

**Investigating the Genetic Basis of Seed
and Leaf Micronutrient Concentrations in
*Brassica napus***

Aoife Grace Sweeney

PhD

University of York

Biology

September 2018

Abstract

With a growing population, changing climate and limited new land available, investigating ways to make crops better at using nutrients and boosting yields is becoming a priority. Such work could lead to improvements in human nutrition (i.e. biofortification); the remediation and/or use of contaminated lands (e.g. phytoremediation), as well as general improvements in crop yields. The current research has focused on investigating micronutrient variation in *Brassica napus* (an important oil seed crop and member of the agriculturally important *Brassicaceae*). An Associative Transcriptomic (AT) approach was used, exploiting the natural variation in gene sequence and expression amongst a diversity panel of *B. napus* to explore differences in the seed and leaf ionome. Candidates from AT were validated by testing their orthologous genes with *Arabidopsis thaliana* T-DNA insertional mutants; if the micronutrient concentration was disrupted relative to a wild type control then the function was validated. After verifying the role of the candidates in *A. thaliana*, the markers from AT analysis could be exploited in marker assisted selection to improve micronutrient use efficiency. In addition to the AT and *A. thaliana* analyses in seed and leaf, two other lines of enquiry were investigated. First, the link between the seed ionome and glucosinolates (GSL) was investigated; this research highlighted the disruption in seed ionome caused by breeding for low GSL lines and has implications for its growth under nutrient deficient conditions. Second, the negative association between time to flowering (prior to floral induction) and leaf ionome was investigated with a leaf ionome timeline. This research highlighted a potential link between the age-dependent flowering pathways and leaf nutrient status, however further research is required to assess whether leaf nutrients play an active role in floral induction. It is hoped that such research will aid in the stabilisation of crop yields and reduce fertiliser inputs.

Table of Contents

Abstract.....	2
Table of Contents.....	3
List of Tables	8
List of Figures	10
List of Appendices	14
Acknowledgements.....	15
Declaration.....	16
1 Introduction	17
1.1 Aim and Scope of PhD thesis	17
1.2 Literature review.....	19
1.2.1 Introduction	19
1.2.2 Plant nutrition	19
1.2.3 What is nutrient use efficiency?	20
1.2.4 The importance of micronutrient use efficiency	21
1.2.5 Elements of interest:.....	23
1.2.6 <i>B. napus</i> in general.....	38
1.2.7 Associative Transcriptomics (AT)	40
1.2.8 Quantitative genetics and Ionomics	44
1.2.9 Quantifying multiple mineral elements	46
1.2.10 Candidate gene functional validation	48
1.2.11 Conclusions	49
2 General methods	51
2.1 Data sets:	51
2.1.1 Pre-existing data sets	51
2.1.2 Selection and processing of pre-existing data	53
2.2 Associative Transcriptomics (AT)	53
2.2.1 AT pipeline	53

2.2.2	Analysis of AT graphed outputs	53
2.3	Testing the predictive capabilities of markers.....	56
2.4	<i>Arabidopsis thaliana</i> T-DNA knock-out analyses of candidate genes.....	57
2.4.1	Growth of plants	57
2.4.2	Genotyping:.....	61
2.5	Acid digestion and ICP-MS analysis.....	62
3	Seed ionome investigation.....	64
3.1	Introduction	64
3.2	Analysis of individual elements within seeds:	66
3.2.1	Associative transcriptomic outputs, predictions and candidates: Cu concentration in seed	66
3.2.2	Associative transcriptomic outputs, predictions and candidates: Cd concentration in seed	71
3.2.3	Associative transcriptomic outputs, predictions and candidates: Mn concentration in seed	76
3.2.4	Associative transcriptomic outputs, predictions and candidates: Zn concentration in seed	80
3.2.5	Associative Transcriptomic outputs, predictions and candidates: Mo concentration in seed	82
3.2.6	Associative Transcriptomic outputs, predictions and candidates: S concentration in seed	85
3.2.7	Summary of AT results and conclusions	91
3.3	Seed candidate gene analysis	92
3.3.1	Seed candidate gene analysis: Cu and Cd seed candidates	93
3.3.2	Seed candidate gene analysis: S seed candidates.....	98
3.3.3	Summary of candidate gene analysis.....	104
3.4	Chapter summary conclusions.....	105
4	Investigating the relationship between the seed ionome and glucosinolates in <i>B. napus</i>	107
4.1	Introduction	107

4.2	Case study specific methods:.....	109
4.2.1	Weighted Gene Co-expression Network Analysis (WGCNA)	109
4.2.2	Seed glucosinolate analysis.....	109
4.2.3	Seed Sulfate analysis.....	109
4.2.4	Leaf senescence analysis.....	110
4.2.5	Pod ionome investigation:	110
4.3	Results.....	114
4.3.1	WGCNA results.....	114
4.3.2	Seed GSL and sulfate analysis	117
4.3.3	Leaf senescence analysis.....	120
4.3.4	Pod ionome investigation	121
4.4	Chapter summary and conclusions.....	127
5	Leaf ionome investigation.....	128
5.1	Introduction	128
5.2	Analysis of individual elements within leaves:	129
5.2.1	Associative transcriptomic outputs, predictions and candidates: Cu concentration in leaf material	129
5.2.2	Associative transcriptomic outputs, predictions and candidates: Cd concentration in leaf material	134
5.2.3	Associative transcriptomic outputs, predictions and candidates: Mn concentration in leaf material	138
5.2.4	Associative transcriptomic outputs, predictions and candidates: Zn concentration in leaf material	142
5.2.5	Associative Transcriptomic outputs, predictions and candidates: Mo concentration in leaf material	146
5.2.6	Associative Transcriptomic outputs, predictions and candidates: S concentration in leaf material	150
5.2.7	Summary of Associative Transcriptomic outputs, predictions and candidates	154
5.3	Leaf candidate gene analysis	156

5.3.1	Leaf candidate gene analysis: Cu and Cd leaf concentration.....	158
5.3.2	Leaf candidate gene analysis: Mo leaf concentration	163
5.3.3	Leaf candidate gene analysis: summary, discussion and conclusions	164
5.4	Chapter summary and conclusions.....	165
6	Investigating the relationship between the wider leaf ionome and flowering	167
6.1	Introduction	167
6.2	Leaf ionome and flowering specific methods:.....	169
6.2.1	Controlling for spring and winter OSR ecotypes: splitting the diversity panel according to flowering time and <i>FLC</i> expression	169
6.2.2	'Total leaf essential ionome' vs 'Coefficient of Variation'	169
6.2.3	<i>B. napus</i> flowering time	170
6.2.4	Candidate gene analysis: growth condition variation in <i>A. thaliana</i>	170
6.2.5	Leaf ionome timeline in <i>B. napus</i>	171
6.3	Results.....	172
6.3.1	Controlling for spring and winter OSR ecotypes: splitting the diversity panel according to <i>FLC</i> expression and flowering time	172
6.3.2	Associative transcriptomic outputs, predictions and candidates: 'Total leaf essential elements'	181
6.3.3	Associative transcriptomic outputs, predictions and candidates: 'Leaf essential coefficient of variation'	186
6.3.4	Associative transcriptomic outputs, predictions and candidates: Flowering time	192
6.3.5	Summary of flowering time related associative transcriptomic outputs, predictions and candidates.....	197
6.3.6	Flowering candidate gene analysis	198
6.3.7	Leaf ionome timeline	201
6.4	Chapter summary and conclusions.....	207
7	Conclusions and General discussions.....	210
7.1	General discussion	210

7.2	Understanding micronutrient concentration in <i>B. napus</i>	216
7.3	Flowering time and leaf element concentration	222
7.4	Overall conclusions	224
	Appendices.....	225
	Abbreviations.....	246
	References	250

List of Tables

Table 2.4.1: Candidate genes found as part of the current project.	59
Table 2.4.2.a variance component analysis from Thomas et al., (2016) for seed mineral composition in <i>B. napus</i>	65
Table 3.2.1.a Predictive capability of markers from Cu concentration in seed AT analysis.	68
Table 3.2.2.a Predictive capability of markers from Cd concentration in seed AT analysis.	74
Table 3.2.3.a Predictive capability of markers from Mn concentration in seed AT analysis. .	78
Table 3.2.5.a Predictive capability of markers from Mo concentration in seed AT analysis. .	84
Table 3.2.6.a Predictive capability of markers from S concentration in seed AT analysis.	87
Table 3.2.7.a list of the candidate genes taken forward for further study in seeds.....	91
Table 4.3.1.a Markers from the most highly correlated module with seed S concentration from WGCNA analysis.	116
Table 4.3.4.a variance component analysis from Thomas et al., (2016) for leaf mineral composition in <i>B. napus</i>	128
Table 5.2.1.a Predictive capability of markers from AT analysis of Cu concentration in leaves.	131
Table 5.2.2.a Predictive capability of markers from AT analysis of Cd concentration in leaves.	136
Table 5.2.3.a Predictive capability of markers from AT analysis of Mn concentration in leaves.	140
Table 5.2.4.a Predictive capability of markers from AT analysis of Zn concentration in leaves.	144
Table 5.2.5.a Predictive capability of markers from AT analysis of Mo concentration in leaves.	148
Table 5.2.6.a Predictive capability of markers from AT analysis of S concentration in leaves.	152
Table 5.2.7.a a list of the candidate genes taken forward for further study from leaf element concentration AT results.	154
Table 5.3.3.a Flowering time correlations (as days relative to the first flowering line, generated by a project student, Mr Martínez Ortuño) against the expression (as RPKM) of all CDS gene models whose orthologues in <i>A. thaliana</i> are annotated as <i>SOC1/FLC</i> . Leaf element concentrations, as mg/kg DW and summed, and 'yield', as g.....	168
Table 6.3.1.a Correlation table between <i>FLC/SOC1</i> expression, flowering time and leaf nutrient concentration when the diversity panel is split according to A3 <i>FLC</i> expression....	174

Table 6.3.1.b Correlation table between <i>FLC/SOC1</i> expression, flowering time and leaf nutrient concentration when the diversity panel is split according to flowering time.	178
Table 6.3.2.a Predictive capability of markers from ‘total leaf essential ionome’ AT analysis.	183
Table 6.3.3.a Predictive capability of markers from AT analysis of ‘ranked leaf coefficient of variation’.	188
Table 6.3.4.a Predictive capability of markers from flowering time (as days) AT analysis. ...	194
Table 6.3.5.a A list of the flowering candidate genes taken forward for further study.	197
Table 6.3.7.a ANOVA results for individual elements, as well as those from elements which have been grouped/summed according to their Apparent Nutrient Remobilisation (ANR%) in <i>B. napus</i> from (Maillard et al., 2015) across vernalisation and development stages.	204

List of Figures

Figure 1.2.5.a Schematic representation of primary and secondary S assimilation (adapted from Mugford et al., (2011) and Koprivova and Kopriva, (2016)).	34
Figure 1.2.6.a The triangle of U: Brassica genome evolution/structure in accordance with the theory presented by (Nagaharu U., 1935).	39
Figure 1.2.7.a The Renewable Industrial Products from Rapeseed (RIPR) population structure from Havlickova et al., (2018).	41
Figure 1.2.7.b A schematic representation from (Trick et al., 2009b) of the various forms of single nucleotide polymorphisms (SNP) within a DH (doubled haploid) line (the polymorphisms are in bold within the sequences).	43
Figure 2.2.2.a An example AT Manhattan output from the current study with key features highlighted to emphasise the analysis process.	55
Figure 2.4.2.a: schematic from the SIGnAL T-DNA primer design tool (SIGnAL).	62
Figure 3.2.1.a Genome wide distribution of mapped markers associating with the Cu concentration in seeds (mg/kg DW) of all 383 accessions.	67
Figure 3.2.1.b Quantile-quantile plot of observed $-\log_{10}P$ values from AT SNP (left) and GEM (right) analysis for Cu concentration in seeds against expected $-\log_{10}P$ values	68
Figure 3.2.2.a Genome wide distribution of mapped markers associating with the Cd concentration in seeds (mg/kg DW) of all 383 accessions.	72
Figure 3.2.2.b Genome wide distribution of mapped markers associating with the Cd concentration in seeds (mg/kg DW) of 274 accessions.	73
Figure 3.2.3.a Genome wide distribution of mapped markers associating with the Mn concentration in seeds (mg/kg DW) of all 383 accessions.	77
Figure 3.2.3.b Quantile-quantile plot of observed $-\log_{10}P$ values from AT SNP analysis (left) and AT GEM analysis (right) for Mn concentration in seeds against expected $-\log_{10}P$ values.	78
Figure 3.2.4.a Genome wide distribution of mapped markers associating with the Zn concentration in seeds (mg/kg DW) of all 383 accessions.	81
Figure 3.2.5.a Genome wide distribution of mapped markers associating with the Mo concentration in seeds (mg/kg DW) of all 383 accessions.	83
Figure 3.2.6.a Genome wide distribution of mapped markers associating with the S concentration in seeds (mg/kg DW) of all 383 accessions.	86
Figure 3.2.7.a The seed 'interactome'	92
Figure 3.3.1.a Candidate gene analyses of Cu (left) and Cd (right) in the seeds of <i>A. thaliana</i> insert mutants, as mg/kg DW of seed tissue.	94

Figure 3.3.1.b Candidate gene analyses of Cu and Cd (as mg/kg DW) in seed <i>A. thaliana</i> insert mutants within stem and pod materials	95
Figure 3.3.1.c Candidate gene analyses of Cu and Cd seed <i>A. thaliana</i> insert mutants within leaves, as mg/kg DW of leaf tissue	97
Figure 3.3.2.a Candidate gene analyses of S seed <i>A. thaliana</i> insert mutants (<i>HAG1/Myb28</i> , <i>GTR2</i> and <i>Per1L</i>), as mg/kg DW of seed tissue.....	99
Figure 3.3.2.b Z-score graph detailing how each seed batch (as a different coloured line) from the <i>myb28/hag1</i> line varied in comparison to the average wildtype control (n= 6 for both, where n is a subsample of seeds from a pooled sample of seeds from 12 plants).	101
Figure 3.3.2.c Z-score graph detailing how leaf material (as a different coloured line) from the <i>myb28/hag1</i> line varied in comparison to the average wild type control (n= 8 for both, where n is an individual plant).	101
Figure 3.3.2.d candidate gene analyses of S seed <i>A. thaliana</i> insert mutants <i>seuss-like2</i> and <i>sbp1</i> , as mg/kg DW of seed tissue	103
Figure 3.3.3.a basic chemical structure of a glucosinolate (GSL) contains two S atoms.	108
Figure 4.2.5.a Experimental design for leaf ionome timeline and pod ionome experiments.	112
Figure 4.3.1.a WGCNA (weighted gene co-expression network analysis) results for the top module (Brown4) from S concentrations in seed.	115
Figure 4.3.2.a Correlation between the total S and total glucosinolate ($\mu\text{mol/g}$) content of <i>B. napus</i> seeds ($r^2= 0.8544$, $p<0.001$).	118
Figure 4.3.2.b Investigating the fidelity of two different extraction protocols for seed sulfate ($\mu\text{mol/g}$) extraction in <i>B. napus</i>	118
Figure 4.3.3.a S (left) and Mo (right) concentrations (as mg/kg DW) within senesced leaves of high S (HS) and low S (LS) seed lines under different N conditions.....	121
Figure 4.3.4.a Difference between the green seed, pod and stem of high S (HS) and low S (LS) seed lines in spring (S) and winter (W) OSR ecotypes for S concentration (mg/kg DW)	123
Figure 4.3.4.b Difference in Mo (as mg/kg DW) in the seeds of high S (HS) and low S (LS) seed lines in spring (S) and winter (W) OSR <i>B. napus</i> ecotypes.....	124
Figure 4.3.4.c Difference in K and Se (as mg/kg) in the pods of high S (HS) and low S (LS) seed lines in spring (S) and winter (W) OSR <i>B. napus</i> ecotypes.	125
Figure 4.3.4.d Difference in K, P and Mn concentration (as mg/kg DW) in the stems of high S (HS) and low S (LS) seed lines in spring (S) and winter (W) <i>B. napus</i> (OSR) ecotypes.	126
Figure 5.2.1.a Genome wide distribution of mapped markers associating with the Cu concentration in leaves (mg/kg DW) of all 383 accessions.	130

Figure 5.2.1.b Quantile-quantile plot of observed $-\log_{10}P$ values from AT SNP (left) and GEM (right) analysis for Cu concentration in leaves against expected $-\log_{10}P$ values.....	131
Figure 5.2.2.a Genome wide distribution of mapped markers associating with the Cd concentration in leaves (mg/kg DW) of all 383 accessions.	135
Figure 5.2.3.a Genome wide distribution of mapped markers associating with the Mn concentration in leaves (mg/kg DW) of all 383 accessions.	139
Figure 5.2.4.a Genome wide distribution of mapped markers associating with the Zn concentration in leaves (mg/kg DW) of all 383 accessions.	143
Figure 5.2.4.b Quantile-quantile plot of observed $-\log_{10}P$ values from AT SNP analysis (left) and AT GEM analysis (right) for Zn in leaf against expected $-\log_{10}P$ values.	144
Figure 5.2.5.a Genome wide distribution of mapped markers associating with the Mo concentration in leaves (mg/kg DW) of all 383 accessions.	147
Figure 5.2.6.a Genome wide distribution of mapped markers associating with the S concentration in leaves (mg/kg DW) of all 383 accessions.	151
Figure 5.2.7.a The leaf ‘interactome’.....	156
Figure 5.3.1.a Candidate gene analysis of Cu (left) and Cd (right) concentrations in the leaves of <i>A. thaliana</i> insert mutants as mg/kg DW of leaf tissue.....	159
Figure 5.3.1.b Z-score graph detailing how leaf material (as a different coloured line) from the <i>Cd leaf hma-2</i> line (At2g36950/SALK_069207C) varied in comparison to the average wildtype control (n: 8 for both, where n is the number of individual plants sampled for analyses).....	160
Figure 5.3.1.c Candidate gene analysis of Cu (left) and Cd (right) concentrations in the stem (top) and seeds (bottom) of <i>A. thaliana</i> insert mutants, as mg/kg DW of leaf tissue	162
Figure 5.3.2.a Candidate gene analysis of Mo in the leaves of <i>A. thaliana</i> insert mutants, as mg/kg DW of leaf tissue.....	163
Figure 6.3.2.a Genome wide distribution of mapped markers associating with the ‘Total essential leaf ionome’ (mg/kg DW), calculated from the sum total of essential element concentrations measured with ICP-MS analysis for each of the 383 accessions.....	182
Figure 6.3.3.a Genome wide distribution of mapped markers associating with the ‘ranked leaf coefficient of variation’ for essential element concentrations.	187
Figure 6.3.4.a Genome wide distribution of mapped markers associating with the days to flowering; calculated from the average of four plants.	193
Figure 6.3.6.a Candidate gene analysis of flowering candidates against ‘total essential elements’ in leaf tissues (as mg/kg DW) of <i>A. thaliana</i> insert mutants.....	199

Figure 6.3.6.b Candidate gene analysis of flowering candidates against P in leaf tissues (as mg/kg DW) of <i>A. thaliana</i> insert mutants	200
Figure 6.3.7.a leaf development timeline (as days from sowing)	202
Figure 6.3.7.b Total essential elements (mg/kg) within leaves at different developmental and vernalisation states	203
Figure 7.1.a temperature timeline for one of the polytunnels used during the growth of the RIPR panel	212
Figure 7.2.a seed: leaf ratios of all elements under investigation across the RIPR diversity panel	217
Figure 7.2.b ratio of S and Mo concentrations across seed and leaf tissues in the RIPR diversity panel	218
Figure 7.3.a Schematic representation of the relationship between leaf nutrient concentration and flowering time	222

List of Appendices

Appendix 1 AT output for Cu (mg/kg DW) seed on the 274 diversity panel	226
Appendix 2 AT output for Cd (mg/kg DW) seed on the 274 diversity panel	227
Appendix 3 AT output for Mn (mg/kg DW) seed on the 274 diversity panel	228
Appendix 4 AT output for Zn (mg/kg DW) seed on the 274 diversity panel.....	229
Appendix 5 AT output for Mo (mg/kg DW) seed on the 274 diversity panel	230
Appendix 6 AT output for S (mg/kg DW) seed on the 274 diversity panel.....	231
Appendix 7 AT output for Cu (mg/kg DW) leaf on the 274 diversity panel.....	232
Appendix 8 AT output for Cd (mg/kg DW) leaf on the 274 diversity panel	233
Appendix 9 AT output for Mn (mg/kg DW) leaf on the 274 diversity panel.....	234
Appendix 10 AT output for Zn (mg/kg DW) leaf on the 274 diversity panel	235
Appendix 11 AT output for Mo (mg/kg DW) leaf on the 274 diversity panel.....	236
Appendix 12 AT output for S (mg/kg DW) leaf on the 274 diversity panel	237
Appendix 13 AT output for Total leaf essential elements (mg/kg DW) on the 274 diversity panel	238
Appendix 14 AT output for Leaf essential coefficient of variation on the 274 diversity panel	239
Appendix 15 AT output for Flowering time (days, data produced by Mr Cándido José Martínez Ortuño) on the 274 diversity panel.....	240
Appendix 16 T-Test outputs for Myb28/HAG1 (AT5G61420/SALK_136312C) in comparison to the wildtype control in seeds (as mg/kg DW for each element).	241
Appendix 17 T-Test outputs for Myb28/HAG1 (AT5G61420/SALK_136312C) in comparison to the wildtype control in leaves (as mg/kg DW for each element).	241
Appendix 18 T-test outputs for Cd leaf HMA2 (AT2G36950/SALK_069207C) in comparison to the wildtype control in leaves (as mg/kg DW for each element).	241
Appendix 19 <i>SOC1</i> correlation tables between flowering (as days), <i>FLC/SOC1</i> expression (as RPKM) and leaf nutrients (as mg/kg DW) when split	242
Appendix 20 ANOVA results for individual elements (in mg/kg DW) across flowering T-DNA lines.....	245

Acknowledgements

This work would not have been possible without funding from the Biotechnology and Biological Sciences Research Council (BBSRC); Renewable Industrial Products from Rapeseed (RIPR: BB/L002124/1).

I will be eternally grateful for the help, support and encouragement provided by my 'super' supervisor, Ian. Thank you for four fantastic years of research and badminton, I have thoroughly enjoyed my time working with you and hope that you have likewise. Similarly, many thanks to Andrea my co-supervisor: your help especially at the start of my PhD was invaluable. Further, your support in all things 'R', be they AT, WGCNA or random 3D PCA, has been essential in the completion of my studies. I would also like to thank my thesis advisory panel, Neil and Simon, for their help and support; your questions have helped shape my thesis into what it is today.

Stan, thank you (and your research group) for help with the sulfur, sulfate and glucosinolate analysis.

To all the members of the RIPR consortium, thank you for your encouragement, questions and support over the course of my studies. Particularly, I would like to thank the research groups at the University of Nottingham for help with the ionic datasets (Prof. Broadley, Prof. Salt and Prof. Young). Lolita, thank you for your patience and help with setting up our own digestion protocol at York.

A massive thank you to all the members (and former members) of the Bancroft lab for making my time at York so enjoyable: Lenka: thank you for answering my infernal questions and always being willing to offer advice on experimental procedures; Natalia: your encouragement has been a constant which I relied upon to get me through some of the tougher times in my PhD; Zhesi: thanks for all your help with AT and computer problems; Lihong and Alison: the best lab-techs a PhD student could dream of, your help with setting up protocols and getting to know the lab were invaluable; Helen, Jevo and Varanya: thank you for all the fun times we have had, as well as the support and encouragement you have provided throughout my studies; Cándido: you helped make my first few months at York a lot less lonely and greatly speeded my settling in; Junhee: you were a great student to teach and I think I learnt just as much from supervising you as I hope you got from my supervision.

Finally thanks to my family for their love and encouragement through everything. Above all, thanks to Joe for his unceasing love, support and proof-reading capabilities.

Declaration

I, Aoife Grace Sweeney, declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

The Ionic data used within the study has been published as part of (Thomas et al., 2016), with the author contributing to discussions surrounding sulfur and molybdenum.

Work on copper and cadmium was performed in part by an undergraduate student, Mr Jun Hee Jung, with supervision from the author as part of his final year undergraduate project at The University of York. Full details and a breakdown of contributions are given throughout the text.

Flowering time data used for reference within this research was generated as part of an Erasmus/Lifelong Learning Programme project by Mr Cándido José Martínez Ortuño. Once again this has been fully acknowledged within the text.

Glucosinolate data was generated by the author under the supervision of Prof. Stanislav Kopriva at the University of Cologne, and in part repeated by Prof. Kopriva and his research group, as is detailed throughout the text.

Finally, leaf senescence analysis was performed on field trial materials sampled by the author but grown by ADAS and the University of Nottingham as part of the Renewable Industrial Products from Rapeseed (RIPR) consortium which the author is a member of. Full details are yet to be published and are not within the remit of the current study, Fraser *et al.*, unpublished.

1 Introduction

1.1 Aim and Scope of PhD thesis

The main aim of this research was to understand whether variation in micronutrient concentration in the seeds and leaves of *B. napus* (Oilseed rape) is caused by underlying genetic loci. The project adopted a quantitative genomics approach, using Associative Transcriptomics (AT) to measure the association between micronutrient concentration and transcript sequence/abundance across a diversity panel of *B. napus* under nutrient sufficient conditions. The associated regions would then be explored for candidate genes which might explain the variation observed in micronutrient concentration. This approach yielded two types of candidates; those which were already known to affect nutrient concentration as well as other, novel candidates. The large T-DNA insert mutant collection within *Arabidopsis thaliana* was then exploited to verify the role of novel candidates. Orthologues of the genes identified in *B. napus* were then tested with the T-DNA insert mutants in *A. thaliana* for disruption in the nutrient of interest (relative to a wild type control). Once verified in *A. thaliana*, these candidates would make promising targets for future research using the *B. napus* diversity panel and could be used as the first step towards breeding for improved micronutrient use efficiency in *B. napus*.

Overall, the project focused on one main hypothesis:

Variation in micronutrient concentration in the seeds and leaves of *B. napus* under nutrient sufficiency is a consequence of underlying genetic variation

This thesis will demonstrate how this hypothesis has been investigated within *B. napus*. To begin, there will be an overview of the background literature considered important for understanding the project. The literature review will start with the basics of plant nutrition and why it is important, both generally and in relation to *B. napus*. Discussion of *B. napus* breeding in general, the diversity panel and AT will follow. Finally, quantitative genetics methods used in other ionomics studies will be reviewed, alongside the various methods of verifying candidates from these analyses. After establishing the background relevant to the project, the general methods used throughout the project will be described. Next, the research investigating seeds will be presented; seed AT outputs will be given, alongside assessment of the trait predictability of markers and an explanation of the candidates tested in *A. thaliana*. The next chapter will investigate the relationship between seed glucosinolates (GSL) and the wider seed ionome which was highlighted as part of initial AT analyses of individual elements within

seeds. It will explain how Weighted Gene Co-expression Network Analysis (WGCNA), seed sulfate/GLS analysis and investigations into the ionome of senescing leaves, stems and pods was used to elucidate the association between the seed ionome and GSLs. In the fifth chapter, work performed as part of individual element AT analyses within the leaves will be detailed, in accordance with those previously described for the seed. As with the analyses on seeds, the final results chapter will elaborate on potential shared mechanisms highlighted in multiple AT outputs; the association between the broader leaf ionome and flowering time. Experiments with splitting the diversity panel, comparing AT results to flowering time data and a leaf ionome timeline will be presented. Finally, all of the results will be collated, establishing key findings and future research directions.

1.2 Literature review

1.2.1 Introduction

Considering the primary aim of this research is to understand the genetic basis of variation in micronutrient concentration within *B. napus* (with implications for micronutrient use efficiency), there is a wide range of topics which needs to be reviewed before any detailed analysis utilising AT can be discussed. Firstly, some general background on plant nutrition will be outlined; this will then be developed into the definition of nutrient use efficiency and highlight why micronutrient use efficiency is important (with an explanation as to why *B. napus* was used in this study). A more detailed background for each of the elements this thesis has considered will then be given; this will aid in the understanding of AT analysis and provide a summary of work performed previously in *B. napus* for the elements under investigation. The literature review will finish with a summary of the methods that have been used to investigate micronutrient concentration in the seeds and leaves of *B. napus*, comparing and contrasting the relative efficacy of these methods to those used in other research.

1.2.2 Plant nutrition

Alongside light, water and air, all plants require a range of mineral nutrients to successfully complete their lifecycle. Generally, these nutrients are split into two groups: the macronutrients (N, P, K, Mg, S and Ca) and the micronutrients (B, Cl, Mn, Fe, Ni, Cu, Zn and Mo). The primary distinction in these groupings is the relative amounts plants require; macronutrients are required in relatively large amounts (>1000s mg/kg) whilst the micronutrients are required in relatively small amounts (~100s- 0.1 mg/kg)(Alloway, 2013). All these nutrients are essential; specifically, an absence of any one of these nutrients will cause death and/or prevent the successful reproduction of the plant (Marschner, 1995c). Alongside the known essential elements are the 'beneficial elements' (e.g. Al, Co, Na, Se and Si); most plants can successfully complete their lifecycle without such elements (i.e. some are essential within specific species, such as Na in C4/CAM plants (Ohnishi et al., 1990)) however they are known to stimulate growth in some way (e.g. replacement of essential elements in their less specific functions, such as Na replacement of K in osmoregulation) (Pilon-Smits et al., 2009). In addition to elements known to play a biologically relevant role within plants, there are many other elements plants are known to take up which have no proven biological role and can be toxic. It is important to note that the term 'ionome' refers to all elements found within an organism (Lahner et al., 2003), irrespective of essentiality or toxicity. The presence of potentially toxic elements (whether essential or not) within plant vegetative tissues is of particular importance to human nutrition (Page and Feller, 2015). Historically, plant nutrient

research has focused on investigating one or two elements at a time; indeed, this is often the best approach for detailing nutrient essentiality and functionality. However an increasing awareness of the interactive nature of most elements within the ionome (e.g. the sharing of uptake pathways or the requirement of one essential element in the biological activation of another), alongside the availability of precise high-throughput and relatively cost-effective methods for multi-element analysis (i.e. ICP-MS/ICP-O/AES), has led to a wave of new research investigating the broader plant ionome (Salt, Baxter and Lahner, 2008; Baxter, 2015).

1.2.3 What is nutrient use efficiency?

The term “nutrient use efficiency” has unfortunately been subject to frequent definition. Generally, the definitions of nutrient use efficiency can be split into two groups: 1) those trying to quantify nutrient inputs in contrast to yield outputs; 2) investigations into plant uptake and internal utilisation. Both definitions have their strengths and limitations (Gourley, Allan and Russelle, 1994; Khoshgoftarmanesh et al., 2010). For example, having a measurable output allows for comparison, however what measurable output (e.g. crop yield, profit or nutrient recovery) and whether there are other confounding variables (e.g. differences in climate, production practices or baseline soil fertility) can all skew this definition of efficiency. In contrast, looking at uptake and utilisation can provide a scientific basis for understanding efficiency but may not translate into field conditions (Fageria, Baligar and Li, 2008). These multiple definitions highlight the complexity of ‘nutrient use efficiency’: it is dependent on many interacting variables, i.e. genotype, environment and management (Dresbøll and Thorup-Kristensen, 2014). In the current study the genetic basis of micronutrient concentration variation was investigated, representing the first step towards breeding crops with improved micronutrient use efficiency. A common theme amongst both types of efficiency definitions is the ability to grow on nutrient deficient soils, i.e. an efficient cultivar can produce the same yield as an inefficient cultivar but with fewer inputs. However, there are limitations to such approaches. For example, crops may yield better under deficiency but may not match the output of an inefficient crop under current fertilisation regimes (Gourley, Allan and Russelle, 1994). These cultivars would require further breeding to boost yields and limit other undesirable traits which were previously selected out of the elite cultivars. Furthermore, there is always the risk that breeding an efficient crop for one or two nutrients will result in an imbalance in other elements due to antagonism (Rietra et al., 2017), i.e. nutrient use efficiency efforts may be better focused on groups of elements and traits (White et al., 2013). Consequently, the current project has adopted a different approach by investigating nutrient concentration, which will likely be of relevance to nutrient use efficiency. It was theorised that exploiting the natural variation in nutrient acquisition/assimilation of plants under nutrient

sufficient conditions would highlight loci which would boost nutrient use efficiency without perturbing the rest of the ionome. The major flaws in this approach include the requirement for maintaining current fertilisation rates (although this is not a particular problem for the micronutrients as they are required in such small quantities, also 1.2.4). Furthermore, reliance upon natural variation under normal element concentrations may not give enough discriminatory power to identify associations and/or limit the efficiency mechanisms uncovered leading to insignificant improvements. Despite these limitations, taking this approach would at the very least provide further background information on plant nutrient concentration, which may aid future efforts in breeding for efficiency.

1.2.4 The importance of micronutrient use efficiency

In general, research on nutrient use efficiency has focused on the macronutrients, specifically N, P and K. For example a simple search of the publication database 'Web of Science' in the plant sciences category yielded 2644 results when "nutrient and efficien* and [nitrogen or phosphorus or potassium]" was searched (although it would appear that nearly half of these were N focused, 1160 results when NOT phosphorus NOT potassium was used), and only 985 when a list of all remaining essential elements was used (separated with OR: S, Ca, Mg, Cl, B, Fe, Mn, Zn, Cu, Ni and Mo) despite there being considerably more of these. It is easy to understand why research has focused on N in particular; it has previously been estimated that chemical N fertilisation supports food production for around half the world's population (Erisman et al., 2008). However, the gain in crop yields associated with fertiliser addition has several monetary and environmental (in terms of production and over application) costs. Given the growing human population, along with the issues of food security, climate change and limited availability of new croplands (and its distribution globally, with implications for nutrient stripping), nutrient use efficient crops for improved yields have been expounded as part of much wider systems based approaches to nutrient cycling (Jones et al., 2013), particularly in relation to the macronutrients.

However, research looking into the micronutrients is also a priority: their dual role in the nutrition of plants and animals, alongside the fact that many can be toxic contaminants, and the increasing awareness surrounding the interdependence of elements within the plant ionome (Baxter, 2015), has resulted in increased interest in micronutrient use efficiency (Cakmak, 2002). The common phrase "two sides of the same coin" has often been applied to this research as a way of illustrating the inter-relation of research on food biofortification and bioremediation (Guerinot and Salt, 2001). Research centring on 'hidden hunger' (i.e. micronutrient deficiency - in humans this includes vitamins as well as mineral elements) has

gained prevalence and is of particular importance in developing countries where most human population growth is expected (UN-DESA, 2017). Combined with the complications of farming in a changing climate with increasingly degraded soils (Lal, 2009; Khoshgoftarmanesh et al., 2010), methods to improve micronutrient use efficiency in crops and increase the concentration or bioavailability of essential elements within edible crop tissues may aid in combatting hidden hunger (alongside other strategies such as dietary diversification, fertilisation strategies and chemical fortification) (White and Broadley, 2009). On the other side of the coin, growth of crops on soils previously considered unproductive has been suggested as a way of boosting yields. Micronutrient use efficiency mechanisms could be applied to this area in a number of ways: e.g. phytoremediation (the use of plants to remove or immobilise contaminants and bring lands back into production), growth of tolerant crops for non-food applications (e.g. biofuels crops which are bred to grow on toxic conditions, freeing croplands for food production) and phytoextraction (using plants to remediate and extract toxic elements, potentially allowing the reapplication of extracted elements for use elsewhere) (Ali, Khan and Sajad, 2013). Furthermore, even within developed countries micronutrient deficiencies and toxicities occur; soil is a heterogeneous matrix and the availability of all elements can vary significantly within a single field (Hinsinger et al., 2009). As such it has been argued that improving micronutrient use efficiencies in particular will never truly be solved with soil amendments (e.g. fertilisers, liming or tilling); a combination of improved crop genotypes and management practices will need to be implemented in order to meet the demands of the 21st century and beyond (Goulding, Jarvis and Whitmore, 2008).

With these issues in mind, the current research has focused on micronutrient use efficiency in *B. napus*. Not only is it one of the world's most important oilseed crops, it is a member of the large and diverse *Brassicaceae*. Members of this family include many vegetable species such as cabbage (*B. oleracea*), turnip (*B. rapa*) and swede (*B. napus*), which would be good targets for biofortification strategies albeit with only limited application (e.g. White et al., (2018)). Alternately, *B. napus* could be used as a biofuel (Milazzo et al., 2013); work investigating micronutrient use efficiency could also produce plants tolerant to soil contamination (either by excluding the contaminants from uptake or through breeding for improved tolerance) and therefore limit the need to grow non-food crops on more fertile lands. Indeed it has often been emphasised that for phytoremediation to be feasible it needs to not only help remediate soils but produce an economic return (due to the long timescales involved in comparison to conventional remediation/extraction technologies)(Robinson et al., 2003). Further, work on micronutrient use efficiency may improve yields of *B. napus* grown in general. Finally, *B. napus* was utilised as a 'model crop' species. In addition to the direct applications of the work being

performed it: demonstrates proof of concept for using Associative Transcriptomics (AT) for micronutrient investigations; provides another demonstration for how the complex genetics of a polyploid crop can be investigated for micronutrient use efficiency (with the potential to expand into other crops, see 1.2.7 for further details on AT) and allows the exploitation of existing genetic resources in *A. thaliana* for quick validation of the genes discovered (as *A. thaliana* is a member of the *Brassicaceae*, allowing orthologous genes to be tested for their role in nutrient concentration, see 1.2.6). Note, there has been an emphasis on ‘breeding better crops’ as part of the current research: this is a consequence of current public opposition to genetic modification (GM) technologies and the stance of the United Kingdom/European Union on GM crops (ECJ, 2018).

1.2.5 Elements of interest:

The current research has primarily focused on the elements within the ionome referred to as micronutrients, specifically Molybdenum (Mo), Manganese (Mn), Copper (Cu) and Zinc (Zn). These elements were selected because across both the seed and leaf datasets analysed as part of the wider RIPR project (2.1.1) they were measured in both datasets (having an average concentration greater than the limit of detection and a good percentage recovery from digestion, >85%). They were not being studied as part of other projects and represented a specific problem in *B. napus* (individual descriptions to follow). However, two other additional elements were included in the current research; Sulfur (S) and Cadmium (Cd). These two elements were included because of their relationship to one or more of the micronutrients initially under assessment. Mo research (particularly in seed) quickly highlighted a close relationship to S nutritional status, eventually leading into a study encompassing S, the wider seed ionome and glucosinolates (GSL, plant defence compounds, see 4). Furthermore, relatively recent research has highlighted links between Mo (one of the elements in which *B. napus* is particularly susceptible to deficiency) and a number of elements (including Cu, Zn and Mn) in *B. napus* (Maillard et al., 2016b). Cd was included pre-emptively; it is well known that Cd can interfere with the uptake and biological functions of many divalent cations, including those which were under investigation as part of this research i.e. Zn, Mn and Cu (Choppala et al., 2014). Considering that the function and specificities of many of the candidates coming from the Zn, Mn and Cu analyses were yet to be determined and the known problems surrounding Cd toxicity (food being the primary source of Cd exposure to the non-smoking human population (Clemens et al., 2013)), it seemed necessary to investigate it simultaneously. If concentration mechanisms were shared between Cd and any of the other elements under investigation it could potentially jeopardise their utility for breeding nutrient efficient food crops, but may have advantages in other applications, e.g. phytoremediation.

Considering the diverse range of elements under assessment as part of this research, and in order to understand the analyses which have taken place, it is important to have a basic understanding of the biological functionality of each element. Therefore, the next section will detail in brief the biological 'life story' of each element under assessment: how they are obtained by the plant from the environment and used within its tissues. Through this analysis it will be easy to understand the range of deficiency and toxicity symptoms plants display when exposed to extremes in bioavailability. Each section will conclude with a brief summary of the work so far in *B. napus* and related species.

1.2.5.1 Molybdenum (Mo)

As with all essential elements, plant accumulation of Mo is dependent on its bioavailability in the soil environment. Mo is thought to be available to plants as the molybdate oxyanion, MoO_4^{2-} , and is therefore most bioavailable under alkaline conditions ($\text{pH} > 6.5$) (Alloway (Eds), 2008). As such, Mo deficiency occurs on highly leached and weathered acidic soils.

Consequently Mo is often deficient on older/weathered soils (those $> 10^6$ years old) (Jones et al., 2013) such as those in Australia (where it is the second most common deficiency after Zn) and China (where Mo deficiency affects 47% of agricultural soils), as well as in acidic sandy soils, such as in Africa (Alloway (Eds), 2008). The symptoms of Mo deficiency are similar to nitrogen deficiency symptoms including leaf margin chlorosis, leaf deformation (including 'whiptail', where the leaves are long and narrow), stunting and destruction of reproductive tissues (Arnon and Stout, 1939; Marschner, 1995b). Toxicity symptoms for Mo in crops are rarely observed as a consequence of the large range in critical deficiency and toxicity levels for Mo (up to a factor of 10^4) (Marschner, 1995b). The symptoms of Mo toxicity are relatively mild in plants, generally resulting in leaf discolouration (Kaiser et al., 2005). However, Mo is toxic to ruminants (causing a disease known as Molybdenosis, which is effectively Cu deficiency). Therefore the concentration within livestock feeds needs to be carefully controlled and generally kept below 2 mg/kg DW of forage (Kaiser et al., 2005; Alloway, 2013).

