

Characterising and Engineering Diversity for Wheat Carotenoid Biofortification

Joshua Oliver Waites



A thesis submitted to the University of East Anglia
for the degree of Doctor of Philosophy

John Innes Centre

June 2024

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

This work is dedicated to my mum, Dr Bridget Waites.
I wish we had had more time together and you had seen me
grow up to follow in your footsteps into research and academia.

Abstract

Plant carotenoids play roles in photosynthesis, photoprotection, phytohormone production, pigmentation, and also contribute to human health as essential components of a healthy diet. In wheat, active grain carotenoid biosynthesis is responsible for the yellow colour of flour, an important quality trait in durum wheat breeding. This thesis aimed to explore and engineer the genetic diversity of wheat grain carotenoid content (GCC), thus contributing towards the development of wheat with advantageous GCC traits.

Using spectrophotometric and HPLC analyses, I measured the GCC diversity of 443 Watkins global landrace accessions from both tetraploid and hexaploid collections. I found considerable variation, including some accessions with very high GCC. These results represent the largest high-resolution analysis of GCC within wheat diversity. Next, I identified 15 marker-trait associations and 14 quantitative trait loci (QTLs) associated with grain carotenoid traits within the Watkins tetraploid collection. Eleven of these were novel QTLs for carotenoid traits and may represent diversity unique to this collection.

Additionally, I examined the role of the *ORANGE* gene in wheat. This gene influences carotenoid biosynthesis in other plants but had not been characterised in wheat before. Using EMS mutants, I found that knocking out *ORANGE* reduced GCC, thereby confirming its role in carotenoid biosynthesis in wheat. Next, I showed that overexpressing a variant of *ORANGE*, containing a polymorphism known to significantly increase carotenoid accumulation in melons, resulted in higher GCC in wheat than the overexpression of the wild-type gene. This demonstrated for the first time that the polymorphism could enhance GCC in a cereal crop. Finally, I attempted to engineer variation within endogenous *ORANGE* using gene editing strategies to increase GCC in wheat. While this was unsuccessful, I identified an *ORANGE* EMS mutant line with increased GCC. Excitingly, this could be a novel gain-of-function mutation within *ORANGE* that increases carotenoid accumulation.

Access Condition and Agreement

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

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

Acknowledgements

I would first like to thank my supervisor, Wendy Harwood, for her incredible support, guidance, patience, and encouragement over the past four years. It has been an absolute joy working in your group. Thank you for all the opportunities you gave me during this time for outreach, discussing my science with visitors and helping with the GE and GM training courses. Thank you for everything you have taught me over the past four years about the world of gene editing and transformation. Equally, I'd like to thank my incredible secondary supervisor, Noam Chayut, for his spectacular support and guidance throughout these four years. I greatly appreciate the meetings and discussions we've had surrounding the project, which were always super engaging and enjoyable. I learned so much from working with you in the field of carotenoid biosynthesis and germplasm collections. I really couldn't have asked for better mentors; you both put the 'super' in supervisor.

I'd also like to thank my other supervisory team members: Sadiye Hayta (for your knowledge and help with gene editing in wheat), Simon Orford (for your help with everything glasshouse growing, TILLING and being the King of Crossing!) and David Seung (for your help and guidance with the biochemical and protein side of my project). I'd also like to thank all the supervisors of my Year 1 Rotation projects: Brande Wulff, David Seung and Janneke Balk. I had an incredibly exciting year learning and experiencing many different research fields, and I appreciate their guidance during this. Thank you also to Cristobal Uauy and Nikolai Adamski who supervised my Master's project here at JIC and my first foray into the world of research here—I had a great time in their supportive lab with supportive mentors.

Very importantly, a MASSIVE thank you to all the people who have helped contribute to this work—I fear crediting you in the methods section doesn't go far enough in expressing my gratitude. An especially noteworthy thanks to Ajay Siluveru, who has done an incredible job extracting carotenoids from the Watkins global landrace collection and who has been a wonderful friend throughout the years we worked together. I know the wheat carotenoid project is in good hands. Also, a massive thank you to Baldeep Kular for her expert help throughout this project on extracting and quantifying carotenoids! I would also like to thank the Horticultural Services, Field Site Team and Lab Support—without you, the JIC could not do the world-leading research it does. Additionally, a big thank you to the JIF rotation programme, the JIF funders and the rotation programme directors during my time here: Steph Bornemann, Antony Dodd, Enrico Coen and Jacob Malone. Thank you for running the programme and providing incredible support and guidance for the students during this.

I was also incredibly lucky during my PhD to be part of the amazing Harwood Lab and Germplasm Resource Unit. Their support, friendship and company during the four years has been incredible. I would also like to thank my wonderful office neighbour Martha, who provided me with such fantastic conversation, friendship and support, keeping me sane throughout my PhD. I hope Plant Josh is keeping you company, and I will greatly miss sharing an office with you. Additionally, thanks to everyone else in the Harwood lab for them welcoming me in and being my lab family. A big thanks to Monika (for your wonderful support and encouragement), Macarena (for your constant joy and help with media making), Kayleigh (for the great conversations and help with media making), Jo (for being my other office neighbour and friend (even if you don't want to admit it)), Azka (also for your fantastic conversation and also being a great office neighbour friend), Alison (for all your expertise with transformation and your updates on your global adventuring), Rachel (for your conversations and introducing me to baby Arthur (even if you didn't let me teach him how to do PCRs)) and Tom (for all your knowledge and interesting conversation on everything gene editing!). An especially big thank you to Penny (for your amazing

friendship, encouragement and support throughout the PhD), Mark (for helping and teaching me the exciting world of cloning and gene jockeying) and Sadiye (for teaching me and introducing me to the world of wheat transformation and gene editing!).

I could not have got through the last four years without the incredible friendships I've had at the Norwich Research Park. I want to thank the Rotation gang, Samuel, Andy, Anson and Sally, for their magnificent friendship and infrequent but enjoyable Rotation lunches. Also, my ground floor Biffen friends and climbing network: Anna, Aura, Jesus, Marina, Max, Isa, Oscar, Daniella, Miki, Jake, Sophie, James, Katie, Rose, Steph, Jen, Liam, Andy and Nikolai. I could write many paragraphs about how much I love each of you, but I am unsure if the acknowledgements are included in the word count. I am very sad to no longer be having lunches in the Biffen coffee room and eating nori (but part of me still questions whether I left the Biffen coffee room; perhaps everything is, in essence, just the Biffen coffee room).

I'd also like to thank my other phenomenal JIC friends: Emma (first for co-chairing the Student Voice with me, second for the skiing holiday, third for all the Bread Source and Café 33 trips and finally for your incredible friendship and support throughout the PhD. You are fantastic, and I am so glad I met you and made you co-chair the JSV with me), Svenja (for first inviting me along to Yarmonics and then your wonderful friendship after that of dancing, ramen and music), Jiawen (for being an incredible friend and all the walks we did during and post-lockdown), Ed & Hannah (for the concerts and introducing me to your incredible cats) Giuseppe (for your permanent optimism and joy, and being my fellow library thesis friend) and Gurpinder (for introducing me into the beautiful world of Sigur Ros and then accompanying me to not one but two Sigur Ros concerts. Your love for music and knowledge of literally everything is awe-inspiring). Also, a massive thank you to The Figs: Franziska, Mau, Alex and Carlos. Learning how to jam, then making music and performing with you has opened my mind to new ways of enjoying life. Each of you has a special place in my heart, and I will greatly miss the music and songs we created. I wish I could include everyone who has been my friend at the NRP, and I am anxious about leaving so many people off. The network of friends I've had around me has been incredible, and I am so lucky to have met and known such amazing people.

I also want to express my great gratitude to my other Norwich friends and family. An especially big thanks to Mel and Sarah, my incredible flatmates, and all the other incredible flatmates I've had over the past four years (Samuel, Magda, Laura, Ella, Fran, Rosh and Matt). A special thanks also to Tom, Ben and Jack for their friendship during a time when I really needed it, the jams we had, and all the dancing and nature exploring we did. I will greatly miss being around you, but I cannot wait to see you all again. Also, thanks to Dan, Stewart, Tasha, Reuben, Tori, Thomas, Sarah, Scott and Magda. I loved the nature and pub trips, and I wish I had had more time to see you all in the last year of being in Norwich.

Additionally, thanks to all my friends outside of Norwich (Eddie, Matt, Nikhil, Doug, Luke, Katie, Ella, Charlotte and Madeleine). It has been fantastic to escape Norfolk and visit you all and thank you for all the times you've made the trip down to visit me. You are the friends I have known the longest, and I cannot express how much I have loved your friendship over the years and growing up with you. I am so grateful to know you and I love you all so much.

Lastly but most importantly, I would like to thank my family for their love and support throughout my PhD and life. Thank you, Ben, for the fun we've had, the phone calls and for being the best brother and friend I could ever ask for. Thank you, Dad, for your support throughout my life and for consoling me whenever I have phoned you up incredibly stressed. Thank you, Granny, for your love and support throughout my life. Finally, a big thank you to my mum and grandpa, who inspired and nurtured a curiosity in the world around me, without which I would not be doing what I am today.

Table of contents

Abstract	<i>i</i>
Acknowledgements	<i>ii</i>
Table of contents	<i>iv</i>
Table of figures	<i>viii</i>
Table of tables	<i>xi</i>
List of abbreviations	<i>xii</i>
1 General Introduction	1
1.1 Carotenoid biosynthesis and their importance	1
1.1.1 The role of carotenoids within plants	1
1.1.2 The carotenoid biosynthesis pathway	1
1.1.3 Degradation of carotenoids	3
1.1.4 Storage of carotenoids.....	4
1.1.5 Importance of carotenoids in human health	5
1.2 The origins and global significance of durum and bread wheat	5
1.2.1 The two main domesticated wheat species.....	5
1.2.2 The shared evolutionary history of durum and bread wheat.....	6
1.2.3 Landraces and the Watkins collections.....	7
1.3 Carotenoids in wheat	9
1.3.1 Carotenoids within wheat grains	9
1.3.2 Quantifying grain carotenoid content in wheat.....	10
1.3.3 Identifying carotenoid-associated quantitative trait loci	11
1.4 Carotenoid biofortification and the <i>ORANGE</i> gene	13
1.4.1 Strategies of carotenoid biofortification	13
1.4.2 The <i>ORANGE</i> gene	17
1.4.3 The cauliflower <i>BoOr^{Mut}</i> mutation	18
1.4.4 The ‘golden SNP’ in the melon <i>ORANGE</i> gene.....	21
1.5 Genome editing in wheat	23
1.5.1 CRISPR/Cas genome editing	23
1.5.2 Base editing	24
1.5.3 Prime editing	27
1.5.4 Wheat transformation and developmental regulators	28
1.6 Thesis Aims	29

2	Materials and Methods	31
2.1	Contributions to this thesis.....	31
2.2	Plant materials	32
2.2.1	Growing conditions.....	32
2.2.2	Watkins global landrace collection	32
2.2.3	Durum wheat <i>Tdor</i> mutants.....	34
2.2.4	Bread and durum wheat <i>TaOR</i> overexpression lines	35
2.2.5	Bread wheat <i>TaOR</i> 'golden SNP' prime editing lines	37
2.2.6	Bread wheat CRISPR/Cas9 <i>BoOr^{Mut}</i> mimic lines.....	37
2.2.7	Field grown <i>OR</i> TILLING lines.....	37
2.3	Carotenoid phenotyping.....	38
2.3.1	High-throughput yellow pigment content (YPC).....	38
2.3.2	High-performance liquid chromatography (HPLC).....	39
2.4	Other phenotyping methods	40
2.4.1	Grain morphometrics	40
2.4.2	Anthesis and height	40
2.4.3	Relative leaf chlorophyll content	40
2.5	Genome-wide association study (GWAS) methods.....	40
2.5.1	GWAS on 35K Axiom® Breeder's Array data	40
2.5.2	GWAS on high-resolution whole-genome sequence data	41
2.6	Sequence analysis and bioinformatics	42
2.6.1	Analysis of QTL regions identified within the Watkins tetraploid collection	42
2.6.2	Identification of durum wheat carotenoid biosynthesis genes.....	42
2.6.3	Analysis of wheat <i>OR</i> homoeologues and <i>OR</i> orthologues	42
2.6.4	Searching for wheat <i>OR</i> allelic diversity.....	43
2.6.5	Investigating <i>TaOR</i> sequence suitability for base editing.....	43
2.6.6	Searching for EMS mutations within K4596	44
2.7	Golden Gate Cloning and construct design.....	44
2.7.1	<i>TaOR^{WT}</i> and <i>TaOR^{His}</i> overexpression constructs.....	46
2.7.2	Prime editing constructs.....	46
2.7.3	CRISPR/Cas9 constructs to mimic <i>BoOr^{Mut}</i>	47
2.8	Transformation of organisms	48
2.8.1	<i>Agrobacterium tumefaciens</i> transformation	48
2.8.2	Wheat transformation	48
2.9	Genotyping.....	49
2.9.1	KASP genotyping.....	49
2.9.2	Transgene copy number analysis.....	50

2.9.3	Sanger sequencing for prime editing and CRISPR/Cas9 editing.....	51
2.9.4	Illumina next-generation sequencing for prime editing	52
2.10	Statistics and data visualisation	54
3	<i>Investigating the grain carotenoid diversity of the Watkins global landrace collection and searching for associated allelic variation</i>	55
3.1	Chapter Introduction	55
3.2	Results	58
3.2.1	The Watkins tetraploid collection has high variability in carotenoid content traits	58
3.2.2	The Watkins hexaploid core collection has a slightly lower grain carotenoid content than the Watkins tetraploid collection	68
3.2.3	Genome-wide association study of grain carotenoid content using the Watkins tetraploid collection reveals marker-trait associations and quantitative trait loci	71
3.3	Discussion	83
3.3.1	Large variation in grain carotenoid content exists within the Watkins tetraploid collection...	83
3.3.2	Higher YPC and carotenoid content in the modern cultivar Miradoux compared to the Watkins landraces may reflect historic durum wheat breeding targets	87
3.3.3	Utilising the Watkins global landrace collection for carotenoid biofortification	88
3.3.4	Novel marker-trait associations and quantitative trait loci associated with carotenoid traits exist within the Watkins tetraploid collection	90
4	<i>Characterising the function of the ORANGE gene and the ‘golden SNP’ substitution in wheat.....</i>	95
4.1	Introduction	95
4.2	Results	97
4.2.1	Wheat orthologues of <i>OR</i> are located on Chromosome 6A, 6B, 6D	97
4.2.2	Knocking out the <i>OR</i> gene reduces grain carotenoid content	101
4.2.3	Overexpressing the ‘golden SNP’ <i>OR</i> gene increases total grain total carotenoid content ...	104
4.2.4	There is very low allelic diversity within <i>OR</i> in wheat.....	112
4.3	Discussion	114
4.3.1	<i>OR^{WT}</i> plays a role in grain carotenoid biosynthesis in wheat	114
4.3.2	<i>OR^{His}</i> increases grain carotenoid content in wheat.....	116
4.3.3	<i>GRF4-GIF1</i> , <i>TaOR^{WT}</i> and <i>TaOR^{His}</i> overexpression affects grain and physiology in wheat	119
4.3.4	Engineered strategies must be used for <i>OR</i> carotenoid biofortification.....	122
5	<i>Engineering and inducing variation into the ORANGE gene in wheat</i>	123
5.1	Introduction	123
5.2	Results	125

5.2.1	Attempting to install the ‘golden SNP’ through prime editing.....	125
5.2.2	Attempting to mimic the <i>BoOr^{Mut}</i> mutation in <i>TaOR</i> using CRISPR/Cas9 editing.....	132
5.2.3	Searching for gain-of-function <i>OR</i> EMS mutations that increase grain carotenoid content ..	138
5.3	Discussion	145
5.3.1	Low Plant Prime Editor 2 efficiency is likely why the ‘golden SNP’ could not be installed in endogenous <i>TaOR</i>	145
5.3.2	The sequence context of <i>TaOR</i> makes mimicking the cauliflower <i>BoOr^{Mut}</i> mutation more challenging than <i>OsOR</i>	147
5.3.3	Extended time in tissue culture may allow for improved gene editing efficiencies.....	150
5.3.4	<i>ORLIKE</i> may also play a role in grain carotenoid biosynthesis in wheat	151
5.3.5	The E141K substitution could be a novel <i>OR</i> gain-of-function mutation that increases grain carotenoid content	152
6	General Discussion	154
6.1	Summary of this thesis	154
6.1.1	Carotenoid diversity within the Watkins global landrace collection.....	154
6.1.2	Allelic diversity within the Watkins tetraploid collection	155
6.1.3	The function of <i>OR</i> and the ‘golden SNP’ in wheat.....	155
6.1.4	Engineering diversity into wheat <i>OR</i> to increase grain carotenoid content	157
6.2	Future approaches for breeding carotenoid biofortified wheat	158
6.2.1	Breeding for high grain carotenoid bread and durum wheat	158
6.2.2	Utilising variation within the carotenoid biosynthesis pathway of the Watkins global landrace collection.....	160
6.2.3	Increasing grain carotenoid content in wheat through transgenesis	160
6.2.4	Using the E141K substitution for increasing grain carotenoid sink strength	161
6.3	GWAS to gene editing: a model for allele replacement in plant breeding	162
6.3.1	A framework for using allele replacement in pre-breeding.....	162
6.3.2	The benefits of using gene editing for allele replacement	164
6.3.3	The UK regulatory framework governing gene editing technologies	165
6.4	Concluding Statement	166
7	References.....	167
Appendices		193
Appendix 1 – List of Appendix Materials		193
Appendix 2 – Appendix Materials of Chapter 3		195
Appendix 3 – Appendix Materials of Chapter 4		216
Appendix 4 – Appendix Materials of Chapter 5		226

Table of figures

Figure 1.1 The carotenoid biosynthesis pathway in plants.....	2
Figure 1.2 The carotenoid content and conversions of plastids.	4
Figure 1.3 Durum wheat and bread wheat evolution and hybridisation events.	7
Figure 1.4 The layers of the wheat grain and the carotenoid compounds found within these.....	10
Figure 1.5 The three common strategies of manipulating the carotenoid biosynthesis pathway for biofortification.....	13
Figure 1.6 The BoOr ^{Mut} mutation in cauliflower causes an overaccumulation of carotenoids in the cauliflower curd.....	19
Figure 1.7 The strategy to mimic BoOr ^{Mut} in rice using CRISPR/Cas9 that was used by Endo and colleagues (2019) and Kim and colleagues (2022).	20
Figure 1.8 The ‘golden SNP’ within the CmOR gene is responsible for the orange-flesh fruit of melons.	21
Figure 1.9 Phenotypes of overexpressing OR ^{WT} and OR ^{His} compared to the non-transgenic control in Arabidopsis calli, tomato fruits and sweet potato tuber.....	22
Figure 1.10 Schematic of genome editing with CRISPR/Cas9.....	24
Figure 1.11 Schematic of genome editing with base editing and prime editing.	27
Figure 2.1 Field plan for the Watkins tetraploid collection sown in the 2020–2021 field season. ..	33
Figure 2.2 Crossing diagram for generating homozygous double knockout Tdor mutant lines and homozygous wild-type lines.	34
Figure 2.3 Lineage and zygosity of the TaOR ^{WT} , TaOR ^{His} and GRF4-GIF1 overexpression lines used in this study.	36
Figure 2.4 KlusterCaller plot for KASP genotyping using specific KASP primers for the K0329 and K4335 mutations.	50
Figure 3.1 Yellow pigment content (YPC) of the field-grown Watkins tetraploid collection and Miradoux field controls, measured using the high-throughput YPC method.....	59
Figure 3.2 Distribution of grain carotenoid content for α -carotene, β -carotene, lutein, zeaxanthin and total carotenoid content in the Watkins tetraploid collection measured by HPLC.....	62
Figure 3.3 Sorted bar charts showing α -carotene, β -carotene, lutein and zeaxanthin content for accessions within the field-grown Watkins tetraploid collection.	63
Figure 3.4 Stacked bar chart of the content of individual carotenoid compounds making up total grain carotenoid content within the Watkins tetraploid collection.....	65
Figure 3.5 HPLC chromatograms of the accessions with the minimum, median and maximum total grain carotenoid content (GCC).	65
Figure 3.6 HPLC chromatogram of accession WAT1180105 with the highest α -carotene content in the Watkins tetraploid collection.	66

Figure 3.7 The proportion of β - β branch carotenoids and provitamin A carotenoids within the Watkins tetraploid collection.....	68
Figure 3.8 Sorted stacked bar chart showing the distribution of total carotenoids in the 106-accession Watkins hexaploid core collection.....	69
Figure 3.9 Comparison of grain carotenoids in the 106-accession Watkins hexaploid core collection and the 337-accession Watkins tetraploid collection measured using HPLC.	70
Figure 3.10 Manhattan plots for GWAS using 35K Axiom® Breeder’s Array genotypic data associated with YPC, α -carotene, β -carotene, lutein, zeaxanthin and total GCC diversity.	75
Figure 3.11 Manhattan plots for carotenoid content traits of a GWAS using the Watkins tetraploid collection genome sequence data.....	79
Figure 3.12 Chromosomal distribution of carotenoid biosynthesis genes in tetraploid wheat, and the MTAs and QTLs associated with carotenoid content traits of the Watkins tetraploid collection.	81
Figure 4.1 Diagram of TaOR homoeologues within the wheat genome.....	98
Figure 4.2 Alignment of OR protein sequences from melon (<i>Cucumis melo</i>) and bread wheat (<i>Triticum aestivum</i>).....	99
Figure 4.3 Phylogenetic tree of the protein sequences of 35 OR orthologues.....	100
Figure 4.4 Grain carotenoid content of Tdor mutants compared to Tdor wild-type plants.	102
Figure 4.5 Preliminary plant growth and grain physiology measurements of Tdor mutants.....	103
Figure 4.6 Plasmid map of the TaOR ^{WT} and TaOR ^{His} overexpression constructs.	105
Figure 4.7 Dark-grown seed-derived callus of pAct-OR ^{WT} and pAct-OR ^{His} T ₁ Cadenza immature embryos.....	106
Figure 4.8 Grain carotenoid content of T ₁ TaOR ^{WT} and TaOR ^{His} overexpression plants compared to GRF control plants presented as box plots.	108
Figure 4.9 Grain carotenoid content of T ₁ TaOR ^{WT} and TaOR ^{His} overexpression lines compared to GRF and wild-type controls presented as stacked bar charts.....	108
Figure 4.10 Grain yield per plant and anthesis date of T ₁ TaOR ^{WT} and TaOR ^{His} overexpression lines compared to GRF and wild-type controls.	110
Figure 4.11 Height, thousand-grain weight, grain area, grain number per plant and relative leaf chlorophyll content of T ₁ TaOR ^{WT} and TaOR ^{His} overexpression lines.....	111
Figure 5.1 Potentially available PAM sites of Cas variants for editing the ‘golden SNP’ cytosine..	126
Figure 5.2 Sequence context of the ‘golden SNP’ in the wheat OR gene and pegRNA locations for prime editing.	127
Figure 5.3 Structure of the prime editing guide RNA #1 (pegRNA#1) for installing the ‘golden SNP’ into the endogenous OR gene in wheat.	128
Figure 5.4 Plasmid map of the prime editing constructs to install the ‘golden SNP’ in the endogenous wheat OR gene.	131

Figure 5.5 Location of sequencing primers over the ‘golden SNP’ nucleotide and example Sanger sequencing chromatograms.	131
Figure 5.6 The PAM site used to mimic BoOr ^{Mut} in rice OsOR using CRISPR/Cas9 and the equivalent PAM sites in wheat TaOR.	133
Figure 5.7 Plasmid map of the TaOR CRISPR/Cas9 editing constructs to mimic the BoOr ^{Mut} mutation from cauliflower.	133
Figure 5.8 Dark-grown calli from immature Cadenza embryos transformed with pTaOR-Cas9. ...	134
Figure 5.9 Staked bar charts showing the carotenoid content of the edited T ₁ pTaOR-Cas9 grains compared to non-edited T ₁ controls of similar copy number.	136
Figure 5.10 Predicted consequences of the editing events found in the pTaOR-Cas9 T ₀ transgenic plants on the TaOR protein.	138
Figure 5.11 Domain annotations and Alpha Fold protein models of TaOR and TaORLIKE.	142
Figure 5.12 Grain carotenoid content of Kronos and Cadenza TILLING lines with EMS mutations in OR and ORLIKE.	144
Figure 5.13 Intron sequence of TaOR is more challenging to mimic BoOr ^{Mut} using CRISPR/Cas9 than OsOR.	148
Figure 5.14 Potential LbCas12a PAM sites within the third exon of TaOR that could be used for a random mutagenesis approach.	149
Figure 6.1 Framework for using allele replacement in pre-breeding.	163

Table of tables

Table 2.1 Names, affiliations and contributions of people who have contributed to this thesis.....	31
Table 2.2 Golden Gate reaction components.	44
Table 2.3 Golden Gate reaction conditions for Eco31I and Bpil reactions.....	45
Table 2.4 Parts used for Golden Gate Cloning within the MoClo system.	45
Table 2.5 pegRNAs designed by PlantPegDesigner for installing the ‘golden SNP’ into the endogenous TaOR homoeologues.....	47
Table 2.6 Oligonucleotide sequences for guide RNAs targeting TaOR homoeologues to mimic the BoOr ^{Mut} mutation.	48
Table 2.7 KASP primers for genotyping the Tdor mutants.	50
Table 2.8 Primer and probe sequences used for quantitative real-time PCR for determining the transgene copy number.	51
Table 2.9 Primer sequences used for the Sanger sequencing of editing events.....	52
Table 2.10 Components for the Sanger sequencing amplicon clean-up reaction.	52
Table 2.11 Primer sequences for Illumina next-generation sequencing of prime editing lines.	53
Table 3.1 Carotenoid content of the Watkins tetraploid collection and Miradoux field controls....	60
Table 3.2 Results of three replicates of HPLC on the high-carotenoid accession WAT1180004.	67
Table 3.3 Results of the three-replicate HPLC screen of the 106-accession Watkins hexaploid core collection.	69
Table 3.4 Significant carotenoid-associated marker-trait associations identified by a GWAS using 35K Axiom® Breeder’s Array genotypic data.	75
Table 3.5 Significant carotenoid-associated quantitative trait loci identified by a GWAS using the Watkins tetraploid collection genome sequence data.	76
Table 4.1 Grain carotenoid content of the Tdor mutants compared to the Tdor wild-type plants.	102
Table 4.2 Preliminary plant growth and grain physiology measurements of Tdor mutants and Tdor wild-type plants.	103
Table 4.3 Grouped average grain carotenoid content of the T ₁ TaOR ^{WT} and TaOR ^{His} overexpression plants compared to GRF control plants.	109
Table 4.4 Grouped averages of other phenotypes of the T ₁ TaOR ^{WT} and TaOR ^{His} overexpression plants compared to the GRF and wild-type controls.	112
Table 5.1 Editing events from 102 T ₀ transgenic pTaOR-Cas9 plants.	135
Table 5.2 Description and carotenoid content of Kronos and Cadenza EMS TILLING mutations identified within OR and ORLIKE.	142

List of abbreviations

ABA	Abscisic acid
<i>Agrobacterium</i>	<i>Agrobacterium tumefaciens</i>
AO	ALDEHYDE OXIDASE
<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>
ARC	ACCUMULATION AND REPLICATION OF CHLOROPLASTS
BLINK	Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway
<i>BoOr^{Mut}</i>	Dominant carotenoid overaccumulation mutation within <i>BoOR</i>
bp	Base pair
CCD	CAROTENOID CLEAVAGE DIOXYGENASE
CHLI	MAGNESIUM CHELATASE SUBUNIT I
CO	CONSTANS-LIKE
CRTISO	CAROTENE ISOMERASE
CSV	Comma-separated value
CT	Cell trays
CYP97A	CYTOCHROME P450-TYPE β -RING HYDROXYLASE
CYP97C	CYTOCHROME P450-TYPE ε -RING HYDROXYLASE
DMAPP	Dimethylallyl diphosphate
DSB	Double-strand break
DXR	DXP REDUCTOISOMERASE
EMS	Ethyl methanesulfonate
epegRNA	Engineered prime editing guide RNA
ePPE	Engineered Plant Prime Editor
ePPEplus	Engineered Plant Prime Editor plus
FarmCPU	Fixed and random model Circulating Probability Unification
GAPIT	Genome Association and Prediction Integrated Tool
GCC	Grain carotenoid content
GGPP	Geranylgeranyl diphosphate
GIF	GRF-INTERACTING FACTOR
GR1	First generation of Golden Rice [®]
GR2	Second generation of Golden Rice [®]
GRF	GROWTH-REGULATING FACTOR

GRU	Germplasm Resource Unit
GWAS	Genome-wide association study
HC106	Watkins Hexaploid Core 106-accession collection
HDR	Homology-directed repair
HPLC	High-performance liquid chromatography
hptII	HYGROMYCIN PHOSPHOTRANSFERASE
HYD	β -CAROTENE HYDROXYLASE
Indels	Insertions or deletions
IPP	Isopentenyl diphosphate
KASP	Kompetitive Allele-Specific PCR
LCYB	LYCOPENE β -CYCLASE
LCYE	LYCOPENE ϵ - CYCLASE
LD	Linkage disequilibrium
M-MLV RT	Moloney murine leukaemia virus reverse transcriptase
MAF	Minor allele frequency
Mbp	Megabase pair
MEP	Methylerythritol 4-phosphate
MgCh	MAGNESIUM-CHELATASE
MLM	Mixed linear model
MTA	Marker-trait association
nCas9	Nickase Cas9
NCED	9-CIS-EPOXYCAROTENOID DIOXYGENASE
NHEJ	Non-homologous end joining
NLS	Nuclear localisation signal
NXS	NEOXANTHIN SYNTHASE
OR	ORANGE
OR^{His}	ORANGE with a histidine at the 'golden SNP' residue
OR^{WT}	ORANGE with an arginine at the 'golden SNP' residue
PAM	Protospacer adjacent motif
PARC	PARALOG OF ARC
PBS	Primer binding sequence
PDS	PHYTOENE DESATURASE
PE	Prime Editor
pegRNA	Prime editing guide RNA

PPE	Plant Prime Editor
PSY	PHYTOENE SYNTHASE
PVA	Provitamin A
qPCR	Quantitative real-time PCR
QTL	Quantitative trait locus
ROS	Reactive oxygen species
RT	Retention time
RT-template	Reverse transcriptase template
SEM	Standard error of the mean
sgRNA	Single-guide RNA
SNP	Single-nucleotide polymorphism
ssp.	Subspecies
TALEN	Transcription activator-like effector nuclease
TGW	Thousand-grain weight
TILLING	Targeted Induced Local Lesions in Genomes
TPM	Transcripts per million
TW337	Watkins Tetraploid 337-accession collection
UGI	Uracil glycosylase inhibitor
VAD	Vitamin A deficiency
var.	Variety
VDE	VIOLAXANTHIN DE-EPOXIDASE
WAT	Watkins collection accession
WSB	Water-saturated butanol
WT	Wild-type
YI	Yellow index
YPC	Yellow pigment content
ZDS	ζ-CAROTENE DESATURASE
ZEP	ZEAXANTHIN EPOXIDASE
ZFN	Zinc finger nuclease
ZISO	ζ-CAROTENE ISOMERASE

1 General Introduction

1.1 Carotenoid biosynthesis and their importance

1.1.1 The role of carotenoids within plants

Carotenoids are a complex group of C₄₀ isoprenoid pigments synthesised in all photosynthetic organisms, including plants, algae, cyanobacteria and some non-photosynthetic fungi and bacteria (Zheng et al. 2020a). In plants, carotenoids are essential antioxidant compounds in protecting the photosynthetic apparatus from photooxidation, where they scavenge and prevent the formation of reactive oxygen species (ROS) that can oxidise and destroy chlorophylls and cellular components (Niyogi 1999). Their role is so crucial for protecting photosynthetic apparatus that carotenoid biosynthesis mutants have bleached photosynthetic tissues (Qin et al. 2007). Moreover, carotenoids play a crucial role in photosynthesis as accessory light-harvesting pigments, absorbing blue-green light and transferring the energy to the chlorophylls (Nisar et al. 2015). Carotenoids also act as the precursors for the biosynthesis of abscisic acid (ABA) and strigolactones, important phytohormones involved in growth, development and response to abiotic stress (Sun et al. 2018). As organic pigments, carotenoids absorb wavelengths of light ranging from 400–550 nm (violet to green light). As such, they are responsible for the many deep yellow, orange and red colours of fruits, roots and flowers in plants. Examples include orange α -carotene and β -carotene from carrots and sweet potatoes, red lycopene from tomatoes and watermelons, and yellow lutein from marigold flowers (Nisar et al. 2015). Carotenoid pigments are also responsible for some of the yellows and oranges of autumn leaves lacking chlorophyll, which normally masks the carotenoids within mature leaves (Sanger 1971). This contribution to the colours of fruits, roots and flowers has significant ecological value, as they serve as attractants for pollinators and frugivores, enhancing the pollination process and seed dispersal (Sun et al. 2022).

1.1.2 The carotenoid biosynthesis pathway

The precursors of carotenoid biosynthesis are the two isoprene isomers, isopentenyl diphosphate (IPP) and its allylic isomer dimethylallyl diphosphate (DMAPP) derived from the methylerythritol 4-phosphate (MEP) pathway (Sun et al. 2022). IPP and DMAPP undergo a series of sequential condensation reactions to produce geranylgeranyl diphosphate (GGPP) mediated by GGPP SYNTHASE (Figure 1.1).

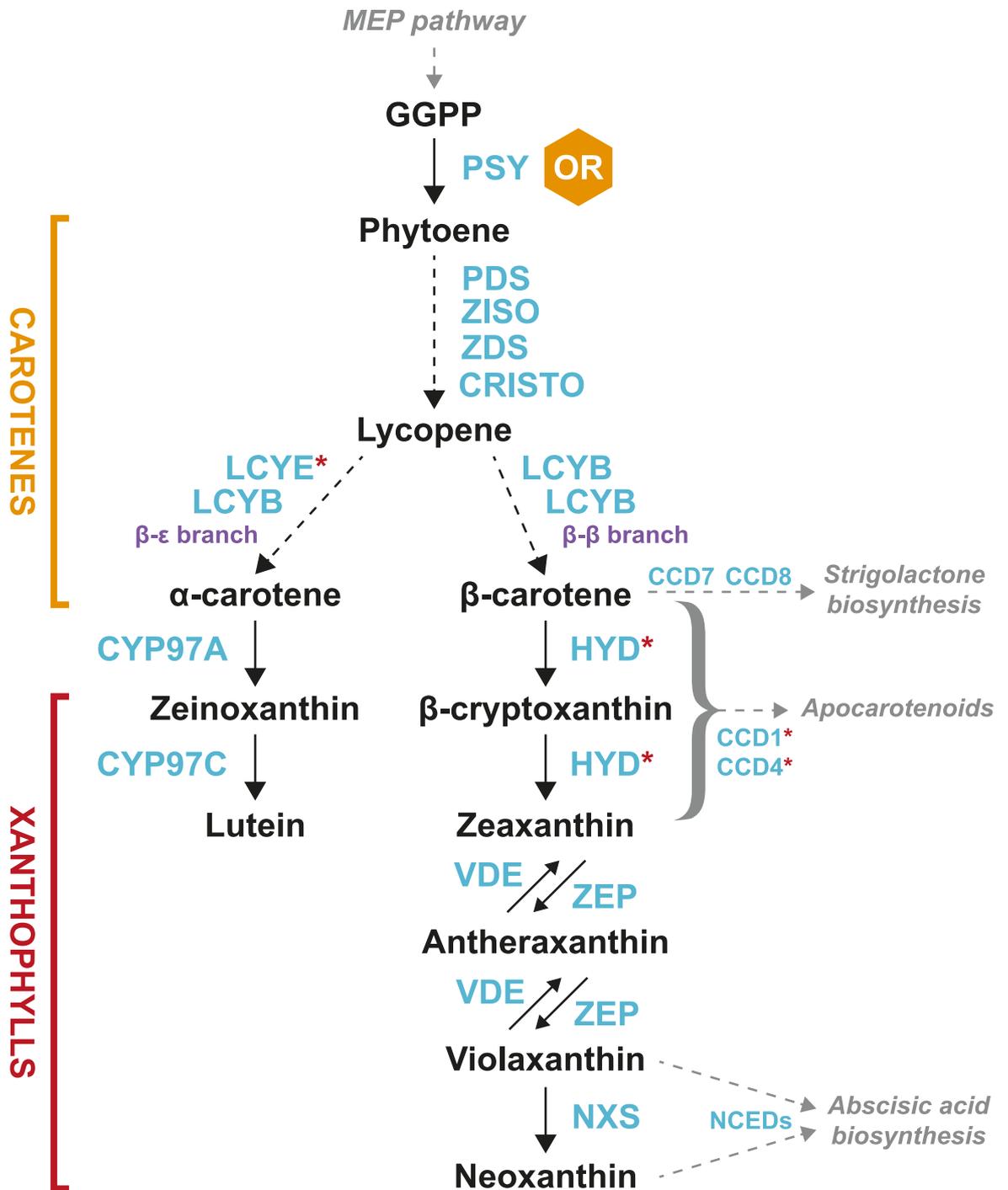


Figure 1.1 The carotenoid biosynthesis pathway in plants. The target enzymes of previous ‘block’ strategies are indicated with red asterisks. The ORANGE protein (OR) is thought to act as a chaperone to regulate PSY activity. Pathway based on those in Sun and colleagues (2022) and Watkins and Pogson (2020).

The first committed step of the carotenoid biosynthesis pathway takes place in plastids, starting with the condensation of two GGPP molecules by PHYTOENE SYNTHASE (PSY) into the carotene, phytoene (Nisar et al. 2015). This reaction is considered the rate-limiting step, and PSY activity is assumed to control the carotenoid metabolic pool size (Zheng et al. 2020a). A series of desaturation and isomerisation reactions follow, converting the colourless phytoene into the red lycopene. In plants, these reactions are catalysed by

PHYTOENE DESATURASE (PDS), ζ -CAROTENE DESATURASE (ZDS), ζ -CAROTENE ISOMERASE (ZISO) and CAROTENE ISOMERASE (CRTISO). In contrast, bacteria use a single phytoene desaturase enzyme (CRTI) to mediate the conversion of phytoene into lycopene (Sandmann 2021; Sun et al. 2022).

At this point, the biosynthesis pathway bifurcates into two branches based on the type of ionone ring introduced into lycopene: β -carotene with two β -rings in the β - β branch of the pathway or α -carotene with one β -ring and one ϵ -ring in the β - ϵ branch of the pathway. These cyclisation reactions are catalysed by LYCOPENE β -CYCLASE (LCYB) and LYCOPENE ϵ -CYCLASE (LCYE). The hydroxylation of these cyclic carotenes leads to the production of xanthophylls, which are carotenoids containing oxygen atoms. Hydroxylation of α -carotene leads to the production of zeinoxanthin and lutein, catalysed by CYTOCHROME P450-TYPE β -RING HYDROXYLASE (CYP97A) and CYTOCHROME P450-TYPE ϵ -RING HYDROXYLASE (CYP97C). The hydroxylation of β -carotene produces β -cryptoxanthin and zeaxanthin, catalysed by two β -CAROTENE HYDROXYLASEs, HYD1 and HYD2 (also referred to as BCH1 and BCH2). Zeaxanthin in the β - β branch is then epoxidated by ZEAXANTHIN EPOXIDASE (ZEP) into violaxanthin. Violaxanthin can be de-epoxidated back into zeaxanthin by VIOLAXANTHIN DE-EPOXIDASE (VDE) within the xanthophyll cycle, which serves a crucial role in protecting plants at high-light intensity (Jahns and Holzwarth 2012). Concluding the core pathway, NEOXANTHIN SYNTHASE (NXS) catalyses the conversion of violaxanthin into neoxanthin, with these two xanthophylls acting as precursors for ABA biosynthesis. Carotenoids can also be esterified in plants, which enhances their stability (Watkins and Pogson 2020).

1.1.3 Degradation of carotenoids

As molecules containing unstable conjugated double bonds, carotenoids continuously degrade in cells and sometimes at a high rate (Sun et al. 2022). Carotenoids undergo specific enzymatic oxidative breakdown catalysed by the carotenoid cleavage oxygenase family of enzymes producing ABA and strigolactones (Figure 1.1). This family comprises of 9-CIS-EPOXYCAROTENOID DIOXYGENASEs (NCEDs) and CAROTENOID CLEAVAGE DIOXYGENASEs (CCDs) in plants. NCEDs specifically cleave violaxanthin and neoxanthin to form xanthoxin, the first step of the ABA biosynthesis pathway (Tan et al. 2003). CCD7 and CCD8, two CCD subfamilies, sequentially cleave β -carotene to produce carlactone for strigolactone biosynthesis (Alder et al. 2012). Other CCDs, such as CCD1 and CCD4, reduce carotenoid content by cleaving a wide range of carotenoids at various double-bond positions into

apocarotenoids (Vogel et al. 2008; Gonzalez-Jorge et al. 2013). Carotenoid degradation can also be carried out by nonspecific enzymes or photochemical oxidation (Sun and Li 2020).

1.1.4 Storage of carotenoids

Carotenoids are stored in plastids, with different types of plastids having dramatically different abilities to accumulate carotenoids (Figure 1.2) (Sun et al. 2018). Amyloplasts are starch-storing plastids found in seeds, roots and tubers, accumulating generally low amounts of the carotenoids lutein, zeaxanthin and violaxanthin (Lopez et al. 2008; Wurtzel et al. 2012). However, high levels of carotenoids can accumulate in amyloplasts when flux into the carotenoid biosynthesis pathway is increased through transgenesis (Paine et al. 2005; Bai et al. 2016). Within chloroplasts, the carotenoids lutein, β -carotene, violaxanthin and neoxanthin accumulate, but chlorophylls mask their colours (Sun et al. 2022). Chromoplasts are the main plastid that synthesise and accumulate very high amounts of carotenoids within flowers, fruits and roots. They contain diverse kinds and amounts of carotenoids that vary based on the tissue and species (Sun et al. 2018). Chromoplasts are fully developed plastids derived from all other plastids. They commonly derive from chloroplasts in fruits and vegetables such as tomatoes (Suzuki et al. 2015) or melon flesh (Tzuri et al. 2015). Within non-photosynthetic tissues, chromoplasts derive from proplastids and amyloplasts, such as in an orange cauliflower mutant (Li et al. 2001), carrot roots (Kim et al. 2010) and papaya (Schweiggert et al. 2011). This plastid conversion into chromoplasts is normally followed by a massive accumulation of carotenoids within these tissues (Sun et al. 2018).

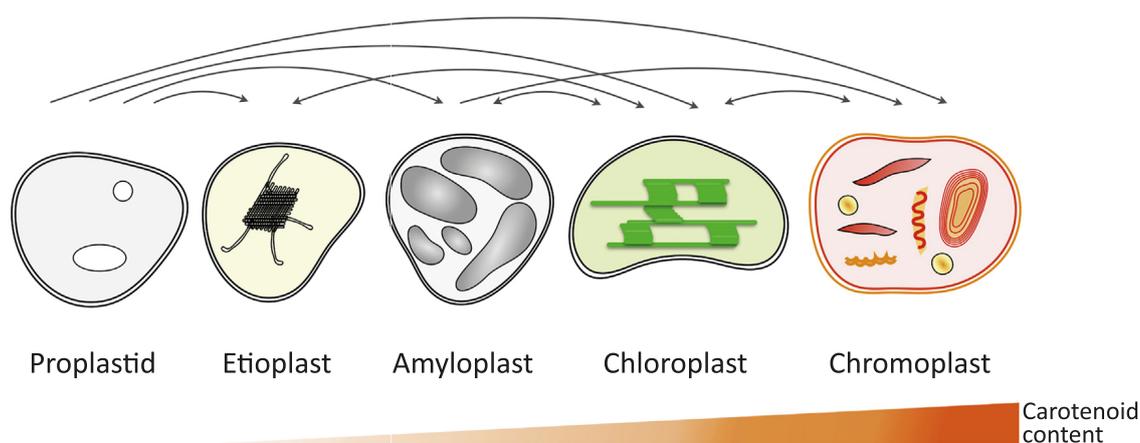


Figure 1.2 The carotenoid content and conversions of plastids. Proplastids are the progenitors of the other plastids. Etioplasts are dark-developing plastids with low carotenoid content. Amyloplasts have a low-medium amount of carotenoid accumulation. Chloroplasts synthesise and store carotenoids and chlorophylls. Chromoplasts are carotenoid-accumulating plastids. Arrows indicate the interconversions of plastid types. Figure from Sun and colleagues. (2018).

1.1.5 Importance of carotenoids in human health

Humans do not synthesise carotenoids, but these compounds play essential roles in human health and are considered crucial components of a healthy diet (Watkins and Pogson 2020). Lutein and zeaxanthin accumulate in the macula of our eyes, the central region responsible for high-resolution colour vision. An increased consumption of these macular carotenoids is thought to decrease the risk of age-related macular degeneration (Ma et al. 2012). Lutein has also been suggested to play a role in brain development and cognitive performance, with post-mortem brain lutein concentrations correlating with pre-mortem cognitive function (Johnson 2014; Erdman et al. 2015). In addition, carotenoids are important antioxidants and ROS scavengers, which may help protect against age-related diseases (Zheng et al. 2020b).

Carotenoids that contain a non-substituted β -ionone ring (namely β -carotene, β -cryptoxanthin and α -carotene) can be cleaved into retinal by the body, the first step in vitamin A (retinol) synthesis (Zheng et al. 2020a). These are known as provitamin A (PVA) carotenoids, with the β - β branch carotenoid β -carotene being the major and most effective PVA carotenoid due to it having two non-substituted β -ionone rings. Insufficient levels of PVA carotenoids in the diet cause vitamin A deficiency (VAD), a severe global health problem that leads to childhood mortality and preventable paediatric blindness (Watkins and Pogson 2020). VAD predominantly affects the stages of life with the highest nutritional demand, such as early childhood and pregnancy. An estimated 250,000–500,000 children lose their sight every year from VAD, and half of them die within 12 months of becoming blind (WHO 2024). VAD is especially prevalent in low-income countries where people rely on starchy staple crops low in PVA carotenoids, such as wheat.

1.2 The origins and global significance of durum and bread wheat

1.2.1 The two main domesticated wheat species

Wheat is an important staple crop providing over 20% of global calorie intake as well as being the most widely cultivated crop, with over 219 million hectares grown worldwide in 2022 (FAO 2023). Two main domesticated species are cultivated today: the hexaploid bread wheat (*Triticum aestivum*) and the tetraploid durum wheat (*T. turgidum* ssp. *durum*). Bread wheat is the main wheat species grown and feeds a considerable proportion of the world's population in products such as bread, chapatis and noodles. Durum wheat only accounts for around 8% of wheat's total cultivated area, but this is still a considerable area

approximately equal to that dedicated to potatoes (Colasuonno et al. 2019; FAO 2023). Durum cultivars tend to have harder grains, a yellow semolina colour and protein content that suits it as an important raw material for finished products such as pasta and couscous (Mastrangelo and Cattivelli 2021).

In contrast to bread wheat's globally distributed cultivation, durum wheat cultivation is primarily located within the Mediterranean basin's semiarid spring habitats. Here, cultivation within Italy, Spain, France, Greece, Morocco, Algeria, Tunisia, Turkey and Syria accounts for around 75% of global production (Broccanello et al. 2023). For much of the population within these countries, durum wheat constitutes the dominant part of the diet, acting as a staple food. Durum wheat also supplies 20–50% of daily caloric intake and 20% of protein intake for 1.2 billion people living in low-income countries, making it an important crop for food security (Broccanello et al. 2023). Historically, durum wheat and its ancestor wild emmer (*T. turgidum* ssp. *dicoccoides*) have played pivotal roles in the diets of various civilisations, from the Neolithic Age through the Greek and Roman periods to today (Martínez-Moreno et al. 2020).

1.2.2 The shared evolutionary history of durum and bread wheat

Wheat traces its origins to ancient hybridisation events among grass species within the Fertile Crescent (Figure 1.3) (Salamini et al. 2002; Katamadze et al. 2023). Initially, around 360,000–500,000 years ago, a hybridisation occurred between two diploid species: the wild grass *T. urartu* ($2n=2x=14$, A^UA^U) and a species closely related to *Aegilops speltoides* ($2n=2x=14$, SS). Through allopolyploidy, this event led to the creation of tetraploid wild emmer wheat (*T. turgidum* ssp. *dicoccoides*; $2n=4x=28$, AABB), characterised by a brittle rachis and hulled grains (Yang et al. 2022). Domestication of wild emmer in the Fertile Crescent, occurring around 10,000–12,000 years ago, resulted in domesticated emmer (*T. turgidum* ssp. *dicoccum*; $2n=4x=28$, AABB), with a non-brittle rachis but still retaining hulled grains (Dubcovsky and Dvorak 2007). Subsequently, between 7,500–9,000 years ago, natural mutations in domesticated emmer led to the development of free-threshing durum wheat (*T. turgidum* ssp. *durum*; $2n=4x=28$, AABB) (Feldman and Kislev 2007). Separately, domesticated emmer underwent a second hybridisation event with the wild goat grass *Ae. tauschii* ($2n=2x=14$, DD), producing hexaploid bread wheat (*T. aestivum*; $2n=6x=42$, AABBDD) approximately 7,000–10,000 years ago (Dubcovsky and Dvorak 2007; Katamadze et al. 2023). The close evolutionary history of durum wheat and bread wheat, both having originated from domesticated emmer wheat, makes it possible to transfer genes between

these tetraploid and hexaploid genotypes through interspecific crosses (Mastrangelo and Cattivelli 2021).

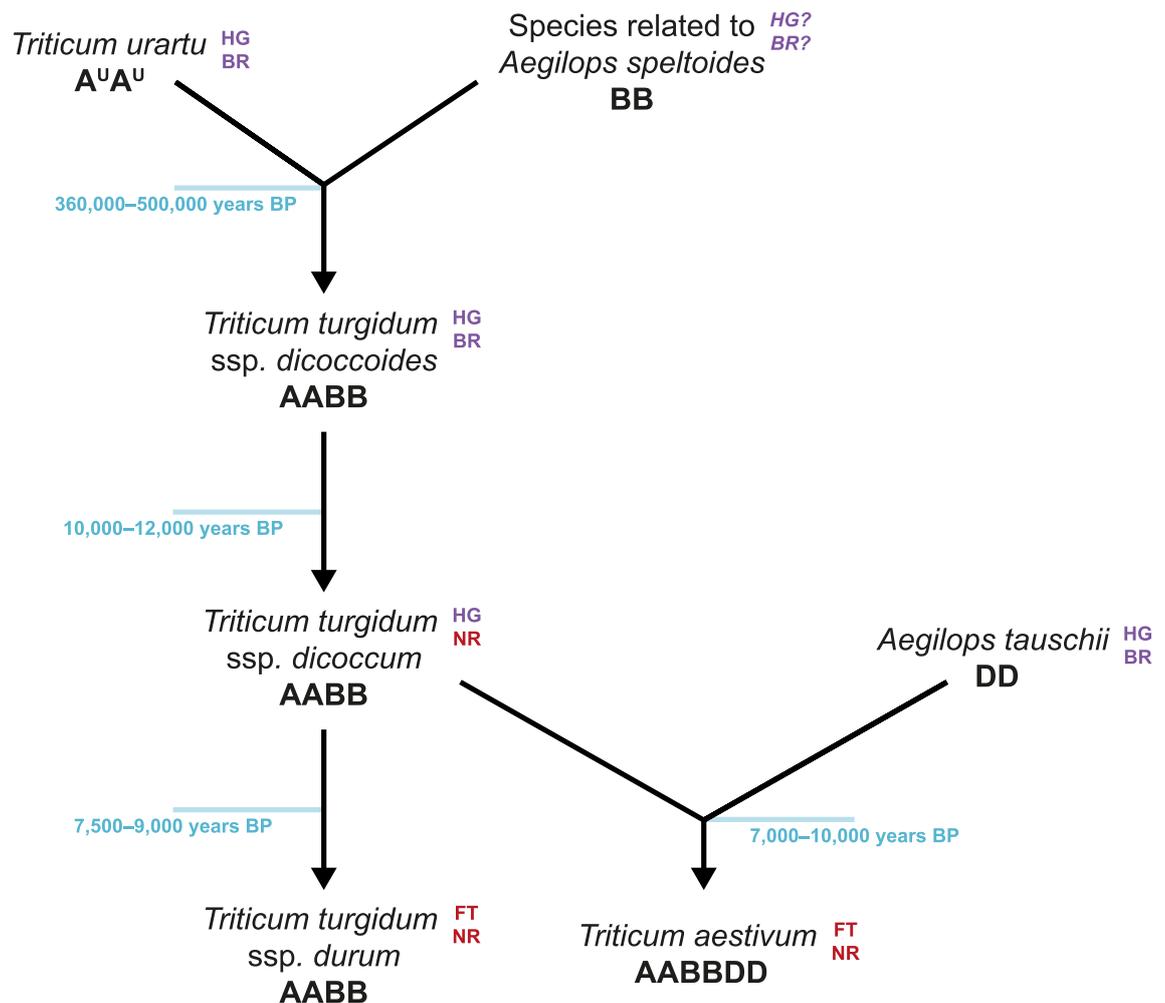


Figure 1.3 Durum wheat and bread wheat evolution and hybridisation events. Dates are based on those provided by Katamadze and colleagues (2023) and Dubcovsky and Dvorak and colleagues (2007). Wild and domesticated traits are indicated in purple and red, respectively. Presumably, the B-genome progenitor species related to *Aegilops speltoides* had wild traits. HG=hulled grains, FT=free-threshing, BR=brittle rachis, NR=non-brittle rachis, BP=before present.

1.2.3 Landraces and the Watkins collections

Durum wheat became a prominent crop around 2,300 years ago after its spread around the Mediterranean basin, replacing previous domesticated emmer wheat cultivation (Broccanello et al. 2023). Durum wheat landraces, which were adapted to their place of origin, were predominantly cultivated from this period until the 1950s (Martínez-Moreno et al. 2020). Landraces are heterogeneous populations of cultivated plants lacking modern crop improvement, often very genetically diverse and locally adapted to their cultivated region (Villa et al. 2005). Modern durum wheat breeding started in Italy with the release of Senatore Cappelli in 1915, which appeared in the pedigree of almost all new varieties until the end of the 1960s (Laidò et al. 2013). After this, breeding activities were increased by

private seed companies and international research centres as efforts were made to release new varieties more often with high yield potentials and other desirable traits (Broccanello et al. 2023).

Over time, the durum wheat landraces grown for centuries were progressively replaced by more genetically homogeneous modern cultivars. This led to a loss of potentially useful genetic variation and environmental adaptations that had evolved for thousands of years within these landraces. The increased genetic diversity of landraces compared to modern cultivars allows them to act as reservoirs of useful allelic diversity that can broaden the variation of important agronomic traits. Due to the efforts of farmers and scientists, wheat landraces have been collected and conserved within seed and gene banks. With modern phenotyping and genotyping technologies, the diversity of desirable traits can be explored within these collections (Nazco et al. 2012; Cheng et al. 2023).

One such landrace collection is the Watkins global landrace collection, consisting of a hexaploid and tetraploid collection. The Watkins tetraploid collection comprises 356 durum wheat landraces from 25 countries, and the Watkins hexaploid collection comprises 828 bread wheat landraces from 32 countries. Assembled by Arthur Ernest Watkins at Cambridge University in the late 1920s to early 1930s, this collection offers a unique snapshot of global bread and durum wheat diversity before the introduction of modern breeding practices (Wingen et al. 2014). The Watkins hexaploid collection has already demonstrated its value in gene discovery and pre-breeding. A substantial effort has been made to characterise both the phenotypic and genotypic diversity within these accessions, revealing a greater variability in agronomically important traits than modern cultivars (Wingen et al. 2014). Moreover, the Watkins hexaploid collection was extensively characterised and whole-genome re-sequenced to allow for high-resolution genome-wide association studies (GWAS). This exploration identified thousands of high-resolution quantitative trait loci (QTLs) and significant marker-trait associations (MTAs) for major traits, many unique to this collection (Cheng et al. 2023). The Watkins hexaploid collection has also been used to derive a core collection, which is a subset of the complete collection, representing its diversity with minimum redundancy (De Beukelaer et al. 2018).

However, the Watkins tetraploid collection has largely been overlooked by the UK crop science community due to a frameshift labelling mistake in a regeneration field around 50 years ago, losing the geographical data for these accessions. Recently, members of the Germplasm Resource Unit at the John Innes Centre have stabilised the heterogeneous

landrace diversity through three single-seed descent generations, reducing the intrinsic diversity. It was then genotyped using the 35K Axiom® Breeder's Array, as well as being whole-genome re-sequenced by Professor Shifeng Cheng's group at the Agricultural Genomics Institute at Shenzhen. Given the Watkin tetraploid collection's potential for uncovering unique allelic diversity, there is now a concerted effort to assess its phenotypic and genotypic diversity more thoroughly.

1.3 Carotenoids in wheat

1.3.1 Carotenoids within wheat grains

Wheat grains contain a variety of carotenoids such as lutein, zeaxanthin, α -carotene, β -carotene and β -cryptoxanthin (Figure 1.4) (Lachman et al. 2017). Here, carotenoids act as important antioxidants that reduce grain ageing by decreasing free radical levels and peroxidase activity (Howitt and Pogson 2006). Within the starchy endosperm of wheat grains, amyloplasts are the main plastids that synthesise and accumulate carotenoids, primarily the macular carotenoids lutein and zeaxanthin (Howitt and Pogson 2006; Wurtzel et al. 2012). Consequently, lutein is typically the most abundant carotenoid in wheat, accounting for 80–90% of total carotenoids, with zeaxanthin making up 7–8% of total carotenoids. The important PVA carotenoids α -carotene and β -carotene account for 3–8% of grain carotenoids and are mainly located within the embryo, which has a higher total carotenoid content than the endosperm (Ndolo and Beta 2013; Shewry and Hey 2015; Colasuonno et al. 2019). The aleurone layer has a lower total carotenoid content than the endosperm and the embryo (Ndolo and Beta 2013; Masisi et al. 2015). The PVA carotenoid β -cryptoxanthin is less abundant than α -carotene or β -carotene within the grain (Digesù et al. 2009).

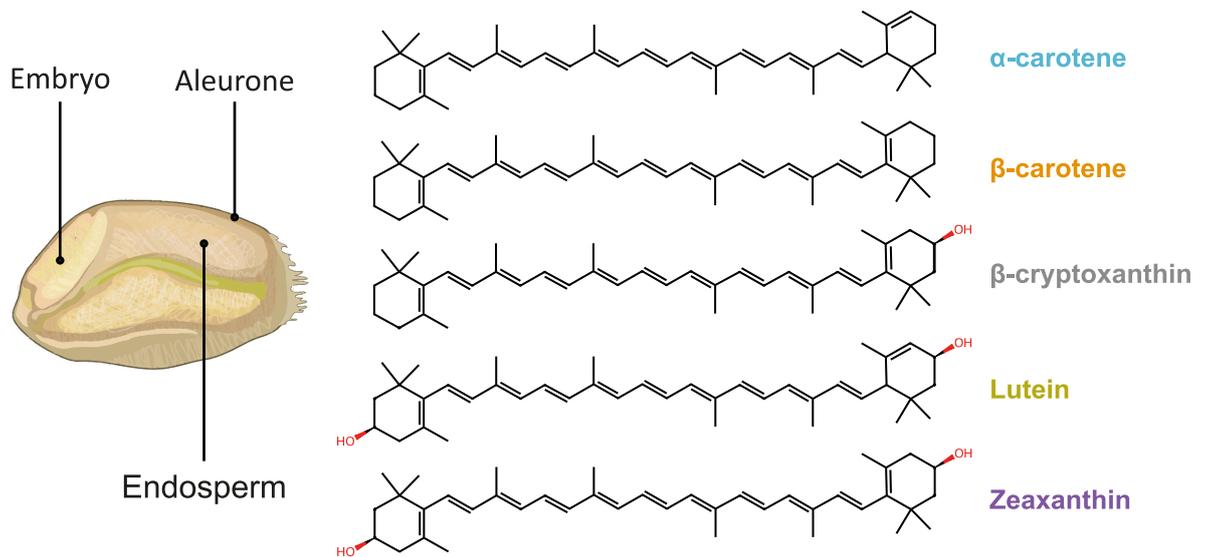


Figure 1.4 The layers of the wheat grain and the carotenoid compounds found within these. The carotenes α -carotene and β -carotene contain no oxygen atoms, whereas the xanthophylls β -cryptoxanthin, lutein and zeaxanthin contain a hydroxyl group. The embryo has the highest total carotenoid content in the grain, followed by the endosperm and the aleurone. The grain diagram was adapted from Brinton and Uauy (2019).

Carotenoid pigments contribute to the yellow colour of the flour (Shewry and Hey 2015). In the past 30 years, durum wheat cultivars have been bred for increased yellow flour colour, a preferred quality trait for pasta production (Digesù et al. 2009; Clarke et al. 2010). Therefore, total grain carotenoid content tends to be higher in modern durum wheat cultivars (2.69–8.38 $\mu\text{g/g}$, this and all following measurements of grain carotenoid content are in grams dry weight) than in modern bread wheat cultivars (1.40–4.90 $\mu\text{g/g}$). In contrast, historic consumer preference for white bread and the avoidance of whole grain products throughout the 20th century contributed to the low carotenoid content of modern bread wheat cultivars (O’Neil et al. 2010; Smith et al. 2019). More recently, there has been a growing trend of consumer preference towards wholemeal and brown bread, allowing for consumer acceptance of healthier high-carotenoid bread wheat lines (Mancino and Kuchler 2012; Lockyer and Spiro 2020).

1.3.2 Quantifying grain carotenoid content in wheat

The carotenoids are responsible for the yellow colour of wheat flour, so analysing the yellowness of flour is a measurement of the relative grain carotenoid content (Colasuonno et al. 2017a). Due to flour colour being an important quality trait in durum wheat breeding, several techniques have been developed to evaluate this parameter over the years. The main methods to do this are measuring the yellow index (YI) or the yellow pigment content (YPC) of flour. YI measures the yellow intensity of flour, which is directly related to carotenoid content. YPC involves spectrophotometric quantification on yellow pigments

extracted from the flour using water-saturated butanol. The absorbance of this extract is measured at 436 nm and used to calculate YPC, which is the wavelength maximum absorption of lutein (Colasuonno et al. 2019). Industry-standard methods exist for YPC measurements (the AACC 14-50 method), although these use significant quantities of flour and water-saturated butanol alongside lengthy extractions, making these impractical for large-scale screening (AACC 2009). Therefore, micro-methods have previously been developed that use less flour, solvent and extraction time to overcome these challenges (Beleggia et al. 2010).

For highly accurate measurements of carotenoid content and the quantities of individual carotenoid compounds within wheat grains, high-performance liquid chromatography (HPLC) analysis is used (Fu et al. 2017). This process begins with extracting carotenoids from wheat flour using specific solvents that effectively dissolve these compounds. Following this, the solution is passed through a column packed with a stationary phase, designed to separate the carotenoid compounds based on their chemical characteristics, such as polarity and molecular weight. This separation occurs because different compounds interact differently with the stationary phase and thus travel at different speeds through the column. As the separated compounds exit the column, they are detected by a photodiode array detector capable of measuring the compounds' absorbance across a range of wavelengths. This allows for their identification and quantification, facilitated by comparing these compounds' absorbances to standard curves generated from reference carotenoid samples, which are known concentrations of carotenoids used to calibrate the HPLC system (Amorim-Carrilho et al. 2014). HPLC gives very precise quantification of the individual carotenoid compounds, but it is more expensive and labour-intensive than alternative methods such as YI or YPC (Colasuonno et al. 2019). For the breeding industry, both YPC and YI methods offer cost-effective means of carotenoid assessment. However, their correlation with HPLC measurements varies significantly: YPC demonstrates a strong correlation ($r=0.89$, $p<0.01$), whereas YI shows a weak correlation ($r=0.30$, $p>0.05$) (Fратиanni et al. 2005).

1.3.3 Identifying carotenoid-associated quantitative trait loci

Grain carotenoid content in wheat is a quantitative trait with high heritability, facilitating the identification of QTLs associated with this trait (Colasuonno et al. 2019). Linkage mapping in biparental populations has been successful in identifying QTLs and candidate genes regulating carotenoid content (Elouafi et al. 2001; Pozniak et al. 2007; Singh et al.

2009; Tsilo et al. 2011; Colasuonno et al. 2014). This technique involves studying the co-segregation of traits and genetic markers within a segregating biparental population, allowing the construction of linkage maps that show the relative positions of carotenoid-associated genetic markers on chromosomes. This gives a high power of QTL detection, but the resolution is often poor due to the strong linkage disequilibrium in wheat, making fine mapping difficult (Gupta et al. 2014). Moreover, developing and utilising biparental populations is costly and time-intensive, discouraging their use for genetic studies and breeding programmes (Shi et al. 2017).

Alternatively, GWAS has provided an effective method to identify QTLs associated with grain carotenoid content (Reimer et al. 2008; Colasuonno et al. 2017a; N'Diaye et al. 2017; Roselló et al. 2018). These involve using diversity panels and estimating the correlations between genotypes and phenotypes based on linkage disequilibrium between alleles of molecular markers and causal genes (Colasuonno et al. 2019). In contrast with biparental mapping, segregating populations need not be established, and more diverse populations can be explored with GWAS, resulting in higher-resolution QTL mapping (Wang et al. 2022b). GWAS uses historical recombination events in diverse populations, which reduce linkage disequilibrium and thereby allow for finer mapping of QTLs. Additionally, it benefits from the consideration of population structure, which can account for genetic diversity and environmental factors, further refining the association between genetic markers and phenotypic traits. Moreover, the availability of high-density single-nucleotide polymorphism (SNP) arrays and cheaper genotype-by-sequencing methods have further improved the utilisation and resolution of GWAS (Torkamaneh and Belzile 2022). Furthermore, the availability of reference genomes of bread wheat (IWGSC RefSeq v1) and durum wheat (Svevo v1) have greatly facilitated the discovery and cloning of genes within associated QTLs (IWGSC et al. 2018; Maccaferri et al. 2019; Sheoran et al. 2022). Most carotenoid-content GWAS have used relative grain carotenoid measurements such as YI or YPC, with only a few studies using HPLC, likely due to the high costs associated with HPLC methods (Guan et al. 2022; Requena-Ramírez et al. 2022). The major QTL associated with carotenoid content is on chromosome 7 of the wheat genomes, which is explained by allelic variations of *PSY1* homoeologues, while minor QTLs have been detected on all chromosomes (Colasuonno et al. 2019).

1.4 Carotenoid biofortification and the *ORANGE* gene

1.4.1 Strategies of carotenoid biofortification

The past three decades of durum wheat breeding have successfully increased the total grain carotenoid content, primarily through increases in the content of the macular carotenoid lutein (Digesù et al. 2009). However, wheat and other staple cereal crops still have low levels of the PVA carotenoids, α -carotene and β -carotene, which remain insufficient to reach significant levels of PVA activity, making other approaches to diversify wheat carotenoid variation necessary (Giuliano 2017; Dias et al. 2018). Given the substantial daily consumption of wheat-based products and wheat's widespread cultivation, even small increases in PVA carotenoid content within bread and durum wheat grains can improve human health worldwide. Efforts to improve the carotenoid content of crops have primarily focused on improving the content of PVA carotenoids such as β -carotene. This has been achieved by manipulating the carotenoid biosynthesis pathway through several strategies: 'push' strategies, 'block' strategies and 'pull' strategies (Figure 1.5).

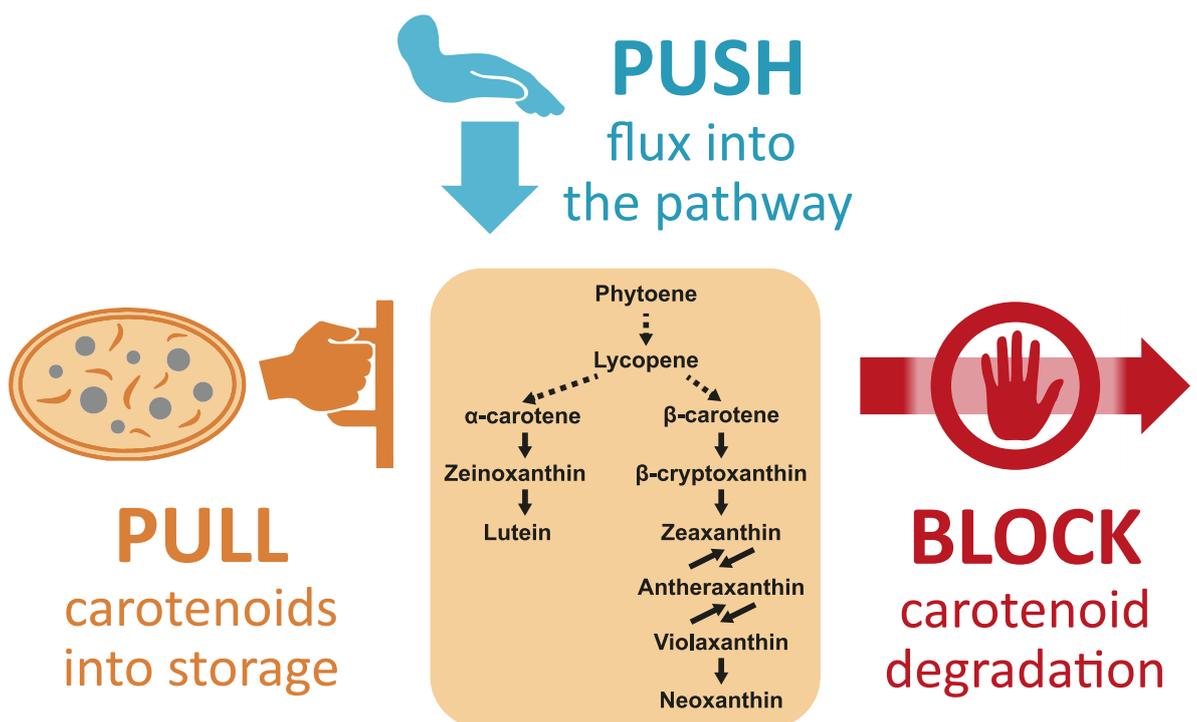


Figure 1.5 The three common strategies of manipulating the carotenoid biosynthesis pathway for biofortification. 'Push' strategies involve increasing the metabolic flux into the carotenoid biosynthesis pathway through the overexpression of one or more biosynthetic enzymes. 'Block' strategies involve reducing the expression of genes downstream to desired carotenoids or genes that degrade carotenoids. 'Pull' strategies involve increasing carotenoid sink capacity and sequestration to promote carotenoid storage.

A common strategy has been to increase metabolic flux into the carotenoid pathway through the overexpression of one or more biosynthetic enzymes, referred to as a 'push'

strategy (Zheng et al. 2020a). In particular, the rate-limiting genes, such as *PSY*, have been frequently overexpressed to produce high-carotenoid crops. The most famous example of this approach used in crops is in the creation of Golden Rice®. The development of rice germplasm with high β -carotene grain accumulation was first decided during a brainstorming meeting in New York in 1993. This workshop was organised by the Rockefeller Foundation in response to the high prevalence of VAD among those populations relying on rice as a staple crop (Welsch and Li 2022). Following this, the first generation of Golden Rice® (GR1) was produced. This was achieved through the endosperm-specific expression of the daffodil (*Narcissus pseudonarcissus*) *NpPSY* gene and the phytoene desaturase gene from the soil bacterium *Erwinia uredovora* (*EuCRTI*) (Ye and Beyer 2000; Beyer et al. 2002). Endosperm expression of *NpPSY* alone resulted in an accumulation of the colourless phytoene, and so *EuCRTI* expression was required to produce lycopene directly from phytoene (Figure 1.1). GR1 plants had a grain carotenoid content of 1.6 $\mu\text{g/g}$; however, this was too low to make a significant contribution to alleviating VAD (Welsch and Li 2022). For the second generation of Golden Rice® (GR2), the maize *ZmPSY1* gene was used instead of *NpPSY* after testing *PSY* genes from different plant species (Paine et al. 2005). In combination with the endosperm-specific expression of *EuCRTI*, this led to a grain carotenoid content of 37 $\mu\text{g/g}$, which was 23 times higher than GR1 and sufficient to reach suitable PVA activity. During the years since the development of GR2, this trait has been introduced into locally adapted rice varieties; however, anti-GMO opposition has hindered its deployment in regions suffering high VAD (De Steur et al. 2022; Welsch and Li 2022). Moreover, despite the significant increase in grain carotenoid content, β -carotene is not a stable compound within rice endosperm. It undergoes oxidative decay and has a half-life of 25 days after harvest in rice grains, reducing the effectiveness of GR2 in alleviating VAD (Welsch and Li 2022). In wheat, a similar combination to that used for GR2 was employed. The endosperm-specific expression of *ZmPSY1* and *EuCRTI* successfully increased endosperm total carotenoid content 10-fold (Cong et al. 2009). A later attempt overexpressed the *Erwinia uredovora* phytoene synthase gene (*EuCRTB*) and *EuCRTI*, which resulted in a 76-fold increase in grain PVA carotenoid content compared with non-transgenic controls (Wang et al. 2014).

'Block' strategies are also used to increase carotenoid content. Here, the expression of genes downstream to desired compounds or enzymes competing for the same substrates is reduced. This decreases flux into competing metabolic routes or reduces the degradation of carotenoids into apocarotenoids (Zheng et al. 2020a). Several genes, including *ZEP*, *LCYE*,

CCDs and the carotenoid hydroxylases (*HYDs* and *CYP97s*), are negative regulators of PVA carotenoid accumulation and are commonly targeted in 'block' strategies (Figure 1.1). In wheat, studies have targeted 'block' strategy genes using ethyl methanesulfonate (EMS) mutations from a TILLING (Targeted Induced Local Lesions in Genomes) population. EMS mutagenesis produces C-to-T transition mutations, resulting in C/G to T/A substitutions (Kumar et al. 2023). Two TILLING populations exist in wheat developed in hexaploid bread wheat (cultivar Cadenza) and tetraploid durum wheat (cultivar Kronos), which have also been exome-captured and sequenced to identify SNPs within coding regions (Krasileva et al. 2017). Using these populations, loss-of-function mutations can be generated through premature truncation variants or missense mutations which are predicted to negatively impact gene function. Two studies have targeted *LCYE* using this wheat TILLING population to increase flux into the β - β branch of the carotenoid biosynthesis pathway, which contains β -carotene, by reducing the flux into the competing β - ϵ branch of the pathway (Figure 1.1). Richaud and colleagues (2018) found a 75% increase in β -carotene in leaves but no differences in grain β -carotene content for their durum wheat single A- and B-genome *lcy*e knockouts compared to controls. Sestili and colleagues (2019) combined *lcy*e mutations in the A- and B-genome of durum wheat and found a 75% increase in grain β -carotene content compared with controls. However, the amount of β -carotene in *lcy*e knockout lines was still low, and the authors suggested the observed upregulation of β -carotene hydroxylase genes (*HYD1* and *HYD2*) could be converting β -carotene into other xanthophylls. Consequently, knocking out *HYD1* using durum wheat TILLING lines increased β -carotene content in the grain by over 70% (Garcia Molina et al. 2021). Moreover, Yu and colleagues (2022) generated several mutant combinations of *lcy*e, *hyd1* and *hyd2*, finding most combinations to significantly increased β -carotene in the endosperm. They also found that eliminating *HYD2* homoeologues is sufficient to prevent β -carotene's conversion into xanthophylls in the grain without compromising leaf xanthophyll production. *CCD4* homoeologues have also been knocked out; however, these did not affect grain carotenoid content (Yu and Tian 2021).

'Push' and 'block' strategies have been combined in wheat through the overexpression of *CRTB* and the RNAi silencing of *HYD1* to both increase flux into the carotenoid biosynthesis pathway and reduce the conversion of β -carotene into xanthophylls (Zeng et al. 2015). Combining these approaches led to a 31-fold increase in grain β -carotene content compared with an increase of 14.6-fold when just *CRTB* was overexpressed or 10.5-fold when just *HYD1* was silenced.

Finally, carotenoid biosynthesis manipulation can be achieved by increasing carotenoid sink capacity and sequestration in a 'pull' strategy. In wheat, the *XAT* gene is responsible for lutein esterification (Watkins et al. 2019). Esterification promotes the sequestration and accumulation of lutein by enhancing its stability in tissues such as the endosperm, so this gene could aid future efforts for carotenoid biofortification. An additional way to increase carotenoid sink capacity could be to increase the formation of chromoplasts, which store massive amounts of carotenoids in highly enriched sequestration substructures (Sun et al. 2018). Dominant gain-of-function mutations of the *ORANGE (OR)* gene have been found to regulate chromoplast formation, and these pose a potential avenue for future carotenoid biofortification attempts in a variety of crops (Watkins and Pogson 2020). The expression of these gain-of-function *OR* variants was found to induce chromoplast biogenesis with enhanced carotenoid accumulation in various crop species (Lopez et al. 2008; Li et al. 2012; Yazdani et al. 2019; Kim et al. 2021). Carotenoids in amyloplast-rich organs, such as the grain endosperm of rice and wheat, are prone to rapid degradation. Due to chromoplasts being better adapted to storing and sequestering carotenoids than amyloplasts, it has been suggested these *OR* variants could be used to promote chromoplast biogenesis and improve carotenoid stability in endosperm tissues (Li et al. 2012). This would be especially applicable to Golden Rice® in preventing the rapid degradation of β -carotene that reduces its efficacy in combating VAD (Welsch and Li 2022). Supporting this idea, the seed-specific expression of the *OR^{His}* variant in *Arabidopsis* seeds induces chromoplast biogenesis and reduces the degradation of carotenoids within seeds when *PSY* and *OR^{His}* are expressed (Sun et al. 2021). Moreover, the overexpression of PLASTID DIVISION 1 increases chromoplast number in *Arabidopsis* seeds, and this could also be used to improve carotenoid sink capacity in crops for 'pull' biofortification strategies (Sun et al. 2020).

In addition to these engineering strategies for carotenoid biofortification, there has been success in generating high PVA orange maize through conventional breeding and marker-assisted selection. Like wheat, the predominant carotenoids in maize are lutein and zeaxanthin, with the PVA carotenoids accounting for only around 10-20% of the total grain carotenoid content (Nkhata et al. 2020). Despite this, conventional breeding efforts by CIMMYT, IITA and HarvestPlus have generated orange maize varieties with PVA carotenoid content between 15–25 $\mu\text{g/g}$, a level sufficient to meet the daily PVA requirement in sub-Saharan Africa (Goredema-Matongera et al. 2021). This is a much greater PVA carotenoid content than in conventional white and yellow maize varieties, which range from undetectable levels to below 2 $\mu\text{g/g}$ (Nkhata et al. 2020). This was achieved due to the

considerable variation in carotenoid content that exists within maize germplasm, with some lines found to have a grain carotenoid content as high as 66 µg/g (Manjeru et al. 2019). However, the carotenoids within the amyloplast-rich maize endosperm also suffer high degradation rates with losses of up to 90% of their grain carotenoid content after a year of storage (Nkhata et al. 2020). Nevertheless, orange maize has sufficient PVA activity to combat VAD in regions relying on maize as a staple crop and is exempt from the restrictive legislation surrounding transgenic crops. Identification of high PVA carotenoid wheat germplasm may allow for a similar conventional breeding approach for carotenoid biofortification.

1.4.2 The *ORANGE* gene

The OR protein is a plastid-localised DnaJ cysteine-rich protein highly conserved among divergent plant species (Lu et al. 2006). OR has chaperone activity and plays a role in carotenoid biosynthesis by directly interacting with and post-transcriptionally stabilising PSY, increasing its protein activity (Figure 1.1) (Zhou et al. 2015; Sun et al. 2023b). In *Arabidopsis thaliana* (*Arabidopsis*) leaves and seed-derived callus, knocking out *AtOR* and its paralogue *AtORLIKE* reduced carotenoid content by lowering PSY protein levels (Zhou et al. 2015; Sun et al. 2023b). Similarly, mutating melon *CmOR* through EMS-induced premature stop codons reduced the carotenoid content of the non-photosynthetic fruit flesh (Chayut et al. 2017). Moreover, the overexpression of *AtOR* increased the carotenoid content in seed-derived callus by promoting PSY protein activity (Yuan et al. 2015; Zhou et al. 2015). The overexpression of *AtOR* in rice calli and white maize endosperm also increased carotenoid content, presumably by increasing PSY protein levels, suggesting that *AtOR* has a similar function between monocots and dicots (Bai et al. 2014; Berman et al. 2017). Additionally, the overexpression of *AtOR*, *ZmPSY1* and *PaCRTI* in rice grains increased carotenoid content more than that of *ZmPSY1* and *PaCRTI* overexpression (Bai et al. 2016). However, *OsOR* overexpression in rice reduced the carotenoid content in leaves and grain-derived calli, and it did not change grain carotenoid content, presumably because of low endogenous carotenoid metabolic flux and enzymatic activity in this tissue (Yu et al. 2021). Prior to this thesis, the function of *OR* in wheat had not been investigated.

In addition to stabilising PSY in the carotenoid biosynthesis pathway, OR has been found to stabilise MAGNESIUM CHELATASE SUBUNIT I (CHLI) in the chlorophyll biosynthesis pathway (Sun et al. 2023b). CHLI is a subunit of MAGNESIUM-CHELATASE (MgCh), a three-component enzyme that catalyses the first committed step in chlorophyll biosynthesis

(Tanaka and Tanaka 2007; Wang and Grimm 2021). Consequently, *Ator Atorlike* mutants in *Arabidopsis* have reduced leaf chlorophyll content (Zhou et al. 2015; Sun et al. 2023b). Therefore, the primary role of *OR* has been suggested to regulate and coordinate the biosynthesis of chlorophyll and carotenoids, both photosynthetic pigments (Sun et al. 2023b). This key role would explain its high level of conservation found within plants. Moreover, *OR* has been found to enhance CHLI and PSY protein stability under heat stress by increasing the levels of carotenoids and chlorophyll during this stress (Sun et al. 2023b). Consequently, the overexpression of *OR* has also been found to increase tolerance to salt, heat and drought stress in *Arabidopsis* (Shan et al. 2022), alfalfa (Wang et al. 2015), tobacco (Wang et al. 2018), and sweet potato (Kim et al. 2013, 2021; Park et al. 2015, 2016; Cho et al. 2016; Kang et al. 2017). The suggested mechanism for this has been attributed to the chaperone activity of *OR* in stabilising CHLI and PSY.

1.4.3 The cauliflower *BoOr^{Mut}* mutation

The *OR* gene was first discovered in a naturally occurring orange curd cauliflower (*Brassica oleracea* var. *botrytis*) mutant discovered in 1971 in a farmer's field (Crisp et al. 1975; Li et al. 2001). The *BoOr^{Mut}* triggers chromoplast differentiation in the cauliflower curd, massively enhancing β -carotene accumulation in this non-photosynthetic tissue (Paolillo et al. 2004; Lopez et al. 2008). When the genotype of *BoOR* was heterozygous for the *BoOr^{Mut}* mutation, the head of the cauliflower had more β -carotene accumulation; however, it was dwarfed in the homozygous state compared to the wild-type and heterozygous cauliflower (Figure 1.6). *BoOr^{Mut}* is caused by a retrotransposon insertion of a *copia* element in the third exon of *BoOR* that disrupts the normal splicing of its transcripts. As a result, three aberrant in-frame transcripts are produced: *BoOr_{Ins}*, with a 13 amino acid insertion; *BoOr_{Del}*, with a 7 amino acid insertion and a 13 amino acid deletion; and *BoOr_{LDel}*, with a 7 amino acid insertion and a 29 amino acid deletion, removing the fourth exon (Figure 1.6). The mechanism for this is thought to be distinct from the role *OR* plays in post-transcriptionally stabilising PSY (Welsch et al. 2020). The overexpression of *BoOr^{Mut}* in potato tubers and white cauliflower also leads to chromoplast differentiation and β -carotene accumulation, suggesting this mutation has a similar role in other plant species (Lu et al. 2006; Lopez et al. 2008; Li et al. 2012).

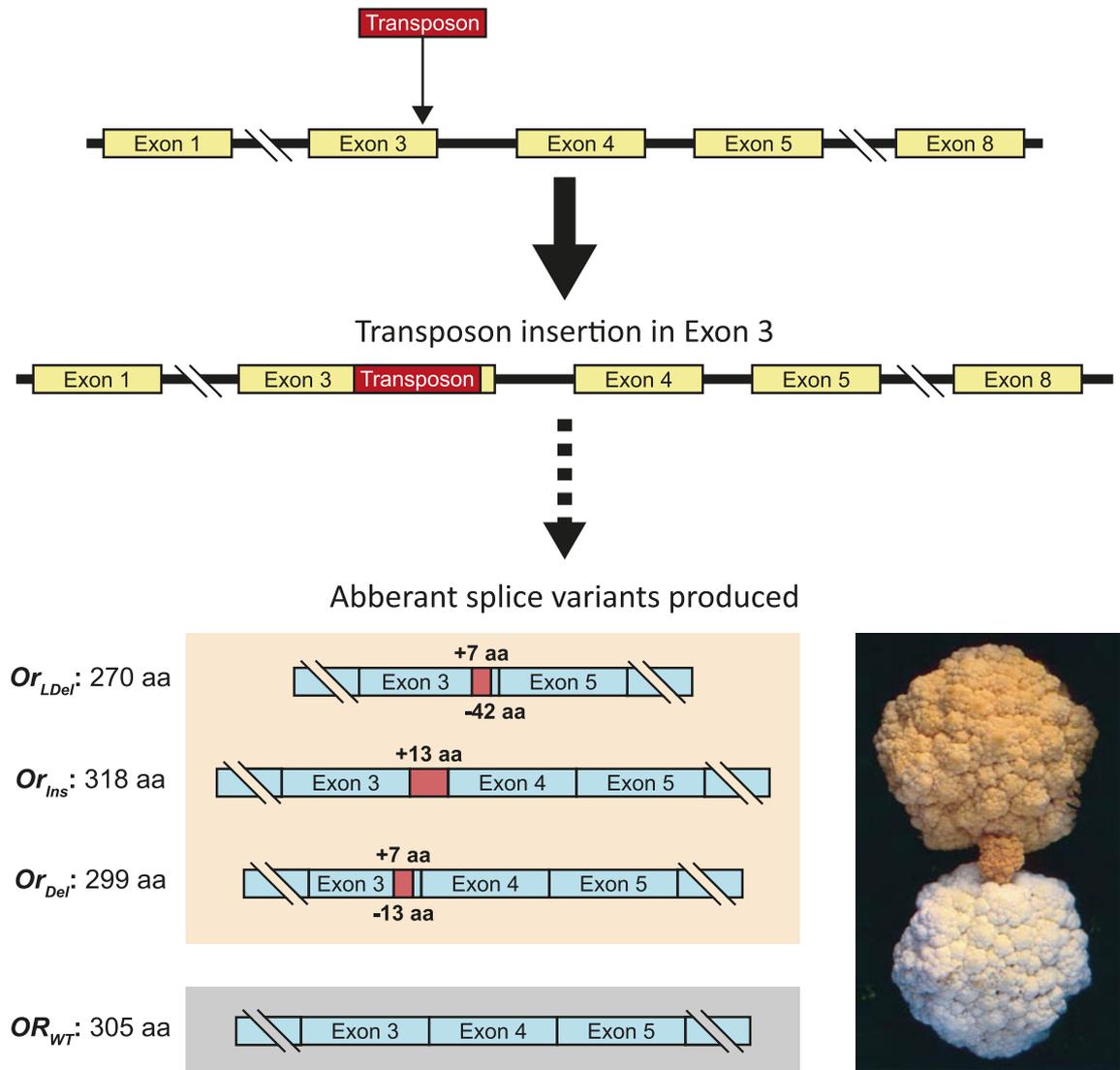


Figure 1.6 The $BoOR^{Mut}$ mutation in cauliflower causes an overaccumulation of carotenoids in the cauliflower curd. This is a result of a transposon integration in the third exon that disrupts the splicing of the $BoOR$ gene producing three aberrant transcripts with an in-frame insertion or deletion of amino acids. The image of the cauliflower shows a wild-type cauliflower on the bottom, a heterozygous $BoOR^{Mut}$ line at the top, and a homozygous $BoOR^{Mut}$ producing a dwarfed phenotype in the middle. Figure adapted from Endo and colleagues (2019) and Li and colleagues (2001). aa=amino acids.

In rice, the $BoOR^{Mut}$ mutation was mimicked in the endogenous $OsOR$ gene using CRISPR/Cas9 gene editing (Endo et al. 2019; Kim et al. 2022). Here, the splice boundary of the third exon and third intron was disrupted, removing the 'GT' splice-donor site and emulating $BoOR_{Ins}$ and $BoOR_{Del}$ transcripts. Carotenoid accumulation was screened by growing callus tissue in the dark, which increases flux into the carotenoid biosynthesis pathway and is commonly used as a visual screen for increased carotenoid biosynthesis (Kim et al. 2013; Bai et al. 2014; Tzuri et al. 2015). Various editing events that removed this splice boundary were generated, but only those that produced aberrant transcripts with an in-frame insertion or deletion of amino acids while retaining the rest of the protein had increased carotenoid accumulation in dark-grown callus (Figure 1.7). These orange dark-

grown calli had 6.8- to 9.7-fold higher total carotenoid content than the wild-type calli. This suggests that *BoOr^{Mut}* functions in cereals; however, when plants were generated from these orange rice calli, grain carotenoid content did not change. This was likely because rice has inactive grain carotenoid biosynthesis, and the *BoOr^{Mut}* is hypothesised to increase sink tissue rather than increase flux into carotenoid biosynthesis. On the other hand, wheat does have active grain carotenoid biosynthesis. Hence, I hypothesised that mimicking *BoOr^{Mut}* may increase grain carotenoid storage capacity, leading to a desirable increase in grain carotenoid content in wheat.

a) Wild-type sequence:

CTTGCCTCCCCTG↓GTAAGTATATACTTC

Aa. sequence: ...IPFLPPLSAANLKI...

b) Edited white callus sequence:

CTTGCCTCCCCTG ----GTATATACTTC

CTTGCCTCCCCTGCGTAAGTATATACTTC

Aa. sequence: ...IPFLPPLKCS*

...IPFLPPLVQLISKSTMLRVSL*

c) Edited orange callus sequence:

CTTGCCTCCCCTG ----ATATACTTC

CTTGCCTCCCCTG ----TATACTTC

Aa. sequence: ...IPFLPPLVQLISKSTMLRVSL*

...IPFLPPLIYTSTSSTISHECS*

...IPFLPPLYIYFYIQHHLTWFIF
KSAANLKI...

...IPFLPPLKYIYFYIQHHLTWFIF
FKSAANLKI...

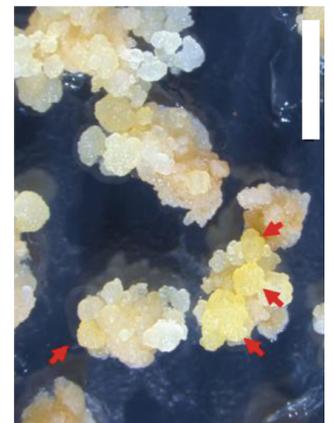
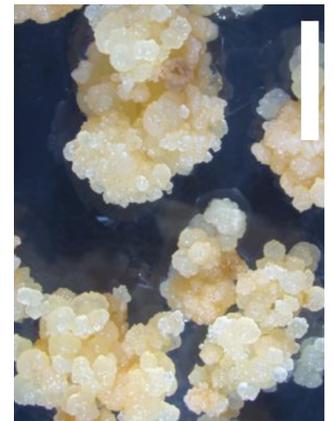


Figure 1.7 The strategy to mimic *BoOr^{Mut}* in rice using CRISPR/Cas9 that was used by Endo and colleagues (2019) and Kim and colleagues (2022). Carotenoid overaccumulating orange segments indicated by red arrows (scale bars=5 mm). (a) Sequence of the wild-type *OsOR* with the CRISPR/Cas9 target site. The third exon sequence is shown in black, and the third intron sequence is shown in blue. The black box highlights the PAM. The 20-nucleotide target site is underlined in red. The red arrow shows the predicted cut site by Cas9. The wild-type amino acid sequence is shown in purple. (b) Sequence of an edited callus that did not overaccumulate carotenoids. The callus segment was mosaic for two editing events, a 4 bp deletion and a 1 bp insertion. Based on mRNA sequences, the predicted amino acid sequences include early stop codons. (c) Sequence of an edited callus showing the carotenoid overaccumulating orange phenotype. The callus segment was mosaic for two events, a 5 bp deletion and a 4 bp deletion. Based on mRNA sequences, the predicted amino acid sequences include two transcripts with an in-frame addition of 16 and 17 amino acids. Sequence information, results and photos are from Endo and colleagues (2019), which this figure was adapted from. Aa.=amino acid.

1.4.4 The 'golden SNP' in the melon *ORANGE* gene

In melons (*Cucumis melo*), a single SNP in *CmOR* was found to be responsible for the massive accumulation of carotenoids in the fruit mesocarp that distinguishes orange-flesh melons from white- or green-fleshed melons (Tzuri et al. 2015). This SNP, named the 'golden SNP', is a G-to-A alteration (CGC-to-CAC) that substitutes a highly conserved arginine residue (OR^{WT}) with a histidine residue (OR^{His}), promoting chloroplast biogenesis and increased β -carotene accumulation in the mesocarp (Figure 1.8). The 'golden SNP' function was shown to be independent of PSY transcriptional, post-translational and enzymatic activity regulation, and how it promotes chloroplast differentiation is still yet to be fully elucidated. In *Arabidopsis*, $AtOR^{His}$ has been shown to interact with ACCUMULATION AND REPLICATION OF CHLOROPLASTS 3 (ARC3) and compete for the binding of ARC3 to PARALOG OF ARC6 (PARC6), whereas $AtOR^{WT}$ does not (Sun et al. 2020). Both ARC3 and PARC6 are crucial regulators of chloroplast division, and their interaction with OR^{His} has been proposed as the mechanism for chloroplast biogenesis. Alternatively, it has been proposed that OR^{His} increases the expression of carotenoid biosynthesis enzymes or reduces the expression of carotenoid degradation enzymes, in turn increasing carotenoid accumulation above a threshold required to stimulate chloroplast biogenesis (Chayut et al. 2017; Kim et al. 2019).

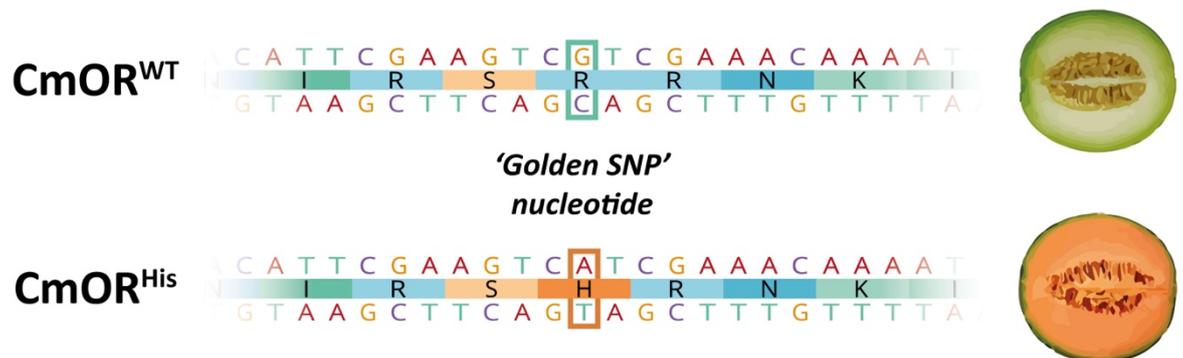


Figure 1.8 The 'golden SNP' within the *CmOR* gene is responsible for the orange-flesh fruit of melons. The 'golden SNP' nucleotide is highlighted within the amino acid sequence of the *CmOR* protein.

In other plants, the overexpression of OR^{His} (containing the 'golden SNP' substitution) has increased carotenoid accumulation more than OR^{WT} overexpression (without the 'golden SNP') (Figure 1.9). OR^{His} overexpression in *Arabidopsis* calli, tomato fruit and sweet potato tubers resulted in higher carotenoid accumulation within these non-photosynthetic tissues than OR^{WT} overexpression (Yuan et al. 2015; Yazdani et al. 2019; Kim et al. 2021). This suggests the 'golden SNP' substitution also increases carotenoid accumulation within other plant species. The 'golden SNP' was introduced to *OR* from the monocot sorghum (*Sorghum*

bicolor), and the overexpression of *SbOR^{His}* in *Arabidopsis* calli resulted in higher carotenoid accumulation compared to the overexpression of wild-type *SbOR^{WT}* (Yuan et al. 2015). This suggests that the ‘golden SNP’ in a monocot *OR* gene has a similar function to the ‘golden SNP’ in a dicot *OR* gene.



Figure 1.9 Phenotypes of overexpressing *OR^{WT}* and *OR^{His}* compared to the non-transgenic control in *Arabidopsis* calli, tomato fruits and sweet potato tuber. *Arabidopsis AtOR^{WT}* and *AtOR^{His}* were overexpressed in *Arabidopsis* calli and tomato fruit. Sweet potato *IbOR^{WT}* and *IbOR^{His}* were overexpressed in sweet potato tuber. In tomato fruit, the orange colour difference is observed at earlier fruit stages before the red lycopene masks the colour. Images from Yuan and colleagues (2015), Yazdani and colleagues (2019) and Kim and colleagues (2021).

Installing the ‘golden SNP’ into the endogenous *OR* gene of staple crops has been suggested as a route for carotenoid biofortification within the grains of these important crops (Li et al. 2012; Sun et al. 2018; Osorio 2019; Watkins and Pogson 2020). Promoting chromoplast biogenesis within this non-photosynthetic tissue is hoped to increase its carotenoid accumulation and storage stability. The only instance of *OR^{His}* overexpression in a cereal grain was in rice, where *OsOR^{His}* overexpression did not increase grain carotenoid content compared to *OsOR^{WT}* or the non-transgenic control (Jung et al. 2021). Rice does not have active carotenoid biosynthesis within the grain due to a lack of *PSY* expression (Beyer et al. 2002); therefore, increasing the sink tissue would likely not affect carotenoid content within the grain, which may explain this finding. However, wheat grains possess active carotenoid metabolic flux, so I hypothesised that if *OR^{His}* has a role in wheat grain carotenoid accumulation, installing the ‘golden SNP’ into the native *OR* gene would likely result in increased carotenoid accumulation within this tissue.

