

Establishing genome editing in amaranth to
understand the betalain pathway regulation and
investigate its evolution



Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Susanne Katja Vollmer
aus Dormagen, Deutschland

Düsseldorf, März 2026

aus dem Institut für Biochemie der Pflanzen,
Center for Plant Genome Engineering
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Markus G. Stetter

2. Prof. Dr. Andreas P. M. Weber

Tag der mündlichen Prüfung:

03. Juni 2026

Zusammenfassung

Betalaine sind mehrfach evolvierte, spezialisierte pflanzliche Metabolite, welche eventuell wichtige Funktionen bei der pflanzlichen Stressreaktion haben, aber auch verschiedene Anwendungen in der Lebensmittelindustrie. Während die Enzyme des zentralen Stoffwechselwegs schon bekannt sind, wissen wir nur wenig über ihre molekulare Regulation. Ein Grund dafür ist, dass Betalain auf die Ordnung der *Caryophyllales* beschränkt sind, für die es generell nur unzureichende Ressourcen gibt. Der erste Teil dieser Dissertation präsentiert die Etablierung von zentralen Methoden für *Amaranthus hypochondriacus* L., eine „Weisennutzpflanze“ mit großer Diversität an Betalainfärbungen. Zuerst wurden mehrere stabile und transiente Transformationsmethoden entwickelt, darunter eine für Kallusgewebe mit einer Transformationseffizienz von bis zu 100%. Des Weiteren wurde die erste gezielte Mutagenese mittels CRISPR/Cas9 in Amaranth gezeigt, bei der erreicht wurde, gleichzeitig zwei Zielgene in etwa 25% aller Proben zu editieren. Diese leistungsfähigen Methoden ermöglichten, die Untersuchung der Betalainregulation in Amaranth, welche im zweiten Teil der Dissertation beschrieben ist, beginnend mit der Identifizierung von AhMYB2, einem Homolog des Anthocyanin-Regulators AtPAP1, als positivem Regulator für die gewebespezifische Betalainregulation. Zusätzliche heterologe Assays haben dann gezeigt, dass die ersten beiden Enzyme des Stoffwechselweges (*AhCYP76AD2* und *AhDODAa1*) die hauptsächlichen Ziele der Regulation sind und dass AhMYB2 ihre Expression hochreguliert. Im Gegensatz zu AtPAP1, suggerieren die Daten, dass AhMYB2 nicht in MYB-bHLH-WD40-Komplexen agiert, sondern stattdessen einen anderen betalainspezifischen Transkriptionsfaktor (TF) benötigt. Vielversprechende Kandidaten, einschließlich WRKY-, ERF-, bZIP- und NAC-TFs wurden durch RNA-Sequenzierung identifiziert. Abschließend, wurden die Ergebnisse mit denen der Literatur verknüpft, um ein mögliches Modell für die evolutionäre Entstehung der AhMYB2-gesteuerten Betalainregulation vorzuschlagen. Hypothesisierend, dass eine der zwei ursprünglichen anthocyaninregulierenden PAP1-Genkopien kooptiert und zur Ansteuerung von Betalaingenen angepasst wurde, erklärt eventuell, warum beide Pigmente einander ausschließen. Dies wird unterstützt durch mehrere konservierte Aminosäuren, spezifisch für PAP1-Homologe von Betalain-produzierenden Pflanzen, welche in der Nähe von Aminosäuren lokalisiert wurden, die bekanntermaßen wichtig für die DNA-Bindung sind, was auf ein geändertes Erkennungsmotiv hindeutet. Die präsentierten Ergebnisse und Protokolle werden die

Betalainforschung und die Züchtung des trockenheitstoleranten C₄ Körner-Amaranth fördern. Zukünftige Studien, welche die Wirkungsweise in anderen Betalainursprüngen und die Integration von anderen Signalwegen in die Betalainregulation aufdecken, könnten eventuell Einsichten in die Evolution von Genregulationsnetzwerken von *de-novo*-evolvierten Stoffwechselpathway geben.

Abstract

Betalains are repeatedly evolved, colored plant specialized metabolites that might have important functions in plant response to stresses but also various applications for the food industry. While the enzymes of the core pathway have been identified, only a little is known about its molecular regulation. One reason is that betalains are restricted to the order *Caryophyllales*, which generally lacks sufficient resources. The first part of this thesis presents the establishment of key methods for *Amaranthus hypochondriacus* L., an orphan crop with a high diversity of betalain pigmentation. First, various stable and transient transformation methods were developed, including one for callus tissue with a transformation efficiency of up to 100%. Furthermore, the first targeted mutagenesis using CRISPR/Cas9 in amaranth was demonstrated, achieving multiplexed editing of two target genes in approximately 25% of all samples tested. These powerful methods enabled the investigation of the betalain regulation in amaranth, which is described in the second part of the thesis, starting with the identification of AhMYB2, a homolog of anthocyanin regulator AtPAP1, as a positive regulator for the tissue-specific betalain regulation. Additional heterologous assays then showed that the first two enzymes in the pathway (*AhCYP76AD2* and *AhDODAa1*) are the main target for regulation and that AhMYB2 upregulates their expression. In contrast to AtPAP1, the data suggest that AhMYB2 does not function in MYB-bHLH-WD40 complexes but instead requires a different betalain-specific transcription factor (TF). Promising candidates, including WRKY, ERF, bZIP, and NAC TFs, were identified using RNA sequencing of transgenic calli. Finally, the results were combined with the literature to propose a model of how the AhMYB2-mediated betalain regulation might have evolved. Hypothesizing that one of the two ancestral anthocyanin-regulating PAP1 gene copies had been co-opted and adapted to target betalain genes, while the other copy has been lost, might explain the mutual exclusiveness of both pigments. This is supported by multiple conserved amino acids specific to PAP1 homologs of betalain-producing species, which were located close to known residues important for DNA binding, suggesting a changed recognition motive. The presented results and protocols will advance betalain research and breeding of the drought-tolerant C₄ grain amaranth. Future studies uncovering the betalain regulation in other origins and the integration of other signaling pathways might provide insight into the evolution of gene regulatory networks for *de novo* evolved pathways.

Contents

1	Introduction	7
1.1	Understanding evolution is the key to the past and future	8
1.2	The betalain pathway as a case study for the evolution of new regulations	10
1.3	Betalain and anthocyanin levels react remarkably similarly to internal and external signals	12
1.3.1	Internal cues	12
1.3.2	External cues	14
1.4	The transcriptional regulation of the betalain pathway could be co-opted from the anthocyanin regulation	16
1.5	The high diversity in pigmentation of grain amaranth is optimal to investigate betalain pigmentation	17
1.6	Genetic resources and stable transformation are the foundation of most state-of-the-art tools, but are restricted to model species	19
1.6.1	Genetic transformation	19
1.6.2	Genome editing	20
1.6.3	Genome sequencing	21
1.7	Aims of the Thesis	22
1.8	References	23
2	First Successful Targeted Mutagenesis Using CRISPR/Cas9 in Stably Transformed Grain Amaranth Tissue	51
2.1	Summary	52
2.2	Main part	52
2.3	Acknowledgments	55
2.4	Conflict of interest	55
2.5	Author contributions	55
2.6	References	55
2.7	Supporting Information	57
2.7.1	Supplementary methods	58
2.7.2	Supplementary figures	60
2.7.3	Supplementary tables	63

2.7.4	Appendix	65
2.8	Supplementary references	72
3	Isoform-resolved genome annotation enables mapping of tissue-specific betalain regulation in amaranth	73
3.1	Summary	74
3.2	Introduction	74
3.3	Materials and Methods	77
3.3.1	Improving the <i>A. hypochondriacus</i> reference genome	77
3.3.2	Plant material and genome annotation datasets	78
3.3.3	Genome annotation	78
3.3.4	Mapping betalain related quantitative trait loci	79
3.3.5	Cloning and overexpression of <i>AhMYB2</i>	81
3.4	Results	82
3.4.1	Improved <i>A. hypochondriacus</i> reference genome	82
3.4.2	Location and conservation of the color regulating MYB transcription factor family	84
3.4.3	Mapping pigment biosynthesis pathway genes	87
3.5	Discussion	93
3.5.1	High quality reference genome annotation for functional genomics	93
3.5.2	Conservation of the flavonoid and betalain pigment pathways in amaranth	95
3.5.3	<i>AhMYB2</i> is a flower-specific regulator of the betalain pathway	96
3.5.4	The rise of mutually exclusive pigment pathways	97
3.6	Acknowledgments	98
3.7	Author contributions	99
3.8	Competing interests	99
3.9	Data availability	99
3.10	ORCID	99
3.11	References	100
3.12	Supporting Information	113
3.12.1	Supplementary methods	115
3.12.2	Supplementary notes	120
3.12.3	Supplementary tables	123
3.12.4	Supplementary figures	137
3.13	Supplementary references	152
4	Uncovering co-option and adaptation of PAP1 for the regulation of the <i>de novo</i> evolved betalain pathway in amaranth	157
4.1	Introduction	158

4.2	Materials and Methods	159
4.2.1	Callus transformation and phenotyping	159
4.2.2	Yeast-two-hybrid to test potential interactors of AhMYB2	161
4.2.3	Protoplast isolation	162
4.2.4	Protoplast transformation and transactivation assays	163
4.2.5	Mammalian cell culture	166
4.2.6	RNA sequencing of transgenic callus	166
4.2.7	Differential expression analysis	167
4.2.8	Data analysis and candidate identification	167
4.3	Results	169
4.3.1	<i>AhCYP76AD2</i> is the main target for the regulation of the betalain biosynthesis	169
4.3.2	In contrast to anthocyanins, betalains seem not to be regulated in MBW complexes	170
4.3.3	Betalain MYBs seem to require another transcription factor to stabilize the DNA-binding	172
4.3.4	AhMYB2 induces betalain accumulation by upregulating <i>AhCYP76AD2</i> and <i>AhDODAA1</i> but lost the ability to regulate flavonoid pathway genes	173
4.4	Discussion	182
4.4.1	AhMYB2 regulates the betalain pathway by controlling the transcription of <i>AhCYP76AD2</i> and <i>AhDODAA1</i>	184
4.4.2	AhMYB2 seems different from its anthocyanin homologs, not acting in MBW complexes	185
4.4.3	AhMYB2 requires additional betalain-specific transcription factors for efficient DNA binding	186
4.4.4	Candidate transcription factors for (co-)regulation of the betalain pathway in amaranth	187
4.5	Acknowledgments	194
4.6	Author contributions	194
4.7	Competing interests	195
4.8	References	195
4.9	Supplement	213
4.9.1	Supplementary figures	214
4.9.2	Supplementary Tables	224
4.9.3	Appendix	236
4.10	Supplementary references	240
5	Overarching Discussion	243
5.1	Conclusion	249

5.2	References	251
	Acknowledgments	258
	Erklärung zur Dissertation	261

List of Figures

1.1	The production of anthocyanins and betalains across the plant kingdom . . .	10
1.2	Overview of the betalain and the anthocyanin pathway, including the involved enzymes	11
1.3	Overview of the MBW complex of <i>A. thaliana</i>	16
2.1	Overview of CRISPR/Cas9-mediated targeted mutagenesis in grain amaranth.	53
S2.1	Plasmid map of the CRISPR/Cas9 vector used for the targeted mutagenesis (CPGE_VEC00447)	60
S2.2	Genotyping of genomic DNA of 74 calli for the presence of the Cas9 gene using primer pairs mentioned in Supplemental Table S2.1	61
S2.3	Gel electrophoresis of target amplicons for <i>AhCYP76AD5</i> (gRNA 4) using primer pairs mentioned in Supplemental Table S2.1	62
3.1	Annotation of the <i>A. hypochondriacus</i> genome v2.2. (a) Simplified schematics of the general phenylpropanoid and flavonoid, and (b) betalain pathways.	76
3.2	Neighbour-joining phylogenetic tree based on multiple sequence alignment of the R2R3, 3R and 4R MYB protein sequences of <i>A. hypochondriacus</i> (blue), <i>A. thaliana</i> and <i>B. vulgaris</i>	85
3.3	Quantification of betalain content for parental <i>A. hypochondriacus</i> accessions used in mapping.	88
3.4	Biosynthesis mapping identifies functional variants in <i>AhCYP76AD2</i>	90
3.5	Regulator mapping identifies the <i>A. hypochondriacus</i> flower color regulator <i>AhMYB2</i>	92
3.6	Functional validation of <i>AhMYB2</i> as betalain regulator.	94
S3.1	Organelle contamination in the <i>A. hypochondriacus</i> assembly.	137
S3.2	Summary of the functional annotation of the <i>A. hypochondriacus</i> reference genome v2.2.	138
S3.3	Phylogenetic analysis of <i>A. hypochondriacus</i> CYP76AD proteins with CYP protein sequences from other species.	139

S3.4	Phylogenetic analysis of <i>A. hypochondriacus</i> DODA proteins with DODA protein sequences from other species.	140
S3.5	Relative gene expression in transcript per million (TPM) of identified <i>A. hypochondriacus</i> betalain pathway genes in different tissues.	141
S3.6	Relative gene expression in transcript per million (TPM) of identified <i>A. hypochondriacus</i> flavonoid pathway genes in different tissues.	142
S3.7	Relative gene expression in transcript per million (TPM) of identified <i>A. hypochondriacus</i> MYB transcription factor genes in different tissues. . . .	143
S3.8	Photometric and LC-MS quantification of betalain pathway metabolites for bulk segregant analysis (BSA) <i>A. hypochondriacus</i> parental genotypes	144
S3.9	Gene expression of <i>A. hypochondriacus</i> betalain pathway genes in flower tissue in green and red bulks of regulator and biosynthesis bulk segregant analyses (BSAs).	145
S3.10	Segregating variants called from <i>A. hypochondriacus</i> whole-genome sequencing data in the biosynthesis BSA and their relative allele depth for the <i>AhCYP76AD2</i> gene.	146
S3.11	Missense and stop alleles in <i>AhCYP76AD2</i> from RNA-seq reads of <i>A. hypochondriacus</i> red flower tissue.	147
S3.12	Sanger sequencing of <i>AhCYP76AD2</i> from biosynthesis bulk segregant analysis (BSA) individuals.	148
S3.13	Relative gene expression in transcript per million (TPM) of identified <i>A. hypochondriacus</i> flavonoid pathway genes in red and green flowers.	149
S3.14	Relative quantification of amaranthin content from transgenic <i>A. hypochondriacus</i> hairy roots.	150
S3.15	Prediction of protein-protein interactions of selected MYB and bHLH proteins.	151
4.1	AhMYB2 seems, in contrast to his homologue AtPAP1, not to form MBW complexes	171
4.2	Y2H experiments found no interaction of AhMYB2 with any of the tested <i>A. thaliana</i> clade IIIb bHLH transcription factors	172
4.3	Transactivation assays in protoplasts indicate that a transcription factor specific for betalain species is required to stabilize the binding of betalain pathway-regulating MYBs	174
4.4	The influence of genetics and the environment on the betalain pigmentation in transgenic <i>A. hypochondriacus</i> callus.	175
4.5	Overview of RNA-seq results from transgenic <i>A. hypochondriacus</i> callus showing the influence on gene expression of both environmental and transgenic-caused activation of betalain production.	178

4.6	Expression of betalain pathway genes in environmentally or transgenically induced betalain accumulation in <i>A. hypochondriacus</i> callus.	180
4.7	Comparison of differentially expressed transcription factors (TFs) and the frequency of main TF classes between samples	181
4.8	Differentially expressed WRKY transcription factors in environmentally or transgenically induced betalain accumulation in <i>A. hypochondriacus</i> callus and the closest related <i>A. thaliana</i> homologues from all WRKY transcription factors in <i>A. hypochondriacus</i>	183
S4.1	Results of all tested yeast-two-hybrid combinations of set 1.	214
S4.2	Results of all tested yeast-two-hybrid combinations of set 2.	215
S4.3	Transactivation assay in mammalian cells and synthetic positive control for amaranth promoter in protoplast.	216
S4.4	Presence, location, and distribution of predicted binding sites of main transcription factor classes across the <i>CYP76AD2</i> promoter and <i>DODAa1</i> promoter.	217
S4.5	Expression of <i>aragonate dehydratase (ADT)</i> and <i>aragonate dehydrogenase (ADH)</i> genes in environmentally or transgenically induced betalain accumulation in <i>A. hypochondriacus</i> callus.	218
S4.6	Expression of flavonoid pathway genes in environmentally or transgenically induced betalain accumulation in <i>A. hypochondriacus</i> callus.	219
S4.7	Expression of differentially expressed AP2/ERF transcription factors in environmentally or transgenically induced betalain accumulation in <i>A. hypochondriacus</i> callus.	220
S4.8	Expression of differentially expressed bZIP transcription factors in environmentally or transgenically induced betalain accumulation in <i>A. hypochondriacus</i> callus.	221
S4.9	Expression of differentially expressed NAC transcription factors in environmentally or transgenically induced betalain accumulation in <i>A. hypochondriacus</i> callus.	222
S4.10	Phylogeny of all <i>A. hypochondriacus</i> (red), <i>A. thaliana</i> (green), and <i>M. domestica</i> ERF transcription factors.	223
5.1	Comparison and conservation of the amino acid sequence of the R2 and R3 domains of PAP1 genes in closely related anthocyanin and betalain species	248

List of Tables

1.1	Published transcription factors involved in betalain regulation	18
S2.1	Primers used for cloning and genotyping	63
S2.2	Comparison of the obtained edit frequency from subcloning and deconvolution.	64
3.1	Comparison of genome annotation statistics of <i>A. cruentus</i> and <i>A. hypochondriacus</i>	83
S3.1	Summary of tissue of origin, read count, and source of long-read and short-read RNA sequencing data used to generate the <i>A. hypochondriacus</i> genome annotation v2.2.	123
S3.2	Number of PacBio Iso-Seq circular consensus sequencing (CCS) and full-length non-concatemer (FLNC) reads for all sequenced <i>A. hypochondriacus</i> tissues.	124
S3.3	Summary of BLAST results for betalain pathway genes.	125
S3.4	Statistics for whole-genome sequencing data used in the bulk segregant analyses (BSAs).	126
S3.5	Statistics for RNA sequencing data generated from bulk segregant analysis (BSA) bulks.	127
S3.6	List of primers used in the study.	128
S3.7	Repetitive element content of the polished <i>A. hypochondriacus</i> genome assembly v2.2.	129
S3.8	BUSCO score summary of different data sets against the Embryophyta reference set.	130
S3.9	Sequences used for phylogenetic analysis of <i>A. hypochondriacus</i> CYP76AD genes.	131
S3.10	Sequences used for phylogenetic analysis of <i>A. hypochondriacus</i> DODA genes.	132
S3.11	Summary of flavonoid pathway gene identification	133
S3.12	MYB transcription factors identified in <i>A. hypochondriacus</i>	134

4.1	Selected top candidates for AhMYB2 co-transcription factors for the regulation of the betalain pathway in grain amaranth.	192
S4.1	Plasmids used in this study	224
S4.2	Primers used for cloning	227
S4.3	Tested yeast-two-hybrid combinations.	230
S4.4	Information about the promoters used for protoplast transactivation assays.	232
S4.5	Sources of the used PAP1 homolog sequences from different species.	233
S4.6	Identified amaranth anthocyanin pathway genes and the conserved amino acid sequences used for identification.	234
S4.7	Identified amaranth aragonate dehydratase (ADT) and aragonate dehydrogenase (ADH α) genes and the conserved amino acid sequences used for identification.	235

Publications

Vollmer, S.K., Stetter, M. G., & Hensel, G. (2026). First Successful Targeted Mutagenesis Using CRISPR/Cas9 in Stably Transformed Grain Amaranth Tissue. *Plant Biotechnology Journal*. doi: 10.1111/pbi.70590.

Winkler, T.S., **Vollmer, S.K.**, Dyballa-Rukes, N., Metzger, S., & Stetter, M. G. (2024). Isoform-resolved genome annotation enables mapping of tissue-specific betalain regulation in amaranth. *New Phytologist*. doi: 10.1111/nph.19736.

These two publications form Chapter 2 and Chapter 3. Small editorial changes had been done for the incorporation in the thesis, mostly for the layout. See <https://doi.org/10.1111/pbi.70590> and <https://doi.org/10.1111/nph.19736> for the journal version.

Contribution

Chapter 2

Experiments: All experimental steps were planned and performed by Vollmer, S.K.. In detail: Vollmer, S.K. designed the guides and cloned the CRISPR construct (CPGE_VEC00447). Vollmer, S.K. performed both *Amaranthus hypochondriacus* callus transformation, including the maintenance of the transgenic calli. Vollmer, S.K. sampled the explants, ground them, and extracted the gDNA. Vollmer, S.K. executed the RE-PCRs, Sanger sequencing, and subcloning. **Analysis:** The complete analysis was done by Vollmer, S.K.. In detail: Vollmer, S.K. analyzed the Sanger sequencing and performed the deconvolution using ICE. Vollmer, S.K. calculated the frequency of edits and the distribution of the different mutations. **Manuskript:** The manuscript draft was written by Vollmer, S.K.. All figures and the supplements were prepared by Vollmer, S.K. with edits from all authors. All authors edited the final version of the manuscript.

Chapter 3

Experiments: Vollmer, S.K. designed the primers for the amplification of *AhMYB2* and for genotyping of *AhCYP76AD2*. Vollmer, S.K. extracted the gDNA and executed the PCR for the *AhCYP76AD2* genotyping. Vollmer, S.K. did the cDNA synthesis and amplified *AhMYB2* from cDNA. Vollmer, S.K. cloned *AhMYB2* in the Green Gate entry module pGGC00 and subsequently in the expression plasmids *p35S:AhMYB2.1-GFP* and *p35S:AhMYB2.2-GFP*. Vollmer, S.K. adapted and established the hairy root protocol for *Amaranthus hypochondriacus* and performed all transformations. Vollmer, S.K. screened for transgenic roots and took the images with the fluorescence binocular. Vollmer, S.K. did the sampling for the betalain extraction of hairy roots. Vollmer, S.K. did the bombardment of leek epidermis cells assisted by Jessica Pietsch and took the images of the bombarded leek epidermis cells with the fluorescence microscope. **Analysis:** Vollmer, S.K. did the image analysis of all microscope and binocular images. **Manuskript:** Vollmer, S.K. wrote the draft of the "Cloning and overexpression of *AhMYB2*" methods

section and assisted in editing the final manuscript together with all authors. Vollmer, S.K. constructed figure 6 with edits by the other authors.

Chapter 4

Experiments: Vollmer, S.K. planned and did all experimental steps for the *Amaranthus hypochondriacus* callus transformation and RNA sequencing and most steps for the transactivation assays, while the yeast-two-hybrid experiments were done by Treder, L.M.. In detail: For the callus transformation and RNA sequencing: Vollmer, S.K. designed the primer for the amplification of *AhCYP76AD2* and cloned the expression plasmid (CPGE_VEC00653). Vollmer, S.K. recloned *AhMYB2* into expression plasmids with a hygromycin resistance cassette (CPGE_VEC00643) and (CPGE_VEC00644), while the CRISPR plasmid (CPGE_VEC00447) was cloned by Vollmer, S.K. in Chapter 2. As mentioned for the CRISPR knockout calli in Chapter 2, Vollmer, S.K. performed both *Amaranthus hypochondriacus* callus transformation including the maintenance of the transgenic calli. Vollmer, S.K. planned and performed the callus phenotyping, including doing the semiquantitative counting and taking the overview pictures and close-up representative images with the camera-equipped binocular. Vollmer, S.K. sampled the explants, ground them, and aliquot them. Vollmer, S.K. extracted the RNA and checked the quality. For the yeast-two-hybrid experiments: Vollmer, S.K. supervised the experiments performed by Treder, L.M.. For the protoplast transactivation assays: Vollmer, S.K. designed all primers. Vollmer, S.K. extracted the gDNA of *Amaranthus hypochondriacus*, *Beta vulgaris* and *Arabidopsis thaliana* and amplified the *AhCYP76AD2* and *BvCYP76AD1* promoters; Treder, L.M. assisted in the amplification of the *AtDFR* promoter. Vollmer, S.K. cloned all promoters and CDS into the transactivation assay plasmids except for the control plasmids pmz824 and pep602. Vollmer, S.K. designed the guides and cloned the PolyGEN synthetic positive control plasmid specific for *Ahc-DOPA5GT*. Vollmer, S.K. did the majority of all midipreps. Vollmer, S.K. prepared the *Amaranthus hypochondriacus* seedling plates for the transactivation assays. Vollmer, S.K. did all the transactivation assays, assisted by Schön, J. in some experiments for the *A. thaliana* protoplast isolation and the plate reader measurements. Vollmer, S.K. did the mammalian transactivation assay assisted by Schön, J.. **Analysis:** Vollmer, S.K. did most of the analysis. In detail: For the callus transformation and RNA sequencing: Vollmer, S.K. analyzed the semiquantitative counting data and did the image analysis of the close-up binocular images. Vollmer, S.K. did the analysis of the differential expression data, i.e., plotted the heatmaps, the Venn diagram, and the upset plot; identified the arag-onate dehydratase and dehydrogenase, flavonoid pathway, and transcription factor (TF) genes and the TF class they belong to and performed the chi-square test. Vollmer, S.K.

collected the PAP1 homolog sequences, plotted the logoplot, and searched for conserved amino acids and their position relative to known residues important for bHLH interaction or DNA binding (specificity). Vollmer, S.K. did the prediction of *cis* element binding sites in the *AhCYP76AD2* and *AhDODAA1* promoters using web tools, curated the obtained data, and plotted the prediction for selected TF classes along the *AhCYP76AD2* and *AhDODAA1* promoter sequences. Vollmer, S.K. identified WRKY transcription factors of *A. hypochondriacus* and *A. thaliana*. Vollmer, S.K. did the multiple sequence alignment, and constructed the phylogenetic tree for WRKY and ERF transcription factors. For the yeast-two-hybrid experiments: Vollmer, S.K. did the analysis of the Y2H data assisted by Treder, L.M.. For the protoplast transactivation assays: Vollmer, S.K. did all the analysis, i.e., calculating the mean of the replicates, if applicable normalizing the data by the second luciferase measurement, combining the data of multiple independent experiments by normalizing by the only_reporter control, calculating the overall mean, and performing ANOVA. **Chapter:** The chapter was written by Vollmer, S.K. with minor edits by Hensel, G., and Stetter, M.G.. All figures and the supplements were prepared by Vollmer, S.K..

Chapter 1

Introduction

1.1 Understanding evolution is the key to the past and future

All life on earth was and is formed by evolution. Wallace published articles about natural selection first (albeit not using this term) (Wallace, 1855; Darwin and Wallace, 1858), but it was Darwin with his groundbreaking publication "*The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*" first published in 1859, who brought evolution to the attention of the scientific society (Darwin, 1859). Yet, evolution was not commonly accepted among scientists in the 19th century, especially after the formation of Mendelian genetics at the beginning of the 20th century (Hall et al., 2008). The formulation of the Hardy-Weinberg law for calculating gene frequencies in populations under natural selection finally combined genetics and evolution and led to the rise of population genetics (Hall et al., 2008; Fisher, 1919). Population genetics, together with systematics and adaptive change, shaped the Modern Synthesis of Evolution (Hall et al., 2008). Biological evolution is now generally understood as inherited changes on the genetic, organismal level and/or the level of population of organisms over time (Hall et al., 2008). The processes underlying evolution are mainly natural selection and genetic drift. Natural selection is based on the hypothesis that individuals who are better fitted to their environment due to heritable variation can reproduce better, thus leading to the stepwise evolution of new species (Darwin, 1859; Hall et al., 2008). Genetic drift, in contrast, describes changes that are caused by random effects (Masel, 2011). Despite impressive advancement since the development of the concept of evolution, there are still many unknown details of this fundamental process.

Evolution and breeding are closely entangled. The artificial selection in breeding was key for Darwin to propose his theory of natural selection (Darwin, 1859). Especially the vast diversity of fancy pigeon breeds, all descending from one species, fascinated him and finally led him to dedicate a section of his book to fancy pigeons to illustrate selection (Darwin, 1859). This was just the start; by studying domestication, researchers have the unique opportunity to analyze the outcome of thousand-year-long repeated evolution, which led to our modern-day crops (Purugganan and Fuller, 2009). Moreover, with advances in archaeology, timelines and past phenotypes can be estimated, and since the emergence of archaeogenetics, even DNA information of past individuals can be compared (Dalal et al., 2023). Domestication research could, for example, gave insights into the constraints of evolution and the pace of speciation (Gaut, 2015; Stetter, 2020). Furthermore, the history of crops is often better documented, and large data sets with multiple hundred different accessions enable the analysis of evolutionary processes with high precision. For instance, the genetic effects of drastic range shifts on genetic diversity, mutation load, and local adaptation could be investigated using the introduction of maize to Europe and grain amaranth to India (Takou et al., 2025; Singh and Stetter, 2025).

vice versa, the understanding of evolutionary processes can guide breeding (Turner-Hissong et al., 2020). Inferring the genetic ancestry, including gene flow, selection, and bottlenecks during domestication, helps to explain the different potential and diversity of modern crops (Stetter, 2020). For example, a strong bottleneck could explain a higher load of deleterious mutation or incomplete domestication observed in some crop species (Stetter, 2020; Nabholz et al., 2014; Gonçalves-Dias et al., 2023; Stetter et al., 2020). Locating domestication genes using selective sweeps and other signs of selection enables identification of breeding targets (Turner-Hissong et al., 2020; Curtin et al., 2022). Furthermore, building on the theory of natural selection, new traits can be designed via targeted evolution (Kababji et al., 2024; Rao et al., 2021). Finally, all this knowledge can be applied for the *de novo* domestication of new crops (Zsögön et al., 2018; Yu et al., 2021; Liu et al., 2025; Lemmon et al., 2018; Li et al., 2018).

The term *de novo* domestication refers to the process during which multiple domestication target traits are introduced or recapitulated in a wild species within a few generations using genome editing (Zhang et al., 2023). The fundamental idea of *de novo* domestication is based on the fact that crops usually have a low genetic diversity, which constrains adaptation, while wild species, in contrast, often have desirable traits and high genetic diversity but perform poorly in agriculture (Rogo et al., 2024; Zhang et al., 2023; Jian et al., 2022). Traditional breeding crossing schemes to transfer the traits from one to the other are very time consuming, since many backcrosses to the elite variety are needed to achieve high agricultural performance (Zhu et al., 2020; Xie et al., 2022). The combined knowledge of decades of plant research, the rise of sequence information, and the development of precise and efficient multiplex genome editing were prerequisites for *de novo* domestication (Rogo et al., 2024; Fernie and Yan, 2019). Previous research on domestication targets revealed the genes responsible for key traits of domesticated species, reference genomes allow us to find the homologous genes in wild relatives, and genome editing enables us to recapitulate the mutations in the target genes within a few generations (Zsögön et al., 2018; Yu et al., 2021; Lemmon et al., 2018; Li et al., 2018; Huang et al., 2022; Rogo et al., 2024; Fernie and Yan, 2019). A prominent example is the study of Lemmon et al. (2018) who showed *de novo* domestication in the orphan crop ground-cherry (*Physalis pruinosa*), a tomato relative. By mutating two genes using CRISPR/Cas (clustered regularly interspaced short palindromic repeats / CRISPR-associated protein), they could achieve a more compact growth, up to 50% higher concentration of fruits along each shoot, and a 24% increase in fruit mass (Lemmon et al., 2018). This showcases how knowledge of crop domestication can be applied using cutting-edge molecular tools to develop new crop species urgently needed to improve our agriculture to feed a growing human population in a changing climate.

Thus, evolution and breeding are closely entangled. Evolution explains our history on earth but is also critical to improving breeding to guarantee future food security.

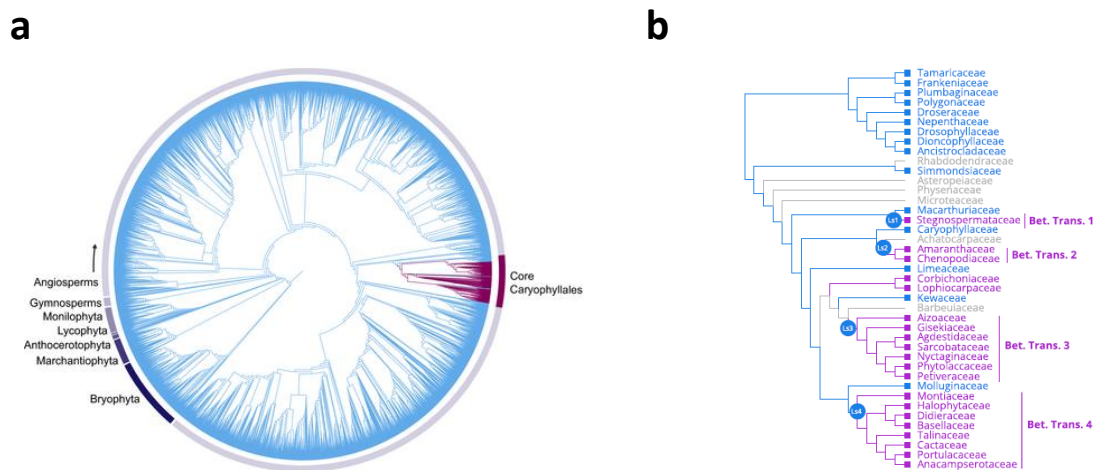


Figure 1.1: Plant phylogeny indicating produced pigments (blue for anthocyanins and purple for betalains)(a, figure modified from Timoneda et al. (2019)). Close up of the *Caryophyllales* showing the four proposed betalain origins (b, figure modified from Pucker et al. (2024)).

1.2 The betalain pathway as a case study for the evolution of new regulations

Despite massive advances, there are still many questions on evolution unanswered. One example is the question of how the regulatory network of new pathways (co-)evolves. While the evolution of the pathway enzymes is relatively well studied, the evolution of the regulation of these new pathways is still unclear. The recruitment model of metabolic evolution hypothesizes that the regulation evolves first and then recruits the enzymes to form the new pathway (Shoji, 2019). Additionally, it proposes that rather the new pathway genes evolve the needed *cis*-elements to adapt to other already existing conserved transcription factors than that new transcription factors evolve from scratch to target the new pathway genes (Shoji, 2019). Plants are characterized by their large number and diversity of secondary metabolism pathways, as they are a key strategy for sessile organisms to adapt to changing environments (Weng et al., 2021). This high rate of new pathway evolution makes plants valuable for evolutionary studies. Many plant orders can be identified by characteristic secondary metabolites (Kroymann, 2011; Dixon and Dickinson, 2024). For instance, only *Brassicales* plants produce glucosinolates, while betalains can be found only in the order *Caryophyllales* (Neugart et al., 2018; Brockington et al., 2011; Timoneda et al., 2019). Throughout the plant kingdom, anthocyanins are ubiquitously present as red pigments, but most *Caryophyllales* plants produce betalain instead (Brockington et al., 2011; Timoneda et al., 2019)(Figure 1.1a, Figure 1.2).

With tyrosine as a building block, the pathway requires only three enzymes and one spontaneous reaction (Figure 1.2a). In the first step, tyrosine is hydroxylated to L-5,6-dihydroxyphenylalanine (L-DOPA) by a cytochrome P450 enzyme (CYP76AD α or β)

(Sunnadeniya et al., 2016; Hatlestad et al., 2012). L-DOPA is subsequently converted to cyclo-DOPA by a CYP76AD α enzyme or by a DODA α 1 enzyme to betalamic acid (Polturak et al., 2016; Christinet et al., 2004; Sasaki et al., 2009; Hatlestad et al., 2012; Chung et al., 2015; DeLoache et al., 2015; Gandía-Herrero and García-Carmona, 2012). The addition of an amine or amino acid to betalamic acid results in yellow betaxanthins (Schliemann et al., 1999). Alternatively, betalamic acid spontaneously reacts with cyclo-DOPA (Strack et al., 2003). Glycosylation of cyclo-DOPA before or after the reaction with betalamic acid, catalyzed by a cyclo-DOPA glucosyltransferase, yields then the red-colored betacyanins (Sasaki et al., 2005; Vogt, 2002; Wyler et al., 1984; Sasaki et al., 2004; Vogt et al., 1999a). The flavonoid pathway, in contrast, uses phenylalanine as a precursor and involves more steps until the formation of anthocyanins (Figure 1.2b)(Tohge et al., 2013; Shoeva et al., 2016; Pandey et al., 2012).

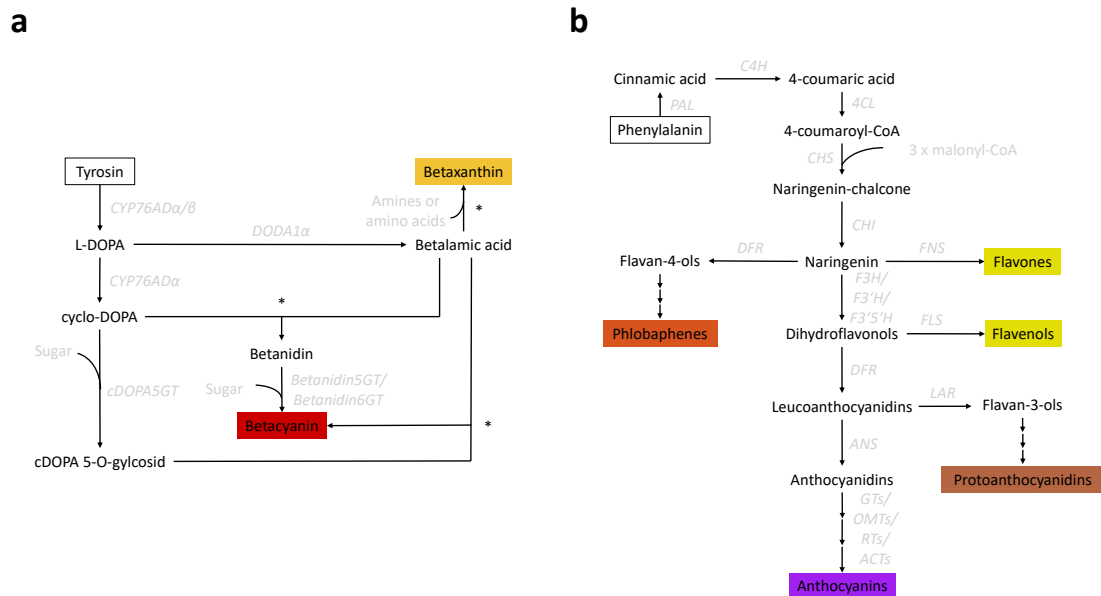


Figure 1.2: Overview of the betalain (a) and the anthocyanin pathway (b), including the involved enzymes. PAL (phenylalanine ammonium lyase), C4H (cinnamate 4-hydroxylase), 4CL (4-coumarate:CoA ligase), CHS (naringenin-chalcone synthase), CHI (chalcone isomerase), FNS (flavone synthase), F3H (flavanone 3-hydroxylase), FLS (flavonol synthase), F3'H (flavonoid 3'-hydroxylase), F3'5'H (flavonoid 3'5'-hydroxylase), DFR (dihydroflavonol 4-reductase), LAR (leucoanthocyanidin reductase), ANS (anthocyanidin synthase), ANR (anthocyanidin reductase), GTs (glucosyltransferases), OMTs (O-methyltransferases), RTs (rhamnosyltransferases), ACTs (acetyltransferases), DODA α 1 (L-3,4-dihydroxyphenylalanine 4,5-dioxygenase, α lineage), cDOPA5GT (*cyclo*-DOPA 5-O-glucosyltransferase), Betanidin5GT (betanidin 5-O-glucosyltransferase), Betanidin6GT (betanidin 6-O-glucosyltransferase).

There are three reasons why betalains and the *Caryophyllales* are ideally suited as a case study for the evolution of new gene regulations. First, genomic studies suggest that the betalain pathway evolved four times within the order (Brockington et al., 2015; Sheehan et al., 2020; Pucker et al., 2024; Walker-Hale et al., 2025)(Figure 1.1b). This enables the comparison of multiple evolutionary "replicates", which increases the robust-

ness of the findings. Second, the anthocyanin pathway, the ancient pathway with the most similar regulation (see the next section 1.3) and thus the most probable source for transcription factor to recruit for the betalain regulation, is very well studied (Cappellini et al., 2021; Winkel-Shirley, 2001; Zhao et al., 2023). This facilitates candidate identification and testing of the hypothesis. Thirdly, betalains as red pigments are easy to detect, extract, and measure (He et al., 2020; Pramanik et al., 2024). Betalains are visible to the naked eye with no special equipment, extraction protocols are simple, and measurement can be done photometrically (He et al., 2020; Pramanik et al., 2024).

1.3 Betalain and anthocyanin levels react remarkably similarly to internal and external signals

For the literature review presented here, more than 50 studies have been found that investigate internal and external cues influencing betalain regulation. However, most studies were performed with species from the *Amaranthaceae*, biasing the results of the literature review towards the proposed betalain origin 2 (according to Pucker et al. (2024), see figure 1.1b)(43, 5 and 10 studies from origins 2, 3, and 4, respectively; no studies on origin 1 found). Nevertheless, if information of more than one origin could be compared, were the observed effects of the respective cues mostly in line with each other.

1.3.1 Internal cues

The most discussed internal cues are phytohormones, involving all classes: For cytokinins, a positive effect on betalain regulation was observed in *A. caudatus*, *A. tricolor*, *Celosia argentea* *Chenopodium quinoa* and *Suaeda salsa* (Kinsman et al., 1975; Bianco-Colomas, 1980; Bianco-Colomas and Hugues, 1990; Sang A Roon et al., 2024; Henarejos-Escudero et al., 2018; Biswas et al., 2013; Elliott, 1983; De Nicola et al., 1975; Kochhar et al., 1981; Zhao et al., 2011; Zheng et al., 2016). For auxin, the effect is less clear, Biswas et al. (2013) reportet a activation by 1-naphthaleneacetic acid (NAA) in *A. tricolor* and Sakuta et al. (1991) an activation by 2,4-dichlorophenoxyacetic acid (2.4D) in *Phytolacca americana*, while Zhao et al. (2011) observed no significant effect by indole-3-acetic acid (IAA) but repression by 2.4D in *Suaeda salsa* and Zheng et al. (2016) observed repression of betalain accumulation by 2.4D in *A. tricolor*. Many studies investigated media with different ratios and combinations of auxins and cytokinins, most finding a slightly higher concentration of cytokinins than auxins as the best inducer of betalains (Pari et al., 2024; Milech et al., 2017; Reis et al., 2017; Girod and Zryd, 1991) suggesting a repression by high auxin concentrations; only one study observed rather the opposite (Reis et al., 2018).

Thus, besides the ratio of auxin and cytokinin, the type of auxin also seems to play a role, as shown by Bhuiyan et al. (2002) who observed a repression in *Portulaca* sp. if

IAA was added instead of 2.4D.

Other phytohormones were shown to influence betalain accumulation too: Ethylene was tested only by Cao et al. (2012) in *A. mangostanus* and Barba-Espin et al. (2018) in *B. vulgaris*, both groups observing an increase of betalain pigmentation upon treatment. Gibberellic acid (GA₃) in contrast, inhibited betalain production in *A. caudatus* (Kinsman et al., 1975; Bianco-Colomas, 1980) and *A. tricolor* (Zheng et al., 2016). Jasmonic acid induced only a mild activation in *Portulaca* sp., but methyl jasmonate induced strong betalain pigmentation also in many other betalain species (Bhuiyan and Adachi, 2003; Lakhotia et al., 2014; Mueangnak et al., 2025; Cao et al., 2012; Suresh et al., 2004). Mostly no significant effect was observed upon salicylic acid treatment (Ceja-López et al., 2022; Mueangnak et al., 2025; Cao et al., 2012; Pari et al., 2024), however Rodrigues-Brandão et al. (2014) and Wee et al. (2018) observed an activation in *Alternanthera tenella* and *Hylocereus polyrhizus*, respectively. Only Hirano et al. (1996) tested the effect of abscisic acid (ABA), finding a repression of the betalain production in *Phytolacca americana*. Unfortunately, no study was found characterizing the effect of strigolactones on the betalain pigmentation.

In summary, betalains seem to be regulated by all investigated phytohormones; only the salicylic acid pathway seems to have little influence. While the regulation by auxins appears to be more complicated, the effect of all other phytohormones is clearer and also seems consistent across all betalain origins. Generally, betalains seem to be positively regulated by methyl jasmonate, ethylene, and cytokinins and negatively by ABA and GA₃. When comparing the observed phytohormone regulation of betalains to the one from anthocyanins, it becomes evident that anthocyanins also get upregulated by methyl jasmonate (Ram et al., 2013; Qu et al., 2011) and downregulated by GA₃ (Loreti et al., 2008; Jiang et al., 2007). Similarly, it also seems the salicylic acid pathway for anthocyanins plays rather a minor role (Ram et al., 2013). Auxins clearly repress anthocyanins (Ji et al., 2015; Wang et al., 2020), but Ji et al. (2015) demonstrated in *Malus sieversii* that as well for anthocyanins, the type of auxin and the ratio to cytokinins are key. In contrast to the one study in betalains observing a repression by ABA, multiple studies show an activation of anthocyanins by ABA (Koyama et al., 2010; Loreti et al., 2008; Pirie and Mullins, 1976; Song et al., 2024; González-Villagra et al., 2017). Interestingly, the effect of cytokinins on the anthocyanin regulation seems to depend on the species: for some, including the model species *Arabidopsis thaliana* and *Nicotiana Benthamiana* there appears to be no link between cytokinins and anthocyanins, while for other species, mostly woody, anthocyanins get strongly upregulated after cytokinin treatment (Zhu et al., 2024; Pirie and Mullins, 1976; Loreti et al., 2008; Ji et al., 2015). Ethylene-mediated anthocyanin regulation also seems to depend on the species, as it was shown to repress anthocyanins in *Fragaria ananassa* and *Brassica oleracea* but activate anthocyanin accumulation in *Vitis vinifera* and *Malus domestica* (Guo et al., 2018; Kang and Burg, 1973; Pirie and Mullins, 1976; An et al., 2018; El-Kereamy et al., 2003). Thus,

the regulation by phytohormones from betalains and anthocyanins is mostly in line for auxin, GA₃, methyl jasmonate, and salicylic acid, only for ABA does lacking data and for cytokinin and ethylene, the strong species dependence prevents a solid conclusion. Not only phytohormones but also nutrients can influence betalain pigmentation. Sucrose was reported both to activate (Ceja-López et al., 2022; Anand et al., 2020; Sang A Roon et al., 2024) and to suppress betalain accumulation (Berlin et al., 1986; Winson et al., 2020). A reduction in total nitrogen (Henarejos-Escudero et al., 2018; Akita et al., 2002; De Aguiar Lage et al., 2015) or of ammonium compared to nitrate could induce betalain in some studies (Akita et al., 2002; De Aguiar Lage et al., 2015), while nitrogen had less influence on betalain pigmentation in others (Trejo-Tapia et al., 1999; Berlin et al., 1986; Contreras-Ovalle et al., 2019). Only one study found that an increase of nitrogen increases the production (Bianco-Colomas and Hugues, 1990). Phosphate had no significant effect on betalain production in the studies from Berlin et al. (1986) and Bianco-Colomas and Hugues (1990), but decreasing phosphate level increased betalain content in the studies from Contreras-Ovalle et al. (2019) and Taya et al. (1994). For anthocyanins, is the regulation by nutrients more clear: high concentrations of sucrose (Loreti et al., 2008; Pirie and Mullins, 1976; Li et al., 2014; Ozden, 2024) and, in contrast, nitrogen (Ito et al., 2015; Zhang et al., 2017; Pirie and Mullins, 1976; Ji et al., 2015) and phosphate (Lei et al., 2011; Ito et al., 2015; Song et al., 2024; Jiang et al., 2007; Dédaldéchamp and Uhel, 1999) depletion lead to upregulation of anthocyanins. The inconsistent data for betalains hamper the comparison, but also the nutrient response of betalains and anthocyanins seems to be quite similar.

1.3.2 External cues

Both abiotic and biotic cues influence the betalain pigmentation. Regarding abiotic factors, light was identified as one of the most important positive signals for betalain production. All studies agree on the positive effect of (white) light (Winson et al., 2021; Cao et al., 2012; Gbrenović, 1985; De Nicola et al., 1975; Kochhar et al., 1981; Shin et al., 2003; Böhm et al., 1991; Kinsman et al., 1975), most finding an even stronger induction by (additional) blue light (Portillo-Nava et al., 2021; Milech et al., 2017; Reis et al., 2018; Berlin et al., 1986; Anand et al., 2020; Kishima et al., 1995; Bhuiyan et al., 2002). Ultraviolet (UV) light yielded a moderate to strong increase in most studies (Bodhipadma et al., 2017; Vogt et al., 1999b; Kishima et al., 1995; Rudat and Göring, 1995), too, only Barba-Espin et al. (2018) and Winson et al. (2021) found no significant effect. For red light, the results were more mixed: activation was found for *A. caudatus*, *Portulaca* sp. and *H. costaricensis* but not *Alternanthera sessilis*, *A. brasiliana*, and *Chenopodium album* (Milech et al., 2017; Reis et al., 2018; Kochhar et al., 1981; Rudat and Göring, 1995; Kishima et al., 1995; Winson et al., 2021). A potential reason might be the differential effect of red light and far-red light, which was only accounted for by Shin et al. (2003)

and French et al. (1973). While Shin et al. (2003) found far-red slightly more positive than red light, did French et al. (1973) found a benefit in oscillating both red and far-red but using red as the last illumination. Like betalains, anthocyanins get upregulated by white light (Maier and Hoecker, 2015), blue light (Liu et al., 2022b; Shi et al., 2014) and UV light (Arakawa et al., 1985). Liu et al. (2015) showed in *A. thaliana* that both far-red and red light induce anthocyanins, but by different signaling cascades.

Other abiotic cues like drought (Casique-Arroyo et al., 2014; Vogt et al., 1999b), salt (Ceja-López et al., 2022; Vogt et al., 1999b; Casique-Arroyo et al., 2014), and low- (Bachelor thesis Maria Rodriguez Esmerodes, Universität zu Köln) and short high-temperature stress (Reyes-Rosales et al., 2023; Ceja-López et al., 2022) have been shown to increase betalain production. Only prolonged strong heat stress leads to a reduction in betalain production (Li et al., 2023; Reyes-Rosales et al., 2023). These observations are in line with the ones for anthocyanins in dicots, which get upregulated by drought (González-Villagra et al., 2017), salt (Kim et al., 2022), and low temperature (Zhang et al., 2011). However, different from the increase of betalains under short-term heat stress (Reyes-Rosales et al., 2023; Ceja-López et al., 2022), anthocyanins get reduced under heat stress (Lin-Wang et al., 2011).

Similar to abiotic cues, do most biotic cues influence betalain accumulation. Fungal (Savitha et al., 2006; Warhade and Badere, 2018), bacterial (Savitha et al., 2006; Sepúlveda-Jiménez et al., 2004), and yeast elicitors (Savitha et al., 2006) have been demonstrated to induce an increase in betalain production. Similarly, do microbial polysaccharides like chitosan or β -glucan increase betalain production in low concentrations (Bhuiyan and Adachi, 2003; Lakhotia et al., 2014; Mueangnak et al., 2025; Savitha et al., 2006). Wounding and herbivory were shown to increase betalain pigmentation, too (Casique-Arroyo et al., 2014; Sepúlveda-Jiménez et al., 2004). Adding H_2O_2 , a common signal involved in defense, increased betalains accumulation in *B. vulgaris* (Sepúlveda-Jiménez et al., 2004) but could only affect it in combination with methyl jasmonate in *A. mangostanus* (Cao et al., 2012). The reaction direction to biotic cues is identical to the one from anthocyanins, which also increase upon exposure to fungal (Liu et al., 2020; Qu et al., 2011), bacterial (Suvarnalatha et al., 1994), and yeast elicitors (Suvarnalatha et al., 1994), as well as to some microbial polysaccharides (Cai et al., 2012; Shpileuski et al., 2024), wounding (Zhang et al., 2024) and H_2O_2 (Barba-Espín et al., 2021).

Putting all these observations together reveals a remarkably similar regulation pattern of betalains and anthocyanins. This continues on the spatial regulation, where betalains and anthocyanins are also often found respectively in the same organs and tissues (Jain and Gould, 2015). Considering that both have similar biochemical properties (antioxidizing, high osmolaric potential, red color) and are postulated to fulfill similar biological functions, it is highly reasonable for betalains and anthocyanins to be similarly regulated (Jain and Gould, 2015).

1.4 The transcriptional regulation of the betalain pathway could be co-opted from the anthocyanin regulation

According to the recruitment model, it would be plausible to expect that for the transcriptional regulation of the betalain pathway, transcription factors of the similarly functioning and regulated anthocyanin pathway were co-opted (Shoji, 2019).

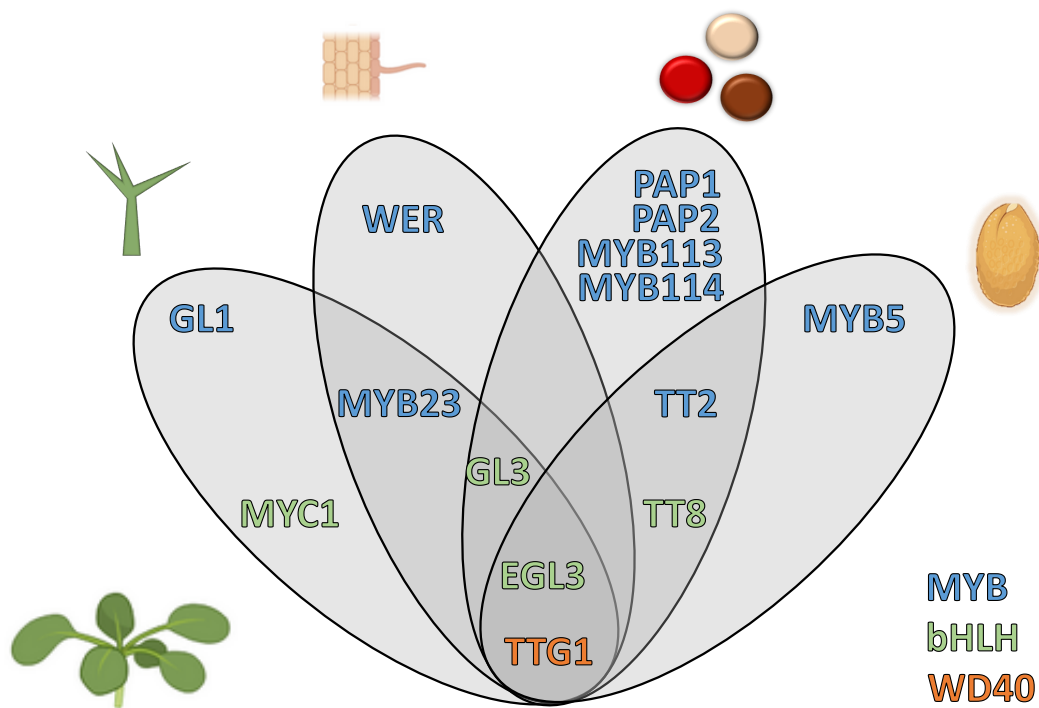


Figure 1.3: Overview of the MYB-bHLH-WD40 (MBW) complex of *A. thaliana* (based on figure 1 of Chen and Wang (2019)). Root hair, *A. thaliana* and seed mucilage icons are from BioRender.

Anthocyanins, together with other pathways, are regulated by complexes from myeloblastosis protein (MYB), basic helix-loop-helix (bHLH) and WD40 transcription factors (MBW complexes) (Lloyd et al., 2017)(Figure 1.3). The WD40 transcription factor is conserved across multiple pathways (Walker et al., 1999), while the bHLH and especially the MYB transcription factors confer specificity to particular pathways and fine-tune the regulation (Lloyd et al., 2017; Chen and Wang, 2019; Li, 2014; Grotewold et al., 2000; Stracke et al., 2001; Lepiniec et al., 2006)(Figure 1.3). While there are a range of studies on internal and external cues influencing betalain regulation, studies investigating transcription factors involved in betalain regulation are scarce. Individual transcription factors had been correlated with betalain production, but there are only a few verified, and for even fewer has the molecular mode of action been investigated (see Table 1.1 for

a list of all publications). Moreover, for none has the regulatory network been characterized. Hatlestad et al. (2015) could identify and verify BvMYB1 as a positive regulator of the betalain pathway by overexpression in transgenic hairy root of *B. vulgaris* and *A. hypochondriacus*. Additionally, they could prove direct binding of BvMYB1 to the *BvCYP76AD1* and *BvDODA1* promoter in Y1H. For pitaya, there had been several WRKY, several MYB, several bHLH, and an SPL transcription factor found to be correlated with betalain pigmentation and binding and/or transactivation of betalain pathway genes shown in heterologous assays (see Table 1.1 for more details). More recently there were also an MYB and a WRKY transcription factor described for the leafy amaranth species *Amaranthus tricolor* (Table 1.1).

1.5 The high diversity in pigmentation of grain amaranth is optimal to investigate betalain pigmentation

One of only a fraction of plants that produce betalains is the orphan crop amaranth. The genus *Amaranthus* contains around 50-70 species of annual or short-lived perennial herbs (Costea and DeMason, 2001; Dorling, 2008; Assad et al., 2017). Being domesticated parallel to maize in Middle and South America (Stetter et al., 2020), amaranths nowadays are used both as ornamental plants and as leaf or grain crops (Sauer, 1967; Assad et al., 2017). Despite their long cultivation history, amaranth still has a high diversity, both genetically and phenotypically (Riggins and Mumm, 2021; Stetter et al., 2025). The high diversity also includes a range of different pigmentation phenotypes, making amaranth well suited to study betalain regulation (Howard et al., 2022). Additionally, amaranth combines a range of different characteristics of high value for breeding climate-resilient crops for future food security (Aderibigbe et al., 2020; Alemayehu et al., 2015; Hoidal et al., 2019; Joshi et al., 2018). First, amaranth is drought and heat tolerant, facilitated by its C₄ photosynthesis (Stallknecht and Schulz-Schaeffer, 1993; Assad et al., 2017). Second, amaranth is highly nutritional, having an amino acid composition that fits perfectly to human needs and containing high amounts of minerals, vitamins, and unsaturated fatty acids (reviewed by Joshi et al. (2018) and Assad et al. (2017)). Finally, are the grains from amaranth gluten-free and thus compatible with the growing demand from people who suffer from gluten-related disorders like celiac disease (Ballabio et al., 2011; King et al., 2020). Thus, amaranth is not only well suited to study the betalain regulation but also of high interest for breeding.

Table 1.1: Published transcription factors involved in betanin regulation. Columns: paper: reference; species: species from which the investigated transcription factor is, *Beta vulgaris* (*B. vulgaris*), *Hylocereus undatus* (*H. undatus*), *Hylocereus polyrhizus* (*H. polyrhizus*) also referred to as *Hylocereus monacanthus*, *Amaranthus tricolor* (*A. tricolor*); TF: name of the investigated transcription factor; type: class of the investigated transcription factor; +/-: indicates the direction as positive (+) or negative (-); binding: indicates if binding to selected promoter of the endogenous betanin pathway was verified heterologously; TA: indicates if transactivation was demonstrated in heterologous system; active: indicates if activity of the transcription factor was shown on the endogenous pathway either by overexpression or VIGS in the investigated species; CRE: indicates if the respective *cis* elements were verified; interactors: indicates if potential interactors were tested heterologously, (h) = homodimerization shown, (+) found interaction to the investigated transcription factor, (-) no interaction to the investigated transcription factor found. yes = it was shown, no = tested but result negative, NA = not tested, (yes) = only indirectly shown. * = no TA ability, **=(represses the activator), *** = to HmoWRKY40 promoter

paper	species	TF	type	+/-	binding	TA	active	CRE	interactors
Hatlestad et al. (2015)	<i>B. vulgaris</i>	BvMYB1	MYB	+	yes	NA	yes	NA	AtEGL3 (-)
Xie et al. (2021)	<i>H. undatus</i>	HuMYB1	MYB	-	(yes)	yes	NA	NA	NA
Xie et al. (2023)	<i>H. undatus</i>	HuMYB132	MYB	+	yes	yes	yes	yes	NA
Xie et al. (2024)	<i>H. undatus</i>	HuMYB9	MYB	-	yes	yes	yes	yes	NA
Xue et al. (2025)	<i>A. tricolor</i>	AtMYB72	MYB	+	yes	yes	yes	NA	NA
Chen et al. (2023)	<i>H. undatus</i>	HuHHLH159	bHLH	-	yes	yes	yes	NA	NA
Cai et al. (2025)	<i>H. polyrhizus</i>	HpbHLH48	bHLH	+	yes	yes	yes	NA	HpbHLH48 (h), HpbHLH64 (-), HpbHLH159 (-)
Cai et al. (2025)	<i>H. polyrhizus</i>	HpbHLH64	bHLH	*	yes	no	yes	NA	HpbHLH64 (h), HpbHLH48 (-), HpbHLH159 (-)
Cheng et al. (2017)	<i>H. polyrhizus</i>	HpWRKY44	WRKY	+	yes	yes	NA	yes	NA
Zhang et al. (2021)	<i>H. polyrhizus</i>	HmoWRKY40	WRKY	+	yes	yes	yes	NA	NA
Chen et al. (2022)	<i>H. undatus</i>	HmoWRKY42	WRKY	-	yes	yes	yes	NA	HmoWRKY42 (h)
Yang et al. (2023)	<i>A. tricolor</i>	AtWRKY42-2	WRKY	+	yes	NA	yes	NA	NA
Zeng et al. (2023)	<i>H. undatus</i>	HuSPL12	SPL	- **	yes ***	yes	NA	NA	HmoMYB1 (+), HmoMYB132 (+), HmoWRKY42 (+), HmoWRKY40 (-), HmoWRKY44 (-)

1.6 Genetic resources and stable transformation are the foundation of most state-of-the-art tools, but are restricted to model species

Scientific progress is largely linked to the development of better and new methods (Cupepus and Queitsch, 2020). A range of new technologies, like single-cell sequencing, super-resolution microscopy, and CRISPR/Cas9, to just mention a few, have enabled previously unprecedented discoveries. By combining multiple single nuclei sequencing experiments, it was possible to construct a database following cell type lineages and their expression profile along human stem cell-based embryos from zygote to gastrula, a highly valuable resource for developmental biology and the development of new therapies (Zhao et al., 2025). The group of Mathew Horrocks developed a method (LIVE-PAINT) to image single protein molecules with a resolution of less than 100 nm within living cells (Oi et al., 2020; Bhaskar et al., 2025). With CRISPR/Cas9, chromosome arms could be relocated (Beying et al., 2020) or specific target genes could be upregulated (e.g (Pan and Qi, 2023)) or silenced (e.g., (Liu et al., 2022a)). But the flip side of the coin is that most of the cutting-edge tools are limited to a few model species. Nicely highlighted by the given examples, showing that new methods often get developed for mammals first and take years until applied to plants, let alone non-model plants. This causes entire research fields not present in model organisms, like betalains which none of the most used model organisms (*A. thaliana*, *N. benthamiana*, *O. sativa*, *H. vulgare*, *S. lycopersicum*) produce, to be unexplored (Rensing, 2017; Roeder et al., 2025). Moreover, it strongly hampers research and subsequent optimization of orphan crops, which are promising candidates to breed climate-resilient, high-yielding, nutritious crops for future food security (Tadele, 2019; Siddique et al., 2021; Kumar and Bhalothia, 2020). This restriction to model species mainly stems from the requirement of stable transformation and high-quality genomic information for most cutting-edge methods (Venezia and Creasey Krainer, 2021; Cardi et al., 2023; Scheben et al., 2017; Yaqoob et al., 2023).

1.6.1 Genetic transformation

Stable transformation describes the process in which DNA is delivered and stably integrated into the genome of an organism (Altpeter et al., 2016). The most common methods for DNA delivery are *Agrobacterium*-mediated transformation (Hwang et al., 2017), particlebombardment (Altpeter et al., 2005), or PEG-mediated transformation (Sahab et al., 2019; Anjanappa and Gruissem, 2021; Reed and Bargmann, 2021)(advantages and disadvantages, e.g., reviewed by Hwang et al. (2017)). To enable stable genetic transformation in most cases requires complicated protocols to regenerate the transformed plant parts (explants) into plants to transmit the genetic change to the next

generation (Altpeter et al., 2016; Hwang et al., 2017). This includes extensive tissue culture, which is highly time- and resource-demanding and strongly genotype-dependent, restricting this method to few species and well-equipped laboratories (Altpeter et al., 2016; Hwang et al., 2017; Anjanappa and Gruissem, 2021). As a consequence, there is a high demand to develop stable transformation methods that are less genotype-dependent, less resource-demanding, and also applicable in yet recalcitrant elite varieties and non-model species (Altpeter et al., 2016; Cardì et al., 2023; Venezia and Creasey Krainer, 2021; Maren et al., 2022). Employing developmental regulators, genes that regulate embryogenesis, regeneration, and stem cells, studies have shown an impressive improvement in regeneration efficiency even in more recalcitrant species (Maren et al., 2022; Gordon-Kamm et al., 2019; Youngstrom et al., 2025). Lian et al. (2022), Maher et al. (2020) and Kelly et al. (2025) have demonstrated that developmental regulators could even enable *in planta* transformation, circumventing the long and resource-demanding tissue culture phase. First studies expanding the range of species tested revealed that there is still a species-specific response, requiring the adaptation of, e.g., the used developmental regulator genes to the target species (Lian et al., 2022; Qi, 2025). Moreover, it is still unclear if endogenous or heterologous developmental regulator genes, and if yes, from which species, are best suited (Bull et al., 2023; Vandeputte et al., 2024). Nevertheless, developmental regulators are very promising, bringing stable transformation also for recalcitrant non-model species within reach (Zhu et al., 2025; Nelson-Vasilchik et al., 2018; Lian et al., 2022). This is often the starting point to establish other important resources like efficient promoters for gene expression or protoplasting, which subsequently enable more complex state-of-the-art tools like targeted mutagenesis using CRISPR/Cas9 and high-throughput assays.

1.6.2 Genome editing

Targeted mutagenesis, or more often used but less precise, genome editing, refers to the targeted induction of single base pair to kilobase pair changes in the genome sequence of an organism (Jung and Till, 2021). There is a range of different systems that can be used, including transcription activator-like effector nucleases (TALENs), zinc fingers, and recombinases (Sun et al., 2025), but the currently most used system is CRISPR/Cas9 (Shamshirgaran et al., 2022). CRISPR/Cas9 was adopted from the bacterial immune system (Nemudryi et al., 2014) and is such an important tool that it was acknowledged with the Nobel Prize in Chemistry 2020 (Jinek et al., 2012). The basic system consists of two parts: the Cas9 protein, which contains an endonuclease domain capable of introducing the double-strand break, and bound to it the guide RNA, which navigates the Cas9 protein to the target region (Nemudryi et al., 2014; Jiang and Doudna, 2017). In theory, can the guide RNA be designed to target any sequence of interest, only limited by the presence of the protospacer adjacent motif (PAM, NGG for SpCas9) next to it

(Jinek et al., 2012; Jiang and Doudna, 2017). Nevertheless, does the sequence, especially of the seed region (12 bp next to PAM, (Uniyal et al., 2019)), strongly influence the guide efficiency (reviewed by Konstantakos et al. (2022)). Nowadays a whole family of different Cas9 proteins and modified CRISPR/Cas9 systems has been developed to, e.g., precisely edit certain bases (base editing), introduce or delete longer DNA fragments, up- or downregulate genes, or modify the DNA methylation (reviewed e.g., by Anzalone et al. (2020), Cardi et al. (2023) and Tuncel et al. (2025)).

1.6.3 Genome sequencing

Constantly faster and higher quality sequencing services and a strong decrease in sequencing cost (from >5000\$/Mb in 2001 to <1 cent/MB in 2021, National Human Genome Research Institute) made genome sequencing generally more accessible (Marks et al., 2021; Hamilton and Robin Buell, 2012). This has led to an increasing coverage of genomic information across the whole plant phylogeny that is already observable (Ye and Fan, 2021; Ellegren, 2014; Marks et al., 2021). One example is the cereal orphan crop tef, where a recent collection of 220 accessions was resequenced, mapped to the previously assembled reference genome, and used to identify important breeding targets using genome-wide association studies (GWAS) (Jones et al., 2025). Nevertheless, while model species have already entered the area of pan-genomics, many important non-model species still lack a sufficient high-quality reference genome (Scheben et al., 2017).

This highlights the importance of including non-model species in genomic projects and establishing new methods for less genotype-dependent plant transformation to emancipate state-of-the-art tools to all plants, unlocking the massive potential of non-model plants (Scheben et al., 2017; Tadele, 2018; Venezia and Creasey Krainer, 2021).

The described lack of knowledge and established cutting-edge tools also hamper research of grain amaranth as a non-model orphan crop. However, efforts in recent times have strongly increased the amount of resources for amaranth above the level of many non-model plants, and thus amaranth belongs to the plants with the most resources within betalain-producing lineages. Currently we have an inbred line reference panel and reference genomes of most grain amaranth species, enabling GWAS and molecular studies. Nevertheless, despite efforts from others (Castellanos-Arévalo et al., 2020) and us in recent times (unpublished), there is no established stable transformation and regeneration protocol, and also targeted mutagenesis using CRISPR/Cas9 was not yet applied in amaranth.

1.7 Aims of the Thesis

Novelty drives the evolution of species, showcased by the high diversity of distinct secondary metabolites in plants that allow rapid response to external cues. How the regulation of new evolving pathways develops and how it is integrated into the existing network is not well understood. Besides its value for science, this knowledge can boost breeding of new traits to better adapt our crops to face the growing demand for food in a changing climate. This study focuses on reconstructing the regulation of the betalain pathway in grain amaranth (*Amaranthus hypochondriacus*) replacing the ubiquitous anthocyanins as red pigments. A well-studied ancient pathway for comparisons, four potential evolutionary origins, and easy detection highlight betalains as a case study for a newly evolved pathway. For an orphan crop and within the *Caryophyllales*, grain amaranth has above-average, high-quality genetic resources. Nevertheless, as most plant species, including other species of this order, amaranth lacks an efficient, stable transformation protocol, the establishment of targeted mutagenesis, and adapted protocols for state-of-the-art molecular methods.

Chapter 2 demonstrates the first successful application of targeted mutagenesis using CRISPR/Cas9 in stable transformed amaranth, which was published in 2026 in the *Plant Biotechnology Journal*.

Chapter 3 capitalizes on the high-quality genetic resources to identify and, in a second step, establishes hairy transformation to validate AhMYB2 as the first transcriptional regulator of the betalain pathway in amaranth, published 2024 in *New Phytologist*.

Chapter 4 combines molecular assays both adapted for amaranth and in heterologous systems with RNA sequencing to reconstruct the molecular framework of AhMYB2-based betalain regulation, presenting yet unpublished results.

1.8 References

- O. R. Aderibigbe, O. O. Ezekiel, S. O. Owolade, J. K. Korese, B. Sturm, and O. Hensel. Exploring the potentials of underutilized grain amaranth (*Amaranthus* spp.) along the value chain for food and nutrition security: A review. *Critical Reviews in Food Science and Nutrition*, pages 1–14, Oct. 2020. ISSN 1040-8398, 1549-7852. doi: 10.1080/10408398.2020.1825323. URL <https://www.tandfonline.com/doi/full/10.1080/10408398.2020.1825323>.
- T. Akita, Y. Hina, and T. Nishi. New Medium Composition for High Betacyanin Production by a Cell Suspension Culture of Table Beet (*Beta vulgaris* L.). *Bioscience, Biotechnology, and Biochemistry*, 66(4):902–905, Jan. 2002. ISSN 0916-8451, 1347-6947. doi: 10.1271/bbb.66.902. URL <https://academic.oup.com/bbb/article/66/4/902-905/5944936>.
- F. R. Alemayehu, M. A. Bendevis, and S.-E. Jacobsen. The Potential for Utilizing the Seed Crop Amaranth (*Amaranthus* spp.) in East Africa as an Alternative Crop to Support Food Security and Climate Change Mitigation. *Journal of Agronomy and Crop Science*, 201(5):321–329, Oct. 2015. ISSN 09312250. doi: 10.1111/jac.12108. URL <http://doi.wiley.com/10.1111/jac.12108>.
- F. Altpeter, N. Baisakh, R. Beachy, R. Bock, T. Capell, P. Christou, H. Daniell, K. Datta, S. Datta, P. J. Dix, C. Fauquet, N. Huang, A. Kohli, H. Mooibroek, L. Nicholson, T. T. Nguyen, G. Nugent, K. Raemakers, A. Romano, D. A. Somers, E. Stoger, N. Taylor, and R. Visser. Particle bombardment and the genetic enhancement of crops: myths and realities. *Molecular Breeding*, 15(3):305–327, Apr. 2005. ISSN 1380-3743, 1572-9788. doi: 10.1007/s11032-004-8001-y. URL <http://link.springer.com/10.1007/s11032-004-8001-y>.
- F. Altpeter, N. M. Springer, L. E. Bartley, A. Blechl, T. P. Brutnell, V. Citovsky, L. Conrad, S. B. Gelvin, D. Jackson, A. P. Kausch, P. G. Lemaux, J. I. Medford, M. Orozco-Cardenas, D. Tricoli, J. VanEck, D. F. Voytas, V. Walbot, K. Wang, Z. J. Zhang, and C. N. Stewart. Advancing Crop Transformation in the Era of Genome Editing. *The Plant Cell*, page tpc.00196.2016, June 2016. ISSN 1040-4651, 1532-298X. doi: 10.1105/tpc.16.00196. URL <https://academic.oup.com/plcell/article/28/7/1510-1520/6098292>.
- J.-P. An, X.-F. Wang, Y.-Y. Li, L.-Q. Song, L.-L. Zhao, C.-X. You, and Y.-J. Hao. EIN3-LIKE1, MYB1, and ETHYLENE RESPONSE FACTOR3 Act in a Regulatory Loop That Synergistically Modulates Ethylene Biosynthesis and Anthocyanin Accumulation. *Plant Physiology*, 178(2):808–823, Oct. 2018. ISSN 0032-0889, 1532-2548. doi: 10.1104/pp.18.00068. URL <https://academic.oup.com/plphys/article/178/2/808-823/6116639>.
- P. Anand, K. P. Singh, S. K. Singh, and K. V. Prasad. Sucrose and light induced betalain biosynthesis in callus cultures of bougainvillea (*Bougainvillea* spp.). *The Indian Journal of Agricultural Sciences*, 90(9):1658–1662, Oct. 2020. ISSN 2394-3319, 0019-5022. doi: 10.56093/ijas.v90i9.106587. URL <https://epubs.icar.org.in/index.php/IJAgS/article/view/106587>.
- R. B. Anjanappa and W. Gruissem. Current progress and challenges in crop genetic trans-

- formation. *Journal of Plant Physiology*, 261:153411, June 2021. ISSN 01761617. doi: 10.1016/j.jplph.2021.153411. URL <https://linkinghub.elsevier.com/retrieve/pii/S017616172100050X>.
- A. V. Anzalone, L. W. Koblan, and D. R. Liu. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology*, 38(7): 824–844, July 2020. ISSN 1087-0156, 1546-1696. doi: 10.1038/s41587-020-0561-9. URL <https://www.nature.com/articles/s41587-020-0561-9>.
- O. Arakawa, Y. Hori, and R. Ogata. Relative effectiveness and interaction of ultraviolet-B, red and blue light in anthocyanin synthesis of apple fruit. *Physiologia Plantarum*, 64(3):323–327, July 1985. ISSN 0031-9317, 1399-3054. doi: 10.1111/j.1399-3054.1985.tb03347.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1399-3054.1985.tb03347.x>.
- R. Assad, Z. A. Reshi, S. Jan, and I. Rashid. Biology of Amaranths. *The Botanical Review*, 83(4):382–436, Dec. 2017. ISSN 0006-8101, 1874-9372. doi: 10.1007/s12229-017-9194-1. URL <http://link.springer.com/10.1007/s12229-017-9194-1>.
- C. Ballabio, F. Uberti, C. Di Lorenzo, A. Brandolini, E. Penas, and P. Restani. Biochemical and Immunochemical Characterization of Different Varieties of Amaranth (*Amaranthus* L. spp.) as a Safe Ingredient for Gluten-free Products. *Journal of Agricultural and Food Chemistry*, 59(24):12969–12974, Dec. 2011. ISSN 0021-8561, 1520-5118. doi: 10.1021/jf2041824. URL <https://pubs.acs.org/doi/10.1021/jf2041824>.
- G. Barba-Espin, S. Glied-Olsen, T. Dzhanfezova, B. Joernsgaard, H. Lütken, and R. Müller. Preharvest application of ethephon and postharvest UV-B radiation improve quality traits of beetroot (*Beta vulgaris* L. ssp. *vulgaris*) as source of colourant. *BMC Plant Biology*, 18(1):316, Dec. 2018. ISSN 1471-2229. doi: 10.1186/s12870-018-1556-2. URL <https://bmcpplantbiol.biomedcentral.com/articles/10.1186/s12870-018-1556-2>.
- G. Barba-Espín, C. Martínez-Jiménez, A. Izquierdo-Martínez, J. R. Acosta-Motos, J. A. Hernández, and P. Díaz-Vivancos. H₂O₂-Elicitation of Black Carrot Hairy Roots Induces a Controlled Oxidative Burst Leading to Increased Anthocyanin Production. *Plants*, 10(12):2753, Dec. 2021. ISSN 2223-7747. doi: 10.3390/plants10122753. URL <https://www.mdpi.com/2223-7747/10/12/2753>.
- J. Berlin, S. Sieg, D. Strack, M. Bokern, and H. Harms. Production of betalains by suspension cultures of *Chenopodium rubrum* L. *Plant Cell, Tissue and Organ Culture*, 5(3):163–174, 1986. ISSN 0167-6857, 1573-5044. doi: 10.1007/BF00040126. URL <http://link.springer.com/10.1007/BF00040126>.
- N. Beying, C. Schmidt, M. Pacher, A. Houben, and H. Puchta. CRISPR–Cas9-mediated induction of heritable chromosomal translocations in *Arabidopsis*. *Nature Plants*, 6(6): 638–645, May 2020. ISSN 2055-0278. doi: 10.1038/s41477-020-0663-x. URL <https://www.nature.com/articles/s41477-020-0663-x>.
- H. Bhaskar, Z. Gidden, G. Viridi, D. Kleinjan, S. J. Rosser, S. Gandhi, L. Regan, and M. H. Horrocks. Super-resolution imaging of proteins inside live mammalian cells with mLIVE-PAINT. *Protein Science*, 34(2):e70008, Feb. 2025. ISSN 0961-8368, 1469-

- 896X. doi: 10.1002/pro.70008. URL <https://onlinelibrary.wiley.com/doi/10.1002/pro.70008>.
- M. N. H. Bhuiyan and T. Adachi. Stimulation of betacyanin synthesis through exogenous methyl jasmonate and other elicitors in suspension-cultured cells of *Portulaca*. *Journal of Plant Physiology*, 160(9):1117–1124, 2003. ISSN 01761617. doi: 10.1078/0176-1617-01044. URL <https://linkinghub.elsevier.com/retrieve/pii/S0176161704705015>.
- M. N. H. Bhuiyan, K. Murakami, and T. Adachi. Variation in betalain content and factors affecting the biosynthesis in *Portulaca* sp. 'Jewel' cell cultures. *Plant Biotechnology*, 19(5):369–376, 2002. ISSN 1347-6114, 1342-4580. doi: 10.5511/plantbiotechnology.19.369. URL http://www.jstage.jst.go.jp/article/plantbiotechnology/19/5/19_5_369/_article.
- J. Bianco-Colomas. Qualitative and quantitative aspects of betalains biosynthesis in *Amaranthus caudatus* L. var. *pendula* seedlings. *Planta*, 149(2):176–180, 1980. Publisher: Springer.
- J. Bianco-Colomas and M. Hugues. Establishment and Characterization of a Betacyanin Producing Cell Line of *Amaranthus tricolor*: Inductive effects of Light and Cytokinin. *Journal of Plant Physiology*, 136(6):734–739, Oct. 1990. ISSN 01761617. doi: 10.1016/S0176-1617(11)81354-4. URL <https://linkinghub.elsevier.com/retrieve/pii/S0176161711813544>.
- M. Biswas, S. S. Das, and S. Dey. Establishment of a stable *Amaranthus tricolor* callus line for production of food colorant. *Food Science and Biotechnology*, 22(S1):1–8, Feb. 2013. ISSN 1226-7708, 2092-6456. doi: 10.1007/s10068-013-0041-9. URL <http://link.springer.com/10.1007/s10068-013-0041-9>.
- K. Bodhipadma, S. Noichinda, P. Wachirabongkoth, A. Thongruang, K. Nathalang, Y. Palapol, and D. W. Leung. Callus induction and influence of UV-C on betalain content in the callus of an amaranth. *Rom Biotech Lett*, 22:12793–12801, 2017.
- S. F. Brockington, R. H. Walker, B. J. Glover, P. S. Soltis, and D. E. Soltis. Complex pigment evolution in the *Caryophyllales*. *New Phytologist*, 190(4):854–864, June 2011. ISSN 0028-646X, 1469-8137. doi: 10.1111/j.1469-8137.2011.03687.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.2011.03687.x>.
- S. F. Brockington, Y. Yang, F. Gandia-Herrero, S. Covshoff, J. M. Hibberd, R. F. Sage, G. K. S. Wong, M. J. Moore, and S. A. Smith. Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in *Caryophyllales*. *New Phytologist*, 207(4):1170–1180, Sept. 2015. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.13441. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.13441>.
- T. Bull, J. Debernardi, M. Reeves, T. Hill, L. Bertier, A. Van Deynze, and R. Michelmore. GRF–GIF chimeric proteins enhance in vitro regeneration and *Agrobacterium*-mediated transformation efficiencies of lettuce (*Lactuca* spp.). *Plant Cell Reports*, Jan. 2023. ISSN 0721-7714, 1432-203X. doi: 10.1007/s00299-023-02980-4. URL <https://link.springer.com/10.1007/s00299-023-02980-4>.
- H. Böhm, L. Böhm, and E. Rink. Establishment and characterization of a betaxanthin-

- producing cell culture from *Portulaca grandiflora*. *Plant Cell, Tissue and Organ Culture*, 26(2):75–82, Aug. 1991. ISSN 0167-6857, 1573-5044. doi: 10.1007/BF00036109. URL <http://link.springer.com/10.1007/BF00036109>.
- J. Cai, L. Shen, X. Cai, Z. Yang, K. Zhou, X. Nie, Q. Hou, L. Wen, and K. Yang. HpbHLH48 and HpbHLH64 facilitated betalain biosynthesis via trans-activation of *HpADH1*, *HpCYP76AD1-1*, *HpDODA1*, and *HpB5GT5* in pitaya. *Plant Physiology and Biochemistry*, page 110426, Aug. 2025. ISSN 09819428. doi: 10.1016/j.plaphy.2025.110426. URL <https://linkinghub.elsevier.com/retrieve/pii/S0981942825009544>.
- Z. Cai, A. Kastell, I. Mewis, D. Knorr, and I. Smetanska. Polysaccharide elicitors enhance anthocyanin and phenolic acid accumulation in cell suspension cultures of *Vitis vinifera*. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 108(3):401–409, Mar. 2012. ISSN 0167-6857, 1573-5044. doi: 10.1007/s11240-011-0051-3. URL <http://link.springer.com/10.1007/s11240-011-0051-3>.
- S. Cao, T. Liu, Y. Jiang, S. He, D. K. Harrison, and D. C. Joyce. The effects of host defence elicitors on betacyanin accumulation in *Amaranthus mangostanus* seedlings. *Food Chemistry*, 134(4):1715–1718, Oct. 2012. ISSN 03088146. doi: 10.1016/j.foodchem.2012.03.129. URL <https://linkinghub.elsevier.com/retrieve/pii/S0308814612006140>.
- F. Cappellini, A. Marinelli, M. Toccaceli, C. Tonelli, and K. Petroni. Anthocyanins: From Mechanisms of Regulation in Plants to Health Benefits in Foods. *Frontiers in Plant Science*, 12:748049, Oct. 2021. ISSN 1664-462X. doi: 10.3389/fpls.2021.748049. URL <https://www.frontiersin.org/articles/10.3389/fpls.2021.748049/full>.
- T. Cardi, J. Murovec, A. Bakhsh, J. Boniecka, T. Brueggemann, S. E. Bull, T. Eeckhaut, M. Fladung, V. Galovic, A. Linkiewicz, T. Lukan, I. Mafra, K. Michalski, M. Kavas, A. Nicolai, J. Nowakowska, L. Sági, C. Sarmiento, K. Yildirim, M. Zlatković, G. Hensel, and K. Van Laere. CRISPR/Cas-mediated plant genome editing: outstanding challenges a decade after implementation. *Trends in Plant Science*, 28(10): 1144–1165, Oct. 2023. ISSN 13601385. doi: 10.1016/j.tplants.2023.05.012. URL <https://linkinghub.elsevier.com/retrieve/pii/S1360138523001644>.
- G. Casique-Arroyo, N. Martínez-Gallardo, L. González de la Vara, and J. P. Délano-Frier. Betacyanin Biosynthetic Genes and Enzymes Are Differentially Induced by (a)biotic Stress in *Amaranthus hypochondriacus*. *PLoS ONE*, 9(6):e99012, June 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0099012. URL <https://dx.plos.org/10.1371/journal.pone.0099012>.
- A. P. Castellanos-Arévalo, A. A. Estrada-Luna, J. L. Cabrera-Ponce, E. Valencia-Lozano, H. Herrera-Ubaldo, S. de Folter, A. Blanco-Labra, and J. P. Délano-Frier. *Agrobacterium rhizogenes*-mediated transformation of grain (*Amaranthus hypochondriacus*) and leafy (*A. hybridus*) amaranths. *Plant Cell Reports*, 39(9):1143–1160, Sept. 2020. ISSN 0721-7714, 1432-203X. doi: 10.1007/s00299-020-02553-9. URL <https://link.springer.com/10.1007/s00299-020-02553-9>.
- J. A. Ceja-López, J. Morales-Morales, J. Araujo-Sánchez, W. G. Kantún, A. Ku, M. D. L. Miranda-Ham, L. C. Rodríguez-Zapata, and E. Castaño. Evaluation of

- Natural Pigments Production in Response to Various Stress Signals in Cell Lines of *Stenocereus queretaroensis*. *Plants*, 11(21):2948, Nov. 2022. ISSN 2223-7747. doi: 10.3390/plants11212948. URL <https://www.mdpi.com/2223-7747/11/21/2948>.
- C. Chen, F. Xie, K. Shah, Q. Hua, J. Chen, Z. Zhang, J. Zhao, G. Hu, and Y. Qin. Genome-Wide Identification of WRKY Gene Family in Pitaya Reveals the Involvement of HmoWRKY42 in Betalain Biosynthesis. *International Journal of Molecular Sciences*, 23(18):10568, Sept. 2022. ISSN 1422-0067. doi: 10.3390/ijms231810568. URL <https://www.mdpi.com/1422-0067/23/18/10568>.
- J. Chen, F. Xie, K. Shah, C. Chen, J. Zeng, J. Chen, Z. Zhang, J. Zhao, G. Hu, and Y. Qin. Identification of HubHLH family and key role of HubHLH159 in betalain biosynthesis by activating the transcription of *HuADH1*, *HuCYP76AD1-1*, and *HuDODA1* in pitaya. *Plant Science*, 328:111595, Mar. 2023. ISSN 01689452. doi: 10.1016/j.plantsci.2023.111595. URL <https://linkinghub.elsevier.com/retrieve/pii/S0168945223000122>.
- S. Chen and S. Wang. GLABRA2, a Common Regulator for Epidermal Cell Fate Determination and Anthocyanin Biosynthesis in *Arabidopsis*. *International Journal of Molecular Sciences*, 20(20):4997, Oct. 2019. ISSN 1422-0067. doi: 10.3390/ijms20204997. URL <https://www.mdpi.com/1422-0067/20/20/4997>.
- M.-n. Cheng, Z.-j. Huang, Q.-z. Hua, W. Shan, J.-f. Kuang, W.-j. Lu, Y.-h. Qin, and J.-y. Chen. The WRKY transcription factor HpWRKY44 regulates *CytP450-like1* expression in red pitaya fruit (*Hylocereus polyrhizus*). *Horticulture Research*, 4(1): 17039, Dec. 2017. ISSN 2052-7276. doi: 10.1038/hortres.2017.39. URL <https://academic.oup.com/hr/article/6447950>.
- L. Christinet, F. X. Burdet, M. Zaiko, U. Hinz, and J.-P. Zrýd. Characterization and Functional Identification of a Novel Plant 4,5-Extradiol Dioxygenase Involved in Betalain Pigment Biosynthesis in *Portulaca grandiflora*. *Plant Physiology*, 134(1):265–274, Jan. 2004. ISSN 1532-2548, 0032-0889. doi: 10.1104/pp.103.031914. URL <https://academic.oup.com/plphys/article/134/1/265/6112200>.
- H.-H. Chung, K. E. Schwinn, H. M. Ngo, D. H. Lewis, B. Massey, K. E. Calcott, R. Crowhurst, D. C. Joyce, K. S. Gould, K. M. Davies, and D. K. Harrison. Characterisation of betalain biosynthesis in Parakeelya flowers identifies the key biosynthetic gene *DOD* as belonging to an expanded *LigB* gene family that is conserved in betalain-producing species. *Frontiers in Plant Science*, 6, July 2015. ISSN 1664-462X. doi: 10.3389/fpls.2015.00499. URL <http://journal.frontiersin.org/Article/10.3389/fpls.2015.00499/abstract>.
- N. I. Contreras-Ovalle, J. Jauregui-Rincon, M. S. Santos-Diaz, and E. Perez-Molphe-Balch. Effect of culture media on the production of betalains from hairy root cultures of four cacti species. *Agrociencia*, 53(5), 2019.
- M. Costea and D. A. DeMason. Stem Morphology and Anatomy in *Amaranthus* L. (*Amaranthaceae*), Taxonomic Significance. *Journal of the Torrey Botanical Society*, 128(3):254, July 2001. ISSN 10955674. doi: 10.2307/3088717. URL <https://www.jstor.org/stable/3088717?origin=crossref>.

- J. T. Cuperus and C. Queitsch. Editorial overview: Technology development as a driver of biological discovery. *Current Opinion in Plant Biology*, 54:A1–A4, Apr. 2020. ISSN 13695266. doi: 10.1016/j.pbi.2020.06.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S1369526620300698>.
- S. Curtin, Y. Qi, L. E. P. Peres, A. R. Fernie, and A. Zsögön. Pathways to *de novo* domestication of crop wild relatives. *Plant Physiology*, 188(4):1746–1756, Mar. 2022. ISSN 0032-0889, 1532-2548. doi: 10.1093/plphys/kiab554. URL <https://academic.oup.com/plphys/article/188/4/1746/6444999>.
- V. Dalal, N. Pasupuleti, G. Chaubey, N. Rai, and V. Shinde. Advancements and Challenges in Ancient DNA Research: Bridging the Global North–South Divide. *Genes*, 14(2):479, Feb. 2023. ISSN 2073-4425. doi: 10.3390/genes14020479. URL <https://www.mdpi.com/2073-4425/14/2/479>.
- C. Darwin. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray, London, 1859. URL <https://darwin-online.org.uk/content/frameset?itemID=F373&viewtype=text&pageseq=1>.
- C. Darwin and A. Wallace. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Journal of the proceedings of the Linnean Society of London. Zoology*, 3(9):45–62, 1858. Publisher: Oxford University Press.
- D. De Aguiar Lage, M. Da Silva Tirado, S. R. Vanicore, K. C. De Carvalho Sabino, and N. Albarello. Production of betalains from callus and cell suspension cultures of *Pereskia aculeata* Miller, an unconventional leafy vegetable. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 122(2):341–350, Aug. 2015. ISSN 0167-6857, 1573-5044. doi: 10.1007/s11240-015-0771-x. URL <http://link.springer.com/10.1007/s11240-015-0771-x>.
- M. G. De Nicola, V. Amico, and M. Piattelli. Effects of light and kinetin on amaranthin synthesis induced by cAMP. *Phytochemistry*, 14(4):989–991, Apr. 1975. ISSN 00319422. doi: 10.1016/0031-9422(75)85173-9. URL <https://linkinghub.elsevier.com/retrieve/pii/0031942275851739>.
- W. C. DeLoache, Z. N. Russ, L. Narcross, A. M. Gonzales, V. J. J. Martin, and J. E. Dueber. An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. *Nature Chemical Biology*, 11(7):465–471, July 2015. ISSN 1552-4450, 1552-4469. doi: 10.1038/nchembio.1816. URL <http://www.nature.com/articles/nchembio.1816>.
- R. A. Dixon and A. J. Dickinson. A century of studying plant secondary metabolism—From “what?” to “where, how, and why?”. *Plant Physiology*, page kiad596, Jan. 2024. ISSN 0032-0889, 1532-2548. doi: 10.1093/plphys/kiad596. URL <https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiad596/7499152>.
- K. Dorling. *RHS AZ Encyclopedia of garden plants*. DK, United Kingdom, 1 edition, 2008. ISBN 978-1-4053-3296-5.

