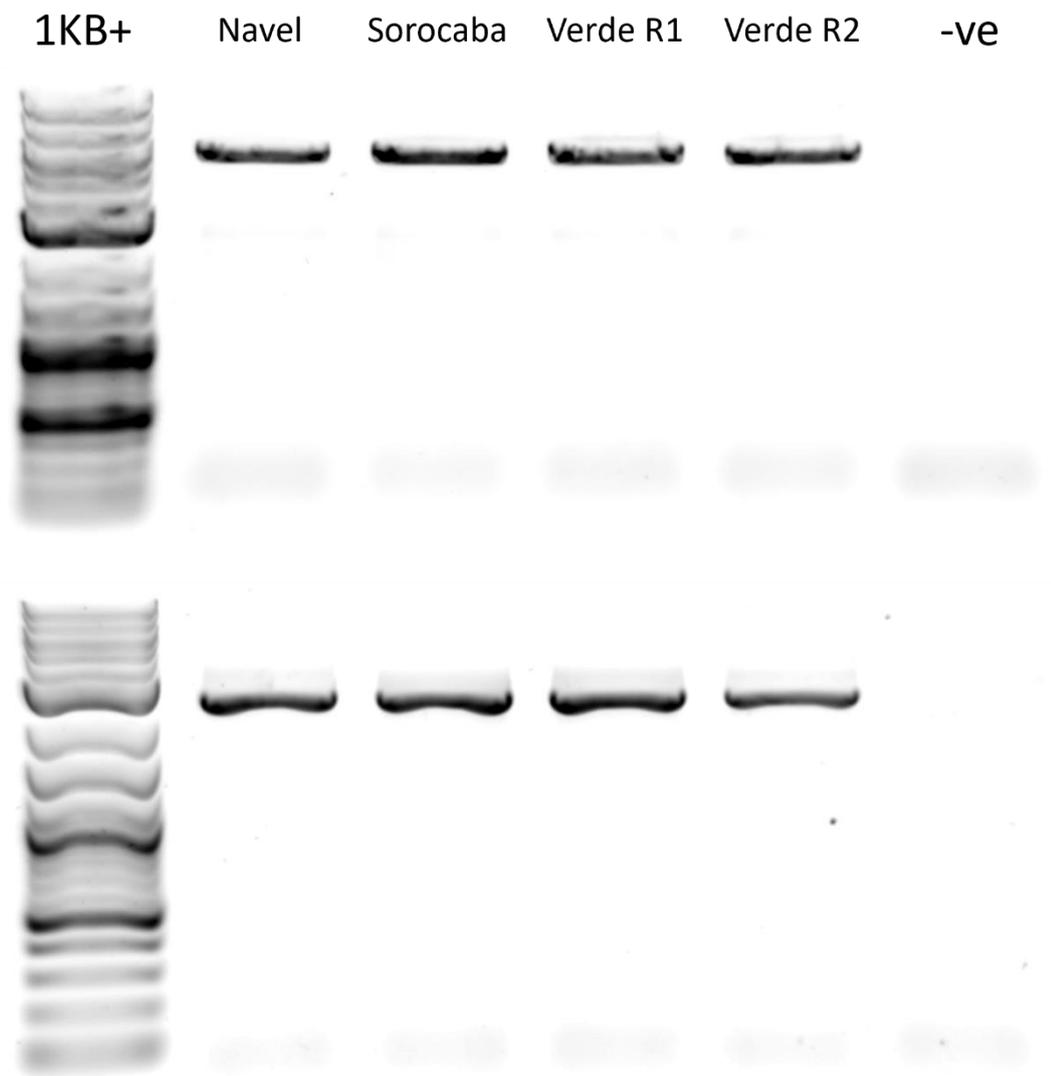
Supplementary Figure 7 PCR amplification of *CsPH3* from gDNA extracted from *C. sinensis* juice. 1KB+: 1KB plus ladder (NEB); -ve: no DNA template.



Supplementary Figure 8 PCR amplification of CsPH1 from gDNA extracted from *C. sinensis* juice. 1KB+: 1KB plus ladder (NEB); -ve: no DNA template.



Supplementary Figure 9 PCR amplification of CsTT12 from gDNA extracted from *C. sinensis* juice. 1KB+: 1KB plus ladder (NEB); -ve: no DNA template.