Within plants, Mo is required as a cofactor in four key enzymes: nitrate reductase (integral for nitrate assimilation, explaining many Mo deficiency symptoms); xanthine dehydrogenase (required in the oxidative metabolism of purines); aldehyde oxidase (important for ABA biosynthesis) and sulfite oxidase (essential for sulfite detoxification/ the catabolism of sulfur containing amino acids) (Bittner, 2014). To perform these roles Mo requires biological activation through incorporation into a pterin complex (the molybdopterin complex/ molybdenum cofactor, MoCo), which is responsible for the catalytic activity of all Mo enzymes within plant tissues (Mendel and Schwarz, 2011). Chemically, molybdate is very similar to

phosphate and sulfate; all are available for plant uptake as divalent anions (Marschner, 1995b). It was this chemical similarity to sulfate which led to the identification of the first molybdate transporter in plants: MOT1 is a member of the large sulfate carrier family, "SULTR" (Tejada-Jiménez, 2007; Tomatsu et al., 2007; Baxter et al., 2008). Unfortunately, conflicting accounts concerning the subcellular localisation of MOT1 have confounded the resolution of its specific functionality (Tomatsu et al., 2007; Baxter et al., 2008). In contrast, the second Mo transporter, MOT2, was shown to have vacuolar localisation in *A. thaliana* (Gasber et al., 2011). *MOT2* transcripts were shown to accumulate within senescing leaves whilst *mot2* T-DNA lines showed an increase of Mo within the leaves and subsequent decrease in seeds (Gasber et al., 2011). Nevertheless it is still unknown how plants take up molybdate from the soil, how expression of the genes involved in Mo accumulation/concentration are regulated or whether molybdate is chelated within plants for storage/movement (Bittner, 2014).

It has previously been shown that *Brassicaceae* and legumes are particularly susceptible to Mo deficiency. Within legumes this is presumably a consequence of the unique role that Mo plays within the symbiotic N fixing bacteria of root nodules, fixing atmospheric N₂ to NH₃ (Kneip et al., 2007; Bittner, 2014). Perhaps the large demand for S in *Brassicaceae*, which is thought to be a consequence of producing large amounts of S rich secondary metabolites (i.e. glucosinolate/GSL defence compounds) (Zhao et al., 1997, 1993), leads to an increased uptake of Mo due to their chemical similarities. It is well documented in *Brassicaceae* that there is an antagonistic relationship between S and Mo (e.g. under S limitation there is an increase in Mo accumulation (Schiavon et al., 2012; Maillard et al., 2016a), whilst under increased S there is a decrease in Mo concentration (Pasricha and Randhawa, 1972; Balík et al., 2006)). Both Cu (Billard et al., 2014) and Zn (Billard et al., 2015) deficiencies have been shown to increase Mo uptake in *B. napus*. It was theorised that this relationship could be a consequence of the dependence of a key enzyme (CNX1) in MoCo biosynthesis on Cu and Zn, which may then regulate expression of *MOT1* (Kuper et al., 2004; Llamas et al., 2006; Billard et al., 2014, 2015). Furthermore, it has been shown that a number of element deficiencies (S, Fe, Cu, Zn, Mn and B) cause an increase in Mo uptake in *B. napus* (Maillard et al., 2016b). This study also suggested the link between Mo activation and the susceptibility to multiple element deficiencies (Maillard et al., 2016b). However, an alternative hypothesis was also suggested: the expression of the sulfate transporter *BnaSultr1.1*, was increased under S, Fe, Mn, Mo and B deprivation. It is known that Mo and S can both utilise this transporter (Fitzpatrick, Tyerman and Kaiser, 2008), indicating that perhaps the effects on Mo are caused indirectly through a disturbance in S metabolism (Maillard et al., 2016b).

1.2.5.2 Manganese (Mn)

Unlike Mo, Mn is most bioavailable to plants under acidic soil conditions (pH < 6) (Alloway (Eds), 2008). It has a variable oxidation state (from Mn¹⁺ to Mn⁷⁺) but is most commonly found in plants as Mn²⁺ (Marschner, 1995b). Mn deficiency is the most common micronutrient deficiency in UK fields, occurring on alkaline, organic rich, sandy and calcareous soils (Alloway (Eds), 2008). The symptoms of Mn deficiency include interveinal chlorosis, tissue necrosis and increased susceptibility to freezing damage. However, Mn deficiency is relatively easy to correct, either through seed coating or foliar sprays (Brennan and Bolland, 2011). Since Mn bioavailability within the soil is heavily dependent on pH and redox status, plants can be subject to extremes in bioavailability. Waterlogging, extremes in temperature and soil compaction significantly increase the availability of Mn, often causing Mn toxicity to develop (Fernando and Lynch, 2015). Toxicity symptoms for Mn are species dependent (El-Jaoual and Cox, 1998), in *B. napus* the symptoms include interveinal and leaf margin chlorosis, leaf shape distortion ('cupping') and necrosis (Moroni, Scott and Wratten, 2003). However, as with deficiency the toxicity of Mn to crops can be managed; either with liming of acidic soils or improved drainage (Brennan and Bolland, 2015, 2011).

The uptake, transport, storage and role of Mn in plants has been well characterised. Mn plays a fundamental role within plant proteins in one of two ways. Firstly it is known to act as a catalytically active metal. The most well-known examples of this are found within the oxygen evolving complex in PSII (catalysing the photolysis of water during photosynthesis) (Barber, 2009) and also as a cofactor within superoxide dismutase (MnSOD) helping to prevent oxidative damage (Allen et al., 2007; Bowler and Slooten, 1991). Secondly, Mn can act as an enzyme activator in numerous enzymes involved in a variety of reactions e.g. oxidation-reduction, hydrolytic and decarboxylation reactions (Marschner, 1995b). However, in many instances Mn can be replaced within its role as an enzyme activator by other metals (such as Mg) (Houtz, Nable and Cheniae, 1988; Christeller, 1981; Jordan and Ogren, 1981; Wildner and Henkel, 1979). It is not just in its role as an enzyme activator that Mn has to compete with other elements; it has been shown that high concentrations of many elements (e.g. Fe, Mg, Ca or P) can induce Mn deficiency, and vice versa (Socha and Guerinot, 2014; Lynch and St.Clair, 2004). This competition is thought to be a result of chemical similarities between the elements (e.g. ionic radius and ligand binding capacity) (Marschner, 1995b). These close chemical similarities are also thought to be responsible for the broad selectivity of most Mn transporters (many of which were previously identified as Fe or Ca transporters, e.g. IRT1 (Vert and Grotz, 2002), AtNramp3 (Thomine et al., 2000), OsYSL2 (Koike et al., 2004), CAX2 (Pittman et al., 2004) and ECA1 (Wu et al., 2002)) (Pittman, 2005; Socha and Guerinot, 2014). One

exception is AtMTP11 which has been identified as a Mn 'specific' transporter (Delhaize et al., 2003, 2007). However, differences in Mn efficiency have been observed in wheat (Jiang, 2006) and barley (Pedas et al., 2008), suggesting that there may be more Mn specific pathways to uncover. Once inside the plant, there appears to be a number of potential ways Mn may be stored. Storage in the apoplast and vacuole has been suggested (Führs et al., 2010; Hirschi et al., 2000; Delhaize et al., 2003), as has chelation with organic acid complexes (Fernando et al., 2010, 2012) and mobilisation within the endoplasmic reticulum/secretory pathway (Peiter et al., 2007; Wu et al., 2002).

In *B. napus* Mn specific research has been limited. When corrected for dry weight, Zn deficient *B. napus* were found to have much higher Mn than control plants (Billard et al., 2015). This increase was attributed to favouring enzyme isoforms with different metal cofactors, e.g. switching to Mn-SOD instead of Zn-Cu SOD (Abreu and Cabelli, 2010; Billard et al., 2015). Similarly, Mg deficient *B. napus* was found to over-accumulate Mn and there appeared to be an upregulation of some Mg dependent proteins capable of utilising Mn (albeit less efficiently)(Billard et al., 2016). Mn deficiency is not the only issue; Mn toxicity is a particular problem for *B. napus* on acid soils. However, natural variation in susceptibility to Mn toxicity is known to exist amongst *B. napus* cultivars (Moroni, Scott and Wratten, 2003) and it was theorised that Mn tolerance was genetically controlled by one locus in *B. napus* (McVittie et al., 2011). QTL analysis supports this hypothesis; a major locus on chromosome A9 was observed, with an orthologue of an *A. thaliana* cation efflux facilitator (*MTP8*, an Mn/Fe transporter) relatively close by (Raman et al., 2017). Other research has linked Mn to Hg accumulation in *B. napus*; it was found that Hg uptake kinetics mimicked those of a low affinity transporter (Esteban, Deza and Zornoza, 2013). It was hypothesised that this might be a Mn transporter since Mn appeared to compete with Hg for uptake (Esteban, Deza and Zornoza, 2013). This could have implications for the bioremedial applications of *Brassicac*s.

1.2.5.3 Copper (Cu)

Just like Mn, Cu has a variable oxidation state; present as both the unstable Cu^+ and the more stable Cu^{2+} , and it is the cycling between these two states that forms the basis of its biological role within plants (Marschner, 1995b). Within the soil environment, most Cu is complexed with organic matter. As such availability is dependent on the pH (lower pH increases availability), redox potential and presence of other competing ions (e.g. Fe, Mn or Zn) (Alloway, 2013; Adrees et al., 2015). Consequently, Cu deficiency is common on sandy, calcareous, weathered and organic rich soils (Alloway (Eds), 2008). Cu deficiency can be a problem in *B. napus* under high N and P fertilisation regimes and where there are high concentrations of Fe, Mn and Zn

(Alloway (Eds), 2008). Cu deficiency symptoms include decreased growth rate, chlorosis of young leaves, necrosis of apical meristems and wilting (Marschner, 1995b). Such deficiencies are important for human nutrition; Cu is an essential element within the human diet (alongside Fe, Zn, Ca, Mg, Se and I) (White and Broadley, 2009). On the other hand Cu toxicity can also be a problem and is primarily a consequence of anthropogenic activities such as overuse of pesticides, fungicides and fertilisers (including slurries), alongside industrial and urban activities which have all led to an increase in Cu in the soil environment (Adrees et al., 2015; Marschner, 1995b). In plants, Cu toxicity leads to reduced yields and tissue chlorosis (Burkhead et al., 2009), while in humans and animals it is rare for toxicity to occur as a consequence of diet/plant consumption (Gupta and Gupta, 1998), i.e. Cu toxicity is more problematic in plants than animals (Adrees et al., 2015).

The uptake, transport, storage and regulation of Cu concentration has been widely investigated. This is a consequence of the multitude of roles Cu plays within plants ranging from photosynthesis and respiration, to cell wall remodelling, ROS metabolism and ethylene signalling (Burkhead et al., 2009). Its role in all of these processes is a consequence of its ability to act as an oxidising or reducing agent (Hänsch and Mendel, 2009). Cu proteins can be split into two groups: those which function in electron transfer, e.g. plastocyanin, and those which act as oxidases, e.g. cytochrome c oxidase (Marschner, 1995b). It is thought plants can take up Cu in a number of ways: ZIP transporters are known to be upregulated in roots under Cu deficiency and are thought to be involved in low affinity uptake of Cu^{2+} (Wintz et al., 2003), however their role has yet to be proven *in vivo* (Peñarrubia et al., 2015). A high affinity uptake system has been suggested that involves the reduction of Cu^{2+} to Cu^+ in the rhizosphere (Bernal et al., 2012) and uptake via CTR-Like transporters, known as COPTs (Sancenon et al., 2003). Alongside the increased transcription of Cu transporters (Wintz et al., 2003; Yamasaki et al., 2009), another deficiency response has been suggested. A transcription factor known as SPL7 is thought to be involved in sensing and signalling Cu deficiency via miRNAs such as miR398 (Yamasaki et al., 2007, 2009; Abdel-Ghany and Pilon, 2008), leading to a down regulation in Cu dependent proteins (being replaced by Fe containing proteins of similar functionality), conserving the remaining Cu for plastocyanin and maintaining photosynthesis (Bernal et al., 2012; Araki et al., 2018). Once within the cytoplasm, Cu is thought to be immediately bound to either metallothioneins (preventing ROS damage etc. (Guo, Bundithya and Goldsbrough, 2003; Guo, Meenam and Goldsbrough, 2008)) or to metallochaperones (responsible for delivering Cu to where it is required (O'Halloran and Culotta, 2000; Puig et al., 2007b)). Any excess Cu is removed from the cytoplasm by P1B-ATPases for detoxification or to the plastids and secretory pathway; (Cu^+ HMA5-8) and (Cu^{2+} HMA1-4)(Puig et al., 2007a).

NRAMPs may also have a role in Cu^{2+} intracellular movement but this has not been shown *in planta* (Liu et al., 1997). The mechanisms surrounding xylem and phloem loading of Cu in dicots are poorly understood; Cu is likely chelated, e.g. with nicotianamine (Curie et al., 2009), and potentially loaded as a complex into the phloem by YSL proteins (as observed in rice (Zheng et al., 2012)) (Printz et al., 2016). There are many reported mechanisms for tolerating excess Cu: e.g. HMA5 detoxification (Andres-Colas et al., 2006; Kobayashi et al., 2008); SIZ1 regulation of YSLs (Chen et al., 2011; Zhan et al., 2018) and Clp protease degradation of PAA2/HMA8 (Tapken et al., 2015).

Cu deficiency or excess has negative effects on seed quality in *B. napus*; reducing seed size, pod number, oil, concentration of protein and carbohydrates, whilst increasing the phenol and non-protein N content of seeds (Khurana, Singh and Chatterjee, 2006). Further, excess Cu was found to be much more toxic to *B. napus* than excess Zn (even when accounting for the relative amounts required of each). No difference could be observed between specific toxic effects of the two elements but Cu seemed to be retained in the roots and lower leaves, potentially responsible for early leaf abscission (which would have major implications for its use in remedial purposes) (Ivanova, Kholodova and Kuznetsov, 2010). Later research implied a role for glutathione (GSH) chelation within the roots of *B. napus* for Cu but not Zn (Zlobin, Kartashov and Shpakovski, 2017). Further interesting interactive effects were observed when *B. napus* was subject to individual and combined Cu/Cd treatments (Mwamba et al., 2016). It was found that Cu was in fact more phytotoxic than Cd, but excess Cd enhanced Cu uptake whilst excess Cu prevented Cd uptake (Mwamba et al., 2016). The short-term effects of Cu excess have also been evaluated in *B. napus*. Early leaf wilting, a reduction in photosynthetic pigments and a change in expression of 10 genes involved in detoxification and concentration were all observed (Zlobin et al., 2015). Interestingly the highest increase in gene expression was in a homologue of *AtNRAMP4* in leaves; it was hypothesised this was a consequence of perturbed Mn and Fe concentration under Cu excess, with the plant remobilising Mn/Fe via NRAMP4 to maintain PSII activity (Zlobin et al., 2015). Further research has investigated *B. napus* in Cu phytoremediation, utilising EDTA (Habiba et al., 2014) and citric acid (Zaheer et al., 2015) as a means of improving Cu uptake. As previously mentioned for Mo, it was found that Cu deficiency in *B. napus* enhanced Mo accumulation (Billard et al., 2014). However it also appeared that under Cu deficiency plants were able to remobilise Cu from older leaves, as well as increase the expression of Cu transporters in the roots (COPT2) and leaves (HMA1) (Billard et al., 2014). These results were in line with broader experiments looking at leaf element remobilisation in a range of species, including *B. napus* where Cu appeared to be remobilised from older tissues during senescence, irrespective of deficiency conditions (Maillard et al.,

2015). Further, it was found that 33 proteins were differentially regulated under Cu deficiency, almost half of which were localised within the chloroplast (fitting with the well-known role of Cu within electron transport chains) (Billard et al., 2014). Interestingly, it was previously shown that miR398 increased in the phloem sap of *B. napus* under Cu starvation, one of the miRNAs involved in the SPL7 Cu deficiency response in *A. thaliana* (Buhtz et al., 2008).

1.2.5.4 Zinc (Zn)

Zn is the micronutrient most commonly deficient in crops (Alloway, 2009). It is available to plants in a number of forms: as the cations Zn^{2+} and $ZnOH^+$, as well as within soluble organic complexes. As such, with increasing pH, organic matter and high mineral contents such as $CaCO_3$, Zn becomes less bioavailable (Alloway, 2009). The presence of other elements is also known to inhibit Zn uptake (e.g. Mg, Mn or Cd). As with most micronutrient deficiencies, the typical symptoms of Zn deficiency are chlorosis and stunting, however unlike other micronutrients these symptoms are present in both older and younger leaves (Alloway, 2013). Since Zn is an essential element in human and animal nutrition, improving Zn accumulation in the edible portion of crops has been a primary aim of both fertilisation and crop breeding strategies (White and Broadley, 2009; Alloway (Eds), 2008). Zn is found to be toxic at high concentrations: causing stunting, chlorosis in young leaves and displacing other elements (e.g. replacing Mg within chlorophyll) (Küpper and Andresen, 2016). As with Cu, the main sources of Zn soil contamination vary from natural to anthropogenic.

The role Zn plays in plants is varied. It is found within all 6 enzyme classes; it has structural roles (in protein folding, from tertiary structures to protein aggregation and interactions with other biological components, e.g. other proteins, DNA/RNA and lipids) and is involved in enzyme catalytic activities and can have simultaneous roles in structural, regulatory and catalytic function of proteins (Maret, 2012; Auld, 2001; Maret, 2005; NC-IUBMB, 2018; Broadley et al., 2007). The most common type of Zn binding proteins are the Zn finger domain containing proteins which can regulate gene expression in a multitude of ways (e.g. DNA/RNA binding, promotion of chromatin modification or RNA metabolism) (Klug, 1999; Englbrecht, Schoof and Böhm, 2004; Broadley et al., 2007). Unlike Cu or Mn, Zn only exists as Zn^{2+} in plants, thus explaining its role in structure and catalysis rather than redox dependent activities (Ricachenevsky et al., 2015). Zn^{2+} uptake from the soil environment is poorly understood. IRT1, which is a member of the ZIP transporter family, is thought to be able to transport a range of divalent cations (Cd^{2+} , Co^{2+} , Fe^{2+} , Mn^{2+} and Zn^{2+}) (Korshunova et al., 1999; Connolly, 2002), has correct localisation in the root and displays metal dependent post-translational regulation (Shin et al., 2013; Connolly, 2002). It was previously proposed that other members of the ZIP

family were responsible for uptake of Zn²⁺ from the soil (Grotz et al., 1998), however *A. thaliana* T-DNA lines of two of the four most highly expressed root ZIP transporters suggested they were important for root-to-shoot Mn (and possibly Zn) transport (Milner et al., 2013). Other transporters with suggested roles in Zn²⁺ concentration and translocation *in planta* include: CDFs/MTPs (e.g. MTP1 is thought to be involved in vacuole sequestration of Zn (Kawachi et al., 2009; Desbrosses-Fonrouge et al., 2005)); HMAs (e.g. HMA2 and 4 are thought to be involved in root-to-shoot translocation of Zn and Cd (Hussain et al., 2004)); PCRs (e.g. PCR2 is thought to play a role in Zn efflux from the roots and knock-out lines display sensitivity to both Zn deficiency and excess (Song et al., 2010)) and VITs (although direct evidence for a role in Zn transport has only been observed in rice (Zhang et al., 2012), i.e. in *A. thaliana* VIT1 is known to be important for Fe localisation in seeds (Kim et al., 2006b)). Transport as chelated Zn has been suggested via YSLs (e.g. Waters et al., (2006) and ZIFLs (e.g. Haydon et al., (2012)).

The genes responsible for regulating Zn deficiency are starting to be uncovered; transcription factors specific to Zn deficiency have been tested, e.g. bZIP19 and bZIP23 (Assuncao et al., 2010), the identification of common regulators between different element deficiencies has begun, e.g. PHR1 (Briat et al., 2015), and work utilising natural variation in Zn, gene expression and various other traits, e.g. ZIP4 and IRT3 in *A. thaliana* (Campos et al., 2017), are all contributing to completing our understanding of Zn regulation. Conversely, research into regulating extremes in Zn concentration has focused on plants known to accumulate excessive Zn concentrations (known as hyperaccumulators), comparing them to non-accumulator species (e.g. van de Mortel et al., 2006), different hyperaccumulators, (e.g. Mishra, Mishra and Küpper, 2017) and different populations of the same hyperaccumulators (Schvartzman et al., 2018). Research with hyperaccumulators has highlighted: upregulation/duplication of transporters, e.g. CDFs (Shahzad et al., 2010), HMAs (Hanikenne et al., 2008) and NRAMPs (Oomen et al., 2009); various metal binding strategies, e.g. non-thiol ligands (Kozhevnikova et al., 2014), nicotianamine (Deinlein et al., 2012) and organic acids (Schneider et al., 2013); and sequestration of excess Zn within the cell wall, vacuole or trichomes, (e.g. Kupper et al., 2000).

Within the *Brassica* genus there has been a lot of research into Zn accumulation. This is a consequence of its role as a nutrient in leafy vegetables and soil contaminant. For example, leafy brassicas have been promoted as an alternative source of Zn from the usual food sources (such as grains, fruit and tubers) as they have greater Zn concentration and lower phytate (which limits the availability of Zn for absorption within the intestines) (White et al., 2018). The investigation of *Brassic*as for phytoremediation has been widely reported (Coolong and Randle, 2003; Soriano and Fereres, 2003; Gisbert et al., 2006; Hamlin and Barker, 2006, 2008;

Yu et al., 2012; Mourato et al., 2015; Podar, Ramsey and Hutchings, 2004; Purakayastha et al., 2008; Belouchrani et al., 2016; Ebbs and Kochian, 1997). Indeed one of the most famous hyperaccumulators (*Noccaea caerulescens*, basionym *Thlaspi caerulescens*) is a member of the *Brassicaceae* (Lasat, 2002; Verbruggen, Hermans and Schat, 2009b). *Brassicaceae* are considered to be promising for phytoremediation of metals due to their relatively high biomass, although the induction of other mineral deficiencies (namely Mn and Fe) by excessive Zn accumulation could limit its application (Ebbs and Kochian, 1997). As previously mentioned, experiments looking into the effects of Zn deficiency on *B. napus* have highlighted poor Zn remobilisation (Billard et al., 2015; Maillard et al., 2015), while those looking at excess Zn have found that GSH does not seem to play a role in tolerating excess Zn within the roots of *B. napus* (Zlobin, Kartashov and Shpakovski, 2017).

1.2.5.5 Sulfur (S)

As a macronutrient, S is required in much larger amounts than the elements previously discussed. Uptake of S from the environment occurs in a number of ways (e.g. accumulation from atmospheric sulfur dioxide (SO₂) or hydrogen sulfide (H₂S), as well as uptake of organic S, e.g. amino acids) however the main form plants take up is sulfate (SO₄²⁻) (Kopriva, 2015). As SO₄²⁻ is the predominant form extracted by the roots from the soil, 'S' is most bioavailable to plants at a pH >6 (similar to Mo, 1.2.5.1) (Curtin and Syers, 1990). Until relatively recently, S deficiency was comparatively unknown in cropping systems due to the deposition of atmospheric S from industrial pollution (Bristow and Garwood, 1984) and the use of S rich fertilisers (i.e. fertilisers which used sulfuric acid as part of the production process) (Kopriva, 2015). The symptoms of S deficiency include yellowing of leaves leading to chlorosis and decreased biomass, alongside reduced protein metabolism/disruption in N metabolism (Marschner, 1995a). On the other hand, S toxicity is very rarely observed in crops but is a problem in lower nutrient input ecosystems exposed to high levels of H₂S and SO₂ pollution, such as forests (Maathuis, 2009; Nakamura et al., 2009). S toxicity results in chlorosis and interveinal necrosis (Chandra and Pandey, 2016; Lee et al., 2017).

The main pathway of S uptake in the plant is via sulfate transporters in the roots. SULTRs are a large sulfate/proton co-transporter family divided into a number of groups (c.f. group 5 SULTR in Mo, 1.2.5.1) (Takahashi et al., 2012). Group 1 transporters are responsible for high affinity sulfate transport, e.g. uptake of sulfate from the soil (Rouached et al., 2008; Yoshimoto et al., 2007, 2018); group 2 are low affinity transporters generally associated with root-to-shoot sulfate transport (Takahashi et al., 2000); group 3 is responsible for transporting sulfate in plastids (Takahashi et al., 2000) while group 4 is responsible for exporting sulfate from the

vacuole (Kataoka et al., 2004; Takahashi et al., 2012; Gigolashvili and Kopriva, 2014; Kopriva, 2015). S is generally highly mobile within the plant and it is theorised that it can be transported throughout the plant in a number of different forms other than sulfate, e.g. as GSH, flavanoids and amino acids (for a review see Gigolashvili and Kopriva, (2014)). Within plant cells however, sulfate which is not assimilated is stored within the vacuole (although the transporter responsible for this is yet to be uncovered). The assimilation of S is generally split into two pathways: primary and secondary S metabolism (**Figure 1.2.5.a**). Primary metabolism involves the gradual reduction of sulfate before becoming incorporated into cysteine. The initial step in sulfate activation requires ATP sulfurylase to adenylate the sulfate (forming adenosine 5'-phosphosulfate/APS), in the cytosol or plastid. This is a branching point between primary and secondary S metabolism; APS can either be reduced (via APS reductase and sulfite reductase to yield sulfide in plastids, which can then be incorporated into O-acetylserine (OAS) via OAS thiollyase to yield cysteine, either in the plastids or cytosol) or phosphorylated (via APS kinase) forming 3'-phosphoadenosine 5'-phosphosulfate (PAPS) which acts as a sulfate donor in many sulfation reactions in the cytosol (i.e. secondary metabolism) (Mugford et al., 2011; Koprivova and Kopriva, 2016). Atmospheric S accumulation can occur, as SO₂ forms sulfite within the leaf aqueous solution when soil S is limiting via leaf stomata, whilst H₂S is assimilated as sulfide (Nakamura et al., 2009).

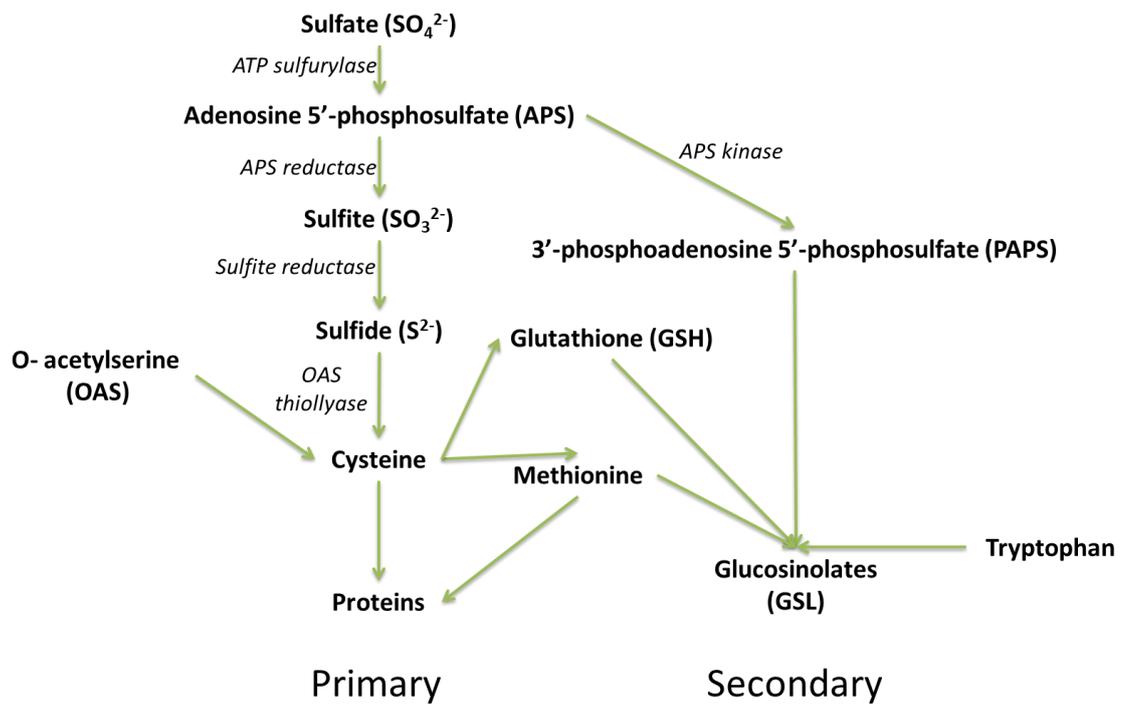


Figure 1.2.5.a Schematic representation of primary and secondary S assimilation (adapted from Mugford et al., (2011) and Koprivova and Kopriva, (2016)).

Arrows indicate reaction steps between each metabolite (given in bold) whilst key enzymes are indicated in italics. Primary and secondary metabolisms have been split, with glucosinolates given as an example of a secondary metabolite (i.e. a metabolite not considered essential for plant survival).

Within the plant S is involved in a range of biological activities; namely through its incorporation into amino acids (cysteine and methionine), but also within glutathione/GSH (Noctor et al., 2012), as sulfolipids (Shimajima, 2011), as a constituent of vitamins and cofactors, via plant hormones (regulation through sulfation) and within numerous secondary metabolites (Takahashi et al., 2011) e.g. glucosinolates (Halkier and Gershenzon, 2006) and aliiinins (Jones et al., 2004). The molecular mechanisms behind the regulation of S assimilation are beginning to be uncovered (reviewed in Koprivova and Kopriva, (2014)). A number of transcription factors has been found which are involved in controlling sulfate uptake and assimilation. For example, mutants with reduced *SLIM1* expression were unable to induce *SULTR1;2* transcription under S deficiency (Maruyama-Nakashita et al., 2006); Post-transcriptional regulation has also been implicated in S regulation, for example miR395 induction by S deficiency and regulation of ATP sulfurylase and *SULTR2;1* (Kawashima et al., 2009, 2011). Further, protein-protein interactions have also been implicated. A good example of this is the cysteine synthase complex: a complex of Serine Acetyl-transferase (SAT) and OAS

thiollyase (OASTL), each regulated by the availability of the others substrate. Within the bi-enzyme complex only SAT is active while OASTL needs to dissociate to free the active site and produce cysteine. Sulfide stabilises the complex allowing the synthesis of OAS, while OAS causes the dissociation of the complex by competing with SAT for binding, allowing cysteine biosynthesis (Droux et al., 1998; Wirtz and Hell, 2006).

It is well known that *B. napus* is highly sensitive to S and N deficiencies and therefore requires high fertilisation rates (Zhao et al., 1997; Dubousset, Etienne and Avise, 2010; Zhao et al., 1993). As previously mentioned for Mo (see 1.2.5.1), it is thought that part of this increased demand in *Brassicacae* is the production of secondary S metabolites, namely the glucosinolates (GSL) (Zhao et al., 1997, 1993). Indeed there is evidence that transcription factors involved in the control of GSL biosynthesis are involved in regulating sulfate assimilation: transactivation assays with MYB28/29/76 (aliphatic GSL transcription factors) and MYB51/34/122 (aromatic GSL transcription factors) highlighted the potential regulation of APS kinase and ATP sulfurylase in *A. thaliana* (Yatusevich et al., 2010). In *B. napus* seeds there is such a strong correlation between seed S and seed GSL that S was originally used as an indirect measure of GSL content (Bloem, Haneklaus and Schnug, 2005). The amount of GSL within the seeds of *B. napus* has however been limited; after oil extraction the seeds of *B. napus* are made into a livestock feed/ protein rich 'cake' and, since GSL have anti-nutritional properties in livestock (Griffiths, Birch and Hillman, 1998), *B. napus* elite varieties were bred to limit the GSL content in seeds (Halkier and Gershenzon, 2006). This reduction in GSLs has been achieved namely through an apparent knock-out in the major aliphatic GSL transcription factor *Myb28/HAG1* (although it also appears to cause a reduction in leaf GSLs, with implications for plant defence) (Harper et al., 2012; Lu et al., 2014). Other research in *B. napus* on S has focused on salt stress and its effect on GSH biosynthesis (e.g. Ruiz and Blumwald, 2002), the interaction of S with metal stresses e.g. effect of miR395 (Zhang et al., 2013) or H₂S (Ali et al., 2014) on Cd, the genes behind sulfate homeostasis (namely Cysteine synthase) via AT analysis in *B. napus* (Koprivova et al., 2014) and modelling approaches to S deficiency (Brunel-Muguet et al., 2015).

1.2.5.6 Cadmium (Cd)

Unlike all of the elements previously discussed, Cd is not an essential element and is highly toxic to plants and animals. Cd is accumulated as Cd²⁺ as consequence of its chemical similarities to other elements (i.e. other divalent cations, namely its analogue Zn²⁺). Therefore, it is most bioavailable to plants under the same conditions that other divalent cations are available, i.e. slightly acidic soils (Alloway, 2013). Furthermore, it has been hypothesised that some of the toxic effects of Cd in plant cells may be a consequence of its replacement of

essential elements within proteins (e.g. upregulation of Zn uptake systems under Cd stress (Weber, Trampczynska and Clemens, 2006)). It is also thought to increase ROS damage as a consequence of depleting reserves of antioxidants, specifically glutathione/GSH through the production of phytochelatins (thiol rich peptides which bind metals) or direct binding to GSH (Schützendübel and Polle, 2002). Many of the specific effects of Cd toxicity are however poorly understood; in general the symptoms observed with Cd toxicity include growth inhibition, wilting and chlorosis (Gallego et al., 2012). It is found naturally in soils (commonly with Zn minerals) at around 0.1-1mg kg⁻¹ (Alloway, 2013), however human activities have increased Cd concentrations within soils, predominantly from P fertilisers, but also through atmospheric deposition from industrial processes, mining and sewage sludge application (Choppala et al., 2014; Alloway, 2013). One of the main issues of Cd contamination is that the levels required to induce toxicity symptoms in plants are higher than those required in humans (Gupta and Gupta, 1998; Alloway, 2013). One of the most famous and extreme examples of this was the break out of *Itai Itai* disease in Japan in the 1960s; a disease which caused severe pain and fractures, particularly in the long bones. It was subsequently found to be a form of osteomalacia (bone softening caused in this case by increased excretion of Ca in Cd damaged kidneys) by the detection of elevated Cd in urine samples. The outbreak of the disease was linked to the irrigation of rice fields with water which had been contaminated with Cd from mining activities (Nordberg, 2009). This case also illustrates the other issue of Cd contamination of food: it is not necessarily a one off exposure to contaminated foods but consistent, low level consumption, as it takes 15-20 years for consumed Cd to be removed from the human body (i.e. background levels of Cd in food needs to be carefully controlled) (Inaba et al., 2005; Alloway, 2013). However, it is important to note that Cd toxicity is not necessarily a factor of total concentration within soil, it is the bioavailability that determines plant uptake (with pH generally considered the most important factor)(Christensen, 1984). Crops differ significantly in their ability to take up Cd, as such Cd toxicity in human populations is avoidable with a diverse diet and suitable management practices (Alloway, 2013). The *Brassicaceae* are known to be relatively good at accumulating Cd which is unfortunate for food production but may have applications in bioremediation (i.e. use of plants to remove or stabilise Cd within contaminated soils) (Rizwan et al., 2018).

As Cd is not an essential nutrient, it has been hypothesised that Cd²⁺ uptake into plants is a consequence of indirect uptake through other nutrient transporters/channels (namely those of divalent cations, such as Ca²⁺, Cu²⁺, Fe²⁺, Mg²⁺ and Zn²⁺)(Clemens, 2006). Transporters and channels which have been implicated in Cd uptake into root cells include: ZIPs (e.g. AtIRT1 under Fe deficiency (Vert and Grotz, 2002)), NRAMPs (e.g. Cd/Mn accumulation in rice via

OsNRAMP5 (Sasaki et al., 2012)), and potentially Ca channels (e.g. tobacco transformed with wheat *LCT1* displayed Ca mitigated Cd tolerance (Antosiewicz and Hennig, 2004)). Cd which does not become bound to root cell walls (Krzyszowska, 2011) and enters the cells is thought to be immediately chelated (with GSH/phytochelatins) e.g. Cd coordination to S containing ligands in *B. juncea* (Salt et al., 1995). In addition to chelation, sequestration of Cd in the vacuole is thought to be another tolerance mechanism, e.g. AtHMA3 (Morel et al., 2009) or AtCAX2 and 4 (Korenkov et al., 2009). However, in some instances Cd can be remobilised from the vacuole, e.g. AtNRAMP3 and 4 (Verbruggen, Hermans and Schat, 2009a; Thomine et al., 2003; Oomen et al., 2009), and loaded into the xylem for long distance transport, e.g. AtHMA4 (Mills et al., 2005). This is often the primary distinction between tolerant plants and hyperaccumulators: tolerant plants will limit root-to-shoot translocation to minimise the cytotoxic effects of metals in the above ground biomass, whereas hyperaccumulators (i.e. those which accumulate concentrations of metals that are normally toxic) usually have efficient root-to-shoot transport, e.g. *N. caerulea* appears to overexpress *NRAMP3* and *4* (Oomen et al., 2009), whilst *A. halleri* appears to have higher *HMA4* expression in comparison to *A. thaliana* (Hanikenne et al., 2008). Once again, within the upper organs, Cd chelation and sequestration are thought to be essential for Cd tolerance.

Brassicaceae species (including *B. juncea*, *B. carinata*, *B. oleracea* and *B. napus*) have been suggested for use in phytoremediation purposes; they have high biomass, can be grown for non-food purposes and are generally tolerant of higher Cd concentrations than other crops (Rizwan et al., 2018). As such there is a relatively large body of recent research which has been performed to understand Cd accumulation (amongst other elements such as Zn, see 1.2.5.4) in various *Brassicaceae* species. It was previously found that exposure of *B. napus* to Cd stress causes an increase in concentration of Fe, Zn, Cu and P in roots (as well as Cd), and a reduction in K. Further, S was shown to increase with Cd exposure, which the authors attributed to a potential increase in GSH or phytochelatin biosynthesis (Larsson, Bornman and Asp, 1998). Follow on experiments highlighted that phytochelatin biosynthesis was playing a role in Cd tolerance in *B. napus* (Selvam and Wong, 2008) and explored the application of this species in phytoremediation (Grispen, Nelissen and Verkleij, 2006). Other experiments emphasized the importance of vacuolar and cell wall sequestration as a long term Cd tolerance mechanism in *B. napus* (Carrier, Baryla and Havaux, 2003). Comparison between the known accumulator *B. juncea* and *B. napus* highlighted that most Cd was retained within the roots of both species but that *B. juncea* had greater accumulation in the shoot and increased lipid content within the leaves (Nouairi et al., 2006). More recently a range of genome wide analyses has investigated Cd tolerance in *B. napus*: Meng et al., (2017) investigated NRAMP expression under Cd

exposure; Chen et al., (2018) performed a GWAS on Cd accumulation and found a number of *A. thaliana* homologues (*IRT1*, *NRAMP6*, *PCS1*, *PCS2* and *GSTs*); Zhang, Zhao and Yang, (2018) investigated Cd responsive ABC transporter genes; whilst Zhang et al., (2018) used the same methods to investigate Cd responsive RNA helicase genes and Li et al., (2018) found Cd responsive HMA transporters. All of this research has focused on the phytoremedial potential of *B. napus*, however it is important to note that the primary use of oil derived from this species is for human consumption (not biofuel). Therefore, careful analysis and regulation of cultivars which can tolerate or accumulate Cd would need to take place to prevent accidental contamination of the food chain. Indeed this is probably one of the most convincing arguments against the use of *B. napus* in phytoremedial applications (c.f. use of *B. napus* for phytostabilisation and biofuel production (Campbell et al., 2017)).

1.2.6 *B. napus* in general

As previously mentioned, *Brassica napus* (Oilseed rape) is a globally important vegetable oil crop. However, it is only within the past 50 years that it has become an economically important oil crop. Breeding to reduce antinutritional components, such as glucosinolates (GSL), and efforts to boost seed yields have resulted in Oilseed rape being the 3rd largest source of vegetable oil worldwide in 2017/18 (Allender and King, 2010; USDA, 2018). In addition to its role as an oil crop, *B. napus* is also an important vegetable crop, with crop types grown for their leaves (e.g. kale and fodders) and roots (e.g. swede) for both human and animal consumption. *B. napus* is an allotetraploid/amphidiploid, derived from the hybridisation of the diploid species *Brassica rapa* (contributing the 'A' genome) and *Brassica oleracea* (contributing the 'C' genome) which diverged from each other ~3.7 Mya (Inaba and Nishio, 2002) (**Figure 1.2.6.a**). It has been theorised that the hybridisation between *B. rapa* and *B. oleracea* took place during human cultivation, sometime less than 10,000 years ago (Trick et al., 2009b). As a further complication, extensive genome triplication has been detected in each of the diploid *Brassica* genomes, i.e. *B. napus* is a paleohexaploid (Lysak et al., 2005, 2007). Importantly, the model organism *A. thaliana* is a member of the same family as *B. napus* (the *Brassicaceae*, diverging ~20 Mya) and as such it is possible to trace orthologous genes between *B. napus* and *A. thaliana* due to their close phylogenetic relationship (Yang et al., 1999).

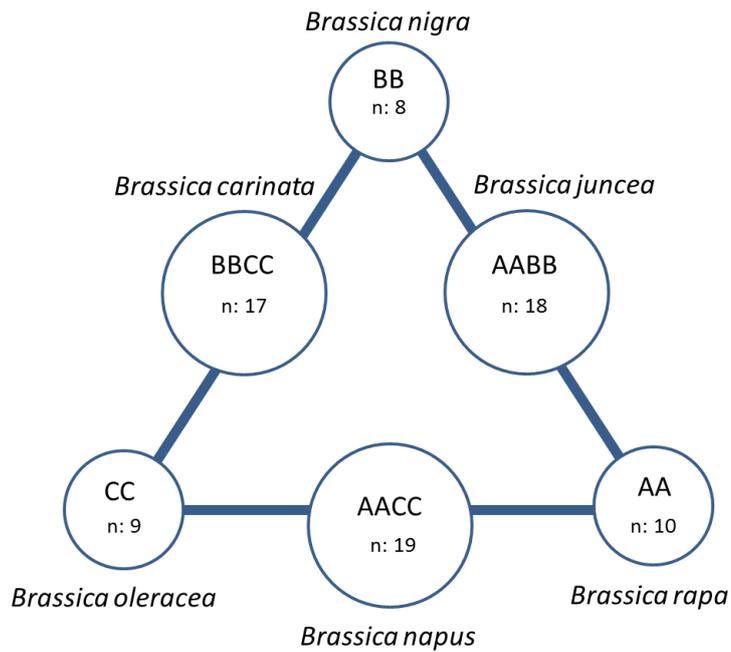


Figure 1.2.6.a The triangle of U: Brassica genome evolution/structure in accordance with the theory presented by (Nagaharu U., 1935).