- F. Dédaldéchamp and C. Uhel. Induction of anthocyanin synthesis in nonpigmented grape cell suspensions by acting on DFR substrate availability or precursors level. *Enzyme and Microbial Technology*, 25(3-5):316–321, Aug. 1999. ISSN 01410229. doi: 10.1016/S0141-0229(99)00047-2. URL <https://linkinghub.elsevier.com/retrieve/pii/S0141022999000472>.
- H. Ellegren. Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, 29(1):51–63, Jan. 2014. ISSN 01695347. doi: 10.1016/j.tree.2013.09.008. URL <https://linkinghub.elsevier.com/retrieve/pii/S0169534713002310>.
- D. C. Elliott. The pathway of betalain biosynthesis: Effect of cytokinin on enzymic oxidation and hydroxylation of tyrosine in *Amaranthus tricolor* seedlings. *Physiologia Plantarum*, 59(3):428–437, Nov. 1983. ISSN 0031-9317, 1399-3054. doi: 10.1111/j.1399-3054.1983.tb04226.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1399-3054.1983.tb04226.x>.
- A. El-Kereamy, C. Chervin, J. Roustan, V. Cheynier, J. Souquet, M. Moutounet, J. Raynal, C. Ford, A. Latché, J. Pech, and M. Bouzayen. Exogenous ethylene stimulates the long-term expression of genes related to anthocyanin biosynthesis in grape berries. *Physiologia Plantarum*, 119(2):175–182, Oct. 2003. ISSN 0031-9317, 1399-3054. doi: 10.1034/j.1399-3054.2003.00165.x. URL <https://onlinelibrary.wiley.com/doi/10.1034/j.1399-3054.2003.00165.x>.
- A. R. Fernie and J. Yan. De Novo Domestication: An Alternative Route toward New Crops for the Future. *Molecular Plant*, 12(5):615–631, May 2019. ISSN 16742052. doi: 10.1016/j.molp.2019.03.016. URL <https://linkinghub.elsevier.com/retrieve/pii/S1674205219301297>.
- R. A. Fisher. XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 52(2):399–433, 1919. Publisher: Royal Society of Edinburgh Scotland Foundation.
- C. J. French, R. C. Peckett, and H. Smith. Effect of light and exogenously applied precursors on amaranthin synthesis in *Amaranthus caudatus*. *Phytochemistry*, 12(12):2887–2891, Dec. 1973. ISSN 00319422. doi: 10.1016/0031-9422(73)80501-1. URL <https://linkinghub.elsevier.com/retrieve/pii/0031942273805011>.
- F. Gandía-Herrero and F. García-Carmona. Characterization of recombinant *Beta vulgaris* 4,5-DOPA-extradiol-dioxygenase active in the biosynthesis of betalains. *Planta*, 236(1):91–100, July 2012. ISSN 0032-0935, 1432-2048. doi: 10.1007/s00425-012-1593-2. URL <http://link.springer.com/10.1007/s00425-012-1593-2>.
- B. S. Gaut. Evolution Is an Experiment: Assessing Parallelism in Crop Domestication and Experimental Evolution: (Nei Lecture, SBE 2014, Puerto Rico). *Molecular Biology and Evolution*, 32(7):1661–1671, July 2015. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msv105. URL <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msv105>.
- S. Gbrenović. Effect of riboflavine and cyanide on blue light induction of betacyanin

- formation in *Amamnthus caudatus* seedlings. *Physiologia Plantarum*, 65(4):418–422, Dec. 1985. ISSN 0031-9317, 1399-3054. doi: 10.1111/j.1399-3054.1985.tb08666.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1399-3054.1985.tb08666.x>.
- P.-A. Girod and J.-P. Zryd. Secondary metabolism in cultured red beet (*Beta vulgaris* L.) cells: Differential regulation of betaxanthin and betacyanin biosynthesis. *Plant Cell, Tissue and Organ Culture*, 25(1):1–12, Apr. 1991. ISSN 0167-6857, 1573-5044. doi: 10.1007/BF00033905. URL <http://link.springer.com/10.1007/BF00033905>.
- J. González-Villagra, L. V. Kurepin, and M. M. Reyes-Díaz. Evaluating the involvement and interaction of abscisic acid and miRNA156 in the induction of anthocyanin biosynthesis in drought-stressed plants. *Planta*, 246(2):299–312, Aug. 2017. ISSN 0032-0935, 1432-2048. doi: 10.1007/s00425-017-2711-y. URL <http://link.springer.com/10.1007/s00425-017-2711-y>.
- J. Gonçalves-Dias, A. Singh, C. Graf, and M. G. Stetter. Genetic Incompatibilities and Evolutionary Rescue by Wild Relatives Shaped Grain Amaranth Domestication. *Molecular Biology and Evolution*, 40(8):msad177, Aug. 2023. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msad177. URL <https://academic.oup.com/mbe/article/doi/10.1093/molbev/msad177/7239273>.
- B. Gordon-Kamm, N. Sardesai, M. Arling, K. Lowe, G. Hoerster, S. Betts, and T. Jones. Using Morphogenic Genes to Improve Recovery and Regeneration of Transgenic Plants. *Plants*, 8(2):38, Feb. 2019. ISSN 2223-7747. doi: 10.3390/plants8020038. URL <https://www.mdpi.com/2223-7747/8/2/38>.
- E. Grotewold, M. B. Sainz, L. Tagliani, J. M. Hernandez, B. Bowen, and V. L. Chandler. Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Proceedings of the National Academy of Sciences*, 97(25):13579–13584, Dec. 2000. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.250379897. URL <https://pnas.org/doi/full/10.1073/pnas.250379897>.
- J. Guo, S. Wang, X. Yu, R. Dong, Y. Li, X. Mei, and Y. Shen. Polyamines Regulate Strawberry Fruit Ripening by Absciscic Acid, Auxin, and Ethylene. *Plant Physiology*, 177(1):339–351, May 2018. ISSN 0032-0889, 1532-2548. doi: 10.1104/pp.18.00245. URL <https://academic.oup.com/plphys/article/177/1/339-351/6117046>.
- B. K. Hall, B. Hallgrímsson, and M. W. Strickberger. *Strickberger’s evolution: the integration of genes, organisms and populations*. Jones and Bartlett, Sudbury, Mass, 4th ed edition, 2008. ISBN 978-0-7637-0066-9.
- J. P. Hamilton and C. Robin Buell. Advances in plant genome sequencing. *The Plant Journal*, 70(1):177–190, Apr. 2012. ISSN 0960-7412, 1365-313X. doi: 10.1111/j.1365-313X.2012.04894.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2012.04894.x>.
- G. J. Hatlestad, R. M. Sunnadeniya, N. A. Akhavan, A. Gonzalez, I. L. Goldman, J. M. McGrath, and A. M. Lloyd. The beet R locus encodes a new cytochrome P450 required for red betalain production. *Nature Genetics*, 44(7):816–820, July 2012. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.2297. URL <http://www.nature.com/articles/ng.2297>.

- G. J. Hatlestad, N. A. Akhavan, R. M. Sunnadaniya, L. Elam, S. Cargile, A. Hembd, A. Gonzalez, J. M. McGrath, and A. M. Lloyd. The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway. *Nature Genetics*, 47(1):92–96, Jan. 2015. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3163. URL <http://www.nature.com/articles/ng.3163>.
- Y. He, T. Zhang, H. Sun, H. Zhan, and Y. Zhao. A reporter for noninvasively monitoring gene expression and plant transformation. *Horticulture Research*, 7(1):152, Dec. 2020. ISSN 2662-6810, 2052-7276. doi: 10.1038/s41438-020-00390-1. URL <https://www.nature.com/articles/s41438-020-00390-1>.
- P. Henarejos-Escudero, B. Guadarrama-Flores, M. A. Guerrero-Rubio, L. R. Gómez-Pando, F. García-Carmona, and F. Gandía-Herrero. Development of Betalain Producing Callus Lines from Colored Quinoa Varieties (*Chenopodium quinoa* Willd). *Journal of Agricultural and Food Chemistry*, 66(2):467–474, Jan. 2018. ISSN 0021-8561, 1520-5118. doi: 10.1021/acs.jafc.7b04642. URL <https://pubs.acs.org/doi/10.1021/acs.jafc.7b04642>.
- H. Hirano, M. Sakuta, and A. Komamine. Inhibition of betacyanin accumulation by abscisic acid in suspension cultures of *Phytolacca americana*. *Zeitschrift für Naturforschung C*, 51(11-12):818–822, 1996. Publisher: Verlag der Zeitschrift für Naturforschung.
- N. Hoidal, M. Díaz Gallardo, S.-E. Jacobsen, and G. Alandia. Amaranth as a Dual-Use Crop for Leafy Greens and Seeds: Stable Responses to Leaf Harvest Across Genotypes and Environments. *Frontiers in Plant Science*, 10:817, June 2019. ISSN 1664-462X. doi: 10.3389/fpls.2019.00817. URL <https://www.frontiersin.org/article/10.3389/fpls.2019.00817/full>.
- J. E. Howard, M. B. Villamil, and C. W. Riggins. Amaranth as a natural food colorant source: Survey of germplasm and optimization of extraction methods for betalain pigments. *Frontiers in Plant Science*, 13:932440, Sept. 2022. ISSN 1664-462X. doi: 10.3389/fpls.2022.932440. URL <https://www.frontiersin.org/articles/10.3389/fpls.2022.932440/full>.
- X. Huang, S. Huang, B. Han, and J. Li. The integrated genomics of crop domestication and breeding. *Cell*, 185(15):2828–2839, July 2022. ISSN 00928674. doi: 10.1016/j.cell.2022.04.036. URL <https://linkinghub.elsevier.com/retrieve/pii/S0092867422005347>.
- H.-H. Hwang, M. Yu, and E.-M. Lai. *Agrobacterium* -Mediated Plant Transformation: Biology and Applications. *The Arabidopsis Book*, 15:e0186, Jan. 2017. ISSN 1543-8120. doi: 10.1199/tab.0186. URL <http://www.bioone.org/doi/10.1199/tab.0186>.
- S. Ito, T. Nozoye, E. Sasaki, M. Imai, Y. Shiwa, M. Shibata-Hatta, T. Ishige, K. Fukui, K. Ito, H. Nakanishi, N. K. Nishizawa, S. Yajima, and T. Asami. Strigolactone Regulates Anthocyanin Accumulation, Acid Phosphatases Production and Plant Growth under Low Phosphate Condition in *Arabidopsis*. *PLOS ONE*, 10(3):e0119724, Mar. 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0119724. URL <https://dx.plos.org/10.1371/journal.pone.0119724>.

- G. Jain and K. S. Gould. Are betalain pigments the functional homologues of anthocyanins in plants? *Environmental and Experimental Botany*, 119:48–53, Nov. 2015. ISSN 00988472. doi: 10.1016/j.envexpbot.2015.06.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S0098847215001112>.
- X.-H. Ji, Y.-T. Wang, R. Zhang, S.-J. Wu, M.-M. An, M. Li, C.-Z. Wang, X.-L. Chen, Y.-M. Zhang, and X.-S. Chen. Effect of auxin, cytokinin and nitrogen on anthocyanin biosynthesis in callus cultures of red-fleshed apple (*Malus sieversii* f. *niedzwetzkyana*). *Plant Cell, Tissue and Organ Culture (PCTOC)*, 120(1):325–337, Jan. 2015. ISSN 0167-6857, 1573-5044. doi: 10.1007/s11240-014-0609-y. URL <http://link.springer.com/10.1007/s11240-014-0609-y>.
- L. Jian, J. Yan, and J. Liu. De Novo Domestication in the Multi-Omics Era. *Plant and Cell Physiology*, 63(11):1592–1606, Nov. 2022. ISSN 0032-0781, 1471-9053. doi: 10.1093/pcp/pcac077. URL <https://academic.oup.com/pcp/article/63/11/1592/6619104>.
- C. Jiang, X. Gao, L. Liao, N. P. Harberd, and X. Fu. Phosphate Starvation Root Architecture and Anthocyanin Accumulation Responses Are Modulated by the Gibberellin-DELLA Signaling Pathway in *Arabidopsis*. *Plant Physiology*, 145(4):1460–1470, Dec. 2007. ISSN 0032-0889, 1532-2548. doi: 10.1104/pp.107.103788. URL <https://academic.oup.com/plphys/article/145/4/1460-1470/6107235>.
- F. Jiang and J. A. Doudna. CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics*, 46(1):505–529, May 2017. ISSN 1936-122X, 1936-1238. doi: 10.1146/annurev-biophys-062215-010822. URL <http://www.annualreviews.org/doi/10.1146/annurev-biophys-062215-010822>.
- M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(6096):816–821, Aug. 2012. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1225829. URL <https://www.science.org/doi/10.1126/science.1225829>.
- M. R. W. Jones, W. Kebede, A. Teshome, A. Girija, A. Teshome, D. Girma, J. K. M. Brown, J. Quiroz-Chavez, C. S. Jones, B. B. H. Wulff, K. Assefa, Z. Tadele, L. A. J. Mur, S. Chanyalew, C. Uauy, and O. Shorinola. Population genomics uncovers loci for trait improvement in the indigenous African cereal tef (*Eragrostis tef*). *Communications Biology*, 8(1):807, May 2025. ISSN 2399-3642. doi: 10.1038/s42003-025-08206-5. URL <https://www.nature.com/articles/s42003-025-08206-5>.
- D. C. Joshi, S. Sood, R. Hosahatti, L. Kant, A. Pattanayak, A. Kumar, D. Yadav, and M. G. Stetter. From zero to hero: the past, present and future of grain amaranth breeding. *Theoretical and Applied Genetics*, 131(9):1807–1823, Sept. 2018. ISSN 0040-5752, 1432-2242. doi: 10.1007/s00122-018-3138-y. URL <http://link.springer.com/10.1007/s00122-018-3138-y>.
- C. Jung and B. Till. Mutagenesis and genome editing in crop improvement: perspectives for the global regulatory landscape. *Trends in Plant Science*, 26(12):1258–1269, Dec. 2021. ISSN 13601385. doi: 10.1016/j.tplants.2021.08.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S1360138521002132>.

- A. M. Kababji, H. Butt, and M. Mahfouz. Synthetic directed evolution for targeted engineering of plant traits. *Frontiers in Plant Science*, 15:1449579, Sept. 2024. ISSN 1664-462X. doi: 10.3389/fpls.2024.1449579. URL <https://www.frontiersin.org/articles/10.3389/fpls.2024.1449579/full>.
- B. G. Kang and S. P. Burg. Role of ethylene in phytochrome-induced anthocyanin synthesis. *Planta*, 110(3):227–235, 1973. Publisher: Springer.
- G. Kelly, E. Plessner, E. Bdolach, M. Arroyave, E. Belausov, A. Doron-Faigenboim, A. Rozen, H. Zemach, Y. Y. Zach, L. Goldenberg, T. Arad, Y. Yaniv, N. Sade, A. Sherman, Y. Eyal, and N. Carmi. In planta genome editing in citrus facilitated by co-expression of CRISPR/Cas and developmental regulators. *The Plant Journal*, 122(2):e70155, Apr. 2025. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.70155. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.70155>.
- D. Kim, S. J. Jeon, S. Yanders, S. Park, H. S. Kim, and S. Kim. MYB3 plays an important role in lignin and anthocyanin biosynthesis under salt stress condition in *Arabidopsis*. *Plant Cell Reports*, 41(7):1549–1560, July 2022. ISSN 0721-7714, 1432-203X. doi: 10.1007/s00299-022-02878-7. URL <https://link.springer.com/10.1007/s00299-022-02878-7>.
- J. A. King, J. Jeong, F. E. Underwood, J. Quan, N. Panaccione, J. W. Windsor, S. Coward, J. deBruyn, P. E. Ronksley, A.-A. Shaheen, H. Quan, J. Godley, S. Veldhuyzen Van Zanten, B. Lebwohl, S. C. Ng, J. F. Ludvigsson, and G. G. Kaplan. Incidence of Celiac Disease Is Increasing Over Time: A Systematic Review and Meta-analysis. *American Journal of Gastroenterology*, 115(4):507–525, Apr. 2020. ISSN 0002-9270, 1572-0241. doi: 10.14309/ajg.0000000000000523. URL <https://journals.lww.com/10.14309/ajg.0000000000000523>.
- L. T. Kinsman, N. J. Pinfield, and A. K. Stobart. The hormonal control of amaranthin synthesis in *Amaranthus caudatus* seedlings. *Planta*, 127(3):207–212, 1975. ISSN 0032-0935, 1432-2048. doi: 10.1007/BF00380717. URL <http://link.springer.com/10.1007/BF00380717>.
- Y. Kishima, A. Shimaya, and T. Adachi. Evidence that blue light induces betalain pigmentation in *Portulaca callus*. *Plant Cell, Tissue and Organ Culture*, 43(1):67–70, Oct. 1995. ISSN 0167-6857, 1573-5044. doi: 10.1007/BF00042673. URL <http://link.springer.com/10.1007/BF00042673>.
- V. K. Kochhar, S. Kochhar, and H. Mohr. Action of Light and Kinetin on Betalain Synthesis in Seedlings of *Amaranthus caudatus* : A Two-Factor Analysis. *Berichte der Deutschen Botanischen Gesellschaft*, 94(1):27–34, Oct. 1981. ISSN 0365-9631. doi: 10.1111/j.1438-8677.1981.tb03378.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1438-8677.1981.tb03378.x>.
- V. Konstantakos, A. Nentidis, A. Krithara, and G. Paliouras. CRISPR–Cas9 gRNA efficiency prediction: an overview of predictive tools and the role of deep learning. *Nucleic Acids Research*, 50(7):3616–3637, Apr. 2022. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkac192. URL <https://academic.oup.com/nar/article/50/7/3616/6555429>.
- K. Koyama, K. Sadamatsu, and N. Goto-Yamamoto. Absciscic acid stimulated ripen-

- ing and gene expression in berry skins of the Cabernet Sauvignon grape. *Functional & Integrative Genomics*, 10(3):367–381, Aug. 2010. ISSN 1438-793X, 1438-7948. doi: 10.1007/s10142-009-0145-8. URL <http://link.springer.com/10.1007/s10142-009-0145-8>.
- J. Kroymann. Natural diversity and adaptation in plant secondary metabolism. *Current Opinion in Plant Biology*, 14(3):246–251, June 2011. ISSN 13695266. doi: 10.1016/j.pbi.2011.03.021. URL <https://linkinghub.elsevier.com/retrieve/pii/S1369526611000331>.
- B. Kumar and P. Bhalothia. Orphan crops for future food security. *Journal of Biosciences*, 45(1):131, Dec. 2020. ISSN 0250-5991, 0973-7138. doi: 10.1007/s12038-020-00107-5. URL <https://link.springer.com/10.1007/s12038-020-00107-5>.
- P. Lakhotia, K. Pal Singh, S. Singh, M. Singh, K. Prasad, and K. Swaroop. Influence of biotic and abiotic elicitors on production of betalain pigments in bougainvillea callus cultures. *Indian Journal of Horticulture*, 71(3):373–378, Sept. 2014. URL https://www.researchgate.net/publication/287446547_Influence_of_biotic_and_abiotic_elicitors_on_production_of_betalain_pigments_in_bougainvillea_callus_cultures?_tp=eyJjb250ZXh0Ijp7ImZpcnNOUGFnZSI6Ii9kaXJlY3QiLCJwYWdlIjoiX2RpcmVjdCJ9fQ.
- M. Lei, C. Zhu, Y. Liu, A. S. Karthikeyan, R. A. Bressan, K. G. Raghothama, and D. Liu. Ethylene signalling is involved in regulation of phosphate starvation-induced gene expression and production of acid phosphatases and anthocyanin in *Arabidopsis*. *New Phytologist*, 189(4):1084–1095, Mar. 2011. ISSN 0028-646X, 1469-8137. doi: 10.1111/j.1469-8137.2010.03555.x. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.2010.03555.x>.
- Z. H. Lemmon, N. T. Reem, J. Dalrymple, S. Soyk, K. E. Swartwood, D. Rodriguez-Leal, J. Van Eck, and Z. B. Lippman. Rapid improvement of domestication traits in an orphan crop by genome editing. *Nature Plants*, 4(10):766–770, Oct. 2018. ISSN 2055-0278. doi: 10.1038/s41477-018-0259-x. URL <http://www.nature.com/articles/s41477-018-0259-x>.
- L. Lepiniec, I. Debeaujon, J.-M. Routaboul, A. Baudry, L. Pourcel, N. Nesi, and M. Caboche. Genetics and Biochemistry of Seed Flavonoids. *Annual Review of Plant Biology*, 57(1):405–430, June 2006. ISSN 1543-5008, 1545-2123. doi: 10.1146/annurev.arplant.57.032905.105252. URL <https://www.annualreviews.org/doi/10.1146/annurev.arplant.57.032905.105252>.
- M. Li, P. He, Z. Zhao, J. Liu, H. Liu, S. Ma, Y. Shen, and B. Li. Effect of temperature on betacyanins synthesis and the transcriptome of *Suaeda salsa*. *Frontiers in Plant Science*, 14:1203089, June 2023. ISSN 1664-462X. doi: 10.3389/fpls.2023.1203089. URL <https://www.frontiersin.org/articles/10.3389/fpls.2023.1203089/full>.
- S. Li. Transcriptional control of flavonoid biosynthesis: Fine-tuning of the MYB-bHLH-WD40 (MBW) complex. *Plant Signaling & Behavior*, 9(1):e27522, Jan. 2014. ISSN 1559-2324. doi: 10.4161/psb.27522. URL <http://www.tandfonline.com/doi/abs/10.4161/psb.27522>.

- T. Li, X. Yang, Y. Yu, X. Si, X. Zhai, H. Zhang, W. Dong, C. Gao, and C. Xu. Domestication of wild tomato is accelerated by genome editing. *Nature Biotechnology*, 36(12):1160–1163, Dec. 2018. ISSN 1087-0156, 1546-1696. doi: 10.1038/nbt.4273. URL <http://www.nature.com/articles/nbt.4273>.
- Y. Li, W. Van Den Ende, and F. Rolland. Sucrose Induction of Anthocyanin Biosynthesis Is Mediated by DELLA. *Molecular Plant*, 7(3):570–572, Mar. 2014. ISSN 16742052. doi: 10.1093/mp/sst161. URL <https://linkinghub.elsevier.com/retrieve/pii/S1674205214602574>.
- Z. Lian, C. D. Nguyen, L. Liu, G. Wang, J. Chen, S. Wang, G. Yi, S. Wilson, P. Ozias-Akins, H. Gong, and H. Huo. Application of developmental regulators to improve *in planta* or *in vitro* transformation in plants. *Plant Biotechnology Journal*, 20(8):1622–1635, Aug. 2022. ISSN 1467-7644, 1467-7652. doi: 10.1111/pbi.13837. URL <https://onlinelibrary.wiley.com/doi/10.1111/pbi.13837>.
- K. Lin-Wang, D. Micheletti, J. Palmer, R. Volz, L. Lozano, R. Espley, R. P. Hellens, D. Chagn , D. D. Rowan, M. Troggio, I. Iglesias, and A. C. Allan. High temperature reduces apple fruit colour via modulation of the anthocyanin regulatory complex. *Plant, Cell & Environment*, 34(7):1176–1190, July 2011. ISSN 0140-7791, 1365-3040. doi: 10.1111/j.1365-3040.2011.02316.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-3040.2011.02316.x>.
- S. Liu, S. Sretenovic, T. Fan, Y. Cheng, G. Li, A. Qi, X. Tang, Y. Xu, W. Guo, Z. Zhong, Y. He, Y. Liang, Q. Han, X. Zheng, X. Gu, Y. Qi, and Y. Zhang. Hypercompact CRISPR–Cas12j2 (CasΦ) enables genome editing, gene activation, and epigenome editing in plants. *Plant Communications*, 3(6):100453, Nov. 2022a. ISSN 25903462. doi: 10.1016/j.xplc.2022.100453. URL <https://linkinghub.elsevier.com/retrieve/pii/S2590346222002905>.
- S. Liu, Y. Xu, K. Yang, Y. Huang, Z. Lu, S. Chen, X. Gao, G. Xiao, P. Chen, X. Zeng, L. Wang, W. Zheng, Z. Liu, G. Liao, F. He, J. Liu, P. Wan, F. Ding, J. Ye, W. Jiao, L. Chai, Z. Pan, F. Zhang, Z. Lin, Y. Zan, W. Guo, R. M. Larkin, Z. Xie, X. Wang, X. Deng, and Q. Xu. Origin and *de novo* domestication of sweet orange. *Nature Genetics*, 57(3):754–762, Mar. 2025. ISSN 1061-4036, 1546-1718. doi: 10.1038/s41588-025-02122-4. URL <https://www.nature.com/articles/s41588-025-02122-4>.
- Y. Liu, M. Li, T. Li, Y. Chen, L. Zhang, G. Zhao, J. Zhuang, W. Zhao, L. Gao, and T. Xia. Airborne fungus-induced biosynthesis of anthocyanins in *Arabidopsis thaliana* via jasmonic acid and salicylic acid signaling. *Plant Science*, 300:110635, Nov. 2020. ISSN 01689452. doi: 10.1016/j.plantsci.2020.110635. URL <https://linkinghub.elsevier.com/retrieve/pii/S0168945220302417>.
- Y. Liu, R. E. Schouten, Y. Tikunov, X. Liu, R. G. Visser, F. Tan, A. Bovy, and L. F. Marcelis. Blue light increases anthocyanin content and delays fruit ripening in purple pepper fruit. *Postharvest Biology and Technology*, 192:112024, Oct. 2022b. ISSN 09255214. doi: 10.1016/j.postharvbio.2022.112024. URL <https://linkinghub.elsevier.com/retrieve/pii/S0925521422001922>.
- Z. Liu, Y. Zhang, J. Wang, P. Li, C. Zhao, Y. Chen, and Y. Bi. Phytochrome-interacting

- factors PIF4 and PIF5 negatively regulate anthocyanin biosynthesis under red light in *Arabidopsis* seedlings. *Plant Science*, 238:64–72, Sept. 2015. ISSN 01689452. doi: 10.1016/j.plantsci.2015.06.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S0168945215001673>.
- A. Lloyd, A. Brockman, L. Aguirre, A. Campbell, A. Bean, A. Cantero, and A. Gonzalez. 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(9):1431–1441, Sept. 2017. ISSN 0032-0781, 1471-9053. doi: 10.1093/pcp/pcx075. URL <https://academic.oup.com/pcp/article/58/9/1431/3858365>.
- E. Loreti, G. Povero, G. Novi, C. Solfanelli, A. Alpi, and P. Perata. Gibberellins, jasmonate and abscisic acid modulate the sucrose-induced expression of anthocyanin biosynthetic genes in *Arabidopsis*. *New Phytologist*, 179(4):1004–1016, Sept. 2008. ISSN 0028-646X, 1469-8137. doi: 10.1111/j.1469-8137.2008.02511.x. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.2008.02511.x>.
- M. F. Maher, R. A. Nasti, M. Vollbrecht, C. G. Starker, M. D. Clark, and D. F. Voytas. Plant gene editing through *de novo* induction of meristems. *Nature Biotechnology*, 38(1):84–89, Jan. 2020. ISSN 1087-0156, 1546-1696. doi: 10.1038/s41587-019-0337-2. URL <http://www.nature.com/articles/s41587-019-0337-2>.
- A. Maier and U. Hoecker. COP1/SPA ubiquitin ligase complexes repress anthocyanin accumulation under low light and high light conditions. *Plant Signaling & Behavior*, 10(1):e970440, Jan. 2015. ISSN 1559-2324. doi: 10.4161/15592316.2014.970440. URL <http://www.tandfonline.com/doi/full/10.4161/15592316.2014.970440>.
- N. A. Maren, H. Duan, K. Da, G. C. Yench, T. G. Ranney, and W. Liu. Genotype-independent plant transformation. *Horticulture Research*, 9:uhac047, Jan. 2022. ISSN 2052-7276. doi: 10.1093/hr/uhac047. URL <https://academic.oup.com/hr/article/doi/10.1093/hr/uhac047/6548281>.
- R. A. Marks, S. Hotelling, P. B. Frandsen, and R. VanBuren. Representation and participation across 20 years of plant genome sequencing. *Nature Plants*, 7(12):1571–1578, Nov. 2021. ISSN 2055-0278. doi: 10.1038/s41477-021-01031-8. URL <https://www.nature.com/articles/s41477-021-01031-8>.
- J. Masel. Genetic drift. *Current Biology*, 21(20):R837–R838, Oct. 2011. ISSN 09609822. doi: 10.1016/j.cub.2011.08.007. URL <https://linkinghub.elsevier.com/retrieve/pii/S0960982211008827>.
- C. Milech, S. R. Lucho, A. M. Kleinowski, D. B. Dutra, M. M. Soares, and E. J. B. Braga. Production of pigments in *Alternanthera sessilis* calli mediated by plant growth regulators and light. *Acta Scientiarum. Biological Sciences*, 39(3):381–388, 2017.
- K. Mueangnak, H. Kitwetcharoen, S. Thanonkeo, P. Klanrit, J. Apiraksakorn, P. Klanrit, P. Klanrit, and P. Thanonkeo. Enhancing betalains production and antioxidant activity in *Celosia argentea* cell suspension cultures using biotic and abiotic elicitors. *Scientific Reports*, 15(1):376, Jan. 2025. ISSN 2045-2322. doi: 10.1038/s41598-024-83096-x. URL <https://www.nature.com/articles/s41598-024-83096-x>.

- B. Nabholz, G. Sarah, F. Sabot, M. Ruiz, H. Adam, S. Nidelet, A. Ghesquière, S. Santoni, J. David, and S. Glémin. Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice *Oryza glaberrima*). *Molecular Ecology*, 23(9):2210–2227, May 2014. ISSN 0962-1083, 1365-294X. doi: 10.1111/mec.12738. URL <https://onlinelibrary.wiley.com/doi/10.1111/mec.12738>.
- K. Nelson-Vasilchik, J. Hague, M. Mookkan, Z. J. Zhang, and A. Kausch. Transformation of Recalcitrant Sorghum Varieties Facilitated by *Baby Boom* and *Wuschel2*. *Current Protocols in Plant Biology*, 3(4):e20076, Dec. 2018. ISSN 2379-8068, 2379-8068. doi: 10.1002/cppb.20076. URL <https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/cppb.20076>.
- A. A. Nemudryi, K. R. Valetdinova, S. P. Medvedev, and S. M. Zakian. TALEN and CRISPR/Cas Genome Editing Systems: Tools of Discovery. *Acta Naturae*, 6(3):19–40, July 2014. ISSN 2075-8251.
- S. Neugart, S. Baldermann, F. S. Hanschen, R. Klopsch, M. Wiesner-Reinhold, and M. Schreiner. The intrinsic quality of brassicaceous vegetables: How secondary plant metabolites are affected by genetic, environmental, and agronomic factors. *Scientia Horticulturae*, 233:460–478, Mar. 2018. ISSN 03044238. doi: 10.1016/j.scienta.2017.12.038. URL <https://linkinghub.elsevier.com/retrieve/pii/S0304423817307616>.
- C. Oi, Z. Gidden, L. Holyoake, O. Kantelberg, S. Mochrie, M. H. Horrocks, and L. Regan. LIVE-PAINT allows super-resolution microscopy inside living cells using reversible peptide-protein interactions. *Communications Biology*, 3(1):458, Dec. 2020. ISSN 2399-3642. doi: 10.1038/s42003-020-01188-6. URL <https://www.nature.com/articles/s42003-020-01188-6>.
- M. Ozden. Secondary metabolite production in callus cultures of *Vitis vinifera*: influence of genotype and sucrose concentration in the medium on antioxidant activity. *Acta Physiologiae Plantarum*, 46(1):6, Jan. 2024. ISSN 0137-5881, 1861-1664. doi: 10.1007/s11738-023-03630-8. URL <https://link.springer.com/10.1007/s11738-023-03630-8>.
- C. Pan and Y. Qi. CRISPR-Combo-mediated orthogonal genome editing and transcriptional activation for plant breeding. *Nature Protocols*, 18(6):1760–1794, June 2023. ISSN 1754-2189, 1750-2799. doi: 10.1038/s41596-023-00823-w. URL <https://www.nature.com/articles/s41596-023-00823-w>.
- A. Pandey, P. Misra, K. Chandrashekar, and P. K. Trivedi. Development of AtMYB12-expressing transgenic tobacco callus culture for production of rutin with biopesticidal potential. *Plant Cell Reports*, 31(10):1867–1876, Oct. 2012. ISSN 0721-7714, 1432-203X. doi: 10.1007/s00299-012-1300-6. URL <http://link.springer.com/10.1007/s00299-012-1300-6>.
- M. Pari, W. Q. Lee, C. K. F. Wong, and C. Y. Teh. Induction of callus culture through plant growth regulators supplementation and the effect of elicitors on enhancement of betalain synthesis using *Gomphrena globosa*. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 156(1):19, Jan. 2024. ISSN 0167-6857, 1573-5044. doi: 10.1007/s11240-023-02628-x. URL <https://link.springer.com/10.1007/s11240-023-02628-x>.

- A. Pirie and M. G. Mullins. Changes in Anthocyanin and Phenolics Content of Grapevine Leaf and Fruit Tissues Treated with Sucrose, Nitrate, and Absciscic Acid. *Plant Physiology*, 58(4):468–472, Oct. 1976. ISSN 0032-0889, 1532-2548. doi: 10.1104/pp.58.4.468. URL <https://academic.oup.com/plphys/article/58/4/468-472/6075204>.
- G. Polturak, D. Breitel, N. Grossman, A. Sarrion-Perdigones, E. Weithorn, M. Pliner, D. Orzaez, A. Granell, I. Rogachev, and A. Aharoni. Elucidation of the first committed step in betalain biosynthesis enables the heterologous engineering of betalain pigments in plants. *New Phytologist*, 210(1):269–283, Apr. 2016. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.13796. URL <https://onlinelibrary.wiley.com/doi/10.1111/nph.13796>.
- C. Portillo-Nava, M. Guerrero-Esperanza, A. Guerrero-Rangel, P. Guevara-Domínguez, N. Martínez-Gallardo, C. Nava-Sandoval, J. Ordaz-Ortiz, L. Sánchez-Segura, and J. Délano-Frier. Natural or light-induced pigment accumulation in grain amaranths coincides with enhanced resistance against insect herbivory. *Planta*, 254(5):101, Nov. 2021. ISSN 0032-0935, 1432-2048. doi: 10.1007/s00425-021-03757-3. URL <https://link.springer.com/10.1007/s00425-021-03757-3>.
- D. Pramanik, K. Lee, and K. Wang. A simple and efficient method for betalain quantification in RUBY-expressing plant samples. *Frontiers in Plant Science*, 15:1449409, Sept. 2024. ISSN 1664-462X. doi: 10.3389/fpls.2024.1449409. URL <https://www.frontiersin.org/articles/10.3389/fpls.2024.1449409/full>.
- B. Pucker, N. Walker-Hale, J. Dzurlic, W. C. Yim, J. C. Cushman, A. Crum, Y. Yang, and S. F. Brockington. Multiple mechanisms explain loss of anthocyanins from betalain-pigmented *Caryophyllales*, including repeated wholesale loss of a key anthocyanidin synthesis enzyme. *New Phytologist*, 241(1):471–489, Jan. 2024. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.19341. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19341>.
- M. D. Purugganan and D. Q. Fuller. The nature of selection during plant domestication. *Nature*, 457(7231):843–848, Feb. 2009. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature07895. URL <https://www.nature.com/articles/nature07895>.
- Y. Qi. Boosting Plant Regeneration via CRISPR Activation-based Reprogramming. In *IN VITRO CELLULAR & DEVELOPMENTAL BIOLOGY-ANIMAL*, volume 61. SPRINGER ONE NEW YORK PLAZA, SUITE 4600, NEW YORK, NY, UNITED STATES, 2025.
- J. Qu, W. Zhang, and X. Yu. A combination of elicitation and precursor feeding leads to increased anthocyanin synthesis in cell suspension cultures of *Vitis vinifera*. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 107(2):261–269, Nov. 2011. ISSN 0167-6857, 1573-5044. doi: 10.1007/s11240-011-9977-8. URL <http://link.springer.com/10.1007/s11240-011-9977-8>.
- M. Ram, K. V. Prasad, S. K. Singh, B. S. Hada, and S. Kumar. Influence of salicylic acid and methyl jasmonate elicitation on anthocyanin production in callus cultures of *Rosa hybrida* L. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 113(3):459–467, June 2013. ISSN 0167-6857, 1573-5044. doi: 10.1007/s11240-013-0287-1. URL <http://link.springer.com/10.1007/s11240-013-0287-1>.

- G. S. Rao, W. Jiang, and M. Mahfouz. Synthetic directed evolution in plants: unlocking trait engineering and improvement. *Synthetic Biology*, 6(1):ysab025, Sept. 2021. ISSN 2397-7000. doi: 10.1093/synbio/ysab025. URL <https://academic.oup.com/synbio/article/doi/10.1093/synbio/ysab025/6366326>.
- K. M. Reed and B. O. R. Bargmann. Protoplast Regeneration and Its Use in New Plant Breeding Technologies. *Frontiers in Genome Editing*, 3:734951, Sept. 2021. ISSN 2673-3439. doi: 10.3389/fgeed.2021.734951. URL <https://www.frontiersin.org/articles/10.3389/fgeed.2021.734951/full>.
- A. Reis, A. M. Kleinowski, F. R. S. Klein, R. T. T. D. Souza, L. D. Amarante, and E. J. B. Braga. Callus induction and betacyanin quantification by HPLC/MS-MS in *Alternanthera brasiliana* (L.) Kuntze. *Hoehnea*, 44(1):90–95, Mar. 2017. ISSN 2236-8906. doi: 10.1590/2236-8906-75/2016. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2236-89062017000100090&lng=en&tlng=en.
- A. Reis, A. M. Kleinowski, R. T. Telles, F. R. S. Klein, L. do Amarante, and E. J. B. Braga. Light quality and plant growth regulators influence pigment production in *Alternanthera brasiliana* calli. *African Journal of Biotechnology*, 17(20):638–648, 2018. Publisher: Academic Journals.
- S. A. Rensing. Why we need more non-seed plant models. *New Phytologist*, 216(2): 355–360, Oct. 2017. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.14464. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.14464>.
- A. Reyes-Rosales, G. Cabrales-Orona, N. A. Martínez-Gallardo, L. Sánchez-Segura, J. P. Padilla-Escamilla, P. A. Palmeros-Suárez, and J. P. Délano-Frier. Identification of genetic and biochemical mechanisms associated with heat shock and heat stress adaptation in grain amaranths. *Frontiers in Plant Science*, 14:1101375, Feb. 2023. ISSN 1664-462X. doi: 10.3389/fpls.2023.1101375. URL <https://www.frontiersin.org/articles/10.3389/fpls.2023.1101375/full>.
- C. W. Riggins and R. H. Mumm. Amaranths. *Current Biology*, 31(13):R834–R835, July 2021. ISSN 09609822. doi: 10.1016/j.cub.2021.05.058. URL <https://linkinghub.elsevier.com/retrieve/pii/S0960982221007569>.
- I. Rodrigues-Brandão, A. M. Kleinowski, A. M. Einhardt, M. C. Lima, L. D. Amarante, J. A. Peters, and E. J. B. Braga. Salicylic acid on antioxidant activity and betacyanin production from leaves of *Alternanthera tenella*. *Ciência Rural*, 44(10):1893–1898, Oct. 2014. ISSN 0103-8478. doi: 10.1590/0103-8478cr20130873. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782014001001893&lng=en&tlng=en.
- A. H. K. Roeder, A. Bent, J. T. Lovell, J. K. McKay, A. Bravo, K. Medina-Jimenez, K. W. Morimoto, S. M. Brady, L. Hua, J. M. Hibberd, S. Zhong, F. Cardinale, I. Visentin, C. Lovisolo, M. A. Hannah, and A. A. R. Webb. Lost in translation: What we have learned from attributes that do not translate from *Arabidopsis* to other plants. *The Plant Cell*, 37(5):koaf036, May 2025. ISSN 1040-4651, 1532-298X. doi: 10.1093/plcell/koaf036. URL <https://academic.oup.com/plcell/article/doi/10.1093/plcell/koaf036/8132777>.

- U. Rogo, S. Simoni, M. Fambrini, T. Giordani, C. Pugliesi, and F. Mascagni. Future-Proofing Agriculture: *De Novo* Domestication for Sustainable and Resilient Crops. *International Journal of Molecular Sciences*, 25(4):2374, Feb. 2024. ISSN 1422-0067. doi: 10.3390/ijms25042374. URL <https://www.mdpi.com/1422-0067/25/4/2374>.
- A. Rudat and H. Göring. Induction of betacyanin formation in cell cultures of *Chenopodium album* under UV-light irradiation. *Journal of Experimental Botany*, 46(1):129–134, 1995. ISSN 0022-0957, 1460-2431. doi: 10.1093/jxb/46.1.129. URL <https://academic.oup.com/jxb/article-lookup/doi/10.1093/jxb/46.1.129>.
- S. Sahab, M. J. Hayden, J. Mason, and G. Spangenberg. Mesophyll Protoplasts and PEG-Mediated Transfections: Transient Assays and Generation of Stable Transgenic Canola Plants. In S. Kumar, P. Barone, and M. Smith, editors, *Transgenic Plants*, volume 1864, pages 131–152. Springer New York, New York, NY, 2019. ISBN 978-1-4939-8777-1 978-1-4939-8778-8. doi: 10.1007/978-1-4939-8778-8_10. URL http://link.springer.com/10.1007/978-1-4939-8778-8_10. Series Title: Methods in Molecular Biology.
- M. Sakuta, H. Hirano, and A. Komamine. Stimulation by 2,4-dichlorophenoxyacetic acid of betacyanin accumulation in suspension cultures of *Phytolacca americana*. *Physiologia Plantarum*, 83(1):154–158, Nov. 1991. ISSN 0031-9317, 1399-3054. doi: 10.1111/j.1399-3054.1991.tb01295.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1399-3054.1991.tb01295.x>.
- T. Sang A Roon, P. Klanrit, P. Klanrit, P. Thanonkeo, J. Apiraksakorn, S. Thanonkeo, and P. Klanrit. Establishment of Betalain-Producing Cell Line and Optimization of Pigment Production in Cell Suspension Cultures of *Celosia argentea* var. *plumosa*. *Plants*, 13(22):3225, Nov. 2024. ISSN 2223-7747. doi: 10.3390/plants13223225. URL <https://www.mdpi.com/2223-7747/13/22/3225>.
- N. Sasaki, T. Adachi, T. Koda, and Y. Ozeki. Detection of UDP-glucose: *cyclo* -DOPA 5- *O* -glucosyltransferase activity in four o’clocks (*Mirabilis jalapa* L.). *FEBS Letters*, 568(1-3):159–162, June 2004. ISSN 0014-5793, 1873-3468. doi: 10.1016/j.febslet.2004.04.097. URL <https://febs.onlinelibrary.wiley.com/doi/10.1016/j.febslet.2004.04.097>.
- N. Sasaki, K. Wada, T. Koda, K. Kasahara, T. Adachi, and Y. Ozeki. Isolation and Characterization of cDNAs Encoding an Enzyme with Glucosyltransferase Activity for *cyclo*-DOPA from Four O’clocks and Feather Cockscombs. *Plant and Cell Physiology*, 46(4):666–670, Apr. 2005. ISSN 1471-9053, 0032-0781. doi: 10.1093/pcp/pci064. URL <http://academic.oup.com/pcp/article/46/4/666/1872998/Isolation-and-Characterization-of-cDNAs-Encoding>.
- N. Sasaki, Y. Abe, Y. Goda, T. Adachi, K. Kasahara, and Y. Ozeki. Detection of DOPA 4,5-Dioxygenase (DOD) Activity Using Recombinant Protein Prepared from *Escherichia coli* Cells Harboring cDNA Encoding DOD from *Mirabilis jalapa*. *Plant and Cell Physiology*, 50(5):1012–1016, May 2009. ISSN 0032-0781, 1471-9053. doi: 10.1093/pcp/pcp053. URL <https://academic.oup.com/pcp/article-lookup/doi/10.1093/pcp/pcp053>.
- J. D. Sauer. The Grain Amaranths and Their Relatives: A Revised Taxonomic and

- Geographic Survey. *Annals of the Missouri Botanical Garden*, 54(2):103, 1967. ISSN 00266493. doi: 10.2307/2394998. URL <https://www.jstor.org/stable/2394998?origin=crossref>.
- B. Savitha, R. Thimmaraju, N. Bhagyalakshmi, and G. Ravishankar. Different biotic and abiotic elicitors influence betalain production in hairy root cultures of *Beta vulgaris* in shake-flask and bioreactor. *Process Biochemistry*, 41(1):50–60, Jan. 2006. ISSN 13595113. doi: 10.1016/j.procbio.2005.03.071. URL <https://linkinghub.elsevier.com/retrieve/pii/S1359511305002448>.
- A. Scheben, F. Wolter, J. Batley, H. Puchta, and D. Edwards. Towards CRISPR/Cas crops – bringing together genomics and genome editing. *New Phytologist*, 216(3): 682–698, Nov. 2017. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.14702. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.14702>.
- W. Schliemann, N. Kobayashi, and D. Strack. The decisive step in betaxanthin biosynthesis is a spontaneous reaction. *Plant Physiology*, 119(4):1217–1232, Apr. 1999. ISSN 1532-2548. doi: 10.1104/pp.119.4.1217.
- G. Sepúlveda-Jiménez, P. Rueda-Benítez, H. Porta, and M. Rocha-Sosa. Betacyanin synthesis in red beet (*Beta vulgaris*) leaves induced by wounding and bacterial infiltration is preceded by an oxidative burst. *Physiological and Molecular Plant Pathology*, 64(3):125–133, Mar. 2004. ISSN 08855765. doi: 10.1016/j.pmpp.2004.08.003. URL <https://linkinghub.elsevier.com/retrieve/pii/S088557650400089X>.
- Y. Shamshirgaran, J. Liu, H. Sumer, P. J. Verma, and A. Taheri-Ghahfarokhi. Tools for Efficient Genome Editing; ZFN, TALEN, and CRISPR. In P. J. Verma, H. Sumer, and J. Liu, editors, *Applications of Genome Modulation and Editing*, volume 2495, pages 29–46. Springer US, New York, NY, 2022. ISBN 978-1-0716-2300-8 978-1-0716-2301-5. doi: 10.1007/978-1-0716-2301-5_2. URL https://link.springer.com/10.1007/978-1-0716-2301-5_2. Series Title: Methods in Molecular Biology.
- H. Sheehan, T. Feng, N. Walker-Hale, S. Lopez-Nieves, B. Pucker, R. Guo, W. C. Yim, R. Badgami, A. Timoneda, L. Zhao, H. Tiley, D. Copetti, M. J. Sanderson, J. C. Cushman, M. J. Moore, S. A. Smith, and S. F. Brockington. Evolution of l-DOPA4,5-dioxygenase activity allows for recurrent specialisation to betalain pigmentation in *Caryophyllales*. *New Phytologist*, 227(3):914–929, Aug. 2020. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.16089. URL <https://onlinelibrary.wiley.com/doi/10.1111/nph.16089>.
- L. Shi, S. Cao, W. Chen, and Z. Yang. Blue light induced anthocyanin accumulation and expression of associated genes in Chinese bayberry fruit. *Scientia Horticulturae*, 179:98–102, Nov. 2014. ISSN 03044238. doi: 10.1016/j.scienta.2014.09.022. URL <https://linkinghub.elsevier.com/retrieve/pii/S0304423814005007>.
- K. Shin, H. Murthy, J. Heo, and K. Paek. Induction of Betalain Pigmentation in Hairy Roots of Red Beet under Different Radiation Sources. *Biologia plantarum*, 46(1): 149–152, July 2003. ISSN 15738264. doi: 10.1023/A:1027313805930. URL <http://bp.ueb.cas.cz/doi/10.1023/A:1027313805930.html>.
- O. Y. Shoeva, H.-P. Mock, T. V. Kukoeva, A. Börner, and E. K. Khlestkina. Regulation

- of the Flavonoid Biosynthesis Pathway Genes in Purple and Black Grains of *Hordeum vulgare*. *PLOS ONE*, 11(10):e0163782, Oct. 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0163782. URL <https://dx.plos.org/10.1371/journal.pone.0163782>.
- T. Shoji. The Recruitment Model of Metabolic Evolution: Jasmonate-Responsive Transcription Factors and a Conceptual Model for the Evolution of Metabolic Pathways. *Frontiers in Plant Science*, 10:560, May 2019. ISSN 1664-462X. doi: 10.3389/fpls.2019.00560. URL <https://www.frontiersin.org/article/10.3389/fpls.2019.00560/full>.
- S. Shpileuski, I. Domanskaya, V. Loucsha, and L. Kabashnikova. B-1, 3-Glucan Effect on the Defence Reaction of Tomato Leaves under *Fusarium* Wilt. *Journal of the Belarusian State University Ecology*, 3:50, 2024.
- K. H. M. Siddique, X. Li, and K. Gruber. Rediscovering Asia’s forgotten crops to fight chronic and hidden hunger. *Nature Plants*, 7(2):116–122, Feb. 2021. ISSN 2055-0278. doi: 10.1038/s41477-021-00850-z. URL <http://www.nature.com/articles/s41477-021-00850-z>.
- A. Singh and M. G. Stetter. Going New Places: Successful Adaptation and Genomic Integrity of Grain Amaranth in India. *Evolutionary Applications*, 18(7):e70124, July 2025. ISSN 1752-4571, 1752-4571. doi: 10.1111/eva.70124. URL <https://onlinelibrary.wiley.com/doi/10.1111/eva.70124>.
- R.-F. Song, X.-Y. Hu, W.-C. Liu, and H.-M. Yuan. ABA functions in low phosphate-induced anthocyanin accumulation through the transcription factor ABI5 in *Arabidopsis*. *Plant Cell Reports*, 43(2):55, Feb. 2024. ISSN 0721-7714, 1432-203X. doi: 10.1007/s00299-024-03146-6. URL <https://link.springer.com/10.1007/s00299-024-03146-6>.
- G. F. Stallknecht and J. Schulz-Schaeffer. Amaranth rediscovered. *New crops*. Wiley, New York, pages 211–218, 1993. ISSN 0-471-59374-5.
- M. G. Stetter. Limits and constraints to crop domestication. *American Journal of Botany*, 107(12):1617–1621, Dec. 2020. ISSN 0002-9122, 1537-2197. doi: 10.1002/ajb2.1585. URL <https://onlinelibrary.wiley.com/doi/10.1002/ajb2.1585>.
- M. G. Stetter, M. Vidal-Villarejo, and K. J. Schmid. Parallel Seed Color Adaptation during Multiple Domestication Attempts of an Ancient New World Grain. *Molecular Biology and Evolution*, 37(5):1407–1419, May 2020. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msz304. URL <https://academic.oup.com/mbe/article/37/5/1407/5682418>.
- M. G. Stetter, D. C. Joshi, and A. Singh. Assessing and mining grain amaranth diversity for sustainable cropping systems. *Theoretical and Applied Genetics*, 138(7):171, July 2025. ISSN 0040-5752, 1432-2242. doi: 10.1007/s00122-025-04940-w. URL <https://link.springer.com/10.1007/s00122-025-04940-w>.
- D. Strack, T. Vogt, and W. Schliemann. Recent advances in betalain research. *Phytochemistry*, 62(3):247–269, Feb. 2003. ISSN 00319422. doi: 10.1016/S0031-9422(02)00564-2. URL <https://linkinghub.elsevier.com/retrieve/pii/S0031942202005642>.

- R. Stracke, M. Werber, and B. Weisshaar. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology*, 4(5):447–456, Oct. 2001. ISSN 13695266. doi: 10.1016/S1369-5266(00)00199-0. URL <https://linkinghub.elsevier.com/retrieve/pii/S1369526600001990>.
- C. Sun, H. Li, Y. Liu, Y. Li, R. Gao, X. Shi, H. Fei, J. Liu, R. Liang, and C. Gao. Iterative recombinase technologies for efficient and precise genome engineering across kilobase to megabase scales. *Cell*, 188(17):4693–4710.e15, Aug. 2025. ISSN 00928674. doi: 10.1016/j.cell.2025.07.011. URL <https://linkinghub.elsevier.com/retrieve/pii/S0092867425008001>.
- R. Sunnadeniya, A. Bean, M. Brown, N. Akhavan, G. Hatlestad, A. Gonzalez, V. V. Symonds, and A. Lloyd. Tyrosine Hydroxylation in Betalain Pigment Biosynthesis Is Performed by Cytochrome P450 Enzymes in Beets (*Beta vulgaris*). *PLOS ONE*, 11(2):e0149417, Feb. 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0149417. URL <https://dx.plos.org/10.1371/journal.pone.0149417>.
- B. Suresh, R. Thimmaraju, N. Bhagyalakshmi, and G. Ravishankar. Polyamine and methyl jasmonate-influenced enhancement of betalaine production in hairy root cultures of *Beta vulgaris* grown in a bubble column reactor and studies on efflux of pigments. *Process Biochemistry*, 39(12):2091–2096, Oct. 2004. ISSN 13595113. doi: 10.1016/j.procbio.2003.10.009. URL <https://linkinghub.elsevier.com/retrieve/pii/S0032959203004072>.
- G. Suvarnalatha, L. Rajendran, G. A. Ravishankar, and L. V. Venkataraman. Elicitation of anthocyanin production in cell cultures of carrot (*Daucus carota* L) by using elicitors and abiotic stress. *Biotechnology Letters*, 16(12):1275–1280, Dec. 1994. ISSN 0141-5492, 1573-6776. doi: 10.1007/BF00149631. URL <https://link.springer.com/10.1007/BF00149631>.
- Z. Tadele. African Orphan Crops under Abiotic Stresses: Challenges and Opportunities. *Scientifica*, 2018:1–19, Jan. 2018. ISSN 2090-908X. doi: 10.1155/2018/1451894. URL <https://www.hindawi.com/journals/scientifica/2018/1451894/>.
- Z. Tadele. Orphan crops: their importance and the urgency of improvement. *Planta*, 250(3):677–694, Sept. 2019. ISSN 0032-0935, 1432-2048. doi: 10.1007/s00425-019-03210-6. URL <http://link.springer.com/10.1007/s00425-019-03210-6>.
- M. Takou, K. Schulz, and M. G. Stetter. Local Selection Shaped the Diversity of European Maize Landraces. *Molecular Ecology*, page e17720, Mar. 2025. ISSN 0962-1083, 1365-294X. doi: 10.1111/mec.17720. URL <https://onlinelibrary.wiley.com/doi/10.1111/mec.17720>.
- M. Taya, K. Yakura, M. Kino-oka, and S. Tone. Influence of medium constituents on enhancement of pigment production by batch culture of red beet hairy roots. *Journal of Fermentation and Bioengineering*, 77(2):215–217, Jan. 1994. ISSN 0922338X. doi: 10.1016/0922-338X(94)90329-8. URL <https://linkinghub.elsevier.com/retrieve/pii/0922338X94903298>.
- A. Timoneda, T. Feng, H. Sheehan, N. Walker-Hale, B. Pucker, S. Lopez-Nieves, R. Guo, and S. Brockington. The evolution of betalain biosynthesis in *Caryophyllales*. *New*

- Phytologist*, 224(1):71–85, Oct. 2019. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.15980. URL <https://onlinelibrary.wiley.com/doi/10.1111/nph.15980>.
- T. Tohge, M. Watanabe, R. Hoefgen, and A. R. Fernie. Shikimate and Phenylalanine Biosynthesis in the Green Lineage. *Frontiers in Plant Science*, 4, 2013. ISSN 1664-462X. doi: 10.3389/fpls.2013.00062. URL <http://journal.frontiersin.org/article/10.3389/fpls.2013.00062/abstract>.
- G. Trejo-Tapia, A. Jimenez-Aparicio, M. Rodriguez, G. Sepulveda, G. Salcedo, B. Martinez, G. Gutierrez, and A. De Jesus. Influence of medium constituents on growth and betalain production in cell suspension cultures of *Beta vulgaris*. *Asia-Pacific Journal of Molecular Biology and Biotechnology*, 7(2):167–172, 1999. URL https://hero.epa.gov/hero/index.cfm/reference/details/reference_id/4485590.
- A. Tuncel, C. Pan, J. S. Clem, D. Liu, and Y. Qi. CRISPR–Cas applications in agriculture and plant research. *Nature Reviews Molecular Cell Biology*, 26(6):419–441, June 2025. ISSN 1471-0072, 1471-0080. doi: 10.1038/s41580-025-00834-3. URL <https://www.nature.com/articles/s41580-025-00834-3>.
- S. D. Turner-Hissong, M. E. Mabry, T. M. Beissinger, J. Ross-Ibarra, and J. C. Pires. Evolutionary insights into plant breeding. *Current Opinion in Plant Biology*, 54:93–100, Apr. 2020. ISSN 13695266. doi: 10.1016/j.pbi.2020.03.003. URL <https://linkinghub.elsevier.com/retrieve/pii/S1369526620300339>.
- A. P. Uniyal, K. Mansotra, S. K. Yadav, and V. Kumar. An overview of designing and selection of sgRNAs for precise genome editing by the CRISPR-Cas9 system in plants. *3 Biotech*, 9(6):223, June 2019. ISSN 2190-572X, 2190-5738. doi: 10.1007/s13205-019-1760-2. URL <http://link.springer.com/10.1007/s13205-019-1760-2>.
- W. Vandeputte, G. Coussens, S. Aesaert, J. Haeghebaert, L. Impens, M. Karimi, J. M. Debernardi, and L. Pauwels. Use of GRF-GIF chimeras and a ternary vector system to improve maize (*Zea mays* L.) transformation frequency. *The Plant Journal*, 119(4):2116–2132, Aug. 2024. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.16880. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.16880>.
- M. Venezia and K. M. Creasey Krainer. Current Advancements and Limitations of Gene Editing in Orphan Crops. *Frontiers in Plant Science*, 12:742932, Sept. 2021. ISSN 1664-462X. doi: 10.3389/fpls.2021.742932. URL <https://www.frontiersin.org/articles/10.3389/fpls.2021.742932/full>.
- T. Vogt. Substrate specificity and sequence analysis define a polyphyletic origin of betanidin 5- and 6-O-glucosyltransferase from *Dortheanthus bellidiformis*. *Planta*, 214(3):492–495, Jan. 2002. ISSN 0032-0935, 1432-2048. doi: 10.1007/s00425-001-0685-1. URL <http://link.springer.com/10.1007/s00425-001-0685-1>.
- T. Vogt, R. Grimm, and D. Strack. Cloning and expression of a cDNA encoding betanidin 5-O-glucosyltransferase, a betanidin- and flavonoid-specific enzyme with high homology to inducible glucosyltransferases from the *Solanaceae*. *The Plant Journal*, 19(5):509–519, Sept. 1999a. ISSN 0960-7412, 1365-313X. doi: 10.1046/j.1365-313X.1999.00540.x. URL <https://onlinelibrary.wiley.com/doi/10.1046/j.1365-313X.1999.00540.x>.

- T. Vogt, M. Ibdah, J. Schmidt, V. Wray, M. Nimtz, and D. Strack. Light-induced betacyanin and flavonol accumulation in bladder cells of *Mesembryanthemum crystallinum*. *Phytochemistry*, 52(4):583–592, Oct. 1999b. ISSN 00319422. doi: 10.1016/S0031-9422(99)00151-X. URL <https://linkinghub.elsevier.com/retrieve/pii/S003194229900151X>.
- A. R. Walker, P. A. Davison, A. C. Bolognesi-Winfield, C. M. James, N. Srinivasan, T. L. Blundell, J. J. Esch, M. D. Marks, and J. C. Gray. The *TRANSPARENT TESTA GLABRA1* Locus, Which Regulates Trichome Differentiation and Anthocyanin Biosynthesis in *Arabidopsis*, Encodes a WD40 Repeat Protein. *The Plant Cell*, 11(7):1337–1349, July 1999. ISSN 1040-4651, 1532-298X. doi: 10.1105/tpc.11.7.1337. URL <https://academic.oup.com/plcell/article/11/7/1337-1349/6008466>.
- N. Walker-Hale, M. A. Guerrero-Rubio, and S. F. Brockington. Multiple transitions to high l-DOPA4,5-dioxygenase activity reveal molecular pathways to convergent betalain pigmentation in *Caryophyllales*. *New Phytologist*, 247(1):341–357, July 2025. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.70177. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.70177>.
- A. Wallace. On the law which has regulated the introduction of new species. *Annals and Magazine of Natural History* 16 (2): 184-196.. 1858. On the tendency of varieties to depart indefinitely from the original type. *Proceedings of the Linnean Society of London (Zoology)*, 3:53–62, 1855.
- C.-K. Wang, P.-L. Han, Y.-W. Zhao, X.-L. Ji, J.-Q. Yu, C.-X. You, D.-G. Hu, and Y.-J. Hao. Auxin regulates anthocyanin biosynthesis through the auxin repressor protein MdIAA26. *Biochemical and Biophysical Research Communications*, 533 (4):717–722, Dec. 2020. ISSN 0006291X. doi: 10.1016/j.bbrc.2020.09.065. URL <https://linkinghub.elsevier.com/retrieve/pii/S0006291X20318131>.
- M. I. Warhade and R. S. Badere. *Fusarium oxysporum* cell elicitor enhances betalain content in the cell suspension culture of *Celosia cristata*. *Physiology and Molecular Biology of Plants*, 24(2):285–293, Mar. 2018. ISSN 0971-5894, 0974-0430. doi: 10.1007/s12298-018-0511-x. URL <http://link.springer.com/10.1007/s12298-018-0511-x>.
- C.-Y. Wee, R. Sekeli, N. H. C. Asari, S. F. Yahya, and C. Machap. Enhancement of bioactive compounds in *Hylocereus polyrhizus* callus mediated by plant growth regulators and elicitors. 10:1–10, 2018. ISSN 1985-0484. Place: Serdang Publisher: Malaysian Society of Plant Physiology (MSPP).
- J.-K. Weng, J. H. Lynch, J. O. Matos, and N. Dudareva. Adaptive mechanisms of plant specialized metabolism connecting chemistry to function. *Nature Chemical Biology*, 17(10):1037–1045, Oct. 2021. ISSN 1552-4450, 1552-4469. doi: 10.1038/s41589-021-00822-6. URL <https://www.nature.com/articles/s41589-021-00822-6>.
- B. Winkel-Shirley. Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. *Plant Physiology*, 126(2):485–493, June 2001. ISSN 1532-2548, 0032-0889. doi: 10.1104/pp.126.2.485. URL <https://academic.oup.com/plphys/article/126/2/485/6102700>.

- K. W. S. Winson, B. L. Chew, K. Sathasivam, and S. Subramaniam. The establishment of callus and cell suspension cultures of *Hylocereus costaricensis* for the production of betalain pigments with antioxidant potential. *Industrial Crops and Products*, 155: 112750, Nov. 2020. ISSN 09266690. doi: 10.1016/j.indcrop.2020.112750. URL <https://linkinghub.elsevier.com/retrieve/pii/S0926669020306671>.
- K. W. S. Winson, B. L. Chew, K. Sathasivam, and S. Subramaniam. Effect of amino acid supplementation, elicitation and LEDs on *Hylocereus costaricensis* callus culture for the enhancement of betalain pigments. *Scientia Horticulturae*, 289:110459, Nov. 2021. ISSN 03044238. doi: 10.1016/j.scienta.2021.110459. URL <https://linkinghub.elsevier.com/retrieve/pii/S0304423821005665>.
- H. Wyler, U. Meuer, J. Bauer, and L. Stravs-Mombelli. Cyclodopa glucoside (= (2S)-5-(β -D-glucopyranosyloxy)-6-hydroxyindoline-2-carboxylic acid) and its occurrence in red beet (*beta vulgaris* var. *rubra* L.). *Helvetica Chimica Acta*, 67(5):1348–1355, Aug. 1984. ISSN 0018-019X, 1522-2675. doi: 10.1002/hlca.19840670520. URL <https://onlinelibrary.wiley.com/doi/10.1002/hlca.19840670520>.
- F. Xie, Q. Hua, C. Chen, Z. Zhang, R. Zhang, J. Zhao, G. Hu, J. Chen, and Y. Qin. Genome-Wide Characterization of R2R3-MYB Transcription Factors in Pitaya Reveals a R2R3-MYB Repressor HuMYB1 Involved in Fruit Ripening through Regulation of Betalain Biosynthesis by Repressing Betalain Biosynthesis-Related Genes. *Cells*, 10(8):1949, July 2021. ISSN 2073-4409. doi: 10.3390/cells10081949. URL <https://www.mdpi.com/2073-4409/10/8/1949>.
- F. Xie, C. Chen, J. Chen, J. Chen, Q. Hua, K. Shah, Z. Zhang, J. Zhao, G. Hu, J. Chen, and Y. Qin. Betalain biosynthesis in red pulp pitaya is regulated via HuMYB132: a R-R type MYB transcription factor. *BMC Plant Biology*, 23(1): 28, Jan. 2023. ISSN 1471-2229. doi: 10.1186/s12870-023-04049-6. URL <https://bmcpplantbiol.biomedcentral.com/articles/10.1186/s12870-023-04049-6>.
- F. Xie, K. Shah, C. Chen, I. A. Sabir, J. Chen, J. Chen, J. Chen, and Y. Qin. Unraveling betalain suppression in pitaya: insights from co-activator HuMYB9 binding at *HuCYP76AD1-1*, *HuADH1*, and *HuDODA1* super-enhancers. *Food Quality and Safety*, 8:fyae016, Jan. 2024. ISSN 2399-1399, 2399-1402. doi: 10.1093/fqsafe/fyae016. URL <https://academic.oup.com/fqs/article/doi/10.1093/fqsafe/fyae016/7625475>.
- Y. Xie, T. Zhang, X. Huang, and C. Xu. A two-in-one breeding strategy boosts rapid utilization of wild species and elite cultivars. *Plant Biotechnology Journal*, 20(5):800–802, May 2022. ISSN 1467-7644, 1467-7652. doi: 10.1111/pbi.13788. URL <https://onlinelibrary.wiley.com/doi/10.1111/pbi.13788>.
- Y. Xue, K. Li, W. Feng, Z. Lai, and S. Liu. Identification of R2R3-MYB Transcription Factor Family Based on *Amaranthus tricolor* Genome and AtrMYB72 Promoting Betalain Biosynthesis by Directly Activating *AtrCYP76AD1* Expression. *Plants*, 14(3):324, Jan. 2025. ISSN 2223-7747. doi: 10.3390/plants14030324. URL <https://www.mdpi.com/2223-7747/14/3/324>.
- R. Yang, T. Huang, W. Song, Z. An, Z. Lai, and S. Liu. Identification of WRKY gene family members in amaranth based on a transcriptome database and functional analysis of AtrWRKY42-2 in betalain metabolism. *Frontiers in Plant Science*, 14:

- 1300522, Dec. 2023. ISSN 1664-462X. doi: 10.3389/fpls.2023.1300522. URL <https://www.frontiersin.org/articles/10.3389/fpls.2023.1300522/full>.
- H. Yaqoob, A. Tariq, B. A. Bhat, K. A. Bhat, I. B. Nehvi, A. Raza, I. Djalovic, P. V. Prasad, and R. A. Mir. Integrating genomics and genome editing for orphan crop improvement: a bridge between orphan crops and modern agriculture system. *GM Crops & Food*, 14(1):1–20, Dec. 2023. ISSN 2164-5698, 2164-5701. doi: 10.1080/21645698.2022.2146952. URL <https://www.tandfonline.com/doi/full/10.1080/21645698.2022.2146952>.
- C.-Y. Ye and L. Fan. Orphan Crops and their Wild Relatives in the Genomic Era. *Molecular Plant*, 14(1):27–39, Jan. 2021. ISSN 16742052. doi: 10.1016/j.molp.2020.12.013. URL <https://linkinghub.elsevier.com/retrieve/pii/S1674205220304445>.
- C. Youngstrom, K. Wang, and K. Lee. Unlocking regeneration potential: harnessing morphogenic regulators and small peptides for enhanced plant engineering. *The Plant Journal*, 121(2):e17193, Jan. 2025. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.17193. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.17193>.
- H. Yu, T. Lin, X. Meng, H. Du, J. Zhang, G. Liu, M. Chen, Y. Jing, L. Kou, X. Li, Q. Gao, Y. Liang, X. Liu, Z. Fan, Y. Liang, Z. Cheng, M. Chen, Z. Tian, Y. Wang, C. Chu, J. Zuo, J. Wan, Q. Qian, B. Han, A. Zuccolo, R. A. Wing, C. Gao, C. Liang, and J. Li. A route to *de novo* domestication of wild allotetraploid rice. *Cell*, 184(5):1156–1170.e14, Mar. 2021. ISSN 00928674. doi: 10.1016/j.cell.2021.01.013. URL <https://linkinghub.elsevier.com/retrieve/pii/S0092867421000131>.
- J. Zeng, J. Chen, K. Shah, F. Xie, C. Chen, J. Chen, J. Zhao, G. Hu, Z. Zhang, and Y. Qin. Identification of *HuSPL* family and key role of HuSPL12 in regulation of betalain biosynthesis in pitaya. *Physiologia Plantarum*, 175(3):e13923, May 2023. ISSN 0031-9317, 1399-3054. doi: 10.1111/ppl.13923. URL <https://onlinelibrary.wiley.com/doi/10.1111/ppl.13923>.
- J. Zhang, H. Yu, and J. Li. *De novo* domestication: retrace the history of agriculture to design future crops. *Current Opinion in Biotechnology*, 81:102946, June 2023. ISSN 09581669. doi: 10.1016/j.copbio.2023.102946. URL <https://linkinghub.elsevier.com/retrieve/pii/S0958166923000563>.
- L. Zhang, C. Chen, F. Xie, Q. Hua, Z. Zhang, R. Zhang, J. Chen, J. Zhao, G. Hu, and Y. Qin. A Novel WRKY Transcription Factor HmoWRKY40 Associated with Betalain Biosynthesis in Pitaya (*Hylocereus monacanthus*) through Regulating *HmoCYP76AD1*. *International Journal of Molecular Sciences*, 22(4):2171, Feb. 2021. ISSN 1422-0067. doi: 10.3390/ijms22042171. URL <https://www.mdpi.com/1422-0067/22/4/2171>.
- Y. Zhang, S. Zheng, Z. Liu, L. Wang, and Y. Bi. Both HY5 and HYH are necessary regulators for low temperature-induced anthocyanin accumulation in Arabidopsis seedlings. *Journal of Plant Physiology*, 168(4):367–374, Mar. 2011. ISSN 01761617. doi: 10.1016/j.jplph.2010.07.025. URL <https://linkinghub.elsevier.com/retrieve/pii/S0176161710004086>.
- Y. Zhang, Z. Liu, J. Liu, S. Lin, J. Wang, W. Lin, and W. Xu. GA-DELTA pathway is involved in regulation of nitrogen deficiency-induced anthocyanin accumulation. *Plant*

- Cell Reports*, 36(4):557–569, Apr. 2017. ISSN 0721-7714, 1432-203X. doi: 10.1007/s00299-017-2102-7. URL <http://link.springer.com/10.1007/s00299-017-2102-7>.
- Z. Zhang, C. Chen, C. Jiang, H. Lin, Y. Zhao, and Y. Guo. VvWRKY5 positively regulates wounding-induced anthocyanin accumulation in grape by interplaying with VvMYBA1 and promoting jasmonic acid biosynthesis. *Horticulture Research*, 11(5):uhae083, May 2024. ISSN 2052-7276. doi: 10.1093/hr/uhae083. URL <https://academic.oup.com/hr/article/doi/10.1093/hr/uhae083/7634760>.
- C. Zhao, A. Plaza Reyes, J. P. Schell, J. Weltner, N. M. Ortega, Y. Zheng, Å. K. Björklund, L. Baqué-Vidal, J. Sokka, R. Trokovic, B. Cox, J. Rossant, J. Fu, S. Petropoulos, and F. Lanner. A comprehensive human embryo reference tool using single-cell RNA-sequencing data. *Nature Methods*, 22(1):193–206, Jan. 2025. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-024-02493-2. URL <https://www.nature.com/articles/s41592-024-02493-2>.
- S.-Z. Zhao, H.-Z. Sun, Y. Gao, N. Sui, and B.-S. Wang. Growth regulator-induced betacyanin accumulation and *dopa-4,5-dioxygenase* (*DODA*) gene expression in euhalophyte *Suaeda salsa* calli. *In Vitro Cellular & Developmental Biology - Plant*, 47(3):391–398, June 2011. ISSN 1054-5476, 1475-2689. doi: 10.1007/s11627-011-9339-6. URL <http://link.springer.com/10.1007/s11627-011-9339-6>.
- Y. Zhao, J. Sun, S. Cheronno, J.-P. An, A. C. Allan, and Y. Han. Colorful hues: insight into the mechanisms of anthocyanin pigmentation in fruit. *Plant Physiology*, 192(3):1718–1732, July 2023. ISSN 0032-0889, 1532-2548. doi: 10.1093/plphys/kiad160. URL <https://academic.oup.com/plphys/article/192/3/1718/7076681>.
- X. Zheng, S. Liu, C. Cheng, R. Guo, Y. Chen, L. Xie, Y. Mao, Y. Lin, Z. Zhang, and Z. Lai. Cloning and expression analysis of betalain biosynthesis genes in *Amaranthus tricolor*. *Biotechnology Letters*, 38(4):723–729, Apr. 2016. ISSN 0141-5492, 1573-6776. doi: 10.1007/s10529-015-2021-z. URL <http://link.springer.com/10.1007/s10529-015-2021-z>.
- H. Zhu, C. Li, and C. Gao. Applications of CRISPR–Cas in agriculture and plant biotechnology. *Nature Reviews Molecular Cell Biology*, 21(11):661–677, Nov. 2020. ISSN 1471-0072, 1471-0080. doi: 10.1038/s41580-020-00288-9. URL <https://www.nature.com/articles/s41580-020-00288-9>.
- L. Zhu, Y. Liao, K. Lin, W. Wu, L. Duan, P. Wang, X. Xiao, T. Zhang, X. Chen, J. Wang, K. Ye, H. Hu, Z.-F. Xu, and J. Ni. Cytokinin promotes anthocyanin biosynthesis via regulating sugar accumulation and *MYB113* expression in *Eucalyptus*. *Tree Physiology*, 44(1):tpad154, Feb. 2024. ISSN 1758-4469. doi: 10.1093/treephys/tpad154. URL <https://academic.oup.com/treephys/article/doi/10.1093/treephys/tpad154/7484621>.
- Q. Zhu, Y. Lu, Y. Yang, J. Song, R. Gong, S. Zhang, J. Xiong, S. Yang, G. Zhang, J. Yang, and S. He. Efficient genetic transformation and gene editing in *Erigeron breviscapus* using developmental regulators. *The Crop Journal*, page S2214514125001473, July 2025. ISSN 22145141. doi: 10.1016/j.cj.2025.05.016. URL <https://linkinghub.elsevier.com/retrieve/pii/S2214514125001473>.

A. Zsögön, T. Čermák, E. R. Naves, M. M. Notini, K. H. Edel, S. Weinl, L. Freschi, D. F. Voytas, J. Kudla, and L. E. P. Peres. *De novo* domestication of wild tomato using genome editing. *Nature Biotechnology*, 36(12):1211–1216, Dec. 2018. ISSN 1087-0156, 1546-1696. doi: 10.1038/nbt.4272. URL <http://www.nature.com/articles/nbt.4272>.

Chapter 2

First Successful Targeted Mutagenesis Using CRISPR/Cas9 in Stably Transformed Grain Amaranth Tissue

[Susanne K. Vollmer](#), Markus G Stetter, and Götz Hensel*

First published: 11 February 2026 | <https://doi.org/10.1111/pbi.70590> | *Plant Biotechnology Journal*

Grain amaranth is a non-model species and lacks key resources and molecular methods used in model plants today, including genetic transformation, a high-quality reference genome, genome editing, and protoplast isolation. In the course of this thesis, two methods for stable transgenic tissue were established. Hairy root transformation (Chapter 3) only requires minimal resources and can be as quick as 4 weeks but yields only a few milligrams of transgenic tissue. Callus transformation produces multiple grams of transgenic tissue and is highly efficient but requires more resources. The implementation of callus transformation enabled the first successful target mutagenesis using CRISPR/Cas9 in amaranth, described in the following chapter. The transformation and targeted mutagenesis will be crucial methods for future research and the study of the betalain pathway.

2.1 Summary

Grain amaranth is a nutritionally rich, stress-tolerant C_4 dicot with considerable potential for climate-resilient agriculture; however, efficient and reproducible protocols for stable transformation, regeneration, and CRISPR/Cas9-mediated editing have not yet been established. CRISPR/Cas-based genome editing is a cornerstone technology for accelerating the development of climate-resilient, high-yielding crops. Its effective application depends on robust, stable transformation procedures and CRISPR/Cas systems optimised for the target species. The absence of such tools remains a critical constraint for the genetic improvement of many promising yet underexplored crops. In this study, we edited key genes of the betalain biosynthesis pathway in grain amaranth (*Amaranthus hypochondriacus* L.) using the CasCADE modular cloning system, thereby demonstrating the feasibility of targeted mutagenesis in an orphan crop. We observed successful edits in up to 49% of transformed calli, resulting in deletions or insertions in the target genes. Our CRISPR/Cas9-mediated editing paves the way for targeted molecular research and breeding of grain amaranth.

2.2 Main part

Genome editing using CRISPR/Cas is a key technology for speeding up breeding for climate-resilient, high-yielding crops (Scheben et al., 2017). However, efficient targeted mutagenesis requires implementing stable transformation methods and establishing a CRISPR/Cas setup suitable for the species of interest (Shan et al., 2020). The availability of such methods is a significant bottleneck to advancing many promising, albeit under-researched, crops. Testing and establishing vectors for efficient application of CRISPR/Cas in non-model crops could boost research and breeding of new valuable crops (Ye and Fan, 2021). We edited key pathway genes in the betalain biosynthesis pathway of grain amaranth, i.e., *A. hypochondriacus* L., to prove how targeted mutagenesis can be implemented in an orphan crop using the CasCADE modular cloning system (Hoffie, 2022). Grain amaranth is a resilient C_4 dicot orphan crop with excellent nutritional composition. These properties make amaranth a well-suited candidate to be bred as climate-resilient crop (Joshi et al., 2018). However, no efficient and reproducible protocol for successful application of CRISPR/Cas9 or stable transformation and regeneration, has been demonstrated in *A. hypochondriacus* (Castellanos-Arévalo et al., 2020).

Amaranth produces red and yellow betalains, which are specialized metabolites in *Caryophyllales* species (Timoneda et al., 2019). Betalains have been employed as reporters in molecular biology using the RUBY cassette, which consists of the three enzymes required to produce red betalains from tyrosine (He et al., 2020). Conversely, the pathway is well-suited to evaluate the knock-out efficiency through targeted mutagenesis in *A.*

hypochondriacus, where the pigments naturally occur. Key enzyme genes in the betalain pathway are *AhCYP76AD2* and *AhCYP76AD5*, which catalyze the initial steps of betalain production. Natural knock-outs of *AhCYP76AD2* have been shown to lack red color (Winkler et al., 2024), suggesting its suitability as a visual reporter.

The foundations of successful CRISPR/Cas-mediated targeted mutagenesis are efficient guide RNAs and suitable components to express the Cas9 enzyme and the gRNAs (Shan et al., 2020). The *Cas9* sequence should be adapted to the codon usage of the target species. The promoter that drives the expression of *Cas9* and the guides need to be highly active in the target species. Finally, the selection cassette has to function efficiently in the target species. To address these requirements in amaranth, we constructed a binary vector containing *Cas9* and a four-guide cassette, using the CasCADE modular cloning system (Figure 2.1a and 2.1b, and Supplemental Figure S2.1, Supplemental Table S2.1, Appendix S3.1; Hoffie (2022)).

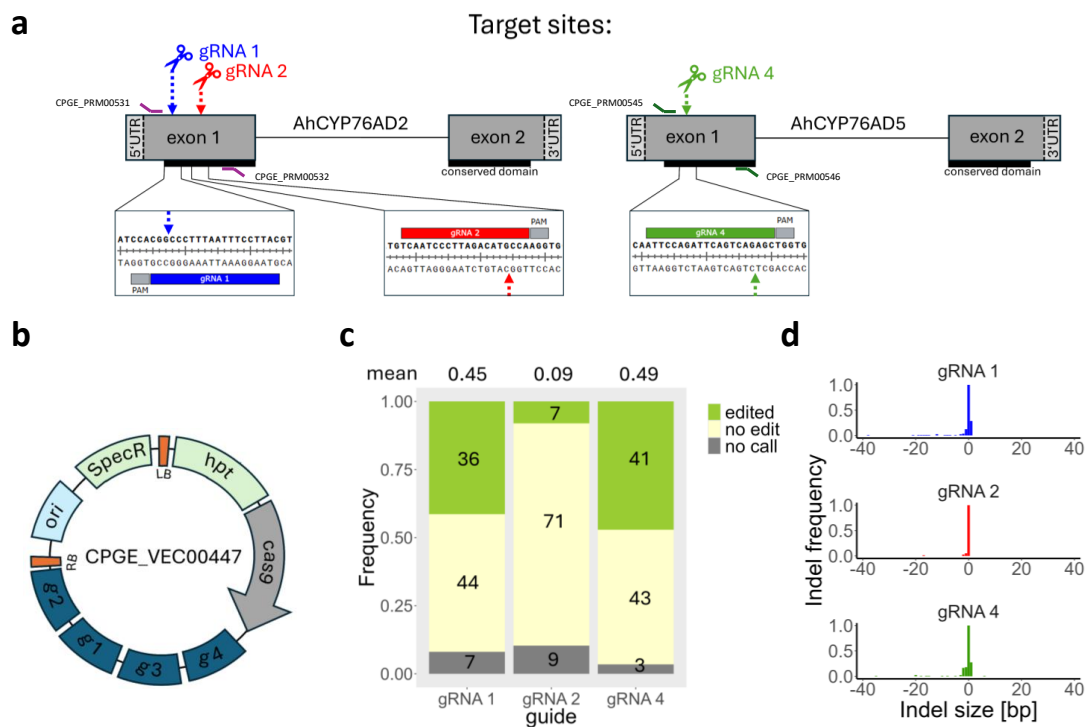


Figure 2.1: Overview of CRISPR/Cas9-mediated targeted mutagenesis in grain amaranth. a) Schematic structure of the target genes with the respective target sites of the guides in amaranth genes and the primers used for analysis. b) Vector with target guides, two targeting the first exon of *AhCYP76AD2* (gRNA 1 and gRNA 2), one the first exon of *AhCYP76AD5* (gRNA 4) and one the *RUBY* reporter (gRNA 3). c) Total number of mutations for all three guides. d) Length and frequency of the observed InDels found in 74 lines.

Amaranth plants successfully mutated in *AhCYP76AD2* should be deficient in betalains, facilitating the later detection of successfully edited regenerates. An additional gRNA not targeting the amaranth genome, but the *CYP76AD1* gene in the RUBY reporter cassette (He et al. (2020); gRNA 3) was included as a control (Figure 2.1). The

Cas9 gene was codon-optimized for the dicot *Arabidopsis thaliana* and driven by the parsley *Ubi4-2* promoter (Hoffie, 2022). Each gRNA was expressed separately by the *AtU6-26* promoter (Hoffie, 2022). For the transgenic tissue selection, the T-DNA contains an intronized *hpt* gene driven by a *CaMV* doubled-enhanced *35S* promoter to confer hygromycin resistance (Hoffie, 2022). We cloned the CRISPR/Cas9 cassette into a binary vector via *SfiI* to enable *Rhizobium*-mediated transformation (Hoffie, 2022) (Figure 2.1b). Next, we established callus transformation, which enables the highly efficient production of stable transgenic tissue in multiple-grain amaranth species. The results were obtained from two independent transformation replicates, batch 1 and batch 2. A random subset of 23 calli from batch 1 and all from batch 2 treated with *R. radiobacter* carrying the CRISPR/Cas9 vector were analyzed for edits. Briefly, 5-week-old calli of *A. hypochondriacus* were transformed with the CRISPR/Cas9 vector using *R. radiobacter* strain GV3101. For co-culture, calli were incubated on filter paper for three days before being transferred to a callus induction medium with selection. All calli treated with *Rhizobium* carrying the CRISPR/Cas9 vector (42/42 for batch 1 and 65/65 for batch 2) grew new resistant calli on the selection medium. In contrast, only two (1/21 for batch 1 and 1/32 for batch 2) of the non-treated calli survived the selection and propagated new callus, indicating the high success of transformation. After multiple rounds of selection, callus material was sampled for genotyping and analysis (Supplemental Figure S2.2). Genomic DNA (gDNA) was extracted from 87 calli of the two independent batches.

We sampled individual calli; however, callus tissue can contain cells from independent transformation events with different edits. Therefore, each gDNA included a pool of wild-type and edited alleles. To increase the sensitivity for detecting edited alleles using Sanger sequencing, we employed restriction enzyme digestion-suppressed PCR (RE-PCR) for two of the guides (gRNA 1 and gRNA 4) in the amaranth genome. Through the deconvolution of the obtained chromatograms, we found edits in 49% (gRNA 4) and 45% (gRNA 1) of the analyzed calli (Figure 2.1c). In contrast, for gRNA 2, a lack of a restriction enzyme recognition site at the Cas9-mediated cut site precluded RE-PCR; edits could still be detected in 9% of the samples. Among all samples with data available for the three sites, 24.3% were edited for both target genes (*AhCYP76AD5* and *AhCYP76AD2*), and 5.4% for all three guide RNA positions. To ensure complete knockouts for downstream analysis of the target gene, targeted mutagenesis should achieve a diverse set of mutations, including larger insertions and deletions (InDels). According to the deconvolution data, most edits resulted in the deletion or insertion of one base pair (Figure 2.1d). However, for each guide, edits with larger deletions were also found (max. 35 bp for gRNA 4, 17 bp for gRNA 2, 38 bp for gRNA 1; Supplemental Figure S2.3). To confirm the results from the deconvolution, we cloned amplicons of the target region of six random calli from gRNA 4 and five calli from gRNA 1 from into pGEM-Teasy (Promega) and sequenced eight colonies from each. In the majority of cases, the frequency for a particular type of InDels was higher in the subcloned colonies compared to the predictions (Supplemental

Table S2.2).

Our established callus transformation protocol enabled us to apply CRISPR/Cas9 in stable transformed tissue. Together with regeneration methods developed for *Amaranthus* (Castellanos-Arévalo et al., 2020), and callus-free methods, our transformation protocol can accelerate research and improvement of the crop (Maher et al., 2020). The presented results show the first successful targeted mutagenesis in grain amaranth using CRISPR/Cas9. Using the CasCADE modular cloning system and our established callus transformation, we achieved an approximately 50% editing efficiency for each of the two target genes. This paves the way for genome editing in grain amaranth for research and breeding of this orphan crop. Moreover, it may also guide the design of CRISPR systems for other species of the *Caryophyllales*, where reports of successful edits are still scarce.

2.3 Acknowledgments

The Deutsche Forschungsgemeinschaft supported this work under Germany’s Excellence Strategy – EXC-2048/1 (project ID 390686111), and STE 2654/4.

2.4 Conflict of interest

The authors declare that they have no competing interests.

2.5 Author contributions

M.G.S. and G.H. supervised the project. S.K.V. performed the experiments and analyzed the data. S.K.V., M.G.S. and G.H. discussed and interpreted the data. S.K.V. wrote the draft of the manuscript and prepared the figures. All authors edited and approved the manuscript.

2.6 References

- A. P. Castellanos-Arévalo, A. A. Estrada-Luna, J. L. Cabrera-Ponce, E. Valencia-Lozano, H. Herrera-Ubaldo, S. de Folter, A. Blanco-Labra, and J. P. Délano-Frier. *Agrobacterium rhizogenes*-mediated transformation of grain (*Amaranthus hypochondriacus*) and leafy (*A. hybridus*) amaranths. *Plant Cell Reports*, 39(9):1143–1160, Sept. 2020. ISSN 0721-7714, 1432-203X. doi: 10.1007/s00299-020-02553-9. URL <https://link.springer.com/10.1007/s00299-020-02553-9>.
- Y. He, T. Zhang, H. Sun, H. Zhan, and Y. Zhao. A reporter for noninvasively monitoring gene expression and plant transformation. *Horticulture Research*, 7(1):152, Dec. 2020. ISSN 2662-6810, 2052-7276. doi: 10.1038/s41438-020-00390-1. URL <https://www.nature.com/articles/s41438-020-00390-1>.

- I. V. O. Hoffie. Entwicklung des modularen CasCADE-Vektorsystems und dessen Verwendung zur gezielten Mutagenese der Stp13-Orthologe von Weizen und Gerste für die Etablierung dauerhafter Resistenz gegen Rost- und Mehltäupilze. 2022. Publisher: Hannover: Institutionelles Repositorium der Leibniz Universität Hannover.
- D. C. Joshi, S. Sood, R. Hosahatti, L. Kant, A. Pattanayak, A. Kumar, D. Yadav, and M. G. Stetter. From zero to hero: the past, present and future of grain amaranth breeding. *Theoretical and Applied Genetics*, 131(9):1807–1823, Sept. 2018. ISSN 0040-5752, 1432-2242. doi: 10.1007/s00122-018-3138-y. URL <http://link.springer.com/10.1007/s00122-018-3138-y>.
- M. F. Maher, R. A. Nasti, M. Vollbrecht, C. G. Starker, M. D. Clark, and D. F. Voytas. Plant gene editing through *de novo* induction of meristems. *Nature Biotechnology*, 38(1):84–89, Jan. 2020. ISSN 1087-0156, 1546-1696. doi: 10.1038/s41587-019-0337-2. URL <http://www.nature.com/articles/s41587-019-0337-2>.
- A. Scheben, F. Wolter, J. Batley, H. Puchta, and D. Edwards. Towards CRISPR/Cas crops – bringing together genomics and genome editing. *New Phytologist*, 216(3):682–698, Nov. 2017. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.14702. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.14702>.
- S. Shan, P. S. Soltis, D. E. Soltis, and B. Yang. Considerations in adapting CRISPR/Cas9 in nongenetic model plant systems. *Applications in Plant Sciences*, 8(1):e11314, Jan. 2020. ISSN 2168-0450, 2168-0450. doi: 10.1002/aps3.11314. URL <https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/aps3.11314>.
- A. Timoneda, T. Feng, H. Sheehan, N. Walker-Hale, B. Pucker, S. Lopez-Nieves, R. Guo, and S. Brockington. The evolution of betalain biosynthesis in *Caryophyllales*. *New Phytologist*, 224(1):71–85, Oct. 2019. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.15980. URL <https://onlinelibrary.wiley.com/doi/10.1111/nph.15980>.
- T. S. Winkler, S. K. Vollmer, N. Dyballa-Rukes, S. Metzger, and M. G. Stetter. Isoform-resolved genome annotation enables mapping of tissue-specific betalain regulation in amaranth. *New Phytologist*, page nph.19736, Apr. 2024. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.19736. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19736>.
- C.-Y. Ye and L. Fan. Orphan Crops and their Wild Relatives in the Genomic Era. *Molecular Plant*, 14(1):27–39, Jan. 2021. ISSN 16742052. doi: 10.1016/j.molp.2020.12.013. URL <https://linkinghub.elsevier.com/retrieve/pii/S1674205220304445>.

2.7 Supporting Information

Supplementary Method 1 – Cloning of binary plasmid CPGE_VEC00447 for targeted mutagenesis

Supplementary Method 2 – Callus transformation

Supplementary Method 3 – Analysis

Supplemental Figure 1 – Plasmid map of the CRISPR/Cas9 vector used for the targeted mutagenesis (CPGE_VEC00447)

Supplemental Figure 2 – Genotyping of genomic DNA of 74 calli for the presence of the Cas9 gene using primer pairs mentioned in Supplemental Table S2.1. Samples with a visible amplification are marked with green (67/74)

Supplemental Figure 3 – Gel electrophoresis of target amplicons for *AhCYP76AD5* (gRNA 4) using primer pairs mentioned in Supplemental Table S2.1. Genomic DNA was amplified from wild-type (WT) and edited callus (PCR). To enrich for the edited alleles, the DNA of the edited callus was digested with Hpy188I before the PCR (RE-PCR) to visualize better the 35 bp deletion at the target site of gRNA 4. MW – 100 bp molecular weight ladder

Supplemental Table 1 – Primers used for cloning and genotyping

Supplemental Table 2 – Comparison of the obtained edit frequency from subcloning and deconvolution. Note that subcloning is a sign of higher quality; however, it is not suitable for precisely determining edit frequencies due to the limited number of samples. The frequency of mutations was calculated as the percentage of samples with edits among all sequenced samples, or refers to the inferred rate of these edits in the deconvoluted Sanger sequencing chromatograms

Appendix S1 – Annotated DNA sequence of the CRISPR/Cas9 vector (CPGE_VEC00447)

2.7.1 Supplementary methods

Methods S1: Cloning of binary plasmid CPGE_VEC00447 for targeted mutagenesis

Cloning was conducted using the CasCADE modular cloning set (Hoffie, 2022). We designed the target sequences with the CRISPOR (<https://crispor.gi.ucsc.edu/>), CRISPRscan (<https://www.crisprscan.org/>), and RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) webtools. The chosen target sequences were ordered with the necessary overhangs for Golden Gate cloning (Engler et al., 2008) using the BbsI Type IIS restriction enzyme as oligos (Biolegio B.V., Nijmegen, The Netherlands)(Supplemental Table S2.1). We mixed 5 µl of each respective forward and reverse oligo (10 µM) and incubated for 5 minutes at 95°C to obtain double stranded fragments for the level 0 cloning. For cloning into Level 0 vectors (IK75-IK78) harbouring the *Arabidopsis thaliana* U6-26 promoter and the Cas9-specific scaffold, we combined 5 µl of the respective duplex and 1 µl of 30 ng/µl Level 0 plasmid with 1 µl BbsI, 1 µl T4 ligase, 2 µl of 10xFD buffer, 1 µl ATP (10 nM), and 9 µl of milliQ water. We executed the Golden Gate cloning in 10 cycles of 5 minutes at 37°C and 10 minutes at 22°C, followed by 30 minutes at 37°C and 15 minutes at 75°C. Five µl of the Golden Gate reactions were used to transform *E. coli* Mach1 or XL1 cells, generating plasmids CPGE_VEC00441-CPGE_VEC00444. For Level 1 cloning we utilised 150 ng of each Level 0 vector and 30 ng of the Level 1 backbone (IK19) for Golden Gate cloning with Esp3I, generating vector CPGE_VEC00445. For Level 2 cloning, we combined each 150 ng of the gRNA assembly (CPGE_VEC00445), the *Arabidopsis thaliana* codon-optimised Cas9 expression cassette (RB20), the auxiliary unit (IK155), and the Level 2 backbone (IK48) for Golden Gate cloning using BsaI generating CPGE_VEC00446. For cloning of the final binary vector, we mixed 20 ng of gel-purified SfiI-digested backbone (CPGE_VEC00113) and 60 ng of SfiI-digested Level 2 vector (CPGE_VEC00446) with 2 µl of T4 ligase buffer, 1 µl of T4 ligase, and 13 µl of milliQ water leading to CPGE_VEC00447. We verified successful cloning via colony PCR, test digest (excluding Level 0 vectors), and Sanger sequencing (Microsynth, Balgach, Switzerland). The final CPGE_VEC00447 was further examined through whole plasmid sequencing (Microsynth, Balgach, Switzerland).

Methods S2: Callus transformation

Seeds of *A. hypochondriacus* (PI 558499) were sterilized (2 min in 70% ethanol, 20 min in 2.4% sodium hypochlorite with one drop of Tween-20, five washes with sterile water) and germinated on germination medium (1/2 MS + 20 g/l sucrose, pH5.8, 8 g/l agar) at 24°C dark to cut hypocotyl explants after 3 days. We cultivated the explants on callus induction medium (full MS (with vitamins) + 2 mg/L 2,4D + 0.5 mg/l BAP + 20 g Sucrose, pH 5.8, 8 g/L agar) (Castellanos-Arévalo et al., 2020; Wang et al., 2001) for approximately

34-36 days at 24°C in the dark, with subculturing every 2-3 weeks, until a callus of at least 2 mm had formed. For transformation, we grew *R. radiobacter* (also named as *Agrobacterium tumefaciens*) GV3101 in CPY medium without antibiotics and with 200 µM acetosyringone and OD⁶⁰⁰ adjusted to 0.7. We incubated the callus in the bacterial solution for 10 minutes at 50 rpm and then dried them on sterile filter paper for 15 minutes. Thereafter, calli were transferred to fresh sterile filter paper on transformation plates (callus induction medium + 100 µM acetosyringone) and co-cultured for 3 days at 21°C in the dark. Without washing, we transferred the explants to callus induction medium with 150 mg/l timentin and 20 mg/l hygromycin for selection and cultivated at 24°C in the dark. On average, explants were transferred to fresh plates every two weeks.

Methods S3: Analysis

For analysis, we flash froze the callus samples in liquid N₂, ground them with a pestle and mortar, and extracted the gDNA using the NucleoSpin Plant II kit (Macherey-Nagel, Düren, Germany). For RE-PCR, we incubated 12 µl of the extracted gDNA with 0.4 µl of enzyme (Hpy188I for g4 and HaeIII for g1), 2 µl of 10x cutsmart, and 5.6 µl of milliQ water. Subsequently, 0.75 µl of the digested gDNA was used in a 30 µl PCR reaction with 15 µl of 2x GoTaq MasterMix (Promega, Madison, Wisconsin, United States of America) and 1.5 µl of respective forward and reverse primer (10 µM). For amplification, the reaction was incubated for 5 minutes at 95°C, followed by 35 cycles of 95°C for 20 seconds, 58°C for 15 seconds, and 72°C for 1 minute, with a final extension of 7 minutes at 72°C. We sent the purified PCR product for Sanger sequencing (Microsynth, Balgach, Switzerland) or used for subcloning in pGEM-Teasy (Promega, Madison, Wisconsin, United States of America).