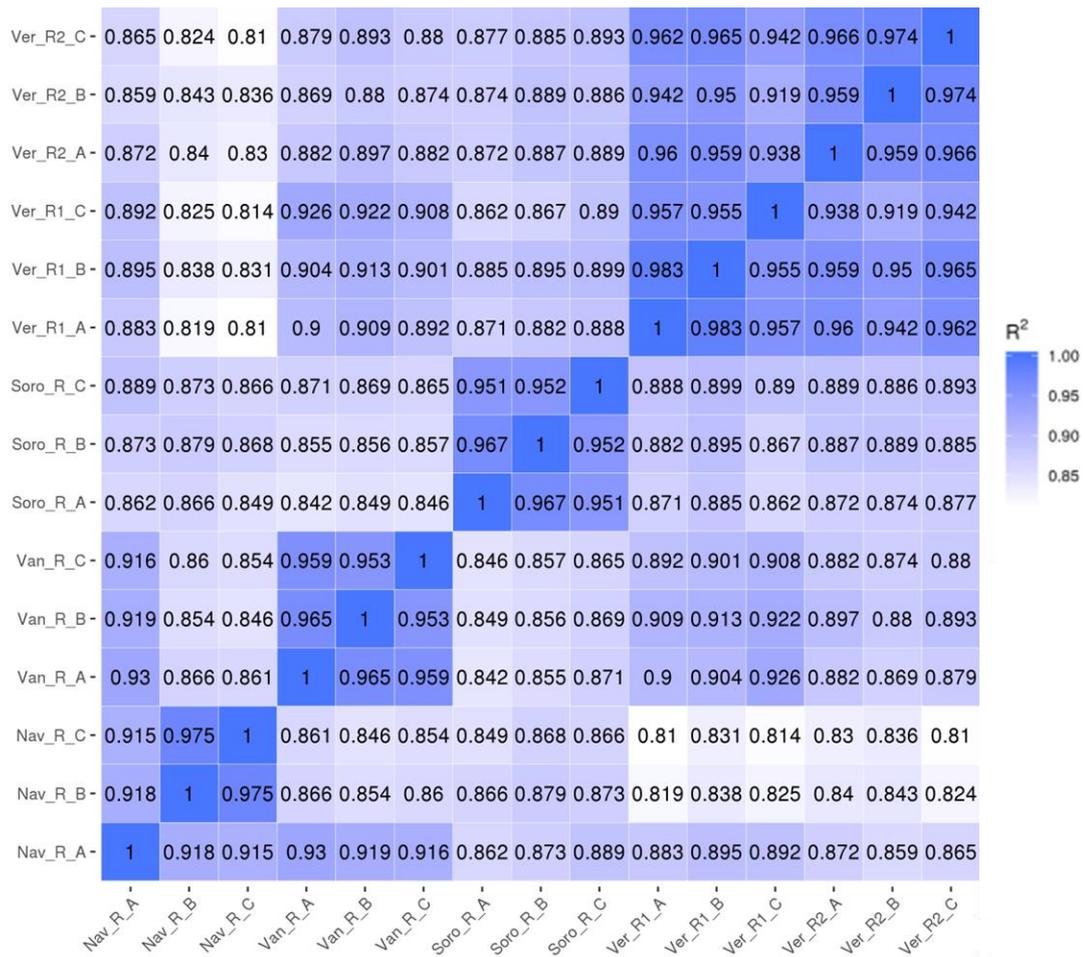


Supplementary Figure 10 PCR amplification of CsPH5 from gDNA extracted from *C. sinensis* juice. 1KB+: 1KB plus ladder (NEB); -ve: no DNA template.

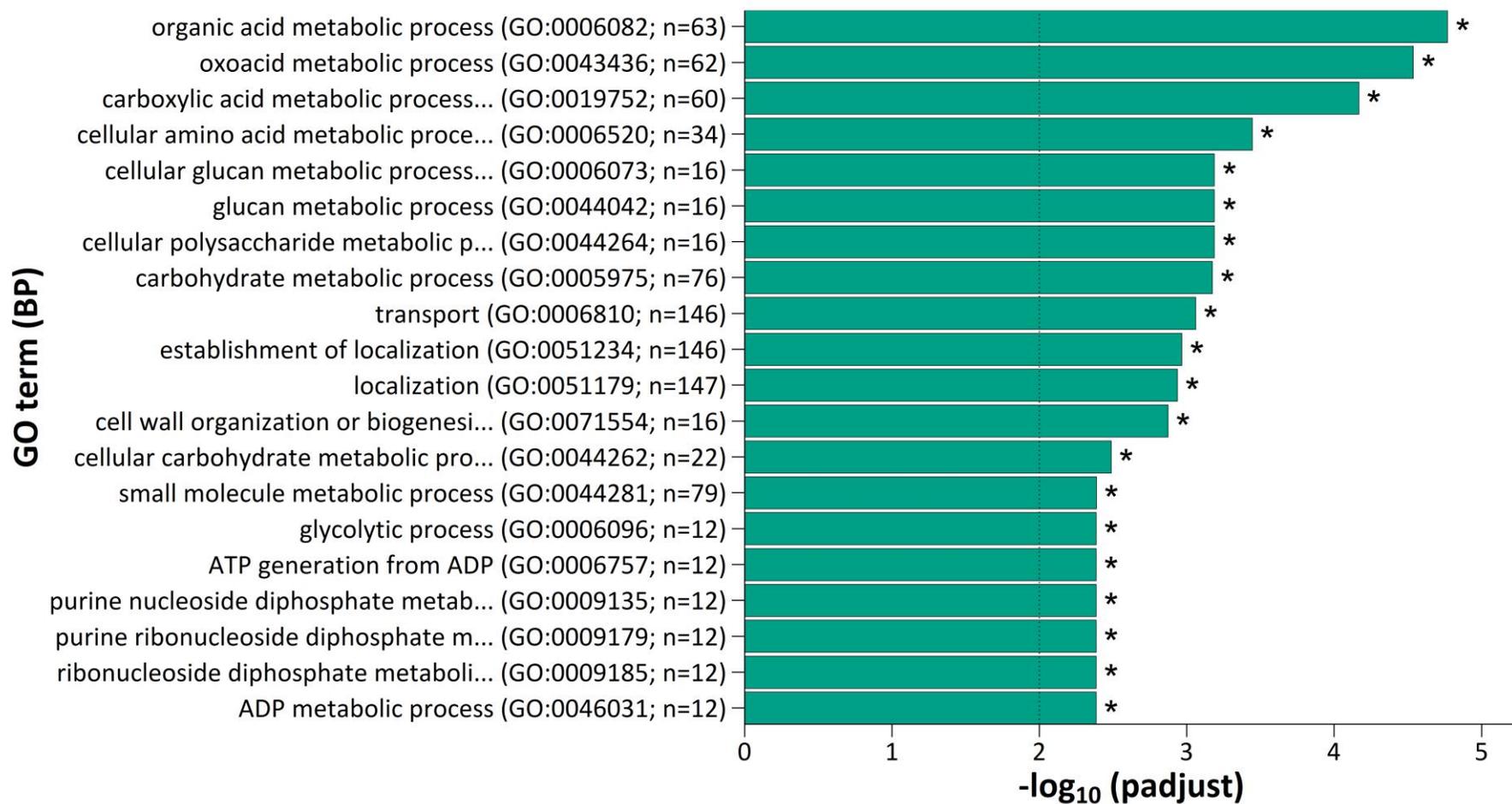
Supplementary Table 5 Sequence of *nicole*^{soro}. Blue nucleotides indicate the exons of *WT Nicole*. Yellow nucleotides indicate the insertion of *Tcs7x*. TAG stop codons are highlighted red. *Italicised TAG* indicates the introduced stop codon.