1.5 Genome editing in wheat

1.5.1 CRISPR/Cas genome editing

Genome editing, also known as gene editing, is a technology that allows the insertion, deletion or alteration of DNA at a specific target location in the genome. Several gene editing tools have been developed: zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and CRISPR/Cas nucleases (Elsharawy and Refat 2023). The complicated and costly protein construction associated with the protein-guided ZFNs and TALENs has impeded their use. Instead, the simpler, cheaper and more precise RNA-guided CRISPR/Cas nucleases are now widely employed.

The most commonly used CRISPR/Cas system is CRISPR/Cas9 from *Streptococcus pyogenes* (Jinek et al. 2012). Here, an engineered single-guide RNA (sgRNA) is used to guide the SpCas9 protein to a 20-nucleotide target site upstream of a protospacer adjacent motif (PAM) where the nuclease generates a double-strand break (DSB) at a specific target site in the genome (Figure 1.10). For SpCas9, the PAM site sequence is 'NGG', but Cas proteins isolated from different bacteria have varied PAM and target site requirements (Wang et al. 2020). DSBs generated by CRISPR/Cas9 are then repaired by endogenous repair mechanisms through the error-prone non-homologous end joining (NHEJ) pathway or accurate homology-directed repair (HDR) in the presence of a template sequence (Elsharawy and Refat 2023). In plants, NHEJ is the predominant repair pathway and commonly introduces random insertions or deletions (indels) at the repair site. Plant gene-editing studies have commonly used these resulting indels to introduce targeted mutations within the genome (Li et al. 2021). CRISPR/Cas9 gene editing using the NHEJ repair pathway was previously used to mimic the *BoOr^{Mut}* in rice by introducing indels to disrupt the 'GT' splice-donor site of the third intron (Figure 1.7) (Endo et al. 2019). However, this is unsuited for more precise editing events such as installing the 'golden SNP' into the endogenous *OR* gene. Studies have shown successful editing events generated through HDR in rice, barley and wheat (Li et al. 2019; Lu et al. 2020; Lawrenson et al. 2021; Luo et al. 2023); but this repair pathway is still impractical for most gene editing attempts due to very low editing efficiencies and difficulty in delivery of the donor DNA templates (Awan et al. 2022b). Instead, two precise gene editing technologies could be used: base editing and prime editing.

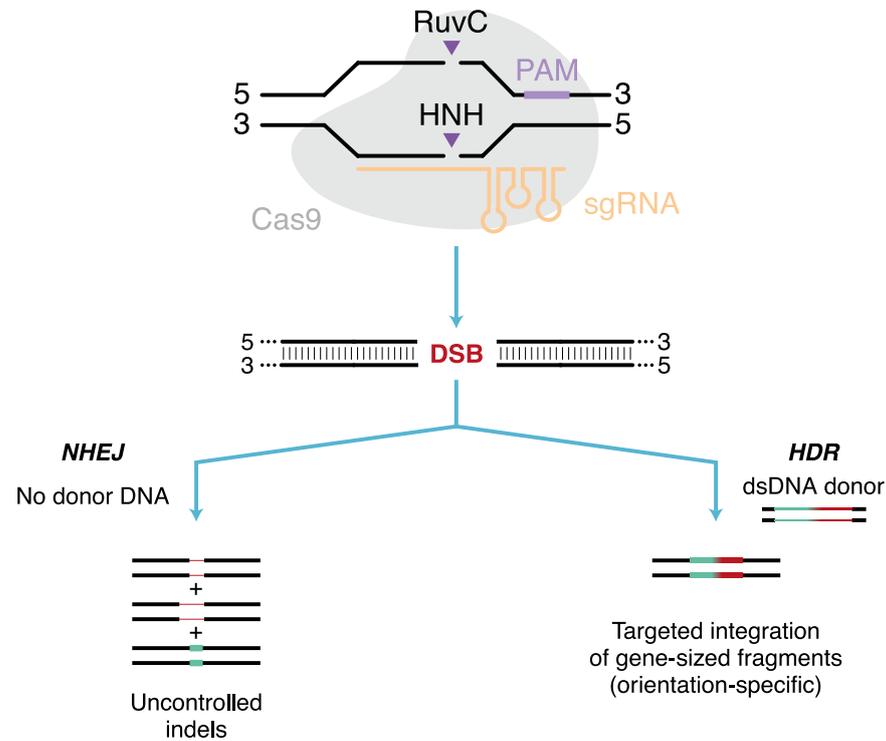


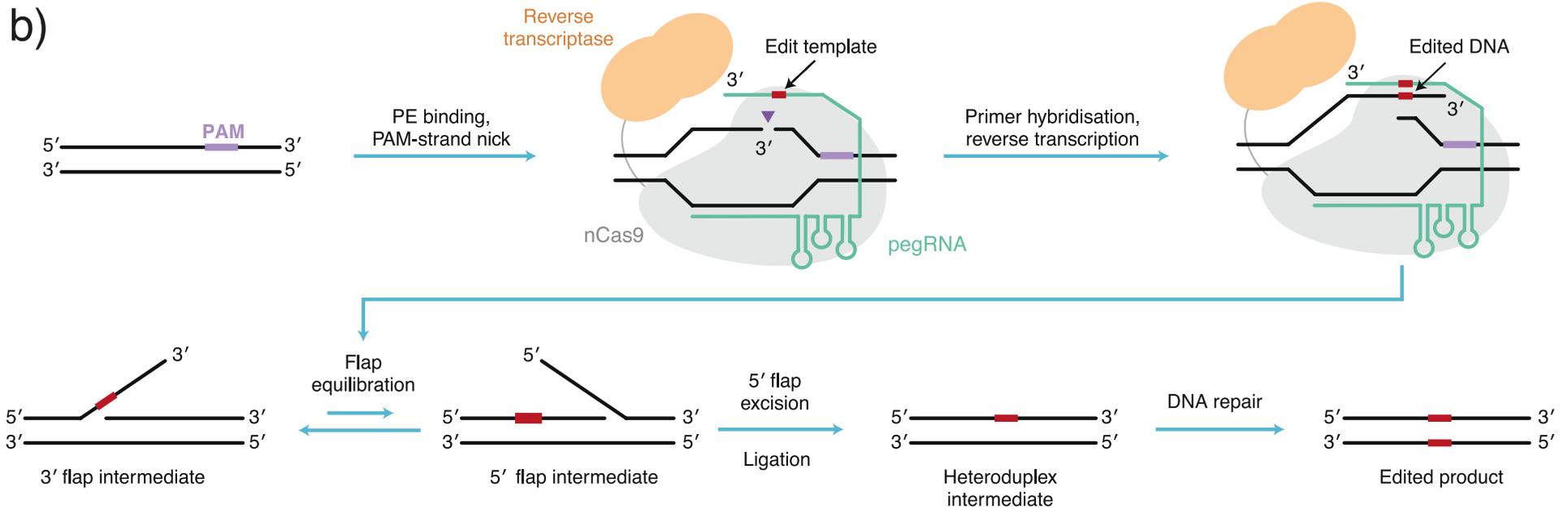
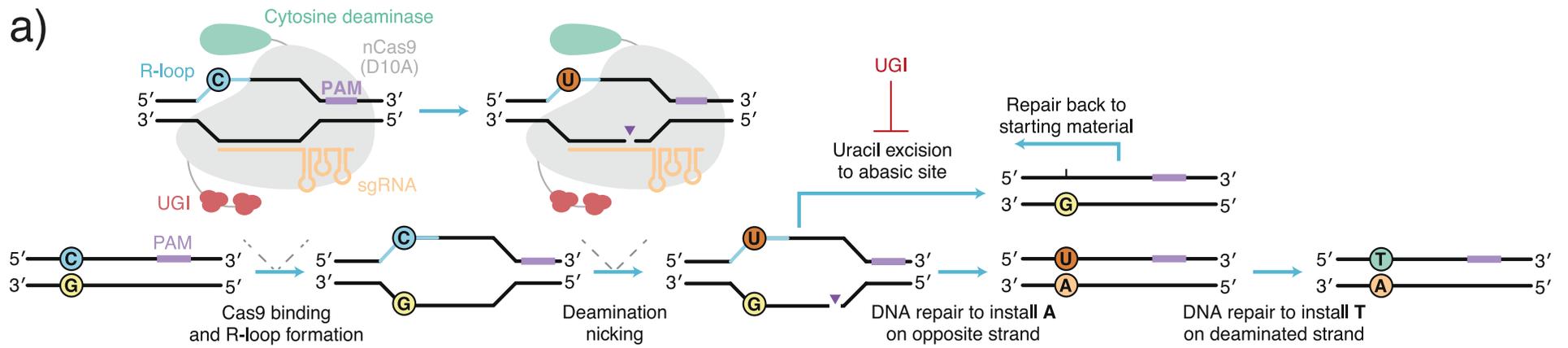
Figure 1.10 Schematic of genome editing with CRISPR/Cas9. RuvC and HNH are the nuclease domains of Cas9. Figure from Anzalone, Koblan and Liu (2020) with modifications. PAM=protospacer adjacent motif, sgRNA=single-guide RNA, DSB=double-strand break, NHEJ=non-homologous end joining, HDR=homology-directed repair.

1.5.2 Base editing

The ‘golden SNP’ is a C-to-T transition on the antisense strand (Figure 1.8), making it possible to introduce this into wheat *OR* using a cytosine base editor. This can precisely install targeted C-to-T point mutations without generating DSBs or requiring DNA donor templates, and base editing has been successfully demonstrated in wheat (Zong et al. 2017, 2018; Zhang et al. 2019). Cytosine base editors contain a catalytically impaired Cas nuclease unable to generate DSBs fused to a cytosine deaminase enzyme (Figure 1.11a), which catalyses the conversion of C/G base pairs to T/A base pairs (Anzalone et al. 2020). Most base editors use Cas nickases such as nCas9 (D10A) to localise the cytosine deaminase enzyme to a specific genomic region of interest (Li et al. 2021). After nCas9 binding, guide RNA hybridisation to the target DNA strand causes the formation of a single-stranded DNA R-loop on the opposite DNA strand. This exposes nucleotides as single-stranded DNA which are accessible to the deaminase domain of the base editor. The cytosine deaminase converts cytosines to uracils within this R-loop (Anzalone et al. 2020). After deamination, stable base editing requires the replacement of the unedited strand to install the corresponding complementary adenine nucleotide opposite the uracil nucleotide. However, uracil is rapidly excised from genomic DNA by uracil DNA N-glycosylase, so cytosine base editors typically include uracil glycosylase inhibitor (UGI) proteins to increase

editing efficiency. Repair of the non-deaminated strand is promoted through the single-strand break introduced by the nCas9. This uses the deaminated strand as a template for resynthesizing the nicked strand (Li et al. 2021).

Only specific nucleotide positions within the R-loop are efficiently deaminated by the base editor, and these are said to be within the 'editing window'. For canonical base editors using nSpCas9 and the APOBEC1 cytosine deaminase enzyme, this window spans positions 4–8 of the protospacer (position 1 being the first nucleotide of the protospacer and position 21–23 being the PAM site). Different Cas variants and deaminase proteins can vary the positions of this editing window (Awan et al. 2022b). Using base editing to install a targeted transition mutation such as the 'golden SNP' requires a suitable PAM site to put the editing window of the cytosine deaminase above the correct target nucleotide. This can limit the suitability of some genomic sites for editing with base editors. Other cytosine nucleotides within the editing window can also be deaminated, leading to unintended bystander editing events at some sites (Anzalone et al. 2020). Moreover, different cytosine deaminases have different sequence context preferences for the cytosines they deaminate, which can also affect sequence context suitability.



On the previous page:

Figure 1.11 Schematic of genome editing with (a) base editing and (b) prime editing. Figure from Anzalone, Koblan and Liu (2020) with modifications. PAM=protospacer adjacent motif, sgRNA=single-guide RNA, DSB=double-strand break, UGI=uracil glycosylase inhibitor, nCas9=nickase Cas9, PE=prime editor, pegRNA=prime editing guide RNA.

1.5.3 Prime editing

Prime editing is a gene editing technology that can introduce all possible types of point mutations (transitions and transversions), small insertions and small deletions in a very precise manner (Anzalone et al. 2019). Prime editors are not limited by the same PAM availability and sequence suitability restrictions as base editors. However, prime editors are reported as having much lower editing efficiencies compared with base editors (Hillary and Ceasar 2022).

Prime editors are a fusion protein between a Cas9 nickase and a reverse transcriptase, usually an engineered Moloney murine leukaemia virus reverse transcriptase (M-MLV RT; Figure 1.11b) (Anzalone et al. 2020). This protein is targeted to the editing site by a prime editing guide RNA (pegRNA), which also encodes the desired edit in an extension at the 3' end of the pegRNA. This extension comprises a reverse transcriptase template (RT-template) with the edit and a primer binding sequence (PBS). Upon nCas9 target binding and nicking of the PAM-containing DNA strand, the PBS hybridises with the 3' end of the nicked target DNA strand. This primes the prime editor to begin reverse transcription using the RT-template in the pegRNA as a template, synthesising the edited DNA strand onto the 3' end of the target DNA strand. Following reverse transcription, the edited DNA strand exists as a 3' DNA flap alongside the original 5' DNA flap containing the non-edited DNA sequence. Endogenous cellular DNA repair processes excise the 5' DNA flap, allowing the edited 3' DNA flap to be incorporated into the target site, generating a DNA heteroduplex containing an edited and non-edited strand. Subsequent replacement of the non-edited strand by endogenous DNA repair mechanisms permanently installs the edit into the genome.

There are three original versions of the prime editing system (Anzalone et al. 2020). Prime Editor 1 contains a fusion of a nCas9 to a wild-type M-MLV RT. Prime Editor 2 (PE2) uses an engineered M-MLV RT that increases editing efficiency 3-fold over the original wild-type version. Prime Editor 3 (PE3) uses the PE2 prime editor and pegRNA alongside an additional sgRNA targeting the non-edited strand for nicking. This nicking of the non-edited strand is thought to promote the resynthesis of this strand using the edited strand as a template. Lin

and colleagues (2020) adapted the prime editors PE2 and PE3 for use in plants, codon-optimising them for cereal plants and naming them Plant Prime Editor 2 and 3 (PPE2 and PPE3). They demonstrated their successful use in installing point mutations, insertions and deletions in wheat and rice protoplasts, and they regenerated prime-edited rice plants at frequencies up to 21.8%. A frequency of 1.5% was reported for installing point mutations in wheat, and PPE2 and PPE3 showed no difference in editing efficiency. Subsequently, a 'dual-pegRNA' strategy was used to improve editing efficiencies in wheat by using two paired pegRNAs that encode the same edit (Lin et al. 2021; Awan et al. 2022a). Following this, many improvements have been made to plant prime editing proteins, such as improved pegRNA expression and design, improved prime editor architecture, and new mutations within the prime editor to greatly increase editing efficiencies (Huang and Liu 2023; Li et al. 2023); however, these were not available at the start of this project.

1.5.4 Wheat transformation and developmental regulators

In addition to new gene editing technologies, advances in wheat transformation methods have improved the ease with which wheat can be gene-edited. An efficient *Agrobacterium*-mediated wheat transformation method was developed in Professor Wendy Harwood's lab and published by Hayta and colleagues (2019). This method improved upon previous *Agrobacterium*-mediated transformation methods, which were inefficient and challenging to implement (Harwood 2012). Consequently, biolistic-mediated transformation was previously favoured for wheat transformation. However, biolistic-mediated transformation tends to produce plants with multiple integration events and random rearrangements of the integrated transgenes, affecting transgene expression. In contrast, *Agrobacterium*-mediated transformation tends to occur in a more predictable and stable manner, often resulting in more low copy insertions and fewer random mutations (Chen et al. 2022). Moreover, *Agrobacterium*-mediated transformation is often more cost-effective, especially for large-scale transformation experiments that require many transgenic plants, such as gene editing experiments. Therefore, the availability of an efficient *Agrobacterium*-mediated method greatly facilitates gene editing in wheat.

While *Agrobacterium*-mediated transformation has improved the ease of efficient delivery of DNA transgenes into wheat explants, the recent use of developmental regulators that promote somatic embryogenesis has greatly improved plant regeneration (Harwood 2023). In wheat, the expression of a fusion protein combining wheat GROWTH-REGULATING FACTOR 4 (GRF4) and its cofactor GRF-INTERACTING FACTOR 1 (GIF1) has greatly increased

the efficiency and speed of regeneration (Debernardi et al. 2020). *GRF4-GIF1* overexpression has also allowed for the transformation of wheat genotypes previously unamenable to transformation and allows an increased number of transgenic wheat plants to be regenerated from each transformed explant. Other developmental regulators, such as *TaWOX5*, have also improved regeneration and overcome genotype dependency in wheat varieties (Wang et al. 2022a). Given these advancements in wheat transformation and regeneration, it is now an opportune moment to target the endogenous wheat *OR* gene for carotenoid biofortification by installing the 'golden SNP' or mimicking the *BoOr^{Mut}* mutation.

1.6 Thesis Aims

The overall aim of this thesis is to explore and engineer the genetic diversity of grain carotenoid content in wheat. To do this, I characterised the diversity of grain carotenoid content within the Watkins global landrace collection and investigated the function of *OR* within wheat to identify opportunities to enhance grain carotenoid content. In doing so, I attempt to answer the following questions:

- What is the grain carotenoid diversity within the Watkins global landrace collection? (Chapter 3)
- Can novel allelic diversity associated with grain carotenoid content be identified within the Watkins tetraploid collection? (Chapter 3)
- Is *OR* involved in grain carotenoid biosynthesis in wheat, and if so, does the 'golden SNP' affect grain carotenoid accumulation? (Chapter 4)
- Can diversity in the wheat *OR* gene be engineered to increase grain carotenoid content? (Chapter 5)

In Chapter 3, to describe the carotenoid diversity within the Watkins tetraploid collection, I first developed a high-throughput method for measuring YPC. I then used this to rapidly and cost-effectively screen the grain YPC of the field-grown Watkins tetraploid collection, a collection of highly diverse global durum wheat landraces. Furthermore, I used HPLC to accurately measure the content of lutein, zeaxanthin, α -carotene and β -carotene of the grains within this field-grown collection. Using a GWAS, I searched for MTAs and QTLs associated with these carotenoid measurements in the Watkins tetraploid collection. In Chapter 4, I investigated the function of *OR* in tetraploid and hexaploid wheat. Using an EMS TILLING population, I knocked out *Tdor* in tetraploid wheat and confirmed its role in

carotenoid biosynthesis. I investigated the effect of the 'golden SNP' substitution in hexaploid wheat by overexpressing *TaOR^{WT}* and *TaOR^{His}*, and then comparing their effect on grain carotenoid content. In Chapter 5, I attempted to install the 'golden SNP' within the endogenous *TaOR* gene using prime editing and mimic the *BoOr^{Mut}* in the endogenous *TaOR* gene using CRISPR/Cas9 gene editing. Following this, I searched the TILLING population to identify EMS-induced amino acid substitutions close to the 'golden SNP' and the *BoOr^{Mut}* mutation site and then screened these for grain carotenoid content.

2 Materials and Methods

2.1 Contributions to this thesis

Table 2.1 Names, affiliations and contributions of people who have contributed to this thesis. JIC=John Innes Centre, NRP=Norwich Research Park, AGIS=Agricultural Genomics Institute at Shenzhen, CAAS=Chinese Academy of Agricultural Sciences.

Name	Affiliation	Contributions
Ajay Siluveru	Germplasm Resource Unit, JIC, NRP, Norwich, NR4 7UH, UK	<ul style="list-style-type: none"> Milled and quantified the Watkins tetraploid collection using the high-throughput YPC method Milled grains and extracted carotenoids for quantification with HPLC Measured plant height, thousand-grain weight, grain area and grain number for <i>TaOR</i> overexpression and <i>Tdor</i> TILLING lines Measured relative chlorophyll content as a SPAD value with Workie Zegeye
Baldeep Kular	Metabolite Services, JIC, NRP, Norwich, NR4 7UH, UK	<ul style="list-style-type: none"> Analysed the extracted carotenoids with HPLC using standard solutions of carotenoid compounds to accurately measure these
Marielle Vigouroux	Informatics Platform, JIC, NRP, Norwich, NR4 7UH, UK	<ul style="list-style-type: none"> Provided example R scripts to run the GWAS using GAPIT
Mark Smedley	The Harwood lab, JIC, NRP, Norwich, NR4 7UH, UK	<ul style="list-style-type: none"> Provided help with designing my <i>TaOR</i> overexpression and prime editing constructs Produced and provided a number of plasmid parts used for assembling my transformation constructs
Mei Jiang	The Cheng lab, AGIS, CAAS, Shenzhen 518124, China	<ul style="list-style-type: none"> Performed the Watkins tetraploid collection's high-resolution GWAS using the whole-genome re-sequenced data Provided the SNP diversity files of the <i>OR</i> and <i>PSY</i> genes in the Watkins global landrace collection
Sadiye Hayta	The Harwood lab, JIC, NRP, Norwich, NR4 7UH, UK	<ul style="list-style-type: none"> Provided help with designing my <i>TaOR</i> overexpression and prime editing constructs Provided the <i>GRF4-GIF1</i> overexpressing T₁ grains used as a control
Simon Orford	Germplasm Resource Unit, JIC, NRP, Norwich, NR4 7UH, UK	<ul style="list-style-type: none"> Helped with the crossing of TILLING lines to create the <i>Tdor</i> mutant lines
Workie Zegeye	Germplasm Resource Unit, JIC, NRP, Norwich, NR4 7UH, UK	<ul style="list-style-type: none"> Measured relative chlorophyll content as a SPAD value with Ajay Siluveru

2.2 Plant materials

2.2.1 Growing conditions

For glasshouse-grown lines, grains were pre-germinated on damp filter paper for 48 hours at 4°C. Seedlings were then sown into 96-cell trays (CT) in 'John Innes F2 Starter soil' (85% fine grade peat, 15% washed grit, 4 kg m⁻³ Maglime, 2.7 kg m⁻³ Osmocote (3–4 months), 1 kg m⁻³ PG Mix 14-16-18 + Te 0.02% and wetting agent). At the 3-leaf stage, plants were transferred into 11 cm diameter pots with 'John Innes Cereal Mix' (40% medium grade peat, 40% sterilised loam, 20% washed horticultural grit, 3 kg m⁻³ Maglime, 1.3 kg m⁻³ PG mix 14-16-18 + Te base fertiliser, 1 kg m⁻³ Osmocote mini 16-8-11 2 mg + Te 0.02% and wetting agent). Plants were grown in standard glasshouse conditions under light for 16 hours of the day, with the remainder determined by the natural photoperiod.

2.2.2 Watkins global landrace collection

The stabilised Watkins tetraploid collection and Watkins hexaploid collection were obtained through the John Innes Centre's Germplasm Resource Unit (GRU), where they are stored (<https://www.seedstor.ac.uk/search-browseaccessions.php?idCollection=39>). Members of the GRU generated these collections by carrying out three single-seed descent rounds of the Watkins Historic Collection of Landrace Wheat (<https://www.seedstor.ac.uk/search-browseaccessions.php?idCollection=4>), a highly heterogeneous landrace collection. A core collection of the Watkins hexaploid collection was developed by Luzie Wingen from the original 1063 accessions using the Core Hunter 3 algorithm (Wingen et al. 2014; De Beukelaer et al. 2018).

The Watkins tetraploid and hexaploid collections were grown over the 2020–2021 field season at Church Farm, Bawburgh (52°38'N 1°10'E). For the Watkins tetraploid collection field trial, 343 accessions were sown by precision plot drill, each in a non-replicated single 1-metre plot. Alongside this, 110 plots of the modern durum cultivar Miradoux were randomly sown to account for environmental effects across the field, kindly supplied for this work by Elsoms Seeds (Elsoms Seeds Ltd, Spalding, United Kingdom). To act as a barcode for navigating the field, 54 plots of the bread wheat variety Paragon were also sown. The layout of the field trial is shown in Figure 2.1 consisting of 10 columns and 60 rows.

For the Watkins tetraploid collection, grain carotenoid content was measured using the high-throughput yellow pigment content (YPC) method (as detailed in Section 2.3.1) and

high-performance liquid chromatography (HPLC; as detailed in Section 2.3.2). A grain sample filling a volumetric cylinder of approximately 20 ml was used to measure the thousand-grain weight (TGW) of the Watkins tetraploid collection (as detailed in 2.4.1). For the Watkins hexaploid core collection, grain carotenoid content was measured using HPLC (as detailed in Section 2.3.2).

	1	2	3	4	5	6	7	8	9	10
1	WAT1180203	Paragon	WAT1180236	WAT1180129	WAT1180123	WAT1180238	Miradoux	WAT1180089	WAT1180043	
2	WAT1180134	WAT1180026	Miradoux	WAT1180016	WAT1180153	WAT1180242	Paragon	Miradoux	Miradoux	
3	Paragon	WAT1180229	WAT1180215	Miradoux	W3070	Paragon	WAT1180165	WAT1180015	WAT1180045	
4	Miradoux	W7291	WAT1180309	W7211	Paragon	Miradoux	WAT1180114	W7353	WAT1180024	
5	WAT1180018	Miradoux	WAT1180263	WAT1180064	WAT1180310	WAT1180079	WAT1180285	WAT1180371	Miradoux	
6	WAT1180098	W8591	Miradoux	W6511	WAT1180326	WAT1180251	W8612	WAT1180149	WAT1180059	
7	Miradoux	WAT1180308	WAT1180013	Paragon	WAT1180269	WAT1180040	Miradoux	W0175	WAT1180312	
8	WAT1180139	W4098	W6435	WAT1180159	Miradoux	Paragon	WAT1180276	Miradoux	WAT1180220	
9	WAT1180316	Miradoux	W0173	WAT1180223	WAT1180188	WAT1180050	WAT1180172	Paragon	Paragon	
10	Miradoux	WAT1180287	Paragon	Miradoux	WAT1180150	WAT1180095	WAT1180346	WAT1180163	Miradoux	
11	WAT1180069	WAT1180175	WAT1180370	WAT1180195	WAT1180052	Miradoux	WAT1180334	WAT1180048	WAT1180339	
12	WAT1180176	WAT1180096	WAT1180383	Paragon	WAT1180391	WAT1180323	Miradoux	W7256	WAT1180110	
13	Paragon	WAT1180218	WAT1180112	WAT1180011	Miradoux	WAT1180211	W7241	Miradoux	WAT1180022	
14	WAT1180245	WAT1180186	WAT1180374	WAT1180094	WAT1180327	WAT1180074	WAT1180266	WAT1180036	Miradoux	
15	WAT1180281	Miradoux	WAT1180292	WAT1180009	WAT1180116	Miradoux	WAT1180314	WAT1180244	WAT1180125	
16	WAT1180054	WAT1180234	Miradoux	Miradoux	Miradoux	WAT1180154	Miradoux	WAT1180106	WAT1180162	
17	WAT1180294	Paragon	Paragon	WAT1180034	Paragon	WAT1180194	WAT1180313	W8582	WAT1180182	
18	WAT1180120	W7354	WAT1180062	W7253	WAT1180042	Miradoux	Paragon	Paragon	WAT1180008	
19	WAT1180007	WAT1180283	WAT1180206	WAT1180147	WAT1180047	WAT1180252	WAT1180151	Miradoux	Paragon	
20	Miradoux	Miradoux	Miradoux	Miradoux	Miradoux	W5661	WAT1180053	WAT1180046	WAT1180328	
21	WAT1180051	Paragon	WAT1180049	Paragon	WAT1180280	WAT1180311	Miradoux	W7259	Miradoux	
22	Paragon	WAT1180184	Miradoux	WAT1180084	Paragon	WAT1180288	WAT1180384	WAT1180202	WAT1180025	
23	Miradoux	Miradoux	WAT1180088	WAT1180190	WAT1180259	Miradoux	WAT1180006	WAT1180085	WAT1180030	
24	WAT1180104	W8645	Paragon	WAT1180261	WAT1180340	W8571	WAT1180315	Miradoux	WAT1180270	
25	Miradoux	WAT1180213	WAT1180158	WAT1180265	WAT1180284	WAT1180180	Paragon	WAT1180290	Paragon	
26	WAT1180174	Miradoux	WAT1180232	WAT1180093	Miradoux	WAT1180142	WAT1180262	WAT1180029	W6613	
27	WAT1180101	WAT1180253	Miradoux	Miradoux	WAT1180002	Paragon	WAT1180208	Paragon	WAT1180109	
28	WAT1180177	WAT1180168	W6566	W8613	WAT1180066	WAT1180010	WAT1180146	WAT1180368	Miradoux	
29	WAT1180239	WAT1180198	WAT1180293	Miradoux	WAT1180012	WAT1180058	WAT1180255	Miradoux	WAT1180014	
30	WAT1180033	Miradoux	Miradoux	WAT1180268	Miradoux	Miradoux	Miradoux	WAT1180279	WAT1180273	
31	Miradoux	WAT1180322	WAT1180119	WAT1180155	WAT1180118	WAT1180210	WAT1180193	WAT1180115	WAT1180275	
32	WAT1180133	WAT1180250	WAT1180126	WAT1180181	WAT1180023	Paragon	WAT1180183	Miradoux	WAT1180298	
33	WAT1180080	WAT1180392	Miradoux	Miradoux	WAT1180222	WAT1180082	WAT1180102	WAT1180107	Paragon	
34	Paragon	Paragon	WAT1180224	WAT1180282	Miradoux	Miradoux	Miradoux	WAT1180233	Miradoux	
35	WAT1180336	W7043	WAT1180167	W0744	WAT1180111	WAT1180003	WAT1180295	WAT1180237	WAT1180161	
36	Miradoux	WAT1180240	WAT1180226	Miradoux	WAT1180325	WAT1180145	Paragon	Paragon	WAT1180122	
37	WAT1180302	WAT1180097	Paragon	WAT1180247	Paragon	Miradoux	Miradoux	Miradoux	W6567	
38	W1449	WAT1180086	WAT1180117	WAT1180345	Miradoux	WAT1180169	WAT1180335	WAT1180249	WAT1180170	
39	WAT1180105	Miradoux	WAT1180318	Paragon	WAT1180031	WAT1180148	WAT1180271	WAT1180087	WAT1180035	
40	WAT1180343	WAT1180337	WAT1180144	WAT1180157	WAT1180138	WAT1180342	WAT1180037	WAT1180303	Miradoux	
41	Paragon	WAT1180228	Paragon	Miradoux	Miradoux	WAT1180332	WAT1180090	Miradoux	WAT1180108	
42	Miradoux	Miradoux	WAT1180205	WAT1180073	WAT1180260	WAT1180274	WAT1180264	WAT1180020	WAT1180277	
43	WAT1180135	WAT1180331	WAT1180072	WAT1180204	WAT1180372	Miradoux	Miradoux	WAT1180121	WAT1180092	
44	W4100	WAT1180065	WAT1180028	WAT1180091	WAT1180254	Paragon	WAT1180207	WAT1180057	WAT1180032	
45	WAT1180075	Paragon	Miradoux	Miradoux	Miradoux	W8576	WAT1180286	Paragon	Miradoux	
46	WAT1180197	WAT1180187	WAT1180256	WAT1180216	WAT1180166	WAT1180347	WAT1180071	Miradoux	WAT1180083	
47	Miradoux	WAT1180219	WAT1180124	WAT1180376	WAT1180278	WAT1180152	Miradoux	WAT1180201	Miradoux	
48	WAT1180127	WAT1180179	WAT1180297	WAT1180128	Paragon	Miradoux	WAT1180113	WAT1180230	Paragon	
49	WAT1180338	Miradoux	Miradoux	Miradoux	WAT1180136	WAT1180227	WAT1180301	WAT1180027	WAT1180103	
50	WAT1180132	WAT1180333	WAT1180060	WAT1180196	Miradoux	WAT1180306	Paragon	WAT1180373	W3001	
51	WAT1180173	Miradoux	W8690	WAT1180296	WAT1180330	W7215	WAT1180164	WAT1180217	Paragon	
52	Miradoux	WAT1180055	WAT1180099	WAT1180140	WAT1180185	WAT1180341	WAT1180289	Miradoux	WAT1180004	
53	WAT1180130	WAT1180231	Miradoux	WAT1180209	Paragon	Miradoux	Miradoux	WAT1180001	Miradoux	WAT1180225
54	WAT1180044	WAT1180212	Paragon	Paragon	W7247	WAT1180171	WAT1180257	WAT1180143	WAT1180021	Miradoux
55	Paragon	WAT1180019	WAT1180189	WAT1180041	WAT1180005	WAT1180081	WAT1180191	WAT1180067	WAT1180178	WAT1180038
56	WAT1180299	WAT1180369	WAT1180248	WAT1180141	WAT1180131	WAT1180061	WAT1180235	WAT1180258	W5662	WAT1180214
57	WAT1180272	Paragon	WAT1180300	Miradoux	Miradoux	Paragon	Miradoux	Miradoux	WAT1180221	WAT1180068
58	Miradoux	WAT1180291	Miradoux	WAT1180329	WAT1180307	Miradoux	WAT1180192	WAT1180246	Miradoux	W4318
59	WAT1180063	WAT1180267	WAT1180321	Paragon	WAT1180200	WAT1180243	WAT1180319	Paragon	WAT1180070	Miradoux
60	WAT1180241	Miradoux	WAT1180100	WAT1180156	WAT1180160	WAT1180199	Paragon	WAT1180317	WAT1180137	WAT1180056

Figure 2.1 Field plan for the Watkins tetraploid collection sown in the 2020–2021 field season. Some accessions within the Watkins tetraploid collection were missing grain samples and these are shown as pink boxes. Control lines Miradoux and Paragon are shown in teal and yellow boxes, respectively.

2.2.3 Durum wheat *Tdor* mutants

Ethyl methanesulfonate (EMS) mutations that affect the function of *TdORANGE* (*TdOR*) were searched for within the Kronos TILLING (Targeting Induced Local Lesions IN Genomes) population using the variant search feature on EnsemblPlants (Krasileva et al. 2017; Yates et al. 2022). EMS mutations producing early stop gain variants within the *TdOR* coding sequence and splice acceptor/donor variants were preferentially searched for. Where these could not be found, EMS mutations that produced missense mutations with a SIFT score of below 0.01 were chosen. These mutations were predicted to be deleterious to protein function based on sequence homology with other species (Ng and Henikoff 2001). Mutations in the lines Kronos 0329 and Kronos 4335 were identified. TILLING lines are referred to as KXXXX for simplicity throughout this thesis (for instance, the TILLING line 'Kronos 0329' is referred to as 'K0329'). Lines were ordered through the GRU (<https://www.seedstor.ac.uk/>). For generating a homozygous double knockout *Tdor* mutant line, an F₂ cross was performed with the lines K0329 and K4335. The crossing scheme for this is illustrated in Figure 2.2. KASP genotyping was used to determine the zygosity of the TILLING lines and track mutations through subsequent crossing strategies (as detailed in Section 2.9.1). Plants were grown in glasshouse conditions (as detailed in Section 2.2.1).

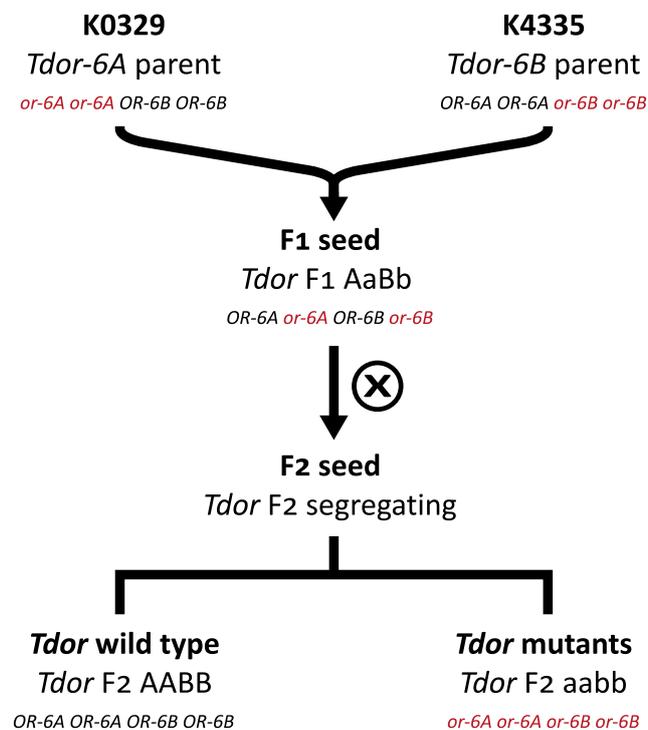


Figure 2.2 Crossing diagram for generating homozygous double knockout *Tdor* mutant lines and homozygous wild-type lines. Self-pollination is represented by '⊗'.

For the *Tdor* mutant and *Tdor* wild-type lines, grain carotenoid content was assessed using HPLC (as detailed in Section 2.3.2). Due to low grain yield per plant, pools of 6 grams from the three highest-yielding plants (2 grams from each plant) were analysed for each line. Plant height at maturity (as detailed in Section 2.4.2) and grain yield per plant were recorded for all plants. Grain area and thousand-grain weight were recorded from the three plants with the highest grain yield per plant (as detailed in Section 2.4.1). This was done because plants with very low grain yield per plant may bias these measurements.

2.2.4 Bread and durum wheat *TaOR* overexpression lines

Overexpression lines were generated in the bread wheat cultivar Cadenza and the durum wheat cultivar Kronos (as detailed in Section 2.8.2) using the constructs pAct-OR^{WT} and pAct-OR^{His} (as detailed in Section 2.7.1). The T₀ durum wheat overexpression lines did not produce enough grain to be used. Bread wheat T₀ plants were assessed for copy number (as detailed in Section 2.9.2), and the T₁ grains of plants with 1, 2 and 4 copies of each transgene were regrown. T₁ plants that only overexpressed the *GRF4-GIF1* developmental regulators under the same promoter were provided by Sadiye Hayta and grown as controls. Plants were grown in glasshouse conditions (as detailed in Section 2.2.1). Lines were referred to by their transgene name: OR^{WT}, OR^{His} and GRF. For each transgene copy number, 48 T₁ grains were sown into 96-CTs. The copy number of T₁ plants was assessed, and plants with 1 copy (referred to as 1C), 2 copies (referred to as 2C) and 3–5 copies (referred to as hiC) were selected for each transgene (OR^{WT}, OR^{His}, GRF). Two separate lines for each transgene copy number were chosen ('-A' and '-B'), totalling six transgenic lines each. Eight plants for each of these lines were grown. Some T₁ grains did not germinate, so T₁ grains of the same copy numbers from other T₀ lines were selected instead. For instance, for the T₀ 1 copy pAct-OR^{WT} 3318-2-01 line, T₁ plants with both 1 copy and 2 copies of the transgene were selected for OR^{WT}-1C-A and OR^{WT}-2C-A, respectively. This means some of the lines are related to each other. The zygosity of these lines was worked out based on the segregation pattern of the transgenes within the 48 T₁ plants. The lineage, naming and zygosity of these lines are illustrated in Figure 2.3.

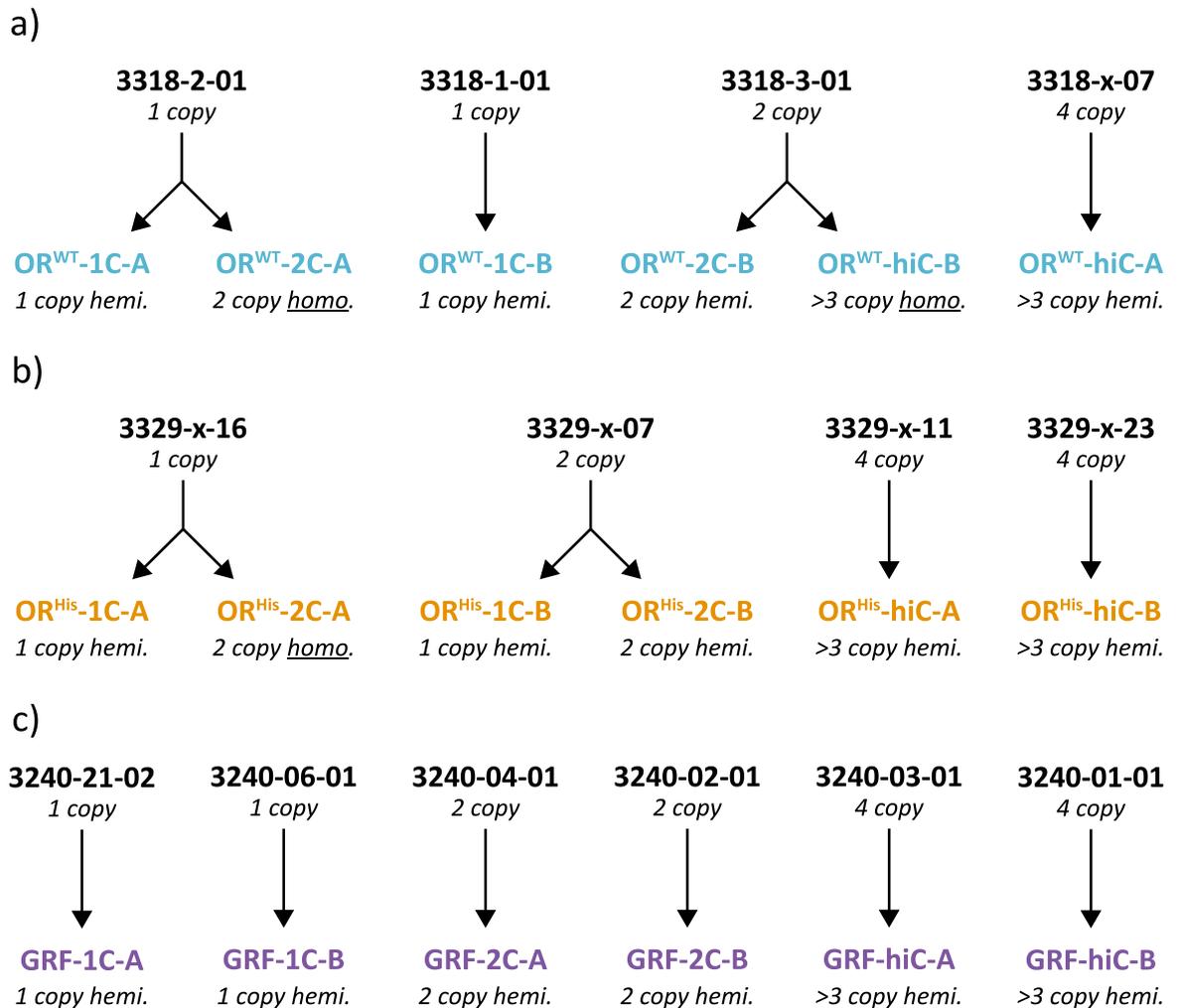


Figure 2.3 Lineage and zygosity of the $TaOR^{WT}$, $TaOR^{His}$ and $GRF4-GIF1$ overexpression lines used in this study. Names of the T_0 plants are shown above in black, with the selected T_1 lines' names coloured in blue, orange and purple depending on their transgene. (a) OR^{WT} T_1 lines containing the pAct- OR^{WT} transgene. (b) OR^{His} T_1 lines containing the pAct- OR^{His} transgene. (c) GRF T_1 control lines containing the $GRF4-GIF1$ and $hptII$ transgenes, acting as controls. Zygosity found below the line's name: hemi.=hemizygous, homo.=homozygous.

The grain carotenoid content of T_2 grains was assessed using HPLC (as detailed in Section 2.3.2). Due to low grain yield per plant, pools of 5 grams from the four plants with the highest grain yield (1.25 grams from each plant) were analysed for each line. Plant height, anthesis date, grain number and grain yield per plant were recorded for all plants (as detailed in Section 2.4.2). Grain area and thousand-grain weight were recorded from the four plants with the highest grain yield per plant (as detailed in Section 2.4.1). This was done because plants with very low grain yield per plant may bias these measurements. The relative leaf chlorophyll content was measured (as detailed in Section 2.4.3) for all plants except those that had not gone through anthesis by the 30th of June.

Dark-grown callus was produced by isolating T_1 immature embryos from the growing bread wheat T_0 $TaOR^{WT}$ and $TaOR^{His}$ overexpression plants, following modifications to the protocol

by Hayta and colleagues (2021): Immature embryos were not inoculated with *Agrobacterium tumefaciens* (*Agrobacterium*), embryos were put on co-cultivation media for 5 days and then moved to resting media in the dark for 10 weeks for callus induction, and calli were moved onto new resting media every 2 weeks.

2.2.5 Bread wheat *TaOR* 'golden SNP' prime editing lines

Prime editing lines to install the 'golden SNP' into the *TaOR* homoeologues were generated in the bread wheat cultivar Cadenza (as detailed in Section 2.8.2) using the constructs pOR-PE#1, pOR-PE#2 and pOR-dualPE (as detailed in Section 2.7.1). Plants were grown in glasshouse conditions (as detailed in Section 2.2.1).

2.2.6 Bread wheat CRISPR/Cas9 *BoOr^{Mut}* mimic lines

CRISPR/Cas9 lines to mimic the *BoOr^{Mut}* mutation in the *TaOR* homoeologues were generated in the bread wheat cultivar Cadenza (as detailed in Section 2.8.2) using the constructs pTaOR-Cas9 (as detailed in Section 2.7.1). Plants were grown in glasshouse conditions (as detailed in Section 2.2.1).

2.2.7 Field grown *OR* TILLING lines

EMS mutations within the Kronos and Cadenza TILLING populations were searched for using the variant search feature on EnsemblPlants (Krasileva et al. 2017; Yates et al. 2022). EMS mutations producing non-synonymous substitutions were searched for in the homoeologues of *TaOR*, *TdOR*, *TaORLIKE* and *TdORLIKE*. Substitutions close to the locations of the 'golden SNP' substitution, the *BoOr^{Mut}* integration site and the *OsOr^{Mut}* edit site (all residing in the first cytoplasmic domain of the protein) were favoured. Moreover, substitutions within the α -helix of the 'golden SNP' were preferred. Substitutions with lower SIFT scores were prioritised.

The Kronos and Cadenza TILLING populations were previously grown in the field over the 2014–2015 field season (at the M₄ generation). Grains from these field bulks were made available for this project. For the TILLING lines containing interesting EMS mutations in *OR* or *ORLIKE* homoeologues, a 5-gram sample was used for measuring grain carotenoid content (as detailed in Section 2.3.2).

2.3 Carotenoid phenotyping

2.3.1 High-throughput yellow pigment content (YPC)

For measuring YPC, a high-throughput YPC method was made to measure absorbance in a 96-well optically clear microplate. This was based on modifications to the industry-standard AACC 14-50 method (AACC 2009) made by Beleggia and colleagues (2010) to use less water-saturated butanol (WSB) extraction solvent and flour. However, their micro-method involved measuring absorbance using a UV-Vis HPLC detector, which required very small aliquots of extract (200 μ l). Instead, the microplate-based reader required more extract; therefore, more WSB was used for the extraction (1500 μ l) while keeping the ratio of flour-to-solvent the same as in the AACC 14-50 YPC method, where 8 grams of flour was extracted in 40 ml of WSB, giving a 1:5 ratio. The flour used in this extraction was increased to 300 mg for 1500 μ l of WSB. A sonication step of 15 minutes and a 16–18 hour wait were included to ensure carotenoids were fully extracted.

The high-throughput YPC method was carried out as follows: Five grams of grain, on a 14% moisture basis, was ground in an A11 IKA lab mill (IKA-Werke, Staufen, Germany) and stored at -20°C to prevent the breakdown of carotenoids. In sets of 96 samples at a time, roughly 300 mg (280-320 mg) of flour was weighed out into amber 2 ml Eppendorf tubes, and the exact flour weight for each sample was recorded. Amber Eppendorfs were used to prevent the UV degradation of carotenoids. To each Eppendorf tube, 1.5 ml of WSB was added, and the sample was mixed in a Genogrinder for 1 minute at 1,500 revolutions per minute. Samples were then sonicated for 15 minutes and left overnight in the dark (for 16–18 hours). The samples were then centrifuged at 5,000 $\times g$ for 15 minutes, after which 350 μ l of supernatant of each sample was transferred to a transparent 96-well flat-bottom microplate (Greiner Bio-One, Kremsmünster, Austria). This gave a path length of 10 mm when read from the top of each well. The absorbance of the extract was then measured at 436 nm by a Varioskan™ LUX multimode microplate reader (Thermo Fisher, Waltham, Massachusetts, US). Five absorbance readings were made for each sample and averaged. YPC was then calculated based on the exact quantity of flour used for extraction and absorbance using the following equation. This equation was adapted for the reduced flour and WSB amount from the AACC 14-50 YPC method's conversion factor of 1.6632 (AACC 2009), defined as the optical density of 1 mg of pigment in 100 ml WSB.

$$\text{Yellow Pigment Content } (\mu\text{g/g}) = \frac{A \text{ (absorbance at 436 nm)}}{W \text{ (weight in grams)} \times 0.11088}$$

2.3.2 High-performance liquid chromatography (HPLC)

For carotenoid measurements using HPLC, grain samples were milled using the RETSCH Cyclone Mill Twister (Retsch, Haan, Germany). This mill required a minimum of 5 grams of grain to ensure consistent flour milling. Carotenoids were extracted from 1 gram of flour with acetone (containing 0.1% butylated hydroxytoluene) following the method used by Atienza and colleagues (2007) with the following modifications: A GeneVac EZ-2 solvent evaporator (Genevac Ltd, Ipswich, United Kingdom) was used for solvent evaporation, and the pigment extract was dissolved in 1 ml of acetone. Following this, samples were stored at -80°C until HPLC analysis.

HPLC was performed by the Metabolic Services platform at the John Innes Centre on an Agilent 1290 Infinity II LC System UHPLC (Agilent Technologies, California, US) following the method used by Mendes-Pinto and colleagues (2005) with the following modifications: A Halo[®] C₃₀ 160 Å, 2.7 µm, 2.1 × 150 mm column (Part No. 92112-730, Element, Strathaven, Scotland) was used. Eluents used were 90% acetonitrile in H₂O (v/v) (solvent A1) and 100% ethyl acetate (solvent B2) with a flow rate of 0.3 ml per minute. The following binary gradient system was used: 0–15 minutes (85% solvent A1 and 15% solvent B2), 15–17 minutes (30% solvent A1 and 70% solvent B2), 17–20 minutes (85% solvent A1 and 15% solvent B2). The sample injection was 2 µl, and absorbance was recorded at 450 nm.

Standard solutions of α-carotene, β-carotene, lutein and zeaxanthin (Sigma-Aldrich, Burlington, Massachusetts, US) were prepared in hexane, ethanol or acetone depending on their solubility. The exact concentration of each stock solution was determined by spectrophotometry using the absorption coefficients of the respective carotenoid. After the determination of the concentrations, the standards were evaporated under nitrogen and dissolved in 100% acetone. Peaks were then identified in the unknown samples by comparing the retention times (RT) and UV-Vis spectral data to those of the corresponding standards. The concentration of each compound was calculated from the calibration curve of the corresponding standard. Carotenoid concentrations were then calculated through a linear regression using the Agilent ChemStation software (Agilent Technologies, California, US). The proportion of β-β branch carotenoids was calculated by summing the content of zeaxanthin and β-carotene and dividing by the total grain carotenoid content (sum of α-carotene, β-carotene, lutein and zeaxanthin). The proportion of provitamin A (PVA) carotenoids was calculated by summing the content of α-carotene and β-carotene and dividing by the total grain carotenoid content.

2.4 Other phenotyping methods

2.4.1 Grain morphometrics

Grain area, grain number per plant and TGW were measured using a MARViN Seed Analyser (MARViTECH GmbH, Wittenburg, Germany). Grain yield per plant was measured as the amount of grain produced by a single plant. It is important to note that this measurement is distinct from 'true yield', which refers to the actual yield achieved under field conditions at a larger scale.

2.4.2 Anthesis and height

Anthesis was scored based on the occurrence of anther extrusion from spikelets. Measurements were made on glasshouse-grown lines three times a week (Monday, Wednesday and Friday). The final date anthesis was scored on was the 30th of June. Plants which had not gone through anthesis by this point were given the date of anthesis as the 1st of July. Height was measured at plant maturity from the base of the plant to the tip of the highest spike, excluding the awns.

2.4.3 Relative leaf chlorophyll content

Relative leaf chlorophyll content, recorded as a SPAD value, was measured by a SPAD-502 meter (Konica Minolta, Tokyo, Japan). Readings were made three times along the flag leaf blade and averaged to obtain a single measurement. The SPAD value was recorded two and three weeks after anthesis to compare leaf chlorophyll at similar stages. These two measurements were then averaged to obtain a final SPAD value.

2.5 Genome-wide association study (GWAS) methods

2.5.1 GWAS on 35K Axiom[®] Breeder's Array data

The Watkins tetraploid collection was previously genotyped using the 35K Axiom[®] Wheat Breeder's Array by members of the GRU, and this data was made available upon request (jic.geneticresources@jic.ac.uk). The genetic map file containing the chromosome and base pair positions (IWGSC RefSeq v1) of each marker was obtained from CerealsDB (<https://www.cerealsdb.uk.net/>) and filtered to exclude the D-genome markers.

For the genome-wide association study (GWAS) using 35K Axiom[®] Wheat Breeder's Array data, the Genome Association and Prediction Integrated Tool (GAPIT) software was used (<https://www.zzlab.net/GAPIT/>) (Wang and Zhang 2021). GAPIT is a widely used genomic

association tool that runs as an R package and integrates several GWAS models. I used the models: mixed linear model (MLM) (Zhu et al. 2008), Fixed and random model Circulating Probability Unification (FarmCPU) (Liu et al. 2016) and Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) (Huang et al. 2018). While MLM is designed to correct for population structure and kinship using both fixed and random effects, the analysis here did not incorporate a kinship matrix, focusing solely on population structure correction. FarmCPU and BLINK were utilised to enhance detection power and computational efficiency. FarmCPU iteratively separates markers into fixed and random effects, evaluating their contributions to trait variance in a segmented, sequential manner. BLINK incorporates Bayesian information and linkage disequilibrium (LD), leveraging LD to better understand marker inheritance to identify causal variants. By integrating these concepts, BLINK aims for precise and efficient association mapping. Each model provides a unique approach to balancing accuracy, computational efficiency, and error management, representing a continuum of innovations for tackling genetic data complexity.

2.5.2 GWAS on high-resolution whole-genome sequence data

The Watkins tetraploid collection was whole-genome re-sequenced by Professor Shifeng Cheng's lab at the Agricultural Genomics Institute at Shenzhen. GWAS using this sequence data was run by Mei Jiang from the Cheng lab following a similar method to that used by Cheng and colleagues (2023). Briefly, 367 accessions of the collection were sequenced. Low-quality reads were filtered using fastp (v0.20.0) (Chen et al. 2018), and the cleaned reads were mapped to Svevo v1 using BWA-MEM (v0.7.17) (Li 2013). SNP calling and filtering were performed using GATK (v4.1.2) (Van der Auwera and O'Connor 2020). A kinship matrix was calculated using GEMMA (v0.98.1) (Zhou and Stephens 2012). The kinship matrix was used as the covariate for the GWAS performed using GEMMA with the following parameters: 'gemma-0.98.1-linux-static -miss 0.9 - gk kinship.txt' and 'gemma-0.98.1-linux-static -miss 0.9 -lmm -k kinship.txt'. In-house R scripts were used to visualise these results.

2.6 Sequence analysis and bioinformatics

2.6.1 Analysis of QTL regions identified within the Watkins tetraploid collection

The Svevo genome browser was used to search for previously annotated durum wheat quantitative trait loci (QTLs) associated with carotenoid traits. This is curated by the International Durum Wheat Genome Sequencing Consortium (Maccaferri et al. 2019; Yao et al. 2022), hosted by GrainGenes (<https://wheat.pw.usda.gov/jb?data=/ggds/whe-svevo2018>).

Knetminer was used to investigate the genes within the associated QTLs (Hassani-Pak et al. 2021). Since Knetminer only works with bread wheat IWGSC RefSeq v1 genes, orthologues to the durum wheat Svevo v1 genes were identified using the EnsemblPlants BioMart orthology search feature (<https://plants.ensembl.org/>) (Kinsella et al. 2011; Yates et al. 2022). These RefSeq v1 genes were submitted to Knetminer with the search term 'carotenoid'. The generated knowledge networks of each gene with a KnetScore above 10 were studied to investigate the links to the searched trait. These links were based on their predicted biochemical function and their orthologous genes linked to this trait in other organisms (such as rice and *Arabidopsis thaliana*).

2.6.2 Identification of durum wheat carotenoid biosynthesis genes

Carotenoid biosynthesis genes were identified through literature searches in bread wheat, rice, and *Arabidopsis thaliana*, where the pathway has been extensively studied (Nisar et al. 2015; Colasuonno et al. 2017a; Sun et al. 2022; Niaz et al. 2023). Durum wheat orthologues were then identified through the EnsemblPlants BioMart orthology search feature and EnsemblPlants BLAST searches (<https://plants.ensembl.org/>) (Kinsella et al. 2011; Yates et al. 2022). This identified carotenoid biosynthesis genes on the Svevo v1 reference genome sequence (Maccaferri et al. 2019).

2.6.3 Analysis of wheat *OR* homoeologues and *OR* orthologues

The BLASTp search to identify the wheat orthologues of melon's CmOR was run on EnsemblPlants (Yates et al. 2022) using the CmOR protein sequence (MELO3C005449) obtained by Tzuri and colleagues (2015) against the *Triticum aestivum* hexaploid wheat genome sequence of the Chinese Spring cultivar (IWGSC et al. 2018), and the *T. turgidum* tetraploid wheat genome sequence of the Svevo cultivar (Maccaferri et al. 2019). Sequence alignments were performed using CLUSTAL O multiple sequence alignment (v1.2.4)

(<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (Madeira et al. 2022). Sequence analysis and comparison were performed using Geneious Prime 2022.2 software (<https://www.geneious.com>). TargetP 2.0 was used to predict and identify N-terminal sorting signals (Almagro Armenteros et al. 2019). The neighbour-joining phylogenetic tree was created with Geneious Prime 2022.2 using default settings with *Capsella rubella* as an outgroup (<https://www.geneious.com>). The amino acid sequences of *OR* orthologues used in building this phylogenetic tree were from Tzuri and colleagues (2015), supplemented by those of *Avena sativa*, *Eragrostis tef* and *T. turgidum*, identified and obtained on EnsemblPlants (Yates et al. 2022). The wheat expression browser, expVIP, was used to analyse the expression of *TaOR* homoeologues within wheat (Borrill et al. 2016; Ramírez-González et al. 2018).

Analysis of the domains of TaOR and TaORLIKE proteins was done using EnsemblPlants (Yates et al. 2022). The transmembrane topology of TaOR and TaORLIKE proteins was analysed using InterPro (Paysan-Lafosse et al. 2023). The 3D predicted structures of TaOR and TaORLIKE were downloaded from the AlphaFold2 Protein Structure Database (Varadi et al. 2024) and imaged using PyMOL software (Schrödinger and DeLano 2020).

2.6.4 Searching for wheat *OR* allelic diversity

A BLASTp search on EnsemblPlants was made against the available bread wheat pangenome to search for *OR* allelic diversity within the bread wheat pangenome (Walkowiak et al. 2020; Yates et al. 2022). For searching for *OR* allelic diversity within the Watkins tetraploid and hexaploid collection, the SNP diversity of *TaOR*, *TdOR*, *TaPSY1*, *TaPSY2*, *TdPSY1* and *TdPSY2* was provided by the lab of Professor Shifeng Cheng as comma-separated value (CSV) files. These were then searched to identify non-synonymous mutations within the coding sequence. The effects of any non-synonymous substitutions were predicted using the Variant Effect Predictor tool on EnsemblPlants (McLaren et al. 2016).

2.6.5 Investigating *TaOR* sequence suitability for base editing

The sequence surrounding the 'golden SNP' nucleotide in the *TaOR* homoeologues was analysed to assess for suitability of base editing this nucleotide. The following Cas variants and their protospacer adjacent motif (PAM) sites were considered: SaCas9 'NNGRRT' PAM (Hua et al. 2019), iSpyMacCas9 'NAA' PAM (Sretenovic et al. 2020), SpCas9-VQR or SpCas9-VRQR 'NGA' PAM (Hu et al. 2016), ScCas9 or ScCas9++ 'NNG' PAM (Chatterjee et al. 2020), St1Cas9 'NNAGAAW' PAM (Agudelo et al. 2020), LbCas12a-RVR 'TATV' PAM (Wang et al.

2020), and SpCas9-NG and xCas9's 'NG' PAM (Zhong et al. 2019; Wang et al. 2020). This sequence analysis was done using Geneious Prime 2022.2 software (<https://www.geneious.com>).

2.6.6 Searching for EMS mutations within K4596

All EMS mutations within the Kronos TILLING collection were downloaded off EnsemblPlants as a CSV file (Yates et al. 2022). The dataset was then filtered to include only entries associated with K4596 using the 'grep' command in Bash. EMS mutations within carotenoid biosynthesis genes that were previously identified (as detailed in Section 2.6.2) were searched for within this filtered set of EMS mutations.

2.7 Golden Gate Cloning and construct design

Golden Gate Cloning (Engler et al. 2008) using the MoClo system (Werner et al. 2012) was used in this thesis. This uses Type IIS restriction enzymes, which cut DNA outside of their recognition sites, and T4 DNA ligase for efficient DNA assembly. Initially, DNA fragments, designed with specific overhangs by Type IIS enzyme recognition sites, are simultaneously cut and ligated in a single reaction. This method enables the precise and orderly assembly of multiple fragments into a vector facilitated by T4 DNA ligase. The reaction components are found in Table 2.2. *Eco31I* (Thermo Fisher, Waltham, Massachusetts, US) has the same cut site as *BsaI* and was used in producing Level 1 plasmids. *Bpil* (Thermo Fisher, Waltham, Massachusetts, US) has the same cut site as *BbsI* and was used in producing Level 2 plasmids. The reaction conditions are found in Table 2.3. The plasmid components used in the Golden Gate constructs are found in Table 2.4.

Table 2.2 Golden Gate reaction components. Additional donor fragments are shown by 'Donor X'.

Component	1x (μl)
Acceptor (100 ng/μl)	1
Donor 1 (100 ng/μl)	1
Donor 2 (100 ng/μl)	1
Donor X (100 ng/μl)	1
T4 Ligase Buffer (10X, NEB)	1.5
T4 DNA Ligase (400 U/μl, NEB)	0.5
<i>Eco31I/Bpil</i> (10 U/μl, Thermo Fisher)	0.5
H ₂ O	Up to 15

Table 2.3 Golden Gate reaction conditions for *Eco31I* and *BpiI* reactions.

Step	Temperature	Duration
Digestion-ligation reaction (x10)	37°C	5 min.
	16°C	5min.
Final digestion	37°C	7 min.
Enzyme inactivation	80°C	10 min.
Hold	4°C	–

Table 2.4 Parts used for Golden Gate Cloning within the MoClo system. ‘LX’ refers to Level X and ‘PX’ refers to Position X of the MoClo system. For example: ‘L1 P2’ refers to Level 1 Position 2. The Sainsburys Lab (TSL) SynBio: <https://synbio.tsl.ac.uk/>; Addgene: <https://www.addgene.org>.

Name	Description	Source	Construct this was used in
pICH41421	L0 P3 nosT	TSL SynBio	pAct-ORWT, pAct-ORHis
pICSL12014	L0 P1 OsActin pro	TSL SynBio	pAct-ORWT, pAct-ORHis
pL0_TaORWT	L0 P2 Synthesised TaORWT CDS	GeneWiz	pAct-ORWT, pAct-ORHis
pL0_TaORHis	L0 P2 Synthesised TaORHis CDS	GeneWiz	pAct-ORWT, pAct-ORHis
pICSL90024	L0 P2 Plant Prime Editor 2	TSL SynBio	pOR-PE#1, pOR-PE#2, pOR-dualPE
pL0_pegRNA#1	L0 P2 Synthesised TaU3P::pegRNA#1	GeneWiz	pOR-PE#1, pOR-dualPE
pL0_pegRNA#1	L0 P2 Synthesised TaU3P::pegRNA#2	GeneWiz	pOR-PE#2, pOR-dualPE
pICH47742	L1 P2 acceptor plasmid	TSL SynBio	pAct-ORWT, pAct-ORHis
pICH47751	L1 P3 acceptor plasmid	TSL SynBio	pOR-PE#1, pOR-PE#2, pOR-dualPE
pICH47761	L1 P4 acceptor plasmid	TSL SynBio	pOR-dualPE
pL1P3_TaU6acc	L1 P3 TaU6 sgRNA acceptor	BRACT, Addgene #165599	pTaOR-Cas9
pL1P4_TaU6acc	L1 P4 TaU6 sgRNA acceptor	BRACT, Addgene #165600	pTaOR-Cas9
pICSL11099	L1 P1 ZmUbiP::Hygromycin::nosT (reverse)	TSL SynBio	pAct-ORWT, pAct-ORHis, pOR-PE#1, pOR-PE#2, pOR-dualPE,
pL1P1_PvUbiHyg	L1 P1 PvUbi2P::Hygromycin::nosT (reverse)	Mark Smedley	pTaOR-Cas9
pL1P2_PPE2	L1 P2 OsUbi3P::Plant Prime Editor 2::nosT	Mark Smedley	pOR-PE#1, pOR-PE#2, pOR-dualPE
pL1P2_Cas9	L1 P2 OsUbiP::Cas9::nosT	BRACT, Addgene #165424	pTaOR-Cas9
pL1P3_GRF-GIF	L1 P3 ZmUbiP::GRF-GIF::nosT	BRACT, Addgene #198047	pAct-ORWT, pAct-ORHis
pL1P4_GRF-GIF	L1 P4 ZmUbiP::GRF-GIF::nosT	BRACT, Addgene #198048	pOR-PE#1, pOR-PE#2
pL1P5_GRF-GIF	L1 P5 ZmUbiP::GRF-GIF::nosT	BRACT, Addgene #198046	pOR-dualPE, pTaOR-Cas9
pICSL 41766	L1 P4 End linker	TSL SynBio	pAct-ORWT, pAct-ORHis
pICH41780	L1 P5 End linker	TSL SynBio	pOR-PE#1, pOR-PE#2
piCH41800	L1 P6 End linker	TSL SynBio	pOR-dualPE, pTaOR-Cas9
pGGG-M	L2 pGoldenGreenGate-M backbone	BRACT, Addgene #165422	pAct-ORWT, pAct-ORHis, pOR-PE#1, pOR-PE#2, pOR-dualPE, pTaOR-Cas9

The assembled plasmids were transformed into *Escherichia coli* Library Efficiency™ DH5α cells (Invitrogen, Waltham, Massachusetts, US), and the QIAprep Spin Miniprep Kit (Qiagen, Venlo, Netherlands) was used for plasmid isolation from *E. coli*. The kits were used according to the manufacturer's instructions, and samples were eluted in distilled water. Restriction digests were performed to confirm the correct assembly of plasmids using *Bpil* for Level 1 plasmids and *EcoRI* (NEB, Ipswich, Massachusetts, US) for Level 2 plasmids. Digestion products were separated by electrophoresis on a 1% agarose gel and imaged using a UV transilluminator. The ligation boundaries of all plasmids were verified using the GeneWiz Sanger Sequencing service (Azenta Life Sciences, Burlington, Massachusetts, US). Standard sequencing primers for the relevant vectors were used. Sanger Sequencing chromatograms were analysed in Geneious Prime 2022.2 software (<https://www.geneious.com>).

2.7.1 *TaOR^{WT}* and *TaOR^{His}* overexpression constructs

The *TaOR-6D* CDS (981 bp) without the 'golden SNP' (pLO_*TaOR*^{WT}) and with the 'golden SNP' (pLO_*TaOR*^{His}) were synthesised using the GeneWiz Gene Synthesis service (Azenta Life Sciences, Burlington, Massachusetts, US). Each *TaOR-6D* CDS was assembled into a Level 1 Position 2 acceptor plasmid (pICH47742) with the rice actin promoter (pICSL12014) to drive the expression of the *TaOR^{WT}* or *TaOR^{His}* transgenes and a nosT terminator (pICH41421). The rice actin promoter was used because it has been found to have higher expression than the CaMV 35S promoter in wheat (Jang et al. 2002; Hayta et al. 2019). Level 2 plasmids were assembled using a pGoldenGreenGate-M backbone (pGGG-M), which contains a kanamycin resistance gene (*nptI*) as well as origins for replication for *E. coli* (colEI ori) and *Agrobacterium* (pSa ori). Included in these Level 2 plasmids were a *HYGROMYCIN PHOSPHOTRANSFERASE* (*hptII*) selection gene (pICSL11099) and the *GRF4-GIF1* developmental regulators (pL1P3_GRF-GIF). The final Level 2 plasmids are referred to as pAct-OR^{WT} (for *TaOR^{His}-6D* overexpression) and pAct-OR^{His} (for *TaOR^{His}-6D* overexpression). Their plasmid maps can be found in Figure 4.6.

2.7.2 Prime editing constructs

The pegRNAs (pegRNA#1 and pegRNA#2) for installing the 'golden SNP' substitution into the endogenous *TaOR* homoeologues were designed using PlantPegDesigner with default settings (<http://www.plantgenomeediting.net>) (Jin et al. 2022). The designed pegRNAs are found in Table 2.5. The pegRNA scaffold and wheat U3 promoter (TaU3 pro) was used from Lin and colleagues' (2020) original Plant Prime Editing 2 paper. The Level 0 pegRNA

expression constructs (pLO_pegRNA#1 and pLO_pegRNA#2) were synthesised using the GeneWiz Gene Synthesis service (Azenta Life Sciences, Burlington, Massachusetts, US). Each was cloned into a Level 1 Position 3 acceptor plasmid (pICH47751) for the single pegRNA construct (pOR-PE#1 and pOR-PE#2). For the dual pegRNA construct (pOR-dualPE), the Level 0 pegRNA#2 expression plasmid (pLO_pegRNA#2) was cloned into a Level 1 Position 4 acceptor plasmid (pICH47761). The Plant Prime Editor 2 (pICSL90024) was domesticated for Golden Gate Cloning and made available through TSL SynBio based on the design by Lin and colleagues (2020). The Level 1 Position 2 Plant Prime Editor 2 expression plasmid (pL1P2_PPE2) had previously been produced within the Harwood lab. Level 2 plasmids were assembled using a pGGG-M backbone. These Level 2 plasmids included a *hptII* selection gene (pICSL11099) and the *GRF4-GIF1* developmental regulators (pL1P4_GRF-GIF or pL1P5_GRF-GIF). The final Level 2 plasmids are referred to as pOR-PE#1 (that just includes pegRNA#1), pOR-PE#2 (that just includes pegRNA#2) and pOR-dualPE (that includes both pegRNA#1 and pegRNA#2). Their plasmid maps can be found in Figure 5.4.

Table 2.5 pegRNAs designed by PlantPegDesigner for installing the ‘golden SNP’ into the endogenous TaOR homoeologues. The edited ‘golden SNP’ nucleotide on the RT-template is indicated in red. RT-template=reverse transcriptase template, PBS=primer binding sequence.

Name	Component	Sequence
pegRNA#1	Target site	TTGATAATATCAGGAGCCGC
	RT-template	ATCTTGTTCCGGTG
	PBS	GCTCCTGATA
pegRNA#2	Target site	GCAAGAATATCTTGTTCCGG
	RT-template	ATATCAGGAGCCACCG
	PBS	GAACAAGATATT

2.7.3 CRISPR/Cas9 constructs to mimic *BoOr^{Mut}*

Constructs for mimicking the *BoOr^{Mut}* in the *TaOR* homoeologues using CRISPR/Cas9 gene editing were produced following the method by Smedley and colleagues (2021), with the following amendment: *GRF4-GIF1* developmental regulators (pL1P5_GRF-GIF) were included on the finished Level 2 pTaOR-Cas9 construct. The *hptII* selection gene was driven by the switchgrass ubiquitin 2 promoter (PvUbi2 pro). As described in this method, for each guide sequence, a sense and complement oligonucleotide were ordered, which are complementary to each other and produce 4 bp overhangs for correct MoClo Golden Gate Cloning into the L1 P3 TaU6 sgRNA acceptor (pL1P3_TaU6acc) or L1 P4 TaU6 sgRNA acceptor (pL1P4_TaU6acc) plasmids. These sequences are found in Table 2.6, with the overhangs indicated in red. These oligonucleotides were ordered through Merck Custom DNA Oligos

synthesis service (Merck, Darmstadt, Germany). The final Level 2 plasmid is referred to as pTaOR-Cas9. Its plasmid map can be found in Figure 5.7.

Table 2.6 Oligonucleotide sequences for guide RNAs targeting TaOR homoeologues to mimic the *BoOr^{Mut}* mutation. Overhangs for MoClo Golden Gate Cloning are indicated in red. Comp.=complement.

Guide RNA	Target gene	Strand	Sequence
sgRNA 1	TaOR-6A and TaOR-6B	Sense	AAACCTGGTTAGTAGGCCCTTACC
sgRNA 1	TaOR-6A and TaOR-6B	Comp.	CTTGGGTAAGGGCCTACTAACCAG
sgRNA2	TaOR-6D	Sense	AAACCTGGTTAGTAGGCCCTCACC
sgRNA2	TaOR-6D	Comp.	CTTGGGTGAGGGCCTACTAACCAG

2.8 Transformation of organisms

2.8.1 *Agrobacterium tumefaciens* transformation

Wheat expression constructs were electroporated into the hypervirulent *Agrobacterium* strain AGL1 (Lazo et al. 1991), as described in Hayta and colleagues (2021). This strain also included the helper plasmid pAL155, which contained an additional *VirG* gene. Standard inoculums of *Agrobacterium* were prepared as described by Hayta and colleagues (2021).

2.8.2 Wheat transformation

For both the bread wheat cultivar Cadenza and durum wheat cultivar Kronos, the transformation was performed using the method described by Hayta and colleagues (2019, 2021). All media recipes used are found within this methods paper. Here, immature wheat embryos were isolated from developing grains under sterile conditions, centrifuged for 10 minutes at 4°C, inoculated with *Agrobacterium* AGL1 containing the wheat expression construct within the pGGG-M plasmid and co-cultivated for 3 days in the dark. After co-cultivation, the embryogenic axes were excised, and the embryos were transferred to callus induction plates. After 5 days, calli were moved to selection media 1 containing hygromycin to select for transgenic callus. After 2 weeks, the calli were transferred to selection media 2 with a higher concentration of hygromycin for another 2 weeks. For the transformation of the pTaOR-Cas9 construct, calli were kept on selection media 2 for 8 weeks in total; during this, they were moved to fresh media every 2 weeks. After the 2 weeks on selection media 2 (or 8 weeks for pTaOR-Cas9 constructs), calli on the selection media 2 were moved to a lit culture room under a 16-hour photoperiod with a single layer of paper towel on top of the plates for low-light conditions. After a week, the calli were transferred to regeneration

media without the tissue paper covering in the same 16-hour photoperiod to promote shoot production. Every 2 weeks, calli were transferred onto new regeneration media. During this time, regenerated shoots 1-2 cm long with visible roots were transferred to De Wit culture tubes with rooting media. After around 10 days in rooting media, regenerated plantlets were moved to 24-CT trays with 'John Innes Cereal Mix' (as detailed in Section 2.2.1). T₀ plants were grown in controlled environment rooms (16 hours of light at 20°C and 8 hours of dark at 15°C), while subsequent generations were grown in standard glasshouse conditions (as detailed in Section 2.2.1). Transgenic plantlets were confirmed by copy number analysis (as detailed in Section 2.9.2).

2.9 Genotyping

2.9.1 KASP genotyping

DNA extraction from wheat leaf tissue was performed by the Genotyping and DNA Extractions platform at the John Innes Centre, following the method by Pallotta and colleagues (2003). Kompetitive Allele-Specific PCR (KASP) genotyping was done using standard protocols previously described by Ramirez-Gonzalez and colleagues (2015a), with the following amendments: assays were performed in FrameStar 384-well skirted PCR plates (4titude® Limited, Surrey, UK); PACE mix (3CR Bioscience, Harlow, United Kingdom) was used instead of KASP mix; 2 µl of PACE mix was used for each reaction. KASP primers were designed with PolyMarker using the *T. turgidum* Kronos reference genome (Ramirez-Gonzalez et al. 2015b). The KASP primers used are found in Table 2.7. Following KASP reactions, the plates were read using the PHERAstar microplate reader (BMC Labtech, Ortenberg, Germany). Samples were grouped into wild-type, mutant and heterozygotes based on the relative FAM and HEX fluorescence levels on the KlusterCaller software (LGC Biosearch Technologies, Hoddesdon, UK). Figure 2.4 shows an example of the KlusterCaller output. KASP genotyping for the segregating F₂ *Tdor* mutant lines was performed by the Genotyping and DNA Extraction platform at the John Innes Centre.

Table 2.7 KASP primers for genotyping the *Tdor* mutants. Primer sequences are given 5'–3'. The wild-type allele had the FAM tag (gaaggtgaccaagttcatgct) on the 5' end. The mutant allele had the HEX tag (gaaggtcggagtcaacggatt) on the 5' end. For the wild-type and mutant alleles, nucleotides in capital letters are the substitutions that discriminate the wild-type and mutant alleles. For the common primer, these are the homoeologous SNPs that make the primer genome specific.

Gene	Line	Mutation	Wild-type allele (FAM)	Mutant allele (HEX)	Common primer
<i>TdOR-6A</i>	K0329	R123K	ctcttttcacacagattcgcaG	ctcttttcacacagattcgcaA	caggggaggcaaaaaCggT
<i>TdOR-6B</i>	K4335	G254E	agattccatgatgttgctgtagG	agattccatgatgttgctgtagA	gttgatcaccatcactgaatgT

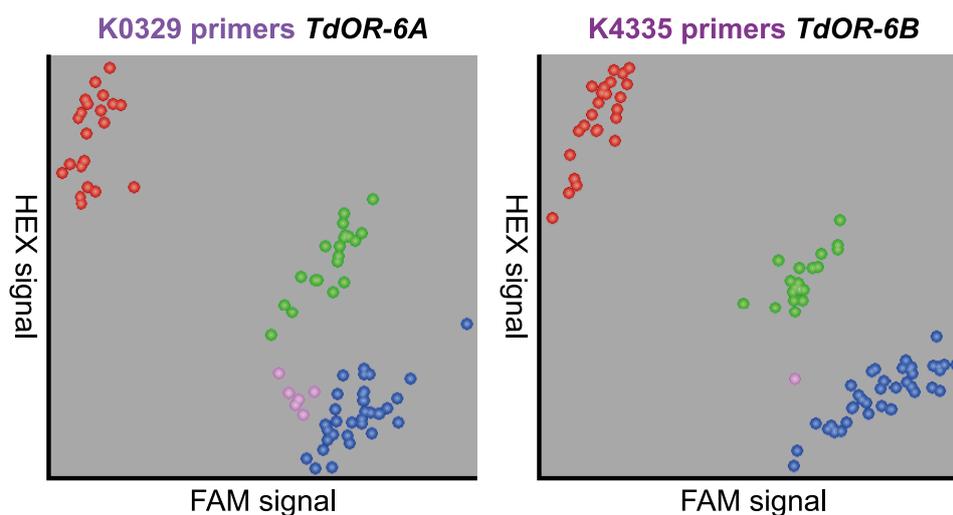


Figure 2.4 KlusterCaller plot for KASP genotyping using specific KASP primers for the K0329 and K4335 mutations. Each dot represents the proportion of fluorescence from the FAM (wild-type) or HEX (mutant) probe. Homozygous mutant plants are clustered in red with a high proportion of the HEX signal. Homozygous wild-type plants are clustered in blue with a high proportion of the FAM signal. Heterozygous plants are clustered in green with balanced FAM and HEX signals. Unknown genotypes are shown in pink.

2.9.2 Transgene copy number analysis

DNA was extracted from 100 mg leaf samples of these seedlings using Qiagen QIAcube DNA Extractions (Qiagen, Venlo, Netherlands) by the Genotyping and DNA Extractions platform at the John Innes Centre. Transgene copy number analysis was performed using quantitative real-time PCR (qPCR) with TaqMan™ probes (Applied Biosystems, Waltham, Massachusetts, US) for the hygromycin resistance gene (*hptII*) and *Constans-like* (*CO2*) genes following the method described by Hayta and colleagues (2019). Primer and probe sequences are found in Table 2.8. qPCR was run on a CFX96 Real-Time Detection System (Bio-Rad, Hercules, California, US). The channels used were FAM-TAMRA and VIC-TAMRA. PCR cycling conditions were: 95°C 15 minutes, then 40 cycles of 95°C 15 seconds and 60°C 60 seconds.

Table 2.8 Primer and probe sequences used for quantitative real-time PCR to determine transgene copy number. The *hptII* hygromycin selection gene was included in the overexpression constructs. Primers ending in ‘F’ and ‘R’ refer to forward and reverse primers, respectively. Primers ending in ‘P’ refer to probes.

<i>hptII</i> gene (on transgene)		<i>CONSTANS</i> -like <i>CO2</i> gene (internal control)	
HygF3	GGATTTCGGCTCCAACAATG	Con2F1	TGCTAACCGTGTGGCATCAC
HygR2	TATTGGGAATCCCCGAACATC	Con2R1	GGTACATAGTGCTGCTGCATCTG
HygP	FAM-CAGCGGTCATTGACTGGAGCGAGG-TAMRA	Con2P	VIC-CATGAGCGTGTGCGTGTCTGCG-TAMRA

2.9.3 Sanger sequencing for prime editing and CRISPR/Cas9 editing

For determining whether the prime editing lines and *BoOr^{Mut}* mimic CRISPR/Cas9 lines had germline edits, Sanger sequencing was used on amplicons containing the editing regions. DNA was extracted from 100 mg leaf samples of these seedlings using Qiagen QIAcube DNA Extractions (Qiagen, Venlo, Netherlands) by the Genotyping and DNA Extractions platform at the John Innes Centre. Homoeologue-specific PCR primers to amplify the ‘golden SNP’ region and the *BoOr^{Mut}* site from each of the three genomes and homoeologue-specific sequencing primers were designed on Geneious Prime 2022.2 software (<https://www.geneious.com>). These primers can be found in Table 2.9. PCR cycling conditions were: 94°C 3 minutes, 40 cycles of 94°C 30 seconds, ‘Ta’ 1 minute, 72°C 90 seconds, followed by a final extension at 72°C 5 minutes. ‘Ta’ is the annealing temperature found in Table 2.9. Following the manufacturer's protocol, PCR was carried out using GoTaq® DNA Polymerase (Promega, Madison, Wisconsin, US), including MgCl₂.

Any leftover primers from the PCR products were removed using *Exonuclease I* (NEB, Ipswich, Massachusetts, US) and rAPid Alkaline Phosphatase (Merck, Darmstadt, Germany). The components for this reaction are in Table 2.10. The conditions for this reaction were: 37°C 30 minutes, 80°C 10 minutes, 20°C 30 seconds, followed by a 10°C hold step. Sanger sequencing standard Big Dye reaction was performed using the BigDye™ Terminator v3.1 Cycling Sequencing Kit (Applied Biosystems, Waltham, Massachusetts, US) following the manufacturer's protocol. The sequencing primers used are found in Table 2.9.

Table 2.9 Primer sequences used for the Sanger sequencing of editing events. ‘PE’ primers were used for amplifying the ‘golden SNP’ region of *TaOR*. ‘Cas9’ primers were used for amplifying the *BoOr^{Mut}* mimic target site within *TaOR*. Ta=annealing temperature, bp=base pairs.

Primer	Primer sequence	Ta used	Size	Specific Genome and use
PE-A-F	GATTTAGGTGCAGTAATGATTGG	54°C	383 bp	Amplifying A-genome
PE-A-R	CCTCACAAGTTCTAGCTCATTAG			Amplifying A-genome
PE-A Seq	CTAACATGGTTACCAATTTACAAGTAT	N/A	N/A	Sequencing A-genome
PE-B-F	GACTGGAGATTTAGGTACAGTAATGAC	54°C	615 bp	Amplifying B-genome
PE-B-R	AGAGGAAAATATGATGCCCGAATGT			Amplifying B-genome
PE-B Seq	CTAACATCGTTACCAATTTACAAGTCC	N/A	N/A	Sequencing B-genome
PE-D-F	GCCAATCAGATTCAGGAGTAAAAAC	54°C	639 bp	Amplifying D-genome
PE-D-R	GAAAGAAGATACACTAGAAGGGCC			Amplifying D-genome
PE-D Seq	CTAACATCGTTACCAATTTACAAGTCT	N/A	N/A	Sequencing D-genome
Cas9-A-F	CCTTCGAGTGAGTTCTGCTA	56°C	492 bp	Amplifying A-genome
Cas9-A-R	ACCATGTTGTTTCATGATGAG			Amplifying and sequencing A-genome
Cas9-B-F	CCTTCGAGTGAGTTCTGCTT	56°C	591 bp	Amplifying B-genome
Cas9-B-R	GGCAGATGAACATTTCCGG			Amplifying and sequencing B-genome
Cas9-D-F	CCTTCGAGTGAGTTCTGCTG	56°C	590 bp	Amplifying D-genome
Cas9-D-R	GGCAAATGAACATTTCCGG			Amplifying and sequencing D-genome

Table 2.10 Components for the Sanger sequencing amplicon clean-up reaction.

Component	1x (µl)
rAPid Alkaline Phosphatase	1
<i>Exonuclease I</i>	0.5
H ₂ O	1.5

Products of the Big Dye sequencing reaction were then run on a 3730xl DNA analyser (Thermo Fisher, Waltham, Massachusetts, US) in Optima DTR 96-well plates (EdgeBio, San Jose, California, US) by the Molecular Genetics platform at the John Innes Centre. Analysis of the Sanger sequencing chromatograms was performed using Geneious Prime 2022.2 software (<https://www.geneious.com>).

2.9.4 Illumina next-generation sequencing for prime editing

Illumina next-generation sequencing was used for detecting somatic editing events in the prime editing plants. First, Illumina sequencing amplicons were produced, barcoded for which plant sample they had come from, and then sequenced. DNA was extracted from 100 mg leaf samples of these seedlings using Qiagen QIAcube DNA Extractions (Qiagen, Venlo, Netherlands) by the Genotyping and DNA Extractions platform at the John Innes Centre. Primers that could target all three homoeologues were designed using Geneious Prime

2022.2 software (<https://www.geneious.com>). Homoeologous SNPs distinguished individual homoeologues within the reads. Primers were barcoded for each plant, and these are found in Table 2.11. PCR cycling conditions were: 98°C 30 seconds, 30 cycles of 98°C 7 seconds, 64°C 15 seconds, 72°C 20 seconds, followed by a final extension at 72°C 2 minutes and a 10°C hold. Following the manufacturer’s protocol, PCR was performed using Q5® High-Fidelity DNA Polymerase (NEB, Ipswich, Massachusetts, US). Sequencing was performed by Novogene Target Region Sequencing service (Novogene, Beijing, China) on a NovaSeq 6000 (Illumina, San Diego, California, US) using paired-end 250 bp reads.

Table 2.11 Primer sequences for Illumina next-generation sequencing of prime editing lines. Barcodes for each primer are indicated in red.

Plant	Primer name	Barcoded Primer Sequence
0 copy control 1	ngsPE_F1	ATCACGGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R1	CGTGATCAAACCAACACACCTCCTCC
0 copy control 2	ngsPE_F2	CGATGTGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R2	ACATCGCAAACCAACACACCTCCTCC
0 copy control 3	ngsPE_F3	TTAGGCGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R3	GCCTAACAAACCAACACACCTCCTCC
Wild-type Cadenza 1	ngsPE_F4	TGACCAAGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R4	TGGTCACAAACCAACACACCTCCTCC
Wild-type Cadenza 2	ngsPE_F5	ACAGTGGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R5	CACTGTCAAACCAACACACCTCCTCC
PE T ₀ 1 copy	ngsPE_F6	GCCAATGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R6	ATTGGCCAAACCAACACACCTCCTCC
PE T ₀ 2 copies 1	ngsPE_F7	CAGATCGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R7	GATCTGCAAACCAACACACCTCCTCC
PE T ₀ 2 copies 2	ngsPE_F8	ACTTGAAGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R8	TCAAGTCAAACCAACACACCTCCTCC
PE T ₀ 6 copies	ngsPE_F9	GATCAGGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R9	CTGATCAAACCAACACACCTCCTCC
PE T ₀ 8 copies	ngsPE_F10	TAGCTTGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R10	AAGCTACAAACCAACACACCTCCTCC
PE T ₁ 1 copy	ngsPE_F11	GGCTACGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R11	GTAGCCCAAACCAACACACCTCCTCC
PE T ₁ 2 copies 1	ngsPE_F12	CTGTAGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R12	TACAAGCAAACCAACACACCTCCTCC
PE T ₁ 2 copies 2	ngsPE_F13	AGTCAAGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R13	TTGACTCAAACCAACACACCTCCTCC
PE T ₁ 4 copies	ngsPE_F14	AGTTCGGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R14	GGAACTCAAACCAACACACCTCCTCC
PE T ₁ 5 copies	ngsPE_F15	ATGTCAAGGATAACCATGATTCTGTTCTTGAAC
	ngsPE_R15	TGACATCAAACCAACACACCTCCTCC

Sequence quality was checked using FastQC (v0.12.0; <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and barcoded sequences were demultiplexed using the Barcode Splitter tool in the FASTX-Toolkit (v0.0.14; http://hannonlab.cshl.edu/fastx_toolkit/index.html). Barcodes were trimmed using Cutadapt (v4.7) (Martin 2011), and sequences were aligned to *TaOR* homoeologues using Bowtie2 (v2.5.3) (Langmead and Salzberg 2012). Indexing of sequences was performed using Samtools (v1.19.2) (Langmead and Salzberg 2012). Variant calling was performed using the Mutect2 tool in GATK (v4.5.0) (Van der Auwera and O'Connor 2020).

2.10 Statistics and data visualisation

Statistical analyses were carried out in R version 4.3 (R Core Team 2023) using the 'emmeans' R package (v1.10.0) (Lenth 2023). Statistical tests carried out for the data are detailed within the results sections. Non-parametric statistical tests were used for data that violated assumptions of normality. Data manipulation was done in R using the packages 'dplyr' (v1.1.4) (Wickham et al. 2023a) and 'tidyr' (v1.3.1) (Wickham et al. 2023b). Data visualisation and figure generation was done in R using the package 'ggplot 2' (v3.5.0) (Wickham 2016). Adobe Illustrator (v28.3, Adobe Inc), Geneious Prime 2022.2 software (<https://www.geneious.com>) and Microsoft Excel (v2312, Microsoft Corporation) were also used for data visualisation and figure generation.

3 Investigating the grain carotenoid diversity of the Watkins global landrace collection and searching for associated allelic variation

3.1 Chapter Introduction

Wheat grains contain a wide range of carotenoids such as lutein, zeaxanthin, α -carotene, β -carotene and β -cryptoxanthin. These compounds act as important antioxidants and are responsible for the yellow colour of wheat flour, playing a crucial role in the quality of wheat products (Lachman et al. 2017). Modern durum wheat has high levels of these pigments compared to bread wheat due to semolina yellowness being an important quality trait in durum wheat breeding (Digesù et al. 2009). Conversely, the consumer preference for white bread has led to the selection of bread wheat varieties with whiter flour and lower carotenoid content (Shewry and Hey 2015).

In addition to affecting flour colour, carotenoids have significant health benefits. The macular carotenoids, lutein and zeaxanthin, accumulate in the macula region of the eye and are associated with reducing the risk of age-related macular degeneration (Ma et al. 2012; Johnson 2014). Lutein is also suggested to support brain development and cognitive performance, with correlations observed between post-mortem brain lutein concentrations and pre-mortem cognitive function (Erdman et al. 2015). Provitamin A (PVA) carotenoids, such as β -carotene, β -cryptoxanthin and α -carotene, are essential for eye development as they can be converted into vitamin A by the body. Among them, β -carotene, with two β -ionone rings, has the highest PVA activity, making it a key target for biofortification efforts (Watkins and Pogson 2020; Zheng et al. 2020a). Consequently, increasing the grain carotenoid content of wheat is desirable, and the increasing consumer preference and awareness of the health benefits associated with wholemeal and brown bread over white bread presents an opportunity to increase the low carotenoid content of bread wheat (Lockyer and Spiro 2020). To achieve this, new material with high carotenoid content and novel allelic diversity associated with carotenoid content should be identified for use in breeding programmes. Achieving this goal requires the identification of new genetic material with high carotenoid content and novel allelic diversity. Such materials can then be utilised in breeding programmes to develop wheat varieties with enhanced nutritional qualities.

Carotenoid content in wheat is a quantitative trait with high heritability, facilitating the identification of quantitative trait loci (QTLs) associated with this trait (Colasuonno et al. 2019). Traditional approaches, such as linkage mapping in biparental populations, have identified loci and candidate genes regulating carotenoid content (Elouafi et al. 2001; Pozniak et al. 2007; Singh et al. 2009; Tsilo et al. 2011; Colasuonno et al. 2014). However, these methods are marked by the lengthy and labour-intensive generation of segregating populations. This often results in low-resolution QTL mapping, decreasing the usefulness of detected QTLs in breeding programmes (Shi et al. 2017).

Instead, several studies have employed association mapping to identify QTLs associated with carotenoid traits (Reimer et al. 2008; Colasuonno et al. 2017a; N'Diaye et al. 2017; Roselló et al. 2018; Requena-Ramírez et al. 2022). This method screens existing populations, enabling the analysis of a broader diversity of material than linkage mapping. It uses past recombination events to achieve higher-resolution mapping, proving more effective for the precise mapping of QTLs and facilitating the identification and characterisation of candidate genes. Furthermore, genome-wide association studies (GWAS) have increasingly been used over linkage mapping for many different traits due to the availability of high-density single-nucleotide polymorphism (SNP) arrays and cheaper sequencing technologies allowing for genotype-by-sequencing methods (Torkamaneh and Belzile 2022).

Various wheat collections, including modern durum wheat collections, Canadian durum wheat collections and landrace collections have been screened for carotenoid content and used in association mapping to identify carotenoid-associated QTLs (Colasuonno et al. 2019; Requena-Ramírez et al. 2022). Notably, the Watkins global landrace collection, comprising 356 durum wheat landraces from 25 countries and 828 bread wheat landraces from 32 countries, remains unexplored. This was assembled in the late 1920s to early 1930s by Arthur Ernest Watkins and offers a unique snapshot of global wheat genetic diversity prior to modern breeding practices, representing a rich source of diversity (Wingen et al. 2014). The Watkins hexaploid collection has high phenotypic diversity, and it has demonstrated its value in pre-breeding through the identification of thousands of high-resolution QTLs and significant marker-trait association (MTAs) for major traits within it (Wingen et al. 2014; Cheng et al. 2023). However, the Watkins tetraploid collection has been largely overlooked until recent efforts by the Germplasm Resource Unit at the John Innes Centre to compartmentalise and further investigate the diversity within this

collection. Given the Watkins global landrace collection's unique diversity, it is a promising resource for identifying novel phenotypic and allelic diversity within wheat. Therefore, characterising the carotenoid diversity of the global landrace collection is desirable as it could provide a valuable source of novel allelic diversity for wheat carotenoid biofortification.

The carotenoid content of wheat is commonly screened by measuring flour's yellow pigment content (YPC), serving as a relative value of total carotenoid content. This is a spectrophotometric quantification where the absorbance of total pigments extracted using water-saturated butanol (WSB) is measured (Colasuonno et al. 2019). The industry-standard YPC method (AACC 14-50) requires significant quantities of flour (8 grams) and WSB (40 ml) alongside a lengthy extraction period (16-18 hours), making it impractical for large-scale screening (AACC 2009). To address these limitations, Beleggia and colleagues (2010) developed a micro-method that significantly reduces both the amount of flour (10–100 mg) and WSB (250–500 μ l) required and shortens the extraction time to just 15 minutes using sonication. However, this adapted method is unsuitable for absorbance readings on a plate reader, which would further increase the speed of this method.

While YPC provides a useful approximation of carotenoid content in wheat flour, it does not differentiate between individual carotenoid compounds or provide an absolute value of carotenoid content. For precise quantification of specific carotenoids, high-performance liquid chromatography (HPLC) is necessary. HPLC can detect carotenoid compounds based on their unique biochemical properties and accurately quantify known compounds using standard curves generated from reference carotenoid samples (Colasuonno et al. 2019). Given the varied health benefits of different carotenoids, knowing the amounts of specific carotenoid compounds within wheat allows for breeding with different biofortification objectives.

In this chapter, I aimed to characterise the diversity of carotenoid content within the Watkins tetraploid collection using a high-throughput YPC method I adapted, followed by an in-depth investigation of the carotenoid content within this collection using HPLC. I asked if useful phenotypic diversity could be found and if it could be associated with genotypic diversity (in two available resolutions). I found considerable variation in YPC and the carotenoid content of α -carotene, β -carotene, lutein and zeaxanthin. I also identified some accessions with high total carotenoid content and some with high content of specific desirable carotenoids for PVA biofortification. I used the Watkins tetraploid collection to

identify MTAs, QTLs and candidate genes associated with these quantified carotenoid traits (YPC, α -carotene content, β -carotene content, lutein content, zeaxanthin content, total grain carotenoid content and proportion of β - β branch carotenoids). I also asked to what extent grain carotenoid content differs between tetraploid and hexaploid global landraces grown prior to the introduction of industrialised breeding. To answer that, I compared the carotenoid diversity within the Watkins tetraploid collection to that found in a core collection of the Watkins hexaploid collection.