2.7.2 Supplementary figures

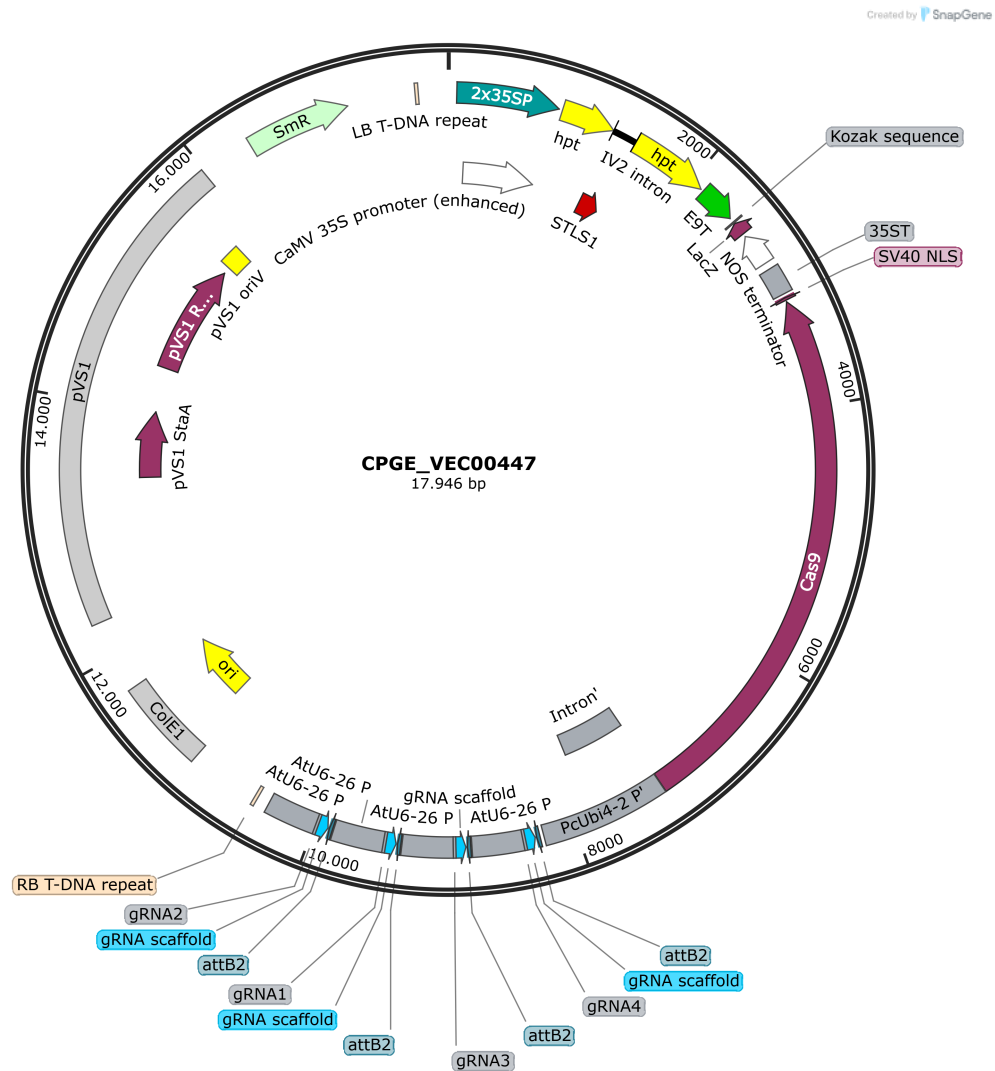


Figure S2.1: Plasmid map of the CRISPR/Cas9 vector used for the targeted mutagenesis (CPGE_VEC00447)

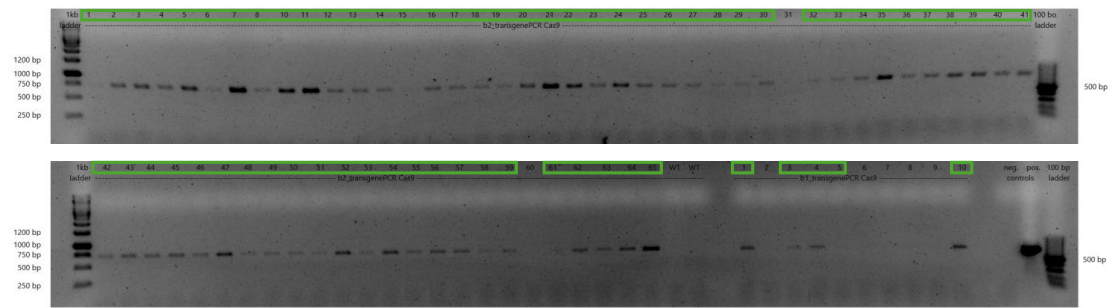


Figure S2.2: Genotyping of genomic DNA of 74 calli for the presence of the Cas9 gene using primer pairs mentioned in Supplemental Table S2.1. Samples with a visible amplification are marked with green (67/74)

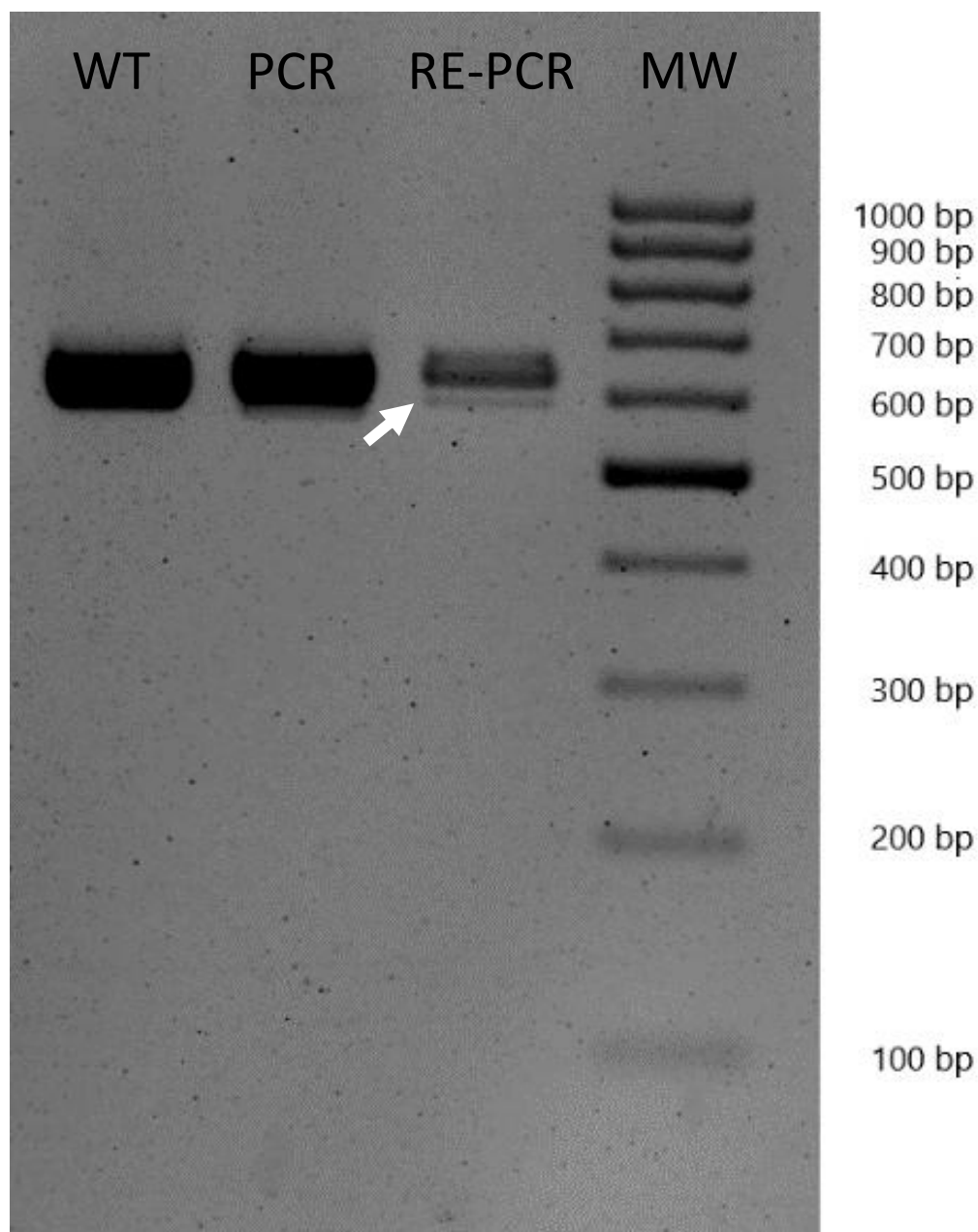


Figure S2.3: Gel electrophoresis of target amplicons for *AhCYP76AD5* (gRNA 4) using primer pairs mentioned in Supplemental Table S2.1. Genomic DNA was amplified from wild-type (WT) and edited callus (PCR). To enrich for the edited alleles, the DNA of the edited callus was digested with Hpy188I before the PCR (RE-PCR) to visualize better the 35 bp deletion at the target site of gRNA 4. MW – 100 bp molecular weight ladder

2.7.3 Supplementary tables

Table S2.1: Primers used for cloning and genotyping

Primer	Sequence (5'- 3')	Purpose
CPGE_PRM00523	ATTGgcaatcccttagacatgcca	To obtain gRNA 2 sequence with overhangs
CPGE_PRM00524	AAACtggcatgtctaagggattgc	
CPGE_PRM00529	AAACcggccctttaatttccttat	To obtain gRNA 1 sequence with overhangs
CPGE_PRM00530	ATTGataaggaaattaaagggccg	
CPGE_PRM00535	AAACccgaccgtgttcttgcctcc	To obtain gRNA 3 sequence with overhangs
CPGE_PRM00536	ATTGgggagcaagaacacggtcgg	
CPGE_PRM00541	ATTGattccagattcagtcagagc	To obtain gRNA 4 sequence with overhangs
CPGE_PRM00542	AAACgctctgactgaatctggaat	
CPGE_PRM00531	tcatggataatgcaaccctagc	Amplification of target region in <i>AhCYP76AD2</i> covering gRNA 1 and gRNA 2
CPGE_PRM00532	tgtttcaattcttgagattcatgaga	
CPGE_PRM00545	tggtttcatctgtagaagtagcga	Amplification of target region in <i>AhCYP76AD5</i> covering gRNA 4
CPGE_PRM00546	cacaaggagatgattaatttcactca	
CPGE_PRM00037	cagctcgtgcagacctacaac	Genotyping for the presence of the <i>Cas9</i> gene in the genome
CPGE_PRM00038	tgccttctaaggatagcgtg	

Table S2.2: Comparison of the obtained edit frequency from subcloning and deconvolution. Note that subcloning is a sign of higher quality; however, it is not suitable for precisely determining edit frequencies due to the limited number of samples. The frequency of mutations was calculated as the percentage of samples with edits among all sequenced samples, or refers to the inferred rate of these edits in the deconvoluted Sanger sequencing chromatograms

Target site_callus number	N subcloned analyzed by Sanger sequencing	Frequency and modification(s) detected in sub-cloned fragments	Results of chromatogram deconvolution using ICE
gRNA1_5	8	25% 1 bp insertion	19% 1 bp insertion
gRNA1_12	6	0% efficiency	5% 1 bp insertion 3% 18 bp deletion
gRNA1_24	6	0% efficiency	4% 7 bp deletion 1% 1 bp deletion
gRNA1_25	7	57% 38 bp deletion	25% 38 bp deletion 2% 1 bp insertion 1% 6 bp deletion 1% 5 bp deletion
gRNA1_64	5	20% 1 bp insertion	19% 1 bp insertion 8% 1 bp deletion
gRNA4_5	8	25% 1 bp insertion 12.5% 35 bp deletion	7% 1 bp insertion 1% 2 bp deletion
gRNA4_12	8	37.5% 1 bp insertion	9% 1 bp insertion
gRNA4_17	8	25% 1 bp deletion	15% 1 bp insertion
gRNA4_25	8	25% 35 bp deletion	8% 35 bp deletion 7% 1 bp insertion
gRNA4_52	8	12.5% 1 bp deletion 12.5% 2 bp deletion 12.5% 1 bp insertion	11% 1 bp deletion 5% 1 bp insertion 1% 2 bp deletion 1% 16 bp deletion
gRNA4_64	6	0% efficiency	0% efficiency

2.7.4 Appendix

Appendix S1: Annotated DNA sequence of the CRISPR/Cas9 vector (CPGE_VEC00447)

Labelling: d35S promoter, intronized hpt, E9 terminator, NOS terminator, 35S terminator, Cas9, PcUbi4-2 promoter, gRNA scaffold, AtU6-26 promoter, gRNA1, gRNA2, gRNA3, gRNA4, SpecR, ColE1 ori, pVS1 ori, left and right border

```
CGTGTCTACATTACGTCCAAATGGGGGCTTAGATGAGAACTTCACGATC
GGCTCTAGATCGGGCCAACATGGTGGAGCAGACACTCTCGTCTACTCCAA
GAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCA
ACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTG
TCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAATGCCA
TCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTGG
TCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGT
TCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATAACATGGTGGAGCA
CGACACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGACCAA
AGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGA
TTCCATTGCCCAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAA
GGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAA
GATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGC
ATCGTGGAAAAAGAAGACGTTCCAAACCACGTCTTCAAAGCAAGTGGATTGA
TGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCCTATCCTTCGC
AAGACCTTCTCTATATAAGGAAGTTCATTTTCATTTGGAGAGGACACGCTG
AAATCACCAGTCTCTCTCTACAAATCTATCTCTCTCGAGGGCCCCGGGGGGC
AATAAGATATGAAAAAGCCTGAACTCACCGCGACGTCTGTGAGAAGTTTC
TGATCGAAAAGTTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCG
AAGAATCTCGTGCTTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGC
GGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGC
ACTTTGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAAT
TCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGGTGTCACGT
TGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTTCGCGG
AGGCCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTTCG
GCCCATTCGGACCGCAAGGTAAGTTTCTGCTTCTACCTTTGATATATATATA
ATAATTATCATTAATTAGTAGTAATATAATATTTCAAATATTTTTTTTCAAAA
TAAAAGAATGTAGTATATAGCAATTGCTTTTCTGTAGTTTATAAGTGTGTA
TATTTTAATTTATAACTTTTCTAATATATGACCAAAATTTGTTGATGTGCAG
GTATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATC
CCCATGTGTATCACTGGCAAACCTGTGATGGACGACACCGTCAGTGCGTCCG
```

TCGCGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAG
TCCGGCACCTCGTGCACGCGGATTTCCGGCTCCAACAATGTCCTGACGGACA
ATGGCCGCATAACAGCGGTTCATTGACTGGAGCGAGGCGATGTTCTGGGGATT
CCCAATACGAGGTGCGCAACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTA
TGGAGCAGCAGACGCGCTACTTTCGAGCGGAGGCATCCGGAGCTTGCAGGAT
CGCCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATC
AGAGCTTGGTTGACGGCAATTTTCGATGATGCAGCTTGGGCGCAGGGTCGAT
GCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTACACAAATCG
CCCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCG
ATAGTGGAACCGACGCCCCAGCACTCGTCCGAGGGCAAAGGAATAGAGTA
GATGCCGACGCGTTCTGAGTATTATGGCATTGGGAAAACCTGTTTTTCTTGTA
CCATTTGTTGTGCTTGTAATTTACTGTGTTTTTTATTCGGTTTTTCGCTATCG
AACTGTGAAATGGAATGGATGGAGAAGAGTTAATGAATGATATGGTCCTT
TTGTTTCATTCTCAAATTAATATTATTTGTTTTTTCTCTTATTTGTTGTGTGT
TGAATTTGAAATTATAAGAGATATGCAAACATTTTGTTTTGAGTAAAAATG
TGTCAAATCGTGGCCTCTAATGACCGAAGTTAATATGAGGAGTAAAACACT
GAAGCCTGCAGGCATGCAAGCTGATCCACTAGAGGCCATGGcggccacggaattacg
ccaagcttgcattgcaggcctctgcagtcgacgggcccgggatccgatatctagatgcattcgcgaggtaccgagctcgaattc
actggccgtttctctactgagcgtaaaaaccggtcccgatctagtaacatagatgacaccgcgcgataatttatectagt
tgcgcgctatattttgtttctatcgcgtattaaatgtataattgcgggacttaatacataaaaccatctcataaataacgtc
atgcattacatgttaattattacatgcttaacgtaattcaacagaaattatatgataatcatcgcaagaccggcaacaggattc
aatcttaagaaactttattgccaatgtttgaacgatcgagctcggggaaattcggatccccaatacttgtatggaggcctga
gccgtacgggtcactggattttggttttaggaattagaaattttattgatagaagtattttacaaatacaatacataactaagg
gtttcttataatgctcaacacatgagcgaaaccctataagaaccctaattcccttatctgggaactactcacacattattctgga
gaaaaatagagagagatagattttagagagagactggtgatttttgcggactctagcggtcggcatctactcgggccgcac
ctcaaaccttctcttcttcttaggatcagcccttgaatcaccaccgagctgtgagagatcgatcctagctcgtagagtccagt
gatagactgatggatgaggtagcatcgagcacttcttggtagaggtgtatctcttctatcgatggttgatcgaagtactt
gaaagcagcaggagcaccgaggttgtaagggtgaagagatggatgatgttctctgcctgttccctgataggttatctctgt
gcttgttgtaagcagacaacaccttatcgaggttgcacgagcaggtacccttttagagaactcagagatctgctcgatg
atctcatccaagtagtgcttgtgctgctcaacgaaaagtgcttctgctcggttatcttctggagatcccttcaacttctcgtagt
agaagcgaggtaaagaaagttaacgtacttagatgggagagcaagctcgttcccttttgaagctcaccagcagaagcgagc
atcctctttctaccgttctcgagttcgaagagttagtacttgggagcttgatgatgagatccttcttaacctcttctgtatccct
agcctcgaggaaatcgattgggttcttctcgaaagatgacctttccatgatagtgattccgagaagttccttaacagacttgag
cttcttactctttcccttctcaaccttagccacaacgagaacagagtaagccacggtaggagaatcgaaaccaccgtatttctt
agggtcccaatcttcttcttagcaatgagcttatcagagttcttcttagggaggatagactctttagagaatccaccggctcg
cacctcgggtttcttaacgatgttcaacctgtggcatagagagcaccttcttaacggtagcgaatcccttcccttatccacacg
atctcacctgtttcaccgtttgtctcgatgagtgccctcttctgatctcaccttagcgagggtaatctcggtcttgaagaaat
tcatgatgttagagtagaagaaatacttagcggtagccttccgatctctgtctcagacttagcgatcatcttctcactcatcgta
caccttgaatcacgtacacgaactctgactcgagcttaggatacttcttgatgagagcgggtccaacaacagcggttaagggt
aagcatcgtgagcgtggtaggtgtgtgatttccctcacctttagaattggaaatcccttctgaaatcagacacgagcttgg

acttgagggtgataaccttcacttccctgatcaacttatcggttctcatcgctacttgggtttcactcctagaatcgaggatctgtgc
aacgtgcttagtgatctgcttggttccacaagctgcctcttgatgaatcctgccttatccaattcagagagtcctccctctca
gccttagtcaagttatcgaaactttctctgagtgatgagcttagcggtgaggagctgcctccaatagttcttcatTTTTCTTcaca
cctcttcacttggcacgttatcactcttaccctgttcttatcagacctgggtgagcaccttggtatcgatagaatcatccttcaag
aatgactgtggcacgatatgatcaacatcgtaatcagagagcctgttgatccaactcttgatccacatacatatcccttccg
ttctggaggtagtagaggtagagcttctcattctggagctgagtggttctcaacagggtgctctttgaggatctgagatccaagc
tctttgataccttctcgatcctcttcatcctttccctagagtcttctgtccctcttgagtggtctgggtctcttagccatttoga
tcacgatgttctcaggcttatgccttcccatcaccttcaccaactcatccacaaccttcacagctcgaggattcccttcttgatt
gcaggagatccagcgaggttagcgatatgctcatggagactatcacctgtcctgaaacctgagccttctggatatcctcttta
aaggtgagagaatcatcggtgatgagctgcatgaagttctgttagcgaatccatcagacttgaggaaatcaaggattgtctt
tccagactgcttatccctgattccgttaatgagctttctgagagccttccccaaccagtgtatcttcttcttcaactgcttca
tcacctatcatcgaagagatgagcgttaggtcttgagcctttcttcaatcatctctctatcttcaaaggagggtgagggttaagaa
cgatatcctccaagatatcctcgttttctcggttatccaagaaatccttatccttaatgatcttgaggagatcggtgtaggttcc
gagagatgcgttgaacctatcctcaacaccagaaatctcaactgaatcgaagcactcgattttcttgaagtaatcctctttgag
ctgcttcacggtcacctttctgttggcttgaacaagagatcaacgatagcctttttgtcctcacgacaaaaaagcagggttc
ctcattccctcggtcacgtacttaaccttggtaactcggtgtacacgggtgaagtactcgtagagcaaagagtgccttagggagc
accttctcggttgaaggttcttatcgaagttggtcatcctctcgatgaaagactgagcactagcaccttatccaccacctctt
cgaagttccaaggggtgatggtttctcagactttctggctcatccaagcgaatcttgagtttctctagcgagaggteccacgt
agtaagggttctgaaggtgagaatcttctcaatctttccctgttatccttgaggaaatgggtagaaatcctcttgccttctaag
gatagcgtgcaactctccgaggtggatctgatgagggatagatccgttatcgaaggctcctctgcttctgagaagatcctctct
attgagcttcacgaggagttcctcggttccatccatcttctcgaggataggcttgatgaactttagaactcttcttgagatgca
ccaccatcgatgtaaccagcgtatccgttcttagactgatcgaagaaaatctcttgtacttctctgggagctgctgtctaaca
agagccttgagaagtgtagatcctggtggtgctcatcgatatcttctgatcatagaagctgagagtgaggccttgggtgatctc
gggttctactctgaggatatcactgaggaggatagcatcagagaggttcttagcagcgaggaacaaatcagcgtactgatctc
cgatctgagcgaggaggttatcgagatcatcatcgtaggtatcctttgagagctggagcttgcacccatcagcgagatcgaag
ttagacttgaagttaggggtgagtcgagagagagagcgatcaagttccgaaaagtcggttcttcttctcaccagggagctg
agcaatgaggttctcaagccttcttgacttagagagccttagcagagaggatagccttagcatccacacctgaagcgttgatag
gggttctcttcgaaaagctggtttaggtctgcacgagctggatgaacaacttatccacatcagagttatcagggttgagatca
cctcgaatgaggaagtgctctgaacttgatcatgtgagcgagagcgaggtagatgagcctgagatcagccttatcagtaga
atcaacgagcttcttctgaggtggtagatagtagggtacttctcggtgatgccacctcatcaacgatgttccgaagatagg
gtgcctctcggtcttcttcttcttccacgaggaatgactcttcgagcctgtggaagaatgaatcatccactttagccatctcg
ttagagaagatctcttgaggtagcagatcctgttcttcttctggtgtaccttctctagcggttctcttgagtctggtagcctc
agcagtttaccagaatcgaagaggagagcaccgataaggttttcttgatagagtgcttatcggtgttccgagaaccttga
acttcttagatggcaccttgactcatcggtgatcacagccatccacagagttagttccgatatcgagtccgatagagtact
tcttatccatgctgcacatacataacatatcaagatcagaacacacatatcacacacaaatacaatcaagtcacaactcca
aaaagtcagatctacatatatacatacgtaataacaaaatcatgtaataatcacaatcatgtaatccagatctatgcaca
tatatatatacacaattaataaaaaaaaaatgatataacagatctatatctatgtatgtataacaacacaatcagatgagagaagt
atgttttcagatctgtatacatacaacacacacagatgaacaattgatacgttagatccatgtatacgtacaattagctaca
cgattaaatgaaaaaatcaacgatttcggattggtacacacaaacgcaacaatatgaagaaattcatatctgattagatata
aacataaccacgtgtagatacacagtcaaatcaacaatttatagcttctaaacggatgagatgaacaagataaagatat

acataaggcatacataagataagcagattaacaaactagcaataatacatacctaattaaaacaaggaataacagagagag
 agagagagagagagatttaccttgaaaatgaagaggagaagagaggatttcttaaaattgggggtagagaaagaaatg
 atgaattgtgagaaaggagagatagaaggggggttgatatataggctgtagaagattatTTTTgtgttgaggcgtgaa
 ggaagaggggatctgactatgacacgtttgcggttacgtatttcgataggagtctttcaacgcttaacgccgttactctatg
 accgtttgggccgtaacggggccgtttgttaacgtgatgttgattctttctttctttctttctttcttttaagaagcaattg
 tacaatcgttgctagctgtcaaacggataattcggatacggatatgcctatatcctatccgtaatttttcaatctacgtggtc
 tactgagtctaattgccaactttgtacaagaaagctgggtctagaaaaaaagcaccgactcggtgccactttttcaagttgata
 acggactagccttattttaacttgctatttctagctctaaaacgctctgactgaatctggaatCAATcactactcgacteta
 gctgtatataaaactcagcttcgttttcttatctaagcgatgtgggacttttgaagattgtttcaacttaaatgggcctatataa
 gaaatactattgttctttccatataaatgggcctgcttctcttctttcagattcccaggggccttttgaagattatcttcatac
 ttaagaatgaagatgttttattcaatcaaattcttgaaggttcgatgcctaatacttctaatacttgggacaaactatgaaacaa
 gatacaaaaactccgaatggaaagttaaaaagaagaaaacgaaagctacggttcaagaaaatgtaagctgataaaca
 aaaaactgtatgaacgaagaagaagaaaaaagaccgtaatgccaactttgtacaagaaagctgggtctagaaaaaaagca
 ccgactcggtgccactttttcaagttgataacggactagccttattttaacttgctatttctagctctaaaacccgaccgtgttct
 tgctcccCAATcactactcgactctagctgtatataaaactcagcttcgttttcttatctaagcgatgtgggacttttgaaga
 ttgttttcaacttaaatgggcctatataagaaatactattgttctttccatataaatgggcctgcttctcttctttcagattccc
 aggggccttttgaagattatcttcatacttaagaatgaagatgttttattcaatcaaattcttgaaggttcgatgcctaatact
 tctaatacttgggacaaactatgaaacaagatacaaaaactccgaatggaaagttaaaaagaagaaaacgaaagctacggtt
 caagaaaatgtaagctgataaacaacaaaaaactgtatgaacgaagaagaagaaaaaagagggtaatgccaactttgt
 caagaaagctgggtctagaaaaaaagcaccgactcggtgccactttttcaagttgataacggactagccttattttaacttgct
 atttctagctctaaaacccggccctttaatttcttatCAATcactactcgactctagctgtatataaaactcagcttcgtttt
 ttatctaagcgatgtgggacttttgaagattgttttcaacttaaatgggcctatataagaaatactattgttctttccatataa
 atgggcctgcttctcttctttcagattcccaggggccttttgaagattatcttcatacttaagaatgaagatgttttattcaat
 caattcttgaaggttcgatgcctaatacttctaatacttgggacaaactatgaaacaagatacaaaaactccgaatggaaagt
 aaaaagaagaaaacgaaagctacggttcaagaaaatgtaagctgataaacaacaaaaaactgtatgaacgaagaaga
 aaaaaagaagctaataatgccaactttgtacaagaaagctgggtctagaaaaaaagcaccgactcggtgccactttttcaagttg
 ataacggactagccttattttaacttgctatttctagctctaaaactggcatgtctaagggttgcCAATcactactcgact
 ctgctgtatataaaactcagcttcgttttcttatctaagcgatgtgggacttttgaagattgttttcaacttaaatgggcctat
 aagaatactattgttctttccatataaatgggcctgcttctcttctttcagattcccaggggccttttgaagattatcttcata
 tcttaagaatgaagatgttttattcaatcaaattcttgaaggttcgatgcctaatacttctaatacttgggacaaactatgaaac
 aagatacaaaaactccgaatggaaagttaaaaagaagaaaacgaaagctacggttcaagaaaatgtaagctgataaaca
 aaaaaaactgtatgaacgaagaagaagaaaaaaggcaatctactgacctaggccttaaGGGCCAGATCTTGG
 GCCGGTACCCGATCAGATTGTCTGTTTCCCGCCTTCGGTTTAACTATCAGT
 GTTTGACAGGATATATTGGCGGGTAAACCTAAGAGAAAAGAGCGTTTATTA
 GAATAATCGGATATTTAAAAGGGCGTGAAAAGGTTTATCCGTTTCGTCCATT
 TGTATGTGCATGCCAACCACAGGGTTCCCTCGGGAGTGCTTGGCATTCCG
 TCGGATAATGACTTCTGTTCAACCACCCAAACGTCGGAAAGCCTGACGACG
 GAGCAGCATTCCAAAAAGATCCCTTGGCTCGTCTGGGTTCGGCTAGAAGGTC
 GAGTGGGCTGCTGTGGCTTGATCCCTCAACGCGGTTCGCGGACGTAGCGCAG
 CGCCGAAAAATCCTCGATCGCAAATCCGACGCTGTTCGAAAAGCGTGATCTG

CTTGTCGCTCTTTTCGGCCGACGTCCTGGCCAGTCATCACGCGCCAAAGTTCC
GTCACAGGATGATCTGGCGCGAGTTGCTGGATCTCGCCTTCAATCCGGGTC
TGTGGCGGGAACCTCCACGAAAATATCCGAACGCAGCAAGATATCGCGGTGC
ATCTCGGTCTTGCCTGGGCAGTCGCCGCCGACGCCGTTGATGTGGACGCCG
AAAAGGATCTAGGTGAAGATCCTTTTTTGATAATCTCATGACCAAATCCCTT
AACGTGAGTTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAG
GATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAA
AAAACCACCGCTACCAGCGGTGGTTTTGTTTGCCGGATCAAGAGCTACCAAC
TCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAATACTGT
TCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCG
CCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCG
ATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG
CGCAGCGGTGCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGAGC
GAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCG
CCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGccCGGCAGGGTCCG
AACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTA
TAGTCCTGTGCGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATG
CTCGTCAGGGGGGCGGAGCCTATGGAAAACAGCCAGCAACGCGGCCCTTTTT
ACGGTTCCTGGCCTTTTTGCTGGCCTTTTTGCTCACATGTTCTTTCCTGCGTTA
TCCCCTGATTCTGTGGATAACCGATTACCGCCTTTGAGTGAGCTGATACCGC
TCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGA
AGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACAC
CGCATATGGTGCACCTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCC
AGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACA
CCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCC
GCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTGAGAGGTTT
TCACCGTCATCACCGAAACGCGCGAGGCAGGGGTACGTCGAGGTCGATCCA
ACCCCTCCGCTGCTATAGTGCAGTCGGCTTCTGACGTTCAAGTGCAGCCGTCT
TCTGAAAAAGACATGTGCGACAAGTCCTAAGTTACGCGACAGGCTGCCGCC
CTGCCCTTTTCCTGGCGTTTTCTTGTCGCGTGTTTTAGTCGCATAAAGTAGA
ATACTTGCGACTAGAAACCGGAGACATTACGCCATGAACAAGAGCGCCGCCG
CTGGCCTGCTGGGCTATGCCCCGCGTCAGCACCGACGACCAGGACTTGACCA
ACCAACGGGCCGAACCTGCACGCGGCCGGCTGCACCAAGCTGTTTTCCGAGA
AGATCACCGGCACCAGGCGCGACCGCCCGGAGCTGGCCAGGATGCTTGACC
ACCTACGCCCTGGCGACGTTGTGACAGTGACCAGGCTAGACCGCCTGGCCC
GCAGCACCCGCGACCTACTGGACATTGCCGAGCGCATCCAGGAGGCGCGG
CGGGCCTGCGTAGCCTGGCAGAGCCGTGGGCCGACACCACCACGCCGGCCG
GCCGCATGGTGTGACCGTGTTCGCCGGCATTTGCCGAGTTCGAGCGTTCCC
TAATCATCGACCGCACCCGGAGCGGGCGCGAGGCCGCCAAGGCGCGAGGCG
TGAAGTTTGGCCCCCGCCCTACCCTCACCCCGGCACAGATCGCGCACGCCCG

CGAGCTGATCGACCAGGAAGGCCGCACCGTGAAAGAGGCGGCTGCACTGCT
TGGCGTGCATCGCTCGACCCTGTACCGCGCACTTGAGCGCAGCGAGGAAGT
GACGCCACCGAGGCCAGGCGGGCGCGGTGCCTTCCGTGAGGACGCATTGAC
CGAGGCCGACGCCCTGGCGGGCCGCGGAGAATGAACGCCAAGAGGAACAAGC
ATGAAACCGCACCAGGACGGCCAGGACGAACCGTTTTTCATTACCGAAGAG
ATCGAGGCGGAGATGATCGCGGCCGGGTACGTGTTTCGAGCCGCCCGCGCAC
GTCTCAACCGTGCGGCTGCATGAAATCCTGGCCGGTTTTGTCTGATGCCAAG
CTCGCGGCCCTGGCCGGCGAGCTTGGCCGCTGAAGAAACCGAGCGCCGCCGT
CTAAAAAGGTGATGTGTATTTGAGTAAAACAGCTTGCGTCATGCGGTCGCT
GCGTATATGATGCGATGAGTAAATAAACAAATACGCAAGGGGAACGCATGA
AGGTTATCGCTGTACTTAACCAGAAAGGCGGGTCAGGCAAGACGACCATCG
CAACCCATCTAGCCCGCGCCCTGCAACTCGCCGGGGCCGATGTTCTGTTAG
TCGATTCCGATCCCCAGGGCAGTGCCCGCGATTGGGCGGGCCGTGCGGGAAG
ATCAACCGCTAACCGTTGTTCGGCATCGACCGCCCGACGATTGACCGCGACG
TGAAGGCCATCGGCCGGCGCGACTTCGTAGTGATCGACGGAGCGCCCCAGG
CGGCGGACTTGGCTGTGTCCGCGATCAAGGCAGCCGACTTCGTGCTGATTC
CGGTGCAGCCAAGCCCTTACGACATATGGGCCACCGCCGACCTGGTGGAGC
TGGTTAAGCAGCGCATTGAGGTCACGGATGGAAGGCTACAAGCGGCCTTTG
TCGTGTTCGCGGGCGATCAAAGGCACGCGCATCGGCGGTGAGGTTGCCGAGG
CGCTGGCCGGGTACGAGCTGCCCATTTCTTGAGTCCCGTATCACGCAGCGCG
TGAGCTACCCAGGCACTGCCGCCGCCGGCACAAACCGTTCTTGAATCAGAAC
CCGAGGGCGACGCTGCCCCGCGAGGTCCAGGCGCTGGCCGCTGAAATTAAT
CAAAACTCATTGAGTTAATGAGGTAAAGAGAAAATGAGCAAAAGCACAAA
CACGCTAAGTGCCGGCCGTCCGAGCGCACGCAGCAGCAAGGCTGCAACGTT
GGCCAGCCTGGCAGACACGCCAGCCATGAAGCGGGTCAACTTTCAGTTGCC
GGCGGAGGATCACACCAAGCTGAAGATGTACGCGGTACGCCAAGGCAAGAC
CATTACCGAGCTGCTATCTGAATACATCGCGCAGCTACCAGAGTAAATGAG
CAAATGAATAAATGAGTAGATGAATTTTAGCGGCTAAAGGAGGCGGCATGG
AAAATCAAGAACAACCAGGCACCGACGCCGTGGAATGCCCCATGTGTGGAG
GAACGGGCGGTTGGCCAGGCGTAAGCGGCTGGGTTGTCTGCCGGCCCTGCA
ATGGCACTGGAACCCCCAAGCCCGAGGAATCGGCGTGAGCGGTGCGAAACC
ATCCGGCCCCGGTACAAATCGGCGCGGCGCTGGGTGATGACCTGGTGGAGAA
GTTGAAGGCGGCGCAGGCCGCCAGCGGCAACGCATCGAGGCAGAAGCACG
CCCCGGTGAATCGTGGCAAGCGGCCGCTGATCGAATCCGCAAAGAATCCCG
GCAACCGCCGGCAGCCGGTGCGCCGTCGATTAGGAAGCCGCCCAAGGGCGA
CGAGCAACCAGATTTTTTCGTTCCGATGCTCTATGACGTGGGCACCCGCGA
TAGTCGCAGCATCATGGACGTGGCCGTTTTCCGTCTGTGGAAGCGTGACCG
ACGAGCTGGCGAGGTGATCCGCTACGAGCTTCCAGACGGGCACGTAGAGGT
TTCCGCAGGGCCGGCCGGCATGGCGAGTGTGTGGGATTACGACCTGGTACT
GATGGCGGTTTTCCCATCTAACCGAATCCATGAACCGATAACCGGGAAGGGAA

GGGAGACAAGCCCGGCCGCGTGTTCCTCCACACGTTGCGGACGTACTCAA
GTTCTGCCGGCGAGCCGATGGCGGAAAGCAGAAAGACGACCTGGTAGAAAC
CTGCATTCGGTTAAACACCACGCACGTTGCCATGCAGCGTACGAAGAAGGC
CAAGAACGGCCCGCCTGGTGACGGTATCCGAGGGTGAAGCCTTGATTAGCCG
CTACAAGATCGTAAAGAGCGAAACCGGGCGGCCGGAGTACATCGAGATCGA
GCTAGCTGATTGGATGTACCGCGAGATCACAGAAGGCAAGAACCCGGACGT
GCTGACGGTTCACCCCGATTACTTTTTTGATCGATCCCGGCATCGGCCGTTTT
CTCTACCGCCTGGCACGCCGCGCCGCAGGCAAGGCAGAAGCCAGATGGTTG
TTCAAGACGATCTACGAACGCAGTGGCAGCGCCGGAGAGTTCAAGAAGTTC
TGTTTCACCGTGCGCAAGCTGATCGGGTCAAATGACCTGCCGGAGTACGAT
TTGAAGGAGGAGGCGGGGCAGGCTGGCCCGATCCTAGTCATGCGCTACCGC
AACCTGATCGAGGGCGAAGCATCCGCCGGTTCCTAATGTACGGAGCAGATG
CTAGGGCAAATTGCCCTAGCAGGGGAAAAAGGTGAAAAGGTCTCTTTTCT
GTGGATAGCACGTACATTGGGAACCCAAAGCCGTACATTGGGAACCGGAAC
CCGTACATTGGGAACCCAAAGCCGTACATTGGGAACCGGTACACACATGTAA
GTGACTGATATAAAAAGAGAAAAAAGGCGATTTTTCCGCCTAAACTCTTTA
AACTTATTAAAACTCTTAAAACCCGCCTGGCCTGTGCATAACTGTCTGGCC
AGCGCACAGCCGAAGAGCTGCAAAAAGCGCCTACCCTTCGGTTCGCTGCGCT
CCCTACGCCCCGCCGCTTCGCGTCGGCCTATCGCGGCCGCTGGCCGCTCAA
AAATGGCTGGCCTACGGCCAGGCAATCTACCAGGGCGCGGACAAGCCGCGC
CGTCGCCACTCGACCGCCGGCGCCACATCAAGGCACCGGTGGGTATGCCT
GACGATGCGTGGAGACCGAAACCTTGCGCTCGTTCGCCAGCCAGGACAGAA
ATGCCTCGACTTCGCTGCTGCCCAAGGTTGCCGGGTGACGCACACCGTGGA
AACGGATGAAGGCACGAACCCAGTGGACATAAGCCTGTTTCGGTTCGTAAGC
TGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCG
AACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCATGGCTTGTTATGACT
GTTTTTTTTGGGGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAGCGCGTT
ACGCCGTGGGTGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGC
AGGGCAGTCGCCCTAAAACAAAGTTAAACATCATGAGGGAAGCGGTGATCG
CCGAAGTATCGACTCAACTATCAGAGGTAGTTGGCGTCATCGAGCGCCATC
TCGAACCGACGTTGCTGGCCGTACATTTGTACGGCTCCGCAGTGGATGGCG
GCCTGAAGCCACACAGTGATATTGATTTGCTGGTTACGGTGACCGTAAGGC
TTGATGAAACAACGCGGCGAGCTTTGATCAACGACCTTTTGGAACCTTCGG
CTTCCCCTGGAGAGAGCGAGATTCTCCGCGCTGTAGAAGTCACCATTGTTG
TGCACGACGACATCATTCCGTGGCGTTATCCAGCTAAGCGCGAACTGCAAT
TTGGAGAATGGCAGCGCAATGACATTCTTGACGGTATCTTCGAGCCAGCCA
CGATCGACATTGATCTGGCTATCTTGCTGACAAAAGCAAGAGAACATAGCG
TTGCCTTGGTAGGTCCAGCGGCGGAGGAACTCTTTGATCCGGTTCCTGAAC
AGGATCTATTTGAGGCGCTAAATGAAACCTTAACGCTATGGAACCTCGCCGC
CCGACTGGGCTGGCGATGAGCGAAATGTAGTGCTTACGTTGTCCCGCATTT

GGTACAGCGCAGTAACCGGCCAAAATCGCGCCGAAGGATGTCGCTGCCGACT
 GGGCAATGGAGCGCCTGCCGGCCCAGTATCAGCCCGTCATACTTGAAGCTA
 GACAGGCTTATCTTGGACAAGAAGAAGATCGCTTGGCCTCGCGCGCAGATC
 AGTTGGAAGAATTTGTCCACTACGTGAAAGGCGAGATCACCAAGGTAGTCG
 GCAAATAATGTCTAACAATTCGTTCAAGCCGACGCCGCTTCGCGGCGCGGC
 TTAACCTCAAGCGTTAGATGCACTAAGCACATAATTGCTCACAGCCAAACTAT
 CAGGTCAAGTCTGCTTTTATTATTTTTTAAGCGTGCATAATAAGCCCTACACA
 AATTGGGAGATATATCATGAAAGGCTGGCTTTTTCTTGTTATCGCAATAGT
 TGGCGAAGTAATCGCAACATAGCTTGCTTGGTCGTTCCGCGTGAACGTCGG
 CTCGATTGTACCTGCGTTCAAATACTTTGCGATCGTGTTGCGCGCCTGCCCG
 GTGCGTCGGCTGATCTCACGGATCGACTGCTTCTCTCGCAACGCCATCCGA
 CGGATGATGTTTTAAAAGTCCCATGTGGATCACTCCGTTGCCCCGTCGCTCAC
 CGTGTTGGGGGGAAGGTGCACATGGCTCAGTTCTCAATGGAAATTATCTGC
 CTAACCGGCTCAGTTCTGCGTAGAAACCAACATGCAAGCTCCACCGGGTGC
 AAAGCGGCAGCGGC **GGCAGGATATATTCAATTGTAAAT** GGCTTCATGTCCG
 GGAAATCTACATGGATCAGCAATGAGTATGATGGTCAATATGGAGAAAAAG
 AAAGAGTAATTACCAATTTTTTTTTCAATTCAAAAATGTAGATGTCCGCAGCG
 TTATTATAAAATGAAAGTACATTTTGTATAAAACGACAAATTACGATCCGTC
 GTATTTATAGGCGAAAGCAATAAACAAATTATTCTAATTCGGAAATCTTTAT
 TTCGA

2.8 Supplementary references

- A. P. Castellanos-Arévalo, A. A. Estrada-Luna, J. L. Cabrera-Ponce, E. Valencia-Lozano, H. Herrera-Ubaldo, S. de Folter, A. Blanco-Labra, and J. P. Délano-Frier. *Agrobacterium rhizogenes*-mediated transformation of grain (*Amaranthus hypochondriacus*) and leafy (*A. hybridus*) amaranths. *Plant Cell Reports*, 39(9):1143–1160, Sept. 2020. ISSN 0721-7714, 1432-203X. doi: 10.1007/s00299-020-02553-9. URL <https://link.springer.com/10.1007/s00299-020-02553-9>.
- C. Engler, R. Kandzia, and S. Marillonnet. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS ONE*, 3(11):e3647, Nov. 2008. ISSN 1932-6203. doi: 10.1371/journal.pone.0003647. URL <https://dx.plos.org/10.1371/journal.pone.0003647>.
- I. V. O. Hoffie. Entwicklung des modularen CasCADE-Vektorsystems und dessen Verwendung zur gezielten Mutagenese der Stp13-Orthologe von Weizen und Gerste für die Etablierung dauerhafter Resistenz gegen Rost-und Mehltäupilze. 2022. Publisher: Hannover: Institutionelles Repositorium der Leibniz Universität Hannover.
- Y. M. Wang, J. B. Wang, D. Luo, and J. F. Jia. Regeneration of plants from callus cultures of roots induced by *Agrobacterium rhizogenes* on *Alhagi pseudoalhagi*. *Cell Research*, 11(4):279–284, Dec. 2001. ISSN 1001-0602, 1748-7838. doi: 10.1038/sj.cr.7290097. URL <http://www.nature.com/articles/7290097>.

Chapter 3

Isoform-resolved genome annotation enables mapping of tissue-specific betalain regulation in amaranth

Tom S. Winkler, [Susanne K. Vollmer](#), Nadine Dyballa-Rukes, Sabine Metzger, and Markus G Stetter*

First published: 08 April 2024 | <https://doi.org/10.1111/nph.19736> | *New Phytologist*

The effective study of molecular processes requires detailed information about an organism's genome. The improved reference genome described in this chapter facilitated the annotation of candidate genes for the betalain pathway. Moreover, this chapter demonstrates the identification of a new candidate transcription factor for the betalain pathway regulation in grain amaranth. The developed hairy root transformation, in combination with transient transformation of leek epidermis cells, allowed to subsequently validate Ah-MYB2 as a positive regulator of the betalain pathway. This chapter exemplifies the power of bioinformatic and molecular resources in grain amaranth and opens important new avenues for further research.

3.1 Summary

- Betalains are coloring pigments produced in some families of the order Caryophyllales, where they replace anthocyanins as coloring pigments. While the betalain pathway itself is well studied, the tissue-specific regulation of the pathway remains mostly unknown.

- We enhance the high-quality *Amaranthus hypochondriacus* reference genome and produce a substantially more complete genome annotation, incorporating isoform details. We annotate betalain and anthocyanin pathway genes along with their regulators in amaranth and map the genetic control and tissue-specific regulation of the betalain pathway.

- Our improved genome annotation allowed us to identify causal mutations that lead to a knock-out of red betacyanins in natural accessions of amaranth. We reveal the tissue-specific regulation of flower color via a previously uncharacterized MYB transcription factor, *AhMYB2*. Downregulation of *AhMYB2* in the flower leads to reduced expression of key betalain enzyme genes and loss of red flower color.

- Our improved amaranth reference genome represents the most complete genome of amaranth to date and a valuable resource for betalain and amaranth research. High similarity of the flower betalain regulator *AhMYB2* to anthocyanin regulators and a partially conserved interaction motif support the co-option of anthocyanin regulators for the betalain pathway as possible reason for mutual exclusiveness of the two pigments.

3.2 Introduction

Coloring pigments fulfill various functions in plants, including UV protection and pollinator attraction (Koes et al., 1994; Młodzieńska and others, 2009). While anthocyanins are commonly observed as red pigments, the functionally similar but biochemically distinct betalains function as pigments in many families of the Caryophyllales (Brockington et al., 2011; Jain and Gould, 2015; Lopez-Nieves et al., 2018; Sunnadeniya et al., 2016). Like anthocyanins, betalains have been associated with photoprotection, abiotic stress resistance and antioxidant activity (Jain and Gould, 2015). It is noteworthy that no plant species producing both anthocyanins and betalains has been identified, suggesting a replacement of widespread anthocyanins by betalains in multiple lineages of the Caryophyllales (Brockington et al., 2013; Timoneda et al., 2019). However, the transition and evolution of the mutually exclusive presence of these two pigment pathways remains poorly understood.

Reference genome assemblies enable the identification of functional elements in genomes and allow to understand the evolution of these elements (Kersey, 2019). In particular, high-quality gene annotations can give insights into gene families, biosynthetic pathways and provide an invaluable resource to study gene expression patterns and genotype-phenotype associations. Still, the annotation and quantification of isoforms is challenging

for computational annotation (Salzberg, 2019). The inclusion of full-length transcript sequencing data improves accuracy and provides information about splicing-variation (Xia et al., 2019; Wang et al., 2016). Sequencing and assembling genomes are becoming common tasks and well-annotated reference genomes are available for most model plants. However, these resources are still scarce for non-model species. As most betalain producing plants are non-model species, only few well-annotated reference genome sequences are available (Dohm et al., 2014; Wang et al., 2022; Chen et al., 2021), hindering the study of the regulation of those pigments.

The anthocyanin biosynthesis pathway, as part of the flavonoid pathway, has been studied in detail and is well understood despite its high complexity and tissue-specific regulation (Fig. 3.1a; Deng and Lu, 2017; Xu et al., 2015; Lai et al., 2013). Betalains and particularly their transcriptional regulation on the other hand, have received far less attention. Both tyrosine, as basis of the betalain pathway, and phenylalanine, as initial substrate of the anthocyanin pathway, are derived from the same initial substrate in aroenate, but their downstream pathways are fundamentally different (Młodzińska and others, 2009; Lopez-Nieves et al., 2018). Betalain production only requires the activity of CYP76AD and DODA (L-3,4-dihydroxyphenylalanine 4,5-dioxygenase) genes to produce yellow betaxanthins (Fig. 3.1b; Hatlestad et al., 2012; Polturak et al., 2016, 2018; Sunnadeniya et al., 2016; Christinet et al., 2004). The activity of specialised CYP76AD enzymes and the subsequent glucosylation through glucosyltransferases allow the synthesis of red betacyanins (Polturak et al., 2016; Sunnadeniya et al., 2016; Brockington et al., 2015; Sasaki et al., 2004; Vogt et al., 1999; Vogt, 2002). Despite the characterization of flavonoid and betalain pathway genes in different species (Chang et al., 2021; Zheng et al., 2016; Casique-Arroyo et al., 2014), understanding of the conservation and regulation of both pathways remains limited.

Although the main enzymes of the betalain pathway have been identified in recent years (Hatlestad et al., 2012; Christinet et al., 2004; Sasaki et al., 2004; Vogt et al., 1999; Vogt, 2002), only few betalain regulators, including MYB transcription factors, have been described (Hatlestad et al., 2015; Xie et al., 2021, 2023; Zhang et al., 2021; Zeng et al., 2023). MYB transcription factors are best known for their involvement in the regulation of the flavonoid and anthocyanin biosynthesis pathway (Deng and Lu, 2017; Li, 2014; Gonzalez et al., 2008; Lloyd et al., 2017). The MYB transcription factors can be classified into subfamilies based on the number of adjacent repeats of the MYB DNA-binding domain (Jin and Martin, 1999). Further classification of the R2R3 subfamily into conserved subgroups has been used to transfer experimentally validated regulatory functions to different species (Stracke et al., 2001; Yanhui et al., 2006; Stracke et al., 2014). Conserved subgroups participate in the regulation of the flavonoid biosynthesis in *Arabidopsis thaliana* (Stracke et al., 2007), grape (Kobayashi et al., 2002), strawberry (Schaart et al., 2013) and many other plant species (Naik et al., 2021; Lai et al., 2013; Brendolise et al., 2017). The MYB transcription factor that has been shown to regulate

betalain biosynthesis in *Beta vulgaris* leaves was most similar to anthocyanin regulating transcription factors from other species (Hatlestad et al., 2015). This finding suggested the co-option of anthocyanin regulating transcription factors for the regulation of the chemically distinct but functionally similar betalain pathway (Hatlestad et al., 2015; Lloyd et al., 2017). However, the genetic basis and evolutionary history of this co-option remains unclear.

In this study, we generate a new reference genome for the betalain producing pseudo-cereal crop, amaranth (*A. hypochondriacus* L.). The crop forms a species complex with two other grain crops (*A. caudatus* and *A. cruentus*) and two wild relatives (*A. hybridus* and *A. quitensis*), which share a complex domestication history (Stetter and Schmid, 2017; Gonçalves-Dias et al., 2023). We remove sequencing errors from the previous assembly (Lightfoot et al., 2017) and combine *ab initio* prediction and long-read isoform sequencing data to create a new, highly complete genome annotation for grain amaranth. We use this novel genome annotation to identify and annotate betalain and flavonoid color pathway genes in the genome of *A. hypochondriacus*. To identify putative regulators of these pigment pathways, we identify MYB transcription factors genome-wide and functionally assign them into conserved subgroups. Using our phylogenetic analysis, we identify multiple candidate MYBs for betalain pathway regulation in amaranth. We utilize different crosses to map genomic regions regulating amaranth tissue color and identify causal variants for color loss in a betalain pathway gene and the flower color regulator *AhMYB2* on color QTLs. *AhMYB2* has high similarity to anthocyanin regulators and regulates multiple enzymes of the betalain pathway. Understanding the tissue-specific regulation of betalain biosynthesis in amaranth can help to elucidate the evolution of this novel pigment pathway and the mutually exclusive nature of betalains and anthocyanins.

3.3 Materials and Methods

3.3.1 Improving the *A. hypochondriacus* reference genome

We first corrected indel errors in the *A. hypochondriacus* L. reference assembly version 2.1 (Lightfoot et al., 2017), using NextPolish (Hu et al., 2020, Methods S3.1). To polish the assembly, we used the publicly available whole-genome paired-end sequencing reads from the second *A. hypochondriacus* genome assembly (Bioproject ID: PRJNA290113, SRA ID: SRR2106212) (Lightfoot et al., 2017). To prepare the whole-genome sequencing (WGS) reads for genome polishing with NextPolish (Hu et al., 2020), we removed all read pairs containing ambiguous bases using *bbduk.sh* from the BBTools suite v38.94 (<https://jgi.doe.gov/data-and-tools/software-tools/bbtools/>). We annotated repetitive elements in the polished genome using Repeatmodeler v2.0.1 (Flynn et al., 2020, Methods S3.1). We identified organellar contamination within the assembly by

mapping the *B. vulgaris* mitochondrial genome (Kubo et al., 2000) and the *A. hypochondriacus* chloroplast genome (Chaney et al., 2016) against the polished assembly using minimap2 v2.17 (Li, 2018).

3.3.2 Plant material and genome annotation datasets

To guide computational prediction and augment the *ab initio* gene prediction, we used short- and long-read RNA sequencing data. Both short- and long-read RNA sequencing data stems from the same accession (PI 558499) that was used to generate the *A. hypochondriacus* reference genome v2.1 (Lightfoot et al., 2017). The short-read RNA-seq data from eight different tissues and conditions from PI 558499 was obtained from the publicly available dataset (BioProject Accession PRJNA263128) generated by Clouse et al. (2016). The RNA-seq dataset consisted of 352,557,994 reads spanning a total of 31.7 Gb (average of 3.97 Gb per tissue type, 90 base pair (bp) paired-end HiSeq Illumina sequencing reads; Clouse et al., 2016).

For the generation of long-read RNA sequencing data, material was harvested from root, cotyledons, flower, pollen, and developing and mature seeds of *A. hypochondriacus* accession PI 558499. Root tissue and cotyledons of approximately 20 seedlings were collected 3 days after germination on wet filter paper. All other tissues were collected from one single plant, grown under controlled short-day conditions. Night and day temperatures were set to 19 °C and 24 °C, respectively. Plants were grown at 60 % air humidity under well-watered conditions. Leaf tissue was collected from a fully developed young leaf at the beginning of flowering. Flower tissue was collected from unopened flowers at the tip of the inflorescence. Pollen was collected by shaking the inflorescence inside a paper bag during full bloom. Pollen collection was repeated three times within one week. Developing seeds were extracted from the flower with forceps to avoid sampling flower tissue. Approximately 20 young seeds were collected at full size, but still soft. Mature seeds were fully ripe and stored before RNA extraction. All other samples were flash frozen in liquid nitrogen and stored at -80 °C until RNA extraction. We extracted RNA using the Invitrogen PureLink Plant RNA Reagent (Thermo Scientific™, Thermo Fisher Scientific Inc., Massachusetts, USA) following the manufacturers protocol. PacBio isoform sequencing (Iso-Seq) was conducted by the West German Genome Center (WGCG) using the extracted RNA samples from different tissues (Table S3.1S1).

3.3.3 Genome annotation

Iso-Seq circular consensus sequencing (CCS) reads were processed using Isoseq3 v3.3.0 (<https://github.com/PacificBiosciences/IsoSeq>; Table S3.2) and cDNACupcake v28.0.0 (https://github.com/Magdoll/cDNA_Cupcake) and corrected with the polished reference genome sequence using SQANTI3 v4.2 (Tardaguila et al., 2018, Methods S3.2).

To assess the effect of genome polishing on the assembly of full-length Iso-Seq transcripts, we also processed the reads using the unpolished reference genome (Lightfoot et al., 2017). We predicted open reading frames (ORFs) in the transcripts using CPC2 (Kang et al., 2017) and evaluated completeness of the transcript set using BUSCO (Simão et al., 2015) in protein mode. We used BRAKER2 v2.1.6 (Brůna et al., 2021) for *ab initio* gene prediction using the softmasked *A. hypochondriacus* genome assembly v2.2 (Methods S3.3 S3). Since it was unlikely for real genes in the genome to be covered neither by RNA sequencing reads nor by protein alignments and these genes contributed only marginally to the BUSCO score, we excluded predicted genes without external support from the annotation. In order to combine the computational prediction of gene structures of BRAKER2 with sequenced full-length transcripts, we used TSEBRA v1.0.3 (Gabriel et al., 2021) to select gene structure annotations from full-length transcripts and the BRAKER2 annotation based on their external support (Methods S3.4). We assessed completeness of the genome annotation v2.2 using the annotated protein sequences as input for BUSCO v5.2.2 (Simão et al., 2015).

Functional annotation of annotated protein sequences was carried out using eggNOG-mapper v2.17 (Cantalapiedra et al., 2021), Mercator4 v.5.0 (Schwacke et al., 2019), and Interproscan v5.56-89.0 (Jones et al., 2014), retrieving pathway and gene ontology (GO) annotations (Methods S3.5). We identified betalain pathway genes using BLAST (Altschul, 1997) searches with previously described betalain pathway genes (Table S3.3, Methods S3.5). We used the flavonoid pathway identification pipeline KIPes v0.35 (Pucker et al., 2020) and protein to genome alignments to identify candidate genes of the flavonoid pathway (Methods S3.5). To identify MYB transcription factors, we searched the predicted protein sequences of the *A. hypochondriacus* genome annotation v2.2 for matches to the MYB DNA-binding domain HMM profile (Pfam ID: PF00249, downloaded 11.01.2022) using HMMer v3.3.2 (Eddy, 2009) and used the identified proteins for phylogenetic analysis (Methods S3.5). We quantified gene expression in different tissues using kallisto v0.48.0 (Bray et al., 2016) with sequence based bias correction and the previously described short-read datasets (Table S3.1).

3.3.4 Mapping betalain related quantitative trait loci

To map the genetic basis of tissue color differences in amaranth, we created two distinct mapping populations. For the color regulator bulk segregant analysis (BSA), we used an F₂ mapping population created from a cross of PI 686465, which represents an F₁ hybrid of PI 568125 and PI 568179 (Brenner, 2019), with PI 538323. For the biosynthesis BSA, the F₂ mapping population was created from a cross between PI 686465 and PI 576485. Because PI 686465 and PI 538323 were not homozygous lines, the mapping populations might carry multiple haplotypes for functional loci. Plants were grown well-watered at 60 % air humidity under controlled short-day conditions with night and day

temperatures set to 19°C and 24°C, respectively. For the regulator BSA, plants were bulked (69 plants per bulk) based on flower color, however all plants produced red leaves. For the biosynthesis BSA, plants were bulked (80 plants per bulk) based on whole plant color and green plants never produced red coloring pigments. We extracted betalains from leaves and flowers of the three parental accessions, using a modified protocol (Chang et al., 2021; von Elbe, 2001), and determined betalain content using photometric quantification and liquid chromatography - mass spectrometry (LC-MS) analysis (Methods S3.6). We extracted bulked DNA from the pools using the DNeasy Plant Pro Kit (Qiagen, Hilden, Germany) using the manufacturers protocol. Extracted DNA was sequenced by the Cologne Center for Genomics (CCG) on an Illumina NovaSeq 6000 platform in 150 bp paired-end reads (Table S3.4). Sequencing quality was assessed using fastQC v0.11.9 (Andrews, 2010). We mapped reads to the *A. hypochondriacus* reference genome v2.2 using bwa-mem2 v2.2.1 (Li, 2013) and followed GATK v4.1.7 "best practices" (McKenna et al., 2010) to call and filter SNPs. Only biallelic SNPs from the 16 largest Scaffolds were included in the analysis and sites were filtered for a maximum of 5 % missing values using vcftools v0.1.16, resulting in a total of 4,367,544 SNPs (Danecek et al., 2011). We predicted the effect of called SNPs on protein coding genes using SnpEff v5.1d (Cingolani et al., 2012). We carried out BSA using the QTLseqR package (Mansfeld and Grumet, 2018) and increased the genome-wide false-discovery G' threshold to 3.

To compare gene expression between the green and red bulks of both BSAs, we newly bulked 29 plants of each bulk for the biosynthesis BSA and 24 plants from each bulk for the regulator BSA. We extracted RNA from bulked flower tissue using the Invitrogen PureLink Plant RNA Reagent (Thermo Scientific™, Thermo Fisher Scientific Inc., Massachusetts, USA) following the manufacturers protocol. Sampled flower tissue was immediately stored at -80°C until RNA extraction. Sequencing libraries were created using the TruSeq stranded mRNA preparation kit and sequenced by the CCG on a Illumina NovaSeq 6000 platform in stranded 100 bp paired-end reads (Table S3.5). We removed adapter sequences using trimmomatic v0.39 (Bolger et al., 2014), evaluated sequencing quality using fastQC v0.11.9 (Andrews, 2010), and quantified gene expression using kallisto v0.48.0 (Bray et al., 2016). We defined a transcript per million (TPM) threshold of 0.5 in at least one sample to define a gene as expressed.

To confirm the presence of the mutually exclusive non-synonymous and stop variants in the *AhCYP76AD2* gene of green plants from the biosynthesis BSA, we Sanger sequenced (LightRun, Eurofins Genomics) one individual with red leaves and four individuals with green leaves using primers SV4 and SV5 (Table S3.6). Sequencing results were mapped to the *AhCYP76AD2* genomic sequence and visualized using CLC DNA Workbench 5.6.1 (QIAGEN).

3.3.5 Cloning and overexpression of *AhMYB2*

We cloned *AhMYB2.1* and *AhMYB2.2* from cDNA of the flower tissue RNA sample of PI 558499 also used for isoform sequencing (see above). We amplified *AhMYB2* CDS by PCR with primers SV1 and SV2 (Table S3.6) and assembled the final vectors, using the pGGZ001, pGGA004, pGGB003, MYB isoforms in pGGC000, pGGD001, pGGE001 and pGGF008 modules of the green gate system (Lampropoulos et al., 2013). The two resulting vectors were *p35S:AhMYB2.1-GFP* and *p35S:AhMYB2.2-GFP*. In addition, we used *p35S:mCherry-NLS* as control for nuclear localisation.

For hairy root transformation, we co-transformed *Agrobacterium rhizogenes* AR1193 (Stougaard et al., 1987) with the expression vectors and the pSoup plasmid (Hellens et al., 2000). We performed hairy root transformation according to Lota et al. (2013) with minor changes. The *A. rhizogenes* cultures for hairy root transformation were grown on plates for 2 days at 28°C. We scraped the bacteria from the plates using sterile loops and re-suspended them in PIPES buffer (Lota et al., 2013), using around 1.5 plates fully covered with *A. rhizogenes* biofilm per milliliter of buffer. Seeds of PI 558499 were sterilized and germinated on solid Murashige-Skoog-medium (MS) (Murashige and Skoog, 1962) with 1 % sucrose. We used a sterile 27 G needle dipped into the respective bacterial solution to wound each 3-4 day-old seedlings hypocotyl several times. *Agrobacterium*-treated plants were then grown in a growth chamber at 22°C with a 16/8 h light/dark cycle. Hairy roots were evaluated visually for their color after 23 days. We evaluated GFP signal under the Leica MZ16 F stereo microscope equipped with a DFC420 C camera and Leica 10447407 GFP long-pass filter to validate successful transformation. For the time series, we imaged side branches of red hairy roots (expressing *p35S:AhMYB2.1-GFP*) 76, 78 and 80 days after transformation. The images were taken using a Leica M165 FC stereo microscope equipped with a DFC425 C camera and Leica 10447407 GFP long-pass filter. The Leica EL6000 UV lamp attached to the stereo microscope was replaced between timepoint 2 and 3 of the time series, leading to slightly higher fluorescence intensity at the third time point. Therefore we adjusted the exposure time for the GFP images from 6.5 ms to 2.0 ms for the third time point. We adjusted the brightness of the GFP images with Fiji (Schindelin et al., 2012), using the same parameters for all images of the series. We extracted betalains from transformed and wild-type like root tissue of the same plant 84 days after transformation and performed relative quantification of betalain content using LC-MS analysis (Methods S3.6).

To study the sub-cellular localisation of AhMYB2, we bombarded leek epidermal cells with DNA-coated gold particles following Pietsch (2022). We coated gold particles simultaneously with *p35S:mCherry-NLS* and either *p35S:AhMYB2.1-GFP* or *p35S:AhMYB2.2-GFP* to evaluate nuclear localisation. We analyzed the bombarded leek cells after 24 to 30 hours under Leica DMRE microscope with a DFC7000 T camera using the AHF F36-504 mRFP (for mCherry) and AHF F46-002 EGFP (for GFP) filter cubes.

3.4 Results

3.4.1 Improved *A. hypochondriacus* reference genome

A high-quality reference genome sequence is a prerequisite for numerous types of analyses. To improve the available *A. hypochondriacus* genome assembly and remove indel errors we polished the genome using 38.8 Gb of Illumina reads (Lightfoot et al., 2017). After filtering of reads with ambiguous basecalls, 384,876,330 reads with length of 100 bp remained. Genome polishing resulted in a length increase of 105,049 bp (from 403,889,442 bp to 403,994,491 bp) of the polished reference genome compared to the unpolished reference genome (Lightfoot et al., 2017). In total, 56.91 % (229,915,110 bp) of the assembly were classified as retroelements (20.57 %), DNA transposons (4.86 %), or simple repeats and low complexity regions (3.08 %; Table S3.7). These elements were distributed across the genome with increased density in regions with lower gene content (Fig. 3.1c). We further identified and annotated regions with organellar sequence similarity (Fig. S3.1). These regions could have resulted from erroneous assembly or could be the result of transfer into the genome (Wang et al., 2018).

In order to improve the annotation with full-length transcript sequencing evidence, we performed PacBio isoform sequencing of seven different tissues of the amaranth reference accession (PI 558499, Table S3.1 S1, Notes S3.1). The assembled Iso-Seq transcriptome showed high reported completeness, which was further increased by reference-based correction using the polished reference genome (BUSCO completeness of 90.7 % and 90.8 %, respectively (Table S3.8). Reference-based correction using the unpolished reference genome strongly decreased the reported BUSCO score, from 90.6 % to 74.6 % complete, instead (Table S3.8). For a total of 43,245 transcript sequences that were predicted to be coding before and after reference-based 'correction' using the unpolished genome, 35.1 % (15,194) differed in predicted ORF length after correction. While 12,289 transcripts had longer predicted ORFs without 'correction', 2,805 were predicted with longer ORF after 'correction' with the unpolished reference genome. In contrast to that, only 10.1 % of transcripts differed in predicted ORF length after correction using the polished genome.

We combined the BRAKER2 genes that were at least partially supported by sequencing evidence with the predicted coding sequence of the full-length transcripts into the new genome annotation v2.2 using TSEBRA (Gabriel et al., 2021). Our final genome annotation has 23,817 annotated genes of which 99.12 % locate to the first 16 Scaffolds (Table 3.1, Notes S3.2). In comparison to the genome annotation of *A. cruentus* (89.8 % complete) and the previously published genome annotation v2.1 (80.7 % complete), the genome annotation v2.2 shows a strong increase in completeness with a BUSCO score of 98.0 % complete (Table S3.8). We were able to assign genes into 5,460 of 5,783 (94.41 %) of available Mercator bins of conserved biological processes and included functional annota-

tions for 20,946 genes (95.56 %) from at least two sources (Fig. S3.2). These annotations will allow the effective use of the *A. hypochondriacus* reference genome for quantitative genetic analysis including genome wide association studies and molecular work.

Table 3.1: Comparison of genome annotation statistics between the *A. cruentus* genome (Ma et al., 2021), the published *A. hypochondriacus* genome v2.1 (Lightfoot et al., 2017) and the *A. hypochondriacus* genome annotation derived from this study (v2.2). Coding sequence (CDS) and exon statistics for both previously published genome annotations were published in Ma et al. (2021). We calculated the BUSCO score using annotated protein sequences with identical parameters for all three genome annotations. Only the reported percentage of complete BUSCO genes (single copy and duplicated) is displayed.

	<i>A. cruentus</i>	<i>A. hypochondriacus</i> v2.1	<i>A. hypochondriacus</i> v2.2
Protein coding genes	25,477	23,883	23,817
Total isoforms	25,477	23,883	28,074
BUSCO score	89.8 %	80.7 %	98.0 %
Mean CDS length (bp)	1,141	1,066	1,258
Mean exons per gene	4.86	4.87	5.39
Mean exon length (bp)	235	219	233

Color pathways in the *A. hypochondriacus* genome

To identify candidate genes for the betalain pathway, we carried out BLAST searches using the sequences of previously characterized pathway genes of different species against our annotated *A. hypochondriacus* protein sequences. For the *CYP76AD* query genes, two main candidate genes were identified, one gene as best hit for *CYP76AD1* from *B. vulgaris* and *CYP76AD2* from *A. cruentus* and one gene as best hit for *CYP76AD5* and *CYP76AD6* from *B. vulgaris* (Table S3.3). For *DODA1*, two different genes were identified as candidates. The candidate gene more similar to the *B. vulgaris* gene was located next to the candidate gene for *CYP76AD1* and *CYP76AD2* on Scaffold 16 (Table S3.3). Since multiple betalain pathway genes underwent gene duplication and neofunctionalization (Brockington et al., 2015), we performed phylogenetic analysis of the multiple *CYP76AD* and *DODA* candidate genes with previously described betalain pathway genes (Table S3.9 and Table S3.10). We termed the identified *CYP76AD* genes *AhCYP76AD2* and *AhCYP76AD5* for clustering with the *CYP76AD* α -clade and β -clade, respectively (Fig. S3.3). Phylogenetic analysis of identified *DODA* candidate genes indicated that both belonged into the *DODA* α -clade with potential *DODA* activity (Fig. S3.4). According to their relationship to BvDODA α 1, which is involved in betalain biosynthesis (Hatlestad et al., 2015), and BvDODA α 2, which displayed only minor *DODA* activity *in planta* and *in vitro* (Bean et al., 2018), we termed the two *DODA* candidates *AhDODA α 1* and *AhDODA α 2*. We further identified three different glucosyltransferases (*AhcDOPA5GT* (cyclo-DOPA 5-O-glucosyltransferase), *AhBetanidin5GT* (betanidin 5-O-glucosyltransferase) and *AhBetanidin6GT* (betanidin 6-O-glucosyltransferase)) based on the published sequences from *A. hypochondriacus* and *Amaranthus tricolor* (Table

S3.3). We found strong expression of *AhCYP76AD2* and *AhDODA α 1* especially in the red flower of PI 558499, as well as in root and stem tissue (Fig. S3.5). The expression patterns of glucosyltransferase genes differed between tissues. While *AhcDOPA5GT* showed constitutive expression in all assessed tissues, *AhBetanidin5GT* and *AhBetanidin6GT* were most strongly expressed in flower and leaf, respectively (Fig. S3.5). We identified candidate genes for all main betalain pathway genes in the genome annotation and annotated their genomic location.

The second major coloring pigment pathway we annotated was the flavonoid biosynthesis pathway that in other species includes the anthocyanin production. We identified flavonoid pathway genes using KIPes (Pucker et al., 2020) and found copies of most flavonoid pathway genes in amaranth (Table S3.11). However, no candidates for *CHI2* (chalcone isomerase), *F3'5'H* (flavonoid 3'5'-hydroxylase), *FNS1* (flavone synthase) and *ANS* (anthocyanidin synthase) could be inferred using KIPes. We identified multiple gene copies displaying all conserved residues for the general phenylpropanoid pathway genes *PAL* (phenylalanine ammonium lyase), *4CL* (4-coumarate:CoA ligase), and *C4H* (cinnamate 4-hydroxylase) as well as for *F3'H* (flavonoid 3'-hydroxylase). Since the *ANR* (anthocyanidin reductase) candidate gene had low sequence similarity to known homologs and missed conserved residues, we defined the gene as missing in amaranth (Table S3.11). Classification into metabolic pathways using Mercator supported all genes identified by KIPes, except for additional *F3'H* candidates, and similarly inferred *F3'5'H*, *ANS* and *ANR* to be missing. We additionally searched for candidate genes for which we had not found candidates using KIPes through alignment of flavonoid pathway proteins to the amaranth genome sequence (Methods S3.5). We identified an unannotated locus with high similarity to ANR genes on Scaffold_9 at position 21,774,236 to 21,777,354, which however had no annotation support from computational and long-read annotation. In contrast, we found high similarity of the annotated gene AHp021795 to ANS proteins from other species and added it as *ANS* candidate gene. In comparison to the alignment of ANS sequences of other species, AHp021795 contained an additional predicted intron which might indicate potential pseudogenisation. We found the general phenylpropanoid pathway genes to be strongly expressed in most tissues (Fig. S3.6). However, the early flavonoid pathway gene *CHS* (naringenin-chalcone synthase) was mainly expressed in flower and water-stressed tissue (Fig. S3.6), possibly limiting the production of later flavonoid compounds in other tissues.

3.4.2 Location and conservation of the color regulating MYB transcription factor family

MYB transcription factors are involved in the regulation of key plant processes, including anthocyanin and betalain pigment biosynthesis (Dubos et al., 2010; Hatlestad et al., 2015; Lloyd et al., 2017; Liu et al., 2015). To identify potential regulators of the pigmentation

pathways, we annotated MYB transcription factor genes in the genome of amaranth. In total, we identified 82 R2R3 MYB candidate genes (92 total isoforms), three 3R MYB candidate genes (three isoforms) and one 4R MYB candidate gene in the genome of amaranth (two isoforms; Table S3.12). Similar to *A. thaliana* and rice (Katiyar et al., 2012), most R2R3 MYB genes featured two (16 genes and isoforms) or three (60 genes and isoforms) exons (Table S3.12). For instance, AHp022773 (*AhMYB2*) showed a conserved gene structure with three exons but two different isoforms (*AhMYB2.1* and *AhMYB2.2*) lead to different proteins and could add functional variation.

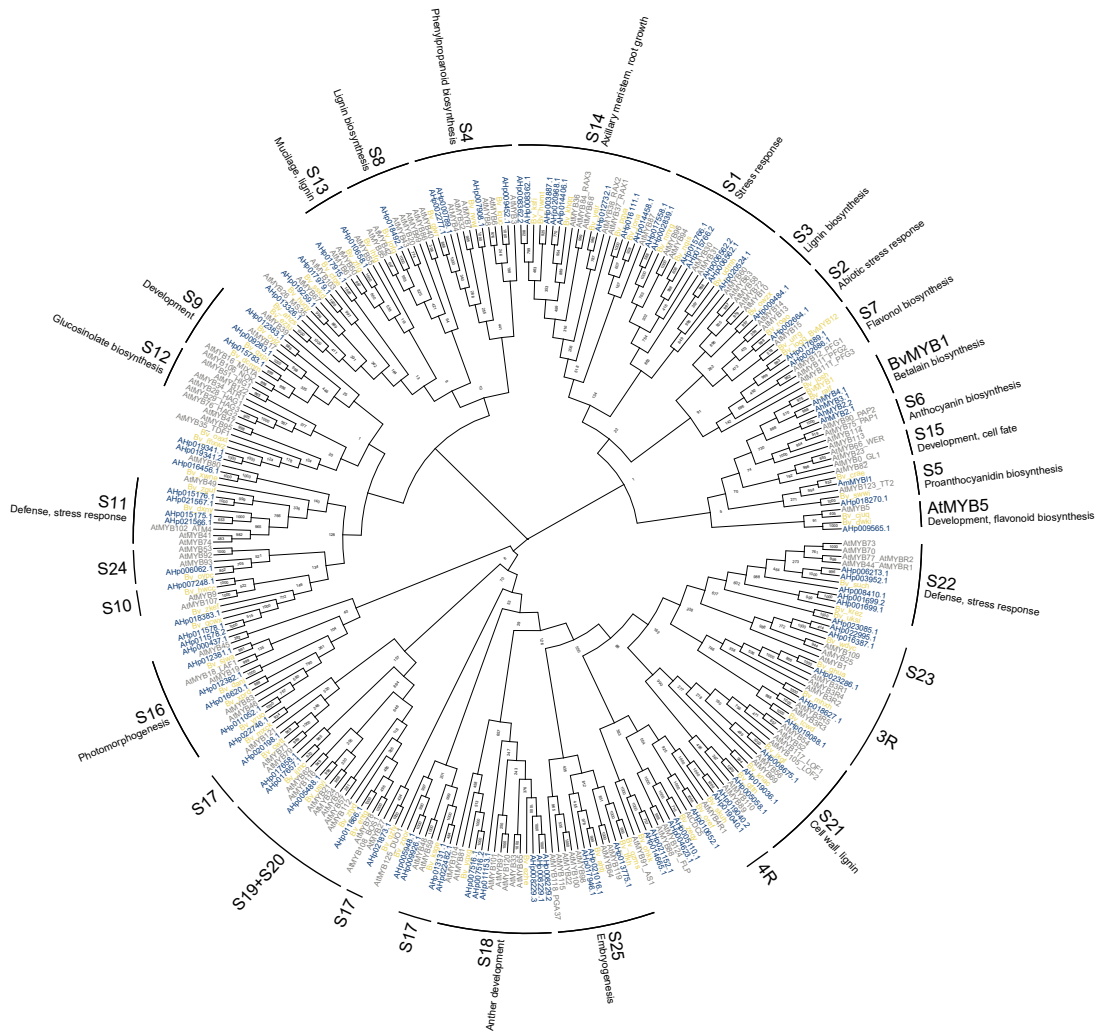


Figure 3.2: Neighbour-joining phylogenetic tree based on multiple sequence alignment of the R2R3, 3R and 4R MYB protein sequences of *A. hypochondriacus* (blue), *A. thaliana* (grey; Stracke et al., 2001) and *B. vulgaris* (yellow; Stracke et al., 2014). Bootstrap values from 1000 replicates, MYB subgroup assignments and functional classifications for subgroups were annotated in the phylogeny.

Out of the 86 identified MYB genes (R2R3, 3R and 4R), 67 could be assigned to estab-

lished subgroups using phylogenetic analysis (Fig. 3.2; [Kranz et al., 1998](#); [Stracke et al., 2001](#)). The identified 3R and 4R MYB genes clustered together with the 3R and 4R MYB genes of *A. thaliana* and *B. vulgaris*, respectively (Fig. 3.2). The conserved functions of subgroups across distant taxa enabled us to functionally assign MYBs to processes like stress responses, developmental processes or metabolite biosynthesis (Table S3.12). No amaranth MYBs were classified into the subgroups S6 (anthocyanin biosynthesis), S12 (glucosinolate biosynthesis), S15 (epidermal cell development) and S19 (stamen development; Fig. 3.2). Multiple MYB subgroups are involved in the regulation of color pathways. We classified four amaranth MYBs (Ahp000789, Ahp002277, Ahp007908 and Ahp009452) into subgroup S4, involved with transcriptional repression of different branches of the phenylpropanoid pathway ([Wang et al., 2020](#)). The amaranth MYBs Ahp002686 and Ahp017689 were classified into the early flavonoid pathway regulator subgroup S7. Ahp018270 and the seed color regulation candidate AmMYB1 ([Stetter et al., 2020](#)) were grouped into subgroup S5 involved with proanthocyanidin regulation. No amaranth or *B. vulgaris* MYBs were classified into the anthocyanin biosynthesis subgroup S6, but the betalain regulator BvMYB1 (Bv_jkk; [Hatlestad et al., 2015](#)) was grouped together with three amaranth MYBs (Ahp022773, Ahp016530 and Ahp016531) next to *A. thaliana* subgroup S6. The amaranth MYBs were assigned to the BvMYB1-clade, potentially associated with the regulation of betalain biosynthesis ([Hatlestad et al., 2015](#)). As candidate betalain pathway regulators, we named the transcription factors *AhMYB2* (Ahp022773), *AhMYB3* (Ahp016530) and *AhMYB4* (Ahp016531). While we observed only low expression of *AhMYB3* and *AhMYB4* across all tissues, the two isoforms of *AhMYB2* were strongly expressed in flower tissue (Fig. S3.7).

Different flavonoid pathway regulating MYB proteins require the formation of the MBW protein complex with two additional transcription factors (bHLH and WDR) to carry out their function ([Li, 2014](#)). The interaction between MYB proteins and the bHLH interaction partners for flavonoid pathway regulation is mediated by a conserved interaction motif in the R3 domain ([Zimmermann et al., 2004](#); [Millard et al., 2019](#)). In contrast to anthocyanin regulators, the betalain regulating *BvMYB1* has lost this interaction motive and functions without known interaction partners ([Hatlestad et al., 2015](#)). We therefore evaluated the presence of the conserved bHLH-interaction motif ([D/E]Lx₂[R/K]x₃Lx₆Lx₃R) in all identified amaranth MYBs. Similar to inferences in other plant species ([Zimmermann et al., 2004](#); [Rajput et al., 2022](#)), we detected the motif in the MYB domain in two out of four genes of S4 (Ahp002277 and Ahp009452), and both genes of S5 (*AmMYB1* and Ahp018270), indicating potential interaction with bHLH proteins and transcriptional regulation through the MBW complex. A slightly altered bHLH-interaction motif in which a conserved leucine residue was replaced by the biochemically similar isoleucine ([D/E]Lx₂[R/K]x₃Lx₆Ix₃R) was found in two *A. hypochondriacus* MYBs, Ahp009565 from the AtMYB5-clade and AhMYB2.1 from the BvMYB1-clade. The motif differentiates AhMYB2.1 from other BvMYB1-clade proteins, in which the

first residues of the motif are not conserved.

3.4.3 Mapping pigment biosynthesis pathway genes

Identifying the evolutionary changes that impact color variation within populations is essential to understand the mutual exclusive nature of anthocyanins and betalains. To study the genetic basis of tissue color and its regulation in amaranth, we conducted two bulk segregant analyses (BSA) from different initial crosses (Fig. 3.3a). We quantified tissue color in the parental accessions and found high betacyanin content (e.g., amaranthin) in red samples, whereas betaxanthin content was low in all assessed tissues and accessions (Figs 3.3, S3.8). In the biosynthesis BSA, 80 green (no red pigmentation in any visible tissue) and 80 red (red in all visible organs) plants were sequenced in two pools (Fig 3.3a). A comparison of allele frequencies between bulks identified a large QTL on Scaffold 16, from 1.1 mb to 14.4 mb (from 1,080,628 bp to 14,358,251 bp; Fig. 3.4a,b). The two betalain pathway genes *AhCYP76AD2* and *AhDODA α 1* were located on the QTL between the two peaks and represent likely candidate genes for loss of coloration (Fig. 3.4b). We quantified gene expression from bulked flower tissue for the green and red bulks of both BSAs using RNA sequencing. However, gene expression levels for *AhCYP76AD2* and *AhDODA α 1* were similar between the red and green bulk of the biosynthesis BSA (Fig. S3.9).

To identify functional variants which could lead to color loss, we used SnpEff (Cingolani et al., 2012) to annotate the effect of each variant which segregated in at least one of the two bulks. We searched betalain pathway genes for segregating 'high impact' variants and identified missense variants and a stop-gain variant in the *AhCYP76AD2* gene (Fig. 3.4c) that we previously annotated as the most likely functional homolog for this enzyme. While most annotated variants occurred at similar frequencies in both bulks, two missense variants (F to I, T1258A and L to F, A1263T) and the stop gain variant at C1387T, all located on exon 2, differed strongly in relative allele depth between the two bulks (Figs 3.4d, S3.10). The SNP C1387T was monomorph for the reference allele in the red bulk of the biosynthesis BSA (allele depth of 0/24), but the alternative allele occurred at intermediate frequency in the green bulk (allele depth of 16/26). Similarly, the reference alleles of T1258A and A1263T were monomorph in the red bulk but the alternative alleles were found at intermediate frequency (relative allele depth 17/40 and 16/38, respectively) in the green bulk. The QTL included two peaks at 3.2 mb (3,205,818 bp) and 7.4 mb (7,425,588 bp) with a decreased G' value in between, which nonetheless did not go below the significance threshold (Fig. 3.4b). A potential explanation might be the presence of multiple causal variants in the bulks, given their multi-parent origin (see Methods).

Since the RNA-seq data had higher coverage of *AhCYP76AD2* than the whole genome sequencing data, we investigated the presence and absence of the three variants in RNA-

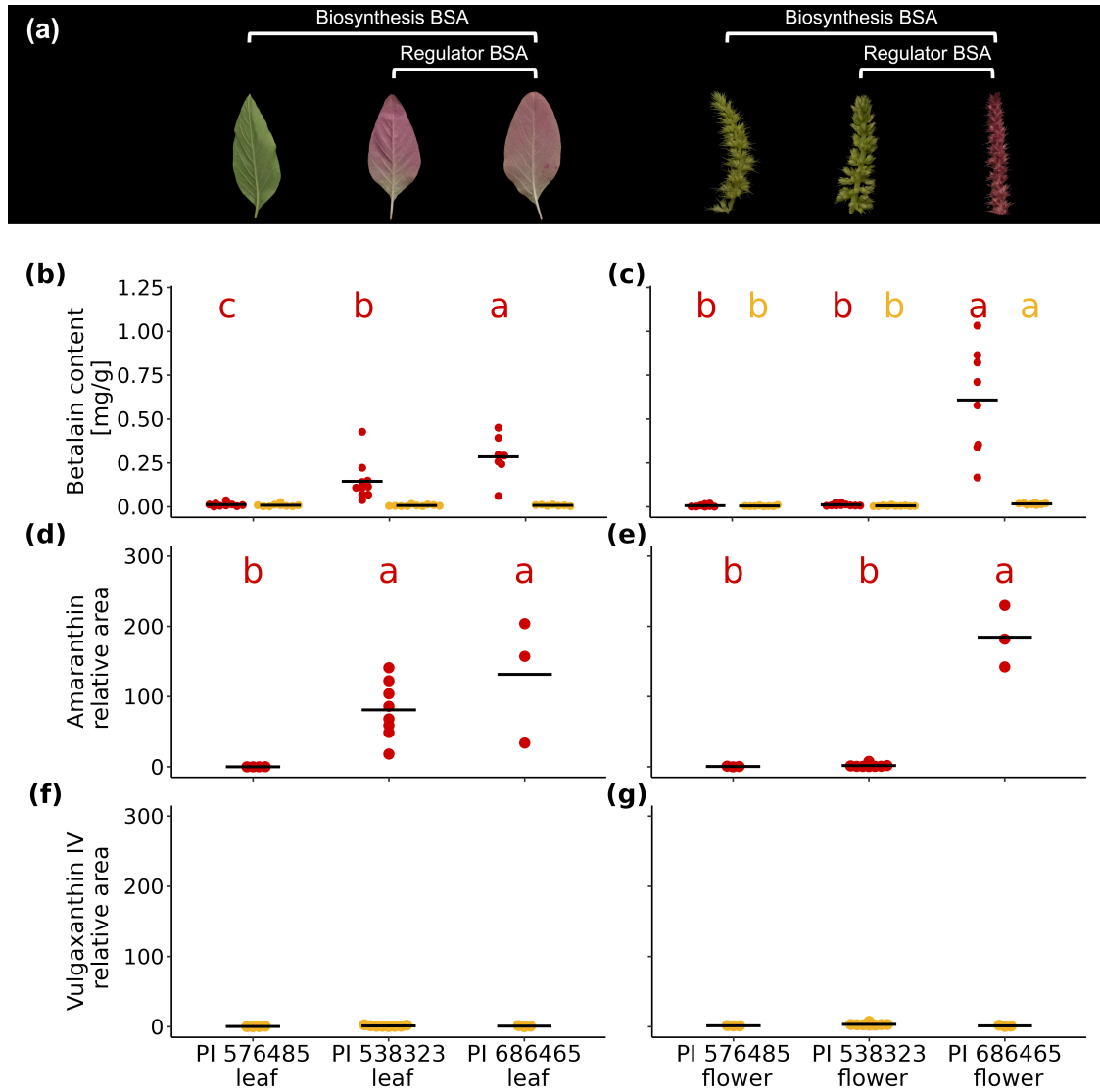


Figure 3.3: Quantification of betalain content for parental *A. hypochondriacus* accessions used in mapping. (a) Representative leaf and flower images for parent plants, with PI 576485 and PI 686465 generating the biosynthesis bulk segregant analysis (BSA) population and PI 538323 and PI 686465 generating the regulator BSA population. (b-c) Photometric quantification of betacyanin (red) and betaxanthin (yellow) content of (b) leaf and (c) flower tissue (mg betalains per g fresh weight), with black bars denoting mean values. (d-e) Relative quantification by LC-MS of amaranthin (betacyanin) content in (d) leaf and (e) flower tissue. (f-g) Relative quantification by LC-MS of vulgaxanthin IV (betaxanthin) content in (f) leaf and (g) flower tissue. Statistical analysis was conducted using ANOVA and significant differences ($p < 0.05$) in betaxanthin or betacyanin content between accessions within tissues were denoted using compact letter display. There were no significant differences in total betaxanthin content between accessions in leaf and no significant differences in vulgaxanthin IV content between accessions in leaf or flower.

seq data from the bulked floral tissue of additional red and green colored pools from the same cross. To resolve *AhCYP76AD2* haplotypes, we considered only read pairs which covered at least one of the non-synonymous variants and the stop variant C1387T. Inspection of read pairs originating from the green bulk revealed that either the alternative alleles of the non-synonymous variants (T1258A and A1263T) or the alternative allele of the stop variant C1387T were found in all read pairs from the green bulk (Fig. 3.4e). The presence of alternative alleles at T1258A and A1263T and at C1387T was almost mutually exclusive in green plants, with only two reads containing alternative alleles at C1387T and one of the missense variants. In contrast to this, less than half of all read pairs covering the *AhCYP76AD2* variants in the red bulk featured either alternative allele at T1258A, A1263T and C1387T (Fig. S3.11), consistent with heterozygous allele frequencies at these variants. Sanger sequencing of four individuals from the green bulk and one individual from the red bulk confirmed the presence of mutually exclusive alleles in the sampled green bulk (Fig. S3.12). Together, this suggests that we mapped two functional haplotypes that both disrupt the first enzyme in the betalain pathway, leading to loss of red betacyanins in the whole plant.

Flower-specific regulation of betalain enzymes by *AhMYB2*

We further intended to identify tissue-specific regulators of betalain pigmentation in amaranth. Therefore, we conducted a second BSA from a different initial cross and pooled 69 plants based on red or green flower color (Fig. 3.3a). All plants from both pools had red leaves and hence could produce betalains (Fig. 3.3). In the regulator BSA, we identified a QTL for flower color at the beginning of Scaffold 16 up to 5.2 mb (1-5,244,109 bp; Fig. 3.5a,b). We found the newly annotated betalain pathway regulator candidate *AhMYB2* to be located at the center of the QTL (Fig. 3.5b). While *AhCYP76AD2* and *AhDODAα1* were also covered by the QTL, we did not find large effect mutations differing between pools and both pools did synthesise red betalains in leaves (Fig. 3.3). We did not identify any 'high impact' mutations in *AhMYB2*, but whole RNA sequencing of flower tissue from pools of diverging flower color individuals revealed a strongly reduced gene expression of both annotated *AhMYB2* isoforms in green flowers compared to red flowers (Fig. 3.5c). This indicates that the loss of *AhMYB2* expression is associated with the loss of flower color. Comparison of betalain pathway enzyme gene expression revealed reduced expression levels of *AhCYP76AD2* and *AhDODAα1* in green flowers (Figs 3.5c, S3.9). Notably, we did not detect expression of *AhCYP76AD5* and downstream betalain pathway gene expression did not differ between red and green flowers, indicating that reduced *AhCYP76AD2* and *AhDODAα1* gene expression caused betalain loss in green flowers (Figs 3.5c, 3.1b). We found similar expression of flavonoid pathway genes between red and green flowers with low expression of DFR (dihydroflavonol 4-reductase) possibly limiting the accumulation of later flavonoid compounds (Fig. S3.13). Given the differ-

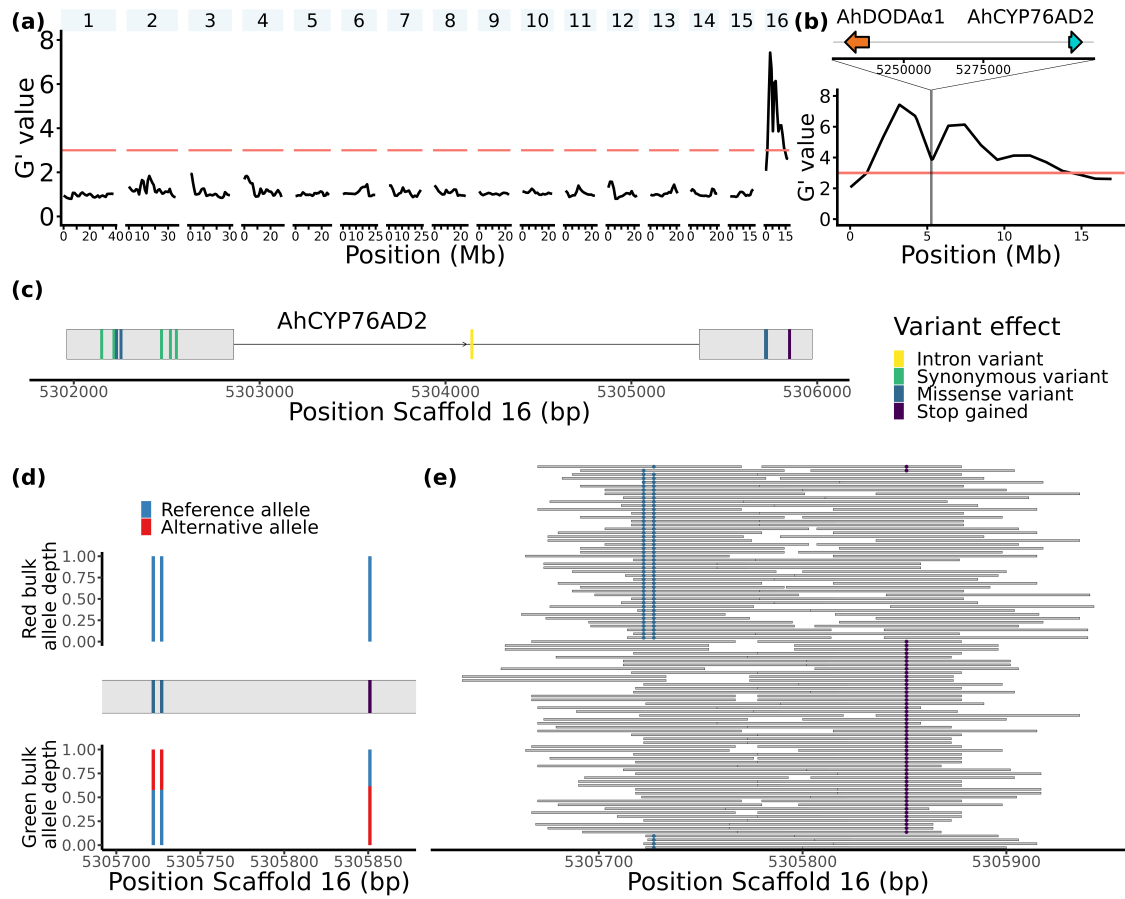


Figure 3.4: Biosynthesis mapping identifies functional variants in *AhCYP76AD2*. (a-b) Biosynthesis bulk segregant analysis (BSA) results (a) genome-wide and (b) on Scaffold 16 show QTL region on *A. hypochondriacus* Scaffold 16. The significance threshold was marked by red horizontal lines, vertical lines denote the location of *AhDODAα1* (cyan) and *AhCYP76AD2* (orange) on the QTL. (c) Variants segregating in whole-genome sequencing data from red and green bulk are displayed with assigned effect of the alternative allele at their position in the exon/intron structure of *AhCYP76AD2*. (d) Position of T1258A and A1263T (missense, dark blue) and C1387T (stop, purple) variants in the second exon of *AhCYP76AD2* and their relative allele depth in the red and green bulk of the biosynthesis BSA. (e) All RNA sequencing read pairs from green bulk flower tissue of the biosynthesis BSA covering both at least one missense and the stop variant, with marked occurrences of T1258A, A1263T and C1387T alternative alleles.

ence in flower color between pools in the regulator BSA (Fig. 3.3), strong expression of *AhMYB2* only in flower tissue (Fig. S3.7) and the correlation of its expression with that of two betalain pathway genes, *AhMYB2* likely represents a flower-specific regulator of the betalain pathway (Fig. 3.5c).