Six Brassica species are portrayed: the three diploid species are shown (*Brassica rapa*; *Brassica nigra* and *Brassica oleracea*) as is their genome structure and chromosome number (i.e. AA (n=10); BB (n=8); CC

The large size (~1.2 Gb) and repetitive nature of the *B. napus* genome has made the production of a draft genome sequence for *B. napus* significantly more complicated (Chalhoub et al., 2014). For example, there can be up to six copies of a single orthologous gene from *A. thaliana* in the *B. napus* genome (giving a total of 12 different allelic variants) (Chalhoub et al., 2014). Furthermore, due to the amphidiploid nature of *B. napus*, many genes will have corresponding homoeologues (which have been defined as “genes or chromosomes in the same species that originated by speciation and were brought back together in the same genome by allopolyploidization” in Glover, Redestig and Dessimoz, (2016)) within each of the A and C sub-genomes. It has been estimated that the transcribed sequences of these homoeologues will vary ~3.5% of the time. Consequently any SNPs detected in such regions can appear to have multiple (>2) base calls (one from each sub-genome, see next sections, 1.2.7)(Trick et al., 2009b). The presence of these homoeologues is further complicated by the extensive homoeologous exchanges (HEs) that have been observed in *B. napus* (Bancroft et al., 2015; He et al., 2016; Chalhoub et al., 2014). Consequently, both SNP calling and the correct assignment of CDS gene models to sub-genomes is significantly more complicated in *B. napus* and other recently formed allopolyploids. The AT approach used in the current research mitigates such issues (see 1.2.7). Crucially, however, these HEs have been shown to segregate

widely in the germplasms used for *B. napus* in crop breeding (He et al., 2016). It has been theorised that these HEs have been placed under selective pressure during the breeding process (Chalhoub et al., 2014). For example, the low glucosinolate (GSL) phenotype (see section 1.2.5.5 for further information on GSL) is now thought to be a consequence of a HE, resulting in the substitution of a region on the C2 chromosome with a homoeologous region from the A genome. This replaces a functional copy of a major positive aliphatic GSL transcription factor (*myb28/HAG1*) with a non-functional one, significantly reducing the concentration of GSL within the seeds and leading to the development of the modern low GSL varieties in the past 50 years (Chalhoub et al., 2014; He et al., 2016). Furthermore, it has been shown that the majority of HEs involve substitutions of C genome sequences by A genome sequences (He et al., 2016). It had previously been theorised that this is a consequence of repeated interspecific crosses between *B. napus* and *B. rapa* during cultivation (Liu et al., 2016), however it may also be explained through the substitution of *B. napus* C genome sequences by the *B. rapa* derived A genome sequences in HEs (He et al., 2016).

1.2.7 Associative Transcriptomics (AT)

Associative Transcriptomics (AT) is a form of GWAS; using historical recombination events and linkage disequilibrium to identify loci associated with the trait under investigation. Therefore, like all GWAS studies, AT requires a large and genetically diverse panel of plants. In the current research, a diversity panel comprising 383 double haploid or inbred accessions of *B. napus* were exploited. The panel consisted of: spring oilseed rape (123 accessions), semi-winter oilseed rape (11 accessions), swede (27 accessions), kale (3 accessions), fodder (6 accessions), winter oilseed rape (169 accessions) and an unassigned group (44 accessions, i.e. crop type information unknown), for further details see 2.1.1 and **Figure 1.2.7.a**. As with all forms of GWAS, the use of a genetic diversity panel in AT can cause spurious marker-trait associations due to underlying population structure (Q) and relatedness (kinship, K) (Zhang et al., 2010a). To control for these associations a number of different approaches have been taken throughout the literature (reviewed in Price et al., (2010)). Population stratification is relatively easy to detect using Genomic Control methods. These methods estimate the inflation of p values caused by underlying population structure (i.e. genome wide inflation of association statistics) and are often used as a correction for underlying population structure (such as the methods used in the current study for GEM analysis, see 2.2.1 and Devlin and Roeder, (1999)). Other methods used to correct for underlying population structure infer genetic ancestry, e.g. Structure Association (SA)/model-based approaches or Principal Component Analyses (PCA)/non-model-based approaches. These methods split the panel into sub-population clusters and then investigate associations within the clusters (Price et al., 2010). The SA

approaches produce a Q matrix (an estimation of the relatedness for each genotype to each sub-population, where $K \geq 2$ populations) but are computationally intensive (e.g. STRUCTURE (Pritchard, Stephens and Donnelly, 2000) or ADMIXTURE (Alexander, Novembre and Lange, 2009)). The PCA approaches are computationally more efficient (e.g. EIGENSTRAT (Price et al., 2006)), using the top principal components as covariates to correct for population stratification, but do not produce a Q matrix. However, neither of these approaches correct for unequal relatedness (K) amongst individuals (Price et al., 2010). Mixed Modelling approaches on the other hand are able to account for both population structure (Q) and relatedness (K) using both random (a phenotypic covariance matrix which represents the sum of heritable and non-heritable random variation, the K matrix) and fixed effects (the functional genotype and Q matrix). The AT approach used in the current research combines all three methodologies: using a PSIKO (Population Structure Inference Using Kernel-PCA and Optimization (Popescu et al., 2014b)) derived Q matrix which utilises PCA and compressed mixed linear modelling (CMLM) through GAPIT (Genome Association and Prediction Integrated tool (Lipka et al., 2012)), which generates the K matrix automatically. The CMLM approach was used as older MLM approaches using TASSEL (Bradbury et al., (2007), see next paragraph) were computationally challenging with the large datasets under analysis (Zhang et al., 2010a).

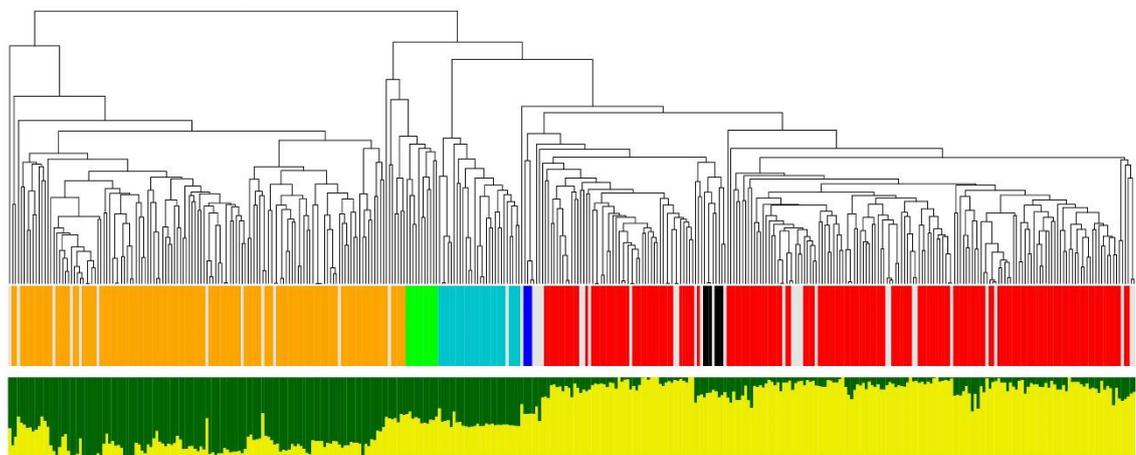


Figure 1.2.7.a The Renewable Industrial Products from Rapeseed (RIPR) population structure from Havlickova et al., (2018).

The top of the figure is a dendrogram produced from a distance matrix showing the relatedness of all 355 536 SNPs in the RIPR panel. In the middle are all the major crop types in the diversity panel colour coded (spring OSR = orange; semi-winter OSR = green; swede = light blue; kale = dark blue; fodder = black; red = winter OSR; grey = unassigned crop type) showing the expected clustering. On the bottom is the population structure as determined using PSIKO (Popescu et al., 2014b), showing a sub-population structure of $k=2$ with mixture across the panel.

However, unlike most other GWAS analyses, AT uses transcriptome data (mRNA-seq data); giving functional genotype data as variation in gene sequence (single nucleotide polymorphisms/SNP) and gene expression (gene expression markers/GEM) in RPKM (reads per kb per million aligned reads). Furthermore, until relatively recently there was no reference sequence for the *B. napus* genome (Chalhoub et al., 2014). As such, before AT could be applied as a technology, methods for calling the functional genotypes and the creation of a linkage map were required. To circumvent the lack of a reference sequence, Trick et al., (2009b) used 94,558 publicly available *Brassica* unigenes (assembled from 810,000 publicly available expressed sequence tags (ESTs)) (Trick et al., 2009a) as the reference sequence against mRNA-seq data from young leaves. SNP identification then exploited the publicly available MAQ software (Li, Ruan and Durbin, 2008) to develop an approach for the identification of SNPs in the absence of a genomic reference sequence (Trick et al., 2009b). Using this methodology, Bancroft et al., (2011), exploited this mRNA-seq approach to screen a small *B. napus* mapping population of 37 double haploid lines, using sequence variation and transcript abundance. SNP linkage maps were then constructed and used to align the *B. napus* genome against the *A. thaliana* reference sequence and the genome sequence assemblies of the two progenitor species, *B. rapa* and *B. oleracea* (Bancroft et al., 2011). The use of this mRNA-seq approach to measure transcript abundance was further validated in the work of Higgins et al., (2012), highlighting differential expression in homoeologous gene pairs after polyploidization (again emphasising the utility of mRNA-seq GWAS analysis for polyploids).

With these resources in place the first AT analysis was published by Harper et al., (2012), highlighting hypothesised deletions corresponding to the low glucosinolate (GSL) phenotype in *B. napus*. This AT approach generated the A and C genome pseudomolecules (sequence scaffolds representing the 19 chromosomes of *B. napus*) from the unigenes and linkage maps developed by Bancroft et al., (2011) but with manual interrogations of genome sequence scaffold assemblies to identify false assemblies (Harper et al., 2012). Another aspect in which the two methodologies differed was in how interhomoeolog polymorphisms (IHP: where sequences differ between homoeologous loci) were inferred (**Figure 1.2.7.b**). Automated SNP detection in polyploids can be confounded by the presence of these IHPs, confusing allelic variation with homoeologue variation. Harper et al., (2012) built upon the “curing” process described by Higgins et al., (2012); instead of utilising a consensus sequence derived from ESTs of the A and C genomes the new approach inferred IHPs based on the diploid A and C genomes of DH *B. rapa* and *B. oleracea* (Harper et al., 2012). Once the A and C pseudomolecules had

been constructed, with the SNPs correctly called and assigned, the mRNA-seq reads were aligned to the newly “cured” reference sequence, quantified and normalised as RPKM for the GEM analysis. AT analysis could then take place: SNPs were analysed with a Mixed Linear Model (MLM) within TASSEL (Bradbury et al., 2007)(using a STRUCTURE derived kinship matrix to account for population structure (Evanno, Regnaut and Goudet, 2005)) against the trait data, whilst a linear regression was performed between the RPKM data (GEMs) and the trait data. The MLM modelling approach was used to account for both population structure and relatedness in the association analysis.

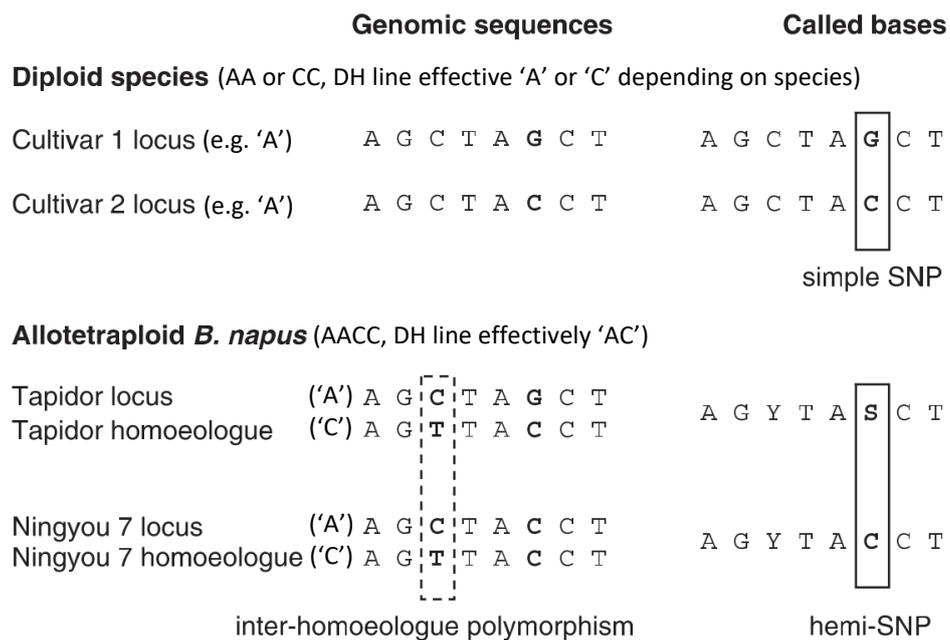


Figure 1.2.7.b A schematic representation from (Trick et al., 2009b) of the various forms of single nucleotide polymorphisms (SNP) within a DH (doubled haploid) line (the polymorphisms are in bold within the sequences).

Two full line boxes highlight where allelic SNPs have been represented, while the broken box represents an inter-homoeologue polymorphism (IHP) which is not an allelic SNP. A “hemi-SNP” occurs when an allelic polymorphism occurs within a homoeologous sequence. The ambiguity codes from the International Union of Biochemistry (IUB) are given

Subsequently, several studies have exploited AT to interrogate the genetic basis of many traits in *B. napus* with slightly modified approaches e.g. with a larger diversity panel (Lu et al., 2014; Koprivova et al., 2014) and/or newer AT methodological approaches and an even bigger diversity panel (Wood et al., 2017; Havlickova et al., 2018). The most recent modifications to the AT approach were recently described in Havlickova et al., (2018). However, as they are relevant for understanding the current research they are detailed in brief here. A major improvement on the original AT methodology is the use of a ‘pan-transcriptome’ (He et al., 2015). Instead of relying on assembled unigenes from across *Brassica* species, this approach

uses the coding DNA sequence (CDS) gene models from the published sequences of the progenitors *B. rapa* and *B. oleracea*, using the published *B. napus* genome sequence to interpolate *B. napus* specific CDS gene models. To be specific, markers were assigned to either the A or C sub-genomes within the pan-transcriptome based on the sequences of the diploid progenitors (and interpolation of the *B. napus* specific CDS gene models within the pan-transcriptome based on the order of the neighbouring CDS gene models (He et al., 2015)). Consequently the developed pan-transcriptome has significantly more CDS gene models than the published *B. napus* genome (by ~35,000) (He et al., 2015; Chalhoub et al., 2014) and enables greater discriminatory power to map markers to specific genomes, resolving a greater number of simple SNPs (Havlickova et al., 2018). Another advance utilised in the current AT approach is the use of PSIKO (Population Structure Inference using Kernel-PCA and Optimisation) (Popescu et al., 2014b) to generate a Q matrix to account for population structure. As previously described, this method combines the best of model and non-model-based approaches to account for population structure within larger datasets. This Q matrix is used with an automatically generated K matrix during CMLM SNP association analysis with the R package GAPIT (Lipka et al., 2012). Once again this enables the advantages of MLM analysis (i.e. controlling for population structure and kinship with a Q + K approach) whilst minimising computational burden from the large dataset (Zhang et al., 2010a). GEM analyses were further improved with the use of the Q matrix (i.e. fixed effect linear modelling) to account for population structure and genomic control methods (described in Devlin and Roeder, (1999)) to account for p value inflation. The results of these analyses are outputted as Manhattan plots (see section 2.2.2). In this research, AT is therefore used for the first time to identify potential candidates underlying micronutrient concentration mechanisms with ICP-MS data from the RIPR diversity panel for a number of elements (Thomas et al., 2016). However, the candidates identified with this approach require validation. The next section will describe alternative quantitative genetic approaches used to investigate ionomics before moving onto the various methods used to validate candidates.

1.2.8 Quantitative genetics and Ionomics

Quantitative genetics relies on the ability to link an observed phenotype to particular genes or loci (i.e. forward genetics). Some of the first quantitative genetic approaches utilised large mutant mapping populations to identify candidate genes. Arguably the first large scale quantitative genetic experiment in 'ionomics' was the screen performed by Lahner et al., (2003) with Fast neutron- mutagenized plants (ionomics being defined as the quantitative study of an organisms ionome in relation to its physiology, development and genetic composition (Salt, Baxter and Lahner, 2008)). This study screened ~12,500 plants with a

combination of ICP-MS and ICP-O/AES. This led to the identification of 51 mutants with altered elemental profile. Indeed this is the study frequently cited for defining the term 'ionome' because the majority of mutants identified had multiple elemental disturbances, highlighting the interdependence of elements in the ionome (Lahner et al., 2003). However, such forward genetic approaches have been limited by the need to clone the genes of interest; i.e. one of the first mutants cloned from the 2003 study (Enhanced Suberin 1-1/ *ESB1-1*) was identified in 2009 with a deletion mapping approach (utilising bulk segregant analysis and DNA microarrays) six years after the initial study (Baxter et al., 2009). Approaches such as this to mapping mutations are not uncommon in ionomics research e.g. (Gong et al., 2004; Chao et al., 2011; Tian et al., 2010) and with the improvement in sequencing technologies (i.e. next generation sequencing methods), mapping by sequencing is now also being applied in ionomics e.g. (Kamiya et al., 2015).

Comparable mapping approaches applying natural variation have also been used to find causative genes. For example, Rus et al., (2006) identified the Na transporter *HKT1;1* from wild populations of *A. thaliana* with DNA microarray BSA (bulk segregant analysis) and reverse genetics (utilising a T-DNA line in validation), the same transporter identified in mutant screens with similar methods (Gong et al., 2004). Other approaches to understanding the link between the genome and ionome have involved screening cDNA libraries of hyperaccumulators in yeast (e.g. Delhaize et al., 2003; Bozdag et al., 2014; Zhou et al., 2018). Natural variation has also been exploited in quantitative trait loci (QTL) mapping. In one experiment, Vreugdenhil et al., (2004) studied the genetic variation of a number of elements (Ca, Fe, K, Mg, Mn, Na and Zn) with recombinant inbred lines (RILs) and QTL analysis. This study found multiple QTL for most elements, some of which co-localised between different elements. This co-localisation was explained as a result of either: pleiotropy (i.e. similar elements sharing accumulation mechanisms) or due to the linkage of multiple genes within a QTL (Vreugdenhil et al., 2004). Other experiments have exploited BSA and RILs Baxter et al., (2008) used BSA to find a rough location for the Mo transporter *MOT1* using a DNA microarray, before using a RIL population to find recombinants and fine map the *MOT1* locus.

Such mapping approaches are very advantageous for identifying rare allelic variants, making it more likely that a novel allele will be found for the phenotype of interest, however this is often at the expense of precision (i.e. a large region of the genome will be highlighted). Further, these mapping and QTL approaches are often based on limited genetic material (e.g. two parental genotypes, although more complex mapping populations are now available, e.g. MAGIC (Kover et al., 2009)) (Bazakos et al., 2017). Genome Wide Association Studies (GWAS)

on the other hand are based on large and genetically diverse panels, i.e. they still use linkage disequilibrium (LD) but this is based on historic recombination events. A good follow-on example is the work in the Na transporter *HKT1;1*: previously identified with mutant (Gong et al., 2004) and natural variation screening approaches (Rus et al., 2006), it was also confirmed with a range of techniques, including GWAS, by Baxter et al., (2010) in a diversity panel of 337 *A. thaliana* accessions. This study built on the previous work of Atwell et al., (2010) which found *HKT1;1* SNP association peaks with Na in a much smaller panel (95 genotypes) as part of a much wider GWAS feasibility study (studying some 107 separate phenotypes, 18 of which were ionic). There are numerous instances in the literature of GWAS being applied to study the genetic basis of mineral element variation in crops e.g. Zn and Fe in maize kernels (Hindu et al., 2018); Mn toxicity in rice (Shrestha et al., 2018) and Ca concentrations in wheat grains (Alomari et al., 2017).

The main advantage of GWASs is that it allows the exploration of the genome in comparatively finer detail compared to traditional mapping approaches, sometimes even identifying specific candidate genes in the process (e.g. Harper et al., 2012). However, this is highly dependent on the extent of LD in the crop under investigation. An excellent example would be rice: LD in wild outcrossing species of rice decays much more quickly in comparison to the self-fertilising varieties under cultivation (e.g. *Oryza rufipogon* decays over ~20kb vs ~150kb in *Oryza sativa*)(Huang et al., 2012, 2010). On the whole, GWAS approaches have been limited in a number of ways: initially by genome sequencing technologies and the requirement for markers mapping the whole genome (particularly amongst polyploid crops); the confounding effects of population structure leading to false positives (i.e. linking causative alleles to non-causative loci due to ancestry) and the lack of detection of rare alleles (i.e. the minimal representation of rare alleles within the diversity panel prevents the detection of significant associations with the phenotype)(Bourke et al., 2018). Associative Transcriptomics is a form of GWAS, which instead of using genomic DNA makes use of mRNA-seq analysis to avoid some of the aforementioned limitations of GWAS, particularly in polyploids.

1.2.9 Quantifying multiple mineral elements

The ability to phenotype large numbers of plants is essential for quantitative genetics and ionomics. By far the technique most commonly applied in plant ionomics is Inductively Coupled Plasma (ICP) spectroscopy; be it ICP-MS (ICP- mass spectrometry) or ICP-OES (ICP-optical emission spectroscopy)/ICP-AES (ICP- atomic emission spectroscopy). The principle behind both techniques is very similar: the samples are ionised with inductively coupled plasma ('ICP') and then the concentration of the ions is determined. ICP-MS utilises mass

spectrometry ('MS') for separation and quantification of ions based on their mass to charge ratios, whilst O/AES rely upon optical/atomic emission spectroscopy (i.e. measuring the emitted electromagnetic radiation from ionised samples to give their concentration). Both techniques rely on thorough digestion of the plant material under investigation to prevent problems with sample introduction or analytical biases (which has implications for the plant materials under investigation, e.g. seeds are harder to digest than leaves)(Husted et al., 2011). These two techniques have been favoured within the plant ionomic community for a number of reasons: the ability to analyse multiple elements at a time; both are relatively cheap and have large sample throughput (Djingova et al., 2013). ICP-MS analysis has significantly lower limits of detection (LOD) in comparison to ICP-O/AES (ICP-MS at $1e^{-6}$ - $1e^{-4}$ $\mu\text{g g}^{-1}$; ICP-O/AES 0.001-10 $\mu\text{g g}^{-1}$ (Djingova et al., 2013)), however ICP-O/AES is easier to run, simple to maintain and more cost-effective (Husted et al., 2011). ICP-MS was the technique used in the current research because of the large number of elements with a wide range of concentrations in the plant materials under investigation. However, ICP techniques are not the only methodologies used in multi-element ionomic research and sometimes a combination of techniques is used to determine a broader range of elements (e.g. Queralt et al., 2005; Phan-Thien, Wright and Lee, 2012; Barbosa et al., 2015). This would not have been a feasible option in the current research considering the large number of samples analysed (383 genotypes with 5 replicates in seeds and leaves, (Thomas et al., 2016)). The phenotype data used in the current study is detailed in the general methods (2.1.1) and was generated as part of the Renewable Industrial Products from Rapeseed consortium (RIPR; (BBSRC, 2014)) as detailed in Thomas et al., (2016).

Other techniques often used in ionomic studies include: Atomic Absorption Spectroscopy (AAS: however this is typically a single element technique); X-ray Fluorescence Spectrometry techniques (XRF: although it has a much higher LOD in comparison to the ICP techniques and can therefore determine fewer elements *in planta*) and Neutron Activation Analysis (NAA: however this is limited by the requirement for an experimental nuclear reactor)(Djingova et al., 2013). Other fast spectroscopic techniques based on secondary indices have also been used for indirect analysis of element composition within plants; such as Ultraviolet, visual, Near and Mid-Infrared Spectroscopy (UV, Vis, NIR and MIR), alongside chlorophyll a fluorescence (van Maarschalkerweerd and Husted, 2015). However, they are limited in a number of ways: they are indirect analyses and therefore require careful calibration; many elements are stored in non-metabolic pools (e.g. the vacuole or cell wall), and therefore will not be analysed and as of yet multi-element determination with such techniques is limited (van Maarschalkerweerd and Husted, 2015). In addition to the mere quantification of elements within tissues, it is important to note that a range of experimental techniques for analysing the

localisation and speciation of elements in plants has been developed, e.g. LA-ICP-MS (Laser Ablation ICP-MS), SIMS (Secondary Ion Mass Spectrometry), S-XRF (Synchrotron XRF) and XAS (X-ray Absorption Spectrometry)(Zhao et al., 2014). Since such techniques have not been used within the current analyses (although based on the same chemistry involved in quantification), they are discussed no further and the reader is referred to Zhao et al., (2014) for a review. Similarly, anion analyses have not been discussed as these were analysed as part of a similar but separate project (i.e. HPLC analysis in Koprivova et al., (2014) and Alcock et al., (2018)).

1.2.10 Candidate gene functional validation

Once a candidate gene has been identified, the role it may play in nutrient concentration needs to be validated. There are a number of approaches which can be used for candidate gene validation, including physiological analyses, genetic transformation and genetic complementation (Pflieger, Lefebvre and Causse, 2001). Physiological analyses provide evidence for the role of the candidate gene in the trait but cannot provide definitive proof. An example of this would be work on the *tb1* gene in maize and teosinte (Doebley, Stec and Hubbard, 1997). In this work expression analyses were performed which highlighted that the expression of *tb1* was two-fold higher in maize than its progenitor species teosinte, implying a role for this gene in the evolution of the unbranched maize phenotype during domestication. However for this role to be definitively proven genetic complementation analyses had to be performed (Doebley, Stec and Hubbard, 1997). Other techniques used in verifying candidate gene function have sought to disrupt the expression of the candidate gene, including traditional mutagenesis approaches (such as TILLING, ‘Targeted Induced Local Lesions In Genomes’(McCallum et al., 2000a, 2000b)), insertional mutagenesis (T-DNA (Feldmann and Marks, 1987)), RNAi (Hannon, 2002) and genome editing (e.g. CRISPR/Cas9, ZFN or TALENs (Bortesi and Fischer, 2015)). However, all of these methods are significantly more complicated in polyploid species due to functionally redundant homoeologues (Fitzgerald, Kazan and Manners, 2012). For example, when TILLING was performed in *B. napus* a novel method needed to be developed in order to distinguish between just two paralogues, screening 1000s of plants to eventually identify 3 functionally compromised mutants (Wang et al., 2008). Considering there can be six copies of a single orthologous gene from *A. thaliana*, this highlights the resource intensive nature and impracticalities of TILLING in *B. napus*, particularly for multiple candidates. CRISPR/Cas9 has also been used to target multiple homoeologues of a gene in *B. napus* (Yang et al., 2018), however the plants produced with this method are considered transgenic. This therefore allows the validation of the candidate gene but limits the agronomic impact of the plant materials produced with such an approach.

The current research has balanced the need for verifying a large number of candidate genes with the practicalities of candidate gene validation in polyploids. Instead of validating the function of the candidate within *B. napus*, the current research has exploited the extensive T-DNA mutant libraries available in *A. thaliana* for quick and cost-effective validation of orthologous genes (O'Malley, Barragan and Ecker, 2015). If the mutant displayed disruption in the concentration of the element of interest relative to the control, it would validate that candidate for a role in the element under investigation and therefore warrant further investigation in *B. napus* (by exploiting the genetic variation within the diversity panel or through TILLING mutants). This method has allowed many candidates to be assessed within a significantly shorter timeframe than any of the other methods mentioned, however it has a number of drawbacks. Firstly, as previously mentioned, *A. thaliana* and *B. napus* diverged ~20 Mya (Yang et al., 1999). Therefore, candidate genes identified in *B. napus* may have functionally diverged from their *A. thaliana* orthologues, meaning results from either species may not be applicable in the other. Other issues have included a lack of suitable *A. thaliana* T-DNA insert lines, e.g. some candidate genes did not have a T-DNA insert line in *A. thaliana* or all plants produced from the 'T-DNA' line were genotyped as wildtype. Furthermore, the T-DNA lines used in the current analyses were not subjected to transcript quantification (e.g. with qRT-PCR) and therefore whether they are true knock-outs or merely knock-downs was not determined (i.e. some candidates may have been rejected incorrectly). Additional issues relate to the growth conditions of *A. thaliana*. All plants were grown in a nutrient rich soil environment and harvested at the same developmental time point. This could have masked phenotypes which only become apparent under nutrient deficient conditions or at certain developmental phases. Finally, only nutrient concentrations were analysed, therefore other important phenotypes could have been missed which may indirectly affect nutrient concentration (e.g. plant size, developmental defects or fertility). The current research should therefore be considered an initial screen, looking at many candidates quickly which can then be taken on for further validation in *B. napus* (e.g. with TILLING).

1.2.11 Conclusions

Investigations into nutrient use efficiency within plants have been complicated by the complexity of the ionome, the various definitions of 'nutrient use efficiency' and the polyploid genomes of many crop plants. Developments in sequencing and phenotyping technologies have led to an explosion in 'Ionomics' research. The current research has focused on using these new sequencing and phenotyping technologies: Associative Transcriptomics was used to circumnavigate the issues limiting genetic research in polyploid crops and ICP-MS analysis was exploited to investigate multiple elements in *B. napus*. As *B. napus* is an economically

important oil crop with high nutrient demands and is closely related to many important crop species, it was deemed to be an excellent candidate for investigating micronutrient concentration. Six elements were subsequently studied: Mo, Mn, Cu and Zn as essential micronutrients, while S and Cd were investigated due to the likelihood of interactions with one or more of the primary micronutrients. Novel candidate genes identified from these analyses were further investigated with *A. thaliana* T-DNA insertion lines, whilst candidates with a known role in micronutrient concentration would make suitable targets for future TILLING or MAS

2 General methods

2.1 Data sets:

2.1.1 Pre-existing data sets

As part of the RIPR (BBSRC, 2014) consortium, a diversity panel of 383 *B. napus* lines was generated (Havlickova et al., 2018) for leaf transcriptome analysis which would detail sequence variation in the form of single nucleotide polymorphisms (SNPs) and as gene expression markers (GEMs), constituting the functional genotypes. Plants from this panel would be further utilised in the generation of various sets of trait data for the consortium, including both the leaf and seed concentration data used within this study. Together, the leaf transcriptomic and trait data would be analysed by Genome Wide Association with an Associative Transcriptomics approach (AT).

As described in Havlickova et al., (2018), the diversity panel consisted of seven main 'crop types': spring oilseed rape (123 accessions), semi-winter oilseed rape (11 accessions), swede (27 accessions), kale (3 accessions), fodder (6 accessions), winter oilseed rape (169 accessions) and an unassigned group (44 accessions, i.e. crop type information unknown). These plants were grown by collaborators at the University of Nottingham, as described in Thomas et al., (2016). In brief, plants were initially grown in fine grade compost-based growth media (<3 mm particle size, Levington Seed and Modular plus sand F2S, Everris Ltd., Ipswich, UK), before being transplanted into individual 5 L pots with Levington C2 compost (Scotts Professional, Ipswich, UK) after approximately 2 months growth (late October 2013- January 2014). The plants were grown in a randomized block design (five plants per accession) in two unheated polytunnels (single skin, Visqueen Luminance Skin, Northern Polytunnels, Colne, UK) with no additional lighting. Plants were watered 3 times daily with an automatic irrigation system (Hunter Irrigation Controller, Hunter Industries, San Marcos, CA, USA, provided by Hortech Systems Ltd., Holbeach, UK). Sampling for RNA extraction was performed on the second true leaf as described in Bancroft et al., (2011), whilst sampling for ionome analysis occurred around the 6-8th true leaf in early March 2014 as described in Thomas et al., (2016) (see below for further details on the functional genotypes and phenotype data). There was no additional fertilisation until after plants were sampled for leaf ionome analysis. The fertiliser was Kristalon Red NPK (Yara, Grimsby, UK) and was applied via a direct feed injector (Dosatron D3GL-2, Tresses, France) from late March to May 2014. Plants were bagged before flowering to prevent cross pollination and harvested for seed in July 2014.

The functional genotypes were generated from the aforementioned leaf RNA, as detailed in Havlickova et al., (2018). In summary, the leaf transcriptome of all lines was sequenced with the Illumina (HiSeq 2000) sequencing platform giving 100 base read length mRNAseq data (by the Earlham Institute, formerly The Genome Analysis Centre). The Illumina reads were then processed by Dr. Z. He (a bioinformatician in the Bancroft group, University of York) according to the methods outlined in Harper et al., (2012) and Bancroft et al., (2011) to give the final SNP and GEM data sets. From the SNP file the genetic architecture of the population was analysed (as detailed in the Introduction (see 1.2.7) and Havlickova et al., (2018)) and a Q matrix was generated using PSIKO (Population Structure Inference using Kernel-PCA and Optimisation (Popescu et al., 2014a)) giving the highest likelihood as a subpopulation K=2. The kinship matrix was automatically generated by GAPIT (see 2.2.1 and Lipka et al., (2012)). For the final AT analysis 9,839 simple SNPs and 246,558 hemi SNPs (which had a second allele frequency >1%) were utilised. Out of 116,098 CDS gene models within the reference pan-transcriptome, 53,889 had significant expression (defined as >0.4 mean reads per kb per million aligned reads/RPKM) and were therefore used in AT analysis.

Two sets of ICP-MS trait data were generated for the purpose of this study (as previously described in Thomas et al., (2016), whilst anion analysis has been described in Alcock et al., (2018) and was not included as part of this project. In brief, five plants per accession were sampled by collaborators at the University of Nottingham: 3-4 leaves were sampled at the 6-8 leaf stage and seed material was harvested at the end of the experiment. Leaves were freeze dried and ground together (i.e. one sample contained multiple leaves from the same plant) and a subsample of ~0.2g DW of leaf material was digested in 70% nitric acid within a microwave digester (as described in Thomas et al., (2016)). For seed analyses, 3-4 seeds were left to predigest in 70% nitric acid overnight. Samples were then heated on hotplates to 115 °C for 4 hours. Weights for seeds were calculated using the method outlined in Danku et al., (2013). Both the leaf and the seed ionome were then analysed with ICP-MS (Inductively Coupled Plasma Mass Spectrometry, leaf ICP-MS analysis performed at the University of Nottingham and seed ICP-MS analysis performed at the University of Aberdeen). Leaves were analysed for 28 separate elements, of which 7 (Ag, Co, Cr, Ni, Pb, U and V) were excluded from further analysis, since their average concentration was either below or very close to the limit of detection (LOD, see 2.5 for further details). Similarly, of the 22 elements analysed in seed, only 15 were carried forward (the 7 elements excluded being As, Co, Cr, Fe, Ni, Pb and Se).

2.1.2 Selection and processing of pre-existing data

As explained previously (1.2.5), only 6 of the elements from the leaf and seed ionome data sets were selected for further analysis, including; Molybdenum (Mo), Copper (Cu), Manganese (Mn), Zinc (Zn), Sulfur (S) and Cadmium (Cd). These 6 elements were selected because; 1) they are essential micronutrients (Mo, Cu, Mn and Zn) or interact with them (S and Cd); 2) they were accurately measured across both the leaf and seed data sets (i.e. Fe and Ni were excluded, see 2.1.1); 3) were not under investigation as part of any other projects associated with the wider RIPR community (i.e. B); 4) displayed similarities or interactions to elements which fitted the first three criteria (S and Cd). The data for each of these elements was treated simply; any element concentration which was more than 5 standard deviations away from the arithmetic mean was removed, an arithmetic mean was then calculated for each of the 383 accessions under analysis. The arithmetic means of the 383 lines was then utilised in AT analysis.

2.2 Associative Transcriptomics (AT)

2.2.1 AT pipeline

As part of the project, AT was performed on a number of element concentrations within the seed and leaf ionic data sets (see 2.1.2). In general, AT was performed in accordance to the methods previously outlined in Harper et al., (2012), with R scripts modified by Dr. Z He. All analysis took place using the statistical package R. For performing SNP associations an edited version of the R script GAPIT (Genomic Association and Prediction Integrated Tool) (Lipka et al., 2012) was used to carry out compressed Mixed Linear Modelling (Zhang et al., 2010b), utilising the PSIKO Q matrix to account for population structure (GAPIT was responsible for automatically generating the K matrix)(Sollars et al., 2017). GEM associations were executed using the R script Regress, to perform fixed effect linear modelling with RPKM data and the PSIKO Q matrix as fixed effects against the leaf/seed ionic trait data. Genomic control measures were applied to GEM analysis to account for p value inflation (Devlin and Roeder, 1999). For an in-depth analysis of this approach and why it was used to investigate nutrient concentration in *B. napus* please refer to the introduction, see 1.2.7.

2.2.2 Analysis of AT graphed outputs

The AT pipeline generated results as Manhattan plots (a specialised scatter plot, **Figure 2.2.2.a**); with the $-\log_{10}P$ values from the SNP and GEM association analyses on the Y-axis, plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes, A1-A10, C1-C9) based on the CDS gene model order using the R script Grapher V11 (modified from Harper

et al., (2012), by Dr. Z. He). Using the graphed outputs from the AT pipeline, it was possible to identify the more highly associated markers, which could be compared to the pseudomolecule database and subsequently related to their orthologous genes in *A. thaliana* (via The Arabidopsis Information Resource (TAIR, 2015) and Araport (Krishnakumar et al., 2015)). Included in these analyses were two type 1 error tests: A) the false discovery rate (highlights the point at which 5% of the most highly associated markers are expected to be false positives); B) the Bonferroni corrected significance threshold ($p=0.05/\text{number of markers scored}$). Usually the use of both these threshold tests would be sufficient for determining whether a biologically relevant association had occurred, providing evidence that a gene of interest would be present within the association peak under analysis. Frequently however, analyses failed to pass either of these thresholds. In these instances, association plots were analysed in a number of ways (**Figure 2.2.2.a**): 1) if a clear clustering of markers was observed, i.e. within ~ 100 CDS gene models and above the general noise of markers, it was considered an association peak worth searching for a gene relevant to the element under assessment, 2) GEMs were specifically analysed for standalone markers above the background noise of the rest of the Manhattan plot as single markers in the GEM analyses may highlight a gene with direct involvement in the trait, i.e. its expression is closely related to the concentration of the element of interest, 3) where no distinct association peaks could be discerned, the markers were sorted according to the significance of their association and the top markers from each chromosome (A1-10, C1-9) were assessed for any involvement in the element under consideration (this method was used so that there was a range of markers from different locations across the genome for assessing trait predictability, see 2.3).

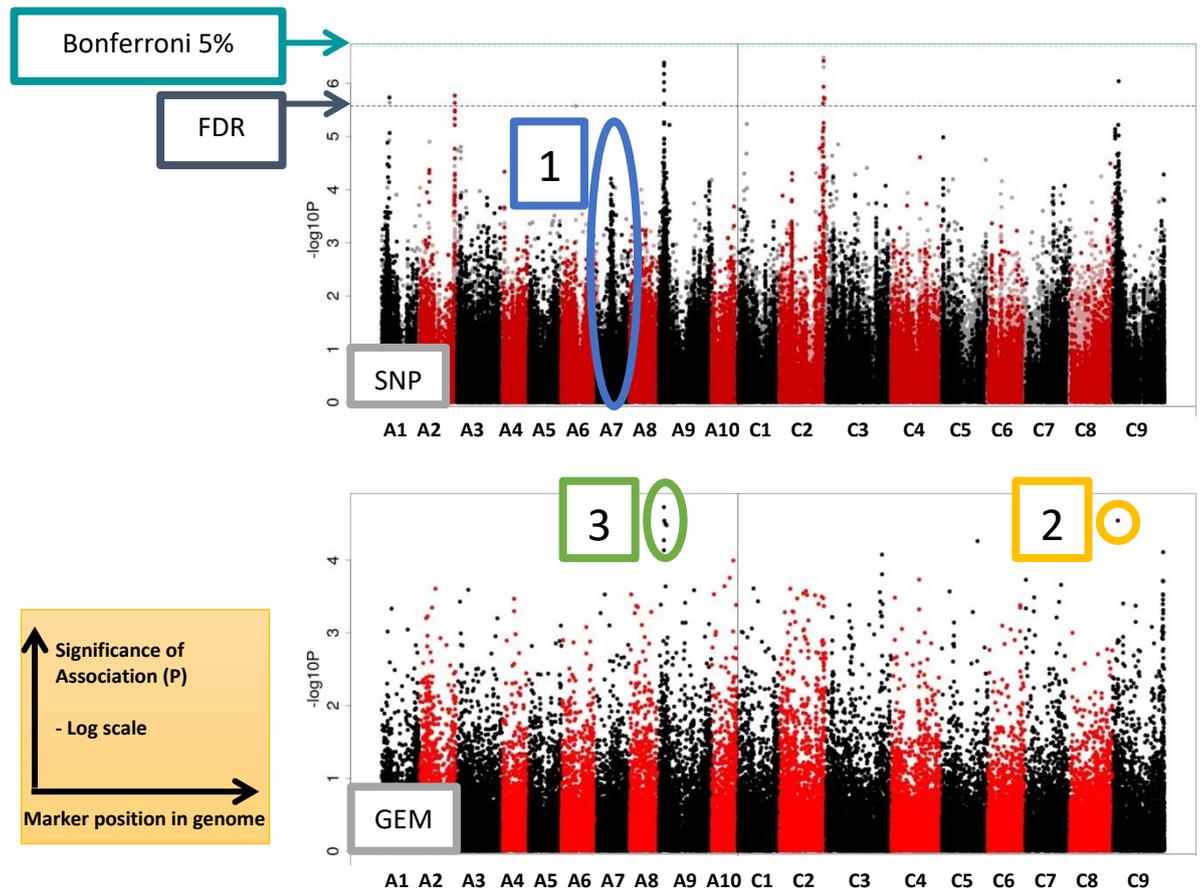


Figure 2.2.2.a An example AT Manhattan output from the current study with key features highlighted to emphasise the analysis process.