3.2 Results

3.2.1 The Watkins tetraploid collection has high variability in carotenoid content traits

To explore the global landrace diversity of YPC that predates modern breeding, I screened 295 tetraploid landraces of the Watkins collection. I grew each accession in 1-metre plots in a field trial, for the analysis of grain carotenoid content (GCC). Prior to harvest, diversity in the colour of floral organs was observed (Appendix Material 1). I developed a high-throughput YPC method to facilitate screening large collections of wheat accessions by using a plate reader for light absorbance measurements. This was based on adaptations made by Beleggia and colleagues (2010) of the industry-standard AACC 14-50 method to use less flour and less extraction solvent. Before screening the whole panel, I tested how the high-throughput YPC method compared to the industry-standard AACC 14-50 method. I found no significant differences between the two methods when performing multiple measurements of the same flour sample ($p=0.38$, $t(8)=0.93$, Student's t-test; Appendix Material 2). This suggested the high-throughput YPC method is comparable to the AACC 14-50 extraction method; therefore, I used this high-throughput screen for measuring the YPC of the Watkins tetraploid collection.

In total, 295 samples of the Watkins tetraploid collection were analysed, each originating from a single plot in the field. Additionally, 93 field control plots of the variety Miradoux were analysed, a modern cultivar of durum wheat bred to have a high YPC content and grown within the UK. A single replicate was performed on each accession, and the results are illustrated in Figure 3.1 and summarised in Table 3.1. The full results are provided in Appendix Material 3. The Miradoux field control plots' average YPC was higher than all the accessions within the Watkins tetraploid collection.

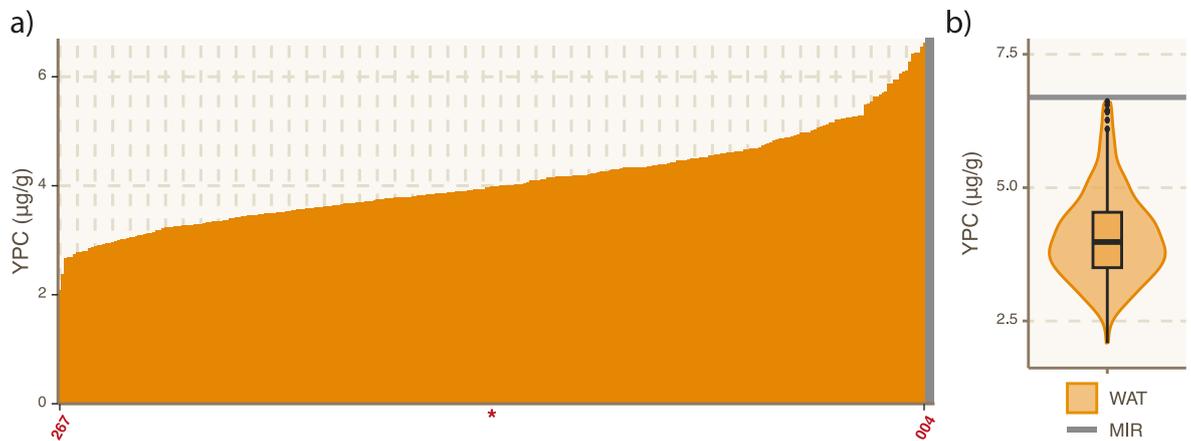


Figure 3.1 Yellow pigment content (YPC) of the field-grown Watkins tetraploid collection and Miradoux field controls, measured using the high-throughput YPC method. (a) Bar graph of accessions ordered for their YPC value. Locations of the minimum, median and maximum accessions are shown by red bold numbering (WAT11800...) or red asterisk. Grey bar on the right shows the mean of the measured Miradoux field controls. (b) Violin plot of YPC values for the Watkins tetraploid collection (WAT). The mean of the Miradoux field controls (MIR) is shown by the horizontal grey bar. Watkins tetraploid collection YPC results n=295; Miradoux field controls YPC results n=93.

Previous studies have found a negative correlation between YPC and grain size, possibly due to a dilution effect in larger grains by starch (Alvarez et al. 1999, 199; Clarke et al. 2006; Zhang et al. 2008; Digesù et al. 2009; Bilgin et al. 2010). To investigate whether YPC also correlated with grain size, the thousand-grain weights (TGW) of the Watkins tetraploid collection and Miradoux field controls were measured. No significant correlation or relationship between YPC and TGW was found for accessions within the Watkins tetraploid collection, as determined by linear regression analysis (coefficient for TGW: $p=0.536$, $\beta=0.005$, $t(293)=0.619$; Appendix Material 4a) or as determined by a Pearson's product-moment correlation ($r(293)=0.036$, $p=0.537$). For the Miradoux field plots, a weak negative correlation was found between YPC and TGW ($r(89)=-0.261$, $p=0.013$, Pearson's product-moment correlation), and a significant negative relationship was found (coefficient for TGW: $p=0.013$, $\beta=-0.086$, $t(89)=-2.547$, linear regression; Appendix Material 4b). This suggests that there is a small effect on YPC by grain weight, although this relationship only explains about 5.7% of the variability. Therefore, while there is a statistically significant negative relationship between grain weight and YPC, the effect is not particularly strong.

Table 3.1 Carotenoid content of the Watkins tetraploid collection and Miradoux field controls. Units for carotenoid measurements are in $\mu\text{g/g}$. Total carotenoid content was calculated by summing the carotenoid compounds identified using standards: α -carotene, β -carotene, lutein, and zeaxanthin. Proportion β - β branch refers to the proportion of zeaxanthin and β -carotene compared to lutein and α -carotene. Proportion PVA refers to the proportion of provitamin A carotenoids α -carotene and β -carotene compared to lutein and zeaxanthin. Watkins tetraploid collection YPC results n=295; Miradoux field controls YPC results n=93; Watkins tetraploid collection HPLC results n=337; Miradoux field controls HPLC results n=16. SD=standard deviation, CV=coefficient of variation, ND.=not detected, Prop.=proportion of.

Measurement	Watkins tetraploid collection				Miradoux field controls			
	Mean (SD)	Range	Median	CV (%)	Mean (SD)	Range	Median	CV (%)
YPC ($\mu\text{g/g}$)	4.083 (0.825)	2.086–6.613	3.981	20.2	6.651 (0.669)	5.208–8.437	6.662	10.1
α -carotene ($\mu\text{g/g}$)	0.028 (0.026)	0.003–0.207	0.020	94.5	0.052 (0.006)	0.042–0.065	0.052	11.4
β -carotene ($\mu\text{g/g}$)	0.031 (0.011)	ND.–0.065	0.031	34.7	0.051 (0.008)	0.038–0.065	0.052	15.3
Lutein ($\mu\text{g/g}$)	0.521 (0.211)	0.121–1.93	0.482	40.4	1.02 (0.145)	0.714–1.329	1.053	14.2
Zeaxanthin ($\mu\text{g/g}$)	0.187 (0.084)	0.027–0.714	0.176	44.9	0.152 (0.056)	0.073–0.234	0.154	36.6
Total GCC ($\mu\text{g/g}$)	0.767 (0.257)	0.287–2.557	0.717	33.4	1.274 (0.158)	0.939–1.542	1.304	12.4
Prop. β - β branch	0.290 (0.081)	0.06–0.595	0.297	27.9	0.16 (0.04)	0.099–0.222	0.162	25.3
Prop. PVA	0.081 (0.043)	0.02–0.279	0.068	53.7	0.081 (0.006)	0.071–0.092	0.08	7.3

YPC is a measurement of the relative carotenoid content of the flour; however, this does not give absolute values of the carotenoid content or what individual carotenoid compounds are present. To measure the absolute values of individual carotenoid compounds, the Watkins tetraploid collection was also screened by HPLC following carotenoid extraction from the whole mill flour. The carotenoids α -carotene, β -carotene, lutein and zeaxanthin were measured, and total GCC was calculated by summing the measurements of these individual carotenoids. β -cryptoxanthin was not quantified as this is present in very low quantities within wheat flour (Digesù et al. 2009). In total, 337 accessions of the Watkins tetraploid collection were analysed, each by a single replicate from a single field plot. For the Miradoux field control, I analysed 16 plots selected to represent the field phenotypic range based on the YPC measurements. The results are summarised in Table 3.1 and Figure 3.2–Figure 3.4. The full HPLC results are provided in Appendix Material 3. A moderate positive correlation was found between YPC and the HPLC-measured total GCC ($r(292)=0.464$, $p<0.001$, Person's product-moment correlation), and a significant positive relationship was found between YPC and total GCC (coefficient for total GCC: $p<0.001$, $\beta=1.445$, $t(292)=8.948$, linear regression; Appendix Material 5).

The Miradoux control plots had a higher total GCC, α -carotene, β -carotene and lutein content compared to the mean of the Watkins tetraploid collection. Zeaxanthin had a higher mean within the Watkins tetraploid collection. The HPLC chromatograms showing the minimum, median and maximum total GCC accessions are shown in Figure 3.5. In addition to the known carotenoids identified through reference standards, several unknown peaks were also detected. These peaks are likely to be carotenoids, given their absorbance at 450 nm, a characteristic wavelength for carotenoids. Three distinctive peaks emerged at the retention times (RT) of 10.193 minutes, 13.637 minutes and 14.803 minutes (referred to as Peak A, Peak B and Peak C) and are highlighted on the chromatograms in Figure 3.5. These unknown compounds were not detected in Miradoux control plots. They were also not present in all the Watkins tetraploid accessions; their distribution across the collection is shown in Appendix Material 6. Consistent with YPC for the Watkins tetraploid collection, grain size did not correlate with total GCC (Appendix Material 7).

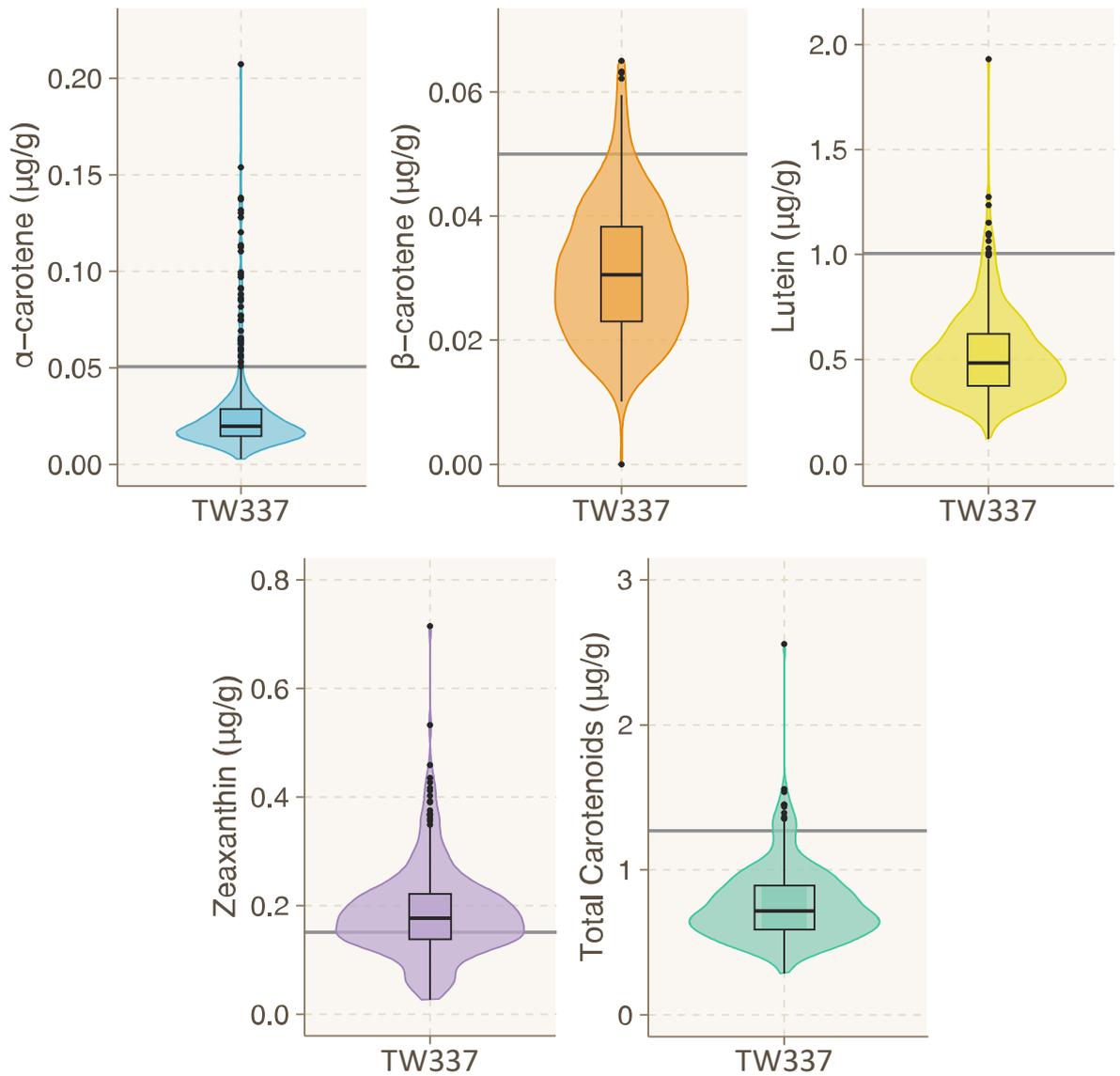


Figure 3.2 Distribution of grain carotenoid content for α -carotene, β -carotene, lutein, zeaxanthin and total carotenoid content in the Watkins tetraploid collection measured by HPLC. The grey horizontal line shows the average of the Miradoux field controls. Watkins tetraploid collection HPLC results n=337, Miradoux field controls HPLC results n=16. TW337=Watkins tetraploid collection.

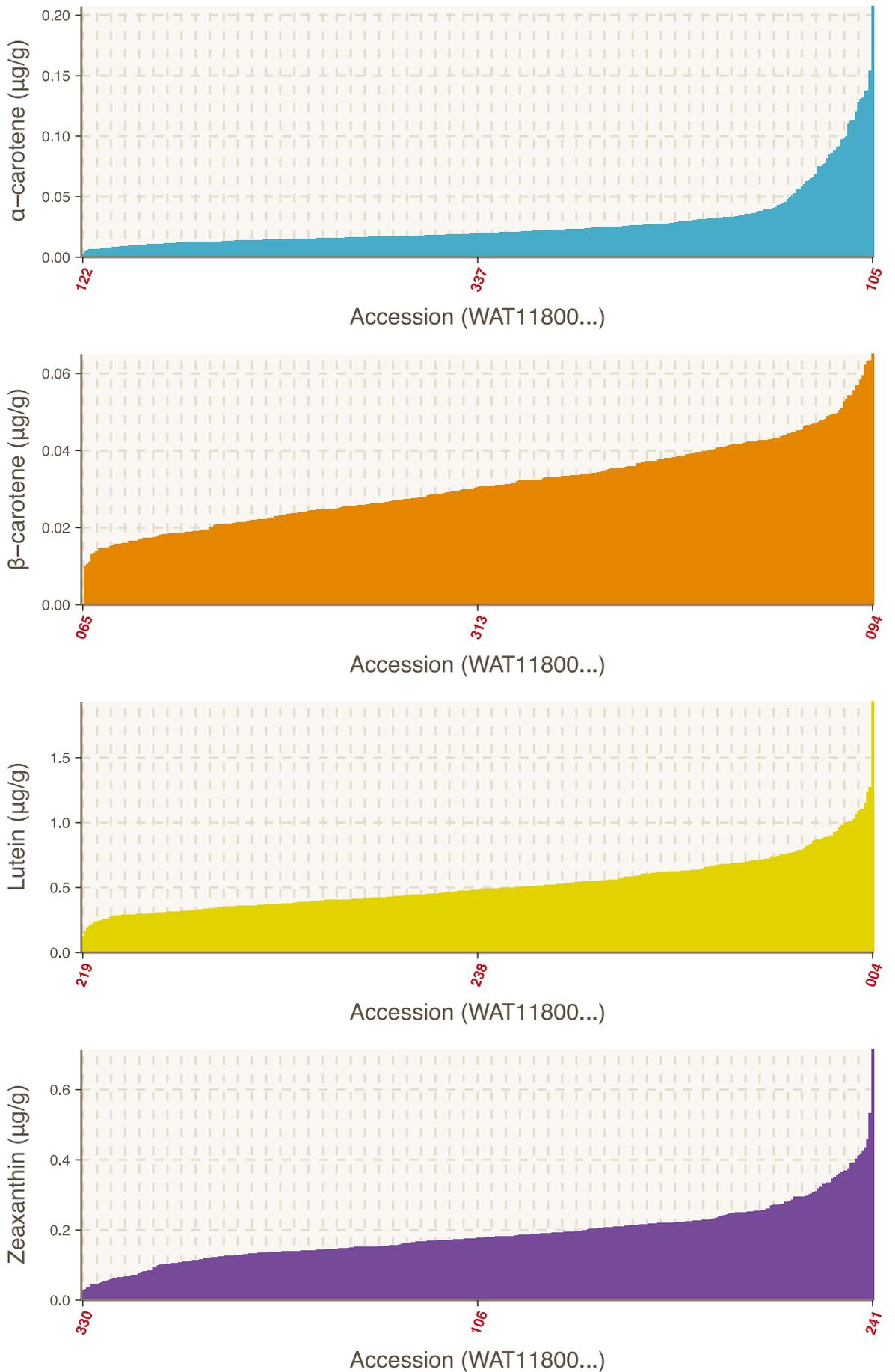
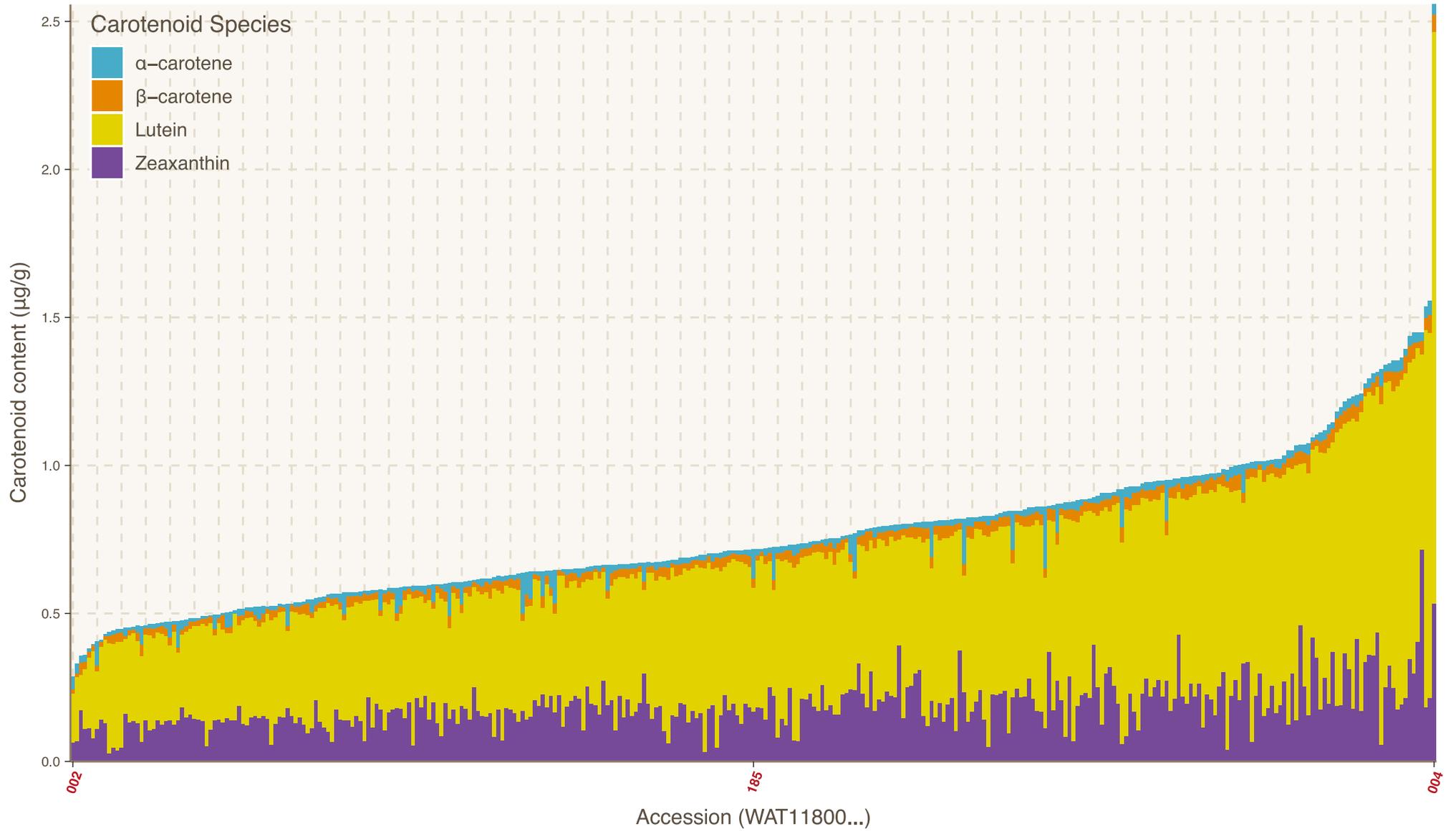


Figure 3.3 Sorted bar charts showing α -carotene, β -carotene, lutein and zeaxanthin content for accessions within the field-grown Watkins tetraploid collection. Minimum, median, and maximum accession numbers are shown in bold red text. "WAT11800..." refers to the Watkins accession ID.



On the previous page:

Figure 3.4 Stacked bar chart of the content of individual carotenoid compounds making up total grain carotenoid content within the Watkins tetraploid collection. Minimum, median, and maximum accession numbers are shown in bold red text.

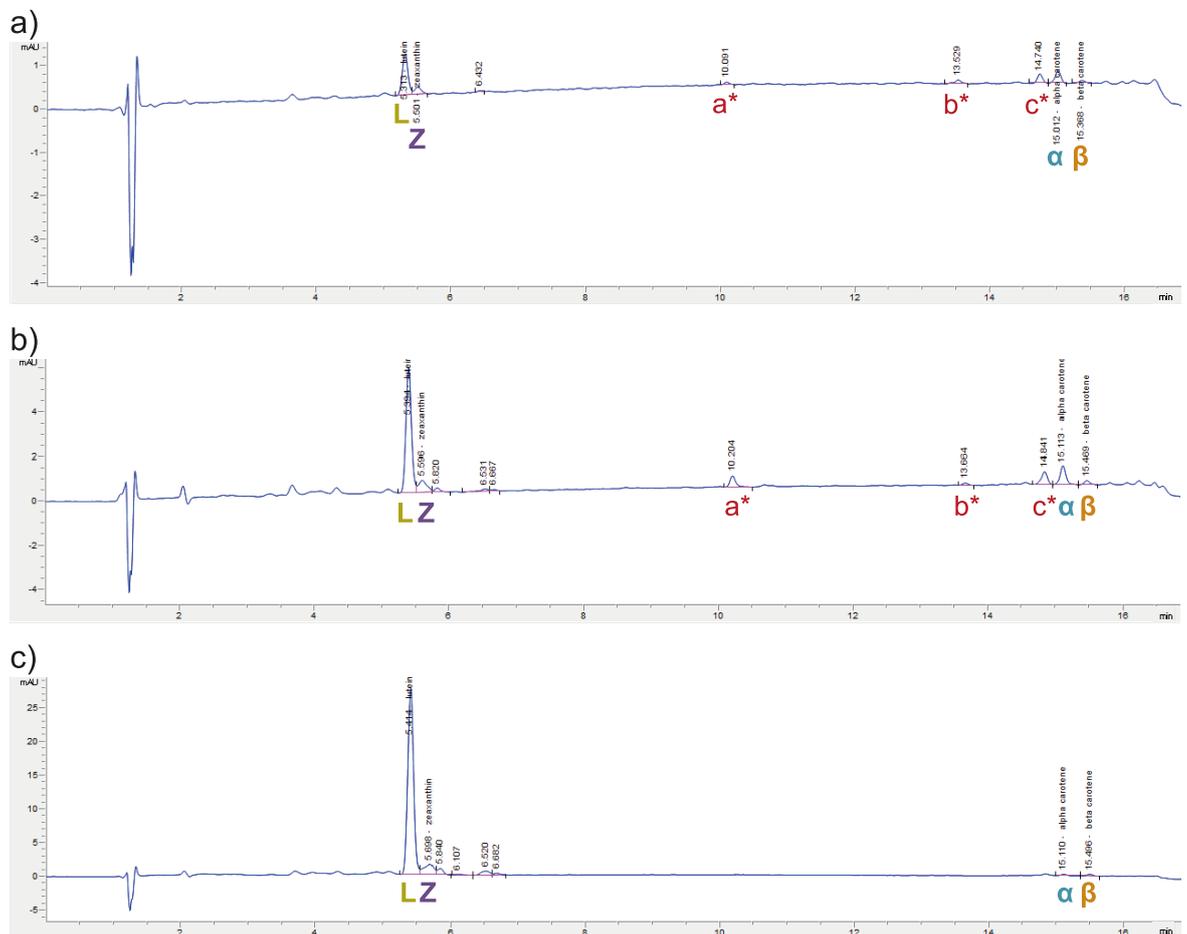


Figure 3.5 HPLC chromatograms of the accessions with the minimum, median and maximum total grain carotenoid content (GCC). (a) The lowest total GCC was found in WAT1180002, (b) WAT1180185 had the median total GCC, and (c) WAT1180004 had the highest total GCC. The peaks of lutein, zeaxanthin, α -carotene and β -carotene are labelled. Three distinct unknown peaks were found around the retention times of 10.193, 13.637 and 14.803 minutes. L=lutein, Z=zeaxanthin, α = α -carotene, β = β -carotene, a*=Peak A, b*=Peak B, c*=Peak C.

Expectedly, the most abundant carotenoid within the Watkins tetraploid collection was lutein, followed by zeaxanthin, β -carotene and α -carotene. Within the Miradoux field control lines, lutein was also the most abundant carotenoid, followed by zeaxanthin, α -carotene and β -carotene. Within both the Watkins tetraploid collection and the Miradoux field controls, the PVA carotenoids α -carotene and β -carotene had similar average amounts (Table 3.1); however, several accessions within the Watkins tetraploid collection had very high α -carotene content (Figure 3.2–Figure 3.3). The HPLC chromatogram of the accession with the highest α -carotene content (WAT1180105) is shown in Figure 3.6. A very high α -

carotene peak is observed at the α -carotene analytical standard RT, confirming the accuracy of this interesting result.

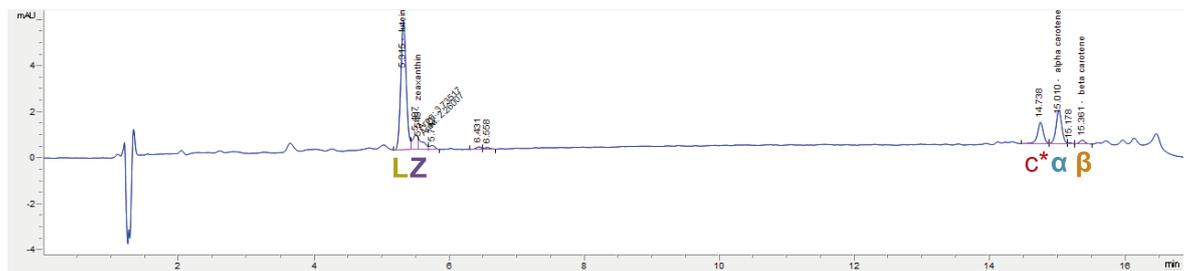
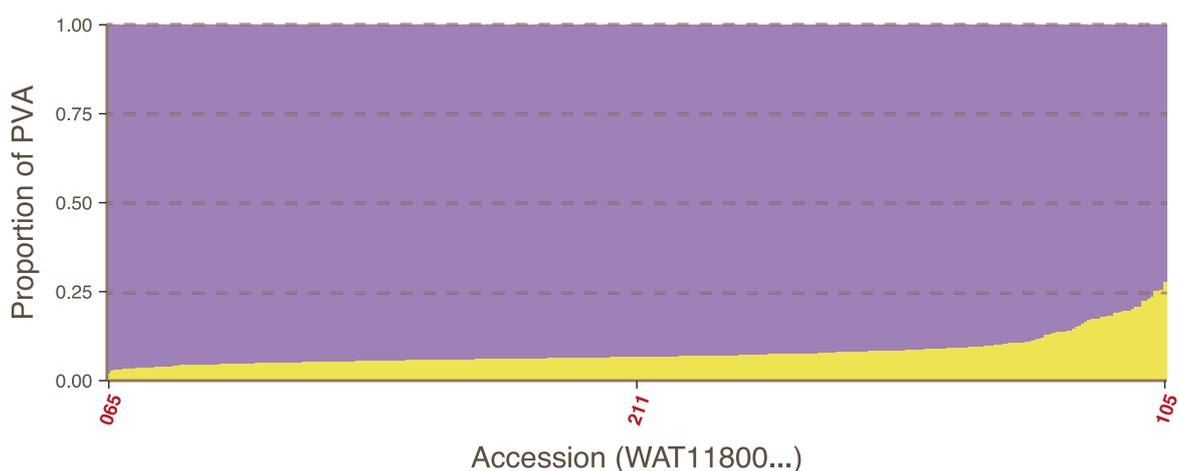
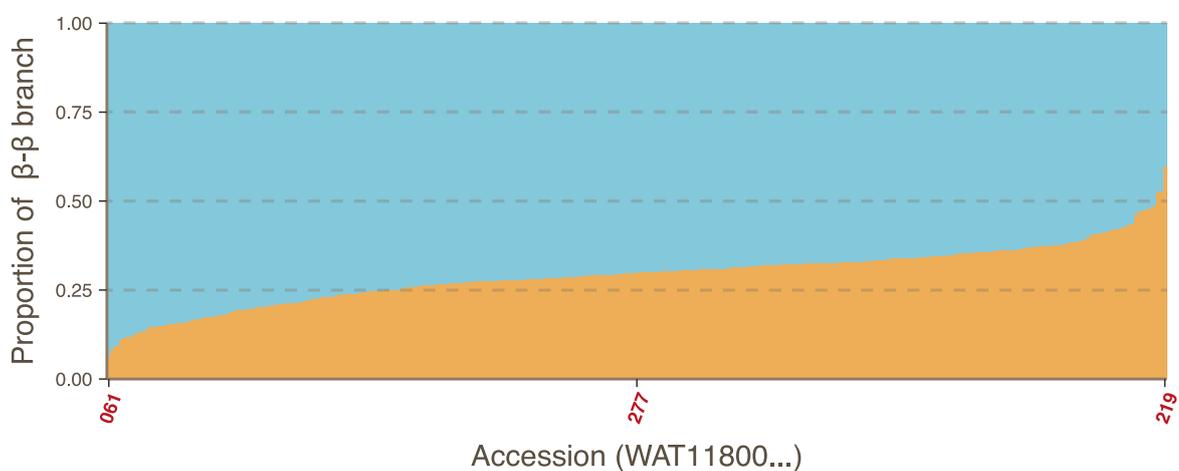


Table 3.2 Results of three replicates of HPLC on the high-carotenoid accession WAT1180004. Carotenoid contents are shown in $\mu\text{g/g}$. Rep.=replicate, SD=standard deviation, CV=coefficient of variation.

Compounds	Rep. 1	Rep. 2	Rep. 3	Mean (SD)	CV (%)
α -carotene	0.037	0.034	0.041	0.037 (0.004)	9.5
β -carotene	0.058	0.061	0.009	0.043 (0.029)	68.5
Lutein	1.930	2.021	1.917	1.956 (0.057)	2.9
Zeaxanthin	0.532	0.611	0.580	0.574 (0.04)	6.9
Total carotenoids	2.557	2.727	2.547	2.610 (0.101)	3.9

To identify accessions with increased flux into carotenoids of interest for biofortification, I investigated the proportion of carotenoids leading to the β - β branch of the carotenoid biosynthesis pathway and the proportion of PVA carotenoids. The results of this are found in Figure 3.7 and Table 3.1. Accession WAT1180219 shows the highest proportion of β - β branch carotenoids (0.595), suggesting this accession has the greatest relative flux into the β - β branch of the carotenoid pathway. The highest proportion of PVA carotenoids was found in accession WAT1180105, which also had the highest α -carotene content (Figure 3.3). The proportions of each carotenoid compound are found in Appendix Material 8.



On the previous page:

Figure 3.7 The proportion of β - β branch carotenoids and provitamin A carotenoids within the Watkins tetraploid collection. Minimum, median, and maximum accession numbers are shown in bold red text.

3.2.2 The Watkins hexaploid core collection has a slightly lower grain carotenoid content than the Watkins tetraploid collection

To compare how carotenoid diversity differed between hexaploid and tetraploid landraces before modern breeding practices, the carotenoid content of the Watkins hexaploid collection was also investigated. I screened 106 accessions from the previously described Watkins hexaploid core collection (Wingen et al. 2014), which I refer to as 'HC106' (Hexaploid Core 106 accessions). These were similarly grown alongside the Watkins tetraploid collection. Three HPLC technical replicates of the HC106 were analysed for each accession's field plot, giving 318 measurements in total, which are summarised in Table 3.3, Figure 3.8 and Figure 3.9. The full results are provided in Appendix Material 9.

Similar to the Watkins tetraploid collection, lutein was the most abundant carotenoid compound within the Watkins hexaploid core collection, followed by zeaxanthin, β -carotene and α -carotene. The Watkins hexaploid core collection had significantly less lutein content ($p < 0.001$, $t(230.43) = -4.969$, 95%CI = -0.134 to -0.058, Student's t-test), zeaxanthin content ($p < 0.001$, $t(370.48) = -10.693$, 95%CI = -0.076 to -0.053, Welch's two-sample t-test) and total GCC ($p < 0.001$, $t(235.67) = -6.408$, 95%CI = -0.194 to -0.103, Student's t-test) than the Watkins tetraploid collection (Figure 3.9). However, as shown by the 95% confidence intervals, this was not a large difference. The Watkins hexaploid core collection had significantly higher α -carotene content ($p < 0.001$, $W = 25135$, 95%CI = 0.008 to 0.013, Wilcoxon rank sum test) and β -carotene content ($p < 0.001$, $t(441) = 4.465$, 95%CI = 0.003 to 0.007, Student's t-test) than the Watkins tetraploid collection (Figure 3.9). Similarly, the Watkins hexaploid core collection had a higher proportion of PVA carotenoids than the Watkins tetraploid collection (Table 3.3).

Table 3.3 Results of the three-replicate HPLC screen of the 106-accession Watkins hexaploid core collection. The results of the single-replicate screen of the 337-accession Watkins tetraploid collection from Section 3.2.1 are also included. Units for carotenoid measurements are in $\mu\text{g/g}$. HC106=Watkins hexaploid core collection, TW337=Watkins tetraploid collection, GCC=grain carotenoid content, PVA=provitamin A carotenoids, SD=standard deviation, CV=coefficient of variation, ND.=not detected.

Compound:	α -caro.	β -caro.	Lutein	Zeaxan.	Total GCC	Prop. PVA	Prop. β - β branch	
HC106	Mean (SD)	0.032 (0.015)	0.036 (0.009)	0.426 (0.159)	0.123 (0.04)	0.617 (0.191)	0.119 (0.038)	0.265 (0.046)
	Range	ND.– 0.081	ND.– 0.067	0.194– 0.901	0.053– 0.27	0.317– 1.213	ND.– 0.194	0.162– 0.415
	Median	0.032	0.037	0.39	0.116	0.589	0.126	0.261
	CV (%)	45.7	25.5	37.4	35.1	31	31.9	17.2
TW337	Mean (SD)	0.028 (0.026)	0.031 (0.011)	0.521 (0.211)	0.187 (0.084)	0.767 (0.257)	0.081 (0.043)	0.290 (0.081)
	Range	0.003– 0.207	ND.– 0.065	0.121– 1.93	0.027– 0.714	0.287– 2.557	0.020– 0.279	0.060– 0.595
	Median	0.02	0.031	0.482	0.176	0.717	0.068	0.297
	CV (%)	94.5	34.7	40.4	44.9	33.4	53.7	27.9

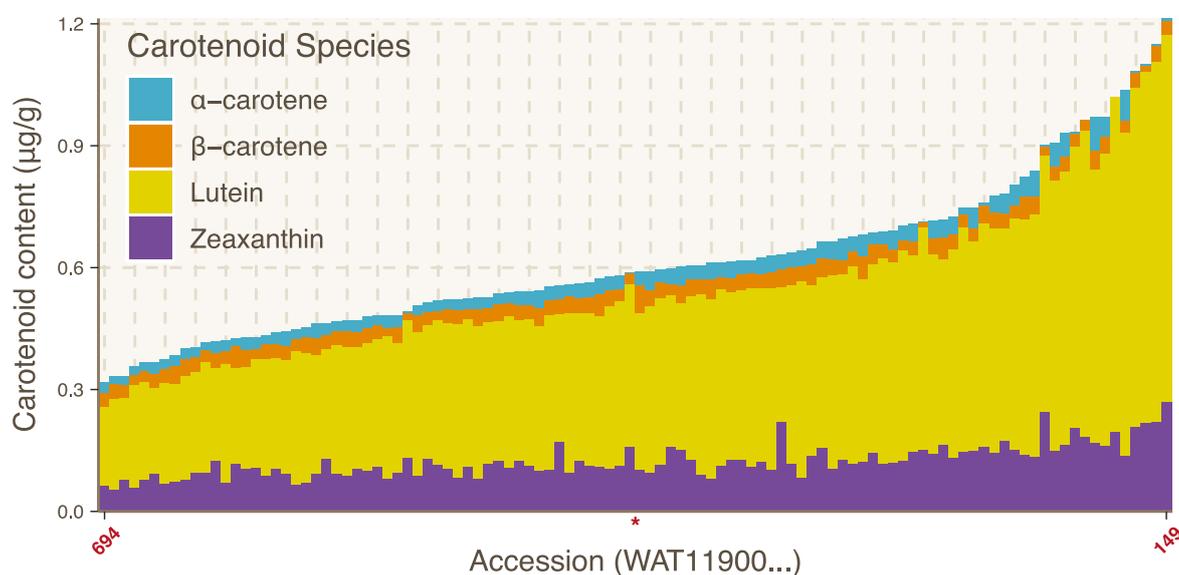


Figure 3.8 Sorted stacked bar chart showing the distribution of total carotenoids in the 106-accession Watkins hexaploid core collection. Minimum and maximum accession numbers are shown in bold red text. The location of the median value is shown by a bold red asterisk.

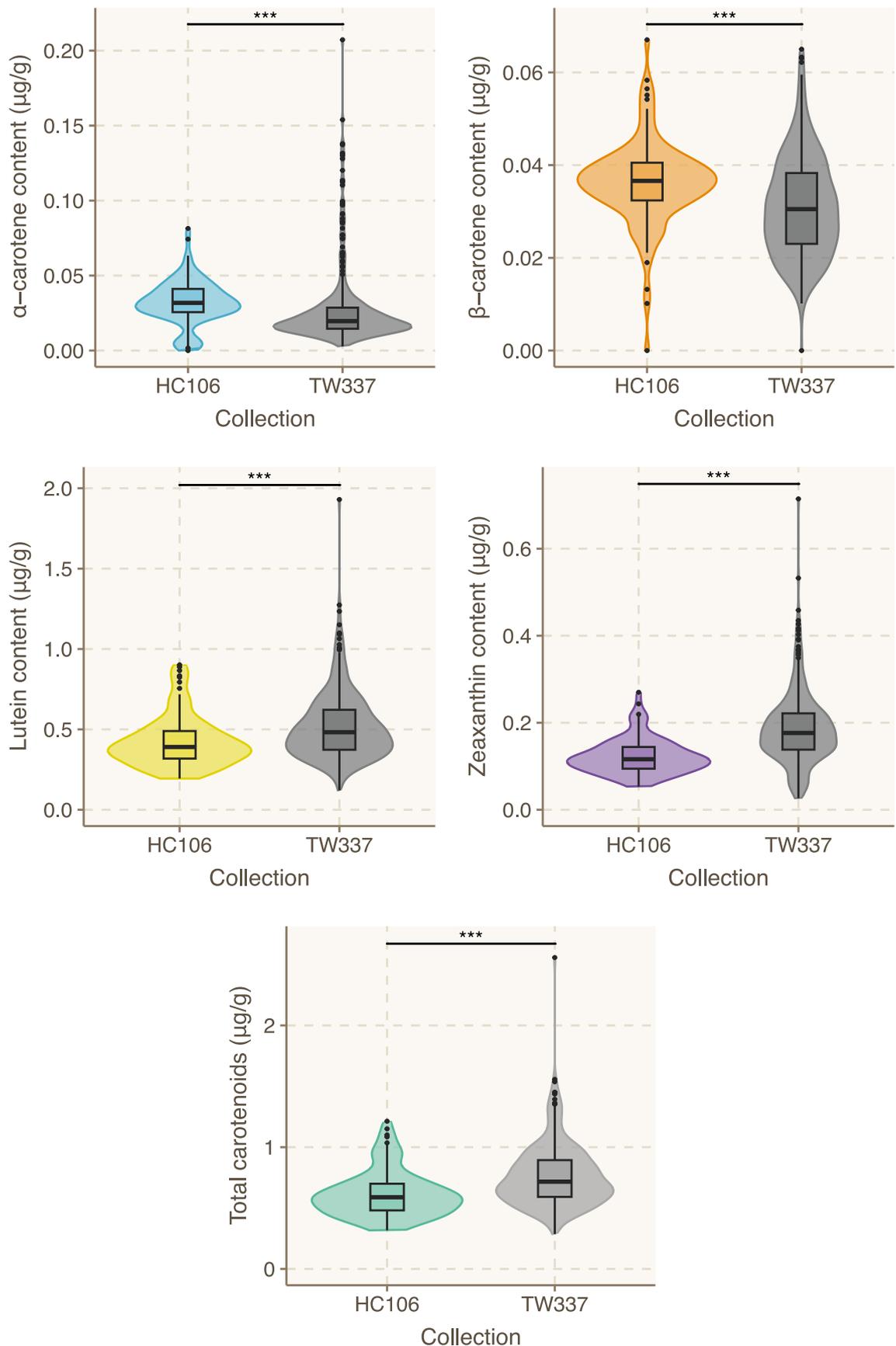
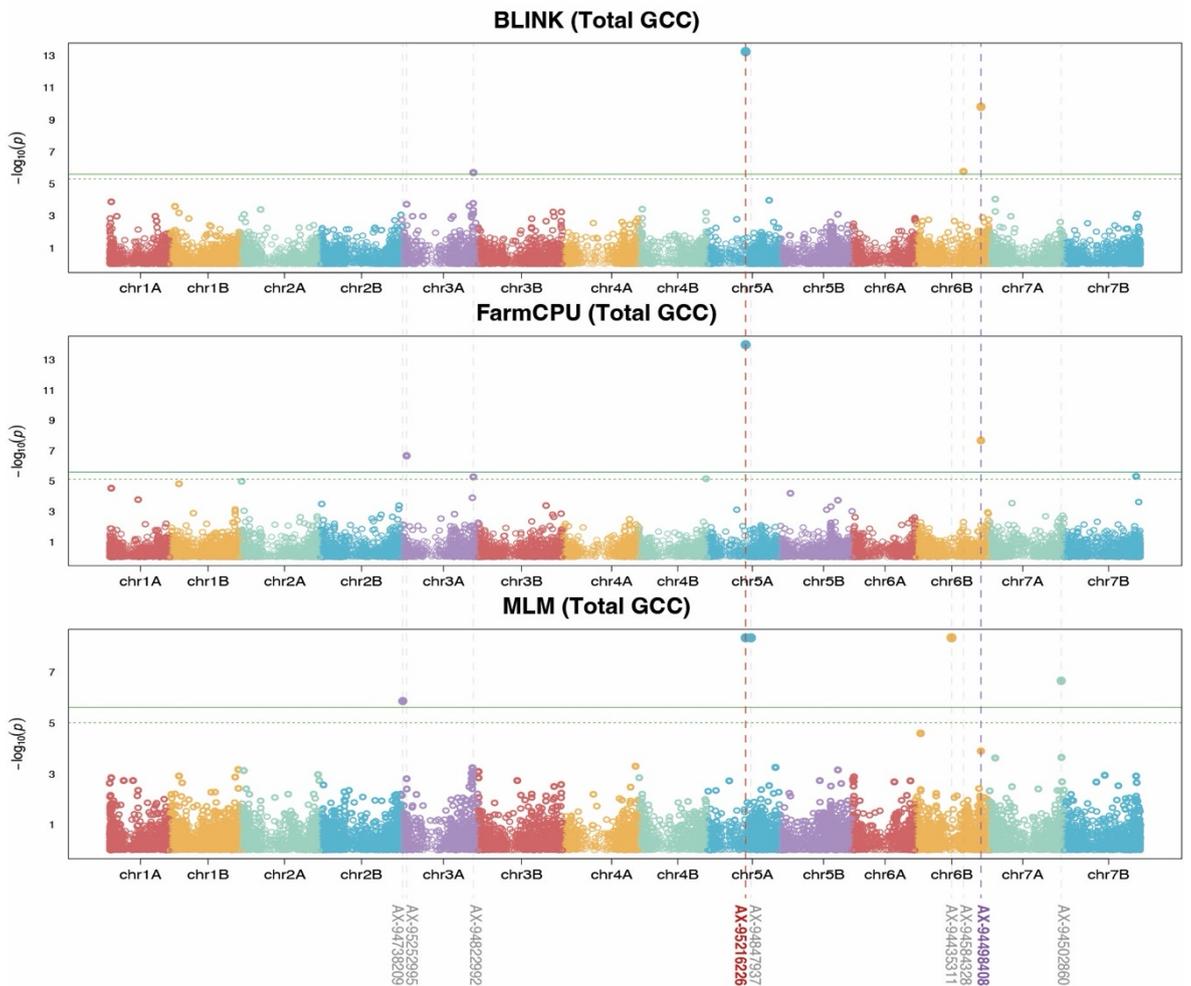
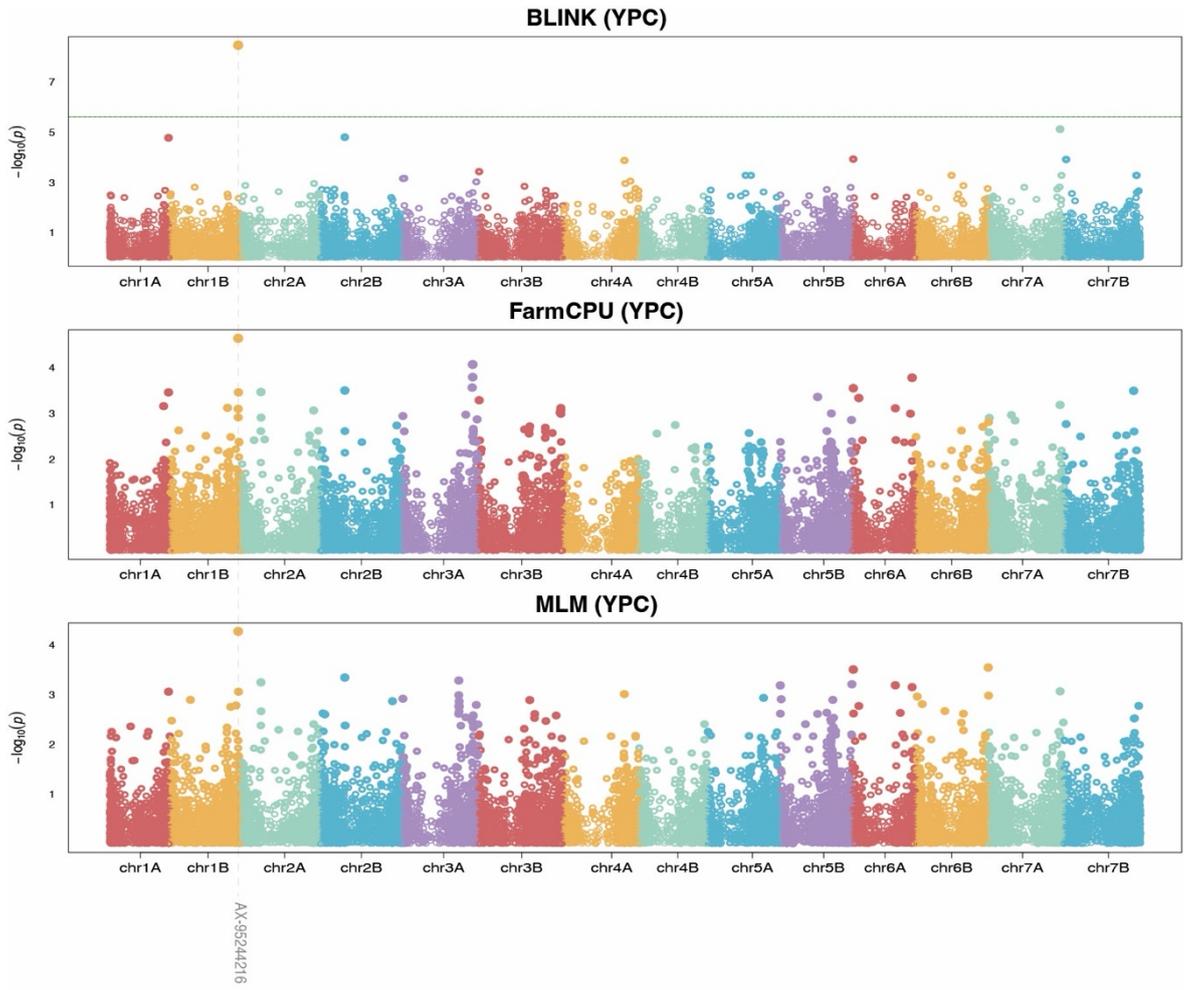


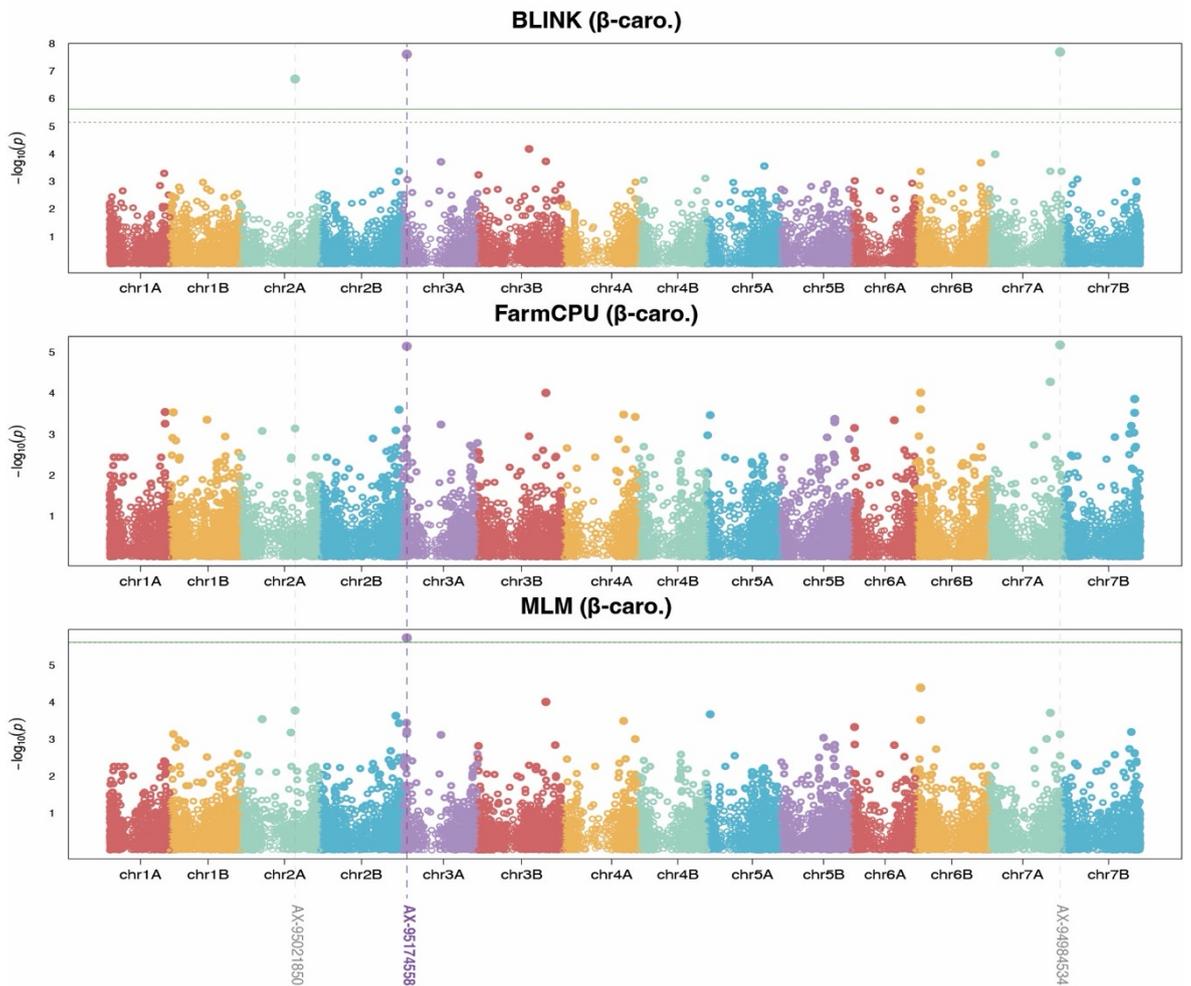
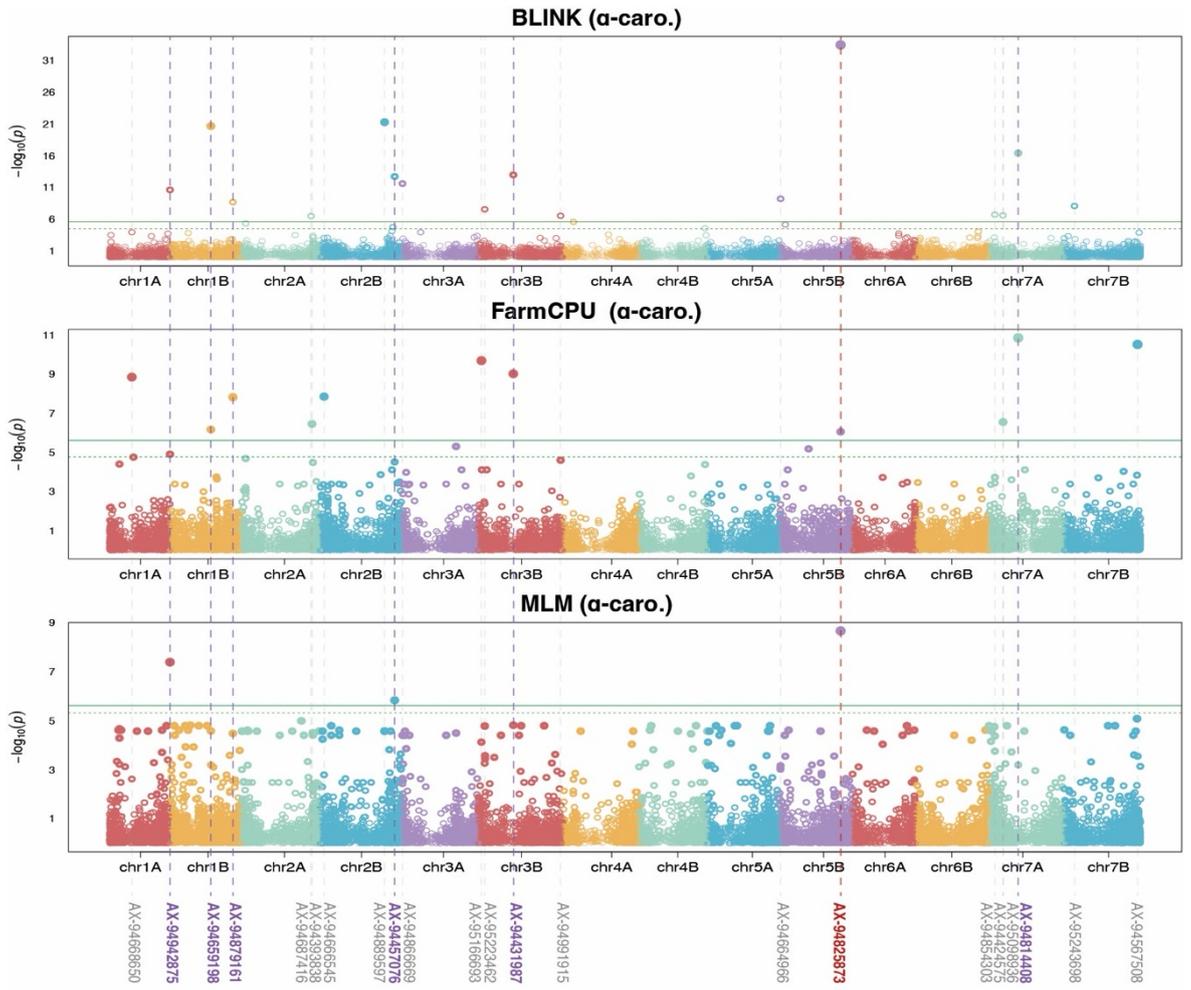
Figure 3.9 Comparison of grain carotenoids in the 106-accession Watkins hexaploid core collection and the 337-accession Watkins tetraploid collection measured using HPLC. The Watkins tetraploid collection results are of a single replicate, and the Watkins hexaploid core collection results are three replicates. HC106=Watkins hexaploid core collection, TW337=Watkins tetraploid collection, NS.=non-significant, *=p<0.05, **=p<0.01, ***=p<0.001.

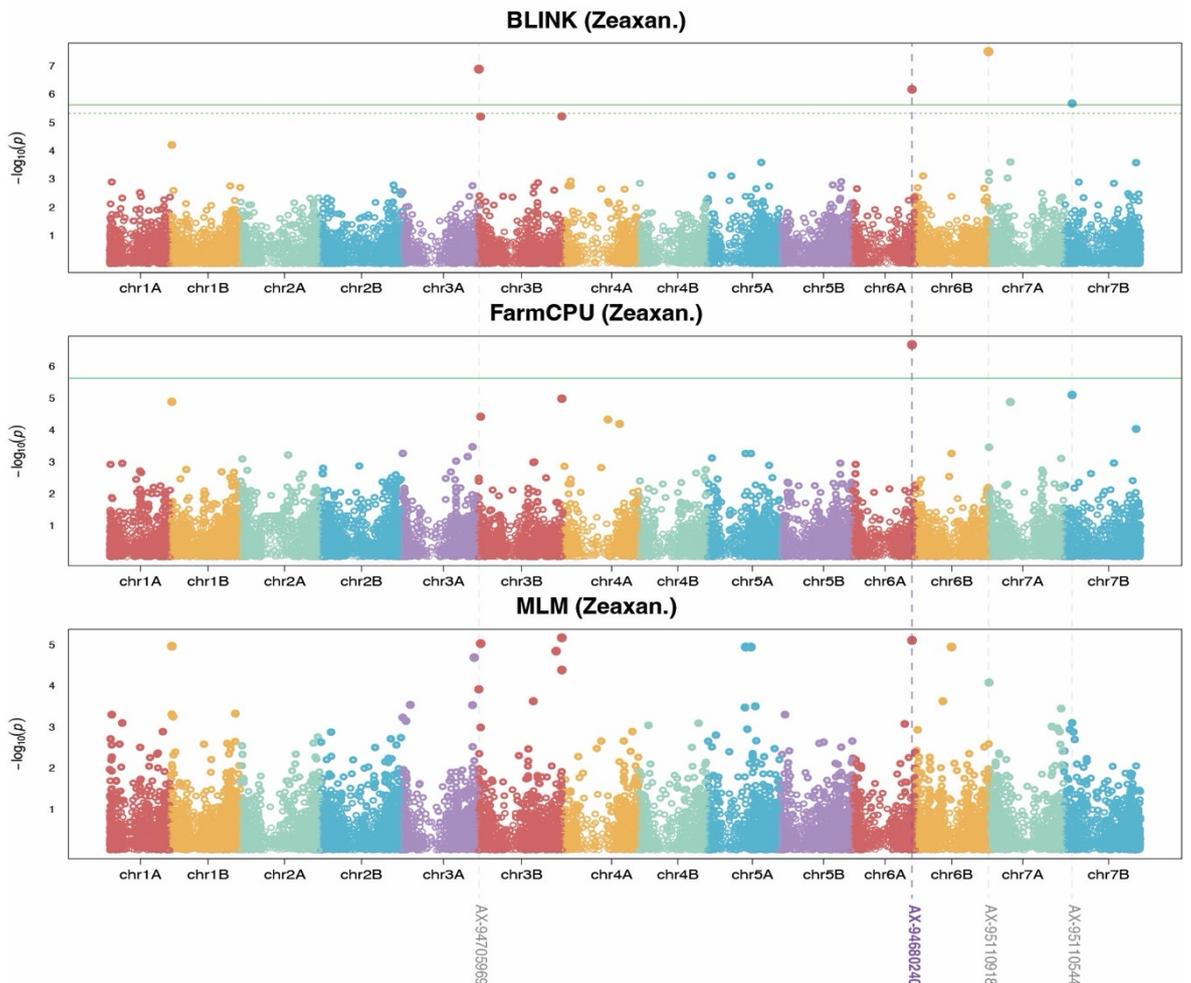
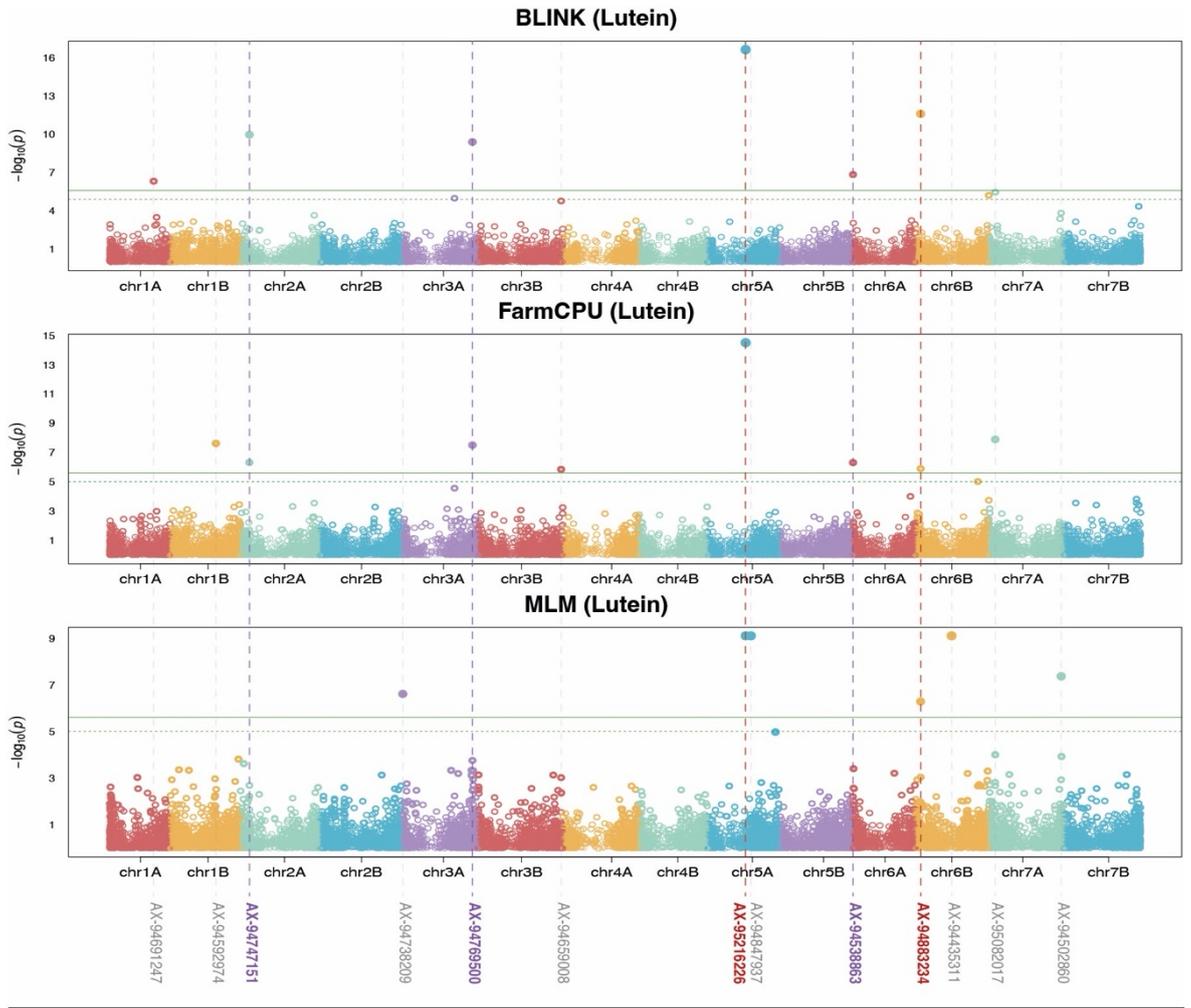
To assess the variability between the three replicates of the Watkins hexaploid core collection, I performed a two-way ANOVA considering both the accession and replicates as factors. The Tukey Honest Significant Difference test results for pairwise comparisons are found in Appendix Material 10 with significance letters above each plot. A table of pairwise comparisons are found in Appendix Material 11. This revealed significant differences between the replicates for β -carotene ($p < 0.001$, $F(2,210) = 27.849$, two-way ANOVA) and zeaxanthin ($p < 0.001$, $F(2,210) = 43.105$, two-way ANOVA). No significant differences were found between the replicates lutein content ($p = 0.413$, $F(2,210) = 0.889$, two-way ANOVA) and α -carotene content ($p = 0.665$, $F(2,210) = 0.408$, two-way ANOVA), strengthening the validity of the results and the overall approach taken in this large scale HPLC field screen.

3.2.3 Genome-wide association study of grain carotenoid content using the Watkins tetraploid collection reveals marker-trait associations and quantitative trait loci

To identify MTAs in the Watkins tetraploid collection with the measurements of YPC, α -carotene, β -carotene, lutein, zeaxanthin and total GCC, I performed a GWAS using the 35K Axiom[®] Wheat Breeder's Array genotypic data of the Watkins tetraploid collection. For this, I used the models Mixed Linear Model (MLM), Fixed and random model Circulating Probability Unification (FarmCPU) and Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK). A total of 47 significant MTAs for these five traits were identified by at least one of these models on all 14 chromosomes, and these can be found in Figure 3.10 and Appendix Material 12. Of these, 1, 22, 3, 13, 4 and 9 MTAs were identified for YPC, α -carotene, β -carotene, lutein, zeaxanthin and total GCC, respectively. To reinforce the reliability of these findings, I filtered MTAs for those identified by more than one model, which identified 15 unique MTAs (AX-95216226 was significant for total GCC and lutein content), which are found in Table 3.4. Nine of these markers had a low minor allele frequency (MAF) of below 0.05, indicating that the less common allele occurs in less than 5% of the collection, and some minor alleles appeared in only one accession (AX-95216226 and AX-94431987). Although low MAF markers are more susceptible to false positives, they are also important for identifying rare, impactful genetic variations.







On the previous page:

Figure 3.10 Manhattan plots for GWAS using 35K Axiom® Breeder’s Array genotypic data associated with YPC, α -carotene, β -carotene, lutein, zeaxanthin and total GCC diversity. Models used for GWAS were BLINK, FarmCPU and MLM. The horizontal green line in each plot shows the significance level; markers above this line are significant. Significant markers are indicated by dashed vertical lines coloured based on the number of models identifying them as significant: grey=1, purple=2 and red=3.

Table 3.4 Significant carotenoid-associated marker-trait associations identified by a GWAS using 35K Axiom® Breeder’s Array genotypic data. MTAs were filtered for those identified as significant by two models. Chr.=chromosome, Pos.=Svevo v1 genomic position, MAF=minor allele frequency, GCC=grain carotenoid content, caro.=carotene, Zeaxan.=zeaxanthin.

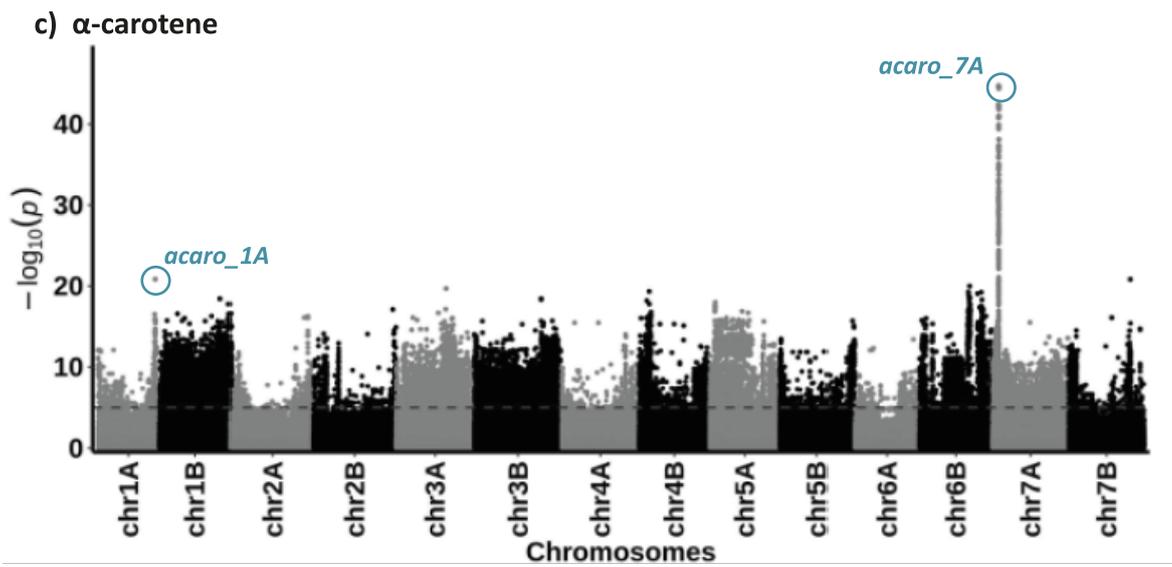
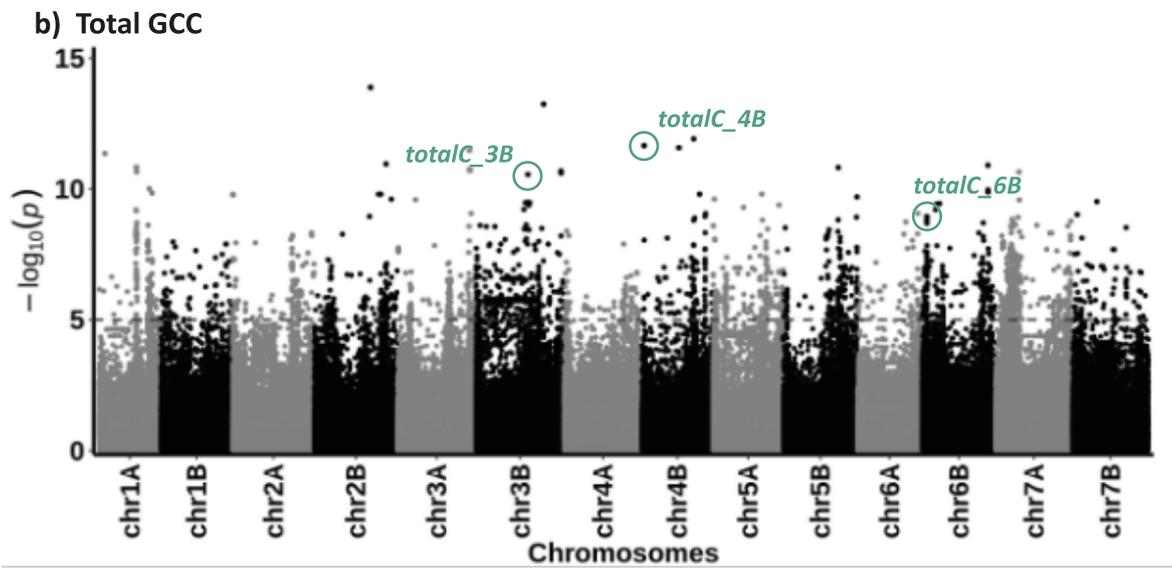
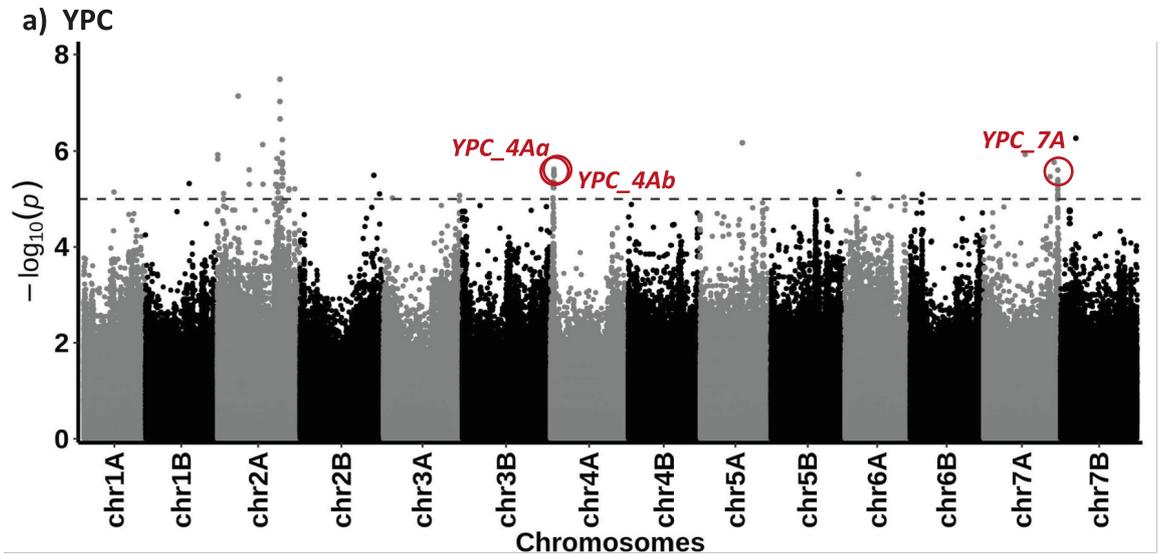
Carotenoid	Marker ID	Marker No.	Model	Chr.	Pos.	P-value	MAF
Total GCC	AX-95216226	12553	BLINK	chr5A	368,652,119	5.31E-14	0.002
Total GCC	AX-95216226	12553	FarmCPU	chr5A	368,652,119	9.84E-15	0.002
Total GCC	AX-95216226	12553	MLM	chr5A	368,652,119	4.39E-09	0.002
Total GCC	AX-94498408	17580	BLINK	chr6B	619,983,512	1.48E-10	0.081
Total GCC	AX-94498408	17580	FarmCPU	chr6B	619,983,512	2.03E-08	0.081
α -caro.	AX-94825873	14828	BLINK	chr5B	594,730,553	3.19E-34	0.003
α -caro.	AX-94825873	14828	FarmCPU	chr5B	594,730,553	8.62E-07	0.003
α -caro.	AX-94825873	14828	MLM	chr5B	594,730,553	2.14E-09	0.003
α -caro.	AX-94942875	1383	BLINK	chr1A	583,460,542	2.25E-11	0.459
α -caro.	AX-94942875	1383	MLM	chr1A	583,460,542	4.03E-08	0.459
α -caro.	AX-94659198	2430	BLINK	chr1B	392,556,289	1.97E-21	0.029
α -caro.	AX-94659198	2430	FarmCPU	chr1B	392,556,289	6.69E-07	0.029
α -caro.	AX-94879161	3054	BLINK	chr1B	596,279,063	1.93E-09	0.003
α -caro.	AX-94879161	3054	FarmCPU	chr1B	596,279,063	1.49E-08	0.003
α -caro.	AX-94457076	6626	BLINK	chr2B	718,930,920	1.76E-13	0.005
α -caro.	AX-94457076	6626	MLM	chr2B	718,930,920	1.45E-06	0.005
α -caro.	AX-94431987	9017	BLINK	chr3B	351,371,824	9.93E-14	0.002
α -caro.	AX-94431987	9017	FarmCPU	chr3B	351,371,824	9.44E-10	0.002
α -caro.	AX-94814408	18747	BLINK	chr7A	284,328,223	3.60E-17	0.003
α -caro.	AX-94814408	18747	FarmCPU	chr7A	284,328,223	1.39E-11	0.003
β -caro.	AX-95174558	7259	BLINK	chr3A	45,600,017	2.45E-08	0.011
β -caro.	AX-95174558	7259	MLM	chr3A	45,600,017	1.80E-06	0.011
Lutein	AX-95216226	12553	BLINK	chr5A	368,652,119	2.09E-17	0.002
Lutein	AX-95216226	12553	FarmCPU	chr5A	368,652,119	3.02E-15	0.002
Lutein	AX-95216226	12553	MLM	chr5A	368,652,119	7.63E-10	0.002
Lutein	AX-94883234	16632	BLINK	chr6B	48,986,624	2.30E-12	0.033
Lutein	AX-94883234	16632	FarmCPU	chr6B	48,986,624	1.22E-06	0.033
Lutein	AX-94883234	16632	MLM	chr6B	48,986,624	5.02E-07	0.033
Lutein	AX-94747151	3795	BLINK	chr2A	87,160,652	9.94E-11	0.327
Lutein	AX-94747151	3795	FarmCPU	chr2A	87,160,652	4.54E-07	0.327
Lutein	AX-94769500	8124	BLINK	chr3A	683,586,078	3.72E-10	0.264
Lutein	AX-94769500	8124	FarmCPU	chr3A	683,586,078	2.98E-08	0.264
Lutein	AX-94538863	15314	BLINK	chr6A	4,920,980	1.34E-07	0.258
Lutein	AX-94538863	15314	FarmCPU	chr6A	4,920,980	4.70E-07	0.258
Zeaxan.	AX-94680240	16099	BLINK	chr6A	580,443,163	6.77E-07	0.495
Zeaxan.	AX-94680240	16099	FarmCPU	chr6A	580,443,163	2.10E-07	0.495

In addition to the 35K Axiom® Breeder’s Array genotypic data, the Watkins tetraploid collection was recently whole-genome re-sequenced by Professor Shifeng Cheng’s group at the Agricultural Genomics Institute at Shenzhen, and I had this data available to me through collaboration. This greatly enhanced the genomic resolution of the carotenoid content GWAS. Overall, using the same phenotypic data, we identified 14 significantly associated QTLs containing a total of 881 candidate genes. No significant QTLs were identified for lutein. Table 3.5 describes these associated QTLs and their position on the Svevo v1 *Triticum turgidum* reference genome sequence, and the Manhattan plots of this GWAS are found in Figure 3.11.

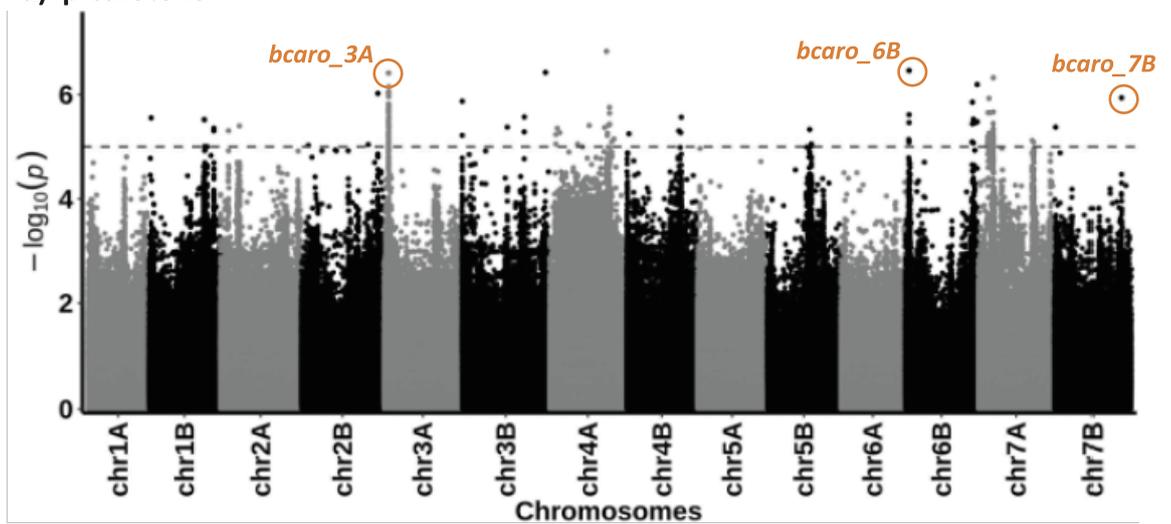
Table 3.5 Significant carotenoid-associated quantitative trait loci identified by a GWAS using the Watkins tetraploid collection genome sequence data. Chr.=chromosome, Pos.=genomic position, YPC=yellow pigment content, GCC=grain carotenoid content, caro.=carotene, zeaxan.=zeaxanthin, Prop. β - β .=Proportion of β - β branch carotenoids.

Traits	QTL ID	Most signif. SNP	P-value*	Chr.	Genomic region	QTL Size	Genes
YPC	YPC_4Aa	chr4A_27310277	4.36E-06	4A	24959991–32911449	4.24Mb	40
YPC	YPC_4Ab	chr4A_34775936	2.39E-06	4A	32911449–35825593	2.91Mb	28
YPC	YPC_7A	chr7A_704231793	2.50E-06	7A	701972109–705988124	4.02Mb	69
Total GCC	totalC_3B	chr3B_596546612	1.66E-08	3B	594946477–598105181	3.16Mb	23
Total GCC	totalC_4B [†]	chr4B_23124193	3.58E-10	4B	22185570–24954101	2.77Mb	44
Total GCC	totalC_6B	chr6B_48013108	5.24E-07	6B	43965733–50367188	6.40Mb	69
α -caro.	acaro_1A	chr1A_547450769	1.42E-21	1A	544281843–551446740	7.16Mb	100
α -caro.	acaro_7A [†]	chr7A_59368406	1.96E-45	7A	56308118–62009496	5.70Mb	89
β -caro.	bcaro_3A	chr3A_45161935	3.87E-07	3A	42187473–48934832	6.75Mb	68
β -caro.	bcaro_6B	chr6B_39413653	3.47E-07	6B	36269580–41776085	5.51Mb	71
β -caro.	bcaro_7B	chr7B_628477799	1.15E-06	7B	625953891–631164665	5.21Mb	33
Zeaxan.	zea_7A [†]	chr7A_29576066	3.52E-13	7A	22812830–33103534	10.29Mb	146
Prop. β - β .	propB_2Ba	chr2B_603634232	2.51E-08	2B	599992756–606994539	7.00Mb	50
Prop. β - β .	propB_2Bb	chr2B_662507210	2.10E-08	2B	659856421–664291240	4.43Mb	51

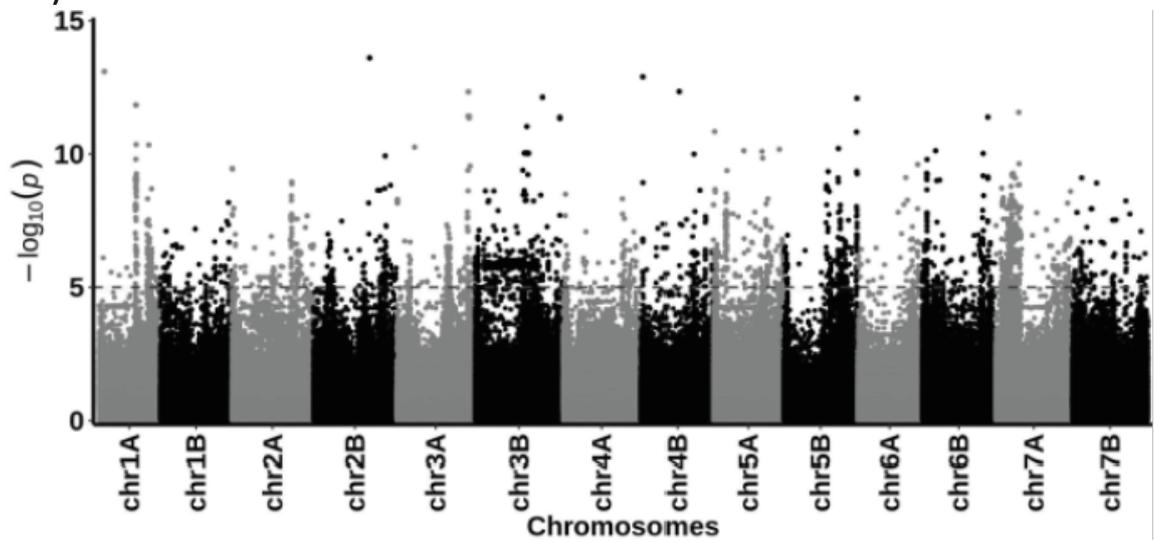
*P-value of the most significant SNP; [†]Overlaps with a QTL previously associated with carotenoid traits.



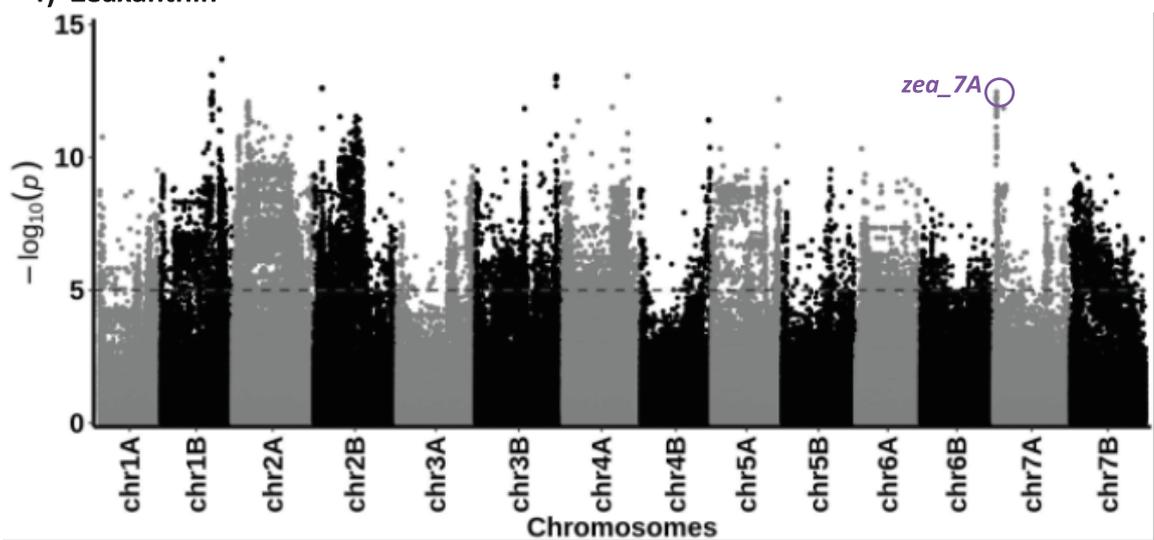
d) β -carotene



e) Lutein



f) Zeaxanthin



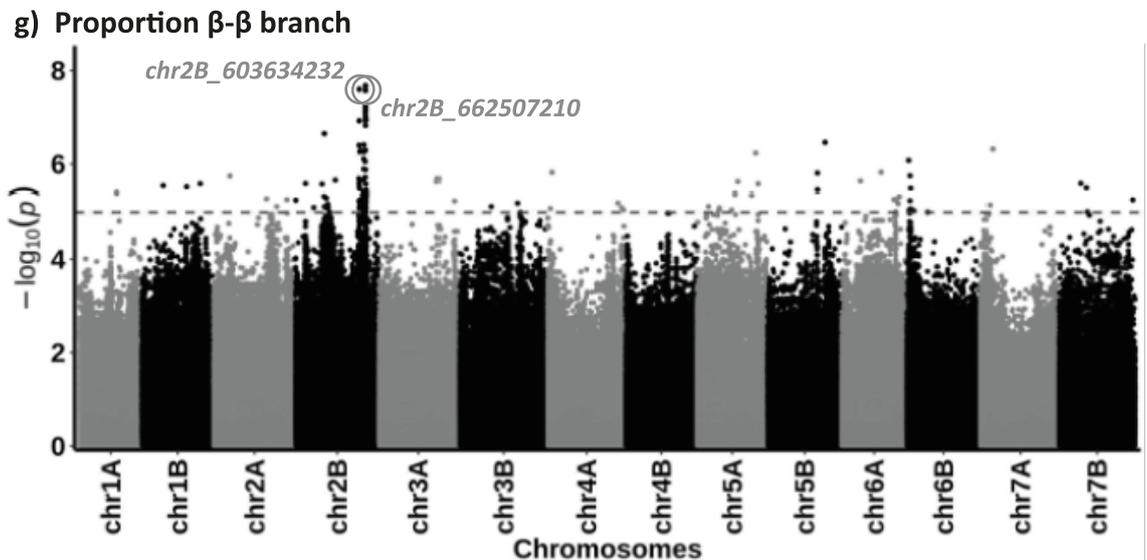


Figure 3.11 Manhattan plots for carotenoid content traits of a GWAS using the Watkins tetraploid collection genome sequence data. The Manhattans show the GWAS for (a) YPC, (b) total GCC, (c) α -carotene, (d) β -carotene, (e) lutein, (f) zeaxanthin, and (g) proportion of β - β branch carotenoids. A total of 14 significantly associated QTLs were identified, and these are labelled on the Manhattan plots.

To investigate whether any of these QTLs have been previously associated with traits related to carotenoid content (such as YPC or yellow index), I searched the QTL track of the Svevo genome browser, which provides the positions of known QTLs curated by the International Durum Wheat Genome Sequencing Consortium (Maccaferri et al. 2019; Yao et al. 2022). This identified four QTL regions previously associated with carotenoid content traits overlapping three QTLs discovered here (Table 3.5): a QTL for pasta yellowness (*QTL0835_PY-N'Diaye_et_al_2017*) (N'Diaye et al. 2017) overlapped the QTL *totalC_4B*; a QTL for yellow index (*QTL0073_SY-Colasuonno_et_al_2017*) (Colasuonno et al. 2017a) overlapped *acaro_7A*; two QTL for YPC (*QTL0995_7A-Colasuonno_et_al_2014*) and yellow index (*QTL0072_SY-Colasuonno_et_al_2017*) (Colasuonno et al. 2014, 2017a) overlapped *zea_7A*. Eleven of the QTLs identified here did not overlap with any described carotenoid-associated QTL on the QTL track of the Svevo genome browser and, to the best of my knowledge, represent novel findings.

To identify putative causal genes located within these significant QTLs, I first identified the *T. aestivum* orthologues of the *T. turgidum* candidate genes found within these QTLs; this filtered out 86 genes (9.76% of the total 881 candidate list) where no *T. aestivum* orthologue was found to the *T. turgidum* gene. I then submitted the *T. aestivum* orthologous genes to Knetminer with the search term 'carotenoid' (Hassani-Pak et al. 2021), which does not accept *T. turgidum* gene names as input. This gene discovery tool creates knowledge networks for each gene based on their biochemical function and orthologous genes in other

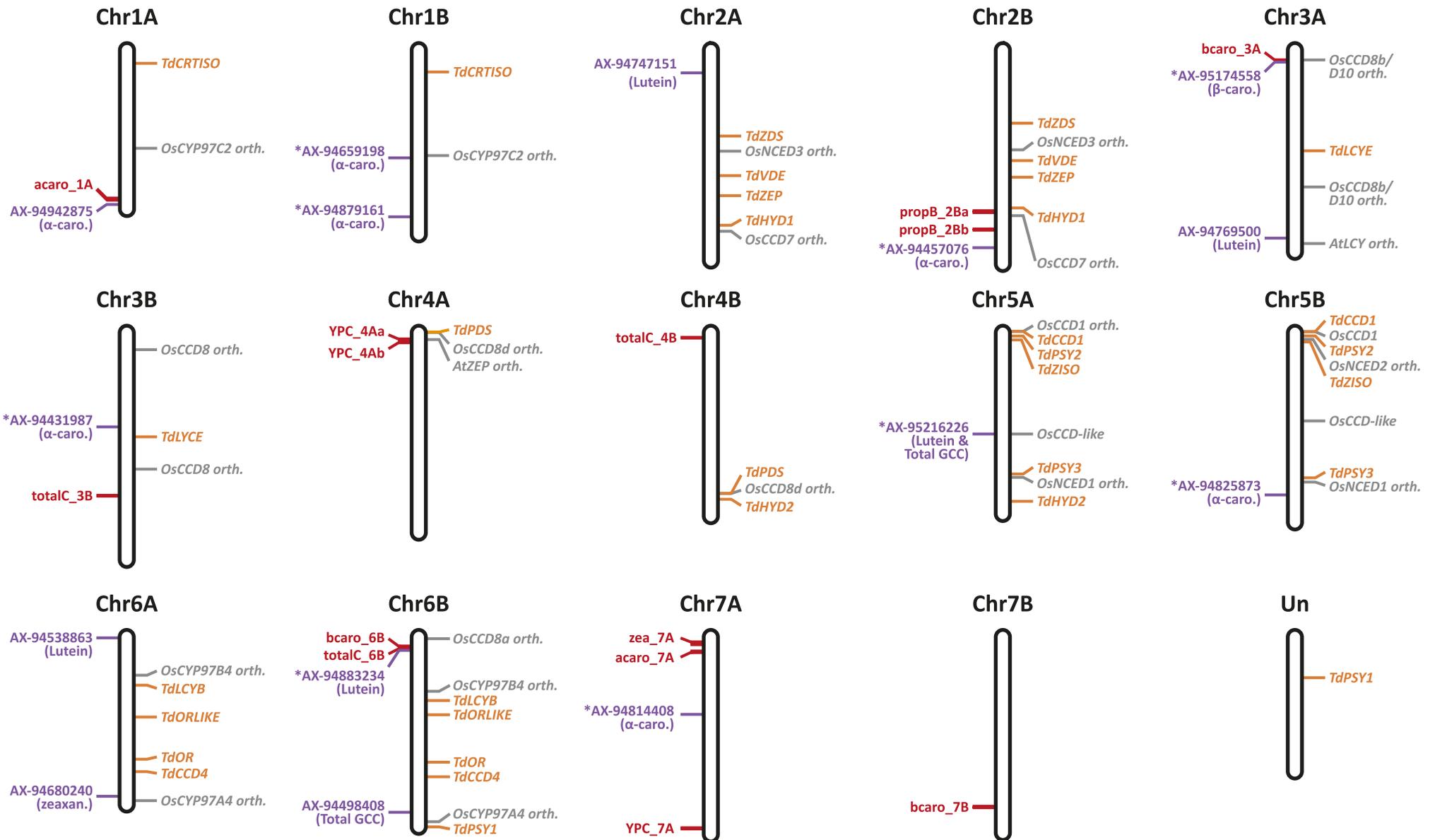
organisms, such as rice and *Arabidopsis thaliana* (*Arabidopsis*). This identified notable candidate genes in the QTL *bcaro_3A*, *YPC_4Aa* and *zea_7A* that are orthologous to carotenoid biosynthesis genes. Within *bcaro_3A*, *TRITD3Av1G020620* is orthologous (one-to-many) to the *Arabidopsis* enzyme *AtCCD8* (*CAROTENOID CLEAVAGE DIOXYGENASE 8*; AT4G32810), which is involved in the production of strigolactones from its precursor β -carotene (Figure 1.1) (Batra et al. 2019). Moreover, *AtCCD8* is believed to be the critical enzyme in strigolactones biosynthesis from β -carotene since its expression level directly determines the level of strigolactones synthesis (Guan et al. 2012). Within *YPC_4Aa*, there are six paralogous genes (*TRITD4Av1G011490*, *TRITD4Av1G011580*, *TRITD4Av1G011600*, *TRITD4Av1G011630*, *TRITD4Av1G012700* and *TRITD4Av1G012720*) that are orthologous (one-to-many) to *AtZEP* (*ZEAXANTHIN EPOXIDASE*; AT5G67030) which catalyses the conversion of zeaxanthin to violaxanthin (Figure 1.1) (Nisar et al. 2015). *AtZEP* regulates carotenoid degradation in maturing *Arabidopsis* seeds, with ZEP-mediated epoxidation targeting carotenoids for degradation by CCD enzymes (Gonzalez-Jorge et al. 2016). Within *zea_7A*, *TRITD7Av1G013490* is orthologous (one-to-many) to *AtDXR* (*DXP REDUCTOISOMERASE*; AT5G62790), which catalyses the first committed step of the MEP pathway that supplies the isoprene building-blocks of carotenoids (Figure 1.1) (Nisar et al. 2015). In *Arabidopsis*, *AtDXR* is a rate-determining enzyme of the MEP pathway whose overexpression increases carotenoid production (Carretero-Paulet et al. 2006).

In addition to these candidate genes that are orthologous to carotenoid biosynthesis genes, Knetminer also highlighted three transcription factors within *YPC_4Ab*, *acaro_1A* and *acaro_7A*, which may play a role in carotenoid biosynthesis. Within *YPC_4Ab*, *TRITD4Av1G014980* is orthologous to the *Arabidopsis* homeodomain leucine zipper protein-encoding genes *AtHB21*, *AtHB40* and *AtHB53*. These transcription factors enhance the expression of *9-CIS-EPOXICAROTENOID DIOXYGENASE 3* (*NCED3*), a key rate-limiting enzyme in the production of abscisic acid (ABA) from carotenoids (González-Grandío et al. 2017). Within *acaro_1A* and *acaro_7A*, *TRITD1Av1G209580* and *TRITD7Av1G027980* are orthologues to *TaWRKY46* (*TRAESCS1A02G401800*) and *TaWRKY61* (*TRAESCS7D02G092400*), respectively. GENIE3 predicts these two WRKY transcription factors to regulate *TaAO1* (*ALDEHYDE OXIDASE*), which catalyses the final steps of carotenoid catabolism producing ABA (Colasuonno et al. 2017b). For the other eight QTLs associated with these carotenoid traits, no candidate genes within these were highlighted by Knetminer.

To investigate whether these associated MTAs and QTLs were close to carotenoid biosynthesis genes, I manually identified *T. turgidum* orthologues of known carotenoid biosynthesis genes from *T. aestivum*, rice and *Arabidopsis*. From this, I identified 62 *T. turgidum* genes, which are orthologous to 29 rice and *Arabidopsis* genes involved in carotenoid biosynthesis. A table of these, their respective orthologues and their genomic locations is found in Appendix Material 13. A chromosome map showing the locations of these 62 *T. turgidum* genes alongside the MTAs and QTLs associated with the carotenoid content traits is found in Figure 3.12. Of the 14 QTLs and 15 MTAs identified from the carotenoid content traits, only 3 MTAs and QTLs overlapped or were located very close together: *acarо_1A* and AX-94942875 (α -*carо.*); *bcarо_3A* and AX-95174558 (β -*carо.*); *totalC_6B* and AX-94883234 (*Lutein*). This shows good evidence for these regions being associated with these carotenoid content traits; however, 11 QTLs and 12 MTAs did not overlap between these two GWAS resolutions.