The protein sequences of the *AhMYB2* isoforms *AhMYB2.1* and *AhMYB2.2* differed only by four additional residues in *AhMYB2.2*, stemming from an earlier start of exon 3 (Fig. 3.5d). The additional residues in *AhMYB2.2* disrupt the conserved R3 domain and the bHLH-interaction motif, which might impair the function of this isoform. To functionally validate *AhMYB2* as regulator of the betalain pathway, we overexpressed both isoforms in transgenic hairy roots in amaranth and investigated their cellular localisation in leek (Fig. 3.6a-g). As expected for transcription factors, both isoforms were localised in the nucleus (Fig. 3.6c,f). In contrast to *AhMYB2.2*, overexpression of *AhMYB2.1* resulted in red coloration of emerging roots, but never in the tip of the root (Fig. 3.6a-d). Transgenic roots showed increased amaranthin accumulation compared to wild-type like roots, validating the suggested involvement of *AhMYB2.1* in betalain pathway regulation (Fig. S3.14 S14). Roots accumulated betalains within 2 days in new tissue, while GFP signal was visible immediately and along the whole transgenic roots (Fig. 3.6d). In *AhMYB2.1*, we identified a slightly altered version of the bHLH-interaction motif compared to that found in anthocyanin regulators, in which a conserved leucine residue was replaced by isoleucine ([D/E]Lx₂[R/K]x₃Lx₆Ix₃R; Fig. 3.5d). Presence of the motif differentiates *AhMYB2.1* from other BvMYB1-clade proteins, in which the first residues are not fully conserved (Fig. 3.5d). Apart from the conserved motif, *AhMYB2.1* diverged from *A. thaliana* anthocyanin regulators at several functionally important sites in the R3 domain (alignment positions 88, 89, 95, 101, 114 and 128, Fig. 3.5d, methods S3.5). To assess the effect of amino acid differences between the *A. thaliana* anthocyanin regulator PAP1 (AtMYB75), the betalain regulators *AhMYB2.1* and BvMYB1, and a mutated BvMYB1 (mutBvMYB1) with restored ability to interact with EGL3 (Hattestad et al., 2015) on interaction capacity with heterologous bHLH proteins, we performed pairwise prediction of protein folding and protein interactions of the MYBs with the known *A. thaliana* PAP1 bHLH interaction partners EGL3, GL3 and TT8 (Zimmermann et al., 2004) using CollabFold (Mirdita et al., 2022, Fig. S3.15, Methods S3.5). For all MYBs, interacting residues were mainly predicted in the R3 MYB domain described to mediate the bHLH-interaction (Zimmermann et al., 2004). For the interaction prediction of EGL3 and GL3 with PAP1, *AhMYB2.1*, and mutBvMYB1 we found the prediction of π -cation interactions to EGL3 residue F124 and GL3 residue F128 and hydrogen bonds to EGL3 residue S123 and GL3 residue S127, which might stabilize an interaction and which were not predicted for BvMYB1 (Fig. S3.15). For the MYB-TT8 prediction, we found the prediction of π -cation interactions to TT8 residue F123 in PAP1, *AhMYB2.1* and mutBvMYB1, whereas BvMYB1 was predicted with ionic interactions not predicted in the other MYB genes. Presence of the altered bHLH-interaction motif in *AhMYB2.1* and

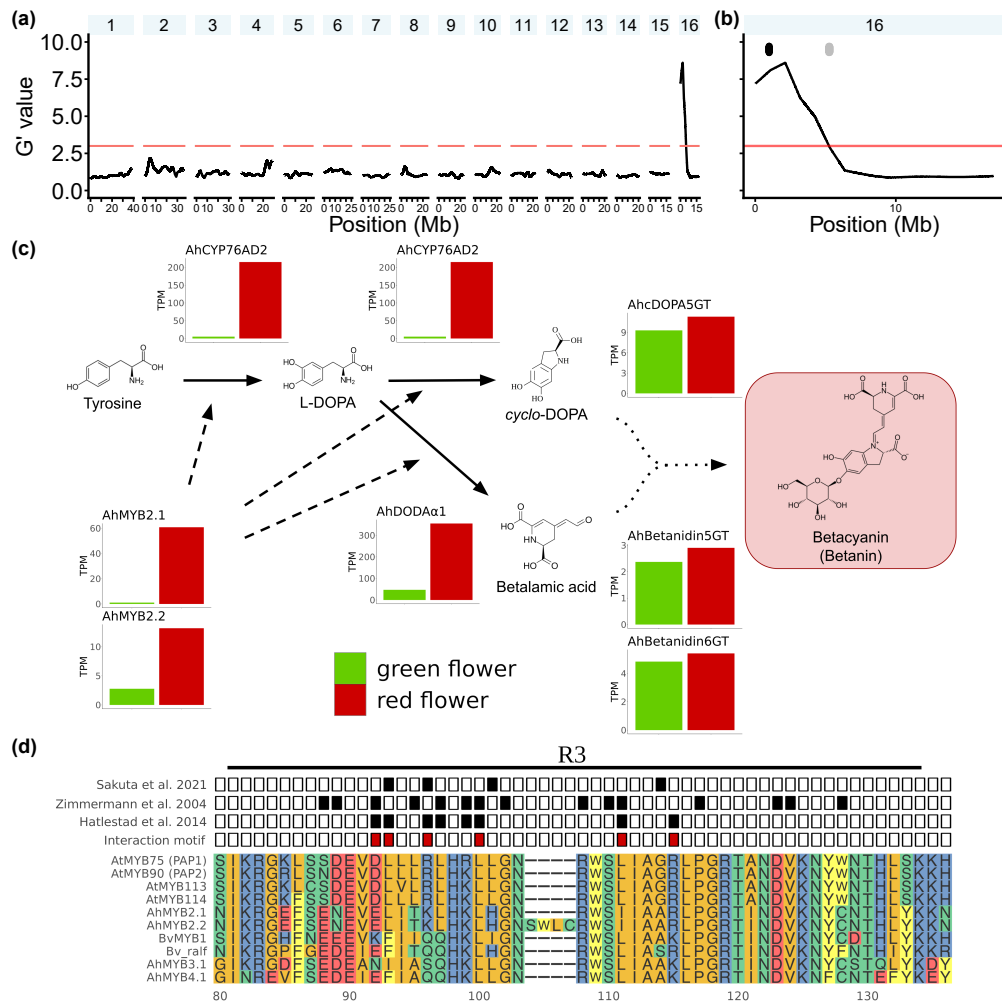


Figure 3.5: Regulator mapping identifies the *A. hypochondriacus* flower color regulator *AhMYB2*. (a-b) Regulator bulk segregant analysis (BSA) results (a) genome-wide and (b) for Scaffold 16 only show the identified QTL region (1-5,244,109 bp) on Scaffold 16. The significance threshold was marked by red horizontal lines. The QTL region includes the betalain pathway genes *AhCYP76AD2* and *AhDODAα1* (grey) as well as the regulator *AhMYB2* (black). (c) Proposed model of regulation of early betalain pathway gene expression by *AhMYB2* in a simplified pathway depiction alongside relative gene expression in transcript per million (TPM) in flower tissue for betalain pathway genes in green and red flower bulk of the regulator BSA. Multiple enzymatic reactions and paths leading to red betacyanin production were simplified with dotted arrows. The structure of betacyanin was displayed as representative betacyanin. The potential betalain pathway genes *AhMYB3*, *AhMYB4* and *AhCYP76AD5* were not expressed. (d) Alignment of the R3 domain of *A. thaliana* anthocyanin regulators with BvMYB1-clade proteins. Above the alignment, the following functional residues were marked: residues of the bHLH-interaction motif as defined by Zimmermann et al. (2004) ([D/E]Lx₂[R/K]x₃Lx₆Lx₃R) (red), BvMYB1 residues required to restore interaction with heterologous bHLH-proteins (Hatlestad et al., 2015), PAP1 residues where amino acid exchanges disturbed bHLH interaction (Zimmermann et al., 2004) and Caryophyllales MYB residues where amino acid exchanges restored transactivation of the *ANS* promoter (Sakuta et al., 2021). Residue 111 of the alignment constitutes the imperfectly conserved interaction motif residue in *AhMYB2.1*.

similarity of predicted interaction residues to the known bHLH-interactor PAP1 could indicate retained ability of AhMYB2.1 to interact with bHLH proteins.

3.5 Discussion

3.5.1 High quality reference genome annotation for functional genomics

We improved the available reference genome of *A. hypochondriacus* (Lightfoot et al., 2017) by genome polishing and subsequent genome annotation. While the *A. hypochondriacus* assembly was highly contiguous with 16 chromosome-scale Scaffolds, even single nucleotide insertions or deletions in coding sequences could have a large impact on gene prediction through frame shifts. The increased completeness of our genome annotation v2.2 compared to the previous annotation likely stems from the removal of a large number of small insertions and deletions in the previous assembly, which impaired protein prediction. Higher similarity of long-read transcript sequences to the polished reference genome and higher overall annotation completeness after genome polishing demonstrates the successful removal of small errors in the assembly (Table S3.8). By using the polished genome for genome annotation and adding additional full-length transcript sequencing data, we were able to produce the most complete genome annotation of amaranth to date (Clouse et al., 2016; Lightfoot et al., 2017; Ma et al., 2021; Wang et al., 2022) and included the annotation of multiple isoforms for 3,484 genes (Table 3.1).

Alternative splicing is involved in regulatory processes and greatly expands on the coding capacity of eukaryotes through the production of alternative isoforms (Wang et al., 2016; Kornblihtt et al., 2013). Our full-length transcript sequencing data allowed for accurate annotation of different splice variants in grain amaranth. These alternative splice variants will enable future functional studies on isoform-specific gene expression, as shown for the newly annotated *AhMYB2* gene for which we found one stronger expressed, and functional splice variant (Figs 3.5c, 3.6). This splicing variation potentially allows for additional functions or differential regulation of the gene in other organs (Huang et al., 2022; Kornblihtt et al., 2013). While we were able to annotate multiple isoforms for over three thousand genes, additional isoform sequencing in further tissues and accessions will be needed to represent the complete splicing variation in the species.

The reported completeness of the new genome annotation is comparable in accuracy to well annotated model plant genomes (Dohm et al., 2014). It represents a valuable resource that will allow more accurate studies of regulatory processes and support functional genetics and future breeding efforts in amaranth.

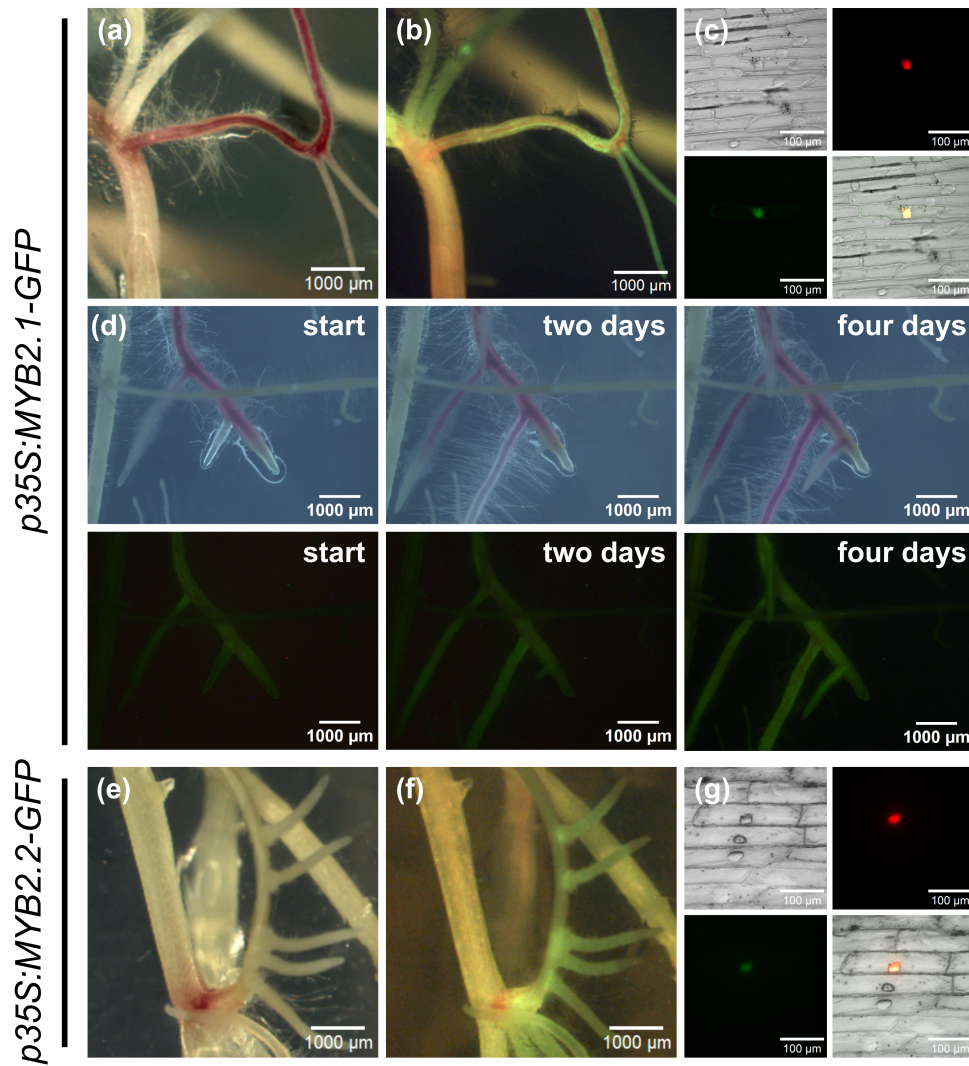


Figure 3.6: Functional validation of *AhMYB2* as betalain regulator. (a-c) Over-expression of *AhMYB2.1* in *A. hypochondriacus* hairy roots, (a) bright-field and (b) GFP-channel show GFP-tag signal and red pigmentation phenotype. (c) Nuclear localisation of *p35S:AhMYB2.1-GFP* in leek epidermins after co-bombardment with *p35S:NLS-mCherry*. (d) Time series images of growth of transgenic roots expressing *p35S:AhMYB2.1-GFP* in bright-field (top row) and GFP channels (bottom row). Images of growing roots were taken in two day intervals. (e-g) Over-expression of *AhMYB2.2* in *A. hypochondriacus* hairy roots, (e) bright-field and (f) GFP-channel show GFP-tag signal but no red coloration. (g) Nuclear localisation of *p35S:AhMYB2.2-GFP* in leek epidermins after co-bombardment with *p35S:NLS-mCherry*. (c, g) Subplots show bright-field (top left), mCherry (top right), GFP (bottom left) and merged (bottom right) channels.

3.5.2 Conservation of the flavonoid and betalain pigment pathways in amaranth

While betalains and anthocyanins appear to be mutually exclusive, the flavonoid pathway upstream of anthocyanin production is reported to be intact in betalain producing species and loss or downregulation of only specific anthocyanin genes appears to have led to the loss of anthocyanins in the Caryophyllales (Sakuta et al., 2021; Pucker et al., 2024). To study the conservation of the flavonoid pathway in amaranth, we identified and annotated genes of the general phenylpropanoid and flavonoid pathway using our improved genome annotation. We found no close homolog of *FNS1*, corresponding to previous reports of *FNS1* homologs being frequently absent outside of Apiaceae species (Pucker et al., 2020; Gebhardt et al., 2007). The amaranth genome did not contain a *F3'5'H* homolog, which is absent in various plant species (Table S3.11; Wessinger and Rausher, 2012). The low similarity of the annotated *ANS* candidate gene to sequences from other species and the unannotated locus with *ANR* similarity might necessitate molecular validation in the future. Still, we found the flavonoid pathway to be generally conserved in amaranth and annotated candidate genes for the key enzymatic reactions.

While the genes of the betalain pathway have been identified in different amaranth species in previous studies (Chang et al., 2021; Zheng et al., 2016; Casique-Arroyo et al., 2014), no comprehensive annotation of the pathway in *A. hypochondriacus* was available that would enable candidate gene association in mapping studies (Table S3.3). The characterization of functional betalain genes was complicated by multiple ancient gene duplications in the Caryophyllales, which gave rise to specific CYP76AD and DODA lineages involved in betalain production (Brockington et al., 2015; Sheehan et al., 2020). Phylogenetic analysis showed that *AhCYP76AD2* and *AhDODA α 1* belong to their respective α -clades and are likely functioning in betalain biosynthesis (Figs S3.3, S3.4; Sunnadeniya et al., 2016; Brockington et al., 2015). In contrast to *AhCYP76AD2*, *AhCYP76AD5* from the CYP76AD β -clade is likely only capable to catalyze the conversion of L-DOPA to cyclo-DOPA (Brockington et al., 2015). Similar to previous reports, we found *AhCYP76AD2* and *AhDODA α 1* to be co-localized on a conserved betalain locus on Scaffold 16 (Fig. 3.4b; Lightfoot et al., 2017; Ma et al., 2021). The biosynthesis BSA identified this locus to be associated with red plant pigmentation (Fig. 3.4), further supporting the involvement of the locus in betacyanin biosynthesis. We identified mutually exclusive alleles in the *AhCYP76AD2* gene, associated with complete loss of red tissue color (Fig. 3.4). Studies on loss of anthocyanin mediated flower color show that, while functional (loss-of-function in pathway genes) and regulatory mutations (leading to e.g. gene downregulation in specific tissues) occur at comparable frequencies, regulatory mutations leading to color loss appear to be preferentially fixed in natural populations, likely because of pleiotropic effects of pathway genes in vegetative tissue (Wessinger and Rausher, 2012; Gates et al., 2018). In our population, betalain pathway genes showed

even higher expression in the green bulk compared to the red bulk (Fig. S3.9). Since betalain genes were still expressed in green tissue, causal mutations for the loss of tissue color likely led to a knock-out of *AhCYP76AD2* (Fig. 3.4). While the knock-out allele at C1387T shows high evidence to lead to a loss of *AhCYP76AD2* function and the loss of betacyanin biosynthesis, the missense alleles at T1258A and A1263T might lead to the loss of function by changed protein folding or through the replacement of functionally important residues (Fig. 3.4c,d). The presence of independent functional mutations and the complete lack of red color might indicate dispensability of betacyanin production, at least in individual populations. Studies on the fitness effect of betacyanin loss and the prevalence of knock-out mutations and their possible pleiotropic effects on plant fitness in natural amaranth populations are currently lacking.

We identified conserved patterns in the betalain and flavonoid pathways through the annotation of pathway genes in the improved gene atlas. Loss of anthocyanin production in amaranth could result from loss-of-function in anthocyanin biosynthesis or transport. Especially late flavonoid pathway genes like *ANS* represent targets for functional validation in future studies. However, even a potential loss-of-function in anthocyanin genes does not have to be causal for mutual exclusiveness, since e.g., previous downregulation or replacement of anthocyanin genes would remove selective constraints to maintain the genes. Since other flavonoids can still be synthesised in betalain producing species (Iwashina, 2015), the emergence of betalain biosynthesis alone is likely not sufficient to explain total replacement of anthocyanins and the observed mutual exclusiveness of the pigments across the large phylogenetic range of the Caryophyllales. To understand the subsequent order of the rise of mutual exclusiveness between betalains and anthocyanins, it might be important to understand the regulation of both pathways in the same study system.

3.5.3 *AhMYB2* is a flower-specific regulator of the betalain pathway

Despite the characterisation of the betalain biosynthesis pathway in recent years (Hatlestad et al., 2012; Vogt et al., 1999; Christinet et al., 2004) and the identification of few regulators (Xie et al., 2023; Zhang et al., 2021; Zeng et al., 2023; Hatlestad et al., 2015), knowledge about tissue-specific regulation of the pathway is limited. We identified the amaranth flower color regulator *AhMYB2* in our regulator mapping population (Fig. 3.5). Strongly reduced gene expression of *AhMYB2*, as well as of the betalain pathway genes *AhCYP76AD2* and *AhDODAα1* in green flowered plants of the regulator BSA (Fig. 3.5c) leads us to propose *AhMYB2* as transcriptional regulator of the early betalain pathway. While we annotated two different *AhMYB2* isoforms and both showed reduced expression in green flower in the regulator BSA, an insertion in the conserved MYB domain leaves only *AhMYB2.1* capable of inducing betalain accumulation (Figs 3.5d, 3.6). Flavonoid

pathway regulating MYB transcription factors commonly control tissue-specific pigment accumulation and frequently determine specificity of the MBW complex (Rajput et al., 2022; Lin-Wang et al., 2010; Xu et al., 2015). While *BvMYB1* was shown to regulate the betalain pathway in *B. vulgaris* leaves (Hatlestad et al., 2015) its ortholog *AhMYB2* regulates betalain biosynthesis in flower, instead. Similar to the tissue-specific regulation by flavonoid pathway MYBs, other BvMYB1-clade proteins might represent regulators in other tissues. While we did not observe high expression for *AhMYB3* and *AhMYB4* in any sampled tissue (Fig. S3.7 S7), the genes might be active in different conditions or genotypes and should be targets of characterization in future studies. The identification of *AhMYB2* as flower-specific color regulator improves the understanding of the tissue-specific regulation of the early betalain pigment pathway.

The flavonoid pathway in plants is regulated by conserved subgroups of MYB transcription factors (Gonzalez et al., 2008; Lloyd et al., 2017; Liu et al., 2015). We found the conservation of phenylpropanoid pathway (S4), early flavonoid pathway (S7) and proanthocyanidin (S5 and AtMYB5-like) MYB regulators in amaranth, in accordance with the conservation of the flavonoid pathway enzymes (Fig. 3.2, Table S3.12). Clustering of AhMYB2 with the betalain regulator BvMYB1 next to anthocyanin MYBs (S6) supports the proposed neofunctionalization of former anthocyanin regulators for betalain pathway regulation (Figs 3.2, 3.5, Table S3.12; Hatlestad et al., 2015; Timoneda et al., 2019). While MBW complex interaction is conserved in S6 anthocyanin regulators across species (Zimmermann et al., 2004; Xu et al., 2015), the ability to interact with bHLH proteins in the MBW complex might have been lost in the betalain regulator BvMYB1 (Hatlestad et al., 2015). In contrast to other *A. hypochondriacus* and *B. vulgaris* BvMYB1-clade MYBs (Hatlestad et al., 2015), AhMYB2.1 includes an altered version of the conserved bHLH-interaction motif in the R3 domain (Fig. 3.5d) and we predicted π -cation interactions to bHLH residues shared among AhMYB2.1 and the known bHLH-interactors PAP1 and mutBvMYB1 but not BvMYB1 (Fig. S3.15), which might indicate retained potential for bHLH-interaction in AhMYB2.1. While overexpression of AhMYB2.1 alone was sufficient to induce red color accumulation in the transformed roots, expression of potential interaction partners in the root cannot be excluded. However, the potential bHLH-interactions and their potential functional relevance for AhMYB2.1-mediated betalain accumulation needs further investigation, especially when considering reports of loss of the EGL3/GL3 bHLH-lineage in the Amaranthaceae (Pucker et al., 2024). Further characterization of AhMYB2.1 in future studies could help to elucidate its capacity to interact with heterologous or native bHLH proteins.

3.5.4 The rise of mutually exclusive pigment pathways

Within the Caryophyllales, betalain producing clades are intermixed with anthocyanin producing lineages, leading recent studies to propose multiple evolutionary origins of be-

talains (Sheehan et al., 2020; Timoneda et al., 2019). Despite this polyphyletic origin of betalain production, the biosynthesis of betalains and anthocyanins appears to be mutual exclusive in all plants (Clement and Mabry, 1996). A deregulated form of the arogenate dehydrogenase gene has been suggested to have enabled the shift from anthocyanins to betalains, by increasing carbon flux of the common substrate arogenate towards the betalain precursor tyrosine and thereby decreasing the availability of phenylalanine for flavonoid (including anthocyanin) production (Lopez-Nieves et al., 2018; Timoneda et al., 2019). Yet, changes in substrate availability cannot sufficiently explain the mutual exclusiveness of the two pigments, since many other flavonoids are still readily synthesised in betalain producing species (Iwashina, 2015). Both, expression differences in anthocyanin biosynthesis genes and loss of anthocyanin transport and flavonoid pathway genes contribute to prevent anthocyanin production in betalain species (Shimada et al., 2007; Pucker et al., 2024). Such regulatory or functional changes of anthocyanin genes can explain how simultaneous production of both betalains and anthocyanins is never observed, but not the evolutionary reason for the alternation. Step-wise models of the shift from anthocyanin to betalain biosynthesis commonly include intermediate phases with both anthocyanin and betalain production and only later loss of the functionally equivalent anthocyanins (Timoneda et al., 2019). A potential key to explain the mutual exclusiveness might be to focus instead on the common components utilized by both pathways, which includes the MYB regulators (Hatlestad et al., 2015). Previous studies found orthologs of anthocyanin MYBs in betalain producing species unable to activate expression of genes required for anthocyanin biosynthesis (Shimada et al., 2007; Sakuta et al., 2021). The co-option of MBW complex interacting anthocyanin MYBs into betalain regulation, followed by loss of key residues for bHLH-interaction could have relieved anthocyanin genes of their key transcriptional regulators (Timoneda et al., 2019; Hatlestad et al., 2015; Sakuta et al., 2021). In this context, the altered bHLH-interaction motif identified in AhMYB2 could represent an interesting intermediate step in the proposed regulator shift. Further characterization of *AhMYB2* and other betalain regulators might help to understand the co-option of former anthocyanin regulators into the regulation of the novel betalain pigment pathway, the molecular prerequisites for this change and the mutual exclusiveness of the two pathways.

3.6 Acknowledgments

We thank David M. Brenner from Iowa State University and a Curator in the US National Plant Germplasm System for providing F₂ mapping populations for pigmentation gene identification. We thank Roswitha Lentz for help with DNA & RNA extraction and Kerstin Becker and Karl Koehrer from the Genomics & Transcriptomics Labor (BMFZ) at HHU for help with the PacBio sequencing. Analysis was conducted on the Regional

Computing Center of the University of Cologne (RRZK) High Performance Computing (HPC) system CHEOPS (Funding number: INST 216/512/1FUGG). We thank Jessica Pietsch, Hanna Bechtel and Swen Schellmann for the help with the bombardment and advice on the microscopy and Jathish Ponnu for providing the *p35S:mCherry-NLS* vector. We acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2048/1 – Project ID 390686111 and STE 2654/4. TSW was supported by the University of Cologne Forum for Plant Biology and Politics of Nutrition (BiPoN).

3.7 Author contributions

TSW and MGS planned and designed the study. TSW and MGS performed experiments and analysed the data. SKV amplified and sequenced *AhCYP76AD2* haplotypes. SKV cloned and overexpressed AhMYB2. NDR and SM performed LC-MS analysis. TSW, NDR and SKV prepared the figures. TSW and MGS wrote the manuscript. All authors read and approved the manuscript.

3.8 Competing interests

None declared.

3.9 Data availability

All sequencing data generated in this study are deposited under BioProject ID PRJEB65083. Code used for the study is available at https://github.com/T-Winkler/Ahyp_v2_2_publication.git. The amaranth genome and annotation v2.2 can be accessed through CoGe under Genome id66602 <https://genomevolution.org/coge/GenomeInfo.pl?gid=66602>. All files, including functional annotations, can be accessed at <https://doi.org/10.6084/m9.figshare.23946765>.

3.10 ORCID

Tom S. Winkler: <https://orcid.org/0009-0008-2483-8488>

Susanne K. Vollmer: <https://orcid.org/0009-0005-8643-1470>

Nadine Dyballa-Rukes: <https://orcid.org/0000-0002-9712-8937>

Sabine Metzger: <https://orcid.org/0000-0002-9722-8495>

Markus G Stetter: <https://orcid.org/0000-0001-7136-0589>

3.11 References

- S. Altschul. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–3402, Sept. 1997. ISSN 13624962. doi: 10.1093/nar/25.17.3389. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/25.17.3389>.
- S. Andrews. FastQC: A Quality Control Tool for High Throughput Sequence Data, 2010. URL <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- A. Bean, R. Sunnadeniya, N. Akhavan, A. Campbell, M. Brown, and A. Lloyd. Gain-of-function mutations in beet *DODA2* identify key residues for betalain pigment evolution. *New Phytologist*, 219(1):287–296, July 2018. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.15159. URL <https://onlinelibrary.wiley.com/doi/10.1111/nph.15159>.
- A. M. Bolger, M. Lohse, and B. Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120, Aug. 2014. ISSN 1367-4811, 1367-4803. doi: 10.1093/bioinformatics/btu170. URL <https://academic.oup.com/bioinformatics/article/30/15/2114/2390096>.
- N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34(5):525–527, May 2016. ISSN 1087-0156, 1546-1696. doi: 10.1038/nbt.3519. URL <https://www.nature.com/articles/nbt.3519>.
- C. Brendolise, R. V. Espley, K. Lin-Wang, W. Laing, Y. Peng, T. McGhie, S. Dejnopratt, S. Tomes, R. P. Hellens, and A. C. Allan. Multiple Copies of a Simple MYB-Binding Site Confers Trans-regulation by Specific Flavonoid-Related R2R3 MYBs in Diverse Species. *Frontiers in Plant Science*, 8:1864, Oct. 2017. ISSN 1664-462X. doi: 10.3389/fpls.2017.01864. URL <http://journal.frontiersin.org/article/10.3389/fpls.2017.01864/full>.
- D. M. Brenner. Registration of DB 199313, Cytoplasmic Male Sterile Grain Amaranth Genetic Stock. *Journal of Plant Registrations*, 13(2):251–253, May 2019. ISSN 1936-5209, 1940-3496. doi: 10.3198/jpr2018.06.0042crgs. URL <https://access.onlinelibrary.wiley.com/doi/10.3198/jpr2018.06.0042crgs>.
- S. F. Brockington, R. H. Walker, B. J. Glover, P. S. Soltis, and D. E. Soltis. Complex pigment evolution in the *Caryophyllales*. *New Phytologist*, 190(4):854–864, June 2011. ISSN 0028-646X, 1469-8137. doi: 10.1111/j.1469-8137.2011.03687.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.2011.03687.x>.
- S. F. Brockington, R. Alvarez-Fernandez, J. B. Landis, K. Alcorn, R. H. Walker, M. M. Thomas, L. C. Hileman, and B. J. Glover. Evolutionary Analysis of the MIXTA Gene Family Highlights Potential Targets for the Study of Cellular Differentiation. *Molecular Biology and Evolution*, 30(3):526–540, Mar. 2013. ISSN 1537-1719, 0737-4038. doi: 10.1093/molbev/mss260. URL <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mss260>.
- S. F. Brockington, Y. Yang, F. Gandia-Herrero, S. Covshoff, J. M. Hibberd, R. F. Sage, G. K. S. Wong, M. J. Moore, and S. A. Smith. Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in *Caryophyllales*. *New Phytologist*, 207

- (4):1170–1180, Sept. 2015. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.13441. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.13441>.
- T. Brůna, K. J. Hoff, A. Lomsadze, M. Stanke, and M. Borodovsky. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, 3(1):lqaa108, Jan. 2021. ISSN 2631-9268. doi: 10.1093/nargab/lqaa108. URL <https://academic.oup.com/nargab/article/doi/10.1093/nargab/lqaa108/6066535>.
- C. P. Cantalapiedra, A. Hernández-Plaza, I. Letunic, P. Bork, and J. Huerta-Cepas. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38(12):5825–5829, Dec. 2021. ISSN 1537-1719. doi: 10.1093/molbev/msab293. URL <https://academic.oup.com/mbe/article/38/12/5825/6379734>.
- G. Casique-Arroyo, N. Martínez-Gallardo, L. González de la Vara, and J. P. Délano-Frier. Betacyanin Biosynthetic Genes and Enzymes Are Differentially Induced by (a)biotic Stress in *Amaranthus hypochondriacus*. *PLoS ONE*, 9(6):e99012, June 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0099012. URL <https://dx.plos.org/10.1371/journal.pone.0099012>.
- L. Chaney, R. Mangelson, T. Ramaraj, E. N. Jellen, and P. J. Maughan. The complete chloroplast genome sequences for four *Amaranthus* species (*Amaranthaceae*). *Applications in Plant Sciences*, 4(9):1600063, Sept. 2016. ISSN 2168-0450, 2168-0450. doi: 10.3732/apps.1600063. URL <https://bsapubs.onlinelibrary.wiley.com/doi/10.3732/apps.1600063>.
- Y.-C. Chang, Y.-C. Chiu, N.-W. Tsao, Y.-L. Chou, C.-M. Tan, Y.-H. Chiang, P.-C. Liao, Y.-C. Lee, L.-C. Hsieh, S.-Y. Wang, and J.-Y. Yang. Elucidation of the core betalain biosynthesis pathway in *Amaranthus tricolor*. *Scientific Reports*, 11(1):6086, Dec. 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-85486-x. URL <http://www.nature.com/articles/s41598-021-85486-x>.
- J.-y. Chen, F.-f. Xie, Y.-z. Cui, C.-b. Chen, W.-j. Lu, X.-d. Hu, Q.-z. Hua, J. Zhao, Z.-j. Wu, D. Gao, Z.-k. Zhang, W.-k. Jiang, Q.-m. Sun, G.-b. Hu, and Y.-h. Qin. A chromosome-scale genome sequence of pitaya (*Hylocereus undatus*) provides novel insights into the genome evolution and regulation of betalain biosynthesis. *Horticulture Research*, 8(1):164, Dec. 2021. ISSN 2662-6810, 2052-7276. doi: 10.1038/s41438-021-00612-0. URL <https://academic.oup.com/hr/article/6446786>.
- L. Christinet, F. X. Burdet, M. Zaiko, U. Hinz, and J.-P. Zrýd. Characterization and Functional Identification of a Novel Plant 4,5-Extradiol Dioxygenase Involved in Betalain Pigment Biosynthesis in *Portulaca grandiflora*. *Plant Physiology*, 134(1):265–274, Jan. 2004. ISSN 1532-2548, 0032-0889. doi: 10.1104/pp.103.031914. URL <https://academic.oup.com/plphys/article/134/1/265/6112200>.
- P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-3. *Fly*, 6(2):80–92, Apr. 2012. ISSN 1933-6934, 1933-6942. doi: 10.4161/fly.19695. URL <http://www.tandfonline.com/doi/abs/10.4161/fly.19695>.

- J. S. Clement and T. J. Mabry. Pigment Evolution in the *Caryophyllales*: a Systematic Overview*. *Botanica Acta*, 109(5):360–367, Oct. 1996. ISSN 09328629. doi: 10.1111/j.1438-8677.1996.tb00584.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1438-8677.1996.tb00584.x>.
- J. W. Clouse, D. Adhikary, J. T. Page, T. Ramaraj, M. K. Deyholos, J. A. Udall, D. J. Fairbanks, E. N. Jellen, and P. J. Maughan. The Amaranth Genome: Genome, Transcriptome, and Physical Map Assembly. *The Plant Genome*, 9(1):plantgenome2015.07.0062, Mar. 2016. ISSN 1940-3372, 1940-3372. doi: 10.3835/plantgenome2015.07.0062. URL <https://access.onlinelibrary.wiley.com/doi/10.3835/plantgenome2015.07.0062>.
- P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158, Aug. 2011. ISSN 1367-4811, 1367-4803. doi: 10.1093/bioinformatics/btr330. URL <https://academic.oup.com/bioinformatics/article/27/15/2156/402296>.
- Y. Deng and S. Lu. Biosynthesis and Regulation of Phenylpropanoids in Plants. *Critical Reviews in Plant Sciences*, 36(4):257–290, July 2017. ISSN 0735-2689, 1549-7836. doi: 10.1080/07352689.2017.1402852. URL <https://www.tandfonline.com/doi/full/10.1080/07352689.2017.1402852>.
- J. C. Dohm, A. E. Minoche, D. Holtgräwe, S. Capella-Gutiérrez, F. Zakrzewski, H. Tafer, O. Rupp, T. R. Sörensen, R. Stracke, R. Reinhardt, A. Goesmann, T. Kraft, B. Schulz, P. F. Stadler, T. Schmidt, T. Gabaldón, H. Lehrach, B. Weisshaar, and H. Himmelbauer. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature*, 505(7484):546–549, Jan. 2014. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature12817. URL <https://www.nature.com/articles/nature12817>.
- C. Dubos, R. Stracke, E. Grotewold, B. Weisshaar, C. Martin, and L. Lepiniec. MYB transcription factors in *Arabidopsis*. *Trends in Plant Science*, 15(10):573–581, Oct. 2010. ISSN 13601385. doi: 10.1016/j.tplants.2010.06.005. URL <https://linkinghub.elsevier.com/retrieve/pii/S1360138510001536>.
- S. R. Eddy. A new generation of homology search tools based on probabilistic inference. *Genome Informatics. International Conference on Genome Informatics*, 23(1):205–211, Oct. 2009. ISSN 0919-9454.
- J. M. Flynn, R. Hubley, C. Goubert, J. Rosen, A. G. Clark, C. Feschotte, and A. F. Smit. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, 117(17):9451–9457, Apr. 2020. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1921046117. URL <https://pnas.org/doi/full/10.1073/pnas.1921046117>.
- L. Gabriel, K. J. Hoff, T. Brûna, M. Borodovsky, and M. Stanke. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics*, 22(1):566, Dec. 2021. ISSN 1471-2105. doi: 10.1186/s12859-021-04482-0. URL <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-021-04482-0>.

- D. J. Gates, B. J. S. C. Olson, T. E. Clemente, and S. D. Smith. A novel R3 MYB transcriptional repressor associated with the loss of floral pigmentation in *Iochroma*. *New Phytologist*, 217(3):1346–1356, Feb. 2018. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.14830. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.14830>.
- Y. H. Gebhardt, S. Witte, H. Steuber, U. Matern, and S. Martens. Evolution of Flavone Synthase I from Parsley Flavanone 3 β -Hydroxylase by Site-Directed Mutagenesis. *Plant Physiology*, 144(3):1442–1454, July 2007. ISSN 1532-2548. doi: 10.1104/pp.107.098392. URL <https://academic.oup.com/plphys/article/144/3/1442/6106752>.
- A. Gonzalez, M. Zhao, J. M. Leavitt, and A. M. Lloyd. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *The Plant Journal*, 53(5):814–827, Mar. 2008. ISSN 0960-7412, 1365-313X. doi: 10.1111/j.1365-313X.2007.03373.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2007.03373.x>.
- J. Gonçalves-Dias, A. Singh, C. Graf, and M. G. Stetter. Genetic Incompatibilities and Evolutionary Rescue by Wild Relatives Shaped Grain Amaranth Domestication. *Molecular Biology and Evolution*, 40(8):msad177, Aug. 2023. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msad177. URL <https://academic.oup.com/mbe/article/doi/10.1093/molbev/msad177/7239273>.
- G. J. Hatlestad, R. M. Sunnadeniya, N. A. Akhavan, A. Gonzalez, I. L. Goldman, J. M. McGrath, and A. M. Lloyd. The beet R locus encodes a new cytochrome P450 required for red betalain production. *Nature Genetics*, 44(7):816–820, July 2012. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.2297. URL <http://www.nature.com/articles/ng.2297>.
- G. J. Hatlestad, N. A. Akhavan, R. M. Sunnadeniya, L. Elam, S. Cargile, A. Hembd, A. Gonzalez, J. M. McGrath, and A. M. Lloyd. The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway. *Nature Genetics*, 47(1):92–96, Jan. 2015. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3163. URL <http://www.nature.com/articles/ng.3163>.
- R. P. Hellens, E. A. Edwards, N. R. Leyland, S. Bean, and P. M. Mullineaux. pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Molecular Biology*, 42(6):819–832, Apr. 2000. ISSN 0167-4412, 1573-5028. doi: 10.1023/A:1006496308160. URL <https://link.springer.com/10.1023/A:1006496308160>.
- J. Hu, J. Fan, Z. Sun, and S. Liu. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*, 36(7):2253–2255, Apr. 2020. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/btz891. URL <https://academic.oup.com/bioinformatics/article/36/7/2253/5645175>.
- C.-K. Huang, W.-D. Lin, and S.-H. Wu. An improved repertoire of splicing variants and their potential roles in Arabidopsis photomorphogenic development. *Genome Biology*, 23(1):50, Feb. 2022. ISSN 1474-760X. doi: 10.1186/s13059-022-02620-2. URL <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-022-02620-2>.

- T. Iwashina. Flavonoid Properties in Plant Families Synthesizing Betalain Pigments (Review). *Natural Product Communications*, 10(6):1934578X1501000675, June 2015. ISSN 1934-578X, 1555-9475. doi: 10.1177/1934578X1501000675. URL <https://journals.sagepub.com/doi/10.1177/1934578X1501000675>.
- G. Jain and K. S. Gould. Are betalain pigments the functional homologues of anthocyanins in plants? *Environmental and Experimental Botany*, 119:48–53, Nov. 2015. ISSN 00988472. doi: 10.1016/j.envexpbot.2015.06.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S0098847215001112>.
- H. Jin and C. Martin. Multifunctionality and diversity within the plant MYB-gene family. *Plant Molecular Biology*, 41(5):577–585, Nov. 1999. ISSN 0167-4412, 1573-5028. doi: 10.1023/A:1006319732410. URL <https://link.springer.com/10.1023/A:1006319732410>.
- P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, and S. Hunter. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9):1236–1240, May 2014. ISSN 1367-4811, 1367-4803. doi: 10.1093/bioinformatics/btu031. URL <https://academic.oup.com/bioinformatics/article/30/9/1236/237988>.
- Y.-J. Kang, D.-C. Yang, L. Kong, M. Hou, Y.-Q. Meng, L. Wei, and G. Gao. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Research*, 45(W1):W12–W16, July 2017. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkx428. URL <https://academic.oup.com/nar/article/45/W1/W12/3831091>.
- A. Katiyar, S. Smita, S. K. Lenka, R. Rajwanshi, V. Chinnusamy, and K. C. Bansal. Genome-wide classification and expression analysis of MYB transcription factor families in rice and *Arabidopsis*. *BMC Genomics*, 13(1):544, Dec. 2012. ISSN 1471-2164. doi: 10.1186/1471-2164-13-544. URL <https://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-544>.
- P. J. Kersey. Plant genome sequences: past, present, future. *Current Opinion in Plant Biology*, 48:1–8, Apr. 2019. ISSN 13695266. doi: 10.1016/j.pbi.2018.11.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S1369526618300918>.
- S. Kobayashi, M. Ishimaru, K. Hiraoka, and C. Honda. *Myb*-related genes of the Kyoho grape (*Vitis labruscana*) regulate anthocyanin biosynthesis. *Planta*, 215(6):924–933, Oct. 2002. ISSN 0032-0935, 1432-2048. doi: 10.1007/s00425-002-0830-5. URL <http://link.springer.com/10.1007/s00425-002-0830-5>.
- R. E. Koes, F. Quattrocchio, and J. N. M. Mol. The flavonoid biosynthetic pathway in plants: Function and evolution. *BioEssays*, 16(2):123–132, Feb. 1994. ISSN 0265-9247, 1521-1878. doi: 10.1002/bies.950160209. URL <https://onlinelibrary.wiley.com/doi/10.1002/bies.950160209>.
- A. R. Kornblihtt, I. E. Schor, M. Alló, G. Dujardin, E. Petrillo, and M. J. Muñoz. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature Reviews Molecular Cell Biology*, 14(3):153–165, Mar. 2013. ISSN 1471-0072, 1471-0080.

- doi: 10.1038/nrm3525. URL <https://www.nature.com/articles/nrm3525>.
- H. D. Kranz, M. Denekamp, R. Greco, H. Jin, A. Leyva, R. C. Meissner, K. Petroni, A. Urzainqui, M. Bevan, C. Martin, S. Smeekeens, C. Tonelli, J. Paz-Ares, and B. Weisshaar. Towards functional characterisation of the members of the *R2R3-MYB* gene family from *Arabidopsis thaliana*. *The Plant Journal*, 16(2):263–276, Oct. 1998. ISSN 0960-7412, 1365-313X. doi: 10.1046/j.1365-313x.1998.00278.x. URL <https://onlinelibrary.wiley.com/doi/10.1046/j.1365-313x.1998.00278.x>.
- T. Kubo, S. Nishizawa, A. Sugawara, N. Itchoda, A. Estiati, and T. Mikami. The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNACys(GCA). *Nucleic Acids Research*, 28(13):2571–2576, July 2000. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/28.13.2571. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/28.13.2571>.
- Y. Lai, H. Li, and M. Yamagishi. A review of target gene specificity of flavonoid R2R3-MYB transcription factors and a discussion of factors contributing to the target gene selectivity. *Frontiers in Biology*, 8(6):577–598, Dec. 2013. ISSN 1674-7984, 1674-7992. doi: 10.1007/s11515-013-1281-z. URL <http://link.springer.com/10.1007/s11515-013-1281-z>.
- A. Lampropoulos, Z. Sutikovic, C. Wenzl, I. Maegele, J. U. Lohmann, and J. Forner. GreenGate - A Novel, Versatile, and Efficient Cloning System for Plant Transgenesis. *PLoS ONE*, 8(12):e83043, Dec. 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0083043. URL <https://dx.plos.org/10.1371/journal.pone.0083043>.
- H. Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, 2013. URL <https://arxiv.org/abs/1303.3997>. Version Number: 2.
- H. Li. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18):3094–3100, Sept. 2018. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/bty191. URL <https://academic.oup.com/bioinformatics/article/34/18/3094/4994778>.
- S. Li. Transcriptional control of flavonoid biosynthesis: Fine-tuning of the MYB-bHLH-WD40 (MBW) complex. *Plant Signaling & Behavior*, 9(1):e27522, Jan. 2014. ISSN 1559-2324. doi: 10.4161/psb.27522. URL <http://www.tandfonline.com/doi/abs/10.4161/psb.27522>.
- D. J. Lightfoot, D. E. Jarvis, T. Ramaraj, R. Lee, E. N. Jellen, and P. J. Maughan. Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biology*, 15(1):74, Dec. 2017. ISSN 1741-7007. doi: 10.1186/s12915-017-0412-4. URL <http://bmcbiol.biomedcentral.com/articles/10.1186/s12915-017-0412-4>.
- K. Lin-Wang, K. Bolitho, K. Grafton, A. Kortstee, S. Karunairetnam, T. K. McGhie, R. V. Espley, R. P. Hellens, and A. C. Allan. An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in *Rosaceae*. *BMC Plant Biology*, 10(1):50, Dec. 2010. ISSN 1471-2229. doi: 10.1186/1471-2229-10-50. URL <https://bmcpplantbiol.biomedcentral.com/articles/10.1186/1471-2229-10-50>.
- J. Liu, A. Osbourn, and P. Ma. MYB Transcription Factors as Regulators of Phenyl-

- propanoid Metabolism in Plants. *Molecular Plant*, 8(5):689–708, May 2015. ISSN 16742052. doi: 10.1016/j.molp.2015.03.012. URL <https://linkinghub.elsevier.com/retrieve/pii/S167420521500180X>.
- A. Lloyd, A. Brockman, L. Aguirre, A. Campbell, A. Bean, A. Cantero, and A. Gonzalez. 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(9):1431–1441, Sept. 2017. ISSN 0032-0781, 1471-9053. doi: 10.1093/pcp/pcx075. URL <https://academic.oup.com/pcp/article/58/9/1431/3858365>.
- S. Lopez-Nieves, Y. Yang, A. Timoneda, M. Wang, T. Feng, S. A. Smith, S. F. Brockington, and H. A. Maeda. Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in *Caryophyllales*. *New Phytologist*, 217(2):896–908, Jan. 2018. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.14822. URL <https://onlinelibrary.wiley.com/doi/10.1111/nph.14822>.
- F. Lota, S. Wegmüller, B. Buer, S. Sato, A. Bräutigam, B. Hanf, and M. Bucher. The *cis*-acting CTTC-P1BS module is indicative for gene function of *LjVTI12*, a Qb-SNARE protein gene that is required for arbuscule formation in *Lotus japonicus*. *The Plant Journal*, 74(2):280–293, Apr. 2013. ISSN 09607412. doi: 10.1111/tpj.12120. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.12120>.
- X. Ma, F. E. Vaistij, Y. Li, W. S. Jansen Van Rensburg, S. Harvey, M. W. Bairu, S. L. Venter, S. Mavengahama, Z. Ning, I. A. Graham, A. Van Deynze, Y. Van De Peer, and K. J. Denby. A chromosome-level *Amaranthus cruentus* genome assembly highlights gene family evolution and biosynthetic gene clusters that may underpin the nutritional value of this traditional crop. *The Plant Journal*, 107(2):613–628, July 2021. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.15298. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.15298>.
- B. N. Mansfeld and R. Grumet. QTLseqr: An R Package for Bulk Segregant Analysis with Next-Generation Sequencing. *The Plant Genome*, 11(2):180006, July 2018. ISSN 1940-3372, 1940-3372. doi: 10.3835/plantgenome2018.01.0006. URL <https://acsess.onlinelibrary.wiley.com/doi/10.3835/plantgenome2018.01.0006>.
- A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytzsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9):1297–1303, Sept. 2010. ISSN 1088-9051. doi: 10.1101/gr.107524.110. URL <http://genome.cshlp.org/lookup/doi/10.1101/gr.107524.110>.
- P. S. Millard, K. Weber, B. B. Kragelund, and M. Burow. Specificity of MYB interactions relies on motifs in ordered and disordered contexts. *Nucleic Acids Research*, 47(18):9592–9608, Oct. 2019. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkz691. URL <https://academic.oup.com/nar/article/47/18/9592/5545731>.
- M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger. ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6):679–682, June 2022. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-022-01488-1. URL

<https://www.nature.com/articles/s41592-022-01488-1>.

- T. Murashige and F. Skoog. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum*, 15(3):473–497, July 1962. ISSN 0031-9317, 1399-3054. doi: 10.1111/j.1399-3054.1962.tb08052.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1399-3054.1962.tb08052.x>.
- E. Młodzińska and others. Survey of plant pigments: molecular and environmental determinants of plant colors. *Acta Biologica Cracoviensia Series Botanica*, 51(1):7–16, 2009.
- J. Naik, R. Rajput, B. Pucker, R. Stracke, and A. Pandey. The R2R3-MYB transcription factor MtMYB134 orchestrates flavonol biosynthesis in *Medicago truncatula*. *Plant Molecular Biology*, 106(1-2):157–172, May 2021. ISSN 0167-4412, 1573-5028. doi: 10.1007/s11103-021-01135-x. URL <https://link.springer.com/10.1007/s11103-021-01135-x>.
- J. Pietsch. *Cross-regulatory interactions of trichome patterning genes in Arabidopsis thaliana and related species*. PhD Thesis, Universität zu Köln, 2022.
- G. Polturak, D. Breitel, N. Grossman, A. Sarrion-Perdigones, E. Weithorn, M. Pliner, D. Orzaez, A. Granell, I. Rogachev, and A. Aharoni. Elucidation of the first committed step in betalain biosynthesis enables the heterologous engineering of betalain pigments in plants. *New Phytologist*, 210(1):269–283, Apr. 2016. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.13796. URL <https://onlinelibrary.wiley.com/doi/10.1111/nph.13796>.
- G. Polturak, U. Heinig, N. Grossman, M. Battat, D. Leshkowitz, S. Malitsky, I. Rogachev, and A. Aharoni. Transcriptome and Metabolic Profiling Provides Insights into Betalain Biosynthesis and Evolution in *Mirabilis jalapa*. *Molecular Plant*, 11(1):189–204, Jan. 2018. ISSN 16742052. doi: 10.1016/j.molp.2017.12.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S1674205217303714>.
- B. Pucker, F. Reiher, and H. M. Schilbert. Automatic Identification of Players in the Flavonoid Biosynthesis with Application on the Biomedical Plant *Croton tiglium*. *Plants*, 9(9):1103, Aug. 2020. ISSN 2223-7747. doi: 10.3390/plants9091103. URL <https://www.mdpi.com/2223-7747/9/9/1103>.
- B. Pucker, N. Walker-Hale, J. Dzurlic, W. C. Yim, J. C. Cushman, A. Crum, Y. Yang, and S. F. Brockington. Multiple mechanisms explain loss of anthocyanins from betalain-pigmented *Caryophyllales*, including repeated wholesale loss of a key anthocyanidin synthesis enzyme. *New Phytologist*, 241(1):471–489, Jan. 2024. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.19341. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19341>.
- R. Rajput, S. Tyagi, J. Naik, B. Pucker, R. Stracke, and A. Pandey. The R2R3-MYB gene family in *Cicer arietinum*: genome-wide identification and expression analysis leads to functional characterization of proanthocyanidin biosynthesis regulators in the seed coat. *Planta*, 256(4):67, Oct. 2022. ISSN 0032-0935, 1432-2048. doi: 10.1007/s00425-022-03979-z. URL <https://link.springer.com/10.1007/s00425-022-03979-z>.

- M. Sakuta, A. Tanaka, K. Iwase, M. Miyasaka, S. Ichiki, M. Hatai, Y. T. Inoue, A. Yamagami, T. Nakano, K. Yoshida, and S. Shimada. Anthocyanin synthesis potential in betalain-producing *Caryophyllales* plants. *Journal of Plant Research*, 134(6):1335–1349, Nov. 2021. ISSN 0918-9440, 1618-0860. doi: 10.1007/s10265-021-01341-0. URL <https://link.springer.com/10.1007/s10265-021-01341-0>.
- S. L. Salzberg. Next-generation genome annotation: we still struggle to get it right. *Genome Biology*, 20(1):92, s13059–019–1715–2, Dec. 2019. ISSN 1474-760X. doi: 10.1186/s13059-019-1715-2. URL <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1715-2>.
- N. Sasaki, T. Adachi, T. Koda, and Y. Ozeki. Detection of UDP-glucose: *cyclo* - DOPA 5- *O* -glucosyltransferase activity in four o’clocks (*Mirabilis jalapa* L.). *FEBS Letters*, 568(1-3):159–162, June 2004. ISSN 0014-5793, 1873-3468. doi: 10.1016/j.febslet.2004.04.097. URL <https://febs.onlinelibrary.wiley.com/doi/10.1016/j.febslet.2004.04.097>.
- J. G. Schaart, C. Dubos, I. Romero De La Fuente, A. M. M. L. Van Houwelingen, R. C. H. De Vos, H. H. Jonker, W. Xu, J. Routaboul, L. Lepiniec, and A. G. Bovy. Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria × ananassa*) fruits. *New Phytologist*, 197(2):454–467, Jan. 2013. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.12017. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.12017>.
- J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682, July 2012. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.2019. URL <https://www.nature.com/articles/nmeth.2019>.
- R. Schwacke, G. Y. Ponce-Soto, K. Krause, A. M. Bolger, B. Arsova, A. Hallab, K. Gruden, M. Stitt, M. E. Bolger, and B. Usadel. MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. *Molecular Plant*, 12(6):879–892, June 2019. ISSN 16742052. doi: 10.1016/j.molp.2019.01.003. URL <https://linkinghub.elsevier.com/retrieve/pii/S1674205219300085>.
- H. Sheehan, T. Feng, N. Walker-Hale, S. Lopez-Nieves, B. Pucker, R. Guo, W. C. Yim, R. Badgami, A. Timoneda, L. Zhao, H. Tiley, D. Copetti, M. J. Sanderson, J. C. Cushman, M. J. Moore, S. A. Smith, and S. F. Brockington. Evolution of l-DOPA 4,5-dioxygenase activity allows for recurrent specialisation to betalain pigmentation in *Caryophyllales*. *New Phytologist*, 227(3):914–929, Aug. 2020. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.16089. URL <https://onlinelibrary.wiley.com/doi/10.1111/nph.16089>.
- S. Shimada, H. Otsuki, and M. Sakuta. Transcriptional control of anthocyanin biosynthetic genes in the *Caryophyllales*. *Journal of Experimental Botany*, 58(5):957–967, Jan. 2007. ISSN 0022-0957, 1460-2431. doi: 10.1093/jxb/erl256. URL <https://academic.oup.com/jxb/article-lookup/doi/10.1093/jxb/erl256>.
- F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov.

- BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19):3210–3212, Oct. 2015. ISSN 1367-4811, 1367-4803. doi: 10.1093/bioinformatics/btv351. URL <https://academic.oup.com/bioinformatics/article/31/19/3210/211866>.
- M. G. Stetter and K. J. Schmid. Analysis of phylogenetic relationships and genome size evolution of the *Amaranthus* genus using GBS indicates the ancestors of an ancient crop. *Molecular Phylogenetics and Evolution*, 109:80–92, Apr. 2017. ISSN 10557903. doi: 10.1016/j.ympev.2016.12.029. URL <https://linkinghub.elsevier.com/retrieve/pii/S105579031630464X>.
- M. G. Stetter, M. Vidal-Villarejo, and K. J. Schmid. Parallel Seed Color Adaptation during Multiple Domestication Attempts of an Ancient New World Grain. *Molecular Biology and Evolution*, 37(5):1407–1419, May 2020. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msz304. URL <https://academic.oup.com/mbe/article/37/5/1407/5682418>.
- J. Stougaard, D. Abildsten, and K. A. Marcker. The *Agrobacterium rhizogenes* pRi TL-DNA segment as a gene vector system for transformation of plants. *Molecular and General Genetics MGG*, 207(2-3):251–255, May 1987. ISSN 0026-8925, 1432-1874. doi: 10.1007/BF00331586. URL <http://link.springer.com/10.1007/BF00331586>.
- R. Stracke, M. Werber, and B. Weisshaar. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology*, 4(5):447–456, Oct. 2001. ISSN 13695266. doi: 10.1016/S1369-5266(00)00199-0. URL <https://linkinghub.elsevier.com/retrieve/pii/S1369526600001990>.
- R. Stracke, H. Ishihara, G. Huep, A. Barsch, F. Mehrrens, K. Niehaus, and B. Weisshaar. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *The Plant Journal*, 50(4):660–677, May 2007. ISSN 0960-7412, 1365-313X. doi: 10.1111/j.1365-313X.2007.03078.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2007.03078.x>.
- R. Stracke, D. Holtgräwe, J. Schneider, B. Pucker, T. Rosleff Sørensen, and B. Weisshaar. Genome-wide identification and characterisation of R2R3-MYB genes in sugar beet (*Beta vulgaris*). *BMC Plant Biology*, 14(1):249, Dec. 2014. ISSN 1471-2229. doi: 10.1186/s12870-014-0249-8. URL <http://bmcpplantbiol.biomedcentral.com/articles/10.1186/s12870-014-0249-8>.
- R. Sunnadeniya, A. Bean, M. Brown, N. Akhavan, G. Hatlestad, A. Gonzalez, V. V. Symonds, and A. Lloyd. Tyrosine Hydroxylation in Betalain Pigment Biosynthesis Is Performed by Cytochrome P450 Enzymes in Beets (*Beta vulgaris*). *PLOS ONE*, 11(2):e0149417, Feb. 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0149417. URL <https://dx.plos.org/10.1371/journal.pone.0149417>.
- M. Tardaguila, L. De La Fuente, C. Marti, C. Pereira, F. J. Pardo-Palacios, H. Del Risco, M. Ferrell, M. Mellado, M. Macchietto, K. Verheggen, M. Edelmann, I. Ezkurdia, J. Vazquez, M. Tress, A. Mortazavi, L. Martens, S. Rodriguez-Navarro, V. Moreno-Manzano, and A. Conesa. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantifi-

- cation. *Genome Research*, 28(3):396–411, Mar. 2018. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.222976.117. URL <http://genome.cshlp.org/lookup/doi/10.1101/gr.222976.117>.
- A. Timoneda, T. Feng, H. Sheehan, N. Walker-Hale, B. Pucker, S. Lopez-Nieves, R. Guo, and S. Brockington. The evolution of betalain biosynthesis in *Caryophyllales*. *New Phytologist*, 224(1):71–85, Oct. 2019. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.15980. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.15980>.
- T. Vogt. Substrate specificity and sequence analysis define a polyphyletic origin of betanidin 5- and 6-O-glucosyltransferase from *Dorotheanthus bellidiformis*. *Planta*, 214(3):492–495, Jan. 2002. ISSN 0032-0935, 1432-2048. doi: 10.1007/s00425-001-0685-1. URL <http://link.springer.com/10.1007/s00425-001-0685-1>.
- T. Vogt, R. Grimm, and D. Strack. Cloning and expression of a cDNA encoding betanidin 5-O-glucosyltransferase, a betanidin- and flavonoid-specific enzyme with high homology to inducible glucosyltransferases from the *Solanaceae*. *The Plant Journal*, 19(5):509–519, Sept. 1999. ISSN 0960-7412, 1365-313X. doi: 10.1046/j.1365-313X.1999.00540.x. URL <https://onlinelibrary.wiley.com/doi/10.1046/j.1365-313X.1999.00540.x>.
- J. H. von Elbe. Betalains. *Current protocols in food analytical chemistry*, (1):F3–1, 2001. Publisher: Wiley Online Library.
- B. Wang, E. Tseng, M. Regulski, T. A. Clark, T. Hon, Y. Jiao, Z. Lu, A. Olson, J. C. Stein, and D. Ware. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nature Communications*, 7(1):11708, June 2016. ISSN 2041-1723. doi: 10.1038/ncomms11708. URL <https://www.nature.com/articles/ncomms11708>.
- H. Wang, D. Xu, S. Wang, A. Wang, L. Lei, F. Jiang, B. Yang, L. Yuan, R. Chen, Y. Zhang, and W. Fan. Chromosome-scale *Amaranthus tricolor* genome provides insights into the evolution of the genus *Amaranthus* and the mechanism of betalain biosynthesis. *DNA Research*, page dsac050, Dec. 2022. ISSN 1340-2838, 1756-1663. doi: 10.1093/dnares/dsac050. URL <https://academic.oup.com/dnaresearch/advance-article/doi/10.1093/dnares/dsac050/6880148>.
- W. Wang, M. Schalamun, A. Morales-Suarez, D. Kainer, B. Schwessinger, and R. Lanfear. Assembly of chloroplast genomes with long- and short-read data: a comparison of approaches using *Eucalyptus pauciflora* as a test case. *BMC Genomics*, 19(1):977, Dec. 2018. ISSN 1471-2164. doi: 10.1186/s12864-018-5348-8. URL <https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-018-5348-8>.
- X. Wang, J. Wu, M. Guan, C. Zhao, P. Geng, and Q. Zhao. *Arabidopsis* MYB4 plays dual roles in flavonoid biosynthesis. *The Plant Journal*, 101(3):637–652, Feb. 2020. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.14570. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.14570>.
- C. A. Wessinger and M. D. Rausher. Lessons from flower colour evolution on targets of selection. *Journal of Experimental Botany*, 63(16):5741–5749, Oct. 2012. ISSN 1460-2431, 0022-0957. doi: 10.1093/jxb/ers267. URL <https://academic.oup.com/jxb/>

- E. Xia, F. Li, W. Tong, H. Yang, S. Wang, J. Zhao, C. Liu, L. Gao, Y. Tai, G. She, J. Sun, H. Cao, Q. Gao, Y. Li, W. Deng, X. Jiang, W. Wang, Q. Chen, S. Zhang, H. Li, J. Wu, P. Wang, P. Li, C. Shi, F. Zheng, J. Jian, B. Huang, D. Shan, M. Shi, C. Fang, Y. Yue, Q. Wu, R. Ge, H. Zhao, D. Li, S. Wei, B. Han, C. Jiang, Y. Yin, T. Xia, Z. Zhang, S. Zhao, J. L. Bennetzen, C. Wei, and X. Wan. The tea plant reference genome and improved gene annotation using long-read and paired-end sequencing data. *Scientific Data*, 6(1):122, July 2019. ISSN 2052-4463. doi: 10.1038/s41597-019-0127-1. URL <https://www.nature.com/articles/s41597-019-0127-1>.
- F. Xie, Q. Hua, C. Chen, Z. Zhang, R. Zhang, J. Zhao, G. Hu, J. Chen, and Y. Qin. Genome-Wide Characterization of R2R3-MYB Transcription Factors in Pitaya Reveals a R2R3-MYB Repressor HuMYB1 Involved in Fruit Ripening through Regulation of Betalain Biosynthesis by Repressing Betalain Biosynthesis-Related Genes. *Cells*, 10(8):1949, July 2021. ISSN 2073-4409. doi: 10.3390/cells10081949. URL <https://www.mdpi.com/2073-4409/10/8/1949>.
- F. Xie, C. Chen, J. Chen, J. Chen, Q. Hua, K. Shah, Z. Zhang, J. Zhao, G. Hu, J. Chen, and Y. Qin. Betalain biosynthesis in red pulp pitaya is regulated via HuMYB132: a R-R type MYB transcription factor. *BMC Plant Biology*, 23(1): 28, Jan. 2023. ISSN 1471-2229. doi: 10.1186/s12870-023-04049-6. URL <https://bmcpplantbiol.biomedcentral.com/articles/10.1186/s12870-023-04049-6>.
- W. Xu, C. Dubos, and L. Lepiniec. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends in Plant Science*, 20(3):176–185, Mar. 2015. ISSN 13601385. doi: 10.1016/j.tplants.2014.12.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S1360138514003124>.
- C. Yanhui, Y. Xiaoyuan, H. Kun, L. Meihua, L. Jigang, G. Zhaofeng, L. Zhiqiang, Z. Yunfei, W. Xiaoxiao, Q. Xiaoming, S. Yunping, Z. Li, D. Xiaohui, L. Jingchu, D. Xing-Wang, C. Zhangliang, G. Hongya, and Q. Li-Jia. The MYB Transcription Factor Superfamily of *Arabidopsis*: Expression Analysis and Phylogenetic Comparison with the Rice MYB Family. *Plant Molecular Biology*, 60(1):107–124, Jan. 2006. ISSN 0167-4412, 1573-5028. doi: 10.1007/s11103-005-2910-y. URL <http://link.springer.com/10.1007/s11103-005-2910-y>.
- J. Zeng, J. Chen, K. Shah, F. Xie, C. Chen, J. Chen, J. Zhao, G. Hu, Z. Zhang, and Y. Qin. Identification of *HuSPL* family and key role of HuSPL12 in regulation of betalain biosynthesis in pitaya. *Physiologia Plantarum*, 175(3):e13923, May 2023. ISSN 0031-9317, 1399-3054. doi: 10.1111/ppl.13923. URL <https://onlinelibrary.wiley.com/doi/10.1111/ppl.13923>.
- L. Zhang, C. Chen, F. Xie, Q. Hua, Z. Zhang, R. Zhang, J. Chen, J. Zhao, G. Hu, and Y. Qin. A Novel WRKY Transcription Factor HmoWRKY40 Associated with Betalain Biosynthesis in Pitaya (*Hylocereus monacanthus*) through Regulating HmoCYP76AD1. *International Journal of Molecular Sciences*, 22(4):2171, Feb. 2021. ISSN 1422-0067. doi: 10.3390/ijms22042171. URL <https://www.mdpi.com/1422-0067/22/4/2171>.
- X. Zheng, S. Liu, C. Cheng, R. Guo, Y. Chen, L. Xie, Y. Mao, Y. Lin, Z. Zhang, and Z. Lai. Cloning and expression analysis of betalain biosynthesis genes in *Ama-*

ranthus tricolor. *Biotechnology Letters*, 38(4):723–729, Apr. 2016. ISSN 0141-5492, 1573-6776. doi: 10.1007/s10529-015-2021-z. URL <http://link.springer.com/10.1007/s10529-015-2021-z>.

- I. M. Zimmermann, M. A. Heim, B. Weisshaar, and J. F. Uhrig. Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *The Plant Journal*, 40(1):22–34, Oct. 2004. ISSN 0960-7412, 1365-313X. doi: 10.1111/j.1365-313X.2004.02183.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2004.02183.x>.

3.12 Supporting Information

Methods S1 Genome polishing and masking of repetitive elements

Methods S2 Assembly of full-length Iso-Seq transcripts

Methods S3 Computational genome annotation using BRAKER2

Methods S4 Combining full-length transcripts and computational annotation

Methods S5 Functional annotation

Methods S6 Betalain quantification

Notes S1 Improvement of the grain amaranth reference assembly

Notes S2 Annotation of the polished *A. hypochondriacus* reference genome increases completeness of annotated genes

Table S1 Summary of tissue of origin, read count, and source of long-read and short-read RNA sequencing data used to generate the *A. hypochondriacus* genome annotation v2.2.

Table S2 Number of PacBio Iso-Seq circular consensus sequencing (CCS) and full-length non-concatemer (FLNC) reads for all sequenced *A. hypochondriacus* tissues.

Table S3 Summary of BLAST results for betalain pathway genes

Table S4 Statistics for whole-genome sequencing data used in the bulk segregant analyses (BSAs)

Table S5 Statistics for RNA sequencing data generated from bulk segregant analysis (BSA) bulks.

Table S6 List of primers used in the study

Table S7 Repetitive element content of the polished *A. hypochondriacus* genome assembly v2.2

Table S8 BUSCO score summary of different data sets against the Embryophyta reference set

Table S9 Sequences used for phylogenetic analysis of *A. hypochondriacus* CYP76AD genes

Table S10 Sequences used for phylogenetic analysis of *A. hypochondriacus* DODA genes

Table S11 Summary of *A. hypochondriacus* flavonoid pathway candidate genes

Table S12 MYB transcription factors identified in *A. hypochondriacus*

Fig. S1 Organellar contamination in the *A. hypochondriacus* assembly

Fig. S2 Summary of the functional annotation of the *A. hypochondriacus* reference genome v2.2

Fig. S3 Phylogenetic analysis of *A. hypochondriacus* CYP76AD proteins with CYP protein sequences from other species

Fig. S4 Phylogenetic analysis of *A. hypochondriacus* DODA proteins with DODA protein sequences from other species

Fig. S5 Relative gene expression in transcript per million (TPM) of identified *A. hypochondriacus* betalain pathway genes in different tissues

Fig. S6 Relative gene expression in transcript per million (TPM) of identified *A.*

hypochondriacus flavonoid pathway genes in different tissues

Fig. S7 Relative gene expression in transcript per million (TPM) of identified *A. hypochondriacus* MYB transcription factor genes in different tissues

Fig. S8 Photometric and LC-MS quantification of betalain pathway metabolites for bulk segregant analysis (BSA) *A. hypochondriacus* parental genotypes

Fig. S9 Gene expression of *A. hypochondriacus* betalain pathway genes in flower tissue in green and red bulks of regulator and biosynthesis bulk segregant analyses (BSAs)

Fig. S10 Segregating variants called from *A. hypochondriacus* whole-genome sequencing data in the biosynthesis BSA and their relative allele depth for the *AhCYP76AD2* gene

Fig. S11 Missense and stop alleles in *AhCYP76AD2* from RNA-seq reads of *A. hypochondriacus* red flower tissue

Fig. S12 Sanger sequencing of *AhCYP76AD2* from biosynthesis bulk segregant analysis (BSA) individuals

Fig. S13 Relative gene expression in transcript per million (TPM) of identified *A. hypochondriacus* flavonoid pathway genes in red and green flowers

Fig. S14 Relative quantification of amaranthin content from transgenic *A. hypochondriacus* hairy roots

Fig. S15 Prediction of protein-protein interactions of selected MYB and bHLH proteins

3.12.1 Supplementary methods

Methods S1: Genome polishing and masking of repetitive elements

To prepare the downloaded reference genome for genome polishing, we removed stretches of ambiguous bases, usually located at the beginning or end of scaffolds, using a custom script (https://github.com/T-Winkler/Ahyp_v2_2_publication.git). Ambiguous bases were merged back into the genome after genome polishing. We ran two cycles of polishing using the two NextPolish core modules, k-mer score chain and k-mer count (Hu et al., 2020). Before each polishing module, we mapped reads using bwa-mem2 v2.2.1 (Li, 2013) and processed bam files using Samtools v1.13 (Li et al., 2009). After genome polishing, we masked the polished genome for repetitive elements in order to prepare the computational gene prediction, using Repeatmodeler v2.0.1 (Flynn et al., 2020) with the *-LTRStruct* option to create a database of repetitive elements from the polished assembly. We reclassified repetitive elements using RepeatClassifier v2.0.4 (Flynn et al., 2020) and used the repetitive element database to annotate repeats in the polished reference genome using Repeatmasker v4.1.1 (Chen, 2004) with the *-e rmbblast* option. We soft-masked the polished genome using bedtools v2.29.2 (Quinlan and Hall, 2010) *maskfasta* with the repetitive element annotation to generate the *A. hypochondriacus* genome assembly v2.2. To identify genomic regions showing mapping bias, we mapped the above described whole-genome sequencing reads to the polished reference genome using bwa-mem2 v2.2.1 (Li, 2013) and calculated read depth using Samtools v1.13 (Li et al., 2009). We identified organellar contamination within the assembly by mapping the *B. vulgaris* mitochondrial genome (Kubo et al., 2000) and the *A. hypochondriacus* chloroplast genome (Chaney et al., 2016) against the polished assembly using minimap2 v2.17 (Li, 2018) with the *asm10* and *asm5* presets, respectively.

Methods S2: Assembly of full-length Iso-Seq transcripts

Circular consensus sequence (CCS) reads were processed using Isoseq3 v3.3.0 (<https://github.com/PacificBiosciences/IsoSeq>; Table S3.2). After primer removal and demultiplexing of the CCS reads, poly-A tails were trimmed and artificial concatemers were removed using *Isoseq3 refine* to create a set of full-length non-concatemer (FLNC) reads (requiring FL reads to include a poly-A tail ≥ 20 bp). We combined the FLNC reads from all sequenced tissues using Samtools v1.13 (Li et al., 2009) *merge*. We clustered full-length non-concatemer reads of the seven different tissues into a set of high quality isoforms with predicted accuracy ≥ 0.99 using *Isoseq3 cluster*. We mapped the clustered high quality isoforms to the polished *A. hypochondriacus* reference genome using minimap2 v2.17 (Li, 2018) with the *splice:hq* preset and the *-secondary=no* and *-uf* options. In order to remove redundant sequences, we collapse the mapped high quality isoforms into unique isoforms using the *collapse_isoforms_by_sam.py* script of cDNAcupcake

v28.0.0 (https://github.com/Magdoll/cDNA_Cupcake). Only mapped high quality isoforms with a coverage of ≥ 0.95 , identity ≥ 0.9 , an identical intron chain and maximum 3' and 5' end differences of 1000 bp were collapsed into unique full-length transcripts. We obtained the number of clustered high quality isoforms supporting each unique full-length transcript and filtered possible 5' degraded transcripts with the *cDNA cupcake* scripts *get_abundance_post_collapse.py* and *filter_away_subset.py*.

We used SQANTI3 v4.2 (Tardaguila et al., 2018) for quality control of the unique isoforms as well as potential sequencing error correction. SQANTI3 uses the genomic positions of the mapped unique full-length transcripts to correct possible sequencing errors using the reference genome sequence (Tardaguila et al., 2018). We predicted the coding potential of the corrected transcript sequences as well as putative open reading frames (ORFs) in the sequences using CPC2 v1.0.1 (Kang et al., 2017). Transcripts classified as protein-coding were required to be labeled "coding" with positive ORF integrity. To estimate the completeness of the generated transcripts, we used BUSCO (Simão et al., 2015) to compare the predicted protein sequences of the transcripts to the embryophyta orthoDBv10 dataset (*-l embryophyta_odb10* option), consisting of 1614 conserved single copy orthologs. Protein-coding transcripts were used to supplement the computational genome annotation.

Methods S3: Computational genome annotation using BRAKER2

In a first run, we used as protein database 3,536,219 protein sequences from 117 Viridiplantae species from orthoDB v10 (Kriventseva et al., 2019), as well as the protein sequences of *Amaranthus cruentus* (Ma et al., 2021), as external evidence to guide the annotation. We used this preliminary run of BRAKER2 to guide mapping of the RNA-seq dataset described above to the reference genome v2.2 (Table S3.1). Quality control of downloaded fastq files was done using FastQC v0.11.9 (Andrews, 2010). We indexed the polished reference genome using STAR v2.7.8a (Dobin and Gingeras, 2015) with the following settings: *sjdbOverhang* 89, *genomeSAindexNbases* 13. We mapped the RNA-seq reads to the indexed genome using the STAR v2.7.8a (Dobin and Gingeras, 2015) with default settings, using the protein-based BRAKER2 annotation as splice-junction database. To generate the final computational gene annotation, we ran BRAKER2 with reference genome v2.2, using as external evidence both the protein database as well as the previously mapped short-read RNA sequencing reads. After computational annotation of the polished genome, we extracted protein and coding sequence (CDS) fasta sequences from the BRAKER2 annotation using the AUGUSTUS script *getAnnoFastaFromJoinGenes.py* (Stanke et al., 2006), filtering out all predicted genes with internal stop codons. We divided the predicted genes into three subcategories based on the amount of external support using the *selectSupportedSubsets.py* script of BRAKER2 (Brûna et al., 2021). The predicted genes were classified as "full support" if all introns of multi-exons genes

were supported by external evidence (protein or RNA-seq alignments) or if both start and stop codon were supported by external evidence in the case of mono-exon genes. Genes were classified into the "any support" subset, if at least one intron or start or stop codon were supported by external evidence. Otherwise, predicted genes were classified into the "no support" subset.

Methods S4: Combining full-length transcripts and computational annotation

We used the supported dataset of predicted genes and the annotation file of external support hints from the BRAKER2 gene prediction together with the genomic positions of the coding sequence of the full-length transcripts, as predicted by CPC2, and the available long-read configuration file as input for TSEBRA. We used gffread v0.12.7 (Perteira and Perteira, 2020) with the *-M* and *-K* options to discard redundant gene models, removing all genes which were fully contained in another gene model, to generate the genome annotation v2.2. We manually corrected the annotation of the *AmMYB11* gene (AHp014591) based on alignment of previous complementary DNA (cDNA) Sanger sequencing results (generated from the dark seeded accession PI 604581) to the polished reference genome using minimap2 v2.17 (Li, 2018). In order to estimate the completeness of the genome annotation v2.2, we used the annotated protein sequences as input for BUSCO v5.2.2 (Simão et al., 2015) with the Embryophyte reference set (*-l embryophyta_odb10* option). For each annotated isoform of the genome annotation v2.2, we assessed support by assembled Iso-Seq transcripts using gffcompare v0.12.6 (Perteira and Perteira, 2020). To compare the annotation completeness with the previously published genome annotations of *A. hypochondriacus* and *A. cruentus*, we submitted the annotated protein sequences from the two genome annotations to BUSCO v5.2.2 using the same parameters as detailed above. We visualized the genome annotation in a circos plot using the circlize R package (Gu et al., 2014), calculating gene and repetitive element density in 1 Mb windows along the genome.

Methods S5: Functional annotation

We additionally annotated sequences using Interproscan v5.56-89.0 (Jones et al., 2014), retrieving pathway and GO-term annotations. We identified betalain pathway genes using BLAST (Altschul, 1997) searches with previously described betalain pathway genes (Table S3.3). Searches were carried out against the annotated amaranth protein sequences using BLAST+ v2.9.0-2, requiring a high-scoring query coverage of 80% (*qcov_hsp_perc 80*) and using the best hit as candidate sequence. For phylogenetic analysis of the candidate sequences, we aligned the amaranth CYP76AD and DODA protein sequences with sequences of other species (Table S3.9 and S3.10) using Clustal Omega v1.2.4 (Sievers et al., 2011) with default settings and constructed neighbour joining phylogenetic trees

using ClustalW v2.1 (Thompson et al., 1994) with 1000 bootstrap replicates. We used the flavonoid pathway identification pipeline KIPes v0.35 (Pucker et al., 2020) to identify candidate genes of the flavonoid pathway. Amaranth genes were defined as candidate genes if they included all of the conserved residues defined by KIPes. If no candidate sequence with all conserved residues could be identified, instead the amaranth gene with the highest percentage similarity and less than three missing conserved residues was defined as candidate gene. For each pathway gene not identified by KIPes, we searched for candidate sequences through alignment KIPes bait sequences against the amaranth genome using BLAST+ v2.9.0-2 (Altschul, 1997) and exonerate v2.4.0 with the *percent 50* option (Slater and Birney, 2005). Few amaranth genes represented the best candidates for multiple different flavonoid pathway genes (e.g. *F3'H* and *F3'5'H* shared the best candidate in AHP017497.1). Since these ambiguous candidate genes could be assigned to specific flavonoid pathway genes with high confidence (displaying all conserved residues), we defined the genes *CHI2*, *F3'5'H* and *FNS1* to be missing in amaranth.

To identify potential MYB transcription factors, we searched the predicted protein sequences of the *A. hypochondriacus* genome annotation v2.2 for matches to the MYB DNA-binding domain HMM profile (Pfam ID: PF00249, downloaded 11.01.2022) using HMMer v3.3.2 (Eddy, 2009). We filtered matching domains based on score (> 25), alignment length (> 20 amino acids) and accuracy (> 0.8) and divided genes with matching domains into the R2R3, 3R and 4R subfamilies based on the number of adjacent repeats of the MYB domain using a custom script. Domains were considered adjacent, if they were located within 15 amino acids of one another. To assign the identified MYB transcription factors to their established subgroups, we aligned amaranth MYB protein sequences with the MYB proteins of *A. thaliana* (Stracke et al., 2001) and *B. vulgaris* (Stracke et al., 2014) using Clustal Omega v1.2.4 (Sievers et al., 2011) with default settings. We constructed a neighbour joining phylogenetic tree with 1000 bootstrap replicates from the Clustal Omega multiple sequence alignment using ClustalW v2.1 (Thompson et al., 1994). MYB proteins were assigned to subgroups based on the subgroup assignment of the closest *A. thaliana* homolog, as reported by Dubos et al. (2010) and Li et al. (2016). For ambiguous placement in the phylogeny, additionally the subgroup assignment of the closest *B. vulgaris* homolog was considered (Stracke et al., 2014). For *A. thaliana* subgroup S8, we considered the classification also used by Tombuloglu (2020). We identified potential bHLH-interacting MYB transcription factors by searches in the protein sequence using the conserved sequence motif ([D/E]Lx₂[R/K]x₃Lx₆Lx₃R), described by Zimmermann et al. (2004). To investigate their sequence conservation, we aligned subgroup S6 anthocyanin regulators and BvMYB1-clade proteins using ClustalOmega with default settings. In addition to the bHLH-interaction motif, we annotated residues required to restore bHLH-interaction in BvMYB1 through site-directed mutagenesis (Hatlestad et al., 2015), residues where amino-acid exchanges resulted in temperature-sensitive defects of the interaction capacity of PAP1 with bHLH proteins (Zimmermann et al., 2004) and residues

where amino acid exchanges resulted in enhanced activation of the *ANS* promoter by Caryophyllales MYBs (Sakuta et al., 2021). To assess the effect of amino acid exchanges in the R3 interaction domain on bHLH-interaction, we performed pairwise prediction of protein-protein interactions of the S6 anthocyanin regulator PAP1 (AtMYB75) and the BvMYB1-clade MYBs AhMYB2.1 and BvMYB1 with three known *A. thaliana* bHLH interaction partners of PAP1 (EGL3, GL3 and TT8) using ColabFold v1.5.2 (Mirdita et al., 2022) on COSMIC² (Cianfrocco et al., 2017) and visualized interacting residues in the MYB R3 domain using iCn3D (Wang et al., 2020).

Methods S6: Betalain quantification

We performed photometric quantification and quantified individual metabolites (amaranthin and betanin (betacyanins), vulgaxanthin IV (betaxanthin), and betalamic acid) using liquid chromatography - mass spectrometry (LC-MS) from the same samples with at least 3 replicates per tissue per accession (4 for leaves of PI 576485, 8 for leaves of PI 538323, 3 for leaves of PI 686465, 3 for flower of PI 576485, 8 for flower of PI 538323 and 3 for flower of PI 686465). We performed additional photometric quantification from 5 leaf samples of PI 576485, 2 leaf samples of PI 538323 and 4 leaf samples of PI 686465, as well as 5 flower samples of PI 576485, 2 flower samples of PI 538323 and 5 flower samples of PI 686465. To confirm *AhMYB2.1* as betalain regulator, we sampled wild-type like and transgenic (expressing *p35S:AhMYB2.1-GFP*) roots from the same plant (two plants in total) and extracted betalains for LC-MS analysis. To quantify extracted betalains, we modified a previously described extraction method (Chang et al., 2021; von Elbe, 2001). We sampled around 30 mg of flower or leaf tissue or around 10 mg of root tissue and immediately stored plant material in liquid nitrogen. Flower tissue was sampled from the tip of the inflorescence, leaf tissue from the tip of the youngest fully developed leaf. We homogenized plant material in liquid nitrogen and extracted betalains in a two step procedure. Betalains were extracted on a shaker in H₂O for 60 min at 8 °C, afterwards samples were centrifuged at 16,000 g for 10 min at 4 °C. After centrifugation, the supernatant was transferred to a new tube and methanol and chloroform were added in a ratio of 1:1:2 H₂O:methanol:chloroform to separate residual chlorophyll from extracted betalains. For usage as internal standard in LC-MS measurements, we added 10 µM sinigrin to methanol in the previous step. Extracts were incubated on a shaker for 10 min at 8 °C and centrifuged again at 16,000 g for 10 min at 4 °C. We collected betalains from the upper aqueous phase for quantification using photometry and LC-MS.

For photometric quantification of betacyanin and betaxanthin content, we measured absorption of the extracts at 536 nm (A_{536}) and 480 nm (A_{480}), respectively. We converted A_{536} to betacyanin content using the molar extinction coefficient $\epsilon = 60,000 \text{ M}^{-1} \text{ cm}^{-1}$ (Jain et al., 2015) and the molecular weight of amaranthin (726.6 g mol^{-1}). To correct for betacyanin absorption at 480 nm, we adjusted our estimation of betaxanthin absorption

using the formula $A_{bx} = A_{480} - \frac{A_{536}}{3.1}$ (von Elbe, 2001). We converted A_{bx} to betaxanthin content using the molar extinction coefficient $\epsilon = 48,000 \text{ M}^{-1} \text{ cm}^{-1}$ (Escribano et al., 2017) and the molecular weight of vulgaxanthin IV ($324.333 \text{ g mol}^{-1}$).

LC of plant extracts was performed on the Dionex UltiMate 3000 UPLC system (Thermo Scientific, Germering, Germany) using a C18 XSelect® HSS T3 (2.5 μm , 3.0x150 mm) column (Waters, Drinagh, Ireland) at a flow rate of 0.35 ml/min. The mobile phase consisted of water/0.1 % formic acid (solution A) and methanol/0.1 % formic acid (solution B). The 35-min gradient conditions were as follows: 0-3 min 1 % B (v/v), 3-20 min linear gradient to 40 %, 20-25 min linear gradient to 99 % B, 25-30 min 99 % B, 30-33 min linear gradient to 1 % B and 33-35 min 1 % B. MS spectra data were acquired using a maXis 4G QTOF-instrument (Bruker Daltonics, Bremen, Germany) equipped with an electrospray ion source. Betalains were analyzed in positive ion mode with following source settings: capillary voltage: 3.5 kV, nebulizer pressure: 1 bar, dry gas flow: 8 l/min, dry temperature: 200 °C. Metabolite identification was achieved from full-scan MS data (mass range m/z 50-1000 m/z) using the DataAnalysis (version 6.0) software (Bruker, Bremen, Germany). Retention time of betalains was confirmed by comparisons to the literature. Peak areas of respective compounds using the extracted ion chromatograms were determined with QuantAnalysis (version 6.0) software (Bruker, Bremen, Germany). Betaxanthin and betacyanin amounts were normalized against sample fresh weight and the internal standard ($\text{Metabolite relative area} = \frac{\text{Metabolite area}}{\text{Internal standard area}}$). We conducted statistical analysis of betalain content differences between samples within tissue using ANOVA.

3.12.2 Supplementary notes

Notes S1: Improvement of the grain amaranth reference assembly

In order to improve the annotation with full-length transcript sequencing evidence, we performed PacBio isoform sequencing of seven different tissues of the amaranth reference accession (PI 558499, Table S3.1). A total of 5,306,288 CCS reads across the seven different tissues were generated, resulting in a total of 163,587 high-quality transcript sequences after filtering and clustering (Table S3.2). Of these high-quality transcripts, 94.23 % (154,949) could be uniquely mapped to the polished reference genome and were further collapsed into 53,638 unique isoforms. We further filtered possible 5' degraded sequences and corrected sequences according to the polished reference genome, resulting in a set of 52,468 transcript sequences from 22,165 loci. Out of these, a total of 42,835 transcripts from 18,593 genes were predicted as coding with intact ORF. Correction of long-read transcript sequences using the polished reference genome resulted in a minor increase in reported BUSCO completeness from 90.7 % to 90.8 % (Table S3.8).

To evaluate the effect of genome polishing on long-read 'error correction', we mapped

HQ long-read clustered isoforms from all tissues to the unpolished reference genome and collapsed the mapped isoforms using identical settings as above. Genome-based correction of long-read transcripts is expected to increase transcript accuracy by correcting possible sequencing errors in the long-reads with the reference genome sequence. A decreased accuracy of transcripts after genome-based correction instead indicates sequencing errors in the assembly not present in the transcripts. We 'corrected' the sequences of the collapsed isoforms using the unpolished reference genome and predict ORFs in both the uncorrected and corrected isoforms. For a total of 43,245 transcript sequences that were predicted to be coding before and after correction using the unpolished reference genome, 35.1 % (15,194) differed in predicted ORF length. While 12,289 transcripts had longer predicted ORFs without 'correction', 2,805 were predicted with longer ORF after 'correction' with the unpolished reference genome. ORFs of transcripts with differential ORF predictions before and after 'correction' differed by an average of 140 amino acids. SQANTI based correction of long-read transcripts using the unpolished reference genome strongly decreased the reported BUSCO score from 90.6 % to 74.6 % complete (Table S3.8). The large difference in BUSCO score and predicted ORF length between Iso-Seq data and unpolished genome indicate that the unpolished reference genome sequence included indels that disrupted the previous annotation. In contrast to that, reference genome based correction using our polished assembly resulted in an increase in reported BUSCO completeness from 90.7 % to 90.8 % (Table S3.8) with only 10.1 % of transcripts differing in predicted ORF length after correction. These results demonstrate the successful removal of sequencing errors from the reference genome through genome polishing and the positive impact for gene annotation.

Notes S2: Annotation of the polished *A. hypochondriacus* reference genome increases completeness of annotated genes

In addition to the long-read RNA annotation, we predicted protein-coding genes in the polished reference genome using BRAKER2 (Brůna et al., 2021) with external evidence in the form of a protein database and short-read RNA sequencing reads from eight different tissues (Table S3.1). Computational gene prediction resulted in a total of 32,604 predicted gene isoforms, 67.5 % (22,012) with full support, 21.5 % (7,013) with partial support and 11 % (3,579) without external support. We excluded all predicted isoforms without support from RNA sequencing or protein alignments. The filtered computational annotation included 29,025 annotated isoforms from 25,405 genes with a BUSCO score of 97.9 % complete (Table S3.8).

After combination of BRAKER2 and Iso-Seq annotations, we found the complete intron chain of 20,959 (74.66 %) annotated isoforms to be supported by full-length transcripts from long-read sequencing data, indicating high-quality of the annotation. As expected, regions with high repeat content show less annotated genes (Fig. 1c). While

the number of predicted genes was comparable to the previously published annotation (Lightfoot et al., 2017) and lower than the number of predicted genes from *A. cruentus* (Ma et al., 2021), the mean CDS length and exon number was slightly increased (Table 1). While the previous *A. hypochondriacus* genome annotation included only a single primary isoform for each gene, our genome annotation v2.2 includes multiple annotated isoforms for 3,484 genes, resulting in a total of 28,074 annotated isoforms.

3.12.3 Supplementary tables

Table S3.1: Summary of tissue of origin, read count, and source of long-read and short-read RNA sequencing data used to generate the *A. hypochondriacus* genome annotation v2.2. For long-read data, the number of reads refers to circular consensus sequencing (CCS) reads. For short-read data, the number of reads refers to read pairs.

Tissue	Sequencing type	Reads	Uniquely mapped (%)	Source
Pollen	long-read	637,760	-	this study
Root	long-read	752,123	-	this study
Root	short-read	22,656,771	93.13	Clouse et al. (2016)
Cotyledons	long-read	928,752	-	this study
Cotyledons	short-read	22,291,775	89.50	Clouse et al. (2016)
Flower	long-read	691,694	-	this study
Flower	short-read	21,501,431	94.93	Clouse et al. (2016)
Leaf	long-read	478,352	-	this study
Leaf	short-read	22,045,070	94.28	Clouse et al. (2016)
Developing seed	long-read	371,693	-	this study
Developing seed	short-read	21,768,342	96.31	Clouse et al. (2016)
Mature seed	long-read	1,445,914	-	this study
Mature seed	short-read	22,242,512	92.17	Clouse et al. (2016)
Stem	short-read	22,162,620	95.61	Clouse et al. (2016)
Water-stressed tissues	short-read	21,610,476	95.37	Clouse et al. (2016)

Table S3.2: Number of PacBio Iso-Seq circular consensus sequencing (CCS) and full-length non-concatemer (FLNC) reads for all sequenced *A. hypochondriacus* tissues. We combined the FLNC reads from all tissues before clustering FLNC reads into a set of high-quality isoforms from all tissues.

	CCS reads	FLNC reads	HQ clustered
Root	752,123	734,212	
Cotyledon	928,752	895,195	
Flower	691,694	676,684	
Leaf	478,352	467,382	
Pollen	637,760	624,434	
Developing seed	371,693	363,574	
Mature seed	1,445,914	1,404,972	
Combined	5,306,288	5,166,453	163,587

Table S3.3: Summary of BLAST results for betalain pathway genes. GenBank or Uniprot entries for query genes are noted alongside their original publication. For each query, the best hit in the annotated *A. hypochondriacus* protein sequences and the percentage identity as reported by blast is displayed. DODA (L-3,4-dihydroxyphenylalanine 4,5-dioxygenase), cDOPA5GT (*cyclo*-DOPA 5-O-glucosyltransferase), Betanidin5GT (betanidin 5-O-glucosyltransferase), Betanidin6GT (betanidin 6-O-glucosyltransferase).

Pathway gene	Query GenBank	Query Uniprot	Reference	Query species	Best hit	Percentage identity
CYP76AD1	-	C76AD_BETVU	Hatlestad et al. (2012)	<i>B. vulgaris</i>	AHp023148	84.14
CYP76AD2	HQ656025.1	-	Hatlestad et al. (2012)	<i>A. cruentus</i>	AHp023148	99.80
CYP76AD5	-	A0A0U1YLX5_BETVU	Sunnadeniya et al. (2016)	<i>B. vulgaris</i>	AHp000674	83.43
CYP76AD6	-	A0A0U1YJA0_BETVU	Sunnadeniya et al. (2016)	<i>B. vulgaris</i>	AHp000674	82.16
DODA1	HQ889614.1	-	Casique-Arroyo et al. (2014)	<i>A. hypochondriacus</i>	AHp010386	100.00
DODA1	-	DOD1U_BETVU	Hatlestad et al. (2012)	<i>B. vulgaris</i>	AHp023147	78.10
cDOPA5GT	KJ136018.1	-	Casique-Arroyo et al. (2014).	<i>A. hypochondriacus</i>	AHp007219	100.00
Betanidin5GT	KJ136017.1	-	Casique-Arroyo et al. (2014)	<i>A. hypochondriacus</i>	AHp001663	99.79
Betanidin6GT	KP174812.1	-	Zheng et al. (2016)	<i>A. tricolor</i>	AHp005940	92.96

Table S3.4: Statistics for whole-genome sequencing data used in the bulk segregant analyses (BSAs). The number of read pairs generated for the green and red bulks of both BSAs and the percentage of read pairs uniquely mapped to the polished *A. hypochondriacus* reference genome is displayed.