In this example, the genome wide distribution of mapped markers associating with the Mo concentration in seeds (mg/kg DW) of all 383 accessions is displayed (the concentration for each accession is based on an average ICP-MS results from the seeds of 5 different plants, see 3.2.5 for a detailed AT analysis). The SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and the seed Mo concentration as the response variable. $-\text{Log}_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which are displayed in the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue. Frequently AT analyses failed to pass either of these thresholds, in these instances no dashed line appears on the AT plot, such as in the GEM output of the current example. Therefore in addition to

using these thresholds to determine association peaks for further analysis, a number of other approaches was taken: 1) clear peaks of ~100 CDS gene models, which did not pass the thresholds were still analysed for relevant candidate genes; 2) stand-alone GEM markers above the noise of the rest of the association plot were always investigated further, as single GEM markers may indicate direct involvement of that gene specifically within the trait; 3) when no distinct association peak could be observed, the top markers of each chromosome were investigated for a potential role in the trait, e.g. those on A9 in the GEMs highlight the deletion close to the gene of interest from the above SNP association peak. Only candidates with a plausible role in the trait under investigation were taken forward for further study in *A. thaliana* (which has obvious limitations, e.g. reliance on author's knowledge of the trait).

From this analysis, candidate genes could be selected and tested with *A. thaliana* T-DNA insertional mutants (see 2.4). This method clearly lacked objective criteria for the selection of candidates and relied upon the author's ability to identify appropriate targets (with implications for the novelty of those candidates identified). Thresholds were not lowered from 5% to maintain consistency across analyses and enable comparison of AT results to previous AT analyses, as well as GWAS analyses within the literature. Furthermore, some of the most highly associated markers or candidates did not have an orthologous *A. thaliana* gene within the pseudomolecule database. In these instances the candidates could only be analysed *in silico*. Briefly, the CDS for the markers were compared to all others within the *B. napus* RIPR (BBSRC, 2014) pan-transcriptome (with a script written by Dr. Z. He) and across plant species in the NCBI database with BLAST (Basic Local Alignment Search Tool (Altschul et al., 1990)). This analysis was useful for unannotated lone GEM markers, making it particularly easy to identify potential uncharacterised transporters in *B. napus*.

2.3 Testing the predictive capabilities of markers

The most highly associated markers from AT analysis were used in the generation of predictive markers, regardless of whether a candidate gene could be identified or if the marker was thought to be directly involved in variation of the element analysed. These predictive markers could then be further exploited by breeders in marker assisted selection (MAS) for improvements in NUE (nutrient use efficiency). To this end, the RIPR diversity panel was split into two subsets: a training panel of 274 lines and a test panel of 109 lines. The panel was split so that there was a representative of each crop type in both diversity panels, with roughly two thirds of all lines in the larger panel. AT analysis was therefore performed twice for each trait analysed: firstly, on the 274 diversity panel for the selection of predictive markers; and again on the full 383 diversity panel for added clarity in candidate gene selection.

Highly associated SNP markers were identified from the 274 AT analysis; SNP data from both the 274 and 109 panels was collected for each of the markers identified, alongside the allelic effects estimate generated as part of the 274 analyses. Trait data for the 274 panel was then aligned against the SNP data and an average trait value for each allele of the SNP was calculated. This calculation enabled the identification of the allele associated with an increase/decrease in the trait. The allelic effect was then divided between the alleles of the 109 panel in accordance to the effect observed in the 274 panel. The 109 SNP panel was recoded with the allelic effect estimate, standardised and rescaled according to trait data from the 274 panel. The observed and predicted trait data for the 109 panel was then compared with a simple correlation to assess the predictive capability of each SNP. Multiple SNPs could be combined to improve trait predictability by simply averaging the predicted trait values prior to comparison with observed values.

It was also possible to make predictions from GEM data. Again, 274 trait data were analysed with AT analysis and the most highly associated GEM markers were selected. From the AT analysis, the gradient and intercept of the marker-trait association were extracted. RPKM data from the 109 panel was then transformed accordingly to generate the predicted trait data. Once again this was compared to the observed trait data with a simple correlation. Again, multiple GEMs could be combined by averaging predicted values and comparing them to the observed data.

In addition to individual predictions, both SNP and GEM predictive markers could be combined into a single prediction, greatly improving their predictive capacity. Furthermore, it was often possible to use the predictive markers across multiple traits (e.g. Mo and S concentrations in seed, see 3.2.5 and 3.2.6) reducing the numbers of markers needed for MAS and indicating potential biological links between traits.

2.4 *Arabidopsis thaliana* T-DNA knock-out analyses of candidate genes

2.4.1 Growth of plants

Once candidate genes had been identified from AT analysis (see 2.2) they were tested with *A. thaliana* T-DNA knock out lines (**Table 2.4.1**). If these mutants displayed disrupted element concentrations relative to a wildtype control, then it would indicate that these genes play a role in plant nutrient concentration and should be investigated further. Mutants were selected through The Arabidopsis Information Resource (TAIR, 2015) and ordered through the National Arabidopsis Stock Centre (Scholl, May and Ware, 2000). If possible, multiple mutant lines were ordered for the same candidate to improve the chances of finding plants homozygous for the

insert and provide greater evidence for the role of each gene in nutrient concentration.

Following the seeds arrival, they were imbibed in water for 24-48 hours before planting. Seeds were sown into F2 soil within P24 pots (24 plants per tray, individual 5 cm³ cell per plant). For each genotype 24 seeds were sown and after 2 weeks thinned so that 12 plants remained per line. Plants were grown in Aracons and Aratubes (Arasystems) to improve harvesting and limit the potential for seed contamination. Plants were grown under 16 hours of light, with day time temperatures of 20°C and night time lows of 16 °C. Plants were genotyped (see 2.4.2) and those which were identified as being homozygous or heterozygous would be left to set seed, harvested and threshed individually.

Once plants had been bulked and segregation of the insertion was observed, the next generation was grown for analysis. For leaf candidates, plants were grown in P24 pots under short day conditions (12 hour day/ night, 16-20 °C temperature range) to improve the volume of leaf material recovered and ensure accurate ICP-MS analysis. For those mutants with suspected/known roles in flowering time induction, plants were transferred to long day conditions (16 hour day) until a week after bolting was observed. Leaf materials were then harvested and dried in the oven at 80-90°C for 48 hours in preparation for the next stage of analysis (see 2.5).

Table 2.4.1: Candidate genes found as part of the current project.

The particular trait under investigation is given in the column “Trait”, followed by the association peak in which the candidate was selected from (“Peak”) and whether the candidate was selected from SNP (S) /GEM (G) /WGCNA (W) analysis (for details of WGCNA see “6.4.1.1 Weighted Gene Co-expression Network Analysis (WGCNA)”). The *A. thaliana* AGI code is given alongside a brief description of gene function. Insert lines were obtained from the National Arabidopsis Stock Centre (NASC(Scholl, May and Ware, 2000)), with their order ID given in “NASC id”. Whether the line was thought to be segregating is indicated (S: segregating, H: homozygous), as well as the supposed “insert site” (E: exon, I: intron, P: promoter, U: unknown. The left border/insert specific primers (“LB”, LB1.3: ATTTTGCCGATTTTCGGAAC; p745_wisc, AACGTCGCAATGTGTTATTAAGTTGTC, SAIL_LB3, TAGCATCTGAATTCATAACCAATCTCGATACAC, o8474_gabi, ATAATAACGCTGCGGACATCTACATTTT), left and right genomic primers used for genotyping the lines are given. Starred (*) primers were generated by an undergraduate project student, Mr Jun Hee Jung, under the supervision of the author, whilst starred lines were analysed in part by the undergraduate (further details given throughout the text).

Trait	Peak	SNP/GEM/WGCNA	AGI	Description	NASC ID	Segregating	Insert site	LB	Left genomic primer	Right genomic primer
Cd leaf*	A3a	S	AT2G36950	Heavy metal transport/detoxification superfamily protein	N657327	H	E	LBb1.3	CGTGGTGCCTTTAAAGATCAG	CATTAGTTGATCGAGAAAATGGC
					N662850	H	E	LBb1.3	CGTGGTGCCTTTAAAGATCAG	CATTAGTTGATCGAGAAAATGGC
Cd seed*	C3	S	AT3G24450	Heavy metal transport/detoxification superfamily protein	N527460	S	I	LBb1.3	CACAAGGCAAAATGTTTTGGAA	GCCCATGTGGAAGAAAAGAG
					N924500	S	I	p745_wisc	CACAAGGCAAAATGTTTTGGAA	GCCCATGTGGAAGAAAAGAG
Cu leaf*	C2	S	AT1G65840	polyamine oxidase 4	N676662	H	P	LBb1.3	CAAGTGGTTGACCAATTCAAC	TCAAATTATTCATGGCGAGG
					N675516	H	P	LBb1.3	CAAGTGGTTGACCAATTCAAC	TCAAATTATTCATGGCGAGG
Cu leaf*	C2	S	AT5G19390	Rho GTPase activation protein (RhoGAP) with PH domain	N684257	H	E	LBb1.3	TTTGTGGATCATCGCCTATC	TACCAAGTCTGCGTCTATCGG
					N678748	H	I	LBb1.3	AGAGAGTCGTAGCTCGGATCC	TGGACCACTCTTGAAAACCTG
Cu leaf*	C2	S	AT1G65930	cytosolic NADP+-dependent isocitrate dehydrogenase (Cu SNP)	N662539	H	E	LBb1.3	GCGTTTGAGAAGATCAAGGTG	ATCAGTGGCACGGTACTGATC
					N509094	S	E	LBb1.3	GCATTCAAAAATTCACATCAA	TTCTTAGTAGCTTCAGCACTTTCA
Cu leaf*	C1	G	AT4G24930	thylakoid lumenal 17.9 kDa protein, chloroplast	N668742	H	I	LBb1.3	CATACCTGCGAAATCGTGAAA	GGGATCGTCAATGGAAGAG
Cu seed*	A7	S	AT1G68100	ZIP metal ion transporter family	N669663	H	U	LBb1.3	GCATAGCAACAAGTATCACCC CTT*	AAACCGAACCGAAACGCAACA*
					N840161	H	U	SAIL_LB3	TGTAGTCAAAAACAACAAAACC CAAG*	TCGGACAACCGAAGAGGTG*

Trait	Peak	SNP/GEM/WGCNA	AGI	Description	NASC ID	Segregating	Insert site	LB	Left genomic primer	Right genomic primer
Cu seed*	A3	S	AT4G05030	Copper transport protein family	N418855	S	I	o8474_gabi	AGGTATTGCTTGACTCATAAGGG*	AGGATGTAGTGCCGTTTCTTTG*
S seed	C2	G/W	AT5G62090	SEUSS-LIKE 2/SLK	N674571	H	P	LBb1.3	TTGCCAAGTTTTGAATGATCC	GCAATCCCTTAAAAATCTCGG
					N65894	H	E	LBb1.3	AGATCACACTGCCATTCATCC	CTGGTGATATGCATAATCCGG
S seed	A8	S	AT4G14030	selenium-binding protein 1 (SBP1)	N698382	H	E	LBb1.3	TCATCTCCAAGACAGGACACC	ACAGTGTACGTGGACACATGG
					N647322	S	E	LBb1.3	TCATCTCCAAGACAGGACACC	ACAGTGTACGTGGACACATGG
Mo/S leaf/flowering	C2/C7/C9	G/W	AT2G40550	E2F-target gene 1	N671574	H	P	LBb1.3	ACAGAGCTCGTAAGCAAGCTG	TAGGGCAAACCTGGGAGATAG
					N684052	H	P	LBb1.3	ACAGAGCTCGTAAGCAAGCTG	TAGGGCAAACCTGGGAGATAG
Mo/S leaf	A10	S/W	AT1G01070	nodulin MtN21 /EamA-like transporter family protein (UMAMIT28)	N676358	H	E	LBb1.3	AATGGTCGATCATTTCGTGTCAG	AAGGCTCAAGAGAGCACATTG
					N667569	H	I	LBb1.3	AATGGTCGATCATTTCGTGTCAG	AAGGCTCAAGAGAGCACATTG
Mo/S seed, total nutrient seed	A2, A9, C2, C7, C9	S/G/W	AT5G61420	Myb28	N686854	H	E	LBb1.3	TCCAACCTCTCCATGTTGGATC	CTCTTCCACACCGTTTCAAC
Mo/S seed, total nutrient seed	A9, C9	S/G	AT5G62680	GTR2	N668259	H	P	LBb1.3	AACAGAGTCAACCGCCGTAAC	TGCAGCCAGCACACTAGATTG
					N572700	S	P	LBb1.3	AACAGAGTCAACCGCCGTAAC	TGCAGCCAGCACACTAGATTG
Mo/S seed, total nutrient seed	A2, C2, C4	G/W	AT5G62130	Per1-like family protein	N656392	H	E	LBb1.3	TGCTCGAGATCAAGAAAGCTC	AATGTCAGAAAACCTGGATGCG
					N554073	S	P	LBb1.3	TCCACCAAACCTGTGAAACTC	AAATTTCTCCAAAAATTCG
Total nutrients leaf/flowering	C3	G	AT2G38480	CASP-LIKE PROTEIN 4B1, CASPL4B1	N685431	H	P	LBb1.3	AAGCAAATACGCCACAATCTG	CAGAAAACACAATCTTCCAATGAG
					N684873	H	P	LBb1.3	AAGCAAATACGCCACAATCTG	CAGAAAACACAATCTTCCAATGAG
Total nutrients leaf/flowering	C4	G	AT2G45660	AGAMOUS-like 20	N657480	H	I	LBb1.3	AAGGATGAGGTTTCAAGCGTC	GAAGAAGATATGGTGAGGGGGC
					N684965	H	I	LBb1.3	AAGGATGAGGTTTCAAGCGTC	TGGCGAATTCATAAAGTTTGC

For seed candidates, the second generation was grown in exactly the same way as the first and allowed to set seed. Seed from multiple individuals of the same genotype were then combined and mixed (individual lines frequently having too few seeds to fulfil weight requirements for accurate element analysis). This was then split into ~6x100mg samples in preparation for the next stage of analysis (see 2.5).

2.4.2 Genotyping:

After 2-5 weeks of growth, plants were sampled for DNA extraction. A leaf was sampled from each plant with forceps and placed into a pre-labelled Eppendorf tube (forceps were cleaned between each sample with 70% ethanol and samples stored on ice throughout). DNA was then extracted following the CTAB method: plant materials were lysed (Qiagen TissueLyser II) in 500 µl of 2x CTAB buffer pre-heated to 65°C (2g hexadecyltrimethylammonium bromide, 28ml 5M NaCl, 10ml 1M Tris-HCl pH8, 4ml 0.5M Na-EDTA pH8, 56ml dH₂O). Lysed samples were incubated at 65°C for 1 hour before the addition of 300 µl chloroform IAA (24:1) and vortexing. After centrifuging for 5 minutes at 14kg, approximately 500 µl of the upper aqueous layer of the sample was transferred to a new tube where it was mixed with 900 µl of ethanol/sodium acetate (960 µl ethanol and 40 µl 3M sodium acetate). Samples were left overnight at 4°C and then centrifuged the following day for 10 minutes at 14Kg. The supernatant was removed and the pellet carefully washed with 70% ethanol. Once all the ethanol had been removed, the pellet was re-suspended in 100 µl of water.

Genotyping utilised the Salk Institute Genomic Analysis Laboratory (SIGnAL) T-DNA Primer Design online tool (SIGnAL) and the ThermoFisher Scientific OligoPerfect™ primer design tool (ThermoFisher) to generate primers (**Table 2.4.1**). Plants were genotyped with the use of three primers in two separate PCR reactions (**Figure 2.4.2.a**). The first PCR reaction determined whether the plant had a WT copy of the candidate gene; by using the left and right genomic primers amplification would only occur if the plant had at least one copy of the gene without the insert (i.e. there would be no amplification for lines homozygous for the insert). The second reaction required the universal insert primer (**Table 2.4.1**) and the right genomic primer; with this reaction amplification would only occur if the plant had at least one copy of the gene with the insert (i.e. there would be no amplification for wild type lines). Using this method, it was possible by a process of elimination to determine if the lines were homozygous, heterozygous or wild type (i.e. wild type lines would only produce a product in PCR1, homozygous lines would only produce a product in PCR2 and heterozygous plants would show amplification in both reactions).

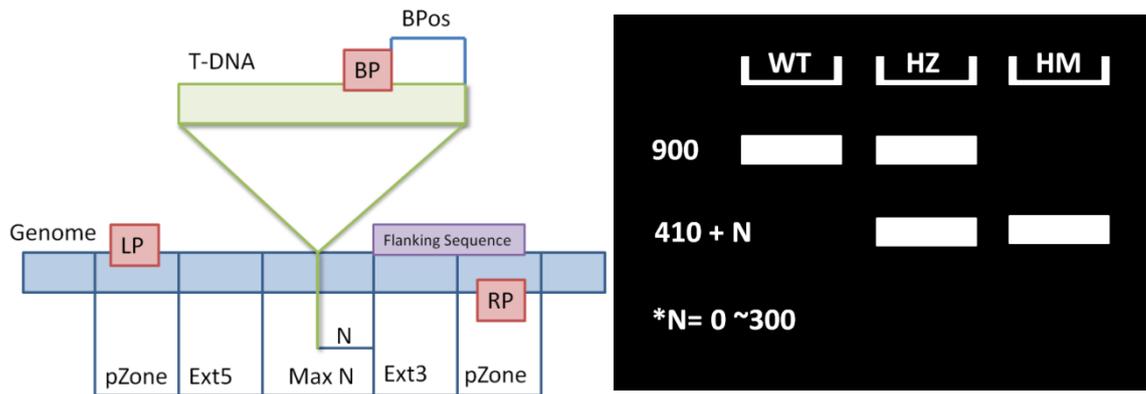


Figure 2.4.2.a: schematic from the SIGnAL T-DNA primer design tool (SIGnAL).

On the left hand side the diagram shows how primers are designed based on the selected T-DNA insert lines used. Parameters can be changed to optimise the design of the primers: N (difference between the insertion site and the flanking sequence, default 0-300bp), MaxN (maximum difference of the insertion site and the sequence, default 300), Ext5/3 (region reserved for not picking primers), Pzone (region used to pick primers, default 100bps), LP/RP (left and right genomic primers), BP (T-DNA border primer) and Bpos (the distance between the insertion site and the left border primer). For this experiment all parameters were left at default. Primers designed using OligoPerfect™ design tool (ThermoFisher) followed the same basic principles applied in the SIGnAL tool (i.e. placement of genomic specific primers away from insert site) wherever primers designed with the SIGnAL tool failed to amplify or were not available. On the right hand side can be seen a representation of how the primer system works on a gel under optimal conditions and a three primer PCR reaction. For the purposes of this experiment two separate PCR reactions were undertaken instead of the 3 primer approach to ensure accurate determination of genotype (e.g. 3 primer PCRs can give inaccurate results when one set of primers has better amplification). Homozygous (HM), heterozygous (HZ) and wild type (WT) plants can be distinguished based upon the size of the product (base pairs, $\sim 900 = \text{WT}$, $\sim 410 + N = \text{HM}$).

DNA was amplified in a 20µl reaction mixture containing 10 µl of 2xMastermix (Thermoscientific 2x Mastermix), 8 µl of water (sigma), 0.5 µl of both left and right primers and 1 µl of DNA. PCR cycler conditions were [94°C for 5 minutes (94°C for 30 seconds, 60°C for 30 seconds {decrease by 1°C every cycle}, 72°C for 1 minute) x 10 cycles, (94°C for 30 seconds, 55°C for 30 seconds, 72°C for 1 minute) x 29 cycles, 72°C for 7 minutes, 7°C forever]. After amplification, PCR products were visualised on a 1% agarose gel and genotyped as described previously (**Figure 2.4.2.a**).

2.5 Acid digestion and ICP-MS analysis

Prior to ICP-MS analysis the plant materials needed to be digested. The dried plant material/seeds were weighed and $\sim 100\text{mg}$ of material was left to soak in 4ml of 70% Trace Analysis Grade HNO_3 for 12 hours. After soaking, the samples were heated on a hot plate at

100°C for 150 minutes. After cooling, the acid was diluted to ~2% with ultrapure water (18.2MΩ cm; PURELAB ultra, ELGA LabWater), filtered and stored at room temperature. Samples were then sent to the University of Nottingham for ICP-MS analysis (machine and standard details described in Thomas et al., (2016)). For each digestion performed, two operational blanks and two samples of certified reference material (CRM) of leaf (WEPAL IPE 132 Broccoli/*Brassica oleracea*, LGC standards UK) were included. This basic digestion protocol was the same for all plant materials analysed in this study.

Results from the ICP-MS analysis were processed; the element specific operational blanks from each digestion were averaged and subtracted from each sample data point. These were then multiplied by initial sample volume and divided by the dry mass of the sample digested. From this value, the concentration of each element under analysis was calculated (as mg/kg). CRM inclusion allowed the percentage recovery of elements from each digestion to be calculated (by comparing returned element concentrations to certified concentrations). Any element which displayed poor recovery (<85%) was removed from analysis. Furthermore, for each run an element specific limit of detection (LOD) was calculated; the standard deviation of the operational blanks was calculated (by assuming a dry weight of 0.1g), and this value was multiplied by three to give the LOD. Any data points below the LOD were also excluded from further analysis. All analysis was performed using R (version 3.1.2) or GenStat (17th edition, VSNI).

3 Seed ionome investigation

3.1 Introduction

In the first part of this chapter results from AT analysis of individual element concentrations will be described (for Mo, S, Mn, Cu, Zn and Cd); giving the relative trait predictability of both SNP and GEM markers (2.3), followed by an in depth analysis of potential candidates within association regions. For reference, the heritability of each element concentration under investigation is included from analyses performed as part of Thomas et al., (2016), see **Table 2.4.2.a**. However, an investigation of individual elements (although common and useful in ascertaining specific biological functionality) fails to consider that elements within the ionome are not necessarily independent of one another. It is widely known within academic literature that oversupply or undersupply of one element can disrupt multiple other elements within the ionome. This is thought to be a consequence of shared uptake/ transport pathways, signalling cross-talk and interactions in the biological activation or roles each element plays (whether positive or negative)(Baxter, 2015). For example, it was previously found within *A. thaliana* that Root System Architecture (RSA) traits were differentially modified depending upon whether an element was deficient on its own or in tandem with other elements, highlighting interaction in signalling between different nutrient uptake mechanisms (Kellermeier et al., 2014). A more specific example in *B. napus* utilised plants grown under various nutrient deficiency conditions. It found many instances where nutrient deficiency caused an increase in another element (a relevant example being Mo, whose uptake was strongly increased under S, Fe, Cu, Zn, Mn and B deprivation (Maillard et al., 2016b)). Conversely, oversupply of an element can perturb the rest of the ionome. For example Cu excess in *A. thaliana* was shown to have broad effects on the whole plant ionome, disrupting both the root and shoot ionome differentially (increased Mg, Ca, Fe and Zn and decreased concentrations of K and S in roots vs a decrease in shoot K, Ca, P and Mn concentration) (Lequeux et al., 2010). Therefore if and when common candidates were identified between individual analyses, the two would be combined and investigated further. A full list of the candidates taken forward to analysis in *A. thaliana* is listed within the summary of AT results and conclusions (3.2.7), with details of which elements they will be tested for in the next section.

Table 2.4.2.a variance component analysis from Thomas et al., (2016) for seed mineral composition in *B. napus*.

Variation (as a %) is shown for each element under investigation associated with genotype, habit, experimental design and residual factors, calculated from Residual Maximum Likelihood (REML) analyses.

Response variate:	Cu	Cd	Mn	Zn	Mo	S
Genotype	32	9	12	21	41	31
Habit	0	8	1	12	21	13
Experimental design	6	40	55	14	7	37
Residual factors	62	42	32	53	31	19

3.2 Analysis of individual elements within seeds:

This section outlines the AT analysis of Cu, Cd, Mn, Zn, Mo and S concentrations in seed. It gives AT results and tests the predictive capability of markers, introducing potential candidate genes to be tested in *A. thaliana*.

3.2.1 Associative transcriptomic outputs, predictions and candidates: Cu concentration in seed

AT analysis for Cu concentration in seeds on the 383 diversity panel revealed three minor SNP association peaks (A3/C3, A7 and A9/C9) which all failed to clear the Bonferroni corrected significance threshold (**Figure 3.2.1.a**), see 2.2 for an overview of how AT Manhattan plots are generated and interpreted. There were no clear associations identified in GEM association analyses. This is unsurprising considering how little Cu concentration observed $-\log_{10}P$ values deviate from the expected $-\log_{10}P$ values within the GEM analysis (**Figure 3.2.1.b**).

Furthermore, when GEM and SNP AT analysis was performed on the smaller 274 diversity panel (see (0) for AT 274 outputs) to test for the predictive capability of markers for Cu concentration in seed, only two markers were found to be predictive to $p < 0.05$, **Table 3.2.1.a**: one SNP Bo3g052290.1:187:G (orthologue of AT4G02790: Ribosome biogenesis GTPase A) and one GEM, Cab022352.2 (with no orthologue information).

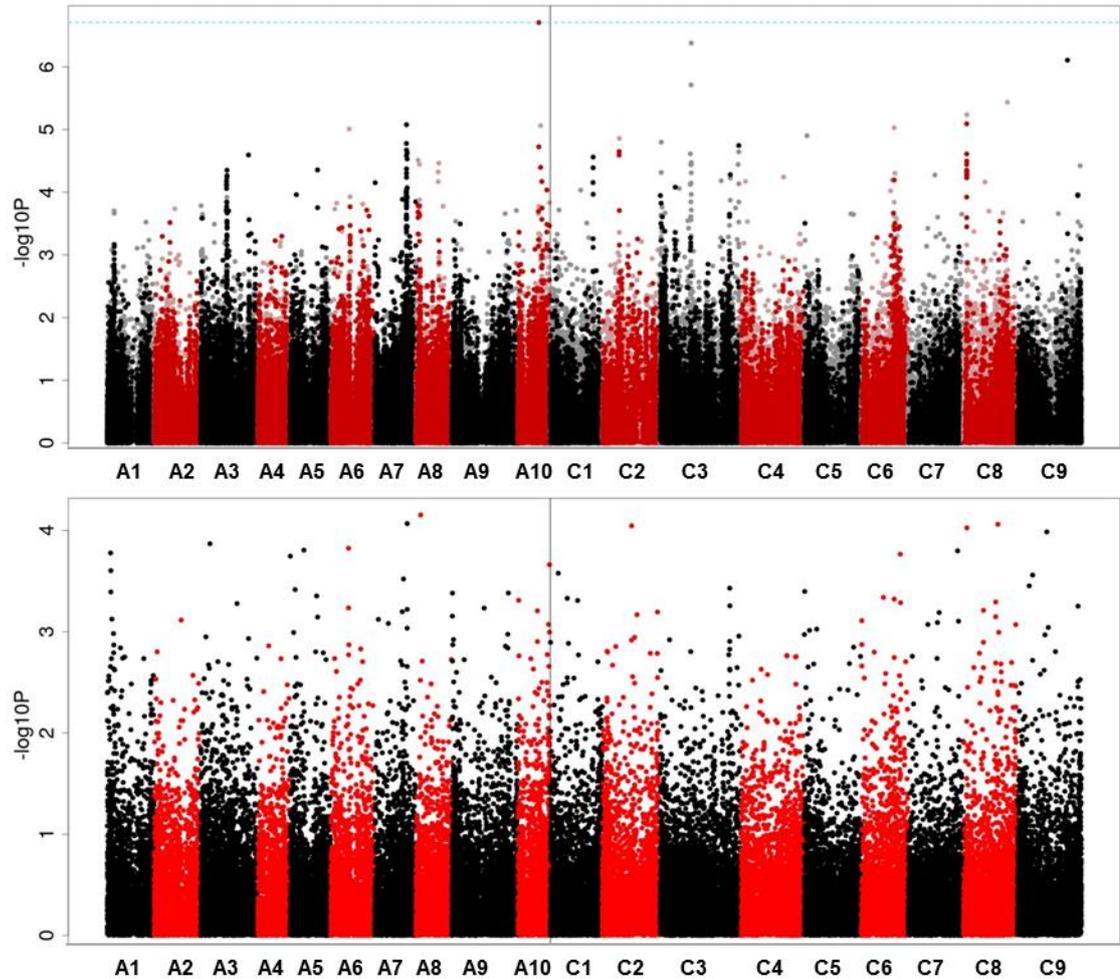


Figure 3.2.1.a Genome wide distribution of mapped markers associating with the Cu concentration in seeds (mg/kg DW) of all 383 accessions

The average concentration of Cu in seeds was calculated from 5 separate plants from ICP-MS analysis for each of the 383 accessions. SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and seed Cu concentration as the response variable. $-\text{Log}_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines only when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

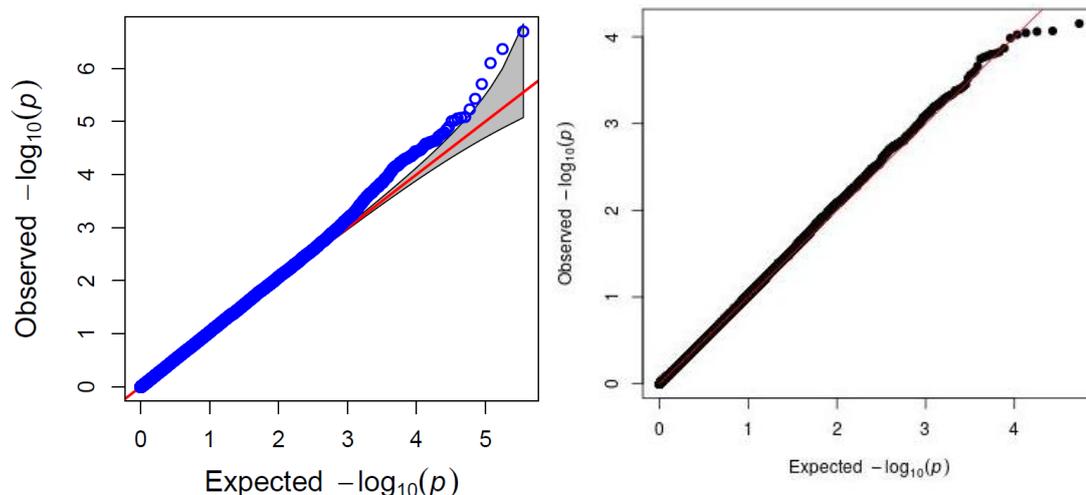


Figure 3.2.1.b Quantile-quantile plot of observed $-\log_{10}P$ values from AT SNP (left) and GEM (right) analysis for Cu concentration in seeds against expected $-\log_{10}P$ values

The red line indicates a theoretical perfect fit of the expected $-\log_{10}P$ values, while the grey area indicates the 95% confidence interval under the null hypothesis of no association between seed Cu and SNP markers.

Table 3.2.1.a Predictive capability of markers from Cu concentration in seed AT analysis.

For assessing the predictive capability of markers the highest SNP markers from discernible association peaks and the most highly associated GEMs were analysed. The marker type is given as either SNP or GEM, alongside their name and position, followed by their $-\log_{10}P$ value from the 274 AT analysis. Finally, the correlation coefficient (R), significance (p) and sample size (n) are given for the predictions made on the 109 diversity panel. Markers that were significantly predictive, $p < 0.05$, are highlighted in bold.

Marker type	Marker	Position	AT 274 $-\log_{10}P$	R	p	n
GEM	Cab007173.1	A10_019817543_019816069	5.03	0.047	0.628	108
GEM	Cab038228.1	A09_004386449_004385508	4.87	-0.032	0.746	108
GEM	Bo8g087210.1	C08_029468363_029471513	4.41	0.039	0.691	108
GEM	Cab022352.2	A08_003714499_003715422	4.39	0.257	0.007	108
SNP	Bo3g052290.1:187:G	C03_020492339_020494548	5.89	0.257	0.007	108
SNP	Bo9g181770.1:1848:G	C09_053777463_053782597	4.85	0.173	0.075	107
SNP	Cab003772.1:1203:A	A03_014714789_014717437	4.33	0.106	0.277	108
SNP	Cab010815.1:783:G	A03_032648901_032646552	3.91	-0.057	0.560	108

Despite only finding minor association peaks (which did not pass either the Bonferroni corrected significance or 5% false discovery rate thresholds) in the Cu concentration in seed AT results, a number of potential candidates was found. It is important to assess clear association peaks even if they do not pass significance thresholds or lack predictive markers due to the possibility of associations with rare alleles (which GWAS techniques have limited capacity to detect) and the limitation of analysing a split diversity panel (see **Error! Reference source not found.**). The best candidate was found within the predictive A3/C3 SNP association peak ($p < 0.05$ for markers under analysis); Bo3g053000.1 is an orthologue of *Heavy metal ATPase 5* in *A. thaliana* (*HMA5*: AT1G63440.1), a known Cu transporter involved in the Cu compartmentalisation and detoxification process (Andres-Colas et al., 2006). However *AtHMA5* is thought to be primarily expressed within the roots, where it is responsible for loading Cu^+ into the xylem (Andres-Colas et al., 2006; Kobayashi et al., 2008; Deng et al., 2013). Analysis of this candidate with *A. thaliana* has already been performed (Andres-Colas et al., 2006) and, although the seed was not tested for disruption to Cu concentration, it was found to be highly expressed in the flowers of *A. thaliana*. The authors of this paper surmised that this was likely to do with high expression within pollen (as per the “functional genomics of plant transporters” database <http://plantst.genomics.purdue.edu/>), however its occurrence in the Cu concentration in seed AT SNP association analyses provides an interesting addendum. Nevertheless, it was determined that assessing the seed of *A. thaliana* T-DNA insert mutants for the effects of this gene would be unhelpful, as even if disruption was observed it would be impossible to distinguish whether this was a seed specific effect or a result of whole plant disruption in Cu concentration (since its primary effects are within the roots). Performing a Cu tracer experiment at seed loading with *HMA5* disrupted would allow its role in seed to be determined.

However, *HMA5* was not the only potential Cu candidate within the A3/C3 SNP association region. A candidate with the description “copper transport protein family” (AT4G05030) was found which would be within the correct AT SNP association region (but does not have an associated CDS gene model within the *B. napus* pan-transcriptome). It is predicted to have metal ion binding capacity but was not identified as a Heavy metal-associated isoprenylated plant protein (HIPP; a type of metallochaperone with a metal binding domain and C-terminal isoprenylation motif) (De Abreu-Neto et al., 2013). Nonetheless, it was decided that it would be worthwhile to test this within *A. thaliana* as it is poorly defined in the literature.

The A7 SNP association peak had a number of potential candidates. The most promising candidate from this region was Cab018316.1, which is an orthologue of *IAR1* (*IAA-Alanine*

resistant 1: AT1G68100), which has metal ion transmembrane transporter activity and is part of the LIV-1 zinc transporter (LZRT) subfamily of the ZIP family (Taylor and Nicholson, 2003). It was originally identified during screening for IAA mutants from chemical (EMS) mutagenesis. Alongside an implied role in IAA- amino acid conjugate metabolism, it was found to share similarities to ZIP transporters but failed to rescue a *Saccharomyces cerevisiae zrt1 zrt2* double mutant when expressed in low Zn conditions (in contrast to other ZIP genes) (Lasswell et al., 2000; Grotz et al., 1998). Subsequent research has suggested that IAR1 works antagonistically with MTP5 (AT3G12100, *Metal Transport Protein 5*) which is a cation diffusion facilitator, regulating metal homeostasis into cellular compartments where IAA- amino acid conjugate hydrolysis occurs (the hydrolases responsible requiring metal cofactors) (Rampey et al., 2013). Further, when *IAR1* T-DNA insertional mutants were assessed they displayed disruption in multiple elements within their leaves, including both Cu and Zn (according to the PiiMS database (Baxter et al., 2007a), <http://www.ionomicshub.org/home/PiiMS>). However, given that the localisation of IAR1 action is still unknown, it was considered prudent to test whether this was a leaf specific effect or whether it would also show disruption within the seed (as suggested by the AT results). It is widely known that transition elements can share and compete for transporters (particularly those present as divalent cations); therefore a range of elements was assessed in the seed of the *A. thaliana* T-DNA lines (including Cu, Zn, Fe, Mn and Cd)(Hall and Williams, 2003).

3.2.2 Associative transcriptomic outputs, predictions and candidates: Cd concentration in seed

Cd concentration in seed AT analysis for the 383 diversity panel (**Figure 3.2.2.a**) produced a number of association peaks in the SNP association analysis which failed to clear the Bonferroni corrected significance threshold of 0.05 (A3/C3 and A8). However, a number of GEMs passed both the 5% false discovery rate (FDR) and Bonferroni thresholds in the GEM association analysis. Despite not clearing the Bonferroni corrected significance threshold, two SNP markers were found that were predictive for Cd concentration in seeds ($p < 0.05$); one on A3 and one on C9 (**Table 3.2.2.a**). For the GEM association analysis, three significantly predictive markers were found ($p < 0.05$); one on A9, C1 and C5. Both these results are in contrast to where the predictive markers would be expected from the 383 GEM association analyses (i.e. within the association peaks displayed on the 383 AT plots). When the AT Manhattan association plots for the 383 diversity panel (**Figure 3.2.2.a**) and the 274 diversity panel (**Figure 3.2.2.b**) were compared; only the SNP association peaks on A3/C3 remained consistent. As such the search for potential candidates was limited to this area, as it was most likely to give candidates with a role related to the trait.

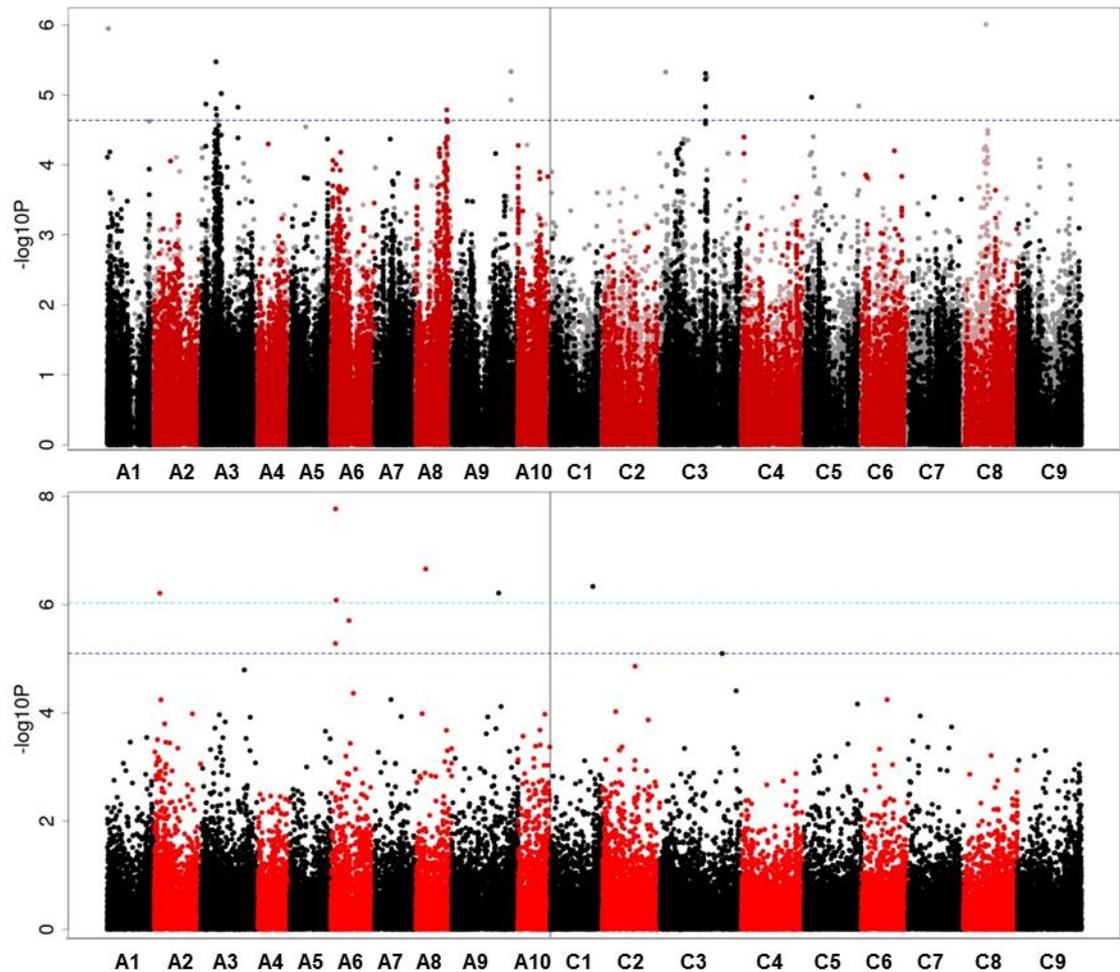


Figure 3.2.2.a Genome wide distribution of mapped markers associating with the Cd concentration in seeds (mg/kg DW) of all 383 accessions.