On the following page:

Figure 3.12 Chromosomal distribution of carotenoid biosynthesis genes in tetraploid wheat, and the MTAs and QTLs associated with carotenoid content traits of the Watkins tetraploid collection. On the left of each chromosome are the MTAs (purple) from the 35K Axiom® Breeder’s Array and QTLs (red) found using a GWAS with whole-genome resequenced data. Asterisks next to the MTA name indicate a minor allele frequency of below 0.05. On the right of each chromosome are *T. turgidum* genes involved with carotenoid biosynthesis located using the Svevo v1 reference genome sequence. Genes in orange are clear one-to-one orthologues of carotenoid biosynthesis genes, and those in grey are one-to-many orthologues of their described genes. Chr=chromosome, Un=undefined chromosome, orth.=orthologue.



3.3 Discussion

3.3.1 Large variation in grain carotenoid content exists within the Watkins tetraploid collection

I adapted the high-throughput YPC method based on the industry standard AACC 14-50 YPC method for screening a large collection of wheat accessions quickly using a plate reader. YPC measurements can vary considerably between studies, complicating direct comparisons between these. For example, Digesù and colleagues (2009) reported YPC values for seven durum cultivars that were approximately twice as high as those documented by Fratianni and colleagues (2005) despite both studies using the same AACC 14-50 YPC method and measuring the same cultivars. Despite this, the YPC of the Watkins tetraploid collection and Miradoux control plots obtained here (ranging from 2.09 $\mu\text{g/g}$ to 8.44 $\mu\text{g/g}$) align well with those found for durum wheat in the literature (ranging from 1.28 $\mu\text{g/g}$ to 12.30 $\mu\text{g/g}$) (Fratianni et al. 2005; Digesù et al. 2009; Beleggia et al. 2010; Fayaz et al. 2013; Li et al. 2015). This alignment suggests that the adapted method provides comparable results to those from previous studies. Moreover, significantly associated QTLs were found using the YPC measurements from this adapted high-throughput method (Figure 3.11a), suggesting its effectiveness in screening panels for association studies.

As with YPC, absolute carotenoid values from HPLC measurements can vary considerably between studies due to differences in the extraction method (Digesù et al. 2009). Abdel-Aal and colleagues (2007) measured the lutein content with HPLC of the same einkorn flour sample using four different extraction solvents and found differing values for each (ranging from 2.16–7.75 $\mu\text{g/g}$). This variability is further highlighted by the varying grain lutein content of the durum cultivar Simeto measured across four separate studies (0.7 $\mu\text{g/g}$ to 3.72 $\mu\text{g/g}$) (Hidalgo et al. 2006; Atienza et al. 2007; Digesù et al. 2009; Mellado-Ortega and Hornero-Méndez 2016), each using a different extraction solvent. In this thesis, acetone was used as the extraction solvent, and the range of measurements for lutein content within the Watkins tetraploid collection (0.121–1.93 $\mu\text{g/g}$, average 0.521 $\mu\text{g/g}$) were similar to measurements of durum wheat (0.6–1.4 $\mu\text{g/g}$) in studies using an acetone solvent extraction (Atienza et al. 2007; Mellado-Ortega and Hornero-Méndez 2016). Similarly, Requena-Ramírez and colleagues (2021) also used an acetone solvent extraction and reported comparable values of lutein, zeaxanthin, α -carotene and β -carotene to those measured here in the Watkins tetraploid collection. Conversely, Giambanelli and colleagues

(2013) found a substantially higher total GCC in Miradoux using a chloroform-methanol extraction (6.78–8.38 $\mu\text{g/g}$) compared to the range found here in the Miradoux control plots (0.939–1.542 $\mu\text{g/g}$); but, the proportion of individual carotenoids like lutein, zeaxanthin and PVA carotenoids closely mirrored those measured here. Thus, despite methodological differences, our HPLC carotenoid measurements are largely consistent with existing literature.

Despite this, there was considerable variability in the YPC measurements of the Miradoux control plots (YPC coefficient of variation was 10.1% for Miradoux control plots compared to 20.2% for the Watkins tetraploid collection; Table 3.1). One possible explanation for this high variation could be environmental effects. Leaf carotenoid biosynthesis, storage and degradation are influenced by changes in light, CO_2 concentration, temperature, drought and soil nutrients (Dhami and Cazzonelli 2020). However, these effects are mostly due to changes in photosynthesis and chloroplast development, which are not applicable in non-photosynthetic grain tissue. Developmental signals rather than environmental effects primarily determine carotenoid content in flowers and fruit, and grain tissues likely operate similarly (Hermanns et al. 2020). Carotenoid production is upregulated in response to high-intensity light, but since these plots were located close to one another in the same field, there are unlikely to be significant differences in light intensity among them (Jahns and Holzwarth 2012; Dhami and Cazzonelli 2020).

Another possible reason for the large variation could be the quantification method. An A11 IKA lab mill was used to mill the samples for YPC analysis. This mill operates by pushing a cutting blade into a grain sample in a grinding chamber, allowing for fast milling; however, obtaining a consistent flour granularity is difficult due to variations in how long the blade remains in the grinding chamber. Given the low quantities of flour used (300 mg), inconsistent granularity could affect the extraction rate of carotenoids between samples. If a flour sample was coarser in consistency than another sample of the same flour, the coarser sample would have a slower extraction rate because the larger particle size results in a lower surface area for extraction. Consequently, fewer carotenoids would be extracted, leading to a lower YPC for the coarser flour compared to the finer sample. A RETSCH Cyclone Mill Twister was used for HPLC analysis to mill grain samples instead of the A11 IKA lab mill. This mill feeds the grains at a constant rate and uses a sieve to produce more consistent flour granularity. For the HPLC measurements of the Miradoux control plots, the coefficient

of variation is artificially inflated because I chose 16 accessions representing the highest, lowest and middle YPC values for measuring with HPLC.

Consistent with previous HPLC studies, lutein was the primary carotenoid in the Watkins tetraploid collection and Watkins hexaploid core collection, followed by zeaxanthin and the PVA carotenoids (Giambanelli et al. 2013; Mellado-Ortega and Hornero-Méndez 2015; Requena-Ramírez et al. 2021; Suriano et al. 2023). Interestingly, in four tetraploid accessions (WAT1180219, WAT1180240, WAT1180241 and WAT1180295), I found a higher amount of zeaxanthin than lutein. A possible explanation for this rare carotenoid ratio could be a low activity of LYCOPENE ϵ -CYCLASE (LCYE), which is responsible for moving flux into the β - ϵ branch of the carotenoid biosynthesis pathway (Figure 1.1) and whose knockout in wheat increases flux into β - β branch carotenoids such as β -carotene and zeaxanthin (Richaud et al. 2018; Sestili et al. 2019). A previous study by Requena-Ramírez and colleagues (2021) screened a Spanish collection of durum wheat landraces for α -carotene, β -carotene, lutein, and zeaxanthin content using HPLC. However, the analysis of the Watkins tetraploid collection in this chapter represents the first investigation of a global durum wheat landrace collection using HPLC. Furthermore, over twice the number of accessions were examined here (337 Watkins tetraploid landraces) compared to the 158 landraces analysed in their research. This research provides the first global analysis and largest dataset of carotenoid diversity within durum wheat landraces, offering novel insights into carotenoid diversity within wheat germplasm.

Several unknown peaks were identified through HPLC (Figure 3.5 and Appendix Material 6), which are likely to be isomers of lutein, zeaxanthin, α -carotene and β -carotene. Previous studies have documented the identification and quantification of such isomers in tetraploid wheat using HPLC (Panfili et al. 2004; Hidalgo et al. 2006; Abdel-Aal et al. 2007; Giambanelli et al. 2013; Requena-Ramírez et al. 2021). Given their lower degradation rates during processing, lutein isomers are considered beneficial for wheat carotenoid biofortification (Paznocht et al. 2019); therefore, future research should quantify their abundance within the Watkins tetraploid collection. Additionally, some of the peaks might represent carotenoids not included as references, such as β -cryptoxanthin, a common component in tetraploid wheat flour (Digesù et al. 2009). Another possibility is these peaks could be from esters of lutein (lutein monoesters or lutein diesters), known to increase carotenoid stability and promote the sequestration and accumulation of carotenoids (Atienza et al. 2007; Watkins et al. 2019). Yet, these tend to be very rare within tetraploid wheat (Ziegler et al.

2015), as evidenced by only four accessions containing lutein monoesters and diesters in a diverse collection of 156 Spanish durum wheat landraces (Requena-Ramírez et al. 2021).

Comparing retention times (RT) between papers using similar HPLC methods might help elucidate the identity of some of these unknown peaks. Abdel-Aal and colleagues (2007) identified minor peaks between RT 3.5–7.5 minutes, surrounding the main lutein and zeaxanthin peaks, as lutein and zeaxanthin isomers (15-cis-lutein, 13-cis-lutein, 13'-cis-lutein, 9-cis-lutein, 9'-cis-lutein, 9-cis-zeaxanthin) within durum wheat flour. They ran this on the same C₃₀ HPLC column as used in this thesis. Consequently, the minor peaks between RT 3.5–7 minutes in my chromatograms (Figure 3.5) could be these isomers. Additionally, the unknown peak A (Figure 3.5a–b) has an RT of around 10 minutes, close to the RT of β -cryptoxanthin (9.5–10 minutes) found by Abdel-Aal and colleagues (2007). This suggests that peak A may indeed be β -cryptoxanthin. To confirm the identities of these unknown carotenoids, further analysis could compare the UV/vis spectra of these peaks to those of known carotenoids or employ liquid chromatography-mass spectrometry to identify them.

I aimed to screen the broadest possible diversity set in an attempt to reveal accessions with extreme phenotypes (such as WAT1180004 with a very high total GCC and accessions with enhanced PVA carotenoids). A limitation of this approach was the number of repeats that could be performed. Hence, the YPC and HPLC measurements of the 356 tetraploid accessions were not replicated. The three HPLC replicates of the Watkins hexaploid core collection and the high-carotenoid Watkins tetraploid accession WAT1180004 (Appendix Material 10 and Table 3.2) showed that zeaxanthin or low abundance carotenoids like β -carotene can differ between measurements made on the same flour. But for other carotenoids, there was consistency between these replicates. Future work measuring more HPLC replicates of the Watkins tetraploid collection would improve the reliability of this data. This work is currently being carried out (March 2024).

3.3.2 Higher YPC and carotenoid content in the modern cultivar Miradoux compared to the Watkins landraces may reflect historic durum wheat breeding targets

The control plots of the modern cultivar Miradoux exhibited a mean YPC of 6.651 $\mu\text{g/g}$, greater than the highest YPC observed within the Watkins tetraploid collection (6.613 $\mu\text{g/g}$; Figure 3.1). Previous studies have also found tetraploid landraces from Iran and the Mediterranean to have lower yellow pigment within flour than modern durum cultivars from the same region (Nazco et al. 2012; Fayaz et al. 2013). Additionally, older cultivars have been found to have lower YPC than modern cultivars of Italian, Spanish, Canadian and Moroccan durum wheat (Digesù et al. 2009; Clarke et al. 2010; Subira et al. 2014; Boussakouran et al. 2022). In these studies, the greatest increase in YPC was observed in cultivars released after 1990 compared to those released earlier. This period marks a shift in breeding objectives, as the yellow colour of durum wheat flour became a valued quality trait for pasta production only in the last three decades (Digesù et al. 2009; Clarke et al. 2010). Moreover, the high heritability of YPC facilitated the effort of breeding programmes to enhance this trait (Ficco et al. 2014). Since Miradoux was released in 2007 (CIMMYT 2023), the finding of greater YPC in Miradoux compared with the Watkins tetraploid collection is expected. Digesù and colleagues (2009) demonstrated that an increase in carotenoid content accompanied the increase in YPC content during durum breeding from pre-1971 to 2008. I also found that the Watkins hexaploid core collection and Watkins tetraploid collection had relatively similar carotenoid content (Figure 3.9 and Table 3.3) compared to the content of modern cultivars of bread and durum wheat found in past studies, which have a greater difference in carotenoid content (1.35–1.79 $\mu\text{g/g}$ average lutein content in bread wheat compared to 3.15–3.26 $\mu\text{g/g}$ in durum wheat) (Hidalgo et al. 2006; Ziegler et al. 2015). Again, this likely reflects the lack of selection for yellower durum wheat flour before the 1990s, with bread and durum wheat landraces historically having a more similar GCC.

Interestingly, despite Miradoux having a higher average YPC than any Watkins tetraploid accession, several Watkins tetraploid accessions had a greater total GCC than the average Miradoux value, with WAT1180004 having nearly double (2.557 $\mu\text{g/g}$) that of the average total GCC of the Miradoux control plots (1.274 $\mu\text{g/g}$). One possibility for this is that the selection for high pigmentation (using YPC or yellow index) in durum breeding has selected for carotenoids and other non-carotenoid pigments that increase the YPC but not the total

GCC. There may be a high amount of these pigments within Miradoux compared to the Watkins tetraploid collection, boosting the YPC of Miradoux. But this contrasts previous studies that have found the YPC of flour to correlate very well with total GCC measured using HPLC (Fратиanni et al. 2005; Abdel-Aal et al. 2007; Digesù et al. 2009), where correlations as high as $r=0.89$ have been reported. Conversely, only a moderate positive correlation was found here between YPC and total GCC of $r=0.46$. This might reflect possible differences in the high-throughput YPC method used here or the lack of repeats in YPC analysis. This discrepancy highlights HPLC's advantage over YPC for a more detailed understanding of GCC, especially when investigating the health benefits of carotenoids in wheat where absolute values of carotenoid compounds are needed.

3.3.3 Utilising the Watkins global landrace collection for carotenoid biofortification

The preference for brown and wholemeal bread over white bread has been on the rise in the UK for the past 45 years, indicating a shift in consumer habits (Lockyer and Spiro 2020). Despite this trend, the carotenoid content in modern bread wheat cultivars remains low, a relic of historical consumer preferences for white bread (Hidalgo et al. 2006; Ziegler et al. 2015). Given the health benefits associated with carotenoids and the rising demand for browner bread, there is a clear incentive to increase their levels in bread wheat. In this study, both bread and durum wheat from the Watkins global landrace collection were found to have relatively similar GCC (Figure 3.9 and Table 3.3). Over the last three decades, durum wheat breeding has successfully enhanced its carotenoid content (Digesù et al. 2009), suggesting that insights into the genetic variation between modern durum cultivars and the Watkins tetraploid collection could help identify important alleles for breeding bread wheat with higher GCC. Previously, *PHYTOENE SYNTHASE 1 (PSY1)* alleles from durum wheat were introduced into bread wheat through interspecific crosses, which increased the carotenoid content of bread wheat (Requena-Ramírez et al. 2023). Likewise, other important genetic determinants for carotenoid content within durum wheat could be introduced into bread wheat. Creating biparental mapping populations of high carotenoid modern cultivars crossed with low carotenoid Watkins tetraploid accessions could help identify these genetic determinants.

There are also high GCC bread wheat accessions already present within the Watkins hexaploid core collection, such as WAT1190149 (total GCC=1.213 $\mu\text{g/g}$), which could provide useful genetic material for improving the carotenoid content of bread wheat

cultivars. Similarly, for durum wheat breeding, WAT1180004 (total GCC=2.557 $\mu\text{g/g}$) could provide useful genetic material for breeding high carotenoid durum wheat or bread wheat cultivars, especially considering it had around double the total GCC compared to the average of the Miradoux control plots (1.274 $\mu\text{g/g}$). Miradoux is described as having a high YPC by its breeding company, and Giambanelli and colleagues (2013) found Miradoux to have the highest total GCC of the three modern cultivars they analysed. They also found Miradoux to have comparable total carotenoid levels to einkorn wheat, which has very high carotenoid content. Again, this emphasises that WAT1180004 has especially high carotenoid content and is a useful breeding material for wheat carotenoid biofortification. WAT1180004 is currently being crossed with Miradoux for high-carotenoid durum wheat pre-breeding.

Screening for individual carotenoid compounds using HPLC, as was done here, can help identify accessions with a high content of desirable carotenoids for different biofortification objectives, given the varied health benefits each carotenoid offers. For example, a focus on increasing lutein and zeaxanthin would be desirable for enhancing macular carotenoid content, while an emphasis on increasing α -carotene and β -carotene would be desirable for enhancing PVA content. Furthermore, within the PVA carotenoids, focusing on increasing the content of β -carotene would be desirable because it has the highest conversion efficiency into vitamin A within the body. Within the Watkins tetraploid collection accession, WAT1180105 and WAT1180094 had the highest α -carotene and β -carotene content, respectively (Figure 3.3), and these could be used as breeding material for improving these PVA carotenoids in durum wheat. Additionally, accessions with a high proportion of β -carotene (WAT1180219; Appendix Material 8) or total PVA carotenoids (WAT1180105; Figure 3.7) could be crossed with an accession with a high total GCC (WAT1180004) to try to increase the carotenoid biosynthesis flux in a background that has a high proportion of desirable PVA carotenoids.

HPLC also allowed for measuring the proportion of carotenoids throughout the pathway within the Watkins tetraploid collection, which provides insights into the flux within the carotenoid biosynthesis pathway. Of specific interest for β -carotene biofortification is the proportion of β - β branch carotenoids. WAT1180219 had the highest proportion of β - β branch carotenoids within the Watkins tetraploid collection (Figure 3.7), and WAT1180241 had the highest zeaxanthin content (Figure 3.3). In these accessions, β -carotene could be elevated by suppressing genes that convert β -carotene into downstream carotenoids. For

example, *β-CAROTENE HYDROXYLASE 1* and *β-CAROTENE HYDROXYLASE 2* (*HYD1* and *HYD2*), which convert β -carotene into zeaxanthin within the β - β branch of the pathway (Nisar et al. 2015; Colasuonno et al. 2017a). Knocking out *TdHYD1* and *TdHYD2* using EMS-mutagenised TILLING lines has recently been shown to increase the amount of β -carotene within wheat grains (Garcia Molina et al. 2021; Yu et al. 2022; Bekkering et al. 2023), and this material could be crossed with these high proportion β - β branch or high zeaxanthin content accessions. During my PhD, I also generated *Tdhyd2* double-knockout lines by crossing EMS-mutagenised TILLING lines (data not shown). These had the same Kronos TILLING parent lines (K0870 and K4420) that Yu and colleagues (2022) used to generate their *Tdhyd2* knockout lines, but I had started these crosses before this paper was published. They showed that these *Tdhyd2* knockout lines had increased β -carotene within the wheat grain. Therefore, future work could cross the high proportion β - β branch accessions identified here with my *Tdhyd2* knockout lines.

3.3.4 Novel marker-trait associations and quantitative trait loci associated with carotenoid traits exist within the Watkins tetraploid collection

Using the measurements of total GCC, α -carotene, β -carotene, lutein, and zeaxanthin, I identified 15 MTAs (35K Axiom® Breeder's Array markers) through a GWAS. These can be used for durum wheat pre-breeding to follow regions associated with increased GCC from the Watkins tetraploid collection into modern durum cultivars through marker-assisted selection. A few of the MTAs appear close to a carotenoid biosynthesis gene that may explain their association (Figure 3.12) and this suggests some reliability of the overall approach. For example, *1B:AX-94659198*, associated with α -carotene content, is close to an orthologue of *OsCYP97C2*, a gene responsible for converting lutein to α -carotene (Niaz et al. 2023). Moreover, *5A:AX-95216226*, associated with lutein and total GCC, is located close to an orthologue of a rice *CAROTENOID CLEAVAGE DIOXYGENASE* gene (*OsCCD*), which is involved in carotenoid degradation (Vallabhaneni et al. 2010).

In addition, 14 QTLs were found to be associated with measurements of total GCC, YPC, the proportion of β - β branch carotenoids and individual carotenoid compounds (α -carotene, β -carotene and zeaxanthin). Of these, three QTLs overlapped regions previously associated with carotenoid content traits, and 11 did not overlap any previously described QTLs. This suggests that the GWAS analysis reliability agreed with and confirmed previously described QTLs, and importantly, it also identified new carotenoid-associated QTL regions not previously described. These novel QTLs could represent genetic material unique to the

Watkins tetraploid collection that is not currently exploited within durum wheat breeding. Previously, an investigation into the Watkins hexaploid collection revealed seven ancestral groups of bread wheat existed within this collection; however, only two of these were represented within modern bread wheat cultivars (Cheng et al. 2023). It is reasonable to assume the Watkins tetraploid collection will also contain genetic diversity not present in breeding germplasm, and future work is being carried out to characterise this genetic diversity.

Moreover, to the best of my knowledge, only two studies have used HPLC carotenoid analysis data to identify carotenoid-associated QTLs using GWAS (Guan et al. 2022; Requena-Ramírez et al. 2022). Instead, past studies have primarily employed less precise YPC and yellow index methods (Colasuonno et al. 2019). Recently, Requena-Ramirez and colleagues (2022) identified DArTSeq MTAs and QTL regions in a collection of 158 Spanish durum landraces, which had been screened with HPLC for grain α -carotene, β -carotene, lutein and zeaxanthin content, as well as a total carotenoid measurement and proportion of β - β branch carotenoids. They identified 28 MTAs, and some of their MTAs were close to associated QTLs found here. For instance, the Watkins tetraploid collection QTL, *prop2Bb* (659.9–664.3 Mbp), was located close to the DArTSeq MTA *2B:4412035* (699.6 Mbp), which was also associated with the proportion of β - β branch carotenoids. Again, this suggests agreement between the GWAS run here and previously described GWAS for carotenoid traits. However, they did not identify any MTAs associated with α -carotene and β -carotene, which were identified in this work. This could be due to the Watkins tetraploid collection being a larger, more diverse collection, and because the high-resolution genome sequence data allowed for rare alleles to be more easily identified. This highlights the novelty of this GWAS using HPLC measurements from the Watkins tetraploid collection.

Knetminer analysis of the genes found within the carotenoid-associated QTLs identified some genes, uncharacterised in wheat, orthologous to carotenoid biosynthesis genes (Section 3.2.3). For example, a gene within *bcaro_3A* was orthologous to *AtCCD8*, and six paralogous genes within *YPC_4Aa* had orthology to *AtZEP*. These are not identified by EnsemblPlants as exact one-to-one orthologue pairs of *AtCCD8* and *AtZEP*. Instead, they have one-to-many orthology with these *Arabidopsis* genes. For instance, the one-to-one orthologues of *AtZEP* (*TdZEP-2A* and *TdZEP-2B*) are found on chromosomes 2A and 2B (Figure 3.12). These one-to-many orthologues likely arose due to duplication events during the evolution of *T. turgidum*, and these genes have not been previously described as playing

a role in carotenoid biosynthesis in cereals. Therefore, this GWAS using high-resolution genome-sequence data possibly enabled me to identify novel genes involved in carotenoid biosynthesis unique to wheat. To confirm these findings, future work could characterise EMS-induced knockout TILLING lines within these genes and investigate their role in carotenoid biosynthesis. Alternatively, fine-mapping populations could narrow these QTLs to their causal genes.

In addition, Knetminer identified other genes found within the carotenoid-associated QTLs that are not a part of the main carotenoid biosynthesis pathway but have links to carotenoid biosynthesis. *TdDXR* was identified within the *zea_7A* QTL, which is a rate-limiting enzyme within the MEP pathway that produces precursor isoprenoids of carotenoids (Figure 1.1). The *zea_7A* QTL (22.8–33.1 Mbp) is close to a previously identified MTA (21.4 Mbp; *7A:AX-94424536*) associated with grain pigment colour (Rathan et al. 2022) and overlaps QTLs associated with YPC (*QTL0995_7A-Colasuonno_et_al_2014*) and yellow index (*QTL0072_SY-Colasuonno_et_al_2017*) (Colasuonno et al. 2014, 2017a). Therefore, the *TdDXR* gene may be the causal gene for these previously associated genomic regions. Additionally, the three transcription factors with possible links to carotenoid content identified by Knetminer within *YPC_4Ab*, *acaro_1A* and *acaro_7A* provide another novel mechanism for modifying carotenoid content. Future work could also characterise these genes to confirm whether they are causal genes within these carotenoid-associated QTLs.

Furthermore, some QTLs appear close to a carotenoid biosynthesis gene that can explain their association (Figure 3.12). For example, *propB_2Ba* is located close to *TdHYD1*, which converts β -carotene into zeaxanthin and functions within the β - β branch of the carotenoid biosynthesis pathway (Figure 1.1) (Nisar et al. 2015; Niaz et al. 2023). Likewise, *bcaro_6B* is located close to an orthologue of *OsCCD8*, a rate-limiting enzyme in the degradation of β -carotene into strigolactones (Guan et al. 2012). This is another QTL associated with β -carotene located close to a *CCD8* orthologue, in addition to *bcaro_3A* discussed above. This suggests that these may be the causal genes for these carotenoid-associated QTLs.

One QTL region frequently identified during GWAS for carotenoid content traits is at the end of chromosomes 7A and 7B (Blanco et al. 2011; Colasuonno et al. 2017a, 2019; Requena-Ramírez et al. 2022). This is where *PSY1* is located, the rate-limiting step in carotenoid biosynthesis and a key determinant for yellow index and YPC in durum wheat (Zhang and Dubcovsky 2008; Niaz et al. 2023). However, in this GWAS on the Watkins tetraploid collection, only *YPC_7A* was found close to this region. Interestingly, the

orthologues of *PSY1* on the Svevo v1 reference genome are found on chromosomes Un (TRITD0Uv1G062430) and 6B (TRITD6Bv1G228570; Figure 3.12). This is likely due to an error in the reference genome since *TdPSY1* was previously demonstrated to be located on chromosomes 7A and 7B in durum wheat (Campos et al. 2016; Vargas et al. 2016). It may be that this discrepancy of where *TdPSY1* is located on the Svevo v1 reference genome is why only one QTL was found close by. Alternatively, it may be possible that the *TdPSY1* alleles that increased total GCC were only introduced during the breeding for high YPC in durum wheat, distinguishing low total GCC landraces from high total GCC modern cultivars. Therefore, these high carotenoid alleles of *TdPSY1* may not be present within the landraces of the Watkins tetraploid collection and could have instead originated as a rare allele brought into breeding programmes during the past 30 years of durum wheat breeding.

Comparing these two GWAS resolutions, three QTLs and MTAs were located very close together or overlapped (Figure 3.12), providing very strong evidence that these regions are associated with their respective carotenoid content traits. Moreover, the overlapping *bcaro_3A* and *AX-95174558* (*β-caro.*) are located close to the *OsCCD8* orthologue discussed previously, which plays a key role in the degradation of β -carotene (Guan et al. 2012). However, 11 QTLs and 12 MTAs did not overlap between the two GWAS resolutions, indicating a low degree of similarity between the results of these methods. It would be expected that, despite their differing resolutions, similar genetic loci would be associated with similar carotenoid content traits. This discrepancy may suggest potential issues with one of the GWAS methods or differences in the approaches used to identify associated regions at these resolutions.

One possible cause for this incongruence could be the different filtering methods employed by each GWAS. The QTLs identified in the high-resolution GWAS were filtered using GATK (Van der Auwera and O'Connor 2020), and a kinship matrix was calculated with GEMMA and used as a covariate for the GWAS (Zhou and Stephens 2012). In contrast, the GWAS models using the 35K Axiom® Breeder's Array genotypic data handled and corrected for population structure differently. These variations in filtering methods may have led to differences in the regions identified as significantly associated with the traits. In the high-resolution GWAS, many markers were identified above the significance cutoff (grey dashed line in Figure 3.11), but after filtering and correcting for kinship, only a few were confidently identified. It is possible that several of these significantly associated markers were also

identified in the 35K Axiom® Breeder's Array GWAS, but then filtered out when correcting for population structure and kinship.

Due to time constraints inherent within a PhD, the GWAS conducted on carotenoid traits in this study was limited to a single harvest year and environment. This raises the possibility that the identified MTAs and QTLs associated with carotenoid traits might exhibit variability under different conditions. The identification of stable QTLs is particularly valuable in plant breeding as it indicates that the genetic influences on the trait are consistent across various environments and years, thereby reliably improving desirable characteristics in crops (Torkamaneh and Belzile 2022). Recognising the importance of stable QTLs, future work is ongoing to examine the consistency of these novel QTLs across another harvest year.

4 Characterising the function of the *ORANGE* gene and the ‘golden SNP’ substitution in wheat

4.1 Introduction

Success has been found in the past three decades of breeding for yellow pigment content in durum wheat, as evidenced by higher carotenoid concentrations of modern durum varieties than those pre-1990 (Digesù et al. 2009). However, previously investigated natural variation in carotenoids is still relatively low and insufficient to reach significant levels of provitamin A (PVA) activity, making other approaches to diversify wheat carotenoid variation necessary (Giuliano 2017). Efforts to improve the carotenoid content of crops have primarily been achieved by manipulating the carotenoid biosynthesis pathway.

One common strategy has been to increase metabolic flux into the carotenoid pathway by overexpressing one or more biosynthetic enzymes, referred to as a ‘push’ strategy (Zheng et al. 2020a). In particular, the rate-limiting enzyme PHYTOENE SYNTHASE (PSY) has been frequently overexpressed to produce high-carotenoid crops. In wheat, *PSY* overexpression alongside a bacterial phytoene desaturase gene (*CRTI*) increased endosperm total carotenoid content 10-fold (Cong et al. 2009). A later attempt overexpressed a bacterial phytoene synthase gene (*CRTB*) as well as *CRTI*, finding a 76-fold increase in grain PVA content compared with non-transgenic controls (Wang et al. 2014). Although this led to a level of PVA carotenoids that was still much lower than found in vegetables, the authors concluded that even small increases in the carotenoid content of wheat grains might help combat vitamin A deficiency due to the huge daily consumption of wheat-based products worldwide. A strategy not explored in wheat is to increase the sequestration and sink capacity of carotenoids, referred to as a ‘pull’ strategy. This could be achieved by increasing the number of chromoplasts, a fully developed plastid that stores massive amounts of PVA carotenoids in highly enriched sequestration substructures (Sun et al. 2018). To date, the only gene found to regulate chromoplast formation is the *ORANGE* gene (*OR*), with mutant forms of *OR* promoting chromoplast formation in non-photosynthetic tissue (Watkins and Pogson 2020).

OR has chaperone activity and plays a role in carotenoid biosynthesis by directly interacting with and post-transcriptionally stabilising PSY, increasing its protein activity (Figure 1.1) (Zhou et al. 2015; Sun et al. 2023b). Knocking out *OR* within *Arabidopsis thaliana*

(*Arabidopsis*) and melons reduces carotenoid content within various tissues such as leaves, callus and fruit flesh (Zhou et al. 2015; Chayut et al. 2017; Sun et al. 2023b). Similarly, the overexpression of *AtOR* increased carotenoid content within *Arabidopsis* callus, rice grains and white maize, suggesting OR has a similar function between monocots and dicots (Bai et al. 2014; Yuan et al. 2015; Berman et al. 2017). Conversely, Yu and colleagues (2021) found that *OsOR* overexpression did not change grain carotenoid content (GCC) but reduced the carotenoid content in rice leaves and grain-derived calli. Additionally, they found that *Osor* mutants did not affect the leaf carotenoid content of rice. However, another study found that *OsOR* overexpression in rice does not affect the leaf carotenoid content, suggesting uncertainty about the role of OsOR in rice (Jung et al. 2021).

In melon, the 'golden SNP' arginine-to-histidine substitution within CmOR was shown to enhance carotenoid levels through a distinct mechanism from PSY post-transcriptional stabilisation (Chayut et al. 2017), most likely by increasing carotenoid sink strength through promoting chromoplast formation (Yuan et al. 2015; Zhou et al. 2015). Tomato, sweet potato and *Arabidopsis* lines overexpressing the 'golden SNP' *OR* sequence (*OR^{His}*) had higher carotenoid content than lines overexpressing the wild-type *OR* (*OR^{WT}*), suggesting the 'golden SNP' substitution increases carotenoid content in these species (Yuan et al. 2015; Kim et al. 2019, 2021; Yazdani et al. 2019). Using the 'golden SNP' within wheat the *OR* gene could present an exciting avenue for carotenoid biofortification; however, the role of *OR^{WT}* and the effect of *OR^{His}* has not been investigated in wheat.

In this chapter, I aimed to investigate the function of *OR* in wheat and the effect that the 'golden SNP' arginine-to-histidine substitution has within the wheat OR protein on GCC. I identify the wheat *OR* orthologues as on chromosomes 6A, 6B and 6D, which share high sequence conservation with the OR proteins of other plants. Using ethyl methanesulfonate (EMS) mutagenised TILLING lines, I found that wheat *or* mutants have reduced GCC, consistent with what was found in *Arabidopsis* and melons. Through the overexpression of wheat *OR^{WT}* and *OR^{His}*, I found that *OR^{His}* increases GCC, indicating the 'golden SNP' substitution functions within wheat. Finally, I found very low allelic diversity within wheat *OR* when searching the pangenome and sequenced Watkins global landrace collection.

4.2 Results

4.2.1 Wheat orthologues of *OR* are located on Chromosome 6A, 6B, 6D

To identify the wheat genes encoding the wheat orthologues of *OR*, I searched for the bread wheat and durum wheat orthologues of the melon *OR* (*CmOR*) protein sequence using a BLASTp search. For *Triticum aestivum*, this identified the proteins TraesCS6A02G241400 (on chromosome 6A), TraesCS6B02G283200 (on chromosome 6B) and TraesCS6D02G223600 (on chromosome 6D), which share over 98% protein sequence identity with each other. These genes were previously labelled as orthologues of the *Arabidopsis AtOR* gene by EnsemblPlants. I will refer to these genes as *TaOR-6A* (TraesCS6A02G241400), *TaOR-6B* (TraesCS6B02G283200) and *TaOR-6D* (TraesCS6D02G223600). For *T. turgidum* ssp. *durum*, this BLASTp identified the proteins TRITD6Av1G155000 (on chromosome 6A) and TRITD6Bv1G140710 (chromosome 6B), which share over 99% protein sequence identity with each other. These genes have also been labelled as orthologues of *CmOR*, *AtOR* and the *TaOR* triad on EnsemblPlants. I refer to these genes as *TdOR-6A* (TRITD6Av1G155000) and *TdOR-6B* (TRITD6Bv1G140710). *TdOR-A* and *TaOR-A* share 100% protein sequence identity, as do *TdOR-B* and *TaOR-B*, showing these proteins have considerable conservation throughout the evolution of bread wheat. Because of this similarity, I will focus on the sequence analysis of *TaOR-6A*, *TaOR-6B* and *TaOR-6D*.

Additionally, this BLASTp identified the proteins TraesCS6A02G197700, TraesCS6B02G218700 and TraesCS6D02G183000 in *T. aestivum* and TRITD6Av1G108670 and TRITD6Bv1G093390 in *T. turgidum* ssp. *durum* as sharing high similarity to *CmOR*. These are the wheat orthologues of *ORLIKE*, a paralog of *OR* which has also been found to be involved in the carotenoid biosynthesis pathway in *Arabidopsis* (Zhou et al. 2015; Sun et al. 2023b). I refer to these genes as *TaORLIKE-6A* (TraesCS6A02G197700), *TaORLIKE-6B* (TraesCS6B02G218700), *TaORLIKE-6D* (TraesCS6D02G183000), *TdORLIKE-6A* (TRITD6Av1G108670) and *TdORLIKE-6B* (TRITD6Bv1G093390). However, in this chapter, I focused on functionally characterising the *TaOR* and *TdOR* homoeologues, which have the highest homology to *CmOR*, where the 'golden SNP' has naturally occurred.

The open reading frames of *TaOR-6A*, *TaOR-6B* and *TaOR-6D* are predicted to encode proteins of 327, 324 and 326 amino acids with an estimated molecular mass of 34.9 kDa, 34.7 kDa, 34.8 kDa, respectively. All three wheat homoeologues have the same eight-exon structure as *CmOR* and *AtOR* (Figure 4.1). Additionally, the amino acid sequences of the three *TaOR* homoeologues are highly conserved with one another (Figure 4.2). There is a

difference in the length of a run of alanine amino acids around position 70. There are also three amino acid differences between the homoeologues, all located within the first 80 amino acids (A-genome/B-genome/D-genome: PSP, DED, SAA). None of these homoeologous polymorphisms are in conserved regions of the protein, and two of the substitutions (DED and SAA) had a positive (+2) BLOSUM score, suggesting these residue substitutions have strong similarity, while PSP had a negative BLOSUM score of (-2). Each of the three *TaOR* homoeologues is predicted to contain an N-terminal chloroplast transit peptide, which is expected since OR is plastid localised in other plants.

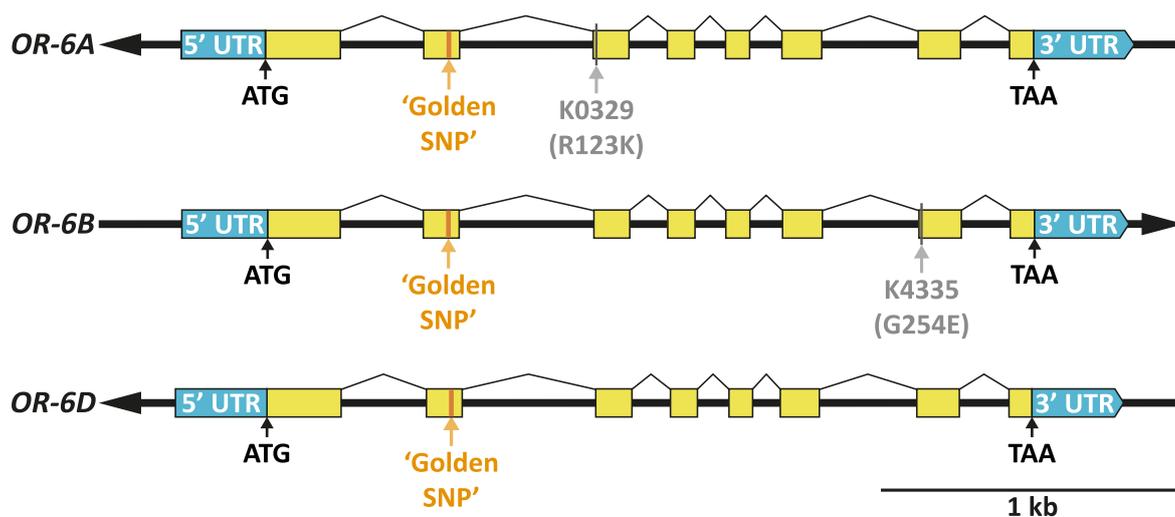


Figure 4.1 Diagram of *TaOR* homoeologues within the wheat genome. Exons are represented by yellow boxes, and light blue boxes represent untranslated regions. The location of the two TILLING line missense mutations that were selected to knock out TdOR in Kronos are shown by grey arrows (Section 4.2.2). The location of the 'golden SNP' residue in each OR homoeologue is shown by the orange arrow. The black arrow shows the direction of the genomic DNA strand.

The *TaOR* homoeologues share 72.8–73.9% amino acid sequence identity with CmOR (Figure 4.2), with the highest identity found at the C-termini of the proteins. They also share identical locations of two CxxCxGxG and CxxCxxxG motifs within the DnaJ cysteine-rich zinc-finger domains. The 'golden SNP' residue is located at positions 110, 107 and 109 in TaOR-6A, TaOR-6B and TaOR-6D, respectively. Like OR proteins in other plant species, the amino acid at this location is a highly conserved arginine residue, which is found in the green-flesh CmOR^{WT}. Therefore, the wheat TaOR proteins do not contain the 'golden SNP' histidine amino acid that increases carotenoid accumulation in orange-flesh melons.

<i>Cucumis melo</i>	---MDRVLVASYPINHLIRPHSFRIDYCWSTCFTRSRLNSGKERQKLSRRWRWRMSADST	57
<i>Triticum aestivum</i> 6A	<u>MLCSGRMLACSG</u> -----LSPGRLRPPRA---YADR-----LRP-PL <u>PARRRWRVAASAAA</u>	45
<i>Triticum aestivum</i> 6B	<u>MLCSGRMLACSG</u> -----LSPGRLRPPRA---YADR-----LRP-PL <u>SARRRWRVAASAAA</u>	45
<i>Triticum aestivum</i> 6D	<u>MLCSGRMLACSG</u> -----LSPGRLRPPRA---YADR-----LRP-PL <u>PARRRWRVAASAAA</u>	45
	.*:..* : * :* . :.:* * ** ** ::	
<i>Cucumis melo</i>	D-----SSSSSFAPSV-----ESDPSDKTSASFCEIEGPETVQDFAKMELQEIQE	103
<i>Triticum aestivum</i> 6A	PGGSPDLPSSSSTPPPFAGDDQAAAAAASSSSSGFCIEGPEVTVQDFDKLDLQEILD	105
<i>Triticum aestivum</i> 6B	PGGSPDLPSSSSTPPPFAGDEQAAAA---AASSSSGFCIEGPEVTVQDFDKLDLQEILD	102
<i>Triticum aestivum</i> 6D	PGGSPDLPSSSSTPPPFAGDDQAAAAA---AASSSSGFCIEGPEVTVQDFDKLDLQEILD	104
	**** * . : .:.*:***** *:**** :	
<i>Cucumis melo</i>	NIRSRNKIFLHMEEVRLRIQQRIKNAELGISKEERENELPNFSPFIFPLPPLSSENLK	163
<i>Triticum aestivum</i> 6A	NIRSRNKIFLHMEEIRRLRIQQRIKNAELGISNEEPEGELPDFSPFIFPLPPLSAANLK	165
<i>Triticum aestivum</i> 6B	NIRSRNKIFLHMEEIRRLRIQQRIKNAELGISNEEPEGELPDFSPFIFPLPPLSAANLK	162
<i>Triticum aestivum</i> 6D	NIRSRNKIFLHMEEIRRLRIQQRIKNAELGISNEEPEGELPDFSPFIFPLPPLSAANLK	164
	*****:*****:* *.*:*****: **	
<i>Cucumis melo</i>	LYYVTCYSLIAGIILFGLLAPTELEKLGIGGTSYEDFIRSVHLPMQLSQVDPIVASFSG	223
<i>Triticum aestivum</i> 6A	VYYATCFSLIAAIMVFGGFLAPILELKLIGGTSYADFIRNVHLPMQLSQVDPIVASFSG	225
<i>Triticum aestivum</i> 6B	VYYATCFSLIAAIMVFGGFLAPILELKLIGGTSYADFIRNVHLPMQLSQVDPIVASFSG	222
<i>Triticum aestivum</i> 6D	VYYATCFSLIAAIMVFGGFLAPILELKLIGGTSYADFIRNVHLPMQLSQVDPIVASFSG	224
	:*.*:****.:*:*** *****:***** *****:*****:*****	
<i>Cucumis melo</i>	GAVGVISALMVVEINNVKQEHKRCYCLGTGYLACARCSNTGALVLTPEVSTLNGEQP	283
<i>Triticum aestivum</i> 6A	GAVGVISALMVVEINNVKQEHKRCYCLGTGYLACARCSSTGAVVLTPEVSTFSDGDQP	285
<i>Triticum aestivum</i> 6B	GAVGVISALMVVEINNVKQEHKRCYCLGTGYLACARCSSTGAVVLTPEVSTFSDGDQP	282
<i>Triticum aestivum</i> 6D	GAVGVISALMVVEINNVKQEHKRCYCLGTGYLACARCSSTGAVVLTPEVSTFSDGDQP	284
	*****:*****:*****:***:* *****:..**	
<i>Cucumis melo</i>	LSLPKTERCONCSGSKVMCPTCLCTGMAMASEHDPRIDPFD 325	
<i>Triticum aestivum</i> 6A	LSAPKTERCPNCSGAGKVMCPTCLCTGMAMASEHDPRIDPFD 327	
<i>Triticum aestivum</i> 6B	LSAPKTERCPNCSGAGKVMCPTCLCTGMAMASEHDPRIDPFD 324	
<i>Triticum aestivum</i> 6D	LSAPKTERCPNCSGAGKVMCPTCLCTGMAMASEHDPRIDPFD 326	
	** ***** ***:*****:*****:*****	

Figure 4.2 Alignment of OR protein sequences from melon (*Cucumis melo*) and bread wheat (*Triticum aestivum*). The N-terminal chloroplast transit peptides are underlined, identified by Tzuri and colleagues (2015) for CmOR and identified using TargetP for TaOR homoeologues. Polymorphisms between the wheat TaOR proteins are highlighted in yellow. The ‘golden SNP’ residue is highlighted in blue at position Arg108 in CmOR, Arg110 in TaOR-6A, Arg107 in TaOR-6B and Arg109 in TaOR-6D. Highly conserved cysteine-rich domains repeats are highlighted in green (two of CxxCxGxG and two of CxxCxxxG). Symbols under the alignment show residues with complete identity (*), highly similar properties (:), or weakly similar properties (.).

To investigate the similarity of the wheat OR proteins to those in other species, I constructed a neighbour-joining tree with the protein sequences of *TaOR* homoeologues and other *OR* orthologues. The wheat *TaOR* proteins clustered closely together on the tree with orthologues from *T. turgidum* and the wild relatives of bread wheat, *T. urartu* and *Ae. tauschii* (Figure 4.3). These clusters were found within the monocot clade of the *OR* proteins. The *OR* protein sequence of the A-genome progenitor *T. urartu* and *TaOR*-6A were 99.7% identical, with only two amino acid polymorphisms between the two. The *OR* protein sequence of the D-genome progenitor, *Ae. Tauschii*, and *TaOR*-6D were 100% identical. Again, this shows high sequence conservation of *OR* throughout the evolution of wheat. Moreover, the *TaOR*-6A, *TaOR*-6B and *TaOR*-6D were found to have very high protein sequence identity with *OR* orthologues from other plants (the lowest sequence similarity of *TaOR* was 56.1% with *Capsella rubella*, followed by 70.6% with *Amborella trichopoda*).

This is concurrent with OR having a highly similar protein sequence among plants (Tzuri et al. 2015; Sun et al. 2021).

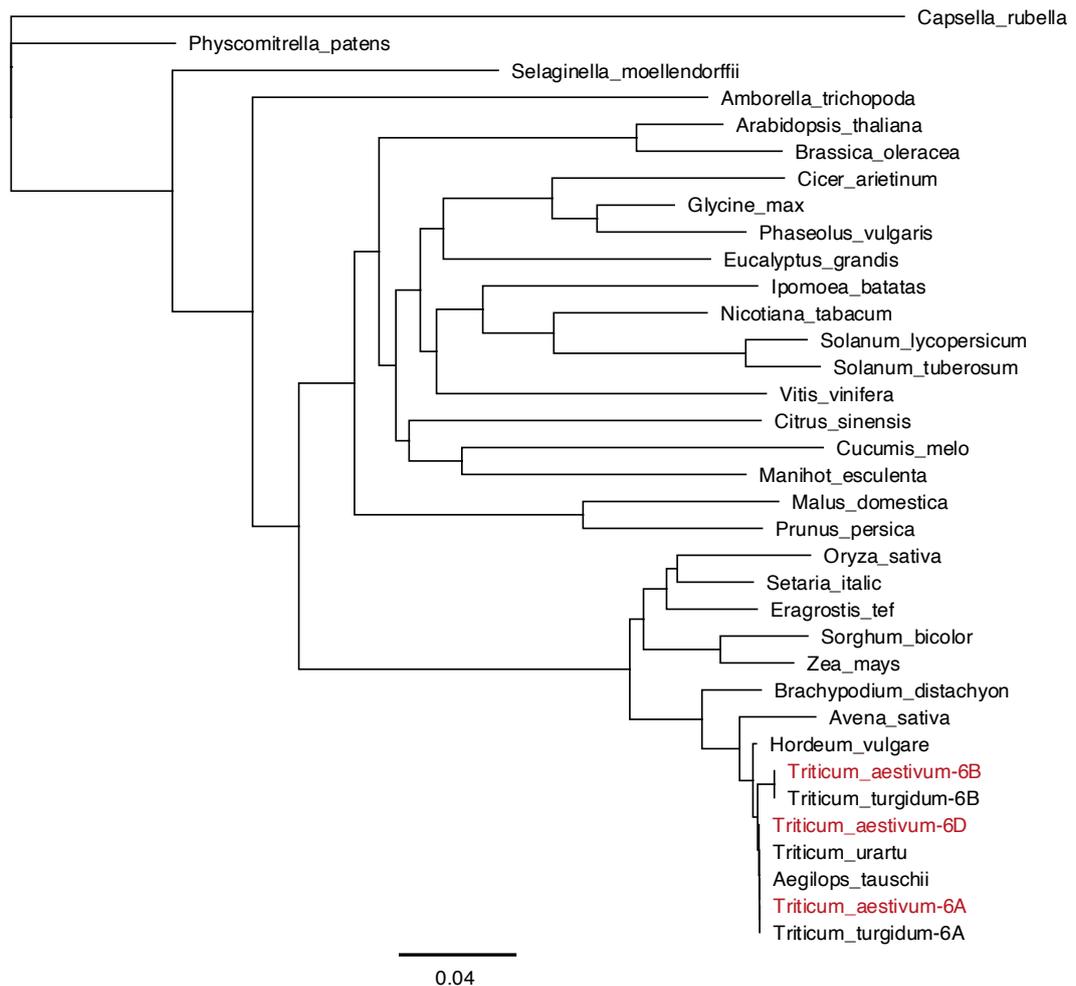


Figure 4.3 Phylogenetic tree of the protein sequences of 35 *OR* orthologues. A neighbour-joining tree was built with *Capsella rubella* as an outgroup. Wheat TaOR proteins are indicated in red. Branch lengths indicate the number of amino acid substitutions per site (as shown by the scale bar).

To investigate whether the *TaOR* homoeologues are expressed within non-photosynthetic tissues, specifically the grains, I used the expVIP wheat expression browser to analyse their expression within wheat (Borrill et al. 2016; Ramírez-González et al. 2018). Expectedly, due to the role OR plays in the chlorophyll biosynthesis pathway (Sun et al. 2023b), the *TaOR* homoeologues are highly expressed within the green tissue of leaves and shoots (\log_2 TPM=5.69–5.86, SEM=5.12–5.49, n=481), as well as in the spike (\log_2 TPM 4.32–4.61, SEM=4.19–4.65, n=280). In addition, I confirmed that *TaOR* is also expressed in the roots (\log_2 TPM 3.61–3.74, SEM=2.54–2.62, n=89) and the grain (\log_2 TPM 2.31–2.68, SEM=1.27–1.75, n=166), suggesting *TaOR* plays a role in these tissues too. I also asked what the pattern of expression between the *TaOR* homoeologues was and found it had a balanced expression where each homoeologue had a similar expression profile, with roughly 70% of wheat homoeologue triads having this profile (Ramírez-González et al. 2018).

4.2.2 Knocking out the *OR* gene reduces grain carotenoid content

To explore the role of OR in the wheat grain, I characterised *Tdor* mutants of a previously generated EMS population within Kronos, a cultivar of durum wheat. I used the *in silico* TILLING platform to identify mutations likely to affect TdOR protein function from an EMS-mutagenised population of Kronos (Krasileva et al. 2017). For *TdOR-6A*, I identified a missense mutation (Arg123Lys substitution) in the line K0329 with a SIFT score below 0.01 (SIFT<0.05 is considered deleterious). For *TdOR-6B*, I identified a missense mutation (Gly254Glu substitution) in the line K4335 with a SIFT score below 0.01. This mutation affects one of the highly conserved glycine residues within the DnaJ cysteine-rich zinc-finger domains (specifically at the end of the first CxxCxGxG motif), as shown in green in Figure 4.2. Notably, the substituted Arg123 and Gly254 residues are conserved across all 35 *OR* orthologues I previously examined. The locations of these TILLING missense substitutions within *TdOR* are illustrated in Figure 4.1.

I generated a homozygous double *Tdor* mutant line in the A-genome and B-genome copies of *TdOR* in an F₂ cross of K0329 and K4335 lines. From these F₂ plants, I also identified a homozygous *TdOR* wild-type line for use as a control. This control has a similar genetic background to that of the mutant line since both originated from the same cross between K0329 and K4335. I refer to the homozygous mutant line as '*Tdor* mutant' and the homozygous wild-type line as '*Tdor* wild-type'. In total, I identified 12 *Tdor* mutant plants and 6 *Tdor* wild-type plants from 192 F₂ progeny of the K0329 and K4335 F₁ cross. These numbers align well with the expected outcome of a dihybrid cross, which anticipates 12 individuals of each type (1/16 *Tdor* mutants and 1/16 *Tdor* wild-types). Furthermore, the distribution of genotypes among these F₂ seedlings closely followed a Mendelian 9:3:3:1 ratio, calculated as 8:35:3.57:3.04:1.04.

To save a year and produce data within a PhD project time frame, I used these F₂ plants in the following analysis. F₃ plants are currently being generated (March 2024) to confirm these findings. Due to the low grain yield per plant within these F₂ *Tdor* plants, I analysed 6-gram pools of F₃ grain for carotenoid content from the three highest-yielding F₂ plants (2 grams from each) of the *Tdor* mutant plants and the *Tdor* wild-type plants. In each pool, the background mutations are expected to be similar as they originated from the same cross, and the only difference is expected to be the genotype of *TdOR*. Additionally, no genetic linkage interference is expected as the diversity arises from EMS random mutagenesis. The pooled samples were used for measuring carotenoid content, and three technical replicates

were conducted for both the *Tdor* mutant and *Tdor* wild-type pooled flours, with the averaged results presented in Figure 4.4 and Table 4.1. Total GCC was reduced by 33.8% in the *Tdor* mutant flour compared to the *Tdor* wild-type flour (average of 0.896 $\mu\text{g/g}$ compared to 1.354 $\mu\text{g/g}$ respectively). The greatest differences were in the PVA carotenoids, with the *Tdor* mutant flour containing 61.0% less α -carotene and 50.7% less β -carotene than the *Tdor* wild-type control flour. Lutein, the predominant carotenoid in wheat grains, was reduced by 36.2%, and zeaxanthin was reduced by 8.5% in the *Tdor* mutant flour compared to the *Tdor* wild-type flour. This suggests *TdOR* plays a role in carotenoid biosynthesis in wheat and, consistent with what has been found for *OR* orthologues in other species, is a potential target for grain carotenoid biofortification.

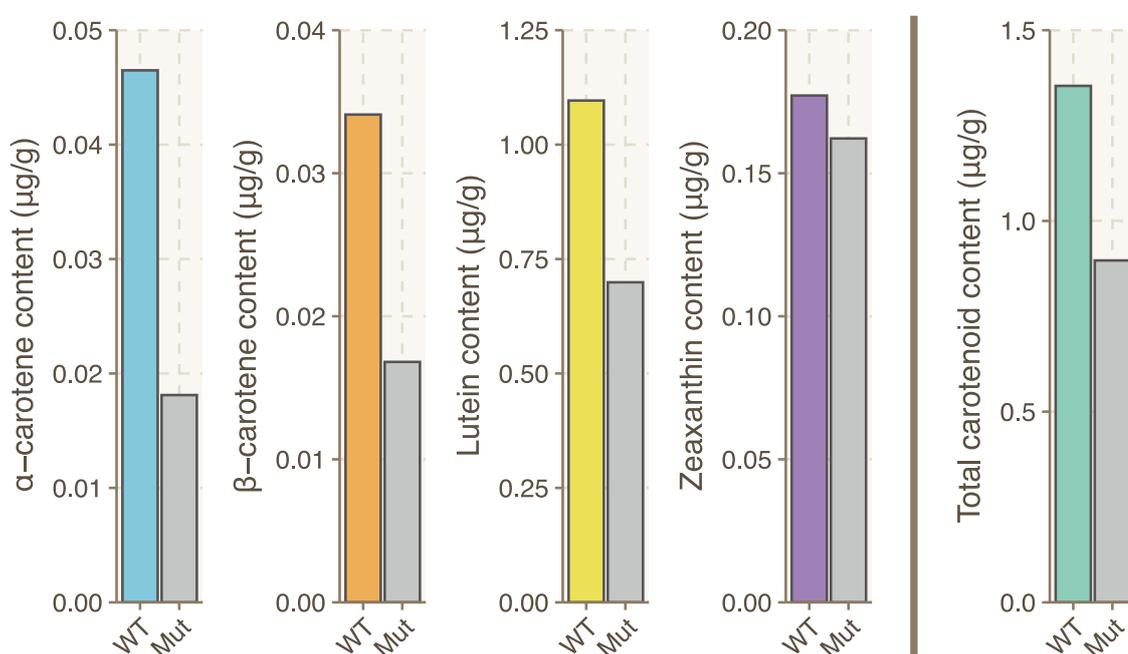


Figure 4.4 Grain carotenoid content of *Tdor* mutants compared to *Tdor* wild-type plants. Due to the low yields of F_2 *Tdor* mutant and F_2 *Tdor* wild-type plants, these measurements were done on the pooled grains of three *Tdor* mutants and three *Tdor* wild-type plants. Each bar represents an average of three technical replicates performed on these pooled grains. Total carotenoid content is the content of α -carotene, β -carotene, lutein and zeaxanthin. WT=*Tdor* wild-type plants; Mut=*Tdor* mutant plants.

Table 4.1 Grain carotenoid content of the *Tdor* mutants compared to the *Tdor* wild-type plants. Measurements were made on pooled grains of three plants for each line. α -caro.= α -carotene, β -caro.= β -carotene, Zeaxanth.=zeaxanthin, GCC=grain carotenoid content.

Line	α -caro.	β -caro.	Lutein	Zeaxanth.	Total GCC
<i>Tdor</i> mutant	0.018	0.017	0.699	0.162	0.896
<i>Tdor</i> wild-type	0.046	0.034	1.096	0.177	1.354

To investigate whether *TdOR* affects physiological growth and has pleiotropic effects on grain growth, I compared the plant height, grain yield per plant, grain area and grain weight between *Tdor* mutants and *Tdor* wild-type plants. The results of this are found in Figure 4.5 and Table 4.2. As expected from F₂ TILLING plants carrying many background mutations, the variation between the plants in the measured traits was sometimes high; for instance, the grain yield per plant in both *Tdor* mutant and *Tdor* wild-type plants varied from 8.44 grams to some producing very little or no grain. There were no statistically significant differences in plant height ($p=0.13$, $t(14)=-1.62$, Student's t-test) and grain yield per plant ($p=0.69$, $t(14)=0.41$, Student's t-test). Additionally, the thousand-grain weight (TGW) and grain area of the three analysed plants (producing the most amount of grain) were similar between *Tdor* mutant plants (39.2 grams and 16.0 mm²) and *Tdor* wild-type plants (41.7 grams and 16.9 mm²).

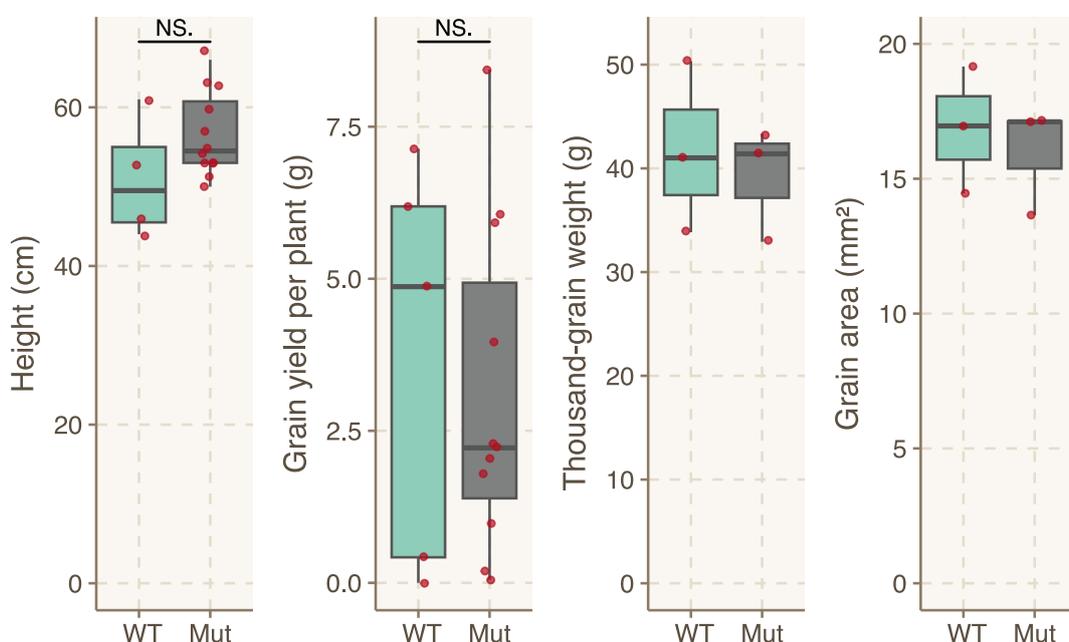


Figure 4.5 Preliminary plant growth and grain physiology measurements of *Tdor* mutants. Plant height and grain yield per plant were measured on all 16 *Tdor* F₂ plants. Thousand-grain weight and grain area were measured on the three highest-yielding *Tdor* mutant and *Tdor* wild-type plants. Red dots are measurements from individual plants. WT=*Tdor* wild-type plants; Mut=*Tdor* mutant plants.

Table 4.2 Preliminary plant growth and grain physiology measurements of *Tdor* mutants and *Tdor* wild-type plants. Values show averages of each line with standard deviation in brackets. Plant height and grain yield per plant were measured on all 16 *Tdor* F₂ plants. Thousand-grain weight and grain area were measured on the three highest-yielding *Tdor* mutant and *Tdor* wild-type plants. TGW=thousand-grain weight.

Line	Height (cm)	Grain yield per plant (g)	TGW (g)	Grain area (mm ²)
<i>Tdor</i> wild-type	51.00 (7.70)	3.72 (3.31)	41.72 (8.26)	16.86 (2.35)
<i>Tdor</i> mutant	56.50 (5.27)	3.09 (2.69)	39.22 (5.55)	15.97 (2.01)

4.2.3 Overexpressing the 'golden SNP' *OR* gene increases total grain total carotenoid content

The TILLING *Tdor* mutant lines suggest that *OR* plays a role in carotenoid biosynthesis in wheat grains. However, it is unknown whether the arginine-to-histidine 'golden SNP' substitution in the *OR* protein will affect the carotenoid content of wheat. To investigate whether the 'golden SNP' substitution increases carotenoid content in wheat, I overexpressed both *TaOR^{WT}* and *TaOR^{His}* in wheat and compared the carotenoid content of these plants. Based on the previous investigation of homoeologue sequence similarity and expression profiles (Section 4.2.1), I chose *TaOR-6D* from the Chinese Spring cultivar to be synthesised and expressed. This had an intermediate number of arginine repeats around position 70 between *TaOR-A* and *TaOR-B* (Figure 4.2). Additionally, of the amino acid polymorphisms between these proteins (PSP, DED, SAA), the D-genome has the more common amino acid (P in PSP, D in DED and A in SAA). The protein sequence of Chinese Spring *TaOR-6D* is also identical to the protein sequence of *TaOR-6A* and *TaOR-6D* in the cultivar Cadenza, which I transformed with this overexpression construct. The 'golden SNP' does not alter the splicing of the transcript (Tzuri et al. 2015); therefore, I overexpressed just the coding sequence of *TaOR-6D*. For the 'golden SNP' *TaOR-6D* sequence, I introduced a single point mutation, changing CGC-to-CAC, to install the arginine-to-histidine substitution at residue 109 (Figure 4.2). I will henceforth refer to the wild-type *TaOR-6D* sequence as *TaOR^{WT}* and the 'golden SNP' *TaOR-6D* sequence as *TaOR^{His}*.

I designed constructs for overexpressing *TaOR^{WT}* and *TaOR^{His}* under the rice actin promoter (OsActin pro). I refer to these plasmids as pAct-OR^{WT} and pAct-OR^{His}, and a plasmid map of these is found in Figure 4.6. I also included the developmental regulators *GRF4-GIF1*, which substantially increase the efficiency of regeneration in wheat and allow the transformation of Cadenza and Kronos (Debernardi et al. 2020). I assembled these constructs into the backbone of the pGoldenGreenGate-M wheat expression vector and transformed these into the tetraploid variety Kronos and hexaploid variety Cadenza. I regenerated 148 T₀ plantlets, 67 Cadenza lines (39 pAct-OR^{WT} and 28 pAct-OR^{His}), and 81 Kronos lines (15 pAct-OR^{WT} and 66 pAct-OR^{His}). The copy number of these ranged from lines with zero copies that had escaped hygromycin selection to lines with 21 copies, with an average copy number of 4. I selected two T₀ plantlets of each of the following copy numbers to keep: 1 copy, 2 copies, 3 copies, 6 copies and >6 copies. This gave 10 transgenic T₀ plantlets for each construct and genotype (Cadenza pAct-OR^{WT}, Cadenza pAct-OR^{His}, Kronos pAct-OR^{WT}, Kronos pAct-OR^{His}).

Several transgenic T₀ Kronos lines had severe spike deformities (Appendix Material 14), severely affecting the seed set of the transgenic T₀ Kronos lines (14/20 lines were sterile, and 4/20 produced below 10 grains). For Cadenza, 2 out of 20 lines were sterile, and both were found in lines with >6 copies of the transgene. Due to the sterility problems with the transgenic T₀ Kronos lines, my subsequent analysis was restricted to the transgenic T₀ Cadenza plants overexpressing *TaOR*^{WT} and *TaOR*^{His}.

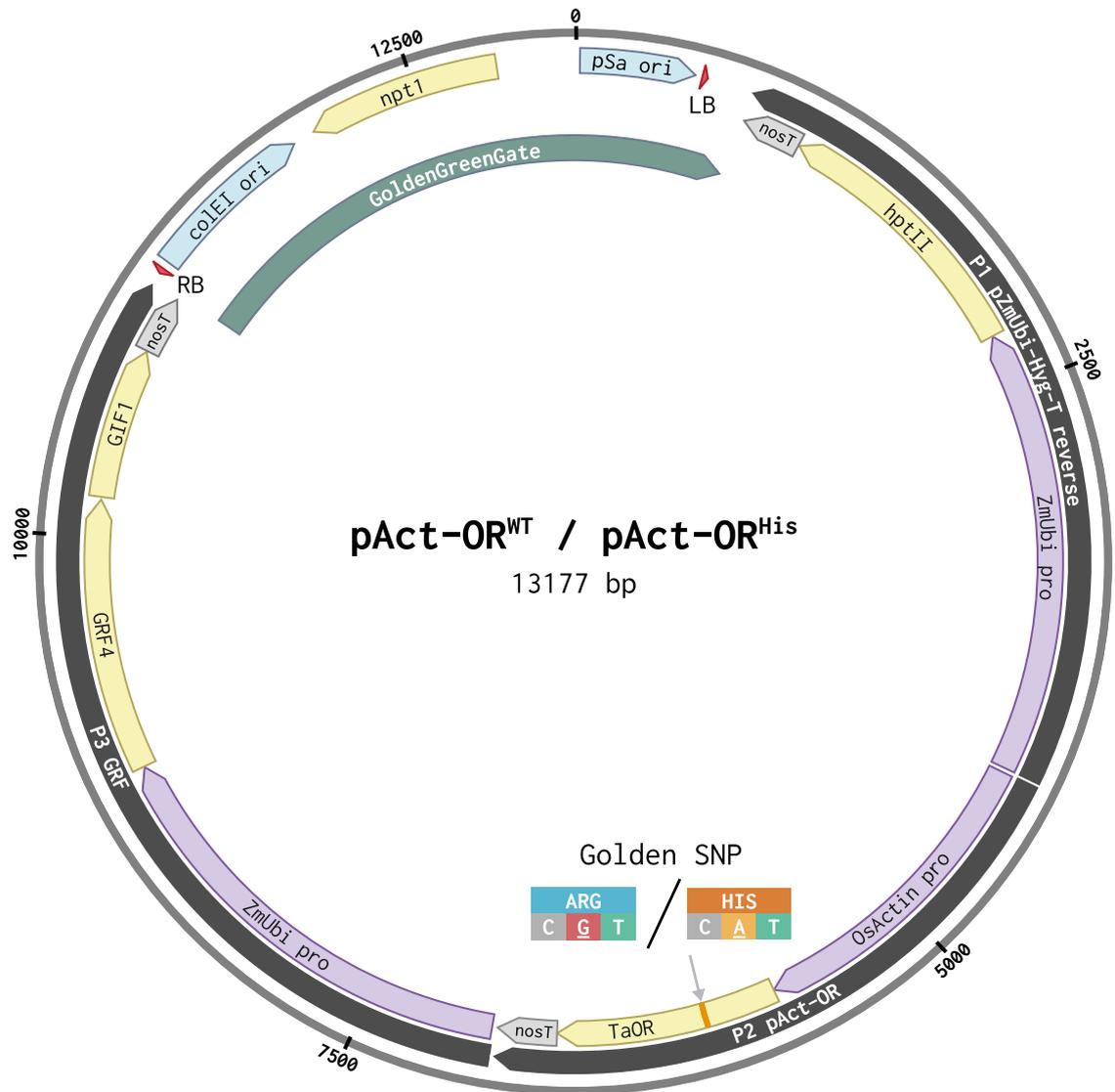


Figure 4.6 Plasmid map of the *TaOR*^{WT} and *TaOR*^{His} overexpression constructs. The pAct-OR^{WT} and pAct-OR^{His} plasmids differ by a single nucleotide at the 'golden SNP' position in the *TaOR* sequence, this is highlighted on the plasmid map. The *TaOR-6D*^{WT} and *TaOR-6D*^{His} are both under the control of the rice actin promoter (*OsActin pro*). The hygromycin selection gene (*hptII*) and developmental regulators (*GRF4-GIF1*) were also included on the plasmid. RB=right border, LB=left border, bp=base pairs.

Growing callus tissue in the dark increases flux into the carotenoid biosynthesis pathway, and this has been used as a visual screen for increased carotenoid biosynthesis in other species (Kim et al. 2013; Bai et al. 2014; Tzuri et al. 2015; Endo et al. 2019). To investigate whether the overexpression of *TaOR^{WT}* or *TaOR^{His}* gives a visual phenotype in the callus, I generated seed-derived callus from immature T₁ embryos of the pAct-OR^{WT} and pAct-OR^{His} Cadenza T₀ lines of 2 and 6 copy numbers. I grew this callus on resting media in the dark for 10 weeks, after which no visible sign of carotenoid accumulation was found for any callus produced from these *TaOR^{WT}* and *TaOR^{His}* overexpression lines (Figure 4.7), suggesting a visual screen cannot identify calli expressing the 'golden SNP' *TaOR* sequence.

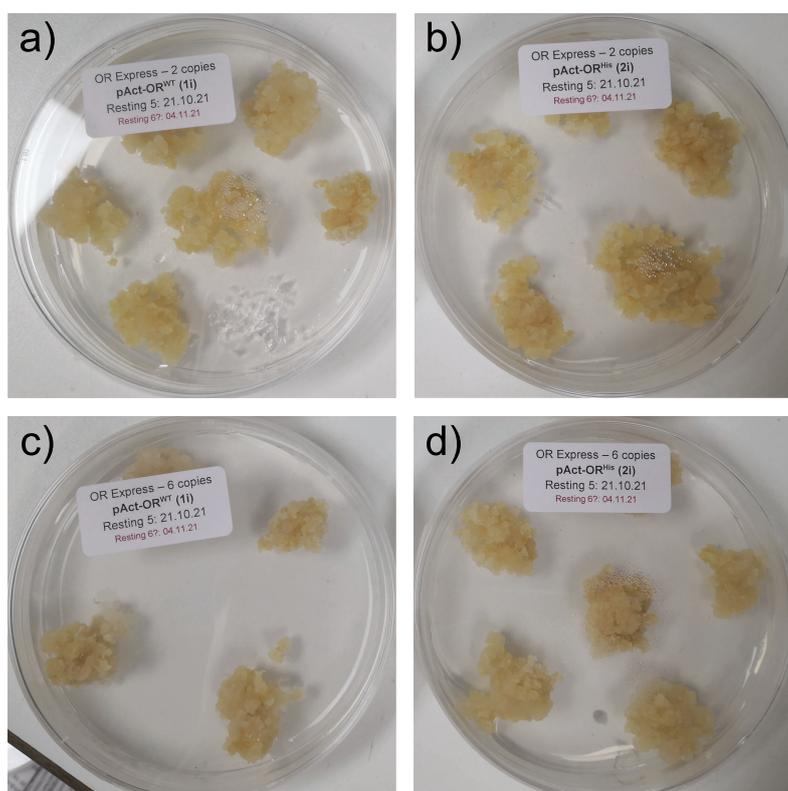


Figure 4.7 Dark-grown seed-derived callus of pAct-OR^{WT} and pAct-OR^{His} T₁ Cadenza immature embryos. (a) pAct-OR^{WT} from 2 copy number plants, (b) pAct-OR^{His} from 2 copy number plants, (c) pAct-OR^{WT} from 6 copy number plants and (d) pAct-OR^{His} from 6 copy number plants. Calli were grown in the dark for 10 weeks on resting media.

There was insufficient T₁ grain to screen for carotenoid content, and the T₁ grain also had a lot of visual variation in grain shape and size, so I grew T₁ plants up to analyse T₂ grain instead. For use as controls, I used T₁ plants that only contained the *GRF4-GIF1* developmental regulator gene under the same promoter used in my constructs. I refer to these as GRF control plants. I germinated T₁ grains from two T₀ plants with 1 copy, 2 copies and 4 copies of each transgene (*TaOR^{WT}*, *TaOR^{His}* and only *GRF4-GIF1*). I selected two T₁ plants with 1 copy (referred to as 1C), 2 copies (referred to as 2C) and 3–5 copies (referred to as hiC) for each of these transgenes, selecting six transgenic lines per transgene. I

referred to these lines by their transgene name (OR^{WT}, OR^{His} or GRF), the number of copies (1C, 2C or hiC) and which of the two chosen lines it is (A or B); for example, OR^{WT}-1C-A. I selected the lines based on copy number rather than zygosity, which meant I selected a mix of homozygous and hemizygous lines. Figure 2.3 shows the lineage, naming and zygosity of these lines. I grew eight plants for each of these lines. In addition, I grew 8 non-transgenic Cadenza plants alongside these lines to act as non-transgenic controls.

To investigate whether *TaOR*^{WT} and *TaOR*^{His} overexpression influences carotenoid biosynthesis, I analysed the GCC of the GRF control plants, OR^{WT} plants, OR^{His} plants and wild-type plants. Due to low grain yield per plant within these T₁ transgenics, I analysed 5-gram pools of flour for carotenoid content from four plants for each line (1.25 grams each). This was necessary because a minimum of 5 grams of grain was required for carotenoid analysis. One line, OR^{His}-2C-A, did not produce enough grain to analyse GCC. Three technical replicates were conducted for each flour pool, with the averaged results presented in Figure 4.8, Figure 4.9 and Table 4.3. The GCC of individual carotenoid compounds for each line is shown in Appendix Material 15.

There was no observed additive effect due to copy number variations on the carotenoid content among the GRF control plants, OR^{WT} plants and OR^{His} plants, as depicted in Figure 4.9. Consequently, these lines, despite varying copy numbers, were grouped based on their respective transgenes for mean comparisons using ANOVA. The *TaOR*^{His} overexpression plants had a significantly higher average total GCC (0.738 µg/g, SD=0.064) compared to the *TaOR*^{WT} overexpression (0.608 µg/g, SD=0.027) and GRF control plants (0.606 µg/g, SD=0.120), as revealed by this ANOVA (p=0.029, F(2,14)=4.598, one-way ANOVA). This suggests that the 'golden SNP' substitution in *TaOR* increases the carotenoid content in wheat grains. The *TaOR*^{His} overexpression plants had a higher content of lutein and zeaxanthin compared to the *TaOR*^{WT} and GRF control plants but showed similar levels of the PVA carotenoids, α-carotene and β-carotene (Table 4.3 and Figure 4.8). ANOVAs performed on these individual carotenoid compounds found no significant differences between the transgenes for lutein (p=0.063, F(2,14)=3.39, one-way ANOVA), α-carotene (p=0.941, F(2,14)=0.061, one-way ANOVA) and β-carotene content (p=0.128, F(2,14)=2.385, one-way ANOVA). An ANOVA revealed significant differences between zeaxanthin content for the *TaOR*^{His} overexpression lines and the GRF control lines (p=0.028, F(2,14)=4.696, one-way ANOVA). Figure 4.8 presents a visual summary of the Tukey Honest Significant Difference test for pairwise comparisons, indicated by significance letters on the boxplots. The tables

of specific pairwise comparisons and statistical significance are detailed in Appendix Material 16.

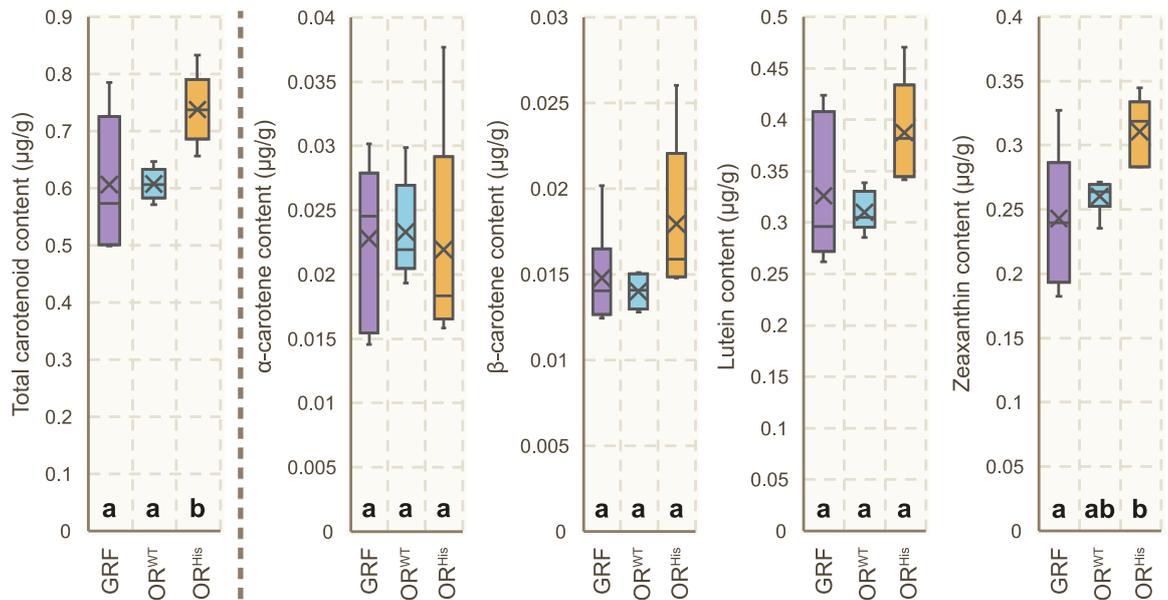


Figure 4.8 Grain carotenoid content of T₁ *TaOR*^{WT} and *TaOR*^{His} overexpression plants compared to GRF control plants presented as box plots. Measurements were done on the pooled grain of four plants from each line, which were grouped based on their transgene. Crosses within the boxplots show the average value for each transgene. Bold letters below boxplots indicate statistical significance between the groups as determined by a one-way ANOVA. GRF=*GRF4-GIF1* transgenic controls; OR^{WT}=pAct-OR^{WT} transgenic plants; OR^{His}=pAct-OR^{His} transgenic plants.

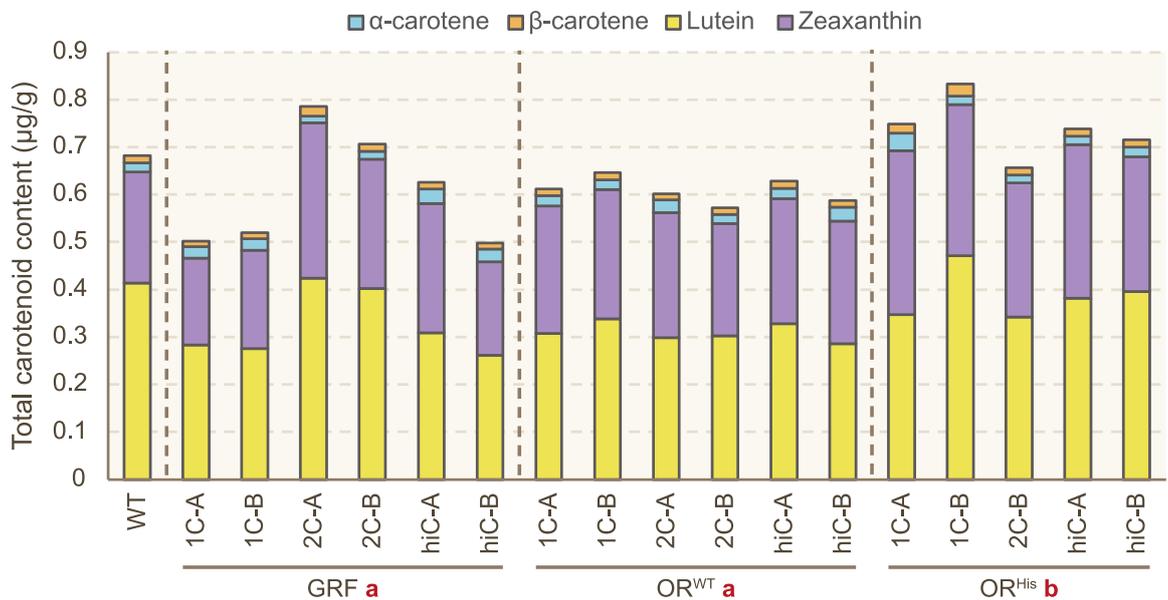


Figure 4.9 Grain carotenoid content of T₁ *TaOR*^{WT} and *TaOR*^{His} overexpression lines compared to GRF and wild-type controls presented as stacked bar charts. Measurements were done on the pooled grain of four plants from each line. Each bar represents an average of three technical replicates performed on these pooled grains. Bold red letters next to the groups' names indicate statistical significance between total grain carotenoid content as determined by a one-way ANOVA. WT=non-transgenic wild-type controls; GRF=*GRF4-GIF1* transgenic controls; OR^{WT}=pAct-OR^{WT} transgenic plants; OR^{His}=pAct-OR^{His} transgenic plants.

Table 4.3 Grouped average grain carotenoid content of the T₁ *TaOR^{WT}* and *TaOR^{His}* overexpression plants compared to GRF control plants. Values show the averages of lines from each transgene with standard deviation in brackets. GRF=*GRF4-GIF1* transgenic controls; OR^{WT}=pAct-OR^{WT} transgenic plants; OR^{His}=pAct-OR^{His} transgenic plants.

Transgene	α-carotene	β-carotene	Zeaxanthin	Lutein	Total GCC
OR ^{WT}	0.023 (0.004)	0.014 (0.001)	0.260 (0.013)	0.310 (0.020)	0.608 (0.027)
OR ^{His}	0.022 (0.009)	0.018 (0.005)	0.311 (0.027)	0.388 (0.052)	0.738 (0.064)
GRF	0.023 (0.006)	0.015 (0.003)	0.243 (0.056)	0.326 (0.070)	0.606 (0.120)

To investigate whether the overexpression of *TaOR^{WT}* and *TaOR^{His}* has pleiotropic effects, I analysed grain yield per plant and anthesis date. For statistical analyses, I grouped the lines by their transgenes. For instance, GRF-1C-A, GRF-1C-B, GRF-2C-A, GRF-2C-B, GRF-hiC-A and GRF-hiC-B plants were grouped as 'GRF'. This was done because each group (GRF, OR^{WT} and OR^{His}) contained similar numbers of plants with a similar combination of copy numbers and, therefore, should show the effect of each transgene. The only exception is the group WT, which contained the results of the 8 wild-type Cadenza plants. There was a lot of variation in grain yield per plant, with some lines having very little to no grains produced per plant (Figure 4.10 and Table 4.4). Several sterile GRF, OR^{WT} and OR^{His} plants displayed deformed spike growth defects. The wild-type Cadenza plants had the highest average grain yield per plant (7.18 grams), followed by OR^{WT} plants (4.06 grams), followed by OR^{His} plants (2.48 grams) and then followed by the GRF controls (2.14 grams). An ANOVA found significant differences existed between the transgene groups ($p < 0.001$, $F(3,147) = 14.04$, one-way ANOVA). The *TaOR^{His}* overexpression lines also had an anthesis date significantly delayed by 6–10 days compared to the OR^{WT}, GRF and wild-type lines, as revealed by an ANOVA ($p < 0.001$, $F(3,147) = 10.34$, one-way ANOVA; Figure 4.10 and Table 4.4). Figure 4.10 presents a visual summary of the Tukey Honest Significant Difference test for pairwise comparisons, indicated by significance letters on the boxplots. The tables of specific pairwise comparisons and statistical significance are detailed in Appendix Material 17.

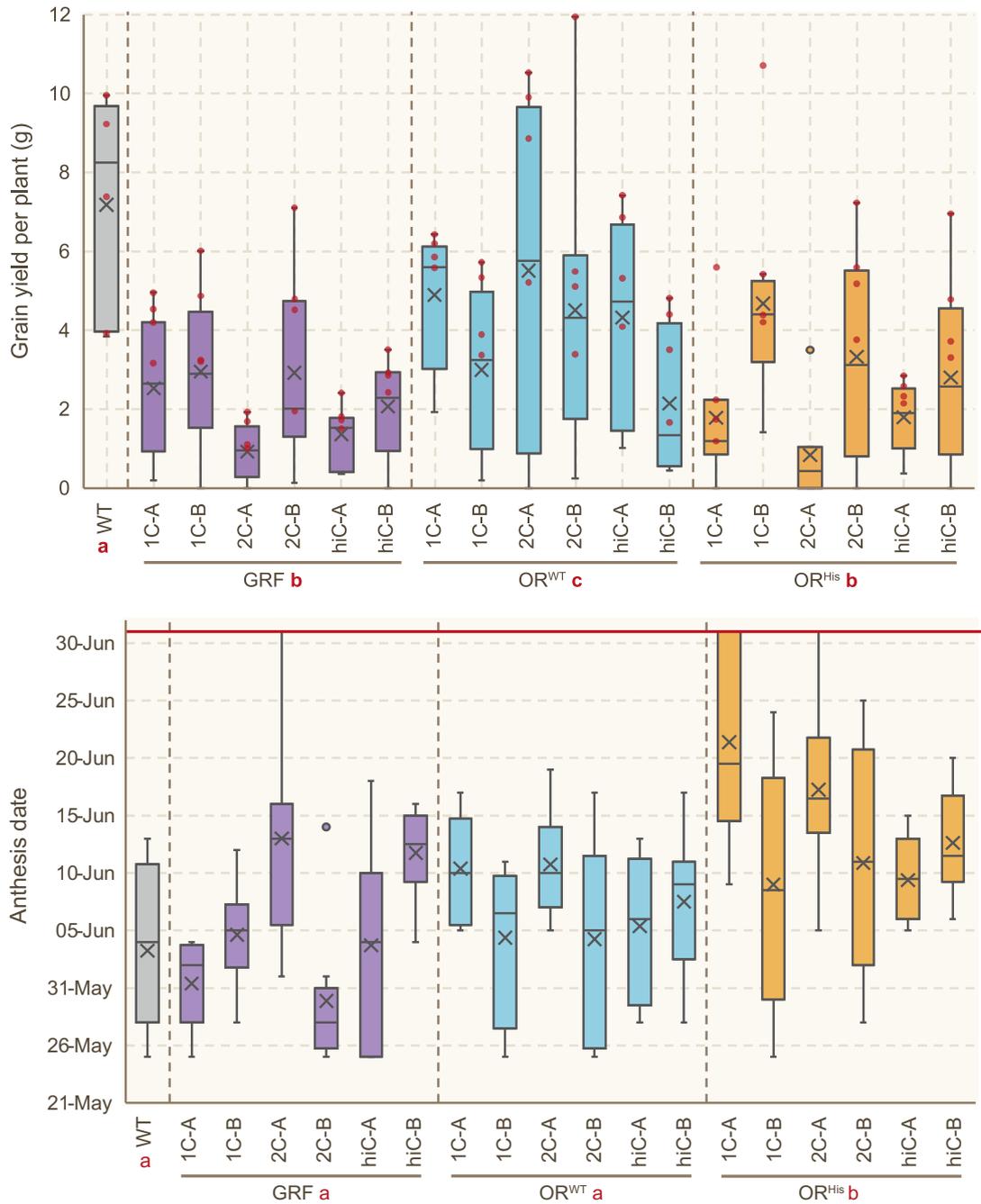


Figure 4.10 Grain yield per plant and anthesis date of T_1 $TaOR^{WT}$ and $TaOR^{His}$ overexpression lines compared to GRF and wild-type controls. Bold red letters next to the groups' names indicate statistical significance between the groups, as determined by a one-way ANOVA. Red dots on the grain yield per plant boxplot show the grain yield of the four plants from each line taken forward for grain morphometric analysis and carotenoid content analysis. The 30th of June was the last day plants were screened for anthesis; plants not yet flowered were recorded as the 1st of July, as shown by the red line on the graph. WT=non-transgenic wild-type controls; GRF= $GRF4-GIF1$ transgenic controls; OR^{WT} = $pAct-OR^{WT}$ transgenic plants; OR^{His} = $pAct-OR^{His}$ transgenic plants.

I also investigated whether the overexpression of $TaOR^{WT}$ and $TaOR^{His}$ influences height, grain morphology and grain number per plant. The averaged results for each transgene are found in Figure 4.11 and Table 4.4. The results of each line are found in Appendix Material 18. An ANOVA on the heights of the GRF, OR^{WT} , OR^{His} and wild-type plants found no significant differences between the transgene groups ($p=0.1$, $F(3,141)=2.12$, one-way

ANOVA), suggesting these lines to have similar plant heights. The GRF plants were found to have significantly fewer grains per plant than the Cadenza wild-type and OR^{WT} plants ($p < 0.001$, $F(3,68) = 7.715$, one-way ANOVA). Due to the large differences in grain yield per plant, the grain of only four plants from each line was measured for TGW and grain area. An ANOVA revealed a significant difference between the TGW of the OR^{His} and GRF plants ($p = 0.035$, $F(3,68) = 3.03$, one-way ANOVA), while the wild-type, OR^{WT} and OR^{His} had comparable TGW. An ANOVA also revealed significant differences existed for grain area between GRF plants and both OR^{WT} and OR^{His} plants ($p < 0.001$, $F(3,68) = 6.198$, one-way ANOVA). The OR protein has also been suggested to be involved in chlorophyll biosynthesis and total chlorophyll (Sun et al. 2023b). To compare leaf chlorophyll at similar stages, the relative amount of leaf chlorophyll (SPAD value) was measured at 2 and 3 weeks after anthesis and averaged. The averaged results for each transgene are found in Figure 4.11 and Table 4.4. The results of each line are found in Appendix Material 19. The *TaOR^{His}* overexpression plants had a slightly lower relative amount of leaf chlorophyll than the OR^{WT}, GRF and wild-type lines, as revealed by an ANOVA ($p = 0.006$, $F(3,142) = 4.35$, one-way ANOVA). The OR^{WT}, GRF and wild-type lines had comparable relative chlorophyll contents. Figure 4.11 presents a visual summary of the Tukey Honest Significant Difference test for pairwise comparisons, indicated by significance letters on the boxplots. Tables of specific pairwise comparisons and statistical significance are detailed in Appendix Material 20.

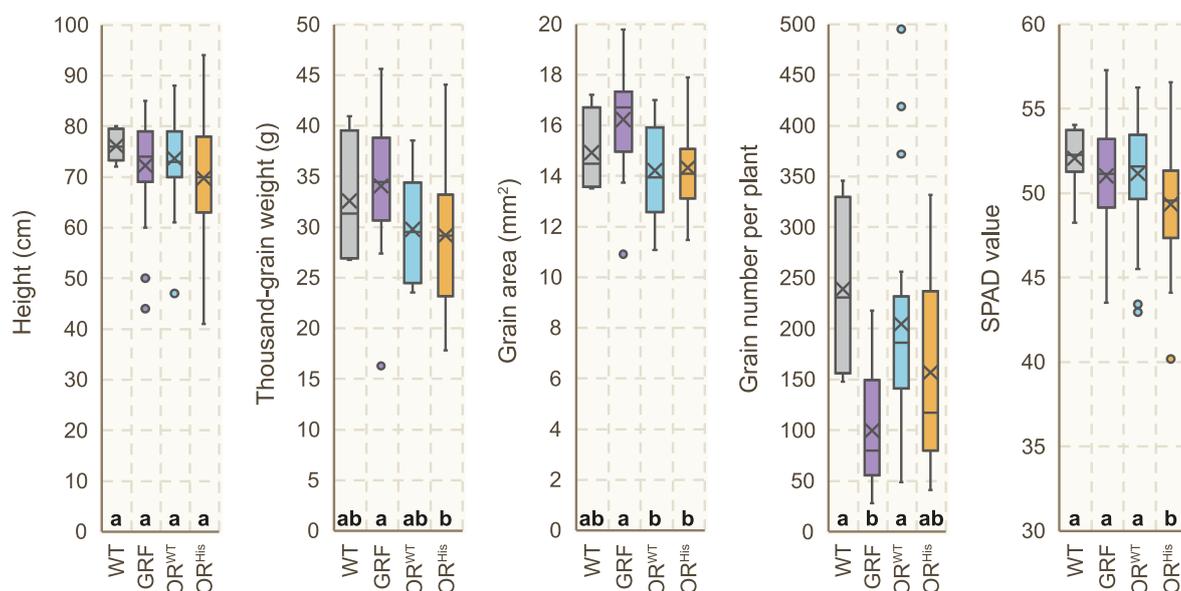


Figure 4.11 Height, thousand-grain weight, grain area, grain number per plant and relative leaf chlorophyll content of T₁ *TaOR^{WT}* and *TaOR^{His}* overexpression lines. SPAD measurements are the average of two measurements made 2 and 3 weeks after anthesis. Bold letters next to the groups' names indicate statistical significance between the groups as determined by a one-way ANOVA. WT=non-transgenic wild-type controls; GRF=*GRF4-GIF1* transgenic controls; OR^{WT}=pAct-OR^{WT} transgenic plants; OR^{His}=pAct-OR^{His} transgenic plants.

Table 4.4 Grouped averages of other phenotypes of the T₁ *TaOR*^{WT} and *TaOR*^{His} overexpression plants compared to the GRF and wild-type controls. Standard deviation values are in brackets. The SPAD value is a measure of relative carotenoid content. TGW=thousand-grain weight, WT=non-transgenic control, GRF=*GRF4-GIF1* transgenic controls; OR^{WT}=pAct-OR^{WT} transgenic plants; OR^{His}=pAct-OR^{His} transgenic plants.

Transgene	Grain yield/plant (g)	Anthesis date	Height (cm)	TGW (g)	Grain area (mm ²)	Grain number/plant	SPAD value
WT	7.18 (2.78)	03-June (7.13 days)	76.1 (3.1)	32.57 (6.79)	14.92 (1.69)	239.0 (91.4)	52.06 (1.89)
GRF	2.14 (1.67)	05-June (8.20 days)	72.2 (8.2)	34.05 (5.90)	16.23 (1.83)	99.5 (55.1)	51.03 (2.85)
OR ^{WT}	4.06 (2.90)	07-June (6.59 days)	73.7 (7.6)	29.77 (5.13)	14.22 (1.81)	204.7 (99.4)	51.16 (2.86)
OR ^{His}	2.48 (2.33)	13-June (8.74 days)	69.8 (10.8)	29.17 (7.08)	14.33 (1.74)	156.9 (88.7)	49.33 (3.20)

4.2.4 There is very low allelic diversity within *OR* in wheat

The protein sequences of the *TaOR* homoeologues share a very high sequence identity (Section 4.2.1). Additionally, the sequence is identical between *TaOR* and *TdOR* for the A-genome copies and the B-genome copies. This suggests there is very low diversity in this gene within wheat. However, the sequences I investigated of the *TaOR* and *TdOR* homoeologues were just from single cultivars (Chinese Spring and Svevo). Therefore, this does not tell us about the allelic diversity of *OR* present within the germplasm of *T. aestivum* or *T. turgidum* ssp. *durum*.

To investigate the diversity of the *TaOR* homoeologues present in different accessions of *T. aestivum*, I ran a BLASTp search on EnsemblPlants using the *TaOR* sequence from the IWGSC Chinese Spring reference genome (IWGSC et al. 2018) against the available bread wheat pangenome (Walkowiak et al. 2020). This identified *TaOR* protein sequences from 11 accessions: ArinaLrFor, Cadenza, Jagger, Julius, Lancer, Landmark, Mace, Norin 61, Renan, Stanley and Sy Mattis. The protein sequences of the *TaOR* homoeologues within the pangenome are highly conserved (Appendix Material 21). The gene annotation identified an alternative start site for two of the 11 pangenome cultivars for *TaOR-6A*, one of the 11 for *TaOR-6B* and one of the 11 for *TaOR-6D*. In each case, the original methionine start site sequence was present, and the sequences between these two methionines were identical; therefore, this is likely due to differences in the gene annotation of these pangenomes. Additionally, for *TaOR-6A*, there was a deletion of a serine residue in a repeat of five serine residues in four of the 11 cultivars. This happens near the start of the gene and is not in a highly conserved region of the OR protein. The rest of the *TaOR* protein sequences from the 11 pangenomes and the Chinese Spring sequence were identical.

These pangenome sequences are from a small number of relatively modern bread wheat cultivars. I also searched for diversity within the Watkins global landrace collection, which is the most highly diverse sequenced panel of wheat accessions to date (Cheng et al. 2023). These accessions were previously sequenced, through a *k*-mer-based approach, by Professor Shifeng Cheng at the Chinese Academy of Agricultural Sciences in collaboration with the John Innes Centre. I searched for SNP diversity in the *TaOR* and *TdOR* homoeologues in these collections. The Watkins hexaploid SNP diversity contained information on 827 Watkins hexaploid landraces and 224 hexaploid cultivars. The Watkins tetraploid SNP diversity contained information on 327 Watkins landraces of durum wheat. I found very low diversity in both the *TaOR* and *TdOR* homoeologues within these collections. Only two non-synonymous variants were found, both missense substitutions in *TaOR-6A* and *TaOR-6D* in the Watkins hexaploid collection. One missense variant was found in exon 1 of *TaOR-6A* at residue 72, which resulted in an alanine-to-threonine substitution and was present in one out of 1046 accessions. The other missense variant was found in exon 1 of *TaOR-6D* at residue 23, which resulted in an alanine-to-valine substitution and was present in seven out of 1041 accessions. Both substitutions were predicted to be tolerated by the Variant Effect Predictor tool on EnsemblPlants (*TaOR-6A* A72T SIFT=0.61, *TaOR-6D* A23V SIFT=0.51), suggesting these were not in a region of conserved homology and likely do not affect protein function.