Sample	Reads	Uniquely mapped (%)
green bulk regulator BSA	59,471,388	98.29
red bulk regulator BSA	68,365,577	98.62
green bulk biosynthesis BSA	63,069,713	98.03
red bulk biosynthesis BSA	71,253,748	97.99

Table S3.5: Statistics for RNA sequencing data generated from bulk segregant analysis (BSA) bulks. Number of read pairs before and after adapter trimming for short-read RNA-seq data generated for gene expression quantification in pooled green and red flower tissue of regulator and biosynthesis BSA. The data was generated from pooled *A. hypochondriacus* flower tissue of 24 plants from each bulk of the biosynthesis BSA and 29 plants from each bulk of the regulator BSA.

Sample	Reads	Trimmed reads
green bulk regulator BSA	25,960,085	25,940,297
red bulk regulator BSA	25,423,535	25,404,128
green bulk biosynthesis BSA	38,659,017	38,560,213
red bulk biosynthesis BSA	25,718,709	25,673,931

Table S3.6: List of primers used in the study.

Name	Function	Sequence
SV1	forward primer <i>AhMYB2</i> amplification	ggggacaagtttgtacaaaaaagcaggc ttaatgctacaaacatttgtatccacac
SV2	reverse primer <i>AhMYB2</i> amplification	ggggaccactttgtacaagaaagctgggt taatactcctcccctaaaattgacaac
SV4	<i>AhCYP76AD2</i> genotyping	gcctaaagatgcacaaatactcg
SV5	<i>AhCYP76AD2</i> genotyping	tcaatgcctaagggatggaatg

Table S3.7: Repetitive element content of the polished *A. hypochondriacus* genome assembly v2.2. Denoted are the classified type of repetitive element with the proportion of genome covered by that class.

Classification	Genome covered [bp]	Genome covered [%]
DNA/CMC-EnSpm	4,098,198	1.01
DNA/MULE-MuDR	5,723,062	1.42
DNA/PIF-Harbinger	1,168,228	0.29
DNA/TcMar-Stowaway	1,239,206	0.31
DNA/Zisupton	800,423	0.2
DNA/hAT-Ac	2,943,534	0.73
DNA/hAT-Tag1	254,129	0.06
DNA/hAT-Tip100	1,342,775	0.33
RC/Helitron	2,071,001	0.51
LINE/CRE	1,014,983	0.25
LINE/L1	8,821,523	2.18
LINE/Penelope	21,568	0.01
LINE/R1	18,480	0
LINE/RTE	504,828	0.12
LINE/RTE-BovB	2,580,488	0.64
LTR	163,433	0.04
LTR/Copia	39,486,653	9.77
LTR/Gypsy	19,135,979	4.74
LTR/Pao	50,089	0.01
LTR/Unknown	10,497,276	2.6
SINE/5S	76,086	0.02
SINE/U	39,697	0.01
SINE/tRNA	697,466	0.17
SINE/tRNA-CR1	38,625	0.01
rRNA	21,280	0.01
tRNA	30,256	0.01
Low complexity	937,694	0.23
Simple repeat	11,495,793	2.85
Unknown	114,642,357	28.38
Total sum	229,915,110	56.91

Table S3.8: BUSCO score summary of different data sets against the Embryophyta reference set. Iso-Seq full-length non-concatemer (FLNC) reads were mapped against the unpolished or polished *A. hypochondriacus* reference genome and corrected using the unpolished or polished reference genome sequence. For Iso-Seq transcript sets, BUSCO was run in protein mode using predicted open reading frames. For genome annotations, BUSCO was run in protein mode on predicted protein sequences.

Dataset	Species	Data source	Complete	Complete (single copy)	Complete (duplicated)	Fragmented	Missing
Genome annotation v2.1	<i>A. hypochondriacus</i>	Lightfoot et al. (2017)	80.7	78.5	2.2	8.8	10.5
Genome annotation	<i>A. cruentus</i>	Ma et al. (2021)	89.8	87.2	2.6	2.5	7.7
Iso-Seq transcripts (uncorrected)	<i>A. hypochondriacus</i>	this study	90.7	58.9	31.8	2.0	7.3
Iso-Seq transcripts (corrected)	<i>A. hypochondriacus</i>	this study	90.8	57.3	33.5	1.4	7.8
Iso-Seq transcripts (unpolished genome, uncorrected)	<i>A. hypochondriacus</i>	this study	90.6	58.7	31.9	2.0	7.4
Iso-Seq transcripts (unpolished genome, corrected)	<i>A. hypochondriacus</i>	this study	74.6	46.7	27.9	11.0	14.4
Braker2 annotation	<i>A. hypochondriacus</i>	this study	97.9	80.9	17.0	0.8	1.3
Genome annotation v2.2	<i>A. hypochondriacus</i>	this study	98.0	83.4	14.6	0.7	1.3

Table S3.9: Sequences used for phylogenetic analysis of *A. hypochondriacus* CYP76AD genes. Species, gene name and GenBank accession of input genes for phylogenetic analysis.

Species	Gene name	GenBank accession
<i>Amaranthus cruentus</i>	AcCYP76AD2	AET43291.1
<i>Arabidopsis thaliana</i>	AtCYP76C2	NP_182081
<i>Basella alba</i>	BaCYP76AD13	AJD87469
<i>Bougainvillea peruviana</i>	BpCYP76AD1-1	BCD59210
<i>Bougainvillea peruviana</i>	BpCYP76AD1-2	BCD59211
<i>Bougainvillea peruviana</i>	BpCYP76AD1-3	BCD59212
<i>Bougainvillea peruviana</i>	BpCYP76AD2	BCD59213
<i>Beta vulgaris</i>	BvCYP76AD1	AET43290
<i>Beta vulgaris</i>	BvCYP76AD5	AJD87473
<i>Beta vulgaris</i>	BvCYP76AD6	AJD87474
<i>Chenopodium quinoa</i>	CqCYP76AD1	XP_021769302
<i>Mirabilis jalapa</i>	MjCYP76AD3	AET43292
<i>Mirabilis jalapa</i>	MjCYP76AD7	AJD87463
<i>Mirabilis jalapa</i>	MjCYP76AD15	AJD87471

Table S3.10: Sequences used for phylogenetic analysis of *A. hypochondriacus* DODA genes. Species, gene name and GenBank accession of input genes for phylogenetic analysis. DODA (L-3,4-dihydroxyphenylalanine 4,5-dioxygenase).

Species	Gene name	GenBank accession
Beta vulgaris	BvDODA α 1	MN153194
Beta vulgaris	BvDODA α 2	MN153195
Beta vulgaris	BvDODA β	MN153198
Chenopodium quinoa	CqDODA1 α 1	XP_021769303
Mirabilis jalapa	MjDOD	AB435372.1
Mesembryanthemum crystallinum	McDODA α 1	MN136225
Mesembryanthemum crystallinum	McDODA β	MN136227

Table S3.11: Summary of *A. hypochondriacus* flavonoid pathway candidate genes. Displayed are candidate gene names from genome annotation v2.2 alongside the percentage similarity, the percentage of conserved residues and the number of missing conserved residues as reported by KIPES. Missing pathway genes (defined by having no unique candidate genes or three or more missing conserved residues) were marked with asterisks and defined as unidentified in *A. hypochondriacus*. We added the ANS candidate gene identified by protein to genome alignments to this table. PAL (phenylalanine ammonium lyase), C4H (cinnamate 4-hydroxylase), 4CL (4-coumarate:CoA ligase), CHS (naringenin-chalcone synthase), CHI (chalcone isomerase), FNS (flavone synthase), F3H (flavanone 3-hydroxylase), FLS (flavonol synthase), F3'H (flavonoid 3'-hydroxylase), F3'5'H (flavonoid 3'5'-hydroxylase), DFR (dihydroflavonol 4-reductase), LAR (leucoanthocyanidin reductase), ANS (anthocyanidin synthase), ANR (anthocyanidin reductase).

Pathway gene	Candidate gene	Percentage similarity	Percentage conserved residues	Missing conserved residues
PAL	AHp012752	79.49	100.0	0
PAL	AHp021980	79.40	100.0	0
C4H	AHp013217	81.06	100.0	0
C4H	AHp022384	80.66	100.0	0
C4H	AHp022382	72.71	100.0	0
4CL	AHp014409	64.73	100.0	0
4CL	AHp020962	65.36	100.0	0
CHS	AHp004305	79.78	94.29	2
CHI1	AHp009962	59.97	100.0	0
CHI2*	no candidate	-	-	-
F3'H	AHp017497	68.15	100.0	0
F3'H	AHp022122	36.88	100.0	0
F3'H	AHp003152	38.80	100.0	0
F3'H	AHp022120	37.24	100.0	0
F3'H	AHp022123	25.78	100.0	0
F3'5'H*	no candidate	-	-	-
FNS1*	no candidate	-	-	-
F3H	AHp006454	79.68	100.0	0
FLS	AHp008991	62.03	100.0	0
DFR	AHp009303	61.99	96.43	1
LAR	AHp017409	57.70	100.0	0
ANS	AHp021795	-	-	-
ANR*	AHp001635	37.14	86.96	3

Table S3.12: MYB transcription factors identified in *A. hypochondriacus*. Genes were assigned into subgroups based on their phylogenetic placement in comparison to *A. thaliana* and *B. vulgaris* MYBs. Subgroup functions were assigned according to conserved subgroup functions.

Gene name	Gene id	Transcript id	Scaffold	Start	End	Strand	Subfamily	Protein length	Exon number	Subgroup	Functional assignment
-	AHp000437	AHp000437.1	Scaffold_1	4347523	4351157	+	R2R3	270	3	S16	Photomorphogenesis
-	AHp000789	AHp000789.1	Scaffold_1	10090637	10092364	+	R2R3	255	2	S4	Repressors of phenylpropanoids, sinapate, lignin
-	AHp001699	AHp001699.1	Scaffold_1	29115865	29116587	+	R2R3	241	1	S22	Defense, stress response
-	AHp001699	AHp001699.2	Scaffold_1	29115865	29116656	+	R2R3	239	2	S22	Defense, stress response
-	AHp002277	AHp002277.1	Scaffold_1	35637528	35639576	-	R2R3	272	2	S4	Repressors of phenylpropanoids, sinapate, lignin
-	AHp002664	AHp002664.1	Scaffold_2	1693532	1697261	+	R2R3	274	3	S2	Abiotic stress response, flower morphogenesis
-	AHp002686	AHp002686.1	Scaffold_2	1885905	1892082	-	R2R3	317	3	S7	Flavonol, flavonoid pathway
-	AHp002839	AHp002839.1	Scaffold_2	3368079	3370731	+	R2R3	322	3	S1	Defense, stress response
-	AHp003887	AHp003887.1	Scaffold_2	26854106	26855787	+	R2R3	309	3	S14	Axillary meristem, root growth
-	AHp003952	AHp003952.1	Scaffold_2	27861374	27862381	-	R2R3	336	1	S22	Defense, stress response
-	AHp004623	AHp004623.1	Scaffold_2	34693851	34698394	-	R2R3	454	12	-	-
-	AHp005058	AHp005058.1	Scaffold_3	3408139	3411828	+	R2R3	359	4	S21	Cell wall, lignin, axillary meristem
-	AHp005110	AHp005110.1	Scaffold_3	3994528	4000152	+	R2R3	454	12	-	-
-	AHp005488	AHp005488.1	Scaffold_3	8254983	8258030	+	R2R3	269	3	S20	Defense, stress response
-	AHp005948	AHp005948.1	Scaffold_3	14823654	14827853	-	R2R3	307	3	-	-
-	AHp006062	AHp006062.1	Scaffold_3	18197344	18199249	-	R2R3	171	4	S24	-
-	AHp006213	AHp006213.1	Scaffold_3	24208195	24209124	+	R2R3	310	1	S22	Defense, stress response
-	AHp006562	AHp006562.1	Scaffold_4	605153	607206	+	R2R3	316	3	S1	Defense, stress response
-	AHp006562	AHp006562.2	Scaffold_4	605153	608233	+	R2R3	178	4	S1	Defense, stress response
-	AHp007248	AHp007248.1	Scaffold_4	11044442	11046030	+	R2R3	359	3	S10	-
-	AHp007516	AHp007516.1	Scaffold_4	18919997	18922556	+	R2R3	540	3	S18	Anther development, stress response
-	AHp007516	AHp007516.2	Scaffold_4	18919997	18922556	+	R2R3	535	3	S18	Anther development, stress response
-	AHp007908	AHp007908.1	Scaffold_4	24322377	24324732	-	R2R3	296	3	S4	Repressors of phenylpropanoids, sinapate, lignin
-	AHp008229	AHp008229.1	Scaffold_4	27535327	27540827	-	R2R3	353	3	S18	Anther development, stress response
-	AHp008229	AHp008229.2	Scaffold_4	27535327	27540827	-	R2R3	375	4	S18	Anther development, stress response
-	AHp008229	AHp008229.3	Scaffold_4	27536451	27540827	-	R2R3	310	2	S18	Anther development, stress response
-	AHp008362	AHp008362.1	Scaffold_5	579584	581724	+	R2R3	223	4	S14	Axillary meristem, root growth
-	AHp008362	AHp008362.2	Scaffold_5	579584	581724	+	R2R3	345	3	S14	Axillary meristem, root growth
-	AHp008410	AHp008410.1	Scaffold_5	1052626	1053435	-	R2R3	270	1	S22	Defense, stress response
-	AHp008675	AHp008675.1	Scaffold_5	4050096	4052532	+	R2R3	348	2	S21	Cell wall, lignin, axillary meristem
-	AHp009283	AHp009283.1	Scaffold_5	19207042	19209867	+	R2R3	301	3	S9	Trichome branching, petal morphogenesis
-	AHp009452	AHp009452.1	Scaffold_5	21916376	21917921	+	R2R3	315	3	S4	Repressors of phenylpropanoids, sinapate, lignin
-	AHp009484	AHp009484.1	Scaffold_5	22269940	22274870	+	R2R3	305	3	S3	Phenylpropanoid pathway, lignin
-	AHp009565	AHp009565.1	Scaffold_5	23545578	23547934	+	R2R3	348	2	AtMYB5-like	Development, proanthocyanidin biosynthesis
-	AHp009926	AHp009926.1	Scaffold_6	2944929	2948000	+	R2R3	330	3	-	-
-	AHp010652	AHp010652.1	Scaffold_6	17732890	17739448	-	R2R3	974	4	-	-
-	AHp010658	AHp010658.1	Scaffold_6	17860532	17862437	+	R2R3	420	3	S13	Mucilage, lignin, stomatal closure
-	AHp011052	AHp011052.1	Scaffold_6	22330077	22334786	-	R2R3	411	2	S16	Photomorphogenesis
-	AHp011153	AHp011153.1	Scaffold_6	23290312	23291464	-	R2R3	300	2	S18	Anther development, stress response
-	AHp011578	AHp011578.1	Scaffold_7	10152532	10155066	-	R2R3	334	3	-	-
-	AHp011578	AHp011578.2	Scaffold_7	10153374	10155066	-	R2R3	334	2	-	-
-	AHp011866	AHp011866.1	Scaffold_7	16613289	16616126	+	R2R3	314	3	S20	Defense, stress response
-	AHp012381	AHp012381.1	Scaffold_7	23842541	23847434	+	R2R3	254	3	S16	Photomorphogenesis
-	AHp012382	AHp012382.1	Scaffold_7	23851791	23855945	-	R2R3	289	3	S16	Photomorphogenesis

Gene name	Gene id	Transcript id	Scaffold	Start	End	Strand	Subfamily	Protein length	Exon number	Subgroup	Functional assignment
-	AHp012383	AHp012383.1	Scaffold_7	23881163	23885856	+	R2R3	250	3	S9	Trichome branching, petal morphogenesis
-	AHp012732	AHp012732.1	Scaffold_8	3120126	3122087	+	R2R3	375	3	S14	Axillary meristem, root growth
-	AHp013131	AHp013131.1	Scaffold_8	6814523	6815510	+	R2R3	255	3	S17	-
-	AHp013326	AHp013326.1	Scaffold_8	9237650	9242324	-	R2R3	403	3	-	-
-	AHp013775	AHp013775.1	Scaffold_8	16914029	16918169	-	R2R3	411	3	S25	Embryogenesis, seed maturation
-	AHp014406	AHp014406.1	Scaffold_9	12055158	12058160	+	R2R3	350	3	S14	Axillary meristem, root growth
-	AHp014458	AHp014458.1	Scaffold_9	12904501	12907181	-	R2R3	242	3	S14	Axillary meristem, root growth
AmMYB11	AHp014591	AHp014591.1	Scaffold_9	14711535	14714080	+	R2R3	283	4	S5	Proanthocyanidin biosynthesis
-	AHp014665	AHp014665.1	Scaffold_9	15476010	15477107	-	R2R3	366	1	-	-
-	AHp015175	AHp015175.1	Scaffold_9	20978046	20979995	-	R2R3	392	3	S11	Defense, stress response
-	AHp015176	AHp015176.1	Scaffold_9	20993200	20994999	-	R2R3	359	3	S11	Defense, stress response
-	AHp015766	AHp015766.1	Scaffold_10	12455733	12458734	+	R2R3	351	3	S1	Defense, stress response
-	AHp015766	AHp015766.2	Scaffold_10	12455733	12458734	+	R2R3	341	3	S1	Defense, stress response
-	AHp015783	AHp015783.1	Scaffold_10	12825187	12827217	-	R2R3	452	3	S9	Trichome branching, petal morphogenesis
-	AHp016111	AHp016111.1	Scaffold_10	16835722	16838292	-	R2R3	325	3	S14	Axillary meristem, root growth
-	AHp016387	AHp016387.1	Scaffold_10	20282177	20286204	+	R2R3	406	2	S23	-
-	AHp016456	AHp016456.1	Scaffold_10	21051032	21053433	+	R2R3	318	4	-	-
AhMYB3	AHp016530	AHp016530.1	Scaffold_10	21762621	21764924	-	R2R3	216	3	BvMYB1-like	Betalain biosynthesis
AhMYB4	AHp016531	AHp016531.1	Scaffold_10	21780565	21783050	-	R2R3	228	3	BvMYB1-like	Betalain biosynthesis
-	AHp016620	AHp016620.1	Scaffold_10	22596330	22599342	-	R2R3	341	2	S16	Photomorphogenesis
-	AHp017558	AHp017558.1	Scaffold_11	17970779	17972903	+	R2R3	293	3	S1	Defense, stress response
-	AHp017657	AHp017657.1	Scaffold_11	18887019	18889370	-	R2R3	289	3	S20	Defense, stress response
-	AHp017658	AHp017658.1	Scaffold_11	18908562	18910154	-	R2R3	248	3	S20	Defense, stress response
-	AHp017689	AHp017689.1	Scaffold_11	19220858	19226127	-	R2R3	341	3	S7	Flavonol, flavonoid pathway
-	AHp017915	AHp017915.1	Scaffold_11	21654178	21657013	-	R2R3	362	3	-	-
-	AHp017946	AHp017946.1	Scaffold_11	22019498	22021190	-	R2R3	294	3	S25	Embryogenesis, seed maturation
-	AHp017979	AHp017979.1	Scaffold_12	139570	144579	-	R2R3	358	3	-	-
-	AHp018270	AHp018270.1	Scaffold_12	3175247	3177848	-	R2R3	188	3	S5	Proanthocyanidin biosynthesis
-	AHp018383	AHp018383.1	Scaffold_12	4698125	4699671	-	R2R3	346	2	-	-
-	AHp018492	AHp018492.1	Scaffold_12	5906028	5907870	-	R2R3	306	2	S8	Lignin
-	AHp018627	AHp018627.1	Scaffold_12	7669283	7672257	+	3R	513	7	3R	-
-	AHp019036	AHp019036.1	Scaffold_12	15482171	15486341	-	R2R3	356	3	S21	Cell wall, lignin, axillary meristem
-	AHp019040	AHp019040.1	Scaffold_12	15614348	15629179	+	4R	900	13	4R	-
-	AHp019040	AHp019040.2	Scaffold_12	15614348	15629179	+	4R	902	13	4R	-
-	AHp019088	AHp019088.1	Scaffold_12	16532558	16544063	+	3R	541	7	3R	-
-	AHp019259	AHp019259.1	Scaffold_12	20323254	20324899	+	R2R3	313	2	-	-
-	AHp019341	AHp019341.1	Scaffold_12	21674277	21675477	-	R2R3	335	3	-	-
-	AHp019341	AHp019341.2	Scaffold_12	21674277	21675477	-	R2R3	325	3	-	-
-	AHp020198	AHp020198.1	Scaffold_13	17051563	17053916	+	R2R3	259	3	S17	-
-	AHp020524	AHp020524.1	Scaffold_13	20104457	20106612	-	R2R3	284	3	S1	Defense, stress response
-	AHp020968	AHp020968.1	Scaffold_14	12366084	12368376	-	R2R3	285	3	S14	Axillary meristem, root growth
-	AHp021016	AHp021016.1	Scaffold_14	13270874	13277249	+	R2R3	412	3	S25	Embryogenesis, seed maturation
-	AHp021152	AHp021152.1	Scaffold_14	14968799	14969911	-	R2R3	371	1	-	-
-	AHp021566	AHp021566.1	Scaffold_14	19210847	19212496	-	R2R3	381	3	S11	Defense, stress response
-	AHp021567	AHp021567.1	Scaffold_14	19222097	19225148	-	R2R3	381	3	S11	Defense, stress response
-	AHp021873	AHp021873.1	Scaffold_15	1897249	1899908	-	R2R3	278	3	S17	-
-	AHp022482	AHp022482.1	Scaffold_15	10193355	10194334	+	R2R3	260	3	S17	-

Gene name	Gene id	Transcript id	Scaffold	Start	End	Strand	Subfamily	Protein length	Exon number	Subgroup	Functional assignment
-	AHp022746	AHp022746.1	Scaffold_16	532237	533511	-	R2R3	274	3	-	-
AhMYB2	AHp022773	AHp022773.1	Scaffold_16	988734	991447	+	R2R3	236	3	BvMYB1-like	Betalain biosynthesis
AhMYB2	AHp022773	AHp022773.2	Scaffold_16	988734	991447	+	R2R3	240	3	BvMYB1-like	Betalain biosynthesis
-	AHp022995	AHp022995.1	Scaffold_16	3273279	3276051	-	R2R3	422	2	S23	-
-	AHp023085	AHp023085.1	Scaffold_16	4378346	4379815	+	R2R3	243	2	-	-
-	AHp023286	AHp023286.1	Scaffold_16	7455684	7465379	-	3R	1017	11	3R	-

3.12.4 Supplementary figures

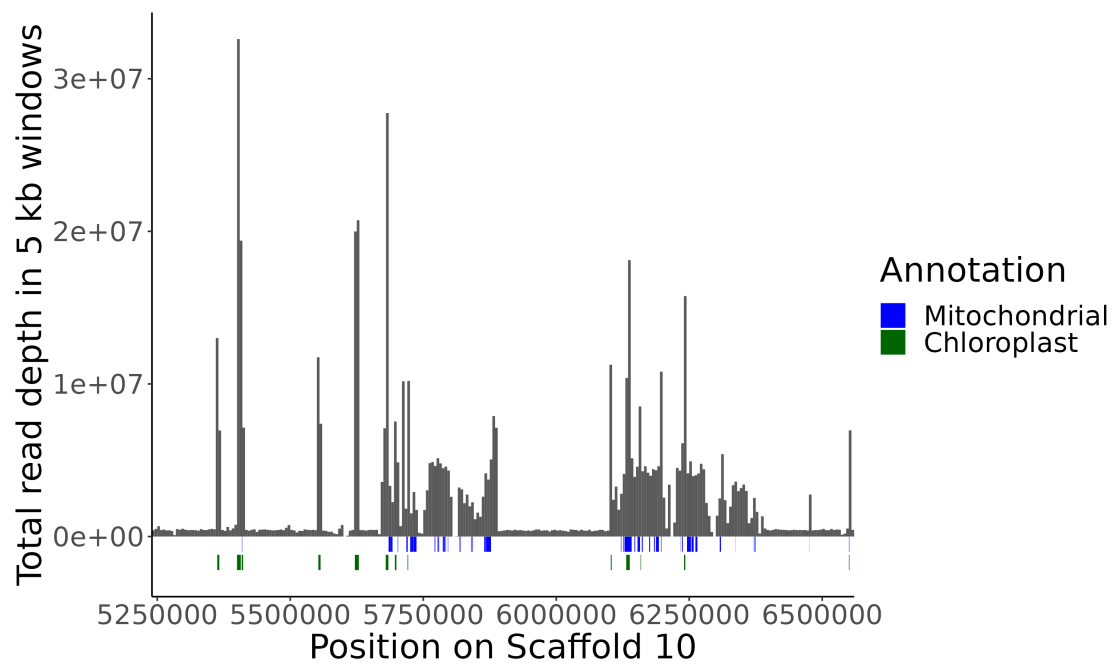


Figure S3.1: Organellar contamination in the *A. hypochondriacus* assembly. Total read depth of whole genome sequencing data in 5 kb windows along Scaffold 10 region with mapping bias. High read depth regions correspond partially to annotated chloroplast and mitochondrial regions in the genome.

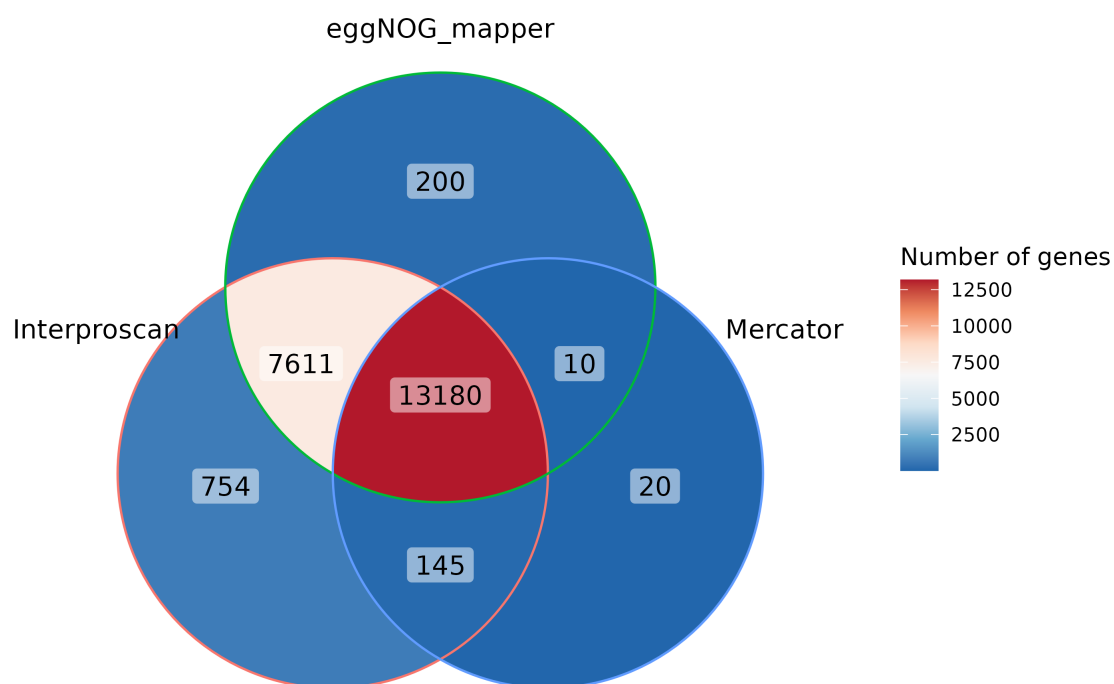


Figure S3.2: Summary of the functional annotation of the *A. hypochondriacus* reference genome v2.2. Overlap between functionally annotated genes using the annotation tools Interproscan (considering only genes with identified Pfam, PANTHER or CDD domains as annotated), eggNOG-mapper and Mercator.

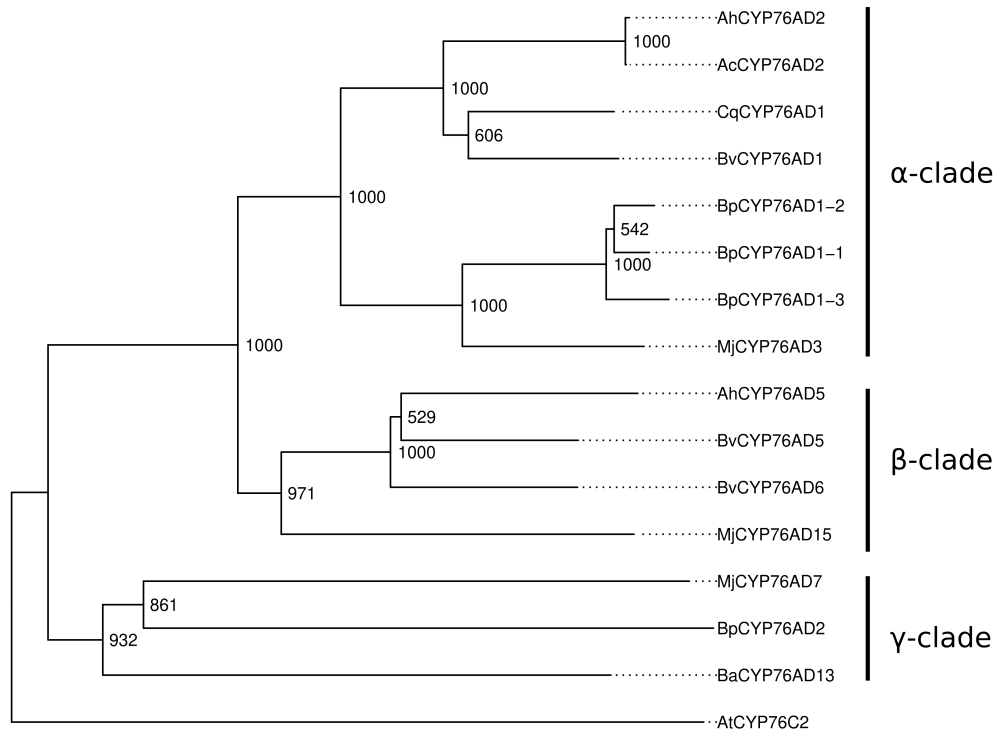


Figure S3.3: Phylogenetic analysis of *A. hypochondriacus* CYP76AD proteins with CYP protein sequences from other species. CYP76AD α -, β - and γ -clades are annotated in the alignment. Bootstrap values from 1000 replicates are depicted on the tree. The following sequences were used in the phylogeny, with GenBank accession in parenthesis: *Amaranthus cruentus*: AcCYP76AD2 (AET43291.1); *Arabidopsis thaliana*: AtCYP76C2 (NP_182081); *Basella alba*: BaCYP76AD13 (AJD87469); *Bougainvillea peruviana*: BpCYP76AD1-1 (BCD59210), BpCYP76AD1-2 (BCD59211), BpCYP76AD1-3 (BCD59212), BpCYP76AD2 (BCD59213); *Beta vulgaris*: BvCYP76AD1 (AET43290), BvCYP76AD5 (AJD87473), BvCYP76AD6 (AJD87474); *Chenopodium quinoa*: CqCYP76AD1 (XP_021769302); *Mirabilis jalapa*: MjCYP76AD3 (AET43292), MjCYP76AD7 (AJD87463), MjCYP76AD15 (AJD87471).

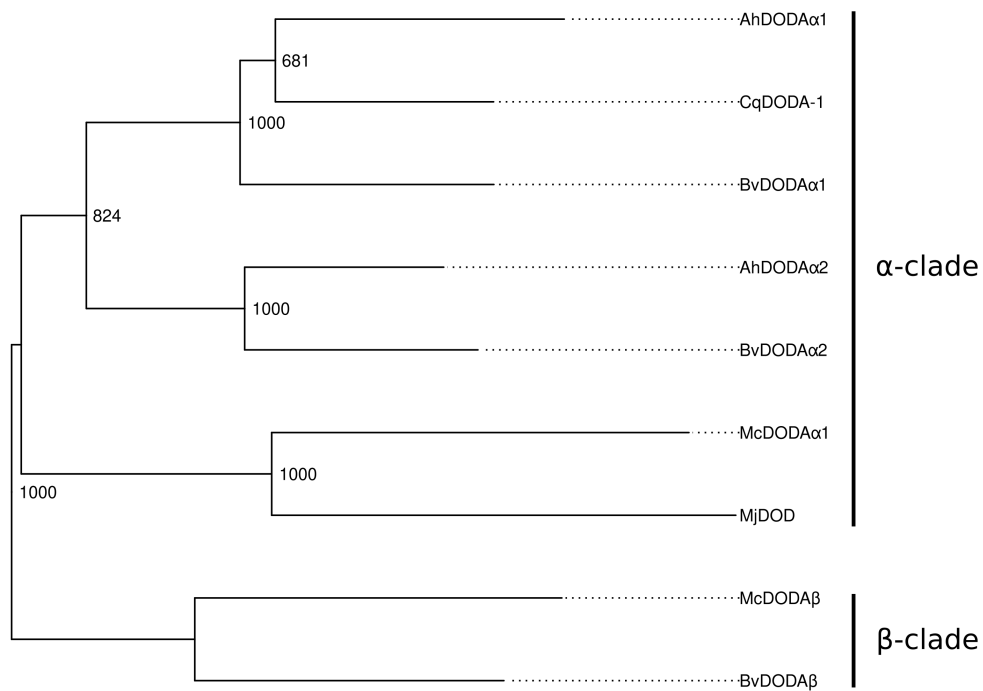


Figure S3.4: Phylogenetic analysis of *A. hypochondriacus* DODA proteins with DODA protein sequences from other species. DODA α - and β -clades are annotated in the alignment. Bootstrap values from 1000 replicates are depicted on the tree. The following sequences were used in the phylogeny, with GenBank accession in parenthesis: *Beta vulgaris*: BvDODA α 1 (MN153194), BvDODA α 2 (MN153195), BvDODA β (MN153198); *Chenopodium quinoa*: CqDODA1 α 1 (XP_021769303); *Mirabilis jalapa*: MjDOD (AB435372.1); *Mesembryanthemum crystallinum*: McDODA α 1 (MN136225), McDODA β (MN136227).

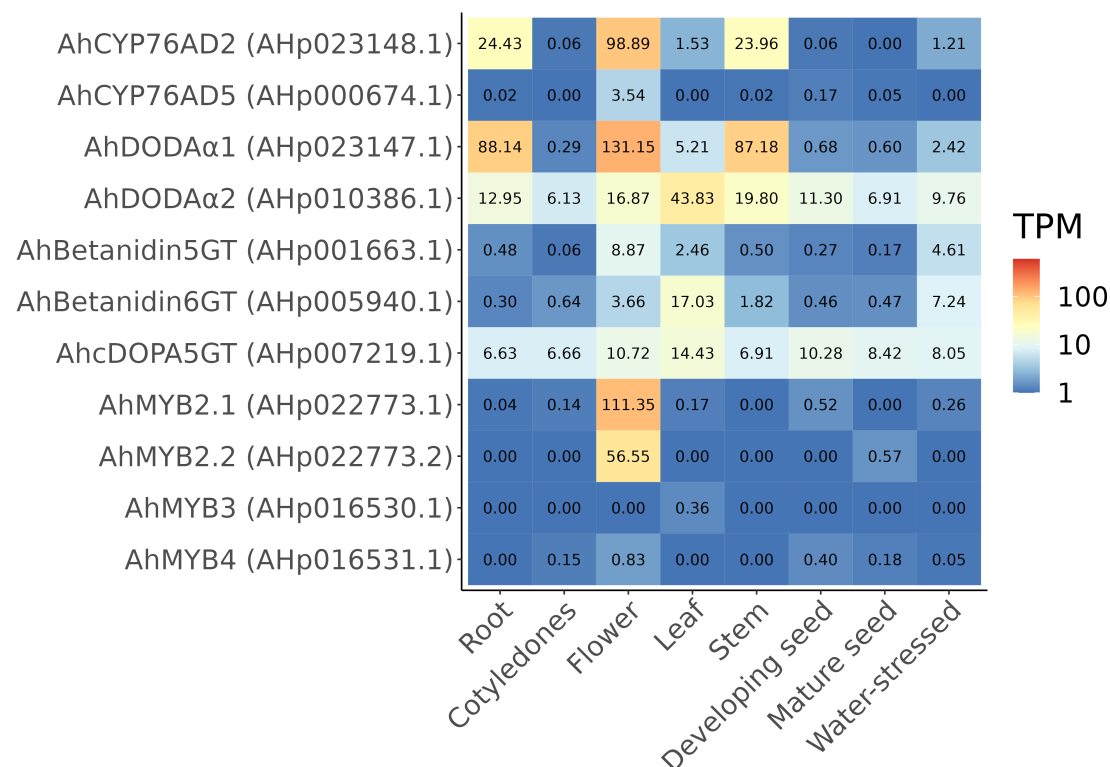


Figure S3.5: Relative gene expression in transcript per million (TPM) of identified *A. hypochondriacus* betalain pathway genes in different tissues. Quantification was done using kallisto from short-read RNA sequencing data. DODA (L-3,4-dihydroxyphenylalanine 4,5-dioxygenase), cDOPA5GT (*cyclo*-DOPA 5-O-glucosyltransferase), Betanidin5GT (betanidin 5-O-glucosyltransferase), Betanidin6GT (betanidin 6-O-glucosyltransferase).

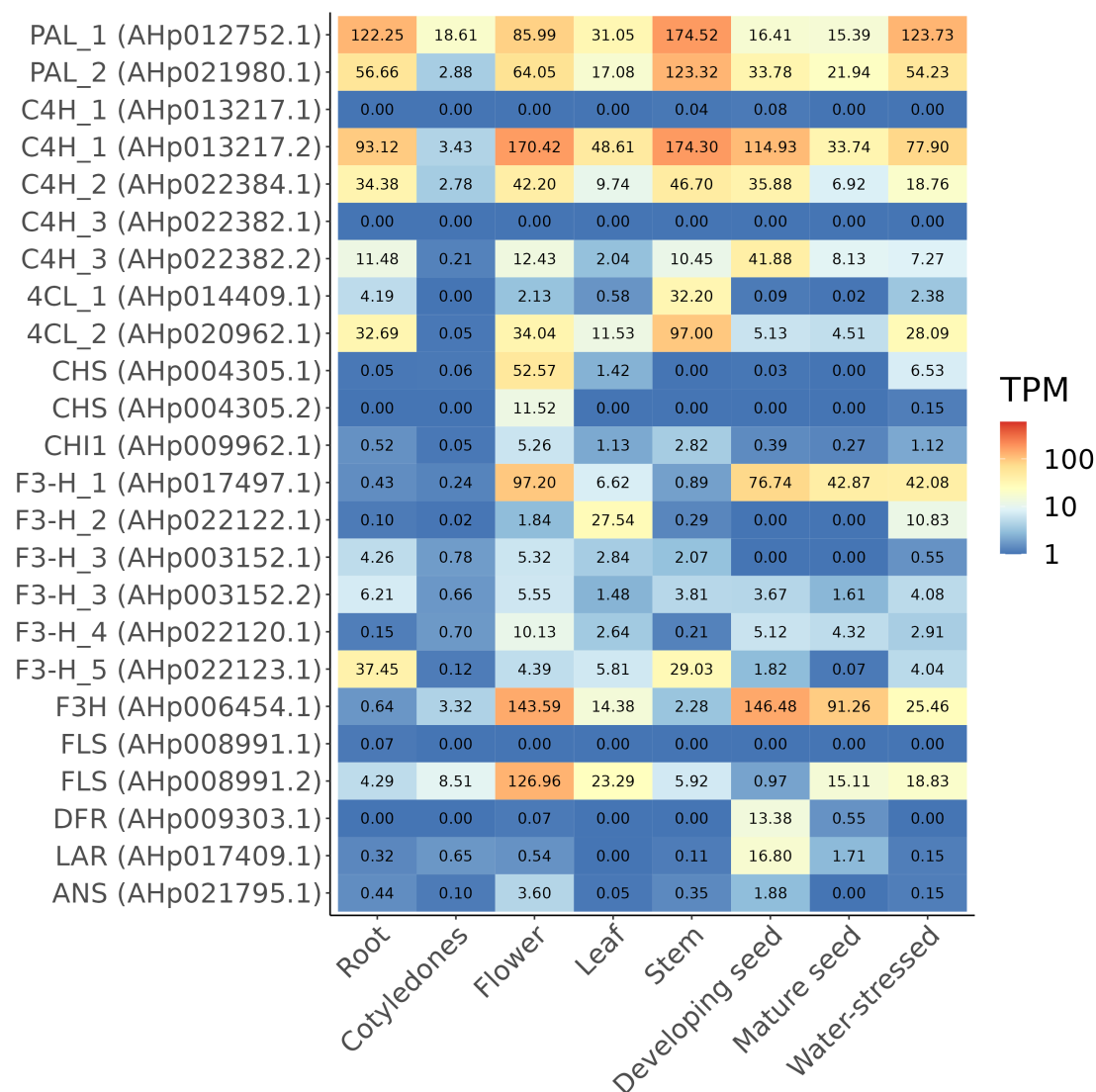


Figure S3.6: Relative gene expression in transcript per million (TPM) of identified *A. hypochondriacus* flavonoid pathway genes in different tissues. Quantification was done using kallisto from short-read RNA sequencing data. PAL (phenylalanine ammonium lyase), C4H (cinnamate 4-hydroxylase), 4CL (4-coumarate:CoA ligase), CHS (naringenin-chalcone synthase), CHI (chalcone isomerase), F3'H (flavonoid 3'-hydroxylase), F3H (flavanone 3-hydroxylase), FLS (flavonol synthase), DFR (dihydroflavonol 4-reductase), LAR (leucoanthocyanidin reductase), ANS (anthocyanidin synthase).

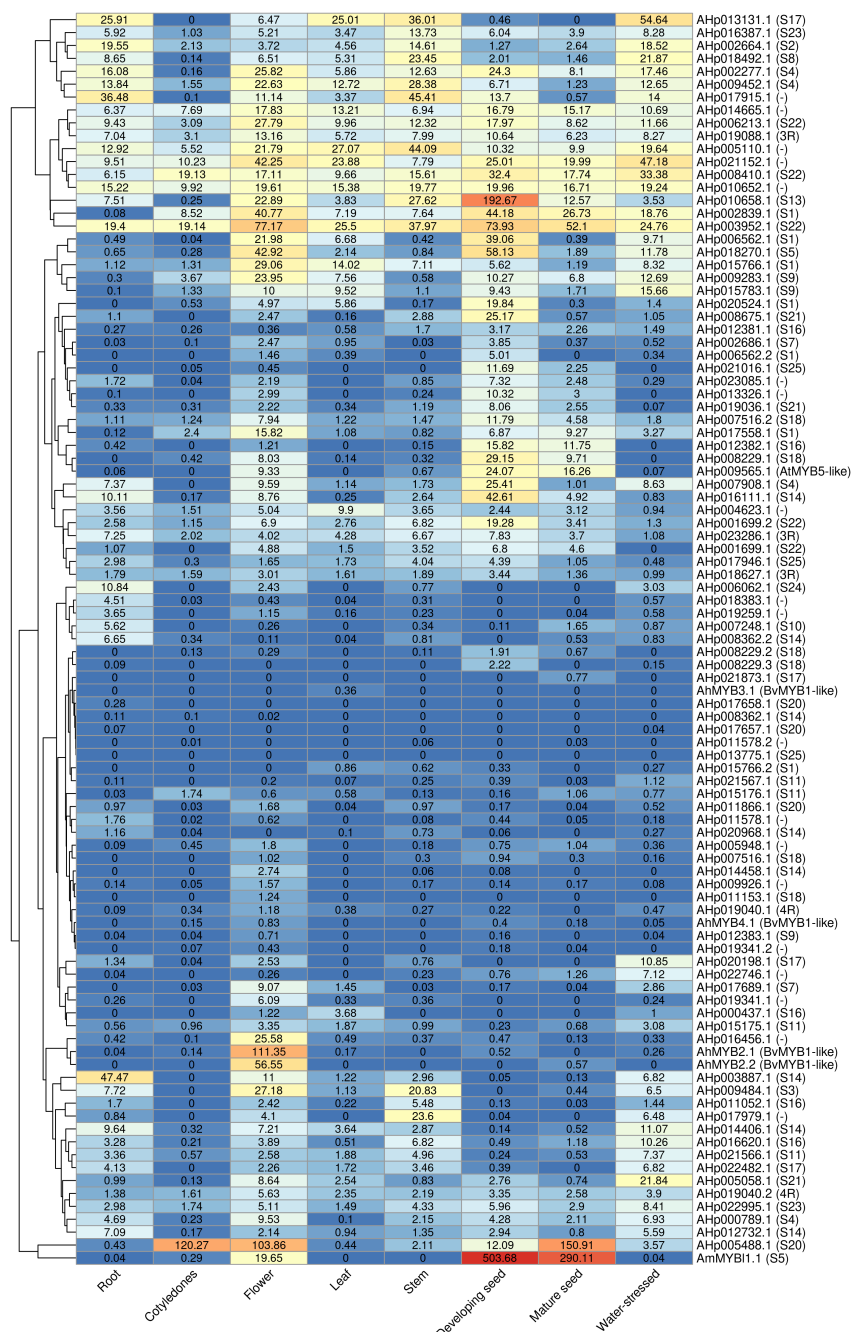


Figure S3.7: Relative gene expression in transcript per million (TPM) of identified *A. hypochondriacus* MYB transcription factor genes in different tissues. Quantification was done using kallisto from short-read RNA sequencing data.

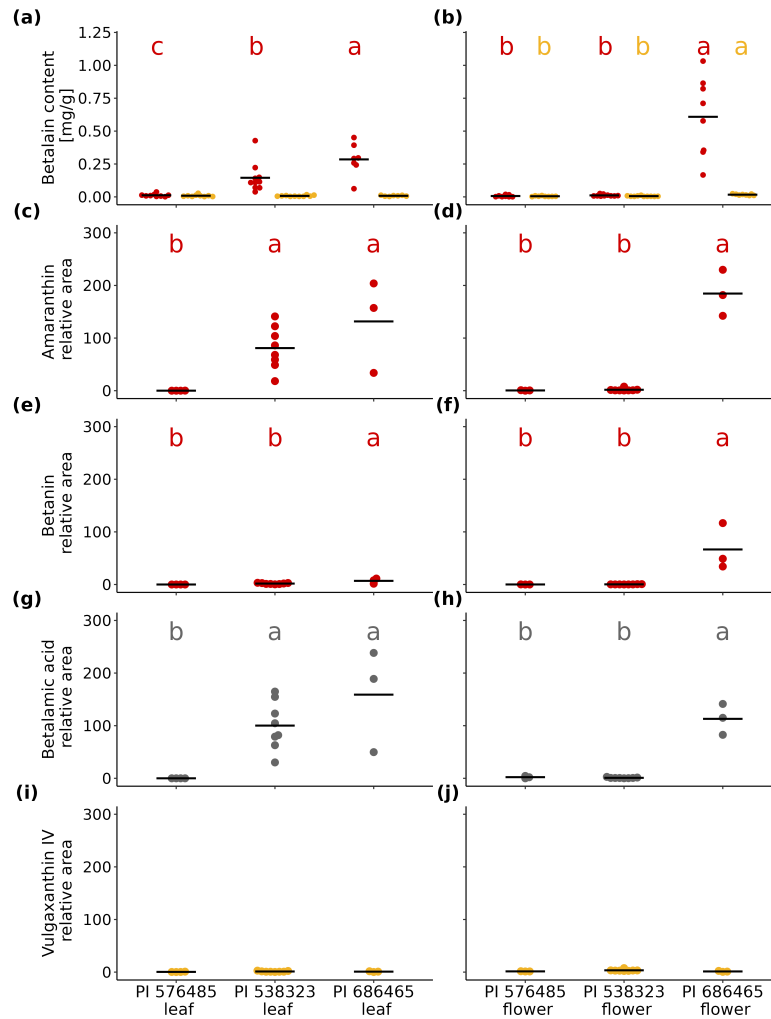


Figure S3.8: Photometric and LC-MS quantification of betalain pathway metabolites for bulk segregant analysis (BSA) *A. hypochondriacus* parental genotypes. (a-b) Photometric quantification of betacyanin (red) and betaxanthin (yellow) content of (a) leaf and (b) flower tissue (mg betalains per g fresh weight), with black bars denoting mean values. (c-f) Relative quantification by LC-MS of the betacyanins amaranthin (c, d) and betanin (e, f) in leaf and flower, respectively. (g-h) Relative quantification by LC-MS of betalamic acid as common chromophore of betacyanins and betaxanthins in (g) leaf and (h) flower. (i-j) Relative quantification by LC-MS of the betaxanthin vulgaxanthin IV in (i) leaf and (j) flower. Statistical analysis was conducted using ANOVA and significant differences ($p < 0.05$) in betaxanthin or betacyanin content between accessions within tissues were denoted using compact letter display. There were no significant differences in total betaxanthin content between accessions in leaf and no significant differences in vulgaxanthin IV content between accessions in leaf or flower.

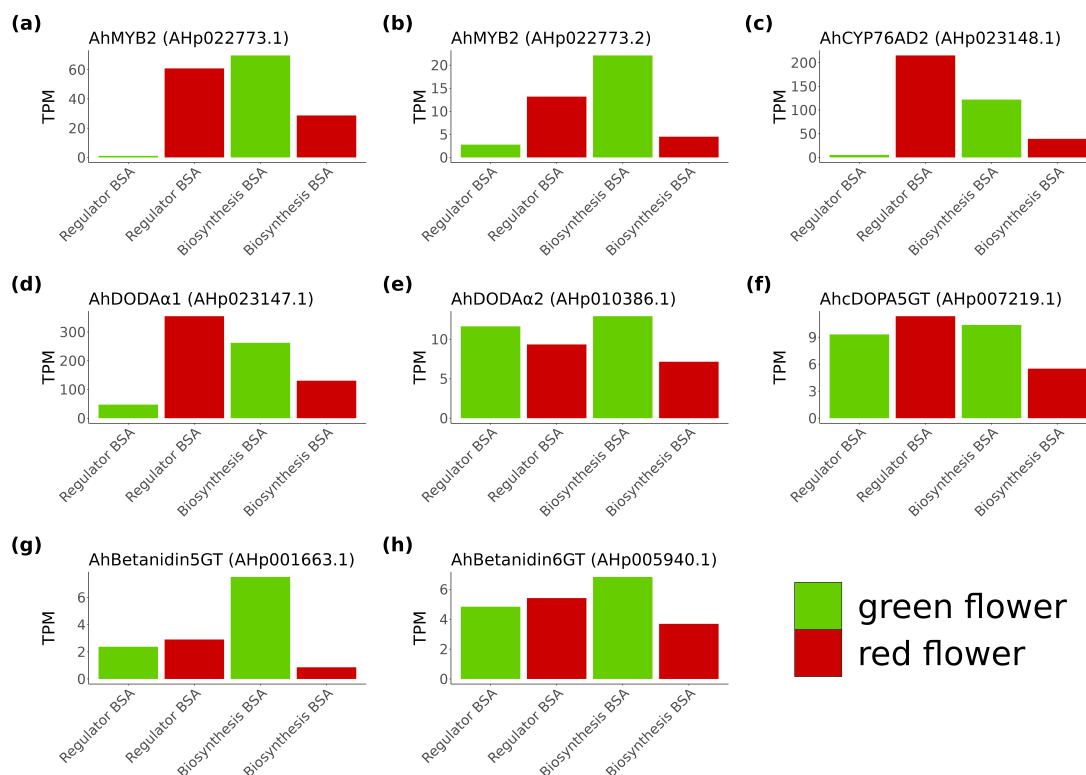


Figure S3.9: Gene expression of *A. hypochondriacus* betalain pathway genes in flower tissue in green and red bulks of regulator and biosynthesis bulk segregant analyses (BSAs). Relative gene expression in transcript per million (TPM) is displayed for *A. hypochondriacus* betalain pathway genes and candidate regulators. *AhCYP76AD5*, *AhMYB3* and *AhMYB4* had expression < 0.5 TPM in flower tissue in all samples and were excluded. DODA (L-3,4-dihydroxyphenylalanine 4,5-dioxygenase), cDOPA5GT (*cyclo*-DOPA 5-O-glucosyltransferase), Betanidin5GT (betanidin 5-O-glucosyltransferase), Betanidin6GT (betanidin 6-O-glucosyltransferase).

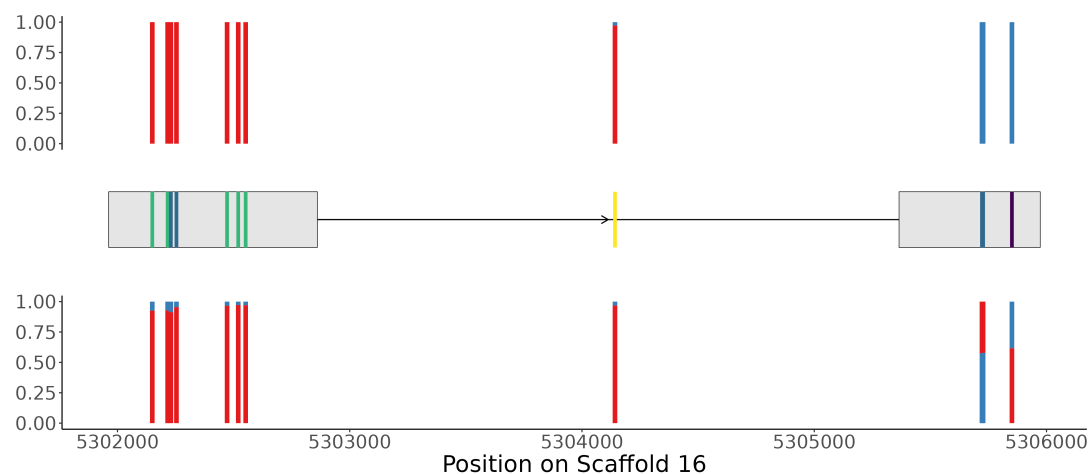


Figure S3.10: Segregating variants called from *A. hypochondriacus* whole-genome sequencing data in the biosynthesis BSA and their relative allele depth for the *AhCYP76AD2* gene. Top row shows the relative depth of reference (blue) and alternative allele (red) for the red bulk, the bottom row the relative depth for the green bulk. Middle part shows the exon structure of the gene and all variants with their predicted effect compared to the reference genome (yellow: intron variant, green: synonymous variant, dark blue: missense variant, purple: stop-gain variant).

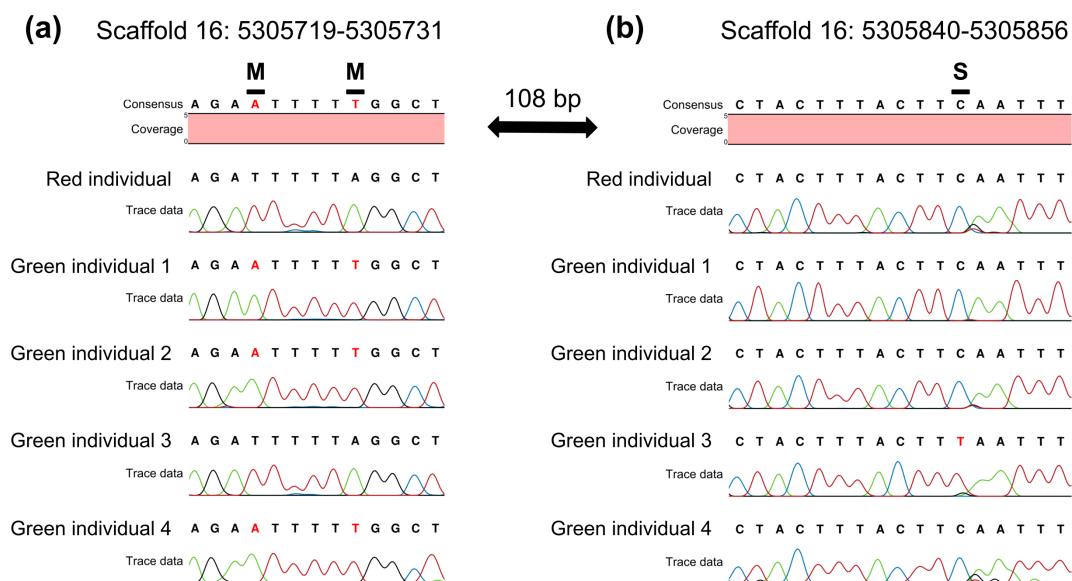


Figure S3.12: Sanger sequencing of *AhCYP76AD2* from biosynthesis bulk segregant analysis (BSA) individuals. Sanger sequencing of *AhCYP76AD2* (primer SV4) from *A. hypochondriacus* individuals from biosynthesis BSA were mapped to the *AhCYP76AD2* genomic sequence. (a-b) One individual was chosen from the red bulk (top), the other four individuals were from the green bulk. Positions differing from the reference genome were marked in red. (a) Positions of the two missense variants T1258A and A1263T were marked with M, (b) the stop variant C1387T was marked with S.

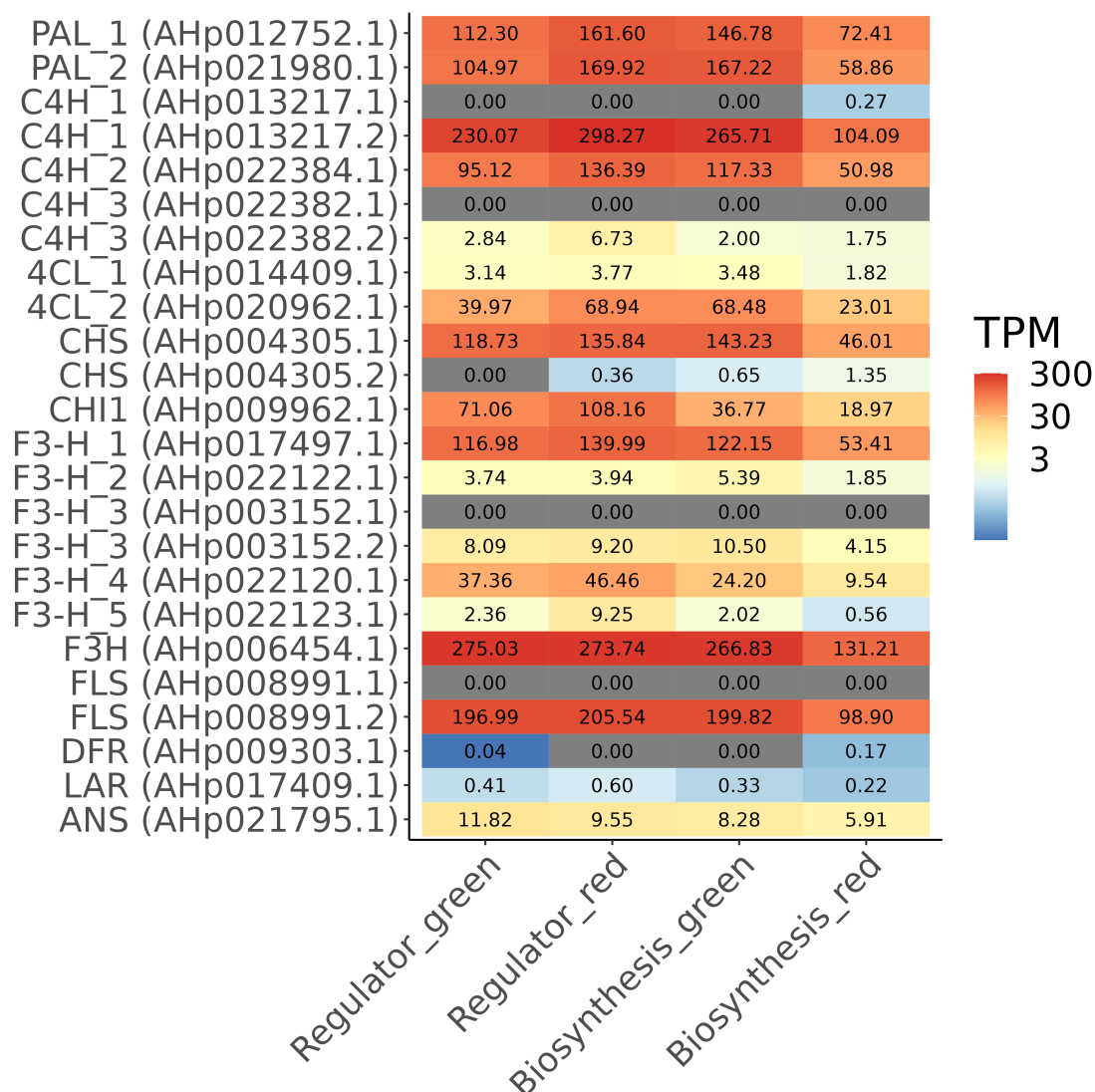


Figure S3.13: Relative gene expression in transcript per million (TPM) of identified *A. hypochondriacus* flavonoid pathway genes in red and green flowers. Quantification was done using kallisto from short-read RNA sequencing data generated from bulked flower tissue from the regulator and biosynthesis bulk segregant analyses (BSAs). PAL (phenylalanine ammonium lyase), C4H (cinnamate 4-hydroxylase), 4CL (4-coumarate:CoA ligase), CHS (naringenin-chalcone synthase), CHI (chalcone isomerase), F3'H (flavonoid 3'-hydroxylase), F3H (flavanone 3-hydroxylase), FLS (flavonol synthase), DFR (dihydroflavonol 4-reductase), LAR (leucoanthocyanidin reductase), ANS (anthocyanidin synthase).

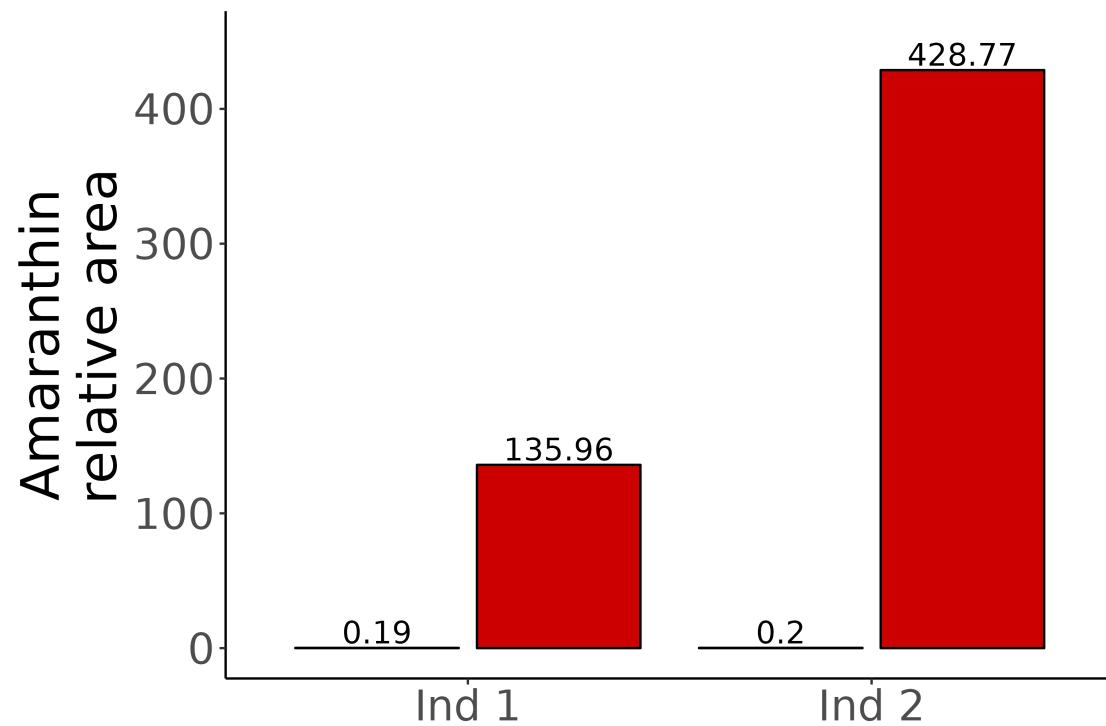


Figure S3.14: Relative quantification of amaranthin content from transgenic *A. hypochondriacus* hairy roots. Relative quantification by liquid chromatography - mass spectrometry (LC-MS) of amaranthin (betacyanin) content in wild-type like (beige) and transgenic (red) hairy roots transformed with *p35S:AhMYB2.1-GFP*. For quantification, both wild-type like and transgenic hairy root tissue was sampled from two successfully transformed individuals.

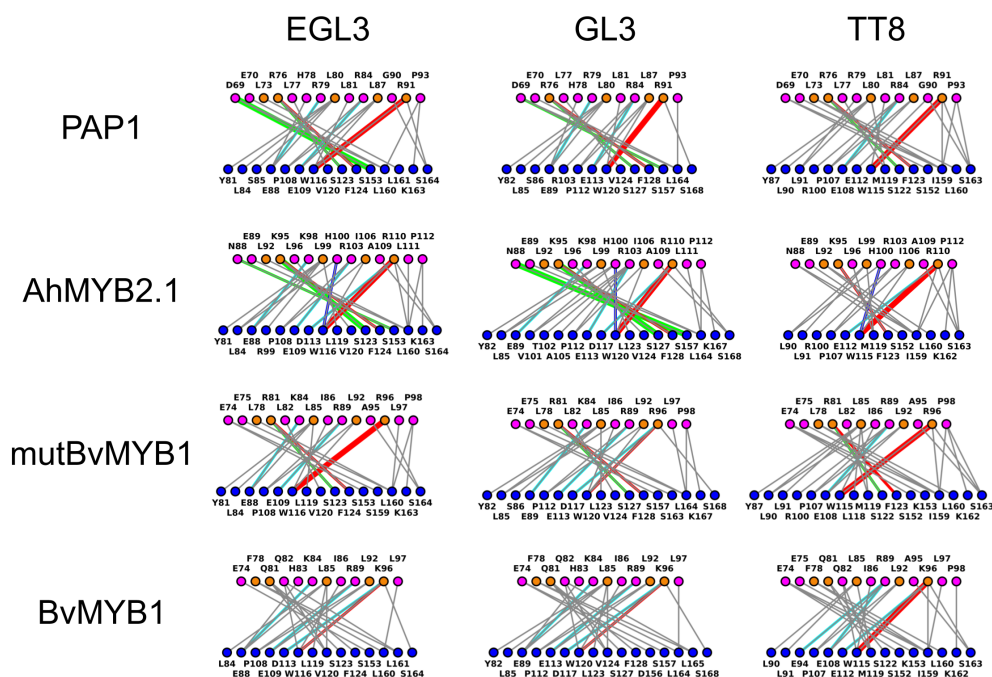


Figure S3.15: Prediction of protein-protein interactions of selected MYB and bHLH proteins. Pairwise prediction of directly interacting residues in the R3 domain of *A. thaliana* PAP1 (At-MYB75), *A. hypochondriacus* AhMYB2.1, *B. vulgaris* BvMYB1, and a mutated BvMYB1 (mut-BvMYB1), with restored ability to interact with bHLH proteins (Hatlestad et al., 2015), with the known *A. thaliana* PAP1 - bHLH interaction partners EGL3, GL3 and TT8. Protein residues are represented by filled circles. The upper row of each prediction depicts MYB R3 residues predicted as interacting, the lower row depicts bHLH residues predicted as interacting with the R3 MYB residues. MYB residues corresponding to the bHLH-interaction motif in PAP1 were marked in orange. Interactions between residues were colored according to the predicted interaction type: grey, direct contacts/interactions; green, hydrogen bonds; light blue, salt bridge/ionic; red, π -cation; dark blue, π -stacking.

3.13 Supplementary references

- S. Altschul. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–3402, Sept. 1997. ISSN 13624962. doi: 10.1093/nar/25.17.3389. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/25.17.3389>.
- S. Andrews. FastQC: A Quality Control Tool for High Throughput Sequence Data, 2010. URL <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- T. Brûna, K. J. Hoff, A. Lomsadze, M. Stanke, and M. Borodovsky. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, 3(1):lqaa108, Jan. 2021. ISSN 2631-9268. doi: 10.1093/nargab/lqaa108. URL <https://academic.oup.com/nargab/article/doi/10.1093/nargab/lqaa108/6066535>.
- G. Casique-Arroyo, N. Martínez-Gallardo, L. González de la Vara, and J. P. Délano-Frier. Betacyanin Biosynthetic Genes and Enzymes Are Differentially Induced by (a)biotic Stress in *Amaranthus hypochondriacus*. *PLoS ONE*, 9(6):e99012, June 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0099012. URL <https://dx.plos.org/10.1371/journal.pone.0099012>.
- L. Chaney, R. Mangelson, T. Ramaraj, E. N. Jellen, and P. J. Maughan. The complete chloroplast genome sequences for four *Amaranthus* species (*Amaranthaceae*). *Applications in Plant Sciences*, 4(9):1600063, Sept. 2016. ISSN 2168-0450, 2168-0450. doi: 10.3732/apps.1600063. URL <https://bsapubs.onlinelibrary.wiley.com/doi/10.3732/apps.1600063>.
- Y.-C. Chang, Y.-C. Chiu, N.-W. Tsao, Y.-L. Chou, C.-M. Tan, Y.-H. Chiang, P.-C. Liao, Y.-C. Lee, L.-C. Hsieh, S.-Y. Wang, and J.-Y. Yang. Elucidation of the core betalain biosynthesis pathway in *Amaranthus tricolor*. *Scientific Reports*, 11(1):6086, Dec. 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-85486-x. URL <http://www.nature.com/articles/s41598-021-85486-x>.
- N. Chen. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Current Protocols in Bioinformatics*, 5(1):4.10.1–4.10.14, 2004. Publisher: Wiley Online Library.
- M. A. Cianfrocco, M. Wong-Barnum, C. Youn, R. Wagner, and A. Leschziner. COSMIC2: A Science Gateway for Cryo-Electron Microscopy Structure Determination. In *Practice and Experience in Advanced Research Computing 2017: Sustainability, Success and Impact*, pages 1–5, New Orleans LA USA, July 2017. ACM. ISBN 978-1-4503-5272-7. doi: 10.1145/3093338.3093390. URL <https://dl.acm.org/doi/10.1145/3093338.3093390>.
- J. W. Clouse, D. Adhikary, J. T. Page, T. Ramaraj, M. K. Deyholos, J. A. Udall, D. J. Fairbanks, E. N. Jellen, and P. J. Maughan. The Amaranth Genome: Genome, Transcriptome, and Physical Map Assembly. *The Plant Genome*, 9(1):plantgenome2015.07.0062, Mar. 2016. ISSN 1940-3372, 1940-3372. doi: 10.3835/plantgenome2015.07.0062. URL <https://acsess.onlinelibrary.wiley.com/doi/10.3835/plantgenome2015.07.0062>.

- A. Dobin and T. R. Gingeras. Mapping RNA-seq reads with STAR. *Current Protocols in Bioinformatics*, 51(1):11–14, 2015. Publisher: Wiley Online Library.
- C. Dubos, R. Stracke, E. Grotewold, B. Weisshaar, C. Martin, and L. Lepiniec. MYB transcription factors in *Arabidopsis*. *Trends in Plant Science*, 15(10):573–581, Oct. 2010. ISSN 13601385. doi: 10.1016/j.tplants.2010.06.005. URL <https://linkinghub.elsevier.com/retrieve/pii/S1360138510001536>.
- S. R. Eddy. A new generation of homology search tools based on probabilistic inference. *Genome Informatics. International Conference on Genome Informatics*, 23(1):205–211, Oct. 2009. ISSN 0919-9454.
- J. Escribano, J. Cabanes, M. Jiménez-Atiénzar, M. Ibañez-Tremolada, L. R. Gómez-Pando, F. García-Carmona, and F. Gandía-Herrero. Characterization of betalains, saponins and antioxidant power in differently colored quinoa (*Chenopodium quinoa*) varieties. *Food Chemistry*, 234:285–294, Nov. 2017. ISSN 03088146. doi: 10.1016/j.foodchem.2017.04.187. URL <https://linkinghub.elsevier.com/retrieve/pii/S0308814617307756>.
- J. M. Flynn, R. Hubley, C. Goubert, J. Rosen, A. G. Clark, C. Feschotte, and A. F. Smit. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, 117(17):9451–9457, Apr. 2020. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1921046117. URL <https://pnas.org/doi/full/10.1073/pnas.1921046117>.
- Z. Gu, L. Gu, R. Eils, M. Schlesner, and B. Brors. *circlize* implements and enhances circular visualization in R. *Bioinformatics*, 30(19):2811–2812, 2014. Publisher: Oxford University Press.
- G. J. Hatlestad, R. M. Sunnadeniya, N. A. Akhavan, A. Gonzalez, I. L. Goldman, J. M. McGrath, and A. M. Lloyd. The beet R locus encodes a new cytochrome P450 required for red betalain production. *Nature Genetics*, 44(7):816–820, July 2012. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.2297. URL <http://www.nature.com/articles/ng.2297>.
- G. J. Hatlestad, N. A. Akhavan, R. M. Sunnadeniya, L. Elam, S. Cargile, A. Hembd, A. Gonzalez, J. M. McGrath, and A. M. Lloyd. The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway. *Nature Genetics*, 47(1):92–96, Jan. 2015. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3163. URL <http://www.nature.com/articles/ng.3163>.
- J. Hu, J. Fan, Z. Sun, and S. Liu. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*, 36(7):2253–2255, Apr. 2020. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/btz891. URL <https://academic.oup.com/bioinformatics/article/36/7/2253/5645175>.
- G. Jain, K. E. Schwinn, and K. S. Gould. Betalain induction by *l* - *DOPA* application confers photoprotection to saline-exposed leaves of *D. isophya australe*. *New Phytologist*, 207(4):1075–1083, Sept. 2015. ISSN 0028-646X, 1469-8137. doi: 10.1111/

- nph.13409. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.13409>.
- P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, and S. Hunter. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9):1236–1240, May 2014. ISSN 1367-4811, 1367-4803. doi: 10.1093/bioinformatics/btu031. URL <https://academic.oup.com/bioinformatics/article/30/9/1236/237988>.
- Y.-J. Kang, D.-C. Yang, L. Kong, M. Hou, Y.-Q. Meng, L. Wei, and G. Gao. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Research*, 45(W1):W12–W16, July 2017. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkx428. URL <https://academic.oup.com/nar/article/45/W1/W12/3831091>.
- E. V. Kriventseva, D. Kuznetsov, F. Tegenfeldt, M. Manni, R. Dias, F. A. Simão, and E. M. Zdobnov. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, 47(D1):D807–D811, 2019. Publisher: Oxford University Press.
- T. Kubo, S. Nishizawa, A. Sugawara, N. Itchoda, A. Estiati, and T. Mikami. The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNACys(GCA). *Nucleic Acids Research*, 28(13):2571–2576, July 2000. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/28.13.2571. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/28.13.2571>.
- H. Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, 2013. URL <https://arxiv.org/abs/1303.3997>. Version Number: 2.
- H. Li. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18):3094–3100, Sept. 2018. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/bty191. URL <https://academic.oup.com/bioinformatics/article/34/18/3094/4994778>.
- H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, 2009. Publisher: Oxford University Press.
- X. Li, C. Xue, J. Li, X. Qiao, L. Li, L. Yu, Y. Huang, and J. Wu. Genome-wide Identification, Evolution and Functional Divergence of MYB Transcription Factors in Chinese White Pear (*Pyrus bretschneideri*). *Plant and Cell Physiology*, 57(4):824–847, 2016. Publisher: Oxford University Press.
- D. J. Lightfoot, D. E. Jarvis, T. Ramaraj, R. Lee, E. N. Jellen, and P. J. Maughan. Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biology*, 15(1):74, Dec. 2017. ISSN 1741-7007. doi: 10.1186/s12915-017-0412-4. URL <http://bmcbiol.biomedcentral.com/articles/10.1186/s12915-017-0412-4>.
- X. Ma, F. E. Vaistij, Y. Li, W. S. Jansen Van Rensburg, S. Harvey, M. W. Bairu, S. L. Venter, S. Mavengahama, Z. Ning, I. A. Graham, A. Van Deynze, Y. Van De Peer, and K. J. Denby. A chromosome-level *Amaranthus cruentus* genome assembly highlights

- gene family evolution and biosynthetic gene clusters that may underpin the nutritional value of this traditional crop. *The Plant Journal*, 107(2):613–628, July 2021. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.15298. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.15298>.
- M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger. ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6):679–682, June 2022. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-022-01488-1. URL <https://www.nature.com/articles/s41592-022-01488-1>.
- G. Pertea and M. Pertea. GFF utilities: GffRead and GffCompare. *F1000Research*, 9, 2020. Publisher: Faculty of 1000 Ltd.
- B. Pucker, F. Reiher, and H. M. Schilbert. Automatic Identification of Players in the Flavonoid Biosynthesis with Application on the Biomedicinal Plant *Croton tiglium*. *Plants*, 9(9):1103, Aug. 2020. ISSN 2223-7747. doi: 10.3390/plants9091103. URL <https://www.mdpi.com/2223-7747/9/9/1103>.
- A. R. Quinlan and I. M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, 2010. Publisher: Oxford University Press.
- M. Sakuta, A. Tanaka, K. Iwase, M. Miyasaka, S. Ichiki, M. Hatai, Y. T. Inoue, A. Yamagami, T. Nakano, K. Yoshida, and S. Shimada. Anthocyanin synthesis potential in betalain-producing *Caryophyllales* plants. *Journal of Plant Research*, 134(6):1335–1349, Nov. 2021. ISSN 0918-9440, 1618-0860. doi: 10.1007/s10265-021-01341-0. URL <https://link.springer.com/10.1007/s10265-021-01341-0>.
- F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, and others. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7:539, 2011. Publisher: John Wiley & Sons, Ltd Chichester, UK.
- F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19):3210–3212, Oct. 2015. ISSN 1367-4811, 1367-4803. doi: 10.1093/bioinformatics/btv351. URL <https://academic.oup.com/bioinformatics/article/31/19/3210/211866>.
- G. S. C. Slater and E. Birney. Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics*, 6:1–11, 2005. Publisher: Springer.
- M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, and B. Morgenstern. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research*, 34(suppl_2):W435–W439, 2006. Publisher: Oxford University Press.
- R. Stracke, M. Werber, and B. Weisshaar. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology*, 4(5):447–456, Oct. 2001. ISSN 13695266. doi: 10.1016/S1369-5266(00)00199-0. URL <https://linkinghub.elsevier.com/retrieve/pii/S1369526600001990>.
- R. Stracke, D. Holtgräwe, J. Schneider, B. Pucker, T. Rosleff Sörensen, and B. Weis-

- shaar. Genome-wide identification and characterisation of R2R3-MYB genes in sugar beet (*Beta vulgaris*). *BMC Plant Biology*, 14(1):249, Dec. 2014. ISSN 1471-2229. doi: 10.1186/s12870-014-0249-8. URL <http://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-014-0249-8>.
- R. Sunnadeniya, A. Bean, M. Brown, N. Akhavan, G. Hatlestad, A. Gonzalez, V. V. Symonds, and A. Lloyd. Tyrosine Hydroxylation in Betalain Pigment Biosynthesis Is Performed by Cytochrome P450 Enzymes in Beets (*Beta vulgaris*). *PLOS ONE*, 11(2):e0149417, Feb. 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0149417. URL <https://dx.plos.org/10.1371/journal.pone.0149417>.
- M. Tardaguila, L. De La Fuente, C. Marti, C. Pereira, F. J. Pardo-Palacios, H. Del Risco, M. Ferrell, M. Mellado, M. Macchietto, K. Verheggen, M. Edelmann, I. Ezkurdia, J. Vazquez, M. Tress, A. Mortazavi, L. Martens, S. Rodriguez-Navarro, V. Moreno-Manzano, and A. Conesa. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Research*, 28(3):396–411, Mar. 2018. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.222976.117. URL <http://genome.cshlp.org/lookup/doi/10.1101/gr.222976.117>.
- J. D. Thompson, D. G. Higgins, and T. J. Gibson. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22):4673–4680, 1994. Publisher: Oxford university press.
- H. Tombuloglu. Genome-wide identification and expression analysis of R2R3, 3R-and 4R-MYB transcription factors during lignin biosynthesis in flax (*Linum usitatissimum*). *Genomics*, 112(1):782–795, 2020. Publisher: Elsevier.
- J. H. von Elbe. Betalains. *Current protocols in food analytical chemistry*, (1):F3–1, 2001. Publisher: Wiley Online Library.
- J. Wang, P. Youkharibache, D. Zhang, C. J. Lanczycki, R. C. Geer, T. Madej, L. Phan, M. Ward, S. Lu, G. H. Marchler, and others. iCn3D, a web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures. *Bioinformatics*, 36(1):131–135, 2020. Publisher: Oxford University Press.
- X. Zheng, S. Liu, C. Cheng, R. Guo, Y. Chen, L. Xie, Y. Mao, Y. Lin, Z. Zhang, and Z. Lai. Cloning and expression analysis of betalain biosynthesis genes in *Amaranthus tricolor*. *Biotechnology Letters*, 38(4):723–729, Apr. 2016. ISSN 0141-5492, 1573-6776. doi: 10.1007/s10529-015-2021-z. URL <http://link.springer.com/10.1007/s10529-015-2021-z>.
- I. M. Zimmermann, M. A. Heim, B. Weisshaar, and J. F. Uhrig. Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *The Plant Journal*, 40(1):22–34, Oct. 2004. ISSN 0960-7412, 1365-313X. doi: 10.1111/j.1365-313X.2004.02183.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2004.02183.x>.

Chapter 4

Uncovering co-option and adaptation of PAP1 for the regulation of the *de novo* evolved betalain pathway in amaranth

[Susanne K. Vollmer](#), [Linda M. Treder](#), [Jonas Schön](#), [Tom S. Winkler](#), [Uriel Urquiza-Garcia](#), [Matias D. Zurbriggen](#), [Götz Hensel*](#), and [Markus G. Stetter*](#)

First published: unpublished

The identification of AhMYB2 as a positive transcriptional regulator described in the previous chapter showed the involvement of the gene in betalain regulation. However, the understanding of the mechanistic regulation of betalains by this transcription factor requires the reconstruction of its binding and co-regulation with other factors. Most transcription factors function in complexes, requiring multiple different transcription factors to assemble and bind to the target DNA to regulate the expression. The following chapter combines yeast-two-hybrid, transactivation assays, and callus over-expression lines to investigate the molecular control of the betalain pathway by AhMYB2. Moreover, provide the presented RNA sequencing dataset previously unknown regulator candidates. This chapter integrates *in vitro*, bioinformatic, and molecular approaches for the betalain regulation in grain amaranth to study a fundamental evolutionary question.

4.1 Introduction

Betalains are a chemically distinct group of pigments that are functionally similar to the ubiquitous phenylalanine-based anthocyanins (Figure 4.6b)([Jain and Gould, 2015](#)). Betalains are only found within the *Caryophyllales* where they are mutually exclusive to anthocyanins, i.e., entire families either produce anthocyanins or betalains but not both ([Brockington et al., 2011](#); [Stafford, 1994](#)). Probably a mutation of the aragonate dehydrogenase releasing it from feedback inhibition at the base of the *Caryophyllales* caused a shift to more tyrosine-rich metabolism, enabling the evolution of the tyrosine-based betalains ([Lopez-Nieves et al., 2018](#)). Interestingly, the required enzymes for the betalain pathway seem to have evolved four times independently ([Sheehan et al., 2020](#); [Pucker et al., 2024](#)). Yet, not much is known about the regulation of the betalain pathway. For the *amaranthaceae* a MYB transcription factor homologous to AtPAP1 (PRODUCTION OF ANTHOCYANIN PIGMENT 1), which regulates the anthocyanin pathway in *A. thaliana* was identified for multiple species, including beetroot (BvMYB1)([Hatlestad et al., 2015](#)) and grain (AhMYB2)([Winkler et al., 2024](#)) and leafy amaranth (AtrMYB72)([Xue et al., 2025](#)).

Thus, it is currently hypothesized that the betalain regulation might have been co-opted from the anthocyanin regulation ([Hatlestad et al., 2015](#); [Winkler et al., 2024](#); [Lloyd et al., 2017](#)). Anthocyanins are common across the plant kingdom ([Timoneda et al., 2019](#); [Campanella et al., 2014](#)), and their regulation is well studied, uncovering a complex network of transcription factors (TFs) to regulate pigmentation depending on tissue, developmental stage, and environmental signals ([Cappellini et al., 2021](#); [Winkel-Shirley, 2001](#); [Zhao et al., 2023](#)). In contrast, besides the identification of some TF genes, the mode of action remains elusive for betalain-regulating TFs. However, for better understanding the evolutionary co-option, it is key to reconstruct the mode of action of the newly evolved TFs.

In this study, the molecular mode of action of TFs refers to the molecular steps required for the activation of the pathway by the TF. This includes (i) how the pathway is mainly regulated (transcriptionally, posttranscriptionally, or -translationally), (ii) which are the main target genes for the regulation, (iii) does the TF bind to the promoters directly or is the regulation indirect, and (iv) are other transcription factors required for the activation of the pathway.

For anthocyanins ([Gonzalez et al., 2008](#); [Tohge et al., 2005](#); [Chhon et al., 2020](#); [Takov et al., 2006](#)) as well as for betalains ([Hatlestad et al., 2015](#); [Cheng et al., 2017](#); [Winkler et al., 2024](#)) it was shown that the expression of key TFs correlates with the expression of the pathway genes, suggesting a mainly transcriptional control for both anthocyanins and betalains. In many metabolic pathways, the first steps are the main targets of regulation ([Blanco and Blanco, 2022](#); [Bhagavan and Ha, 2015](#)). The results from Hatlestad et al. (2015) who demonstrated that BvMYB1 can bind to the promoter from the first two en-

zymes of the pathway suggest that this seems to be the case for betalains, too. Moreover, these results suggest that at least MYB transcription factors homologous to BvMYB1, like AhMYB2, directly bind to the pathway promoters and do not regulate the pathway only indirectly. MYB transcription factors regulating anthocyanins mainly function in complexes with bHLH and WD 40 transcription factors (MBW complexes). In contrast, a small-scale test by Hatlestad et al. (2015) and bioinformatic analysis by Pucker et al. (2024) rather point to an MYB complex-independent regulation by BvMYB1 homologs. Hence, while there are hypotheses for some aspects of the mode of action, most of them have not been validated, and the complex formation as the main part remains unclear. Due to the complexity, multiple different methods are required to tackle the different aspects of the mode of action. We used stable transgenic *A. hypochondriacus* callus to demonstrate *AhCYP76AD2* and *DODAA1* as main regulatory targets and transcriptional control as the main regulatory strategy. We employed yeast-two-hybrid (Y2H) (Fields and Song, 1989) to investigate the interaction of AhMYB2 with multiple class III bHLH transcription factors from *Arabidopsis thaliana* and their homologs in amaranth. To test for direct promoter binding, we adapted protoplast isolation, transformation, and the transactivation assay for *A. hypochondriacus* and compared the results to the ones from transactivation assays in *A. thaliana* protoplasts. Finally, results from RNA sequencing on the stable transgenic *A. hypochondriacus* callus further supported our previous results and enabled us to construct a valuable database for the identification of additional betalain (co-)regulators in grain amaranth. Albeit we could not reconstruct the complete mode of action of AhMYB2, this study is the first systematic investigation for a betalain-regulating TF in the *Amaranthaceae*, offering insights into the evolution and co-option of a TF to new target genes.

4.2 Materials and Methods

All cloned plasmids and primers can be found in Supplementary Table S4.1 and S4.2, respectively.

4.2.1 Callus transformation and phenotyping

The *AhCYP76AD2* and *AhCYP76AD5* CRISPR knockout plasmid and the empty vector control were cloned/taken from the CasCADE modular cloning system (Hoffie, 2022) as described by (Vollmer et al., 2026). The *35S:RUBY* vector was cloned previously by transferring the RUBY cassette (He et al., 2020) to CPGE_VEC00113 (empty vector control). *35S:AhMYB2.1*, *35S:AhMYB2.2* and *35S:AhCYP76AD2* were cloned using the GreenGate system (Lampropoulos et al., 2013). Cloning of *35S:AhMYB2.1* and *35S:AhMYB2.2* coding sequences (CDS) into the entry vector is described in (Winkler

et al., 2024). The respective expression vectors were cloned using the following GreenGate modules: pGGZ001, pGGA004, pGGB003, the respective *AhMYB2* CDS (no stop codon) in pGGC000, pGGD001, pGGE001, and pGGF005. For *35S:AhCYP76AD2*, we used cDNA from green leaves from AM-00332 (a cross between PI 686465 and PI 576485) to amplify the CDS without a stop codon using primers SV6 and SV7 and cloned it via Gibson cloning into pGGC000 from the GreenGate kit (Lampropoulos et al., 2013). The expression vector was then assembled using pGGZ001, pGGA004, pGGB001, *AhCYP76AD2* CDS in pGGC000, pGGD002, pGGE001, and pGGF005 modules. For plant transformation, we obtained *Agrobacterium tumefaciens* (also named *Rhizobium radiobacter*) GV3101 strains (Holsters et al., 1980) harboring each one of the plasmids using electroporation. For the GreenGate plasmids, *Agrobacteria* strains were co-transformed with pSoup (Hellens et al., 2000).

We did the callus transformation as previously described in (Vollmer et al., 2026). We generally refer to all calli that were transformed with the same plasmid and that had been subcultured multiple times as a callus line. For the phenotyping experiments, a subset of 12 (18 for the CRISPR knockout line) calli (each obtained from different callus explants used for transformation) were selected per line. Each selected callus was then separated into three parts: one transferred to fresh callus induction medium (full Murashige-Skoog medium (MS) (Murashige and Skoog, 1962)(with vitamins), 2 mg/l 2.4D, 0.5 mg/l BAP, 20 g sucrose, pH 5.8, 8 g/l agar) (Castellanos-Arévalo et al., 2020; Wang et al., 2001), one to betalain medium (full MS (with vitamins), 0.2 mg/l NAA, 3 mg/l BAP, 20 g sucrose, pH 5.8, 8 g/l agar)(Taipova et al., 2019; Pal et al., 2013), and one part sampled with liquid nitrogen and kept at -70°C. For the three pigmented lines (*35S:RUBY*, *35S:AhMYB2.1*, and *35S:AhCYP76AD2*) each 10 callus parts per color classification (yellow, mixed, and red) were selected. The different pigmented parts were taken either all from one or from different calli. Six calli were cultivated per plate; in the case of the pigmented lines, two from each color class were on each plate. Plates with callus induction medium were kept as before at 24°C in the dark ("non-inducing conditions"), while the betalain medium plates were transferred to 24°C 16 h white light/8 h dark cycle ("inducing conditions"). In total, 6 genetically different lines (empty vector (CPGE_VEC00113), MYB2.1 (CPGE_VEC00643), MYB2.2 (CPGE_VEC00644), CYP67AD2 (CPGE_VEC00653), RUBY (CPGE_VEC00256), and the CRISPR line (CPGE_VEC00447)) on two different conditions (callus induction medium and dark (non_inducing) and betalain medium and light (inducing)) with an additional three color classes each (yellow, intermediate, and red) for MYB2.1, CYP76AD2, and RUBY lead to a final 24 callus line x condition combinations. We phenotyped them after less than one day (d0), five days (d5), and ten days (d10). For each phenotyping, we took overview pictures of all calli and scored all calli manually with the binocular. For scoring, only healthy-looking parts of the callus were considered. The coverage of the color as visible from the top was classified as follows: 0 = 0%, 1 = >0%, ≤10%, 2 = >10%, ≤50%, 3 = >50%, ≤90%, 4 = >90%, ≤100%. Color

intensity as visible from the top was classified as 0 = yellow (no pigmentation), 1 = low, 2 = medium, and 3 = strong. Additionally, two representative calli of each line, condition, and, if applicable, color subclass were imaged as close-ups under the camera-equipped binocular (Leica MZ10F, equipped with a CoolLED pE-300^{lite} light source and a Leica FLEXACAM C1 camera). After phenotyping at d10, we sampled a representative part of each callus with liquid nitrogen and kept it at -70°C.

We analyzed the semiquantitative counting data (color coverage and color intensity) in R (R Core Team, 2024) using functions from the base and from tidyverse (Wickham et al., 2019). To obtain the frequency of each category, the total number of times each category was noted for each line, condition, and, if applicable, color subclass and time point was divided by the total number of the calli from this group. The data was visualized using ggplot (Wickham, 2016).

Images were first opened in Fiji (Schindelin et al., 2012) to add a scale bar and then arranged, if needed, orientation adjusted, and the background manually removed using PowerPoint.

4.2.2 Yeast-two-hybrid to test potential interactors of Ah-MYB2

For each interaction investigated, we tested both reciprocal pairs (each transcription factor fused to the activation and the DNA-binding domain). All used vectors can be found in Supplemental Table S4.1 and all tested combinations in Supplemental Table S4.3. *Ah-MYB2.1* and *AhMYB2.2* CDS (Winkler et al., 2024) were cloned first into pDONR201 and then into pAS and pACT using Gateway cloning (developed by Invitrogen/Thermo Fisher Scientific) to obtain the final Y2H vectors.

Saccharomyces cerevisiae AH109 was used for all Y2H experiments. Y2H was done according to the protocol from the AG Hülkamp, University of Cologne (applying a modified yeast transformation from Gietz and Schiestl (2007)). We inoculated the main cultures using 1 ml of the 10 ml YPAD (20 g/l peptone, 10 g/l yeast extract, 100 mg/l adenine, 20 g/l glucose (added sterile after autoclaving), pH = 5.8) overnight culture, and which was incubated at 30°C with 200 rpm. For every 30 yeast transformations, we prepared two main cultures with 50 ml YPAD. Yeast cells were harvested when they reached an optical density (OD₆₀₀) of 0.6-0.8 by transferring them into 50 ml falcon tubes and centrifuging for 5 min at 4000 rpm (Eppendorf 5910R). We removed the supernatant and resuspended the pellet in 10 ml of 0.1 M lithium acetate (LiAc) to wash the cells. Centrifugation was repeated, the supernatant discarded, and cells resuspended in yeast transformation master mix (per one transformation: 240 µl of 50% polyethylene glycol (PEG), 36 µl of 1 M LiAc, 10 µl of ssDNA (10 mg/ml boiled at 99°C for 10 min), and 61 µl sterile water). For each transformation, we added 350 µl of the resuspended yeast to the previously prepared 1.5 ml Eppendorf tube containing the respective DNA (around

400 ng from each plasmid). The yeast was then immediately vortexed for 30 sec and incubated for 40 min at 42°C. After incubation, we centrifuged the yeast for 30 sec at 6000 rpm, resuspended it in 100 µl sterile water, and plated 5 µl on synthetic dropout plates lacking leucine and tryptophan (1.7 g/l yeast nitrogen base, 5 g/l ammonium sulfate, dropout mix without leucine and tryptophan (SD-LW) (added after autoclaving), 100 mg/l adenine, 18 g/l agar, pH = 5.8). Plates were sealed with parafilm and incubated upside down for 3-4 days at 30°C.

For the dropout, three colonies per transformation were picked and resuspended in one well of a sterile round-bottom 96-well plate containing 100 µl sterile water. To prepare a serial dilution, 10 µl of resuspension was added to another well with 90 µl of sterile water. This was repeated another time to have a total of three different concentrated resuspensions per construct. We dropped 5 µl of each of these on each one SD-LW and on SD-LWH (SD plates lacking histidine) plates with 0, 5, 10, 15, and 20 mM 3-aminotriazol (3-AT, dissolved in water, stored at -20°C). Plates were sealed with parafilm and incubated at 30°C for 3 days. For analysis, plates were scanned using a flatbed scanner (Epson Perfection V600 Photo). All Y2H experiments were repeated three times (transformation and dropping). The tested TF combination was considered as interacting if the respective yeast lines were growing at 15 mM 3-AT or higher for at least two out of the three repetitions. The tested TF combination was considered as non-interacting if the respective yeast lines were growing maximally at a 3-AT concentration of 5 mM in two out of three repetitions.