>*nicole*^{soro}

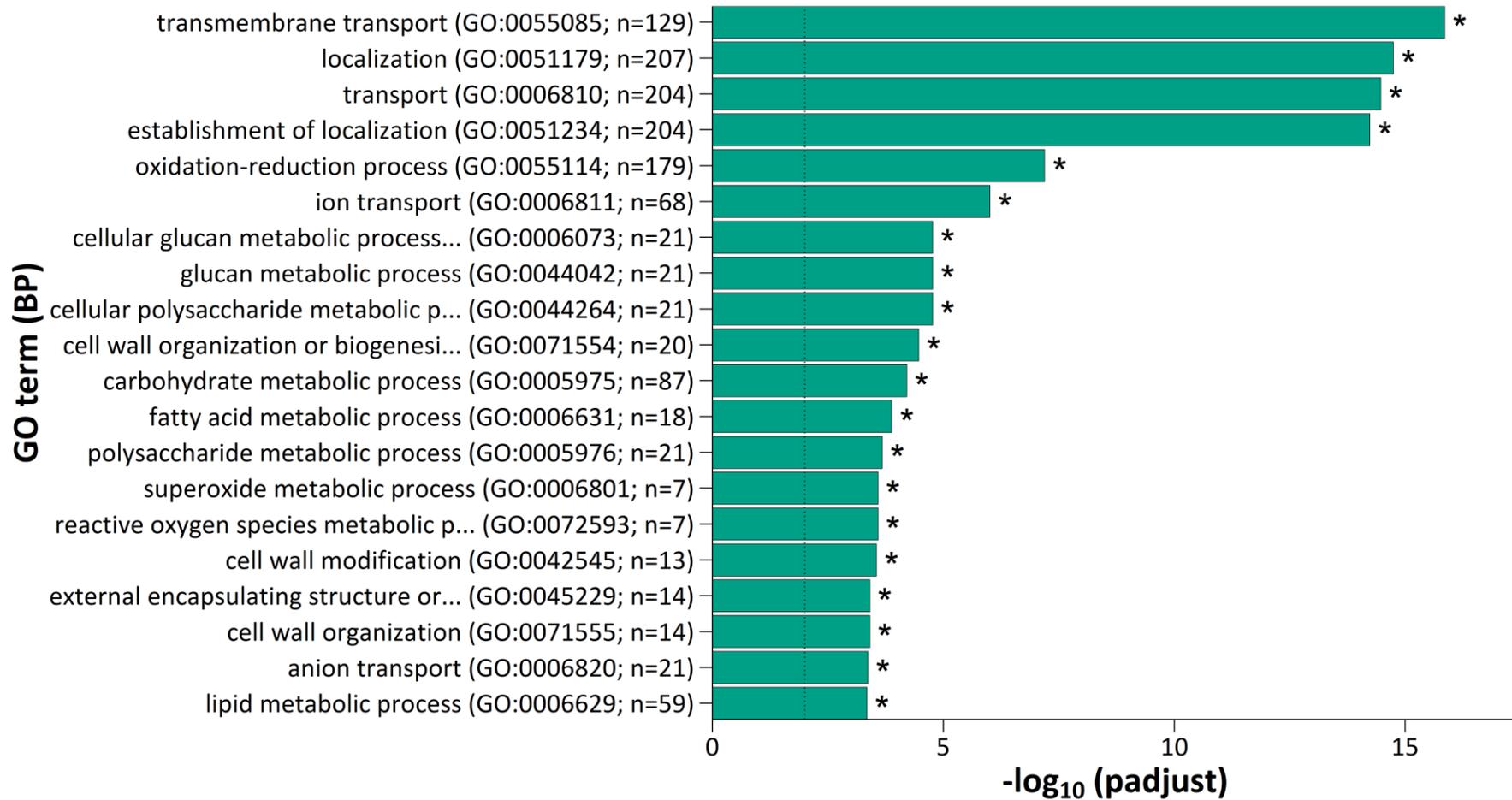
ATGAGGAACCCATCAACATCACCATCATCAACAGCAGCAGCAGCAGCAGCAACCAATAAGAGCACGCCATGTTGCA
GCAAGGTAGGGTTAAAGAGAGGGCCATGGACGCCAGAGGAAGACGAGCTTCTGGCCAACACATCAATAAAGAAGCGAAGG
CCGGTGGCGAAGCTTCGCCAAAACGGCCGGATTGCTCCGCTGCGGCAAGAGTTGCCGGCTTCGTTGGATGAACATCTGAGA
CCCTCCGTTAAACGAGGACATATCGCCCCGATGAAGAAGATCTCATTCTTCGCCTACATCGCCTTCTCGGTAACCGGTAAT
AAGAAATAATAACCCAGCAATACCTTAGGGTTCCATGGATTTTATTTACTCGGTTTGTGTTTGGTTTGTACTTCAGAGTT
AGGGCTAGGGTTTTTGTAACTGAATTTGGTTTTGTGTTTGTGCTTTATAGATGGTCTCTGATAGCGGGGAGGATTCGGGG
AGAACAGATAATGAGATAAAGAATTACTGGAACACTCACCTGAGTAAGAAGCTGATAAGCCAAGGGATTGATCCAAGAATC
ATAAGCCATTGAATCAAGAACTTGATCCTTCTTCTGCTGATCAAGTTACTAATAGCAACAGCAAAAGCTTCAACTTCGAAAGC
AATACTAACTCGAGCAGCTCAAACCCATAATCTCACTCCAATGACCGTTTCATCTGGTCATTTAGATCAACGTCATACTCT
GCTGGCTGTGGTGAATGATCTCGTCGATCATGATGATCAATAAGGAAAAATGGGTATTACCGAAGCCTTTAGTTGATGATC
ATGACAGTGAGTATCATCAAAATGGGATGATGGAGAACCCGTATACGAGTTTATCGAATTGTGATCATCATGACGATGA
TGGGGGTTGGGTTTGAGAAGCAATAACGTGAATAACGTTTAAACGAAGGGCTTAGCTATGCAAAAGCTGATGATGATCAAT
TACTGCAACGACGATGTTTTCTCTCGTTTTCTCAATTCGTTGATCAATGAAGATGCTTTTGTAGCCAGCATAATCAACAAG
TACTGCAACAACCTGTCAAAGTTTTTGGCTGCAACTTCTAGAAGCTACATC **TA**GAAAGCGTTGATGCAGCAAAATGTTGAAGGGA
CCGCCAAGGAAGAATGAAGAAGATGTTGCAGAAGTTGATCTTGACAAATGATGGTTGAAACGCCAAAATGGAAGAAAAGAAA
ATGTTGGAACACCGCTTTGAAGAAGCCACGAGAAGAATAAAGAAAATATTAATAATAGCAAAATTTGCTATATAAAGAA
AGCTCCCCATTTTTGGTTTTTGCATCTAATCCTCGGCTTCTTTTCTTCAATTCAGAGAGTATTTCTTTGGGGTGTATTTGGG
GCTTAGGTGAGAGAAAATATTTCTGAGAGTGTGGTTGTAATAATTTCCACATAGTGAATATTTTTCTCTGGTTGTCTTT
TTGACAACGGCCGTGGTTTTTCTCCGATTTGGAGTTTTCCACGTAATCTTGTGTTGTGTGATTGGTGTATTTCTCCATTA
ATTTTTCTGTAAATTTGTGCTTGACAAATGCTTAGAGTGATCTTGGGAGGAAGCTCAATTTCCATAACAGTGGTATCAGA
GCCATTGATTTAAGTTTTGGTGTGGGGCACTGTTACGTTATACGGTACTGTTACGTTATACGGTACTATTCACGTGAAGCA
GTGGGAGCCAATCCAACAGTCTGTGGTGAAGAGTAAAGCTTATCTGCAAGCAAAATATGTCAGGATGAAATTTTCAAGT
CCGGTGAAATTTGAAATAGAAAAATTCGATGGGAGAATTAACTTGGCTTGTGGCAAGTTCAAGTCAAAGATGTGTTAATC
AATCTGGGTTACACAAGGCATTGAAGGGGAAGCCATCCCCGCTTCCAGTAGTGGCTCTGGAAAACTAGTATAAAGTATGA
AGATTGGGAAGAATTAGATGATAGAGCTGCAAGTGCCATACGACTGTGTCTAGCAAGAATGTTCTTGCAAAATGTAGGAAAA
ATTTCTACAGCGAAGAAGCTTTGGGAGAAGCTAGAAAAGTTGATCAGACAAAGAGCATCTCAAATCGATTGTACCTGAAGG
AGCGATTTACACACTGCGAATGGCTGAAGGTACAAAAATTTCCGATCACCTCAGTGTCTCAATGGTATTGTGTGAGAAGT
AGAAGCCATTGGAGTTAAATTTGAAGATGAGGACAAGGCCCTTAGGTTACTATGGTCACTTCCAATTTCTACAAACACTTG
TTACTACTTTGATGATGGGAAGGAGACAGTAGATCTTGAAGAAGTTACTAGTACTTTACTCTCAGAAGAAAGGAGACTGG
GTGGTGAAAGTACTAAAACATATAGATGTCTCGGCTTTGGCAGTTGTAGGGAATTGGCAGAAAGATAAATCTAAGAAGAAAGG
AGTCTGCTGGGGGTGTGGACAATCGGGGCACTTAAAAAGAGATTGTCATAGTAGAAAATGGAGCAGGATCGGCAAGTGGCTCC
AGATCAGATACTGATAGTATTGCTAGTGGTAAAGTCTCTCATCATCGTGGGAGACGATGATCCCTTGTAAAAATGGATGATGAT
GACATCCTCATGGTATACCGCTAGTACCATGAAAGGGGATATGTTACTACTAGCGGGTCCACAAGATTTACACACAAGGCAT
GGTTGGCATTGATGCAGGGTGTGTGGTGAATTTATGTCGATGGCTGACAAACTCCAGGAAGGCCAACATGGAAAGTTGCAC
CATAAATTTGAGCAGGATATTTGACATGTGCCGACGTAATAATCTTAGAATTTGGTAATTAATTTCAAGTGGTATACTCTTT
TATGTTGGGGTATGATAATTTCTCTATGGTGAGGAAAATAATAAACTTGGTGTGAAGATTGATGGTTCTCAATCAAATCTCC
AAGTGGGAGAATGTCAAAGTTTTTGGCTGCAACTTCTAGAAGCTACATCTAGAAGCGTTGATGCAGCAAAATGTTGAAGGGAC
CGCCAAGGAAGAATGAAGAAGATGTTGCAGAAGTTGATCTTGACAAATGATGGTTGAAACGCCAAAATGGAAGAAAAGAAAA
TGTTGGAACACGGCTTTGAAGAAGCCACGAGAAGAATAAAGAAAATATTAATAATAGCCAATTTTGGCTATATAAAGAAA
GCTCCCCATTTTTGGTTTTTGCATCTAATCCTCGGCTTCTTTTCTTCAATTCAGAGAGTATTTCTTTGGGGTGTATTTGGGG
CTTAGGTGAGAGAAAATATTTCTGAGAGTGTGGTTGTAATAATTTCCACATAGTGAATATTTTTCTCTGGTTGTCTTTT
TGACAACGGCCGTGGTTTTTCTCCGATTTGGAGTTTTCCACGTAATCTTGTGTTGTGATGGTGTATTTCTCCATTA
TTTTTCTGTTAATTTGTTGCTTGACAAATGCTTAGAGTGATCTTGGGAGGAAGCTCAATTTCTAA **CAACAACAACA**
GCAGCACCTATCAAATGAGACGATTGCATTGCCGAATACAATTAAGCTGGCTCATCATCGGATCCTTTGGTTTTGACTGCAGCG
GCATCAACTTTGGCCTTGAAGCAAACTGGGAATCTCAATCATGGCTTCTTCTTGAACCAAGATGAGTCCAGGAGGGTTG
ATGAACACGTTGAG**TAG**