To compare this to the diversity of another gene, I also investigated the diversity within *PSY1* and *PSY2* in the Watkins global landrace collection. The length of *TaPSY1* and *TaPSY2* proteins is similar to the *TaOR* proteins (for instance, *TaPSY1-7A* is 284 amino acids long, *TaPSY2-5A* is 396 amino acids long, and *TaOR* proteins are 324–327 amino acids long). Therefore, it is expected that these proteins should have a similar natural mutation rate. For *PSY1*, there were 22 non-synonymous variants within the Watkins hexaploid collection (*TaPSY1-7A*=5, *TaPSY1-7B*=9, *TaPSY1-7D*=8) and 11 within the Watkins tetraploid collection (*TdPSY1-0U*=8, *TdPSY1-6B*=3). For *PSY2*, there were 16 non-synonymous variants within the Watkins hexaploid collection (*TaPSY2-5A*=6, *TaPSY2-5B*=4, *TaPSY2-5D*=6) and 7 within the Watkins tetraploid collection (*TdPSY2-5A*=1, *TdPSY2-5B*=6). Compared to only two non-synonymous *OR* variants within the Watkins global landrace collection, this underscores the lack of diversity within wheat *OR*.

4.3 Discussion

4.3.1 *OR*^{WT} plays a role in grain carotenoid biosynthesis in wheat

One aim of this chapter was to investigate the role that *TaOR* plays in wheat through an analysis of *Tdor* knockout lines in the tetraploid variety Kronos and the overexpression of *TaOR*^{WT} in Cadenza and Kronos. Consistent with knocking out *OR* in *Arabidopsis* and melons (Zhou et al. 2015; Chayut et al. 2017; Sun et al. 2023b), the *Tdor* mutant lines had 33.8% lower GCC than their wild-type controls, suggesting *OR* plays a role in grain carotenoid biosynthesis in wheat. In *Arabidopsis* and melon *OR* mutants, it was found that the decrease in carotenoid content is associated with decreasing PSY protein levels. Additionally, the sweet potato IbOR was shown to play a key role in regulating PSY stability in sweet potatoes and leading to carotenoid accumulation (Park et al. 2016). Welsch and colleagues (2018) hypothesised that the formation of a membrane complex with OR produces the active form of PSY, while Clp proteases degrade non-associated misfolded PSY. Presumably, OR plays the same role in stabilising or activating PSY in wheat, and knocking out *TdOR* reduces the protein levels of TdPSY, the rate-limiting enzyme within the wheat carotenoid biosynthesis pathway (He et al. 2008). A very high protein sequence conservation between *TaOR* homoeologues and other *OR* orthologues (72.8–73.9% similarity with CmOR) also supports *OR* playing a similar role in wheat and its function being conserved among plants. Future work looking at the protein levels of PSY within the grain of *Tdor* mutants would help elucidate whether this is the cause of the reduced GCC.

When flux into the carotenoid biosynthesis pathway was increased in wheat grains through the overexpression of *CRTB*, which plays the same role as PSY converting GGPP to phytoene, the PVA carotenoids had the greatest percentage increase compared to other non-PVA carotenoids (Wang et al. 2014). This led to the transgenic lines having a PVA carotenoid proportion of 26–80% versus 6% in the non-transgenic control. Similarly, Zeng and colleagues (2015) found a 16-fold increase in PVA carotenoids compared to a 6-fold increase in total carotenoids when *CRTB* was overexpressed within wheat grains. This suggests that affecting flux into the biosynthesis pathway has the strongest effect on the PVA carotenoids, which are found early in the pathway. Equally, in this study, the PVA carotenoids α -carotene and β -carotene had the greatest percentage reduction in the *Tdor* mutants compared to the wild-type controls (61.0% and 50.7%, respectively). If OR acts to stabilise PSY in wheat, then knocking *OR* out would reduce flux into the carotenoid biosynthesis pathway, consistent with the idea that variations in flux have the greatest impact on PVA carotenoids.

Interestingly, the reduction of zeaxanthin in the *Tdor* mutants was relatively small compared to the reductions of other carotenoids (Figure 4.4). This could be due to the tighter regulation of zeaxanthin content, given its important role in the xanthophyll cycle, which protects cells during intense light and high temperatures (Jahns and Holzwarth 2012; Dhami and Cazzonelli 2020).

The *TaOR^{WT}* overexpression lines did not show an increased GCC compared to the GRF controls, nor was any effect on GCC observed with an increased number of *TaOR^{WT}* transgenes. However, these overexpression lines have not been validated for correct *TaOR^{WT}* overexpression using quantitative PCR, but future work is ongoing to investigate this (March 2024). Despite this, based on previous wheat overexpression analyses using similar constructs and promoters, it is assumed that the constructs are expressing their transgenes correctly. This suggests that PSY protein level is not limited by the amount of endogenous OR within the wheat grains, so boosting OR protein level further does not affect PSY activity. This is consistent with *OsOR^{WT}* overexpression in rice, where no change in GCC was found, suggesting that endogenous OR does not limit the PSY protein level here (Yu et al. 2021). Instead, grain carotenoid biosynthesis is limited by the level of *PSY* expression in both wheat and rice (Qin et al. 2016; Zhou et al. 2022). Another study in rice found no change in GCC when just *AtOR^{WT}* was overexpressed; however, when combined with *CRTI* and *PSY* overexpression, it increased the GCC more than *CRTI* and *PSY* overexpression alone (Bai et al. 2016). Therefore, *OR^{WT}* overexpression can help boost GCC in combination with *PSY* overexpression, likely by stabilising and increasing the overexpressed PSY protein level. A similar approach could be used in wheat to further boost GCC than just with *PSY* overexpression alone. In contrast, the overexpression of *AtOR^{WT}* in a white maize variety that normally accumulates only trace amounts of carotenoids significantly increased the GCC here (Berman et al. 2017). This suggests that in this variety, PSY stability and post-transcriptional regulation by OR is the rate-limiting step for the carotenoid biosynthesis pathway rather here than the expression of *PSY*. In the *Tdor* mutants, grain carotenoid biosynthesis is presumably limited in a similar fashion, where knocking out endogenous *OR* reduces PSY level and acts as a rate-limiting step.

The effect on carotenoid content of knocking out *Tdor* or overexpressing *TaOR^{WT}* in other tissues could also be investigated. *TaOR* expression is higher within the green photosynthetic tissues of leaves and shoots than in the grain (Section 4.2.1), and the role it plays here would be of interest. In *Arabidopsis*, knocking out *Ator* and *Atorlike* together

reduced leaf carotenoid content, suggesting these proteins may play a role in stabilising PSY here (Zhou et al. 2015; Sun et al. 2023b). However, knocking out *Ator* or *Atorlike* alone did not affect leaf carotenoid content, suggesting AtORLIKE also plays the same role as AtOR. Similarly, Yu and colleagues (2021) found that *Osor* mutants did not affect leaf carotenoid content, which could be because OsORLIKE also compensates for OsOR. A question yet to be addressed is whether wheat ORLIKE plays a similar role, and if so, whether a *Tdor* *Tdorlike* mutant would have a greater reduction of GCC. Similar to *TaOR^{WT}* overexpression in wheat grains, overexpression of *AtOR^{WT}* in *Arabidopsis* and *OsOR^{WT}* in rice did not increase carotenoid content in leaves (Zhou et al. 2015; Jung et al. 2021; Sun et al. 2023b). This could be because carotenoids rarely overaccumulate in plant leaves, unlike in non-photosynthetic tissues such as fruits and roots, instead, they are continuously synthesised and degraded to maintain optimal photosynthesis (Dhami and Cazzonelli 2020). Alternatively, PSY activity may not be limited by the action of OR stabilisation.

4.3.2 *OR^{His}* increases grain carotenoid content in wheat

Through the overexpression of *TaOR^{His}*, I aimed to understand whether the ‘golden SNP’ arginine-to-histidine substitution would affect GCC. The overexpression of *AtOR^{His}*, *SbOR^{His}* and *IbOR^{His}*, with arginine-to-histidine substitutions at the equivalent ‘golden SNP’ site, resulted in higher total carotenoid levels than *OR^{WT}* overexpression in non-photosynthetic tissue (Yuan et al. 2015; Yazdani et al. 2019; Sun et al. 2020; Kim et al. 2021). Consistent with this, I found *TaOR^{His}* overexpression increased the GCC by 21.6% more than the *TaOR^{WT}* overexpression lines and GRF controls, suggesting that the ‘golden SNP’ plays a role in wheat carotenoid accumulation in the grain. However, the variation in the GCC of the GRF controls is still quite high; therefore, this should be repeated on more lines with greater replication. This increase in carotenoid content was smaller than in these other studies, possibly due to differences in how *TaOR^{His}* acts in grain tissues compared to the other non-photosynthetic tissues. This comparatively small increase in carotenoids might be why no visible colour difference in dark-grown *TaOR^{His}* overexpressing callus was observed (Figure 4.7), as the small difference may only be observable with a high-resolution technique like HPLC. The only other instance of *OR^{His}* overexpression in a monocot grain was in rice, which did not lead to an increase in GCC in contrast to my results (Jung et al. 2021). However, unlike in wheat, there is no active carotenoid biosynthesis in rice due to no *PSY* expression, so increasing sink tissue would not affect GCC (Beyer et al. 2002).

The 'golden SNP' arginine-to-histidine mutation in *OR* is thought to promote carotenoid accumulation by activating chromoplast differentiation in non-green tissue, producing a single large chromoplast here (Tzuri et al. 2015; Yuan et al. 2015; Yazdani et al. 2019; Sun et al. 2020). This increases the number of carotenoid-sequestering structures, creating a greater metabolic sink that enhances carotenoid accumulation (Li and Van Eck 2007; Sun et al. 2018). The mechanism of how *OR^{His}* promotes chromoplast differentiation is still not fully elucidated. Sun and colleagues (2020) showed *AtOR^{His}* interacts with ACCUMULATION AND REPLICATION OF CHLOROPLASTS 3 (*ARC3*), a crucial regulator of chloroplast division, whereas *AtOR^{WT}* does not. This interaction interferes with the binding of *ARC3* to PARALOG OF *ARC 6* (*PARC6*), another crucial regulator of chloroplast division, and they suggested that this results in the formation of a single large chromoplast in non-photosynthetic tissue. Alternatively, it has been proposed that *IbOR^{His}* increases the expression of carotenoid biosynthesis genes in sweet potatoes, increasing carotenoid accumulation (Kim et al. 2019). Additionally, *CmOR^{His}* in melons was suggested to stabilise β -carotene by inhibiting its degradation by *HYD* enzymes (Chayut et al. 2017). In these cases, *OR^{His}* might act to increase carotenoid accumulation above a threshold required to stimulate plastid differentiation and chromoplast biogenesis (Bai et al. 2014; Sun et al. 2018; Welsch et al. 2018). Analysing the expression of these carotenoid biosynthesis genes within the *TaOR^{His}* overexpressing wheat grains could help elucidate whether a similar mechanism is responsible here.

It is unclear whether the promotion of chromoplast biogenesis is responsible for the increase in carotenoid content found in my *TaOR^{His}* overexpression lines. In previous cases of *OR^{His}* overexpression, β -carotene was the primary carotenoid increased, which chromoplasts are adept at storing (Sun et al. 2018). However, for the *TaOR^{His}* overexpression lines, the main increase was found for lutein and zeaxanthin (Table 4.3), with β -carotene showing only a minor non-significant increase. In the starchy endosperm of staple crops like wheat, amyloplasts are the main plastid that synthesises and accumulates carotenoids, primarily as the xanthophylls lutein, zeaxanthin and violaxanthin (Howitt and Pogson 2006; Wurtzel et al. 2012). Therefore, it could be that *TaOR^{His}* overexpression in wheat grain increases carotenoid storage as xanthophylls in amyloplasts by producing more carotenoid-sequestering plastoglobuli here instead of promoting chromoplast biogenesis. However, in starchy potatoes containing amyloplasts, the overexpression of the cauliflower orange-inflorescence mutant gene (*BoOr^{Mut}*) was found to initiate chromoplast formation within the tuber, increasing β -carotene content, whereas high carotenoid potato varieties do not typically form chromoplasts here (Lopez et al. 2008). Carotenoids within amyloplasts are

prone to degradation during the final stages of maturation and post-harvest storage (Farré et al. 2013; De Moura et al. 2015; Che et al. 2016; Schaub et al. 2017). It has been suggested that using *OR^{His}* to promote chromoplast biogenesis in staple crops could enhance storage stability due to the specific characteristics of chromoplasts being better at storing carotenoids (Li et al. 2012). Consequently, whether *TaOR^{His}* overexpression promotes chromoplast biogenesis within wheat grains is of specific interest to carotenoid biofortification in cereals, and future work should investigate the types of plastids in these developing grains.

A crucial limiting factor for high carotenoid production within starchy tissue like wheat grains is the low transcription and activity of key enzymes in the carotenoid biosynthesis pathway, especially *PSY* (Vallabhaneni and Wurtzel 2009; Bai et al. 2016). Low *PSY* gene expression is directly associated with the carotenoid content within the starchy organs of wheat grains (Rodríguez-Suárez et al. 2014; Flowerika et al. 2016; Qin et al. 2016; Vargas et al. 2016), and overexpressing *PSY*, or bacterial phytoene synthase *CRTB*, results in large increases in wheat GCC in a ‘push’ carotenoid biofortification strategy (Cong et al. 2009; Wang et al. 2014). Carotenoid degradation in amyloplasts is also responsible for the low carotenoid levels within starchy tissues (Schaub et al. 2017). Increased wheat GCC was achieved in a ‘block’ carotenoid biofortification strategy by knocking out the β -CAROTENE HYDROXYLASE genes (*HYD1* and *HYD2*), which degrade β -carotene into β -cryptoxanthin (Garcia Molina et al. 2021; Yu et al. 2022; Bekkering et al. 2023). It would be very interesting to combine *PSY* overexpression and *HYD* knockouts with *OR^{His}* overexpression, combining ‘push’, ‘block’ and ‘pull’ strategies to increase carotenoid content. ‘Push’ and ‘block’ strategies have been combined in wheat previously through the overexpression of a *CRTB* and RNAi silencing of *HYD1* to both increase flux into carotenoid biosynthesis and reduce the conversion of β -carotene into β -cryptoxanthin (Zeng et al. 2015). Combining these approaches led to a 31-fold increase in grain β -carotene content compared with an increase of 14.6-fold when just *CRTB* was overexpressed or 10.5-fold when just *HYD1* was silenced.

A limitation of the results presented here analysing the GCC of *Tdor* mutants and *TaOR* overexpression lines is that these have come from pools of grain from multiple plants rather than treating each plant as a biological replicate due to the limited grain yield per plant for both the *Tdor* mutant lines and the *TaOR* overexpression lines. Future work should bulk up the next generation of F₃ *Tdor* lines and T₂ *TaOR* overexpression lines to confirm these results. Additionally, the *Tdor* mutants are a cross of TILLING lines that contain many

background mutations that may be responsible for the reduced GCC. However, these F₂ plants are segregants selected from F₁ heterozygous lines based on their *TdOR* genotype, so the mutations in other genes between the different *Tdor* F₂ mutant and *Tdor* F₂ wild-type plants would have segregated randomly (unless linked to *TdOR*). Nevertheless, backcrossing the mutant *Tdor* lines to Kronos would help reduce the number of background mutations. For the *TaOR^{WT}* and *TaOR^{His}* overexpression lines, many hemizygous lines were selected (Figure 2.3), which could have differences in the location of their T-DNA insertion and, therefore, expression between individual plants. Since each of these lines represents a different integration of the *TaOR^{WT}* and *TaOR^{His}* overexpressing transgenes, these can be viewed as different biological replicates. Therefore, the increase in GCC consistently observed in the *TaOR^{His}* overexpression lines compared to *TaOR^{WT}* overexpression lines supports the ‘golden SNP’ producing this increased GCC in wheat (Figure 4.9). Moreover, there was no additive increase in carotenoid content between the different copy numbers of the *TaOR^{WT}* and *TaOR^{His}* lines, the latter consistent with the ‘golden SNP’ being a dominant gain-of-function mutation (Tzuri et al. 2015). This also suggests that installing the ‘golden SNP’ into just one of the endogenous *OR* homoeologues is hopefully enough to give a measurable increase in carotenoid content.

4.3.3 *GRF4-GIF1*, *TaOR^{WT}* and *TaOR^{His}* overexpression affects grain and physiology in wheat

The Cadenza T₁ *OR^{WT}*, *OR^{His}*, and GRF plants had a significantly lower grain yield per plant than the Cadenza non-transgenic control (Figure 4.10), and most Kronos T₀ pAct-*OR^{WT}* and pAct-*OR^{His}* plants were sterile or produced below 10 grains. These problems are most likely due to the *GRF4-GIF1* developmental regulators included in these constructs to boost regeneration efficiency and allow for the transformation of unamenable wheat varieties like Kronos and Cadenza. In plants, developmental regulator genes control the growth of multiple tissues and organs, and their overexpression in *Arabidopsis*, rice, potatoes and maize has produced varying impacts on plant development and physiology (Liebsch and Palatnik 2020). Debernardi and colleagues (2020) first demonstrated that *GRF4-GIF1* improves regeneration during wheat transformation; however, they also found that this affected grain morphology and number. Moreover, other labs using the *GRF4-GIF1* developmental regulators have also found issues with plant sterility in Kronos (Andy Chen, personal communications). Interestingly, I also found that the GRF controls, only overexpressing *GRF4-GIF1*, had a reduced GCC compared to the Cadenza wild-type controls,

which had not been previously demonstrated (Figure 4.9). These pleiotropic phenotypes associated with *GRF4-GIF1* overexpression made it slightly more challenging to investigate the difference between plants overexpressing *TaOR^{WT}*, *TaOR^{His}*, or not expressing these at all. For this reason, I included the GRF controls to compare with the *TaOR^{WT}* and *TaOR^{His}* overexpressing plants. Future studies on the role *TaOR^{WT}* and *TaOR^{His}* overexpression plays in wheat could not include these developmental regulators on the transformation constructs. Alternatively, there is ongoing work to develop systems to express *GRF4-GIF1* only within wheat calli during the transformation procedure. This includes using calli-specific promoters, a heat-shock inducible Cre-Lox recombination system to remove the *GRF4-GIF1* transgenes (Harrington et al. 2020) or a chemical inducible promoter system (Mark Smedley and Tom Lawrenson, personal communications).

GRF4-GIF1 overexpression was previously found to significantly lower the number of grains per spike and increase grain area, as well as non-significantly increase TGW (Debernardi et al. 2020). Similarly, my GRF controls had a significantly lower grain number per plant compared to the non-transgenic control plants and a higher but non-significant increase in grain area and TGW (Figure 4.11 and Table 4.4). This supports the finding that *GRF4-GIF1* overexpression impacts grain number and morphology. Interestingly, however, the *TaOR^{WT}* overexpression plants had a significantly higher number of grains per plant than the GRF controls (Figure 4.11 and Table 4.4), suggesting *TaOR^{WT}* overexpression in wheat can reverse the effect of *GRF4-GIF1* overexpression on grain number. This also suggests that *TaOR^{WT}* overexpression may increase grain number. A similar result was found in tomatoes where the overexpression of *AtOR^{WT}* significantly enhanced fruit set and increased seed number compared to non-transgenic controls (Yazdani et al. 2019). This suggests that *TaOR^{WT}* overexpression may have a similar effect in wheat, enhancing grain number; however, the use of *GRF4-GIF1* developmental genes makes it difficult to investigate fully in these lines. There also seems to be a compensatory effect associated with the increased grain number, with the *TaOR^{WT}* plants showing decreased grain area and TGW compared to the GRF controls (Figure 4.11 and Table 4.4). Interestingly, grain yield per plant was significantly higher for the *TaOR^{WT}* plants compared to the GRF controls, which might be due to the increased grain number per plant (Figure 4.10). Future work should investigate the grain number per plant, grain yield per plant and grain morphology of plants overexpressing *TaOR^{WT}* without *GRF4-GIF1*.

The *TaOR^{His}* plants had a later anthesis date than the wild-type, GRF and *TaOR^{WT}* plants (Figure 4.10). This could be due to the increased carotenoid content associated with *TaOR^{His}* overexpression impacting levels of abscisic acid (ABA) produced from the β - β branch of the carotenoid biosynthesis pathway (Figure 1.1), and these plants had the greatest increase in the β - β branch carotenoid zeaxanthin (Figure 4.8; Table 4.3). ABA application was found to delay flowering transition in *Arabidopsis* (Barrero et al. 2005; Domagalska et al. 2010; Wang et al. 2013), and hypersensitive ABA mutants in wheat have significant delays in flowering time (Schramm et al. 2013). However, the role of ABA in flowering is complex, and its effect on flowering is suggested to be differentially modulated under short and long days (Domagalska et al. 2010; Riboni et al. 2013). In contrast to my overexpression lines, the overexpression of both *AtOR^{WT}* and *AtOR^{His}* was found to promote early flowering in tomatoes by altering the expression of genes involved in flowering (Yazdani et al. 2019).

OR also stabilises MAGNESIUM CHELATASE SUBUNIT I (CHLI) in the chlorophyll biosynthesis pathway (Sun et al. 2023b). CHLI is a subunit of MAGNESIUM-CHELATASE (MgCh), which catalyses the first committed step in chlorophyll biosynthesis (Tanaka and Tanaka 2007; Wang and Grimm 2021). Consequently, *Ator Atorlike* mutants in *Arabidopsis* have reduced leaf chlorophyll content (Zhou et al. 2015; Sun et al. 2023b). In wheat, the *TaOR* homoeologues had the greatest expression within the photosynthetic tissue of the leaves, stem, and spike (Section 4.2.1); therefore, TaOR may play a similar role within this tissue. I did not investigate the relative leaf chlorophyll content within the F₂ *Tdor* mutants, and future work should do this to see if TdOR might play a similar role here. For my overexpression lines, there was no significant difference in relative chlorophyll content between *TaOR^{WT}* plants and the GRF and wild-type plants (Figure 4.11 and Table 4.4), consistent with *AtOR* overexpression in *Arabidopsis* (Zhou et al. 2015; Sun et al. 2023b). However, the *TaOR^{His}* overexpression lines were found to have a lower relative chlorophyll content than *TaOR^{WT}*, GRF or wild-type plants (Figure 4.11 and Table 4.4). This may be a pleiotropic result of these plants having delayed anthesis. In contrast, rice plants overexpressing *OsOR^{His}* had no difference in chlorophyll content to *OsOR^{WT}* and non-transgenic control lines under normal conditions (Jung et al. 2021).

The OR protein has holdase chaperone activity and stabilises PSY under heat stress (Park et al. 2016). Among the subunits of MgCh, CHLI has been found to be the most vulnerable under heat stress (Rocco et al. 2013; Echevarría-Zomeño et al. 2016). In *Arabidopsis*, *AtOR* was also shown to enhance the stability of CHLI under heat stress, safeguarding

photosynthetic pigment biosynthesis and enhancing thermotolerance in *Arabidopsis* and tomatoes (Sun et al. 2023b). Similarly, the overexpression of *OsOR^{WT}* and *OsOR^{His}* in rice led to significantly higher leaf chlorophyll content under heat stress than non-transgenic controls (Jung et al. 2021). *OR^{WT}* overexpression in sweet potato, *Arabidopsis*, alfalfa and potatoes also maintained higher chlorophyll content and photosystem II efficiency under abiotic stress conditions than wild-type plants (Kim et al. 2013, 2021; Park et al. 2015, 2016; Wang et al. 2015, 2018; Cho et al. 2016). Therefore, the overexpression of *OR^{WT}* within crops could play a potential role in improving the thermotolerance of crops to address the challenges climate change brings by increased heatwave frequencies. Consequently, it would be very interesting to compare the relative chlorophyll content and grain yield per plant of *TaOR^{WT}* and *TaOR^{His}* overexpression lines in wheat under heat stress. Due to the pleiotropic effects *GRF4-GIF1* overexpression brings, it would be best to test this role without these developmental regulators. If this maintains chlorophyll content under heat stress, *OR* overexpression could be combined with HB4 wheat, the first approved genetically modified drought-tolerant wheat (Gupta 2023), creating a wheat variety resilient to heat and drought stress.

4.3.4 Engineered strategies must be used for *OR* carotenoid biofortification

In addition to the high conservation in *OR* orthologues, limited allelic diversity was found within the Watkins global landrace collection, the most highly diverse sequenced collection available. This suggests that *OR* is under strong selective pressure, with variants potentially being disadvantageous and selected against through natural or artificial selection (Doebley et al. 2006). As discussed, *OR* plays a role in stabilising both *PSY* and *CHL1* in plants, integrating and regulating the chlorophyll and carotenoid biosynthesis pathways (Sun et al. 2023b). As such, it has a highly adapted role with loss-of-function mutations likely severely affecting the photosynthetic potential of plants, and this could be the reason for this low allelic diversity observed within wheat *OR*. Within the Watkins global landrace collection, allelic diversity was much greater within the *PSY* genes, suggesting this as a potential route for carotenoid biofortification. However, for a carotenoid biofortification strategy engineering variation into endogenous *OR* homoeologues, such as by utilising *OR* mutations such as the 'golden SNP', or ectopically expressing *OR^{His}* is likely the only way.

5 Engineering and inducing variation into the *ORANGE* gene in wheat

5.1 Introduction

The *ORANGE* (*OR*) protein plays a role in carotenoid biosynthesis, stabilising PHYTOENE SYNTHASE (*PSY*) and increasing its protein level (Zhou et al. 2015; Sun et al. 2023b). In addition, dominant gain-of-function mutations within *OR* have been found that increase carotenoid accumulation within non-photosynthetic tissues through a mechanism thought to be separate from its interaction with *PSY* (Li et al. 2001; Tzuri et al. 2015; Zhou et al. 2015). One such mutation is the ‘golden SNP’, discovered to be responsible for a massive accumulation of carotenoids in melon fruit mesocarp, distinguishing orange-fleshed melons from white- or green-fleshed melons (Tzuri et al. 2015). This is a G-to-A substitution (CGC-to-CAC) that produces a histidine substitution (*OR^{His}*) at a conserved arginine residue in the *OR* protein that causes chromoplast biogenesis and increased carotenoid accumulation in non-photosynthetic tissues (Yuan et al. 2015; Yazdani et al. 2019). It has been suggested that installing the ‘golden SNP’ into the *OR* gene of staple crops could promote chromoplast biogenesis in non-photosynthetic starchy tissues like grains, thereby enhancing carotenoid accumulation and storage stability within these (Li et al. 2012; Sun et al. 2018; Osorio 2019; Watkins and Pogson 2020). Wheat grains possess active carotenoid metabolic flux, and the suggested mechanisms of the ‘golden SNP’ propose a way to boost grain carotenoid content (GCC) and enhance grain carotenoid stability. I previously found that the overexpression of *TaOR^{His}* increased the GCC of wheat compared to *TaOR^{WT}* overexpression (Section 4.2.3), suggesting that the ‘golden SNP’ in the endogenous wheat *OR* gene would boost the carotenoid content of wheat compared to the wild-type sequence.

Since the ‘golden SNP’ is a C-to-T transition on the antisense strand, one possible way to introduce it into *OR* would be using a cytosine base editor. Base editing involves a denatured Cas nuclease unable to generate double-strand breaks fused to a nucleobase deaminase enzyme (Anzalone et al. 2020). The Cas nuclease brings the enzyme to a sequence specified by a corresponding single-guide RNA (sgRNA); upon binding, the hybridisation of the sgRNA to the target DNA strand causes the formation of a single-stranded DNA R-loop on the opposite DNA strand. Nucleotides within this region are then accessible to the deaminase domain of the base editor. A cytosine deaminase converts cytosines within this R-loop to

uracils, and the cytosine deaminases APOBEC1 and APOBEC3A have been demonstrated to work in wheat (Zong et al. 2017, 2018; Zhang et al. 2019). However, this strategy relies on a suitable PAM site to put the editing window of the cytosine deaminase above the correct target cytosine nucleotide. This limits the suitability of some genomic sites for editing with base editors. Moreover, other suitable nucleotides within this editing window can be deaminated, leading to unintended bystander editing events at some sites.

Another possible method to introduce the 'golden SNP' into endogenous *OR* is to use prime editing. This is a gene editing technology that can introduce all types of point mutations and small insertions or deletions in a precise and targeted manner (Anzalone et al. 2019). Prime editors are a fusion protein between a Cas9 nickase (nCas9) and an engineered reverse transcriptase targeted to the editing site by a prime editing guide RNA (pegRNA) (Anzalone et al. 2020). The pegRNA guides the prime editor to the target site and has a 3' extension containing a primer binding sequence (PBS) that anneals to the nicked 3' strand, and a reverse transcription template (RT-template) that encodes the desired edit to be installed. Upon binding and nicking of the target site by nCas9, the pegRNA extension template is incorporated into the newly liberated 3' end of the nicked DNA strand through reverse transcription. Prime editing does not have the same issues with PAM availability as base editing; however, lower efficiencies can make this technology more difficult to implement. Prime editing has previously been demonstrated to work in wheat, and improvements have increased editing efficiencies by using a dual prime editing guide RNA system where two pegRNAs are used to install the same edit (Lin et al. 2020, 2021; Awan et al. 2022a; Zong et al. 2022; Ni et al. 2023).

Other mutations that increase the carotenoid content of non-photosynthetic tissue have been found within *OR* (Li et al. 2001; Ellison et al. 2018). These provide other avenues to modify the endogenous *OR* gene within wheat. In cauliflower (*Brassica oleracea* var. *botrytis*), a naturally occurring dominant mutation within *BoOR* (*BoOr^{Mut}*) activates chromoplast biogenesis within the cauliflower curd leading to an overaccumulation of β -carotene in this non-photosynthetic tissue (Lu et al. 2006). The causal mutation was identified as a retrotransposon insertion in the third exon of *BoOR*, which disrupts normal splicing of its transcripts leading to various aberrant in-frame transcripts with amino acid insertions or deletions (Lu et al. 2006).

Parallel studies in rice by Endo and colleagues (2019) and Kim and colleagues (2022) successfully mimicked this mutation using CRISPR/Cas9. They both used the same target

sequence to induce a deletion mutation in *OsOR* (*OsOR^{Mut}*) that disrupted the splice boundary between the third exon and third intron. The PAM site they used sits three bases away from the exon-intron boundary, making it ideal for disrupting this because SpCas9 typically cuts three nucleotides upstream of the PAM. Doing so produced orange-coloured dark-grown rice callus with a total carotenoid level 6.8–9.7 times higher than wild-type dark-grown callus. Further analysis revealed that the edited orange-coloured calli produced mRNA transcripts of *OsOR^{Mut}* with in-frame insertions or deletions of amino acids, mirroring the *BoOR^{Mut}* mutation (Endo et al. 2019). Editing events were also found in white-coloured calli that did not have an overaccumulation of carotenoids; however, these events produced early stop codons within *OsOR*.

Based on knowledge gained in other plant species (Li et al. 2001; Lopez et al. 2008; Park et al. 2015; Tzuri et al. 2015; Wang et al. 2015; Yuan et al. 2015; Bai et al. 2016; Yazdani et al. 2019), I hypothesised that *OR* could be a target for carotenoid biofortification in wheat. I confirmed this by overexpressing *TaOR^{His}* and found an increase in the GCC compared to *TaOR^{WT}* overexpression (Section 4.2.3), suggesting that variation within *OR* can boost GCC in wheat. I then searched the most highly diverse sequenced wheat panel for existing variation within *OR* (Section 4.2.4); however, no such variation was detected. Having no natural allelic variation to study prompted me to induce changes within *OR* through gene editing and search for changes within a mutagenised population. In this chapter, I used the knowledge acquired on *OR* function in melons and cauliflower to engineer new allelic variation, which I hypothesised could boost carotenoid accumulation within the non-photosynthetic wheat grains. I was unsuccessful in installing the ‘golden SNP’ substitution or mimicking *BoOR^{Mut}* within the endogenous *OR* through gene editing. But by screening ethyl methanesulfonate (EMS) mutagenised TILLING lines with substitutions close to the ‘golden SNP’ and *BoOR^{Mut}* integration site, I found a line with increased GCC. This could represent a new gain-of-function *OR* mutation that leads to an increased carotenoid accumulation phenotype.

5.2 Results

5.2.1 Attempting to install the ‘golden SNP’ through prime editing

To install the ‘golden SNP’ into the native *TaOR* gene, I first analysed the sequence context of the ‘golden SNP’ to see if I could use a cytosine base editor. There were no canonical SpCas9 ‘NGG’ PAM sites that put the editing window of the cytidine deaminase proteins

APOBEC1 (protospacer position 4–8) or APOBEC3A (protospacer positions 2–14) above the target cytosine (Figure 5.1). Instead, I identified several possible Cas variants that put the target cytosine within the editing windows of APOBEC1 and APOBEC3A, as shown in Figure 5.1. At the time of designing this project, only SpCas9-NG and xCas9’s ‘NG’ PAM were demonstrated in wheat (Zhong et al. 2019; Wang et al. 2020); however, these disfavour an ‘NGA’ PAM that would be used to place the base editing machinery over the ‘golden SNP’ (‘AGA’ or ‘TGA’ PAM sequences). I also assessed the possible bystander edits that APOBEC3A could produce with these Cas variants. Six non-‘golden SNP’ cytosines exist within the potential editing windows (Figure 5.1), and five out of the eight potential mutations that could be produced would lead to non-synonymous amino acid changes. APOBEC1 has a smaller editing window than APOBEC3A and so would produce fewer bystander edits. However, it has a sequence preference of strongly disfavours deaminating cytosines within a ‘GC’ context (Anzalone et al. 2020), and the ‘golden SNP’ falls within a GGCGG context, which APOBEC1 would strongly disfavour. Additionally, the cytosine deaminase editing window’s position can differ between Cas variants (Anzalone et al. 2020). Therefore, it is difficult to accurately predict whether the Cas variants I analysed in Figure 5.1 would put the ‘golden SNP’ nucleotide within the editing window. Due to these problems with bystander edits, deaminase sequence context preferences, difficulty in predicting editing windows and most Cas variants not being demonstrated in wheat, I decided a base editing strategy would be a risky strategy to install the ‘golden SNP’ substitution into the native *TaOR* gene.

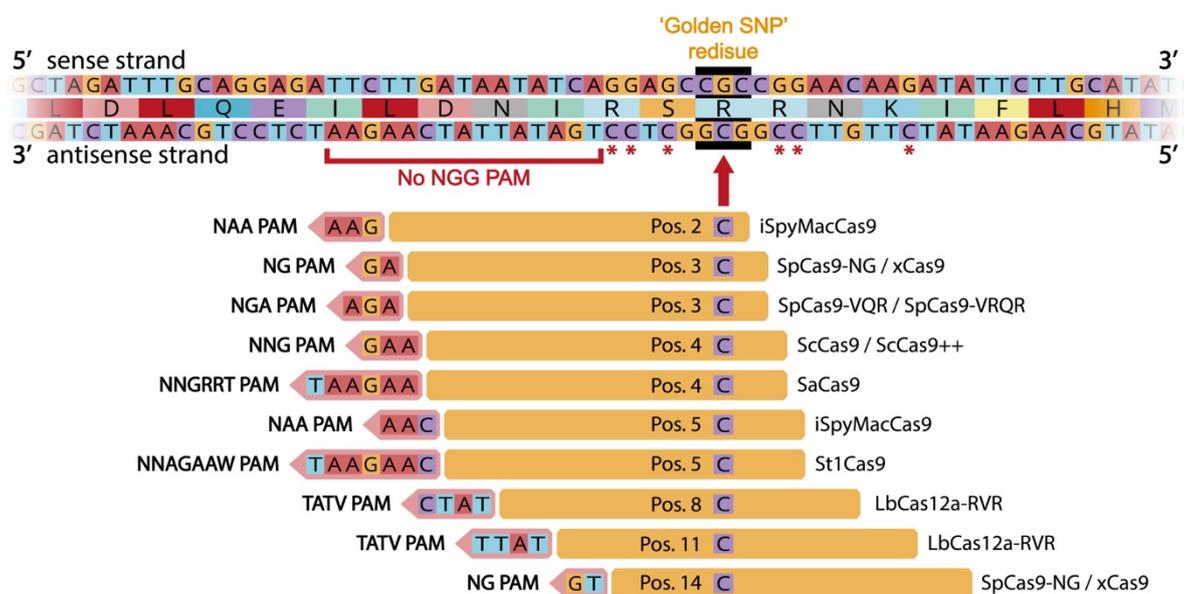


Figure 5.1 Potentially available PAM sites of Cas variants for editing the ‘golden SNP’ cytosine. The editing windows of APOBEC1 and APOBEC3A are upstream of the PAM on the antisense strand. The red bracket shows where a canonical ‘NGG’ PAM site would be required on the antisense strand.

The red arrow on the antisense strand indicates the location of the ‘golden SNP’ cytosine. The yellow rectangles show the protospacer for each Cas variant and the position number of the ‘golden SNP’ cytosine within it. The red asterisks indicate the six non-‘golden SNP’ cytosines, which could be bystander edits of these Cas variant base editors.

I then analysed the sequence context of the ‘golden SNP’ to see if a prime editing approach would be better suited. The ‘golden SNP’ sits close to two SpCas9 ‘NGG’ PAM sites (Figure 5.2a–b), allowing very short RT-templates to be used on the pegRNA. Short RT-templates were found to have higher prime editing efficiencies than longer ones (Anzalone et al. 2020; Lin et al. 2020), making this a promising target for prime editing. Moreover, the two PAM sites are close together, allowing for a dual prime editing approach, where two pegRNAs target editing to both the sense and antisense strands (Figure 5.2c). This was found to significantly improve prime editing efficiencies more than just using a single pegRNA (Lin et al. 2021). Therefore, I attempted a prime editing strategy to install the ‘golden SNP’.

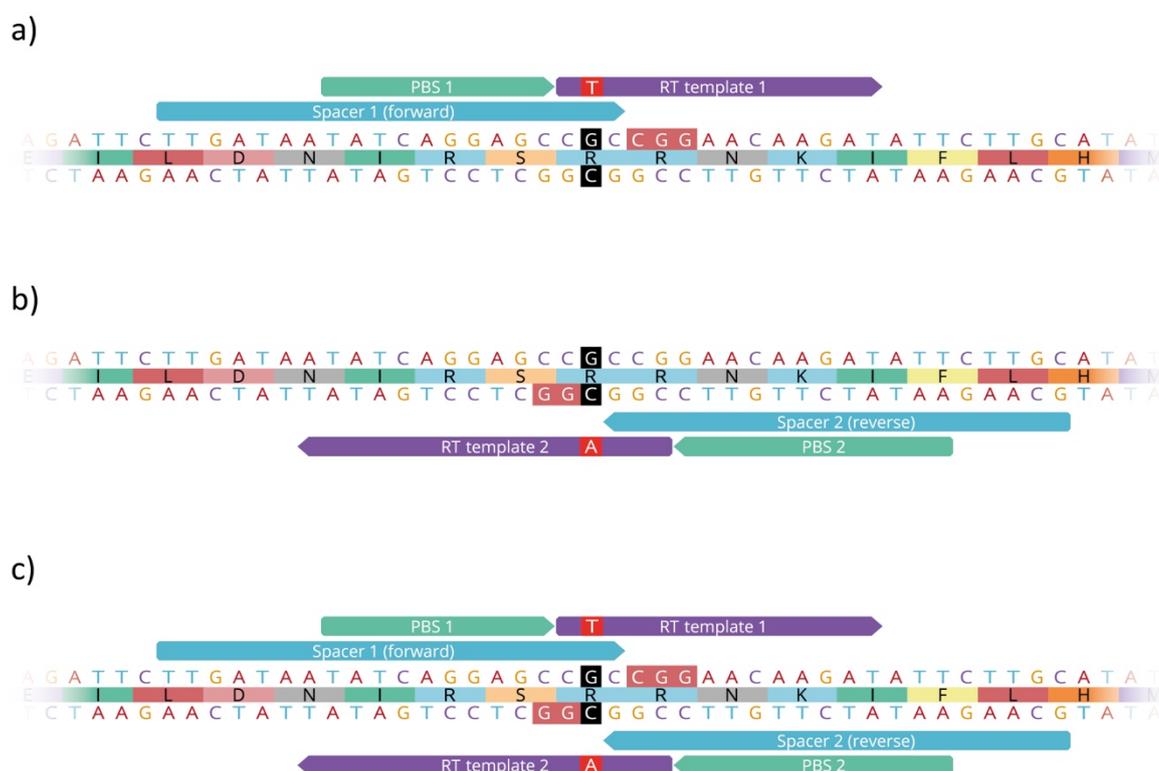


Figure 5.2 Sequence context of the ‘golden SNP’ in the wheat *OR* gene and pegRNA locations for prime editing. The positions of (a) pegRNA#1, (b) pegRNA#2 and (c) the pegRNA#1 pegRNA#2 dual pegRNA strategy are shown. The ‘golden SNP’ nucleotide is shown in the black box. PAM sites are shown on their strand by pink boxes. The desired edit on the RT-template is shown in the red box. RT template=reverse transcriptase template, PBS=primer binding sequence.

To design pegRNA constructs suitable for installing the ‘golden SNP’ substitution, I used the PlantPegDesigner to generate two pegRNAs that could be used in this dual pegRNA strategy (Jin et al. 2022). Figure 5.2 shows where these pegRNAs would anneal to in the *TaOR* sequence. The main outputs of this program are the protospacer sequence that targets the

prime editing machinery to the correct genomic location, the PBS that anneals to the nicked 3' strand, and the RT-template that contains the desired edit to be installed. The scaffold of the pegRNA is the same as a standard single-guide RNA. Figure 5.3 shows the structure of pegRNA#1 and a diagram of how this pegRNA anneals to the 'golden SNP' genomic region to install the desired edit. Due to sequence similarity between the homoeologues of *TaOR* and *TdOR*, these constructs could target all *OR* homoeologues.

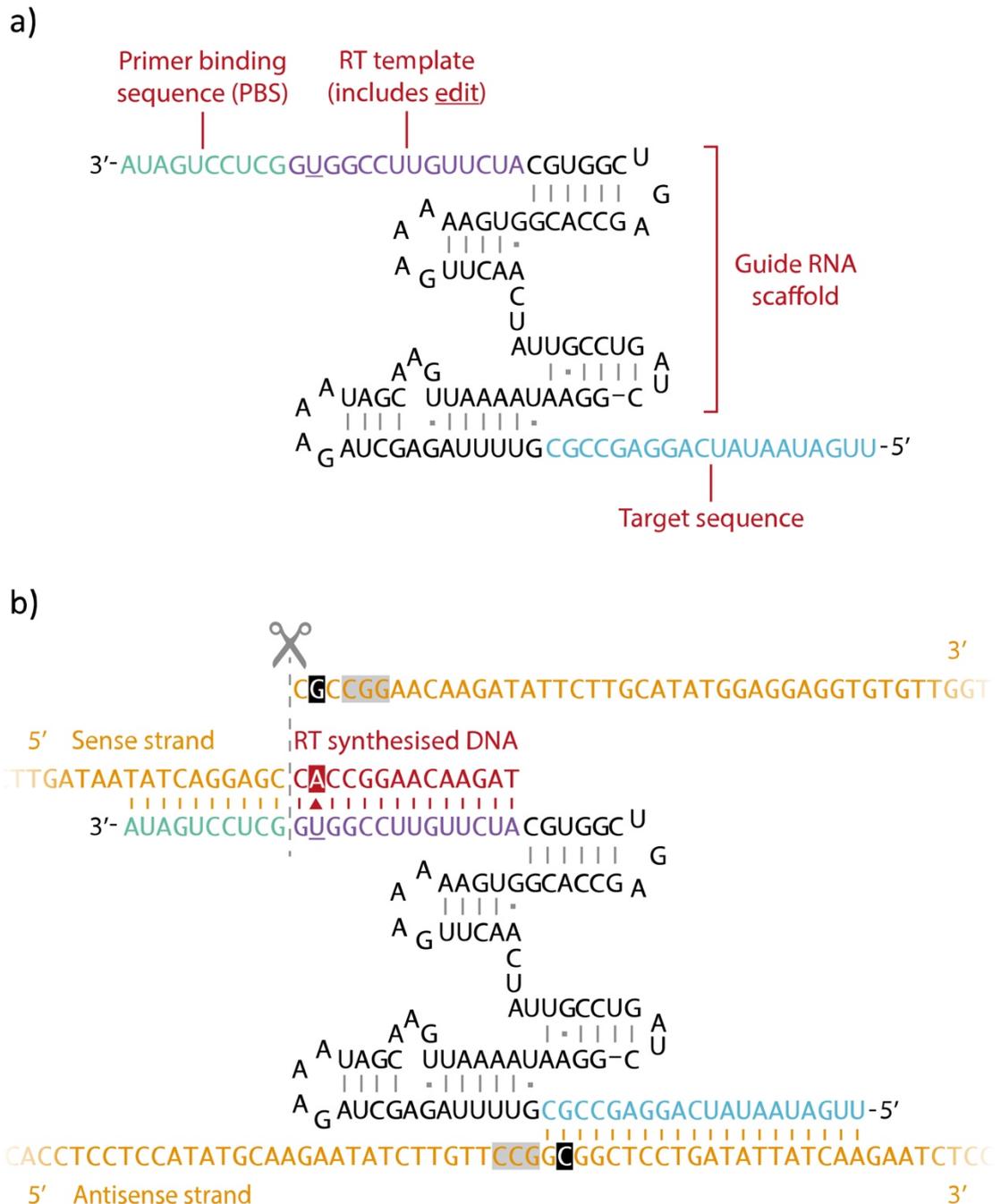
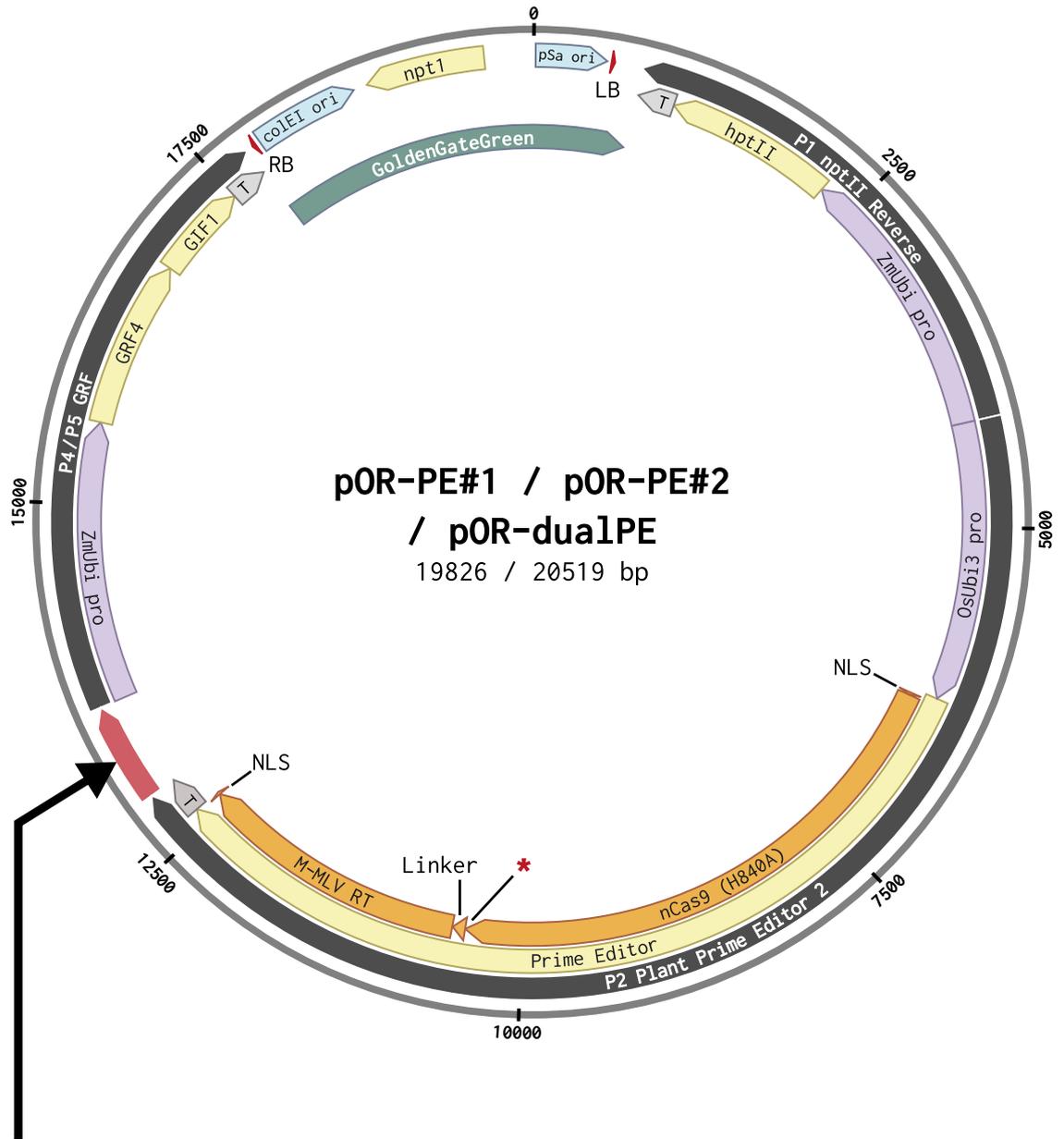


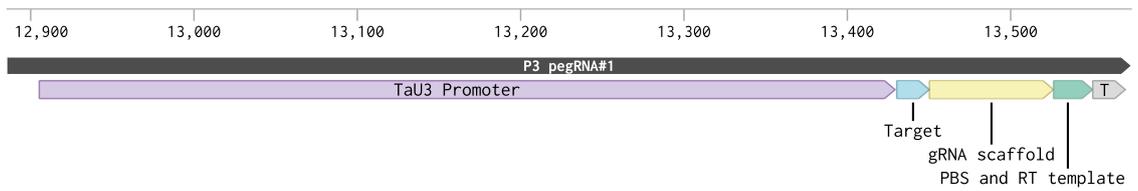
Figure 5.3 Structure of the prime editing guide RNA #1 (pegRNA#1) for installing the 'golden SNP' into the endogenous *OR* gene in wheat. (a) The RNA structure of pegRNA#1 with each component labelled. (b) A diagram of how pegRNA#1 anneals to the genomic DNA (shown in orange). The black box shows the location of the 'golden SNP' nucleotide on both strands. The grey box shows the PAM site on both strands. The grey dashed line and scissors show the cut site of the nCas9 (H840A). The red arrow shows the desired edit to be inserted. RT=reverse transcriptase.

I designed constructs for expressing the prime editor protein and pegRNAs for installing the 'golden SNP' into the wheat *TaOR* gene. In total, three constructs were designed: one with pegRNA#1 that uses a protospacer and PAM sequence on the sense strand, one with pegRNA#2 that uses a protospacer and PAM sequence on the antisense strand, and a dual-pegRNA construct with both pegRNA#1 and pegRNA#2, illustrated in Figure 5.4. The prime editor protein used was produced and made available by TSL SynBio, based on the Plant Prime Editor 2 used by Lin and colleagues (2020). This is a fusion protein of a nickase SpCas9 (H840A), a Moloney murine leukaemia virus reverse transcriptase (M-MLV RT) protein and three nuclear localisation signals (NLSs) located throughout the protein. It was later discovered that the prime editor protein I used was missing one of the NLSs in the middle of this fusion protein (shown by a red asterisk in Figure 5.4). I also included the *GRF4-GIF1* developmental regulators to boost regeneration efficiency. I assembled these constructs into the backbone of the pGoldenGreenGate-M wheat expression vector, and I refer to these plasmids as pOR-PE#1, pOR-PE#2 and pOR-dualPE. Due to the sterility issues found in Kronos when transforming the *OR* overexpression constructs containing *GRF4-GIF1* (Section 4.2.3), I decided to only transform these constructs into Cadenza. I regenerated 127 T₀ plantlets, 48 from pOR-PE#1 transformed embryos, 8 from pOR-PE#2 transformed embryos, and 71 from pOR-dualPE transformed embryos. I assessed only 81 of these plantlets for copy number and found 5 zero copy number escapes and 76 transgenic plants. Copy numbers ranged from 1 to 28 copies, with an average copy number of 7. The remaining 46 plantlets were not analysed for copy number but were analysed for editing events due to these likely being transgenic.

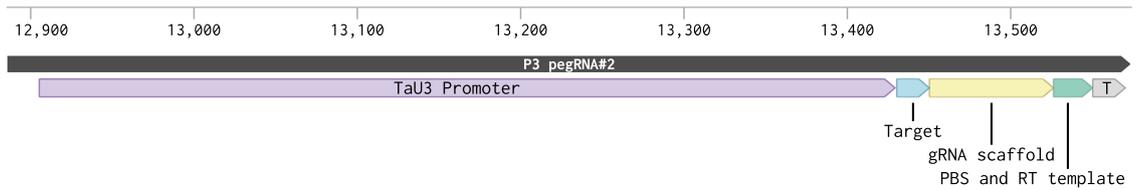
I initially hoped a visual screen could identify plantlets originating from orange calli segments where the 'golden SNP' was installed; however, due to no visible differences found between the *TaOR*^{WT} and *TaOR*^{His} overexpression callus (Section 4.2.3). Instead, screening for germline editing events was done by Sanger sequencing. I designed homoeologue-specific primers to amplify and sequence the 'golden SNP' region of *TaOR-6A*, *TaOR-6B* and *TaOR-6D* (Figure 5.5a) from the 76 confirmed transgenic plantlets and 46 plantlets not screened for copy number. I analysed each of the 366 chromatograms, looking for adenine nucleotides or double peaks at the 'golden SNP', suggesting a successful edit in a homozygous or heterozygous state (Figure 5.5b). I found no editing events in any of the screened T₀ plantlets, suggesting a lack of germline editing events in these.



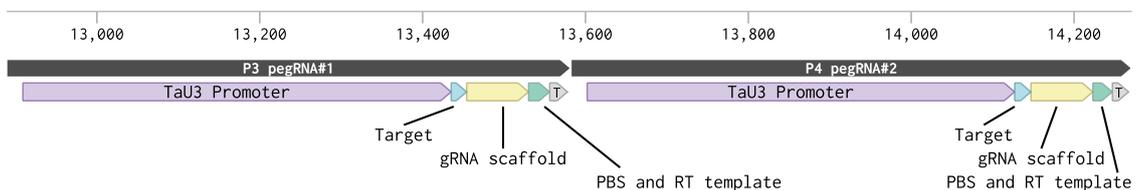
P3 pOR-PE#1 (689 bp)



P3 pOR-PE#2 (689 bp)



P3 and P4 pOR-dualPE (1382 bp)



On the previous page:

Figure 5.4 Plasmid map of the prime editing constructs to install the 'golden SNP' in the endogenous wheat *OR* gene. The prime editor is a fusion of nSpCas9 and M-MLV reverse transcriptase; this is under the expression of the rice ubiquitin 3 promoter (OsUbi3 pro). The red asterisk within the prime editor fusion protein shows the location of the missing nuclear localisation signal compared to the original protein in Lin and colleagues (2020). The three plasmids differ by their pegRNA region, shown by the red arrow within the circular plasmid. The differing pegRNA regions are shown below this. NLS=nuclear localisation signal, T=nos terminator, RB=right border, LB=left border, bp=base pairs, gRNA=single-guide RNA, PBS=primer binding sequence, RT=reverse transcriptase.

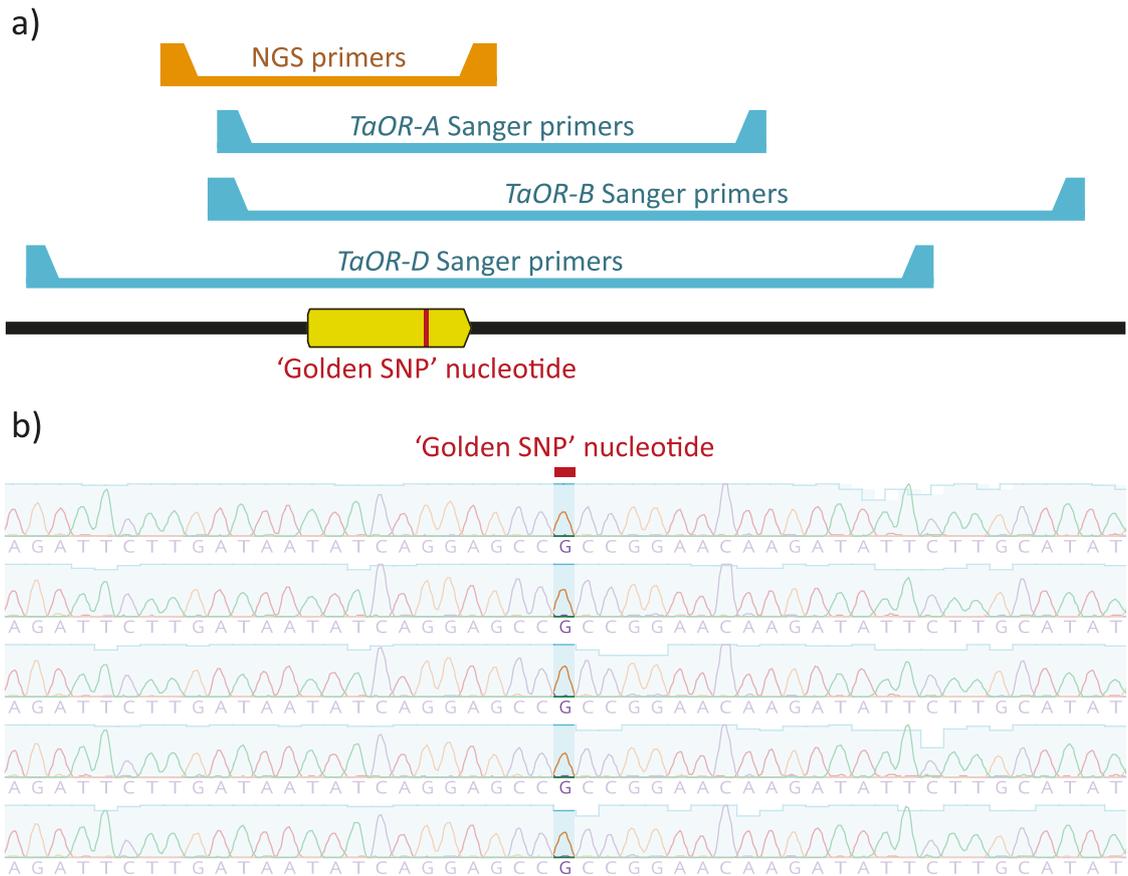


Figure 5.5 Location of sequencing primers over the 'golden SNP' nucleotide and example Sanger sequencing chromatograms. (a) Location of the primer sets for Sanger sequencing and Illumina NGS sequencing over the 'golden SNP' nucleotide. The black bar shows the *TaOR* intron sequence with the yellow rectangle showing the *TaOR* second exon. (b) Example Sanger chromatograms of *TaOR*-A at the 'golden SNP'. The presence of an adenine nucleotide or double guanine-adenine peak was searched for in the chromatograms of the 366 'golden SNP' regions.

To investigate whether editing was occurring in somatic cells at levels below the detection threshold of Sanger sequencing, I used Illumina sequencing to provide a more sensitive analysis. If low-level editing events were found to be occurring, regenerating additional transgenic T_0 plantlets or growing and analysing T_1 grain might enable the detection of the 'golden SNP' in germline tissue. I screened five T_0 pOR-dualPE lines with copy numbers 1, 2, 2, 6 and 8, and I grew several T_1 pOR-dualPE plants and screened five of these lines with

copy numbers 1, 2, 2, 4 and 5. Additionally, five zero-copy number lines from both T_0 and T_1 plants were included as controls to account for sequencing errors. I amplified a 230-233 bp region encompassing the 'golden SNP' across the three wheat genomes with non-homoeologue-specific primers (Figure 5.5a) and sequenced these using Illumina NovoSeq PE250 next-generation sequencing. In total, each plant had a very high coverage (an average of 122,127 reads with a range of 25,902–174,312). To detect somatic editing events, I used Mutect2, which was designed to identify somatic mutations in heterogeneous samples using normal tissue samples as a control (Van der Auwera and O'Connor 2020). Despite the high resolution of this sequencing, this analysis revealed no evidence of somatic editing events at or near the 'golden SNP' nucleotide of the *TaOR* homoeologues. The average frequency of an adenosine nucleotide at the 'golden SNP' guanine position within the alignment was 0.017% in the five controls (a range of 0.016–0.019%), 0.014% in the five T_0 pOR-dualPE plants (a range of 0.007–0.018%) and 0.018% in the five T_1 pOR-dualPE plants (a range of 0.013–0.021%). The absence of detectable edits in the extensive dataset of paired-end reads suggests that no editing events occurred in the somatic tissue of transgenic plants with the prime editing constructs.

5.2.2 Attempting to mimic the *BoOr^{Mut}* mutation in *TaOR* using CRISPR/Cas9 editing

My attempt to introduce the 'golden SNP' into the endogenous wheat *TaOR* gene using prime editing was unsuccessful, so I explored alternative approaches to modify the wheat *TaOR* gene to influence carotenoid accumulation. Previous studies in rice have mimicked the cauliflower *BoOr^{Mut}* using CRISPR/Cas9 (Endo et al. 2019; Kim et al. 2022), and the same PAM site exists in wheat (Figure 5.6). Therefore, I attempted to mimic the *BoOr^{Mut}* mutation in *TaOR* using this PAM site. To target all homoeologues in wheat, two different sgRNAs are necessary: one pair targets *TaOR-A* and *TaOR-B* (sgRNA 1), and another specifically for *TaOR-D* (sgRNA 2). I designed a wheat CRISPR/Cas9 construct for editing this third exon-intron boundary that expressed these two sgRNAs for targeting all *TaOR* homoeologues. I refer to this plasmid as pTaOR-Cas9 and this is illustrated in Figure 5.7.



Figure 5.6 The PAM site used to mimic *BoOr^{Mut}* in rice *OsOR* using CRISPR/Cas9 and the equivalent PAM sites in wheat *TaOR*. The predicted CRISPR/Cas9 cut site is shown by the red bar within the protospacer sequence.

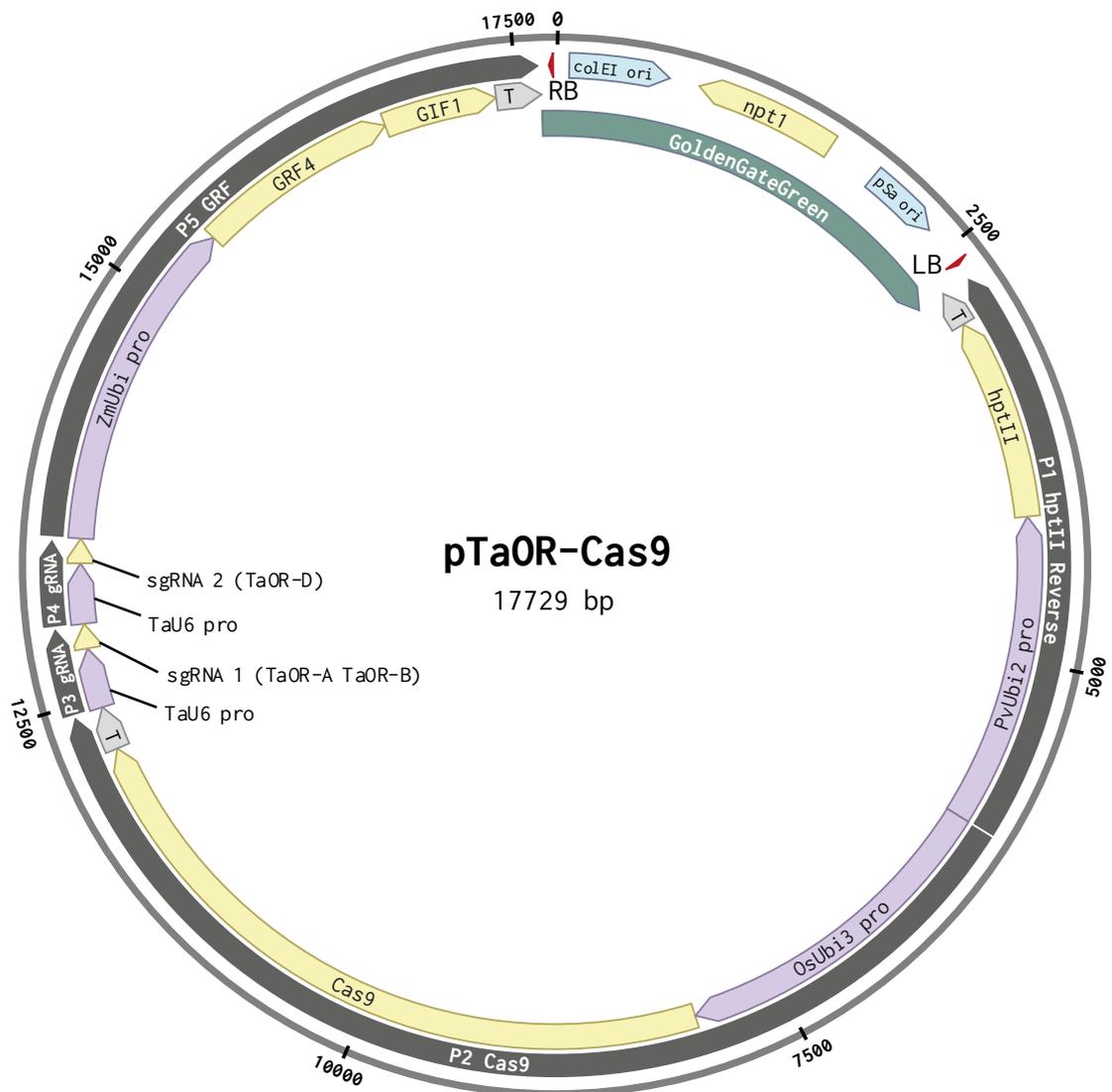


Figure 5.7 Plasmid map of the *TaOR* CRISPR/Cas9 editing constructs to mimic the *BoOr^{Mut}* mutation from cauliflower. The hygromycin selection gene (*hptII*) and developmental regulators (*GRF4-GIF1*) were also included on the plasmid. T=nos terminator, RB=right border, LB=left border, bp=base pairs, sgRNA=single-guide RNA.

A dark-grown calli screen was used to identify callus with *OsOr^{Mut}* installed (Endo et al. 2019; Kim et al. 2022). I previously found a visual screen could not identify dark-grown calli overexpressing the ‘golden SNP’ *TaOR* sequence (Section 4.2.3), suggesting modifying the endogenous *TaOR* gene would not result in an observable colour difference in wheat callus. However, in rice, the overexpression of the rice *OR* gene with the ‘golden SNP’ installed (*OsOR^{His}*) also did not produce a visible orange-coloured callus (Jung et al. 2021), whereas installing *OsOr^{Mut}* did. Therefore, *OsOr^{Mut}* and *OsOR^{His}* may work through different mechanisms, and successfully installing *TaOr^{Mut}* in wheat may still produce orange-coloured dark-grown calli. Consequently, I screened for editing events using dark-grown callus. I transformed these constructs into Cadenza immature embryos. I then kept growing them on selection media in the dark, instead of moving them onto regeneration media in the light after 5 weeks of selection. During this time, no orange-coloured callus was observed (Figure 5.8). After 12 weeks on selection media, I moved these calli onto regeneration media under light conditions to regenerate plantlets. In total, 102 T₀ plantlets were regenerated, and these were all transgenic. Copy numbers ranged from 1 to 57 copies, with an average copy number of 6. This suggests that leaving them on selection media for longer prevented non-transgenic calli from regenerating non-transgenic plantlets.



Figure 5.8 Dark-grown calli from immature Cadenza embryos transformed with pTaOR-Cas9. This photo was taken after 10 weeks in the dark on selection media. No obvious orange-coloured callus segments were observed.

To investigate whether editing events were occurring, I designed homoeologue-specific primers to amplify and sequence the *TaOr^{Mut}* cut site of *TaOR-6A*, *TaOR-6B* and *TaOR-6D* from the 102 transgenic T₀ pTaOR-Cas9 plantlets. Sanger sequencing revealed six

heterozygous editing events out of 306 potential editing sites, with the location and type of these edits found in Table 5.1. This suggests the pTaOR-Cas9 sgRNAs were correctly targeting the three *TaOR* homoeologues; however, this editing efficiency was very low at 1.47% for the sgRNA targeting *TaOR-6A* and *TaOR-6B*, and 2.94% for the sgRNA targeting *TaOR-6D*. This low editing occurred despite Cas9 presumably having had longer to produce edits before plantlets were regenerated due to being grown at the callus stage on selection media for longer than the normal transformation protocol.

Table 5.1 Editing events from 102 T₀ transgenic pTaOR-Cas9 plants. CN=copy number, bp=base pairs.

Plant ID	CN	Yield/plant (g)	Homoeologue	Editing event	Predicted Consequence on <i>TaOR</i>
3476-4-09	2	17.5	<i>TaOR-6A</i>	-15 bp	Introduces amino acids from intron but produces an early stop codon
3476-7-02	6	23.3	<i>TaOR-6A</i>	+1 bp (C)	Frameshift and an early stop codon
3475-4-02	10	0	<i>TaOR-6B</i>	-14 bp	Introduces amino acids from intron but produces an early stop codon
3476-4-05	2	8.8	<i>TaOR-6D</i>	+1 bp (A)	Frameshift and an early stop codon
3476-5-05	3	0	<i>TaOR-6D</i>	+1 bp (G)	Frameshift and an early stop codon
3476-7-06	14	1.7	<i>TaOR-6D</i>	+1 bp (A)	Frameshift and an early stop codon

To assess whether these editing events affected GCC, I grew these edited lines to maturity. *BoOr^{Mut}* is a dominant mutation, so if any of these T₀ heterozygous edited lines mimicked this, there would be an increase in GCC. I also grew seven non-edited lines of similar copy numbers to act as controls, which had also been through the same tissue culture process. These contained the *GRF4-GIF1* developmental regulators and came from the same calli segments as the edited lines (four 2-copy lines, two 10-copy lines, and one 15-copy line). They still had the CRISPR/Cas9 transgene and could have edits occurring after they were genotyped; however, since the editing efficiency was so low, edits would likely be in small chimeric segments. Three of the edited lines produced insufficient grain for grain carotenoid analysis (Table 5.1). Three technical replicates were conducted for each line, and the averaged results are found in Figure 5.9. No edited line had an increase in GCC compared to the controls, either for total carotenoid content or the content of any individual carotenoid species, suggesting that these *TaOR*-edited lines did not overaccumulate carotenoids in the grain.

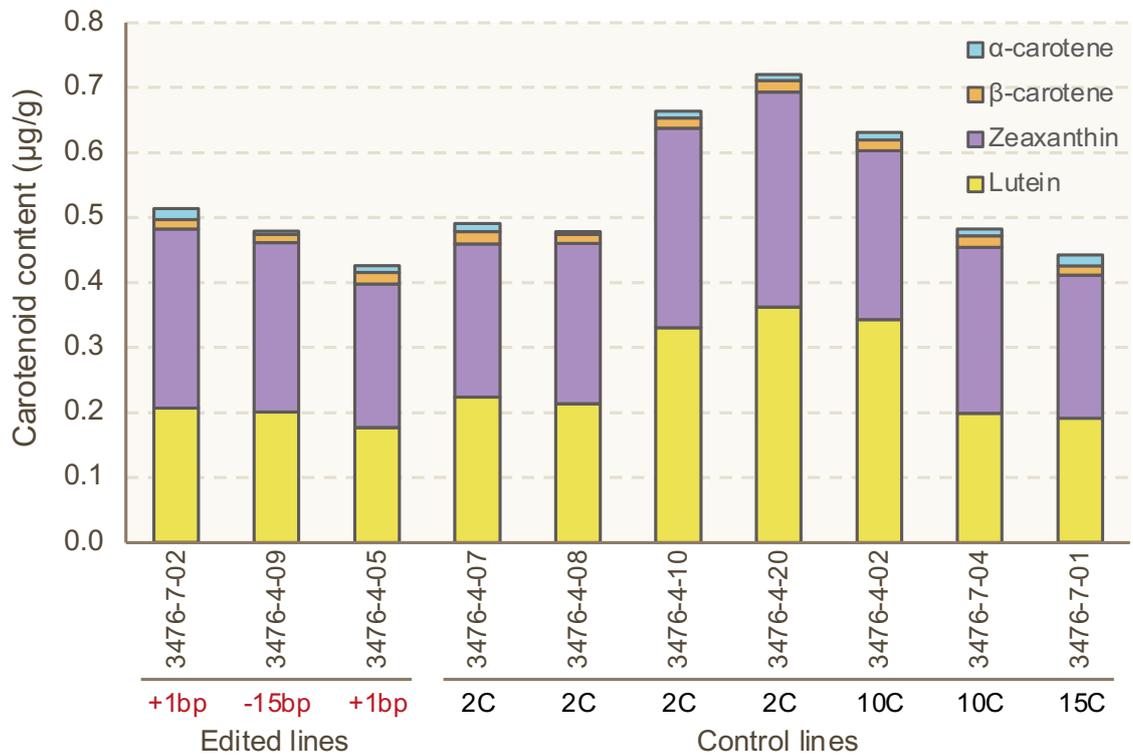
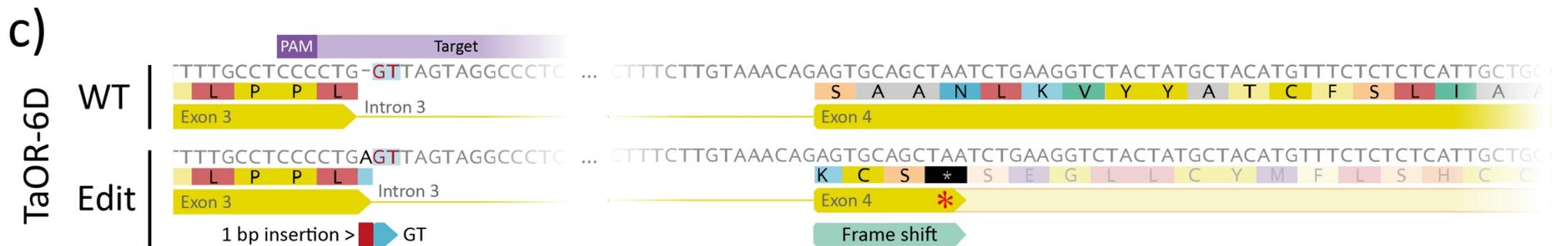
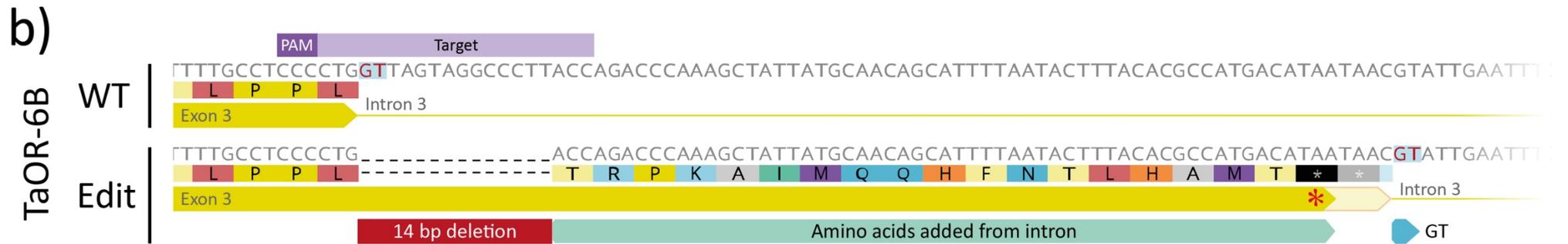
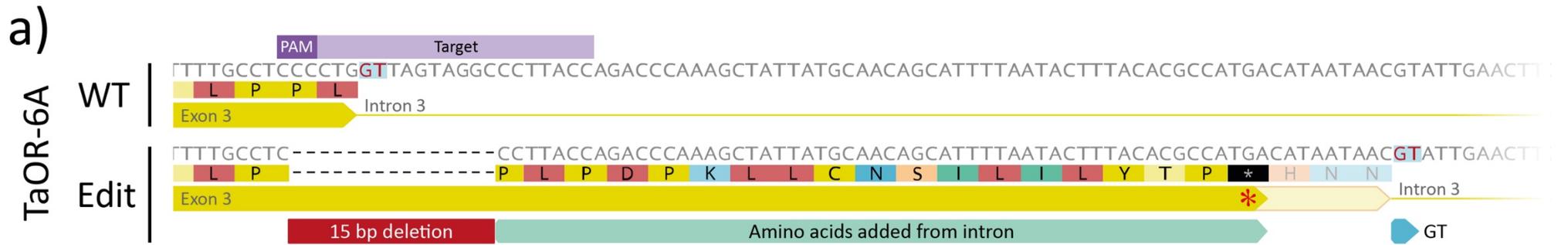


Figure 5.9 Stacked bar charts showing the carotenoid content of the edited T_1 pTaOR-Cas9 grains compared to non-edited T_1 controls of similar copy number.

To predict if any editing events I found in *TaOR* could lead to an in-frame change, I analysed the impact of these edits on *TaOR* transcripts. The 15 bp deletion in plant 3476-4-09 and the 14 bp deletion in plant 3475-4-02 removed the 'GT' splice-donor site of the third intron. Therefore, this would likely produce a transcript that includes some of the third intron (Figure 5.10a–b). I searched within the intron region for the next 'GT' that could be used as a splice-donor site; however, in both deletion events, an early stop codon was produced before this 'GT'. The presence of this 'GT' site does not necessarily mean it would be used as a splice-donor site since splicing is difficult to predict; however, since a stop codon is produced before this, even if splicing were to occur here, it would likely produce a truncated, non-functional protein. The 1 bp insertion in either *TaOR-6A* or *TaOR-6D* retains the 'GT' splice-donor site and produces a frameshift mutation in the *TaOR* transcript (Figure 5.10c). This creates a stop codon near the start of the fourth exon in both *TaOR-6A* and *TaOR-6D*. This suggests these editing events likely produce a truncated TaOR protein rather than an in-frame addition or deletion of amino acids, which was desired to mimic *BoOR^{Mut}*.



On the previous page:

Figure 5.10 Predicted consequences of the editing events found in the pTaOR-Cas9 T₀ transgenic plants on the TaOR protein. (a) The editing event of plant 3476-4-09, a 15 bp deletion in *TaOR-6A*. (b) The editing event of plant 3475-4-02, a 14 bp deletion in *TaOR-6B*. (c) The editing event of plant 3476-4-05, an insertion of an adenine nucleotide. Plants 3476-7-02, 3476-5-05 and 3476-7-06 also had single nucleotide insertions, which are predicted to cause similar frameshift mutations. All mutations are predicted to result in an early stop codon (red asterisk).

5.2.3 Searching for gain-of-function *OR* EMS mutations that increase grain carotenoid content

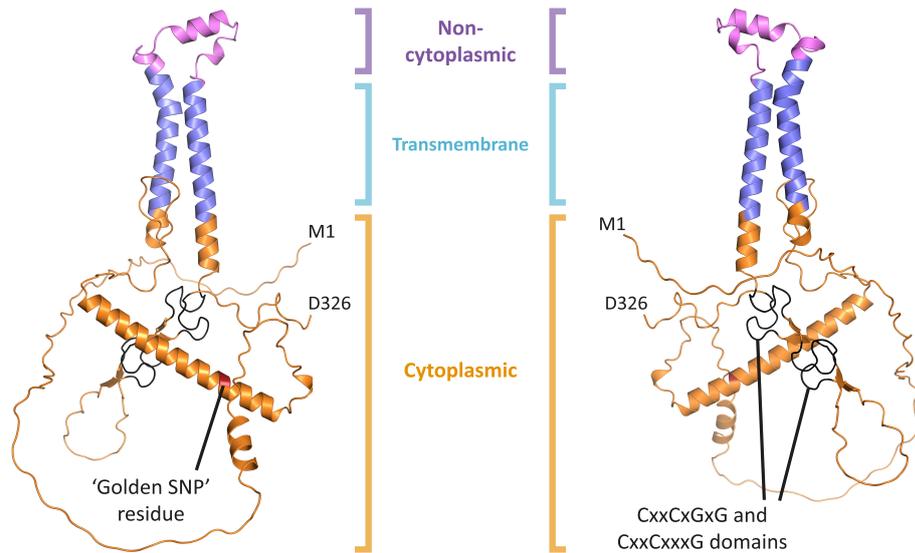
The cauliflower *BoOr^{Mut}* mutation and the rice *OsOr^{Mut}* mutations suggest that mutations within *OR* besides the ‘golden SNP’ can result in an overaccumulation of carotenoids. Therefore, I hypothesised that I could select EMS TILLING lines with amino acid substitutions close to the site of the ‘golden SNP’, *BoOr^{Mut}* and *OsOr^{Mut}* within *TaOR*, and then screen field bulks of these lines to hopefully identify a gain-of-function mutation with increased GCC. The TILLING collection would be an especially good resource to look for these mutations because non-synonymous mutations caused by G-to-C or A-to-T EMS-mutagenesis are primarily amino acid substitutions rather than protein truncations. This is similar to how the ‘golden SNP’, *BoOr^{Mut}* and *OsOr^{Mut}* are in-frame alterations of the *OR* protein sequence that lead to a dominant gain-of-function mutation.

To identify TILLING mutations similar to the ‘golden SNP’, *BoOr^{Mut}* and *OsOr^{Mut}* mutations, I analysed the domains, transmembrane topology and structure of *TaOR*. The ‘golden SNP’ is located within a predicted cytoplasmic α -helix at the start of the protein, outside the predicted DnaJ zinc finger domain (Figure 5.11a-b). The integration site of the *BoOr^{Mut}* retrotransposon and CRISPR/Cas9 cut site of *OsOr^{Mut}* are within this predicted cytoplasmic domain, close to this α -helix (37 residues and 50 residues away from the ‘golden SNP’ residue, respectively). Therefore, I focused my search for TILLING mutations within this first cytoplasmic domain and close to or within the α -helix where the ‘golden SNP’ residue is located. I prioritised mutations with low SIFT scores, indicating these are in conserved regions. I identified 11 mutations within Kronos and 7 mutations within Cadenza, the locations of which are found in Figure 5.11b and their descriptions are found in Table 5.2. Interestingly, one of the Kronos lines, K0685, had TILLING mutations in both the A-genome and B-genome copies of *TdOR*.

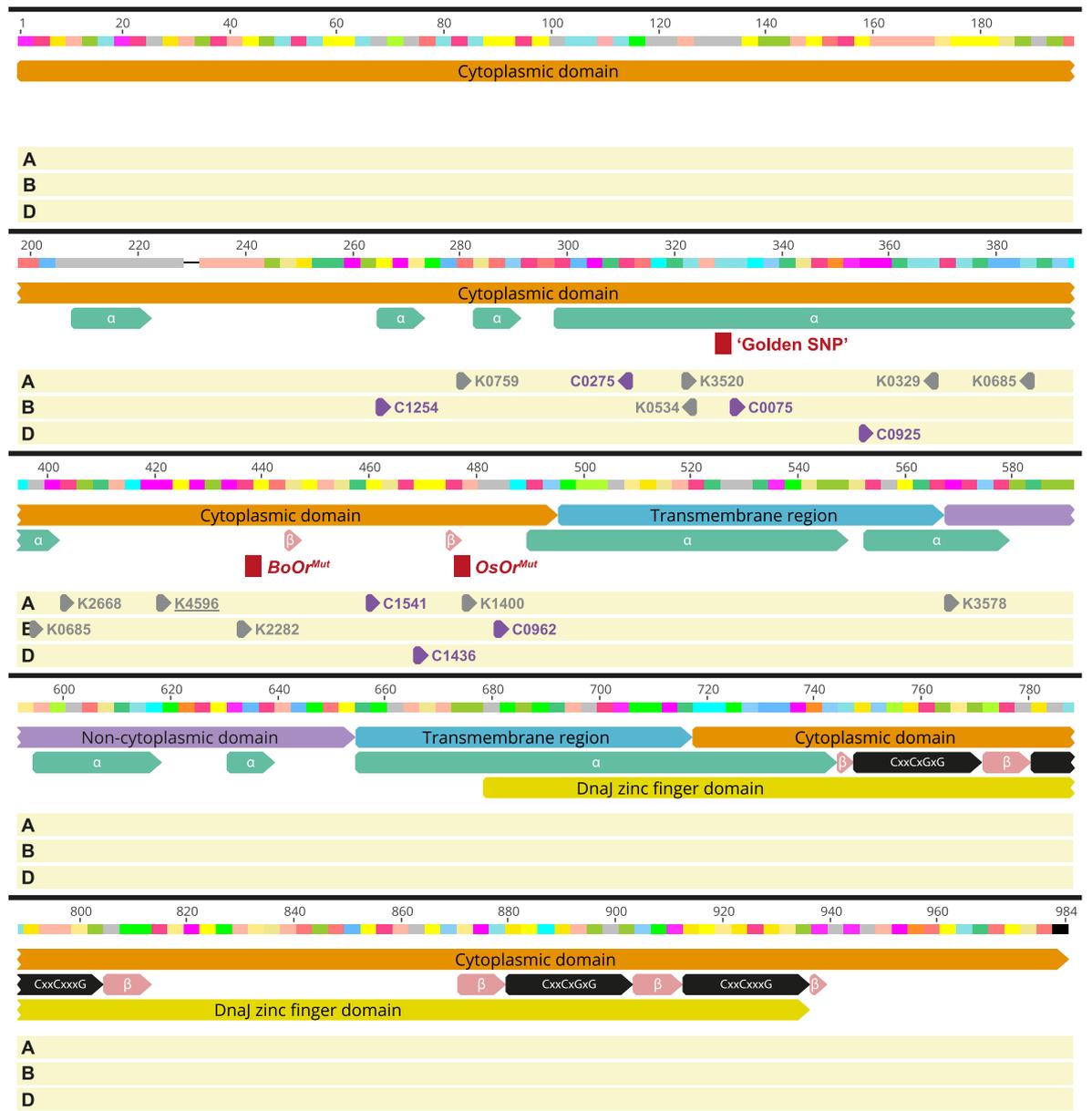
I also analysed *TaORLIKE* because this is thought to be involved in the carotenoid biosynthesis process (Zhou et al. 2015; Sun et al. 2023b), and it is possible a mutation in

this gene would increase carotenoid accumulation as well. The TaORLIKE predicted structure and domains are similar to TaOR, and the α -helix that the 'golden SNP' residue is within is also present in this protein (Figure 5.11c-d). The major difference between TaORLIKE and TaOR is that one of the conserved CxxCxxxG motifs within the DnaJ zinc finger domain is missing. I identified 8 mutations within Kronos and 3 mutations within Cadenza, the locations of which are found in Figure 5.11d and their descriptions are found in Table 5.2. Two Cadenza mutations are at the equivalent 'golden SNP' residue in TaORLIKE (C0773 and C1233). I hypothesised that most of these mutations within *TaOR* and *TaORLIKE* would reduce carotenoid content due to the role these proteins are thought to play in stabilising PSY in carotenoid biosynthesis. However, a dominant gain-of-function mutation like the 'golden SNP' would lead to an increase in carotenoid content, and this is what I hoped to find.

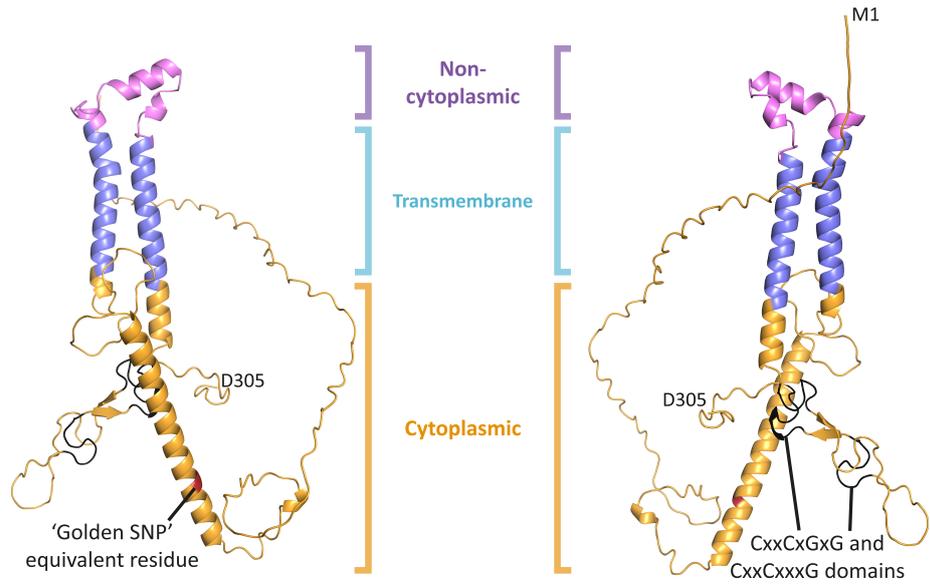
a) TaOR AlphaFold prediction



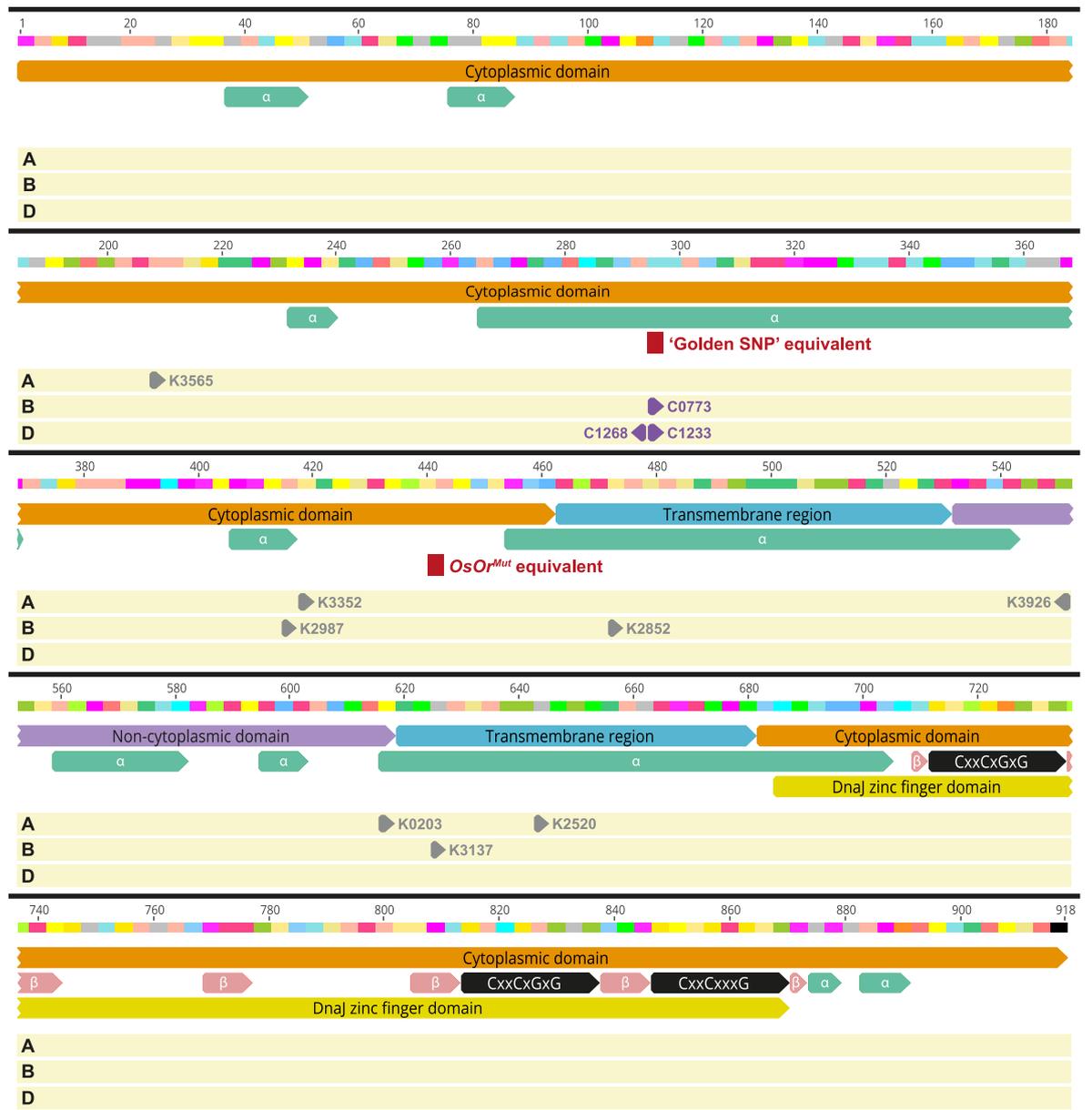
b) TaOR domain annotations and locations of EMS mutations



c) TaORLIKE AlphaFold prediction



d) TaORLIKE domain annotations and locations of EMS mutations



On the previous page:

Figure 5.11 Domain annotations and Alpha Fold protein models of TaOR and TaORLIKE. The Alpha Fold models (a) and (c) have been coloured based on the predicted location of the region, either non-cytoplasmic, transmembrane or cytoplasmic. In the domain analysis (b) and (d), the three yellow bars represent the A, B and D-genomes with the location of TILLING mutations displayed by the grey arrows for the tetraploid Kronos variety and purple arrows for the hexaploid Cadenza variety. α = α -helix, β = β -sheet.

Table 5.2 Description and carotenoid content of Kronos and Cadenza EMS TILLING mutations identified within *OR* and *ORLIKE*. Carotenoid content was measured from field bulks and is in $\mu\text{g/g}$. The high grain carotenoid content line, K4596, is highlighted in grey. ID=line identifier, Sub.=amino acid substitution, Coord.=coordinate of amino acid within the protein, SIFT=SIFT score of substitution, Geno.=EMS genotype, Hom.=homozygous, Het.=heterozygous, TGCC=total grain carotenoid content, Kro.=Kronos, Cad.=Cadenza.

Gene	ID	Sub.	Coord.	SIFT	Geno.	α -caro.	β -caro.	Zeaxan.	Lutein	TGCC
Kro. Control	KWT'15	–	–	–	–	0.052	0.060	0.167	2.227	2.505
Kro. Control	KWT'16	–	–	–	–	0.066	0.049	0.134	2.237	2.486
<i>TdOR-A</i>	K0759	D/N	94	0.02	Hom.	0.057	0.049	0.187	1.650	1.943
<i>TdOR-A</i>	K3520	R/K	108	0.23	Het.	0.053	0.054	0.094	1.412	1.612
<i>TdOR-A</i>	K0329	R/K	123	0	Het.	0.036	0.035	0.217	1.224	1.514
<i>TdOR-A</i>	K2668	L/F	135	0.07	Het.	0.054	0.058	0.201	1.589	1.902
<i>TdOR-A</i>	K4596	E/K	141	0.05	Het.	0.070	0.076	0.426	2.819	3.390
<i>TdOR-A</i>	K1400	S/N	160	0.18	Het.	0.083	0.072	0.140	2.179	2.474
<i>TdOR-A</i>	K3578	E/K	190	0	Hom.	0.047	0.059	0.307	1.800	2.213
<i>TdOR-A*</i>	K0685	R/K	129	0.03	Het.	0.022	0.041	0.167	0.847	1.077
<i>TdOR-B*</i>	K0685	A/T	130	0.2	Het.	0.022	0.041	0.167	0.847	1.077
<i>TdOR-B</i>	K0534	R/K	105	0.26	Het.	0.034	0.048	0.081	1.149	1.312
<i>TdOR-B</i>	K2282	L/F	143	0.07	Het.	0.049	0.039	0.345	1.725	2.158
<i>TdORLIKE-A</i>	K3565	S/F	70	0	Het.	0.057	0.060	0.188	2.008	2.314
<i>TdORLIKE-A</i>	K2520	A/V	215	0.01	Het.	0.041	0.045	0.179	2.063	2.329
<i>TdORLIKE-A</i>	K3926	G/D	184	0	Hom.	0.051	0.051	0.095	1.556	1.754
<i>TdORLIKE-A</i>	K0203	P/L	206	0	Het.	0.035	0.033	0.102	0.974	1.144
<i>TdORLIKE-A</i>	K3352	T/I	140	0.18	Het.	0.040	0.041	0.168	1.039	1.287
<i>TdORLIKE-B</i>	K2987	S/F	139	0	Hom.	0.036	0.055	0.018	1.416	1.525
<i>TdORLIKE-B</i>	K2852	T/I	158	0.09	Het.	0.048	0.048	0.105	1.241	1.442
<i>TdORLIKE-B</i>	K3137	A/T	209	0	Hom.	0.035	0.045	0.101	1.414	1.596
Cad. Control	CWT	–	–	–	–	0.603	0.257	0.015	0.014	0.889
<i>TaOR-A</i>	C0275	L/F	104	0.01	Het.	0.469	0.294	0.013	0.009	0.785
<i>TaOR-A</i>	C1541	P/S	154	0	Het.	0.472	0.320	0.019	0.011	0.822
<i>TaOR-B</i>	C0075	R/W	108	0	Het.	0.422	0.208	0.011	0.012	0.653
<i>TaOR-B</i>	C1254	P/L	86	0.01	Het.	0.656	0.076	0.013	0.016	0.760
<i>TaOR-B</i>	C0962	A/T	159	0.45	Het.	0.691	0.113	0.014	0.015	0.832
<i>TaOR-D</i>	C0925	E/K	118	0	Het.	0.475	0.252	0.011	0.012	0.750
<i>TaOR-D</i>	C1436	P/L	156	0	Het.	0.399	0.266	0.013	0.010	0.687
<i>TaORLIKE-B</i>	C0773	R/Q	99	0	Het.	0.556	0.311	0.013	0.016	0.896
<i>TaORLIKE-D</i>	C1268	S/N	98	0.01	Hom.	0.218	0.316	0.012	0.005	0.552
<i>TaORLIKE-D</i>	C1233	R/Q	99	0	Hom.	0.637	0.156	0.017	0.014	0.824

*K0685 has mutations within *TdOR-A* and *TdOR-B*.