4.2.3 Protoplast isolation

Protoplast isolation was done according to the AG Zurbriggen (Heinrich Heine University Düsseldorf) protocol for *Arabidopsis thaliana* and, with slight modification, also for amaranth (([Ochoa-Fernandez et al., 2020](#); [Andres and Zurbriggen, 2022](#)) Urquiza-Garcia et al., in preparation): For *Arabidopsis thaliana* Col-0 seeds were sterilized and sown on a strip of Whatman paper in squared Petri dishes containing SCA agar. For *Amaranthus hypochondriacus* PI558499 seeds were sterilized and sown on solid $\frac{1}{2}$ MS with 20 g/l sucrose (see 4.2.1). *A. thaliana* seedlings were kept at long days (16 h white light/8 h dark cycle) at 21°C for 12-14 d until harvest. Amaranth seedlings were kept for 5 days at 24°C in the dark until moved to light (16 h white light/8 h dark cycle) at 24°C for another 6 d before harvest. For harvesting, we cut the *A. thaliana* shoots with a sterile scalpel and collected them in 6 ml of MMC buffer (10 mM MES, 40 mM $\text{CaCl}_2 \cdot \text{H}_2\text{O}$ in ddH₂O adding mannitol until 550 mOsm, adjusted to pH 5.8, sterile filtered). Amaranth seedlings were harvested with sterile forceps and scissors and collected in 2 ml of MMC. To assess protoplast yield, we weighed the plates with MMC before and after harvest. For *A. thaliana*, 2.6-5.4 g, and for amaranth, 0.4-1.2 g fresh weight were used per digest. Harvested tissue we then manually cut with sterile one-time-use scalpels for around 4-6

min per digest before transferring them to a fresh Petri dish containing the remaining MMC for a total volume of 10 ml. Per digest, around 2 ml of enzyme solution (a mix of 5% cellulase Onozuka R10 and 5% macerozyme R10 in MMC buffer) was added, plates gently swirled, sealed with parafilm, wrapped with aluminum foil, and incubated at 21°C for 15.0-17.3 h. The digest solution was pipetted up and down four times using a 25 ml serological pipette to release the protoplasts after the incubation time. Afterwards, we cleared the solution through a 70 μ M filter on an angled 50 ml Falcon tube. To dilute the enzyme solution, we filled up the final volume through the filter to 50 ml using MMC. The suspension was subsequently carefully centrifuged for 20 min at 100 g with an acceleration and deceleration of 1, and the supernatant was removed using a vacuum pump, leaving only about 1-2 ml. After gently fully resuspending the pellet by rotating the tube, we added 10 or 20 ml of MSC buffer (10 mM MES, 0.4 M sucrose, 20 mM $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$, 467 mM mannitol, pH 5.8, sterile filtered), depending on the amount of protoplasts expected. For lower amounts of protoplasts (e.g. only one square Petri dish of *A. thaliana* seedlings), 10 ml was used, and for higher amounts (e.g. two square Petri dishes of *A. thaliana* seedlings), 20 ml of MSC. The homogenous resuspension was then divided into one (if 10-12 ml total volume) or two (if 20-22 ml total volume) round-bottom 15 ml tubes. To obtain two phases, we first inverted the tubes and then carefully overlayed the resuspension with 3 ml of MMM (15 mM MgCl_2 , 5 mM MES, 467 mM mannitol, pH 5.8, sterile filtered) using double-tip pipetting. Centrifugation at 80 g for 10 min with an acceleration and deceleration of 1 caused viable protoplasts to accumulate at the interphase as a visible layer. Per interphase tube, we collected 1.5 ml of protoplasts from the interphase in a fresh round-bottom tube containing W5 (150 mM NaCl, 125 mM CaCl_2 , 5 mM KCl, 2 mM MES, 5 mM glucose, pH 5.7, sterile filtered). Wide-open tips were used for pipetting protoplasts. If two interphase tubes were prepared per sample, then protoplasts were collected in 9 ml of W5; if only one interphase tube was used, then only 7 ml of W5 was added to the collection tube. The centrifugation and collection are repeated once to maximize protoplast yield. Note that for amaranth, green and red protoplasts are obtained. The red protoplasts tend to float above and sometimes are not restricted to the interphase. To increase the compactness of the interphase for amaranth, only 2 ml of MMM can be added for the first centrifugation, and 1 ml of W5 can carefully be added on top of the MMM for the second centrifugation. We then counted the protoplasts using a Fuchs-Rosenthal chamber (0.200 mm, 0.0625 mm²).

4.2.4 Protoplast transformation and transactivation assays

All transactivation assay plasmids were cloned with Gibson. More information about the cloned promoter fragments and their sequences can be found in supplements (Supplemental Table S4.4 and Appendix S3.1). The used backbones (pmz824, pep602, pep602_sw) were linearized with primers CPGE_PRM00659 and CPGE_PRM00660 for pmz824

and primers CPGE_PRM00661 and CPGE_PRM00662 for pep602 and pep602_sw. We amplified the *AtPAP1* CDS without a stop codon from BP311 (Zhang, 2018) using primers CPGE_PRM01190 and CPGE_PRM01192 and cloned it in pmz824. The *AtDFR* promoter fragment we amplified from *A. thaliana* Col-0 genomic DNA (gDNA) using primers CPGE_PRM01193 (position from Guo et al. (2023)) and CPGE_PRM01194 (position from He et al. (2021)) and cloned it in pep602. The *BvMYB1* sequence as published by Hatlestad et al. (2015) was synthesized and amplified with primers CPGE_PRM01111 and CPGE_PRM01113 without a stop codon and inserted it in pmz824. We amplified *BvCYP76AD1* from *B. vulgaris* Bull's Blood gDNA using primers CPGE_PRM01187 and CPGE_PRM01189 and inserted it into pep602. *AhMBY2.1* cDNA was obtained as described in Winkler et al. (2024), amplified with primers CPGE_PRM00663 and CPGE_PRM00664, and cloned into pmz824. The *AhCYP76AD2* promoter fragment we amplified with primers CPGE_PRM00681 and CPGE_PRM01196 from gDNA of *A. hypochondriacus* PI 558499. The synthetic positive controls (PolyGEN, Urquiza-Garcia *et al.*, in preparation) we cloned using Golden Gate (Engler et al., 2008). The guides were obtained by amplification of the universal template (pUU017) with the specific oligos (guide selection done as described in (Vollmer et al. (2026) section S3.1, oligos designed with sbPolyGEN (Urquiza-Garcia *et al.*, in preparation)). The guide cassette was cloned via Golden Gate in the entry vector (pUU385), which was subsequently used to assemble the final plasmid (CPGE_VEC01022) with the modules GB0037 and GB0223 in the backbone pdgb3alpha1. We designed the guides to target the *AhcDOPA5GT* promoter, which was amplified using CPGE_PRM00672 and CPGE_PRM00673 primers from gDNA of *A. hypochondriacus* PI 558499 and cloned in the pep602_sw backbone, yielding CPGE_VEC01029. For the PolyGEN synthetic positive control, CPGE_VEC01029 and CPGE_VEC01022 were co-transformed with GB2047 carrying the dCas9 activator. Plasmids used for protoplast transformation were extracted from *Escherichia coli* using the NucleoBond Xtra Midi Kit (MACHEREY-NAGEL, Düren, Germany) for high purity.

Protoplast transformation and measurement were done according to the AG Zurbriggen protocol for *Arabidopsis thaliana* (Heinrich Heine University Düsseldorf) and, with slight modification, also for amaranth ((Ochoa-Fernandez et al., 2020; Andres and Zurbriggen, 2022) Urquiza-Garcia *et al.*, in preparation): We adjusted the protoplast density to 500 protoplasts per μl for *A. thaliana* and 250-600 per μl for amaranth (depending on the total yield, 500 was not always possible) by centrifuging for 10 min at 100 g with an acceleration and deceleration of 1, removal of the supernatant via vacuum pump, and resuspension in the respective amount of MMM. For transformation, DNA was pipetted in sterile 6-well plates. We used 10 μg of DNA from each plasmid (one for the reporter and one with the transcription factor). Total DNA content per transformation was kept the same for each sample by using stuffer plasmids (*35:GFP* or using AtEIN3 as an unrelated transcription factor), and total volume was fixed to 20 μl (filled up with MMM

if required). 100 µl protoplasts were then added per sample and mixed with the DNA by pipetting up and down. After around 5-15 min of incubation time, we collected the protoplasts at the edge of the wells to form a half-moon. For transformation, we added 120 µl of PEG solution (40% (w/v) PEG4000, 400 mM D-mannitol, 100 mM CaCl₂) dropwise using double-tips. After 8 min, we added 120 µl of MMM and 2x 720 µl of W5 (+50 mg/l ampicillin) and carefully pipetted up and down a few times to stop the reaction and release the PEG stress. Plates were then sealed with parafilm, wrapped in aluminum foil, and incubated at 21°C for 21-23 h before measuring.

For measurement of the luciferase expression, we carefully resuspended the protoplasts if sedimented and pipetted 80 µl of each into white, flat-bottom 96-well plates. Both luciferases were measured separately on different protoplast aliquots in different 96-well plates. For *A. thaliana*, 4 or 6 replicates per luciferase (so either 8 or 12 wells in total split into two 96-well plates) and for amaranth, 4 replicates per luciferase were prepared. Due to the low signal intensity, amaranth protoplasts were concentrated by centrifuging for 10-20 min at 80 g with an acceleration and deceleration of 1 and removing 900 µl of supernatant. The protoplast pellet we then carefully resuspended in a small volume to be just enough for the 8 replicates. Directly before measuring in the plate readers (Berthold Centro XS3 LB 960 microplate luminometer), we added 20 µl/well of the respective luciferase substrate: firefly (0.47 mM d-luciferin (Biosynth AG), 20 mM tricine, 2.67 mM MgSO₄·7H₂O, 0.1 mM EDTA·2H₂O, 33.3 mM dithiothreitol, 0.52 mM adenosine 5'-triphosphate, 0.27 mM acetyl-coenzyme A, 5 mM NaOH, 264 µM MgCO₃·5H₂O in H₂O, pH 8) or renilla (0.472 mM coelenterazine stock solution in methanol, diluted directly before use, 1:15 in PBS). Light signal was recorded at 29 time points over a course of 20 min.

We analyzed the luciferase measurements in R ([R Core Team, 2024](#)) using functions from the base and from tidyverse ([Wickham et al., 2019](#)). First, the mean of all 4-6 replicates was calculated and, if applicable, normalized by the second *35S*-expressed luciferase (i.e., by dividing and multiplying by 100) per tested combination for each transactivation experiment separately. To combine data from multiple independent transactivation assays, we normalized each of the replicate means by the respective only_reporter (only stuffer, but no binding TF) sample to obtain one value per transactivation assay per tested treatment combination. An overall mean was then subsequently calculated using these normalized means from each transactivation assay. For statistical analysis, we used the multcomp package ([Hothorn et al., 2008](#)): `do(Model = cld(emmeans(aov(mean_measurement_norm ~ TF, data=.), "TF", adjust="tukey"), alpha=.05, Letters=letters))` and for visualization `ggplot2` ([Wickham, 2016](#)).

4.2.5 Mammalian cell culture

AhMYB2.1 was recloned in pJS174 using CPGE_PRM01050, CPGE_PRM01051, CPGE_PRM00659, CPGE_PRM00664 for expression in mammalian cells. We used a shorter *AhCYP76AD2* promoter, which we amplified with CPGE_PRM00681 and CPGE_PRM00682 and cloned it in pep602. For mammalian cell culture, Chinese ovary hamster (CHO-K1, ACC 110; DSMZ Braunschweig, Germany) cells were cultivated in Ham's F12 medium at 37°C. Transfection was done using polyethylenimine (PEI; Polysciences Inc. Europe, Hirschberg, Germany; no. 23966-1) with 50000 cells in 500 µl per well in 24-well plates seeded 24 h before, according to the AG Zurbruggen (Heinrich Heine University Düsseldorf) protocol. For the transfection, we mixed 0.75 µg of DNA with 2.5 µg of PEI, both diluted to a total volume of 50 µl with OptiMEM (Gibco), and after vortexing, incubated at room temperature for about 15 min. The resulting 100 µl of this mix were added dropwise per well. To reduce toxicity, we exchanged the medium after 4-5 h. 24 h after transfection, the cells were lysed and firefly luciferase was measured as described for protoplasts. We analyzed firefly luciferase measurements in R (R Core Team, 2024) using functions from the base and from tidyverse (Wickham et al., 2019). No normalization was possible, as *35S:renilla* is not expressed in mammalian cells. We did the statistical analysis grouped using the multcomp package (Hothorn et al., 2008): `do(Model = cld(emmeans(aov(measurement ~ TF, data=.), "TF", adjust="tukey"), alpha=.05, Letters=letters))` and visualized the results using ggplot2 (Wickham, 2016).

4.2.6 RNA sequencing of transgenic callus

Six callus line x condition combinations were selected for RNA-seq: EV_inducing, EV_non_inducing, MYB2.1_non_inducing (red color class), MYB2.2_non_inducing, CYP76AD2_non_inducing (red color class) and RUBY_non_inducing (red color class). Each 4 calli were selected randomly, but preferring healthy-looking calli and taking care that not all selected calli were cultivated on the same plate. Frozen callus samples were ground with a pestle and mortar. We aliquot roughly 100 mg of ground tissue for RNA extraction with the RNeasy Plat Mini kit (Qiagen, Venlo, Netherlands) according to the manufacturer's protocol with minor modifications. As a first step, we added 450 µl of Buffer RLT (with freshly added 10 µl β-Mercaptoethanol) to the ground samples and vortexed for around 20 sec. Samples were transferred to the shredder column and centrifuged for 2 min at 21,000 rcf. We transferred the supernatant to a fresh tube, mixed it with 210 µl ethanol (analytical grade), and loaded it on the spin column. After 15 sec centrifugation with 8000 rcf, 350 µl RW1 buffer was used to wash the column with 15 sec centrifugation with 8000 rcf. To digest remaining gDNA we added 80 µl of DNase incubation mix (freshly prepared 70 µl buffer RDD and 10 µl DNase I stock solution) the

column and incubated it for 15 min at room temperature. The column was then washed again with 350 µl RW1 buffer and 15 sec centrifugation at 8000 rcf. As a final washing step, we washed the column twice with 500 µl RPE buffer, first with 15 sec at 8000 rcf, then with 2 min and 8000 rcf. The remaining buffer was removed by centrifuging for 1 min at 21.000 rcf before the column was transferred to a fresh tube. For elution we added 20 µl of RNase-free water, followed by centrifugation for 1 min at 8000 rcf. For maximal yield, elution was repeated a second time with a 5 min incubation at room temperature with the RNase-free water before the centrifugation. Centrifuging at higher speed for the elution led to slightly higher impurities at A230, without significantly increasing the yield. Final extracted RNA was kept on ice or at -20 °C for short-term handling and stored at -70 °C. For RNA-seq, the in total 24 RNA samples were sequenced on an Illumina NovaSeq X Plus platform in 150 bp paired-end reads.

4.2.7 Differential expression analysis

We trimmed adapter sequences from the reads using Trimmomatic v0.39 (Bolger et al., 2014) using the ILLUMINACLIP module with the following settings: TruSeq3-PE-2-GGGGGG.fa:2:30:10. We performed quality control of the trimmed sequencing reads using FastQC v0.12.1 (Andrews, 2010). Next, we mapped the trimmed reads to the *A. hypochondriacus* reference genome v3 (Graf et al., 2025) using STAR v2.7.11b (Dobin and Gingeras, 2015) with default settings. We quantified fragments based on the exon annotation, only counting read pairs for which both reads were aligned using featureCounts v2.1.1 (Liao et al., 2014). To exclude very lowly expressed genes from the analysis, we required genes to have at least 10 assigned fragments in at least three samples to be included in the analysis. We performed differential expression analysis using DESeq2 (Love et al., 2014) for pairwise comparison between conditions and the empty vector control (EV_non_inducing) and considered genes with absolute logfold2 change > 1 and false discovery rate < 0.01 to be differentially expressed.

4.2.8 Data analysis and candidate identification

Data analysis was done in R (R Core Team, 2024) using functions from the base and from tidyverse (Wickham et al., 2019), including tidyr (Wickham et al., 2024) and dplyr (Wickham et al., 2023). If required, we used openxlsx (Schauberger and Walker, 2025) to import the data. For visualization, we used, besides ggplot (Wickham, 2016), pheatmap (Kolde, 2025) for the construction of heatmaps (and ggplotify (Yu, 2025) to save them), ggvenn (Yan, 2025) for Venn diagrams, and UpSetR (Gehlenborg, 2019) for upset graphs. The source of the sequences used for the logoplot (5 for anthocyanin and 6 for betalains) and for the R2R3 domain comparison in the overall discussion are listed in Supplemental Table S4.5. The logoplot we did using ggseqlogo (Wagih, 2025) with the custom color

scheme from Walker-Hale et al. (2025) only exchanging #2a7fff with #3D6DB5 for better color matching.

To identify the anthocyanin pathway genes, we identified conserved amino acid sequences based on the literature and used them as bait to find homologous proteins in the set of all proteins based on the *A. hypochondriacus* genome annotation v3 (Graf et al., 2025). A list of all identified amaranth anthocyanin genes and the used conserved amino acid sequences can be found in Supplementary Table S4.6. Note that no candidate was identified for F3'5'H, as already observed by Pucker et al. (2024). With the same approach, we identified ADT and ADR genes (Supplementary Table S4.7).

To identify TF classes, functional annotation was added using the current *A. hypochondriacus* annotation found on <https://amaranthgdb.org/downloads.html> (version 3, (Graf et al., 2025)). Genes belonging to the investigated TF classes we found based on the presence of the respective TF class or domain name in either the "Description" or in the "PFAMs" column. To obtain an estimation of all TF genes expressed, we made a subset from the total data set by selecting for the presence of the GO-Term for TF (GO:0003700) and complementing it, if required, with the TFs identified by classes that are not in the GO-term filtered list. The estimation of all differentially expressed genes we obtained by filtering this subset for only differentially expressed genes. Enrichment of certain TF classes in the pool of all genes, which are differentially regulated in at least one sample, we did using `chisq.test()` and `chisq.posthoc.test(data, method = "bonferroni", round = 6)` from the `chisq.posthoc.test` package (Ebbert, 2019). To find *cis*-elements in the *AhCYP76AD2* and *AhDODAA1* promoters, we used the 2 kb upstream of the 5'UTR and the 5'UTR as input sequences (2125 bp and 2064 bp respectively) and submitted them in December 2025 to PlantCARE (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) (Lescot, 2002), PlantPAN4.0 (https://plantpan.its.ncku.edu.tw/plantpan4/promoter_analysis.php, with all available species at this time) (Chow et al., 2024) and PLACE (<https://www.dna.affrc.go.jp/PLACE/?action=newplace>) (Higo et al., 1999). The position and the associated TF class or description from all database results were combined in one Excel list, noting the respective database. In R, all motives that were on the exact same position identified from the same database were considered as duplicates and removed. Subsequently, to facilitate the removal of duplicates stemming from the identification of the same motive by different databases, we summarized the TF description and subclasses to fewer TF classes (WRKY, TATA-box, bHLH, AP2/ERF, B3/ARF, bZIP/Homeodomain, MADS-box, MYB, Zinc finger, GATA, NAC/NAM, CAAT-box; and all remaining were classified as "others"). Duplicated TF motifs belonging to the same class with the same or overlapping position (irrespective of the database it was identified from) were then removed manually. Potential LOB transcription factor binding sites we additionally identified manually by searching for the known LOB element ((G) CGGC (G)/(C) GCCG (C)) (Rong et al., 2024) and for CATTTAT, which was identified as a binding site for an *A.*

thaliana LOB TF (Ohashi-Ito et al., 2018). Identified motives from selected TF classes from the final list were then visualized using ggplot (Wickham, 2016). For phylogenies, we imported and handled sequencing data using Biostrings (Pagès et al., 2024) and seqinr (Charif and Lobry, 2007). Either lists of all TFs from certain species were used from PlantTFDB version 5.0 (<https://planttfdb.gao-lab.org/index.php>) (Tian et al., 2019) (for ERF) or were created by filtering a list of all proteins from selected species using the respective PFAM domain with `vmatchPattern("WRKYGQK", total_prot_data)` (for WRKY). *A. thaliana* data was downloaded from TAIR (https://www.arabidopsis.org/download/list?dir=Proteins%2FAraport11_protein_lists) and *A. hypochondriacus* data from AmaranthGDB (<https://amaranthgdb.org/downloads.html>) (Graf et al., 2025). For WRKY, we additionally included AtrWRKY42 (Yang et al., 2023) as a comparison. Multiple sequence alignment we did with `msa` (Bodenhofer et al., 2015) with the default settings, which uses ClustalOmega (Sievers et al., 2011). To get the distance matrix (`dist.aa(as.AAbin(alignment_all_WRKY))`) and to construct the phylogenetic tree (`nj(dist_matrix_all_WRKY)`) we used `ape` (Paradis and Schliep, 2019) based on the multiple sequence alignment. We plotted the trees using `ggtree` (Yu et al., 2017).

4.3 Results

4.3.1 *AhCYP76AD2* is the main target for the regulation of the betalain biosynthesis

To start the elucidation of the downstream signaling of AhMYB2, we first aimed to identify the main regulatory target of the betalain pathway. Since the expression of *AhCYP76AD2* and *AhDODA1* was highly correlated with the betalain pigmentation and *AhMYB2* expression (Winkler et al., 2024), both were good candidates. The first committed step in a metabolic pathway is often important for the regulation of the pathway (Blanco and Blanco, 2022; Bhagavan and Ha, 2015). Thus, we chose *AhCYP76AD2* to test as a potential (main) target of the betalain pathway regulation. We make use of our established and very efficient callus transformation method (see 4.2.1) to produce a stable transformed *A. hypochondriacus* callus line that overexpresses *AhCYP76AD2*. We compared its pigmentation to an empty vector (EV) control, observing that 66 of 67 calli of the *AhCYP76AD2* overexpression line showed different degrees of red pigmentation, while none of the 46 empty vector control calli had any red pigmentation in the dark.

These results suggest that indeed *AhCYP76AD2* seems to be the main target to regulate the activation of the betalain biosynthesis, since overexpression of *AhCYP76AD2* (thus escaping the endogenous transcriptional regulation) is sufficient to activate the pathway.

4.3.2 In contrast to anthocyanins, betalains seem not to be regulated in MBW complexes

To characterize AhMYB2 and as a starting point to decipher its molecular mode of function, the protein sequence was analyzed. The phylogeny from (Winkler et al., 2024) Figure 3.2 shows that the closest homologs of AhMYB2 from *B. vulgaris* are BvMYB1 and Bv_ralf, where the first was shown to regulate the betalain pathway positively (Hatlestad et al., 2015) (close up in Figure 4.1a). The closest *A. thaliana* homologues are AtPAP1, AtPAP2, AtMYB113, and AtMYB114, which regulate the flavonoid pathway in MBW complexes. To evaluate the probability of AhMYB2 interacting with bHLHs, required to form MBW complexes, its protein sequence of the R3 domain was compared to the generally conserved PAP1 sequence from anthocyanin- and betalain-producing species (Figure 4.1b)(for a more comprehensive conserved PAP1 R3 domain sequence analysis, see Pucker et al. (2024)). Additionally, the sequence was compared to the one from AtPAP1, which is known to interact with bHLHs and BvMYB1, where experiments suggest that it does not interact with bHLHs. The comparison showed that most of the R3 domain is conserved across both anthocyanin- and betalain-producing species, including AhMYB2 and BvMYB1. However, both AhMYB2 and BvMYB1 have, respectively, six and seven amino acids altered in the analyzed 40 amino acid long R3 domain. When considering the 20 amino acids that were identified or proposed to be important for bHLH interaction, notably, most of the altered amino acids (5 from each) are in the potential bHLH interaction regions. Nevertheless, the majority of the potential bHLH interaction residues are still conserved in both AhMYB2 and BvMYB1, leaving it unclear if they generally can interact with bHLHs or not. Thus, we performed a yeast-two-hybrid (Y2H) experiment (Fields and Song, 1989) to test it experimentally and compared it to the literature-reported interaction of the *A. thaliana* homologues (TRANSPARENT TESTA 8 (TT8), TRANSPARENT TESTA GLABRA 1 (TTG1), Figure 4.1c,d). AhMYB2 did not interact with AhTT8 (Figure 4.1c, supplemental Figure S4.1). Moreover, AhTT8 and AhTTG1a did not show interaction, further supporting the hypothesis that AhMYB2 does not act in an MBW complex. As AhMYB2 showed autoactivation, indicating strong transactivation ability, all combinations were repeated with a version that has a 12 bp insertion that abolishes the transactivation activity. To clarify which version was used in the experiments, we refer to the full-length functional gene (=AhMYB2) as AhMYB2.1 and the 12 bp insertion variant with abolished transactivation activity as AhMYB2.2. Using AhMYB2.2, we could confirm the results of AhMYB2.1 and find a weak homodimerization of AhMYB2.

To further confirm the hypothesis that AhMYB2 can't interact with bHLHs, we tested the interaction of AhMYB2 with additional clade IIIb bHLH transcription factors (Figure 4.2, Supplemental Figure S4.2). We choose AtMYC1, AtTT8, AtGL3 (GLABRA3), and AtEGL3 (Enhancer of GLABRA3), which were all shown to interact with AtPAP1 and

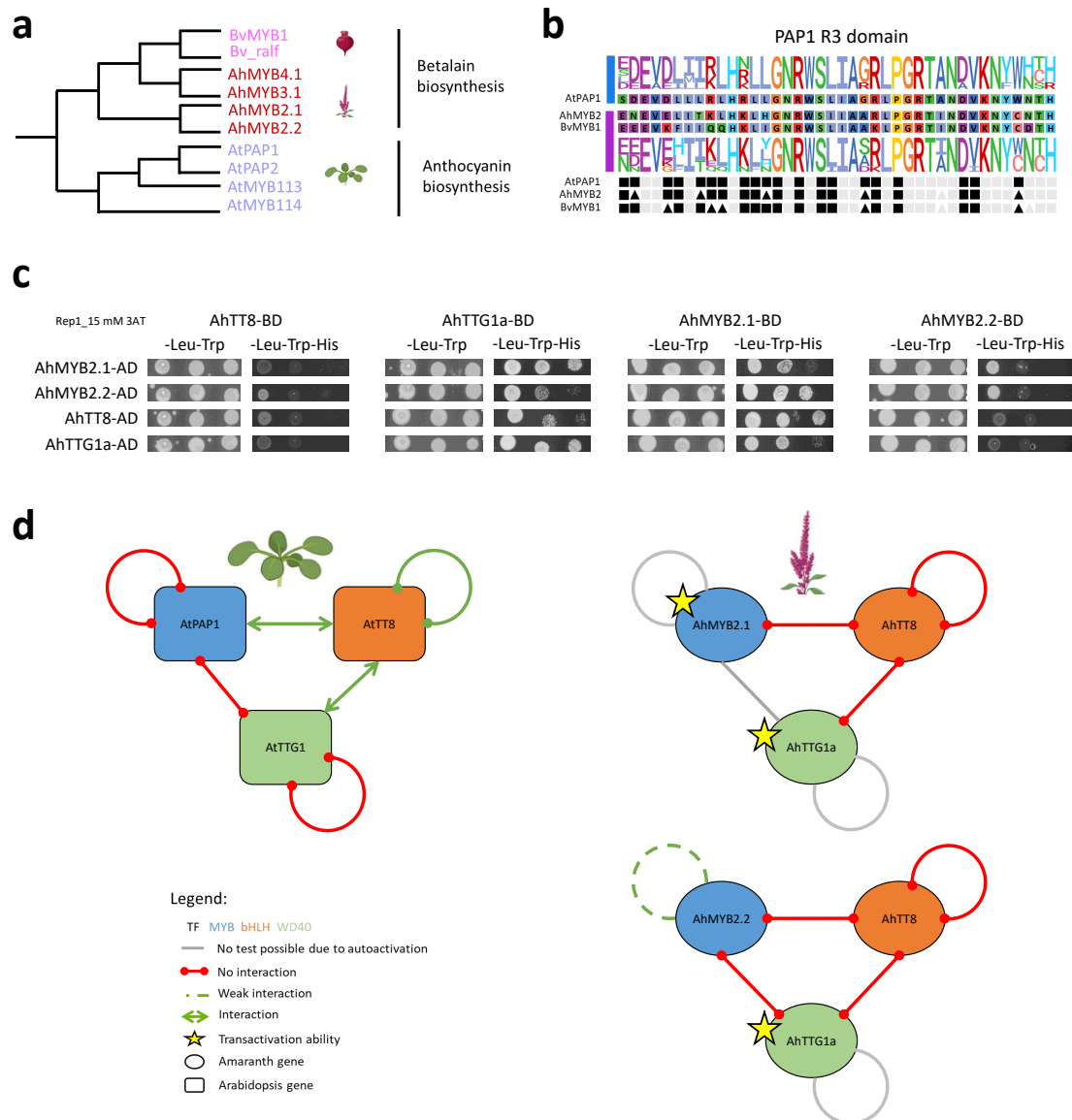
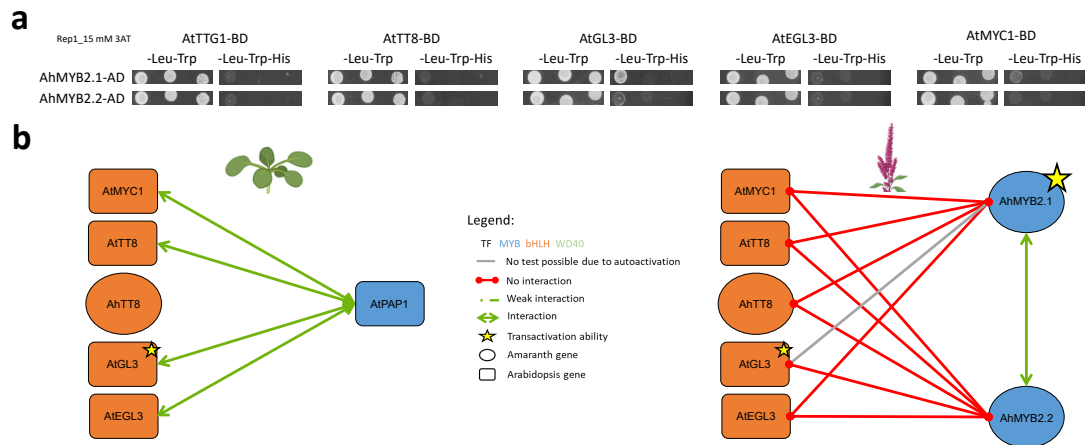


Figure 4.1: AhMYB2 seems, in contrast to his homologue AtPAP1, not to form MBW complexes. Phylogeny showing the closest related homologs of *B. vulgaris* and *A. thaliana* from AhMYB2 (data from (Winkler et al., 2024) Figure 3.2)(a). Comparison of the amino acid sequence of the R3 domain of AtPAP1, AhMYB2.1, and BvMYB1 (b). Sequence logos show the conserved sequence of PAP1 proteins from anthocyanin (blue) and betalain (purple) species. Residues shown or proposed to be involved in bHLH interaction are indicated by black (important) and grey (not relevant) symbols below (data from (Pucker et al., 2024; Hatlestad et al., 2015; Zimmermann et al., 2004; Sakuta et al., 2021)). Squares indicate conservation of the anthocyanin PAP1 amino acid sequence, and triangles indicate deviations from the conserved sequence. Y2H data from the *A. hypochondriacus* homologues AhMYB2, AhTT8, and AhTTG1a (c). AhMYB2.1 is the full-length gene, while AhMYB2.2 has a 12 bp insertion in the R3 domain, which abolishes the autoactivation. For each interaction were both reciprocal combinations tested, i.e., once as fusion with the GAL4 activation domain (AD) on the first gene and fusion of the GAL4 DNA binding domain (BD) on the second gene and once *vice versa*. Summarized interactions of AhMYB2 with AhTT8 and AhTTG1a and, as a comparison, the AtPAP1 MBW complex in *A. thaliana*(information from literature, e.g., summarized by Zhang and Schrader (2017))(d). Red beet, *A. thaliana* and amaranth icons are from BioRender.

to be part of MBW complexes in *A. thaliana* (e.g., summarized by Zhang and Schrader (2017)). In contrast to AtPAP1, AhMYB2 could not interact with any of the tested bHLHs (no results for AtGL3 as it had slide autoactivation). Supporting our hypothesis that the ability to interact with (clade IIIf) bHLHs was lost in AhMYB2.



4.3.3 Betalain MYBs seem to require another transcription factor to stabilize the DNA-binding

Our Y2H data suggest that AhMYB2 has transactivation ability and that it seems not to interact with bHLH transcription factors. To test if AhMYB2 can bind to the pathway promoters directly and, if yes, to which and how strongly, we performed transactivation assays in protoplasts (Figure 4.3a). By fusing the transcription factor of interest with VP16 (a viral activator domain), we could exclude variability due to differences in the transactivation activity and only test for binding to the potential target promoter. First, as a positive control, we showed activation of the *A. thaliana dihydroflavonol 4-reductase* (*AtDFR*) promoter by AtPAP1-VP16 in *A. thaliana* protoplasts, replicating the previously shown results (Guo et al., 2023; He et al., 2021)(Figure 4.3b). Next, we tested whether AhMYB2.1-VP16 can bind to the *AhCYP76AD2* promoter, but observed only weak and nonsignificant activation in *protoplasts of A. thaliana* (Figure 4.3c). We excessively tested other possible explanations for the lack of binding. We could rule out missed ATG-adjacent *cis*-elements as we cloned the *AhCYP76AD2* promoter until the ATG and we could rule out secondary structure of the promoters, and blockages of the *cis*-elements by *A. thaliana* transcription factors by using a synthetic positiv control (PolyGEN, Urquiza-Garcia et al. in preparation) specific for an amaranth promoter

(Supplementary Figure S4.3c) Thus, we hypothesized that perhaps *in planta* binding of AhMYB2 requires an additional transcription factor specific to betalain-producing species to stabilize the binding, which is lacking in *A. thaliana* protoplasts. To test this hypothesis, we established the protoplasting, transformation, and transactivation assay in *A. hypochondriacus*. However, also in *A. hypochondriacus* protoplasts we could not detect significant binding of AhMYB2.1-VP16, maybe caused by the low signal strength in the amaranth protoplasts. We repeated the test with BvMYB1, which was shown to upregulate the betalain pathway *in planta* and to bind to the *BvCYP76AD1* promoter in yeast-one-hybrid (Y1H) (Hatlestad et al., 2015). In line with our hypothesis, we could see binding of BvMYB1-VP16 in *A. hypochondriacus*, but not in *A. thaliana* protoplasts (Figure 4.3d). To conclude, we could not prove binding of AhMYB2 to the *AhCYP76AD2* promoter, but our transactivation data with the similar BvMYB1 transcription factor suggests that betalain-regulating transcription factors might require an additional transcription factor to stabilize the binding *in planta*.

4.3.4 AhMYB2 induces betalain accumulation by upregulating *AhCYP76AD2* and *AhDODAα1* but lost the ability to regulate flavonoid pathway genes

We validated AhMYB2.1 as a positive regulator of the betalain pathway in grain amaranth (Winkler et al., 2024) and identified *AhCYP76AD2* as the key target for the betalain regulation (see section 4.3.1). Although most likely, we did not yet confirm that AhMYB2.1 is regulating the betalain pathway by upregulating *AhCYP76AD2*. Moreover, our Y2H data showed that AhMYB2.1 seems not to interact with bHLHs and form MBW-complexes (section 4.3.2), but our transactivation data suggests that AhMYB2.1 still requires (a) yet unknown additional partner(s) for sufficient DNA binding (section 4.3.3). Betalain biosynthesis is strongly regulated by environmental signals, including light, but also by internal signals, especially phytohormones. This enables control of the betalain pathway activity by cultivating tissue of amaranth either in conditions that induce betalain pigmentation (“inducing conditions”, high cytokinin and low auxin concentration in light) or not (“non-inducing conditions”, low cytokinin and high auxin in dark). To test the involvement of AhMYB2.1 in the environmental betalain regulation, to prove upregulation of *AhCYP76AD2* by AhMYB2.1, and to find potential candidates that co-regulate the betalain pathway in grain amaranth with AhMYB2.1, we produced a set of transgenic *A. hypochondriacus* calli, which we analyzed for their pigmentation and transcription in both inducing and non-inducing conditions. Our set of transgenic callus lines is centered around an *AhMYB2.1* overexpression line and, as a comparison, one overexpressing *AhMYB2.2*, an insertion mutant abolishing the transcriptional activation activity of AhMYB2. To disentangle the effect of strong overaccumulation of betalains,

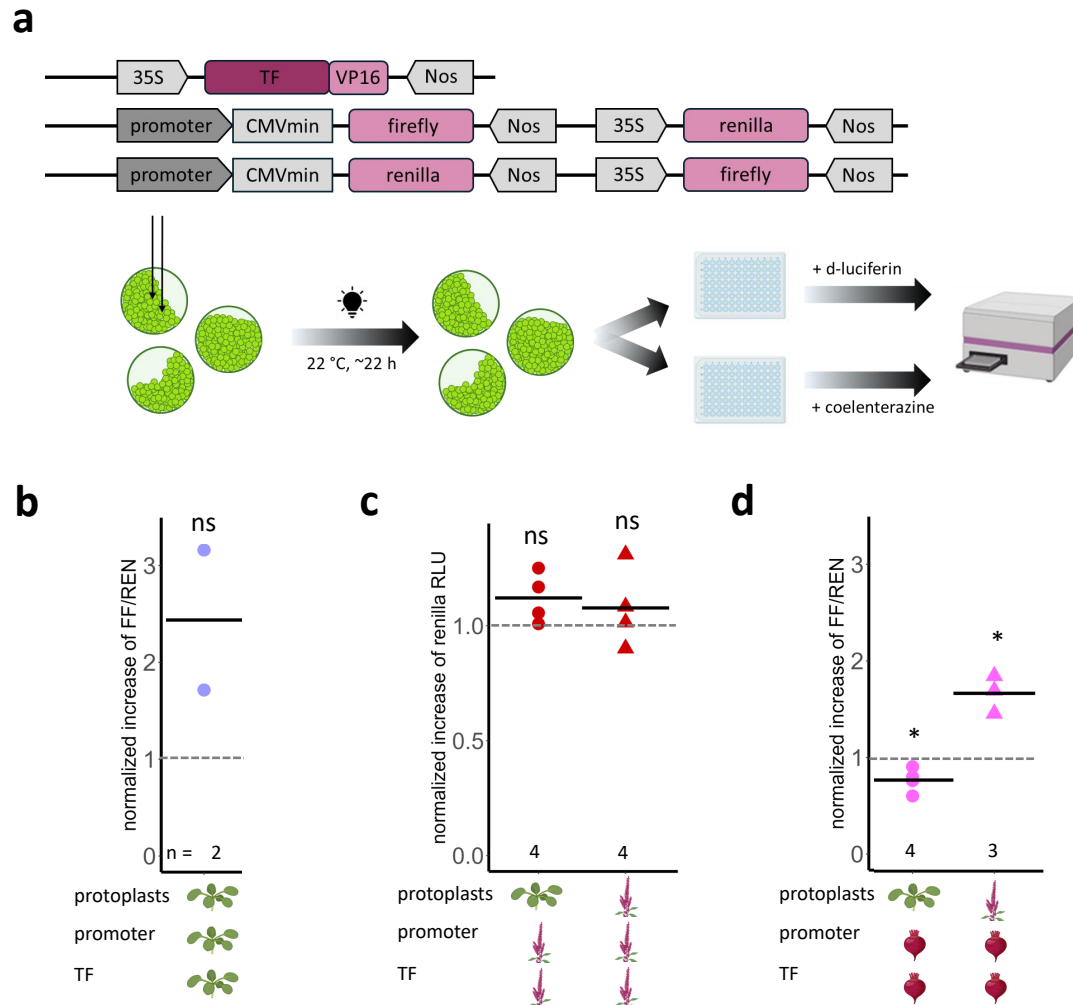


Figure 4.3: Transactivation assays in protoplasts indicate that a transcription factor specific for betalain species is required to stabilize the binding of betalain pathway-regulating MYBs. Schematic overview of the constructs and the protocol used for the transactivation assays in protoplasts (a). Grey indicates promoters and terminators, and purple translated sequences; faded parts are fixed, while the dark color highlights the parts that were specific for each tested combination. For each assay, one effector construct (with the transcription factor (TF) of interest) and one reporter construct (with the potential target promoter) were co-transformed in protoplasts. Transactivation assay in *A. thaliana* protoplasts with *AtDFR* and *AtPAP1*-VP16 showed an activation of the promoter (b). Transactivation assay with *AhCYP76AD2* and *AhMYB2.1*-VP16 showed no significant activation of the promoter neither in *A. thaliana* nor *A. hypochondriacus* protoplasts (c). Note that renilla was used as readout as it showed stronger luminescence in amaranth protoplast than firefly (pep602_sw backbone). No normalization with firefly was possible for *AhMYB2*-VP16 as a deletion in the firefly cassette in a plasmid used in some assays abolished the firefly signal in these. Transactivation assay with *BvCYP76AD1* and *BvMYB1*-VP16 showed a slight significant deactivation of the promoter in *A. thaliana* and a significant activation of the promoter in *A. hypochondriacus* protoplasts (d). Luminescence data was averaged across 4-6 protoplast aliquots from one experiment and normalized by the respective control (same reporter but a non-binding effector, dashed grey line). Data from 2-4 independent experiments are shown; a black line indicates the overall mean. Significant deviation from the respective control was tested with ANOVA with a significance threshold of 0.05. CMVmin = minimal promoter from the cauliflower mosaic virus, FF = firefly, REN = renilla. Red beet, *A. thaliana*, amaranth, plate and plate reader icons are from BioRender.

we also produced a RUBY line containing the expression cassette from He et al. (2020) with a *CYP76AD2*, *DODA1*, and a *glucosyltransferase* gene from different species. We complemented the set with an empty vector control, one line just overexpressing the first enzyme of the pathway (AhCYP76AD2), and a T0 heterogenic knockout line (Ahcyp76ad2; the line was transformed with a CRISPR cassette targeting *AhCYP76AD2* and *AhCYP76AD5*, see Vollmer et al. (2026)).

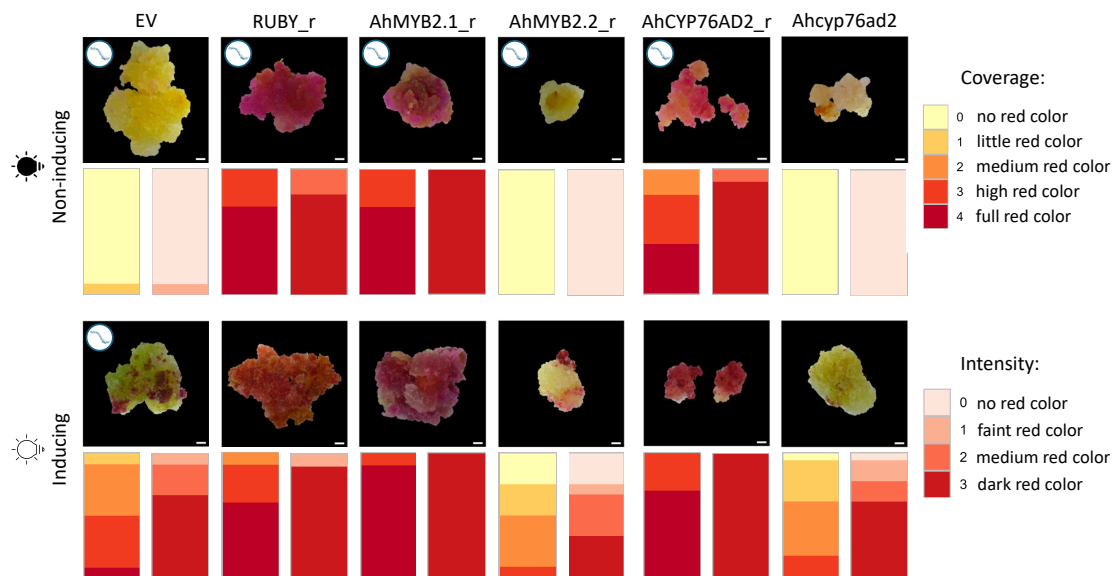


Figure 4.4: The influence of genetics and the environment on the betalain pigmentation in transgenic *A. hypochondriacus* callus. The top panel shows the result after ten days (d10) in non-inducing conditions, and the bottom panel shows the result under inducing conditions. For RUBY, AhMYB2.1, and AhCYP76AD2 lines, only the results of the red lines are shown. The pictures show one representative callus per line, while the graph shows the data of all 10-18 calli assessed per line. The coverage of the color as visible from the top was classified as follows: 0 = 0%, 1 = >0%, ≤10%, 2 = >10%, ≤50%, 3 = >50%, ≤90%, 4 = >90%, ≤100%. Color intensity as visible from the top was classified as 0 = yellow (no pigmentation), 1 = low, 2 = medium, and 3 = strong. The scale bar in the image represents 2 mm. RNA symbol in the upper left corner of the image indicate callus lines used for RNA sequencing.

Three lines (RUBY, AhMYB2.1, and AhCYP76AD2) showed strong betalain accumulation; however, there were variations in the amount of colored tissue per callus (penetration) and the intensity of the red color (expressivity). To verify that the coloration is stable and to reduce the heterogeneity of the callus lines, we split up these three lines further into a mostly yellow, a mostly red, and an intermediate line. The coloration proved to be stable, but the yellow lines tend to get more red over time. Due to the growth on the selection medium, we could assume the successful transformation of all callus parts. However, different parts of the same callus could arise from independent transformation events leading to differences in copy number or silencing. Moreover, calli could show heterogeneity with different metabolic or developmental stages in different callus parts. Both might explain the color variation we observed in the primary grown

transgenic callus. Those 10-18 calli from each callus line were then transferred to both non-inducing and inducing conditions and phenotyped at day zero (d0), day five (d5), and day ten (d10) by taking overview pictures of all plates, taking close-up pictures with a camera-equipped binocular, and manual semiquantitative counting using the binocular (details in 4.2.1). Additionally, samples were taken at d0 and d10.

Betalain production is suppressed in dark and high auxin condition, but AhMYB2.1 can still activate the production

Generally, the betalain production is suppressed in dark and high auxin conditions (non-inducing conditions) and activated by light and high cytokinin conditions (inducing conditions), as shown by the empty vector control (Figure 4.4). As expected, we could observe a strong coloration of *AhMYB2.1* but not *AhMYB2.2* overexpressing callus, in line with our previous experiments in hairy roots (Winkler et al., 2024)(Figure 4.4). Interestingly, the callus lines overexpressing *AhMYB2.2* were less red-colored, having less color coverage and intensity, than the empty vector control (Figure 4.4). We showed that the transactivation ability is abolished in AhMYB2.2 (Figure 4.1c), but the rest of the protein is identical to AhMYB2.1, meaning that it should still be able to bind to DNA. One could speculate that this causes AhMYB2.2 to be a dominant negative regulator as it blocks the *cis*-element and prevents binding of activating transcription factors (Herskowitz, 1987). It had been demonstrated in mammals that dominant negative transcription factor mutants with altered activation domains can have severe effects, in some cases causing stronger phenotypes than null mutants (McCann et al., 2021). While RUBY lines had nearly the same coloration as the AhMYB2.1 lines, the *AhCYP76AD2* overexpression lines were slightly less strongly colored (Figure 4.4). All three lines were colored irrespective of the environment, suggesting that the environmental regulation mostly controls the transcription of *AhCYP76AD2*. Since deregulating *AhCYP76AD2* transcription by transgenically overexpressing it is sufficient to overcome the suppression by the dark and high auxin condition (see section 4.3.1). However, the slightly lower coloration of the AhCYP76AD line indicates that regulation of more than just AhCYP76AD2 is required for high betalain production. This supports our hypothesis that AhMYB2.1 upregulates the betalain production by increasing *AhCYP76AD2* and *AhDODAa1* expression. The callus line expressing a CRISPR cassette targeting *AhCYP76AD2* (and *AhCYP76AD5*, see Vollmer et al. (2026)) had a reduced coloration in inducing conditions, underlining the importance of AhCYP76AD2 for the betalain pigmentation. The pigmentation was not completely abolished, since the selection with hygromycin only enabled the selection of transgenic tissue but not for successful editing, thus causing callus tissue that is heterogenous in terms of the *AhCYP76AD2* and *AhCYP76AD5* alleles. To validate our observation and investigate the transcriptional betalain regulation in more detail, we selected 6 lines (marked in Figure 4.4 in the upper left corner) for RNA sequencing

(RNA-seq).

AhMYB2.1 has more influence on the transcriptome than just activating the betalain pathway, but less compared to the non-endogenous RUBY or the environment

The environment (different phytohormones in the media and light) had by far the strongest effect on overall gene expression (Figure 4.5a). With 1203 unique differentially expressed genes (DEGs), clusters EV_inducing and next RUBY_non_inducing with 629 DEGs furthest away from both the *AhMYB2.1* and the *AhCYP76AD2* overexpression callus lines (Figure 4.5a, b). With overall only 109 and 300 DEGs, the influence on gene expression in the *AhCYP76AD2* and *AhMYB2.2* overexpression lines was only low. The biggest overlaps of DEGs are between the *RUBY* and *AhMYB2.1* overexpression lines, RUBY_non_inducing and EV_inducing, and MYB2.1_non_inducing and EV_inducing. Considering that these are the three samples with a significantly higher amount of DEGs, is this observation not surprising. The low number of differentially expressed genes in CYP76AD2_non_inducing suggests that the transcriptional activation of the endogenous betalain pathway and the subsequent accumulation of betalains has a rather low effect on the global transcriptome. In contrast, the strong expression of all three required betalain pathway enzymes by non-endogenous genes in the RUBY cassette caused many times more DEGs than observed in CYP76AD2_non_inducing or even MYB2.1_non_inducing. Raising the question of what could explain the strong effect of RUBY on the transcriptome if high betalain production *per se* seems to have only little influence on the global transcriptome. The strong effect of RUBY on the transcriptome, even in a betalain-producing species, might explain why other phenotypes besides coloration (e.g., stunted growth (Prasad et al., 2024; Bi et al., 2026)) were reported in many species transformed with the RUBY cassette. Only 10 genes were differentially expressed in all five samples.

Also, light and cytokinin activate the betalain pathway by upregulating *AhCYP76AD2* and *DODAα1* but do not require upregulation of *AhMYB2.1*

Investigating the expression of the betalain pathway genes in all samples solidified previous results and verified multiple hypotheses (Figure 4.6a). Firstly, the data show the successful and strong overexpression (around 600-1000 times higher compared to control) of *AhMYB2.1*, *AhMYB2.2*, and *AhCYP76AD2* by the *35S* promoter in the respective transgenic lines, validating the *Agrobacterium*-mediated stable transformation and the functionality of the *35S* promoter in grain amaranth. The upregulation of *AhCYP76AD2* in EV_inducing confirms the observation that *AhCYP76AD2* is the main target for reg-

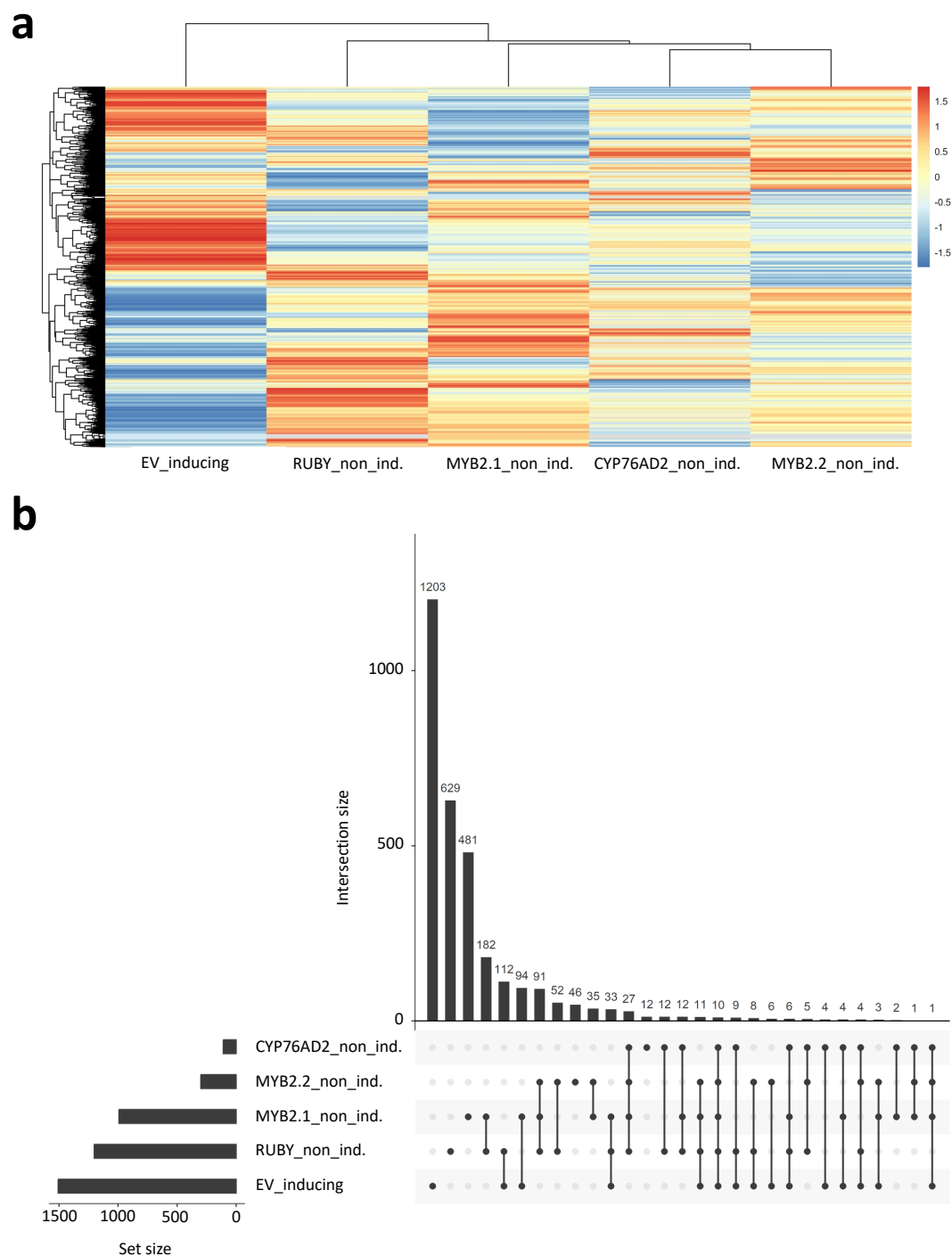


Figure 4.5: Overview of RNA-seq results from transgenic *A. hypochondriacus* callus showing the influence on gene expression of both environmental and transgenic-caused activation of betalain production. Heatmap showing relative expression of all expressed genes in all samples compared to the control (EV_non_inducing)(a). Logfold2 values were row-normalized to highlight the difference between samples within each gene, red indicating higher and blue indicating lower expression. Upsetplot showing the overlap of differentially expressed genes compared to the control (EV_non_inducing) between different samples (b).

ulation, i.e., its expression is usually kept low to prevent betalain biosynthesis and is upregulated if the betalain production is induced. Moreover, the significant upregulation of *AhCYP76AD2* and *AhDODAA1* in the MYB2.1_non_inducing verifies that MYB2 is upregulating both genes and thus inducing betalain accumulation. As the upregulation of *AhCYP76AD2* and *AhDODAA1* could only be observed in MYB2.1_non_inducing, but their expression remained unchanged in MYB2.2_non_inducing, indicates that the R3 domain of MYB2.1 is required for the upregulation. That *AhDODAA1* is not upregulated in the *AhCYP76AD2* overexpression line excludes that it only gets upregulated indirectly by higher activity of the *AhCYP76AD2* gene, but that both genes are directly upregulated by MYB2.1. Interestingly, *AhMYB2* is not upregulated in EV_inducing, suggesting that other transcription factors (TFs) are responsible for the betalain accumulation induced by light and phytohormones. Since also in the environmentally induced betalain accumulation both *AhCYP76AD2* and *AhDODAA1* were upregulated, this suggests that both genes get generally regulated together. This might hint at a co-regulator of AhMYB2 that is also involved in the environmental regulation of the betalain pathway. To better estimate the redirection of the aragonate metabolisms to betalains we assessed the expression of all betalain and flavonoid genes in the MYB2.1_non_inducing line (Figure 4.6b). The upregulation of *AhCYP76AD2* and *AhDODAA1* increases the flow from tyrosine to the betalain precursors. However, one of the two *ADHa* copies got significantly downregulated, which might reduce the flux from aragonate to tyrosine. Flavonoid genes were not differentially regulated in the MYB2.1_non_inducing line except for one potential *CHS* gene, which had a very low base-level expression. Despite its evolutionary origin from PAP1 and other anthocyanin regulator MYBs, AhMYB2 is thus highly specific and has lost the ability to regulate the anthocyanin pathway genes.

No transcription factor class is significantly enriched or depleted in the set of differentially expressed genes in MYB2.1_non_inducing callus

As a first step to find candidates for potential MYB2 co-regulators, we checked the number of genes from the main transcription factor classes in the entire data set with all expressed genes and the subset containing all genes that are differentially regulated in at least one of the samples (Figure 4.7a). All included main classes together covered most of the TFs found (86% and 82% in the expressed genes and in the DEGs, respectively). A chi-square test identified only IAA transcription factors as significantly enriched in the combined DEGs of all samples. No significant enriched or deleted TF class was found in the chi-square test comparing only the MYB2.1_non_inducing DEGs compared to all expressed genes. The overlap of differentially expressed TFs mostly mirrors the one for all DEGs: most unique differentially expressed genes have EV_inducing, RUBY_non_inducing, and MYB2.1_non_inducing, and the largest overlaps are between the pairs of each of them (Figure 4.7b). The presence, location, and distribution of potential binding sites of

selected main transcription factor classes were predicted for the *AhCYP76AD2* (Figure S4.4a) and *AhDODAA1* promoter (Figure S4.4b) using web tools (see 4.2.8). Binding sites for many main transcription factor classes were found in both promoters, suggesting potential (co-)regulation by other transcription factor classes.

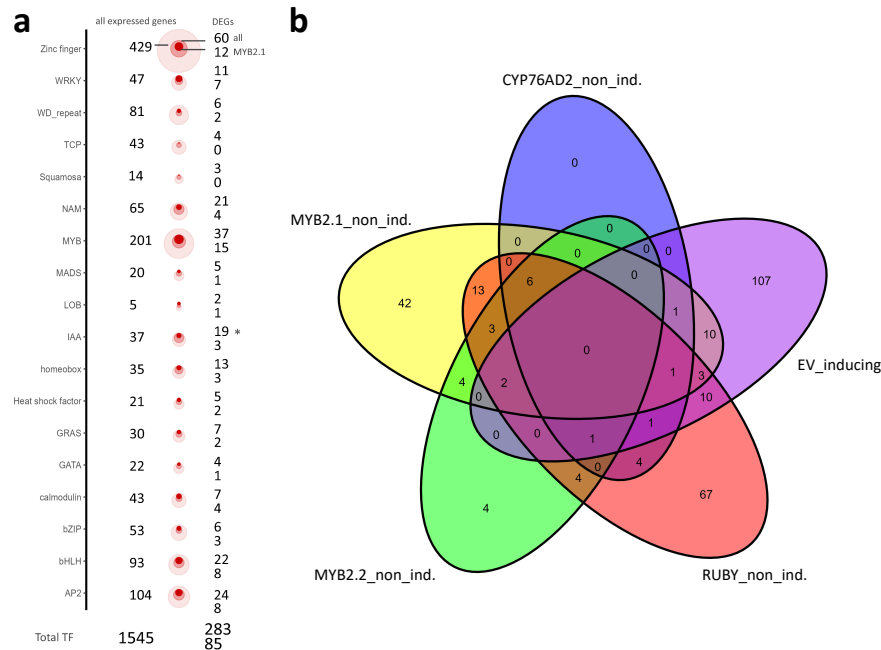


Figure 4.7: Comparison of differentially expressed transcription factors (TFs) and the frequency of main TF classes between samples. Number of genes from main transcription factor classes in the data set containing all expressed genes and in the set containing only differentially expressed genes (DEGs) from all samples or in MYB2.1_non_inducing (a). Overrepresentation of certain classes was tested with a chi-square test, finding only IAA transcription factors significantly overrepresented in the set of all DEGs, indicated by the asterisk. Venn diagram showing the overlap of differentially expressed TFs across all samples (b).

Not the homolog of *AtrWRKY42-2* and *AtTTG2*, but several other WRKY transcription factors are differentially expressed in MYB2.1_non_inducing callus

Except for AhMYB2 and its homologs AtrMYB72 from *A. tricolor* (Xue et al., 2025) and BvMYB1 from *B. vulgaris* (Hatlestad et al., 2015) is only AtrWRKY42-2 from *A. tricolor* (Yang et al., 2023) reported to regulate betalains in plants belonging to the betalain origin 2 (according to Pucker et al. (2024)). Thus, we took a closer look at the WRKY transcription factors from amaranth and their differential expression in the different samples (Figure 4.8a). Regarding the differential expression cluster EV_inducing and MYB2.1_non_inducing closest to each other, as most of the differentially expressed WRKY genes get similarly up- or downregulated in both samples. To find closely related WRKY genes and, as a comparison, the closest homologs from *A. thaliana* and

AtrWRKY42, we extracted all WRKY genes from *A. hypochondriacus* and *A. thaliana* and constructed a phylogenetic tree (Figure 4.8b). Most differentially expressed genes in EV_inducing and MYB2.1_non_inducing can be found close to each other in the phylogeny: AHq024477 and AHq017085 form a tip in the phylogeny, and both got down-regulated in both EV_inducing and MYB2.1_non_inducing. Their next closely related neighbors in the tree are AHq000664 and AT1G80840, where AHq000664 was also differentially regulated in both EV_inducing and MYB2.1_non_inducing. However, different from AHq024477 and AHq017085, AHq000664 got upregulated. All three genes had the same direction also in all other samples. The same pattern of similar expression and high sequence similarity is also the case for AHq001528 and AHq000417, who both got upregulated in EV_inducing and MYB2.1_non_inducing and, to a much lesser extent, in the other samples. Their closest homolog in *A. thaliana* is AT3G58710, a clade IIe WRKY involved in thermomorphogenesis. Additionally, AHq016908 and AHq022840 got similarly regulated in both EV_inducing and MYB2.1_non_inducing, while AHq018915, AHq010170, and AHq004776 got downregulated in MYB2.1_non_inducing but (not significantly) upregulated in EV_inducing. Notably, AHq018915 and AHq016908 both have an altered WRKY domain (WRKYGQK, the glutamine was replaced by glutamic acid or lysine, respectively). Interestingly, the homolog of AtrWRKY42-2 ([Xue et al., 2025](#))/AtTTG2 (AT2G37260) ([Gonzalez et al., 2016](#))/PH3 ([Verweij et al., 2016](#)), which were shown to influence betalain, protoanthocyanin, and anthocyanin pigmentation, respectively, AHq005364, was not differentially expressed in any of the samples.

To conclude, the RNA-seq data validated *AhCYP76AD2* as the main regulatory target for the regulation of the betalain pathway and that MYB2 is activating the betalain pathway by upregulating *AhCYP76AD2* and *AhDODAa1* expression. Moreover, indicate the gene expression data from EV_inducing that MYB2 is not involved in the betalain pathway activation by light and phytohormones (i.e., auxin and cytokinin). Additionally, did the gene expression analysis of flavonoid genes in the MYB2.1_non_inducing transgenic callus lines suggest that, despite evolutionary origins from flavonoid regulators, AhMYB2 now precisely targets the betalain pathway genes and not the flavonoid genes.

4.4 Discussion

We previously identified AhMYB2 as the first positive transcriptional regulator of the betalain pathway in grain amaranth. However, how AhMYB2 is regulating the betalain pathway in grain amaranth remained unknown. Thus, the goal of this study was to reconstruct the molecular mode of action by combining *in vivo* experiments in grain amaranth with high-throughput molecular experiments in heterologous organisms. Using stably transformed *A. hypochondriacus* callus we could validate that *AhCYP76AD2* is

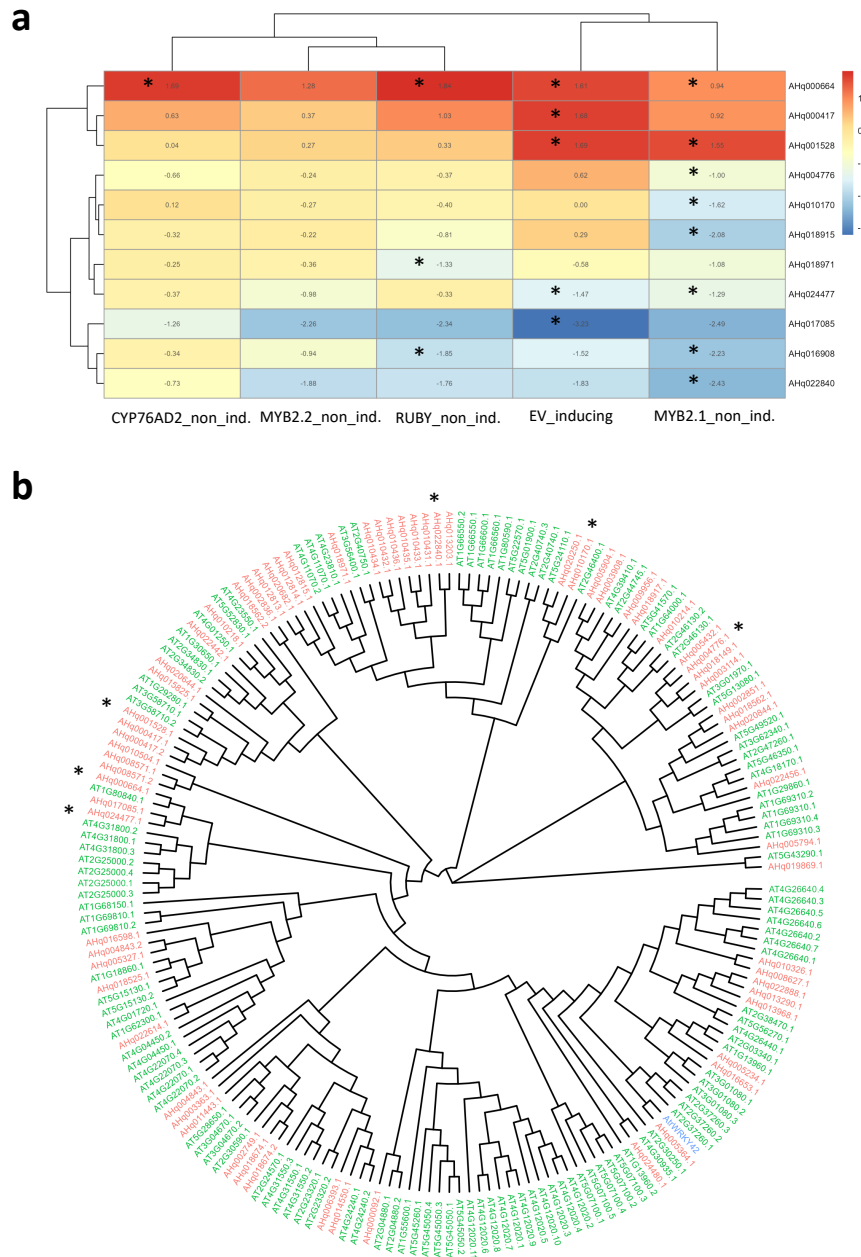


Figure 4.8: Differentially expressed WRKY transcription factors in environmentally or transgenically induced betalain accumulation in *A. hypochondriacus* callus and the closest related *A. thaliana* homologues from all WRKY transcription factors in *A. hypochondriacus*. Heatmap showing the logfold2 values from WRKY transcription factors, which are differentially regulated in at least one sample (a). Differentially expressed genes of each sample are marked with asterisks. Phylogeny of all *A. hypochondriacus* (red) and *A. thaliana* (green) WRKY transcription factors and additionally, the only published WRKY transcription factor involved in betalain regulation in amaranth, AtrWRKY42 (blue)(b). Genes differentially expressed in MYB2.1_non_inducing transgenic callus lines are highlighted with asterisks. Note that the DEGs AHq018915 and AHq016908 are not included in the tree, as they have an altered WRKY domain (WRKYGQK, the glutamine was replaced by glutamic acid or lysine, respectively).

the main regulatory target of the betalain pathway both in tissue-specific regulation by AhMYB2 and in environmentally regulated betalain accumulation by light and phytohormones. However, the environmental regulations seem not to require AhMYB2. We could confirm that AhMYB2 is increasing the betalain production by activating *AhCYP76AD2* and *AhDODAa1* expression. AhMYB2 has transactivation ability but could not interact with various bHLH transcription factors in Y2H. Transactivation assays in protoplasts suggest that AhMYB2 might bind to its target promoter directly but require an additional betalain-specific TF to stabilize the DNA binding. Finally, we performed RNA-seq to obtain a valuable database for the identification of additional (co-)regulators of the betalain pathway in grain amaranth.

4.4.1 AhMYB2 regulates the betalain pathway by controlling the transcription of *AhCYP76AD2* and *AhDODAa1*

The first or an early enzyme is the point of regulation in many metabolic pathways (Blanco and Blanco, 2022; Bhagavan and Ha, 2015). The expression of *AhCYP76AD2* and *AhDODAa1*, the first two enzymes of the betalain pathway, correlates with betalain levels in multiple species (Hatlestad et al., 2015; Cheng et al., 2017)(Figure 3.5). Nevertheless, it was not yet demonstrated that the transcriptional control of *AhCYP76AD2* is the main mechanism for regulation of the betalain biosynthesis. Stable overexpression of *AhCYP76AD2* in grain amaranth induced strong coloration (Figure 4.4), indicating that overwriting the transcriptional downregulation of *AhCYP76AD2* is sufficient to activate the betalain biosynthesis pathway. Both AhMYB2-mediated and light- and phytohormone-mediated induction of the betalain pathway lead to a higher expression of *AhCYP76AD2* and *AhDODAa1* (Figure 4.6a). Thus, our results clearly suggest that the tissue-specific (AhMYB2) and at least some environmental (light) and internal signals (phytohormones) regulate the betalain production by controlling the transcription of *AhCYP76AD2*. This is in line with the regulatory network controlling anthocyanin biosynthesis in anthocyanin-producing plants: Tissue-specific accumulation of anthocyanins is achieved by spatial control of transcription factor expression, which in turn regulates the expression of the genes encoding for the pathway enzymes (Wang et al., 2019; Deluc et al., 2008; Yao et al., 2017; Martins et al., 2013). Likewise, do light-induced and phytohormone-mediated regulation of anthocyanin accumulation act via controlling either the expression of pathway genes directly or by regulating transcription factors, which in turn influence the expression of pathway genes (Araguirang and Richter, 2022). Besides the evidence for a transcriptional control of *AhCYP76AD2*, many enzymes are regulated biochemically and/or by post-transcriptional or -translational modification. For PAL, the first enzyme of the anthocyanin pathway, it was shown that its activity is regulated by phosphorylation and ubiquitination (Allwood et al., 1999; Zhang et al., 2014) and biochemically by its product, cinnamic acid. Moreover, multiple microRNAs (miRNAs),

especially the conserved *miR156/157b*, are known to regulate anthocyanin biosynthesis post-transcriptionally (Sharma et al., 2024). Similarly, it was shown in *H. undatus* that *Hmo-miR156/157b* expression correlates with fruit ripening and influences the expression of SQUAMOSA promoter binding protein-like (SPL) transcription factors, which interact with betalain regulators in pitaya (Zeng et al., 2023). Post-transcriptional silencing by small RNAs was shown for CYP76AD in bougainvillea (Ohno et al., 2021). Thus, future studies need to investigate the possibility of post-translational modifications on enzyme activity, post-transcriptional silencing, and biochemical feedback loops being involved in betalain regulation.

4.4.2 AhMYB2 seems different from its anthocyanin homologs, not acting in MBW complexes

MYB transcription factors regulating the anthocyanin pathway are well known to function mainly in MBW complexes. To test if this also applies to the homologous AhMYB2, we performed Y2H with the *A. hypochondriacus* homologs of AtTT8 and AtTTG1. With neither the bHLH nor the WD40 transcription factor, we found interactions with AhMYB2 or each other (Figure 4.1d). A comparison of the conserved residues of the R3 domain in betalain and anthocyanin PAP1 homologs indicated that both AhMYB2 and BvMYB1 have a partially altered bHLH domain (Figure 4.1b). We additionally tested multiple *A. thaliana* bHLH from the clade III_f, which are all involved in MBW complexes, but could not find any interaction with AhMYB2, suggesting that AhMYB2 might have lost the ability to interact with (clade III_f) bHLH transcription factors. This is in line with the observation from Hatlestad et al. (2015), who tested the interaction of AtEGL3 and BvMYB1 and could detect interaction only after mutating 6 residues of BvMYB1 that differed from AtPAP1. Fitting to this observation was AHq020375, the closest homolog of AtMYC1, not expressed in any callus sample investigated in section 4.3.4 and barely expressed in any tissue except for a low expression in mature seeds (tissue-specific RNA-seq data set from Clouse et al. (2016) with annotation from Graf et al. (2025), for details see section 3.4 and Supplementary Table S3.1 of Winkler et al. (2024)). Similarly, AHq013577, the closest homolog of AtTT8, was only expressed lowly in seeds and not in any other tissue (Graf et al., 2025; Clouse et al., 2016) and was also not differentially regulated in any callus sample. Moreover, the homolog of AtEGL3 and AtGL3 seems to have been lost in the whole *Amaranthaceae* (Pucker et al., 2024). Nevertheless, AhMYB2.1 showed strong autoactivation in Y2H, indicating transactivation ability. That the activation of the pathway can be uncoupled from the interaction in the MBW complex was also shown for a few flavonoid-regulating MYBs that activate the pathway genes independently of the MBW complex (Grotewold et al., 1994; Premathilake et al., 2020; Jiang et al., 2021). Interestingly, all three have altered R2R3 domains ranging from a few amino acids to a deletion of half of the R3 domain. To conclude, our Y2H data

provide an important piece of evidence to the hypothesis that AhMYB2 kept its function in activating the enzyme's gene expression, but without the involvement of a class IIIb bHLH transcription factor. It could be speculated that this evolved to uncouple betalain-regulating AtPAP1 homologs from the MBW complex. Thus avoiding pleiotropism with anthocyanin, trichome, root hair, and mucilage regulation, which is controlled by MBW complexes involving PAP1 interacting class IIIb bHLH transcription factors (Zhang and Hülskamp, 2019; Heim, 2003; Schuh and Morris, 2025).

4.4.3 AhMYB2 requires additional betalain-specific transcription factors for efficient DNA binding

Motivated by the transactivation ability of AhMYB2.1 in Y2H and to test for direct binding to the AhCYP76AD2 promoter, we performed transactivation assays in protoplasts. The positive control using AtPAP1-VP16 caused stronger reporter expression compared to the background in *A. thaliana* protoplasts (Figure 4.3b), indicating binding of AtPAP1 to the *AtDFR* promoter as demonstrated previously (Guo et al., 2023; He et al., 2021). In contrast, for AhMYB2.1-VP16 and the *B. vulgaris* homolog BvMYB1-VP16, we could not detect significant activation in *A. thaliana* protoplasts (Figure 4.3c,d). Since we used a positive control for the assay in each experiment, we could exclude technical issues with the assay. Results with a synthetic positive control (Urquiza-Garcia et al., in preparation) specific to an amaranth promoter suggest that the promoters are not blocked by other plant transcription factors and are generally accessible (Supplementary Figure S4.3c). While we only hypothesize that AhMYB2 can directly bind to the *AhCYP76AD2* promoter, was direct binding of BvMYB to the *BvCYP76AD1* and *BvDODA1* promoters already validated in yeast-one-hybrid (Y1H) (Hatlestad et al., 2015). It was thus surprising that we could not detect binding in *A. thaliana* protoplasts. However, after implementing the transactivation assay in *A. hypochondriacus* protoplasts, we could detect binding of BvMYB1-VP16 to the *BvCYP76AD1* promoter. We thus hypothesize that betalain-regulating PAP1 homologs might require co-regulators specific to betalain-producing plants to stabilize the DNA binding *in planta*. Supporting our hypothesis, a preliminary test in mammalian cells showed a small, non-significant activation of the *AhCYP76AD2* promoter by AhMYB2-VP16, suggesting binding but with low efficiency (Supplementary Figure S4.3b). Compared to the highly optimized *A. thaliana* protoplast system, the signal strength in *A. hypochondriacus* protoplasts was much lower, which might explain why we could not detect significant binding of AhMYB2.1-VP16 even in *A. hypochondriacus* protoplasts. Further optimizing the transactivation assay in *A. hypochondriacus* protoplasts should increase the sensitivity and thus the probability of detecting weak binding. Similarly, it might help to increase the amount of effector DNA used for protoplast transformation to increase sensitivity for weakly binding TF. To conclude, although our transactivation assay data suggest direct binding of AhMYB2.1

to *AhCYP76AD2* depending on betalain-specific co-regulators, we were unable to prove it. Thus, additional studies are required, using an optimized *A. hypochondriacus* protoplast transformation protocol for transactivation assays or other methods, e.g., Y1H, to validate promoter binding *in vivo*.

4.4.4 Candidate transcription factors for (co-)regulation of the betalain pathway in amaranth

The transactivation assay data suggested that AhMYB2 requires a betalain-specific co-regulator for efficient DNA binding. This is highly probable, as many MYB transcription factors act in complexes with other TFs (Zhang et al., 2025a; Mariyam et al., 2025). According to BAR (<https://bar.utoronto.ca/eplant/>), there were already 64 proteins, mostly TFs, found that AtPAP1, a homolog of AhMYB2 in *A. thaliana*, interacts with. Besides the MBW complex, many other transcription factors from various classes were shown to influence transactivation activity from certain PAP1 homologous MYBs (see section below, e.g., (Ni et al., 2019; Yao et al., 2017; An et al., 2020; Malacarne et al., 2016; An et al., 2018, 2019; Zhang et al., 2020; Xu et al., 2023; Zhu et al., 2025; Zhang et al., 2024; Li and Zachgo, 2013; An et al., 2020b; Lee et al., 2024; Liu et al., 2023b; Qi et al., 2011)). Combining the Y2H results, the RNA-seq data, and the literature, we selected four TFs (Table 4.1) representing diverse classes as the most promising candidates to co-regulate the betalain pathway in grain amaranth with AhMYB2:

A promising TF class for AhMYB2 co-regulators is the plant-specific APETALA2/Ethylene-Responsive element binding Factor (AP2/ERF) family, characterized by the presence of one or two APETALA2 (AP2) domains (Riechmann and Meyerowitz, 1998). In *A. thaliana* it was shown that AtERF4 and AtERF8 are involved in light-induced anthocyanin regulation (Koyama and Sato, 2018). However, direct interaction of AtERF4 and AtERF8 with AtPAP1 or AtPAP2 was not investigated (Koyama and Sato, 2018). In pear blue light-induced Pp4ERF24 and Pp12ERF96 were shown to interact directly with PpMYB114 and increase its upregulation of an anthocyanin pathway gene, thus probably playing a role in the light-induced anthocyanin regulation (Ni et al., 2019). Similarly, does PpERF3 interact with PpMYB114 and increase its upregulation of several anthocyanin pathway genes (Yao et al., 2017). In apple, MdERF38 was shown to interact with the apple PAP1 homolog MdMYB1, influencing its DNA-binding strength (An et al., 2020). MdERF38 expression is strongly drought responsive, and it is thus thought that MdERF38 mediates the drought-dependent upregulation of anthocyanin biosynthesis in apple (An et al., 2020). Besides MdERF38, it had been shown that MdERF1B interacts with MdMYB1 (and other anthocyanin-regulating MYBs), although for this ERF it is hypothesized that it regulates anthocyanin expression by regulating the expression of other MYB anthocyanin-regulating TFs (Zhang et al.,

2018). Also for litchi and citron, ERFs were identified that influence the expression of anthocyanin pathway genes, but interaction with PAP1 homologs was not tested (Zhuo et al., 2024; Zhao et al., 2021). Several AP2/ERF transcription factors were differentially expressed in the transgenic callus lines (Supplemental Figure S4.7). Four AP2/ERF genes were upregulated in MYB2.1_non_inducing: from these, AHq018982 was only differentially regulated in MYB2.1_non_inducing, and AHq017795 was additionally also upregulated in EV_inducing (Supplemental Figure S4.7). The closest *A. thaliana* homolog of AHq018982 is AT1G64380 (AtERF61), which is not characterized yet. Based on its phylogenetic classification, AtERF61 could be involved in drought response but was also shown to be upregulated upon elicitor treatment (Libault et al., 2007; Wang et al., 2024). The expression of AHq018982 was rather low across tissues (<6 transcripts per million (TPM)), being highest in leaves, flowers, and water-stressed tissue (Graf et al., 2025; Clouse et al., 2016). Additionally, four AP2/ERF genes were downregulated in MYB2.1_non_inducing, including the neighboring genes AHq019220 and AHq019221, which were both upregulated, albeit not significantly, in EV_inducing (Supplemental Figure S4.7). Sequence-wise, they are closely related to Pp4ERF24 and Pp12ERF96 and cluster also not far away from MdERF1B (Supplemental Figure S4.10). Both had the highest expression in flowers (13.62 TPM and 22.60 TPM for AHq019220 and AHq019221, respectively) but were also expressed in root and stem (> 5 TPM), which are the three main expression domains of *AhCYP76AD2* and *AhDODAa1*. The closest homolog of MdERF38 is AHq019203, which was downregulated in all callus samples, but significantly only in MYB2.2_non_inducing and RUBY_non_inducing (Supplemental Figures S4.7 and S4.10). AHq019203 also had the highest expression in flowers, but only with 7.76 TPM, and had in no other tissue an expression of 5 TPM or higher (Graf et al., 2025; Clouse et al., 2016). Since ERFs had been found to influence PAP1 homolog-mediated anthocyanin regulation in diverse species, ERFs are promising candidates for an AhMYB2 co-regulator. However, most of these ERFs do not cluster close to each other in the phylogeny, hampering the identification of specific *A. hypochondriacus* candidate ERF genes (Supplemental Figure S4.10). Based on the combination of the literature with the RNA-seq data from various tissues and transgenic callus, are AHq019220 and AHq019221 among the best candidates.

Another class of TF, which was shown to be involved in anthocyanin regulation in various species, are bZIP (basic region/leucine zipper) transcription factors (An et al., 2018; Malacarne et al., 2016). A characteristic of bZIP TF is an α -helix that contains a basic N-terminal region for DNA binding and a leucine-zipper region at the C-terminal for dimerization (Jakoby et al., 2002). bZIP TFs were found in all eukaryotes but have expanded in plants (Riechmann et al., 2000; Dröge-Laser et al., 2018). One interesting bZIP gene identified in the RNA-seq data set is AHq004055. The closest *A. thaliana* homolog is AT1G75390 (AtbZIP44), which belongs to the lowly characterized subclass

S, which might be involved in carbohydrate demand/supply balancing (Jakoby et al., 2002). However, the closest homolog for both in apple is MdbZIP44, which was demonstrated to mediate ABA-induced anthocyanin regulation by increasing the DNA-binding of MdMYB1 and thus the upregulation of anthocyanin pathway genes (An et al., 2018). Similarly, it was shown in grapes that UV-light-induced VvibZIPC22 increases the transactivation of PAP1 homologs (VviMYBF1 and VviMYBPA1) of anthocyanin pathway genes when cotransformed (Malacarne et al., 2016). AHq004055 was significantly down-regulated in MYB2.1_non_inducing but not differentially expressed in any other investigated callus line (Supplemental Figure S4.8). The expression of AHq004055 across tissues was rather low (<16 TPM), but highest in flowers, reaching a moderate expression of 25 TPM (Graf et al., 2025; Clouse et al., 2016). VvibZIPC22 belongs to the grape C subclass of bZIP transcription factors, which corresponds to the S subclass in *A. thaliana*, and thus belongs to the same class as MdbZIP44 and AHq004055 (Malacarne et al., 2016). Subclass S bZIPs from *A. thaliana* are known to form heterodimers with subclass C bZIPs; this interaction network seems to be highly conserved and is hypothesized to regulate diverse (metabolic) pathways dependent on the plant energy level (Ehlert et al., 2006; Dröge-Laser et al., 2018). One potential amaranth subclass C bZIP is AHq008107, which was not differentially expressed in any investigated callus line (Supplemental Figure S4.7) but was highest expressed in flowers, albeit only with 13 TPM. A bZIP TF is hence a promising candidate for an AhMYB2 co-regulator. Based on the homology, are AHq004055 and, as a potential heterodimerization partner, AHq008107, good candidates.

Another promising TF class are plant-specific WRKY transcription factors, which could be identified by the presence of one or two WRKYGQK domains at the N-terminus and a zinc-finger-like domain at the C-terminus (Rushton et al., 1995). For *Arabidopsis thaliana* (Devaiah et al., 2007; Duan et al., 2018), *Brassica napus* (Duan et al., 2018), petunia (Verweij et al., 2016), grape (Amato et al., 2019), kiwifruit (Peng et al., 2020), and apple (An et al., 2019), WRKY transcription factors were found to be involved in anthocyanin pathway regulation. Moreover, a WRKY TF is the only other TF class reported for betalain regulation in amaranth besides MYB (Yang et al., 2023). For the wound-induced MdWRKY40 in apple, it was shown that it interacts directly with the PAP1 homolog MdMYB1 and increases its DNA binding and thus upregulation of anthocyanin pathway genes (An et al., 2019). In contrast, for AtTTG2 in *A. thaliana* it was demonstrated that it interacts with the MBW complex only via AtTTG1, not the MYB transcription factor, but similarly amplifying the transactivation of the MBW complex (Gonzalez et al., 2016; Pesch et al., 2014). Likewise, VvWRKY26 in grape (Amato et al., 2019), PH3 in petunia (Verweij et al., 2016), and potentially also WKRY44 in kiwifruit (Peng et al., 2020) seem to regulate anthocyanin production by interacting with the WD40 TF of the anthocyanin-regulating MBW complex and increasing its transactivation strength. Also, the identified WRKY TF from *A. tricolor*, AtrWRKY42-2 is the

homolog from AtTTG2 (Yang et al., 2023)(Figure 4.8b, At2G37260). However, the closest homolog of AtTTG2 in *A. hypochondriacus* is AHq005364, which was not differentially expressed in any of the investigated callus lines (Figure 4.8), nor was it highly expressed in any tissue (max 8 TPM in young seeds) (Graf et al., 2025; Clouse et al., 2016). In contrast, the closest homolog of MdWKRY40 is AHq000664, which was significantly up-regulated in all callus lines except the non-colored MYB2.2_non_inducing (Figure 4.8) and was highest expressed in flowers and leaves with 14 and 25 TPM. Moreover, we found multiple predicted binding sites for WRKY transcription factors in the *AhCYP76AD2* and *AhDODAA1* upstream regions (Supplemental Figure S4.4). Thus, WKRY TFs and especially AHq000664 are promising candidates as an AhMYB2 betalain co-regulator in grain amaranth.

NAC TFs (named after NO APICAL MERISTEM (NAM), ARABIDOPSIS TRANSCRIPTION ACTIVATOR FACTOR 1/2 (ATAF1/2), and CUP-SHAPED COTYLEDON (CUC2)) represent another plant-specific family of transcription factors that are involved in anthocyanin regulation in various species. TF belonging to the NAC family can be identified by the highly conserved N-terminal NAC domain, which mediates dimerization, DNA binding, and specificity, while the C-terminal part is important for the biological function, often containing a transactivation domain (Mohanta et al., 2020). NAC TFs were reported to be involved in anthocyanin regulation in many species; however, most of them showed or implied that the regulation is achieved by regulating the expression of PAP1 homologs or other anthocyanin-regulating TFs (Liu et al., 2023a; Sun et al., 2019; Zou et al., 2023; Geng et al., 2022; Zhou et al., 2015; Jiang et al., 2019; Li et al., 2023; Wang et al., 2023a, 2025). While others did not test the interaction with PAP1 proteins or their promoter (Mahmood et al., 2016; Morishita et al., 2009; Qu et al., 2022; Shi et al., 2025; Wang et al., 2018). Nevertheless, there are a few NAC proteins that have been demonstrated to regulate the anthocyanin pathway by interacting with a PAP1 homolog and increasing their transactivation of anthocyanin pathway genes (Zhang et al., 2020; Xu et al., 2023; Zhu et al., 2025; Zhang et al., 2024). Two of them (VrNAC25 in mung bean (Zhu et al., 2025) and MdNAC42 in apple (Zhang et al., 2020)) are both homologs of *A. thaliana* AtNAC025, while the other ones are homologous to AtNAC097 (MdNAC14-Like, (Xu et al., 2023)) or AtNAC019 (MdNAC33, (Zhang et al., 2024)). The homolog of AtNAC025 in *A. hypochondriacus* is AHq022750, which was highly upregulated in MYB2.1_non_inducing but not in any other callus line (Supplemental Figure S4.9). Regarding the tissue-specific expression, it was highest expressed in young and mature seeds (291 and 115 TPM, respectively), followed by flowers with 25 TPM, while there was nearly no expression in all other investigated tissues (<1 TPM) (Graf et al., 2025; Clouse et al., 2016). Furthermore, were multiple potential binding sites for NAC transcription factors predicted in the *AhCYP76AD2* and *AhDODAA1* upstream regions (Supplemental Figure S4.4). Thus, NAC TFs and in particular, AHq022750, are promis-

ing candidates as AhMYB2 betalain co-regulators.

Besides the four above-described TF families, there are also members from other TF families known to regulate anthocyanins in some species. *teosinte branched1/cycloidea/proliferating* (TCP) TFs, identified by having a TCP domain containing a non-canonical bHLH motif (Cubas et al., 1999), were found to be involved in anthocyanin regulation in *A. thaliana* and apple (Viola et al., 2016; An et al., 2020b; Li and Zachgo, 2013; Meng et al., 2022). AtTCP3 had been shown to interact with several R2R3-MYBs involved in MBW complexes (but not with the bHLH or WD-TF in the complex), including AtPAP1, and enhance the transactivation of anthocyanin pathway genes (Li and Zachgo, 2013). In apple it was demonstrated that the high-light-regulated MdTCP46 interacts with MdMYB1 and increases the binding to anthocyanin pathway genes, thus increasing their upregulation (An et al., 2020b). In an apple relative, MhTCP4, which belongs to a different TCP subclass than MdTCP46, was found to regulate anthocyanin pathway genes directly, possibly independently from PAP1 homologs (Meng et al., 2022). Amaranth homologs of AtTCP3 were generally very lowly expressed and not differentially regulated in any callus sample. One homolog of MdTCP46, AHq018501, was significantly downregulated in MYB2.2_non_inducing and in none of the investigated tissues expressed with more than 7 TPM (Graf et al., 2025; Clouse et al., 2016). In summary, also considering the inability of AhMYB2 to interact with bHLHs, TCP TFs seems less promising as AhMYB2 betalain co-regulators.

The brassinosteroid-response-TF, AtBZR1, was demonstrated to interact with AtPAP1 and increase its activation of anthocyanin pathway genes important for the low-nitrogen-induced anthocyanin accumulation (Lee et al., 2024). However, in apple, it was shown that MdBZR1 interacts with other TFs to repress anthocyanin (Li et al., 2025; Wang et al., 2023b). The amaranth homolog of AtBZR1 is AHq009144, which was not differentially expressed in any callus sample, albeit the highest upregulated in MYB2.1_non_inducing and moderately expressed across tissues with a peak of 36 TPM in flowers (Graf et al., 2025; Clouse et al., 2016). The tissue expression pattern would fit a betalain regulator, but due to the mixed results in literature, AHq009144 is rather a low-priority candidate.

For many species, MADS-box (named after MINICHROMOSOME MAINTENANCE1 (MCM1, *Saccharomyces cerevisiae*), AGAMOUS (AG, *Arabidopsis thaliana*), DEFICIENS (DEF, *Antirrhinum majus*), and SERUM RESPONSE FACTOR (SRF, *Homo sapiens*)) transcription factors were found to be involved in anthocyanin regulation. However, for most of them, this appears to be independent of the MBW complex or was not investigated (Lalusin et al., 2006; Li et al., 2021; Zhang et al., 2025b; Ma et al., 2019; Qi et al., 2022; Lu et al., 2018; Qi et al., 2020). Only a handful of MADS-box TFs were found or are implicated to regulate the expression of PAP1 homologs (Jaakola et al., 2010; Zhang et al., 2025b; Wu et al., 2014), and only AcMADS68 from kiwifruit was shown to interact

with PAP1 homologs directly and increase the transactivation of anthocyanin genes of the MBW complex (Liu et al., 2023b). Hence, there is the possibility that MADS-box TFs are involved in betalain regulation, but probably not by directly interacting with AhMBY2.1.

Moreover, JASMONATE ZIM-DOMAIN (JAZ) TFs are known to interact with PAP1 homologs, but only have a repressive function and are thus not candidates for positive AhMYB2 co-regulators (Qi et al., 2011).

Table 4.1: Selected top candidates for AhMYB2 co-transcription factors for the regulation of the betalain pathway in grain amaranth. Selection was done based on literature, previous results, and RNA-seq data; see methods 4.2.8.

candidate	TF class	<i>A. thaliana</i> homologue(s)	Why selected
AHq019220 and AHq019221	ERF/AP2	AT4G17490, ATG5G47230 and AT4G17500, AT5G47220	Significantly downregulated in MYB2.1_non_ind. and AHq019221 also in MYB2.2_non_ind.. Closely related to Pp4ERF24 and Pp12ERF96 which were shown to interact with Pp-MYB114 and increase its upregulation of an anthocyanin pathway gene in pear (Ni et al., 2019).
AHq004055	bZIP	AT1G75390	Upregulated in MYB2.1_non_ind. Closest homolog of MdbZIP44, which is shown to interact with MdMYB1 in apple, influencing the binding to anthocyanin structural genes (An et al., 2018).
AHq000664	WRKY	AT1G80840	Upregulated in all except MYB2.2_non_ind. Closest homolog of MdWRKY40, which is shown to interact with MdMYB1 in apple, influencing the binding to anthocyanin structural genes (An et al., 2019).
AHq022750	NAC	AT1G61110	Upregulated only MYB2.1_non_ind. Closest homolog of MdNAC42, which is shown to interact with MdMYB1 in apple, influencing the binding to anthocyanin structural genes (Zhang et al., 2020).

This literature review does not cover all possible PAP1 homolog-interacting TFs found across all plant species, as there are more TFs from smaller TF families that were found in only one or a few species (e.g., He et al. (2021)). Nevertheless, all major TF classes reported to be involved in anthocyanin or betalain regulation were investigated to identify the most promising candidates for AhMYB2 co-regulators. As the next step, potential

co-regulators need to be tested for interaction (e.g., via Y2H) with AhMYB2 and if they increase the transactivation of AhCYP76AD2 if co-transformed with AhMYB2 (e.g., in transactivation assays).

The biggest limitation of the reported approach is that all candidates have been selected due to their reported involvement in anthocyanin regulation, assuming that if betalain-regulating PAP1 homologs have been co-opted from the anthocyanin pathway, co-regulators might have been co-opted as well. This approach risks missing TFs that had been recruited from other pathways or evolved *de novo*. Future studies should include additional unbiased methods like co-immunoprecipitation to identify potential betalain-regulating TFs that are not homologous to anthocyanin regulators.