The average concentration of Cd in seeds was calculated from 5 separate plants from ICP-MS analysis for each of the 383 accessions. SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and seed Cd concentration as the response variable. $-\text{Log}_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

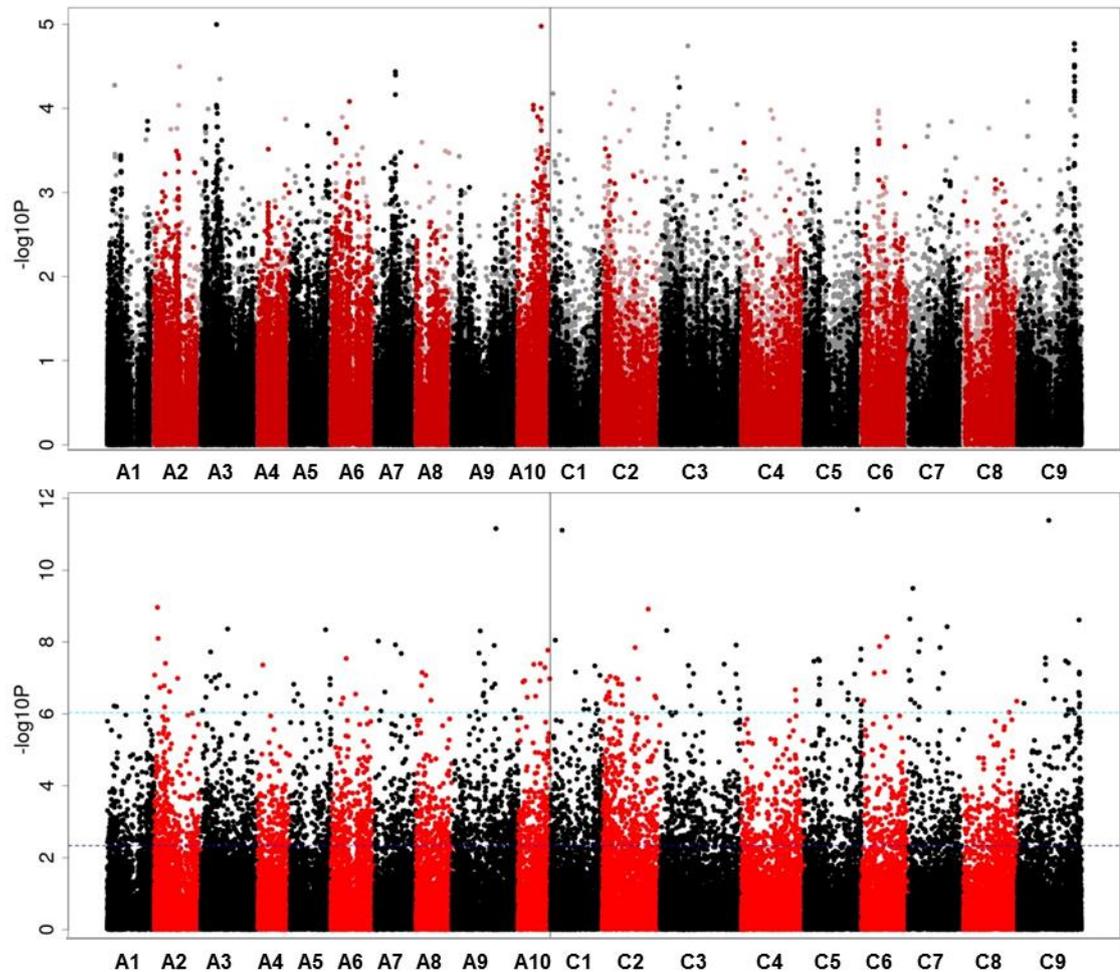


Figure 3.2.2.b Genome wide distribution of mapped markers associating with the Cd concentration in seeds (mg/kg DW) of 274 accessions.

The average concentration of Cd in seeds was calculated from 5 separate plants from ICP-MS analysis for each of the 274 accessions. SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and seed Cd concentration as the response variable. $-\text{Log}_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

Table 3.2.2.a Predictive capability of markers from Cd concentration in seed AT analysis.

For assessing the predictive capability of markers the highest SNP markers from discernible association peaks and the most highly associated GEMs were analysed. The marker type is given as either SNP or GEM, alongside their name and position, followed by their $-\log_{10}P$ value from the 274 AT analysis. Finally, the correlation coefficient (R), significance (p) and sample size (n) are given for the predictions made on the 109 diversity panel. Markers that were significantly predictive ($p < 0.05$) are highlighted in bold.

Marker type	Marker	Position	AT 274 $-\log_{10}P$	R	p	n
GEM	Bo5g143470.1	C05_044454166_044456461	11.72	0.199	0.039	108
GEM	Bo9g088240.1	C09_027093568_027093892	11.32	0.186	0.054	108
GEM	Cab000382.1	A09_033333496_033335711	11.31	0.194	0.044	108
GEM	Bo1g021550.1	C01_007680298_007683081	11.03	0.196	0.042	108
SNP	Bo9g171570.1:189:C	C09_050570395_050570844	4.95	0.343	<0.001	102
SNP	Cab008036.4:797:T	A10_015636261_015631086	4.93	0.182	0.091	87
SNP	Cab023006.1:1851:A	A06_003183873_003190307	3.65	0.140	0.149	107
SNP	Cab002648.1:216:C	A03_008522315_008523559	4.94	0.487	<0.001	98
SNP	Cab015851.3:354:T	A03_002505534_002507914	3.86	0.092	0.374	96

Within the A3/C3 SNP association region, the best candidate found was Bo3g083230.1 which is an orthologue of AT3G24450 (described as a “heavy metal transport/detoxification superfamily protein”). It contains a heavy metal associated domain (HMA, IPR006121) which is thought to be involved in Cu ion transport/binding. There has been no research focused upon this gene in the literature (i.e. it only appears to have been studied as part of genome wide analyses, such as Fabro et al., (2008)) and it does not seem to have been characterised in *A. thaliana*. Cu is known to indirectly affect plant Cd sensitivity (Carrió-Seguí et al., 2015); as such it was decided to test this candidate further with *A. thaliana* T-DNA insert mutants for effects on both Cu and Cd concentration.

GEM association outputs from the 383 diversity panel were assessed despite their inconsistency with the 274 GEM association outputs (again emphasising the issues of analysing a split diversity panel, see 7.1). However all of the GEMs which passed the Bonferroni and 5% FDR thresholds were unannotated for orthologues in *A. thaliana*. As such they were compared against the rest of the pan-transcriptome and across species in the NCBI database with BLAST (see 2.2.2). Two of the GEMs assessed this way came up with similarities to Zn ion binding proteins. Bo1g124760 was similar to 10 other CDS gene models within the pan-transcriptome whose orthologues were described in *A. thaliana* as AT1G52300 zinc-binding ribosomal protein family protein (varying from an alignment score of S' 183, E: 5e-45 (Cab022661) down to S' 91.6, E: 3e-17 (Cab013044)). Cab036431 aligned well with 3 other CDS gene models

whose orthologues are described in *A. thaliana* as AT5G16470.1 zinc finger (C2H2 type) family protein (varying from an alignment score of S' 577, E: 3e-163 (Cab008105.2) to S' 449, E: 6e-125 (Bo3g012060.1)). Cd and Zn are chemically very similar and thought to be able to compete for shared transport pathways and binding sites (Clemens, 2006), although the literature on Cd incorporation into Zn fingers is sparse (Kluska, Adamczyk and Krężel, 2018). Considering that these analyses took place under non-stress Cd conditions, it perhaps highlights the close association between the two elements or the sensitivity of seed tissues to Cd levels (particularly since AT5G16470.1 is thought to be involved in the ROS response/ stress tolerance (Grotz et al., 1998)). Despite this, as the actual GEM markers from the association analysis were unannotated, analysis of these related CDS gene model orthologues was not taken any further.

3.2.3 Associative transcriptomic outputs, predictions and candidates: Mn concentration in seed

As previously observed in Cu and Cd seed concentration AT analyses, SNP and GEM AT analyses of Mn concentration in seeds failed to pass the Bonferroni corrected significance threshold of 0.05. Nevertheless, three distinct association peaks can be observed within the SNP association outputs (**Figure 3.2.3.a**) on A1, A3 and A5 which pass the 5% FDR threshold. The presence of these peaks in the SNP association analyses is easy to understand when Q-Q plots are observed (**Figure 3.2.3.b**) which show a marked deviation from the expected p values for the SNP markers and minimal aberration for the GEMs (which show no clear associations within AT outputs). Note, the difference in p values between the Q-Q plot and Manhattan AT output is a consequence of filtering out SNPs with a minor allele frequency of <1%. However when SNP associations were investigated further in the AT 274 diversity panel, it was found that the top 42 most highly associated SNP markers had a minor allele frequency of less than 5%, making it impossible to test the predictive capacity of the SNP markers or find candidates in these regions. The most highly associated GEMs were assessed for trait predictability and two were found to correlate with the Mn concentration in seeds, $p < 0.05$ (**Table 3.2.3.a**).

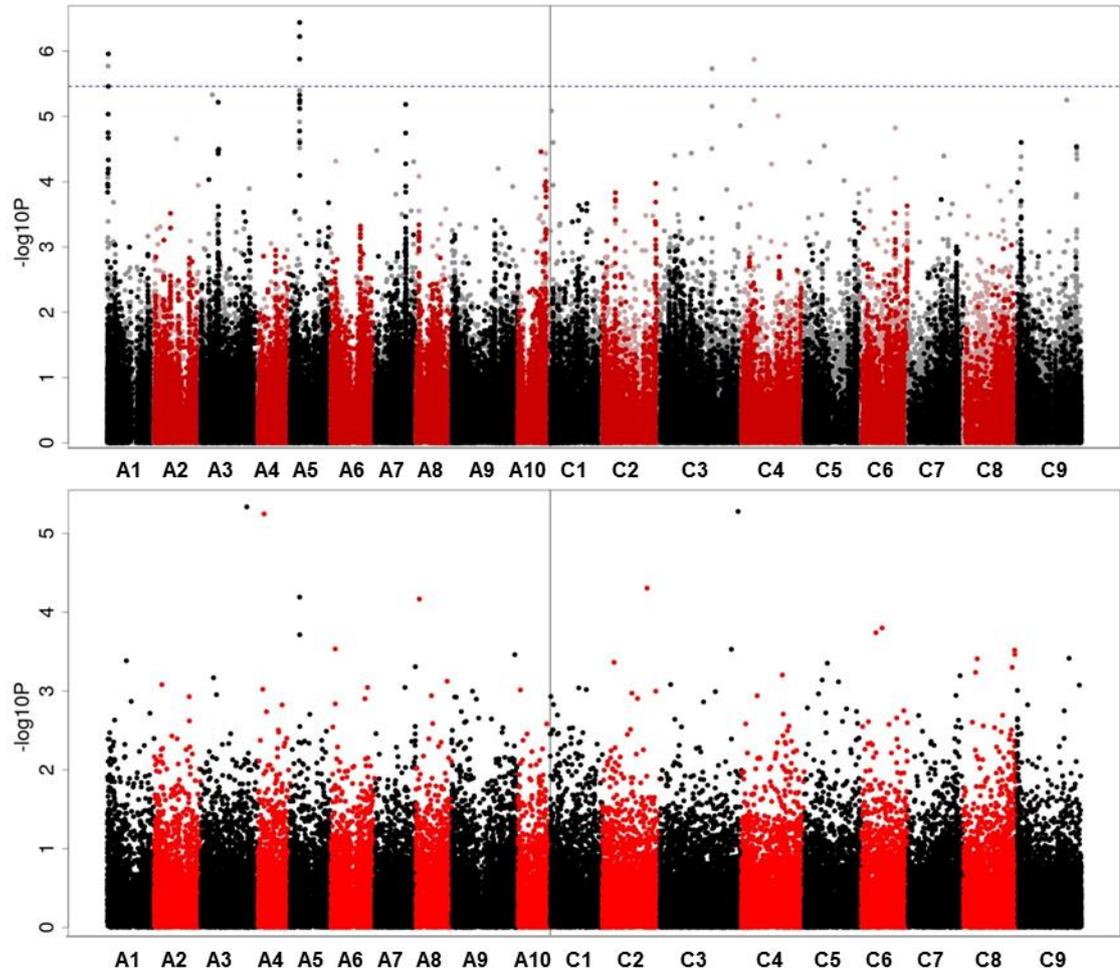


Figure 3.2.3.a Genome wide distribution of mapped markers associating with the Mn concentration in seeds (mg/kg DW) of all 383 accessions.

The average concentration of Mn in seeds was calculated from 5 separate plants from ICP-MS analysis for each of the 383 accessions. SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and seed Mn concentration as the response variable. $-\text{Log}_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

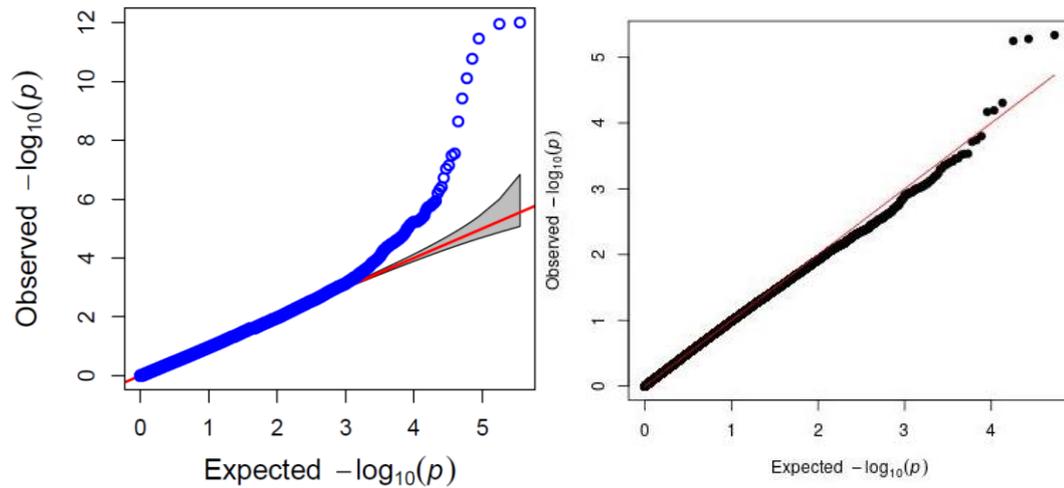


Figure 3.2.3.b Quantile-quantile plot of observed $-\log_{10}P$ values from AT SNP analysis (left) and AT GEM analysis (right) for Mn concentration in seeds against expected $-\log_{10}P$ values.

The red line indicates a theoretical perfect fit of the expected $-\log_{10}P$ values, while the grey area indicates the 95% confidence interval under the null hypothesis of no association between seed Mn and the SNPs/GEMs.

Table 3.2.3.a Predictive capability of markers from Mn concentration in seed AT analysis.

No predictive markers could be found from SNP association analysis, while the top 6 GEMs were assessed. The marker type is indicated (GEM), alongside their name and position, followed by their $-\log_{10}P$ value from the 274 AT analysis. Finally, the correlation coefficient (R), significance (p) and sample size (n) are given for the predictions made on the 109 panel. Markers that were significantly predictive ($p < 0.05$) are highlighted in bold.

Marker type	Marker	Position	AT 274 $-\log_{10}P$	R	p	n
GEM	Bo3g178370.1	C03_062203601_062211708	7.84	0.017	0.864	108
GEM	Cab002086.1	A03_026944047_026945855	5.88	0.207	0.032	108
GEM	Cab033883.1	A05_006148761_006148075	5.51	0.037	0.708	108
GEM	Bo2g134920.1	C02_042824094_042824996	5.16	0.157	0.104	108
GEM	BnaA08g02500D	A08_002633555_002631912	5.084	0.0674	0.488	108
GEM	Cab019042.1	A04_004690992_004686381	4.84	0.223	0.021	108

Despite the low predictive capability of the SNP markers, it was observed that the most highly associated SNP marker was on C3 (for both AT 383 and 274), denoted as Bo3g032060 ($-\log_{10}P$: 12.02, minor allele frequency: 0.004). This marker corresponds to an orthologue in *A. thaliana* called *COPT4* (At2g37920.1), a copper ion transmembrane transporter (Sancenon et al., 2003). It has been disputed whether this is in fact a Cu transporter in *A. thaliana* (as it lacks key methionine residues to perform its function as a Cu^+ transporter (Puig et al., 2007a)), however in rice it is thought to act cooperatively with other COPT proteins to mediate high affinity Cu transport (Yuan et al., 2011). There is no evidence of it functioning in the transport of other elements; however it has been shown to be transcriptionally suppressed in roots by Mn and Zn deficiency (Yuan et al., 2011). However as *COPT4* is thought to be involved in embryo development, leading to the development arrest of mutant lines, it was not analysed further in this study (TAIR, 2015).

The two weakly predictive GEMs ($p < 0.05$) corresponded to candidates with only minimal biological significance for seed Mn concentration. The marker known as Cab002086.1, has an orthologue in *A. thaliana* annotated as a *sodium bile acid symporter family (BASS6: AT4G22840.1)*, which when disrupted shows decreased photosynthesis and growth under ambient CO_2 (South et al., 2017). Mn plays an essential role in photosynthesis as part of the oxygen-evolving complex of photosystem II, which could explain the link to *BASS6*. However, when T-DNA lines were assessed on the PiiMs database (Baxter et al., 2007a), Mn concentrations were not found to be significantly disrupted in leaves. As this is the primary site of photosynthesis, which was the hypothesised link between *BASS6* and Mn concentrations in the seeds, it was decided not to pursue investigation of this within the seed. Furthermore, the second GEM did not have a clear biological role in Mn concentration variation either. Cab019042.1, orthologue of AT3G53090.2/ *ubiquitin-protein ligase 7 (UPL7)*, could potentially be involved in Mn concentration by targeting Mn containing proteins for degradation (which would be important for seed Mn concentrations as it is a highly immobile element within plants). However this was considered too tangential to warrant further investigation of this gene.

3.2.4 Associative transcriptomic outputs, predictions and candidates: Zn concentration in seed

AT outputs for the Zn concentration in seeds did not show any clear association peaks and failed to clear the Bonferroni corrected significance threshold (**Figure 3.2.4.a**). No markers could be found within the SNP association analysis for predictions and only the highest GEM showed weak predictive capacity in GEM association analyses (Bo3g171060.1, Localisation: C03_060372458_060372949, AT 274 $-\log_{10}P$: 7.97, R: 0.213, p: 0.0266, n:108). Unfortunately this GEM is unannotated and no functional information is available in *A. thaliana*. Interestingly, three minor association peaks within the GEM AT Manhattan outputs on A9/C9 and C2 occur within a region previously identified in an AT study of glucosinolates (GSL) in seed (Lu et al., 2014). These regions also appears to a greater extent in Mo and S seed concentration AT analyses (see 3.2.5 and 3.2.6) and has been analysed in greater depth for candidates as part of these studies. It is interesting that the Zn AT outputs did not show any relation to the Cd AT outputs as the relationship between Zn and Cd concentration in plant tissues is well documented (see 1.2.5.6). This could be a consequence of *B. napus* being relatively tolerant to Cd, possibly through effective exclusion mechanisms (e.g. ineffective root to shoot translocation (Nouairi et al., 2006)). There is no correlation between Zn and Cd concentrations in the seeds (R: -0.05, p=0.38, n: 380) but there is a positive correlation in the leaves (see 5.1 for the leaf interactome, R: 0.41, p<0.001, n: 385). This perhaps fits with the theory that *B. napus* has poor shoot Cd translocation, as when the ratio of the two element concentrations in the seeds and leaves are compared the net accumulation of Zn is increased in comparison to Cd (Zn:Cd 200:1, where $([Zn]_{seed}/[Cd]_{seed})/([Zn]_{leaf}/[Cd]_{leaf})$, as explained in Thomas et al., (2016))

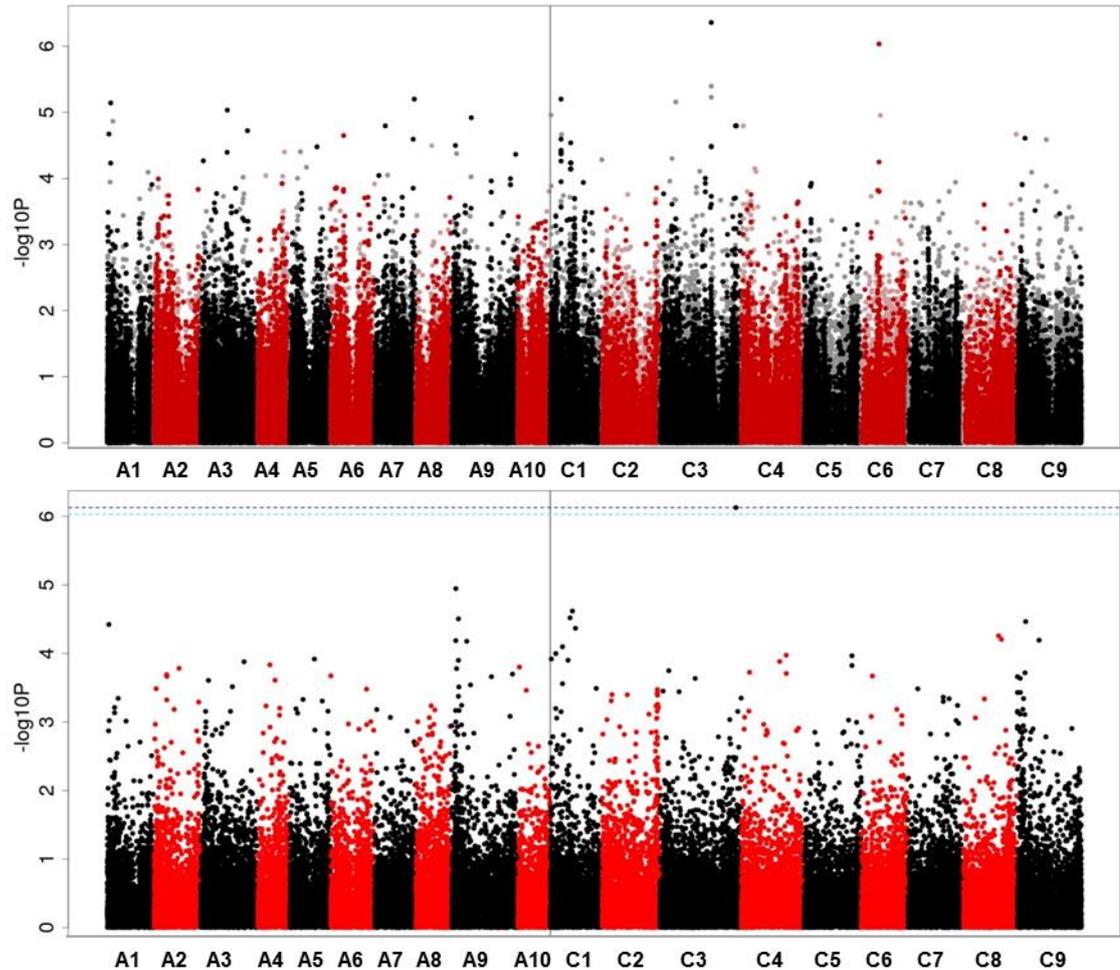


Figure 3.2.4.a Genome wide distribution of mapped markers associating with the Zn concentration in seeds (mg/kg DW) of all 383 accessions.

The average concentration of Zn in seeds was calculated from 5 separate plants from ICP-MS analysis for each of the 383 accessions. SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and seed Zn concentration as the response variable. $-\text{Log}_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

3.2.5 Associative Transcriptomic outputs, predictions and candidates: Mo concentration in seed

When AT analysis for Mo concentrations within seeds was performed (**Figure 3.2.5.a**), some distinct similarities were observed to both seed S concentrations (see **Figure 3.2.6.a**) and previous work on GSL in seeds (Lu et al., 2014). Association peaks observed on A2/C2 and A9/C9 in the SNP association analysis (and to a much lesser extent in the GEM association analysis) closely resemble those of the major aliphatic glucosinolate regulator *HAG1/Myb28*. Despite not passing the Bonferroni corrected significance threshold, both SNP and GEM markers were found which were capable of predicting not only for Mo, but for S, Mg, B and Zn concentrations in seed in more than one instance (**Table 3.2.5.a**). In addition to the known GSL association peaks an additional association peak on A7 was observed within the SNPs. However when further investigated the minor allele frequency of all the most highly associated SNPs assessed were significantly small, preventing predictive assessment for this association region. Since no Mo specific candidates could be found and the similarities observed between the AT associations in SNP and GEM analyses, candidates for Mo concentration in seeds will be discussed as part of S concentration in seed AT analyses in the next section.

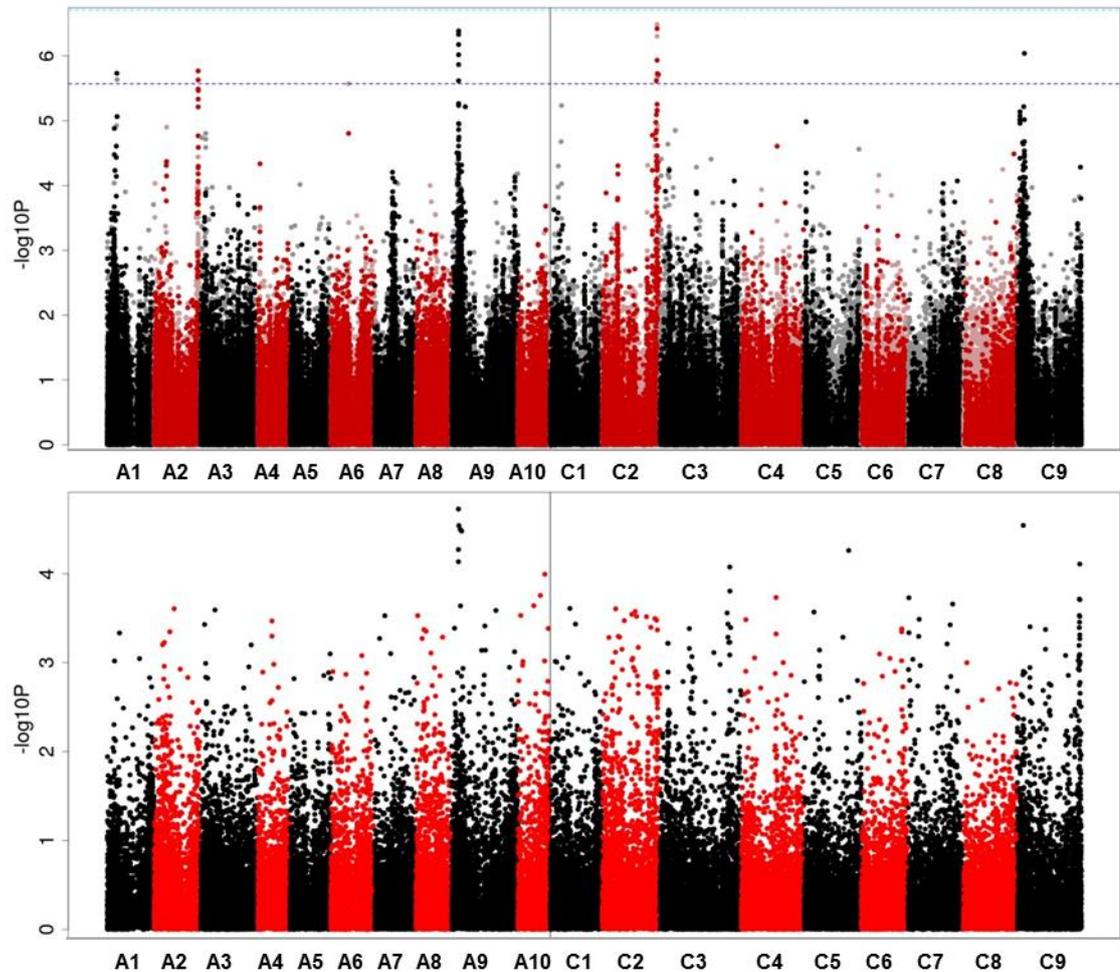


Figure 3.2.5.a Genome wide distribution of mapped markers associating with the Mo concentration in seeds (mg/kg DW) of all 383 accessions.

The average concentration of Mo in seeds was calculated from 5 separate plants from ICP-MS analysis for each of the 383 accessions. SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and seed Mo concentration as the response variable. $-\text{Log}_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

Table 3.2.5.a Predictive capability of markers from Mo concentration in seed AT analysis.

For assessing the predictive capability of markers the highest SNP markers from discernible association peaks and the most highly associated GEMs were analysed.. The marker type is given as either SNP or GEM, alongside their name and position, followed by their $-\log_{10}P$ value from the 274 AT analysis. The correlation coefficient (R) and sample size (n) are given for the predictions made on the 109 panel for Mo, S, Mg, B and Zn. Markers that were significantly predictive ($p < 0.05$) are indicated with: $p < 0.001$ (red***), $p < 0.01$ (blue**) and $p < 0.05$ (*).

Marker type	Marker	Position	AT 274 - $\log_{10}P$	Mo (R)	S (R)	B (R)	Mg (R)	Zn (R)	n
GEM	Bo9g012580.1	C09_003794670_003806441	12.02	0.434* **	- 0.526* **	0.199*	0.239*	-0.202*	108
GEM	Bo9g181240.1	C09_053471789_053472822	12.79	0.407* **	-0.123	0.036	0.268* *	-0.227*	108
GEM	Cab038259.1	A09_004138078_004136068	14.46	0.339* **	- 0.557* **	0.354* **	0.227*	- 0.268* *	108
GEM	Cab038300.1	A09_003880147_003877462	15.69	0.310* *	- 0.703* **	0.342* **	0.205*	-0.215*	108
GEM	Cab038301.1	A09_003874397_003872142	14.62	0.284* *	- 0.661* **	0.361* **	0.240*	-0.233*	108
SNP	Bo2g161640.1:1011:T	C02_050931684_050934267	5.16	0.417* **	- 0.483* **	0.271* *	0.171	-0.125	108
SNP	Bo9g014770.1:2001:T	C09_004562333_004565949	4.38	0.395* **	- 0.586* **	0.264* *	0.212*	-0.212*	108
SNP	Cab038280.1:2250:T	A09_003976556_003972924	5.88	0.379* **	- 0.676* **	0.319* **	0.200*	- 0.312* *	106
SNP	Cab038280.1:1671:A	A09_003976556_003972924	5.94	0.371* **	- 0.669* **	0.314* *	0.197*	- 0.294* *	106
SNP	Bo2g161640.1:228:C	C02_050931684_050934267	4.82	0.363* **	- 0.603* **	0.248*	0.279* *	-0.15	106
SNP	Bo2g163990.1:1105:T	C02_051188221_051191814	5.85	0.361* **	- 0.514* **	0.211*	0.248*	-0.169	105
SNP	Cab038279.1:738:T	A09_003982037_003978463	5.52	0.353* **	- 0.617* **	0.286* *	0.147	- 0.273* *	106
SNP	Cab038280.1:2481:C	A09_003976556_003972924	5.21	0.346* **	- 0.655* **	0.315* *	0.158	- 0.266* *	105
SNP	Cab038279.1:2649:A	A09_003982037_003978463	6.51	0.340* **	- 0.646* **	0.352* **	0.157	- 0.294* *	106
SNP	Bo9g014770.1:1455:C	C09_004562333_004565949	4.3	0.339* **	- 0.610* **	0.330* **	0.211*	- 0.253* *	105
SNP	Cab021711.1:1953:A	A02_033148871_033144959	4.12	0.313* *	- 0.503* **	0.225*	0.277* *	-0.14	103
SNP	Cab038280.1:1212:T	A09_003976556_003972924	5.6	0.251*	- 0.563* **	0.259*	0.149	-0.165	98
SNP	Cab021724.1:1682:G	A02_033083524_033079765	4.06	0.227*	- 0.535* **	0.212*	0.038	0.028	108

3.2.6 Associative Transcriptomic outputs, predictions and candidates: S concentration in seed

AT analysis for S concentration in seed material was the only analysis within the seeds to pass the Bonferroni corrected significance threshold for both SNP and GEM association outputs (**Figure 3.2.6.a**). Five clear association peaks are observable on A2/C2, A9/C9 and C7 in both the SNP and GEM AT Manhattan outputs, seemingly coinciding with GSL association peaks containing *HAG1/Myb28* (a major aliphatic GSL regulator). Numerous predictive markers could be found from the AT analysis for both SNPs and GEMs ($p < 0.05$), however they were not only predictive for S seed concentration, they were also capable of predicting for a number of other element concentrations within the seed (**Table 3.2.6.a**): namely Mo (which displays distinct similarities in AT outputs), as well as B, Mg, Mn, Sr and Zn (see AT Zn GEM outputs) in more than one instance for both the SNPs and GEMs.

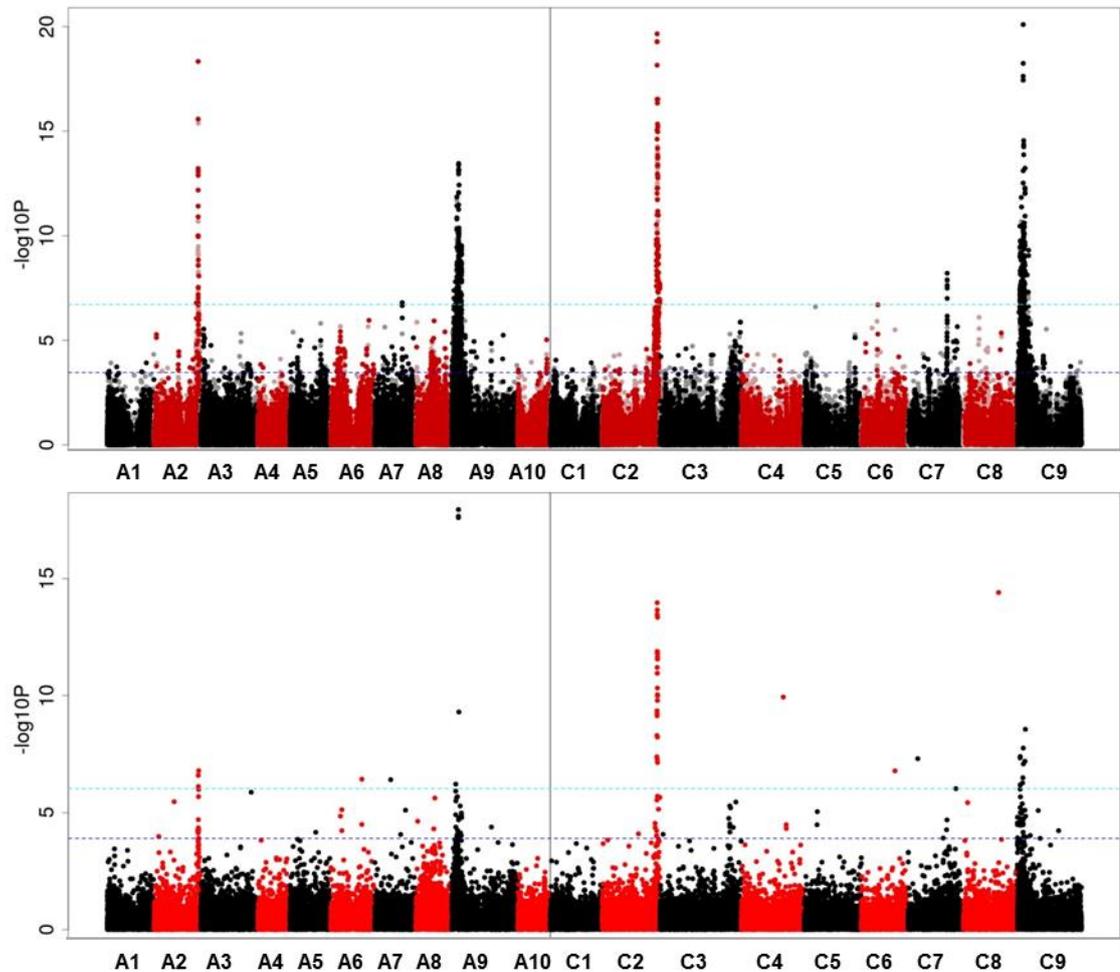


Figure 3.2.6.a Genome wide distribution of mapped markers associating with the S concentration in seeds (mg/kg DW) of all 383 accessions.

The average concentration of S in seeds was calculated from 5 separate plants from ICP-MS analysis for each of the 383 accessions. SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and seed S concentration as the response variable. $-\text{Log}_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

Table 3.2.6.a Predictive capability of markers from S concentration in seed AT analysis.

For assessing the predictive capability of markers the highest SNP markers from discernible association peaks and the most highly associated GEMs were analysed. The marker type is given as either SNP or GEM, alongside their name and position, followed by their $-\log_{10}P$ value from the 274 AT analysis. The correlation coefficient (R) and sample size (n) are given for the predictions made on the 109 panel for S, Mo, B, Mg, Mn, Sr and Zn concentration in seeds. Markers that were significantly predictive are indicated with: $p < 0.001$ (red***), $p < 0.01$ (blue**) and $p < 0.05$ (*).

Marker type	Marker	Position	AT 274 -log10P	S (R)	Mo (R)	B (R)	Mg (R)	Mn (R)	Sr (R)	Zn (R)	n
GEM	Cab038300 .1	A09_003880147 _003877462	31.84	0.703 ***	- 0.31* *	- 0.342 ***	- 0.20 5*	- 0.08 8	- 0.01 9	0.21 5*	1 0 8
GEM	Cab038301 .1	A09_003874397 _003872142	33.06	0.661 ***	- 0.284 **	- 0.361 ***	- 0.24 *	- 0.02 3	0.00 2	0.23 3*	1 0 8
GEM	Cab038298 .3	A09_003885476 _003883159	35.89	0.635 ***	- 0.344 ***	- 0.239 *	- 0.24 *	0.05 5	- 0.07 8	0.26 8**	1 0 8
GEM	Bo2g16179 0.1	C02_051022984 _051026770	34.51	0.559 ***	- 0.248 *	- 0.262 **	- 0.21 4*	- 0.15 5	0.26 5**	0.11 6	1 0 8
GEM	Bo2g16185 0.1	C02_051074261 _051077445	35.73	0.537 ***	- 0.249 **	- 0.209 *	- 0.19 8*	- 0.10 6	0.19 7*	0.07	1 0 8
GEM	Bo4g14236 0.1	C04_038656985 _038658185	23.44	0.522 ***	- 0.277 **	- 0.174	- 0.23 5*	- 0.06 7	0.16 7	0.17 3	1 0 8
GEM	Bo8g09126 0.1	C08_030818898 _030819173	36.59	0.514 ***	- 0.313 ***	- 0.182	- 0.16 5	- 0.19 1*	0.07 7	0.19 7*	1 0 8
GEM	Cab021710 .1	A02_033160073 _033155393	15.17	0.462 ***	- 0.343 ***	- 0.176	- 0.14 2	- 0.23 *	0.02	0.06 6	1 0 8
GEM	Bo9g01776 0.1	C09_005223700 _005226066	21.79	0.435 ***	- 0.129	- 0.113	- 0.18 8	- 0.19 1*	0.28 8**	0.12 9	1 0 8
GEM	Cab021700 .1	A02_033237611 _033234397	17.47	0.341 ***	- 0.151	- 0.116	- 0.05 9	- 0.16 4	0.14 4	- 0.05 3	1 0 8
GEM	Cab021665 .1	A02_033436338 _033434395	19.45	0.315 ***	- 0.268 **	- 0.093	- 0.22 2*	- 0.21 5*	0.04 3	0.17 5	1 0 8
SNP	Bo9g01154 0.1:93:A	C09_003764308 _003765093	13.45	0.79* **	- 0.276 **	- 0.292 **	- 0.25 5*	- 0.25 *	0.13 8	0.26 2*	8 9
SNP	Bo9g01153 0.1:329:T	C09_003756156 _003757704	13.89	0.772 ***	- 0.368 ***	- 0.267 **	- 0.30 9**	- 0.17 3	0.07 2	0.24 8*	1 0 0
SNP	Cab038280 .1:2481:C	A09_003976556 _003972924	10.21	0.656 ***	- 0.322 ***	- 0.33* **	- 0.16 9	- 0.12 5	0.03 4	0.28 2**	1 0 5
SNP	Cab038279 .1:2649:A	A09_003982037 _003978463	10.66	0.645 ***	- 0.309 **	- 0.335 ***	- 0.17 9	- 0.06 5	0.05 9	0.30 7**	1 0 6
SNP	Cab021725 .1:2274:G	A02_033078961 _033071276	13.03	0.642 ***	- 0.512 ***	- 0.314 **	- 0.25 3*	- 0.26 5**	0.14	0.37 4***	1 0 2

Marker type	Marker	Position	AT 274 -log10P	S (R)	Mo (R)	B (R)	Mg (R)	Mn (R)	Sr (R)	Zn (R)	n
SNP	Cab038235 .2:849:G	A09_004336787 _004332190	10	0.64* **	- 0.322 **	- 0.28* *	- 0.27 8**	- 0.11 2	- 0.03 4	0.39 1***	9 7
SNP	Bo2g16164 0.1:147:C	C02_050931684 _050934267	16.36	0.583 ***	- 0.241 *	- 0.28* *	- 0.14 2	- 0.11 4	0.12 4	0.16 5	1 0 5
SNP	Bo9g01739 0.1:357:G	C09_004997946 _004999585	11.48	0.552 ***	- 0.179	- 0.241 *	- 0.24 5*	- 0.18 8	0.17 4	0.16 4	8 5
SNP	Bo2g16164 0.1:246:A	C02_050931684 _050934267	16.83	0.55* **	- 0.263 **	- 0.264 **	- 0.17 8	- 0.12 3	0.15 6	0.14 8	1 0 6
SNP	Cab021724 .1:682:G	A02_033083524 _033079765	12.49	0.535 ***	- 0.227 *	- 0.212 *	- 0.03 8	- -0.1	0.21 2*	- 0.02 8	1 0 8
SNP	Cab021711 .1:1953:A	A02_033148871 _033144959	13.2	0.503 ***	- 0.313 **	- 0.225 *	- 0.27 7**	- 0.19 8*	0.09 7	0.14	1 0 3
SNP	Cab038394 .1:345:C	A09_003376623 _003374357	9.8	0.432 ***	- 0.199 *	- 0.166	- 0.18 4	- 0.15 9	0.19 3*	0.21 8*	1 0 8
SNP	Bo7g09866 0.1:933:G	C07_038634261 _038636719	6.53	0.384 ***	- 0.339 ***	- 0.162	- 0.24 6*	- 0.09 8	0.11 5	0.21 5*	1 0 7
SNP	Bo7g09866 0.1:989:A	C07_038634261 _038636719	6.26	0.376 ***	- 0.334 ***	- 0.154	- 0.24 8*	- 0.09 7	0.10 4	0.21 2*	1 0 7
SNP	Bo7g09870 0.1:2103:C	C07_038647541 _038654259	7.03	0.345 ***	- 0.342 ***	- 0.219 *	- 0.33 1**	- 0.09 4	0.04 9	0.23 1*	9 2
SNP	Cab038464 .1:2103:G	A09_003008607 _003003998	9.1	0.316 **	- 0.222 *	- 0.054	0.04 8	0.14 7	- 0.12 4	0.05 8	1 0 2
SNP	Cab045525 .1:1764:C	A05_023161730 _023166769	6.5	0.194 *	- 0.264 **	- 0.081	- 0.13 4	- 0.14 8	0.05	0.13 2	1 0 8

The most obvious candidate for S concentration in seeds is *HAG1/Myb28* (At5G61420), since S is a constituent of GSL and it is well documented that variation in the GSL concentration of seeds also affects the S concentration, to the extent that S concentration can be used to estimate GSL content in the seeds of *B. napus* (Bloem, Haneklaus and Schnug, 2005). The close similarity of the Mo AT results to the association regions around the CDS gene models whose orthologues correspond to *HAG1/Myb28* could be explained in a number of ways. For example, chemical similarity of SO_4^{2-} and MoO_4^{2-} could result in increased Mo concentration in seeds as a consequence of sulfate deficiency or via the role of Mo as a cofactor in aldehyde oxidase 4 (AT1G04580) which is involved in the generation of aromatic GSL (Ibdah et al., 2009). However, exactly why the regions (e.g. A2/C2 and A9/C9) should have SNP and GEM markers capable of predicting for the concentration of other elements in seed is unknown (i.e. B, and to a much lesser extent Mg, Mn, Sr and Zn). As such it was deemed necessary to test the

seed of an *A. thaliana* T-DNA insert mutant for *HAG1/Myb28* to determine if element concentrations other than S would be disrupted.