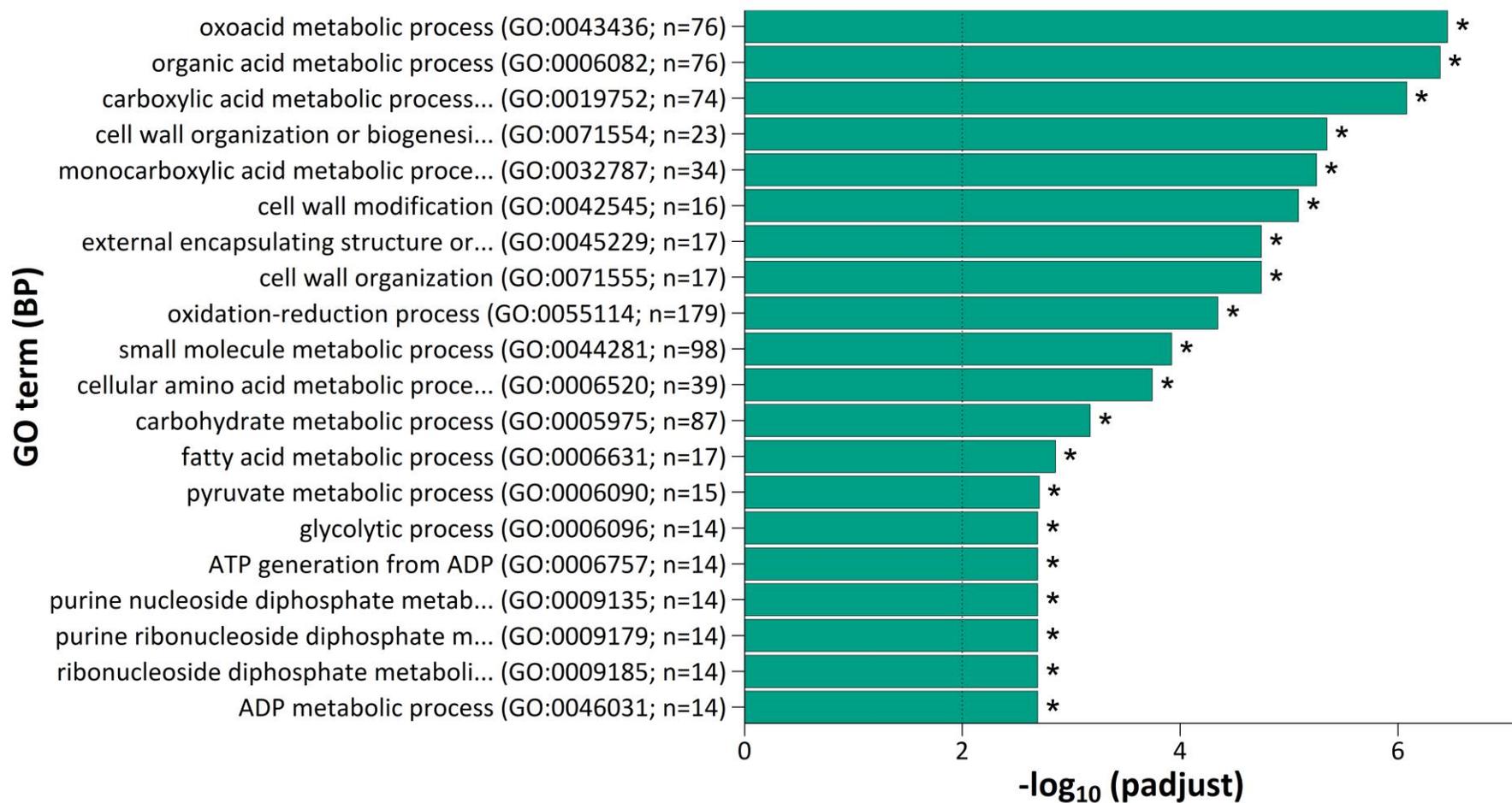
Supplementary Figure 11 Pearson correlation heatmap of all *C. sinensis* RNAseq biological replicates.