We previously identified AhMYB2 as a positive transcriptional regulator of the betalain pathway in amaranth. However, the molecular mode of how AhMYB2 activates the betalain pathway remained elusive. Using transgenic *A. hypochondriacus* callus we could demonstrate that the betalain pathway seems to be mainly regulated by controlling the transcription of the first enzyme, *AhCYP76AD2*. RNA-seq validated that AhMYB2 up-regulates the expression of *AhCYP76AD2* and *AhDODAA1*. For efficient transactivation of *AhCYP76AD2*, AhMYB2 seems to require an additional betalain-specific co-regulator. Different from the anthocyanin regulation, which requires the formation of MBW complexes, AhMYB2 does not interact with class IIIb bHLH TFs or the homolog of AtTTG1 in Y2H. This suggests that, albeit AhMYB2 has been co-opted from the anthocyanin regulation, the molecular mode of action and the target genes are different to its homolog AtPAP1 and had been adapted for the betalain regulation.

This study focused on the AhMYB2-mediated transcriptional regulation of the betalain pathway in grain amaranth. However, for the anthocyanin pathway, it was shown that also posttranscriptional regulation (e.g., via miRNAs) and posttranslational regulation (e.g., phosphorylations, ubiquitinations) play an important role (reviewed by Khan and Abbas (2023); Zhang et al. (2025c)). For instance, does the COP1/SPA complex regulate the stability of AtPAP1 and AtPAP2 in a light-dependent manner (Maier et al., 2013). Li et al. (2016) demonstrated that the phosphorylation of AtPAP1 by AtMPK4 increases its stability and is required for the full function of AtPAP1 in upregulating anthocyanin pathway genes. Posttranscriptional and posttranslational regulation were beyond the scope of this study but should be considered when further unraveling the betalain regulation. Future studies could investigate the protein stability of AhMYB2 in various conditions, but the absence of the typical VP motif in AhMYB2 suggests rather not a regulation by COP1/SPA.

Our results suggest that co-option of PAP1 for the betalain pathway changed the mode of action from involving the MBW complex to a yet unknown betalain-specific TF. It could be speculated that this enabled specific targeting of the betalain instead of the

anthocyanin pathway. Moreover, independence from the MBW complex probably contributed to reducing pleiotropy, leaving other MBW complex-regulated pathways, including protoanthocyanin-, mucilage-, trichome-, and root hair-regulation, unaffected. Generally, our study demonstrated that the betalain regulation seems to be more complex than previously thought (Hatlestad et al., 2015). Understanding the interplay of co-opted, adapted, and potential new evolved regulations in the betalain regulator network might improve our understanding of the evolutionary flexibility of gene regulatory networks.

4.5 Acknowledgments

We thank Andrii Zinchenko (University of Cologne) for providing the AtEG3, AtGL3, and AtMYC1 Y2H plasmids and for sharing the Y2H protocol, Julio Martinez (previously, University of Cologne) for providing the AhTT8, AhTTG1a, AtTT8, and AtTTG1 Y2H plasmids. We thank Martin Hülskamp for sharing the *Saccharomyces cerevisiae* AH109 strain and BP311 with the AtPAP1 CDS, Gabriele Buchmann for CPGE_VEC00256, Jonas Schön for providing pJS078 as stuffer for transactivation assays, Estefania Pavesi for pep602 and Uriel Urquiza-Garcia for the backbones and modules for PolyGEN. We thank Daniel Wasim Djamriani for his help with developing the callus transformation protocol. We thank Raphael Schönlé for technical assistance with the transactivation assay. We thank Thea Rütjes for advice regarding the RNA extraction. We acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2048/1 – Project ID 390686111.

4.6 Author contributions

SKV, MZ, GH, and MGS planned and designed the study. SKV performed callus transformation and phenotyping experiments, cloned the respective plasmids, and analyzed the data. LMT performed the Y2H assays and cloned missing plasmids. LMT and SKV analyzed the Y2H data. SKV performed the protoplast isolation, transformation, mammalian transactivation assay, luciferase measurements, cloned the respective plasmids, and analyzed the data. JS assisted in performing the protoplast isolation, transformation, mammalian transactivation assay, and luciferase measurements. UU provided the PolyGEN system and assisted in designing the PolyGEN constructs. SKV isolated the RNA and analyzed the differential expression data. TSW did the differential expression analysis and wrote the respective method section. SKV prepared the figures. SKV wrote the manuscript.

4.7 Competing interests

None declared.

4.8 References

- E. G. Allwood, D. R. Davies, C. Gerrish, B. E. Ellis, and G. Bolwell. Phosphorylation of phenylalanine ammonia-lyase: evidence for a novel protein kinase and identification of the phosphorylated residue. *FEBS Letters*, 457(1):47–52, Aug. 1999. ISSN 0014-5793, 1873-3468. doi: 10.1016/S0014-5793(99)00998-9. URL <https://febs.onlinelibrary.wiley.com/doi/10.1016/S0014-5793%2899%2900998-9>.
- A. Amato, E. Cavallini, A. R. Walker, M. Pezzotti, M. Bliet, F. Quattrocchio, R. Koes, B. Ruperti, E. Bertini, S. Zenoni, and G. B. Tornielli. 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*, 99(6):1220–1241, Sept. 2019. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.14419. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.14419>.
- J. An, J. Yao, R. Xu, C. You, X. Wang, and Y. Hao. Apple bZIP transcription factor MdbZIP44 regulates abscisic acid-promoted anthocyanin accumulation. *Plant, Cell & Environment*, 41(11):2678–2692, Nov. 2018. ISSN 0140-7791, 1365-3040. doi: 10.1111/pce.13393. URL <https://onlinelibrary.wiley.com/doi/10.1111/pce.13393>.
- J. An, X. Zhang, C. You, S. Bi, X. Wang, and Y. Hao. MdWRKY40 promotes wounding-induced anthocyanin biosynthesis in association with MdMYB1 and undergoes MdBT2-mediated degradation. *New Phytologist*, 224(1):380–395, Oct. 2019. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.16008. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.16008>.
- J. An, X. Zhang, S. Bi, C. You, X. Wang, and Y. Hao. The ERF transcription factor MdERF38 promotes drought stress-induced anthocyanin biosynthesis in apple. *The Plant Journal*, 101(3):573–589, Feb. 2020a. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.14555. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.14555>.
- J.-P. An, Y.-J. Liu, X.-W. Zhang, S.-Q. Bi, X.-F. Wang, C.-X. You, and Y.-J. Hao. Dynamic regulation of anthocyanin biosynthesis at different light intensities by the BT2-TCP46-MYB1 module in apple. *Journal of Experimental Botany*, 71(10):3094–3109, May 2020b. ISSN 0022-0957, 1460-2431. doi: 10.1093/jxb/eraa056. URL <https://academic.oup.com/jxb/article/71/10/3094/5717769>.
- J. Andres and M. D. Zurbriggen. Genetically Encoded Biosensors for the Quantitative Analysis of Auxin Dynamics in Plant Cells. In *Methods in Molecular Biology*, pages 183–195. Springer US, New York, NY, 2022. ISBN 978-1-0716-1790-8 978-1-0716-1791-5. doi: 10.1007/978-1-0716-1791-5_11. URL https://link.springer.com/10.1007/978-1-0716-1791-5_11. ISSN: 1064-3745, 1940-6029.
- S. Andrews. FastQC: A Quality Control Tool for High Throughput Sequence Data, 2010. URL <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

- G. E. Araguirang and A. S. Richter. Activation of anthocyanin biosynthesis in high light – what is the initial signal? *New Phytologist*, 236(6):2037–2043, Dec. 2022. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.18488. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.18488>.
- N. Bhagavan and C.-E. Ha. Enzymes and Enzyme Regulation. In *Essentials of Medical Biochemistry*, pages 63–84. Elsevier, 2015. ISBN 978-0-12-416687-5. doi: 10.1016/B978-0-12-416687-5.00006-3. URL <https://linkinghub.elsevier.com/retrieve/pii/B9780124166875000063>.
- Y. Bi, Q. Zhang, Y. He, Y. Ding, H. Tian, F. Zeng, K. Lyu, H. Li, S. Lyu, and Y. Fan. Creation and resistance evaluation of a new soybean germplasm rich in betalain. *Frontiers in Plant Science*, 16:1743684, Jan. 2026. ISSN 1664-462X. doi: 10.3389/fpls.2025.1743684. URL <https://www.frontiersin.org/articles/10.3389/fpls.2025.1743684/full>.
- A. Blanco and G. Blanco. Enzymes. In *Medical Biochemistry*, pages 165–190. Elsevier, 2022. ISBN 978-0-323-91599-1. doi: 10.1016/B978-0-323-91599-1.00029-8. URL <https://linkinghub.elsevier.com/retrieve/pii/B9780323915991000298>.
- U. Bodenhofer, E. Bonatesta, C. Horejs-Kainrath, and S. Hochreiter. msa: an R package for multiple sequence alignment. *Bioinformatics*, 31(24):3997–3999, 2015. doi: 10.1093/bioinformatics/btv494.
- A. M. Bolger, M. Lohse, and B. Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120, Aug. 2014. ISSN 1367-4811, 1367-4803. doi: 10.1093/bioinformatics/btu170. URL <https://academic.oup.com/bioinformatics/article/30/15/2114/2390096>.
- S. F. Brockington, R. H. Walker, B. J. Glover, P. S. Soltis, and D. E. Soltis. Complex pigment evolution in the *Caryophyllales*. *New Phytologist*, 190(4):854–864, June 2011. ISSN 0028-646X, 1469-8137. doi: 10.1111/j.1469-8137.2011.03687.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.2011.03687.x>.
- J. J. Campanella, J. V. Smalley, and M. E. Dempsey. A phylogenetic examination of the primary anthocyanin production pathway of the *Plantae*. *Botanical Studies*, 55(1):10, Dec. 2014. ISSN 1999-3110. doi: 10.1186/1999-3110-55-10. URL <https://as-botanicalstudies.springeropen.com/articles/10.1186/1999-3110-55-10>.
- F. Cappellini, A. Marinelli, M. Toccaceli, C. Tonelli, and K. Petroni. Anthocyanins: From Mechanisms of Regulation in Plants to Health Benefits in Foods. *Frontiers in Plant Science*, 12:748049, Oct. 2021. ISSN 1664-462X. doi: 10.3389/fpls.2021.748049. URL <https://www.frontiersin.org/articles/10.3389/fpls.2021.748049/full>.
- A. P. Castellanos-Arévalo, A. A. Estrada-Luna, J. L. Cabrera-Ponce, E. Valencia-Lozano, H. Herrera-Ubaldo, S. de Folter, A. Blanco-Labra, and J. P. Délano-Frier. *Agrobacterium rhizogenes*-mediated transformation of grain (*Amaranthus hypochondriacus*) and leafy (*A. hybridus*) amaranths. *Plant Cell Reports*, 39(9):1143–1160, Sept. 2020. ISSN 0721-7714, 1432-203X. doi: 10.1007/s00299-020-02553-9. URL <https://link.springer.com/10.1007/s00299-020-02553-9>.
- D. Charif and J. R. Lobry. SeqinR 1.0-2: a contributed package to the R project for

- statistical computing devoted to biological sequences retrieval and analysis. In U. Bastolla, M. Porto, H. E. Roman, and M. Vendruscolo, editors, *Structural approaches to sequence evolution: Molecules, networks, populations*, Biological and Medical Physics, Biomedical Engineering, pages 207–232. Springer Verlag, New York, 2007.
- M.-n. Cheng, Z.-j. Huang, Q.-z. Hua, W. Shan, J.-f. Kuang, W.-j. Lu, Y.-h. Qin, and J.-y. Chen. The WRKY transcription factor HpWRKY44 regulates *CytP450-like1* expression in red pitaya fruit (*Hylocereus polyrhizus*). *Horticulture Research*, 4(1): 17039, Dec. 2017. ISSN 2052-7276. doi: 10.1038/hortres.2017.39. URL <https://academic.oup.com/hr/article/6447950>.
- S. Chhon, J. Jeon, J. Kim, and S. U. Park. Accumulation of Anthocyanins through Overexpression of *AtPAP1* in *Solanum nigrum* Lin. (Black Nightshade). *Biomolecules*, 10(2):277, Feb. 2020. ISSN 2218-273X. doi: 10.3390/biom10020277. URL <https://www.mdpi.com/2218-273X/10/2/277>.
- C.-N. Chow, C.-W. Yang, N.-Y. Wu, H.-T. Wang, K.-C. Tseng, Y.-H. Chiu, T.-Y. Lee, and W.-C. Chang. PlantPAN 4.0: updated database for identifying conserved non-coding sequences and exploring dynamic transcriptional regulation in plant promoters. *Nucleic Acids Research*, 52(D1):D1569–D1578, Jan. 2024. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkad945. URL <https://academic.oup.com/nar/article/52/D1/D1569/7332064>.
- J. W. Clouse, D. Adhikary, J. T. Page, T. Ramaraj, M. K. Deyholos, J. A. Udall, D. J. Fairbanks, E. N. Jellen, and P. J. Maughan. The Amaranth Genome: Genome, Transcriptome, and Physical Map Assembly. *The Plant Genome*, 9(1): plantgenome2015.07.0062, Mar. 2016. ISSN 1940-3372, 1940-3372. doi: 10.3835/plantgenome2015.07.0062. URL <https://access.onlinelibrary.wiley.com/doi/10.3835/plantgenome2015.07.0062>.
- P. Cubas, N. Lauter, J. Doebley, and E. Coen. The TCP domain: a motif found in proteins regulating plant growth and development. *The Plant Journal*, 18(2):215–222, Apr. 1999. ISSN 0960-7412, 1365-313X. doi: 10.1046/j.1365-313X.1999.00444.x. URL <https://onlinelibrary.wiley.com/doi/10.1046/j.1365-313X.1999.00444.x>.
- L. Deluc, J. Bogs, A. R. Walker, T. Ferrier, A. Decendit, J.-M. Merillon, S. P. Robinson, and F. Barrieu. The Transcription Factor VvMYB5b Contributes to the Regulation of Anthocyanin and Proanthocyanidin Biosynthesis in Developing Grape Berries. *Plant Physiology*, 147(4):2041–2053, Aug. 2008. ISSN 1532-2548. doi: 10.1104/pp.108.118919. URL <https://academic.oup.com/plphys/article/147/4/2041/6107699>.
- B. N. Devaiah, A. S. Karthikeyan, and K. G. Raghothama. WRKY75 Transcription Factor Is a Modulator of Phosphate Acquisition and Root Development in *Arabidopsis*. *Plant Physiology*, 143(4):1789–1801, Apr. 2007. ISSN 1532-2548. doi: 10.1104/pp.106.093971. URL <https://academic.oup.com/plphys/article/143/4/1789/6106831>.
- A. Dobin and T. R. Gingeras. Mapping RNA-seq reads with STAR. *Current Protocols in Bioinformatics*, 51(1):11–14, 2015. Publisher: Wiley Online Library.
- W. Dröge-Laser, B. L. Snoek, B. Snel, and C. Weiste. The *Arabidopsis* bZIP transcription factor family — an update. *Current Opinion in Plant Biology*, 45:36–49, Oct.

2018. ISSN 13695266. doi: 10.1016/j.pbi.2018.05.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S1369526617302157>.
- S. Duan, J. Wang, C. Gao, C. Jin, D. Li, D. Peng, G. Du, Y. Li, and M. Chen. Functional characterization of a heterologously expressed *Brassica napus* *WRKY41-1* transcription factor in regulating anthocyanin biosynthesis in *Arabidopsis thaliana*. *Plant Science*, 268:47–53, Mar. 2018. ISSN 01689452. doi: 10.1016/j.plantsci.2017.12.010. URL <https://linkinghub.elsevier.com/retrieve/pii/S0168945217309342>.
- D. Ebbert. *chisq.posthoc.test: A Post Hoc Analysis for Pearson's Chi-Squared Test for Count Data*. 2019. URL <https://CRAN.R-project.org/package=chisq.posthoc.test>.
- A. Ehlert, F. Weltmeier, X. Wang, C. S. Mayer, S. Smeekeens, J. Vicente-Carbajosa, and W. Dröge-Laser. Two-hybrid protein–protein interaction analysis in *Arabidopsis* protoplasts: establishment of a heterodimerization map of group C and group S bZIP transcription factors. *The Plant Journal*, 46(5):890–900, June 2006. ISSN 0960-7412, 1365-313X. doi: 10.1111/j.1365-313X.2006.02731.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2006.02731.x>.
- C. Engler, R. Kandzia, and S. Marillonnet. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS ONE*, 3(11):e3647, Nov. 2008. ISSN 1932-6203. doi: 10.1371/journal.pone.0003647. URL <https://dx.plos.org/10.1371/journal.pone.0003647>.
- S. Fields and O.-k. Song. A novel genetic system to detect protein–protein interactions. *Nature*, 340(6230):245–246, July 1989. ISSN 0028-0836, 1476-4687. doi: 10.1038/340245a0. URL <https://www.nature.com/articles/340245a0>.
- N. Gehlenborg. *UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets*. 2019. URL <https://CRAN.R-project.org/package=UpSetR>.
- M. Geng, Y. Shao, M. Zhang, X. Zheng, B. Tan, W. Wang, L. Zhang, X. Ye, M. Li, J. Li, J. Cheng, and J. Feng. Overexpression of peach *NAC25* promotes anthocyanin biosynthesis in poplar shoots. *Fruit Research*, 2(1):1–9, 2022. ISSN 2769-4615. doi: 10.48130/FruRes-2022-0021. URL <https://www.maxapress.com/article/doi/10.48130/FruRes-2022-0021>.
- R. D. Gietz and R. H. Schiestl. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, 2(1):31–34, Jan. 2007. ISSN 1754-2189, 1750-2799. doi: 10.1038/nprot.2007.13. URL <https://www.nature.com/articles/nprot.2007.13>. Publisher: Springer Science and Business Media LLC.
- A. Gonzalez, M. Zhao, J. M. Leavitt, and A. M. Lloyd. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *The Plant Journal*, 53(5):814–827, Mar. 2008. ISSN 0960-7412, 1365-313X. doi: 10.1111/j.1365-313X.2007.03373.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2007.03373.x>.
- A. Gonzalez, M. Brown, G. Hatlestad, N. Akhavan, T. Smith, A. Hembd, J. Moore, D. Montes, T. Mosley, J. Resendez, H. Nguyen, L. Wilson, A. Campbell, D. Sudar-

- shan, and A. Lloyd. TTG2 controls the developmental regulation of seed coat tannins in *Arabidopsis* by regulating vacuolar transport steps in the proanthocyanidin pathway. *Developmental Biology*, 419(1):54–63, Nov. 2016. ISSN 00121606. doi: 10.1016/j.ydbio.2016.03.031. URL <https://linkinghub.elsevier.com/retrieve/pii/S0012160616301737>.
- C. Graf, T. S. Winkler, P. J. Maughan, and M. G. Stetter. Domestication shaped the chromatin landscape of grain amaranth. *Nature Communications*, 16(1):10407, Nov. 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-66445-w. URL <https://www.nature.com/articles/s41467-025-66445-w>.
- E. Grotewold, B. J. Drummond, B. Bowen, and T. Peterson. The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. *Cell*, 76(3):543–553, Feb. 1994. ISSN 00928674. doi: 10.1016/0092-8674(94)90117-1. URL <https://linkinghub.elsevier.com/retrieve/pii/0092867494901171>.
- Y. Guo, D. Li, T. Liu, Y. Li, J. Liu, M. He, X. Cui, Z. Liu, and M. Chen. Genome-Wide Identification of PAP1 Direct Targets in Regulating Seed Anthocyanin Biosynthesis in *Arabidopsis*. *International Journal of Molecular Sciences*, 24(22):16049, Nov. 2023. ISSN 1422-0067. doi: 10.3390/ijms242216049. URL <https://www.mdpi.com/1422-0067/24/22/16049>.
- G. J. Hatlestad, N. A. Akhavan, R. M. Sunnadeniya, L. Elam, S. Cargile, A. Hembd, A. Gonzalez, J. M. McGrath, and A. M. Lloyd. The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway. *Nature Genetics*, 47(1):92–96, Jan. 2015. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3163. URL <http://www.nature.com/articles/ng.3163>.
- Y. He, T. Zhang, H. Sun, H. Zhan, and Y. Zhao. A reporter for noninvasively monitoring gene expression and plant transformation. *Horticulture Research*, 7(1):152, Dec. 2020. ISSN 2662-6810, 2052-7276. doi: 10.1038/s41438-020-00390-1. URL <https://www.nature.com/articles/s41438-020-00390-1>.
- Y. He, X. Zhang, L. Li, Z. Sun, J. Li, X. Chen, and G. Hong. SPX4 interacts with both PHR1 and PAP1 to regulate critical steps in phosphorus-status-dependent anthocyanin biosynthesis. *New Phytologist*, 230(1):205–217, Apr. 2021. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.17139. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.17139>.
- M. A. Heim. The Basic Helix-Loop-Helix Transcription Factor Family in Plants: A Genome-Wide Study of Protein Structure and Functional Diversity. *Molecular Biology and Evolution*, 20(5):735–747, May 2003. ISSN 07374038, 15371719. doi: 10.1093/molbev/msg088. URL <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msg088>.
- R. P. Hellens, E. A. Edwards, N. R. Leyland, S. Bean, and P. M. Mullineaux. pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Molecular Biology*, 42(6):819–832, Apr. 2000. ISSN 0167-4412, 1573-5028. doi: 10.1023/A:1006496308160. URL <https://link.springer.com/10.1023/A:1006496308160>.

- I. Herskowitz. Functional inactivation of genes by dominant negative mutations. *Nature*, 329(6136):219–222, Sept. 1987. ISSN 0028-0836, 1476-4687. doi: 10.1038/329219a0. URL <https://www.nature.com/articles/329219a0>.
- K. Higo, Y. Ugawa, M. Iwamoto, and T. Korenaga. Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research*, 27(1):297–300, Jan. 1999. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/27.1.297. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/27.1.297>.
- I. V. O. Hoffie. Entwicklung des modularen CasCADE-Vektorsystems und dessen Verwendung zur gezielten Mutagenese der Stp13-Orthologe von Weizen und Gerste für die Etablierung dauerhafter Resistenz gegen Rost-und Mehltaupilze. 2022. Publisher: Hannover: Institutionelles Repositorium der Leibniz Universität Hannover.
- M. Holsters, B. Silva, F. Van Vliet, C. Genetello, M. De Block, P. Dhaese, A. Depicker, D. Inzé, G. Engler, R. Villarroel, M. Van Montagu, and J. Schell. The functional organization of the nopaline *A. tumefaciens* plasmid pTiC58. *Plasmid*, 3(2):212–230, Mar. 1980. ISSN 0147619X. doi: 10.1016/0147-619X(80)90110-9. URL <https://linkinghub.elsevier.com/retrieve/pii/0147619X80901109>.
- T. Hothorn, F. Bretz, and P. Westfall. Simultaneous Inference in General Parametric Models. *Biometrical Journal*, 50(3):346–363, 2008.
- L. Jaakola, M. Poole, M. O. Jones, T. Kämäräinen-Karppinen, J. J. Koskimäki, A. Hottola, H. Häggman, P. D. Fraser, K. Manning, G. J. King, H. Thomson, and G. B. Seymour. A SQUAMOSA MADS Box Gene Involved in the Regulation of Anthocyanin Accumulation in Bilberry Fruits. *Plant Physiology*, 153(4):1619–1629, Aug. 2010. ISSN 1532-2548. doi: 10.1104/pp.110.158279. URL <https://academic.oup.com/plphys/article/153/4/1619/6111251>.
- G. Jain and K. S. Gould. Are betalain pigments the functional homologues of anthocyanins in plants? *Environmental and Experimental Botany*, 119:48–53, Nov. 2015. ISSN 00988472. doi: 10.1016/j.envexpbot.2015.06.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S0098847215001112>.
- M. Jakoby, B. Weisshaar, W. Dröge-Laser, J. Vicente-Carbajosa, J. Tiedemann, T. Kroj, and F. Parey. bZIP transcription factors in *Arabidopsis*. *Trends in Plant Science*, 7(3):106–111, Mar. 2002. ISSN 13601385. doi: 10.1016/S1360-1385(01)02223-3. URL <https://linkinghub.elsevier.com/retrieve/pii/S1360138501022233>.
- G. Jiang, Z. Li, Y. Song, H. Zhu, S. Lin, R. Huang, Y. Jiang, and X. Duan. LcNAC13 Physically Interacts with LcR1MYB1 to Coregulate Anthocyanin Biosynthesis-Related Genes during Litchi Fruit Ripening. *Biomolecules*, 9(4):135, Apr. 2019. ISSN 2218-273X. doi: 10.3390/biom9040135. URL <https://www.mdpi.com/2218-273X/9/4/135>.
- S. Jiang, Q. Sun, T. Zhang, W. Liu, N. Wang, and X. Chen. MdMYB114 regulates anthocyanin biosynthesis and functions downstream of MdbZIP4-like in apple fruit. *Journal of Plant Physiology*, 257:153353, Feb. 2021. ISSN 01761617. doi: 10.1016/j.jplph.2020.153353. URL <https://linkinghub.elsevier.com/retrieve/pii/S0176161720302431>.
- R. A. Khan and N. Abbas. Role of epigenetic and post-translational modifications in an-

- thocyanin biosynthesis: A review. *Gene*, 887:147694, Dec. 2023. ISSN 03781119. doi: 10.1016/j.gene.2023.147694. URL <https://linkinghub.elsevier.com/retrieve/pii/S0378111923005358>.
- R. Kolde. *pheatmap: Pretty Heatmaps*. 2025. URL <https://CRAN.R-project.org/package=pheatmap>.
- T. Koyama and F. Sato. The function of *ETHYLENE RESPONSE FACTOR* genes in the light-induced anthocyanin production of *Arabidopsis thaliana* leaves. *Plant Biotechnology*, 35(1):87–91, 2018. ISSN 1342-4580, 1347-6114. doi: 10.5511/plantbiotechnology.18.0122b. URL https://www.jstage.jst.go.jp/article/plantbiotechnology/35/1/35_18.0122b/_article.
- A. G. Lalusin, K. Nishita, S.-H. Kim, M. Ohta, and T. Fujimura. A new MADS-box gene (IbMADS10) from sweet potato (*Ipomoea batatas* (L.) Lam) is involved in the accumulation of anthocyanin. *Molecular Genetics and Genomics*, 275(1):44–54, Jan. 2006. ISSN 1617-4615, 1617-4623. doi: 10.1007/s00438-005-0080-x. URL <http://link.springer.com/10.1007/s00438-005-0080-x>.
- A. Lampropoulos, Z. Sutikovic, C. Wenzl, I. Maegele, J. U. Lohmann, and J. Forner. GreenGate - A Novel, Versatile, and Efficient Cloning System for Plant Transgenesis. *PLoS ONE*, 8(12):e83043, Dec. 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0083043. URL <https://dx.plos.org/10.1371/journal.pone.0083043>.
- S.-H. Lee, S.-H. Kim, T.-K. Park, Y.-P. Kim, J.-W. Lee, and T.-W. Kim. Transcription factors BZR1 and PAP1 cooperate to promote anthocyanin biosynthesis in *Arabidopsis* shoots. *The Plant Cell*, 36(9):3654–3673, Sept. 2024. ISSN 1040-4651, 1532-298X. doi: 10.1093/plcell/koae172. URL <https://academic.oup.com/plcell/article/36/9/3654/7692851>.
- M. Lescot. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*, 30(1):325–327, Jan. 2002. ISSN 13624962. doi: 10.1093/nar/30.1.325. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/30.1.325>.
- D. Li, H. Ma, X. Chen, Z. Wu, and S. Feng. MdLBD37 and MdBZR1 inhibit anthocyanin biosynthesis in apple under high temperature. *Plant Science*, 359:112621, Oct. 2025. ISSN 01689452. doi: 10.1016/j.plantsci.2025.112621. URL <https://linkinghub.elsevier.com/retrieve/pii/S0168945225002390>.
- J. Li, N. Ma, Y. An, and L. Wang. FcMADS9 of fig regulates anthocyanin biosynthesis. *Scientia Horticulturae*, 278:109820, Feb. 2021. ISSN 03044238. doi: 10.1016/j.scienta.2020.109820. URL <https://linkinghub.elsevier.com/retrieve/pii/S0304423820306488>.
- S. Li and S. Zachgo. TCP3 interacts with R2R3-MYB proteins, promotes flavonoid biosynthesis and negatively regulates the auxin response in *Arabidopsis thaliana*. *The Plant Journal*, 76(6):901–913, Dec. 2013. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.12348. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.12348>.
- S. Li, W. Wang, J. Gao, K. Yin, R. Wang, C. Wang, M. Petersen, J. Mundy, and J.-L. Qiu. MYB75 Phosphorylation by MPK4 Is Required for Light-Induced Anthocyanin

- Accumulation in *Arabidopsis*. *The Plant Cell*, 28(11):2866–2883, Nov. 2016. ISSN 1040-4651, 1532-298X. doi: 10.1105/tpc.16.00130. URL <https://academic.oup.com/plcell/article/28/11/2866-2883/6098261>.
- X. Li, C. Martín-Pizarro, L. Zhou, B. Hou, Y. Wang, Y. Shen, B. Li, D. Posé, and G. Qin. Deciphering the regulatory network of the NAC transcription factor FvRIF, a key regulator of strawberry (*Fragaria vesca*) fruit ripening. *The Plant Cell*, 35(11):4020–4045, Oct. 2023. ISSN 1040-4651, 1532-298X. doi: 10.1093/plcell/koad210. URL <https://academic.oup.com/plcell/article/35/11/4020/7233020>.
- Y. Liao, G. K. Smyth, and W. Shi. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930, Apr. 2014. ISSN 1367-4811, 1367-4803. doi: 10.1093/bioinformatics/btt656. URL <https://academic.oup.com/bioinformatics/article/30/7/923/232889>.
- M. Libault, J. Wan, T. Czechowski, M. Udvardi, and G. Stacey. Identification of 118 *Arabidopsis* Transcription Factor and 30 Ubiquitin-Ligase Genes Responding to Chitin, a Plant-Defense Elicitor. *Molecular Plant-Microbe Interactions*®, 20(8):900–911, Aug. 2007. ISSN 0894-0282, 1943-7706. doi: 10.1094/MPMI-20-8-0900. URL <https://apsjournals.apsnet.org/doi/10.1094/MPMI-20-8-0900>.
- W. Liu, Z. Mei, L. Yu, T. Gu, Z. Li, Q. Zou, S. Zhang, H. Fang, Y. Wang, Z. Zhang, X. Chen, and N. Wang. The ABA-induced NAC transcription factor MdNAC1 interacts with a bZIP-type transcription factor to promote anthocyanin synthesis in red-fleshed apples. *Horticulture Research*, 10(5):uhad049, May 2023a. ISSN 2052-7276. doi: 10.1093/hr/uhad049. URL <https://academic.oup.com/hr/article/doi/10.1093/hr/uhad049/7077840>.
- Y. Liu, G. Lv, Y. Yang, K. Ma, X. Ren, M. Li, and Z. Liu. Interaction of AcMADS68 with transcription factors regulates anthocyanin biosynthesis in red-fleshed kiwifruit. *Horticulture Research*, 10(2):uhac252, Feb. 2023b. ISSN 2052-7276. doi: 10.1093/hr/uhac252. URL <https://academic.oup.com/hr/article/doi/10.1093/hr/uhac252/6827844>.
- A. Lloyd, A. Brockman, L. Aguirre, A. Campbell, A. Bean, A. Cantero, and A. Gonzalez. 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(9):1431–1441, Sept. 2017. ISSN 0032-0781, 1471-9053. doi: 10.1093/pcp/pcx075. URL <https://academic.oup.com/pcp/article/58/9/1431/3858365>.
- S. Lopez-Nieves, Y. Yang, A. Timoneda, M. Wang, T. Feng, S. A. Smith, S. F. Brockington, and H. A. Maeda. Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in *Caryophyllales*. *New Phytologist*, 217(2):896–908, Jan. 2018. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.14822. URL <https://onlinelibrary.wiley.com/doi/10.1111/nph.14822>.
- M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, Dec. 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0550-8. URL <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8>.

- W. Lu, J. Chen, X. Ren, J. Yuan, X. Han, L. Mao, T. Ying, and Z. Luo. One novel strawberry MADS-box transcription factor FaMADS1a acts as a negative regulator in fruit ripening. *Scientia Horticulturae*, 227:124–131, Jan. 2018. ISSN 03044238. doi: 10.1016/j.scienta.2017.09.042. URL <https://linkinghub.elsevier.com/retrieve/pii/S0304423817305940>.
- G. Ma, Q. Zou, X. Shi, D. Tian, and Q. Sheng. Ectopic expression of the AaFUL1 gene identified in *Anthurium andraeanum* affected floral organ development and seed fertility in tobacco. *Gene*, 696:197–205, May 2019. ISSN 03781119. doi: 10.1016/j.gene.2019.02.050. URL <https://linkinghub.elsevier.com/retrieve/pii/S0378111919301829>.
- K. Mahmood, Z. Xu, A. El-Kereamy, J. A. Casaretto, and S. J. Rothstein. The Arabidopsis Transcription Factor ANAC032 Represses Anthocyanin Biosynthesis in Response to High Sucrose and Oxidative and Abiotic Stresses. *Frontiers in Plant Science*, 7, Oct. 2016. ISSN 1664-462X. doi: 10.3389/fpls.2016.01548. URL <http://journal.frontiersin.org/article/10.3389/fpls.2016.01548/full>.
- A. Maier, A. Schrader, L. Kokkelink, C. Falke, B. Welter, E. Iniesto, V. Rubio, J. F. Uhrig, M. Hülskamp, and U. Hoecker. Light and the E3 ubiquitin ligase COP1/SPA control the protein stability of the MYB transcription factors PAP1 and PAP2 involved in anthocyanin accumulation in *Arabidopsis*. *The Plant Journal*, 74(4):638–651, May 2013. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.12153. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.12153>.
- G. Malacarne, E. Collier, S. Czernmel, U. Vrhovsek, K. Engelen, V. Goremykin, J. Bogs, and C. Moser. The grapevine VvibZIPC22 transcription factor is involved in the regulation of flavonoid biosynthesis. *Journal of Experimental Botany*, 67(11):3509–3522, May 2016. ISSN 0022-0957, 1460-2431. doi: 10.1093/jxb/erw181. URL <https://academic.oup.com/jxb/article-lookup/doi/10.1093/jxb/erw181>.
- S. Mariyam, V. Kumar, A. Roychoudhury, G. S. Ghodake, S. Muneer, J. S. Duhan, F. Ahmad, R. K. Sharma, J. Singh, and C. S. Seth. Functional Diversification and Mechanistic Insights of MYB Transcription Factors in Mediating Plant Growth and Development, Secondary Metabolism, and Stress Responses. *Journal of Plant Growth Regulation*, 44(4):1465–1484, Apr. 2025. ISSN 0721-7595, 1435-8107. doi: 10.1007/s00344-025-11622-w. URL <https://link.springer.com/10.1007/s00344-025-11622-w>.
- T. R. Martins, J. J. Berg, S. Blinka, M. D. Rausher, and D. A. Baum. Precise spatio-temporal regulation of the anthocyanin biosynthetic pathway leads to petal spot formation in *Clarkia gracilis* (Onagraceae). *New Phytologist*, 197(3):958–969, Feb. 2013. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.12062. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.12062>.
- A. J. McCann, J. Lou, M. Moustaqil, M. S. Graus, A. Blum, F. Fontaine, H. Liu, W. Luu, P. Rudolffi-Soto, P. Koopman, E. Sierecki, Y. Gambin, F. Meunier, Z. Liu, E. Hinde, and M. Francois. A dominant-negative SOX18 mutant disrupts multiple regulatory layers essential to transcription factor activity. *Nucleic Acids Research*, 49(19):10931–10955, Nov. 2021. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkab820. URL <https://academic.oup.com/nar/article/49/19/10931/6376034>.

- J. Meng, J. Yin, H. Wang, and H. Li. A TCP Transcription Factor in *Malus halimifolia*, MhTCP4, Positively Regulates Anthocyanins Biosynthesis. *International Journal of Molecular Sciences*, 23(16):9051, Aug. 2022. ISSN 1422-0067. doi: 10.3390/ijms23169051. URL <https://www.mdpi.com/1422-0067/23/16/9051>.
- T. K. Mohanta, D. Yadav, A. Khan, A. Hashem, B. Tabassum, A. L. Khan, E. F. Abd_Allah, and A. Al-Harrasi. Genomics, molecular and evolutionary perspective of NAC transcription factors. *PLOS ONE*, 15(4):e0231425, Apr. 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0231425. URL <https://dx.plos.org/10.1371/journal.pone.0231425>.
- T. Morishita, Y. Kojima, T. Maruta, A. Nishizawa-Yokoi, Y. Yabuta, and S. Shigeoka. *Arabidopsis* NAC Transcription Factor, ANAC078, Regulates Flavonoid Biosynthesis under High-light. *Plant and Cell Physiology*, 50(12):2210–2222, Dec. 2009. ISSN 1471-9053, 0032-0781. doi: 10.1093/pcp/pcp159. URL <https://academic.oup.com/pcp/article-lookup/doi/10.1093/pcp/pcp159>.
- T. Murashige and F. Skoog. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum*, 15(3):473–497, July 1962. ISSN 0031-9317, 1399-3054. doi: 10.1111/j.1399-3054.1962.tb08052.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1399-3054.1962.tb08052.x>.
- J. Ni, S. Bai, Y. Zhao, M. Qian, R. Tao, L. Yin, L. Gao, and Y. Teng. Ethylene response factors Pp4ERF24 and Pp12ERF96 regulate blue light-induced anthocyanin biosynthesis in ‘Red Zaosu’ pear fruits by interacting with MYB114. *Plant Molecular Biology*, 99(1-2):67–78, Jan. 2019. ISSN 0167-4412, 1573-5028. doi: 10.1007/s11103-018-0802-1. URL <http://link.springer.com/10.1007/s11103-018-0802-1>.
- R. Ochoa-Fernandez, N. B. Abel, F.-G. Wieland, J. Schlegel, L.-A. Koch, J. B. Miller, R. Engesser, G. Giuriani, S. M. Brandl, J. Timmer, W. Weber, T. Ott, R. Simon, and M. D. Zurbriggen. Optogenetic control of gene expression in plants in the presence of ambient white light. *Nature Methods*, 17(7):717–725, July 2020. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-020-0868-y. URL <https://www.nature.com/articles/s41592-020-0868-y>.
- K. Ohashi-Ito, K. Iwamoto, and H. Fukuda. LOB DOMAIN-CONTAINING PROTEIN 15 Positively Regulates Expression of VND7, a Master Regulator of Tracheary Elements. *Plant and Cell Physiology*, 59(5):989–996, May 2018. ISSN 0032-0781, 1471-9053. doi: 10.1093/pcp/pcy036. URL <https://academic.oup.com/pcp/article/59/5/989/4850576>.
- S. Ohno, R. Makishima, and M. Doi. Post-transcriptional gene silencing of *CYP76AD* controls betalain biosynthesis in bracts of bougainvillea. *Journal of Experimental Botany*, 72(20):6949–6962, Oct. 2021. ISSN 0022-0957, 1460-2431. doi: 10.1093/jxb/erab340. URL <https://academic.oup.com/jxb/article/72/20/6949/6323944>.
- H. Pagès, P. Aboyoun, R. Gentleman, and S. DebRoy. *Biostrings: Efficient manipulation of biological strings*. 2024. doi: 10.18129/B9.bioc.Biostrings. URL <https://bioconductor.org/packages/Biostrings>.
- A. Pal, S. S. Swain, A. B. Das, A. K. Mukherjee, and P. K. Chand. Stable germ line

- transformation of a leafy vegetable crop amaranth (*Amaranthus tricolor* L.) mediated by *Agrobacterium tumefaciens*. *In Vitro Cellular & Developmental Biology - Plant*, 49(2):114–128, Apr. 2013. ISSN 1054-5476, 1475-2689. doi: 10.1007/s11627-013-9489-9. URL <http://link.springer.com/10.1007/s11627-013-9489-9>.
- E. Paradis and K. Schliep. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35:526–528, 2019. doi: 10.1093/bioinformatics/bty633.
- Y. Peng, A. H. Thrimawithana, J. M. Cooney, D. J. Jensen, R. V. Espley, and A. C. Allan. The proanthocyanin-related transcription factors MYBC1 and WRKY44 regulate branch points in the kiwifruit anthocyanin pathway. *Scientific Reports*, 10(1):14161, Aug. 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-70977-0. URL <https://www.nature.com/articles/s41598-020-70977-0>.
- M. Pesch, B. Dartan, R. Birkenbihl, I. E. Somssich, and M. Hülskamp. *Arabidopsis* TTG2 Regulates *TRY* Expression through Enhancement of Activator Complex-Triggered Activation. *The Plant Cell*, 26(10):4067–4083, Nov. 2014. ISSN 1532-298X, 1040-4651. doi: 10.1105/tpc.114.129379. URL <https://academic.oup.com/plcell/article/26/10/4067/6101552>.
- K. V. S. K. Prasad, A. Cheema, W. Scanlon, A. Matthews, S. Sharifova, E. Huq, and A. S. N. Reddy. A simple method to visualize pre-mRNA splicing with the naked eye using a genetically encoded visual splicing reporter. *Plant Physiology*, 196(2):726–730, Oct. 2024. ISSN 0032-0889, 1532-2548. doi: 10.1093/plphys/kiae396. URL <https://academic.oup.com/plphys/article/196/2/726/7721312>.
- A. T. Premathilake, J. Ni, S. Bai, R. Tao, M. Ahmad, and Y. Teng. R2R3-MYB transcription factor PpMYB17 positively regulates flavonoid biosynthesis in pear fruit. *Planta*, 252(4):59, Oct. 2020. ISSN 0032-0935, 1432-2048. doi: 10.1007/s00425-020-03473-4. URL <https://link.springer.com/10.1007/s00425-020-03473-4>.
- B. Pucker, N. Walker-Hale, J. Dzurlic, W. C. Yim, J. C. Cushman, A. Crum, Y. Yang, and S. F. Brockington. Multiple mechanisms explain loss of anthocyanins from betalain-pigmented *Caryophyllales*, including repeated wholesale loss of a key anthocyanidin synthesis enzyme. *New Phytologist*, 241(1):471–489, Jan. 2024. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.19341. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19341>.
- F. Qi, Y. Liu, Y. Luo, Y. Cui, C. Lu, H. Li, H. Huang, and S. Dai. Functional analysis of the ScAG and ScAGL11 MADS-box transcription factors for anthocyanin biosynthesis and bicolour pattern formation in *Senecio cruentus* ray florets. *Horticulture Research*, 9:uhac071, Jan. 2022. ISSN 2052-7276. doi: 10.1093/hr/uhac071. URL <https://academic.oup.com/hr/article/doi/10.1093/hr/uhac071/6552746>.
- T. Qi, S. Song, Q. Ren, D. Wu, H. Huang, Y. Chen, M. Fan, W. Peng, C. Ren, and D. Xie. The Jasmonate-ZIM-domain proteins interact with the WD-Repeat/bHLH/MYB complexes to regulate Jasmonate-mediated anthocyanin accumulation and trichome initiation in *Arabidopsis thaliana*. *The Plant Cell*, 23(5):1795–1814, May 2011. ISSN 1532-298X. doi: 10.1105/tpc.111.083261.

- X. Qi, C. Liu, L. Song, and M. Li. PaMADS7, a MADS-box transcription factor, regulates sweet cherry fruit ripening and softening. *Plant Science*, 301:110634, Dec. 2020. ISSN 01689452. doi: 10.1016/j.plantsci.2020.110634. URL <https://linkinghub.elsevier.com/retrieve/pii/S0168945220302405>.
- S. Qu, G. Wang, M. Li, W. Yu, and S. Zhu. LcNAC90 transcription factor regulates biosynthesis of anthocyanin in harvested litchi in response to ABA and GA3. *Postharvest Biology and Technology*, 194:112109, Dec. 2022. ISSN 09255214. doi: 10.1016/j.postharvbio.2022.112109. URL <https://linkinghub.elsevier.com/retrieve/pii/S0925521422002770>.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024. URL <https://www.R-project.org/>.
- J. L. Riechmann and E. M. Meyerowitz. The AP2/EREBP family of plant transcription factors. *Biological chemistry*, 379:633–646, 1998.
- J. L. Riechmann, J. Heard, G. Martin, L. Reuber, C.-Z. Jiang, J. Keddie, L. Adam, O. Pineda, O. J. Ratcliffe, R. R. Samaha, R. Creelman, M. Pilgrim, P. Broun, J. Z. Zhang, D. Ghandehari, B. K. Sherman, and G. L. Yu. *Arabidopsis* Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes. *Science*, 290(5499): 2105–2110, Dec. 2000. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.290.5499.2105. URL <https://www.science.org/doi/10.1126/science.290.5499.2105>.
- M. Rong, S.-X. Gao, D. Wen, Y.-H. Xu, and J.-H. Wei. The LOB domain protein, a novel transcription factor with multiple functions: A review. *Plant Physiology and Biochemistry*, 214:108922, Sept. 2024. ISSN 09819428. doi: 10.1016/j.plaphy.2024.108922. URL <https://linkinghub.elsevier.com/retrieve/pii/S0981942824005904>.
- P. J. Rushton, H. Macdonald, A. K. Huttly, C. M. Lazarus, and R. Hooley. Members of a new family of DNA-binding proteins bind to a conserved *cis*-element in the promoters of α -Amy2 genes. *Plant Molecular Biology*, 29(4):691–702, Nov. 1995. ISSN 0167-4412, 1573-5028. doi: 10.1007/BF00041160. URL <http://link.springer.com/10.1007/BF00041160>.
- M. Sakuta, A. Tanaka, K. Iwase, M. Miyasaka, S. Ichiki, M. Hatai, Y. T. Inoue, A. Yamagami, T. Nakano, K. Yoshida, and S. Shimada. Anthocyanin synthesis potential in betalain-producing *Caryophyllales* plants. *Journal of Plant Research*, 134(6):1335–1349, Nov. 2021. ISSN 0918-9440, 1618-0860. doi: 10.1007/s10265-021-01341-0. URL <https://link.springer.com/10.1007/s10265-021-01341-0>.
- P. Schauburger and A. Walker. *openxlsx: Read, Write and Edit xlsx Files*. 2025. URL <https://CRAN.R-project.org/package=openxlsx>.
- J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682, July 2012. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.2019. URL <https://www.nature.com/articles/nmeth.2019>.

- A. Schuh and G. Morris. Loss of Pleiotropic Regulatory Functions in *Tannin1*, the Sorghum Ortholog of Arabidopsis Master Regulator *TTG1*. *Plant Direct*, 9(3):e70055, Mar. 2025. ISSN 2475-4455, 2475-4455. doi: 10.1002/pld3.70055. URL <https://onlinelibrary.wiley.com/doi/10.1002/pld3.70055>.
- H. Sharma, P. Sharma, A. Kumar, N. Chawla, and A. S. Dhath. Multifaceted Regulation of Anthocyanin Biosynthesis in Plants: A Comprehensive Review. *Journal of Plant Growth Regulation*, 43(9):3048–3062, Sept. 2024. ISSN 0721-7595, 1435-8107. doi: 10.1007/s00344-024-11306-x. URL <https://link.springer.com/10.1007/s00344-024-11306-x>.
- H. Sheehan, T. Feng, N. Walker-Hale, S. Lopez-Nieves, B. Pucker, R. Guo, W. C. Yim, R. Badgami, A. Timoneda, L. Zhao, H. Tiley, D. Copetti, M. J. Sanderson, J. C. Cushman, M. J. Moore, S. A. Smith, and S. F. Brockington. Evolution of l-DOPA 4,5-dioxygenase activity allows for recurrent specialisation to betalain pigmentation in *Caryophyllales*. *New Phytologist*, 227(3):914–929, Aug. 2020. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.16089. URL <https://onlinelibrary.wiley.com/doi/10.1111/nph.16089>.
- M. Shi, Y. Zhang, T. Zhang, W. Zhang, S. Wang, M. Wei, S. Wang, and L. Zhao. The NAC activator, MdNAC77L, regulates anthocyanin accumulation in red flesh apple. *Horticultural Plant Journal*, 11(6):2009–2023, Nov. 2025. ISSN 24680141. doi: 10.1016/j.hpj.2024.06.007. URL <https://linkinghub.elsevier.com/retrieve/pii/S2468014124001900>.
- F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, and others. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7:539, 2011. Publisher: John Wiley & Sons, Ltd Chichester, UK.
- H. A. Stafford. Anthocyanins and betalains: evolution of the mutually exclusive pathways. *Plant Science*, 101(2):91–98, Jan. 1994. ISSN 01689452. doi: 10.1016/0168-9452(94)90244-5. URL <https://linkinghub.elsevier.com/retrieve/pii/0168945294902445>.
- Q. Sun, S. Jiang, T. Zhang, H. Xu, H. Fang, J. Zhang, M. Su, Y. Wang, Z. Zhang, N. Wang, and X. Chen. Apple NAC transcription factor MdNAC52 regulates biosynthesis of anthocyanin and proanthocyanidin through MdMYB9 and MdMYB11. *Plant Science*, 289:110286, Dec. 2019. ISSN 01689452. doi: 10.1016/j.plantsci.2019.110286. URL <https://linkinghub.elsevier.com/retrieve/pii/S0168945219308295>.
- R. M. Taipova, K. G. Musin, and B. R. Kuluev. *Agrobacterium*-Mediated Transformation of *Amaranthus cruentus* L. Epicotils. *Journal of Siberian Federal University. Biology*, pages 1–9, Mar. 2019. ISSN 19971389, 23135530. doi: 10.17516/1997-1389-0292. URL <http://journal.sfu-kras.ru/en/article/110078>.
- A. M. Takos, F. W. Jaffé, S. R. Jacob, J. Bogs, S. P. Robinson, and A. R. Walker. Light-Induced Expression of a *MYB* Gene Regulates Anthocyanin Biosynthesis in Red Apples. *Plant Physiology*, 142(3):1216–1232, Nov. 2006. ISSN 0032-0889, 1532-2548. doi: 10.1104/pp.106.088104. URL <https://academic.oup.com/plphys/article/142/3/1216-1232/6106456>.

- F. Tian, D.-C. Yang, Y.-Q. Meng, J. Jin, and G. Gao. PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Research*, page gkz1020, Nov. 2019. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkz1020. URL <https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkz1020/5614566>.
- A. Timoneda, T. Feng, H. Sheehan, N. Walker-Hale, B. Pucker, S. Lopez-Nieves, R. Guo, and S. Brockington. The evolution of betalain biosynthesis in *Caryophyllales*. *New Phytologist*, 224(1):71–85, Oct. 2019. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.15980. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.15980>.
- T. Tohge, Y. Nishiyama, M. Y. Hirai, M. Yano, J. Nakajima, M. Awazuhara, E. Inoue, H. Takahashi, D. B. Goodenowe, M. Kitayama, M. Noji, M. Yamazaki, and K. Saito. Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an *MYB* transcription factor. *The Plant Journal*, 42(2):218–235, Apr. 2005. ISSN 0960-7412, 1365-313X. doi: 10.1111/j.1365-313X.2005.02371.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2005.02371.x>.
- W. Verweij, C. E. Spelt, M. Bliet, M. De Vries, N. Wit, M. Faraco, R. Koes, and F. M. Quattrocchio. Functionally Similar WRKY Proteins Regulate Vacuolar Acidification in *Petunia* and Hair Development in *Arabidopsis*. *The Plant Cell*, 28(3):786–803, Mar. 2016. ISSN 1040-4651, 1532-298X. doi: 10.1105/tpc.15.00608. URL <https://academic.oup.com/plcell/article/28/3/786-803/6100934>.
- I. L. Viola, A. Camoirano, and D. H. Gonzalez. Redox-Dependent Modulation of Anthocyanin Biosynthesis by the TCP Transcription Factor TCP15 during Exposure to High Light Intensity Conditions in *Arabidopsis*. *Plant Physiology*, 170(1):74–85, Jan. 2016. ISSN 0032-0889, 1532-2548. doi: 10.1104/pp.15.01016. URL <https://academic.oup.com/plphys/article/170/1/74-85/6114309>.
- S. K. Vollmer, M. G. Stetter, and G. Hensel. First Successful Targeted Mutagenesis Using CRISPR/Cas9 in Stably Transformed Grain Amaranth Tissue. *Plant Biotechnology Journal*, page pbi.70590, Feb. 2026. ISSN 1467-7644, 1467-7652. doi: 10.1111/pbi.70590. URL <https://onlinelibrary.wiley.com/doi/10.1111/pbi.70590>.
- O. Wagih. *ggseqlogo: A 'ggplot2' Extension for Drawing Publication-Ready Sequence Logos*. 2025. URL <https://CRAN.R-project.org/package=ggseqlogo>.
- N. Walker-Hale, M. A. Guerrero-Rubio, and S. F. Brockington. Multiple transitions to high l-DOPA 4,5-dioxygenase activity reveal molecular pathways to convergent betalain pigmentation in *Caryophyllales*. *New Phytologist*, 247(1):341–357, July 2025. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.70177. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.70177>.
- J. Wang, W. Lian, Y. Cao, X. Wang, G. Wang, C. Qi, L. Liu, S. Qin, X. Yuan, X. Li, S. Ren, and Y.-D. Guo. Overexpression of *BoNAC019*, a NAC transcription factor from *Brassica oleracea*, negatively regulates the dehydration response and anthocyanin biosynthesis in *Arabidopsis*. *Scientific Reports*, 8(1):13349, Sept. 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-31690-1. URL <https://www.nature.com/articles/s41598-018-31690-1>.

- J. Wang, X. Wang, B. Ma, P. Leng, J. Wu, and Z. Hu. SoNAC72-SoMYB44/SobHLH130 module contributes to flower color fading via regulating anthocyanin biosynthesis by directly binding to the *SoUFGT1* promoter in lilac (*Syringa oblata*). *Horticulture Research*, 12(3):uhae326, Feb. 2025. ISSN 2052-7276. doi: 10.1093/hr/uhae326. URL <https://academic.oup.com/hr/article/doi/10.1093/hr/uhae326/7906584>.
- K. Wang, H. Guo, and Y. Yin. AP2/ERF transcription factors and their functions in *Arabidopsis* responses to abiotic stresses. *Environmental and Experimental Botany*, 222:105763, June 2024. ISSN 00988472. doi: 10.1016/j.envexpbot.2024.105763. URL <https://linkinghub.elsevier.com/retrieve/pii/S0098847224001217>.
- L. Wang, W. Tang, Y. Hu, Y. Zhang, J. Sun, X. Guo, H. Lu, Y. Yang, C. Fang, X. Niu, J. Yue, Z. Fei, and Y. Liu. A MYB/bHLH complex regulates tissue-specific anthocyanin biosynthesis in the inner pericarp of red-centered kiwifruit *Actinidia chinensis* cv. Hongyang. *The Plant Journal*, 99(2):359–378, July 2019. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.14330. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.14330>.
- Y. Wang, Y. Wang, L.-J. Zhou, J. Peng, C. Chen, S. Liu, A. Song, J. Jiang, S. Chen, and F. Chen. CmNAC25 targets *CmMYB6* to positively regulate anthocyanin biosynthesis during the post-flowering stage in chrysanthemum. *BMC Biology*, 21(1):211, Oct. 2023a. ISSN 1741-7007. doi: 10.1186/s12915-023-01719-7. URL <https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-023-01719-7>.
- Y. Wang, Y. Zhu, H. Jiang, Z. Mao, J. Zhang, H. Fang, W. Liu, Z. Zhang, X. Chen, and N. Wang. The regulatory module MdBZR1–MdCOL6 mediates brassinosteroid- and light-regulated anthocyanin synthesis in apple. *New Phytologist*, 238(4):1516–1533, May 2023b. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.18779. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.18779>.
- Y. M. Wang, J. B. Wang, D. Luo, and J. F. Jia. Regeneration of plants from callus cultures of roots induced by *Agrobacterium rhizogenes* on *Alhagi pseudoalhagi*. *Cell Research*, 11(4):279–284, Dec. 2001. ISSN 1001-0602, 1748-7838. doi: 10.1038/sj.cr.7290097. URL <http://www.nature.com/articles/7290097>.
- H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL <https://ggplot2.tidyverse.org>.
- H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemond, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019. doi: 10.21105/joss.01686.
- H. Wickham, R. François, L. Henry, K. Müller, and D. Vaughan. *dplyr: A Grammar of Data Manipulation*. 2023. URL <https://CRAN.R-project.org/package=dplyr>.
- H. Wickham, D. Vaughan, and M. Girlich. *tidyr: Tidy Messy Data*. 2024. URL <https://CRAN.R-project.org/package=tidyr>.
- B. Winkel-Shirley. Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. *Plant Physiology*, 126(2):485–493, June 2001. ISSN

- 1532-2548, 0032-0889. doi: 10.1104/pp.126.2.485. URL <https://academic.oup.com/plphys/article/126/2/485/6102700>.
- T. S. Winkler, S. K. Vollmer, N. Dyballa-Rukes, S. Metzger, and M. G. Stetter. Isoform-resolved genome annotation enables mapping of tissue-specific betalain regulation in amaranth. *New Phytologist*, page np.19736, Apr. 2024. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.19736. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19736>.
- R. Wu, T. Wang, T. McGie, C. Voogd, A. C. Allan, R. P. Hellens, and E. Varkonyi-Gasic. Overexpression of the kiwifruit *SVP3* gene affects reproductive development and suppresses anthocyanin biosynthesis in petals, but has no effect on vegetative growth, dormancy, or flowering time. *Journal of Experimental Botany*, 65(17):4985–4995, Sept. 2014. ISSN 1460-2431, 0022-0957. doi: 10.1093/jxb/eru264. URL <https://academic.oup.com/jxb/article-lookup/doi/10.1093/jxb/eru264>.
- T. Xu, L. Yu, N. Huang, W. Liu, Y. Fang, C. Chen, L. Jiang, T. Wang, J. Zhao, Z. Zhang, Y. Xu, N. Wang, and X. Chen. The regulatory role of MdNAC14-Like in anthocyanin synthesis and proanthocyanidin accumulation in red-fleshed apples. *Plant Physiology and Biochemistry*, 204:108068, Nov. 2023. ISSN 09819428. doi: 10.1016/j.plaphy.2023.108068. URL <https://linkinghub.elsevier.com/retrieve/pii/S098194282300579X>.
- Y. Xue, K. Li, W. Feng, Z. Lai, and S. Liu. Identification of R2R3-MYB Transcription Factor Family Based on *Amaranthus tricolor* Genome and AtrMYB72 Promoting Betalain Biosynthesis by Directly Activating *AtrCYP76AD1* Expression. *Plants*, 14(3):324, Jan. 2025. ISSN 2223-7747. doi: 10.3390/plants14030324. URL <https://www.mdpi.com/2223-7747/14/3/324>.
- L. Yan. *ggvenn: Draw Venn Diagram by 'ggplot2'*. 2025. URL <https://github.com/yanlinlin82/ggvenn>.
- R. Yang, T. Huang, W. Song, Z. An, Z. Lai, and S. Liu. Identification of WRKY gene family members in amaranth based on a transcriptome database and functional analysis of AtrWRKY42-2 in betalain metabolism. *Frontiers in Plant Science*, 14:1300522, Dec. 2023. ISSN 1664-462X. doi: 10.3389/fpls.2023.1300522. URL <https://www.frontiersin.org/articles/10.3389/fpls.2023.1300522/full>.
- G. Yao, M. Ming, A. C. Allan, C. Gu, L. Li, X. Wu, R. Wang, Y. Chang, K. Qi, S. Zhang, and J. Wu. Map-based cloning of the pear gene *MYB114* identifies an interaction with other transcription factors to coordinately regulate fruit anthocyanin biosynthesis. *The Plant Journal*, 92(3):437–451, Nov. 2017. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.13666. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.13666>.
- G. Yu. *ggplotify: Convert Plot to 'grob' or 'ggplot' Object*. 2025. URL <https://CRAN.R-project.org/package=ggplotify>.
- G. Yu, D. K. Smith, H. Zhu, Y. Guan, and T. T. Lam. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1):28–36, Jan. 2017. ISSN 2041-210X, 2041-210X. doi: 10.1111/2041-210X.12628. URL <https://besjournals.onlinelibrary>.

wiley.com/doi/10.1111/2041-210X.12628.

- J. Zeng, J. Chen, K. Shah, F. Xie, C. Chen, J. Chen, J. Zhao, G. Hu, Z. Zhang, and Y. Qin. Identification of *HuSPL* family and key role of HuSPL12 in regulation of betalain biosynthesis in pitaya. *Physiologia Plantarum*, 175(3):e13923, May 2023. ISSN 0031-9317, 1399-3054. doi: 10.1111/ppl.13923. URL <https://onlinelibrary.wiley.com/doi/10.1111/ppl.13923>.
- B. Zhang. *Evolutionary Analysis of MYBs-bHLH-WD40 Protein Complexes Formation and Their Functional Relationship in Planta*. PhD thesis, Universität zu Köln, May 2018.
- B. Zhang and M. Hülskamp. Evolutionary Analysis of MBW Function by Phenotypic Rescue in *Arabidopsis thaliana*. *Frontiers in Plant Science*, 10:375, Mar. 2019. ISSN 1664-462X. doi: 10.3389/fpls.2019.00375. URL <https://www.frontiersin.org/article/10.3389/fpls.2019.00375/full>.
- B. Zhang and A. Schrader. TRANSPARENT TESTA GLABRA 1-Dependent Regulation of Flavonoid Biosynthesis. *Plants*, 6(4):65, Dec. 2017. ISSN 2223-7747. doi: 10.3390/plants6040065. URL <https://www.mdpi.com/2223-7747/6/4/65>.
- D. Zhang, H. Zhou, Y. Zhang, Y. Zhao, Y. Zhang, X. Feng, and H. Lin. Diverse roles of MYB transcription factors in plants. *Journal of Integrative Plant Biology*, 67(3): 539–562, Mar. 2025a. ISSN 1672-9072, 1744-7909. doi: 10.1111/jipb.13869. URL <https://onlinelibrary.wiley.com/doi/10.1111/jipb.13869>.
- J. Zhang, H. Xu, N. Wang, S. Jiang, H. Fang, Z. Zhang, G. Yang, Y. Wang, M. Su, L. Xu, and X. Chen. The ethylene response factor MdERF1B regulates anthocyanin and proanthocyanidin biosynthesis in apple. *Plant Molecular Biology*, 98(3):205–218, Oct. 2018. ISSN 0167-4412, 1573-5028. doi: 10.1007/s11103-018-0770-5. URL <http://link.springer.com/10.1007/s11103-018-0770-5>.
- L. Zhang, J. Zhang, B. Wei, Y. Li, X. Fang, Y. Zhong, and L. Wang. Transcription factor MdNAC33 is involved in ALA-induced anthocyanin accumulation in apples. *Plant Science*, 339:111949, Feb. 2024. ISSN 01689452. doi: 10.1016/j.plantsci.2023.111949. URL <https://linkinghub.elsevier.com/retrieve/pii/S0168945223003667>.
- S. Zhang, Y. Chen, L. Zhao, C. Li, J. Yu, T. Li, W. Yang, S. Zhang, H. Su, and L. Wang. A novel NAC transcription factor, MdNAC42, regulates anthocyanin accumulation in red-fleshed apple by interacting with MdMYB10. *Tree Physiology*, 40(3):413–423, Mar. 2020. ISSN 1758-4469. doi: 10.1093/treephys/tpaa004. URL <https://academic.oup.com/treephys/article/40/3/413/5728680>.
- X. Zhang, M. Gou, and C.-J. Liu. *Arabidopsis* Kelch Repeat F-Box Proteins Regulate Phenylpropanoid Biosynthesis via Controlling the Turnover of Phenylalanine Ammonia-Lyase. *The Plant Cell*, 25(12):4994–5010, Jan. 2014. ISSN 1532-298X, 1040-4651. doi: 10.1105/tpc.113.119644. URL <https://academic.oup.com/plcell/article/25/12/4994/6098483>.
- Y. Zhang, S. Wang, X. Wang, T. Qian, Y. Sun, M. Jaudal, R. G. Atkinson, J. Zhao, G. Zheng, P. Sun, C. Fang, and X. Xie. FaMADS6 negatively regulates anthocyanin accumulation in strawberry by inhibiting *FaMYB10* expression and ABA biosynthesis.

- Horticultural Plant Journal*, page S2468014125000986, May 2025b. ISSN 24680141. doi: 10.1016/j.hpj.2025.02.013. URL <https://linkinghub.elsevier.com/retrieve/pii/S2468014125000986>.
- Z. Zhang, X. Wang, Y. Gao, X. Xian, D. Zhang, W. Zhao, X. Wang, and Y. Wang. Orchestrating anthocyanin biosynthesis in fruit of fruit trees: Transcriptional, post-transcriptional, and post-translational regulation. *International Journal of Biological Macromolecules*, 307:141835, May 2025c. ISSN 01418130. doi: 10.1016/j.ijbiomac.2025.141835. URL <https://linkinghub.elsevier.com/retrieve/pii/S0141813025023864>.
- C. Zhao, X. Liu, Q. Gong, J. Cao, W. Shen, X. Yin, D. Grierson, B. Zhang, C. Xu, X. Li, K. Chen, and C. Sun. Three AP2/ERF family members modulate flavonoid synthesis by regulating type IV chalcone isomerase in citrus. *Plant Biotechnology Journal*, 19(4):671–688, Apr. 2021. ISSN 1467-7644, 1467-7652. doi: 10.1111/pbi.13494. URL <https://onlinelibrary.wiley.com/doi/10.1111/pbi.13494>.
- Y. Zhao, J. Sun, S. Cherono, J.-P. An, A. C. Allan, and Y. Han. Colorful hues: insight into the mechanisms of anthocyanin pigmentation in fruit. *Plant Physiology*, 192(3): 1718–1732, July 2023. ISSN 0032-0889, 1532-2548. doi: 10.1093/plphys/kiad160. URL <https://academic.oup.com/plphys/article/192/3/1718/7076681>.
- H. Zhou, K. Lin-Wang, H. Wang, C. Gu, A. P. Dare, R. V. Espley, H. He, A. C. Allan, and Y. Han. Molecular genetics of blood-fleshed peach reveals activation of anthocyanin biosynthesis by NAC transcription factors. *The Plant Journal*, 82(1): 105–121, Apr. 2015. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.12792. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.12792>.
- Y. Zhu, Y. Liu, F. You, Z. Wan, M. Guo, M. Lu, L. Yang, X. Wang, J. Yang, L. Jia, and N. Su. VrNAC25 Promotes Anthocyanin Synthesis in Mung Bean Sprouts Synergistically with VrMYB90. *Plants*, 14(23):3667, Dec. 2025. ISSN 2223-7747. doi: 10.3390/plants14233667. URL <https://www.mdpi.com/2223-7747/14/23/3667>.
- M.-G. Zhuo, T.-Y. Wang, X.-M. Huang, G.-B. Hu, B.-Y. Zhou, H.-C. Wang, and F. Abbas. ERF transcription factors govern anthocyanin biosynthesis in litchi pericarp by modulating the expression of anthocyanin biosynthesis genes. *Scientia Horticulturae*, 337:113464, Nov. 2024. ISSN 03044238. doi: 10.1016/j.scienta.2024.113464. URL <https://linkinghub.elsevier.com/retrieve/pii/S0304423824006216>.
- I. M. Zimmermann, M. A. Heim, B. Weisshaar, and J. F. Uhrig. Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *The Plant Journal*, 40(1):22–34, Oct. 2004. ISSN 0960-7412, 1365-313X. doi: 10.1111/j.1365-313X.2004.02183.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2004.02183.x>.
- S.-C. Zou, M.-G. Zhuo, F. Abbas, G.-B. Hu, H.-C. Wang, and X.-M. Huang. Transcription factor LcNAC002 coregulates chlorophyll degradation and anthocyanin biosynthesis in litchi. *Plant Physiology*, 192(3):1913–1927, July 2023. ISSN 0032-0889, 1532-2548. doi: 10.1093/plphys/kiad118. URL <https://academic.oup.com/plphys/article/192/3/1913/7058562>.

4.9 Supplement

Supplemental Figure 1 – Results of all tested yeast-two-hybrid combinations of set 1.

Supplemental Figure 2 – Results of all tested yeast-two-hybrid combinations of set 2.

Supplemental Figure 3 – Transactivation assay in mammalian cells and synthetic positive control for amaranth promoter in protoplast.

Supplemental Figure 4 – Presence, location, and distribution of predicted binding sites of main transcription factor classes across the *CYP76AD2* promoter and *DODAA1* promoter.

Supplemental Figure 5 – Expression of *aragonate dehydratase* (*ADT*) and *aragonate dehydrogenase* (*ADH*) genes in environmentally or transgenically induced betalain accumulation in *A. hypochondriacus* callus.

Supplemental Figure 6 – Expression of flavonoid pathway genes in environmentally or transgenically induced betalain accumulation in *A. hypochondriacus* callus.

Supplemental Figure 7 – Expression of differentially expressed AP2/ERF transcription factors in environmentally or transgenically induced betalain accumulation in *A. hypochondriacus* callus.

Supplemental Figure 8 – Expression of differentially expressed bZIP transcription factors in environmentally or transgenically induced betalain accumulation in *A. hypochondriacus* callus.

Supplemental Figure 9 – Expression of differentially expressed NAC transcription factors in environmentally or transgenically induced betalain accumulation in *A. hypochondriacus* callus.

Supplemental Figure 10 – Phylogeny of all *A. hypochondriacus* (red), *A. thaliana* (green), and *M. domestica* ERF transcription factors.

Supplemental Table 1 – Plasmids used in this study.

Supplemental Table 2 – Primers used for cloning.

Supplemental Table 3 – Tested yeast-two-hybrid combinations.

Supplemental Table 4 – Information about the promoters used for protoplast transactivation assays.

Supplemental Table 5 – Sources of the used PAP1 homolog sequences from different species

Supplemental Table 6 – Identified amaranth anthocyanin pathway genes and the conserved amino acid sequences used for identification.

Supplemental Table 7 – Identified amaranth aragonate dehydratase (*ADT*) and aragonate dehydrogenase (*ADH α*) genes and the conserved amino acid sequences used for identification.

Appendix S1 - DNA sequences of the promoters used for protoplast transactivation assays.

4.9.1 Supplementary figures

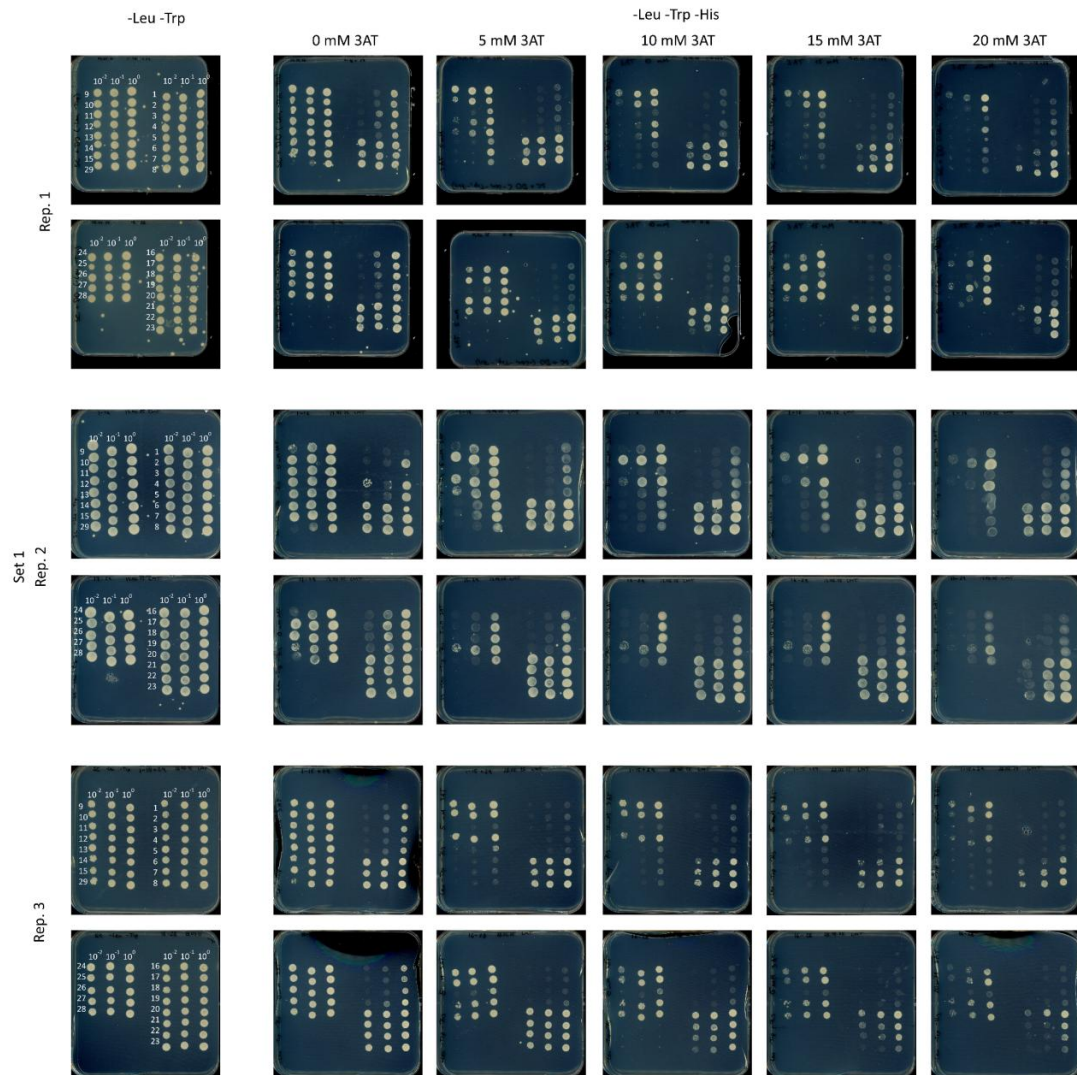


Figure S4.1: Results of all tested yeast-two-hybrid combinations of set 1. See Table S4.3 for which number refers to which combination. Set 1 contains combinations 1-29, and set 2 contains combinations 1, 7, 8, 12, 13, and 26-49.

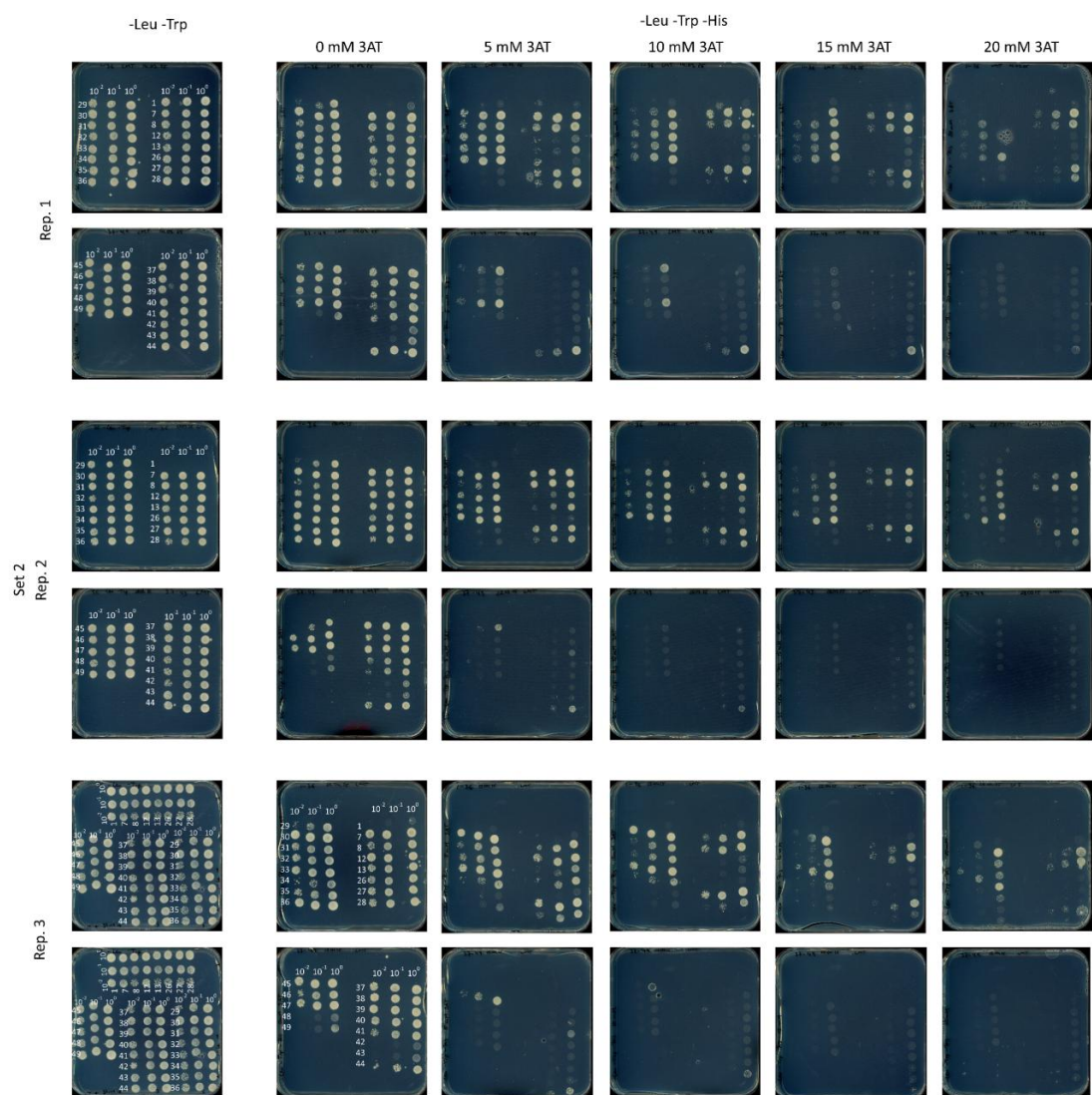
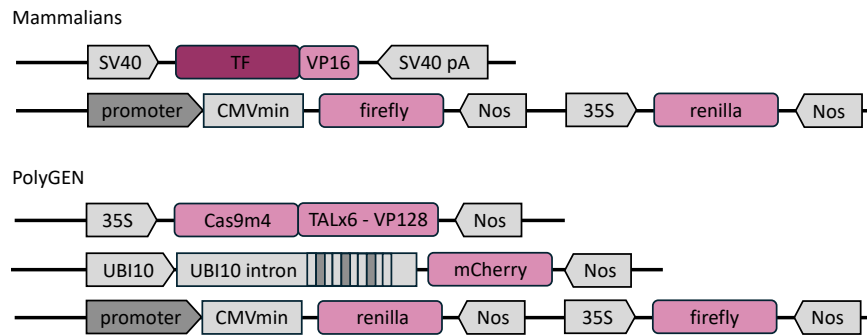
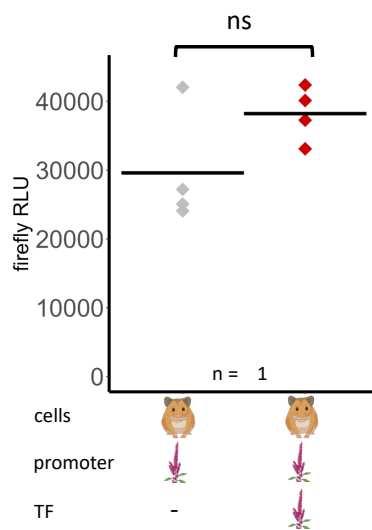


Figure S4.2: Results of all tested yeast-two-hybrid combinations of set 2. See Table S4.3 for which number refers to which combination. Set 1 contains combinations 1-29, and set 2 contains combinations 1, 7, 8, 12, 13, and 26-49.

a



b



c

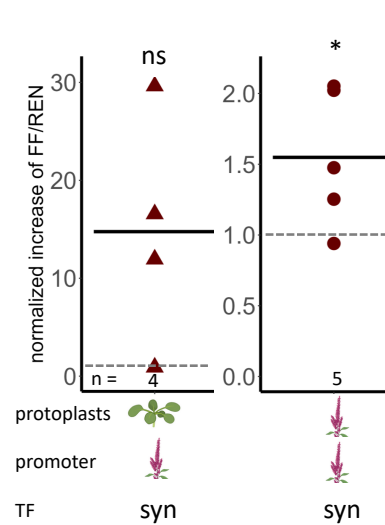


Figure S4.3: Transactivation assay in mammalian cells and synthetic positive control for amaranth promoter in protoplast. Schematic overview of the constructs and the protocol used for the transactivation assays in mammalian cells and for the PolyGEN synthetic positive control (a). Grey indicates promoters and terminators, and purple translated sequences; faded parts are fixed, while the dark color highlights the parts that were specific for each tested combination. For mammalian assays, one effector construct (with the transcription factor (TF) of interest) and one reporter construct (with the potential target promoter) were co-transformed. For PolyGEN synthetic positive control, three plasmids were co-transformed: one to constitutively express the dead Cas9 activator, one carrying a UBI10 promoter with an intron containing the polycistronic guide cassette and mCherry and one with the respective reporter construct for which the guides have been designed. Transactivation assay in CHO cells with the shorter *AhCYP76AD2* promoter and AhMYB2.1-VP16 showed a non-significant activation of the promoter (b). Note that the data is only based on one experiment. Transactivation assay in *A. thaliana* or *A. hypochondriacus* protoplasts with the *AhcDOPA5GT* promoter and a synthetic positive control designed to target the promoter (PolyGEN, Urquiza-Garcia *et al.* in preparation) showed binding and activation in both species (only in amaranth significant) (c). Luminescence data was averaged across 4-6 protoplast aliquots from one experiment and normalized by the respective control (same reporter but a non-binding effector, dashed grey line). Data from 1-5 independent experiments are shown; a black line indicates the overall mean. Significant deviation from the respective control was tested with ANOVA with a significance threshold of 0.05. CMVmin = minimal promoter from the cauliflower mosaic virus, FF = firefly, REN = renilla, SV40 = simian virus 40 promoter and polyadenylation signal (pA), UBI10 = ubiquitin 10 promoter and intron. Hamster, *A. thaliana* and amaranth icons are from BioRender.

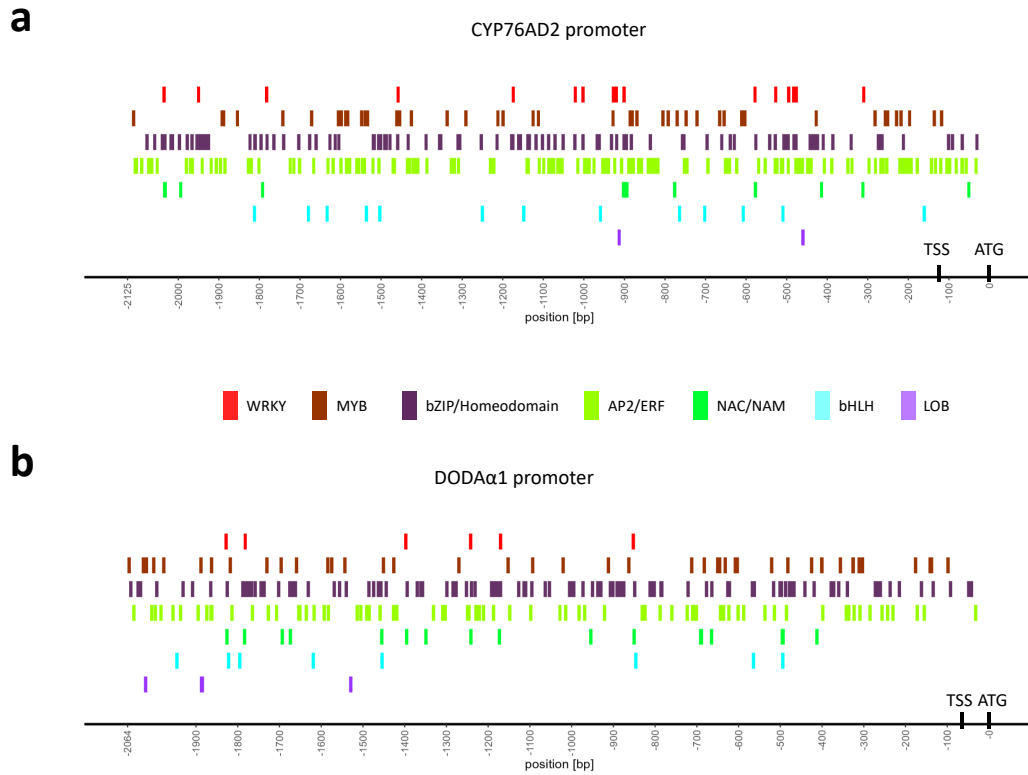


Figure S4.4: Presence, location, and distribution of predicted binding sites of main transcription factor classes across the *CYP76AD2* promoter and *DODAA1* promoter. Only the predicted binding sites of WRKY, MYB, bZip/Homeodomain, AP2/ERF, NAC/NAM, bHLH and LOB are shown. For both *CYP76AD2* (a) and *DODAA1* (b) the 2 kb upstream region from the transcriptional start site (TSS) and the 5' untranslated region (UTR) were taken for the analysis.

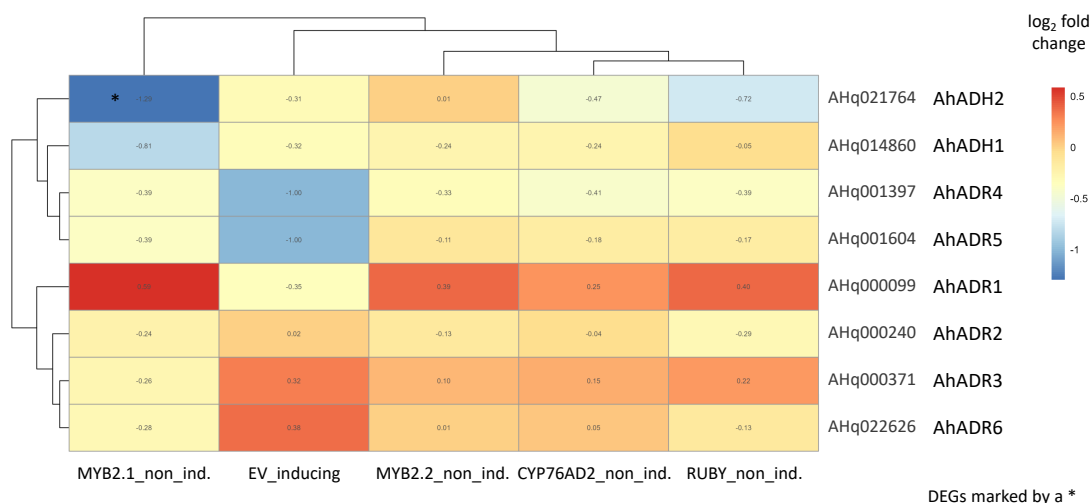


Figure S4.5: Expression of *aragonate dehydratase* (*ADT*) and *aragonate dehydrogenase* (*ADH*) genes in environmentally or transgenically induced betalain accumulation in *A. hypochondriacus* callus. Heatmap showing the expression of all found ADT and ADH genes of all samples as logfold2 values compared to the control (EV_non_inducing). Differentially expressed genes (DEGs) of each sample are marked with asterisks.

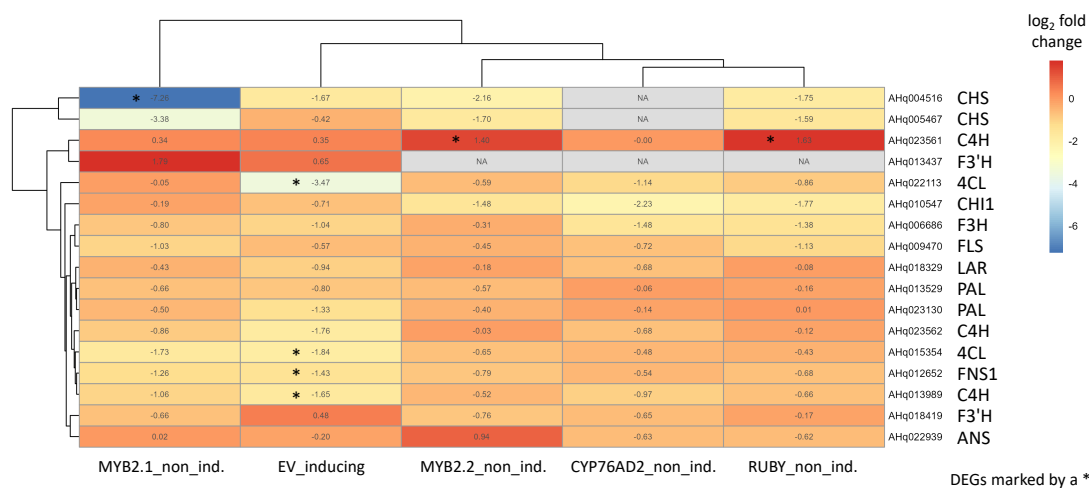


Figure S4.6: Expression of flavonoid pathway genes in environmentally or transgenically induced betalain accumulation in *A. hypochondriacus* callus. Heatmap showing the expression of all found flavonoid pathway genes of all samples as logfold2 values compared to the control (EV_non_inducing). Note that the found DFR homolog (AHq009850) was not expressed in any of the callus samples and that we did not find a candidate gene for F3'5'H. Differentially expressed genes of each sample are marked with asterisks. PAL (phenylalanine ammonium lyase), C4H (cinnamate 4-hydroxylase), 4CL (4-coumarate:CoA ligase), CHS (naringenin-chalcone synthase), CHI (chalcone isomerase), FNS (flavone synthase), F3H (flavanone 3-hydroxylase), FLS (flavonol synthase), F3'H (flavonoid 3'-hydroxylase), F3'5'H (flavonoid 3'5'-hydroxylase), DFR (dihydroflavonol 4-reductase), LAR (leucoanthocyanidin reductase), ANS (anthocyanidin synthase).

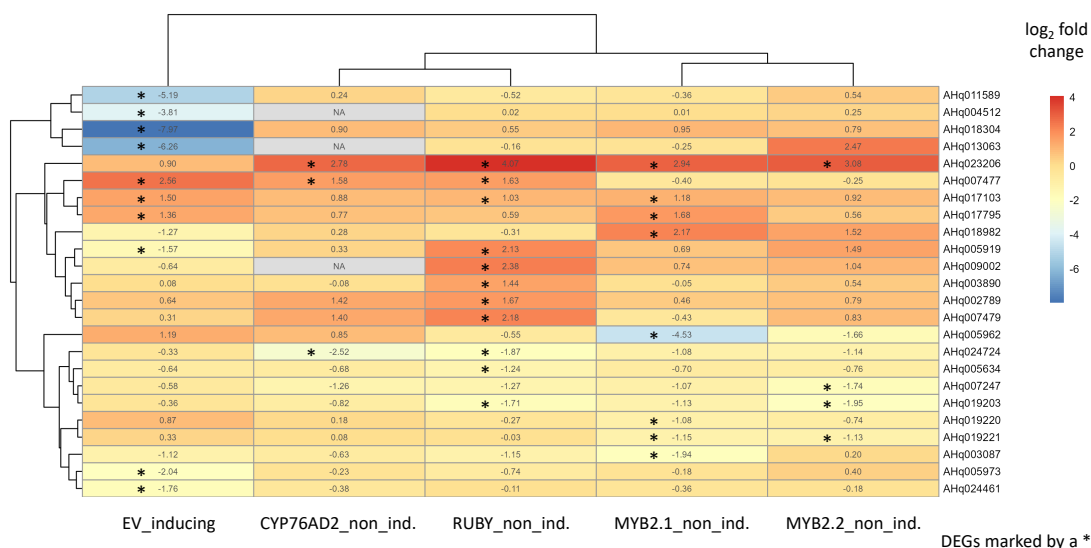


Figure S4.7: Expression of differentially expressed AP2/ERF transcription factors in environmentally or transgenically induced betalain accumulation in *A. hypochondriacus* callus. Heatmap showing the expression of all identified AP2/ERF transcription factors that are differentially expressed in at least one sample as logfold2 values compared to the control (EV_non_inducing). See section 4.2.8 for more details on how AP2/ERF transcription factor genes were identified. Differentially expressed genes (DEGs) of each sample are marked with asterisks. AP2/ERF (APETALA2/Ethylene-Responsive element binding Factor).

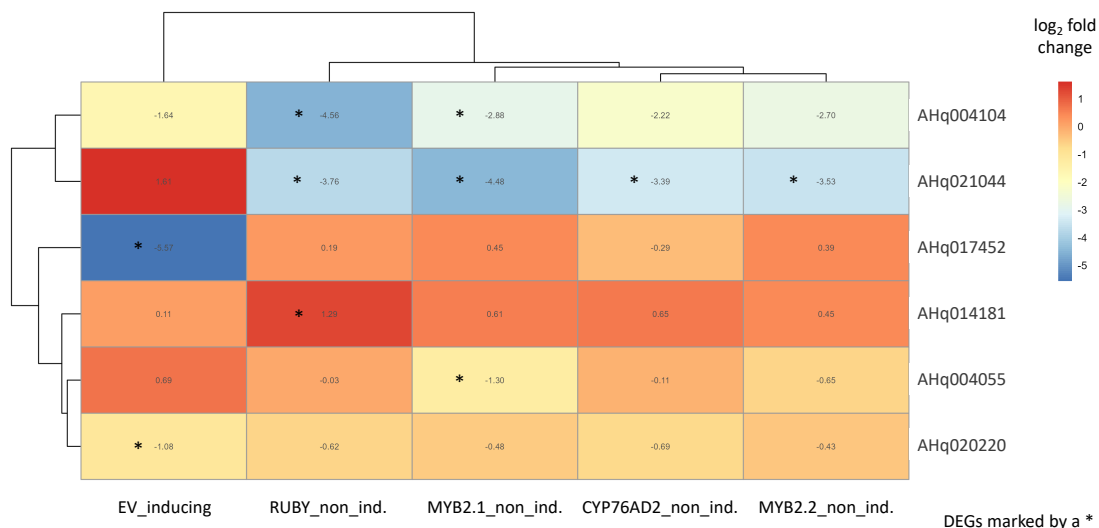


Figure S4.8: Expression of differentially expressed bZIP transcription factors in environmentally or transgenically induced betalain accumulation in *A. hypochondriacus* callus. Heatmap showing the expression of all identified bZIP transcription factors that are differentially expressed in at least one sample as logfold2 values compared to the control (EV_non_inducing). See section 4.2.8 for more details on how bZIP transcription factor genes were identified. Differentially expressed genes (DEGs) of each sample are marked with asterisks. bZIP (basic region/leucine zipper).

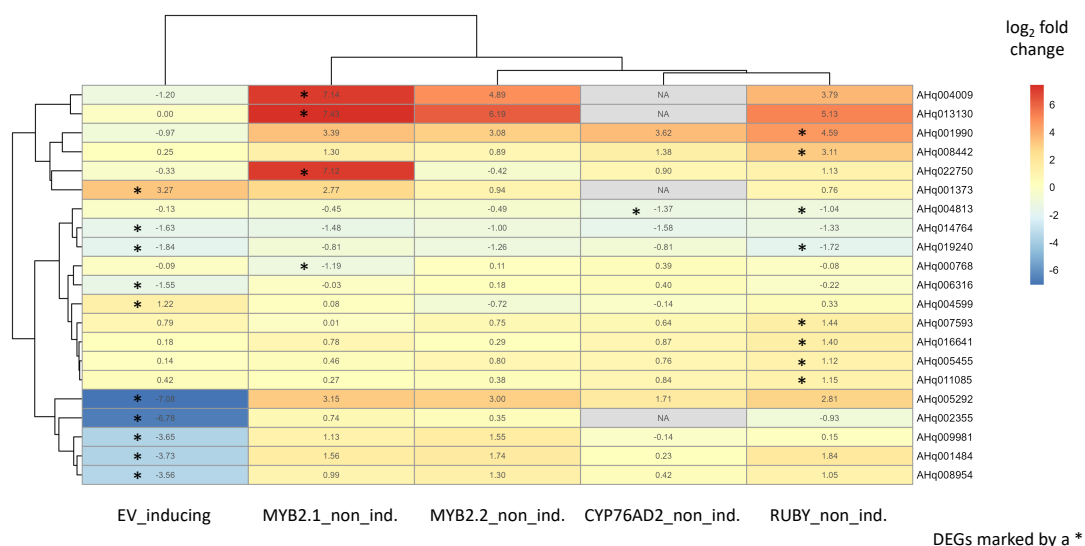


Figure S4.9: Expression of differentially expressed NAC transcription factors in environmentally or transgenically induced betalain accumulation in *A. hypochondriacus* callus. Heatmap showing the expression of all identified NAC transcription factors that are differentially expressed in at least one sample as logfold2 values compared to the control (EV_non_inducing). See section 4.2.8 for more details on how NAC transcription factor genes were identified. Differentially expressed genes (DEGs) of each sample are marked with asterisks. NAC (named after NO APICAL MERISTEM (NAM), ARABIDOPSIS TRANSCRIPTION ACTIVATOR FACTOR 1/2 (ATAF1/2), and CUP-SHAPED COTYLEDON (CUC2)).