One reason why these regions may be coinciding with GSL for multiple element concentrations is that they could contain several different genes which have been selected/bred with the loci for low GSL in seeds (as the association regions are broad, e.g. the A9/C9 association peak contains ~870 CDS gene models). For example, in the S concentration of seed A9 association peak, there are 14 CDS gene models which have 'zinc' within their *A. thaliana* orthologous gene descriptions, one of which (AT5G61510/ Cab038296.1) is only two gene models away from a *HAG1/Myb28* CDS gene model. For Mg, a potential candidate in this region would be Cab040264.1, whose orthologous gene is an Mg transporter (*MGT9*: AT5G64560.1) which appears within the same AT association peak on A9 in the SNP and GEM outputs. However, narrowing down candidates in this region is extremely difficult for this reason and it was therefore not possible to assess all potential candidates as part of this study. Furthermore, given the large size of the association regions involved it is likely that many candidate genes related to the phenotype would be identified by chance. Therefore only a reduced set of the most likely candidates was carried forward with *HAG1/Myb28*.

The first of these was *GTR2* (AT5G62680), present in 3 association peaks for S seed concentration as the CDS gene models Cab038257.1/Cab038255.1/Bo9g015100.1. It was identified as part of an AT study of GSL in seeds as potentially being involved in loading GSL into seeds (Nour-Eldin et al., 2012; Lu et al., 2014). It would be interesting to investigate whether disruption in this gene would affect the concentration of multiple nutrients in the seed, as it would likely disrupt S concentrations through its impact on GSL transport. Another candidate within the same association regions (i.e. A2/C2), as well as a stand-alone GEM on C4 in GEM association analyses, were the CDS gene models Cab047952.2/Cab021700.1/Bo2g162860.1/Bo4g142360.1, whose orthologous gene is described as a *Per1-like family protein* (AT5G62130). There is very little information on this gene within the literature and it was therefore selected for further investigation. Another candidate in the C2 association region which could have an effect on multiple elements was Bo2g161850.1, whose orthologue is *SEUSS-LIKE2/SLK* (AT5G62090) which is involved in embryogenesis and post-embryonic development (Lee et al., 2014), and could therefore cause disruption in the seed ionome. Finally, Cab010445.6 was found in the S seed SNP association region on chromosome A8, whose orthologue corresponds to *SBP1* (i.e. *selenium binding protein 1*, AT4G14030). Its expression is known to be tightly bound to S concentration in *A. thaliana* (as they are chemically similar Se is able to replace S in proteins and other

compounds, causing toxicity). However, SBP1 proteins have been shown to accumulate in the presence of Cd and other stresses requiring GSH for tolerance (Hugouvieux et al., 2009), representing a good candidate for further testing in *A. thaliana* seeds for effects on multiple element concentrations in seeds.

3.2.7 Summary of AT results and conclusions

Analysis of seed element concentration data with AT yielded a number of potential candidates to take forward to analyse with *A. thaliana* T-DNA lines (summarised in **Table 3.2.7.a**). Many candidates identified as part of the initial AT analysis were already known within the literature to play a role within the trait under study (at least within *A. thaliana*) and were not investigated further as part of this project. These previously identified candidates represent good targets for further study with the TILLING (McCallum et al., 2000b). If lines could be found with mutations in the candidates highlighted, they could be assessed directly for a role within the trait in *B. napus* to confirm the AT results. For the candidates taken forward into *A. thaliana* T-DNA lines, many could have roles within the concentration of multiple elements (either by appearing in more than one AT output or through previous roles identified within the literature). Where possible therefore these would be analysed for an effect in elements other than the one they were originally identified for. For example, *HAG1/Myb28* and *GTR2* appeared in multiple AT outputs and markers within association peaks could predict for multiple element concentrations within seeds.

Table 3.2.7.a list of the candidate genes taken forward for further study in seeds.

Detailed are the original AT trait analysis the candidate was found for, its marker within the pan-transcriptome, AGI code, description in *A. thaliana*, line ordered from NASC and other potential element interactions are listed. Cu and Cd lines were analysed as part of an undergraduate project.

Trait AT	Marker	AGI	Description	Line ordered	Interaction ?
Cu seed*	NA	AT4G05030	Copper transport protein family	GK-197D03	?
Cu seed*	Cab018316.1	AT1G68100	ZIP metal ion transporter family	SALK_047876C SAIL_891_H08.v1	Zn, Mn, Cd, Fe
Cd seed*	Bo3g083230.1	AT3G24450	Heavy metal transport/detoxification superfamily protein	SALK_027460.35.15.x WiscDsLox481-484J7	Cu
Mo/S seed	Cab021728.1/Cab038298.3/Bo2g161590.1/Bo7g098590.1/Bo9g014610.1	AT5G61420	Myb28	SALK_136312C	Mo, S, B, Mg, Zn, Sr
Mo/S seed	Cab038257.1/Cab038255.1/Bo9g015100.1	AT5G62680	GTR2	SALK_052178C SALK_072700.40.35.x	Mo, S, B, Mg, Zn, Sr
Mo/S seed	Cab047952.2/Cab021700.1/Bo2g162860.1/Bo4g142360.1	AT5G62130	Per1-like family protein	SALK_039375C SALK_054073	Mo, S, B, Mg, Zn, Sr
S seed	Bo2g161850.1	AT5G62090	SEUSS-LIKE 2 (SLK)	SALK_039276C slk2-1/SALK_089954	?
S seed	Cab010445.6	AT4G14030	selenium-binding protein 1 (SBP1)	SALK_147323C	Mo, Se?

3.3 Seed candidate gene analysis

Many of the candidate genes identified from AT analysis were applicable to other elements within the seed ionome (see previous section for a summary of the candidates taken forward for analysis in *A. thaliana*, **Table 3.2.7.a**). This applicability was either a result of the candidate being shared across multiple elements (e.g. *HAG1/Myb28*) or because previous descriptions of the selected candidate have implied a role in another element within the literature (other than the one tested by AT). For the latter, it is important to evaluate whether the candidate has an effect on either of the elements, as it is possible that the diversity panel may not be displaying enough variability for the element the candidate was originally described for (thus it is important to rule out if it could possibly be playing a role in multiple nutrients). Furthermore, the concentration of many elements is known to correlate significantly within seeds (**Figure 3.2.7.a**), which could imply some level of interaction between many of the elements under assessment and must be considered as part of the analysis.

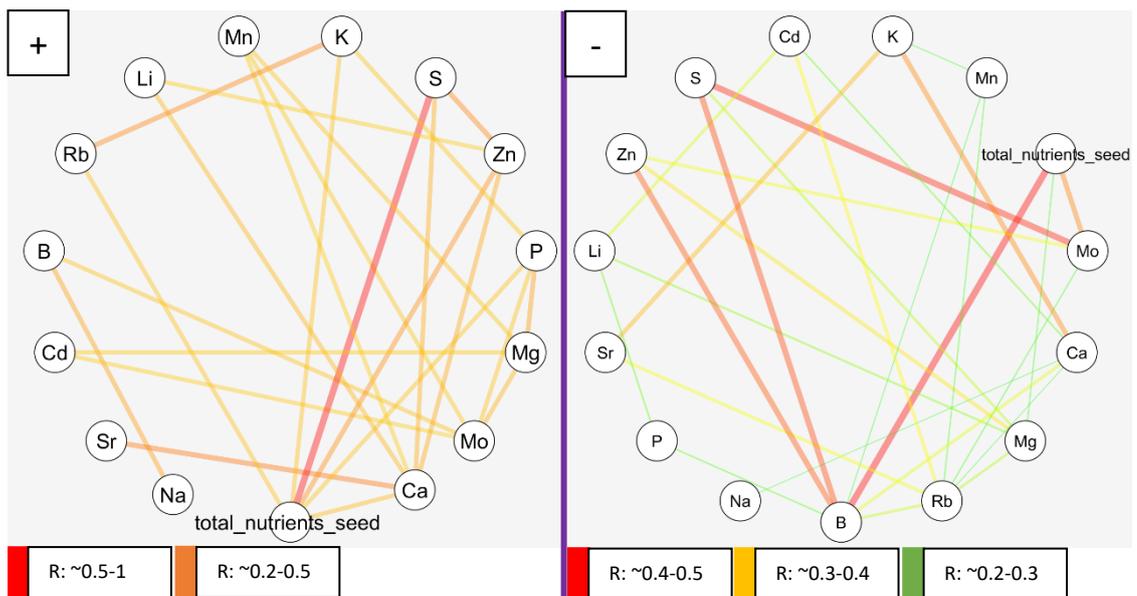


Figure 3.2.7.a The seed ‘interactome’

Significant correlations ($p < 0.001$) for all element concentrations within the seed are displayed, with the R value representing the lines connecting the elements (thicker and darker lines are more highly correlated, key for each displayed directly below). Element concentrations which show a significant positive correlation are displayed on the left whilst those which are negatively correlated are displayed on the right. Both positive and negative associations are compared to the sum total of the element concentrations (‘total nutrients seed’) to highlight the effect individual elements may have on the ionome as a whole (e.g. large variation in seed S concentration). Diagrams were generated with Cytoscape 3.2.1, using the R correlation coefficient as edges.

This section will start with the analysis of Cu and Cd seed candidates originally used as part of an undergraduate project. These candidates were chosen from AT analysis performed as part of this research and investigated further under the supervision of the author by an undergraduate student who designed primers, grew the plants and analysed seed and stem/pod materials with ICP-MS. For the purposes of this project, the samples were re-diluted and analysed with ICP-MS after initial results were found to be too variable to analyse. Additionally, Cu/Cd lines were regrown as part of the research presented here to assess the effect within leaves and confirm previous analyses (from the literature and PiiMs (Baxter et al., 2007a)). Following on from the Cu/Cd analysis, the next section will focus on those candidates from S/Mo AT analyses. These candidates were only analysed within seed tissues as there was no evidence within the literature for a role in other tissues. The only exceptions to this are *GTR2* and *HAG1/Myb28*. There is evidence within the literature that *GTR2* is involved in phloem loading, however since its primary effect would be observed in the seed only this tissue was analysed (Nour-Eldin et al., 2012). *HAG1/Myb28* on the other hand was assessed in leaves as well as seed, since the primary site of its action is disputed.

3.3.1 Seed candidate gene analysis: Cu and Cd seed candidates

Given that for the three candidate genes picked for further analysis from Cu and Cd could all potentially show variation for another element, it was decided that these candidates would be analysed together across multiple elements. For seed and stem analyses, all element concentrations under assessment had averages greater than the LOD and had a percentage recovery >80% (Mn had 100% recovery, Fe 80%, Cu 88%, Zn 90% and Cd 96%). Unfortunately, one of the Cu *iar1* mutant lines (SALK_047876C) failed to yield any homozygote or heterozygote plants for analysis (i.e. all plants were genotyped as wild type). Therefore only one mutant (SAIL_891_H08.v1) could be analysed for this line. When seed tissues were analysed no disturbance in elemental concentrations was found for any of the candidates under any of the elements specifically tested (Cu, Cd, Mn, Fe and Zn; **Figure 3.3.1.a** shows only Cu and Cd analyses). This could be a result of inaccuracies in weighing, as all lines involved produced very little seed. As a result, subsequent seed analyses would combine seeds from multiple plants to improve the accuracy of ICP-MS results.

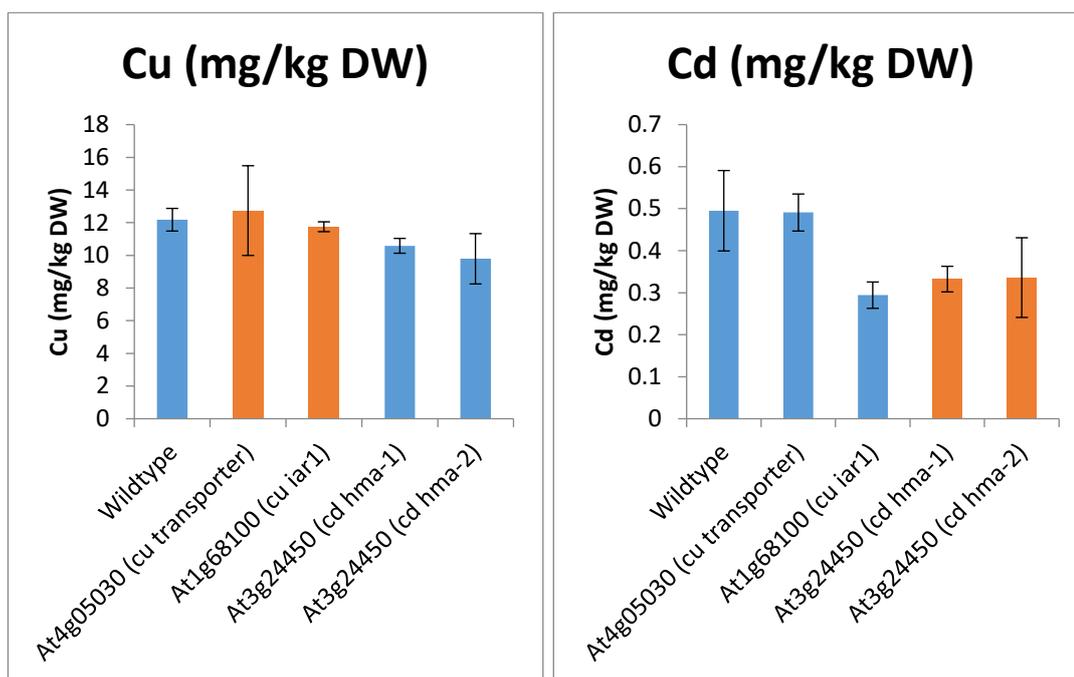


Figure 3.3.1.a Candidate gene analyses of Cu (left) and Cd (right) in the seeds of *A. thaliana* insert mutants, as mg/kg DW of seed tissue.

Wildtype; *A. thaliana* Col 0, Cu: AT4G05030 = *Cu transporter* (GK-197D03), AT1G68100 = *iar1* (SAIL_891_H08.v1); Cd: AT3G24450 = *Cd hma-1* (SALK_027460.35.15.x) and *Cd hma-2* (WiscDsLox481-484J7). The mean and standard error are shown for each line, n = 4 in all instances except Cu transporter where n = 5, n being the number of individual plants sampled for seed material. No significant differences were found between any of the candidates under any element concentration (displayed for only Cu on the left and Cd on the right). Highlighted in orange are the candidates which were picked out specifically from AT analysis of the element concentration being portrayed. Cu ANOVA: F:0.53, p:0.718, df 4,16; Cd ANOVA: F 2.23, p:0.111, df 4,16

When the stem materials were assessed, a number of distinct differences was observed. Within the stem and pod materials Cu *iar1* (AT1G68100) was found to have significantly lower shoot Cd concentration (**Figure 3.3.1.b**). This was unsurprising considering previous research has found disruption within the leaves for Cd in this line (PiiMs, (Baxter et al., 2007a)). However, other element concentrations (including Zn and Cu) have been shown to vary in this line in leaf material. Why only Cd concentrations would vary is uncertain; perhaps this is part of its suggested role in the compartmentalisation of cations (differing between the essential Cu and Zn, in comparison to the toxic and lower concentrations of Cd). *Cd hma-1* (AT3G24450) was found to have lower Cd and Zn concentrations in stems (**Figure 3.3.1.b**). Interestingly *Cd hma-2* was not significantly different from either the wildtype control or the other mutants. This is potentially to do with the variability of the results obtained. It would likely be significant

with further testing and a larger sample size. Considering that Cd *HMA* is meant to have Cu ion binding/transport activity, it is surprising that it was found to vary for both Cd and Zn and not Cu concentrations (Cu ANOVA: F 0.86, df 4, 16 p 0.511). This may be because this function has been inferred from sequence information (i.e. it is a conserved protein domain) and as such could have binding capacity for a number of elements, such as Cu, Cd, Co and Zn (NCBI, <https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=238219>). Furthermore this gene is thought to be primarily expressed within the stem in *A. thaliana* (“the bioanalytic resource for plant biology” <http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi/>), coinciding with the ionic results. However this candidate was identified in seed AT outputs. This could be a consequence of the difference between identifying a gene in *B. napus* and testing its orthologue in *A. thaliana*. For example, this gene may have a slightly different expression pattern in *B. napus* making it more important during seed loading or it could be a consequence of morphological differences between the stems of *B. napus* and *A. thaliana*.

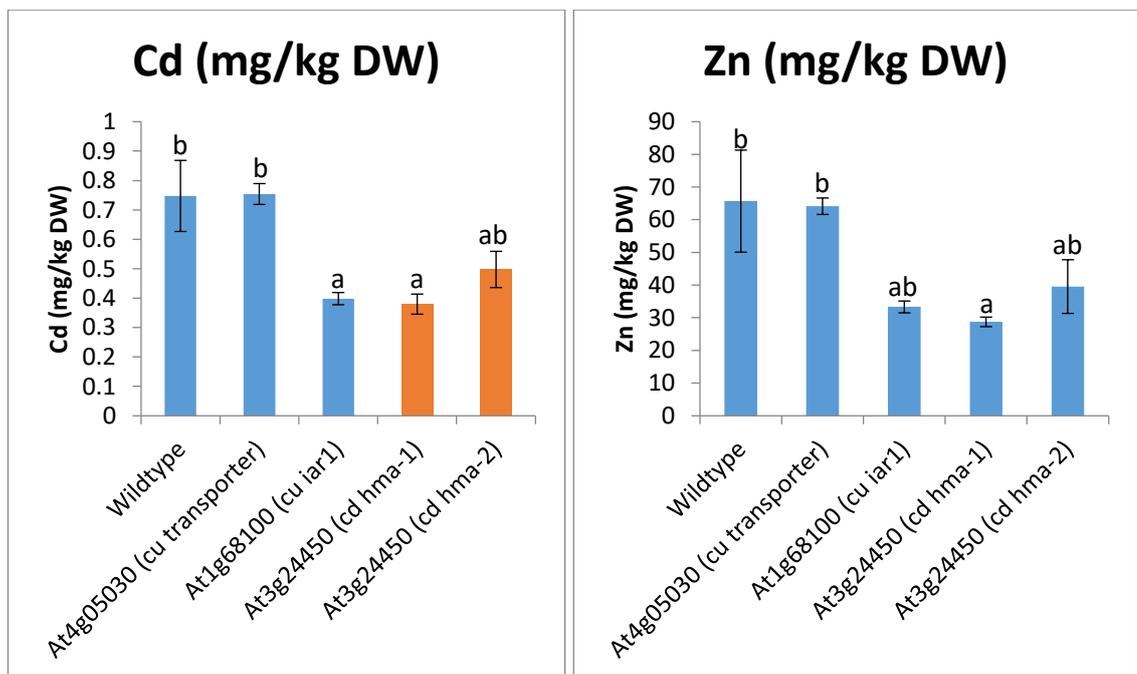


Figure 3.3.1.b Candidate gene analyses of Cu and Cd (as mg/kg DW) in seed *A. thaliana* insert mutants within stem and pod materials

Wildtype; *A. thaliana* Col 0, Cu: AT4G05030= *Cu transporter* (GK-197D03), AT1G68100 = *iar1* (SAIL_891_H08.v1); Cd: AT3G24450 = *Cd hma-1* (SALK_027460.35.15.x) and *Cd hma-2* (WiscDsLox481-484J7). The mean and standard error are shown for each line, n = 4 in all instances except Cu transporter where n = 5, n being the number of individual plants sampled for stem/pod materials. Elements are displayed as mg/kg DW of stem. Highlighted in orange are the candidates which were picked out specifically from AT analysis of the element being investigated. Only Cd (left) and Zn (right) showed any

significant difference in comparison to the wildtype control. Cd ANOVA: F 8.73, df 4,16, $p < 0.001$; Zn ANOVA: F 5.27, df 4,16, $p 0.007$.

Leaf tissues were analysed as part of a separate ICP run and the recovery of all elements in the following analysis was significantly lower (Mn had 93% recovery, Fe 65%, Cu 75%, Zn 81% and Cd 93%) but all elements had averages above the LOD. The *Cu transporter* protein family candidate (AT4G05030) showed a significant decrease in Cu concentration and a significant increase in Cd and Zn concentrations within the leaves in comparison to the wildtype control (**Figure 3.3.1.c**). It did not show any variation in the stem or seeds (**Figure 3.3.1.a** and **Figure 3.3.1.b**). Cu concentrations gave a negative result in this candidate as there was apparently less Cu within the digested material than within the water controls. This is likely a consequence of the very small weight of plants produced giving inaccurate element concentrations and a very low (~75%) recovery of this element during digestion. The results for this mutant are therefore highly questionable for Cu concentrations. Visually this mutant was much smaller in comparison to the wildtype plants and displayed purple discolouration (indicative of ROS damage), implying the candidate gene may have an essential role in broader leaf nutrient concentration regulation, consistent with the observed disruption in the ionome of the plants. Alternatively, another mutation in the line could have caused the stunted appearance. Interestingly, a significant reduction ($p < 0.05$) in Cu concentration was observed in the leaves of the *Cd hma 2* T-DNA line (**Figure 3.3.1.c**). This is in contrast to the stem where it did not show any significant differences, unlike *Cd hma 1*. This could be a consequence of the insertion in each line being in a different location, perhaps disrupting expression patterns. *iar1* did not show any variation between elements in the leaves in contrast to what has previously been observed (PiiMs (Baxter et al., 2007a)). Fe had very poor recovery (~65%) and showed no significant difference between any of the lines (Fe ANOVA; F 0.35, df 4,40, $p 0.84$).

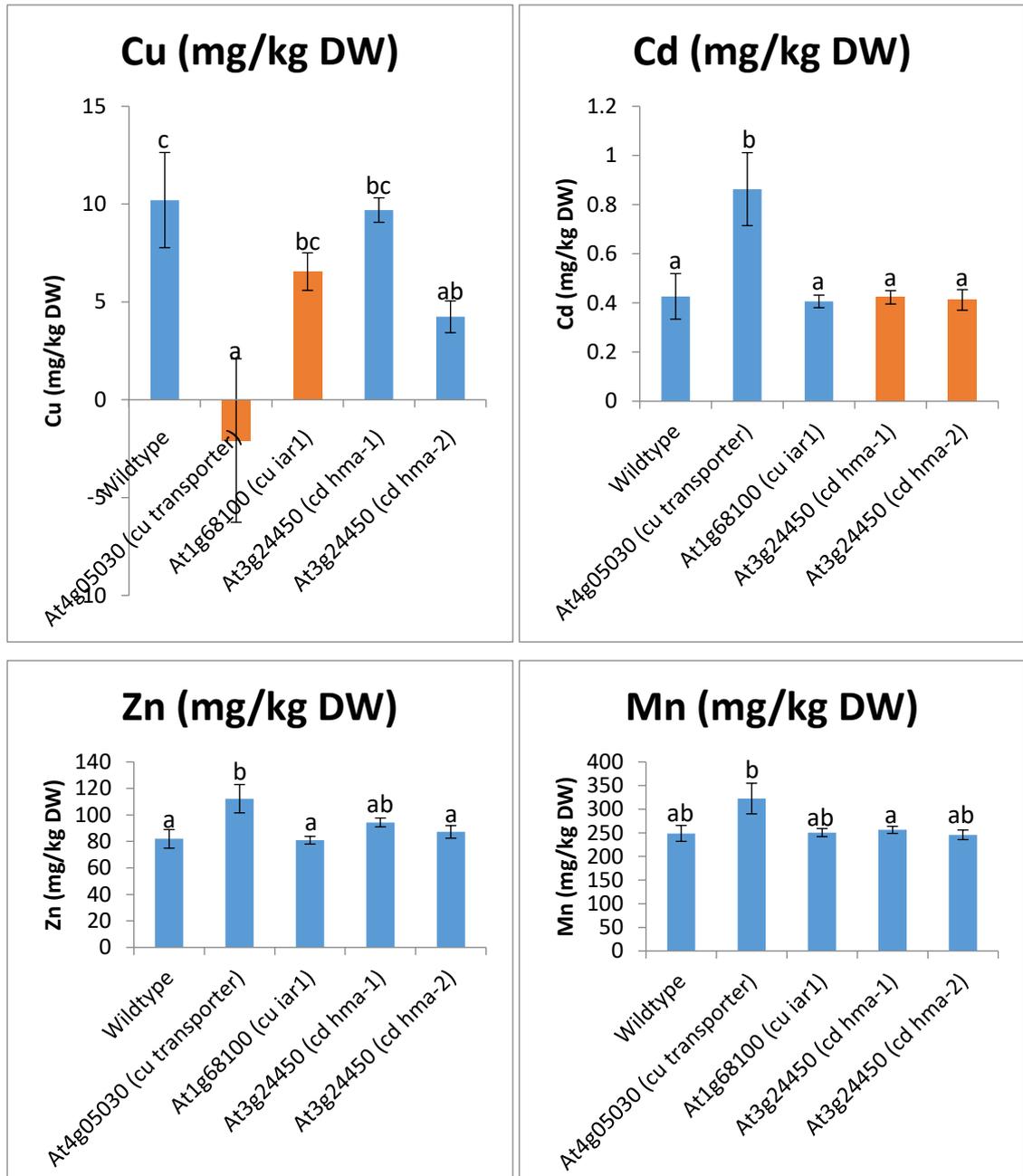


Figure 3.3.1.c Candidate gene analyses of Cu and Cd seed *A. thaliana* insert mutants within leaves, as mg/kg DW of leaf tissue

Wildtype; *A. thaliana* Col 0, Cu: AT4G05030= *Cu transporter* (GK-197D03), AT1G68100= *iar1*

(SAIL_891_H08.v1); Cd: AT3G24450 = *Cd hma-1* (SALK_027460.35.15.x) and *Cd hma-2* (WiscDsLox481-

484J7). The mean and standard error are shown for each line: Col n = 8; *Cu transporter* n= 3; *Cu iar1* n= 11, *Cd hma-1* n= 5, *Cd hma-2* n=12, n being the number of individual plants sampled for leaf material.

Elements are displayed as mg/kg DW of leaf. Highlighted in orange are the candidates which were picked out specifically from AT analysis of the element being portrayed. Only the *Cu transporter* protein (At4g05030) showed any significant difference to the wildtype control (in Cu, Cd and Zn), Cu ANOVA (top left): F 6.39, df 4,34, p<0.001; Cd ANOVA (top right): F 9.04, df 4,34, p<0.001; Zn ANOVA (bottom left): F 4.46, df 4,34, p 0.005; Mn ANOVA (bottom right): F 2.75, df 4,34, p 0.044.

3.3.2 Seed candidate gene analysis: S seed candidates

Many of the candidates from S AT analysis could affect other element concentrations within the seed ionome. This is because they also appeared as part of other element AT outputs and were predictive ($p < 0.05$) for other element concentrations in the seed ionome (e.g. B, Mg, Zn, Sr and Mo). The analysis was split into two, with three candidates (*Myb28/HAG1*, *GTR2* and *Per1L*) analysed together and the remaining two (*SLK* and *SBP1*) run separately. All candidates displayed are homozygous for the mutation unless otherwise stated. It is important to note that seeds were combined from multiple plants in order to achieve weights to give accurate digestion results. For the first round (*myb28/hag1*, *gtr2* and *per1l*) percentage recovery for the elements under investigation (B, Mg, Zn, Sr and Mo) ranged from 90%- 103%. Only Sr was excluded due to having an average concentration below the LOD. For the second round (*seuss-like2* and *sbp1*) all elements which were successfully analysed were tested. This was because *seuss-like2* could have broad implications for the seed ionome as it is involved in embryo development. Of the 20 elements which could be accurately analysed (excluding elements without CRM values) only Al was excluded from analysis for having an average concentration across mutants lower than the LOD. Of the remaining 19 elements, 3 had a low percentage recovery (Fe 69%, Co 82% and Ni 84%) leaving 16 elements to be analysed across seed tissues (recovery ranged from 94%-111%; B, Na, Mg, P, S, K, Ca, Mn, Cu, Zn, As, Se, Sr, Mo, Cd and Ba).

Within S analyses (**Figure 3.3.2.a** top left), all candidates showed a reduction in S concentration in comparison to the wildtype control (except *per1L-1*). All *gtr2* lines showed similar reductions in S concentration, regardless of whether they were homozygous (HM) or heterozygous (HZ). The effect, although significant, was not a large drop likely because only *GTR2* was disrupted. Perhaps the effect would have been greater if *GTR1* was disrupted as well, as demonstrated for GSL (*gtr 2* mutant displayed lower GSL within seed, but significantly lower when both *gtr 1* and *2* were disrupted (Nour-Eldin et al., 2012)). *gtr 2-1* HZ and *gtr 2-2* HM also displayed a reduction in Mo concentration. An explanation for this could be a potential role for Mo in aromatic GSL biosynthesis. If there was a reduction in GSL concentration within the seeds it would result in reduced demand for Mo containing aldehyde oxidase 4, which is required for converting benzaldehyde to benzoic acid in aromatic GSL production (Ibdah et al., 2009)). *Per1l-2* also displayed disrupted S concentration unlike *per1l-1*. It has been surmised this could be due to alternative splicing, as many alternatively spliced forms of this candidate are possible and the two inserts within each mutant are in different locations, i.e. *per1l-1* may have been spliced out.

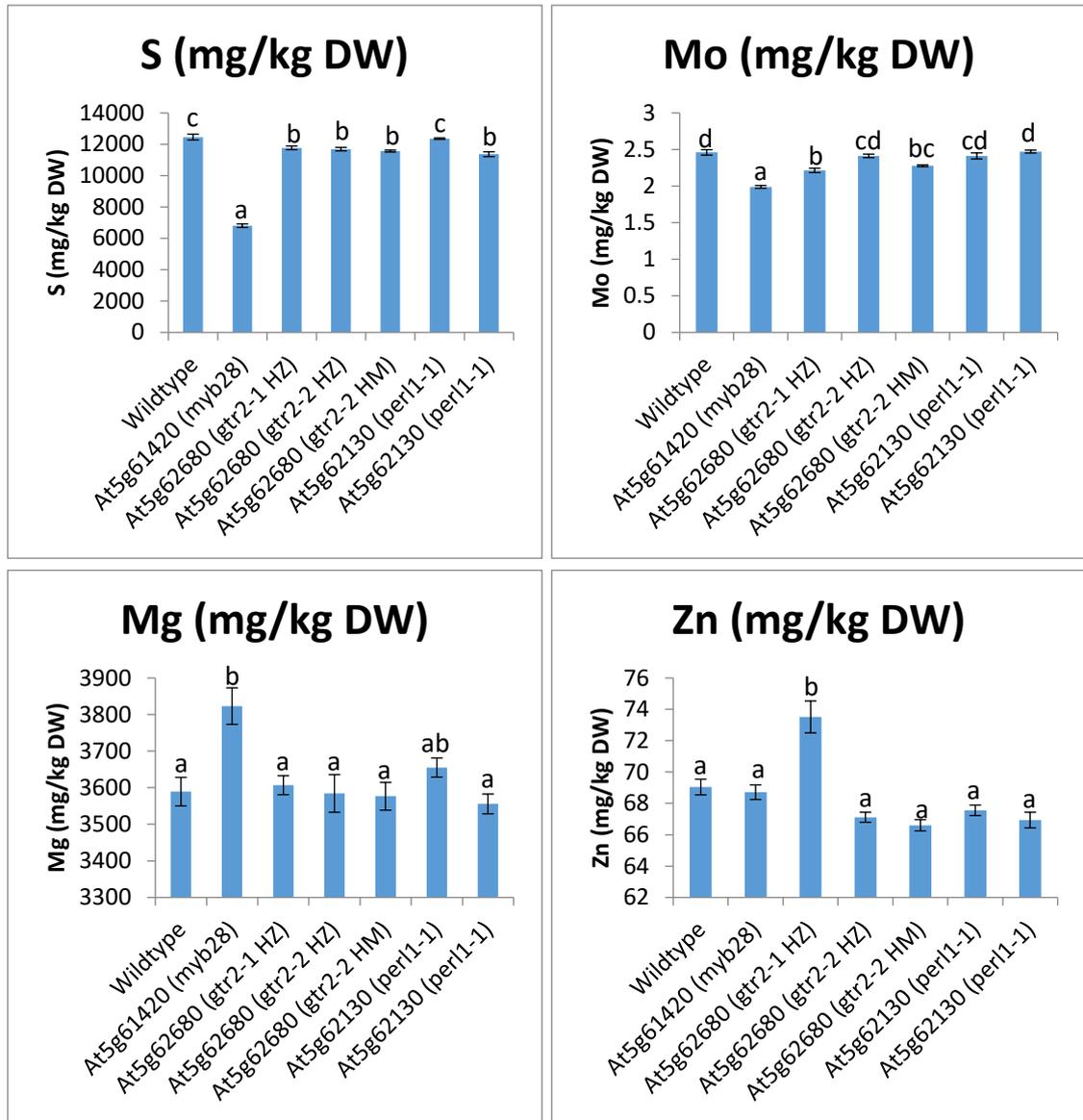


Figure 3.3.2.a Candidate gene analyses of S seed *A. thaliana* insert mutants (*HAG1/Myb28*, *GTR2* and *Per1L*), as mg/kg DW of seed tissue

Wildtype; *A. thaliana* Col 0; AT5G6142 = *myb28/hag1* (SALK_136312C); AT5G62680 = *gtr2-1* (SALK_052178C) and *gtr2-2* (SALK_072700.40.35.x); AT5G62130 = *per1-1* (SALK_052178C) and *per1-2* (SALK_054073). The mean and standard error of 6 batches of seed per line are shown (6 sub-samples of seed from a pool of seeds from 12 plants) for all lines except *gtr2A* HZ where 12 batches were included (12 sub-samples of seed from a pool of seeds from 24 plants). Elements are displayed as mg/kg DW of seed. Where there are the letters HZ or HM within a line it indicates that heterozygote and homozygote plants respectively have been analysed. Significant differences between lines are indicated when letters are not shared from post-hoc analysis Bonferroni $p < 0.05$. S ANOVA: $F 209.63$, $df 6, 41$, $p < 0.001$; Mo ANOVA: $F 31.41$, $df 6, 41$, $p < 0.001$; Mg ANOVA: $F 5.58$, $df 6, 41$, $p < 0.001$; Zn ANOVA: $F 13.72$, $df 6, 41$, $p < 0.001$.

The most significant effect was observed within the *myb28/hag1* mutant line (**Figure 3.3.2.a** and **Figure 3.3.2.b**). This line also showed significantly lower Mo and significantly higher Mg concentrations in comparison to the wildtype control and other mutants. Furthermore, when just the Columbia wildtype control and *myb28/hag1* mutant were compared many other element concentrations within the seed ionome appear to be disrupted (**Figure 3.3.2.b**; Mg, P, S, K, Co, Cu, As, Mo, Cd, Pb, $p < 0.05$). As an extension of this analysis, it was decided to assess leaf tissues (since *Myb28/HAG1* is meant to regulate aliphatic GSL throughout the plants). Once again a multitude of element concentrations appears to be disrupted within this line (**Figure 3.3.2.c**; B, S, As, Se, Mo, Cd and Ba, $p < 0.05$). Why *myb28/hag1* would affect multiple element concentrations within the ionome across both leaf and seed tissues is not known. One theory suggests that *Myb28/HAG1* may play a role in the S deficiency response. Previous research has suggested that 'sulfur deficiency induced 1' (*SDI1*) may inhibit the transcription of aliphatic GSL genes by interacting with *Myb28/HAG1* and down regulating GSL biosynthesis in favour of sulfate use in primary metabolism (Aarabi et al., 2016). This is in line with previous research which has shown a general reduction in GSL production under S deficiency conditions (Falk, Tokuhisa and Gershenzon, 2007). As such *A. thaliana* T-DNA lines and *B. napus* low GSL lines (effective *Myb28/HAG1* knock-outs) could be perceiving the reduced GSL concentrations as S deficiency, perturbing the rest of the ionome as they respond to rectify the imbalance (e.g. stimulation of root formation or increased transporter capacity). Whether this has a positive effect on the nutrient use efficiency of the plants requires further study. Perhaps simulating nutrient deficiency under sufficiency conditions could improve plant nutrient use efficiency, or it could cause ionic disturbance perturbing plant metabolism with wide and unexpected consequences. Indeed, whether this would even constitute nutrient use efficiency is dependent on the definition of 'efficiency'. A plant with simulated deficiency could potentially use more of the nutrients supplied, but whether this would be turned into an increased yield is debatable (and could be considered an oversimplification of the argument), and would require much further research (Yu et al., 2016a). Furthermore, there is some evidence that root morphology is disrupted in *myb28myb29* knock-out lines (Martinez-Ballesta et al., 2015). Alternatively, given that GSL are plant defence compounds, perhaps the plants were more susceptible to pest damage (Beekweelder et al., 2008) which in turn disrupted the ionome. The most simplified explanation could be increased root/leaf damage within the low GSL lines. This may be one of the reasons why previous research on this line failed to see any differences between the macronutrients (under low or high S) as they were grown in hydroponics (Martinez-Ballesta et al., 2015) in contrast to the current experiment which was performed in soil.

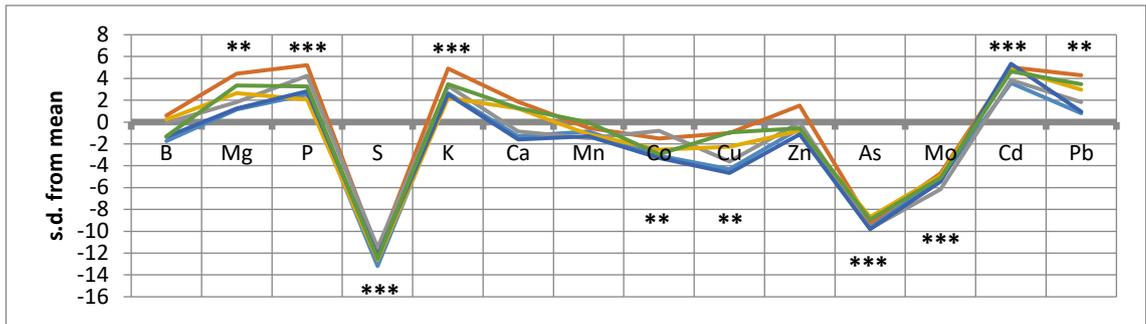


Figure 3.3.2.b Z-score graph detailing how each seed batch (as a different coloured line) from the *myb28/hag1* line varied in comparison to the average wildtype control (n= 6 for both, where n is a subsample of seeds from a pooled sample of seeds from 12 plants).

Only elements which had a recovery >85% and an average concentration greater than the LOD were analysed. The mean and standard deviation of the wildtype controls are used to calculate the number of standard deviations each mutant deviates from the average of the wildtype control in each element concentration, in accordance with methods previously outlined (Lahner et al., 2003) and in use by PiiMs (Baxter et al., 2007a). A t-test was used to determine whether the difference between wildtype and mutant lines was significant for each element ($p < 0.001$ ***, $p < 0.01$ **), as detailed in the appendices (0).

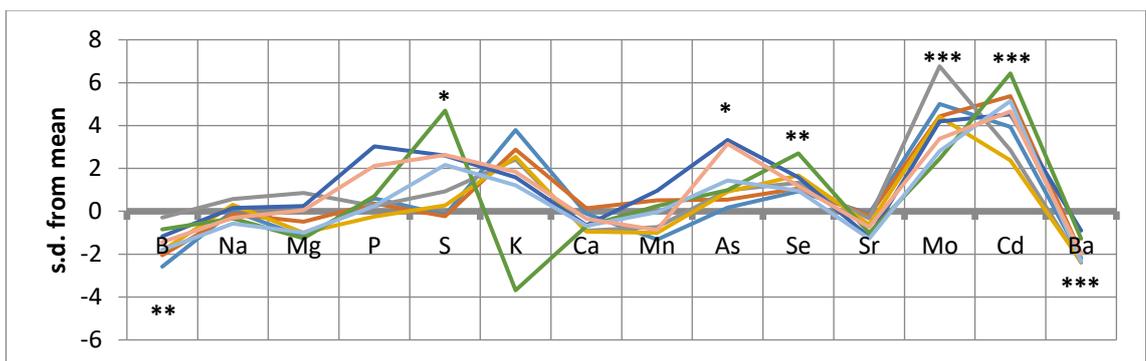


Figure 3.3.2.c Z-score graph detailing how leaf material (as a different coloured line) from the *myb28/hag1* line varied in comparison to the average wild type control (n= 8 for both, where n is an individual plant).

Only elements which had a recovery >85% and an average concentration greater than the LOD were analysed. The mean and standard deviation of the wildtype controls are used to calculate the number of standard deviations each mutant deviates from the average of the wildtype control in each element concentration, in accordance with methods previously outlined (Lahner et al., 2003) and in use by PiiMs (Baxter et al., 2007a). A t-test was used to determine whether the difference between wildtype and mutant lines was significant for each element ($p < 0.001$ ***, $p < 0.01$ **, $p < 0.05$ *), as detailed in the appendices (0).