The M₄ field bulks of these TILLING lines were screened for their carotenoid content, with only one measurement for each line. For the Kronos bulks, controls from a 2015 bulk and a 2016 bulk were available, and these were both screened. The results of this are presented in Figure 5.12 and Table 5.2. As expected, most of the mutations within *OR* and *ORLIKE* in Kronos and Cadenza varieties reduced total carotenoid content compared to the controls, suggesting that both proteins play a role in carotenoid biosynthesis. K0685, which had mutations within *TdOR-6A* and *TdOR-6B*, had the lowest total carotenoid content despite one of these mutations having a relatively high SIFT score (0.2). Some mutations appeared to have little effect on total carotenoid content compared to the controls, such as K1400 and C0773.

Excitingly, the Kronos line K4596 had a higher total GCC compared to the Kronos wildtype controls (Figure 5.12 and Table 5.2). This line had a total carotenoid content of 3.390 µg/g compared to 2.496 µg/g in the controls, an increase of 35.8%. The content of lutein, zeaxanthin, β-carotene and α-carotene were also all higher in K4596 than the Kronos control line (Table 5.2). The EMS mutation within this line is a glutamate-to-lysine (E/K) substitution in *TaOR-6A* in between the 'golden SNP' residue (31 residues away) and the site of the *BoOr^{Mut}* retrotransposon integration site (5 residues away; Figure 5.11b). I refer to this as the 'E141K substitution'. It has a SIFT score of 0.05, suggesting it to be a deleterious mutation to *TaOR* function based on conservation within other species. I performed an *in silico* search for EMS mutations located in carotenoid biosynthesis genes within the line K4596 (searching for mutations within *PSY*, *PDS*, *ZDS*, *Z-ISO*, *CRTISO*, *LCYE*, *LCYB*, *HYDs*, *CCDs*, *NCEDs*, *ORLIKE*, *ZEP* and *LOX*). This did not identify any other mutations within these major carotenoid biosynthesis genes. I aligned the protein sequences of the 35 OR orthologues that I previously used to produce the gene tree of OR and found that at this residue, the majority of these have the acidic amino acid glutamate that *TaOR* has (25/35) with the acidic amino acid aspartic acid as the next most common residue (4/25) (Appendix Material 22). The only orthologue with lysine at this residue is *IbOR* from sweet potato (*Ipomoea batatas*). I designed KASP primers to follow the E141K substitution through crossing programmes, and K4596 lines are being grown in the glasshouse to confirm this effect (March 2024).

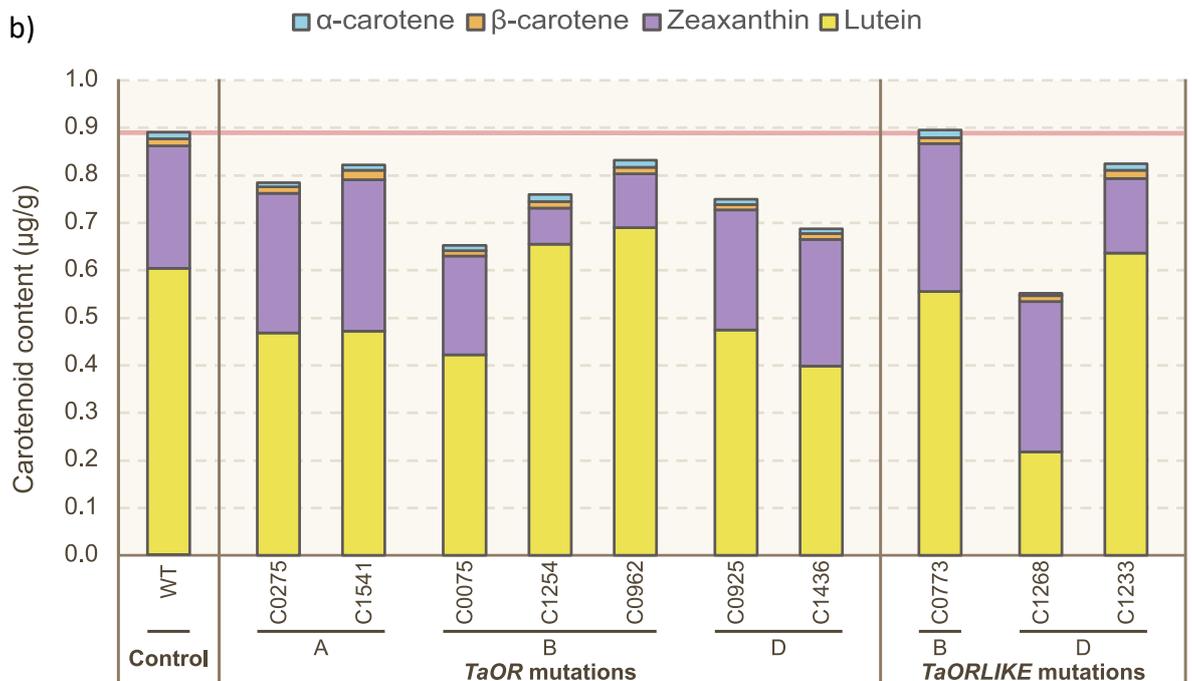
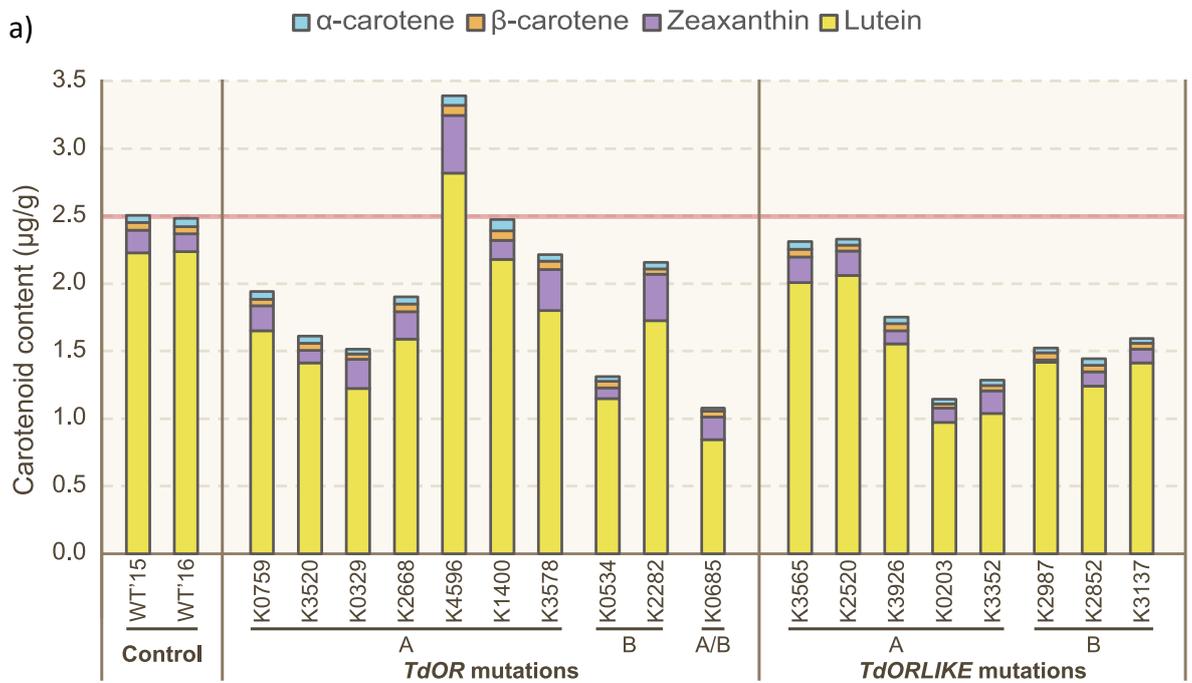


Figure 5.12 Grain carotenoid content of (a) Kronos and (b) Cadenza TILLING lines with EMS mutations in *OR* and *ORLIKE*. TILLING lines were screened to identify a gain-of-function mutation with increased carotenoid accumulation. Carotenoid content was measured from field bulks, and each bar represents a single replicate. The red line displays the total grain carotenoid content of the wild-type Kronos or Cadenza control.

5.3 Discussion

5.3.1 Low Plant Prime Editor 2 efficiency is likely why the ‘golden SNP’ could not be installed in endogenous *TaOR*

My attempt to install the ‘golden SNP’ into the endogenous *TaOR* gene using prime editing was unsuccessful, and I found no evidence of germline or somatic editing events in the transgenic plants (Section 5.2.1). I used the Plant Prime Editor 2 (PPE2) produced by Lin and colleagues (2020), which they demonstrated to work in wheat; however, this was only within a protoplast transformation system. They introduced PPE2 and pegRNA expression constructs to wheat protoplasts by PEG-mediated transfection and analysed protoplasts for editing events with deep sequencing. This is a very different method to generating and screening *Agrobacterium*-mediated stably transformed lines, and prime editing efficiencies are likely to differ between them. Transgene expression levels can be much lower in *Agrobacterium*-based stable transformation than protoplast transfection, reducing editing efficiencies. For instance, the T-DNA can randomly integrate into transcriptionally inactive regions of the genome, and multiple T-DNA insertions may result in transgene silencing (Gelvin 2017). The efficiencies of standard non-homologous end joining (NHEJ) CRISPR/Cas9 editing within protoplasts have previously been found to be higher in plants than in *Agrobacterium*-mediated transformed lines (González et al. 2021), and it is probable that prime editing efficiencies are also higher within protoplast systems.

Within the wheat protoplasts, Lin and colleagues (2020) found the maximum prime editing efficiency of seven targets to be around 1.4%, with some sites showing editing efficiencies below 0.2%. These are very low editing efficiencies, and it is possible that the *Agrobacterium*-mediated transformation I used pushed editing efficiencies below the detection limit of deep sequencing. Additionally, this shows that editing efficiencies vary between different target sites, and it is possible that the pegRNAs that I used to install the ‘golden SNP’ in *TaOR* had very low efficiency. For a future attempt, I could use a protoplast system to test a variety of pegRNAs capable of installing the ‘golden SNP’ into *TaOR* and then select the most efficient one for stable transformation. Importantly, the prime editor protein I used was missing one of the three NLSs found between the nCas9 and the M-MLV RT (Figure 5.4), which was included in the original PPE2 protein (Lin et al. 2020). Although this NLS was not included in the original Prime Editor 2 protein (Anzalone et al. 2019), it is possible that within wheat, this third NLS is required to transport the large prime editor fusion protein to the nucleus correctly. Additional NLSs on PPE2 have increased its editing

efficiency (Chen et al. 2021); and so it is possible that removing existing NLS domains would reduce editing efficiency. It may be that this reduced editing efficiency below a detectable level within somatic cells.

Following the PPE2 system many improvements have been made to the prime editors and pegRNAs to improve editing efficiencies in plants (Huang and Liu 2023; Li et al. 2023). Improved efficiency was found by increasing the expression of pegRNAs in maize plants by doubling the expression cassette or using a U6 composite promoter (Jiang et al. 2020). Editing efficiency was also improved by optimising the architecture of the prime editor by fusing the M-MLV RT protein at the N-terminus of the prime editor rather than the C-terminus that PPE2 uses (Xu et al. 2021). For single nucleotide edits, such as introducing the 'golden SNP', including synonymous mutations within the RT-template has also been found to improve editing efficiencies as, after nicking, these prevent the RT-template from binding to the nicked DNA strand, which might occur when there is only a single nucleotide difference in the RT-template (Xu et al. 2021). The pegRNAs are also prone to degradation at their 3' end, which contains the PBS and RT-template. Therefore, RNA motifs with specific secondary structures (pseudoknots) have been introduced to their 3' end to prevent degradation and enhance their stability (Jiang et al. 2022; Li et al. 2022). These engineered pegRNAs (epegRNA) increased prime editing efficiency by 2.35- to 29.22-fold compared to pegRNAs. Additionally, plant prime editing efficiency was improved by using a PEmax prime editor, an improvement of PE2 (Jiang et al. 2022; Li et al. 2022). This includes several substitutions in nSpCas9 that improve its efficiency, alongside additional NLSs and a new linker between nSpCas9 and the M-MLV RT (Chen et al. 2021). Another engineered Plant Prime Editor (ePPE) was produced by deleting the reverse transcriptase RNase H domain within the M-MLV RT and incorporating a viral nucleocapsid protein with nucleic acid chaperone activity to the prime editor, which also improved prime editing efficiencies in plants (Zong et al. 2022).

Ni and colleagues (2023) combined several of these improvements to achieve prime editing efficiencies of 6.5- to 503.6-fold (average 33.0-fold) higher than PPE2 in wheat protoplasts. Excitingly, they were able to regenerate prime edited plants through *Agrobacterium*-mediated transformation with editing frequencies up to 74.5%. They used epegRNAs in combination with a new engineered Plant Prime Editor plus (ePPEplus) that combines the PEmax and ePPE modifications and introduces a mutation within the M-MLV RT to enhance DNA synthesis during prime editing. Furthermore, they used a Cys-type ribonuclease 4

processing system to express multiple epegRNAs from the same promoter, allowing multiplex prime editing. Using nine epegRNAs to target eight wheat genes, they found 48 out of 51 transgenic plants harbouring mutations in at least one target gene. This is the first published example of successfully regenerating prime edited wheat plants, and therefore, future attempts to install the 'golden SNP' into the endogenous *TaOR* gene could use this system. It is also likely that in the future, new modifications to prime editing will further improve its editing efficiency, facilitating the installation of short insertions or deletions.

5.3.2 The sequence context of *TaOR* makes mimicking the cauliflower *BoOr^{Mut}* mutation more challenging than *OsOR*

In rice, NHEJ CRISPR/Cas9 editing was used to install mutations mimicking the *BoOr^{Mut}* cauliflower mutation, which led to an increased carotenoid accumulation in dark-grown calli (Endo et al. 2019; Kim et al. 2022). In this chapter, I attempted the same in wheat; however, I found very low editing efficiencies, and the three edited plants I analysed did not show an increase in GCC (Figure 5.9). Instead, these lines showed a decrease in GCC compared to the average of the non-edited transformed controls, likely because these edits produced frameshift mutations that generated early stop codons (Figure 5.10). One reason for this difference between what was found in rice is that the intron sequence context of *OsOR* appears to be more favourable than *TaOR* to produce an editing event that retains an in-frame insertion or deletion of amino acids when removing the 'GT' splice-donor site of the third exon, as illustrated in Figure 5.13. For *TaOR*, stop codons exist in the three possible reading frames before the next 'GT' that could function as a splice-donor site. Meanwhile, for *OsOR*, only one of the reading frames produces a stop codon before the next 'GT', and *OsOR* contains more possible 'GT' splice-donor sites than *TaOR*. Therefore, only a few specific editing events in *TaOR* could produce a transcript predicted to retain an in-frame insertion or deletion of amino acids. Since the editing efficiencies of these sgRNAs were low, it was unlikely these specific mutations would be found. Interestingly, Kumagai and colleagues (2022) used the same sgRNA sequence to target *TaOR-6A* and *TaOR-6B* as mine (sgRNA 1), and they also found a low editing efficiency of 1.9%, similar to the 1.47% efficiency found here. They used this sgRNA to demonstrate editing was working in a new *in planta* CRISPR/Cas9 delivery method, but it is likely this lab originally designed this sgRNA to attempt to mimic the *BoOr^{Mut}* in wheat. This was not mentioned within this paper or subsequently published, so it is possible they also faced the same problems associated with mimicking *BoOr^{Mut}* in wheat.

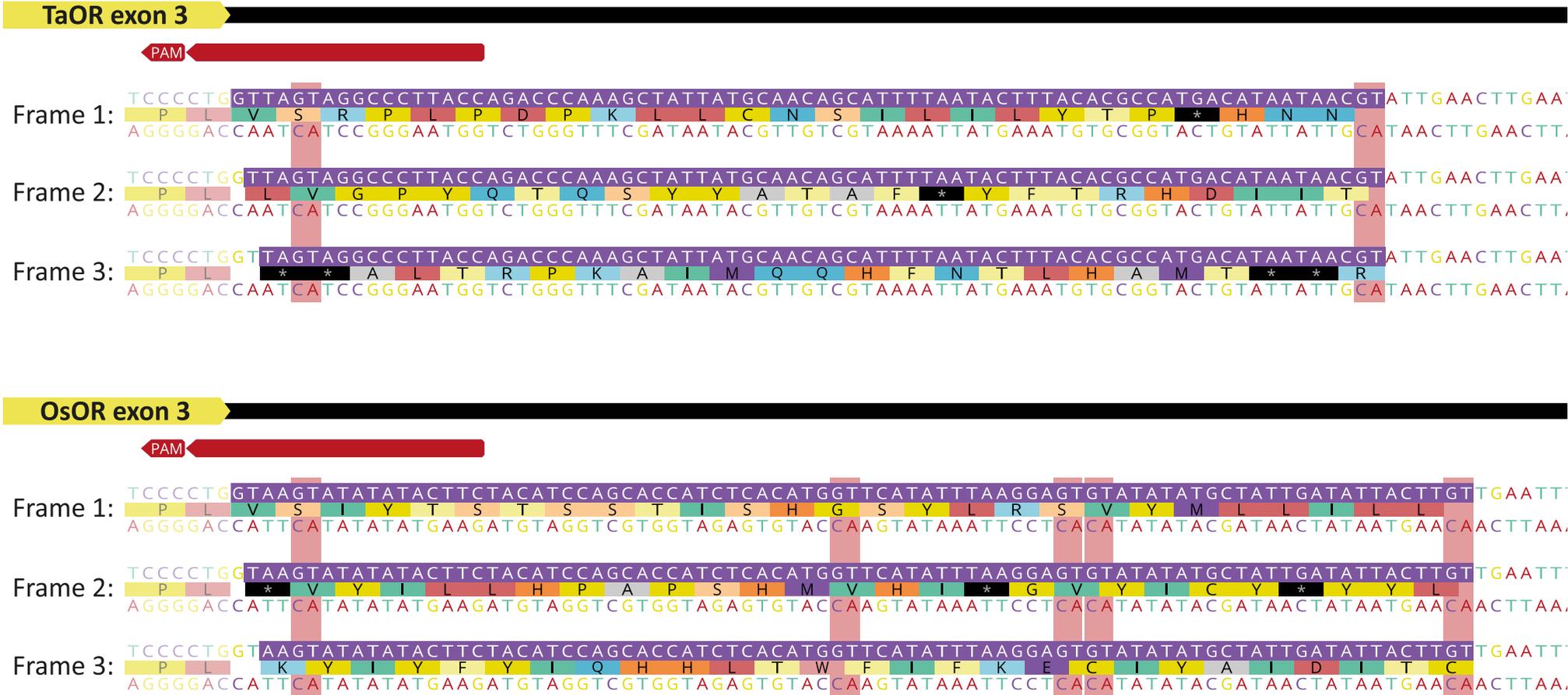


Figure 5.13 Intron sequence of *TaOR* is more challenging to mimic *BoOR^{Mut}* using CRISPR/Cas9 than *OsOR*. Possible 'GT' splice-donor sites are shown by pink vertical bars. Stop codons (black bars with an asterisk) appear in the three possible frames of *TaOR* and only one reading frame of *OsOR*. Additionally, *OsOR* contains more possible 'GT' splice-donor sites.

Recently, improvements have been made to the editing efficiency of Cas9 in plants by using a version of Cas9 which includes 13 introns within the coding sequence (Grützner et al. 2021; Lawrenson et al. 2022). This was found to greatly improve the editing efficiency of target sites with very low efficiencies. For example, no transformants were edited when targeting *AtTRY* or *AtCPC* with Cas9 lacking introns in *Arabidopsis thaliana* (*Arabidopsis*); however, between 70–100% of transformants were edited when targeting the same sites using 13-intron Cas9. As such, 13-intron Cas9 could be used to edit this third exon-intron boundary within *TaOR*, hopefully boosting editing efficiencies enough to find the desired editing events. Alternatively, a random mutagenesis approach using LbCas12a could target the sequence between the *BoOr^{Mut}* insertion site and the *OsOr^{Mut}* mutation site, as there are two PAM sites here that could be used (Figure 5.14). Cas12a utilises a T-rich ‘TTTV’ PAM and produces staggered end cuts with 4-5 nucleotide overhangs downstream of this PAM (Zetsche et al. 2015). This leads to Cas12a editing commonly producing larger deletions than Cas9, which normally generates short indels (Swarts and Jinek 2018). Therefore, LbCas12a editing is likely better suited to removing several amino acids at this site than SpCas9 editing, and LbCas12a has been demonstrated to work efficiently within plants (Schindele and Puchta 2020; Lawrenson et al. 2022).

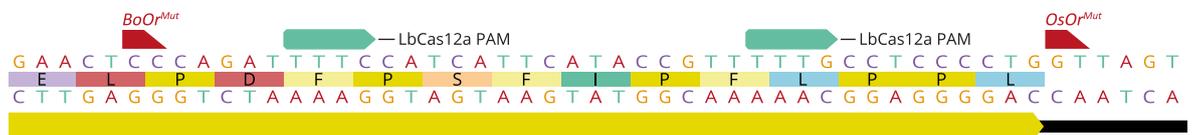


Figure 5.14 Potential LbCas12a PAM sites within the third exon of *TaOR* that could be used for a random mutagenesis approach. These LbCas12a PAMs lie between the cauliflower *BoOr^{Mut}* transposon integration site (*BoOr^{Mut}*) and the third exon-intron boundary targeted in the rice *OsOr^{Mut}* strategy to mimic the *BoOr^{Mut}* mutation (*OsOr^{Mut}*). The cut site of LbCas12a is downstream of its ‘TTTV’ PAM site.

However, it is still unclear whether mutations mimicking *BoOr^{Mut}* or targeting this third exon would cause an increased carotenoid accumulation in wheat as they do in rice. Overexpressing *TaOr^{Mut}* with some of the mutations that increased carotenoid content in the rice *OsOr^{Mut}* editing strategy would help confirm whether the same strategy works in wheat. A similar approach was taken in sweet potatoes, where the overexpression of *IbOR^{Ins}* that mimicked the *BoOr^{Mut}* mutation within the sweet potato IbOR protein led to a higher tuber carotenoid content than *IbOR^{WT}* overexpression (Kim et al. 2013; Park et al. 2015). Interestingly, overexpressing *IbOR^{His}* has also increased carotenoid content more than overexpressing *IbOR^{Ins}* within sweet potato calli (Kim et al. 2019). This suggests that these mutations act differently, and it may be possible that they could act additively to increase

carotenoid content further. Overexpressing *TaOr^{Mut}* would also show whether dark-grown wheat calli are expected to show a visible colour change, possibly allowing this screening method to be used.

5.3.3 Extended time in tissue culture may allow for improved gene editing efficiencies

During this attempt to mimic *BoOr^{Mut}* in wheat, the callus grown on selection media for 12 weeks, rather than the standard 5 weeks, still regenerated plantlets. This was unexpected because regeneration efficiency tends to reduce with the time callus spends in tissue culture (Wendy Harwood, personal communication). The *GRF4-GIF1* developmental regulators may have been responsible for this as they have been found to allow more flexibility in the standard *Agrobacterium*-mediated wheat transformation procedure (Debernardi et al. 2020). Moreover, all these plantlets were transgenic, likely due to this extended time on hygromycin media killing or severely impeding the growth of non-transformed callus, so these could not regenerate. This could allow for selecting transgenic plantlets without needing copy number analysis or confirming transgene integration.

Another possibility this might allow for is extending the time CRISPR/Cas9 has to edit transformed calli before regenerating plantlets, hopefully increasing the time for Cas9 to edit and, therefore, increasing editing efficiencies. Moreover, Cas9 and Cas12a have been shown to have higher editing efficiencies at increased temperatures in plants (LeBlanc et al. 2018; Malzahn et al. 2019; Blomme et al. 2022). Milner and colleagues (2020) demonstrated that increased heat during the selection stage of regeneration from 25.5/23.5°C to 28.5/25.5°C (day/night) increased editing efficiency in wheat when Cas9 was driven by the *ZmUbi* promoter. Prime editing efficiencies were also significantly higher in rice protoplasts grown at 37°C compared to 26°C (Lin et al. 2020), and elevated temperatures increased gene targeting efficiencies in *Arabidopsis* seedlings (Rahavi and Kovalchuk 2013). Wheat callus has been found to survive and grow at temperatures as high as 40°C, likely due to their lack of photosynthetic machinery, which is sensitive to heat stress (Benderradji et al. 2012; El-Beltagi et al. 2016). The *GRF4-GIF1* developmental regulators may also further improve regeneration efficiency at these high temperatures. Therefore, it may be possible that combining the heating of the callus with an extended time on selection media would increase gene editing efficiencies further. This could be especially useful in boosting the efficiency of editing strategies that are typically very low, such as prime editing and gene targeting. Therefore, it might be possible to increase gene editing efficiencies

further by combining growing wheat callus at higher temperatures with extending the time for gene editing to take place at the callus stage.

5.3.4 *ORLIKE* may also play a role in grain carotenoid biosynthesis in wheat

Consistent with my *Tdor* mutant lines (Section 4.2.1), most Kronos and Cadenza EMS TILLING lines with substitutions within *OR* showed a reduced GCC in the M₄ field bulks compared to the wild-type (Figure 5.12). This suggests knocking *OR* out in both a tetraploid and hexaploid background affects GCC. Moreover, the majority of these had a heterozygous EMS genotype when genotyped at the M₂ stage, so the effect could be masked by plants that were homozygous for the wild-type allele. K0685, which had mutations in both *TdOR-6A* and *TdOR-6B*, showed a reduction in GCC of 56.8% compared to the Kronos control. The double *Tdor* mutant plants I analysed in Section 4.2.1 had a reduction of 33.8% compared to the *Tdor* wild-type plants. This could suggest that the effect of knocking out *OR* function is greater under field conditions compared to glasshouse-grown lines. Alternatively, TILLING lines may suffer from lower carotenoid content due to background mutations stressing these plants; therefore, comparing them to a Kronos wild-type control may result in a greater decrease in GCC. It would be good to include TILLING control lines that do not contain mutations within *OR* or *ORLIKE* to compare the variation of carotenoid content in these lines. The mutations in K0685 may also be more severe than those in *Tdor* mutants.

The Kronos and Cadenza TILLING lines with substitutions within *ORLIKE* also showed a reduced GCC in the field bulks (Figure 5.12), suggesting wheat *ORLIKE* also plays a role in carotenoid biosynthesis. This is consistent with work in *Arabidopsis*, finding that *OR* and *ORLIKE* play similar roles in carotenoid biosynthesis by stabilising *PSY* (Zhou et al. 2015; Sun et al. 2023b). Additionally, *ORLIKE* was also found to interact with and stabilise *MAGNESIUM CHELATASE SUBUNIT I (CHLI)* in the chlorophyll biosynthesis pathway like *OR* does (Sun et al. 2023b), suggesting similar roles for these proteins. It would be interesting to see if a *Tdor Tdorlike* mutant line has an even greater reduction in GCC in wheat, and future work could investigate this possibility. Two *ORLIKE* TILLING lines (C0773 and C1233) had substitutions at the equivalent 'golden SNP' residue in *TdORLIKE*; however, no GCC difference was found for either of these lines (Figure 5.12). This suggests that substitutions at the equivalent 'golden SNP' residue in *ORLIKE* do not affect GCC. Consistent with this, the overexpression of *AtORLIKE^{His}* with the 'golden SNP' arginine-to-histidine substitution at the equivalent residue resulted in no change to carotenoid content in *Arabidopsis* dark-grown calli (Yuan et al. 2015).

5.3.5 The E141K substitution could be a novel *OR* gain-of-function mutation that increases grain carotenoid content

Previous natural mutations within the *OR* gene have been associated with increased carotenoid accumulation (Lu et al. 2006; Tzuri et al. 2015; Ellison et al. 2018). In this chapter, I selected EMS TILLING lines with substitutions close to where these natural mutations are in the OR protein and screened field bulks of these lines. This identified the line K4596, with a 35.8% increased total GCC, containing an E141K substitution located between the 'golden SNP' residue and the location of the *BoOr^{Mut}* retrotransposon integration site (Figure 5.11b). However, these results are just from a single replication, which must be repeated on the lines grown under glasshouse conditions to confirm this effect. Moreover, these TILLING lines contain a lot of other background mutations. Although K4596 contained no EMS mutations in the carotenoid biosynthesis genes I searched for, this cannot rule out that background mutations in other genes were responsible for the GCC increase. Future work backcrossing K4596 to Kronos could be done to reduce the proportion of background mutations. Nevertheless, the increase in GCC associated with K4596 is quite large, suggesting that the E141K substitution within *TdOR* leads to an increase in GCC within wheat. If this is the case, it is especially exciting because this material is from an EMS-mutagenised population; therefore, it does not come under laws and regulations surrounding material produced through transgenesis or gene editing. As such, it could be used directly within global breeding programmes to improve carotenoid content within wheat grains. Furthermore, Kronos is a durum wheat variety, and this would facilitate its introduction into durum wheat breeding programmes where increased yellow flour colour is a breeding target.

If the E141K substitution is responsible for this GCC increase, the effect might be greater in lines homozygous for the E141K substitution. The M₂ EMS genotype of the E141K substitution within K4596 was heterozygous (Table 5.2), so the M₄ field bulk was segregating and contained homozygous mutants and homozygous wild-type plants for the E141K substitution. Therefore, the wild-type plants might be masking the effect of the E141K substitution. Future work analysing genotyped K4596 TILLING lines will show the effect of the E141K substitution in homozygous lines, which are currently being regrown (March 2024). Furthermore, K4596 TILLING lines can be selected in the first generation that are E141K homozygous mutants or E141K homozygous wild-types due to this segregating EMS genotype. If the effect on GCC is only seen in E141K homozygous mutant lines, this will

provide good evidence that the E141K substitution is responsible for a GCC increase rather than any background mutations.

As well as the 'golden SNP' arginine-to-histidine substitution, other single amino acid changes within OR proteins have also resulted in carotenoid accumulation, so it is conceivable that this single amino acid E141K substitution could be doing the same in wheat TdOR. In carrots (*Daucus carota*), a serine-to-leucine substitution within DcOR is associated with increased carotenoid accumulation in the non-photosynthetic taproot (Ellison et al. 2018). It was hypothesised that this has been selected for during carrot domestication to increase carotenoid formation and storage in the taproot. Additionally, one of the aberrant transcripts produced in gene-edited rice calli mimicking *BoOr^{Mut}* that resulted in increased carotenoid accumulation was a deletion of a single amino acid within OsOR (Endo et al. 2019). At the 'golden SNP' residue, Yuan and colleagues (2015) demonstrated that an arginine-to-alanine substitution (*AtOR^{Ala}*) could also increase carotenoid accumulation in *Arabidopsis*, again suggesting other *OR* mutations or substitutions can produce dominant gain-of-function mutations.

6 General Discussion

6.1 Summary of this thesis

This thesis aimed to explore and enhance the genetic diversity of grain carotenoid content in wheat. To do this, I characterised the Watkins global landrace collection (Chapter 3), identified new carotenoid-associated genetic diversity within the Watkins tetraploid collection (Chapter 3), investigated the function of *ORANGE* (*OR*) and the ‘golden SNP’ in wheat (Chapter 4), and attempted to engineer diversity into wheat *OR* to increase grain carotenoid content (Chapter 5). Throughout these chapters, I investigated the questions:

- What is the grain carotenoid diversity within the Watkins global landrace collection?
- Can novel allelic diversity associated with grain carotenoid content be identified within the Watkins tetraploid collection?
- Is *OR* involved in grain carotenoid biosynthesis in wheat, and if so, does the ‘golden SNP’ affect grain carotenoid accumulation?
- Can diversity in the wheat *OR* gene be engineered to increase grain carotenoid content?

In this chapter, I discuss how my findings address these questions. Then, I discuss the implications of my findings concerning two main topics: approaches for breeding carotenoid biofortified wheat and the potential role of allele replacement in pre-breeding.

6.1.1 Carotenoid diversity within the Watkins global landrace collection

To answer what is the grain carotenoid diversity in the Watkins global landrace collection, I analysed the grain carotenoid content (GCC) of the Watkins tetraploid collection using my high-throughput yellow pigment content (YPC) method and high-performance liquid chromatography (HPLC; Section 3.2.1). I found high variation within the panel for GCC, and I identified some accessions with a high total GCC and a high content of desirable carotenoid compounds for human health. These accessions warrant further investigation due to their potential for wheat carotenoid biofortification. This was the first time GCC had been analysed using HPLC on a global tetraploid landrace collection and in a collection this large. I also investigated the carotenoid content of the Watkins hexaploid core collection using HPLC (Section 3.2.2), which was the first time a diversity collection of hexaploid landraces had been analysed for GCC using HPLC. Doing so allowed me to compare the GCC of bread wheat and durum wheat landraces, revealing that the tetraploid landraces had a slightly higher total GCC, although this difference was not great. This suggests that historically,

bread wheat and durum wheat landraces had a similar GCC, aligning with evidence that the higher GCC in modern durum wheat results from plant breeding over the last 30 years rather than durum wheat inherently having a higher GCC.

6.1.2 Allelic diversity within the Watkins tetraploid collection

To see if novel allelic diversity associated with GCC could be identified in the Watkins tetraploid collection, I conducted two genome-wide association studies (GWAS) at different resolutions to find marker-trait associations (MTAs) and quantitative trait loci (QTLs) associated with YPC, α -carotene, β -carotene, lutein, zeaxanthin and total GCC (Section 3.2.3). In total, 15 MTAs were identified with the 35K Axiom[®] Wheat Breeder's Array GWAS and 14 QTLs were identified with the high-resolution GWAS. Of these QTLs, three overlapped with previously identified carotenoid-associated QTLs, and 11 were novel QTLs for carotenoid traits. One of these QTLs (*zea_7A*) overlapped with two previously identified carotenoid-associated QTLs (Colasuonno et al. 2014, 2017a) and was close to a significant MTA previously associated with grain colour (Rathan et al. 2022), demonstrating the agreement of this GWAS with previous literature. The novel QTLs could represent new allelic variation for carotenoid content only present within the Watkins tetraploid collection. Therefore, this may be novel variation that can be brought into durum breeding programmes for increasing GCC. Within three of the carotenoid-associated QTLs identified here, candidate genes were found with orthology to carotenoid biosynthesis within other plants: *bcaro_3A* had an orthologue of *AtCCD8*, *YPC_4A* had six orthologues of *AtZEP* and *zea_7A* had an orthologue of *AtDXR*. The orthologues of *AtCCD8* and *AtZEP* have not previously been studied or associated with wheat carotenoid biosynthesis, and investigating their function could open new doors for exploring and understanding carotenoid biosynthesis in wheat. Additionally, this study was the highest-resolution GWAS run on carotenoid traits to date. As such, the allelic diversity identified through this GWAS offers novel and noteworthy insights into wheat carotenoid biosynthesis. However, since the results are based on data from a single year and environment, conducting the analysis with data from an additional year would enhance the reliability of these findings and allow for stable QTL to be found.

6.1.3 The function of *OR* and the 'golden SNP' in wheat

OR is a chaperone protein that has been found to play a role in plant carotenoid biosynthesis by post-transcriptionally stabilising PHYTOENE SYNTHASE (PSY), increasing its protein activity (Zhou et al. 2015); however, it has not been studied in wheat. To investigate whether

the *OR* gene is involved in carotenoid biosynthesis within wheat, I generated complete knockouts of *Tdor* within durum wheat and examined the GCC of these (Section 4.2.2). These knockout lines had a 33.8% lower total GCC than the wild-type controls, suggesting that *OR* plays a role in grain carotenoid biosynthesis in wheat. In Chapter 5, I selected several Kronos and Cadenza lines with EMS mutations in *OR* to hopefully identify a gain-of-function *OR* mutation with increased GCC. The field bulks of these lines, except for K4596, had reduced total GCC compared to the control, further supporting the role of *OR* in grain carotenoid biosynthesis in wheat. These results are consistent with studies that show knocking out *OR* reduces carotenoid content in *Arabidopsis thaliana* and melons (Zhou et al. 2015; Chayut et al. 2017; Sun et al. 2023b). Presumably, *OR* in wheat stabilises PSY as it does in other plants, and future work could investigate the protein level of PSY within the grain of these *Tdor* mutants to confirm this.

The 'golden SNP' is a single nucleotide polymorphism (SNP) in the melon *CmOR* gene that leads to a massive overaccumulation of carotenoids within melon fruit flesh (Tzuri et al. 2015). This overaccumulation is thought to be due to an increase in the number of chromoplasts in this tissue, a carotenoid storage plastid. However, the effect of the 'golden SNP' in wheat *OR* was unknown. To investigate whether the same 'golden SNP' installed in the wheat *OR* gene would affect carotenoid accumulation, I overexpressed both the wild-type *TaOR* sequence (*TaOR^{WT}*) and *TaOR* with the 'golden SNP' installed in it (*TaOR^{His}*), and then compared the GCC between these lines (Section 4.2.3). The *TaOR^{His}* overexpression lines had a 21.6% higher total GCC compared to the *TaOR^{WT}* overexpression lines. This suggests that the 'golden SNP' substitution in *TaOR* does increase the carotenoid content in wheat grains. Future work should investigate whether an increase in the number of carotenoid-sequestering bodies in the grain accompanies this increase in total GCC because the 'golden SNP' is thought to increase sink capacity and promote chromoplast biogenesis.

This was the first time the 'golden SNP' has been demonstrated to increase grain carotenoid content in cereals. *OsOR^{His}* was previously overexpressed in rice grains, but this did not lead to an increase in GCC (Jung et al. 2021). This is likely because, unlike in wheat, rice has no active grain carotenoid biosynthesis due to a lack of *PSY* expression, so increasing carotenoid storage in rice grains would not increase GCC (Beyer et al. 2002). Based on my results and how the 'golden SNP' is thought to act, it is reasonable to hypothesise that the overexpression of *OR^{His}* in rice grains may increase carotenoid accumulation when the flux in the grain is turned on. This has implications for the Golden Rice® project, where the

carotenoid flux in rice grains was switched on through the grain expression of *PSY* and *CRTI* (Beyer et al. 2002; Paine et al. 2005). Golden Rice® has massively increased β -carotene content within the rice grain; however, it suffers from problems of high degradation of β -carotene during storage (Gayen et al. 2015). The β -carotene degradation rate is even greater under high temperatures and humidity, such as in India or Southeast Asia, where Golden Rice® was hoped to have the most impact (Bollinedi et al. 2019). This is likely because amyloplasts are unsuited to storing provitamin A (PVA) carotenoids like β -carotene, as they generally accumulate low levels of macular carotenoids (Lopez et al. 2008; Wurtzel et al. 2012). On the other hand, chromoplasts are adept at accumulating and storing all kinds of carotenoids (Sun et al. 2018). If the 'golden SNP' promotes chromoplast biogenesis within cereal grains, this would be incredibly beneficial for improving β -carotene storage and stability in Golden Rice®. Again, this suggests that future work should investigate whether there is an increase in grain chromoplast biogenesis within the *TaOR^{His}* overexpression lines.

6.1.4 Engineering diversity into wheat *OR* to increase grain carotenoid content

Based on my results showing that the 'golden SNP' increases GCC in wheat, I next asked whether I could engineer diversity in the wheat *OR* gene to increase GCC. To do this, I originally attempted to use gene editing (Section 5.2.1). First, I tried installing the 'golden SNP' within endogenous *TaOR* using prime editing; however, I found no evidence of germline or somatic editing events. This was likely due to the low editing efficiency of the original Plant Prime Editor 2 protein I used and because of its missing nuclear localisation signal. The new advancements in prime editing may allow the 'golden SNP' to be successfully installed into *TaOR* to increase GCC (Ni et al. 2023). Next, I tried mimicking the *BoOr^{Mut}* mutation within *TaOR* using CRISPR/Cas9 editing (Section 5.2.2). This had previously been achieved in the rice *OsOR* gene, which increased the carotenoid accumulation of dark-grown calli (Endo et al. 2019; Kim et al. 2022). However, I was not able to mimic the *BoOr^{Mut}* mutation in *TaOR* using CRISPR/Cas9, likely because *TaOR* has a poor sequence context for doing this compared to *OsOR*.

I next attempted to find gain-of-function ethyl methanesulfonate (EMS) mutations located close to the 'golden SNP' and *BoOr^{Mut}* integration site (Section 5.2.3). I hypothesised that these mutations might be within an important functional domain of *TaOR* that the 'golden SNP' modifies. I measured the GCC of a number of field-grown Kronos and Cadenza TILLING

lines that contained these EMS mutations, and I identified a line, K4596, with a 35.8% increase in total GCC compared to the wild-type Kronos field control. K4596 contains an E141K substitution within TaOR-6A, and this could be an exciting novel *OR* mutation that increases carotenoid accumulation similar to the ‘golden SNP’ or *BoOR^{Mut}*. Since these GCC measurements were from a single replicate of K4596, future work should further study this line and backcross the E141K substitution into Kronos and Miradoux, a modern durum wheat cultivar. Excitingly, since the E141K substitution was discovered in an EMS-mutagenised population and not developed through gene editing, it can be used directly within breeding programmes to improve carotenoid content within wheat grains.

6.2 Future approaches for breeding carotenoid biofortified wheat

Increased consumer awareness of dietary health has opened the new possibility to breed high GCC bread and durum wheat varieties that could be marketed for their health benefits (Mancino and Kuchler 2012; Lockyer and Spiro 2020). Moreover, because of the high widespread consumption of wheat-based products, even small increases in GCC may have large impacts on improving human health worldwide. This is especially true of improving the content of PVA carotenoids, which are typically found only at a low content within durum and bread wheat grains. In this section, I explore potential future work building upon the findings of my thesis, aiming to facilitate the breeding of wheat with a high GCC.

6.2.1 Breeding for high grain carotenoid bread and durum wheat

Modern bread wheat cultivars have lower GCC compared to modern durum wheat cultivars (Shewry and Hey 2015). This is attributed to the past three decades of durum wheat breeding that aimed at increasing flour yellowness to meet consumer preferences (Digesù et al. 2009). For bread wheat breeding, the opposite has occurred due to the historic consumer preference for white bread (Hidalgo et al. 2006; Ziegler et al. 2015). Correspondingly, the GCC of the durum and bread wheat landraces within the Watkins global landrace collection are similar (Section 3.2.2) since this collection was assembled before these opposing bread and durum wheat breeding targets were introduced. Consequently, by comparing modern cultivars of durum wheat with high GCC to older cultivars with lower GCC, the allelic diversity responsible for this increased GCC might be identified. This diversity could then be introduced into bread wheat through interspecific crosses to increase the GCC here. Requena-Ramírez and colleagues (2023) previously achieved this by introducing high GCC alleles of durum wheat *TdPSY1* into bread wheat

through interspecific crossing, which led to an increase in total GCC of 16-23%. Nevertheless, it might be more simple to introduce already existing bread wheat diversity for high GCC into modern cultivars.

Since the Watkins hexaploid collection was assembled, breeding for whiter flour likely resulted in a loss of alleles for high GCC from modern bread wheat germplasm. Therefore, this landrace collection may be a good source of genetic diversity for breeding for high GCC in bread wheat. Consistent with this, some accessions within the Watkins hexaploid core collection had a high total GCC (Section 3.2.2). For instance, WAT1190149 had a total GCC of 1.213 $\mu\text{g/g}$, similar to the total GCC of Miradoux (1.274 $\mu\text{g/g}$), a high carotenoid modern durum wheat cultivar (Giambanelli et al. 2013). These accessions may prove to be very useful pre-breeding material, and future work could cross these with modern bread wheat cultivars. Additionally, only the core collection of the Watkins hexaploid collection was screened for GCC here. While screening the whole 828 accessions of the Watkins hexaploid collection might be unfeasible using HPLC, the collection could be screened using the high-throughput YPC method to identify accessions with high YPC content that could then be further analysed with HPLC. Doing so could identify bread wheat accessions with very high total GCC, like WAT1180004 in the tetraploid collection, that could be further used as pre-breeding material.

As previously discussed, modern durum wheat cultivars have high GCC due to breeding targets for increased pasta yellowness. However, some Watkins tetraploid collection accessions were identified with higher total GCC than Miradoux (1.274 $\mu\text{g/g}$), and one accession (WAT1180004) had double the total GCC of Miradoux (2.557 $\mu\text{g/g}$). Therefore, the Watkins tetraploid collection may also act as an extremely useful source of genetic material for increasing the total GCC of modern durum cultivars further. Additionally, 11 carotenoid-associated QTLs identified within the Watkins tetraploid collection had not been previously associated with carotenoid traits. These could represent allelic variation unique to the Watkins tetraploid collection that could be exploited within breeding programmes. Future work should investigate these QTLs further to identify novel stable QTLs associated with an increase in GCC.

6.2.2 Utilising variation within the carotenoid biosynthesis pathway of the Watkins global landrace collection

In addition to high total GCC accessions, I identified bread and durum wheat accessions with a high content of the desirable PVA carotenoids, α -carotene and β -carotene, which could also be used as useful pre-breeding material. I also identified accessions with a high proportion of β - β branch carotenoids, which is of specific interest for PVA biofortification because β -carotene is located within the β - β branch of the pathway. Within these accessions, the β -carotene hydroxylase genes (*HYD1* and *HYD2*) could be knocked out to prevent the degradation of β -carotene further down the pathway (Figure 1.1). However, because these genes are upstream from abscisic acid biosynthesis, this could cause pleiotropic effects on plant growth and physiology. Previously, *HYD1* and *HYD2* expression was targeted using endosperm-specific RNAi silencing, which increased grain β -carotene content by 10.5-fold (Zeng et al. 2015), and a similar approach could be taken for these high β - β branch accessions. An alternative to using a transgenic RNAi approach could be to use gene editing to target regulatory regions within *HYD1* and *HYD2* to affect their expression. Up- and down-regulation of gene expression has been achieved by introducing mutations within different regulatory control elements such as promoters, introns, alternative splicing sites and untranslated regions (Dong 2024). For example, a deletion using CRISPR/Cas9 within the 5' UTR region of *CAROTENOID ISOMERASE (CRTISO)* resulted in a downregulation of *CRTISO* expression within tomato (Lakshmi Jayaraj et al. 2021). This produced an intermediate phenotype between a *crtiso* knockout and wild-type plant. Similarly, editing events within regulatory regions of *HYD1* and *HYD2* could be sought that specifically reduce the expression within the grain so as not to affect their role in the biosynthesis of important downstream molecules in other parts of the plant.

6.2.3 Increasing grain carotenoid content in wheat through transgenesis

Increasing GCC within wheat grains could also be achieved through a transgenic approach by overexpressing *PSY1* with *TaOR^{WT}* or *TaOR^{His}*. In rice grains, the overexpression of *AtOR^{WT}*, *ZmPSY1* and *PaCRTI* led to 2.1- to 4.6-fold higher total GCC than just the overexpression of *ZmPSY1* and *PaCRTI*, likely due to *AtOR^{WT}* post-transcriptionally stabilising the activity of *ZmPSY1* (Bai et al. 2016). Previously, *ZmPSY1* and *CRTI* were overexpressed within wheat grains, which led to a 10.8-fold increase in total carotenoid content (Cong et al. 2009), and this could be combined with *TaOR^{WT}* in an attempt to increase this further. If the increase in GCC associated with *TaOR^{His}* overexpression (Section 4.2.3) is due to an

increased carotenoid sink strength, then combining *ZmPSY1* and *TaOR^{His}* overexpression may also further improve GCC than just *ZmPSY1* or *TaOR^{His}* overexpression alone. This would hopefully push more flux into the pathway and increase the storage of carotenoids within the wheat grains. Increasing sink strength with *OR^{His}* overexpression has been suggested to enhance the stability of carotenoids during processing and post-harvest storage (Li et al. 2012). This approach could prove advantageous in wheat grains since carotenoids here are susceptible to degradation during processing (Colasuonno et al. 2017a). Furthermore, the *GRF4-GIF1* developmental regulators open up the possibility of transforming modern cultivars or accessions from the Watkins global landrace collection directly because they reduce the genotype dependency associated with *Agrobacterium*-mediated transformation. Therefore, accessions with already high GCC could be transformed with *TaOR^{His}* to improve their grain carotenoid storage and stability. However, as highlighted by the problem of low grain yield associated with *GRF4-GIF1* overexpression (Section 4.2.3), having an inducible *GRF4-GIF1* system to express these developmental regulators only during regeneration may be necessary to avoid pleiotropic phenotypes.

Moreover, combining ‘push’, ‘pull’ and ‘block’ strategies of carotenoid biofortification (Figure 1.5) could be achieved by the endosperm-specific overexpression of *ZmPSY1*, *TaOR^{His}* and RNAi to silence *HYD1*. This would increase the flux going into the carotenoid biosynthesis pathway, improve the carotenoid storage and sequestration and reduce the enzymatic degradation of carotenoids within the grain. Additionally, knocking out or silencing *LCYE* may also be carried out to push flux into the β - β branch of the carotenoid biosynthesis pathway. Combining these various strategies has the possibility of greatly improving β -carotene content within wheat grains and could lead to biofortified wheat with very high PVA activity. This could be very beneficial in combating vitamin A deficiency (VAD) in regions of both high wheat consumption and high prevalence of VAD, such as in sub-Saharan African countries like Ethiopia (Li et al. 2024).

6.2.4 Using the E141K substitution for increasing grain carotenoid sink strength

In Chapter 5, I attempted to install gain-of-function mutations within endogenous *TaOR* using gene editing; however, these attempts were unsuccessful. Instead, I discovered a novel EMS mutation (E141K) within *TaOR*, which led to a 35.8% increase in total GCC. Work is ongoing to confirm the effect of E141K since this result came from a single replicate of a field-grown bulk. If E141K is confirmed to be associated with an increase in GCC, it is

reasonable to assume that it may enhance sink strength, similar to the way other gain-of-function *OR* mutations have been found to function (Sun et al. 2018; Osorio 2019). The E141K substitution could be backcrossed into high GCC modern cultivars and accessions of the Watkins tetraploid collection to improve their carotenoid sink strength. As discussed previously, improving the storage of carotenoids within wheat grains is very desirable for reducing carotenoid degradation. Since this is an EMS-induced mutation and not from a transgenic or gene-edited approach, this would not come under the same regulations and restrictions as these approaches and could be used directly in global breeding programmes.

6.3 GWAS to gene editing: a model for allele replacement in plant breeding

A naturally occurring allele of *CmOR*, responsible for a massive accumulation of carotenoids within melon fruit flesh, was identified within a biparental mapping population of orange- and green-flesh melons (Tzuri et al. 2015). This overaccumulation was traced back to a causal G-to-A substitution within *CmOR*, named the ‘golden SNP’. In my thesis, I demonstrated *TaOR*’s role in wheat carotenoid biosynthesis by knocking out the gene (Section 4.2.2). Subsequently, I confirmed that incorporating the ‘golden SNP’ into the *TaOR* gene increased GCC through overexpression (Section 4.2.3). Using gene editing, I then attempted to install the ‘golden SNP’ within the endogenous *TaOR* gene to replace the original *TaOR*^{WT} allele with a new *TaOR*^{His} allele that increases GCC (Section 5.2.1). This approach serves as an illustrative model for how precise gene editing tools, such as prime editing, might be used in the future to utilise allele replacement for pre-breeding.

6.3.1 A framework for using allele replacement in pre-breeding

Transitioning from the specific case of the ‘golden SNP’, I will now outline a framework of how allele replacement could be utilised in pre-breeding, integrating modern genomics and gene editing technologies. First, GWAS would be employed to identify natural genetic variation associated with traits of agronomic importance within crop diversity collections (Figure 6.1a). New high-resolution GWAS, such as that used in Chapter 3 to identify carotenoid-associated QTLs, will be especially useful here. This is due to their ability to identify small QTL regions, which facilitates the identification of causal allelic variation. Large diversity collections will prove to be invaluable sources of new allelic variation. For bread wheat, thousands of high-resolution QTLs for major agronomic traits have already been identified within the Watkins hexaploid collection, many unique to this collection

(Cheng et al. 2023). Moreover, the exploration of the Watkins tetraploid collection is underway, promising further discovery of high-resolution QTLs. Germplasm banks like those maintained at CIMMYT and ICARDA, which house around 185,000 accessions encompassing wild relatives, landraces and modern cultivars of wheat, offer additional resources to be searched (Sansaloni et al. 2020). Once QTLs for agronomic traits are identified, candidate genes within these QTLs would be pinpointed through the *in silico* analysis of available genome sequences or additional fine mapping to narrow down the associated intervals (Figure 6.1b). The alleles of candidate genes would then undergo functional characterisation (Figure 6.1c), either through mutagenesis to knock these out or via transgenic expression to ascertain their role.

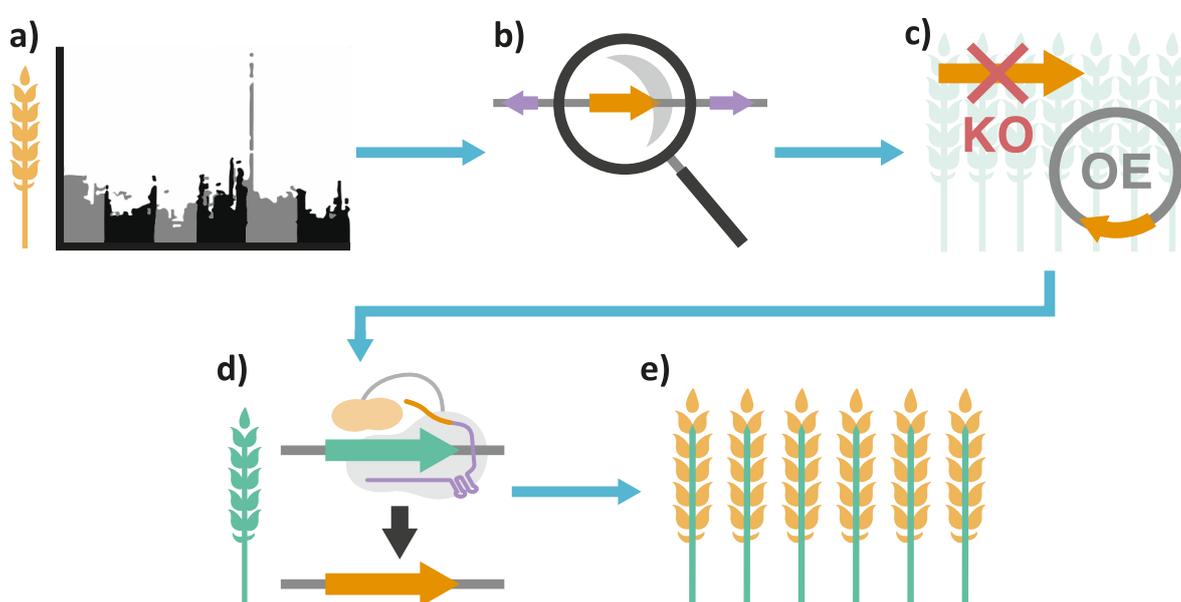


Figure 6.1 Framework for using allele replacement in pre-breeding. (a) Quantitative trait loci (QTLs) associated with traits of agronomic interest are identified within diversity collections. (b) QTL regions are examined using *in silico* analyses to identify candidate genes. (c) Causal gene and allelic variation are identified using functional analysis of genes. (d) Gene editing is used to replace the endogenous allele of modern cultivars with the superior agronomic allele. (e) Phenotypes of the improved cultivars are investigated. KO=gene knockout, OE=gene overexpression.

Finally, modern cultivars would be enhanced by replacing their endogenous alleles with those identified as causative for superior agronomic traits (Figure 6.1d–e). This replacement process could be achieved using precise gene editing tools, such as prime editing, that enable the precise manipulation of gene sequences at specific DNA sites. Currently, prime editing is limited to the replacement and modification of short sequences; however, advancements in gene editing technologies are expanding these capabilities to include the replacement of entire genes. An example of such progress is PrimeRoot, a novel tool enabling the precise integration of large DNA segments into plant genomes (Sun et al.

2023a). This technique employs prime editing to introduce recombination sites into the genome and then uses a tyrosine recombinase to insert or replace a desired DNA fragment at the newly established recombinase site. With the improvements made to prime editing efficiency in wheat (Ni et al. 2023), PrimeRoot has the potential to revolutionise wheat breeding by facilitating the replacement of alleles linked to beneficial agronomic traits. This could, for example, enable the replacement of *TaPSY1* alleles within bread wheat with durum wheat *TdPSY1* alleles that increase GCC, thereby bypassing the need for interspecific crossing and backcrosses.

Alternatively, homology-directed repair (HDR) presents another approach for allele replacement. HDR has been demonstrated in cereals like rice, maize, barley and wheat (Li et al. 2019; Barone et al. 2020; Lu et al. 2020; Lawrenson et al. 2021; Luo et al. 2023); however, the efficiency of HDR within monocots is currently very low. Future improvements in HDR and PrimeRoot's editing efficiencies will likely significantly enhance their applicability in allele replacement, paving the way for more precise and efficient crop improvement strategies.

6.3.2 The benefits of using gene editing for allele replacement

The main advantages of using gene editing for allele replacement are the accelerated breeding timelines and the precision with which advantageous alleles can be introduced to modern cultivars during pre-breeding. Traditional backcrossing is a slow and laborious process, often requiring multiple backcrossing generations to introduce a desired QTL region with an acceptable level of linked foreign genetic variation. For instance, when introducing desired material into an adapted modern cultivar, 10 backcrosses with the modern cultivar are required to retain 99.951% of the modern cultivar. In contrast, gene editing would facilitate the rapid integration of beneficial alleles into modern cultivars without lengthy backcrossing steps. This improved speed is especially pertinent due to the pressing demands of adapting our crops to a rapidly changing climate. Furthermore, backcrossing is marred by linkage drag, where undesirable traits may be inadvertently introduced into modern cultivars from the parent containing the desired allelic variation. For example, after 10 backcrosses with the modern cultivar, there would still be 0.049% of the other parent's genome. For bread wheat, 0.049% is around 8 megabases of foreign DNA, which may contain hundreds of undesirable non-adapted genes. Gene editing circumvents this issue by facilitating the precise insertion of only the identified causal

sequences, ensuring the enhancement of agronomic traits without the accompanying burden of unwanted genetic material.

Allele replacement using gene editing may also enable plant breeders to harness the knowledge gained from years of QTL searches for agronomically important allelic diversity. The wealth of QTLs and causal allelic variation identified through these studies may be directly exploited using this, helping to bridge the gap between discovery and application. Additionally, efficient allele replacement could offer an unprecedented opportunity to test and utilise the vast diversity contained within germplasm collections, such as the Watkins global landrace collection, without the need for backcrossing. Genetic variation identified within these diversity collections could be unlocked for plant breeding through the precise introduction or replacement of agronomically important allelic variation into modern cultivars.

6.3.3 The UK regulatory framework governing gene editing technologies

In the context of using gene editing in plant breeding, the regulatory framework governing gene editing technologies is a crucial consideration. The UK's 'Genetic Technology (Precision Breeding) Act', introduced in March 2023, allows for the release of precision bred organisms (UK Parliament 2023). It defines precision bred organisms as those with modifications achievable through traditional breeding, differentiating between these and transgenic organisms. Under this definition, cultivars developed through allele replacement that mimic outcomes achievable by conventional breeding and backcrossing would be classified as precision bred organisms. This classification would exempt them from the stringent regulations applied to transgenic plants, streamlining their path to farmers' fields and the market.

Under this legislation, the introduction of alleles from other species would likely still be categorised as transgenic. Consequently, it is interesting to consider how a wheat cultivar with the 'golden SNP' installed in the endogenous *OR* gene would be regulated. Although the allele was initially identified and originated from melon *CmOR*, only the single causal SNP would be introduced through gene editing. This G-to-A substitution could be achieved through traditional EMS mutagenesis, which this UK legislation defines as a traditional breeding technique. Therefore, a wheat cultivar modified with a single SNP would likely be considered a precision bred organism, and its release would be allowed under this new legislation. Therefore, this new approach to regulation facilitates the deployment of gene

editing innovations in crop improvement with a more accommodating legal framework, emphasising the importance of aligning gene editing advancements with regulatory policies for improving agriculture.

6.4 Concluding Statement

Overall, this thesis represents a significant step forward in our understanding of the carotenoid diversity present within wheat germplasm and the genes involved in the grain carotenoid biosynthesis pathway. To achieve this, I explored the carotenoid diversity of the Watkins global landrace collection and identified new carotenoid-associated QTLs within the Watkins tetraploid collection. These QTLs warrant further investigation as I found evidence to suggest that they contain genes not previously implicated in grain carotenoid biosynthesis in wheat. I also found that *OR* is involved in grain carotenoid biosynthesis in wheat, consistent with its role in other plants. Furthermore, I discovered that the 'golden SNP' installed in the wheat *TaOR* gene increases total grain carotenoid content, representing the first evidence of this in cereal crops. Utilising EMS-mutagenised populations, I identified a promising gain-of-function mutation within the durum wheat *TdOR* gene, which holds significant potential for improving carotenoid storage and stability within the grain. The findings from this research not only advance our understanding of the genetic mechanisms underpinning carotenoid biosynthesis in wheat but also set the stage for breeding strategies to further increase grain carotenoid content in wheat. Additionally, through integrating genomics, GWAS, and precise gene editing tools, I have outlined a path forward for rapidly incorporating beneficial traits into modern cultivars. This could contribute to the rapid development of nutritionally enhanced cultivars and environmentally resilient wheat varieties.

7 References

- AACC.** Determination of Pigments AACC 14-50. AACC International Approved Methods. 2009:2–3. <https://doi.org/10.1094/aaccintmethod-14-50.01>
- Abdel-Aal E-SM, Young JC, Rabalski I, Hucl P, and Fregeau-Reid J.** Identification and Quantification of Seed Carotenoids in Selected Wheat Species. *J Agric Food Chem.* 2007;**55**(3):787–794. <https://doi.org/10.1021/jf062764p>
- Agudelo D, Carter S, Velimirovic M, Durringer A, Rivest JF, Levesque S, Loehr J, Mouchiroud M, Cyr D, Waters PJ, et al.** Versatile and robust genome editing with *Streptococcus thermophilus* CRISPR1-Cas9. *Genome Research.* 2020;**30**(1):107–117. <https://doi.org/10.1101/gr.255414.119>
- Alder A, Jamil M, Marzorati M, Bruno M, Vermathen M, Bigler P, Ghisla S, Bouwmeester H, Beyer P, and Al-Babili S.** The Path from β -Carotene to Carlactone, a Strigolactone-Like Plant Hormone. *Science.* 2012;**335**(6074):1348–1351. <https://doi.org/10.1126/science.1218094>
- Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A, and Nielsen H.** Detecting sequence signals in targeting peptides using deep learning. *Life Sci Alliance.* 2019;**2**(5):e201900429. <https://doi.org/10.26508/lsa.201900429>
- Alvarez JB, Martín LM, and Martín A.** Genetic variation for carotenoid pigment content in the amphiploid *Hordeum chilense* \times *Triticum turgidum* conv. *durum*. *Plant Breeding.* 1999;**118**(2):187–189. <https://doi.org/10.1046/j.1439-0523.1999.118002187.x>
- Amorim-Carrilho KT, Cepeda A, Fente C, and Regal P.** Review of methods for analysis of carotenoids. *TrAC Trends in Analytical Chemistry.* 2014;**56**:49–73. <https://doi.org/10.1016/j.trac.2013.12.011>
- Anzalone AV, Koblan LW, and Liu DR.** Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology.* 2020;**38**(7):824–844. <https://doi.org/10.1038/s41587-020-0561-9>
- Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, et al.** Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature.* 2019;**576**(7785):149–157. <https://doi.org/10.1038/s41586-019-1711-4>
- Atienza SG, Ballesteros J, Martín A, and Hornero-Méndez D.** Genetic Variability of Carotenoid Concentration and Degree of Esterification among Tritordeum (\times *Tritordeum* Ascherson et Graebner) and Durum Wheat Accessions. *J Agric Food Chem.* 2007;**55**(10):4244–4251. <https://doi.org/10.1021/jf070342p>
- Awan MJA, Ali Z, Amin I, and Mansoor S.** Twin prime editor: seamless repair without damage. *Trends in Biotechnology.* 2022a;**40**(4):374–376. <https://doi.org/10.1016/j.tibtech.2022.01.013>

- Awan MJA, Pervaiz K, Rasheed A, Amin I, Saeed NA, Dhugga KS, and Mansoor S.** Genome edited wheat- current advances for the second green revolution. *Biotechnology Advances*. 2022b;**60**:108006. <https://doi.org/10.1016/j.biotechadv.2022.108006>
- Bai C, Capell T, Berman J, Medina V, Sandmann G, Christou P, and Zhu C.** Bottlenecks in carotenoid biosynthesis and accumulation in rice endosperm are influenced by the precursor-product balance. *Plant Biotechnology Journal*. 2016;**14**(1):195–205. <https://doi.org/10.1111/pbi.12373>
- Bai C, Rivera SM, Medina V, Alves R, VilaprinYO E, Sorribas A, Canela R, Capell T, Sandmann G, Christou P, et al.** An in vitro system for the rapid functional characterization of genes involved in carotenoid biosynthesis and accumulation. *The Plant Journal*. 2014;**77**(3):464–475. <https://doi.org/10.1111/tpj.12384>
- Barone P, Wu E, Lenderts B, Anand A, Gordon-Kamm W, Svitashv S, and Kumar S.** Efficient Gene Targeting in Maize Using Inducible CRISPR-Cas9 and Marker-free Donor Template. *Molecular Plant*. 2020;**13**(8):1219–1227. <https://doi.org/10.1016/j.molp.2020.06.008>
- Barrero JM, Piqueras P, González-Guzmán M, Serrano R, Rodríguez PL, Ponce MR, and Micol JL.** A mutational analysis of the ABA1 gene of *Arabidopsis thaliana* highlights the involvement of ABA in vegetative development. *Journal of Experimental Botany*. 2005;**56**(418):2071–2083. <https://doi.org/10.1093/jxb/eri206>
- Batra R, Agarwal P, Tyagi S, Saini DK, Kumar V, Kumar A, Kumar S, Balyan HS, Pandey R, and Gupta PK.** A study of CCD8 genes/proteins in seven monocots and eight dicots. *PLoS ONE*. 2019;**14**(3):e0213531. <https://doi.org/10.1371/journal.pone.0213531>
- Bekkering CS, Yu S, Isaka NN, Sproul BW, Dubcovsky J, and Tian L.** Genetic dissection of the roles of β -hydroxylases in carotenoid metabolism, photosynthesis, and plant growth in tetraploid wheat (*Triticum turgidum* L.). *Theor Appl Genet*. 2023;**136**(1):1–15. <https://doi.org/10.1007/s00122-023-04276-3>
- Beleggia R, Platani C, Nigro F, and Cattivelli L.** A micro-method for the determination of Yellow Pigment Content in durum wheat. *Journal of Cereal Science*. 2010;**52**(1):106–110. <https://doi.org/10.1016/j.jcs.2010.03.008>
- Benderradji L, Brini F, Kellou K, Ykhlef N, Djekoun A, Masmoudi K, and Bouzerzour H.** Callus Induction, Proliferation, and Plantlets Regeneration of Two Bread Wheat (*Triticum aestivum* L.) Genotypes under Saline and Heat Stress Conditions. *ISRN Agronomy*. 2012;**2012**:1–8. <https://doi.org/10.5402/2012/367851>
- Berman J, Zorrilla-López U, Medina V, Farré G, Sandmann G, Capell T, Christou P, and Zhu C.** The *Arabidopsis* ORANGE (AtOR) gene promotes carotenoid accumulation in transgenic corn hybrids derived from parental lines with limited carotenoid pools. *Plant Cell Reports*. 2017;**36**(6):933–945. <https://doi.org/10.1007/s00299-017-2126-z>
- Beyer P, Al-Babili S, Ye X, Lucca P, Schaub P, Welsch R, and Potrykus I.** Golden Rice: Introducing the β -Carotene Biosynthesis Pathway into Rice Endosperm by Genetic

Engineering to Defeat Vitamin A Deficiency. *The Journal of Nutrition*. 2002;**132**(3):506S-510S. <https://doi.org/10.1093/jn/132.3.506S>

- Bilgin O, Korkut KZ, Başer I, Dağlıoğlu O, Öztürk I, Kahraman T, and Balkan A.** Variation and heritability for some semolina characteristics and grain yield and their relations in durum wheat (*Triticum durum* Desf.). *World Journal of Agricultural Sciences*. 2010;**6**(3):301–308.
- Blanco A, Colasuonno P, Gadaleta A, Mangini G, Schiavulli A, Simeone R, Digesù AM, De Vita P, Mastrangelo AM, and Cattivelli L.** Quantitative trait loci for yellow pigment concentration and individual carotenoid compounds in durum wheat. *Journal of Cereal Science*. 2011;**54**(2):255–264. <https://doi.org/10.1016/j.jcs.2011.07.002>
- Blomme J, Develtere W, Köse A, Arraiza Ribera J, Brugmans C, Jaraba-Wallace J, Decaestecker W, Rombaut D, Baekelandt A, Daniel Fernández Fernández Á, et al.** The heat is on: a simple method to increase genome editing efficiency in plants. *BMC Plant Biol*. 2022;**22**(1):142. <https://doi.org/10.1186/s12870-022-03519-7>
- Bollinedi H, Dhakane-Lad J, Gopala Krishnan S, Bhowmick PK, Prabhu KV, Singh NK, and Singh AK.** Kinetics of β -carotene degradation under different storage conditions in transgenic Golden Rice® lines. *Food Chemistry*. 2019;**278**(August 2018):773–779. <https://doi.org/10.1016/j.foodchem.2018.11.121>
- Borrill P, Ramirez-Gonzalez R, and Uauy C.** expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform. *Plant Physiology*. 2016;**170**(4):2172–2186. <https://doi.org/10.1104/pp.15.01667>
- Boussakouran A, El Yamani M, Sakar EH, Taghouti M, and Rharrabti Y.** Quality attributes associated with breeding progress in Moroccan durum wheat released between 1984 and 2007. *Genet Resour Crop Evol*. 2022. <https://doi.org/10.1007/s10722-022-01464-w>
- Brinton J and Uauy C.** A reductionist approach to dissecting grain weight and yield in wheat. *Journal of Integrative Plant Biology*. 2019;**61**(3):337–358. <https://doi.org/10.1111/jipb.12741>
- Broccanello C, Bellin D, DalCorso G, Furini A, and Taranto F.** Genetic approaches to exploit landraces for improvement of *Triticum turgidum* ssp. durum in the age of climate change. *Frontiers in Plant Science*. 2023;**14**.
- Campos KM, Royo C, Schulthess A, Villegas D, Matus I, Ammar K, and Schwember AR.** Association of phytoene synthase Psy1-A1 and Psy1-B1 allelic variants with semolina yellowness in durum wheat (*Triticum turgidum* L. var. durum). *Euphytica*. 2016;**207**(1):109–117. <https://doi.org/10.1007/s10681-015-1541-x>
- Carretero-Paulet L, Cairó A, Botella-Pavía P, Besumbes O, Campos N, Boronat A, and Rodríguez-Concepción M.** Enhanced flux through the methylerythritol 4-phosphate pathway in Arabidopsis plants overexpressing deoxyxylulose 5-phosphate reductoisomerase. *Plant Mol Biol*. 2006;**62**(4–5):683–695. <https://doi.org/10.1007/s11103-006-9051-9>

- Chatterjee P, Lee J, Nip L, Koseki SRT, Tysinger E, Sontheimer EJ, Jacobson JM, and Jakimo N.** A Cas9 with PAM recognition for adenine dinucleotides. *Nature Communications*. 2020;**11**(1):1–6. <https://doi.org/10.1038/s41467-020-16117-8>
- Chayut N, Yuan H, Ohali S, Meir A, Sa'ar U, Tzuri G, Zheng Y, Mazourek M, Gepstein S, Zhou X, et al.** Distinct Mechanisms of the ORANGE Protein in Controlling Carotenoid Flux. *Plant Physiology*. 2017;**173**(1):376–389. <https://doi.org/10.1104/pp.16.01256>
- Che P, Zhao Z-Y, Glassman K, Dolde D, Hu TX, Jones TJ, Gruis DF, Obukosia S, Wambugu F, and Albertsen MC.** Elevated vitamin E content improves all- *trans* β -carotene accumulation and stability in biofortified sorghum. *Proc Natl Acad Sci USA*. 2016;**113**(39):11040–11045. <https://doi.org/10.1073/pnas.1605689113>
- Chen PJ, Hussmann JA, Yan J, Knipping F, Ravisankar P, Chen P-F, Chen C, Nelson JW, Newby GA, Sahin M, et al.** Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell*. 2021;**184**(22):5635–5652.e29. <https://doi.org/10.1016/j.cell.2021.09.018>
- Chen S, Zhou Y, Chen Y, and Gu J.** fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;**34**(17):i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Chen Z, Debernardi JM, Dubcovsky J, and Gallavotti A.** Recent advances in crop transformation technologies. *Nat Plants*. 2022. <https://doi.org/10.1038/s41477-022-01295-8>
- Cheng S, Feng C, Wingen LU, Cheng H, Riche AB, Jiang M, Leverington-Waite M, Huang Z, Collier S, Orford S, et al.** Harnessing Landrace Diversity Empowers Wheat Breeding for Climate Resilience. *bioRxiv*. 2023. <https://doi.org/10.1101/2023.10.04.560903>
- Cho KS, Han EH, Kwak SS, Cho JH, Im JS, Hong SY, Sohn HB, Kim YH, and Lee SW.** Expressing the sweet potato orange gene in transgenic potato improves drought tolerance and marketable tuber production. *Comptes Rendus - Biologies*. 2016;**339**(5–6):207–213. <https://doi.org/10.1016/j.crv.2016.04.010>
- CIMMYT.** Wheat Atlas. *Wheat Atlas*. 2023. <http://wheatatlas.org>. Retrieved November 12, 2023
- Clarke FR, Clarke JM, McCaig TN, Knox RE, and DePauw RM.** Inheritance of yellow pigment concentration in seven durum wheat crosses. *Can J Plant Sci*. 2006;**86**(1):133–141. <https://doi.org/10.4141/P05-083>
- Clarke JM, Clarke FR, and Pozniak CJ.** Forty-six years of genetic improvement in Canadian durum wheat cultivars. *Can J Plant Sci*. 2010;**90**(6):791–801. <https://doi.org/10.4141/cjps10091>
- Colasuonno P, Gadaleta A, Giancaspro A, Nigro D, Giove S, Incerti O, Mangini G, Signorile A, Simeone R, and Blanco A.** Development of a high-density SNP-based linkage map and detection of yellow pigment content QTLs in durum wheat. *Molecular Breeding*. 2014;**34**(4):1563–1578. <https://doi.org/10.1007/s11032-014-0183-3>

- Colasuonno P, Lozito ML, Marcotuli I, Nigro D, Giancaspro A, Mangini G, De Vita P, Mastrangelo AM, Pecchioni N, Houston K, et al.** The carotenoid biosynthetic and catabolic genes in wheat and their association with yellow pigments. *BMC Genomics*. 2017a;**18**(1). <https://doi.org/10.1186/s12864-016-3395-6>
- Colasuonno P, Marcotuli I, Blanco A, Maccaferri M, Condorelli GE, Tuberosa R, Parada R, de Camargo AC, Schwember AR, and Gadaleta A.** Carotenoid Pigment Content in Durum Wheat (*Triticum turgidum* L. var durum): An Overview of Quantitative Trait Loci and Candidate Genes. *Frontiers in Plant Science*. 2019;**10**:1347. <https://doi.org/10.3389/fpls.2019.01347>
- Colasuonno P, Marcotuli I, Lozito ML, Simeone R, Blanco A, and Gadaleta A.** Characterization of Aldehyde Oxidase (AO) Genes Involved in the Accumulation of Carotenoid Pigments in Wheat Grain. *Front Plant Sci*. 2017b;**8**:863. <https://doi.org/10.3389/fpls.2017.00863>
- Cong L, Wang C, Chen L, Liu H, Yang G, and He G.** Expression of phytoene synthase1 and Carotene Desaturase crtI Genes Result in an Increase in the Total Carotenoids Content in Transgenic Elite Wheat (*Triticum aestivum* L.). *Journal of Agricultural and Food Chemistry*. 2009;**57**(18):8652–8660. <https://doi.org/10.1021/jf9012218>
- Crisp P, Walkey DGA, Bellman E, and Roberts E.** A mutation affecting curd colour in cauliflower (*Brassica oleracea* L. var. *Botrytis* DC). *Euphytica*. 1975;**24**(1):173–176. <https://doi.org/10.1007/BF00147182>
- De Beukelaer H, Davenport GF, and Fack V.** Core Hunter 3: flexible core subset selection. *BMC Bioinformatics*. 2018;**19**(1):203. <https://doi.org/10.1186/s12859-018-2209-z>
- De Moura FF, Miloff A, and Boy E.** Retention of Provitamin A Carotenoids in Staple Crops Targeted for Biofortification in Africa: Cassava, Maize and Sweet Potato. *Critical Reviews in Food Science and Nutrition*. 2015;**55**(9):1246–1269. <https://doi.org/10.1080/10408398.2012.724477>
- De Steur H, Stein AJ, and Demont M.** From Golden Rice to Golden Diets: How to turn its recent approval into practice. *Global Food Security*. 2022;**32**:100596. <https://doi.org/10.1016/j.gfs.2021.100596>
- Debernardi JM, Tricoli DM, Ercoli MF, Hayta S, Ronald P, Palatnik JF, and Dubcovsky J.** A chimera including a GROWTH-REGULATING FACTOR (GRF) and its cofactor GRF. *bioRxiv*. 2020:2020.08.23.263905. <https://doi.org/10.1101/2020.08.23.263905>
- Dhami N and Cazzonelli CI.** Environmental impacts on carotenoid metabolism in leaves. *Plant Growth Regul*. 2020;**92**(3):455–477. <https://doi.org/10.1007/s10725-020-00661-w>
- Dias MG, Olmedilla-Alonso B, Hornero-Méndez D, Mercadante AZ, Osorio C, Vargas-Murga L, and Meléndez-Martínez AJ.** Comprehensive Database of Carotenoid Contents in Ibero-American Foods. A Valuable Tool in the Context of Functional Foods and the Establishment of Recommended Intakes of Bioactives. *J Agric Food Chem*. 2018;**66**(20):5055–5107. <https://doi.org/10.1021/acs.jafc.7b06148>

- Digesù AM, Platani C, Cattivelli L, Mangini G, and Blanco A.** Genetic variability in yellow pigment components in cultivated and wild tetraploid wheats. *Journal of Cereal Science*. 2009;**50**(2):210–218. <https://doi.org/10.1016/j.jcs.2009.05.002>
- Doebley JF, Gaut BS, and Smith BD.** The Molecular Genetics of Crop Domestication. *Cell*. 2006;**127**(7):1309–1321. <https://doi.org/10.1016/j.cell.2006.12.006>
- Domagalska MA, Sarnowska E, Nagy F, and Davis SJ.** Genetic Analyses of Interactions among Gibberellin, Abscisic Acid, and Brassinosteroids in the Control of Flowering Time in *Arabidopsis thaliana*. *PLoS ONE*. 2010;**5**(11):e14012. <https://doi.org/10.1371/journal.pone.0014012>
- Dong H.** Application of genome editing techniques to regulate gene expression in crops. *BMC Plant Biology*. 2024;**24**(1):100. <https://doi.org/10.1186/s12870-024-04786-2>
- Dubcovsky J and Dvorak J.** Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science*. 2007;**316**(5833):1862–1866. <https://doi.org/10.1126/science.1143986>
- Echevarría-Zomeño S, Fernández-Calvino L, Castro-Sanz AB, López JA, Vázquez J, and Castellano MM.** Dissecting the proteome dynamics of the early heat stress response leading to plant survival or death in *Arabidopsis*. *Plant Cell & Environment*. 2016;**39**(6):1264–1278. <https://doi.org/10.1111/pce.12664>
- El-Beltagi HS, Ahmed OK, and Hegazy AE.** Protective Effect of Nitric Oxide on High Temperature Induced Oxidative Stress in Wheat (*Triticum aestivum*) Callus Culture. *Not Sci Biol*. 2016;**8**(2):192–198. <https://doi.org/10.15835/nsb.8.2.9807>
- Ellison SL, Luby CH, Corak KE, Coe KM, Senalik D, Iorizzo M, Goldman IL, Simon PW, and Dawson JC.** Carotenoid presence is associated with the *or* gene in domesticated carrot. *Genetics*. 2018;**210**(4):1497–1508. <https://doi.org/10.1534/genetics.118.301299>
- Elouafi I, Nachit MM, and Martin LM.** Identification of a microsatellite on chromosome 7B showing a strong linkage with yellow pigment in durum wheat (*Triticum turgidum* L. var. durum). *Hereditas*. 2001;**135**(2–3):255–261. <https://doi.org/10.1111/j.1601-5223.2001.t01-1-00255.x>
- Elsharawy H and Refat M.** CRISPR/Cas9 genome editing in wheat: enhancing quality and productivity for global food security—a review. *Funct Integr Genomics*. 2023;**23**(3):265. <https://doi.org/10.1007/s10142-023-01190-1>
- Endo A, Saika H, Takemura M, Misawa N, and Toki S.** A novel approach to carotenoid accumulation in rice callus by mimicking the cauliflower Orange mutation via genome editing. *Rice*. 2019;**12**(1):81. <https://doi.org/10.1186/s12284-019-0345-3>
- Engler C, Kandzia R, and Marillonnet S.** A one pot, one step, precision cloning method with high throughput capability. *PLoS One*. 2008;**3**(11):e3647. <https://doi.org/10.1371/journal.pone.0003647>
- Erdman JW, Smith JW, Kuchan MJ, Mohn ES, Johnson EJ, Rubakhin SS, Wang L, Sweedler JV, and Neuringer M.** Lutein and brain function. *Foods*. 2015;**4**(4):547–564. <https://doi.org/10.3390/foods4040547>

- FAO.** FAOSTAT statistics database, Crops. [Online]. 2023. <http://www.fao.org/faostat/en/#data/QC>
- Farré G, Maiam Rivera S, Alves R, Vilaprinyo E, Sorribas A, Canela R, Naqvi S, Sandmann G, Capell T, Zhu C, et al.** Targeted transcriptomic and metabolic profiling reveals temporal bottlenecks in the maize carotenoid pathway that may be addressed by multigene engineering. *The Plant Journal*. 2013;**75**(3):441–455. <https://doi.org/10.1111/tpj.12214>
- Fayaz F, Mardi M, Aghaee M, Darvish F, and Talebi R.** Phenotypic diversity analysis of grain yield and yellow pigment content in germplasm collected from Iranian durum wheat (*Triticum turgidum* L.) landraces. *Archives of Agronomy and Soil Science*. 2013;**59**(10):1339–1357. <https://doi.org/10.1080/03650340.2012.708927>
- Feldman M and Kislev ME.** Domestication of emmer wheat and evolution of free-threshing tetraploid wheat. *Israel Journal of Plant Sciences*. 2007;**55**(3):207–221. <https://doi.org/10.1560/IJPS.55.3-4.207>
- Ficco DBM, Mastrangelo AM, Trono D, Borrelli GM, De Vita P, Fares C, Beleggia R, Platani C, and Papa R.** The colours of durum wheat: A review. *Crop and Pasture Science*. 2014;**65**(1):1–15. <https://doi.org/10.1071/CP13293>
- Flowerika, Alok A, Kumar J, Thakur N, Pandey A, Pandey AK, Upadhyay SK, and Tiwari S.** Characterization and Expression Analysis of Phytoene Synthase from Bread Wheat (*Triticum aestivum* L.). *PLOS ONE*. 2016;**11**(10):e0162443. <https://doi.org/10.1371/journal.pone.0162443>
- Fратиани A, Irano M, Panfili G, and Acquistucci R.** Estimation of color of durum wheat. Comparison of WSB, HPLC, and reflectance colorimeter measurements. *Journal of Agricultural and Food Chemistry*. 2005;**53**(7):2373–2378. <https://doi.org/10.1021/jf040351n>
- Fu BX, Chiremba C, Pozniak CJ, Wang K, and Nam S.** Total phenolic and yellow pigment contents and antioxidant activities of durum wheat milling fractions. *Antioxidants*. 2017;**6**(4). <https://doi.org/10.3390/antiox6040078>
- Garcia Molina MD, Botticella E, Beleggia R, Palombieri S, De Vita P, Masci S, Lafiandra D, and Sestili F.** Enrichment of provitamin A content in durum wheat grain by suppressing β -carotene hydroxylase 1 genes with a TILLING approach. *Theoretical and Applied Genetics*. 2021. <https://doi.org/10.1007/s00122-021-03944-6>
- Gayen D, Ali N, Sarkar SN, Datta SK, and Datta K.** Down-regulation of lipoxygenase gene reduces degradation of carotenoids of golden rice during storage. *Planta*. 2015;**242**(1):353–363. <https://doi.org/10.1007/s00425-015-2314-4>
- Gelvin SB.** Integration of *Agrobacterium* T-DNA into the Plant Genome. *Annu Rev Genet*. 2017;**51**(1):195–217. <https://doi.org/10.1146/annurev-genet-120215-035320>
- Giambanelli E, Ferioli F, Koçaoglu B, Jorjadze M, Alexieva I, Darbinyan N, and D'Antuono LF.** A comparative study of bioactive compounds in primitive wheat populations from Italy, Turkey, Georgia, Bulgaria and Armenia. *J Sci Food Agric*. 2013;**93**(14):3490–3501. <https://doi.org/10.1002/jsfa.6326>

- Giuliano G.** Provitamin A biofortification of crop plants: a gold rush with many miners. *Current Opinion in Biotechnology*. 2017;**44**:169–180. <https://doi.org/10.1016/j.copbio.2017.02.001>
- González MN, Massa GA, Andersson M, Décima Oneto CA, Turesson H, Storani L, Olsson N, Fält A-S, Hofvander P, and Feingold SE.** Comparative potato genome editing: *Agrobacterium tumefaciens*-mediated transformation and protoplasts transfection delivery of CRISPR/Cas9 components directed to StPPO2 gene. *Plant Cell Tissue Organ Cult*. 2021;**145**(2):291–305. <https://doi.org/10.1007/s11240-020-02008-9>
- González-Grandío E, Pajoro A, Franco-Zorrilla JM, Tarancón C, Immink RGH, and Cubas P.** Abscisic acid signaling is controlled by a *BRANCHED1/HD-ZIP I* cascade in *Arabidopsis* axillary buds. *Proc Natl Acad Sci USA*. 2017;**114**(2). <https://doi.org/10.1073/pnas.1613199114>
- Gonzalez-Jorge S, Ha S-H, Magallanes-Lundback M, Gilliland LU, Zhou A, Lipka AE, Nguyen Y-N, Angelovici R, Lin H, Cepela J, et al.** CAROTENOID CLEAVAGE DIOXYGENASE4 Is a Negative Regulator of β -Carotene Content in *Arabidopsis* Seeds. *The Plant Cell*. 2013;**25**(12):4812–4826. <https://doi.org/10.1105/tpc.113.119677>
- Gonzalez-Jorge S, Mehrshahi P, Magallanes-Lundback M, Lipka AE, Angelovici R, Gore MA, and DellaPenna D.** ZEAXANTHIN EPOXIDASE Activity Potentiates Carotenoid Degradation in Maturing Seed. *Plant Physiol*. 2016;**171**(3):1837–1851. <https://doi.org/10.1104/pp.16.00604>
- Goredema-Matongera N, Ndhlela T, Magorokosho C, Kamutando CN, Van Biljon A, and Labuschagne M.** Multinutrient Biofortification of Maize (*Zea mays* L.) in Africa: Current Status, Opportunities and Limitations. *Nutrients*. 2021;**13**(3):1039. <https://doi.org/10.3390/nu13031039>
- Grützner R, Martin P, Horn C, Mortensen S, Cram EJ, Lee-Parsons CWT, Stuttmann J, and Marillonnet S.** High-efficiency genome editing in plants mediated by a Cas9 gene containing multiple introns. *Plant Communications*. 2021;**2**(2):100135. <https://doi.org/10.1016/j.xplc.2020.100135>
- Guan JC, Koch KE, Suzuki M, Wu S, Latshaw S, Petruff T, Goulet C, Klee HJ, and McCarty DR.** Diverse Roles of Strigolactone Signaling in Maize Architecture and the Uncoupling of a Branching-Specific Subnetwork. *Plant Physiology*. 2012;**160**(3):1303–1317. <https://doi.org/10.1104/pp.112.204503>
- Guan P, Li X, Zhuang L, Wu B, Huang J, Zhao J, Qiao L, Zheng J, Hao C, and Zheng X.** Genetic dissection of lutein content in common wheat via association and linkage mapping. *Theor Appl Genet*. 2022;**135**(9):3127–3141. <https://doi.org/10.1007/s00122-022-04175-z>
- Gupta PK.** Drought-tolerant transgenic wheat HB4®: a hope for the future. *Trends in Biotechnology*. 2023:S0167779923003463. <https://doi.org/10.1016/j.tibtech.2023.12.007>
- Gupta PK, Kulwal PL, and Jaiswal V.** Chapter Two - Association Mapping in Crop Plants: Opportunities and Challenges. . In. *Advances in Genetics*, T Friedmann, JC Dunlap,

and SF Goodwin, eds. (Academic Press), pp. 109–147.
<https://doi.org/10.1016/B978-0-12-800271-1.00002-0>

Harrington SA, Backhaus AE, Fox S, Rogers C, Borrill P, Uauy C, and Richardson A. A heat-shock inducible system for flexible gene expression in cereals. *Plant Methods*. 2020;**16**(1):137. <https://doi.org/10.1186/s13007-020-00677-3>

Harwood W. Breaking transformation barriers. *Nat Plants*. 2023;**9**(6):854–855.
<https://doi.org/10.1038/s41477-023-01420-1>

Harwood WA. Advances and remaining challenges in the transformation of barley and wheat. *Journal of Experimental Botany*. 2012;**63**(5):1791–1798.
<https://doi.org/10.1093/jxb/err380>

Hassani-Pak K, Singh A, Brandizi M, Hearnshaw J, Parsons JD, Amberkar S, Phillips AL, Doonan JH, and Rawlings C. KnetMiner: a comprehensive approach for supporting evidence-based gene discovery and complex trait analysis across species. *Plant Biotechnology Journal*. 2021;**19**(8):1670–1678. <https://doi.org/10.1111/pbi.13583>

Hayta S, Smedley MA, Clarke M, Forner M, and Harwood WA. An Efficient Agrobacterium-Mediated Transformation Protocol for Hexaploid and Tetraploid Wheat. *Current Protocols*. 2021;**1**(3). <https://doi.org/10.1002/cpz1.58>

Hayta S, Smedley MA, Demir SU, Blundell R, Hinchliffe A, Atkinson N, and Harwood WA. An efficient and reproducible Agrobacterium-mediated transformation method for hexaploid wheat (*Triticum aestivum* L.). *Plant Methods*. 2019;**15**(1):121.
<https://doi.org/10.1186/s13007-019-0503-z>

He XY, Zhang YL, He ZH, Wu YP, Xiao YG, Ma CX, and Xia XC. Characterization of phytoene synthase 1 gene (Psy1) located on common wheat chromosome 7A and development of a functional marker. *Theoretical and Applied Genetics*. 2008;**116**(2):213–221. <https://doi.org/10.1007/s00122-007-0660-8>

Hermanns AS, Zhou X, Xu Q, Tadmor Y, and Li L. Carotenoid Pigment Accumulation in Horticultural Plants. *Horticultural Plant Journal*. 2020;**6**(6):343–360.
<https://doi.org/10.1016/j.hpj.2020.10.002>

Hidalgo A, Brandolini A, Pompei C, and Piscozzi R. Carotenoids and tocopherols of einkorn wheat (*Triticum monococcum* ssp. *monococcum* L.). *Journal of Cereal Science*. 2006;**44**(2):182–193. <https://doi.org/10.1016/j.jcs.2006.06.002>

Hillary VE and Ceasar SA. Prime editing in plants and mammalian cells: Mechanism, achievements, limitations, and future prospects. *BioEssays*. 2022;**44**(9):2200032.
<https://doi.org/10.1002/bies.202200032>

Howitt CA and Pogson BJ. Carotenoid accumulation and function in seeds and non-green tissues. *Plant, Cell and Environment*. 2006;**29**(3):435–445.
<https://doi.org/10.1111/j.1365-3040.2005.01492.x>

Hu X, Wang C, Fu Y, Liu Q, Jiao X, and Wang K. Expanding the Range of CRISPR/Cas9 Genome Editing in Rice. *Molecular Plant*. 2016;**9**(6):943–945.
<https://doi.org/10.1016/j.molp.2016.03.003>

- Hua K, Tao X, and Zhu JK.** Expanding the base editing scope in rice by using Cas9 variants. *Plant Biotechnology Journal*. 2019;**17**(2):499–504. <https://doi.org/10.1111/pbi.12993>
- Huang M, Liu X, Zhou Y, Summers RM, and Zhang Z.** BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience*. 2018;**8**(2):giy154. <https://doi.org/10.1093/gigascience/giy154>
- Huang Z and Liu G.** Current advancement in the application of prime editing. *Front Bioeng Biotechnol*. 2023;**11**:1039315. <https://doi.org/10.3389/fbioe.2023.1039315>
- IWGSC, Appels R, Eversole K, Stein N, Feuillet C, Keller B, Rogers J, Pozniak CJ, Choulet F, Distelfeld A, et al.** Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*. 2018;**361**(6403):eaar7191. <https://doi.org/10.1126/science.aar7191>
- Jahns P and Holzwarth AR.** The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2012;**1817**(1):182–193. <https://doi.org/10.1016/j.bbabi.2011.04.012>
- Jang I-C, Choi W-B, Lee K-H, Song SI, Nahm BH, and Kim J-K.** High-Level and Ubiquitous Expression of the Rice Cytochrome *c* Gene *OsCc1* and Its Promoter Activity in Transgenic Plants Provides a Useful Promoter for Transgenesis of Monocots. *Plant Physiology*. 2002;**129**(4):1473–1481. <https://doi.org/10.1104/pp.002261>
- Jiang Y, Chai Y, Qiao D, Wang J, Xin C, Sun W, Cao Z, Zhang Y, Zhou Y, Wang X-C, et al.** Optimized prime editing efficiently generates glyphosate-resistant rice plants carrying homozygous TAP-IVS mutation in EPSPS. *Molecular Plant*. 2022;**15**(11):1646–1649. <https://doi.org/10.1016/j.molp.2022.09.006>
- Jiang Y-Y, Chai Y-P, Lu M-H, Han X-L, Lin Q, Zhang Y, Zhang Q, Zhou Y, Wang X-C, Gao C, et al.** Prime editing efficiently generates W542L and S621I double mutations in two ALS genes in maize. *Genome Biol*. 2020;**21**(1):257. <https://doi.org/10.1186/s13059-020-02170-5>
- Jin S, Lin Q, Gao Q, and Gao C.** Optimized prime editing in monocot plants using PlantPegDesigner and engineered plant prime editors (ePPEs). *Nat Protoc*. 2022;1–23. <https://doi.org/10.1038/s41596-022-00773-9>
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, and Charpentier E.** A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. 2012;**337**(6096):816–821. <https://doi.org/10.1126/science.1225829>
- Johnson EJ.** Role of lutein and zeaxanthin in visual and cognitive function throughout the lifespan. 2014. <https://doi.org/10.1111/nure.12133>
- Jung YJ, Go JY, Lee HJ, Park JS, Kim JY, Lee YJ, Ahn M, Kim M-S, Cho Y, Kwak S, et al.** Overexpression of Orange Gene (*OsOr-R115H*) Enhances Heat Tolerance and Defense-Related Gene Expression in Rice (*Oryza sativa* L.). *Genes*. 2021;**12**(12):1891. <https://doi.org/10.3390/genes12121891>

- Kang L, Kim HS, Kwon YS, Ke Q, Ji CY, Park SC, Lee HS, Deng X, and Kwak SS.** IbOr regulates photosynthesis under heat stress by stabilizing ibpsbp in sweetpotato. *Frontiers in Plant Science*. 2017;**8**(June):1–13. <https://doi.org/10.3389/fpls.2017.00989>
- Katamadze A, Vergara-Díaz O, Uberegui E, Yoldi-Achalandabaso A, Araus JL, and Vicente R.** Evolution of wheat architecture, physiology, and metabolism during domestication and further cultivation: Lessons for crop improvement. *The Crop Journal*. 2023;**11**(4):1080–1096. <https://doi.org/10.1016/j.cj.2023.06.006>
- Kim HK, Kim JY, Kim JH, Go JY, Jung Y-S, Lee HJ, Ahn M-J, Yu J, Bae S, Kim HS, et al.** Biochemical Characterization of Orange-Colored Rice Calli Induced by Target Mutagenesis of OsOr Gene. *Plants*. 2022;**12**(1):56. <https://doi.org/10.3390/plants12010056>
- Kim JE, Rensing KH, Douglas CJ, and Cheng KM.** Chromoplasts ultrastructure and estimated carotene content in root secondary phloem of different carrot varieties. *Planta*. 2010;**231**(3):549–558. <https://doi.org/10.1007/s00425-009-1071-7>
- Kim SE, Kim HS, Wang Z, Ke Q, Lee CJ, Park SU, Lim YH, Park WS, Ahn MJ, and Kwak SS.** A single amino acid change at position 96 (Arg to His) of the sweetpotato Orange protein leads to carotenoid overaccumulation. *Plant Cell Reports*. 2019;**38**(11):1393–1402. <https://doi.org/10.1007/s00299-019-02448-4>
- Kim S-E, Lee C-J, Park S-U, Lim Y-H, Park WS, Kim H-J, Ahn M-J, Kwak S-S, and Kim HS.** Overexpression of the Golden SNP-Carrying Orange Gene Enhances Carotenoid Accumulation and Heat Stress Tolerance in Sweetpotato Plants. *Antioxidants*. 2021;**10**(1):51. <https://doi.org/10.3390/antiox10010051>
- Kim SH, Ahn YO, Ahn MJ, Jeong JC, Lee HS, and Kwak SS.** Cloning and characterization of an Orange gene that increases carotenoid accumulation and salt stress tolerance in transgenic sweetpotato cultures. *Plant Physiology and Biochemistry*. 2013;**70**:445–454. <https://doi.org/10.1016/j.plaphy.2013.06.011>
- Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, et al.** Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database*. 2011;**2011**:bar030. <https://doi.org/10.1093/database/bar030>
- Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, Clissold L, Simmonds J, Ramirez-Gonzalez RH, Wang X, Borrill P, et al.** Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences*. 2017;**114**(6):E913–E921. <https://doi.org/10.1073/pnas.1619268114>
- Kumagai Y, Liu Y, Hamada H, Luo W, Zhu J, Kuroki M, Nagira Y, Taoka N, Katoh E, and Imai R.** Introduction of a second “Green Revolution” mutation into wheat via in planta CRISPR/Cas9 delivery. *Plant Physiology*. 2022;**188**(4):1838–1842. <https://doi.org/10.1093/plphys/kiab570>
- Kumar V, Rahman MH, AlMomin S, and Shajan A.** Future of TILLING in Plant Breeding. . In. *TILLING and Eco-TILLING for Crop Improvement*, A Bhattacharya, V Parkhi, and B Char, eds. (Springer Nature: Singapore), pp. 185–216. https://doi.org/10.1007/978-981-99-2722-7_10