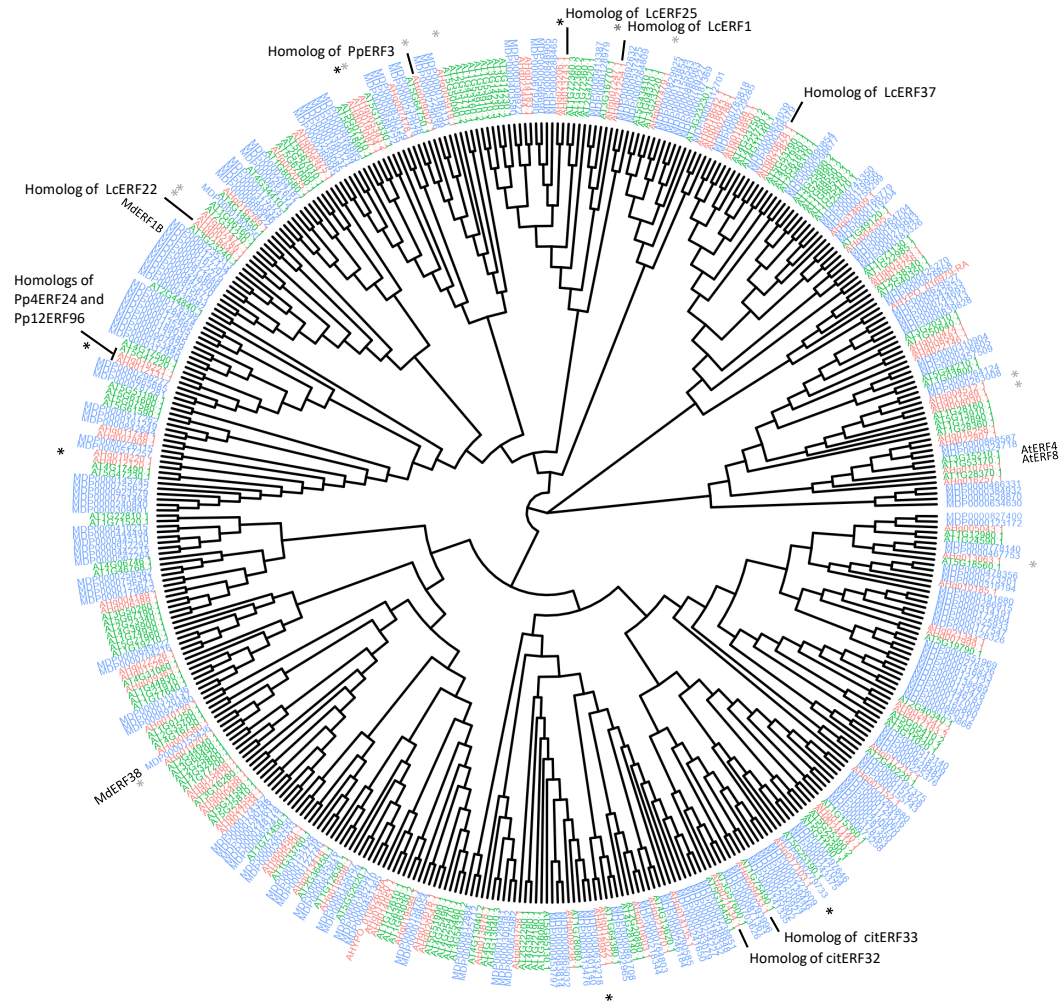


Figure S4.10: Phylogeny of all *A. hypochondriacus* (red), *A. thaliana* (green), and *M. domestica* ERF transcription factors. Genes differentially expressed (DEGs) in MYB2.1_non_inducing transgenic callus lines are highlighted with black asterisks, and DEGs of the other sample with grey asterisks. Homologous amaranth genes of known anthocyanin-regulating ERFs from other species (Ni et al., 2019; Zhuo et al., 2024; Zhao et al., 2021; Koyama and Sato, 2018; Yao et al., 2017; An et al., 2020; Zhang et al., 2018) were found by BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Note that the DEGs AHq017795, AHq005962, AHq003890, AHq024724, AHq005634, AHq005973, and AHq024461 are not included in the tree, as they were not part of the ERF data set from PlantTFDB. AHq005634, AHq003890, and AHq024724 are in the AP2 data set, and AHq005962 is in the RAV data set from PlantTFDB. AHq017795 and AHq005973 are classified as AP2-like TF homologs to AtANT (AT4G37750). Lastly, for AHq024461, it is not clear why it is not included in the PlantTFDB database, but the *A. thaliana* homolog AT5G25190 can be found in the tree.

4.9.2 Supplementary Tables

Table S4.1: Plasmids used in this study

Name	Description	Source
CPGE_VEC00113	empty vector control	DNA cloning service
CPGE_VEC00447	<i>AhCYP76AD1</i> and <i>AhCYP76AD5</i> CRISPR knock-out vector	Vollmer et al. (2026)
CPGE_VEC00256	35S:RUBY	Gabriele Buchmann
CPGE_VEC00594	pZ001	Lampropoulos et al. (2013)
CPGE_VEC00599	pGGA004	Lampropoulos et al. (2013)
CPGE_VEC00602	pGGB001	Lampropoulos et al. (2013)
CPGE_VEC00508	pGGC000	Lampropoulos et al. (2013)
CPGE_VEC00617	pGGD002	Lampropoulos et al. (2013)
CPGE_VEC00623	pGGE001	Lampropoulos et al. (2013)
CPGE_VEC00632	pGGF005	Lampropoulos et al. (2013)
CPGE_VEC00604	pGGB003	Lampropoulos et al. (2013)
CPGE_VEC00616	pGGD001	Lampropoulos et al. (2013)
CPGE_VEC00509	pGGC_AhMYB2.2	Winkler et al. (2024)
CPGE_VEC00510	pGGC_AhMYB2.1	Winkler et al. (2024)
CPGE_VEC00647	pGGC_AhCYP76AD2	this study
CPGE_VEC00643	35S:AhMYB2.1_GFP_hyg	this study
CPGE_VEC00644	35S:AhMYB2.2_GFP_hyg	this study
CPGE_VEC00653	35S:AhCYP76AD2	this study
CPGE_VEC00638	pSOUP (helper plasmids for GreenGate plasmids)	Hellens et al. (2000)
CPGE_VEC00826	BP311	Zhang (2018)

CPGE_VEC00896	pDONR201_AhMYB2.2	this study
CPGE_VEC00897	pDONR201_AhMYB2.1	this study
CPGE_VEC00881	pAS	Harper et al. (1993)
CPGE_VEC00900	pAS_AhMYB2.1	this study (Linda Marie Treder)
CPGE_VEC00899	pAS_AhMYB2.2	this study (Linda Marie Treder)
CPGE_VEC00903	pAS_AhTT8	Julio Martinez
CPGE_VEC00904	pAS_AhTTG1a	Julio Martinez
CPGE_VEC00905	pAS_AtTT8	Julio Martinez
CPGE_VEC00946	pAS_AtGL3	Andrii Zinchenko
CPGE_VEC00947	pAS_AtEGL3	Andrii Zinchenko
CPGE_VEC00948	pAS_AtMYC1	Andrii Zinchenko
CPGE_VEC00906	pAS_AtTTG1	Julio Martinez
CPGE_VEC00882	pACT	Harper et al. (1993)
CPGE_VEC00902	pACT_AhMYB2.1	this study (Linda Marie Treder)
CPGE_VEC00901	pACT_AhMYB2.2	this study (Linda Marie Treder)
CPGE_VEC00907	pACT_AhTT8	Julio Martinez
CPGE_VEC00908	pACT_AhTTG1a	Julio Martinez
CPGE_VEC00909	pACT_AtTT8	Julio Martinez
CPGE_VEC00949	pACT_AtGL3	Andrii Zinchenko
CPGE_VEC00950	pACT_AtEGL3	Andrii Zinchenko
CPGE_VEC00951	pACT_AtMYC1	Andrii Zinchenko
CPGE_VEC00910	pACT_AtTTG1	Julio Martinez
CPGE_VEC00513	pmz824	Müller et al. (2014)
CPGE_VEC00514	pep602	Estefania Pavesi (Pavesi <i>et al. in preparation</i>)
CPGE_VEC01025	pep602_sw	this study
CPGE_VEC00960	pJS078 (stuffer)	Jonas Schön
CPGE_VEC00979	pmz824_AtPAP1_vp16-NLS	this study

CPGE_VEC00980	pep602_AtDFR-1132bp	this study
CPGE_VEC01024	pep602_AHp023148_UTR_2125bp_sw	this study
CPGE_VEC00518	pmz824_AhMYB2.1_vp16	this study
CPGE_VEC00977	pep602_BvCYP76AD1_BB_551bp	this study
CPGE_VEC00888	pmz824_BvMYB1_vp16-NLS	this study
CPGE_VEC00829	pJS174_AhMYB2.1-VP16	this study
CPGE_VEC00527	pep602_AHp023148_1966bp	this study
CPGE_VEC00828	pJS174	Jonas Schön
CPGE_VEC01022	pdgb3alpha1_PTG-UBI-cDOPA5GT-NOS_mCherry	this study
CPGE_VEC01029	pep602_AHp007219_1158bp_sw	this study
CPGE_VEC00953	GB2047 (Cas9m4-TALx6-VP128)	Selma et al. (2019)
CPGE_VEC00932	pdgb3alpha1	Vazquez-Vilar et al. (2017)
CPGE_VEC00931	pUU017	Urquiza-Garcia <i>et al. in preparation</i>
CPGE_VEC00929	GB0223	Vazquez-Vilar et al. (2017)
CPGE_VEC00930	GB0037	Vazquez-Vilar et al. (2017)
CPGE_VEC01018	pUU385	Urquiza-Garcia <i>et al. in preparation</i>
CPGE_VEC01020	pUU385_PTG-cDOPA5GT	this study

Table S4.2: Primers used for cloning

Primer	Sequence (5'- 3')	Purpose
SV6	tggtctcaggctccggatccatgtttaatcat ttcatcatggataaatgcaaccc	amplification of <i>Ah-CYP76AD2</i> for cloning in greengate entry vector pGGC000
SV7	cggctctactgaggtaccatgcctaagggat ggaatgatttcaag	
CPGE_PRM00659	cgcgcgctacgaaaaac	backbone (pmz824) amplification for Gibson cloning of the transactivation plasmids
CPGE_PRM00660	ggtgtgtgtcatatgcgtcg	
CPGE_PRM00661	gatgctagccctatataagcagagct	backbone (pep602) amplification for Gibson cloning of the transactivation plasmids
CPGE_PRM00662	gtttaaacttcgaagctttttgcaaaagcc	
CPGE_PRM01190	cgcatatgacaacacaccatggagggttcgt cca	amplification of <i>AtPAP1</i> for Gibson cloning of the transactivation plasmids
CPGE_PRM01192	tttcgtacgcgcgatcaaatttcacagtct ctccatcgaaaag	
CPGE_PRM01193	caaaaagcttcgaagtttaaacttcaccaga at ttgctaacgcaatg	amplification of the <i>AtDFR</i> promoter for Gibson cloning of the transactivation plasmids
CPGE_PRM01194	gcttatatagggtagcatcgattgtgctttg ggaaagattcaactatgtatg	
CPGE_PRM01111	cgcatatgacaacacaccatgtaccagcaga atagtgaaactggt	amplification of <i>BvMYB1</i> for Gibson cloning of the transactivation plasmids
CPGE_PRM01113	tttcgtacgcgcgtgccacaaagttcacaa catcaaaat	
CPGE_PRM01187	caaaaagcttcgaagtttaaacgttggtcata taaggtaggaaaaaaaaaatcgtttgt	amplification of the <i>Bv-CYP76AD1</i> promoter for Gibson cloning of the transactivation plasmids
CPGE_PRM01189	gcttatatagggtagcatcgcatgtgatagt ggatcctagagaaagg	
CPGE_PRM00663	cgcatatgacaacacaccatgtacaaacat ttgtatccacactgaac	amplification of <i>AhMYB2</i> for Gibson cloning of the transactivation plasmids
CPGE_PRM00664	ttttcgtacgcgcgatactcctccctaa aattgacaacc	
CPGE_PRM00681	caaaaagcttcgaagtttaaactgattttga atacttttagatctagagcctaaaaatagagg tg	amplification of the <i>Ah-CYP76AD2</i> promoter for Gibson cloning of the transactivation plasmids
CPGE_PRM01196	gcttatatagggtagcatcttttttttaaat aaatagtactacaaattaataagaagaaa aaag	

CPGE_PRM01324	gaacaataacccgtggtccccgc	to construct pep602_sw using gibbon cloning
CPGE_PRM01325	gccgtgtaatgagcttggaatggatcttcgatcccg	
CPGE_PRM01326	tctagtcgaaatggaagacgcaaaaacataaagaaagg	
CPGE_PRM01327	ccaagctcattacacggcgatctttccgc	
CPGE_PRM01328	gtcttcatttcgactagaatagtaaattgtaatgttgttgttgttgtt	
CPGE_PRM01329	cgaagtcatgctagcgccagcgc	
CPGE_PRM01330	gaccacgggttattgttcatttttgagaactcgtcaacgaacgatttga	
CPGE_PRM01331	ggcgctagcatgacttcgaaagttatgatccagaacaaagga	
CPGE_PRM01050	ggtggaattcgatccgggac	with CPGE_PRM00664 amplify <i>AhMYB2.1</i> for gibbon cloning of the mammalian transactivation plasmids
CPGE_PRM01051	cggatcgaattccaccatgctacaaacattgtatccacactgaac	with CPGE_PRM00659 amplify the backbone (pJS174) for Gibson cloning of the mammalian transactivation plasmids
CPGE_PRM00682	gcttatatagggctagcatcagtagtgcattgacaatgtatggctatgtgttgattct	with CPGE_PRM00681 amplify part of the <i>AhCYP76AD2</i> promoter for Gibson cloning of the transactivation plasmids
PBC_PRM00183	ttgaagacttggaacaaagcaccagtggctagtggttag	to construct the PolyGEN guide cassette for <i>Ahc-DOPA5GT</i>
CPGE_PRM01162	ttgaagacttaaggctcactgcaccagccgggaatcg	
CPGE_PRM01163	ttgaagacttccttgaagattaatgggttttagagctagaaatagcaagttaa	
CPGE_PRM01164	ttgaagacttgaccggttctcacggtttgcaccagccgggaatcg	
CPGE_PRM01165	ttgaagacttggtcactgttttagagctagaaatagcaagttaa	
CPGE_PRM01166	ttgaagactttccaatgcaccagccgggaatcg	

CPGE_PRM01167	ttgaagactttggaaaatgaccaccattagtt ttagagctagaaatagcaagttaa	
PBC_PRM00184	ttgaagacttaaactgcaccagccgggaatc gaac	
CPGE_PRM00672	caaaaagcttcgaagtttaaactaatcaaga agatcaatttcgagcataaattaaaatgga	amplification of the <i>Ahc-DOPA5GT</i> promoter for Gibson cloning of the transactivation plasmid
CPGE_PRM00673	gcttatatagggctagcatcaattagtaatta atcattatgattttgaaatcactatcaaaaga t	

Table S4.3: Tested yeast-two-hybrid combinations. Genes in the pAS backbone have the binding domain (BD) and genes in the pACT backbone have the activation domain (AD). - indicate the empty pAS or pACT backbones included as additional controls to test, e.g., for autoactivation. Nr. 1 and 26 functioned as negative controls, and nr. 27 and 28 as positive controls.

Combination	pAS (BD)	pACT (AD)
1	-	-
2	-	<i>AhMYB2.1</i>
3	-	<i>AhMYB2.2</i>
4	-	<i>AhTT8</i>
5	-	<i>AhTTG1a</i>
6	<i>AhMYB2.1</i>	-
7	<i>AhMYB2.1</i>	<i>AhMYB2.1</i>
8	<i>AhMYB2.1</i>	<i>AhMYB2.2</i>
9	<i>AhMYB2.1</i>	<i>AhTT8</i>
10	<i>AhMYB2.1</i>	<i>AhTTG1a</i>
11	<i>AhMYB2.2</i>	-
12	<i>AhMYB2.2</i>	<i>AhMYB2.1</i>
13	<i>AhMYB2.2</i>	<i>AhMYB2.2</i>
14	<i>AhMYB2.2</i>	<i>AhTT8</i>
15	<i>AhMYB2.2</i>	<i>AhTTG1a</i>
16	<i>AhTT8</i>	-
17	<i>AhTT8</i>	<i>AhMYB2.1</i>
18	<i>AhTT8</i>	<i>AhMYB2.2</i>
19	<i>AhTT8</i>	<i>AhTT8</i>
20	<i>AhTT8</i>	<i>AhTTG1a</i>
21	<i>AhTTG1a</i>	-
22	<i>AhTTG1a</i>	<i>AhMYB2.1</i>
23	<i>AhTTG1a</i>	<i>AhMYB2.2</i>
24	<i>AhTTG1a</i>	<i>AhTT8</i>
25	<i>AhTTG1a</i>	<i>AhTTG1a</i>
26	<i>AtTTG1</i>	<i>AtTTG1</i>
27	<i>AtTTG1</i>	<i>AtTT8</i>
28	<i>AtTT8</i>	<i>AtTTG1</i>

29	<i>AtTT8</i>	<i>AtTT8</i>
30	<i>AhMYB2.1</i>	<i>AtTTG1</i>
31	<i>AhMYB2.1</i>	<i>AtTT8</i>
32	<i>AhMYB2.1</i>	<i>AtGL3</i>
33	<i>AhMYB2.1</i>	<i>AtEGL3</i>
34	<i>AhMYB2.1</i>	<i>AtMYC1</i>
35	<i>AhMYB2.2</i>	<i>AtTTG1</i>
36	<i>AhMYB2.2</i>	<i>AtTT8</i>
37	<i>AhMYB2.2</i>	<i>AtGL3</i>
38	<i>AhMYB2.2</i>	<i>AtEGL3</i>
39	<i>AhMYB2.2</i>	<i>AtMYC1</i>
40	<i>AtTTG1</i>	<i>AhMYB2.1</i>
41	<i>AtTTG1</i>	<i>AhMYB2.2</i>
42	<i>AtTT8</i>	<i>AhMYB2.1</i>
43	<i>AtTT8</i>	<i>AhMYB2.2</i>
44	<i>AtGL3</i>	<i>AhMYB2.1</i>
45	<i>AtGL3</i>	<i>AhMYB2.2</i>
46	<i>AtEGL3</i>	<i>AhMYB2.1</i>
47	<i>AtEGL3</i>	<i>AhMYB2.2</i>
48	<i>AtMYC1</i>	<i>AhMYB2.1</i>
49	<i>AtMYC1</i>	<i>AhMYB2.2</i>

Table S4.4: Information about the promoters used for protoplast transactivation assays. Start and end position each refers to the number of base pairs upstream of the ATG, e.g., the *AhCYP76AD2* promoter fragment includes the 2125 bp upstream of the ATG, while for the *BvCYP76AD1* promoter fragment, the 112 bp directly upstream of the ATG were excluded.

Gene name	Species	Accession	length	start	end	Plasmid
<i>AhCYP76AD2</i>	<i>Amaranthus hypochondriacus</i>	PI558499	2125	0	2125	CPGE_VEC01024
<i>AhCYP76AD2</i>	<i>Amaranthus hypochondriacus</i>	PI558499	1966	159	2125	CPGE_VEC00527
<i>AhcDOPA5GT</i>	<i>Amaranthus hypochondriacus</i>	PI558499	1158	257	1415	CPGE_VEC01029
<i>BvCYP76AD1</i>	<i>Beta vulgaris</i>	Bulls Blood	551	112	663	CPGE_VEC00977
<i>AtDFR</i>	<i>Arabidopsis thaliana</i>	Col-0	1132	18	1150	CPGE_VEC00980

Table S4.5: Sources of the used PAP1 homolog sequences from different species. Associated gene IDs or NCBI database identifiers are listed in square or round brackets, respectively.

Gene name	Species	Source
<i>AhMYB2</i>	<i>Amaranthus hypochondriacus</i>	Winkler et al. (2024) [AHp022773.1]
<i>BvMYB1</i>	<i>Beta vulgaris</i>	Hatlestad et al. (2015) (JF432080)
<i>AmMYB1</i>	<i>Amaranthus tricolor</i>	Liyang et al. (KU557504.1)
<i>PgPAP</i>	<i>Portulaca grandiflora</i>	Sakuta et al. (2021)
<i>BaPAP</i>	<i>Basella alba</i>	Sakuta et al. (2021)
<i>MzPAP</i>	<i>Mammillaria zeilmanniana</i>	Sakuta et al. (2021)
<i>DcaPAP.1</i>	<i>Dianthus caryophyllus</i>	one of the two proteins with "LPGRTA" in the aminoacid sequence from the genome assembly from Yagi et al. (2014) [Dca7400.1]
<i>DcaPAP.2</i>	<i>Dianthus caryophyllus</i>	one of the two proteins with "LPGRTA" in the aminoacid sequence from the genome assembly from Yagi et al. (2014) [Dca21863.1]
<i>VvMYBA1</i>	<i>Vitis vinifera</i>	Deluc et al. (2008)
<i>MdMYB1</i>	<i>Malus domestica</i>	(NP_001315777.1)
<i>AtPAP1</i>	<i>Arabidopsis thaliana</i>	TAIR [AT1G56650]
<i>AhMYBL1</i>	<i>Amaranthus hypochondriacus</i>	Graf et al. (2025) [AHq015545.1]
<i>MdMYB9</i>	<i>Malus sieversii</i>	Wang et al. (2018)
<i>AtTT2</i>	<i>Arabidopsis thaliana</i>	TAIR [AT5G35550]

Table S4.6: Identified amaranth anthocyanin pathway genes and the conserved amino acid sequences used for identification. An underscore in the amino acid sequence denotes a not (highly) conserved position. PAL (phenylalanine ammonium lyase), C4H (cinnamate 4-hydroxylase), 4CL (4-coumarate:CoA ligase), CHS (naringenin-chalcone synthase), CHI (chalcone isomerase), FNS (flavone synthase), F3H (flavanone 3-hydroxylase), FLS (flavonol synthase), F3'H (flavonoid 3'-hydroxylase), F3'5'H (flavonoid 3'5'-hydroxylase), DFR (dihydroflavonol 4-reductase), LAR (leucoanthocyanidin reductase), ANS (anthocyanidin synthase), ANR (anthocyanidin reductase).

Gene name	Amaranth gene ID	conserved sequence
<i>PAL</i>	AHq023130 AHq013529	ASGDLVPL
<i>C4H</i>	AHq013989 AHq023561 AHq023562	RMAIPLL
<i>4CL</i>	AHq015354 AHq022113	SSGTTGLPK
<i>CHS</i>	AHq004516 AHq005467	GGTVLRL
<i>CHI1</i>	AHq010547	LAGAG_RGL
<i>F3'H</i>	AHq018419 AHq013437	IPFGAGRR_CAG
<i>DFR</i>	AHq009850	GRY_S____TI
<i>F3H</i>	AHq006686	SRLSIATF
<i>FLS</i>	AHq009470	RMSW_VF
<i>LAR</i>	AHq018329	ICCNSIA
<i>ANS</i>	AHq022939	ALGVEAHTD
<i>FNS1</i>	AHq012652	GQHMAVNYYYP

Table S4.7: Identified amaranth aragonate dehydratase (ADT) and aragonate dehydrogenase (ADH α) genes and the conserved amino acid sequences used for identification.

Gene name	Amaranth gene ID	conserved sequence
<i>ADHa</i>	AHq014860 AHq021764	GFGNFGQF
<i>ADR</i>	AHq000099 AHq000240 AHq000371 AHq001397 AHq001604 AHq022626	I_L_KIESR

4.9.3 Appendix

Appendix S1: DNA sequences of the promoters used for protoplast transactivation assays. Promoter fragment of *AtDFR* was based on the primer position from Guo et al. (2023) (forward primer) and He et al. (2021)(reverse primer).

AhCYP76AD2:

TGATTTTGAATACTTTAGATCTAGAGCCTAAAAATAGAGGTGGGCGTTTAA
TAGAGTCTTAGACCCATTATCATCTTTGAGTGTATGTCATGGTCAAAATAA
AAAAAGAAATAATTCAATGGGATGAATTAAGGAAAAATAAGCTAATTC
ATAGTTATGTACAAAGAATGATTGACATTATTGATGATTAAtgattataatgtagaaa
tagctttttctagggcatggcggccctatatatatttggggccctgggcgaaatagtacaataggcccttttccttttctttatc
atttataatctatgtggacgttctccaactattaatttggctttttgacatgggaccttgagcgacagccctcttggccaccct
tagggctggctctgGACTTCTAGGCTATCAAAGAAAAAATTAGACAAAATGCACAAT
AGCAAATAACGTGGCCAAATCTTCCTTCACCCACTTCTTCTCTTCCAAAACA
AATGTATGACACTCATCTTTTAATGTCCCTCAACCGTTGCTCTCGAATCTA
ACATAGAGGAACTTTCCCATTAGTTTAGACTTTTTAGAGTAGACCAATTGCC
AAATTCTGTAGACGGGTGTAATTATCTACGTTTGTTGCCTTGTAATGATCG
TAATTTTAACTATATGCTTGTCACCGTTTCTTTCCAAAAAATAACTAATCTC
AAAAATCTATTGATATAACTATAAAAAACATTTCAAATTAAATTCCAAAATA
TTGAAACTCAAATTGAATAATTAATGTGCTAAAATTGCCTAAAAATTATATT
TGATTTTATATAAAAGATTAAACCCAAACAATAGGAATATTGGCAAAAACCTT
TatacttccttcgtttcataaacttgctactgagtgtgatatgttgatcgaaattgtcaatatcagtaAAACGGAGA
AAATAAGAACGAAATAAGTCAGCATTGCTTCAATCATTAACCCAACTGCCA
TGTAACCCCATGTCTGAATATCCTTCATCGGTTGCCTCTTAATCTCATACTTC
ATAATCAATAAAGCACCGACAAATTCTATatcatacattacacaagaagcttgaaaattttaagt
gtcacaatttaagagagtttgtcaactagaacttatttataggetatagaaaatgactgtcaattgatcttgatcaattttc
taccattaaaatatctttgacaaacatttatttatagtcaaattaagtaaataagataataactgcagaatatattgttgata
atttctaaaaatagagttaattttatacataaaaactatcaaaattgagtgttttactgcaaaatatcactgatttttttctta
atcatatgtttatatatatatagatactaaaaaaccgtgcgttgcaacggatttctacactaCTACTAGTATTAAA
TTGAACACTATTTGTTTCTCGATCGGTGTTTCGTTATAATATCTCTAAAAAAC
TAATACAATTTAATATTGATGAATTTTTTAACCAAGTGATTTCGAGTCTCGCCC
AGCCTCTCCAGGTCAAAATCAATGGTGTGTCATGATACAATGCGACATA
ACTTACAGGTCATCTCAAGCCGAGCTATATGGGTTTACGGGTCATGGCTTG
CTTGACGGGTCATGTCTAAGAACtagttgtgagacaatctctattagactgacataatagattacaatc
ttaaagtaatcaatttttacaatttttaataatcacttttatagctcttaaaatgatcgcttataaTAAATAACATTA
CATCATTTTTCGCTGTTTTTCTCTTTCTTGACTTTTATTTTTTGATGGACCCC
ACTATAGATTTTATAATTAAATTATCAAAAATATCCAGTATCTACACACTGC
AGTGTACACCAAATTTATTATCTTCATATAACACTCTAGAATCAACACATAG

CCATACATTGTCCATGCACTACTCCTATATAAAAAGACTAAATGAGCCTAAA
TAAGCACACATCAACAGTTCTTTTTTCTATATATATCGCCTCATAACTTCTA
TAATTTCTTATATATTATTTAAATGCTTTACTTTTTTCTTCTTATTTAATT
TGTAAGTACTATTTATTTAAAAA

AhCYP76AD2_1966 bp:

TGATTTTGAATACTTTAGATCTAGAGCCTAAAAATAGAGGTGGGCGTTTAA
TAGAGTCTTAGACCCATTATCATCTTTGAGTGTATGTCATGGTCAAAATAA
AAAAAGAAATAATTCAATGGGATGAATTTAAAGGAAAAATAAGCTAATTC
ATAGTTATGTACAAAGAATGATTGACATTATTGATGATTAAtgattataatgtagaaa
tagctttttctagggcatggccggccctatatattttggggccctgggcgaaatagtacaatagggcccttttccttttcctttatc
atttataatctatgtggacgttctccaactattaattttgctttttgacatgggaccttgagcgacagcccccttggccaccct
tagggctggctctgGACTTCTAGGCTATCAAAGAAAAAATTAGACAAAATGCACAAT
AGCAAATAACGTGGCCAAATCTTCCTTCACCCACTTCTTCTCTTCCAAAACA
AATGTATGACACTCATCTTTTAATGTCCCCTCAACCGTTGCTCTCGAATCTA
ACATAGAGGAACCTTCCCATTAGTTTAGACTTTTTAGAGTAGACCAATTGCC
AAATTCTGTAGACGGGTGTAATTATCTACGTTTGTTGCCTTGTAATGATCG
TAATTTTAACTATATGCTTGTACCGTTTCTTTCCAAAAAATAACTAATCTC
AAAAATCTATTGATATAACTATAAAAAACATTTCAAATTAATTTCCAAAATA
TTGAAACTCAAATTGAATAATTAATGTGCTAAAATTGCCTAAAAATTATATT
TGATTTATATAAAAGATTAAACCCAAACAATAGGAATATTGGCAAAAACCT
TatacttccttcgtttcataacttgctactgagtgtgatatggtgatcgaaattgtcaatatcagtaAAACGGAGA
AAATAAGAACGAAATAAGTCAGCATTGCTTCAATCATTAACCCAACTGCCA
TGTAACCCCATGTCTGAATATCCTTCATCGGTTGCCTCTTAATCTCATACTTC
ATAATCAATAAAGCACCGACAAATTCTATatcacattacacaagaagcttgaaaaatttaagt
gtcacaatttaagagagtttgtcaactagaacttatttataggctatagaaaaatgacttgtcaattgatcttgatcaatttttc
taccattaaaatatctttgacaaacatttatttatagtcataaataagataataactgcagaatatattgttgata
atttctaaaaatagagtttaatttatacataaaaactatcaaaattgagtgtttctactgcaaaatatcactgatttttttcttaat
atcatatgtttatatatatagatactaaaaaacgtgcgttgcaacggatttctacactaCTACTAGTATTAAA
TTGAACACTATTTGTTTCTCGATCGGTGTTTCGTTATAATATCTCTAAAAAAC
TAATACAATTTAATATTGATGAATTTTAAACCAAGTGATTTCGAGTCTCGCCC
AGCCTCTCCAGGTCAAAATCAATGGTGTGTCATTGATACAATGCGACATA
ACTTACAGGTCATCTCAAGCCGAGCTATATGGGTTTACGGGTCATGGCTTG
CTTGACGGGTCATGTCTAAGAACTagttgtgagacaatctctattagactgacataatagattacaatc
ttaaagtaatacaatttttacaatttttaataatcacttttatagtccttaaaatgatcgcttataaTAAATAACATTA
CATCATTTTGGCGTGTTTTTCTCTTTCTTGACTTTTATTTTTTGATGGACCCC
ACTATAGATTTTATAATTAAATTATCAAAAATATCCAGTATCTACACACTGC
AGTGTACACCAAATTTATTATCTTCATATAACACTCTAGAATCAACACATAG
CCATACATTGTCCATGCACTACT

AhcDOPA5GT:

TAATCAAGAAGATCAATTTTCGAGCATAAATTAAAATGGATTCCTTCAAGAA
GATCAATTTCTTCTTGTGGTTGCTTTTACTTGATTCTGATGGTTTGATATTT
TTGCATTTAACAGTCTTCATACCAAGGAAAAAATGGAAGAGAGAAAAATAA
ACTCTCTTCATGAATGGAAGACAGAGAACATAAGAGAGAGATGAATCGTCC
GAAGAACAATTTGAGAAACCTAAAATTTTCCCAGAACATTTCCCCCAGATA
CAATTGGGTACTACATCACACTATTAGTTAACTTTTTATGGGCTTAATGATT
TCCATCTAATCCAAACATTCAACTCAAGCTGTATTATTCAAGCAATTTTAAG
GAGTAAACAATTTGAGAAACCTAAAATTTGTAAAGAATTCTAAAATCAAAC
ACTGCAAAAAGTTGACTAAGAGTAAAGAAGTCGCCTCAATTCCTTGAAAC
ATCACCTTAAATAGCACTAACACTATAAAAAAATGCAGCATTAAAGTTCTT
TTACATTTTTCAAATATTACTGCCTTTTTTATGAATTTACTACCTTATGTAA
TTTCGGAAAAGTTTTTCCTTCGTAATCCAAACTTATGATTTTGCCATATTT
GACTGTGATTCAGCAGTTTCTCTCCTTCTTTCTTGAATTTCTGGGTTATG
GCTTGATTTGGGAATCTCTCTGGGTatttggtattctttgattagagttcttgaaaagggaagaaa
tctagagtttgacttttagagagtttggtctttgaagattaatggtgactttgaagattaatggaggatgaagaggattaag
aggatgaagaagaaggaggaagaggataaagatgtgaagaggatgaagaaattaaattaaaaaaggaaaaagaaaaa
aaactcaaaaaaaaaaagtaaaatcaggaaaaaagaaagtaaaagtcattttgctggtcaacggtgagaaccggtcac
ttggccgtaatggtggtcattttcaaaagttggtggctgaattcataaaaaaaaaacttgcggtgtaattcacaaaacgac
caaactttgaggctgtaattcacaaaaattcttttattatttttAATCTTTTGATAGTGATTTCAAAA
TCATAATGATTAATTACTAATT

BvCYP76AD1:

GTTGTTTCATATAAGGTTAGGAAAAAAAAAATCGTTTGTTTGTTTTTACCAA
GGACCTTGGGTTCAATGGTATGACTTCTTATCTCCCACACGAGAAACCAAA
GTTCAAATCTTAATGGAAGCAAAATTTCAATATTCTTGGGAAGAAGGGGGT
TGAGACTTGTTGGGCCTCTGGGTAAGCAGCCCAAAGAAAATTCATCTTATC
AGGGGTTTTTCGAATGGGGCCAAAAAGAATAAAAAAATACAAGGTTAGGAA
AAAAAAGGTGGTGAAGCACCTCATTGGAGAAATAGGAAGAACTTACATATA
CAAAGGAATAAATAGTAGCAAGCAATAAAAAAATTTGGAATTAAAAAAA
TTGGTAGTATATAAATTATGAGAGAGGGATTGTGATTTTTTACATCAGCACT
ACTATGTGCAAATGCATAAATTTAATTCAGATACTCTTTTTAAAAATATTAT
ACTGTCCATTCTTTTAACCCCCCAAAAAAAAAAAAAAAAAAACTGTGCGTAATAC
ACCATTTAATTCCTTTCTCTAGGATCCACTATCACATGC

AtDFR:

ttcaccagaattttgctaacgcaatgcaattgaatagagaacgaagatttctagtgggtgaagaagaagaaggagagctttg
aagtattcttctctctctcgctcgagctcaaacgtgggtggactacaccaagacgcttggttactttgtctctctgtttg
aggagagtcaaaattaagctgacctcttctctacgtcttacgacacaacaaattgaaaataaaagagttttattgattctgtgt
tttagaatccacgtggacgaggaaccaccacgtgttattcttactttatgagattaagtgattcactgtccttctaattttatt

ttattttctttgcgaaaactggaaataaataaatttttgaaaacaatatTTTtaacatatcaattatttttataataattcgtgaa
ataataatgaaaaagctgacatgggacacaattggataattacatcttttaaagttttagtttatgtattttatgtaaatgctta
taatatttagattgcaataaatctagaagtcataattataaattaatatgttttataattttacaattttgtaataagaaactcct
aattcataaatctaaaaatataaacatatTTTtcatttaagcttttcataatttataattatttttaggtgtctgatttttagattt
caattaaaattaaaaatattattttaagtaaaatgtatttctgtatatattctatcaaaatgtagttttagacaaaatttgat
ttatttcgtaaaagtgggtggggaacaaaaacaaaaacaaactgaactgaagtcacccacacgtctaccaaaacaaatcgaa
gtcaacgtatttcacccaccggtacaacaacaaaatacacacctaaggaaataaaaaatcaacttaccagattgttacgtac
cacacatctcttagtccttcgtcaaccaacgttccccacgtgcttctccggttggtactcacgtgaccggcagcttctcgttctt
attatctgttttctcaataacgattcataatctctagtgcttatttataatgtcttcacatcacaagatttgtaccgaacata
catagttgaatctttcccaaagcacaaac

4.10 Supplementary references

- J. An, X. Zhang, S. Bi, C. You, X. Wang, and Y. Hao. The ERF transcription factor MdERF38 promotes drought stress-induced anthocyanin biosynthesis in apple. *The Plant Journal*, 101(3):573–589, Feb. 2020. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.14555. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.14555>.
- L. Deluc, J. Bogs, A. R. Walker, T. Ferrier, A. Decendit, J.-M. Merillon, S. P. Robinson, and F. Barrieu. The Transcription Factor VvMYB5b Contributes to the Regulation of Anthocyanin and Proanthocyanidin Biosynthesis in Developing Grape Berries. *Plant Physiology*, 147(4):2041–2053, Aug. 2008. ISSN 1532-2548. doi: 10.1104/pp.108.118919. URL <https://academic.oup.com/plphys/article/147/4/2041/6107699>.
- C. Graf, T. S. Winkler, P. J. Maughan, and M. G. Stetter. Domestication shaped the chromatin landscape of grain amaranth. *Nature Communications*, 16(1):10407, Nov. 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-66445-w. URL <https://www.nature.com/articles/s41467-025-66445-w>.
- Y. Guo, D. Li, T. Liu, Y. Li, J. Liu, M. He, X. Cui, Z. Liu, and M. Chen. Genome-Wide Identification of PAP1 Direct Targets in Regulating Seed Anthocyanin Biosynthesis in *Arabidopsis*. *International Journal of Molecular Sciences*, 24(22):16049, Nov. 2023. ISSN 1422-0067. doi: 10.3390/ijms242216049. URL <https://www.mdpi.com/1422-0067/24/22/16049>.
- J. W. Harper, G. Adami, N. Wei, K. Keyomarsi, and S. J. Elledge. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell*, 75(4):805–816, Nov. 1993. ISSN 00928674. doi: 10.1016/0092-8674(93)90499-G. URL <https://linkinghub.elsevier.com/retrieve/pii/009286749390499G>.
- G. J. Hatlestad, N. A. Akhavan, R. M. Sunnadeniya, L. Elam, S. Cargile, A. Hembd, A. Gonzalez, J. M. McGrath, and A. M. Lloyd. The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway. *Nature Genetics*, 47(1):92–96, Jan. 2015. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3163. URL <http://www.nature.com/articles/ng.3163>.
- Y. He, X. Zhang, L. Li, Z. Sun, J. Li, X. Chen, and G. Hong. SPX4 interacts with both PHR1 and PAP1 to regulate critical steps in phosphorus-status-dependent anthocyanin biosynthesis. *New Phytologist*, 230(1):205–217, Apr. 2021. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.17139. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.17139>.
- R. P. Hellens, E. A. Edwards, N. R. Leyland, S. Bean, and P. M. Mullineaux. pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Molecular Biology*, 42(6):819–832, Apr. 2000. ISSN 0167-4412, 1573-5028. doi: 10.1023/A:1006496308160. URL <https://link.springer.com/10.1023/A:1006496308160>.
- T. Koyama and F. Sato. The function of *ETHYLENE RESPONSE FACTOR* genes in the light-induced anthocyanin production of *Arabidopsis thaliana* leaves. *Plant Biotechnology*, 35(1):87–91, 2018. ISSN 1342-4580, 1347-6114. doi: 10.5511/plantbiotechnology.18.0122b. URL <https://www.jstage.jst.go.jp/article/plantbiotechnology/35/>

- A. Lampropoulos, Z. Sutikovic, C. Wenzl, I. Maegele, J. U. Lohmann, and J. Forner. GreenGate - A Novel, Versatile, and Efficient Cloning System for Plant Transgenesis. *PLoS ONE*, 8(12):e83043, Dec. 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0083043. URL <https://dx.plos.org/10.1371/journal.pone.0083043>.
- X. Liyang, L. Shengcai, B. Yu, L. Yan, L. Mi, C. Shunliang, Z. Xueli, X. Xiaoqing, F. Xin, C. Chunzhen, C. Yukun, and L. Zhongxiong. Cloning and Expression Analysis of Betalain-related Transcription Factor Gene AmMYB1 in *Amaranthus tricolor* L. *Acta Botanica Boreali-Occidentalia Sinica*, 36(6):1080–1090. URL xbzwxben/article/abstract/20160602.
- K. Müller, D. Siegel, F. Rodriguez Jahnke, K. Gerrero, S. Wend, E. L. Decker, R. Reski, W. Weber, and M. D. Zurbriggen. A red light-controlled synthetic gene expression switch for plant systems. *Mol. Biosyst.*, 10(7):1679–1688, 2014. ISSN 1742-206X, 1742-2051. doi: 10.1039/C3MB70579J. URL <http://xlink.rsc.org/?DOI=C3MB70579J>.
- J. Ni, S. Bai, Y. Zhao, M. Qian, R. Tao, L. Yin, L. Gao, and Y. Teng. Ethylene response factors Pp4ERF24 and Pp12ERF96 regulate blue light-induced anthocyanin biosynthesis in ‘Red Zaosu’ pear fruits by interacting with MYB114. *Plant Molecular Biology*, 99(1-2):67–78, Jan. 2019. ISSN 0167-4412, 1573-5028. doi: 10.1007/s11103-018-0802-1. URL <http://link.springer.com/10.1007/s11103-018-0802-1>.
- M. Sakuta, A. Tanaka, K. Iwase, M. Miyasaka, S. Ichiki, M. Hatai, Y. T. Inoue, A. Yamagami, T. Nakano, K. Yoshida, and S. Shimada. Anthocyanin synthesis potential in betalain-producing *Caryophyllales* plants. *Journal of Plant Research*, 134(6):1335–1349, Nov. 2021. ISSN 0918-9440, 1618-0860. doi: 10.1007/s10265-021-01341-0. URL <https://link.springer.com/10.1007/s10265-021-01341-0>.
- S. Selma, J. M. Bernabé-Orts, M. Vazquez-Vilar, B. Diego-Martin, M. Ajenjo, V. Garcia-Carpintero, A. Granell, and D. Orzaez. Strong gene activation in plants with genome-wide specificity using a new orthogonal CRISPR/Cas9-based programmable transcriptional activator. *Plant Biotechnology Journal*, 17(9):1703–1705, Sept. 2019. ISSN 1467-7644, 1467-7652. doi: 10.1111/pbi.13138. URL <https://onlinelibrary.wiley.com/doi/10.1111/pbi.13138>.
- M. Vazquez-Vilar, A. Quijano-Rubio, A. Fernandez-del Carmen, A. Sarrion-Perdigones, R. Ochoa-Fernandez, P. Ziarsolo, J. Blanca, A. Granell, and D. Orzaez. GB3.0: a platform for plant bio-design that connects functional DNA elements with associated biological data. *Nucleic Acids Research*, page gkw1326, Jan. 2017. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkw1326. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw1326>.
- S. K. Vollmer, M. G. Stetter, and G. Hensel. First Successful Targeted Mutagenesis Using CRISPR/Cas9 in Stably Transformed Grain Amaranth Tissue. *Plant Biotechnology Journal*, page pbi.70590, Feb. 2026. ISSN 1467-7644, 1467-7652. doi: 10.1111/pbi.70590. URL <https://onlinelibrary.wiley.com/doi/10.1111/pbi.70590>.
- N. Wang, C. Qu, S. Jiang, Z. Chen, H. Xu, H. Fang, M. Su, J. Zhang, Y. Wang, W. Liu, Z. Zhang, N. Lu, and X. Chen. The proanthocyanidin-specific transcription factor

- MdMYBPA1 initiates anthocyanin synthesis under low-temperature conditions in red-fleshed apples. *The Plant Journal*, 96(1):39–55, Oct. 2018. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.14013. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.14013>.
- T. S. Winkler, S. K. Vollmer, N. Dyballa-Rukes, S. Metzger, and M. G. Stetter. Isoform-resolved genome annotation enables mapping of tissue-specific betalain regulation in amaranth. *New Phytologist*, page nph.19736, Apr. 2024. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.19736. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19736>.
- M. Yagi, S. Kosugi, H. Hirakawa, A. Ohmiya, K. Tanase, T. Harada, K. Kishimoto, M. Nakayama, K. Ichimura, T. Onozaki, H. Yamaguchi, N. Sasaki, T. Miyahara, Y. Nishizaki, Y. Ozeki, N. Nakamura, T. Suzuki, Y. Tanaka, S. Sato, K. Shirasawa, S. Isobe, Y. Miyamura, A. Watanabe, S. Nakayama, Y. Kishida, M. Kohara, and S. Tabata. Sequence Analysis of the Genome of Carnation (*Dianthus caryophyllus* L.). *DNA Research*, 21(3):231–241, June 2014. ISSN 1340-2838, 1756-1663. doi: 10.1093/dnares/dst053. URL <https://academic.oup.com/dnares/article-lookup/doi/10.1093/dnares/dst053>.
- G. Yao, M. Ming, A. C. Allan, C. Gu, L. Li, X. Wu, R. Wang, Y. Chang, K. Qi, S. Zhang, and J. Wu. Map-based cloning of the pear gene *MYB114* identifies an interaction with other transcription factors to coordinately regulate fruit anthocyanin biosynthesis. *The Plant Journal*, 92(3):437–451, Nov. 2017. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.13666. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.13666>.
- B. Zhang. *Evolutionary Analysis of MYBs-bHLH-WD40 Protein Complexes Formation and Their Functional Relationship in Planta*. PhD thesis, Universität zu Köln, May 2018.
- J. Zhang, H. Xu, N. Wang, S. Jiang, H. Fang, Z. Zhang, G. Yang, Y. Wang, M. Su, L. Xu, and X. Chen. The ethylene response factor MdERF1B regulates anthocyanin and proanthocyanidin biosynthesis in apple. *Plant Molecular Biology*, 98(3):205–218, Oct. 2018. ISSN 0167-4412, 1573-5028. doi: 10.1007/s11103-018-0770-5. URL <http://link.springer.com/10.1007/s11103-018-0770-5>.
- C. Zhao, X. Liu, Q. Gong, J. Cao, W. Shen, X. Yin, D. Grierson, B. Zhang, C. Xu, X. Li, K. Chen, and C. Sun. Three AP2/ERF family members modulate flavonoid synthesis by regulating type IV chalcone isomerase in citrus. *Plant Biotechnology Journal*, 19(4):671–688, Apr. 2021. ISSN 1467-7644, 1467-7652. doi: 10.1111/pbi.13494. URL <https://onlinelibrary.wiley.com/doi/10.1111/pbi.13494>.
- M.-G. Zhuo, T.-Y. Wang, X.-M. Huang, G.-B. Hu, B.-Y. Zhou, H.-C. Wang, and F. Abbas. ERF transcription factors govern anthocyanin biosynthesis in litchi pericarp by modulating the expression of anthocyanin biosynthesis genes. *Scientia Horticulturae*, 337:113464, Nov. 2024. ISSN 03044238. doi: 10.1016/j.scienta.2024.113464. URL <https://linkinghub.elsevier.com/retrieve/pii/S0304423824006216>.

Chapter 5

Overarching Discussion

Evolution is the driving force, influencing all organisms fate on earth, from the first cell to the complex organisms found today. The complex and highly diverse networks of plant secondary metabolism represent a stunning demonstration of the abilities of evolution yet unmatched by human engineering. Deciphering secondary metabolism pathways and reconstructing their evolution enable insights into general evolutionary principles. The power of this approach has been documented in many publications (Dixon and Dickinson, 2024; Larter et al., 2018; Mughal and Caetano-Anollés, 2019; Wheeler et al., 2022). However, most research has focused on the evolution of whole pathways and their enzymes, while the evolution of the associated regulation is still less understood (Voordeckers et al., 2015) and has just recently come into focus (Peng et al., 2025; Jones and Vandepoele, 2020). Considering the high importance of regulatory elements and transcription factors for breeding, having a more profound understanding of regulatory evolutionary principles has a high impact (Leong et al., 2025). This study focuses on reconstructing the regulation of the *de novo* and repeatedly evolved betalain pathway in the grain amaranth *A. hypochondriacus*. Betalains are chemically different but functionally similar pigments that replaced the otherwise ubiquitous anthocyanins in most families of the *Caryophyllales* (Brockington et al., 2011; Timoneda et al., 2019; Jain and Gould, 2015; Lopez-Nieves et al., 2018; Sunnadeniya et al., 2016). Their higher pH stability and health benefits, e.g., antioxidant properties, make them of interest to the food industry (Azeredo, 2009). Since around 10 years ago, the core biochemical pathway, including the two main enzymes CYP76AD and DODA1 as well as several glyceroltransferases, has been identified (Polturak and Aharoni, 2018). In contrast, the regulation of betalains is mostly unknown (Polturak and Aharoni, 2018)(see section 1.4). Following the current hypothesis, was a defeedback regulation of the aragonate dehydrogenase at the base of *Caryophyllales* key (Lopez-Nieves et al., 2018), while the enzymes evolved convergently four times in the order (Sheehan et al., 2020; Pucker et al., 2024). Deciphering the molecular betalain regulation in grain amaranth could pave the way to understanding how the betalain regulation evolved. By comparing it to the regulation of the other betalain origins, it might provide a better understanding of how generally new regulatory networks evolve.

We identified and verified AhMYB2 as the first positive transcriptional regulator in *A. hypochondriacus* homologous to BvMYB1 (Hatlestad et al., 2015) and AtrMYB72 (Xue et al., 2025) in *B. vulgaris* and *A. tricolor* (Chapter 3). We further demonstrated that AhMYB2 activates the betalain pathway by upregulating the expression of the first two enzymes of the pathway, AhCYP76AD2 and AhDODA α 1 (Chapter 4 section 4.3.4, Figure 4.6). While this was also shown for BvMYB1 and AtrMYB72 by RT-PCR (Hatlestad et al., 2015; Xue et al., 2025), is our RNA-seq data more comprehensive. Moreover, does our RNA-seq data of transgenic callus tissue overexpressing *AhCYP76AD2* reveal that AhDODA α 1 is not upregulated by the higher expression of AhCYP76AD2, but

that AhMYB2 is upregulating both genes (Chapter 4, section 4.3.4, Figure 4.6). Since direct binding was shown for BvMYB1 using Y1H ([Hatlestad et al., 2015](#)) and in a transactivation assay for AtrMYB72 ([Xue et al., 2025](#)), we expected to observe binding of AhMYB2 to *AhCYP76AD2*. However, we were unable to prove direct binding of AhMYB2 to both promoters by transactivation assays in protoplast (Chapter 4 section 4.3.3). As discussed in more detail in section 4.4.3 we hypothesize that a lack of visible binding in our transactivation data is caused by the low signal strength in amaranth protoplasts and a weak binding strength by AhMYB2 alone. It is well known that TFs, including MYBs often function in complexes ([Zhang et al., 2025](#); [Mariyam et al., 2025](#)). AtPAP1, an *A. thaliana* homolog of AhMYB2, can't bind its target genes alone but requires its partner bHLH and WD40 transcription factors ([Zimmermann et al., 2004](#)). Thus, it is highly possible that AhMYB2 also requires an additional TF for efficient binding and that alone the binding is too weak to be detected in our transactivation assay. This is supported by our transactivation data for BvMYB1, showing that binding was only found in betalain-producing *A. hypochondriacus* but not in the anthocyanin-producing *A. thaliana* protoplasts (Chapter 4, Figure 4.3). We tested if AhMYB2 can interact with bHLH TFs and the WD40 TF TTG1 homologous to AtPAP1 but could not find any interaction in Y2H (Chapter 4, Figure 4.1 and 4.1). Hence, suggesting that during co-option of AtPAP1 for the betalain regulation, new betalain-specific co-regulators have evolved, while the interaction with the anthocyanin-regulating co-regulators has been lost. This is an important update to the previous model that proposed that AtPAP1 homologs like BvMYB1 are regulating the pathway without co-regulators ([Hatlestad et al., 2015](#)). Our RNA-seq data set in combination with the conducted literature review for AtPAP1 homolog interacting TF across the angiosperms is a valuable starting point to identify possible betalain co-regulators of PAP1 homologs like AhMYB2 (Chapter 4).

To conclude, our study found AhMYB2 as a positive transcriptional regulator of the betalain regulation in grain amaranth, in line with previously identified MYB transcription factors in the *Amaranthaceae*. However, while previous studies focused on the identification of the TF, the studies performed in the frame of this thesis go further and provide important insights into the molecular mode of action. Furthermore, so far the investigation has been limited to the tissue-specific regulation of betalains, but the comprehensive literature review in the introduction evinces the complex regulation of betalains by diverse internal and external cues (1.3). Albeit not the focus of the studies, we found that AtMYB2 can override the repression of the betalain pathway by dark and high auxin concentration or, rather, activate the pathway independent of light and phytohormones (Chapter 4). Our RNA-seq data also show that light- and phytohormone-mediated activation of the betalain pathway acts via the transcriptional upregulation of AhCYP76AD2 and AhDODA α 1, too, but differently to the anthocyanin regulation, independently of AhMYB2 (Chapter 4, Figure 4.6). Regarding the high similarity of the regulatory pattern of anthocyanin and betalain regulation, it could be speculated that

many signaling pathways involved in the anthocyanin regulation had been co-opted for the betalain regulation. It will be of high interest to investigate which signaling pathways are involved in the regulation of betalain by internal and external cues, if they were co-opted, and how they had been integrated during the evolution of the betalain regulation.

The accumulating data about betalain regulation in the *Amaranthaceae* and the additional information about the molecular mode of action found in the studies of this thesis, enabled to build a hypothesis about how AhMYB2 and other betalain-regulating PAP1 homologs might have evolved. PAP1 homologs are highly conserved for anthocyanin regulation across the plant kingdom (e.g., in monocots: (Li et al., 2020; Zheng et al., 2021), and diverse dicots: (Talos et al., 2006; Chen et al., 2025; Mano et al., 2007)). Although there are also PAP1-independent regulations (e.g. (Wei et al., 2024)), it has been shown for *Arabidopsis* (He et al., 2021; Lee et al., 2024; Qin et al., 2022) and apple (An et al., 2019, 2020b, 2018, 2020a) that PAP1 homologs often have a central role integrating signaling pathways of multiple internal and external cues. Considering the functional similarity of betalains and anthocyanins, one could expect an evolutionary advantage if the co-option of PAP1 included a network of regulatory pathways already adapted to integrate various internal and external signals. Co-option of PAP1 required the adaptation to new target genes, which could either happen by *trans* mutations leading to a TF that recognizes a different motif or by *cis* mutations causing the binding motifs to degenerate in the original target promoter and to arise in the new target promoter. *cis* mutations are expected to be evolutionarily more likely, as motifs are short and can arise with few mutations (Stone and Wray, 2001) and are mostly additive and less likely to be pleiotropic (Meiklejohn et al., 2014). In contrast, *trans* mutations are usually dominant/repressive and are more likely to be pleiotropic (Meiklejohn et al., 2014). Moreover, the (core) binding motif of TFs is highly conserved for most TF classes (Zenker et al., 2025). Nevertheless, there is a varying degree of conservation with some TF families, such as C2C2 with a more than 450 million-year-old conserved motif, while others show high diversity (e.g., C2H2) or semi-conservation (e.g., MYB), which are thus more likely to evolve to bind to new DNA sequences (Zenker et al., 2025). Despite the common assumption that *cis* mutations are more common, in the case of PAP1 homologs and the betalain regulation, there is evidence that *trans* mutations had been important. According to Pucker et al. (2024) there had been a duplication of PAP1 leading to two paralogs in ancestral *Caryophyllales* species. In all betalain-producing species, only one of both copies was retained (PAP1 α for betalain transition 2, including BvMYB1 and AhMYB2, and PAP1 β for the other three betalain transitions) (Pucker et al., 2024). We could hypothesize that in the ancestors of betalain-producing lineages, one copy had been free to evolve to adapt to new target genes. After adaptation to target betalains was the anthocyanin-regulating paralog lost, maybe explaining the mutation exclusiveness of both pigments. The second observation supporting the hypothesis of causal *trans* mutations, is that for both PAP1 paralogous

lines, it was shown that the betalain PAP1 homolog can't regulate anthocyanin genes, neither endogenous nor the one from anthocyanin-producing species. Sakuta et al. (2021) showed that PAP1 homologs from *Portulaca grandiflora*, *Basella alba*, and *Mammillaria zeilmanniana*, all belonging to betalain transition 4, can't activate the ANS promoter neither from *Arabidopsis* nor the betalain-producing spinach. Similarly, our RNA-seq data found no activation of the endogenous flavonoid pathway genes by AhMYB2 overexpression (Chapter 4, Figure 4.6). We also did not observe any visible pigmentation after transiently overexpressing AhMYB2 in *N. benthamiana* in line which Peng et al. (2019) for AmMYB2. Finally, by comparing conserved amino acids in the R2 and R3 domains for both PAP1 paralogs in betalain and anthocyanin-producing species, we found betalain-specific amino acids for both paralogs (8 for PAP1 α and 5 for PAP1 β) that are not present in the closely related anthocyanin-producing *Dianthus caryophyllus* (Figure 5.1, see also the logoplot by Pucker et al. (2024) for the R3 domain only). Some of these amino acids are close to residues that are known to be involved in DNA binding (3/8 for PAP1 α and 1/5 for PAP1 β , see Figure 5.1). It was previously shown that often the residues surrounding the ones in direct contact with the DNA influence the specificity of the bound DNA sequence (Heppel et al., 2013). For two of them and for two residues that are not directly located to an already known DNA contact residue, it has even been demonstrated that these residues can cause a switch in target genes if mutated together with others, backing our hypothesis (Sakuta et al. (2021) for positions 47, 66, and 83 and Heppel et al. (2013) for position 90) (Figure 5.1). However, some betalain-specific amino acids are also at or close by to residues involved in bHLH interaction (4/8 for PAP1 α and 3/5 for PAP1 β , see Figure 5.1), and could, as discussed in Chapter 4, section 4.4.2 explain the lost bHLH interaction. The lost bHLH interaction, at least shown for one PAP1 paralogous line (BvMYB1 (Hatlestad et al., 2015) and AhMYB2 by us), could have multiple effects. On the one hand, it might have helped to reduce pleiotropy, as the bHLHs are involved in trichome, mucilage, protoanthocyanin, and root hair regulation too (Marin-Recinos and Pucker, 2024; Schuh and Morris, 2025). On the other hand, contribute the bHLH in the MYB complex to DNA binding (Morohashi et al., 2007; Bulanov et al., 2025), and might be able to draw it to the anthocyanin genes if still interacting similarly as it was shown for the MYB protein P in maize, which was able to activate additional promoters if mutated to interact with a bHLH TF known to be involved in regulating these genes (Grotewold et al., 2000). The fact that the GL3/EGL3 gene lineage has been completely lost in the species of the betalain transition 2 and, complementarily, the MYC1 gene lineage in species of the betalain transition 3 supports a strong selection against the interaction of PAP1 with class III β bHLH during betalain evolution (Pucker et al., 2024). That two residues (at positions 12 and 83) had been altered in both PAP1 lineages, but to different amino acids, and that in at least two different betalain transitions, different class III β bHLH gene lineages have been lost, strongly hints at convergent evolution. That both amino acids, which (potentially) play a role in DNA binding speci-

ficity as well as residues important for bHLH interaction, had been changed in both PAP1 lineages suggests that both traits (different TF binding site and lost bHLH interaction) are important for the switch to new target genes. This is supported by the observation that the mutated BvMYB1 version of Hatlestad et al. (2015), despite having restored bHLH interaction, is unable to activate the anthocyanin pathway in Arabidopsis (Lloyd et al., 2017).

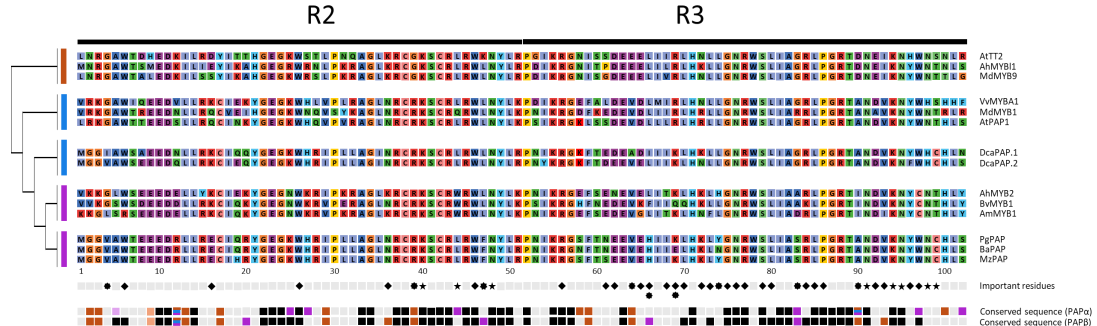


Figure 5.1: Comparison and conservation of the amino acid sequence of the R2 and R3 domains of PAP1 genes in closely related anthocyanin and betalain species. The top panel represents three protoanthocyanin regulators from betalain and anthocyanin species used as an outgroup for the PAP1 gene lineage. The blue bar marks 3 sequences from anthocyanin regulating PAP1s outside (2nd from the top) *Caryophyllales* and 2 copies from a PAP1 homolog of a *Caryophyllales* species (3rd from the top). Each of the 3 PAP1 sequences from the PAP1 α gene lineage (4th from top) and from the PAP1 β gene lineage (5th from top) are marked with purple. Note that the three PAP1 α sequences are from verified betalain regulators, while three PAP1 β sequences are from betalain-producing species but were not verified to regulate betalains. The third panel from the bottom indicates residues that were found to play a role in bHLH interaction (black diamond, data from Zimmermann et al. (2004); Hichri et al. (2010); Hatlestad et al. (2015)), DNA binding (black 4-point star), or specificity (black 8-point star, data for both from Hichri et al. (2010); Sakuta et al. (2021); Heppel et al. (2013)). Bottom panels show the conserved (black), variable (grey), and specific (purple, blue, and brown for betalains, anthocyanins, and protoanthocyanins, respectively; light color indicates that only in this group the residue is not conserved) residues. Sequences from At (*Arabidopsis thaliana*), Ah (*Amaranthus hypochondriacus*), Md (*Malus domestica*), Dc (*Dianthus caryophyllus*), Bv (*Beta vulgaris*), Am (*Amaranthus tricolor*), Pg (*Portulaca grandiflora*), Ba (*Basella alba*), and Mz (*Mammillaria zeilmanniana*).

In the course of this thesis, I implemented a range of valuable molecular, *in-vitro*, and transcriptomic resources in the non-model, orphan crop grain amaranth, which will facilitate future research and breeding. Today there is still no efficient and reproducible protocol for regeneration of stably transformed grain amaranth, which strongly hampers breeding and research of this versatile crop (Castellanos-Arévalo et al., 2020, 2021). Thus, represents our established efficient callus transformation protocol, a breakthrough that enables the study of stable transformed amaranth tissue in high quantities (Chapter 2, 4). Using this as a foundation, we could, for the first time, demonstrate successful targeted mutagenesis using CRISPR/Cas9 in amaranth, a key method in modern breeding and research (Chapter 2, (Vollmer et al., 2026)). Additionally, we adapted hairy root transformation as a comparable, easy, and fast stable transformation method, speeding

up gene function experiments drastically (Chapter 3, (Winkler et al., 2024)). Moreover, we presented a protocol for isolation and, to our knowledge, the first protocol for the transformation of *A. hypochondriacus* protoplast (Chapter 4). Although the protocol requires further optimization, it represents a valuable starting point, putting cutting-edge, high-throughput methods like single-cell omics for this orphan crop within reach. We implemented hygromycin for efficient selection and demonstrated sufficient activity of several promoters and terminators (promoters: *AtUBQ10*, *AtU6-26*, and *PcUbi4-2*; terminators: *OCS*, *E9*, *HSP18.2* and *NOS*) and high activity of the *35S* promoter in grain amaranth (Chapter 2, 3, 4). This paves the way to design efficient constructs for transgene expression in grain amaranth. Furthermore, does the new *A. hypochondriacus* genome assembly reach the quality of model organism reference genomes, lifting genomic research to a new level (Chapter 3, (Winkler et al., 2024)). Complemented with the generated RNA-seq dataset, this will significantly facilitate candidate selection, primer design, and planning of future experiments. Together, the methods presented in this thesis vastly extend the molecular toolbox for amaranth research, unlocking key state-of-the-art methods for this orphan, non-model crop.

5.1 Conclusion

The scientific advancement in the last decade enabled to move from the simple identification of genes to the reconstruction of whole gene regulatory networks. Engineering of gene regulatory networks could be a powerful tool to breed crops that are better adapted to the extreme conditions caused by climate change (Leong et al., 2025; Borowsky and Bailey-Serres, 2024). Optimizing gene regulatory networks requires a deep understanding of their structure and evolution, which we just started to elucidate. However, evolutionary studies need to investigate species across the tree of life beyond the limited number of model species, but reports of functional evolutionary studies in non-model species are scarce due to the lack of key molecular methods outside model species (Pabón-Mora et al., 2023; Rensing, 2017). The studies presented in this thesis vastly extended the toolbox for the non-model crop grain amaranth. We increased the available genomic resources by providing a new assembly of the *A. hypochondriacus* genome with reference genome quality (Chapter 3, (Winkler et al., 2024)) and a RNA-seq data set (Chapter 4). We established various *in vitro* methods for endogenous overexpression and present the first targeted mutagenesis using CRISPR/Cas9 in amaranth (Chapter 2, (Vollmer et al., 2026)). Additionally, we tested a protocol for protoplast isolation and presented the first protoplast transformation protocol for amaranth (Chapter 4). The established methods paved the way for functional studies in grain amaranth, representing the understudied *Caryophyllales*. The probably four-times evolved betalain pathway, replacing the otherwise ubiquitous anthocyanins in parts of the *Caryophyllales* (Brockington et al., 2011;

Timoneda et al., 2019; Pucker et al., 2024), represents an optimal model to investigate the evolution of gene regulatory networks for *de novo* evolved pathways. I present in this thesis how we reconstructed the tissue-specific regulation of the betalain pathway in the orphan crop grain amaranth by combining computational, molecular, and *in vitro* methods. We showed that upregulation of the first enzyme in the pathway, AhCYP76AD2, is sufficient for the induction of the betalain pathway. We then demonstrated that AhMYB2, a homolog of the anthocyanin regulating AtPAP1, mediates the tissue-specific betalain regulation by upregulating *AhCYP76AD2* and *AhDODAA1*, the second enzyme in the pathway in flowers. Extensive testing in Y2H proved that AhMYB2, similar to what was suggested for BvMYB1, which regulated betalains in the closely related *B. vulgaris* (Hatlestad et al., 2015), is unable to interact with class IIIb bHLH TFs. However, while it was previously hypothesized that betalain-regulating PAP1 might function alone (Hatlestad et al., 2015), does our transactivation data indicate that they require a betalain-specific additional TF for efficient binding to the promoter. This TF is yet to be identified, but our RNA-seq data enabled us to identify a list of promising candidates for AhMYB2 co-regulators, including WRKY, AP2/ERF, bZIP, and NAC TFs. Finally, comparing the amino acid sequence of the R2 and R3 domains of AhMYB2 and other betalain-regulating PAP1 homologs (Sakuta et al., 2021; Hatlestad et al., 2015; Xue et al., 2025) led us to the discovery of multiple betalain-specific mutations. Combining our data with published information, we could observe a strong indication that *trans* mutation had been key during the evolution of the betalain pathway regulation. We propose that amino acid changes in one of the two PAP1 copies in ancestral *Caryophyllales* species (Pucker et al., 2024) lead to a loss of the bHLH interaction and to a shift of bound DNA sequences together, achieving the regulation of a different set of genes. This thesis combines the first systematic investigation of the mode of action of a betalain-regulating TF in the *Amaranthaceae*, laying a strong foundation to reconstruct the evolution of the betalain pathway regulation. Testing the binding and transactivation ability of AhMYB2 with mutation either restoring the bHLH interaction or the binding to the PAP1 binding motif in heterologous assays will validate our hypothesis and might even provide general insight into the flexibility and evolution of (MYB) transcription factor DNA binding specificity. We found that light and cytokinin induce the betalain pathway similarly to AhMYB2 by upregulating *AhCYP76AD2* and *DODAA1*, but this appears to be independent of AhMYB2. Future studies exploring how the range of internal and external signals known to influence betalain accumulation are integrated (for instance, by capitalizing on RNA-seq data sets from different conditions) could unravel if more than just PAP1 was co-opted from the anthocyanin regulation. More data also about the other betalain origins could uncover general principles of how gene regulatory networks of *de novo* evolved pathways evolve.

5.2 References

- J. An, J. Yao, R. Xu, C. You, X. Wang, and Y. Hao. Apple bZIP transcription factor MdbZIP44 regulates abscisic acid-promoted anthocyanin accumulation. *Plant, Cell & Environment*, 41(11):2678–2692, Nov. 2018. ISSN 0140-7791, 1365-3040. doi: 10.1111/pce.13393. URL <https://onlinelibrary.wiley.com/doi/10.1111/pce.13393>.
- J. An, X. Zhang, C. You, S. Bi, X. Wang, and Y. Hao. MdWRKY40 promotes wounding-induced anthocyanin biosynthesis in association with MdMYB1 and undergoes MdBT2-mediated degradation. *New Phytologist*, 224(1):380–395, Oct. 2019. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.16008. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.16008>.
- J. An, X. Zhang, S. Bi, C. You, X. Wang, and Y. Hao. The ERF transcription factor MdERF38 promotes drought stress-induced anthocyanin biosynthesis in apple. *The Plant Journal*, 101(3):573–589, Feb. 2020a. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.14555. URL <https://onlinelibrary.wiley.com/doi/10.1111/tpj.14555>.
- J.-P. An, Y.-J. Liu, X.-W. Zhang, S.-Q. Bi, X.-F. Wang, C.-X. You, and Y.-J. Hao. Dynamic regulation of anthocyanin biosynthesis at different light intensities by the BT2-TCP46-MYB1 module in apple. *Journal of Experimental Botany*, 71(10):3094–3109, May 2020b. ISSN 0022-0957, 1460-2431. doi: 10.1093/jxb/eraa056. URL <https://academic.oup.com/jxb/article/71/10/3094/5717769>.
- H. M. Azeredo. Betalains: properties, sources, applications, and stability – a review. *International Journal of Food Science & Technology*, 44(12):2365–2376, Dec. 2009. ISSN 0950-5423, 1365-2621. doi: 10.1111/j.1365-2621.2007.01668.x. URL <https://academic.oup.com/ijfst/article/44/12/2365/7865206>.
- A. T. Borowsky and J. Bailey-Serres. Rewiring gene circuitry for plant improvement. *Nature Genetics*, 56(8):1574–1582, Aug. 2024. ISSN 1061-4036, 1546-1718. doi: 10.1038/s41588-024-01806-7. URL <https://www.nature.com/articles/s41588-024-01806-7>.
- S. F. Brockington, R. H. Walker, B. J. Glover, P. S. Soltis, and D. E. Soltis. Complex pigment evolution in the *Caryophyllales*. *New Phytologist*, 190(4):854–864, June 2011. ISSN 0028-646X, 1469-8137. doi: 10.1111/j.1469-8137.2011.03687.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.2011.03687.x>.
- A. N. Bulanov, E. A. Andreeva, N. V. Tsvetkova, and P. A. Zykin. Regulation of Flavonoid Biosynthesis by the MYB-bHLH-WDR (MBW) Complex in Plants and Its Specific Features in Cereals. *International Journal of Molecular Sciences*, 26(2):734, Jan. 2025. ISSN 1422-0067. doi: 10.3390/ijms26020734. URL <https://www.mdpi.com/1422-0067/26/2/734>.
- A. P. Castellanos-Arévalo, A. A. Estrada-Luna, J. L. Cabrera-Ponce, E. Valencia-Lozano, H. Herrera-Ubaldo, S. de Folter, A. Blanco-Labra, and J. P. Délano-Frier. *Agrobacterium rhizogenes*-mediated transformation of grain (*Amaranthus hypochondriacus*) and leafy (*A. hybridus*) amaranths. *Plant Cell Reports*, 39(9):1143–1160, Sept. 2020. ISSN 0721-7714, 1432-203X. doi: 10.1007/s00299-020-02553-9. URL <https://link.springer.com/10.1007/s00299-020-02553-9>.

- A. P. Castellanos-Arévalo, J. L. Cabrera-Ponce, C. Nava-Sandoval, and J. P. Délano-Frier. How to Overcome Recalcitrance? Novel Strategies and Recent Advances in the Genetic Transformation of Grain Amaranth. In D. Adhikary, M. K. Deyholos, and J. P. Délano-Frier, editors, *The Amaranth Genome*, pages 125–149. Springer International Publishing, Cham, 2021. ISBN 978-3-030-72364-4 978-3-030-72365-1. doi: 10.1007/978-3-030-72365-1_8. URL https://link.springer.com/10.1007/978-3-030-72365-1_8. Series Title: Compendium of Plant Genomes.
- D. Chen, C. Wang, Y. Liu, W. Shen, Q. Cuimu, D. Zhang, B. Zhu, L. Chen, and C. Tan. Systematic identification of R2R3-MYB S6 subfamily genes in *Brassicaceae* and its role in anthocyanin biosynthesis in Brassica crops. *BMC Plant Biology*, 25(1):290, Mar. 2025. ISSN 1471-2229. doi: 10.1186/s12870-025-06296-1. URL <https://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-025-06296-1>.
- R. A. Dixon and A. J. Dickinson. A century of studying plant secondary metabolism—From “what?” to “where, how, and why?”. *Plant Physiology*, page kiad596, Jan. 2024. ISSN 0032-0889, 1532-2548. doi: 10.1093/plphys/kiad596. URL <https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiad596/7499152>.
- E. Grotewold, M. B. Sainz, L. Tagliani, J. M. Hernandez, B. Bowen, and V. L. Chandler. Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Proceedings of the National Academy of Sciences*, 97(25):13579–13584, Dec. 2000. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.250379897. URL <https://pnas.org/doi/full/10.1073/pnas.250379897>.
- G. J. Hatlestad, N. A. Akhavan, R. M. Sunnadeniya, L. Elam, S. Cargile, A. Hembd, A. Gonzalez, J. M. McGrath, and A. M. Lloyd. The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway. *Nature Genetics*, 47(1):92–96, Jan. 2015. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3163. URL <http://www.nature.com/articles/ng.3163>.
- Y. He, X. Zhang, L. Li, Z. Sun, J. Li, X. Chen, and G. Hong. SPX4 interacts with both PHR1 and PAP1 to regulate critical steps in phosphorus-status-dependent anthocyanin biosynthesis. *New Phytologist*, 230(1):205–217, Apr. 2021. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.17139. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.17139>.
- S. C. Heppel, F. W. Jaffé, A. M. Takos, S. Schellmann, T. Rausch, A. R. Walker, and J. Bogs. Identification of key amino acids for the evolution of promoter target specificity of anthocyanin and proanthocyanidin regulating MYB factors. *Plant Molecular Biology*, 82(4-5):457–471, July 2013. ISSN 0167-4412, 1573-5028. doi: 10.1007/s11103-013-0074-8. URL <http://link.springer.com/10.1007/s11103-013-0074-8>.
- I. Hichri, S. C. Heppel, J. Pillet, C. Léon, S. Czemmel, S. Delrot, V. Lauvergeat, and J. Bogs. The Basic Helix-Loop-Helix Transcription Factor MYC1 Is Involved in the Regulation of the Flavonoid Biosynthesis Pathway in Grapevine. *Molecular Plant*, 3(3):509–523, May 2010. ISSN 16742052. doi: 10.1093/mp/ssp118. URL <https://linkinghub.elsevier.com/retrieve/pii/S1674205214607292>.
- G. Jain and K. S. Gould. Are betalain pigments the functional homologues of antho-

- cyanins in plants? *Environmental and Experimental Botany*, 119:48–53, Nov. 2015. ISSN 00988472. doi: 10.1016/j.envexpbot.2015.06.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S0098847215001112>.
- D. M. Jones and K. Vandepoele. Identification and evolution of gene regulatory networks: insights from comparative studies in plants. *Current Opinion in Plant Biology*, 54: 42–48, Apr. 2020. ISSN 13695266. doi: 10.1016/j.pbi.2019.12.008. URL <https://linkinghub.elsevier.com/retrieve/pii/S1369526619301232>.
- M. Larter, A. Dunbar-Wallis, A. E. Berardi, and S. D. Smith. Convergent Evolution at the Pathway Level: Predictable Regulatory Changes during Flower Color Transitions. *Molecular Biology and Evolution*, 35(9):2159–2169, Sept. 2018. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msy117. URL <https://academic.oup.com/mbe/article/35/9/2159/5034462>.
- S.-H. Lee, S.-H. Kim, T.-K. Park, Y.-P. Kim, J.-W. Lee, and T.-W. Kim. Transcription factors BZR1 and PAP1 cooperate to promote anthocyanin biosynthesis in Arabidopsis shoots. *The Plant Cell*, 36(9):3654–3673, Sept. 2024. ISSN 1040-4651, 1532-298X. doi: 10.1093/plcell/koae172. URL <https://academic.oup.com/plcell/article/36/9/3654/7692851>.
- R. Leong, X. He, B. S. Beijen, T. Sakai, J. Goncalves, and P. Ding. Unlocking gene regulatory networks for crop resilience and sustainable agriculture. *Nature Biotechnology*, 43(8):1254–1265, Aug. 2025. ISSN 1087-0156, 1546-1696. doi: 10.1038/s41587-025-02727-4. URL <https://www.nature.com/articles/s41587-025-02727-4>.
- B.-J. Li, B.-Q. Zheng, J.-Y. Wang, W.-C. Tsai, H.-C. Lu, L.-H. Zou, X. Wan, D.-Y. Zhang, H.-J. Qiao, Z.-J. Liu, and Y. Wang. New insight into the molecular mechanism of colour differentiation among floral segments in orchids. *Communications Biology*, 3(1):89, Feb. 2020. ISSN 2399-3642. doi: 10.1038/s42003-020-0821-8. URL <https://www.nature.com/articles/s42003-020-0821-8>.
- A. Lloyd, A. Brockman, L. Aguirre, A. Campbell, A. Bean, A. Cantero, and A. Gonzalez. 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(9):1431–1441, Sept. 2017. ISSN 0032-0781, 1471-9053. doi: 10.1093/pcp/pcx075. URL <https://academic.oup.com/pcp/article/58/9/1431/3858365>.
- S. Lopez-Nieves, Y. Yang, A. Timoneda, M. Wang, T. Feng, S. A. Smith, S. F. Brockington, and H. A. Maeda. Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in *Caryophyllales*. *New Phytologist*, 217(2):896–908, Jan. 2018. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.14822. URL <https://onlinelibrary.wiley.com/doi/10.1111/nph.14822>.
- H. Mano, F. Ogasawara, K. Sato, H. Higo, and Y. Minobe. Isolation of a Regulatory Gene of Anthocyanin Biosynthesis in Tuberous Roots of Purple-Fleshed Sweet Potato. *Plant Physiology*, 143(3):1252–1268, Mar. 2007. ISSN 0032-0889, 1532-2548. doi: 10.1104/pp.106.094425. URL <https://academic.oup.com/plphys/article/143/3/1252-1268/6106663>.

- M. F. Marin-Recinos and B. Pucker. Genetic factors explaining anthocyanin pigmentation differences. *BMC Plant Biology*, 24(1):627, July 2024. ISSN 1471-2229. doi: 10.1186/s12870-024-05316-w. URL <https://bmcpplantbiol.biomedcentral.com/articles/10.1186/s12870-024-05316-w>.
- S. Mariyam, V. Kumar, A. Roychoudhury, G. S. Ghodake, S. Muneer, J. S. Duhan, F. Ahmad, R. K. Sharma, J. Singh, and C. S. Seth. Functional Diversification and Mechanistic Insights of MYB Transcription Factors in Mediating Plant Growth and Development, Secondary Metabolism, and Stress Responses. *Journal of Plant Growth Regulation*, 44(4):1465–1484, Apr. 2025. ISSN 0721-7595, 1435-8107. doi: 10.1007/s00344-025-11622-w. URL <https://link.springer.com/10.1007/s00344-025-11622-w>.
- C. D. Meiklejohn, J. D. Coolon, D. L. Hartl, and P. J. Wittkopp. The roles of *cis* - and *trans* -regulation in the evolution of regulatory incompatibilities and sexually dimorphic gene expression. *Genome Research*, 24(1):84–95, Jan. 2014. ISSN 1088-9051. doi: 10.1101/gr.156414.113. URL <http://genome.cshlp.org/lookup/doi/10.1101/gr.156414.113>.
- K. Morohashi, M. Zhao, M. Yang, B. Read, A. Lloyd, R. Lamb, and E. Grotewold. Participation of the *Arabidopsis* bHLH Factor GL3 in Trichome Initiation Regulatory Events. *Plant Physiology*, 145(3):736–746, Nov. 2007. ISSN 1532-2548. doi: 10.1104/pp.107.104521. URL <https://academic.oup.com/plphys/article/145/3/736/6107095>.
- F. Mughal and G. Caetano-Anollés. MANET 3.0: Hierarchy and modularity in evolving metabolic networks. *PLOS ONE*, 14(10):e0224201, Oct. 2019. ISSN 1932-6203. doi: 10.1371/journal.pone.0224201. URL <https://dx.plos.org/10.1371/journal.pone.0224201>.
- N. Pabón-Mora, H. Suárez-Baron, Y. Madrigal, J. F. Alzate, and F. González. Expression and Functional Studies of Leaf, Floral, and Fruit Developmental Genes in Non-model Species. In J. L. Riechmann and C. Ferrándiz, editors, *Flower Development*, volume 2686, pages 365–401. Springer US, New York, NY, 2023. ISBN 978-1-0716-3298-7 978-1-0716-3299-4. doi: 10.1007/978-1-0716-3299-4_19. URL https://link.springer.com/10.1007/978-1-0716-3299-4_19. Series Title: Methods in Molecular Biology.
- J. Peng, B.-J. Wang, N. Svetec, and L. Zhao. Gene regulatory networks and essential transcription factors for *de novo*-originated genes. *Nature Ecology & Evolution*, 9(8):1487–1498, July 2025. ISSN 2397-334X. doi: 10.1038/s41559-025-02747-y. URL <https://www.nature.com/articles/s41559-025-02747-y>.
- L. Peng, Y. Wang, X. Sun, X. Wang, C. Zhao, C. Wang, L. Xiang, J. Chen, Z. Lai, and S. Liu. Expression and Functional Analysis of AmMYB2 Related to Betalain Metabolism of *Amaranthus tricolor*. *Acta Horticulturae Sinica*, 46(3):473–485, 2019. doi: 10.16420/j.issn.0513-353x.2018-0572iijZ.
- G. Polturak and A. Aharoni. “La Vie en Rose”: Biosynthesis, Sources, and Applications of Betalain Pigments. *Molecular Plant*, 11(1):7–22, Jan. 2018. ISSN 16742052. doi: 10.1016/j.molp.2017.10.008. URL <https://linkinghub.elsevier.com/retrieve/pii/S1674205217303076>.

- B. Pucker, N. Walker-Hale, J. Dzurlic, W. C. Yim, J. C. Cushman, A. Crum, Y. Yang, and S. F. Brockington. Multiple mechanisms explain loss of anthocyanins from betalain-pigmented *Caryophyllales*, including repeated wholesale loss of a key anthocyanidin synthesis enzyme. *New Phytologist*, 241(1):471–489, Jan. 2024. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.19341. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19341>.
- J. Qin, C. Zhao, S. Wang, N. Gao, X. Wang, X. Na, X. Wang, and Y. Bi. PIF4-PAP1 interaction affects MYB-bHLH-WD40 complex formation and anthocyanin accumulation in *Arabidopsis*. *Journal of Plant Physiology*, 268:153558, Jan. 2022. ISSN 01761617. doi: 10.1016/j.jplph.2021.153558. URL <https://linkinghub.elsevier.com/retrieve/pii/S0176161721001978>.
- S. A. Rensing. Why we need more non-seed plant models. *New Phytologist*, 216(2): 355–360, Oct. 2017. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.14464. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.14464>.
- M. Sakuta, A. Tanaka, K. Iwase, M. Miyasaka, S. Ichiki, M. Hatai, Y. T. Inoue, A. Yamagami, T. Nakano, K. Yoshida, and S. Shimada. Anthocyanin synthesis potential in betalain-producing *Caryophyllales* plants. *Journal of Plant Research*, 134(6):1335–1349, Nov. 2021. ISSN 0918-9440, 1618-0860. doi: 10.1007/s10265-021-01341-0. URL <https://link.springer.com/10.1007/s10265-021-01341-0>.
- A. Schuh and G. Morris. Loss of Pleiotropic Regulatory Functions in *Tannin1*, the Sorghum Ortholog of Arabidopsis Master Regulator *TTG1*. *Plant Direct*, 9(3):e70055, Mar. 2025. ISSN 2475-4455, 2475-4455. doi: 10.1002/pld3.70055. URL <https://onlinelibrary.wiley.com/doi/10.1002/pld3.70055>.
- H. Sheehan, T. Feng, N. Walker-Hale, S. Lopez-Nieves, B. Pucker, R. Guo, W. C. Yim, R. Badgami, A. Timoneda, L. Zhao, H. Tiley, D. Copetti, M. J. Sanderson, J. C. Cushman, M. J. Moore, S. A. Smith, and S. F. Brockington. Evolution of l-DOPA 4,5-dioxygenase activity allows for recurrent specialisation to betalain pigmentation in *Caryophyllales*. *New Phytologist*, 227(3):914–929, Aug. 2020. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.16089. URL <https://onlinelibrary.wiley.com/doi/10.1111/nph.16089>.
- J. R. Stone and G. A. Wray. Rapid Evolution of cis-Regulatory Sequences via Local Point Mutations. *Molecular Biology and Evolution*, 18(9):1764–1770, Sept. 2001. ISSN 1537-1719, 0737-4038. doi: 10.1093/oxfordjournals.molbev.a003964. URL <http://academic.oup.com/mbe/article/18/9/1764/1000770>.
- R. Sunnadeniya, A. Bean, M. Brown, N. Akhavan, G. Hatlestad, A. Gonzalez, V. V. Symonds, and A. Lloyd. Tyrosine Hydroxylation in Betalain Pigment Biosynthesis Is Performed by Cytochrome P450 Enzymes in Beets (*Beta vulgaris*). *PLOS ONE*, 11(2):e0149417, Feb. 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0149417. URL <https://dx.plos.org/10.1371/journal.pone.0149417>.
- A. M. Takos, F. W. Jaffé, S. R. Jacob, J. Bogs, S. P. Robinson, and A. R. Walker. Light-Induced Expression of a *MYB* Gene Regulates Anthocyanin Biosynthesis in Red Apples. *Plant Physiology*, 142(3):1216–1232, Nov. 2006. ISSN 0032-0889, 1532-2548. doi: 10.1104/pp.106.088104. URL <https://academic.oup.com/plphys/article/142/3/>

- A. Timoneda, T. Feng, H. Sheehan, N. Walker-Hale, B. Pucker, S. Lopez-Nieves, R. Guo, and S. Brockington. The evolution of betalain biosynthesis in *Caryophyllales*. *New Phytologist*, 224(1):71–85, Oct. 2019. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.15980. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.15980>.
- S. K. Vollmer, M. G. Stetter, and G. Hensel. First Successful Targeted Mutagenesis Using CRISPR/Cas9 in Stably Transformed Grain Amaranth Tissue. *Plant Biotechnology Journal*, page pbi.70590, Feb. 2026. ISSN 1467-7644, 1467-7652. doi: 10.1111/pbi.70590. URL <https://onlinelibrary.wiley.com/doi/10.1111/pbi.70590>.
- K. Voordeckers, K. Pougach, and K. J. Verstrepen. How do regulatory networks evolve and expand throughout evolution? *Current Opinion in Biotechnology*, 34: 180–188, Aug. 2015. ISSN 09581669. doi: 10.1016/j.copbio.2015.02.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S0958166915000208>.
- J. Wei, Y. Zhang, Y. Xu, Z. Wang, J. Meng, H. Wang, Y. Yang, Y. Wang, M. Chen, and H. Li. MsLBD41-like bridges auxin response factors MsARF5 and MsF3'H function to regulate low nitrogen-induced anthocyanin accumulation in *Malus spectabilis* leaves. *Industrial Crops and Products*, 214:118572, Aug. 2024. ISSN 09266690. doi: 10.1016/j.indcrop.2024.118572. URL <https://linkinghub.elsevier.com/retrieve/pii/S0926669024005491>.
- L. C. Wheeler, J. F. Walker, J. Ng, R. Deanna, A. Dunbar-Wallis, A. Backes, P. H. Pezzi, M. V. Palchetti, H. M. Robertson, A. Monaghan, L. B. De Freitas, G. E. Barboza, E. Moyroud, and S. D. Smith. Transcription Factors Evolve Faster Than Their Structural Gene Targets in the Flavonoid Pigment Pathway. *Molecular Biology and Evolution*, 39(3):msac044, Mar. 2022. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msac044. URL <https://academic.oup.com/mbe/article/doi/10.1093/molbev/msac044/6536971>.
- T. S. Winkler, S. K. Vollmer, N. Dyballa-Rukes, S. Metzger, and M. G. Stetter. Isoform-resolved genome annotation enables mapping of tissue-specific betalain regulation in amaranth. *New Phytologist*, page nph.19736, Apr. 2024. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.19736. URL <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19736>.
- Y. Xue, K. Li, W. Feng, Z. Lai, and S. Liu. Identification of R2R3-MYB Transcription Factor Family Based on *Amaranthus tricolor* Genome and AtrMYB72 Promoting Betalain Biosynthesis by Directly Activating AtrCYP76AD1 Expression. *Plants*, 14(3):324, Jan. 2025. ISSN 2223-7747. doi: 10.3390/plants14030324. URL <https://www.mdpi.com/2223-7747/14/3/324>.
- S. Zenker, D. Wulf, A. Meierhenrich, P. Viehöver, S. Becker, M. Eisenhut, R. Stracke, B. Weisshaar, and A. Bräutigam. Many transcription factor families have evolutionarily conserved binding motifs in plants. *Plant Physiology*, 198(2):kiaf205, May 2025. ISSN 0032-0889, 1532-2548. doi: 10.1093/plphys/kiaf205. URL <https://academic.oup.com/plphys/article/doi/10.1093/plphys/kiaf205/8142508>.
- D. Zhang, H. Zhou, Y. Zhang, Y. Zhao, Y. Zhang, X. Feng, and H. Lin. Diverse roles

- of MYB transcription factors in plants. *Journal of Integrative Plant Biology*, 67(3): 539–562, Mar. 2025. ISSN 1672-9072, 1744-7909. doi: 10.1111/jipb.13869. URL <https://onlinelibrary.wiley.com/doi/10.1111/jipb.13869>.
- J. Zheng, H. Wu, M. Zhao, Z. Yang, Z. Zhou, Y. Guo, Y. Lin, and H. Chen. OsMYB3 is a R2R3-MYB gene responsible for anthocyanin biosynthesis in black rice. *Molecular Breeding*, 41(8):51, Aug. 2021. ISSN 1380-3743, 1572-9788. doi: 10.1007/s11032-021-01244-x. URL <https://link.springer.com/10.1007/s11032-021-01244-x>.
- I. M. Zimmermann, M. A. Heim, B. Weisshaar, and J. F. Uhrig. Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *The Plant Journal*, 40(1):22–34, Oct. 2004. ISSN 0960-7412, 1365-313X. doi: 10.1111/j.1365-313X.2004.02183.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2004.02183.x>.

Acknowledgments

First of all, I would like to thank my both supervisors **Dr. Götz Hensel** and **Prof. Dr. Markus Stetter**, for deciding to entrust me with this challenging but fascinating project and guiding me through all its turns over the years. I am grateful for introducing me to amaranth and the fascinating field of evolution, **Markus**, and for everything you have taught me, from the first steps in data analysis to how to refine and present my project and efficient project management. I am grateful for the freedom that you let me have, **Götz**, but also for guiding me to my first CRISPR mutants and giving me the opportunity to supervise students, asking me to step in giving lectures for you, and being on stage with you. Your both motivation to constantly improve and your trust have let me grow.

I highly appreciate the feedback and support by all my thesis advisory board members: besides **Dr. Götz Hensel** and **Prof. Dr. Markus Stetter**, **Prof. Dr. Andreas Weber**, **Prof. Dr. Samuel Brockington**, and **Prof. Dr. Maria von Korff Schmising**. Special thanks to **Prof. Dr. Andreas Weber** for being my second examiner.

I am honored to be part of the **Cluster of Excellence in Plant Sciences (CEPLAS)** and their grad school. Besides the mere fact of funding, **CEPLAS** has an excellent community, a perfect environment for plant science, and is united by the deep motivation to push science forward and advance human society. Thanks to the **CEPLAS office** and your various support over time.

I had the great chance of working in many different labs, and I want to thank all members of the **AG Weber**, the **Centre for Plant Genome Engineering (CPGE)**, the **AG Stetter**, the **AG Hülkamp** and the **AG Zurbriggen** for their warm welcome. A special thanks to all the help I got from postdocs, technicians, and PhD students by providing materials, sharing protocols, discussing ideas, and giving advice. While there were too many to list them here, I would like to mention **Dr. Uriel Urquiza-Garcia**, **Dr. Kirill Schenstniy**, **Dr. Hanna Mozghovska**, **Dr. Margarita Takou** for giving me feedback on the thesis introduction, **Dr. Vanessa Valencia**, **Tom Winkler**, **Corbinian Graf**, **Roswitha Lenz**, and **Dr. Katarzyna Makowska**. Moreover,

thanks for all the work from my students and student helpers, **Lucas** and **Linda**, **Wasim** and **Khadija**.

A PhD is so much more than just a job, and all the colleagues and friends who made this journey together with me were the ones who made it not just an incisive experience but also a wonderful one. We had a lot of fun during group events of the CPGE, the AG Stetter, and the AG Weber, in particular the gaming evenings with **Vanessa** in Düsseldorf, but also with **Tom**, **Margarita**, **Akanksha**, **Yuna**, **Maxime**, **Corbinian**, **Ella**, and the AG Stetter in Cologne, and the boulder sessions with **Jan**, **Moritz**, **Izzy**, and others from the AG Weber were regular highlights in my schedule. Thanks to everyone who listened and supported me, who celebrated the peaks with me and helped me through the valleys, especially **Amelie**, **Theresa**, and **Ilyse**.

Finally, to **my parents**, these lines are too short to express my deep gratitude. From my first steps to writing this thesis, you have always been there for me. Your encouragement lets me overcome doubts and move forward, your love keeps me going and teaches me to believe in myself, and your support trains me not to give up but how to grow with challenges.

Eidesstattliche Erklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Die Dissertation habe ich in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht.

Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Susanne Katja Vollmer

Düsseldorf, den 04. März 2026