Of the other two candidates that were tested, one line of '*seuss-like2*' failed to produce any homozygous plants (SALK_039276C) and was not carried forward (i.e. it was wildtype). The

other *seuss-like2* line showed altered B, Mo and Mg concentrations within the seed, whilst *sbp1* only varied for Mo concentration (**Figure 3.3.2.d**). Neither of these mutants displayed variation in S concentration. This was unexpected for *seuss-like2* as it was the most highly connected marker from WGCNA analysis (as detailed in the next chapter, 4.3.1 WGCNA); one of the highest GEM hits from seed S AT analysis and a predictive GEM marker (see 3.2.6). This likely reflects the genetic background of the low GSL phenotype in *B. napus*: a deletion/homeologous exchange could cause an association with multiple GEM markers as all their gene expression levels would be equally disrupted in a deletion type event. That *seuss-like2* causes disruption in element concentrations other than S within *A. thaliana* provides further evidence for the theory that multiple elements may show disruption in the low GSL *B. napus* lines as an indirect result of how the low GSL lines were bred (i.e. the broad association peaks may contain multiple candidates capable of causing multi-element phenotypes, with breeding for the low GSL phenotype indirectly perturbing the ionome, see 1.2.6). Why *slk* would show disruption in B, Mo and Mg concentrations is yet to be established. Interestingly, the *sbp1* knock-out T-DNA line showed lowered Mo concentration in comparison to the wildtype control. The expression of *SBP1* is thought to be linked to the S status of cells with its promoter being strongly induced under S starvation conditions (Hugouvieux et al., 2009). The biological function of *SBP1* is as yet unknown. However the suggested role in S and Se concentration could link it to Mo via the chemical similarities of these elements (Marschner, 1995c).

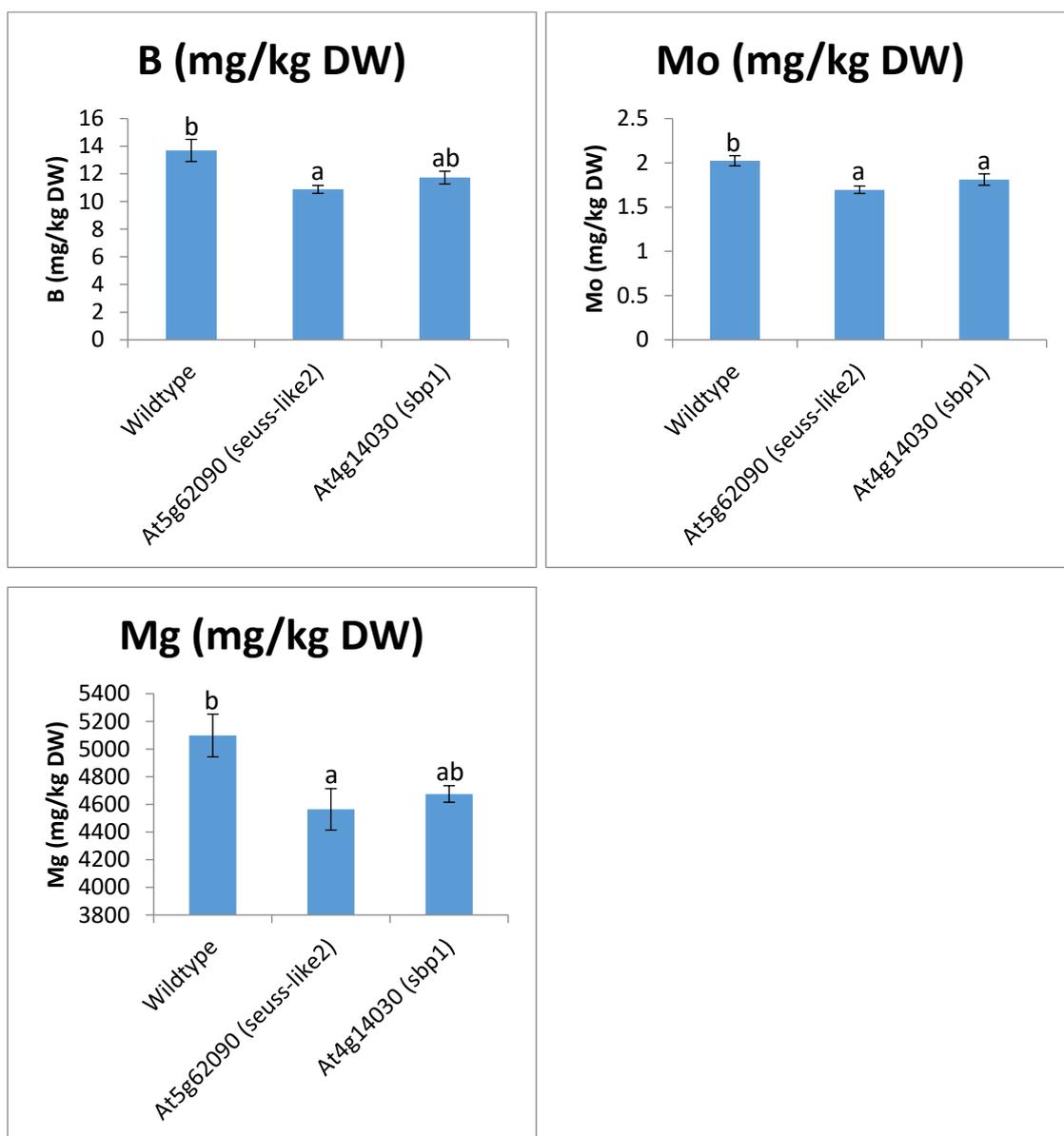


Figure 3.3.2.d candidate gene analyses of S seed *A. thaliana* insert mutants *seuss-like2* and *sbp1*, as mg/kg DW of seed tissue

Wildtype; *A. thaliana* Col 0; AT5G62090= *seuss-like2/slk* (SALK_089954); AT4G14030= *sbp1*

(SALK_147323C). The mean and standard error of 6 batches of seed per line are shown for all lines (i.e. 6 sub-samples of seed from a pool of seeds from 12 plants); elements are displayed as mg/kg DW of seed (for B, Mo and Mg). Significant differences between lines are indicated when letters are not shared from post-hoc analysis Bonferroni $p < 0.05$. B ANOVA (top left): F 6.65, df 2, 15, p 0.009; Mo ANOVA (top right): F 9.03, df 2, 15, p 0.003; Mg ANOVA (bottom left): F 4.78, df 2, 15, p 0.025.

3.3.3 Summary of candidate gene analysis

A number of candidates was successfully tested with *A. thaliana* T-DNA lines. Cu/Cd candidates identified from seed AT analysis failed to reveal any significant differences within the seed tissues. However this was likely a result of the small weights used initially in this study. When analysis was expanded to include stem and leaf materials, some significant differences were observed. The *Cu transporter family protein* plants were very stunted and slow growing, producing very little biomass for accurate analysis of leaf tissues. Nevertheless, this line seemed to display reduced Cu concentrations combined with increased Cd and Zn concentrations in leaf tissues. *IAR1* did not display any elemental variability within the seeds or leaves of the *A. thaliana* T-DNA lines investigated, in contrast to previous research which found variation in Cu and Zn concentrations within the leaves but did show reduced stem/pod Cd concentration. Results from the HMA protein identified from Cd seed AT analyses were contradictory; *Cd HMA1* seemed to show a reduction of Zn and Cd concentration in the stem/pods, whilst *Cd HMA2* showed a reduction in Cu concentrations within leaves. It was hypothesised that this may be a consequence of differences between the two insertion lines and requires further testing in *A. thaliana*.

S analysis also revealed some unexpected results. *Myb28/HAG1* analysis caused disruption to multiple element concentrations within the seed and leaf tissues. It was expected to cause disruption in seed S concentration via its connections to GSL biosynthesis but not to have such a wide impact on the rest of the ionome. Alongside *Myb28/HAG1*, the other GSL related gene tested was *GTR2*. *gtr2* T-DNA lines revealed a slight reduction in S and Mo concentration. *Per1L* displayed some disruption to S concentration but given the little information present within the literature on this gene an adequate explanation could not be found. *slk* T-DNA lines showed disrupted Mo, B and Mg concentrations in seed, whilst *sbp1* had a slight reduction in Mo concentration in comparison to the wildtype control. Given that these candidates were all found within the S seed AT analyses and that many show multi-element disruptions emphasises the difficulty of analysing such broad association regions. Many elements may be showing similar associations as a consequence of how *B. napus* was bred for low GSLs rather than any common biological mechanism. However, the link between the wider plant ionome and GSL was previously unsuspected and was therefore carried forward into further analysis (as detailed in the next section, 4).

3.4 Chapter summary conclusions

This chapter has focused on investigating the genes behind nutrient concentration in the seeds of *B. napus*. It has been explored for a number of elements (including Mo, S, Mn, Cu, Zn and Cd) using an AT approach. Utilising AT, a number of individual candidate genes was highlighted; some of which were already known to play a role in nutrient concentration, while others represented potentially novel candidates for further testing. Analyses performed on Cu/Cd candidates from seed did not show any disruption in the concentration of the elements under investigation. One explanation for this could be that the weights of seed analysed prevented accurate measurement of element concentrations. Alternately, this could be a consequence of analysing orthologues in *A. thaliana* instead of analysing the candidates directly in *B. napus* (either as a consequence of functional redundancy or divergence). Both of these conclusions are supported by the observation that some of the candidates under investigation were found to display disruption in element concentration in different tissues. For example, the *Cu transporter family protein* plants were stunted with purple discoloration and displayed increased Cd and Zn concentration within leaf tissues. This candidate may have a different expression pattern or a divergent function in *A. thaliana* in comparison to *B. napus*, whilst it could also be that the larger weights of samples used in the other tissues improved the accuracy in which ICP-MS results could be interpreted. Furthermore, this also highlights the importance of scoring for other important phenotypes, such as plant size, which may provide further explanation of the candidate gene's biological role. Similarly, the current research on *iar1* failed to validate previous research which had shown variation within the leaves for Cu and Zn concentrations (PiiMs, Baxter et al., (2007)). This could be a consequence of differences in the developmental time point analysed between the two experiments and the growth conditions (i.e. perhaps the plants were sampled at different growth stages and/or the growth medium differed, such as growth on nutrient rich soil or inert substrate and liquid fertiliser).

S seed AT analysis gave a number of candidates. However considering the large size of the association peaks, many candidates may have been present purely by chance. Therefore only a subset of the most likely candidates were analysed. Given the similarities observed between the S AT outputs to those of previous GSL analyses (Lu et al., 2014), it was important to test both *Myb28/HAG1* and *GTR2* for their effects on the seed ionome (i.e. the candidates thought to be responsible for the GSL association peaks). The *myb28/hag1* line displayed disruption in multiple element concentrations, whilst *gtr2* displayed only a slight reduction in S and Mo concentration. Variation in S concentration in these lines had been expected as a consequence

of the close relationship between seed S and the GSL. There were numerous other candidates in the broad association peaks in S concentration from seed AT analysis and many of the other AT outputs seemed to mirror the association peaks observed in S. It is thought this might be a consequence of breeding for the low GSL phenotype in *B. napus*, with breeding resulting in the selection of other traits (including differences in multiple element concentrations). The low GSL phenotype is thought to be a consequence of a homoeologous exchange (replacing a functional copy of *Myb28/HAG1* with a non-functional copy) it is possible that multiple genes have been indirectly selected or show variation in expression as a consequence of this. Nevertheless, since the *Myb28/HAG1* mutant appeared to have a multi-element phenotype, the link between GSL and the wider seed ionome was therefore investigated further.

4 Investigating the relationship between the seed ionome and glucosinolates in *B. napus*

4.1 Introduction

From the seed candidate gene analysis and AT results, it became apparent that there may be a link between the wider seed ionome and the glucosinolate (GSL) content of seeds in *B. napus*. GSL are the secondary S metabolites (see 1.2.5.5) responsible for the distinctive bitter taste of *Brassicaceae* vegetables (Engel et al., 2002; Halkier and Gershenzon, 2006). They are thought to act as plant defence compounds. Upon tissue damage GSL' are hydrolysed by myrosinases into a number of different toxic compounds, e.g. isothiocyanates, thiocyanates and nitriles (Halkier and Gershenzon, 2006). The biosynthesis of GSL is thought to occur namely in rosette leaves and silique walls of the plants (Jørgensen, Nour-Eldin and Halkier, 2015). The GSL are then transported throughout the plant via phloem specific transporters (GTR1 and GTR2)(Nour-Eldin et al., 2012; Andersen et al., 2013; Andersen and Halkier, 2014). The biosynthesis of GSL can be broken down into three steps: 1) amino acid chain elongation by insertion of methylene groups into side chains; 2) the formation of the core GSL structure (**Figure 3.3.3.a**); 3) secondary transformation/ modification of initial GSL. Almost all of the genes involved in core GSL biosynthesis have been identified in *A. thaliana* (Halkier and Gershenzon, 2006). Understanding GSL biosynthesis, tissue specificity and regulation is one of the main aims of *Brassica* breeders for many reasons. The primary reason in Oilseed rape is to limit the concentration of GSL in the seed (as the GSL breakdown products are harmful to animals, Griffiths, Birch and Hillman, (1998) and see 1.2.5.5) whilst maintaining high GSL concentrations in the leaves for effective plant defence (Lu et al., 2014; Nour-Eldin et al., 2017). In other members of the *Brassicaceae*, whose primary purpose is for human consumption, the main aim has been to increase specific GSL concentrations (e.g. sulforaphane in broccoli, *B. oleracea cv. Italic* (Zhang et al., 1992)) in the edible portions of the crop for anti-carcinogenic purposes (Halkier and Gershenzon, 2006).

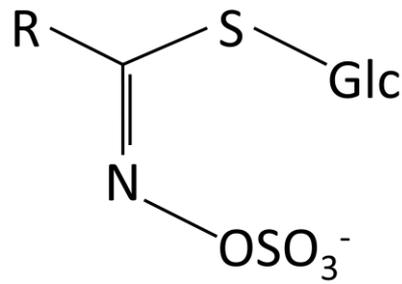


Figure 3.3.3.a basic chemical structure of a glucosinolate (GSL) contains two S atoms.

The 'R' group is a variable side chain that differs between the various types of glucosinolate (aliphatic, indolic and aromatic, depending on the amino acid the GSL is derived from). All GSL have a thioglucose moiety and a sulfonated oxime.

To investigate the potential link between S, GSL and the seed ionome, a number of approaches was taken. Firstly, WGCNA analysis was performed for S and Mo concentrations in seed (the two elements under analysis whose AT plots gave the most closely associated results) to assess whether there were similarities between the networks of genes being uncovered corresponding to the different elements. Next a subset of high and low S concentration seed lines was compared for GSL and sulfate content to confirm previous observations within the literature. Then the senescing leaves of high and low S seed *B. napus* from a field trial under high and low nitrogen fertilisation (N) were assessed for differences in the leaf ionome because it was suspected that N status and nutrient remobilisation may differ between the lines (i.e. perhaps the seeds were signalling a S deficiency resulting in a change in how the plants were remobilising nutrients). Finally the pod, green seed and stem ionome of high and low S seed lines was investigated. Since the primary site of GSL biosynthesis is disputably the pod wall it would be interesting to assess how the ionome differed. This section will start by looking at the specific methods required in these investigations, before moving on to discuss the outcome of each experiment in turn.

4.2 Case study specific methods:

4.2.1 Weighted Gene Co-expression Network Analysis (WGCNA)

The WGCNA (Langfelder and Horvath, 2008) was performed in accordance to the methods outlined in Harper et al., (2012) for Mo and S concentrations within seed as these elements showed the greatest similarities to seed GSL AT outputs. For the method in brief, the RPKM expression data from all 53,889 CDS gene models was used to cluster the 383 *B. napus* lines. The scale-free topology of the network was then approximated with a soft thresholding power of $\beta=12$. The topological overlap (a basic measurement of the number of shared connections between any two genes) was calculated for all CDS gene models based on the scaled correlations from soft thresholding. These were then clustered based upon their topological overlap to produce a dendrogram. The dynamic tree-cutting algorithm was used to cut the dendrogram at 0.25, giving a network consisting of 330 modules. From these modules a network could be constructed by employing the function `blockwiseModules` which would use the 'module eigengene' (the first principle component of gene expression, i.e. the most representative expression for the group of genes/module) in a correlation against the trait data. This allowed networks of genes with close trait associations to be discovered.

4.2.2 Seed glucosinolate analysis

For total glucosinolate extraction four replicates (seeds from 4 plants) of 18 accessions representing the 9 highest and 9 lowest total S seed lines were analysed. For analysis, 50 mg of seeds was homogenised in 70% methanol before being left to incubate at 70°C for 45 minutes, this was to prevent myrosinase degradation of the GSL. The samples were then centrifuged and the supernatant added to a pre-prepared sephadex column. Once the samples had been washed with water and 0.02M sodium acetate buffer, 75 μ l of sulfatase was added and the samples left to incubate at room temperature for 24 hours. The next day the glucosinolates were eluted with water and run on the HPLC (ThermoScientific, UltiMate 3000 Standard LC Systems) quantified using 229nm UV relative to a sinigrin control.

4.2.3 Seed Sulfate analysis

Seed sulfate analysis was initially performed alongside GSL analysis, utilising the same 18 accessions representing the 9 highest and 9 lowest total S seed lines. 50mg of seed was homogenised in cold water and incubated for 1 hour at 4°C, and then transferred to 95°C for 15 minutes. Samples were centrifuged and the supernatant taken and diluted 1:2 before being analysed with Elemental Analyser-Isotope ratio mass spectrometry (EA-IRM, ThermoScientific, Dionex ICS-1100). However, when preliminary results were analysed it was thought that this

method may not be adequately controlling for myrosinase degradation of GSL (Bones and Rossiter, 1996) (see 0). As such, the experiment was repeated by colleagues at the University of Cologne for 6 accessions (i.e. the 3 highest and the 3 lowest total seed S lines) but utilising hot water and incubating at 95°C for 1 hour instead of 4°C. All other aspects of the protocol were kept the same.

4.2.4 Leaf senescence analysis

Alongside use of plant materials grown under glasshouse conditions, materials grown under field conditions were used to assess differences in the ionome of senescing leaves. Leaves were collected from a field trial in Bessingby (Bridlington, UK, 54°04'36.1"N 0°13'58.5"W) on 19th June 2015. 18 accessions representing 9 high and 9 low S seed lines were sampled across two treatment types: normal N application (300 kg N/ha) and low N application (60 kg N/ha). A clearly senescing leaf was taken from four plants of each accession/trial plot. Full details of the field trial are yet to be published, Fraser *et al.*, unpublished.

Leaf senescence analysis took place over two ICP-MS runs, each digested in two different ways. The first run comprised 40 leaf samples from the HN conditions (representing 10 accessions with a leaf from 4 plants of the same accession) and was digested broadly in accordance to the protocols outlined for leaf analysis in previous research using Microwave digestion (Thomas et al., 2016). The protocol deviated slightly from that published in three ways; firstly individual leaves were analysed (because samples were below 200mg in weight), these were oven dried (rather than freeze dried) and a different CRM was used (Cabbage 1-NCS ZC73012, LGC standards, UK). The second round of analysis was performed under the conditions previously detailed for acid digestion and ICP-MS analysis in the General Methods on hotplates (2.5), for the remaining 104 samples from HN and LN conditions (18 accession on LN, 8 accessions on HN, all with a leaf from 4 separate plants of the same accession). Analysis was split as part of the development of the new digestion protocol testing at The University of York, detailed in General Methods (2).

4.2.5 Pod ionome investigation:

4.2.5.1 Growth of *B. napus*

For both the leaf ionome timeline (6.3.7) and the pod ionome experiments the same *B. napus* plants were used with the experiment designed so that the remaining plants from the leaf ionome timeline could be used within the pod ionome experiment. As such, 12 accessions with four individual replicate plants were grown (**Figure 4.2.5.a**); this encompassed 6 winter Oilseed rape (OSR), 6 spring OSR, 6 high S seed and 6 low S seed lines (i.e. 3 high S winter OSR (HS/W) ,

3 low S winter OSR (LS/W), 3 high S spring OSR (HS/S) and 3 low S spring OSR (LS/S)). 48 plants were grown per P60 tray with F2 soil. There were 5 trays of plants (240 plants in total) for the two experiments and plants were arranged randomly within each tray. The trays were kept within a growth cabinet (SANYO MLR-350) under 16 hour days, with a day time temperature of 20°C and a night time minimum of 14 °C. All 5 trays were grown under these conditions until they reached the 4th true leaf. Once individual plants of trays 4 and 5 had reached the 4th true leaf, they were moved to vernalisation conditions: 6 weeks of growth at a constant temperature of 4°C with 8 hour days (SANYO Fitotron SGC065). After this treatment the plants were re-potted into 4 inch pots and grown under glasshouse conditions mimicking those of the original growth chamber (16 hour days, temperature range of 14-20°C).

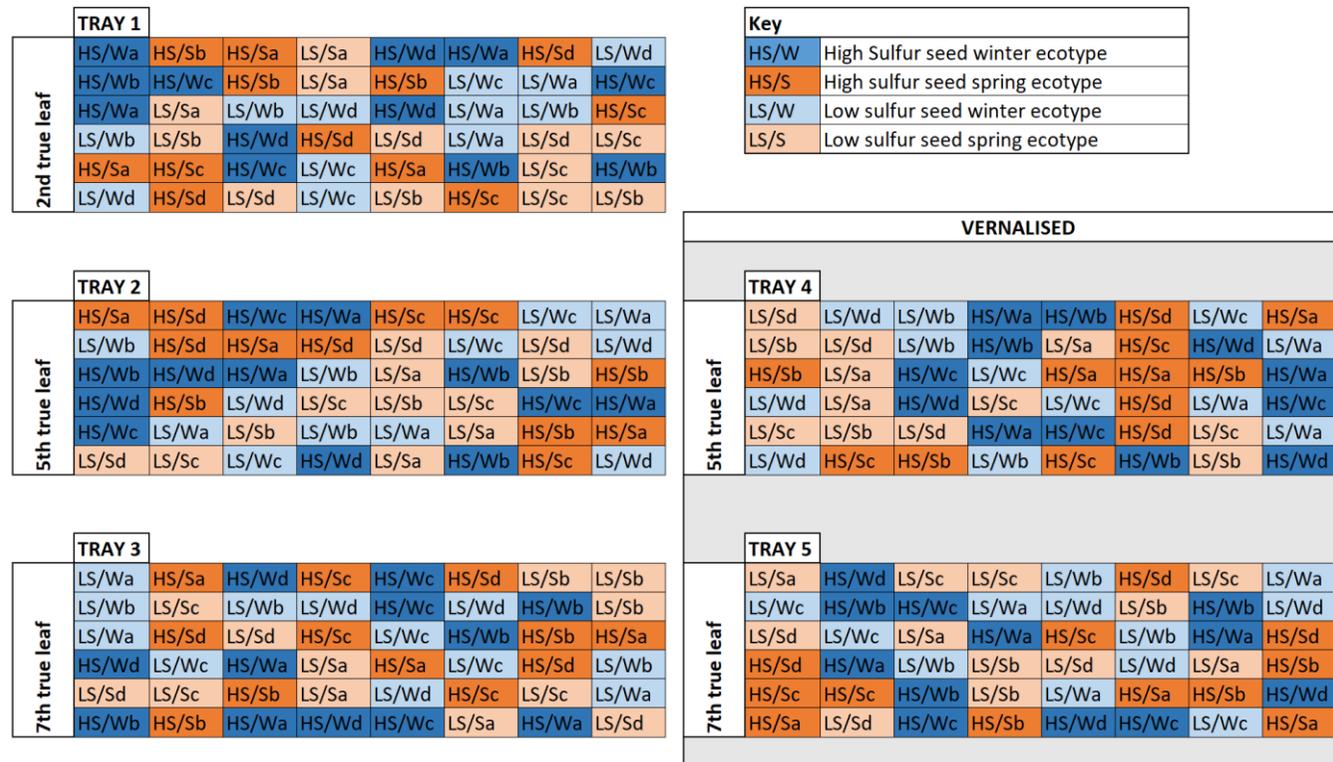


Figure 4.2.5.a Experimental design for leaf ionome timeline and pod ionome experiments.

Five trays containing 48 individual plants were grown with a randomised layout. These 48 plants represented 12 accessions: 3 high sulfur seed winter oilseed rape (OSR) accessions (HS/W, dark blue); 3 high sulfur seed spring OSR (HS/S, dark orange); 3 low sulfur seed OSR (LS/W, light blue) and 3 low sulfur seed spring OSR (LS/S, light orange). 4 plants were grown per accession (as indicated with a, b, c and d, although there is no differentiation in the figure between accessions within a group, i.e. there are 3 'a's for each HS/W, HS/S, LS/W and LS/S). The leaf each plant was sampled for in the leaf ionome developmental timeline (see 6.2.5) is indicated next to the tray label, e.g. plants in tray 1 were sampled for their 2nd true leaf. Trays 4 and 5 were subject to vernalisation (growth at 4°C for 6 weeks, with 8 hour day length) after reaching the 4th true leaf. Only the plants of trays 4 and 5 were re-potted to allow for further growth and flowering as part of the pod ionome experiment (see 4.2.5.2).

4.2.5.2 *Sampling for variation in pods*

It was observed that seed S and Mo concentrations (amongst other elements) appeared to be varying with seed glucosinolate (GSL) content. To investigate whether the link between S/Mo and GSL was a result of how seeds were loaded, an experiment was designed to compare the pod ionome between high and low S seed lines. Plants from trays 4 and 5 were allowed to bolt, flower and develop pods, leaving 8 plants per accession to be analysed (96 plants in total). After roughly 35 days of development after flowering, pods were sampled while green. Pods were counted from the bottom of the stem upwards, the first ten pods being discarded, with the next five sampled for analysis. Thus the stem was cut just above the 10th pod and just below the 16th pod. Pod, stem and any spare materials collected were then photographed as a visual reference. Each of the five pods was then cut open and the seeds within were separated from the pod walls. Stem, pod and seed materials were then stored separately at -80 °C until they could be freeze dried and analysed with ICP-MS. Some plants were incompletely fertilised (i.e. the pods contained no seeds); in these instances it was usually impossible to sample a set range of pods and therefore pods which visibly contained seeds were selected off the stem. Of the 96 plants sampled, the sampling method was altered in this way for 9 plants. These were highlighted during data analysis so that if any appeared as outliers they could be removed (outliers were classified as +/- 5 standard deviations from the average of each element concentration), no outliers were removed in this way. The only outliers removed were some of the values for stem LS/S ecotypes in B which were ran in a separate ICP-MS run and were significantly different to the other LS/S ecotypes analysed ($p < 0.01$) and one B concentration in the seeds which was negative relative to the water controls.

4.3 Results

4.3.1 WGCNA results

The most highly correlated module for seed S concentrations ($r^2=0.51$, $p=1.06 \times 10^{-69}$) was also the third highest module for seed Mo (negatively correlated, $r^2=0.14$, $p=7.10 \times 10^{-14}$). **Figure 4.3.1.a** shows the most highly connected genes from the module (a 'degree' sorted circle) and each gene is described in **Table 4.3.1.a**. Looking at the module in this way is useful for a number of reasons. Firstly, this module contains an orthologue of *HAG1/myb28* (AT5G61420), further confirming the assertion made from AT analysis that this could be the gene primarily responsible for the correlation observed between GSL, S, Mo and a number of other element concentrations. Secondly, it can be observed from the markers in **Table 4.3.1.a** that many of the genes selected for this module are all relatively close together. This is a pattern commonly observed following a deletion. Since all the genes within such regions are removed, they all have similarly disrupted expression patterns and are therefore all found within the same co-expression module. This further corroborates the hypothesis that the similarities of all the AT outputs is related to *HAG1/myb28*, as the low GSL phenotype observed in *B. napus* is a consequence of a deletion/homologous exchange (Harper et al., 2012; Lu et al., 2014; He et al., 2016). Thirdly, this approach has highlighted some genes which were not initially detected in the AT analysis. For example, the marker with the highest degree (or connectedness) is an orthologue of *SEUSS-like 2* (SLK: AT5G62090, Bo2g161850.1). This gene is known to be involved during embryogenesis and floral development through the coordination of auxin distribution (Bao, Azhakanandam and Franks, 2010) and could therefore play a role in the nutrient status of the developing seed. Finally, the WGCNA analysis has also been able to provide additional evidence to support some of the less well annotated genes flagged by AT analysis. For example, *Per1-like* (*PerL*: AT5G62130 orthologue of Bo2g162860.1/ Bo4g142360.1); is a gene that is poorly characterised (Moriyama et al., 2006) with no known link to S or glucosinolates, which has yet been identified within the most closely associated module for seed S concentration and is one of the most highly associated GEMs from AT analysis.

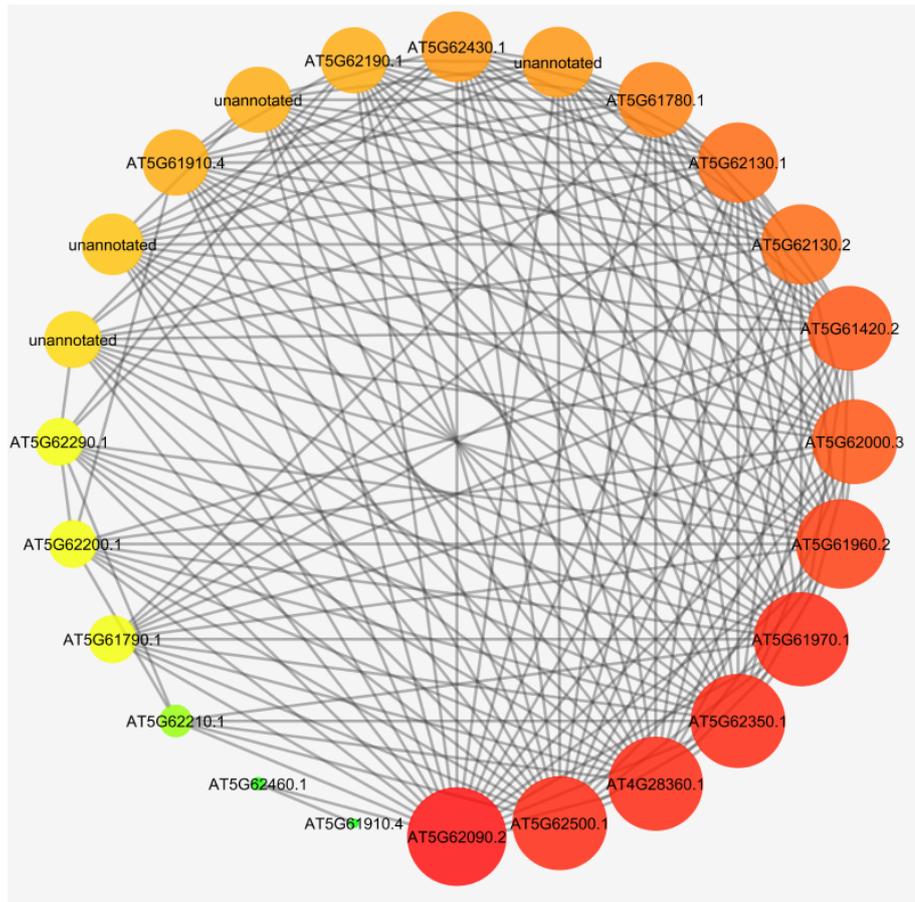


Figure 4.3.1.a WGCNA (weighted gene co-expression network analysis) results for the top module (Brown4) from S concentrations in seed.

Results are presented as a degree sorted circle representing those genes with the greatest degree/connectivity as larger red nodes, while the least connected genes are represented by smaller green nodes. The gene within the node is indicated by its AGI code, while those markers which are unannotated in *A. thaliana* have been labelled as such.

Table 4.3.1.a Markers from the most highly correlated module with seed S concentration from WGCNA analysis.

Pseudomolecule marker names, AGI codes and a brief description of gene functionality (from The Arabidopsis Information Resource(TAIR, 2015)) are provided in a degree sorted order. Markers with higher degree are the most highly connected.

Marker	ATG	Description	Degree
Bo2g161850.1	AT5G62090.2	SEUSS-like 2	22
Bo2g164170.1	AT4G28360.1	Ribosomal protein L22p/L17e family protein	21
Bo2g161790.1	AT5G61970.1	signal recognition particle-related / SRP-related	21
Bo2g164070.1	AT5G62350.1	Plant invertase/pectin methylesterase inhibitor superfamily protein	21
Bo2g164190.1	AT5G62500.1	end binding protein 1B	21
Bo2g161770.1	AT5G61960.2	MEI2-like protein 1	20
Bo2g161590.1	AT5G61420.2	myb domain protein 28	19
Bo2g161810.1	AT5G62000.3	auxin response factor 2	19
Bo2g162860.1	AT5G62130.1	Per1-like family protein	18
Bo4g142360.1	AT5G62130.2	Per1-like family protein	18
Bo2g161630.1	AT5G61780.1	TUDOR-SN protein 2	17
Bo2g164130.1	AT5G62430.1	cycling DOF factor 1	16
Bo8g091260.1	unannotated		16
Bo2g161730.1	AT5G61910.4	DCD (Development and Cell Death) domain protein	15
Bo2g163990.1	AT5G62190.1	DEAD box RNA helicase (PRH75)	15
Bo2g162870.1	unannotated		15
Bo2g161500.1	unannotated		14
Bo2g164140.1	unannotated		13
Bo2g161640.1	AT5G61790.1	calnexin 1	11
Bo2g164000.1	AT5G62200.1	Embryo-specific protein 3, (ATS3)	11
Bo2g164050.1	AT5G62290.1	nucleotide-sensitive chloride conductance regulator (ICln) family protein	11
Bo2g164020.1	AT5G62210.1	Embryo-specific protein 3, (ATS3)	7
Bo2g164150.1	AT5G62460.1	RING/FYVE/PHD zinc finger superfamily protein	2
Bo2g161720.1	AT5G61910.4	DCD (Development and Cell Death) domain protein	1

4.3.2 Seed GSL and sulfate analysis

For the analysis of seed GSL 18 accessions (representing the 9 highest and 9 lowest total S concentration lines, with 4 replicates/seeds from 4 separate plants) were analysed with HPLC. A correlation was found between total seed S and GSL in seed, helping to support the hypothesis that the main source of S concentration variation in seed comes from differences in GSL content (**Figure 4.3.2.a**, $r^2=0.854$, $n=18$, $p<0.001$). Based on the stoichiometry of S to GSL, with the basic structure of a GSL having two S molecules (Halkier and Gershenzon, 2006), a linear relationship between S and GSL in the seeds could be expected, as is broadly observed in the current research. The relative contribution of proteins and lipids to the S concentration of seeds has previously been deemed relatively constant between high S and low S lines (Bloem, Haneklaus and Schnug, 2005), sulfate however does not seem to have been studied in depth. Therefore it was deemed important to measure the seed sulfate concentration to prove that it was purely variation in seed GSL content responsible for the relationship between total S, Mo and the GSLs. To measure seed sulfate two different extraction protocols were used (see 4.2.3); the cold protocol was performed initially utilising the same 18 accessions that were analysed for GSL content, while the hot protocol analysed a subset of this group, measuring sulfate in the 3 highest and 3 lowest total S lines (again with 4 replicates). The repetition of the experiment with different methods was deemed necessary as it was thought the original cold protocol had inadequately controlled for myrosinase degradation of GSL into sulfate which would skew analysis (Bones and Rossiter, 1996). However when the two methods were directly compared based on the same six accessions it was found that there was no significant difference between the results of the two protocols (t-test: $t=0.32$, $df=10$, $p=0.753$) and the cold extraction protocol was in fact showing less variation in comparison to the hot extraction method for high S lines (**Figure 4.3.2.b**). Considering then that more accessions were measured utilising the initial cold extraction protocol it was decided that these results would be used.

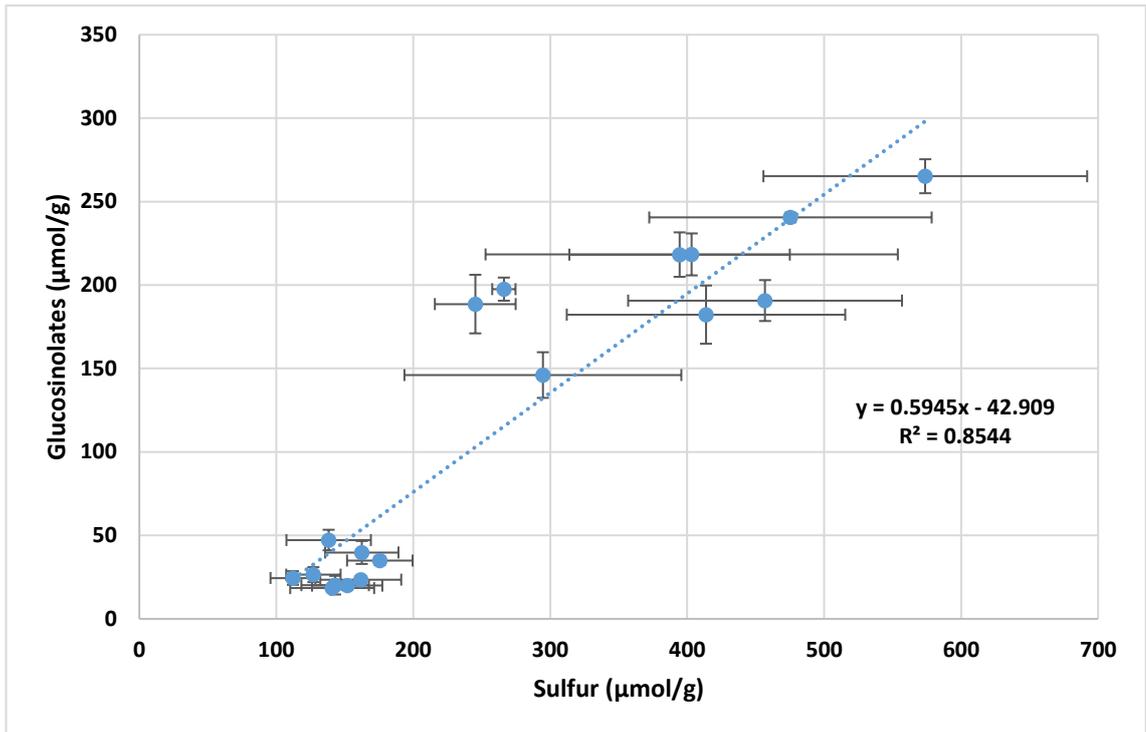


Figure 4.3.2.a Correlation between the total S and total glucosinolate ($\mu\text{mol/g}$) content of *B. napus* seeds ($r^2 = 0.8544$, $p < 0.001$).

Four independent seed samples of the 18 accessions representing 9 of the highest and 9 of the lowest total seed S lines were analysed for glucosinolate content with HPLC (average per accession displayed with standard error bars for each measurement, S data comes from the original RIPR measurement of seed S concentration).

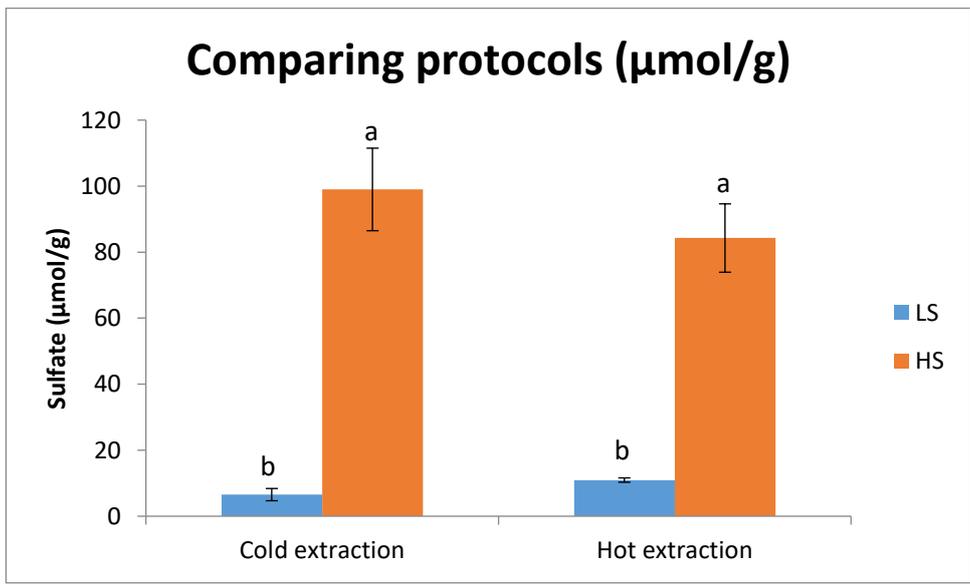


Figure 4.3.2.b Investigating the fidelity of two different extraction protocols for seed sulfate ($\mu\text{mol/g}$) extraction in *B. napus*.

Six accessions with 4 independent seed samples were analysed for seed sulfate with two different extraction protocols: one utilising a cold incubation, the other a hot incubation. Average sulfate content for high S (orange) and low S (blue) accessions are presented with +/- one standard error as error bars. The two methods were compared with a one way ANOVA with post hoc Tukey and Bonferroni tests (ANOVA: F 52.46, df 3, 8, p <0.001). As expected there was a significant difference between the high and low S lines, however there was no significant difference between the two extraction methods. As such the larger data set from the cold extraction protocol was carried forward into further analysis.