- Lachman J, Martinek P, Kotíková Z, Orsák M, and Šulc M.** Genetics and chemistry of pigments in wheat grain – A review. *Journal of Cereal Science*. 2017;**74**:145–154. <https://doi.org/10.1016/j.jcs.2017.02.007>
- Laidò G, Mangini G, Taranto F, Gadaleta A, Blanco A, Cattivelli L, Marone D, Mastrangelo AM, Papa R, and De Vita P.** Genetic Diversity and Population Structure of Tetraploid Wheats (*Triticum turgidum* L.) Estimated by SSR, DArT and Pedigree Data. *PLoS One*. 2013;**8**(6):e67280. <https://doi.org/10.1371/journal.pone.0067280>
- Lakshmi Jayaraj K, Thulasidharan N, Antony A, John M, Augustine R, Chakravartty N, Sukumaran S, Uma Maheswari M, Abraham S, Thomas G, et al.** Targeted editing of tomato carotenoid isomerase reveals the role of 5' UTR region in gene expression regulation. *Plant Cell Rep*. 2021;**40**(4):621–635. <https://doi.org/10.1007/s00299-020-02659-0>
- Langmead B and Salzberg SL.** Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;**9**(4):357–359. <https://doi.org/10.1038/nmeth.1923>
- Lawrenson T, Chhetry M, Clarke M, Hundleby P, and Harwood W.** Improved SpCas9 and LbCas12a genome editing systems in *Brassica oleracea* and *Brassica napus* (Genomics). <https://doi.org/10.1101/2022.05.16.492057>
- Lawrenson T, Hinchliffe A, Clarke M, Morgan Y, and Harwood W.** In-planta Gene Targeting in Barley Using Cas9 With and Without Geminiviral Replicons. *Front Genome Ed*. 2021;**3**:663380. <https://doi.org/10.3389/fgeed.2021.663380>
- Lazo GR, Stein PA, and Ludwig RA.** A DNA Transformation–Competent Arabidopsis Genomic Library in Agrobacterium. *Nat Biotechnol*. 1991;**9**(10):963–967. <https://doi.org/10.1038/nbt1091-963>
- LeBlanc C, Zhang F, Mendez J, Lozano Y, Chatpar K, Irish VF, and Jacob Y.** Increased efficiency of targeted mutagenesis by CRISPR/Cas9 in plants using heat stress. *The Plant Journal*. 2018;**93**(2):377–386. <https://doi.org/10.1111/tpj.13782>
- Lenth RV.** emmeans: Estimated marginal means, aka least-squares means.
- Li H.** Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. <https://doi.org/10.48550/arXiv.1303.3997>
- Li J, Chen L, Liang J, Xu R, Jiang Y, Li Y, Ding J, Li M, Qin R, and Wei P.** Development of a highly efficient prime editor 2 system in plants. *Genome Biol*. 2022;**23**(1):161. <https://doi.org/10.1186/s13059-022-02730-x>
- Li J, Martin C, and Fernie A.** Biofortification's contribution to mitigating micronutrient deficiencies. *Nat Food*. 2024;**5**(1):19–27. <https://doi.org/10.1038/s43016-023-00905-8>
- Li J, Zhang C, He Y, Li S, Yan L, Li Y, Zhu Z, and Xia L.** Plant base editing and prime editing: The current status and future perspectives. *JIPB*. 2023;**65**(2):444–467. <https://doi.org/10.1111/jipb.13425>
- Li L, Paolillo DJ, Parthasarathy MV, DiMuzio EM, and Garvin DF.** A novel gene mutation that confers abnormal patterns of β -carotene accumulation in cauliflower (*Brassica*

oleracea var. botrytis). *Plant Journal*. 2001;**26**(1):59–67.
<https://doi.org/10.1046/j.1365-313X.2001.01008.x>

Li L and Van Eck J. Metabolic engineering of carotenoid accumulation by creating a metabolic sink. *Transgenic Research*. 2007;**16**(5):581–585.
<https://doi.org/10.1007/s11248-007-9111-1>

Li L, Yang Y, Xu Q, Owsiany K, Welsch R, Chitchumroonchokchai C, Lu S, Van Eck J, Deng XX, Failla M, et al. The *or* gene enhances carotenoid accumulation and stability during post-harvest storage of potato tubers. *Molecular Plant*. 2012;**5**(2):339–352.
<https://doi.org/10.1093/mp/ssr099>

Li S, Li J, He Y, Xu M, Zhang J, Du W, Zhao Y, and Xia L. Precise gene replacement in rice by RNA transcript-templated homologous recombination. *Nat Biotechnol*. 2019;**37**(4):445–450. <https://doi.org/10.1038/s41587-019-0065-7>

Li S, Zhang C, Li J, Yan L, Wang N, and Xia L. Present and future prospects for wheat improvement through genome editing and advanced technologies. *Plant Communications*. 2021;**2**(4):100211. <https://doi.org/10.1016/j.xplc.2021.100211>

Li YF, Wu Y, Wang T, Li LR, Lu L, Zhang CY, Li JM, Zhang L, Liu ZH, and Zheng SG. Polyphenol oxidase activity and yellow pigment content in *Aegilops tauschii*, *Triticum turgidum*, *Triticum aestivum*, synthetic hexaploid wheat and its parents. *Journal of Cereal Science*. 2015;**65**:192–201.
<https://doi.org/10.1016/j.jcs.2015.07.011>

Liebsch D and Palatnik JF. MicroRNA miR396, GRF transcription factors and GIF co-regulators: a conserved plant growth regulatory module with potential for breeding and biotechnology. *Current Opinion in Plant Biology*. 2020;**53**:31–42.
<https://doi.org/10.1016/j.pbi.2019.09.008>

Lin Q, Jin S, Zong Y, Yu H, Zhu Z, Liu G, Kou L, Wang Y, Qiu JL, Li J, et al. High-efficiency prime editing with optimized, paired pegRNAs in plants. *Nature Biotechnology*. 2021. <https://doi.org/10.1038/s41587-021-00868-w>

Lin Q, Zong Y, Xue C, Wang S, Jin S, Zhu Z, Wang Y, Anzalone AV, Raguram A, Doman JL, et al. Prime genome editing in rice and wheat. *Nature Biotechnology*. 2020;**38**(5):582–585. <https://doi.org/10.1038/s41587-020-0455-x>

Liu X, Huang M, Fan B, Buckler ES, and Zhang Z. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLoS Genet*. 2016;**12**(2):e1005767. <https://doi.org/10.1371/journal.pgen.1005767>

Lockyer S and Spiro A. The role of bread in the UK diet: An update. *Nutrition Bulletin*. 2020;**45**(2):133–164. <https://doi.org/10.1111/nbu.12435>

Lopez AB, Van Eck J, Conlin BJ, Paolillo DJ, O’Neill J, and Li L. Effect of the cauliflower or transgene on carotenoid accumulation and chromoplast formation in transgenic potato tubers. *Journal of Experimental Botany*. 2008;**59**(2):213–223.
<https://doi.org/10.1093/jxb/erm299>

Lu S, Van Eck J, Zhou X, Lopez AB, O’Halloran DM, Cosman KM, Conlin BJ, Paolillo DJ, Garvin DF, Vrebalov J, et al. The cauliflower *Or* gene encodes a DnaJ cysteine-rich

domain-containing protein that mediates high levels of β -carotene accumulation. *Plant Cell*. 2006;**18**(12):3594–3605. <https://doi.org/10.1105/tpc.106.046417>

- Lu Y, Tian Y, Shen R, Yao Q, Wang M, Chen M, Dong J, Zhang T, Li F, Lei M, et al.** Targeted, efficient sequence insertion and replacement in rice. *Nat Biotechnol*. 2020;**38**(12):1402–1407. <https://doi.org/10.1038/s41587-020-0581-5>
- Luo W, Suzuki R, and Imai R.** Precise *in planta* genome editing via homology-directed repair in wheat. *Plant Biotechnology Journal*. 2023;**21**(4):668–670. <https://doi.org/10.1111/pbi.13984>
- Ma L, Yan SF, Huang YM, Lu XR, Qian F, Pang HL, Xu XR, Zou ZY, Dong PC, Xiao X, et al.** Effect of lutein and zeaxanthin on macular pigment and visual function in patients with early age-related macular degeneration. *Ophthalmology*. 2012;**119**(11):2290–2297. <https://doi.org/10.1016/j.ophtha.2012.06.014>
- Maccaferri M, Harris NS, Twardziok SO, Pasam RK, Gundlach H, Spannagl M, Ormanbekova D, Lux T, Prade VM, Milner SG, et al.** Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat Genet*. 2019;**51**(5):885–895. <https://doi.org/10.1038/s41588-019-0381-3>
- Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A, and Lopez R.** Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*. 2022;**50**(W1):W276–W279. <https://doi.org/10.1093/nar/gkac240>
- Malzahn AA, Tang X, Lee K, Ren Q, Sretenovic S, Zhang Y, Chen H, Kang M, Bao Y, Zheng X, et al.** Application of CRISPR-Cas12a temperature sensitivity for improved genome editing in rice, maize, and Arabidopsis. *BMC Biol*. 2019;**17**(1):9. <https://doi.org/10.1186/s12915-019-0629-5>
- Mancino L and Kuchler F.** Demand for Whole-grain Bread Before and After the Release of Dietary Guidelines. *Applied Economic Perspectives and Policy*. 2012;**34**(1):76–101. <https://doi.org/10.1093/aep/pper035>
- Manjeru P, Van Biljon A, and Labuschagne M.** The development and release of maize fortified with provitamin A carotenoids in developing countries. *Critical Reviews in Food Science and Nutrition*. 2019;**59**(8):1284–1293. <https://doi.org/10.1080/10408398.2017.1402751>
- Martin M.** Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*. 2011;**17**(1):10–12. <https://doi.org/10.14806/ej.17.1.200>
- Martínez-Moreno F, Solís I, Noguero D, Blanco A, Özberk İ, Nsarellah N, Elias E, Mylonas I, and Soriano JM.** Durum wheat in the Mediterranean Rim: historical evolution and genetic resources. *Genet Resour Crop Evol*. 2020;**67**(6):1415–1436. <https://doi.org/10.1007/s10722-020-00913-8>
- Masisi K, Diehl-Jones WL, Gordon J, Chapman D, Moghadasian MH, and Beta T.** Carotenoids of Aleurone, Germ, and Endosperm Fractions of Barley, Corn and Wheat Differentially Inhibit Oxidative Stress. *J Agric Food Chem*. 2015;**63**(10):2715–2724. <https://doi.org/10.1021/jf5058606>

- Mastrangelo AM and Cattivelli L.** What Makes Bread and Durum Wheat Different? Trends in Plant Science. 2021;**26**(7):677–684.
<https://doi.org/10.1016/j.tplants.2021.01.004>
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, and Cunningham F.** The Ensembl Variant Effect Predictor. Genome Biol. 2016;**17**(1):122.
<https://doi.org/10.1186/s13059-016-0974-4>
- Mellado-Ortega E and Hornero-Méndez D.** Carotenoids in cereals: an ancient resource with present and future applications. Phytochem Rev. 2015;**14**(6):873–890.
<https://doi.org/10.1007/s11101-015-9423-3>
- Mellado-Ortega E and Hornero-Méndez D.** Carotenoid evolution during short-storage period of durum wheat (*Triticum turgidum* conv. durum) and tritordeum (*×Tritordeum* Ascherson et Graebner) whole-grain flours. Food Chemistry. 2016;**192**:714–723. <https://doi.org/10.1016/j.foodchem.2015.07.057>
- Mendes-Pinto MM, Silva Ferreira AC, Caris-Veyrat C, and Guedes de Pinho P.** Carotenoid, Chlorophyll, and Chlorophyll-Derived Compounds in Grapes and Port Wines. J Agric Food Chem. 2005;**53**(26):10034–10041. <https://doi.org/10.1021/jf0503513>
- Milner MJ, Craze M, Hope M, and Wallington EJ.** Turning Up The Temperature On CRISPR: Increased Temperature Can Improve The Editing Efficiency Of Wheat Using CRISPR/Cas9. Frontiers in Plant Science. 2020;**11**(November):1780.
<https://doi.org/10.3389/fpls.2020.583374>
- Nazco R, Villegas D, Ammar K, Peña RJ, Moragues M, and Royo C.** Can Mediterranean durum wheat landraces contribute to improved grain quality attributes in modern cultivars? Euphytica. 2012;**185**(1):1–17. <https://doi.org/10.1007/s10681-011-0588-6>
- N’Diaye A, Haile JK, Cory AT, Clarke FR, Clarke JM, Knox RE, and Pozniak CJ.** Single marker and haplotype-based association analysis of semolina and pasta colour in elite durum wheat breeding lines using a high-density consensus map. PLoS ONE. 2017;**12**(1):e0170941. <https://doi.org/10.1371/journal.pone.0170941>
- Ndolo VU and Beta T.** Distribution of carotenoids in endosperm, germ, and aleurone fractions of cereal grain kernels. Food Chemistry. 2013;**139**(1):663–671.
<https://doi.org/10.1016/j.foodchem.2013.01.014>
- Ng PC and Henikoff S.** Predicting Deleterious Amino Acid Substitutions. Genome Res. 2001;**11**(5):863–874. <https://doi.org/10.1101/gr.176601>
- Ni P, Zhao Y, Zhou X, Liu Z, Huang Z, Ni Z, Sun Q, and Zong Y.** Efficient and versatile multiplex prime editing in hexaploid wheat. Genome Biol. 2023;**24**(1):156.
<https://doi.org/10.1186/s13059-023-02990-1>
- Niaz M, Zhang B, Zhang Y, Yan X, Yuan M, Cheng Y, Lv G, Fadlalla T, Zhao L, Sun C, et al.** Genetic and molecular basis of carotenoid metabolism in cereals. Theor Appl Genet. 2023;**136**(3):63. <https://doi.org/10.1007/s00122-023-04336-8>
- Nisar N, Li L, Lu S, Khin NC, and Pogson BJ.** Carotenoid Metabolism in Plants. Molecular Plant. 2015;**8**(1):68–82. <https://doi.org/10.1016/j.molp.2014.12.007>

- Niyogi KK.** PHOTOPROTECTION REVISITED: Genetic and Molecular Approaches. *Annu Rev Plant Physiol Plant Mol Biol.* 1999;**50**:333–359. <https://doi.org/10.1146/annurev.arplant.50.1.333>
- Nkhata SG, Chilungo S, Memba A, and Mponela P.** Biofortification of maize and sweetpotatoes with provitamin A carotenoids and implication on eradicating vitamin A deficiency in developing countries. *Journal of Agriculture and Food Research.* 2020;**2**:100068. <https://doi.org/10.1016/j.jafr.2020.100068>
- O’Neil CE, Nicklas TA, Zanutto M, and Cho S.** Whole-Grain Consumption Is Associated with Diet Quality and Nutrient Intake in Adults: The National Health and Nutrition Examination Survey, 1999-2004. *Journal of the American Dietetic Association.* 2010;**110**(10):1461–1468. <https://doi.org/10.1016/j.jada.2010.07.012>
- Osorio CE.** The Role of Orange Gene in Carotenoid Accumulation: Manipulating Chromoplasts Toward a Colored Future. *Frontiers in Plant Science.* 2019;**10**:1235. <https://doi.org/10.3389/fpls.2019.01235>
- Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, et al.** Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nature Biotechnology.* 2005;**23**(4):482–487. <https://doi.org/10.1038/nbt1082>
- Pallotta MA, Warner P, Fox RL, Kuchel H, Jefferies SJ, and Langridge P.** Marker assisted wheat breeding in the southern region of Australia. *PROCEEDINGS OF THE INTERNATIONAL WHEAT GENETICS SYMPOSIUM.* 2003;**2**:789–791.
- Panfili G, Fratianni A, and Irano M.** Improved Normal-Phase High-Performance Liquid Chromatography Procedure for the Determination of Carotenoids in Cereals. *J Agric Food Chem.* 2004;**52**(21):6373–6377. <https://doi.org/10.1021/jf0402025>
- Paolillo DJ, Garvin DF, and Parthasarathy MV.** The chromoplasts of Or mutants of cauliflower (*Brassica oleracea* L. var. botrytis). *Protoplasma.* 2004;**224**(3–4):245–253. <https://doi.org/10.1007/s00709-004-0059-1>
- Park S, Kim HS, Jung YJ, Kim SH, Ji CY, Wang Z, Jeong JC, Lee HS, Lee SY, and Kwak SS.** Orange protein has a role in phytoene synthase stabilization in sweetpotato. *Scientific Reports.* 2016;**6**:1–12. <https://doi.org/10.1038/srep33563>
- Park SC, Kim SH, Park S, Lee HU, Lee JS, Park WS, Ahn MJ, Kim YH, Jeong JC, Lee HS, et al.** Enhanced accumulation of carotenoids in sweetpotato plants overexpressing IbOr-Ins gene in purple-fleshed sweetpotato cultivar. *Plant Physiology and Biochemistry.* 2015;**86**:82–90. <https://doi.org/10.1016/j.plaphy.2014.11.017>
- Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, et al.** InterPro in 2022. *Nucleic Acids Research.* 2023;**51**(D1):D418–D427. <https://doi.org/10.1093/nar/gkac993>
- Paznocht L, Kotíková Z, Orsák M, Lachman J, and Martinek P.** Carotenoid changes of colored-grain wheat flours during bun-making. *Food Chemistry.* 2019;**277**:725–734. <https://doi.org/10.1016/j.foodchem.2018.11.019>

- Pozniak CJ, Knox RE, Clarke FR, and Clarke JM.** Identification of QTL and association of a phytoene synthase gene with endosperm colour in durum wheat. *Theoretical and Applied Genetics*. 2007;**114**(3):525–537. <https://doi.org/10.1007/s00122-006-0453-5>
- Qin G, Gu H, Ma L, Peng Y, Deng XW, Chen Z, and Qu L-J.** Disruption of phytoene desaturase gene results in albino and dwarf phenotypes in *Arabidopsis* by impairing chlorophyll, carotenoid, and gibberellin biosynthesis. *Cell Res*. 2007;**17**(5):471–482. <https://doi.org/10.1038/cr.2007.40>
- Qin X, Fischer K, Dubcovsky J, and Tian L.** Endosperm carotenoid concentrations in wheat are better correlated with PSY1 transcript levels than enzyme activities. *Crop Science*. 2016;**56**(6):3173–3184. <https://doi.org/10.2135/cropsci2016.06.0520>
- R Core Team.** R: A language and environment for statistical computing (R Foundation for Statistical Computing: Vienna, Austria).
- Rahavi SMR and Kovalchuk I.** Changes in homologous recombination frequency in *Arabidopsis thaliana* plants exposed to stress depend on time of exposure during development and on duration of stress exposure. *Physiol Mol Biol Plants*. 2013;**19**(4):479–488. <https://doi.org/10.1007/s12298-013-0197-z>
- Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A, et al.** The transcriptional landscape of polyploid wheat. *Science*. 2018;**361**(6403):aar6089. <https://doi.org/10.1126/science.aar6089>
- Ramirez-Gonzalez RH, Segovia V, Bird N, Fenwick P, Holdgate S, Berry S, Jack P, Caccamo M, and Uauy C.** RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. *Plant Biotechnology Journal*. 2015a;**13**(5):613–624. <https://doi.org/10.1111/pbi.12281>
- Ramirez-Gonzalez RH, Uauy C, and Caccamo M.** PolyMarker: A fast polyploid primer design pipeline. *Bioinformatics*. 2015b;**31**(12):2038–2039. <https://doi.org/10.1093/bioinformatics/btv069>
- Rathan ND, Krishna H, Ellur RK, Sehgal D, Govindan V, Ahlawat AK, Krishnappa G, Jaiswal JP, Singh JB, Sv S, et al.** Genome-wide association study identifies loci and candidate genes for grain micronutrients and quality traits in wheat (*Triticum aestivum* L.). *Sci Rep*. 2022;**12**(1):7037. <https://doi.org/10.1038/s41598-022-10618-w>
- Reimer S, Pozniak CJ, Clarke FR, Clarke JM, Somers DJ, Knox RE, and Singh AK.** Association mapping of yellow pigment in an elite collection of durum wheat cultivars and breeding lines. *Genome*. 2008;**51**(12):1016–1025. <https://doi.org/10.1139/G08-083>
- Requena-Ramírez MD, Hornero-Méndez D, Rodríguez-Suárez C, and Atienza SG.** Durum Wheat (*Triticum durum* L.) Landraces Reveal Potential for the Improvement of Grain Carotenoid Esterification in Breeding Programs. *Foods*. 2021;**10**(4):757. <https://doi.org/10.3390/foods10040757>
- Requena-Ramírez MD, Rodríguez-Suárez C, Ávila CM, Palomino C, Hornero-Méndez D, and Atienza SG.** Bread Wheat Biofortification for Grain Carotenoid Content by

Inter-Specific Breeding. *Foods*. 2023;**12**(7):1365.
<https://doi.org/10.3390/foods12071365>

Requena-Ramírez MD, Rodríguez-Suárez C, Flores F, Hornero-Méndez D, and Atienza SG. Marker-Trait Associations for Total Carotenoid Content and Individual Carotenoids in Durum Wheat Identified by Genome-Wide Association Analysis. *Plants*. 2022;**11**(15):2065. <https://doi.org/10.3390/plants11152065>

Riboni M, Galbiati M, Tonelli C, and Conti L. GIGANTEA Enables Drought Escape Response via Abscisic Acid-Dependent Activation of the Florigens and SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1. *PLANT PHYSIOLOGY*. 2013;**162**(3):1706–1719.
<https://doi.org/10.1104/pp.113.217729>

Richaud D, Stange C, Gadaleta A, Colasuonno P, Parada R, and Schwember AR. Identification of Lycopene epsilon cyclase (LCYE) gene mutants to potentially increase β -carotene content in durum wheat (*Triticum turgidum* L.ssp. durum) through TILLING. *PLOS ONE*. 2018;**13**(12):e0208948.
<https://doi.org/10.1371/journal.pone.0208948>

Rocco M, Arena S, Renzone G, Scippa GS, Lomaglio T, Verrillo F, Scaloni A, and Marra M. Proteomic analysis of temperature stress-responsive proteins in *Arabidopsis thaliana* rosette leaves. *Mol BioSyst*. 2013;**9**(6):1257.
<https://doi.org/10.1039/c3mb70137a>

Rodríguez-Suárez C, Mellado-Ortega E, Hornero-Méndez D, and Atienza SG. Increase in transcript accumulation of *Psy1* and *e-Lcy* genes in grain development is associated with differences in seed carotenoid content between durum wheat and tritordeum. *Plant Mol Biol*. 2014;**84**(6):659–673. <https://doi.org/10.1007/s11103-013-0160-y>

Roselló M, Royo C, Álvaro F, Villegas D, Nazco R, and Soriano JM. Pasta-making quality QTLome from mediterranean durum wheat landraces. *Frontiers in Plant Science*. 2018;**871**:1512. <https://doi.org/10.3389/fpls.2018.01512>

Salamini F, Özkan H, Brandolini A, Schäfer-Pregl R, and Martin W. Genetics and geography of wild cereal domestication in the near east. *Nature Reviews Genetics*. 2002;**3**(6):429–441. <https://doi.org/10.1038/nrg817>

Sandmann G. Diversity and origin of carotenoid biosynthesis: its history of coevolution towards plant photosynthesis. *New Phytologist*. 2021;**232**(2):479–493.
<https://doi.org/10.1111/nph.17655>

Sanger JE. Quantitative Investigations of Leaf Pigments From Their Inception in Buds Through Autumn Coloration to Decomposition in Falling Leaves. *Ecology*. 1971;**52**(6):1075–1089. <https://doi.org/10.2307/1933816>

Sansaloni C, Franco J, Santos B, Percival-Alwyn L, Singh S, Petroli C, Campos J, Dreher K, Payne T, Marshall D, et al. Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. *Nat Commun*. 2020;**11**(1):4572. <https://doi.org/10.1038/s41467-020-18404-w>

- Schaub P, Wüst F, Koschmieder J, Yu Q, Virk P, Tohme J, and Beyer P.** Nonenzymatic β -Carotene Degradation in Provitamin A-Biofortified Crop Plants. *J Agric Food Chem.* 2017;**65**(31):6588–6598. <https://doi.org/10.1021/acs.jafc.7b01693>
- Schindele P and Puchta H.** Engineering CRISPR/ *Lb* Cas12a for highly efficient, temperature-tolerant plant gene editing. *Plant Biotechnology Journal.* 2020;**18**(5):1118–1120. <https://doi.org/10.1111/pbi.13275>
- Schramm EC, Nelson SK, Kidwell KK, and Steber CM.** Increased ABA sensitivity results in higher seed dormancy in soft white spring wheat cultivar ‘Zak.’ *Theor Appl Genet.* 2013;**126**(3):791–803. <https://doi.org/10.1007/s00122-012-2018-0>
- Schrödinger L and DeLano W.** PyMOL. 2020.
- Schweiggert RM, Steingass CB, Heller A, Esquivel P, and Carle R.** Characterization of chromoplasts and carotenoids of red- and yellow-fleshed papaya (*Carica papaya* L.). *Planta.* 2011;**234**(5):1031–1044. <https://doi.org/10.1007/s00425-011-1457-1>
- Sestili F, Garcia-Molina MD, Gambacorta G, Beleggia R, Botticella E, De Vita P, Savatin DV, Masci S, and Lafiandra D.** Provitamin A Biofortification of Durum Wheat through a TILLING Approach. *International Journal of Molecular Sciences.* 2019;**20**(22):5703. <https://doi.org/10.3390/ijms20225703>
- Shan Y-J, Li D, Cao J-J, Zhang L, Han L-Q, Zhang M-P, and Shen Z-G.** Over-expression of Arabidopsis ORANGE gene enhances drought stress tolerance through ABA-dependent pathway in Arabidopsis thaliana. *Plant Growth Regul.* 2022;**96**(1):91–101. <https://doi.org/10.1007/s10725-021-00760-2>
- Sheoran S, Jaiswal S, Raghav N, Sharma R, Sabhyata, Gaur A, Jaisri J, Tandon G, Singh S, Sharma P, et al.** Genome-Wide Association Study and Post-genome-Wide Association Study Analysis for Spike Fertility and Yield Related Traits in Bread Wheat. *Frontiers in Plant Science.* 2022;**12**.
- Shewry PR and Hey S.** Do “ancient” wheat species differ from modern bread wheat in their contents of bioactive components? *Journal of Cereal Science.* 2015;**65**:236–243. <https://doi.org/10.1016/j.jcs.2015.07.014>
- Shi W, Hao C, Zhang Y, Cheng J, Zhang Z, Liu J, Yi X, Cheng X, Sun D, Xu Y, et al.** A combined association mapping and linkage analysis of kernel number per spike in common wheat (*Triticum aestivum* L.). *Frontiers in Plant Science.* 2017;**8**(August):1–13. <https://doi.org/10.3389/fpls.2017.01412>
- Singh A, Reimer S, Pozniak CJ, Clarke FR, Clarke JM, Knox RE, and Singh AK.** Allelic variation at Psy1-A1 and association with yellow pigment in durum wheat grain. *Theoretical and Applied Genetics.* 2009;**118**(8):1539–1548. <https://doi.org/10.1007/s00122-009-1001-x>
- Smedley MA, Hayta S, Clarke M, and Harwood WA.** CRISPR-Cas9 Based Genome Editing in Wheat. *Current Protocols.* 2021;**1**(3):1–15. <https://doi.org/10.1002/cpz1.65>
- Smith BM, Ramsay SA, Roe A, Ferrante MJ, and Brooks S (Worden).** Reducing Visual Differences in Whole Grain Bread Prepared with Hard Red and Hard White Wheat:

Application for Sensory Studies. *Journal of Food Science*. 2019;**84**(8):2325–2329.
<https://doi.org/10.1111/1750-3841.14722>

Sretenovic S, Yin D, Levav A, Selengut JD, Mount SM, and Qi Y. Expanding Plant Genome-Editing Scope by an Engineered iSpyMacCas9 System That Targets A-Rich PAM Sequences. *Plant Communications*. 2020:100101.
<https://doi.org/10.1016/j.xplc.2020.100101>

Subira J, Peña RJ, Álvaro F, Ammar K, Ramdani A, and Royo C. Breeding progress in the pasta-making quality of durum wheat cultivars released in Italy and Spain during the 20th Century. *Crop Pasture Sci*. 2014;**65**(1):16.
<https://doi.org/10.1071/CP13238>

Sun C, Lei Y, Li B, Gao Q, Li Y, Cao W, Yang C, Li H, Wang Z, Li Y, et al. Precise integration of large DNA sequences in plant genomes using PrimeRoot editors. *Nat Biotechnol*. 2023a. <https://doi.org/10.1038/s41587-023-01769-w>

Sun T and Li L. Toward the ‘golden’ era: The status in uncovering the regulatory control of carotenoid accumulation in plants. *Plant Science*. 2020;**290**(November 2019).
<https://doi.org/10.1016/j.plantsci.2019.110331>

Sun T, Rao S, Zhou X, and Li L. Plant carotenoids: recent advances and future perspectives. *Mol Horticulture*. 2022;**2**(1):3. <https://doi.org/10.1186/s43897-022-00023-2>

Sun T, Wang P, Rao S, Zhou X, Wrightstone E, Lu S, Yuan H, Yang Y, Fish T, Thannhauser T, et al. Co-chaperoning of chlorophyll and carotenoid biosynthesis by ORANGE family proteins in plants. *Molecular Plant*. 2023b;**16**(6):1048–1065.
<https://doi.org/10.1016/j.molp.2023.05.006>

Sun T, Yuan H, Cao H, Yazdani M, Tadmor Y, and Li L. Carotenoid Metabolism in Plants: The Role of Plastids. *Molecular Plant*. 2018;**11**(1):58–74.
<https://doi.org/10.1016/j.molp.2017.09.010>

Sun T, Yuan H, Chen C, Kadirjan-Kalbach DK, Mazourek M, Osteryoung KW, and Li L. ORHis, a Natural Variant of OR, Specifically Interacts with Plastid Division Factor ARC3 to Regulate Chromoplast Number and Carotenoid Accumulation. *Molecular Plant*. 2020;**13**(6):864–878. <https://doi.org/10.1016/j.molp.2020.03.007>

Sun T, Zhu Q, Wei Z, Owens LA, Fish T, Kim H, Thannhauser TW, Cahoon EB, and Li L. Multi-strategy engineering greatly enhances provitamin A carotenoid accumulation and stability in Arabidopsis seeds. *aBIOTECH*. 2021;**2**(3):191–214.
<https://doi.org/10.1007/s42994-021-00046-1>

Suriano S, Codianni P, and Iannucci A. Carotenoids and tocopherols comparison in different Subspecies of *Triticum turgidum* and *aestivum*. *Food Research International*. 2023:113620. <https://doi.org/10.1016/j.foodres.2023.113620>

Suzuki M, Takahashi S, Kondo T, Dohra H, Ito Y, Kiriwa Y, Hayashi M, Kamiya S, Kato M, Fujiwara M, et al. Plastid Proteomic Analysis in Tomato Fruit Development. *PLOS ONE*. 2015;**10**(9):e0137266. <https://doi.org/10.1371/journal.pone.0137266>

- Swarts DC and Jinek M.** Cas9 versus Cas12a/Cpf1: Structure–function comparisons and implications for genome editing. *WIREs RNA*. 2018;**9**(5):e1481. <https://doi.org/10.1002/wrna.1481>
- Tan B-C, Joseph LM, Deng W-T, Liu L, Li Q-B, Cline K, and McCarty DR.** Molecular characterization of the Arabidopsis 9-cis epoxy-carotenoid dioxygenase gene family. *The Plant Journal*. 2003;**35**(1):44–56. <https://doi.org/10.1046/j.1365-313X.2003.01786.x>
- Tanaka R and Tanaka A.** Tetrapyrrole Biosynthesis in Higher Plants. *Annu Rev Plant Biol*. 2007;**58**(1):321–346. <https://doi.org/10.1146/annurev.arplant.57.032905.105448>
- Torkamaneh D and Belzile F eds.** *Genome-Wide Association Studies* (Springer US: New York, NY). <https://doi.org/10.1007/978-1-0716-2237-7>
- Tsilo TJ, Hareland GA, Chao S, and Anderson JA.** Genetic mapping and QTL analysis of flour color and milling yield related traits using recombinant inbred lines in hard red spring wheat. *Crop Science*. 2011;**51**(1):237–246. <https://doi.org/10.2135/cropsci2009.12.0711>
- Tzuri G, Zhou X, Chayut N, Yuan H, Portnoy V, Meir A, Sa'ar U, Baumkoler F, Mazourek M, Lewinsohn E, et al.** A 'golden' SNP in CmOr governs the fruit flesh color of melon (*Cucumis melo*). *The Plant Journal*. 2015;**82**(2):267–279. <https://doi.org/10.1111/tpj.12814>
- UK Parliament.** Genetic Technology (Precision Breeding) Act. 2023. <https://www.legislation.gov.uk/ukpga/2023/6/contents/enacted>
- Vallabhaneni R, Bradbury LMT, and Wurtzel ET.** The carotenoid dioxygenase gene family in maize, sorghum, and rice. *Archives of Biochemistry and Biophysics*. 2010;**504**(1):104–111. <https://doi.org/10.1016/j.abb.2010.07.019>
- Vallabhaneni R and Wurtzel ET.** Timing and Biosynthetic Potential for Carotenoid Accumulation in Genetically Diverse Germplasm of Maize. *Plant Physiology*. 2009;**150**(2):562–572. <https://doi.org/10.1104/pp.109.137042>
- Van der Auwera GA and O'Connor BD.** Genomics in the cloud: using Docker, GATK, and WDL in Terra (O'Reilly Media).
- Varadi M, Bertoni D, Magana P, Paramval U, Pidruchna I, Radhakrishnan M, Tsenkov M, Nair S, Mirdita M, Yeo J, et al.** AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research*. 2024;**52**(D1):D368–D375. <https://doi.org/10.1093/nar/gkad1011>
- Vargas VH, Schulthess A, Royo C, Matus I, and Schwember AR.** Transcripts levels of Phytoene synthase 1 (Psy-1) are associated to semolina yellowness variation in durum wheat (*Triticum turgidum* L. ssp. durum). *Journal of Cereal Science*. 2016;**68**:155–163. <https://doi.org/10.1016/j.jcs.2016.01.011>
- Villa TCC, Maxted N, Scholten M, and Ford-Lloyd B.** Defining and identifying crop landraces. *Plant Genetic Resources*. 2005;**3**(3):373–384. <https://doi.org/10.1079/PGR200591>

- Vogel JT, Tan B-C, McCarty DR, and Klee HJ.** The Carotenoid Cleavage Dioxygenase 1 Enzyme Has Broad Substrate Specificity, Cleaving Multiple Carotenoids at Two Different Bond Positions*. *Journal of Biological Chemistry*. 2008;**283**(17):11364–11373. <https://doi.org/10.1074/jbc.M710106200>
- Walkowiak S, Gao L, Monat C, Haberer G, Kassa MT, Brinton J, Ramirez-Gonzalez RH, Kolodziej MC, Delorean E, Thambugala D, et al.** Multiple wheat genomes reveal global variation in modern breeding. *Nature* 2020. 2020:(April):1–7. <https://doi.org/10.1038/s41586-020-2961-x>
- Wang C, Zeng J, Li Y, Hu W, Chen L, Miao Y, Deng P, Yuan C, Ma C, Chen X, et al.** Enrichment of provitamin A content in wheat (*Triticum aestivum* L.) by introduction of the bacterial carotenoid biosynthetic genes *CrtB* and *CrtI*. *Journal of Experimental Botany*. 2014;**65**(9):2545–2556. <https://doi.org/10.1093/jxb/eru138>
- Wang J and Zhang Z.** GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genomics Proteomics Bioinformatics*. 2021;**19**(4):629–640. <https://doi.org/10.1016/j.gpb.2021.08.005>
- Wang K, Shi L, Liang X, Zhao P, Wang W, Liu J, Chang Y, Hiei Y, Yanagihara C, Du L, et al.** The gene *TaWOX5* overcomes genotype dependency in wheat genetic transformation. *Nature Plants*. 2022a;**8**(2):110–117. <https://doi.org/10.1038/s41477-021-01085-8>
- Wang P and Grimm B.** Connecting Chlorophyll Metabolism with Accumulation of the Photosynthetic Apparatus. *Trends in Plant Science*. 2021;**26**(5):484–495. <https://doi.org/10.1016/j.tplants.2020.12.005>
- Wang P, Tian T, Ma J, Liu Y, Zhang P, Chen T, Shahinnia F, and Yang D.** Genome-Wide Association Study of Kernel Traits Using a 35K SNP Array in Bread Wheat (*Triticum aestivum* L.). *Frontiers in Plant Science*. 2022b;**13**.
- Wang W, Tian B, Pan Q, Chen Y, He F, Bai G, Akhunova A, and Harold N.** Expanding the range of editable targets in the wheat genome using the variants of the *Cas12a* and *Cas9* nucleases. 2020:1–27.
- Wang Y, Li L, Ye T, Lu Y, Chen X, and Wu Y.** The inhibitory effect of ABA on floral transition is mediated by *ABI5* in *Arabidopsis*. *Journal of Experimental Botany*. 2013;**64**(2):675–684. <https://doi.org/10.1093/jxb/ers361>
- Wang Z, Ke Q, Kim MD, Kim SH, Ji CY, Jeong JC, Lee HS, Park WS, Ahn MJ, Li H, et al.** Transgenic alfalfa plants expressing the sweetpotato orange gene exhibit enhanced abiotic stress tolerance. *PLoS ONE*. 2015;**10**(5):1–17. <https://doi.org/10.1371/journal.pone.0126050>
- Wang Z, Xu W, Kang J, Li M, Huang J, Ke Q, Kim HS, Xu B, and Kwak SS.** Overexpression of alfalfa Orange gene in tobacco enhances carotenoid accumulation and tolerance to multiple abiotic stresses. *Plant Physiology and Biochemistry*. 2018;**130**(August):613–622. <https://doi.org/10.1016/j.plaphy.2018.08.017>
- Watkins JL, Li M, McQuinn RP, Chan KX, McFarlane HE, Ermakova M, Furbank RT, Mares D, Dong C, Chalmers KJ, et al.** A GDSL esterase/lipase catalyzes the esterification of

lutein in bread wheat. *Plant Cell*. 2019;**31**(12):3092–3112.
<https://doi.org/10.1105/tpc.19.00272>

Watkins JL and Pogson BJ. Prospects for Carotenoid Biofortification Targeting Retention and Catabolism. *Trends in Plant Science*. 2020;**25**(5):501–512.
<https://doi.org/10.1016/j.tplants.2019.12.021>

Welsch R and Li L. Golden Rice—Lessons learned for inspiring future metabolic engineering strategies and synthetic biology solutions. . In. *Methods in Enzymology*. (Elsevier), pp. 1–29. <https://doi.org/10.1016/bs.mie.2022.03.014>

Welsch R, Zhou X, Koschmieder J, Schlossarek T, Yuan H, Sun T, and Li L. Characterization of Cauliflower OR Mutant Variants. *Frontiers in Plant Science*. 2020;**10**(January):1–13. <https://doi.org/10.3389/fpls.2019.01716>

Welsch R, Zhou X, Yuan H, Álvarez D, Sun T, Schlossarek D, Yang Y, Shen G, Zhang H, Rodriguez-Concepcion M, et al. Clp Protease and OR Directly Control the Proteostasis of Phytoene Synthase, the Crucial Enzyme for Carotenoid Biosynthesis in Arabidopsis. *Molecular Plant*. 2018;**11**(1):149–162.
<https://doi.org/10.1016/j.molp.2017.11.003>

Werner S, Engler C, Weber E, Gruetzner R, and Marillonnet S. Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioeng Bugs*. 2012;**3**(1):38–43. <https://doi.org/10.4161/bbug.3.1.18223>

WHO. Vitamin A deficiency. Vitamin and Mineral Nutrition Information System (VMNIS). 2024. <https://www.who.int/data/nutrition/nlis/info/vitamin-a-deficiency>. Retrieved March 16, 2024

Wickham H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York).

Wickham H, François R, Henry L, Müller K, and Vaughan D. *dplyr: A grammar of data manipulation*.

Wickham H, Vaughan D, and Girlich M. *tidyr: Tidy messy data*.

Wingen LU, Orford S, Goram R, Leverington-Waite M, Bilham L, Patsiou TS, Ambrose M, Dicks J, and Griffiths S. Establishing the A. E. Watkins landrace cultivar collection as a resource for systematic gene discovery in bread wheat. *Theoretical and Applied Genetics*. 2014;**127**(8):1831–1842. <https://doi.org/10.1007/s00122-014-2344-5>

Wurtzel ET, Cuttriss A, and Vallabhaneni R. Maize Provitamin A Carotenoids, Current Resources, and Future Metabolic Engineering Challenges. *Front Plant Sci*. 2012;**3**. <https://doi.org/10.3389/fpls.2012.00029>

Xu W, Yang Y, Yang B, Krueger CJ, Xiao Q, Zhao S, Zhang L, Kang G, Wang F, Yi H, et al. A design optimized prime editor with expanded scope and capability in plants. *Nature Plants*. 2021. <https://doi.org/10.1038/s41477-021-01043-4>

Yang F, Zhang J, Liu Q, Liu H, Zhou Y, Yang W, and Ma W. Improvement and Re-Evolution of Tetraploid Wheat for Global Environmental Challenge and Diversity

Consumption Demand. IJMS. 2022:**23**(4):2206.
<https://doi.org/10.3390/ijms23042206>

Yao E, Blake VC, Cooper L, Wight CP, Michel S, Cagirici HB, Lazo GR, Birkett CL, Waring DJ, Jannink J-L, et al. GrainGenes: a data-rich repository for small grains genetics and genomics. Database. 2022:**2022**:baac034.
<https://doi.org/10.1093/database/baac034>

Yates AD, Allen J, Amode RM, Azov AG, Barba M, Becerra A, Bhai J, Campbell LI, Carbajo Martinez M, Chakiachvili M, et al. Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. Nucleic Acids Research. 2022:**50**(D1):D996–D1003. <https://doi.org/10.1093/nar/gkab1007>

Yazdani M, Sun Z, Yuan H, Zeng S, Thannhauser TW, Vrebalov J, Ma Q, Xu Y, Fei Z, Van Eck J, et al. Ectopic expression of ORANGE promotes carotenoid accumulation and fruit development in tomato. Plant Biotechnology Journal. 2019:**17**(1):33–49.
<https://doi.org/10.1111/pbi.12945>

Ye X and Beyer P. Engineering the provitamin A (β -carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science. 2000:**287**(5451):303–305.
<https://doi.org/10.1126/science.287.5451.303>

Yu S, Li M, Dubcovsky J, and Tian L. Mutant combinations of *lycopene ϵ -cyclase* and *β -carotene hydroxylase 2* homoeologs increased β -carotene accumulation in endosperm of tetraploid wheat (*Triticum turgidum* L.) grains. Plant Biotechnology Journal. 2022:**20**(3):564–576. <https://doi.org/10.1111/pbi.13738>

Yu S and Tian L. Assessing the Role of Carotenoid Cleavage Dioxygenase 4 Homoeologs in Carotenoid Accumulation and Plant Growth in Tetraploid Wheat. Front Nutr. 2021:**8**:740286. <https://doi.org/10.3389/fnut.2021.740286>

Yu Y, Yu J, Wang Q, Wang J, Zhao G, Wu H, Zhu Y, Chu C, and Fang J. Overexpression of the rice ORANGE gene OsOR negatively regulates carotenoid accumulation, leads to higher tiller numbers and decreases stress tolerance in Nipponbare rice. Plant Science. 2021:**310**(January):110962.
<https://doi.org/10.1016/j.plantsci.2021.110962>

Yuan H, Owsiany K, Sheeja TE, Zhou X, Rodriguez C, Li Y, Welsch R, Chayut N, Yang Y, Thannhauser TW, et al. A Single Amino Acid Substitution in an ORANGE Protein Promotes Carotenoid Overaccumulation in Arabidopsis. Plant Physiology. 2015:**169**(1):421–431. <https://doi.org/10.1104/pp.15.00971>

Zeng J, Wang X, Miao Y, Wang C, Zang M, Chen X, Li M, Li X, Wang Q, Li K, et al. Metabolic Engineering of Wheat Provitamin A by Simultaneously Overexpressing CrtB and Silencing Carotenoid Hydroxylase (*TaHYD*). Journal of Agricultural and Food Chemistry. 2015:**63**(41):9083–9092.
<https://doi.org/10.1021/acs.jafc.5b04279>

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell. 2015:**163**(3):759–771.
<https://doi.org/10.1016/j.cell.2015.09.038>

- Zhang R, Liu J, Chai Z, Chen S, Bai Y, Zong Y, Chen K, Li J, Jiang L, and Gao C.** Generation of herbicide tolerance traits and a new selectable marker in wheat using base editing. *Nat Plants*. 2019;**5**(5):480–485. <https://doi.org/10.1038/s41477-019-0405-0>
- Zhang W, Chao S, Manthey F, Chicaiza O, Brevis JC, Echenique V, and Dubcovsky J.** QTL analysis of pasta quality using a composite microsatellite and SNP map of durum wheat. *Theoretical and Applied Genetics*. 2008;**117**(8):1361–1377. <https://doi.org/10.1007/s00122-008-0869-1>
- Zhang W and Dubcovsky J.** Association between allelic variation at the Phytoene synthase 1 gene and yellow pigment content in the wheat grain. *Theoretical and Applied Genetics*. 2008;**116**(5):635–645. <https://doi.org/10.1007/s00122-007-0697-8>
- Zheng X, Giuliano G, and Al-Babili S.** Carotenoid biofortification in crop plants: citius, altius, fortius. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*. 2020a;**1865**(11):158664. <https://doi.org/10.1016/j.bbalip.2020.158664>
- Zheng X, Kuijter HNJ, and Al-Babili S.** Carotenoid Biofortification of Crops in the CRISPR Era. *Trends in Biotechnology*. 2020b;**xx**(xx):1–4. <https://doi.org/10.1016/j.tibtech.2020.12.003>
- Zhong Z, Sretenovic S, Ren Q, Yang L, Bao Y, Qi C, Yuan M, He Y, Liu S, Liu X, et al.** Improving Plant Genome Editing with High-Fidelity xCas9 and Non-canonical PAM-Targeting Cas9-NG. *Molecular Plant*. 2019;**12**(7):1027–1036. <https://doi.org/10.1016/j.molp.2019.03.011>
- Zhou X, Rao S, Wrightstone E, Sun T, Lui ACW, Welsch R, and Li L.** Phytoene Synthase: The Key Rate-Limiting Enzyme of Carotenoid Biosynthesis in Plants. *Front Plant Sci*. 2022;**13**:884720. <https://doi.org/10.3389/fpls.2022.884720>
- Zhou X and Stephens M.** Genome-wide efficient mixed-model analysis for association studies. *Nat Genet*. 2012;**44**(7):821–824. <https://doi.org/10.1038/ng.2310>
- Zhou X, Welsch R, Yang Y, Álvarez D, Riediger M, Yuan H, Fish T, Liu J, Thannhauser TW, and Li L.** Arabidopsis OR proteins are the major posttranscriptional regulators of phytoene synthase in controlling carotenoid biosynthesis. *Proceedings of the National Academy of Sciences*. 2015;**112**(11):3558–3563. <https://doi.org/10.1073/pnas.1420831112>
- Zhu X, Li S, Cooper RS, and Elston RC.** A Unified Association Analysis Approach for Family and Unrelated Samples Correcting for Stratification. *Am J Hum Genet*. 2008;**82**(2):352–365. <https://doi.org/10.1016/j.ajhg.2007.10.009>
- Ziegler JU, Wahl S, Würschum T, Longin CFH, Carle R, and Schweiggert RM.** Lutein and Lutein Esters in Whole Grain Flours Made from 75 Genotypes of 5 *Triticum* Species Grown at Multiple Sites. *J Agric Food Chem*. 2015;**63**(20):5061–5071. <https://doi.org/10.1021/acs.jafc.5b01477>
- Zong Y, Liu Y, Xue C, Li B, Li X, Wang Y, Li J, Liu G, Huang X, Cao X, et al.** An engineered prime editor with enhanced editing efficiency in plants. *Nature Biotechnology*. 2022. <https://doi.org/10.1038/s41587-022-01254-w>

Zong Y, Song Q, Li C, Jin S, Zhang D, Wang Y, Qiu JL, and Gao C. Efficient c-to-t base editing in plants using a fusion of ncas9 and human apobec3a. *Nature Biotechnology*. 2018;**36**(10):950. <https://doi.org/10.1038/nbt.4261>

Zong Y, Wang Y, Li C, Zhang R, Chen K, Ran Y, Qiu J-L, Wang D, and Gao C. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nature Biotechnology*. 2017;**35**(5):438–440. <https://doi.org/10.1038/nbt.3811>

Appendices

Appendix 1 – List of Appendix Materials

Appendix Material 1 Photo of field-grown Watkins tetraploid collection before harvest.....	195
Appendix Material 2 A comparison of the high-throughput YPC method and the industry-standard AACC 14-50 method.	195
Appendix Material 3 Full results of the carotenoid content of the Watkins tetraploid collection analysis.	196
Appendix Material 4 Relationship between yellow pigment content (YPC) and thousand-grain weight (TGW) in wheat.	204
Appendix Material 5 Relationship between yellow pigment content (YPC) and HPLC total grain carotenoid content (Total GCC).	204
Appendix Material 6 Number of Watkins tetraploid accessions with the presence or absence of the three unknown peaks.	205
Appendix Material 7 Relationship between HPLC-measured total carotenoid content and thousand-grain weight (TGW) in the Watkins tetraploid collection and Miradoux field controls.	205
Appendix Material 8 Proportions of each carotenoid compound within the Watkins tetraploid collection.	206
Appendix Material 9 Full results of the carotenoid content of the Watkins hexaploid core collection analysis.	207
Appendix Material 10 Comparison of α -carotene, β -carotene, lutein, and zeaxanthin content across three replicates from the 106-accession Watkins hexaploid core collection.	210
Appendix Material 11 Pairwise comparisons of grain carotenoid content of the three replicates from the 106-accession Watkins hexaploid core collection.	211
Appendix Material 12 Associated 35K Breeder's Array markers with carotenoid content in the Watkins tetraploid collection. Pos.=IWGSC RefSeq v1.0 genomic position.	212
Appendix Material 13 <i>T. turgidum</i> genes identified as orthologous to carotenoid biosynthesis genes in rice and Arabidopsis.	214
Appendix Material 14 Deformities in Kronos spike architecture leading to sterility.....	216
Appendix Material 15 Bar charts of individual carotenoid compounds of the T_1 TaOR ^{WT} and TaOR ^{His} overexpression lines compared to the GRF and wild-type controls.	218
Appendix Material 16 Pairwise comparisons of grain carotenoid content of T_1 TaOR ^{WT} and TaOR ^{His} overexpression lines compared to GRF controls.	218
Appendix Material 17 Pairwise comparisons of grain yield per plant and anthesis date of T_1 TaOR ^{WT} and TaOR ^{His} overexpression lines compared to GRF controls.	219

Appendix Material 18 Thousand-grain weight and grain area of T ₁ TaOR ^{WT} and TaOR ^{His} overexpression lines compared to GRF and wild-type controls.	221
Appendix Material 19 Relative leaf chlorophyll content of T ₁ TaOR ^{WT} and TaOR ^{His} overexpression lines compared to GRF and wild-type controls.....	221
Appendix Material 20 Pairwise comparisons of height, grain number per plant, TGW, grain area and SPAD of T ₁ TaOR ^{WT} and TaOR ^{His} overexpression lines compared to GRF controls and non-transgenic Cadenza plants.....	222
Appendix Material 21 Alignment of the TaOR-6A, TaOR-6B and TaOR-6D protein sequences within the wheat pangenome.	225
Appendix Material 22 Protein sequence alignment of 35 OR orthologues at the E141K residue identified within durum wheat TdOR TILLING lines.....	226

Appendix 2 – Appendix Materials of Chapter 3



Appendix Material 1 Photo of field-grown Watkins tetraploid collection before harvest. Diversity in the colour of spikelets can be observed.

Appendix Material 2 A comparison of the high-throughput YPC method and the industry-standard AACC 14-50 method. Five YPC measurements were made on the same Miradoux flour sample. Absorbance was measured in a standard lab spectrophotometer using a 10 mm cuvette rather than a plate reader due to the low number of samples being tested. The AACC 14-50 method gave a slightly higher average value for Yellow Pigment Content (YPC) of 10.18 $\mu\text{g/g}$, compared to the high-throughput YPC method, 9.99 $\mu\text{g/g}$; however, there was no significant difference between the two methods ($p=0.38$, $t(8)=0.93$, Student's t-test).

Method	YPC 1	YPC 2	YPC 3	YPC 4	YPC 5	Mean RRC ($\mu\text{g/g}$)	CV (%)
High-throughput YPC	10.46	9.50	10.19	9.99	9.80	9.99 (SD=0.37)	3.67
AACC 14-50	9.78	10.36	10.41	10.21	10.11	10.17 (SD=0.25)	2.46

Appendix Material 3 Full results of the carotenoid content of the Watkins tetraploid collection analysis. YPC, α -carotene, β -carotene, lutein, zeaxanthin and total grain carotenoid content are measured in $\mu\text{g/g}$. YPC=yellow pigment content, α -caro.= α -carotene, β -caro.= β -carotene, Zeaxanth.=zeaxanthin, Total GCC=total grain carotenoid content, Prop β - β .=proportion of β - β carotenoids.

Accession	YPC	α -caro.	β -caro.	Lutein	Zeaxanth.	Total GCC	Prop β - β .
WAT1180001	3.0186	0.1126	0.0417	0.3066	0.1417	0.6026	0.3043
WAT1180002	3.8593	0.0436	0.0158	0.1642	0.0631	0.2866	0.2751
WAT1180003	3.9614	0.0328	0.0376	0.5963	0.0685	0.7352	0.1444
WAT1180004	6.6128	0.0366	0.0584	1.9299	0.5325	2.5574	0.2310
WAT1180005	3.9676	0.0187	0.0308	0.7867	0.1803	1.0165	0.2077
WAT1180006	6.0994	0.0251	0.0569	1.0983	0.2131	1.3933	0.1938
WAT1180007	5.7166	0.0293	0.0352	0.7890	0.1144	0.9679	0.1545
WAT1180008	6.4406	0.0234	0.0402	0.7550	0.1551	0.9736	0.2006
WAT1180009	4.3934	0.0171	0.0256	0.3961	0.1355	0.5742	0.2805
WAT1180010	5.9433	0.0336	0.0188	0.9946	0.4028	1.4498	0.2908
WAT1180011	5.2741	0.0279	0.0394	0.6963	0.2699	1.0334	0.2993
WAT1180012		0.0189	0.0468	0.6216	0.1415	0.8289	0.2272
WAT1180013	3.5668	0.0182	0.0432	0.5628	0.1911	0.8154	0.2874
WAT1180014	4.6206	0.0200	0.0255	0.4673	0.2286	0.7414	0.3428
WAT1180015	3.3295	0.0138	0.0226	0.2912	0.1564	0.4840	0.3699
WAT1180016	3.2405	0.0225	0.0333	0.2713	0.1871	0.5142	0.4286
WAT1180018	2.9107	0.0327	0.0139	0.2161	0.0674	0.3301	0.2462
WAT1180019	3.4816	0.0164	0.0261	0.3493	0.1467	0.5385	0.3208
WAT1180020	4.2177	0.0200	0.0437	0.7616	0.1036	0.9290	0.1586
WAT1180021	2.7825	0.0227	0.0247	0.2885	0.1354	0.4713	0.3396
WAT1180022	5.9442	0.0221	0.0451	0.8610	0.1382	1.0664	0.1719
WAT1180023	3.5148	0.0205	0.0195	0.2895	0.1482	0.4777	0.3510
WAT1180024	3.2841	0.0125	0.0260	0.5206	0.1848	0.7439	0.2833
WAT1180025	3.7843	0.0131	0.0271	0.5004	0.0824	0.6230	0.1758
WAT1180026	4.1208	0.0138	0.0236	0.6856	0.2514	0.9744	0.2823
WAT1180027	5.8608	0.0771	0.0396	0.3818	0.1414	0.6399	0.2828
WAT1180028	2.6913	0.0134	0.0172	0.2403	0.1101	0.3810	0.3342
WAT1180029	5.1505	0.0323	0.0570	1.0015	0.3455	1.4363	0.2802
WAT1180030		0.0066	0.0209	0.3641	0.1991	0.5907	0.3724
WAT1180031	4.5213	0.0091	0.0182	0.2954	0.1295	0.4522	0.3266
WAT1180032		0.0126	0.0330	0.4907	0.2254	0.7618	0.3393
WAT1180034	3.4856	0.0098	0.0147	0.3300	0.1351	0.4897	0.3060
WAT1180035		0.0477	0.0215	0.2882	0.1375	0.4949	0.3214
WAT1180036	4.1625	0.0438	0.0246	0.3085	0.1461	0.5230	0.3264
WAT1180037		0.0129	0.0446	0.3629	0.1656	0.5860	0.3587
WAT1180038	3.0448	0.0456	0.0180	0.3720	0.1120	0.5476	0.2374
WAT1180040	3.0705	0.0194	0.0417	0.4224	0.2480	0.7315	0.3960
WAT1180041	3.7505	0.0068	0.0135	0.2820	0.1285	0.4308	0.3296
WAT1180042	2.7974	0.0264	0.0343	0.3426	0.1968	0.6001	0.3850
WAT1180043	3.0841	0.0124	0.0172	0.4435	0.1965	0.6695	0.3191

WAT1180044	5.2403	0.0179	0.0269	0.5426	0.2389	0.8264	0.3217
WAT1180045	4.6937	0.0266	0.0220	0.2043	0.1081	0.3610	0.3604
WAT1180046		0.0225	0.0372	0.6302	0.1196	0.8096	0.1937
WAT1180047	4.9007	0.0170	0.0302	0.4193	0.1932	0.6596	0.3386
WAT1180048		0.0161	0.0322	0.4389	0.1854	0.6725	0.3235
WAT1180049	4.2135	0.0343	0.0373	0.5484	0.1707	0.7907	0.2630
WAT1180050	3.0396	0.0140	0.0339	0.5017	0.1027	0.6524	0.2095
WAT1180051	5.2199	0.0263	0.0340	0.9619	0.3229	1.3451	0.2654
WAT1180052	4.4895	0.0913	0.0161	0.2298	0.1375	0.4746	0.3235
WAT1180053	4.8802	0.1372	0.0245	0.3039	0.1706	0.6361	0.3066
WAT1180054	4.9734	0.0209	0.0468	0.7089	0.3684	1.1450	0.3626
WAT1180055		0.0320	0.0480	0.7567	0.2824	1.1191	0.2952
WAT1180056	5.6681	0.0311	0.0491	0.8641	0.1935	1.1377	0.2132
WAT1180057	4.3273	0.0153	0.0239	0.7152	0.3490	1.1034	0.3380
WAT1180058	3.2276	0.0144	0.0336	0.3490	0.0461	0.4432	0.1800
WAT1180059	3.7787	0.0132	0.0311	0.4489	0.2560	0.7490	0.3832
WAT1180060	3.1255	0.0149	0.0337	0.4035	0.1735	0.6256	0.3312
WAT1180061	3.8468	0.0412	0.0211	0.8858	0.0382	0.9862	0.0601
WAT1180062	2.7905	0.0314	0.0187	0.2456	0.1817	0.4774	0.4198
WAT1180063	3.3246	0.0107	0.0146	0.3575	0.1481	0.5309	0.3064
WAT1180064	4.0030	0.0182	0.0237	0.4033	0.1706	0.6159	0.3156
WAT1180065	2.3810	0.0103	NF	0.3609	0.1381	0.5094	0.2712
WAT1180066	3.8186	0.0120	0.0381	0.4057	0.1403	0.5961	0.2992
WAT1180067	3.8843	0.0186	0.0479	0.5511	0.1964	0.8140	0.3001
WAT1180068	3.9895	0.0137	0.0303	0.4737	0.1485	0.6661	0.2683
WAT1180069	5.2745	0.0217	0.0348	0.6056	0.2232	0.8854	0.2915
WAT1180070	3.5497	0.0155	0.0329	0.4209	0.0561	0.5253	0.1694
WAT1180071	4.6602	0.0401	0.0411	0.6186	0.2281	0.9279	0.2901
WAT1180072	4.5094	0.0076	0.0208	0.3514	0.1511	0.5308	0.3238
WAT1180073	4.4577	0.0655	0.0383	0.2903	0.0645	0.4586	0.2242
WAT1180074		0.0110	0.0321	0.3747	0.1226	0.5404	0.2862
WAT1180075	3.4373	0.0185	0.0244	0.4925	0.1619	0.6973	0.2671
WAT1180079		0.0097	0.0164	0.3560	0.1422	0.5243	0.3026
WAT1180080		0.0086	0.0470	0.3156	0.1305	0.5017	0.3537
WAT1180082	3.5847	0.0102	0.0111	0.2987	0.1345	0.4545	0.3204
WAT1180083	5.4957	0.0145	0.0358	0.6064	0.0465	0.7032	0.1171
WAT1180084	3.6843	0.0148	0.0175	0.3744	0.1291	0.5358	0.2737
WAT1180085	3.6713	0.0288	0.0306	0.8001	0.1615	1.0210	0.1882
WAT1180086	3.5766	0.0209	0.0256	0.5109	0.1318	0.6892	0.2283
WAT1180087	3.3366	0.0114	0.0286	0.3056	0.1238	0.4693	0.3245
WAT1180088		0.0121	0.0192	0.5168	0.1456	0.6938	0.2376
WAT1180089	4.5591	0.0151	0.0189	0.4119	0.1799	0.6258	0.3177
WAT1180090	4.1174	0.0188	0.0319	0.5475	0.1568	0.7550	0.2500
WAT1180091	3.5293	0.0171	0.0276	0.6468	0.1683	0.8598	0.2279
WAT1180092	3.9317	0.0097	0.0165	0.5116	0.1802	0.7180	0.2740
WAT1180093		0.0129	0.0159	0.6906	0.2953	1.0147	0.3066
WAT1180094	4.5628	0.0110	0.0650	0.5281	0.1820	0.7861	0.3142

WAT1180095	4.3022	0.0206	0.0474	0.3872	0.1627	0.6179	0.3400
WAT1180096	3.8429	0.0273	0.0433	0.3877	0.2061	0.6643	0.3754
WAT1180097	4.5670	0.1201	0.0394	0.4472	0.2039	0.8106	0.3001
WAT1180098		0.0126	0.0292	0.4126	0.1742	0.6286	0.3235
WAT1180099	2.9495	0.0286	0.0101	0.2537	0.1587	0.4511	0.3742
WAT1180100	4.2941	0.0163	0.0259	0.4035	0.2195	0.6652	0.3689
WAT1180101	3.6705	0.0126	0.0146	0.3145	0.1427	0.4844	0.3246
WAT1180102	3.8706	0.0147	0.0206	0.3941	0.1749	0.6044	0.3236
WAT1180103	4.0949	0.0242	0.0310	0.5298	0.2150	0.8001	0.3075
WAT1180104	4.0479	0.0149	0.0337	0.4662	0.0833	0.5982	0.1957
WAT1180105	5.0186	0.2072	0.0331	0.5098	0.1099	0.8600	0.1663
WAT1180106	4.0044	0.0625	0.0267	0.3226	0.1765	0.5883	0.3453
WAT1180107		0.0102	0.0186	0.4368	0.0987	0.5645	0.2079
WAT1180108	4.3395	0.0273	0.0331	0.7195	0.1704	0.9502	0.2141
WAT1180109	4.8339	0.0172	0.0172	0.4440	0.1937	0.6720	0.3138
WAT1180110	3.5100	0.0114	0.0152	0.5069	0.2312	0.7646	0.3222
WAT1180111	4.1661	0.0118	0.0147	0.4988	0.1551	0.6804	0.2496
WAT1180112	4.3384						
WAT1180113	6.5522	0.0316	0.0346	0.7657	0.1897	1.0216	0.2195
WAT1180114	4.1498	0.0144	0.0173	0.4229	0.1869	0.6416	0.3183
WAT1180115	6.2673	0.0338	0.0154	0.8860	0.3600	1.2952	0.2899
WAT1180116	5.0655	0.0292	0.0322	0.6358	0.2489	0.9461	0.2971
WAT1180117	3.8867	0.0125	0.0241	0.3627	0.1212	0.5204	0.2791
WAT1180118	6.4194	0.0275	0.0275	0.6082	0.2789	0.9421	0.3252
WAT1180119	3.9013	0.0179	0.0245	0.4863	0.2079	0.7365	0.3155
WAT1180120	4.3502	0.0253	0.0273	0.4611	0.1881	0.7019	0.3069
WAT1180121	3.8014	0.0159	0.0189	0.4805	0.1522	0.6675	0.2563
WAT1180122	3.5621	0.0029	0.0173	0.4038	0.2230	0.6470	0.3715
WAT1180123		0.0129	0.0387	0.3822	0.1372	0.5711	0.3081
WAT1180124		0.0076	0.0132	0.5105	0.1163	0.6476	0.2000
WAT1180125	3.7026	0.0185	0.0195	0.5387	0.2206	0.7973	0.3011
WAT1180126	4.8636	0.0220	0.0259	0.5626	0.2962	0.9067	0.3552
WAT1180127	3.7004	0.0362	0.0542	1.0639	0.2954	1.4497	0.2412
WAT1180128		0.0128	0.0306	0.6110	0.2547	0.9091	0.3138
WAT1180129	3.2871	0.0124	0.0324	0.3576	0.0448	0.4472	0.1727
WAT1180130	5.4797	0.0065	0.0214	0.3605	0.1062	0.4947	0.2580
WAT1180131	5.1254	0.0191	0.0300	0.4768	0.2102	0.7362	0.3263
WAT1180132	3.2624	0.0135	0.0222	0.3161	0.1224	0.4742	0.3049
WAT1180133	3.1424	0.0140	0.0308	0.3995	0.1456	0.5899	0.2991
WAT1180134	3.6790	0.0244	0.0272	0.2783	0.1284	0.4583	0.3395
WAT1180135		0.0212	0.0292	0.5098	0.1048	0.6650	0.2015
WAT1180136	4.3547	0.0181	0.0221	0.3724	0.1736	0.5862	0.3339
WAT1180137	3.7588	0.0079	0.0148	0.3678	0.2719	0.6624	0.4328
WAT1180138	3.2474	0.0127	0.0342	0.4199	0.1504	0.6171	0.2991
WAT1180139	3.1847	0.0073	0.0238	0.3537	0.1933	0.5782	0.3756
WAT1180140	3.4591	0.0168	0.0298	0.5657	0.2274	0.8396	0.3063
WAT1180141	4.5037	0.0235	0.0337	0.5313	0.2866	0.8750	0.3660

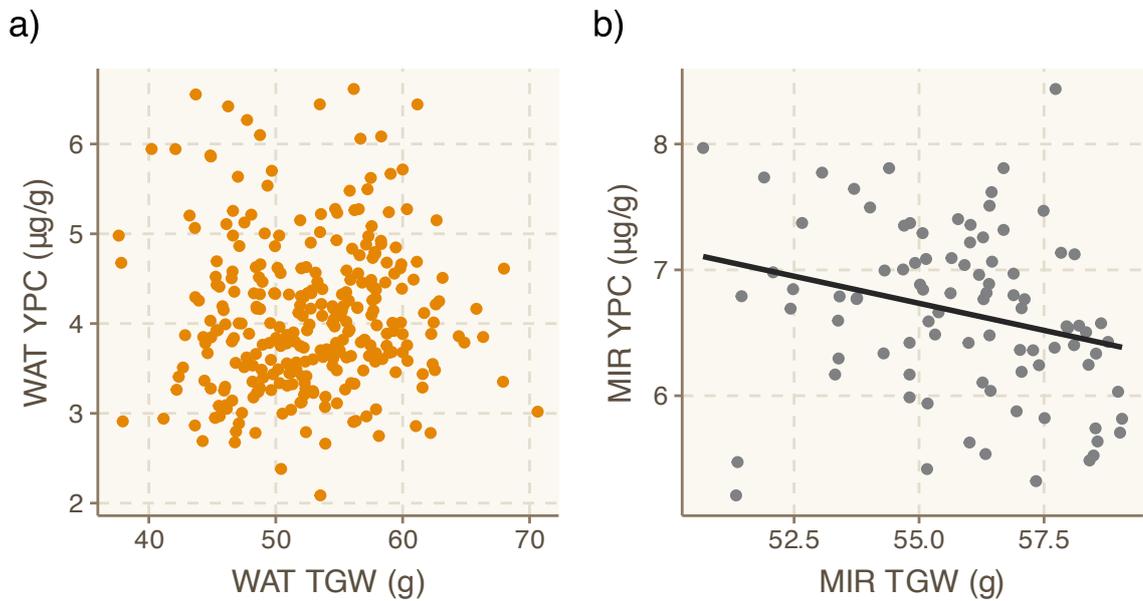
WAT1180142	2.7480	0.0173	0.0250	0.3868	0.2217	0.6507	0.3791
WAT1180143		0.0249	0.0428	0.5111	0.1720	0.7507	0.2861
WAT1180144	4.3686	0.0160	0.0208	0.6669	0.2613	0.9650	0.2923
WAT1180145	5.8728	0.0182	0.0316	0.6616	0.2269	0.9382	0.2755
WAT1180147	5.0048	0.0113	0.0272	0.5625	0.2447	0.8457	0.3215
WAT1180148	4.2036	0.0178	0.0225	0.6247	0.2064	0.8714	0.2626
WAT1180149	4.3884	0.0392	0.0422	1.2738	0.1824	1.5376	0.1460
WAT1180150	5.2333	0.0170	0.0331	0.4087	0.2084	0.6672	0.3619
WAT1180151	3.4774	0.0129	0.0184	0.2957	0.1386	0.4655	0.3373
WAT1180152	4.4322	0.0179	0.0311	0.4729	0.1931	0.7150	0.3136
WAT1180153	3.9812	0.0142	0.0233	0.5215	0.1539	0.7128	0.2486
WAT1180154	3.5956	0.0169	0.0263	0.8970	0.3357	1.2760	0.2837
WAT1180155	4.2556	0.0216	0.0253	0.5469	0.2007	0.7945	0.2845
WAT1180156	3.8177	0.0318	0.0527	0.6402	0.2734	0.9981	0.3267
WAT1180157	4.0049	0.0168	0.0248	0.4994	0.1714	0.7124	0.2754
WAT1180158	4.9803	0.0343	0.0384	0.6217	0.1867	0.8811	0.2554
WAT1180159	3.9253	0.0650	0.0288	0.4292	0.2434	0.7664	0.3552
WAT1180160	3.4880	0.0207	0.0335	0.4333	0.2236	0.7110	0.3615
WAT1180161	3.7530	0.0208	0.0161	0.4589	0.1926	0.6883	0.3032
WAT1180162	3.6756	0.0251	0.0300	0.7378	0.0812	0.8741	0.1272
WAT1180163	3.3533	0.0090	0.0230	0.2883	0.1391	0.4594	0.3529
WAT1180164	4.8618	0.1539	0.0397	0.3942	0.2323	0.8201	0.3317
WAT1180166		0.0322	0.0427	0.5841	0.1659	0.8250	0.2529
WAT1180167	3.4904	0.0970	0.0427	0.4607	0.2706	0.8709	0.3597
WAT1180168		0.0234	0.0380	1.0278	0.2506	1.3398	0.2154
WAT1180169	2.9405	0.0599	0.0230	0.3147	0.1998	0.5973	0.3729
WAT1180170	4.5152	0.1316	0.0445	0.4522	0.2168	0.8451	0.3092
WAT1180171		0.0747	0.0208	0.3366	0.1390	0.5711	0.2798
WAT1180172	6.4420	0.0277	0.0356	0.7529	0.2529	1.0691	0.2699
WAT1180173	5.1506	0.0162	0.0293	0.4151	0.1828	0.6433	0.3297
WAT1180174	3.8824	0.0353	0.0429	0.5830	0.0706	0.7318	0.1551
WAT1180175	5.6223	0.0431	0.0307	1.0991	0.1907	1.3635	0.1623
WAT1180176	3.9102	0.0167	0.0324	0.5805	0.2161	0.8457	0.2938
WAT1180177	3.8389	0.0165	0.0185	0.3858	0.1331	0.5539	0.2737
WAT1180178	3.6389	0.0239	0.0299	0.5248	0.1404	0.7190	0.2368
WAT1180179	3.6045	0.0327	0.0466	0.8369	0.3104	1.2266	0.2910
WAT1180180	4.7790	0.0175	0.0224	0.5432	0.2083	0.7915	0.2915
WAT1180181		0.0158	0.0258	0.5815	0.2021	0.8252	0.2761
WAT1180182	4.3308	0.0233	0.0398	0.6911	0.2951	1.0493	0.3191
WAT1180183	3.4345	0.0218	0.0437	0.6804	0.2247	0.9706	0.2765
WAT1180184	4.4105	0.0860	0.0377	0.2938	0.2252	0.6428	0.4090
WAT1180185	5.2553	0.0995	0.0313	0.4182	0.1676	0.7167	0.2776
WAT1180186	3.6405	0.0143	0.0352	0.4056	0.1162	0.5713	0.2650
WAT1180187	4.0863	0.1103	0.0355	0.3565	0.2216	0.7239	0.3551
WAT1180188	3.8729	0.0880	0.0271	0.3586	0.1895	0.6631	0.3265
WAT1180189	3.7134	0.0770	0.0276	0.3137	0.1873	0.6056	0.3548
WAT1180190	4.6279	0.0192	0.0240	0.4615	0.1564	0.6610	0.2729

WAT1180191	4.6607	0.0239	0.0339	0.5742	0.2183	0.8503	0.2966
WAT1180192	3.2656	0.0157	0.0185	0.3384	0.1479	0.5205	0.3197
WAT1180193		0.0280	0.0395	0.6239	0.1321	0.8235	0.2084
WAT1180194		0.0155	0.0174	0.3271	0.2156	0.5757	0.4048
WAT1180195	4.1645	0.0073	0.0165	0.2468	0.1408	0.4113	0.3822
WAT1180196	4.0870	0.0291	0.0191	0.3231	0.1515	0.5227	0.3263
WAT1180197	4.0319	0.0065	0.0214	0.3161	0.1404	0.4844	0.3340
WAT1180198	3.1124	0.0068	0.0213	0.3535	0.1521	0.5337	0.3249
WAT1180199	5.6353	0.0170	0.0509	0.4922	0.2356	0.7957	0.3601
WAT1180200		0.0187	0.0187	0.3722	0.1828	0.5923	0.3402
WAT1180201	3.4577	0.0141	0.0201	0.3858	0.1527	0.5727	0.3018
WAT1180202	4.7963	0.0294	0.0423	0.6305	0.2610	0.9631	0.3149
WAT1180203	4.0881	0.0353	0.0400	0.7663	0.0848	0.9264	0.1348
WAT1180204	3.1202	0.0109	0.0257	0.5382	0.1699	0.7446	0.2626
WAT1180205	2.9687	0.0085	0.0180	0.4092	0.2051	0.6408	0.3481
WAT1180206	2.9162	0.0233	0.0417	0.6826	0.2172	0.9647	0.2683
WAT1180207	3.8503	0.0150	0.0230	0.3534	0.1752	0.5667	0.3498
WAT1180208	2.8649	0.0148	0.0269	0.3358	0.2143	0.5918	0.4075
WAT1180209	2.8878	0.0060	0.0219	0.4421	0.1275	0.5975	0.2500
WAT1180210	3.8399	0.0106	0.0106	0.4499	0.1894	0.6604	0.3028
WAT1180211	3.4142	0.0178	0.0366	0.4572	0.2948	0.8065	0.4109
WAT1180212	2.6772	0.0137	0.0284	0.4050	0.2025	0.6496	0.3554
WAT1180213	3.3375	0.0507	0.0185	0.2868	0.1463	0.5024	0.3282
WAT1180214	4.7312	0.0271	0.0321	0.7220	0.1245	0.9058	0.1729
WAT1180215	3.2535	0.0233	0.0359	0.4744	0.1824	0.7159	0.3049
WAT1180216	4.9220	0.0267	0.0426	0.5782	0.2505	0.8979	0.3264
WAT1180217	4.0120	0.0207	0.0424	0.4029	0.1836	0.6496	0.3478
WAT1180218	3.7583	0.0170	0.0246	0.3579	0.1269	0.5265	0.2878
WAT1180219	3.7885	0.0233	0.0409	0.1209	0.1711	0.3562	0.5953
WAT1180220	2.8583	0.0690	0.0212	0.3149	0.1749	0.5800	0.3381
WAT1180221	3.3519	0.0255	0.0358	0.4290	0.1820	0.6723	0.3240
WAT1180222	4.3254	0.0364	0.0488	0.6889	0.2201	0.9942	0.2705
WAT1180224		0.0248	0.0328	0.8686	0.1231	1.0492	0.1485
WAT1180225	6.0838	0.0276	0.0466	0.9343	0.1884	1.1968	0.1963
WAT1180226	4.3386	0.0224	0.0386	0.5866	0.2114	0.8590	0.2911
WAT1180227	3.7845	0.0202	0.0193	0.3925	0.1761	0.6080	0.3213
WAT1180229	3.3369	0.0530	0.0210	0.2950	0.1380	0.5069	0.3136
WAT1180230		0.0157	0.0246	0.4613	0.0639	0.5655	0.1565
WAT1180231	4.2485	0.1380	0.0477	0.4934	0.2710	0.9502	0.3354
WAT1180232	4.6120	0.0909	0.0247	0.3736	0.1512	0.6405	0.2747
WAT1180233		0.0312	0.0445	0.7740	0.3653	1.2150	0.3372
WAT1180234	4.6790	0.0197	0.0334	0.5562	0.2231	0.8323	0.3081
WAT1180235	4.4647	0.0156	0.0313	0.4196	0.1629	0.6294	0.3085
WAT1180236	4.5914	0.0207	0.0273	0.4022	0.3297	0.7799	0.4577
WAT1180237	4.7613	0.0128	0.0236	0.4065	0.3751	0.8179	0.4874
WAT1180238	4.8864	0.0176	0.0296	0.4824	0.4268	0.9564	0.4773
WAT1180239	3.8195	0.0201	0.0440	0.5457	0.4586	1.0685	0.4704

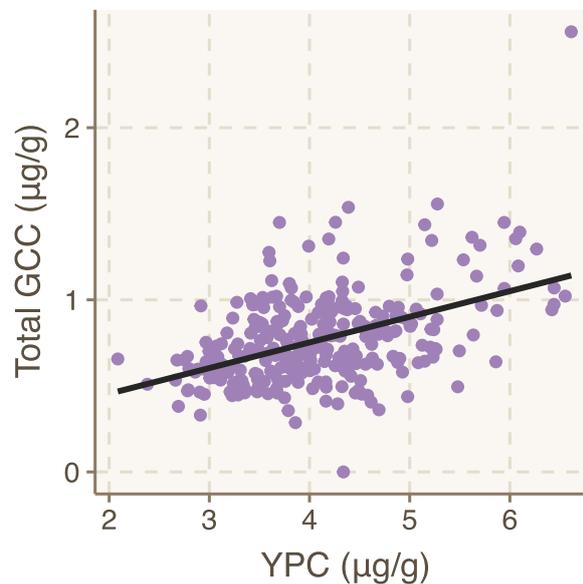
WAT1180240	4.6881	0.0200	0.0310	0.3591	0.3901	0.8002	0.5263
WAT1180241	4.2597	0.0316	0.0454	0.6592	0.7145	1.4507	0.5238
WAT1180242	4.4860	0.0121	0.0372	0.4468	0.3683	0.8644	0.4692
WAT1180243	3.9234	0.0220	0.0345	0.4236	0.1246	0.6047	0.2631
WAT1180244	5.0851	0.0181	0.0372	0.4466	0.1326	0.6346	0.2677
WAT1180245	4.1795	0.0153	0.0248	0.5471	0.1258	0.7130	0.2112
WAT1180246		0.0106	0.0222	0.4318	0.1128	0.5774	0.2337
WAT1180247	3.4581	0.0147	0.0423	0.7066	0.2093	0.9729	0.2586
WAT1180248	4.4156	0.0155	0.0228	0.4038	0.1133	0.5555	0.2451
WAT1180249	4.3760	0.0246	0.0404	0.6594	0.2178	0.9422	0.2741
WAT1180250	2.9664	0.0137	0.0330	0.5187	0.1879	0.7533	0.2932
WAT1180251	5.5360	0.0272	0.0494	0.9784	0.1775	1.2326	0.1841
WAT1180252	3.3088	0.0177	0.0242	0.4977	0.0521	0.5917	0.1289
WAT1180253		0.0165	0.0389	0.4761	0.1489	0.6805	0.2761
WAT1180254	3.4071	0.0213	0.0184	0.4969	0.1811	0.7178	0.2780
WAT1180255	3.2960	0.0137	0.0220	0.4935	0.1428	0.6721	0.2452
WAT1180256	4.0005	0.0137	0.0333	0.5599	0.1204	0.7273	0.2113
WAT1180257	4.1921	0.0395	0.0633	1.0028	0.2477	1.3533	0.2298
WAT1180258	5.2638	0.0112	0.0184	0.4905	0.1927	0.7128	0.2961
WAT1180260	3.6611	0.0223	0.0407	0.5539	0.0600	0.6769	0.1488
WAT1180261	3.3991	0.0141	0.0263	0.4821	0.1692	0.6917	0.2826
WAT1180262		0.0207	0.0266	0.6209	0.0306	0.6987	0.0818
WAT1180263	3.2755	0.0172	0.0383	0.6297	0.3005	0.9857	0.3437
WAT1180264	3.6131	0.0850	0.0277	0.2947	0.1792	0.5866	0.3528
WAT1180265	3.5569	0.0983	0.0321	0.5434	0.3298	1.0037	0.3606
WAT1180266	4.1403	0.0269	0.0389	0.5136	0.2791	0.8585	0.3704
WAT1180267	2.0863	0.0155	0.0262	0.3617	0.2531	0.6565	0.4254
WAT1180268	3.5426	0.0262	0.0424	0.6762	0.2099	0.9548	0.2643
WAT1180270	5.1077	0.0248	0.0200	0.6791	0.1946	0.9185	0.2336
WAT1180271		0.0165	0.0214	0.6184	0.1456	0.8019	0.2082
WAT1180272	2.9968	0.0089	0.0218	0.4932	0.1776	0.7016	0.2843
WAT1180273	3.5326	0.0270	0.0280	0.7387	0.2201	1.0139	0.2447
WAT1180274	3.6225	0.0171	0.0251	0.6750	0.1004	0.8176	0.1536
WAT1180275	4.2558	0.0261	0.0280	0.4113	0.1516	0.6169	0.2911
WAT1180276		0.0164	0.0265	0.6890	0.2175	0.9495	0.2570
WAT1180277	3.4461	0.0093	0.0333	0.6135	0.2304	0.8865	0.2975
WAT1180278	5.7020	0.0170	0.0368	0.8284	0.4349	1.3171	0.3582
WAT1180279	3.7995	0.0144	0.0287	0.6350	0.4166	1.0947	0.4068
WAT1180280	4.1931	0.0149	0.0308	0.7188	0.2536	1.0181	0.2793
WAT1180281	4.3369	0.0253	0.0379	1.0102	0.1692	1.2426	0.1667
WAT1180282		0.0253	0.0448	0.9260	0.1862	1.1824	0.1954
WAT1180283	4.1665	0.0234	0.0273	0.8955	0.0654	1.0116	0.0916
WAT1180284	3.7645	0.1279	0.0248	0.3768	0.2409	0.7704	0.3449
WAT1180285	5.2769	0.0495	0.0595	1.2345	0.2134	1.5569	0.1753
WAT1180286	6.0589	0.0394	0.0497	1.0908	0.1752	1.3551	0.1660
WAT1180287	3.6225	0.0329	0.0357	0.8692	0.1738	1.1116	0.1885
WAT1180288	4.2963	0.0292	0.0367	0.3762	0.2097	0.6518	0.3781

WAT1180289	4.6148	0.0815	0.0211	0.1951	0.1081	0.4059	0.3183
WAT1180290	3.8877	0.0251	0.0533	0.6886	0.1940	0.9610	0.2573
WAT1180291	4.0087	0.0202	0.0386	0.4334	0.1956	0.6877	0.3405
WAT1180292	3.5107	0.0142	0.0427	0.6365	0.1726	0.8661	0.2486
WAT1180293	4.1458	0.0219	0.0409	0.6988	0.1666	0.9282	0.2236
WAT1180294	4.1936	0.0227	0.0463	0.5881	0.1522	0.8093	0.2453
WAT1180295	2.7814	0.0640	0.0286	0.2835	0.2954	0.6715	0.4824
WAT1180296	3.4799	0.0300	0.0237	0.3732	0.1402	0.5670	0.2889
WAT1180297	4.3055	0.0111	0.0292	0.4445	0.3034	0.7881	0.4220
WAT1180298	3.9245	0.0101	0.0221	0.4759	0.0673	0.5755	0.1554
WAT1180299	3.7508	0.0243	0.0395	0.5271	0.3173	0.9082	0.3929
WAT1180300		0.0155	0.0291	0.4253	0.2509	0.7209	0.3884
WAT1180301	4.2817	0.0046	0.0191	0.2953	0.0766	0.3956	0.2419
WAT1180302	3.6837	0.0252	0.0369	0.6309	0.1669	0.8599	0.2370
WAT1180303	3.7919	0.0186	0.0333	0.4178	0.1399	0.6096	0.2841
WAT1180306	4.4562	0.0196	0.0451	0.6103	0.2737	0.9488	0.3361
WAT1180307	4.6296	0.0145	0.0407	0.5493	0.2257	0.8302	0.3209
WAT1180308	4.4655	0.0176	0.0353	0.5208	0.1012	0.6750	0.2022
WAT1180309	4.4877	0.0380	0.0631	0.8180	0.1549	1.0740	0.2029
WAT1180310	4.3236	0.0113	0.0160	0.3281	0.1401	0.4955	0.3150
WAT1180311	4.9283	0.0143	0.0286	0.4332	0.1040	0.5801	0.2286
WAT1180312	3.6015	0.0212	0.0344	0.5483	0.2124	0.8164	0.3024
WAT1180313	4.1737	0.0153	0.0305	0.6178	0.1478	0.8113	0.2197
WAT1180314	3.2593	0.0240	0.0157	0.3505	0.1264	0.5166	0.2750
WAT1180315	4.3845	0.0181	0.0322	0.4514	0.1673	0.6690	0.2982
WAT1180316	3.9179	0.0159	0.0309	0.5047	0.2509	0.8024	0.3511
WAT1180317	3.5763	0.0167	0.0373	0.4980	0.2284	0.7803	0.3405
WAT1180318	3.4123	0.0203	0.0324	0.6189	0.3353	1.0069	0.3652
WAT1180319		0.0185	0.0342	0.5523	0.2387	0.8437	0.3235
WAT1180321	3.2342	0.0216	0.0324	0.4474	0.3923	0.8937	0.4752
WAT1180322	4.0117						
WAT1180323	4.1662	0.0095	0.0255	0.4075	0.0501	0.4926	0.1536
WAT1180325	4.0216	0.0148	0.0259	0.6340	0.2129	0.8876	0.2690
WAT1180326	4.8480	0.0203	0.0415	0.6213	0.1322	0.8152	0.2130
WAT1180327	4.5564	0.0184	0.0312	0.7019	0.0936	0.8450	0.1477
WAT1180328		0.0211	0.0422	0.4889	0.1541	0.7062	0.2779
WAT1180329	4.6151	0.0590	0.0261	0.3386	0.2206	0.6443	0.3829
WAT1180330	4.9799	0.0152	0.0238	0.3718	0.0266	0.4375	0.1152
WAT1180331		0.0249	0.0211	0.3000	0.2061	0.5520	0.4115
WAT1180332	4.9803	0.0396	0.0495	0.7359	0.4120	1.2370	0.3731
WAT1180333	5.2127	0.0311	0.0412	0.7087	0.0485	0.8296	0.1082
WAT1180335	4.6765	0.0186	0.0419	0.4945	0.0708	0.6258	0.1801
WAT1180336	3.2752	0.0259	0.0288	0.3168	0.1526	0.5242	0.3462
WAT1180337	5.2034	0.0196	0.0290	0.5336	0.1430	0.7252	0.2371
WAT1180338		0.1302	0.0503	0.6815	0.0582	0.9202	0.1179
WAT1180339	2.6624	0.0746	0.0191	0.2593	0.1799	0.5330	0.3734
WAT1180341	3.8501	0.0121	0.0353	0.5014	0.2552	0.8040	0.3614

WAT1180342	3.9916	0.0329	0.0432	0.8799	0.3559	1.3119	0.3042
WAT1180343	3.2076	0.0168	0.0336	0.3066	0.1084	0.4655	0.3052
WAT1180345	3.0053	0.0147	0.0376	0.3990	0.0945	0.5458	0.2420
WAT1180346	4.5799	0.0128	0.0293	0.3687	0.0349	0.4457	0.1440
WAT1180347	3.7013	0.0379	0.0359	0.6669	0.2193	0.9600	0.2658
WAT1180368	2.9043	0.0133	0.0257	0.3201	0.1048	0.4639	0.2813
WAT1180369	3.0913	0.0315	0.0278	0.3157	0.2194	0.5944	0.4159
WAT1180370	3.3645	0.0557	0.0264	0.2539	0.1369	0.4730	0.3453
WAT1180371		0.0356	0.0543	0.5080	0.1562	0.7540	0.2792
WAT1180372	3.7822	0.1135	0.0353	0.3327	0.1649	0.6464	0.3097
WAT1180373	3.1769	0.0163	0.0316	0.4522	0.3076	0.8076	0.4199
WAT1180374	3.0503	0.0082	0.0165	0.3392	0.2496	0.6135	0.4337
WAT1180376	3.7318	0.0264	0.0401	0.7455	0.0665	0.8786	0.1214
WAT1180383		0.0225	0.0176	0.6053	0.0791	0.7244	0.1334
WAT1180391	4.1871	0.0308	0.0556	0.7091	0.2066	1.0021	0.2616
WAT1180392	3.6281	0.0308	0.0417	0.7900	0.1528	1.0153	0.1916



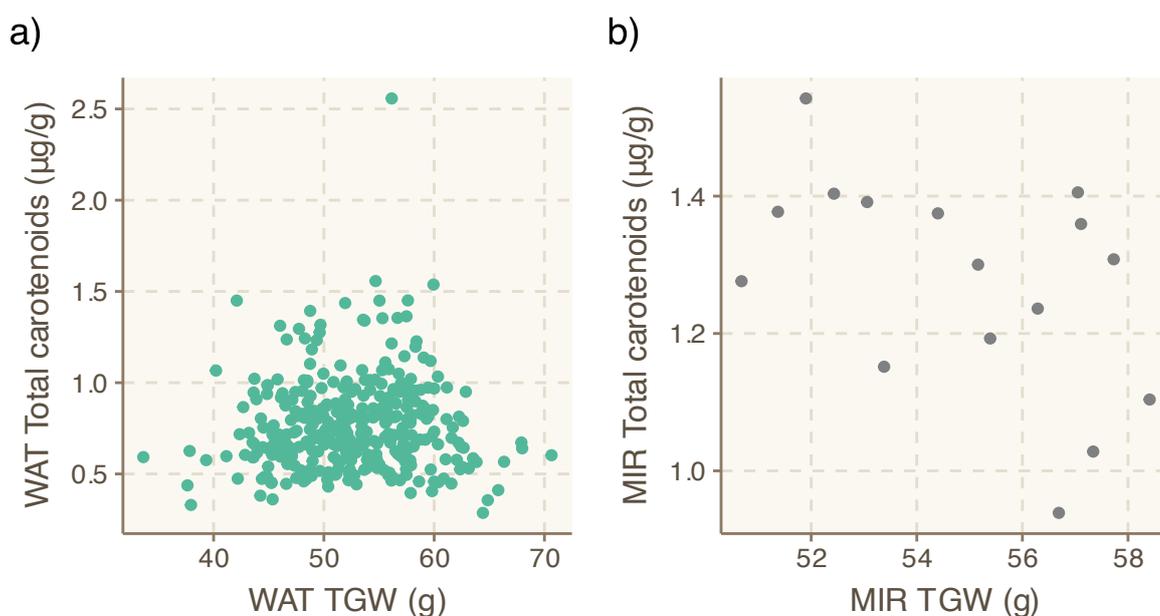
Appendix Material 4 Relationship between yellow pigment content (YPC) and thousand-grain weight (TGW) in wheat. (a) Scatterplot showing the relationship between YPC and TGW for accessions within the Watkins tetraploid collection (WAT). Linear regression analysis indicated no significant relationship. (b) Scatterplot depicting the relationship between YPC and TGW for the Miradoux field controls (MIR). A significant negative relationship was observed ($p=0.013$, $\beta=-0.086$), represented by the equation: $\text{YPC} = 11.469 - 0.086 \times \text{TGW}$.



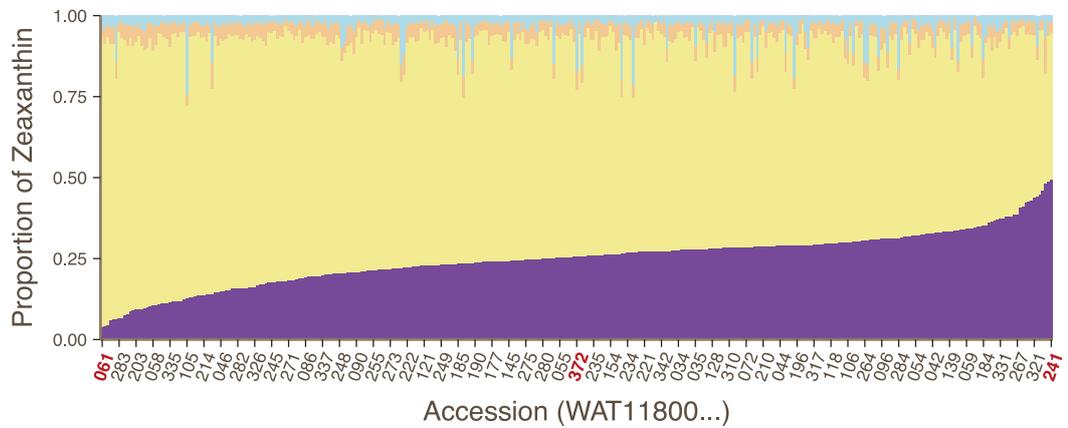
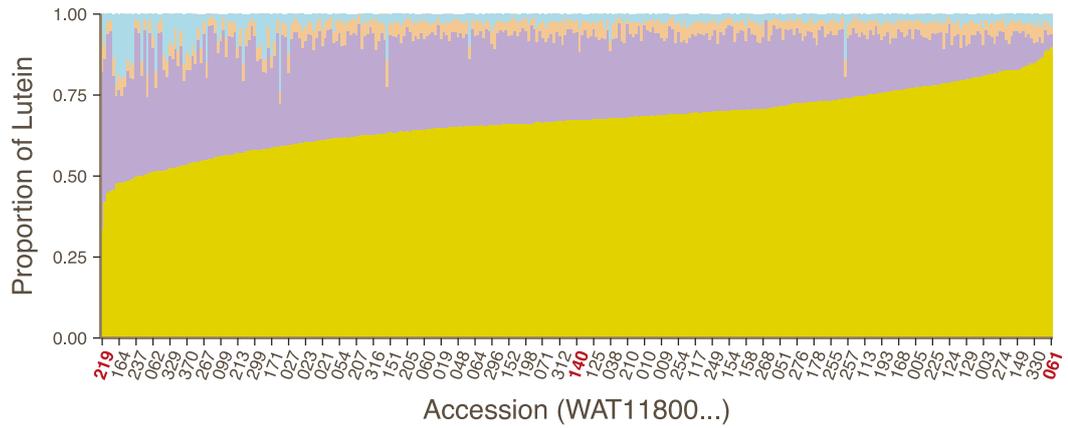
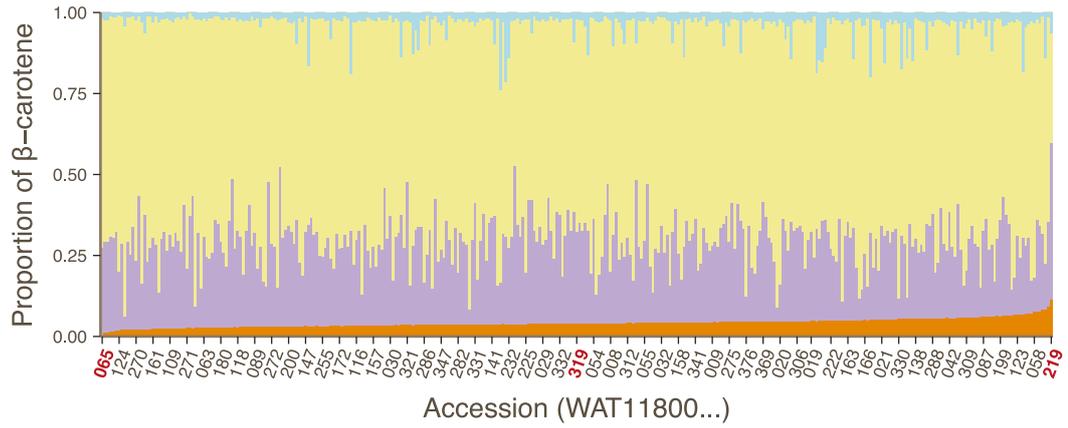
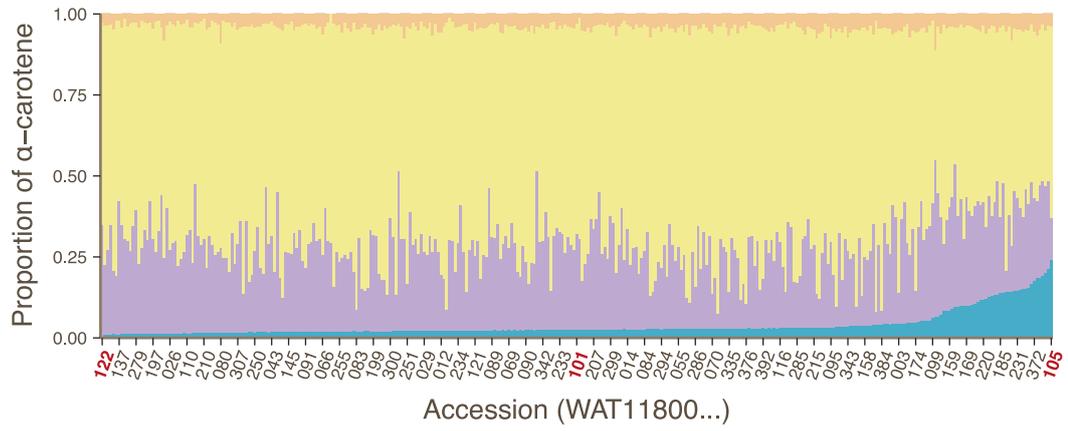
Appendix Material 5 Relationship between yellow pigment content (YPC) and HPLC total grain carotenoid content (Total GCC). A significant positive relationship was observed ($p<0.001$, $\beta=1.445$), represented by the equation: $\text{YPC} = 2.978 + 1.445 \times \text{Total GCC}$.