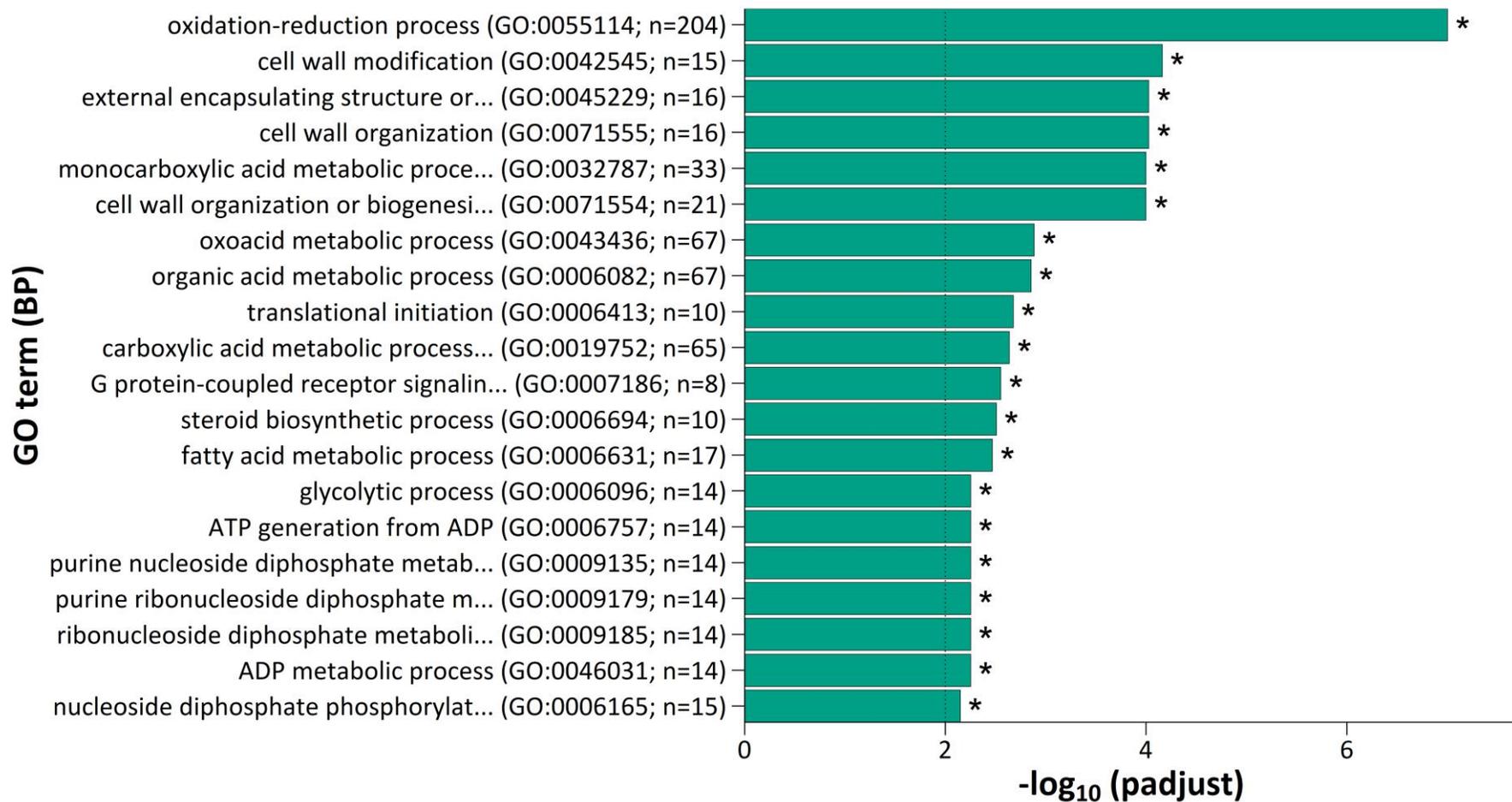
Supplementary Figure 12 Biological process gene ontology enrichment in significantly downregulated genes in Vaniglia relative to Navel. Asterisks indicate significant enrichment.