Appendix Material 6 Number of Watkins tetraploid accessions with the presence or absence of the three unknown peaks. Peaks were identified in the HPLC analysis of the Watkins tetraploid collection analysis.

Peak A presence	✓	✓	✓	✓	✓	✓	✓	
Peak B presence		✓	✓		✓		✓	
Peak C presence			✓		✓	✓	✓	
Number of accessions	25	89	29	7	117	11	9	50



Appendix Material 7 Relationship between HPLC-measured total carotenoid content and thousand-grain weight (TGW) in the Watkins tetraploid collection and Miradoux field controls. (a) Scatterplot showing the relationship between total carotenoids and TGW for accessions within the Watkins tetraploid collection (WAT). Linear regression analysis indicated no significant relationship. (b) Scatterplot showing the relationship between total carotenoids and TGW for the Miradoux field controls (MIR). A near-significant relationship was observed ($p=0.0532$, $\beta=-0.031$).



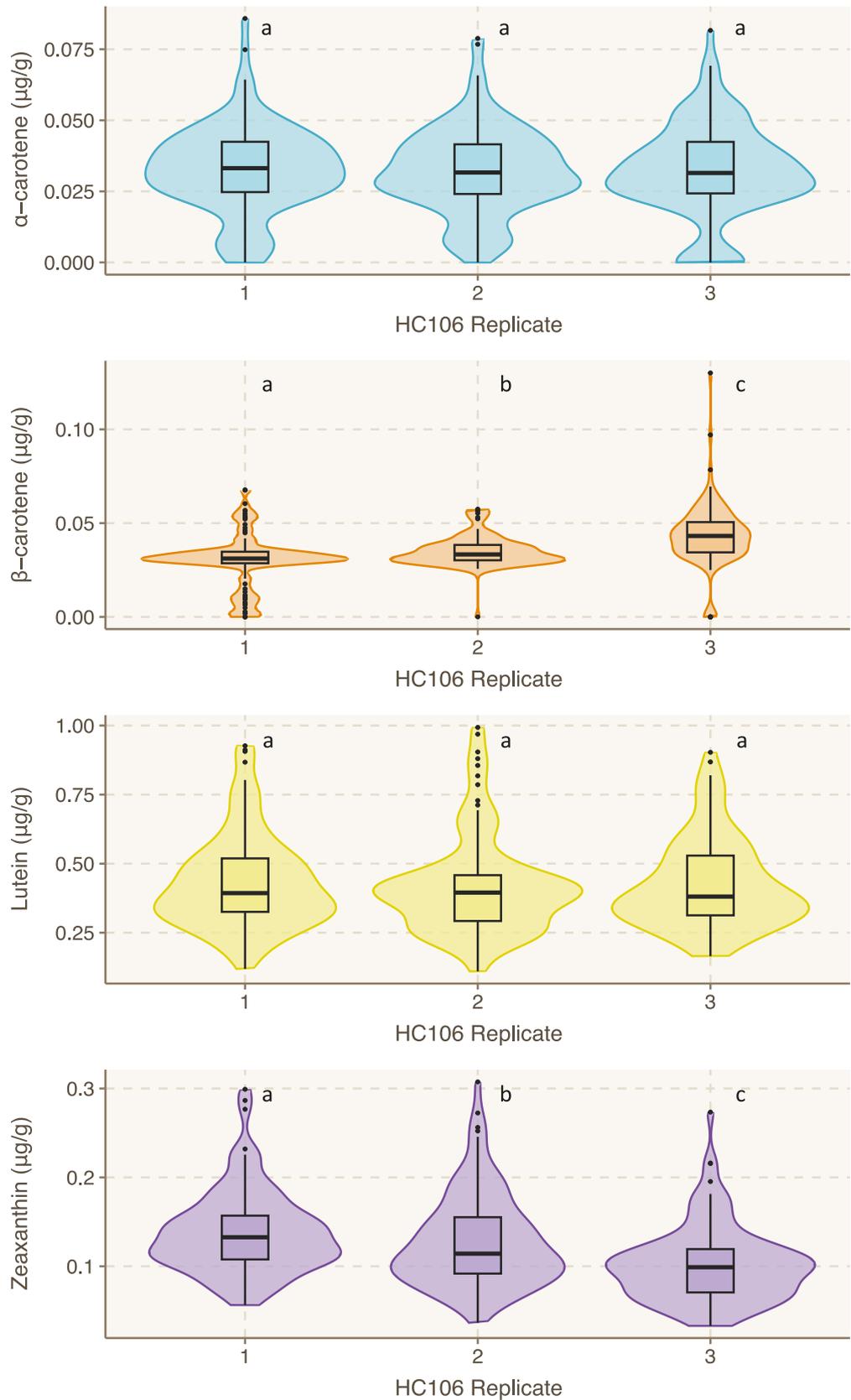
Appendix Material 8 Proportions of each carotenoid compound within the Watkins tetraploid collection. Minimum, median, and maximum accession numbers are shown in bold red text.

Appendix Material 9 Full results of the carotenoid content of the Watkins hexaploid core collection analysis. α -carotene, β -carotene, lutein, zeaxanthin and total grain carotenoid content are measured in $\mu\text{g/g}$. α -caro.= α -carotene, β -caro.= β -carotene, Zeaxanth.=zeaxanthin, Total GCC=total grain carotenoid content, Prop β - β .=proportion of β - β carotenoids.

Accession	α -caro.	β -caro.	Lutein	Zeaxanth.	Total GCC	Prop β - β .
WAT11900004	0.0272	0.0396	0.3150	0.0874	0.4692	0.2707
WAT11900007	0.0304	0.0211	0.3516	0.0788	0.4820	0.2073
WAT11900023	0.0408	0.0510	0.4199	0.1363	0.6480	0.2890
WAT11900032	0.0242	0.0340	0.3577	0.1050	0.5209	0.2668
WAT11900034	0.0218	0.0347	0.2237	0.0533	0.3334	0.2639
WAT11900040	0.0327	0.0391	0.4443	0.0887	0.6048	0.2113
WAT11900042	0.0353	0.0390	0.4007	0.1040	0.5791	0.2470
WAT11900044	0.0278	0.0339	0.3781	0.0822	0.5220	0.2224
WAT11900045	0.0287	0.0423	0.2405	0.0731	0.3845	0.2999
WAT11900079	0.0307	0.0414	0.2523	0.1034	0.4278	0.3385
WAT11900081	0.0384	0.0434	0.3321	0.2194	0.6333	0.4149
WAT11900103	0.0277	0.0341	0.3486	0.1161	0.5265	0.2852
WAT11900110	0.0052	0.0132	0.8661	0.2166	1.1010	0.2087
WAT11900127	0.0420	0.0368	0.5525	0.1441	0.7754	0.2333
WAT11900139	0.0247	0.0340	0.2478	0.0686	0.3750	0.2734
WAT11900141	0.0256	0.0310	0.2943	0.0694	0.4202	0.2388
WAT11900145	0.0341	0.0407	0.4041	0.1255	0.6044	0.2750
WAT11900149	0.0077	0.0344	0.9014	0.2699	1.2134	0.2508
WAT11900160	0.0287	0.0411	0.3764	0.0795	0.5258	0.2295
WAT11900181	0.0343	0.0355	0.4386	0.1104	0.6188	0.2358
WAT11900209	0.0364	0.0413	0.2930	0.0913	0.4620	0.2870
WAT11900216	0.0256	0.0419	0.3535	0.0866	0.5076	0.2532
WAT11900218	0.0261	0.0320	0.3291	0.1282	0.5154	0.3107
WAT11900219	0.0352	0.0266	0.5162	0.1248	0.7028	0.2155
WAT11900223	0.0295	0.0373	0.2983	0.1048	0.4699	0.3025
WAT11900231	0.0298	0.0379	0.3213	0.0940	0.4829	0.2730
WAT11900238	0.0481	0.0388	0.4097	0.0946	0.5911	0.2256
WAT11900239	0.0412	0.0341	0.4133	0.1266	0.6151	0.2611
WAT11900246	0.0029	0.0190	0.3390	0.1304	0.4912	0.3041
WAT11900254	0.0290	0.0500	0.4620	0.1444	0.6854	0.2836
WAT11900264	0.0333	0.0395	0.4082	0.1148	0.5958	0.2590
WAT11900273	0.0416	0.0360	0.4263	0.1217	0.6256	0.2521
WAT11900291	0.0266	0.0328	0.3179	0.0916	0.4690	0.2652
WAT11900292	0.0342	0.0308	0.4029	0.1125	0.5805	0.2469
WAT11900299	0.0345	0.0670	0.3874	0.1009	0.5899	0.2847
WAT11900300	0.0814	0.0467	0.6719	0.1693	0.9693	0.2228
WAT11900305	0.0374	0.0296	0.4346	0.1117	0.6133	0.2305
WAT11900308	0.0033	0.0253	0.4001	0.1589	0.5876	0.3135
WAT11900313	0.0017	0.0255	0.7551	0.1819	0.9641	0.2151
WAT11900324	0.0297	0.0354	0.2685	0.1292	0.4629	0.3556
WAT11900325	0.0349	0.0359	0.2779	0.0929	0.4416	0.2916

WAT11900349	0.0385	0.0385	0.4837	0.0818	0.6425	0.1872
WAT11900352	0.0496	0.0565	0.5797	0.1375	0.8232	0.2356
WAT11900355	0.0240	0.0303	0.3275	0.0656	0.4474	0.2143
WAT11900360	0.0309	0.0398	0.3935	0.0943	0.5584	0.2401
WAT11900387	0.0633	0.0459	0.5952	0.1341	0.8385	0.2146
WAT11900396	0.0490	0.0372	0.5231	0.1723	0.7815	0.2680
WAT11900397	0.0176	0.0307	0.5523	0.1461	0.7467	0.2368
WAT11900398	0.0295	0.0353	0.2708	0.1051	0.4408	0.3187
WAT11900406	0.0047	0.0342	0.6924	0.2040	0.9353	0.2547
WAT11900420	0.0256	0.0330	0.3732	0.1062	0.5381	0.2588
WAT11900433	0.0222	0.0349	0.2895	0.0859	0.4326	0.2793
WAT11900440	0.0515	0.0316	0.5688	0.1517	0.8037	0.2282
WAT11900444	0.0571	0.0353	0.6642	0.1494	0.9059	0.2038
WAT11900451	0.0424	0.0451	0.4196	0.1564	0.6635	0.3037
WAT11900471	0.0288	0.0382	0.2270	0.1247	0.4187	0.3892
WAT11900474	0.0219	0.0297	0.2709	0.0950	0.4175	0.2988
WAT11900481	0.0345	0.0386	0.3627	0.1243	0.5600	0.2908
WAT11900483	0.0743	0.0304	0.7940	0.1371	1.0358	0.1617
WAT11900496	0.0333	0.0378	0.3811	0.1010	0.5531	0.2508
WAT11900507	0.0037	0.0102	0.5489	0.1500	0.7128	0.2246
WAT11900546	0.0414	0.0404	0.4446	0.1026	0.6290	0.2273
WAT11900551	0.0333	0.0329	0.5060	0.1168	0.6889	0.2173
WAT11900560	0.0058	0.0376	0.8883	0.2190	1.1507	0.2230
WAT11900562	0.0292	0.0250	0.2665	0.1079	0.4287	0.3102
WAT11900566	0.0083	0.0432	0.5505	0.1578	0.7598	0.2645
WAT11900568	0.0234	0.0323	0.2001	0.0776	0.3335	0.3298
WAT11900579	0.0228	0.0273	0.2383	0.0780	0.3663	0.2873
WAT11900580	0.0045	0.0217	0.6325	0.2434	0.9022	0.2939
WAT11900591	0.0405	0.0261	0.3735	0.1579	0.5980	0.3078
WAT11900605	0.0048	0.0376	0.8337	0.2084	1.0846	0.2269
WAT11900624	0.0441	0.0392	0.5118	0.1319	0.7270	0.2354
WAT11900627	0.0255	0.0372	0.3190	0.0707	0.4524	0.2385
WAT11900629	0.0205	0.0265	0.2513	0.0581	0.3564	0.2373
WAT11900637	0.0373	0.0419	0.4414	0.1160	0.6365	0.2480
WAT11900639	0.0453	0.0334	0.4932	0.1188	0.6907	0.2203
WAT11900651	0.0358	0.0384	0.3778	0.1110	0.5630	0.2653
WAT11900652	0.0402	0.0522	0.3723	0.1084	0.5731	0.2802
WAT11900662	0.0532	0.0366	0.4548	0.1272	0.6718	0.2438
WAT11900670	0.0416	0.0477	0.4436	0.0794	0.6123	0.2077
WAT11900671	0.0340	0.0356	0.3159	0.1696	0.5552	0.3697
WAT11900680	0.0210	0.0551	0.2339	0.1169	0.4270	0.4029
WAT11900683	0.0250	0.0321	0.3149	0.1094	0.4814	0.2940
WAT11900685	0.0455	0.0394	0.4913	0.1402	0.7164	0.2507
WAT11900694	0.0256	0.0359	0.1942	0.0615	0.3171	0.3069
WAT11900698	0.0366	0.0371	0.4170	0.1275	0.6182	0.2663
WAT11900704	0.0347	0.0387	0.4846	0.1174	0.6755	0.2311
WAT11900707	0.0519	0.0583	0.4477	0.1227	0.6806	0.2659

WAT11900722	0.0462	0.0440	0.3613	0.1511	0.6026	0.3237
WAT11900729	0.0272	0.0413	0.2553	0.0770	0.4008	0.2952
WAT11900731	0.0298	0.0346	0.2098	0.0929	0.3671	0.3474
WAT11900732	0.0000	0.0000	0.8244	0.1963	1.0206	0.1923
WAT11900740	0.0341	0.0352	0.3610	0.1107	0.5409	0.2696
WAT11900742	0.0269	0.0235	0.3562	0.1135	0.5201	0.2635
WAT11900749	0.0241	0.0366	0.2492	0.0943	0.4041	0.3238
WAT11900750	0.0285	0.0406	0.3447	0.1240	0.5377	0.3060
WAT11900753	0.0290	0.0383	0.3134	0.0999	0.4806	0.2875
WAT11900771	0.0514	0.0314	0.5174	0.1479	0.7481	0.2397
WAT11900777	0.0612	0.0373	0.6705	0.1638	0.9328	0.2155
WAT11900784	0.0492	0.0423	0.7181	0.1614	0.9710	0.2098
WAT11900788	0.0425	0.0542	0.4568	0.1635	0.7169	0.3036
WAT11900811	0.0454	0.0435	0.3570	0.0986	0.5444	0.2609
WAT11900814	0.0427	0.0409	0.4755	0.1045	0.6636	0.2191
WAT11900816	0.0450	0.0325	0.4849	0.1456	0.7080	0.2515
WAT11900827	0.0336	0.0366	0.3472	0.1232	0.5405	0.2956
WAT11900912	0.0277	0.0255	0.3619	0.1098	0.5248	0.2577



Appendix Material 10 Comparison of α -carotene, β -carotene, lutein, and zeaxanthin content across three replicates from the 106-accession Watkins hexaploid core collection. The letters above the violin plots indicate statistical significance between the groups, as determined by a two-way ANOVA. The pairwise comparisons for the ANOVAs are found in Appendix Material 11.

Appendix Material 11 Pairwise comparisons of grain carotenoid content of the three replicates from the 106-accession Watkins hexaploid core collection. The tables show the differences of pairwise comparisons between the replicates based on a two-way ANOVA with a Tukey post-hoc test. Significant comparisons ($p < 0.05$) are highlighted in grey. SE=standard error, df=degrees of freedom, 95% CI=95% confidence interval.

a) Total grain carotenoid content

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
Rep1-Rep2	0.0187	0.0141	210	1.328	0.381	-0.01451	0.0519
Rep1-Rep3	0.0305	0.0141	210	2.171	0.0787	-0.00267	0.0637
Rep2-Rep3	0.0118	0.0141	210	0.842	0.6773	-0.02135	0.045

b) α -carotene content

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
Rep1-Rep2	0.0004	0.000974	210	0.411	0.9112	-0.0019	0.0027
Rep1-Rep3	0.000879	0.000974	210	0.902	0.6394	-0.00142	0.00318
Rep2-Rep3	0.000479	0.000974	210	0.492	0.8754	-0.00182	0.00278

c) β -carotene content

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
Rep1-Rep2	-0.00475	0.00173	210	-2.747	0.0179	-0.00883	-0.000669
Rep1-Rep3	-0.01277	0.00173	210	-7.383	<.0001	-0.01685	-0.008688
Rep2-Rep3	-0.00802	0.00173	210	-4.636	<.0001	-0.0121	-0.003936

d) Lutein content

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
Rep1-Rep2	0.01353	0.0102	210	1.333	0.3787	-0.0104	0.0375
Rep1-Rep3	0.00645	0.0102	210	0.635	0.8008	-0.0175	0.0304
Rep2-Rep3	-0.00708	0.0102	210	-0.697	0.7654	-0.0311	0.0169

e) Zeaxanthin content

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
Rep1-Rep2	0.0095	0.00401	210	2.366	0.0493	0.0000239	0.019
Rep1-Rep3	0.036	0.00401	210	8.959	<.0001	0.0265	0.0454
Rep2-Rep3	0.0265	0.00401	210	6.592	<.0001	0.017	0.0359

Appendix Material 12 Associated 35K Breeder's Array markers with carotenoid content in the Watkins tetraploid collection. Pos.=IWGSC RefSeq v1.0 genomic position.

Carotenoid	Marker ID	Marker No.	Model	Chr.	Pos.	P-value	MAF
YPC	AX-95244216	3249	BLINK	chr1B	667089514	3.26E-09	0.433
Total GCC	AX-94738209	7065	MLM	chr3A	10303016	1.34E-06	0.006
Total GCC	AX-95252995	7249	FarmCPU	chr3A	47496444	2.04E-07	0.009
Total GCC	AX-94822992	8171	BLINK	chr3A	702491594	1.90E-06	0.003
Total GCC	AX-95216226	12553	BLINK	chr5A	375376060	5.31E-14	0.002
Total GCC	AX-95216226	12553	FarmCPU	chr5A	375376060	9.84E-15	0.002
Total GCC	AX-95216226	12553	MLM	chr5A	375376060	4.39E-09	0.002
Total GCC	AX-94847937	12633	MLM	chr5A	427317500	4.39E-09	0.002
Total GCC	AX-94435311	17130	MLM	chr6B	355423938	4.39E-09	0.002
Total GCC	AX-94584328	17264	BLINK	chr6B	472609840	1.63E-06	0.224
Total GCC	AX-94498408	17580	BLINK	chr6B	643567733	1.48E-10	0.081
Total GCC	AX-94498408	17580	FarmCPU	chr6B	643567733	2.03E-08	0.081
Total GCC	AX-94502860	19362	MLM	chr7A	709641487	2.15E-07	0.002
α-carot.	AX-94668650	577	FarmCPU	chr1A	218217702	1.39E-09	0.005
α-carot.	AX-94942875	1383	BLINK	chr1A	592089005	2.25E-11	0.459
α-carot.	AX-94942875	1383	MLM	chr1A	592089005	4.03E-08	0.459
α-carot.	AX-94659198	2430	BLINK	chr1B	399335719	1.97E-21	0.029
α-carot.	AX-94659198	2430	FarmCPU	chr1B	399335719	6.69E-07	0.029
α-carot.	AX-94879161	3054	BLINK	chr1B	614793764	1.93E-09	0.003
α-carot.	AX-94879161	3054	FarmCPU	chr1B	614793764	1.49E-08	0.003
α-carot.	AX-94687416	4443	BLINK	chr2A	692633149	3.05E-07	0.021
α-carot.	AX-94393838	4481	FarmCPU	chr2A	701866893	3.43E-07	0.099
α-carot.	AX-94666545	5301	FarmCPU	chr2B	39205391	1.38E-08	0.008
α-carot.	AX-94889597	6254	BLINK	chr2B	632917375	4.89E-22	0.002
α-carot.	AX-94457076	6626	BLINK	chr2B	731895094	1.76E-13	0.005
α-carot.	AX-94457076	6626	MLM	chr2B	731895094	1.45E-06	0.005
α-carot.	AX-94866669	7014	BLINK	chr3A	7125368	2.33E-12	0.002
α-carot.	AX-95166693	8577	FarmCPU	chr3B	31789286	2.00E-10	0.002
α-carot.	AX-95223462	8657	BLINK	chr3B	64696881	2.63E-08	0.038
α-carot.	AX-94431987	9017	BLINK	chr3B	344976572	9.93E-14	0.002
α-carot.	AX-94431987	9017	FarmCPU	chr3B	344976572	9.44E-10	0.002
α-carot.	AX-94991915	9889	BLINK	chr3B	808874202	2.76E-07	0.005
α-carot.	AX-94664966	13625	BLINK	chr5B	9256996	5.73E-10	0.009
α-carot.	AX-94825873	14828	BLINK	chr5B	597813593	3.19E-34	0.003
α-carot.	AX-94825873	14828	FarmCPU	chr5B	597813593	8.62E-07	0.003
α-carot.	AX-94825873	14828	MLM	chr5B	597813593	2.14E-09	0.003
α-carot.	AX-94854303	18292	BLINK	chr7A	58993570	1.80E-07	0.372
α-carot.	AX-94424575	18522	BLINK	chr7A	138011393	2.36E-07	0.009
α-carot.	AX-95098936	18527	FarmCPU	chr7A	139098696	2.76E-07	0.005
α-carot.	AX-94814408	18747	BLINK	chr7A	287639086	3.60E-17	0.003
α-carot.	AX-94814408	18747	FarmCPU	chr7A	287639086	1.39E-11	0.003
α-carot.	AX-95243698	19765	BLINK	chr7B	103115932	7.82E-09	0.012
α-carot.	AX-94567508	20910	FarmCPU	chr7B	721819562	2.99E-11	0.005

Continued from the previous page:

Carotenoid	Marker ID	Marker No.	Model	Chr.	Pos.	P-value	MAF
β-carot.	AX-95021850	4210	BLINK	chr2A	536421906	1.92E-07	0.006
β-carot.	AX-95174558	7259	BLINK	chr3A	49482119	2.45E-08	0.011
β-carot.	AX-95174558	7259	MLM	chr3A	49482119	1.80E-06	0.011
β-carot.	AX-94984534	19312	BLINK	chr7A	698687787	2.02E-08	0.267
Lutein	AX-94691247	878	BLINK	chr1A	432198658	4.43E-07	0.114
Lutein	AX-94592974	2562	FarmCPU	chr1B	448742632	2.29E-08	0.110
Lutein	AX-94747151	3795	BLINK	chr2A	87884303	9.94E-11	0.327
Lutein	AX-94747151	3795	FarmCPU	chr2A	87884303	4.54E-07	0.327
Lutein	AX-94738209	7065	MLM	chr3A	10303016	2.37E-07	0.006
Lutein	AX-94769500	8124	BLINK	chr3A	693176804	3.72E-10	0.264
Lutein	AX-94769500	8124	FarmCPU	chr3A	693176804	2.98E-08	0.264
Lutein	AX-94659008	9916	FarmCPU	chr3B	813903114	1.36E-06	0.182
Lutein	AX-95216226	12553	BLINK	chr5A	375376060	2.09E-17	0.002
Lutein	AX-95216226	12553	FarmCPU	chr5A	375376060	3.02E-15	0.002
Lutein	AX-95216226	12553	MLM	chr5A	375376060	7.63E-10	0.002
Lutein	AX-94847937	12633	MLM	chr5A	427317500	7.63E-10	0.002
Lutein	AX-94538863	15314	BLINK	chr6A	6735017	1.34E-07	0.258
Lutein	AX-94538863	15314	FarmCPU	chr6A	6735017	4.70E-07	0.258
Lutein	AX-94883234	16632	BLINK	chr6B	52172895	2.30E-12	0.033
Lutein	AX-94883234	16632	FarmCPU	chr6B	52172895	1.22E-06	0.033
Lutein	AX-94883234	16632	MLM	chr6B	52172895	5.02E-07	0.033
Lutein	AX-94435311	17130	MLM	chr6B	355423938	7.63E-10	0.002
Lutein	AX-95082017	18302	FarmCPU	chr7A	62838984	1.22E-08	0.488
Lutein	AX-94502860	19362	MLM	chr7A	709641487	4.15E-08	0.002
Zeaxan.	AX-94705969	8471	BLINK	chr3B	7189529	1.31E-07	0.213
Zeaxan.	AX-94680240	16099	BLINK	chr6A	585428558	6.77E-07	0.495
Zeaxan.	AX-94680240	16099	FarmCPU	chr6A	585428558	2.10E-07	0.495
Zeaxan.	AX-95110918	17913	BLINK	chr6B	717862009	3.14E-08	0.245
Zeaxan.	AX-95110544	19725	BLINK	chr7B	79942170	2.13E-06	0.003

Appendix Material 13 *T. turgidum* genes identified as orthologous to carotenoid biosynthesis genes in rice and *Arabidopsis*. Genes are named with their Svevo v1 IDs. Genes are hyperlinked to their EnsemblPlants gene summary. Chr.=chromosome, orth.=orthologue

Gene name	Chr.	Svevo v1	Genomic Location	Orthologue (rice or <i>Arabidopsis</i>)
<i>TdOR</i>	6A	TRITD6Av1G155000	449939186–449941730	Os02g0651300 (OsOR)
<i>TdOR</i>	6B	TRITD6Bv1G140710	460492496–460495030	Os02g0651300 (OsOR)
<i>TdORLIKE</i>	6A	TRITD6Av1G108670	295968051–295970322	Os02g0535000 (OsORLIKE)
<i>TdORLIKE</i>	6B	TRITD6Bv1G093390	287374418–287376548	Os02g0535000 (OsORLIKE)
<i>TdPSY1</i>	0U	TRITD0Uv1G062430	153508012–153510987	Os06g0729000 (OsPSY1)
<i>TdPSY1</i>	6B	TRITD6Bv1G228570	696979833–696983031	Os06g0729000 (OsPSY1)
<i>TdPSY2</i>	5A	TRITD5Av1G008310	17616395–17617884	Os12g0626400 (OsPSY2)
<i>TdPSY2</i>	5B	TRITD5Bv1G006680	17026598–17106302	Os12g0626400 (OsPSY2)
<i>TdPSY3</i>	5A	TRITD5Av1G193290	521568398–521569956	Os09g0555500 (OsPSY3)
<i>TdPSY3</i>	5B	TRITD5Bv1G185680	535321177–535322686	Os09g0555500 (OsPSY3)
<i>TdPDS</i>	4A	TRITD4Av1G000480	982975–987347	Os03g0184000 (OsPDS)
<i>TdPDS</i>	4B	TRITD4Bv1G174000	590725185–590729523	Os03g0184000 (OsPDS)
<i>TdZISO</i>	5A	TRITD5Av1G014480	31756668–31760416	Os12g0405200 (OsZISO)
<i>TdZISO</i>	5B	TRITD5Bv1G013720	37226691–37227542	Os12g0405200 (OsZISO)
<i>TdZDS</i>	2A	TRITD2Av1G117040	318754381–318760919	Os07g0204900 (OsZDS)
<i>TdZDS</i>	2B	TRITD2Bv1G099850	273582665–273589363	Os07g0204900 (OsZDS)
<i>TdCRTISO</i>	1A	TRITD1Av1G024310	52035230–52039912	Os11g0572700 (OsCRTISO)
<i>TdCRTISO</i>	1B	TRITD1Bv1G031990	85697655–85702444	Os11g0572700 (OsCRTISO)
<i>TdLCYB</i>	6A	TRITD6Av1G072890	179700824–179702293	Os02g0190600 (OsLCYB)
<i>TdLCYB</i>	6B	TRITD6Bv1G077610	235396899–235398359	Os02g0190600 (OsLCYB)
<i>TdLCYE</i>	3A	TRITD3Av1G132870	374071991–374078978	Os01g0581300 (OsLCYE)
<i>TdLCYE</i>	3B	TRITD3Bv1G121980	381299870–381302109	Os01g0581300 (OsLCYE)
<i>TdLCYE</i>	3B	TRITD3Bv1G121950	381295285–381299057	Os01g0581300 (OsLCYE)
<i>AtLCY orth.</i>	3A	TRITD3Av1G269690	712962126–712964081	AT2G32640 (LCYE/B)
<i>TdHYD1</i>	2A	TRITD2Av1G234270	644481998–644483538	Os04g0578400 (OsBCH2)
<i>TdHYD1</i>	2B	TRITD2Bv1G196580	582998858–583000343	Os04g0578400 (OsBCH2)
<i>TdHYD2</i>	5A	TRITD5Av1G238340	621330186–621331610	Os10g0533500 (OsBCH3)
<i>TdHYD2</i>	4B	TRITD4Bv1G181720	612337034–612338569	Os10g0533500 (OsBCH3)
<i>OsCYP97A4 orth.</i>	6A	TRITD6Av1G219960	601483422–601488171	Os02g0817900 (OsCYP97A4)
<i>OsCYP97A4 orth.</i>	6B	TRITD6Bv1G220480	677873268–677878411	Os02g0817900 (OsCYP97A4)
<i>OsCYP97A4 orth.</i>	6B	TRITD6Bv1G221900	681512321–681517174	Os02g0817900 (OsCYP97A4)
<i>OsCYP97B4 orth.</i>	6A	TRITD6Av1G059950	143276949–143281016	Os02g0173100 (OsCYP97B4)
<i>OsCYP97B4 orth.</i>	6B	TRITD6Bv1G069020	201634597–201639587	Os02g0173100 (OsCYP97B4)
<i>OsCYP97C2 orth.</i>	1A	TRITD1Av1G134470	362510882–362515806	Os10g0546600 (OsCYP97C2)
<i>OsCYP97C2 orth.</i>	1B	TRITD1Bv1G128420	390991753–390996815	Os10g0546600 (OsCYP97C2)
<i>TdZEP</i>	2A	TRITD2Av1G193300	536414458–536420090	Os04g0448900 (OsABA1/OsZEP)
<i>TdZEP</i>	2B	TRITD2Bv1G158920	470884435–470890420	Os04g0448900 (OsABA1/OsZEP)
<i>TdVDE</i>	2A	TRITD2Av1G167010	463148531–463150378	Os04g0379700 (OsVDE)
<i>TdVDE</i>	2B	TRITD2Bv1G139140	409999946–410001789	Os04g0379700 (OsVDE)

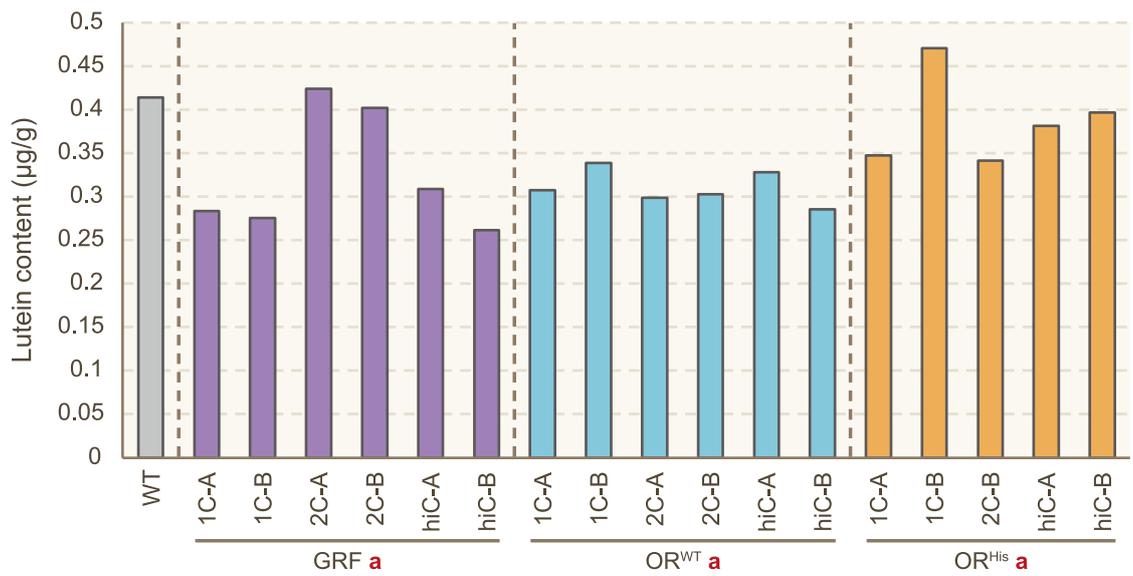
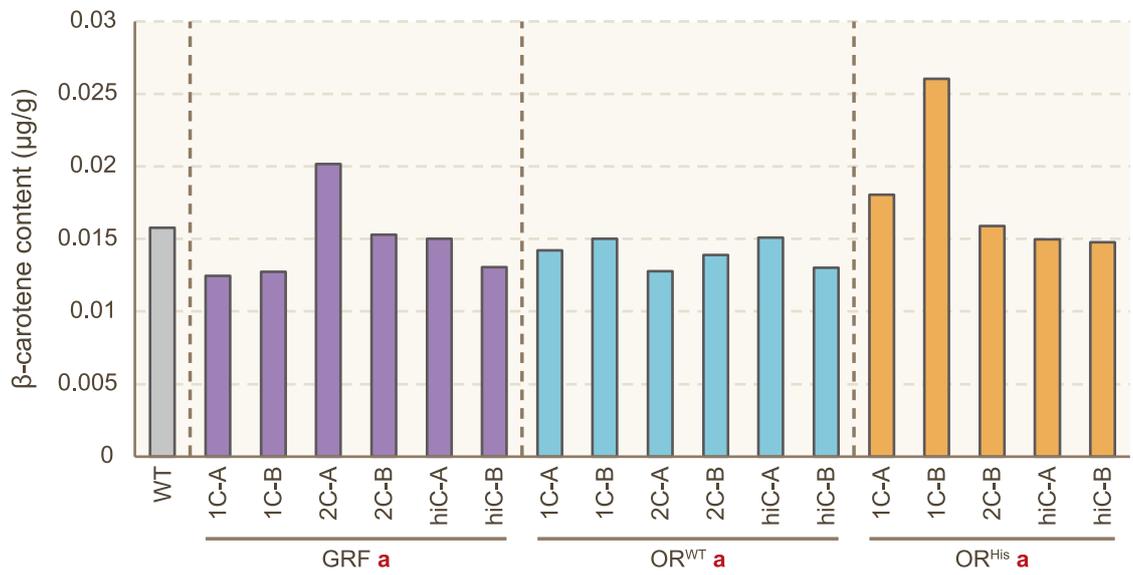
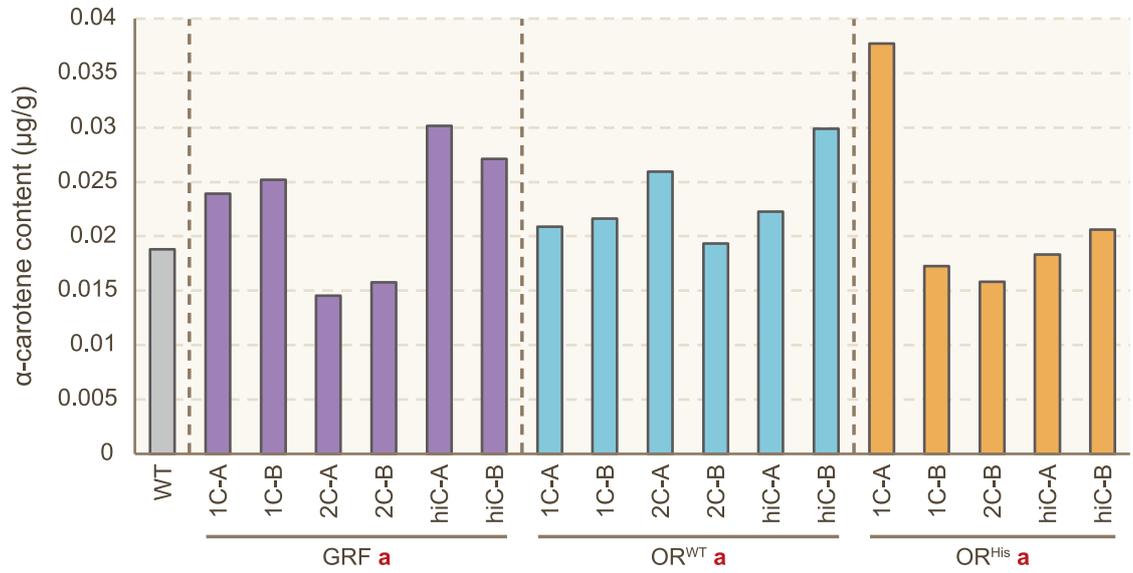
Continued from the previous page:

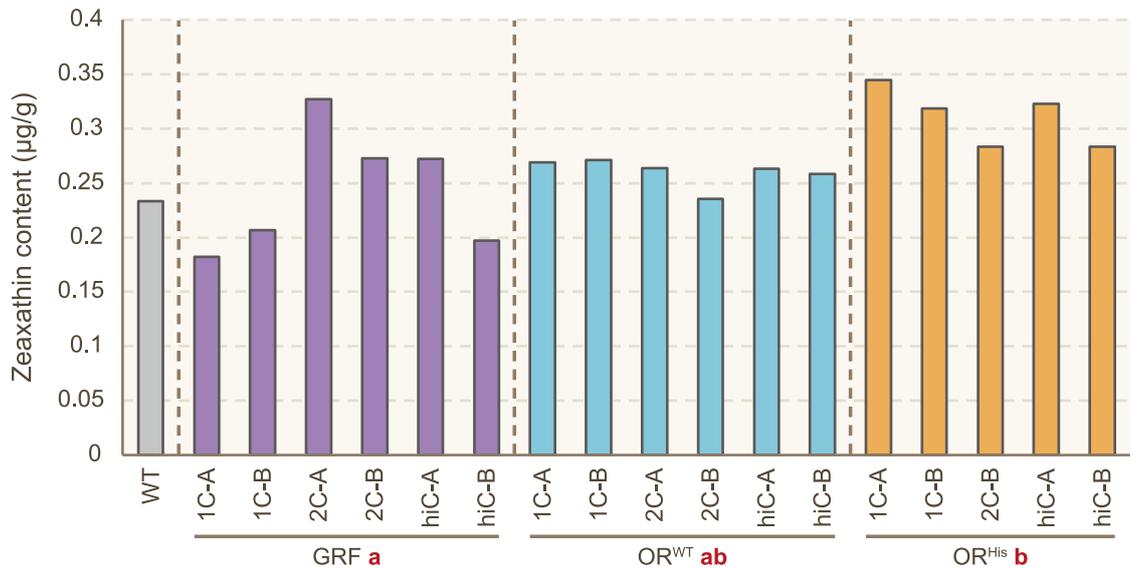
Gene name	Chr.	Svevo v1	Location	Orthologue (rice or <i>Arabidopsis</i>)
<i>TdCCD1</i>	5A	TRITD5Av1G000820	1541224–1548567	Os12g0640600 (OsCCD1)
<i>TdCCD1</i>	5B	TRITD5Bv1G000390	1242083–1247568	Os12g0640600 (OsCCD1)
<i>TdCCD4</i>	6A	TRITD6Av1G171740	495470962–495472881	Os02g0704000 (OsCCD4/OsNCED1)
<i>TdCCD4</i>	6B	TRITD6Bv1G159060	513857345–513859276	Os02g0704000 (OsCCD4/OsNCED1)
<i>OsCCD1 orth.</i>	5A	TRITD5Av1G000470	739559–743551	Os12g0640600 (OsCCD1)
<i>OsCCD1 orth.</i>	5B	TRITD5Bv1G000430	1773355–1777324	Os12g0640600 (OsCCD1)
<i>OsCCD1 orth.</i>	5A	TRITD5Av1G000560	891362–895223	Os12g0640600 (OsCCD1)
<i>OsCCD7 orth.</i>	2A	TRITD2Av1G243980	666451682–666454218	Os04g0550600 (OsCCD7/HTD1)
<i>OsCCD7 orth.</i>	2B	TRITD2Bv1G204570	611277535–611280084	Os04g0550600 (OsCCD7/HTD1)
<i>OsCCD8a orth.</i>	6B	TRITD6Bv1G003130	9466430–9467822	Os01g0566500 (OsCCD8a)
<i>OsCCD8b orth.</i>	3A	TRITD3Av1G020620	42285809–42286516	Os01g0746400 (OsCCD8b/D10)
<i>OsCCD8b orth.</i>	3B	TRITD3Bv1G024020	62996637–63000302	Os01g0746400 (OsCCD8b/D10)
<i>OsCCD8b orth.</i>	3A	TRITD3Av1G181880	506289587–506292803	Os01g0746400 (OsCCD8b/D10)
<i>OsCCD8b orth.</i>	3B	TRITD3Bv1G162510	499423853–499425692	Os01g0746400 (OsCCD8b/D10)
<i>OsCCD8d orth.</i>	4A	TRITD4Av1G000580	1126289–1136194	Os08g0369800 (OsCCD8d)
<i>OsCCD8d orth.</i>	4B	TRITD4Bv1G174150	591172373–591181604	Os08g0369800 (OsCCD8d)
<i>OsCCD-like</i>	5B	TRITD5Bv1G111550	327609086–327613069	Os09g0321200 (OsCCD-like)
<i>OsCCD-like</i>	5A	TRITD5Av1G132540	375355951–375361220	Os09g0321200 (OsCCD-like)
<i>OsNCED1 orth.</i>	5A	TRITD5Av1G198650	533847699–533848721	Os03g0645900 (OsNCED1)
<i>OsNCED1 orth.</i>	5B	TRITD5Bv1G192230	550826581–550827603	Os03g0645900 (OsNCED1)
<i>OsNCED2 orth.</i>	5B	TRITD5Bv1G011050	29602428–29604272	Os12g0617400 (OsNCED2)
<i>OsNCED3 orth.</i>	2A	TRITD2Av1G136080	373936882–373938663	Os07g0154100 (OsNCED3)
<i>OsNCED3 orth.</i>	2B	TRITD2Bv1G125530	370723071–370725675	Os07g0154100 (OsNCED3)

Appendix 3 – Appendix Materials of Chapter 4



Appendix Material 14 Deformities in Kronos spike architecture leading to sterility.





Appendix Material 15 Bar charts of individual carotenoid compounds of the T₁ *TaOR*^{WT} and *TaOR*^{His} overexpression lines compared to the GRF and wild-type controls. Bold red letters next to the groups' names indicate statistical significance between the groups as determined by a one-way ANOVA. WT=non-transgenic wild-type controls; GRF=*GRF4-GIF1* transgenic controls; OR^{WT}=pAct-OR^{WT} transgenic plants; OR^{His}=pAct-OR^{His} transgenic plants.

Appendix Material 16 Pairwise comparisons of grain carotenoid content of T₁ *TaOR*^{WT} and *TaOR*^{His} overexpression lines compared to GRF controls. The tables show the differences of pairwise comparisons between the lines based on a one-way ANOVA with a Tukey post-hoc test. Significant comparisons (p<0.05) are highlighted in grey. SE=standard error, df=degrees of freedom, 95% CI=95% confidence interval, GRF=*GRF4-GIF1* transgenic controls; OR^{WT}=pAct-OR^{WT} transgenic plants; OR^{His}=pAct-OR^{His} transgenic plants.

a) Total grain carotenoid content

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
GRF-OR ^{His}	-0.13155	0.0491	14	-2.678	0.0447	-0.26013	-0.00296
GRF-OR ^{WT}	-0.00118	0.0468	14	-0.025	0.9996	-0.12378	0.12142
OR ^{His} -OR ^{WT}	0.13037	0.0491	14	2.654	0.0467	0.00178	0.25895

b) α-carotene content

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
GRF-OR ^{His}	0.000835	0.00395	14	0.211	0.9757	-0.00951	0.01118
GRF-OR ^{WT}	-0.00054	0.00377	14	-0.143	0.9888	-0.0104	0.00932
OR ^{His} -OR ^{WT}	-0.001375	0.00395	14	-0.348	0.9357	-0.01172	0.00897

c) β-carotene content

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
GRF-OR ^{His}	-0.003158	0.00188	14	-1.677	0.2482	-0.008086	0.00177
GRF-OR ^{WT}	0.000774	0.0018	14	0.431	0.9034	-0.003925	0.00547
OR ^{His} -OR ^{WT}	0.003931	0.00188	14	2.088	0.1284	-0.000997	0.00886

d) Lutein content

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
GRF-OR ^{His}	-0.0617	0.0311	14	-1.987	0.152	-0.14297	0.0196
GRF-OR ^{WT}	0.0158	0.0296	14	0.532	0.857	-0.06173	0.0932
OR ^{His} -OR ^{WT}	0.0775	0.0311	14	2.494	0.0627	-0.00382	0.1587

e) Zeaxanthin content

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
GRF-OR ^{His}	-0.0675	0.0227	14	-2.981	0.0252	-0.12683	-0.00823
GRF-OR ^{WT}	-0.0172	0.0216	14	-0.795	0.7122	-0.07371	0.03937
OR ^{His} -OR ^{WT}	0.0504	0.0227	14	2.223	0.1018	-0.00894	0.10966

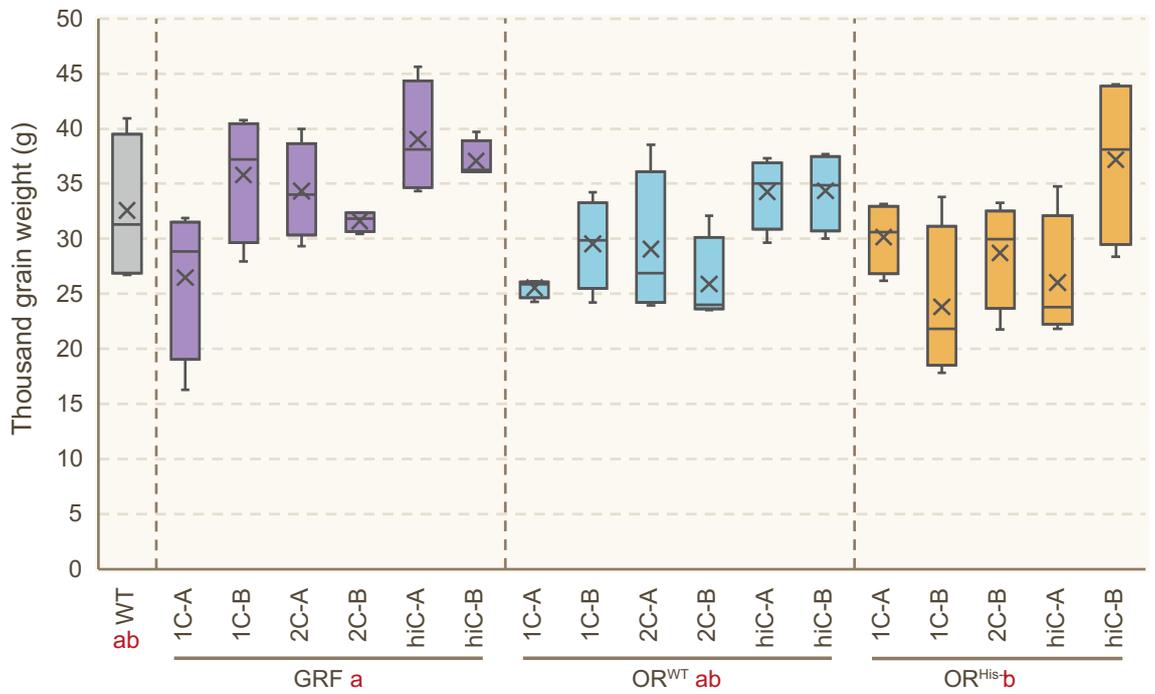
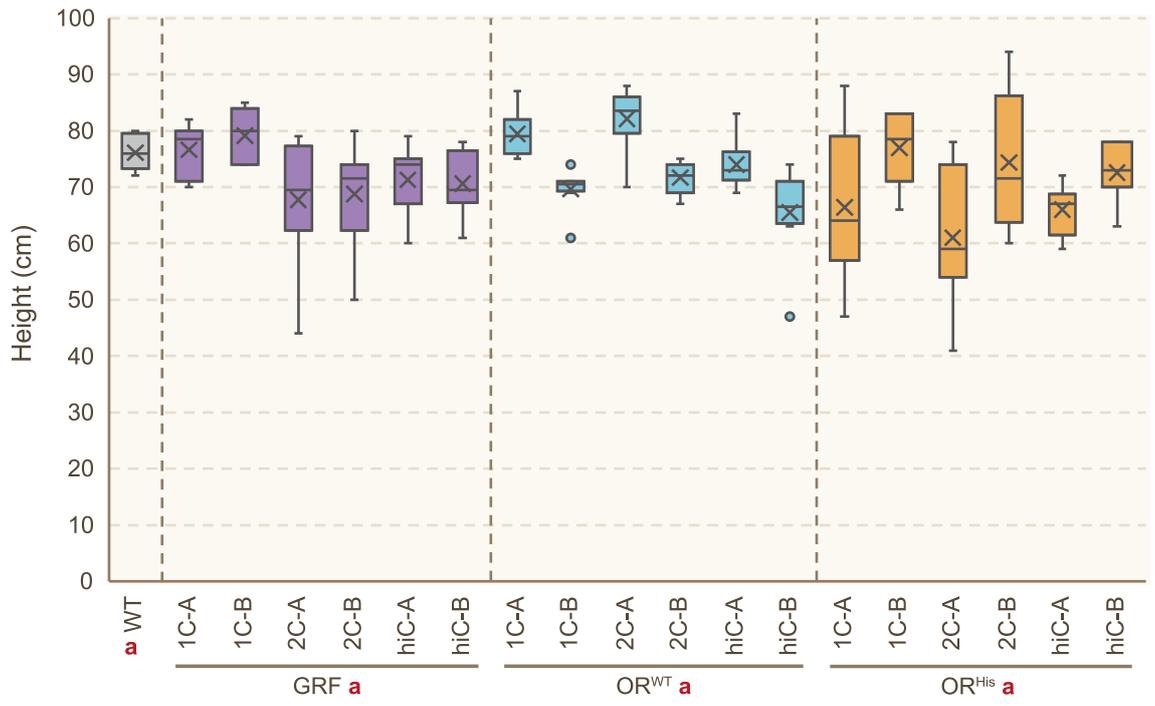
Appendix Material 17 Pairwise comparisons of grain yield per plant and anthesis date of T₁ *TaOR*^{WT} and *TaOR*^{His} overexpression lines compared to GRF controls. The tables show the differences of pairwise comparisons between the lines based on a one-way ANOVA with a Tukey post-hoc test. Significant comparisons (p<0.05) are highlighted in grey. SE=standard error, df=degrees of freedom, 95% CI=95% confidence interval, GRF=*GRF4-GIF1* transgenic controls; OR^{WT}=pAct-OR^{WT} transgenic plants; OR^{His}=pAct-OR^{His} transgenic plants.

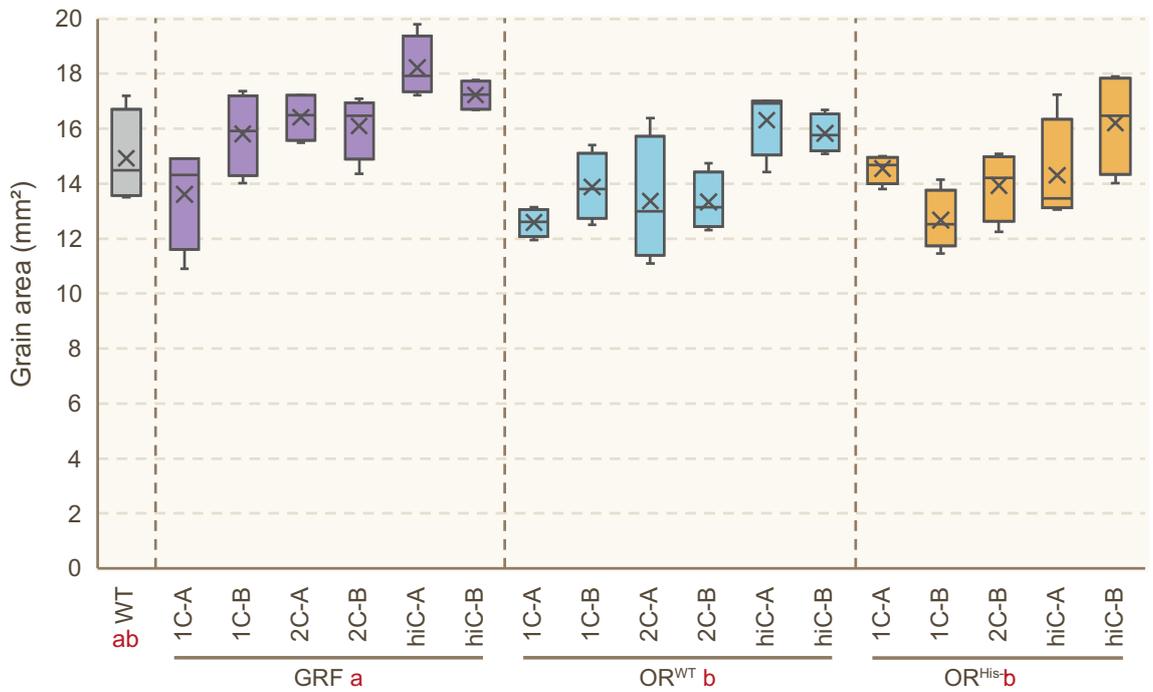
a) Grain yield per plant

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
GRF-OR ^{His}	-0.338	0.489	147	-0.693	0.8997	-1.61	0.931
GRF-OR ^{WT}	-1.919	0.489	147	-3.929	0.0007	-3.19	-0.65
GRF-WT	-5.034	0.91	147	-5.529	<.0001	-7.4	-2.668
OR ^{His} -OR ^{WT}	-1.581	0.486	147	-3.253	0.0077	-2.84	-0.318
OR ^{His} -WT	-4.695	0.909	147	-5.165	<.0001	-7.06	-2.333
OR ^{WT} -WT	-3.115	0.909	147	-3.426	0.0044	-5.48	-0.752

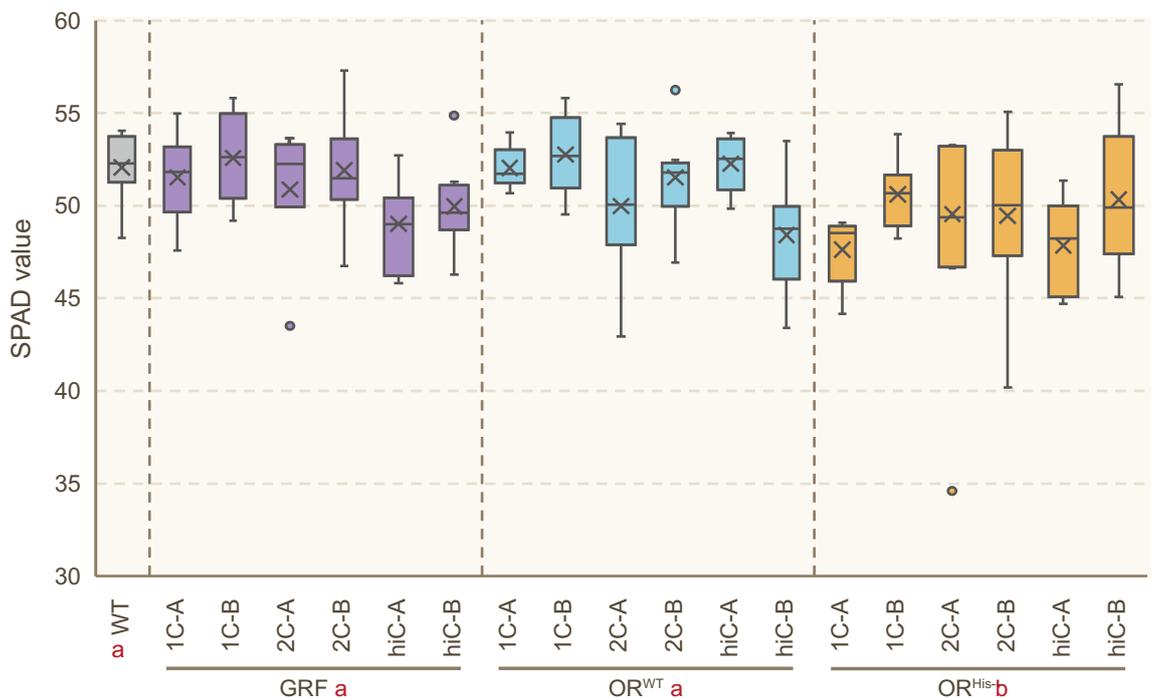
b) Anthesis date

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
GRF-OR ^{His}	-7.99	1.61	147	-4.954	<.0001	-12.18	-3.8
GRF-OR ^{WT}	-1.68	1.61	147	-1.041	0.7259	-5.87	2.51
GRF-WT	2.18	3.01	147	0.724	0.8875	-5.64	9.99
OR ^{His} -OR ^{WT}	6.31	1.6	147	3.934	0.0007	2.14	10.48
OR ^{His} -WT	10.17	3	147	3.387	0.0050	2.37	17.97
OR ^{WT} -WT	3.85	3	147	1.284	0.5746	-3.95	11.65





Appendix Material 18 Thousand-grain weight and grain area of T₁ *TaOR*^{WT} and *TaOR*^{His} overexpression lines compared to GRF and wild-type controls. Measurements were done on four plants for each of the lines. Bold red letters next to the groups' names indicate statistical significance between the groups as determined by a one-way ANOVA. WT=non-transgenic wild-type controls; GRF=*GRF4-GIF1* transgenic controls; OR^{WT}=pAct-OR^{WT} transgenic plants; OR^{His}=pAct-OR^{His} transgenic plants.



Appendix Material 19 Relative leaf chlorophyll content of T₁ *TaOR*^{WT} and *TaOR*^{His} overexpression lines compared to GRF and wild-type controls. SPAD measurements are the average of two measurements made 2 and 3 weeks after anthesis. Bold red letters next to the groups' names indicate statistical significance between the groups as determined by a one-way ANOVA. WT=non-transgenic wild-type controls; GRF=*GRF4-GIF1* transgenic controls; OR^{WT}=pAct-OR^{WT} transgenic plants; OR^{His}=pAct-OR^{His} transgenic plants.

Appendix Material 20 Pairwise comparisons of height, grain number per plant, TGW, grain area and SPAD of T_1 $TaOR^{WT}$ and $TaOR^{His}$ overexpression lines compared to GRF controls and non-transgenic Cadenza plants. The tables show the differences of pairwise comparisons between the lines based on a one-way ANOVA with a Tukey post-hoc test. Significant comparisons ($p < 0.05$) are highlighted in grey. SE=standard error, df=degrees of freedom, 95% CI=95% confidence interval, WT=non-transgenic wild-type controls, GRF=GRF4-GIF1 transgenic controls; OR^{WT} =pAct- OR^{WT} transgenic plants; OR^{His} =pAct- OR^{His} transgenic plants.

a) Height

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
GRF- OR^{His}	2.48	1.84	141	1.351	0.5319	-2.29	7.262
GRF- OR^{WT}	-1.41	1.83	141	-0.773	0.8665	-6.16	3.339
GRF-WT	-3.89	3.36	141	-1.157	0.6546	-12.62	4.843
OR^{His} - OR^{WT}	-3.9	1.84	141	-2.12	0.1517	-8.67	0.881
OR^{His} -WT	-6.37	3.36	141	-1.894	0.2353	-15.11	2.374
OR^{WT} -WT	-2.47	3.36	141	-0.737	0.8822	-11.2	6.256

b) Grain number per plant

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
WT-GRF	139.5	45	68	3.1	0.0146	21	258.01
WT- OR^{His}	82.2	45.6	68	1.8	0.2822	-38	202.34
WT- OR^{WT}	34.3	45	68	0.763	0.8707	-84.2	152.84
GRF- OR^{His}	-57.4	25.2	68	-2.274	0.1144	-123.8	9.09
GRF- OR^{WT}	-105.2	24.1	68	-4.373	0.0002	-168.5	-41.82
OR^{His} - OR^{WT}	-47.8	25.2	68	-1.896	0.2396	-114.3	18.62

c) Thousand-grain weight

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
WT-GRF	-1.48	3.27	68	-0.452	0.969	-10.1	7.14
WT- OR^{His}	3.4	3.32	68	1.024	0.7359	-5.3426	12.14
WT- OR^{WT}	2.8	3.27	68	0.855	0.8276	-5.8202	11.42
GRF- OR^{His}	4.88	1.83	68	2.659	0.0469	0.0472	9.71
GRF- OR^{WT}	4.28	1.75	68	2.446	0.078	-0.3278	8.89
OR^{His} - OR^{WT}	-0.6	1.83	68	-0.327	0.9878	-5.4323	4.23

d) Grain area

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
WT-GRF	-1.311	0.968	68	-1.354	0.5323	-3.862	1.24
WT- OR^{His}	0.587	0.982	68	0.597	0.9326	-2	3.17
WT- OR^{WT}	0.7	0.968	68	0.723	0.8876	-1.85	3.25
GRF- OR^{His}	1.898	0.543	68	3.496	0.0045	0.468	3.33
GRF- OR^{WT}	2.012	0.518	68	3.886	0.0013	0.648	3.37
OR^{His} - OR^{WT}	0.114	0.543	68	0.209	0.9967	-1.316	1.54

e) SPAD

Contrast	Difference	SE	df	t.ratio	p.value	-95% CI	+95% CI
GRF- OR^{His}	1.694	0.616	142	2.748	0.0339	0.0914	3.296
GRF- OR^{WT}	-0.139	0.603	142	-0.23	0.9957	-1.7068	1.429
GRF-WT	-1.037	1.12	142	-0.926	0.791	-3.9478	1.874
OR^{His} - OR^{WT}	-1.833	0.61	142	-3.004	0.0164	-3.4188	-0.247
OR^{His} -WT	-2.731	1.124	142	-2.43	0.076	-5.6515	0.19
OR^{WT} -WT	-0.898	1.116	142	-0.804	0.8523	-3.7999	2.004

A-genome pangenome alignment

ChineseSpring_TraesCS6A02G241400.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Arinalrfor_TraesARI6A03G03310500.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Jagger_TraesJAG6A03G03347390.1	-----MLACSGLSGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	54
Julius_TraesJUL6A03G03380250.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Landmark_TraesLDM6A03G03356980.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Mace_TraesMAC6A03G03353130.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Stanley_TraesSTA6A03G03344170.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Renan_TraesRN6A0100634400.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Lancer_TraesLAC6A03G03309930.1	-----MLACSGLSGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	54
Norin61_TraesNOR6A03G03386850.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
SyMattis_TraesSYM6A03G03295900.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Cadenza_TraesCAD_scaffold_025161_01G000100.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60

ChineseSpring_TraesCS6A02G241400.1	PFAGDDQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEE	120
Arinalrfor_TraesARI6A03G03310500.1	PFAGDDQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEE	120
Jagger_TraesJAG6A03G03347390.1	PFAGDDQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEE	114
Julius_TraesJUL6A03G03380250.1	PFAGDDQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEE	120
Landmark_TraesLDM6A03G03356980.1	PFAGDDQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEE	120
Mace_TraesMAC6A03G03353130.1	PFAGDDQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEE	120
Stanley_TraesSTA6A03G03344170.1	PFAGDDQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEE	120
Renan_TraesRN6A0100634400.1	PFAGDDQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEE	120
Lancer_TraesLAC6A03G03309930.1	PFAGDDQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEE	113
Norin61_TraesNOR6A03G03386850.1	PFAGDDQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEE	119
SyMattis_TraesSYM6A03G03295900.1	PFAGDDQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEE	119
Cadenza_TraesCAD_scaffold_025161_01G000100.1	PFAGDDQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEE	119

ChineseSpring_TraesCS6A02G241400.1	IRRLRIQQRKNAELGINSNEEPEGLPDPFSPFPLPPLSAANLKVVYATCFSLIAAIMV	180
Arinalrfor_TraesARI6A03G03310500.1	IRRLRIQQRKNAELGINSNEEPEGLPDPFSPFPLPPLSAANLKVVYATCFSLIAAIMV	180
Jagger_TraesJAG6A03G03347390.1	IRRLRIQQRKNAELGINSNEEPEGLPDPFSPFPLPPLSAANLKVVYATCFSLIAAIMV	174
Julius_TraesJUL6A03G03380250.1	IRRLRIQQRKNAELGINSNEEPEGLPDPFSPFPLPPLSAANLKVVYATCFSLIAAIMV	180
Landmark_TraesLDM6A03G03356980.1	IRRLRIQQRKNAELGINSNEEPEGLPDPFSPFPLPPLSAANLKVVYATCFSLIAAIMV	180
Mace_TraesMAC6A03G03353130.1	IRRLRIQQRKNAELGINSNEEPEGLPDPFSPFPLPPLSAANLKVVYATCFSLIAAIMV	180
Stanley_TraesSTA6A03G03344170.1	IRRLRIQQRKNAELGINSNEEPEGLPDPFSPFPLPPLSAANLKVVYATCFSLIAAIMV	180
Renan_TraesRN6A0100634400.1	IRRLRIQQRKNAELGINSNEEPEGLPDPFSPFPLPPLSAANLKVVYATCFSLIAAIMV	180
Lancer_TraesLAC6A03G03309930.1	IRRLRIQQRKNAELGINSNEEPEGLPDPFSPFPLPPLSAANLKVVYATCFSLIAAIMV	173
Norin61_TraesNOR6A03G03386850.1	IRRLRIQQRKNAELGINSNEEPEGLPDPFSPFPLPPLSAANLKVVYATCFSLIAAIMV	179
SyMattis_TraesSYM6A03G03295900.1	IRRLRIQQRKNAELGINSNEEPEGLPDPFSPFPLPPLSAANLKVVYATCFSLIAAIMV	179
Cadenza_TraesCAD_scaffold_025161_01G000100.1	IRRLRIQQRKNAELGINSNEEPEGLPDPFSPFPLPPLSAANLKVVYATCFSLIAAIMV	179

ChineseSpring_TraesCS6A02G241400.1	FGGFLAPILELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEIN	240
Arinalrfor_TraesARI6A03G03310500.1	FGGFLAPILELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEIN	240
Jagger_TraesJAG6A03G03347390.1	FGGFLAPILELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEIN	234
Julius_TraesJUL6A03G03380250.1	FGGFLAPILELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEIN	240
Landmark_TraesLDM6A03G03356980.1	FGGFLAPILELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEIN	240
Mace_TraesMAC6A03G03353130.1	FGGFLAPILELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEIN	240
Stanley_TraesSTA6A03G03344170.1	FGGFLAPILELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEIN	240
Renan_TraesRN6A0100634400.1	FGGFLAPILELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEIN	240
Lancer_TraesLAC6A03G03309930.1	FGGFLAPILELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEIN	233
Norin61_TraesNOR6A03G03386850.1	FGGFLAPILELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEIN	239
SyMattis_TraesSYM6A03G03295900.1	FGGFLAPILELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEIN	239
Cadenza_TraesCAD_scaffold_025161_01G000100.1	FGGFLAPILELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEIN	239

ChineseSpring_TraesCS6A02G241400.1	NVKQQEHKRCYKCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGA	300
Arinalrfor_TraesARI6A03G03310500.1	NVKQQEHKRCYKCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGA	300
Jagger_TraesJAG6A03G03347390.1	NVKQQEHKRCYKCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGA	294
Julius_TraesJUL6A03G03380250.1	NVKQQEHKRCYKCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGA	300
Landmark_TraesLDM6A03G03356980.1	NVKQQEHKRCYKCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGA	300
Mace_TraesMAC6A03G03353130.1	NVKQQEHKRCYKCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGA	300
Stanley_TraesSTA6A03G03344170.1	NVKQQEHKRCYKCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGA	300
Renan_TraesRN6A0100634400.1	NVKQQEHKRCYKCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGA	300
Lancer_TraesLAC6A03G03309930.1	NVKQQEHKRCYKCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGA	293
Norin61_TraesNOR6A03G03386850.1	NVKQQEHKRCYKCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGA	299
SyMattis_TraesSYM6A03G03295900.1	NVKQQEHKRCYKCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGA	299
Cadenza_TraesCAD_scaffold_025161_01G000100.1	NVKQQEHKRCYKCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGA	299

ChineseSpring_TraesCS6A02G241400.1	GKVMCPTCLCTGMAMASEHDPRIIDPFD	327
Arinalrfor_TraesARI6A03G03310500.1	GKVMCPTCLCTGMAMASEHDPRIIDPFD	327
Jagger_TraesJAG6A03G03347390.1	GKVMCPTCLCTGMAMASEHDPRIIDPFD	321
Julius_TraesJUL6A03G03380250.1	GKVMCPTCLCTGMAMASEHDPRIIDPFD	327
Landmark_TraesLDM6A03G03356980.1	GKVMCPTCLCTGMAMASEHDPRIIDPFD	327
Mace_TraesMAC6A03G03353130.1	GKVMCPTCLCTGMAMASEHDPRIIDPFD	327
Stanley_TraesSTA6A03G03344170.1	GKVMCPTCLCTGMAMASEHDPRIIDPFD	327
Renan_TraesRN6A0100634400.1	GKVMCPTCLCTGMAMASEHDPRIIDPFD	327
Lancer_TraesLAC6A03G03309930.1	GKVMCPTCLCTGMAMASEHDPRIIDPFD	320
Norin61_TraesNOR6A03G03386850.1	GKVMCPTCLCTGMAMASEHDPRIIDPFD	326
SyMattis_TraesSYM6A03G03295900.1	GKVMCPTCLCTGMAMASEHDPRIIDPFD	326
Cadenza_TraesCAD_scaffold_025161_01G000100.1	GKVMCPTCLCTGMAMASEHDPRIIDPFD	326

B-genome pangenome alignment

ChineseSpring_TraesCS6B02G283200.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPPLSARRWRVAASAAAPGGSPDLPSSSSSTPP	60
Arinalrfor_TraesARI6B03G03517030.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPPLSARRWRVAASAAAPGGSPDLPSSSSSTPP	60
Jagger_TraesJAG6B03G03546680.1	-----MLACSGLSGRLRPPRAYADRLRPPLSARRWRVAASAAAPGGSPDLPSSSSSTPP	54
Julius_TraesJUL6B03G03587620.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPPLSARRWRVAASAAAPGGSPDLPSSSSSTPP	60
Lancer_TraesLAC6B03G03511950.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPPLSARRWRVAASAAAPGGSPDLPSSSSSTPP	60
Landmark_TraesLDM6B03G03559050.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPPLSARRWRVAASAAAPGGSPDLPSSSSSTPP	60
Mace_TraesMAC6B03G03556420.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPPLSARRWRVAASAAAPGGSPDLPSSSSSTPP	60
Norin61_TraesNOR6B03G03593130.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPPLSARRWRVAASAAAPGGSPDLPSSSSSTPP	60
Stanley_TraesSTA6B03G03546990.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPPLSARRWRVAASAAAPGGSPDLPSSSSSTPP	60
Renan_TraesRN6B0100765800.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPPLSARRWRVAASAAAPGGSPDLPSSSSSTPP	60
SyMattis_TraesSYM6B03G03499080.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPPLSARRWRVAASAAAPGGSPDLPSSSSSTPP	60
Cadenza_TraesCAD_scaffold_030968_01G000100.1	MLCSGRMLACSGLSGRLRPPRAYADRLRPPLSARRWRVAASAAAPGGSPDLPSSSSSTPP	60

ChineseSpring_TraesCS6B02G283200.1	PFAGGDEQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEIIR	120
Arinalrfor_TraesARI6B03G03517030.1	PFAGGDEQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEIIR	120
Jagger_TraesJAG6B03G03546680.1	PFAGGDEQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEIIR	114
Julius_TraesJUL6B03G03587620.1	PFAGGDEQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEIIR	120
Lancer_TraesLAC6B03G03511950.1	PFAGGDEQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEIIR	120
Landmark_TraesLDM6B03G03559050.1	PFAGGDEQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEIIR	120
Mace_TraesMAC6B03G03556420.1	PFAGGDEQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEIIR	120
Norin61_TraesNOR6B03G03593130.1	PFAGGDEQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEIIR	120
Stanley_TraesSTA6B03G03546990.1	PFAGGDEQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEIIR	120
Renan_TraesRN6B0100765800.1	PFAGGDEQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEIIR	120
SyMattis_TraesSYM6B03G03499080.1	PFAGGDEQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEIIR	120
Cadenza_TraesCAD_scaffold_030968_01G000100.1	PFAGGDEQAAAAAASSSSGFCIEGPEVQDFDKLDLQELDNIIRSRNKIFLHMEIIR	120

ChineseSpring_TraesCS6B02G283200.1	LRIQQRIKNAELGISNEEPGELPDPFSPFIPPLPPLSAANLKVYYATCFSLIAAIMVFGG	180
Arinalrfor_TraesARI6B03G03517030.1	LRIQQRIKNAELGISNEEPGELPDPFSPFIPPLPPLSAANLKVYYATCFSLIAAIMVFGG	180
Jagger_TraesJAG6B03G03546680.1	LRIQQRIKNAELGISNEEPGELPDPFSPFIPPLPPLSAANLKVYYATCFSLIAAIMVFGG	174
Julius_TraesJUL6B03G03587620.1	LRIQQRIKNAELGISNEEPGELPDPFSPFIPPLPPLSAANLKVYYATCFSLIAAIMVFGG	180
Lancer_TraesLAC6B03G03511950.1	LRIQQRIKNAELGISNEEPGELPDPFSPFIPPLPPLSAANLKVYYATCFSLIAAIMVFGG	180
Landmark_TraesLDM6B03G03559050.1	LRIQQRIKNAELGISNEEPGELPDPFSPFIPPLPPLSAANLKVYYATCFSLIAAIMVFGG	180
Mace_TraesMAC6B03G03556420.1	LRIQQRIKNAELGISNEEPGELPDPFSPFIPPLPPLSAANLKVYYATCFSLIAAIMVFGG	180
Norin61_TraesNOR6B03G03593130.1	LRIQQRIKNAELGISNEEPGELPDPFSPFIPPLPPLSAANLKVYYATCFSLIAAIMVFGG	180
Stanley_TraesSTA6B03G03546990.1	LRIQQRIKNAELGISNEEPGELPDPFSPFIPPLPPLSAANLKVYYATCFSLIAAIMVFGG	180
Renan_TraesRN6B0100765800.1	LRIQQRIKNAELGISNEEPGELPDPFSPFIPPLPPLSAANLKVYYATCFSLIAAIMVFGG	180
SyMattis_TraesSYM6B03G03499080.1	LRIQQRIKNAELGISNEEPGELPDPFSPFIPPLPPLSAANLKVYYATCFSLIAAIMVFGG	180
Cadenza_TraesCAD_scaffold_030968_01G000100.1	LRIQQRIKNAELGISNEEPGELPDPFSPFIPPLPPLSAANLKVYYATCFSLIAAIMVFGG	180

ChineseSpring_TraesCS6B02G283200.1	FLAPILELKLIGGTSYADFIRNVHLPQLSQVDPVIVASFSGGAVGVISALMVVEINNVK	240
Arinalrfor_TraesARI6B03G03517030.1	FLAPILELKLIGGTSYADFIRNVHLPQLSQVDPVIVASFSGGAVGVISALMVVEINNVK	240
Jagger_TraesJAG6B03G03546680.1	FLAPILELKLIGGTSYADFIRNVHLPQLSQVDPVIVASFSGGAVGVISALMVVEINNVK	234
Julius_TraesJUL6B03G03587620.1	FLAPILELKLIGGTSYADFIRNVHLPQLSQVDPVIVASFSGGAVGVISALMVVEINNVK	240
Lancer_TraesLAC6B03G03511950.1	FLAPILELKLIGGTSYADFIRNVHLPQLSQVDPVIVASFSGGAVGVISALMVVEINNVK	240
Landmark_TraesLDM6B03G03559050.1	FLAPILELKLIGGTSYADFIRNVHLPQLSQVDPVIVASFSGGAVGVISALMVVEINNVK	240
Mace_TraesMAC6B03G03556420.1	FLAPILELKLIGGTSYADFIRNVHLPQLSQVDPVIVASFSGGAVGVISALMVVEINNVK	240
Norin61_TraesNOR6B03G03593130.1	FLAPILELKLIGGTSYADFIRNVHLPQLSQVDPVIVASFSGGAVGVISALMVVEINNVK	240
Stanley_TraesSTA6B03G03546990.1	FLAPILELKLIGGTSYADFIRNVHLPQLSQVDPVIVASFSGGAVGVISALMVVEINNVK	240
Renan_TraesRN6B0100765800.1	FLAPILELKLIGGTSYADFIRNVHLPQLSQVDPVIVASFSGGAVGVISALMVVEINNVK	240
SyMattis_TraesSYM6B03G03499080.1	FLAPILELKLIGGTSYADFIRNVHLPQLSQVDPVIVASFSGGAVGVISALMVVEINNVK	240
Cadenza_TraesCAD_scaffold_030968_01G000100.1	FLAPILELKLIGGTSYADFIRNVHLPQLSQVDPVIVASFSGGAVGVISALMVVEINNVK	240

ChineseSpring_TraesCS6B02G283200.1	QQEHRCKYCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGAGKV	300
Arinalrfor_TraesARI6B03G03517030.1	QQEHRCKYCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGAGKV	300
Jagger_TraesJAG6B03G03546680.1	QQEHRCKYCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGAGKV	294
Julius_TraesJUL6B03G03587620.1	QQEHRCKYCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGAGKV	300
Lancer_TraesLAC6B03G03511950.1	QQEHRCKYCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGAGKV	300
Landmark_TraesLDM6B03G03559050.1	QQEHRCKYCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGAGKV	300
Mace_TraesMAC6B03G03556420.1	QQEHRCKYCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGAGKV	300
Norin61_TraesNOR6B03G03593130.1	QQEHRCKYCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGAGKV	300
Stanley_TraesSTA6B03G03546990.1	QQEHRCKYCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGAGKV	300
Renan_TraesRN6B0100765800.1	QQEHRCKYCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGAGKV	300
SyMattis_TraesSYM6B03G03499080.1	QQEHRCKYCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGAGKV	300
Cadenza_TraesCAD_scaffold_030968_01G000100.1	QQEHRCKYCLGTGYLACARCSSTGAVVLEPVSTFSDGQPLSAPKTERCPNCSGAGKV	300

ChineseSpring_TraesCS6B02G283200.1	MCPTCLCTGMAMASEHDPRIIDPFD 324	
Arinalrfor_TraesARI6B03G03517030.1	MCPTCLCTGMAMASEHDPRIIDPFD 324	
Jagger_TraesJAG6B03G03546680.1	MCPTCLCTGMAMASEHDPRIIDPFD 318	
Julius_TraesJUL6B03G03587620.1	MCPTCLCTGMAMASEHDPRIIDPFD 324	
Lancer_TraesLAC6B03G03511950.1	MCPTCLCTGMAMASEHDPRIIDPFD 324	
Landmark_TraesLDM6B03G03559050.1	MCPTCLCTGMAMASEHDPRIIDPFD 324	
Mace_TraesMAC6B03G03556420.1	MCPTCLCTGMAMASEHDPRIIDPFD 324	
Norin61_TraesNOR6B03G03593130.1	MCPTCLCTGMAMASEHDPRIIDPFD 324	
Stanley_TraesSTA6B03G03546990.1	MCPTCLCTGMAMASEHDPRIIDPFD 324	
Renan_TraesRN6B0100765800.1	MCPTCLCTGMAMASEHDPRIIDPFD 324	
SyMattis_TraesSYM6B03G03499080.1	MCPTCLCTGMAMASEHDPRIIDPFD 324	
Cadenza_TraesCAD_scaffold_030968_01G000100.1	MCPTCLCTGMAMASEHDPRIIDPFD 324	

D-genome pangenome alignment

ChineseSpring_TraesCS6D02G223600.1	MLCSGRMLACSLSPGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Arinalrfor_TraesARI6D03G03704790.1	MLCSGRMLACSLSPGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Jagger_TraesJAG6D03G03723550.1	-----MLACSLSPGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	54
Julius_TraesJUL6D03G03773610.1	MLCSGRMLACSLSPGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Lancer_TraesLAC6D03G03691300.1	MLCSGRMLACSLSPGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Landmark_TraesLDM6D03G03744560.1	MLCSGRMLACSLSPGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Mace_TraesMAC6D03G03738630.1	MLCSGRMLACSLSPGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Norin61_TraesNOR6D03G03781230.1	MLCSGRMLACSLSPGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Stanley_TraesSTA6D03G0373760.1	MLCSGRMLACSLSPGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Renan_TraesRN6D0100581600.1	MLCSGRMLACSLSPGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
SyMattis_TraesSYM6D03G03688230.1	MLCSGRMLACSLSPGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60
Cadenza_TraesCAD_scaffold_098052_01G000200.1	MLCSGRMLACSLSPGRLRPPRAYADRLRPLPARRWRVAASAAAPGGSPDLPSSSSTPP	60

ChineseSpring_TraesCS6D02G223600.1	PFAGDDQAAAAAAAASSSGFCIEGPEVQDFDKLDLQEIILDNIRSRNKIFLHMEEI	120
Arinalrfor_TraesARI6D03G03704790.1	PFAGDDQAAAAAAAASSSGFCIEGPEVQDFDKLDLQEIILDNIRSRNKIFLHMEEI	120
Jagger_TraesJAG6D03G03723550.1	PFAGDDQAAAAAAAASSSGFCIEGPEVQDFDKLDLQEIILDNIRSRNKIFLHMEEI	114
Julius_TraesJUL6D03G03773610.1	PFAGDDQAAAAAAAASSSGFCIEGPEVQDFDKLDLQEIILDNIRSRNKIFLHMEEI	120
Lancer_TraesLAC6D03G03691300.1	PFAGDDQAAAAAAAASSSGFCIEGPEVQDFDKLDLQEIILDNIRSRNKIFLHMEEI	120
Landmark_TraesLDM6D03G03744560.1	PFAGDDQAAAAAAAASSSGFCIEGPEVQDFDKLDLQEIILDNIRSRNKIFLHMEEI	120
Mace_TraesMAC6D03G03738630.1	PFAGDDQAAAAAAAASSSGFCIEGPEVQDFDKLDLQEIILDNIRSRNKIFLHMEEI	120
Norin61_TraesNOR6D03G03781230.1	PFAGDDQAAAAAAAASSSGFCIEGPEVQDFDKLDLQEIILDNIRSRNKIFLHMEEI	120
Stanley_TraesSTA6D03G0373760.1	PFAGDDQAAAAAAAASSSGFCIEGPEVQDFDKLDLQEIILDNIRSRNKIFLHMEEI	120
Renan_TraesRN6D0100581600.1	PFAGDDQAAAAAAAASSSGFCIEGPEVQDFDKLDLQEIILDNIRSRNKIFLHMEEI	120
SyMattis_TraesSYM6D03G03688230.1	PFAGDDQAAAAAAAASSSGFCIEGPEVQDFDKLDLQEIILDNIRSRNKIFLHMEEI	120
Cadenza_TraesCAD_scaffold_098052_01G000200.1	PFAGDDQAAAAAAAASSSGFCIEGPEVQDFDKLDLQEIILDNIRSRNKIFLHMEEI	120

ChineseSpring_TraesCS6D02G223600.1	RRLRQQRIKNAELGISNEEPEGLPDPFSPFIFPLPPLSAANLKVYYATCFSLIAAIMVF	180
Arinalrfor_TraesARI6D03G03704790.1	RRLRQQRIKNAELGISNEEPEGLPDPFSPFIFPLPPLSAANLKVYYATCFSLIAAIMVF	180
Jagger_TraesJAG6D03G03723550.1	RRLRQQRIKNAELGISNEEPEGLPDPFSPFIFPLPPLSAANLKVYYATCFSLIAAIMVF	174
Julius_TraesJUL6D03G03773610.1	RRLRQQRIKNAELGISNEEPEGLPDPFSPFIFPLPPLSAANLKVYYATCFSLIAAIMVF	180
Lancer_TraesLAC6D03G03691300.1	RRLRQQRIKNAELGISNEEPEGLPDPFSPFIFPLPPLSAANLKVYYATCFSLIAAIMVF	180
Landmark_TraesLDM6D03G03744560.1	RRLRQQRIKNAELGISNEEPEGLPDPFSPFIFPLPPLSAANLKVYYATCFSLIAAIMVF	180
Mace_TraesMAC6D03G03738630.1	RRLRQQRIKNAELGISNEEPEGLPDPFSPFIFPLPPLSAANLKVYYATCFSLIAAIMVF	180
Norin61_TraesNOR6D03G03781230.1	RRLRQQRIKNAELGISNEEPEGLPDPFSPFIFPLPPLSAANLKVYYATCFSLIAAIMVF	180
Stanley_TraesSTA6D03G0373760.1	RRLRQQRIKNAELGISNEEPEGLPDPFSPFIFPLPPLSAANLKVYYATCFSLIAAIMVF	180
Renan_TraesRN6D0100581600.1	RRLRQQRIKNAELGISNEEPEGLPDPFSPFIFPLPPLSAANLKVYYATCFSLIAAIMVF	180
SyMattis_TraesSYM6D03G03688230.1	RRLRQQRIKNAELGISNEEPEGLPDPFSPFIFPLPPLSAANLKVYYATCFSLIAAIMVF	180
Cadenza_TraesCAD_scaffold_098052_01G000200.1	RRLRQQRIKNAELGISNEEPEGLPDPFSPFIFPLPPLSAANLKVYYATCFSLIAAIMVF	180

ChineseSpring_TraesCS6D02G223600.1	GGFLAPLLELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEINN	240
Arinalrfor_TraesARI6D03G03704790.1	GGFLAPLLELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEINN	240
Jagger_TraesJAG6D03G03723550.1	GGFLAPLLELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEINN	234
Julius_TraesJUL6D03G03773610.1	GGFLAPLLELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEINN	240
Lancer_TraesLAC6D03G03691300.1	GGFLAPLLELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEINN	240
Landmark_TraesLDM6D03G03744560.1	GGFLAPLLELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEINN	240
Mace_TraesMAC6D03G03738630.1	GGFLAPLLELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEINN	240
Norin61_TraesNOR6D03G03781230.1	GGFLAPLLELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEINN	240
Stanley_TraesSTA6D03G0373760.1	GGFLAPLLELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEINN	240
Renan_TraesRN6D0100581600.1	GGFLAPLLELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEINN	240
SyMattis_TraesSYM6D03G03688230.1	GGFLAPLLELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEINN	240
Cadenza_TraesCAD_scaffold_098052_01G000200.1	GGFLAPLLELKLIGGTSYADFIRNVHLPMLQSQVDPIVASFSGGAVGVISALMVVEINN	240

ChineseSpring_TraesCS6D02G223600.1	VKQEHKRCYKCLGTGYLACARCSSTGAVVLTEPVSTFSDGDQPLSAPKTERCPNCSGAG	300
Arinalrfor_TraesARI6D03G03704790.1	VKQEHKRCYKCLGTGYLACARCSSTGAVVLTEPVSTFSDGDQPLSAPKTERCPNCSGAG	300
Jagger_TraesJAG6D03G03723550.1	VKQEHKRCYKCLGTGYLACARCSSTGAVVLTEPVSTFSDGDQPLSAPKTERCPNCSGAG	294
Julius_TraesJUL6D03G03773610.1	VKQEHKRCYKCLGTGYLACARCSSTGAVVLTEPVSTFSDGDQPLSAPKTERCPNCSGAG	300
Lancer_TraesLAC6D03G03691300.1	VKQEHKRCYKCLGTGYLACARCSSTGAVVLTEPVSTFSDGDQPLSAPKTERCPNCSGAG	300
Landmark_TraesLDM6D03G03744560.1	VKQEHKRCYKCLGTGYLACARCSSTGAVVLTEPVSTFSDGDQPLSAPKTERCPNCSGAG	300
Mace_TraesMAC6D03G03738630.1	VKQEHKRCYKCLGTGYLACARCSSTGAVVLTEPVSTFSDGDQPLSAPKTERCPNCSGAG	300
Norin61_TraesNOR6D03G03781230.1	VKQEHKRCYKCLGTGYLACARCSSTGAVVLTEPVSTFSDGDQPLSAPKTERCPNCSGAG	300
Stanley_TraesSTA6D03G0373760.1	VKQEHKRCYKCLGTGYLACARCSSTGAVVLTEPVSTFSDGDQPLSAPKTERCPNCSGAG	300
Renan_TraesRN6D0100581600.1	VKQEHKRCYKCLGTGYLACARCSSTGAVVLTEPVSTFSDGDQPLSAPKTERCPNCSGAG	300
SyMattis_TraesSYM6D03G03688230.1	VKQEHKRCYKCLGTGYLACARCSSTGAVVLTEPVSTFSDGDQPLSAPKTERCPNCSGAG	300
Cadenza_TraesCAD_scaffold_098052_01G000200.1	VKQEHKRCYKCLGTGYLACARCSSTGAVVLTEPVSTFSDGDQPLSAPKTERCPNCSGAG	300

ChineseSpring_TraesCS6D02G223600.1	KVMCPTCLCTGMAMASEHDPRI DPF 326	
Arinalrfor_TraesARI6D03G03704790.1	KVMCPTCLCTGMAMASEHDPRI DPF 326	
Jagger_TraesJAG6D03G03723550.1	KVMCPTCLCTGMAMASEHDPRI DPF 320	
Julius_TraesJUL6D03G03773610.1	KVMCPTCLCTGMAMASEHDPRI DPF 326	
Lancer_TraesLAC6D03G03691300.1	KVMCPTCLCTGMAMASEHDPRI DPF 326	
Landmark_TraesLDM6D03G03744560.1	KVMCPTCLCTGMAMASEHDPRI DPF 326	
Mace_TraesMAC6D03G03738630.1	KVMCPTCLCTGMAMASEHDPRI DPF 326	
Norin61_TraesNOR6D03G03781230.1	KVMCPTCLCTGMAMASEHDPRI DPF 326	
Stanley_TraesSTA6D03G0373760.1	KVMCPTCLCTGMAMASEHDPRI DPF 326	
Renan_TraesRN6D0100581600.1	KVMCPTCLCTGMAMASEHDPRI DPF 326	
SyMattis_TraesSYM6D03G03688230.1	KVMCPTCLCTGMAMASEHDPRI DPF 326	
Cadenza_TraesCAD_scaffold_098052_01G000200.1	KVMCPTCLCTGMAMASEHDPRI DPF 326	

Appendix Material 21 Alignment of the TaOR-6A, TaOR-6B and TaOR-6D protein sequences within the wheat pangenome. The sequences within this pangenome are highly conserved.

Appendix 4 – Appendix Materials of Chapter 5

Kronos TILLING line K4596	EILDNIRSRRNKIFLHMEE--IRRLRIQQRIKNAELGISNEEPE-	143
Capsella_rubella	EIQDNIRSRRNKIFLLMEE--VRLRVQQLKSVQS--INELYSEL	128
Selaginella_moellendorffii	EIRDNIMSRRNKIFLLMEEASVRLRLRIQLRIKNAEQGVEDD---	62
Physcomitrella_patens	EIRDNITSRRNKIFLLMEE--VRLRLRIQQRIKSAEQGLSDNDPSN	47
Amborella_trichopoda	EIQDNIRSRRNKIFLHMEE--VRLRLRIQQRIKSAELGVLKEEKE-	123
Malus_domestica	EIQDNIRSRRNKIFLHTEE--VRLRLRIQQRIKRAELGAVNEDQE-	137
Prunus_persica	EIQDNIRSRRNKIFLHMEE--VRLRLRIQQRIKSAELGMLSEDEQE-	128
Brachypodium_distachyon	EILDNIRSRRNKIFLHMEE--IRRLRIQQRIKNAELGISVEEHE-	135
Avena_sativa	EILDNIRSRRNKIFLHMEE--IRRLRIQQRIKNAELGIAVEEPE-	147
Aegilops_tauschii	EILDNIRSRRNKIFLHMEE--IRRLRIQQRIKNAELGISNEEPE-	142
Triticum_aestivum-6D	EILDNIRSRRNKIFLHMEE--IRRLRIQQRIKNAELGISNEEPE-	142
Triticum_urartu	EILDNIRSRRNKIFLHMEE--IRRLRIQQRIKNAELGISNEEPE-	142
Triticum_aestivum-6A	EILDNIRSRRNKIFLHMEE--IRRLRIQQRIKNAELGISNEEPE-	143
Triticum_turgidum-6A	EILDNIRSRRNKIFLHMEE--IRRLRIQQRIKNAELGISNEEPE-	143
Hordeum_vulgare	EILDNIRSRRNKIFLHMEE--IRRLRIQQRIKNAELGISIEEPE-	140
Triticum_aestivum-6B	EILDNIRSRRNKIFLHMEE--IRRLRIQQRIKNAELGISNEEPE-	140
Triticum_turgidum-6B	EILDNIRSRRNKIFLHMEE--IRRLRIQQRIKNAELGISNEEPE-	140
Oryza_sativa	EIQDNIRSRRNKIFLHMEE--IRRLRIQQRIKKNVELGISVDVPE-	148
Sorghum_bicolor	EIQDNIRSRRNKIFLHMEE--IRRLRIQQRIKKNVELGISDEESD-	137
Zea_mays	EIQDNIRSRRNKIFLHMEE--IRRLRIQQRIKKNVELGISDEERD-	137
Setaria_italic	EIQDNIRSRRNKIFLHMEE--IRRLRIQQRIKNAELGISVDEPD-	138
Eragrostis_tef	EIQDNIRSRRNKIFLHMEE--IRRLRIQQRIKKNVELGISVEEPL-	136
Citrus_sinensis	EIHDNIRSRRNKIFLHMEE--VRLRLRIQQRIKNAELGISKEEQD-	127
Vitis_vinifera	EIQDNIRSRRNKIFLHMEE--VRLRLRIQQRIKNAELGILKEQ-E-	123
Manihot_esculenta	EIRDNIRSRRNKIFLQMEE--VRLRLRIQQRIKSAELGILKEDHE-	131
Cucumis_melo	EIQENIRSRRNKIFLHMEE--VRLRLRIQQRIKNAELGISKEERE-	141
Arabidopsis_thaliana	EIQDNIRSRRNKIFLHMEE--VRLRLRIQQRIKNTLGIINEEQE-	123
Brassica_oleracea	EIQDNIRSRRNKIFLHMEE--VRLRLRIQQRIKNTLGIIDEEQE-	121
Ipomoea_batatas	EIQDNIRSRRNKIFLHMEE--VRLRLRIQQRIKNAELGNLNEKQE-	129
Nicotiana_tabacum	EIRDNIRSRRNKIFLHMEE--VRLRLRIQQRLKSAELGILTDEQE-	127
Solanum_lycopersicum	EIRDNIRSRRNKIFLHMEE--VRLRLRIQQRIKSAELGIITEAQE-	128
Solanum_tuberosum	EIRDNIRSRRNKIFLHMEE--VRLRLRIQQRIKSAELGILTDAQE-	129
Eucalyptus_grandis	EIQDNIRSRRNKIFLHMEE--VRLRLRIQQRIKSAELGVLKDEQE-	131
Daucus_carota	EIQDNIRSRRNKIFLHMEE--VRLRLRIQQRIKNAELGISNEEPE-	85
Cicer_arietinum	EIQDNIRSRRNKIFLHMEE--VRLRLRIQQRIKNAELGIFKEEQE-	129
Glycine_max	EIQDNIRSRRNKIFLHMEE--VRLRLRIQQRIKSAELGILNEEQE-	117
Phaseolus_vulgaris	EIQDNIRSRRNKIFLQMEE--VRLRLRIQQRIKNAELGIIEEQE-	127
	** : ** ***** ** : ***** * * : : :	

Appendix Material 22 Protein sequence alignment of 35 *OR* orthologues at the E141K residue identified within durum wheat *TdOR* TILLING lines. The equivalent E141K residue in each orthologue is highlighted in yellow.