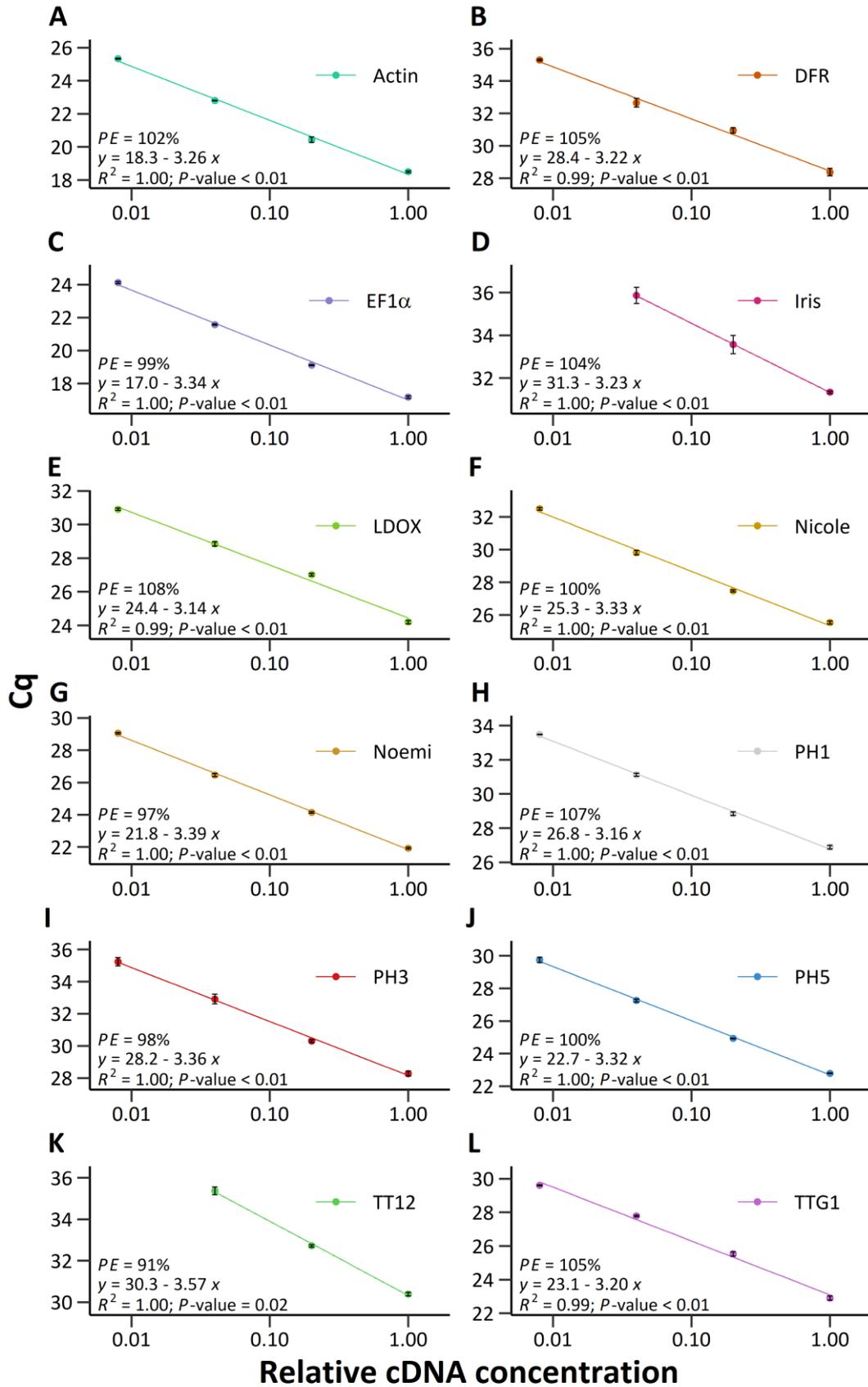
Supplementary Figure 13 Biological process gene ontology enrichment in significantly downregulated genes in Sorocaba relative to Navel. Asterisks indicate significant enrichment.



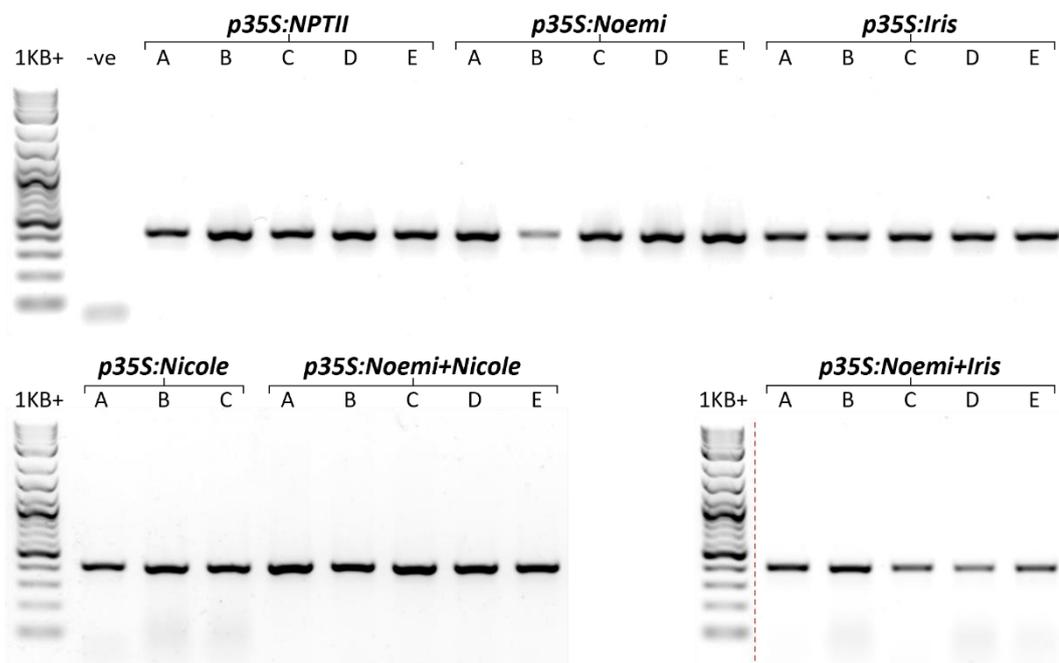
Supplementary Figure 14 Biological process gene ontology enrichment in significantly downregulated genes in Verde R1 relative to Navel. Asterisks indicate significant enrichment.



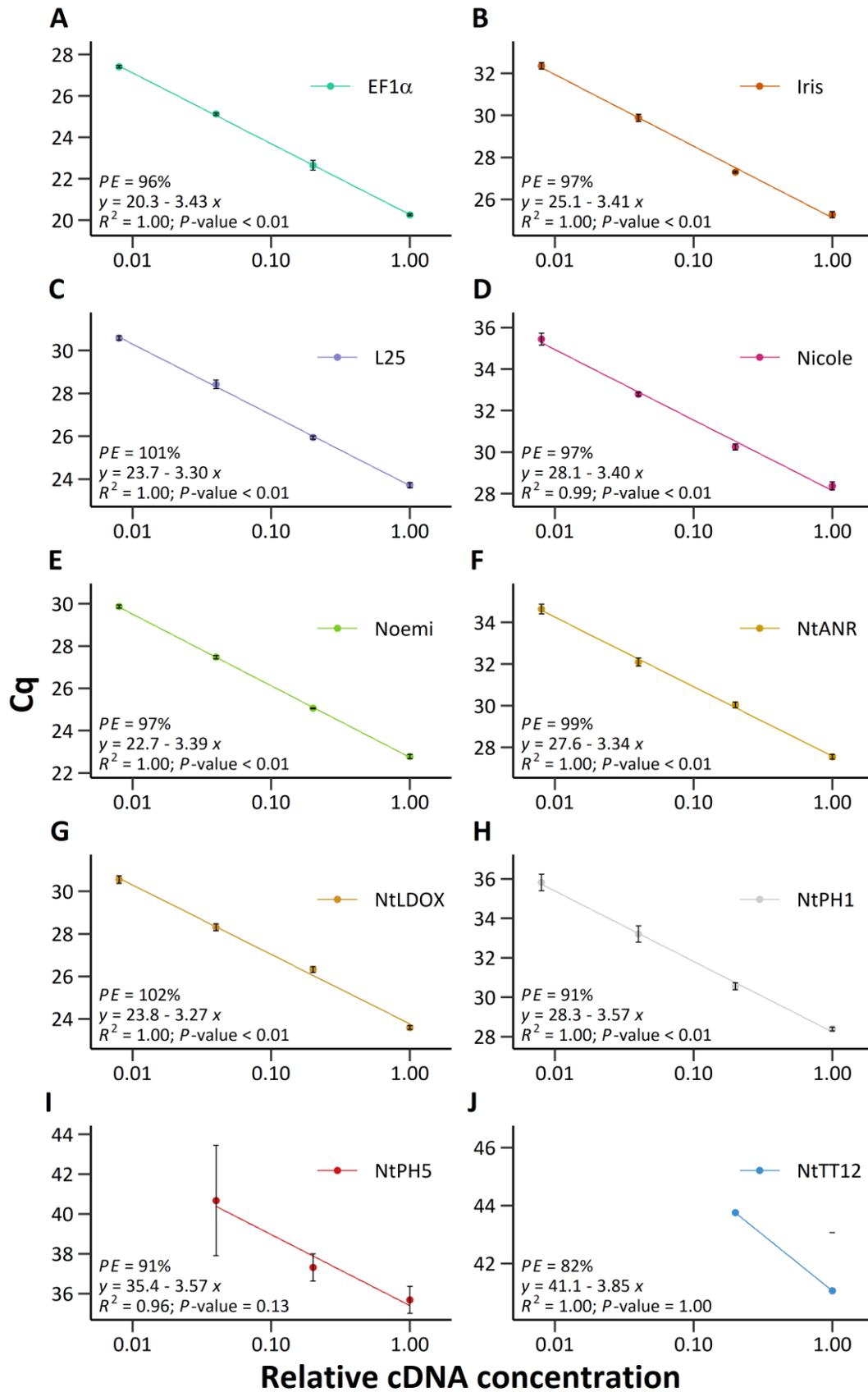
Supplementary Figure 15 Biological process gene ontology enrichment in significantly downregulated genes in Verde R2 relative to Navel. Asterisks indicate significant enrichment.



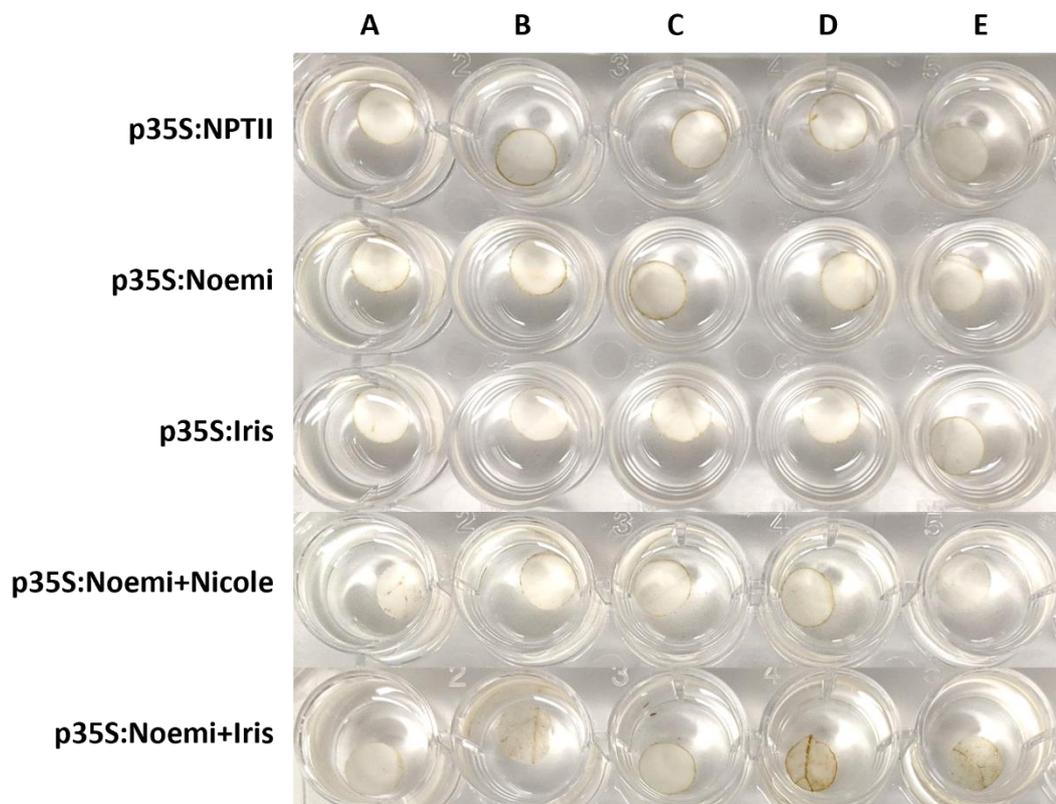
Supplementary Figure 16 Primer efficiency analyses for RT-qPCR. Primers were tested on a dilution series of a cDNA mix containing all *C. sinensis* samples. Line represents linear regression model.



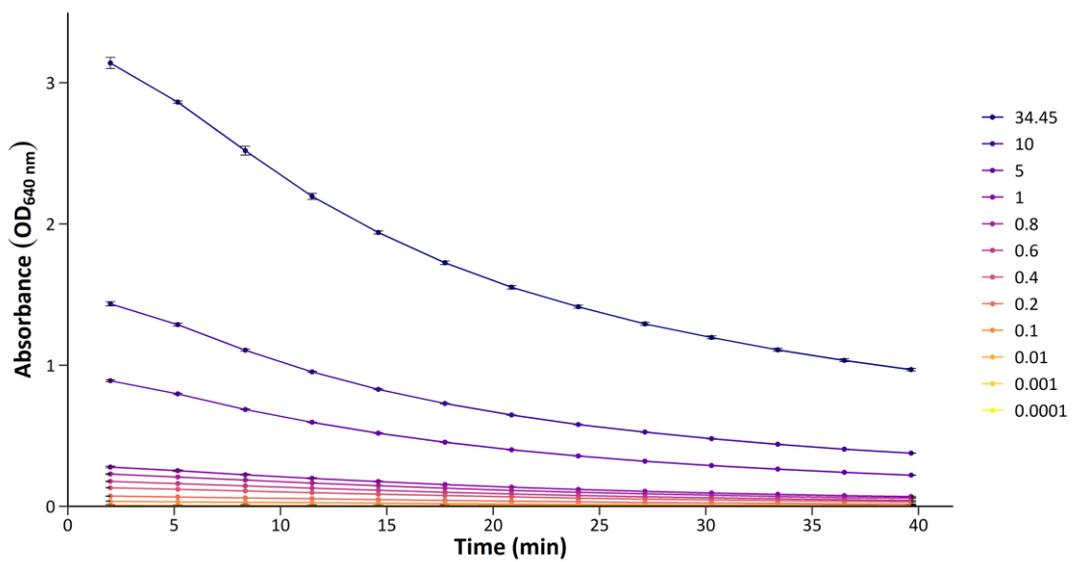
Supplementary Figure 17 PCR amplification of NPTII T-DNA from gDNA extracted from overexpression *N. tabacum* transformed lines. 1KB+: 1KB plus ladder (NEB); -ve: no DNA template; red dashed line: indicates gel location where unrelated samples were cropped from the image and remaining areas spliced together.



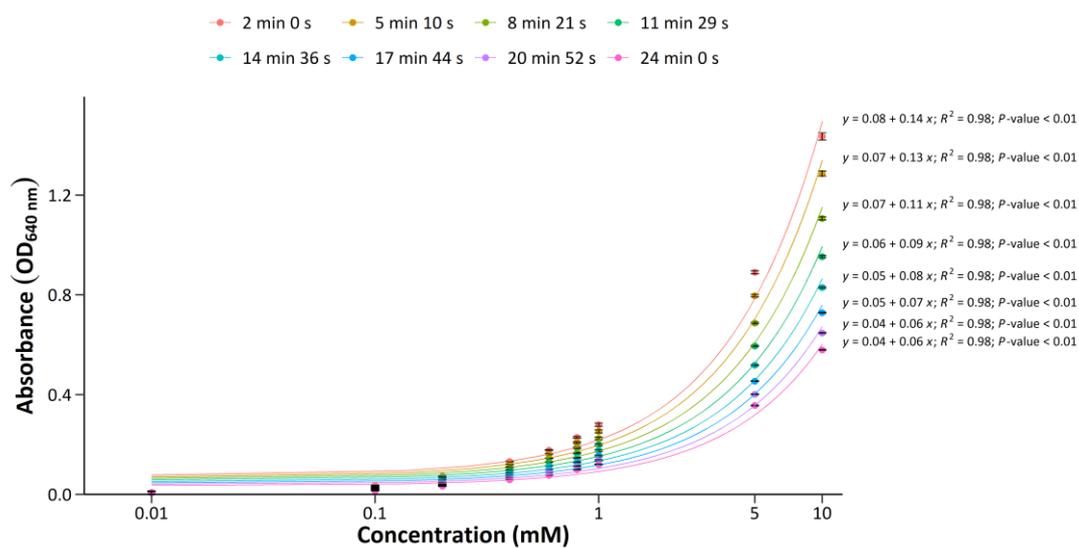
Supplementary Figure 18 Primer efficiency analyses for RT-qPCR. Primers were tested on a dilution series of a cDNA mix containing all *N. tabacum* samples. Line represents linear regression model.



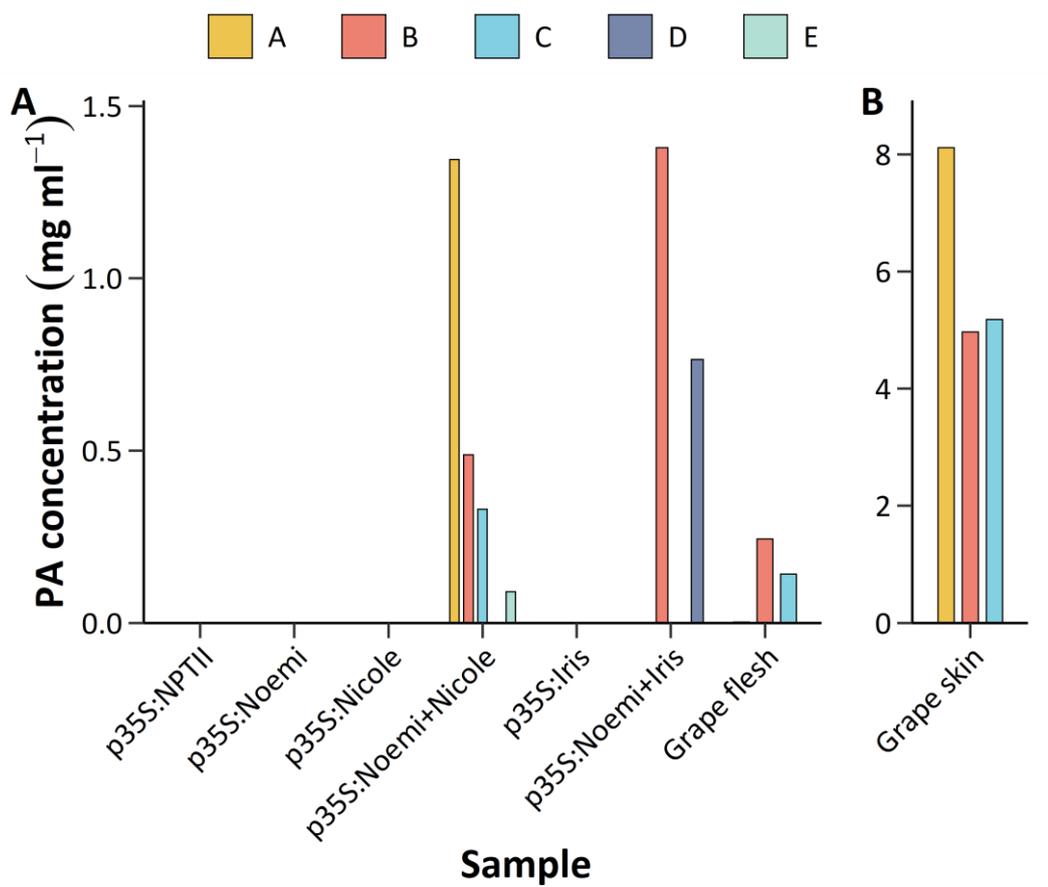
Supplementary Figure 19 Bleached leaf discs from overexpression *N. tabacum* lines, prior to DMACA staining. Leaf discs from five biological replicates (A-E).



Supplementary Figure 20 Absorbance ($OD_{640\text{ nm}}$) of various concentrations of $4\ \mu\text{l}$ (+)-catechin standard reacted with $96\ \mu\text{l}$ 0.3% DMACA over time. Values and error bars presented represent the mean of 3 technical reps \pm se.



Supplementary Figure 21 (+)-Catechin standard curves of absorbance ($OD_{640\text{ nm}}$) after varying lengths of time reacting with 0.3% DMACA reagent. Lines represent linear regression models. X-axis is log10 transformed. Values and error bars presented represent the mean of 3 technical reps \pm se.



Supplementary Figure 22 Proanthocyanidin quantification of via colorimetric DMACA assays. A: overexpression of *N. tabacum* leaf and grape flesh extracts; B, grape skin extracts as a positive control. Values plotted are measurements of individual biological replicates.