The considerable variation observed between the sulfate concentration of high total S lines and low S lines (t-test: t=13.10, df ~9.18, p<0.001, n 18) was surprising for a number of reasons. Firstly, sulfate was not expected to vary significantly within seeds (Bloem, Haneklaus and Schnug, 2005) and was in fact significantly correlating with the GSL content ($r^2=0.974$, p<0.001, n 18). This was unexpected as it has been suggested within the literature that the main site of GSL biosynthesis is actually the pod wall (Bloem, Haneklaus and Schnug, 2007). This observation could be explained simply by a supply and demand effect: plants which produce more GSL require a greater supply of free sulfate to the pod walls which may also give the seeds access to a greater supply of sulfate. These observations would fit in well with the Mo hypothesis, that Mo was correlating negatively with seed sulfate ($r^2=0.476$, p<0.01, n 18) as would be expected if the two were competing for a shared transporter (Bittner, 2014). Alternatively, a more controversial theory has been suggested: that there may be some level of GSL biosynthesis within seeds. It has been suggested that the seeds are capable of reductive assimilation of sulfate and could incorporate them into the sulfonate moiety of GSL (Toroser, Griffiths and Thomas, 1995). This again would provide an explanation as to why seed sulfate was high in the high S lines, with Mo following as a structural analogue of sulfate. On the other hand, other researchers have advocated that lines with low GSL have both a metabolic block in GSL production and seeds with selectivity against 'incomplete GSL' (Bloem, Haneklaus and Schnug, 2007; Josefsson, 1971, 1973). This research found that lines with low GSL seed showed an increase in S within the pod walls and suggested that the S becomes 'immobilised' (Josefsson, 1971, 1973; Bloem, Haneklaus and Schnug, 2007). This may explain why lines with low total S had reduced concentrations of sulfate (i.e. with the S being trapped in a 'dead end' process/the vacuole) and the negative relationship to Mo as it would have increased access to any shared transporters. The only other link between GSL/S and Mo would require Mo to play a biosynthetic role in GSL production. This is known to be the case in *A. thaliana* for benzyl glucosinolates (Ibdah et al., 2009) although these are thought not to occur in *B. napus* seeds in large concentrations (Toroser, Griffiths and Thomas, 1995). Finally, it could be that neither protocol effectively controls for GSL breakdown into sulfate. This could explain why there

appears to be little information on the sulfate/GSL relationship within the literature, i.e. as it is difficult to conclusively separate GSL from sulfate concentrations (due to myrosinase degradation of GSL) they have not been widely reported within the literature.

4.3.3 Leaf senescence analysis

HS and LS lines were compared for differences in remobilisation by comparing their senesced leaves from a field trial with a low (LN, 60 kg N/ha) and 'high' (HN, 300 kg N/ha) nitrogen conditions. This analysis was split across two different digestion protocols and ICP-MS runs (see 4.2.4). As such, two different CRMs were used in comparison with the data outputted, restricting the number of elements which could be assessed across both analyses. ICP-MS analysis generated information for 31 elements, however 12 did not have a CRM values (Ag, Al, Be, Cr, Cs, Li, Pb, Rb, Ti, Tl, U and V), 6 had a recovery <85% in one of the runs (Ba, Ca, Co, Fe, Na and Ni) and the average measurement for Cu concentrations across samples in the second run was below the LOD. Therefore, 12 elements were left for direct comparison (As, B, Cd, K, Mg, Mn, Mo, P, S, Se, Sr and Zn). The two ICP-MS runs were compared for the different HN lines. A two sided t-test was used on each of the elements analysed. Only Cd was highlighted as being significantly different across the two ICP-MS runs and therefore excluded from further analysis (Cd t-test: $t = -2.79$, $df = 16$, $p < 0.05$). Overall, 11 elements (As, B, K, Mg, Mn, Mo, P, S, Se, Sr and Zn) were compared across senescing leaves for difference in remobilisation between HS/LS and HN/LN.

It was observed that the LS lines on LN had significantly higher S concentrations remaining in senesced materials in comparison to either HS or LS lines on HN (**Figure 4.3.3.a**). This result corroborates the previously reported interaction between S and N status of plants, i.e. where N is limiting it can have adverse effects on other elements within the ionome. This may imply that under LN condition the LS lines are less efficient at remobilising nutrients from leaves or rather that under LN conditions LS plants accumulate more S within their leaf tissues. It is clear, however, that there is a difference in how the HS and LS lines are using S within their tissues, but that it does not have a significant effect under normal N fertilisation conditions. This could impact NUE where the aim is to grow plants with less fertiliser. However, this requires further investigation to understand whether there is an impact on crop yield and if either the HS or LS lines display an advantage under the lower N conditions. The different treatments also showed variation in Mo concentration (**Figure 4.3.3.a**) however in this instance the difference occurs between HS/HN plants and HS/LN plants. The HS/HN lines had significantly higher Mo concentrations in comparison to the HS/LN (and LS/LN lines). This is likely to be a consequence of the dependence of N assimilation on the Mo containing enzyme

Nitrate reductase (required in the first and rate limiting step in N assimilation (Schwarz and Mendel, 2006)). When there is an abundance of N, Mo will be required to a much greater extent to enable the effective assimilation of the additional N. Whether this could have an effect on GSL concentrations (which also require N) needs further study.

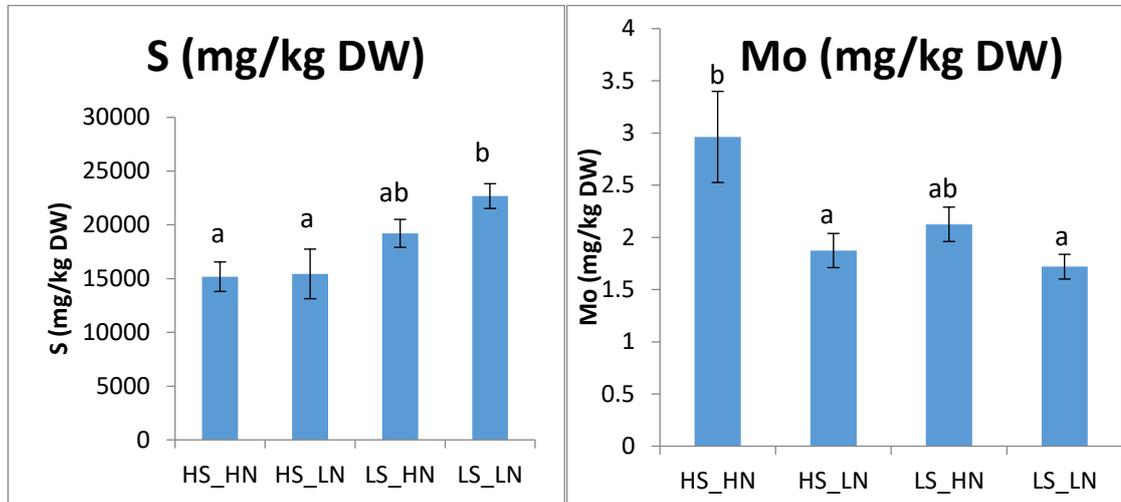


Figure 4.3.3.a S (left) and Mo (right) concentrations (as mg/kg DW) within senesced leaves of high S (HS) and low S (LS) seed lines under different N conditions.

Mean and standard errors are displayed for 9 ecotypes within each grouping (HS/HN, HS/LN, LS/HN and LS/LN). Significant differences are indicated where the letters are not shared. S ANOVA: F 4.94, df 3, 32, p 0.006; Mo ANOVA: F 4.75, df 3, 32; p 0.007.

Of the other elements assessed, only Sr concentrations showed a significant difference between the groups, with LS_LN having a significantly larger concentrations in senesced tissues in comparison to the HS_HN leaves (ANOVA: F 4.26, df 3,32, p<0.05). This does not provide much biological detail as the difference is across both treatments (i.e. not between HS/LS or HN/LN) and was therefore not further investigated. All other element concentrations analysed showed no significant difference between any of the groups.

4.3.4 Pod ionome investigation

As the pod is meant to be the primary site of GSL biosynthesis in *B. napus* and the seed and leaf ionome of *A. thaliana* seem to be perturbed by variation in *HAG1/Myb28* (3.3.2), it was decided that HS and LS *B. napus* lines would be assessed for variation within the pod and stem. Of the 23 elements analysed (B, Na, Mg, P, S, K, Ca, Li, Be, Al, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Sr, Mo, Cd), 14 had a percentage recovery $\geq 85\%$ (B, Na, Mg, P, S, K, Ca, Mn, Zn, As, Se, Sr, Mo, Cd). Some of the essential micronutrients had poor recoveries; Cu at 79%, Ni at 76% and Fe at 69%. If the average concentration of an element was below the LOD in one tissue (seed, stem or pod) it was excluded from further analysis as the aim was to look for differences

across tissues. This includes 11 elements: Na, Li, Be, Al, V, Cr, Fe, Ni, As, Se and Sr. As a result, 12 elements were taken forward into analysis for variation in green seeds, pods and stems (B, Mg, P, S, K, Ca, Mn, Co, Cu, Zn, Mo, Cd).

S concentrations were found to significantly vary within the green seed, pod and stems of HS and LS lines (**Figure 4.3.4.a**). Within seeds the relationship was the same as that previously observed in the RIPR panel, i.e. the high S seed lines had high S and vice versa. However, the pod data was somewhat more complicated with the low S lines having slightly higher S concentration than the high S lines (i.e. the opposite of what was observed in seed with the LS_S lines having significantly more S, $p < 0.05$). The stems also mirrored this dynamic, with the LS_S ecotypes having significantly higher S concentrations than all the others ($p < 0.05$). This variation between winter OSR/spring OSR/LS/HS ecotypes is in line with previous research which investigated various stages of seed and pod development in winter and spring OSR ecotypes (Bloem, Haneklaus and Schnug, 2007). This study showed that winter OSR ecotypes seemed to accumulate S and GSL much more slowly and consistently than spring OSR ecotypes throughout pod development. This may account for the differences observed between LS_S and LS_W ecotypes in the pod and stem data sets. If the winter OSR ecotypes take longer to accumulate GSL the difference in S concentrations between the low and high S lines may not have fully established at the time of sampling. This suggests a link between how the pods develop and how the GSL' are being synthesised/transported into the seeds. The current study aimed to assess the pods of spring and winter OSR ecotypes at the same developmental time point; plants were sampled after ~35 days post flowering when the pods were still green and looked approximately of equivalent developmental stage (e.g. similar size and colouration). However, since differences were observed between the LS_W and LS_S ecotypes this was unsuccessful. Nevertheless, it was interesting that the increase in S concentration within the pods of LS_S lines was accompanied with an increase in the S concentration within the stems. This was previously unreported within the literature and implies a wider level of disruption in S concentration across low glucosinolate spring OSR varieties. It would be interesting to measure the sulfate and GSL concentrations within the stems of these plants to see if they were also perturbed, as this would have implications for theories concerning the movement of GSL into developing seeds.

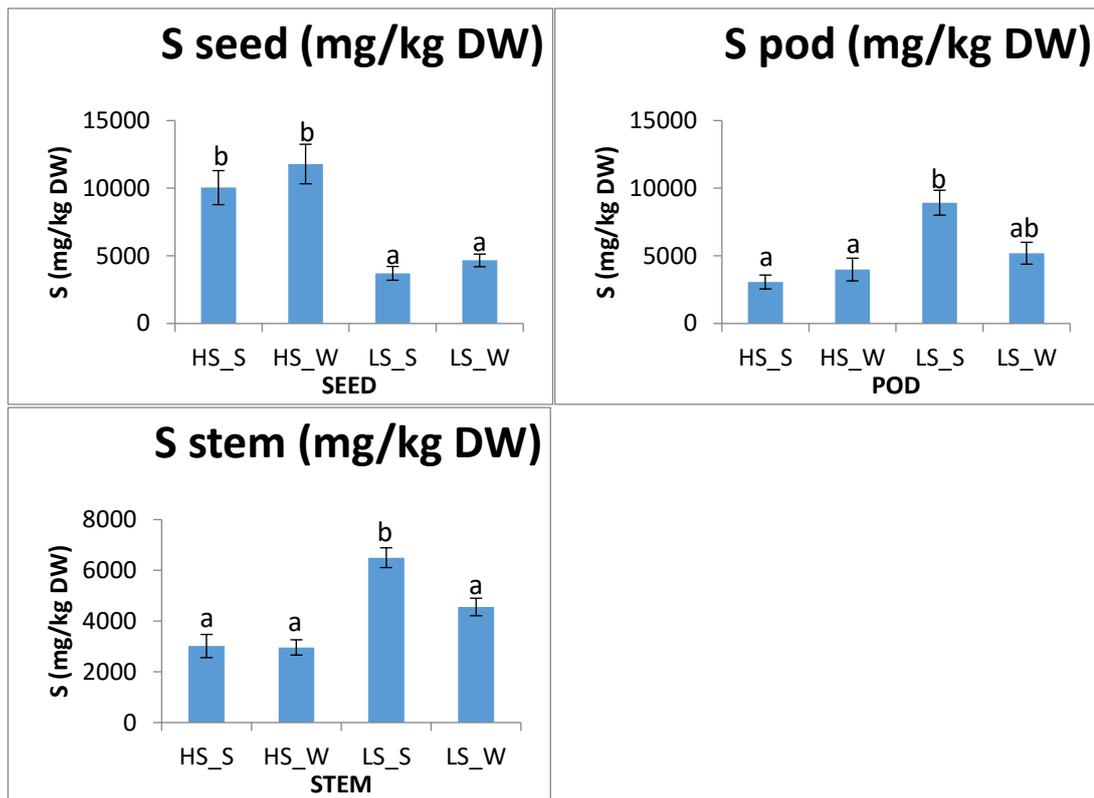


Figure 4.3.4.a Difference between the green seed, pod and stem of high S (HS) and low S (LS) seed lines in spring (S) and winter (W) OSR ecotypes for S concentration (mg/kg DW)

Values for S (as mg/kg) are given for green seed (top left), pod (top right) and stem (bottom). The mean and standard error are displayed for each, 3 accessions for each (HS_S; high S spring OSR, HS_W; high S winter OSR, LS_S; low S spring OSR, LS_W; low S winter OSR). Different letters indicate a significant difference between means calculated from a post hoc bonferroni test ($p < 0.05$) of ANOVA results: SEED [F 29.15, df 3,8, n 12, $p < 0.001$]; POD [F 10.73, df 3,8, n 12, $p < 0.004$]; STEM [F 19.44, df 3,8, n 12, $p < 0.001$].

Although S was the only nutrient to vary between the HS and LS lines across all tissues, variation was observed within the green seeds for Mo concentrations. This was exactly the same pattern as previously observed in the rest of the diversity panel (i.e. an increased concentration of Mo in the low S lines, particularly LS_S, **Figure 4.3.4.b**). However, since there is not a concomitant increase in the Mo concentration within pods it is clear that this is a seed specific effect. It is interesting that there is a significant difference in Mo concentration within the seeds of the LS spring and winter OSR ecotypes. The pattern of Mo and S concentrations could imply a shared sulfate transporter, i.e. higher concentration of Mo within the seeds of the LS_S ecotypes because of the increased S concentrations in the LS_S pods/stem. If the plants are moving more S to the pods but are for some reason unable to transport this S into the seeds it would enable an increase in movement of Mo into the seed via a shared

transporter. It would be interesting to assess if the flux of Mo into the seeds was higher in LS accessions. Alternatively, perhaps Mo is being used for the biosynthesis of aromatic GSL, in which case an analysis of the composition of GSL' in HS and LS accessions would be relevant.

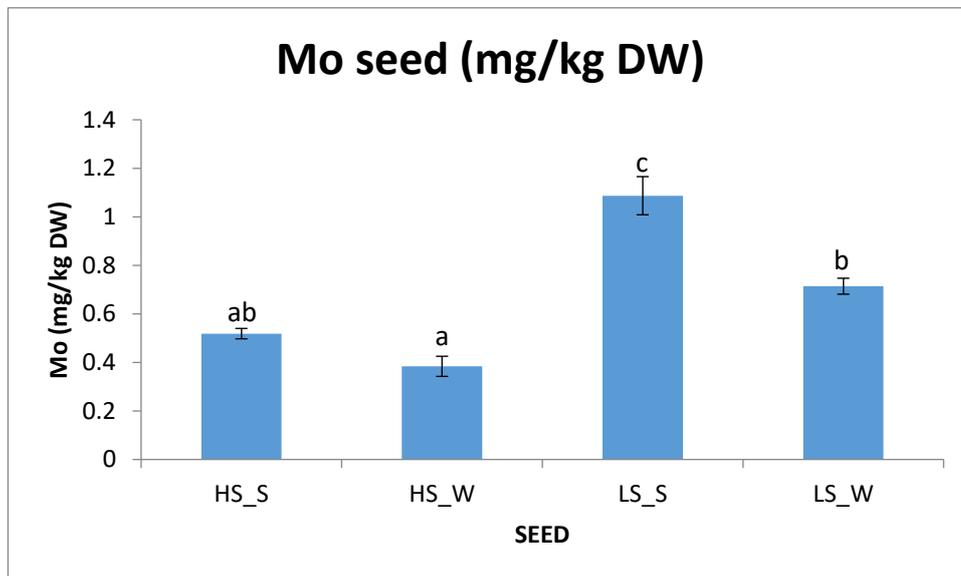


Figure 4.3.4.b Difference in Mo (as mg/kg DW) in the seeds of high S (HS) and low S (LS) seed lines in spring (S) and winter (W) OSR *B. napus* ecotypes.

The mean and standard error are displayed, 3 accessions for each (HS_S; high S spring OSR, HS_W; high S winter OSR, LS_S; low S spring OSR, LS_W; low S winter OSR). Different letters indicate a significant difference between means calculated from a post hoc bonferroni test ($p < 0.05$) of ANOVA results: [F 39.61, df 3, 8, n 12, $p < 0.001$]

Of all the other element concentrations highlighted within AT analysis (B, Zn, Mg and Sr) none showed any variation between HS or LS accessions in seeds, pods or stems. This was unsurprising as selection of lines was based on their S seed (high in comparison to low) and flowering time (spring/early in comparison to winter/late) phenotypes. Testing the green seed was merely a way of confirming that the phenotype (i.e. high and low S concentrations) was occurring within the seeds when the pods and stems were being analysed for ionic variation. Nevertheless the results may provide some information at least for B. Within the RIPR diversity panel mature seed data there is a strong negative correlation ($p < 0.001$) between seed S concentration and B concentration (**Figure 3.2.7.a** seed interactome). When the green seeds, pods and stems were analysed there was no significant difference between the HS and LS groups for B concentrations (ANOVA F0.16, df 3, 8, n 12, p 0.922). This could reflect partitioning of B into the seed coat (which at the time of pod sampling was yet to develop), which is thought to be the primary site of B storage in the seed (Eggert and von Wirén, 2016).

On the other hand, it could be that the lines selected for the current analysis did not display a significant difference between the HS and LS lines used (ANOVA F 3.36, df 3, 8, n 12, p 0.076). The link between B, S and GSL would be easy to test across the diversity panel merely by analysing the seed coats separately with ICP-MS. Why B and potentially the seed coat would be linked to the GSL/S content of seed is unknown, but may support the developmental hypothesis suggested by (Bloem, Haneklaus and Schnug, 2007) or other biological traits that have been progressively bred into crop plants (e.g. low GSL lines have been bred for increased yield, perhaps this has affected seed size and therefore B concentration as part of breeding programmes).

Within the pods, Se and K were shown to vary between the four treatments (**Figure 4.3.4.c**). K, however, did not show any significant difference between treatments when post-hoc analysis was performed (ANOVA: F 5.06, df 3, 8, n 12, p 0.03). Se showed a similar pattern to S within the pods, with the LS_S lines having the highest concentration. Like Mo, Se is also known to be a structural analogue of sulfate, which may be an additional indicator that perhaps sulfate movement within the plants has been perturbed. Nevertheless, the difference was not significant when compared to HS_S ecotypes, again perhaps indicating that the way that winter and spring OSR ecotypes are accumulating elements is different.

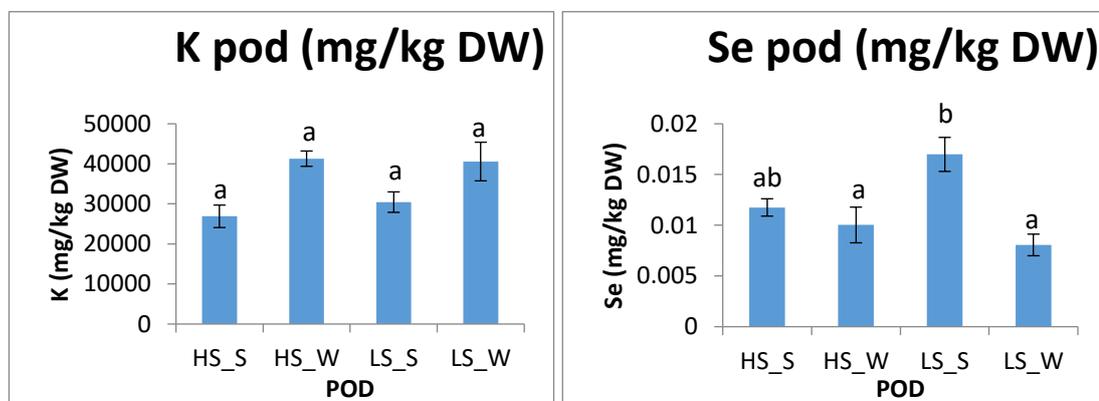


Figure 4.3.4.c Difference in K and Se (as mg/kg) in the pods of high S (HS) and low S (LS) seed lines in spring (S) and winter (W) OSR *B. napus* ecotypes.

The mean and standard error are displayed, 3 accessions for each (HS_S; high S spring OSR, HS_W; high S winter OSR, LS_S; low S spring OSR, LS_W; low S winter OSR). Different letters indicate a significant difference between means calculated from a post hoc bonferroni test ($p < 0.05$) of ANOVA results: K [F 5.06, df 3, 8, n 12, $p < 0.05$] and Se [F 7.4, df 3, 8, n 12, $p < 0.01$].

Within the stems only S concentrations showed a consistent pattern among ecotypes: P, K and Mn concentrations all showed significant differences between the various ecotypes under assessment, however the differences highlighted with post hoc Bonferroni testing provided little additional insight into how the seed ionome might be varying with the GSL' (**Figure 4.3.4.d**). For example, K concentrations once again provided no informative data; only the LS_W and HS_S lines showed significant differences ($p < 0.05$). P concentrations displayed a significant difference between spring and winter OSR ecotypes in LS lines, while Mn concentrations showed a significant difference between spring and winter OSR ecotypes of HS lines. Once again this may substantiate the observation that winter and spring OSR ecotypes acquire/utilise nutrients differently (see 6).

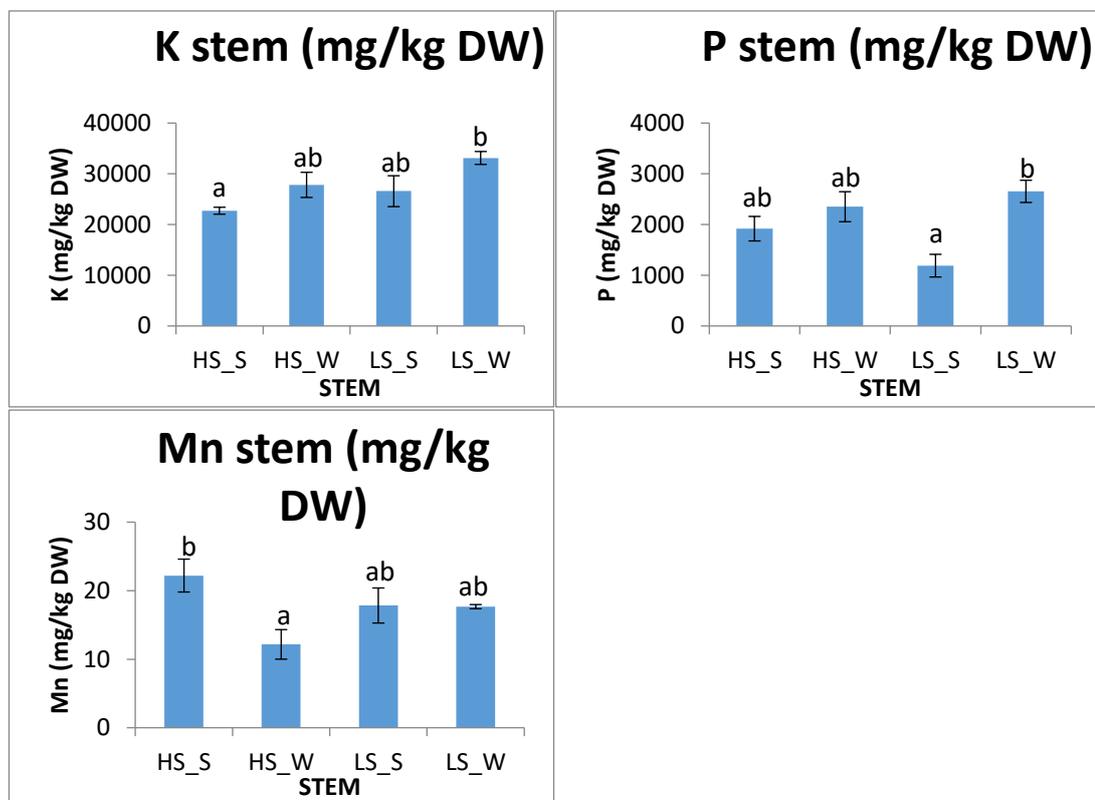


Figure 4.3.4.d Difference in K, P and Mn concentration (as mg/kg DW) in the stems of high S (HS) and low S (LS) seed lines in spring (S) and winter (W) *B. napus* (OSR) ecotypes.

The mean and standard error are displayed, 3 accessions for each (HS_S; high S spring OSR, HS_W; high S winter OSR, LS_S; low S spring OSR, LS_W; low S winter OSR). Different letters indicate a significant difference between means calculated from a post hoc bonferroni test ($p < 0.05$) of ANOVA results: K [F 4.22, df 3, 8, n 12, $p < 0.05$]; P [F 6.63, df 3, 8, n 12, $p < 0.05$] and Mn [F 5.93, df 3, 8, n 12, $p < 0.05$]

4.4 Chapter summary and conclusions

The aim of this chapter was to investigate the relationship between seed GSL and the wider seed ionome. It began by verifying the observations made as part of AT analysis with WGCNA analysis of seed S and Mo concentrations. The most highly correlated module for S and one of the most highly correlated modules for Mo contained an orthologue of *HAG1/Myb28* (the transcription factor involved in regulating aliphatic GSL biosynthesis). This gene had previously been shown to disrupt the seed and leaf ionome of *A. thaliana* T-DNA lines discussed in the previous chapter (3.3.2). To further support previous research on S and GSL, a subset of high and low S (HS and LS) seeds was assessed for its sulfate and GSL concentration. There was a significant correlation between S and GSL concentrations in seed, as well as sulfate. It was initially thought that the remobilisation of nutrients could be different between the HS and LS lines. To test if there were differences in remobilisation, the senescing leaves of HS and LS lines were tested on high (HN) and low N (LN) conditions. There was a significant difference between the HS and LS lines under LN conditions with the LS seed lines having a greater concentration of S within the senesced materials. This demonstrated the close relationship between S and N nutrition in plants and could have implications for the growth of LS seed lines under limited N. Finally, the pod ionome investigation was performed, contributing to the leaf senescence analysis. Once again the LS lines generally contained significantly higher S concentrations, supporting the observation that plant wide S concentration had been perturbed during breeding for low seed GSL. However, no wider effect on the rest of the seed, pod or stem ionome was observed. The seed ionome therefore requires further investigation in relation to *HAG1/myb28* and the GSLs. This could be investigated with specific research focusing on sulfate concentrations in stems and pods. It is possible that the linking of multiple elements to the GSL can be explained by the interlinking nature of nutrient status (abiotic) and biotic responses (GSL being defence compounds). Alternatively, the close associations could be the result of years of breeding for low GSL' indirectly selecting candidates in the surrounding region which could affect the nutrient concentration in the seeds.

5 Leaf ionome investigation

5.1 Introduction

Within leaf a total of 28 elements was analysed with ICP-MS analysis (see General methods, 2.1.1). Of these, only Mo, S, Mn, Cu, Zn and Cd were assessed as part of this study, as detailed in the introduction (see 1.2.5). As with the seed ionome investigation (see 3), this chapter on the leaf ionome will detail the leaf AT results for the elements under study individually. Again for reference the heritability of each element under investigation is included from analyses performed as part of Thomas et al., (2016), see **Table 4.3.4.a**. It will consider the predictive capacity of markers, discussing potential candidates within association regions and introducing those which have been taken forward to further analysis with *A. thaliana* T-DNA lines. Once again, many elements displayed similarities between AT results/selected candidates. Where this occurs it will be discussed in detail and candidates analysed accordingly. Distinct similarities were observed between almost all GEM association analyses of leaf element concentrations; orthologues of the floral regulators *SOC1* and *FLC* were consistently within the top GEM hits. A lot of previous research in the area of flowering has suggested links between individual elements and flowering time, c.f. (Hall, Savin and Slafer, 2014). However, within this study, multiple element concentrations displayed a close association to *SOC1* and *FLC*, implying a link between the wider ionome and flowering (not just individual macronutrients). As such, these relationships were analysed further as part of a separate chapter (see 6).

Table 4.3.4.a variance component analysis from Thomas et al., (2016) for leaf mineral composition in *B. napus*.

Variation (as a %) is shown for each element under investigation associated with genotype, habit, experimental design and residual factors, calculated from Residual Maximum Likelihood (REML) analyses.

Response variate:	Cu	Cd	Mn	Zn	Mo	S
Genotype	17	11	15	24	22	40
Habit	5	1	5	6	10	15
Experimental design	29	59	48	30	6	17
Residual factors	49	29	32	40	61	28

5.2 Analysis of individual elements within leaves:

This section outlines the AT analysis of Cu, Cd, Mn, Zn, Mo and S concentrations in the leaf. It gives AT results and assesses the predictive capabilities of markers, introducing potential candidate genes to be tested in *A. thaliana*.

5.2.1 Associative transcriptomic outputs, predictions and candidates: Cu concentration in leaf material

AT analysis of Cu concentration in leaves of the 383 diversity panel revealed two relatively small SNP association peaks on A2/C2, with the association peak on C2 just passing the Bonferroni corrected significance threshold (**Figure 5.2.1.a**). None of the GEMs cleared the Bonferroni corrected significance threshold and they all failed to give any clear association peaks. However, when Q-Q plots were assessed for Cu concentration in leaves within the GEM analyses it becomes apparent that the AT model is overcorrecting for type 2 errors, as the observed p values are much lower than those the model expects (**Figure 5.2.1.b**). This is perhaps emphasised with the predictions made from the 274 diversity panel; despite the clear association peaks identified within the SNP association analyses no predictive markers were found, while four predictive markers were found within the GEM AT association analyses (**Table 5.2.1.a**).

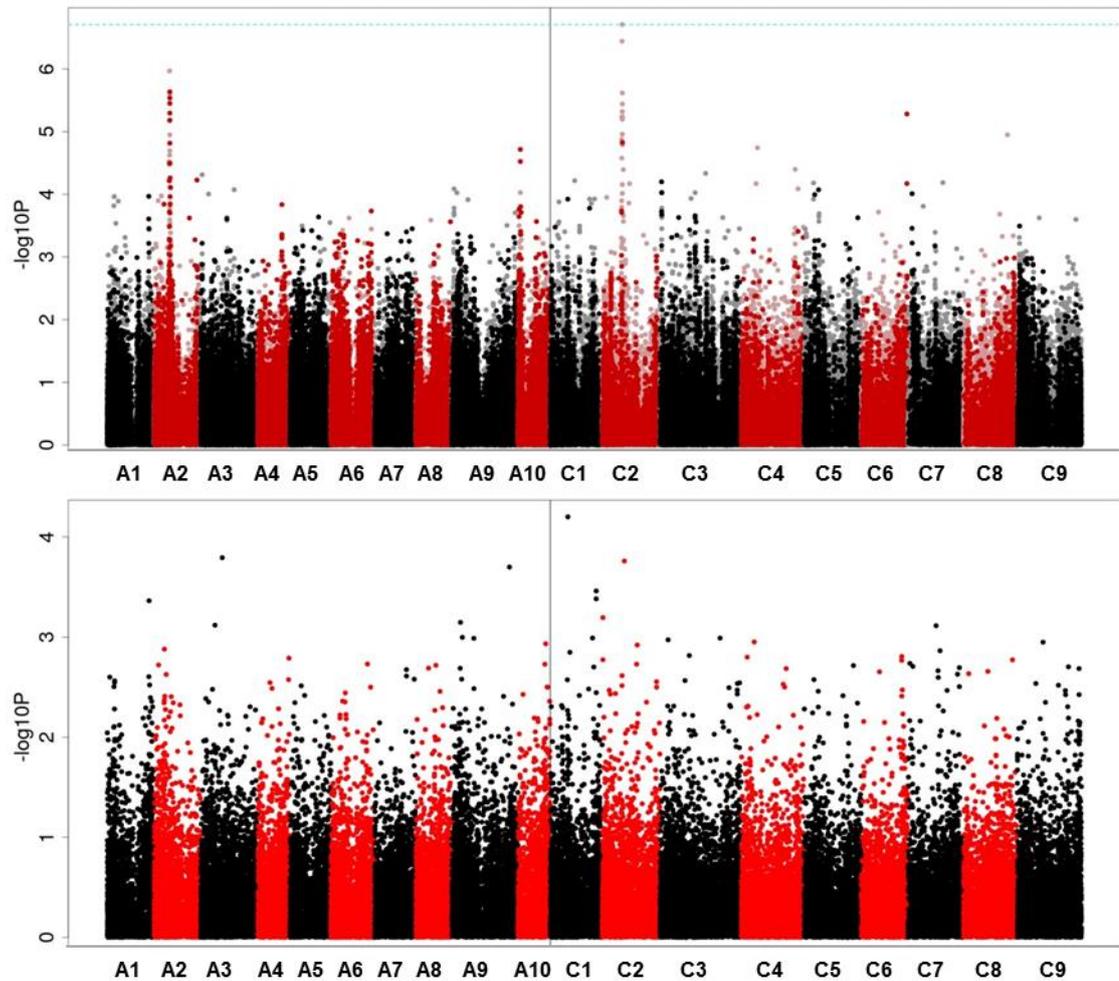


Figure 5.2.1.a Genome wide distribution of mapped markers associating with the Cu concentration in leaves (mg/kg DW) of all 383 accessions.

The average concentration of Cu in leaves was calculated from 5 separate plants from ICP-MS analysis for each of the 383 accessions. SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and leaf Cu concentration as the response variable. $-\text{Log}_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

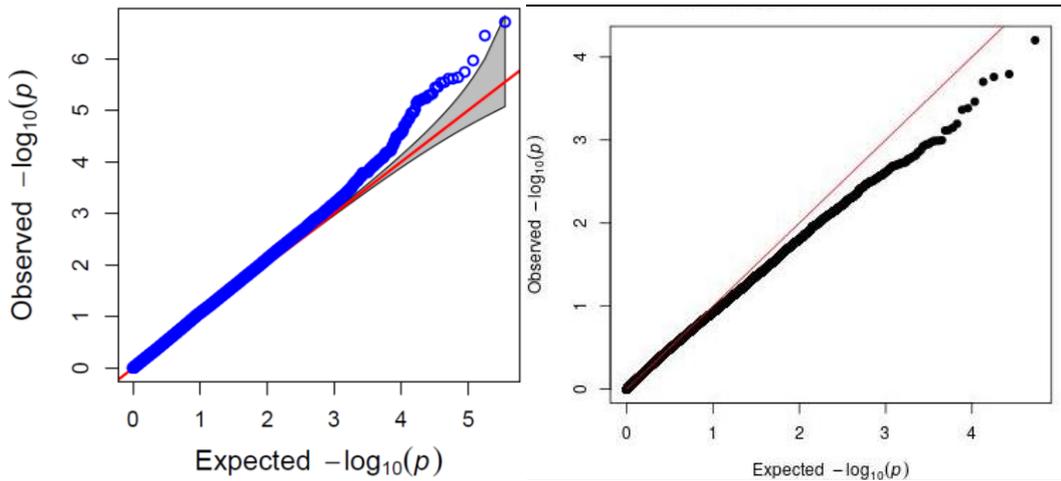


Figure 5.2.1.b Quantile-quantile plot of observed $-\log_{10}P$ values from AT SNP (left) and GEM (right) analysis for Cu concentration in leaves against expected $-\log_{10}P$ values

The red line indicates a theoretical perfect fit of the expected $-\log_{10}P$ values, while the grey area indicates the 95% confidence interval under the null hypothesis of no association between the Cu concentration in leaves and SNPs/GEMs.

Table 5.2.1.a Predictive capability of markers from AT analysis of Cu concentration in leaves.

For assessing the predictive capability of markers the highest SNP markers from discernible association peaks and the most highly associated GEMs were analysed. The marker type is given as either SNP or GEM, alongside their name and position, followed by their $-\log_{10}P$ value from the 274 AT analysis. Finally, the correlation coefficient (R), significance (p) and sample size (n) are given for the predictions made on the 109 diversity panel. Markers that were significantly predictive are highlighted in bold.

Marker type	Marker	Position	AT 274 $-\log_{10}P$	R	p	n
GEM	Bo2g057690.1	C02_017047389_017048454	9.16	0.284	0.003	109
GEM	Bo1g039440.1	C01_011993289_011994233	9.09	0.326	<0.001	109
GEM	Bo1g139020.1	C01_039714822_039716964	8.01	0.281	0.003	109
GEM	Cab040172.1	A09_005225883_005223304	7.95	0.206	0.032	109
SNP	Bo2g051300.1:1383:T	C02_014686167_014690727	5.02	0.185	0.069	97
SNP	Cab031385.2:756:T	A02_009517638_009520274	4.42	0.150	0.126	106
SNP	Cab039789.1:318:T	A03_006053883_006054242	4.11	-0.019	0.847	106
SNP	Bo2g018010.1:1818:A	C02_005166106_005171591	3.38	0.152	0.152	91

Despite finding no predictive markers within the SNPs a number of potential candidates was observed. This lack of predictive capacity was common within these analyses, highlighting the limitations of GWAS for identifying rare allelic variants and of using a split diversity panel for testing the predictive capability of markers (see 7.1). Within the A2/C2 association region the candidate AT1G66240 known as *ATX1* (*antioxidant like 1*) should occur, however it does not

have a CDS gene model within the pan-transcriptome. Nevertheless, based on the AGI information of the CDS gene models within the A2/C2 region it should be found within this region. *ATX1* is a known Cu chaperone and is required for Cu homeostasis (Shin, Lo and Yeh, 2012). It is interesting that *ATX1* should be present within the leaf SNP association analyses when *HMA5* is present within the seed SNP association analyses (see 3.2.1) as it has been suggested the two interact (Andres-Colas et al., 2006; Puig et al., 2007b). It has been suggested that *ATX1* is responsible for delivering Cu to *HMA5* for detoxification within the roots and delivery to the shoot (Andres-Colas et al., 2006; Puig et al., 2007b). The presence of *ATX1* within the leaf SNP association analyses and *HMA5* within the seed SNP association analyses could be highlighting the primary site of *ATX1/HMA5* action. Perhaps *ATX1* works redundantly with other chaperones in the roots to deliver Cu to *HMA5* (with this root to shoot translocation being essential for seed Cu concentrations), while *ATX1* works specifically with some other transporter in the leaves (e.g. *COPT5*) playing an essential role in leaf Cu concentration (Shin, Lo and Yeh, 2012). As *ATX1* insertional mutants have already been well characterised in *A. thaliana* (Shin, Lo and Yeh, 2012) it was not investigated further as part of this study.

Given that *ATX1* does not have a CDS gene model within the pan-transcriptome, the A2/C2 association region was explored for other candidates. There were no other candidates found within the region with a described role in Cu concentration. As such, analysis focused on the most highly associated candidates in the region whose functional annotations did not rule them out of a potential role in Cu concentration. The first of these was Bo2g052580.1, whose orthologue in *A. thaliana* is described as Polyamine Oxidase 4 (PAO4: AT1G65840.1); while the second was Bo2g052640.1, whose orthologue in *A. thaliana* is described as Cytosolic NADP+ dependent Isocitrate Dehydrogenase (CICDH: AT1G65930.1). PAO4 is thought to catalyse the oxidation of spermine into spermidine within the root peroxisome (i.e. polyamine catabolism (Kamada-Nobusada et al., 2008)) and is thought to play a role in cell oxidative balance (Sequera-Mutiozabal et al., 2016). Polyamines such as spermine/spermidine have been implicated in many physiological processes including abiotic stress tolerance (Gill and Tuteja, 2010). CICDH is known to be involved in the redox homeostasis of cells (Mhamdi et al., 2010). A final candidate was selected from a minor association peak on C2; Bo2g018010.1, whose orthologue in *A. thaliana* is described as Rho GTPase activation protein (RhoGAP : AT5G19390.2) with PH domain (PHGAP2), which may be involved in cell polarity and mitosis (Stockle et al., 2016; Hwang et al., 2008).

The CDS gene models for which GEMs were identified did not have annotations that suggested involvement in Cu concentration. These GEMs were: Bo2g057690.1, whose *A. thaliana* orthologue is an uncharacterised Reticulon family protein (AT1G68230.2); Bo1g039440.1, whose orthologue is described as a thylakoid luminal 17.9 kDa protein in *A. thaliana* (AT4G24930.1) and is the most highly associated GEM; Bo1g139020.1, whose *A. thaliana* orthologue is a Eukaryotic aspartyl protease family protein (AT3G12700.1); Cab040172.1, whose orthologue is described on TAIR (Lamesch et al., 2012) as a “Pyridoxal phosphate (PLP)-dependent transferases superfamily protein” (AT5G66950.1). It was decided therefore that only the top GEM (AT4G24930) would be tested for disruption within the leaf ionome of *A. thaliana* insert lines. Interestingly, the second highest GEM from the AT outputs for Cu concentrations in leaf on the 383 diversity panel was Cab003267.1, whose orthologue in *A. thaliana* is the major floral integrator *SOC1* (AT2G45660). However, there were many instances where *SOC1* and *FLC* orthologues were the most highly associated GEMs in multiple leaf element concentration AT outputs. This did not always correspond to the same CDS gene model, e.g. for Cd concentration in leaf in the next section it is an orthologue of ‘*SOC1*’ on C4 that is amongst the most highly associated GEMs). As this became the focus of a follow up experiment it will be discussed as part of the next chapter (see 6).

5.2.2 Associative transcriptomic outputs, predictions and candidates: Cd concentration in leaf material

Of all the leaf AT analyses, Cd concentrations in leaf AT results were arguably the most successful. A number of association peaks was observed in SNP association analysis. Two association peaks on A3 and two minor association peaks on C9, one association peak on A6, A7, A8 and C8 (**Figure 5.2.2.a**) were detected. The GEM association analysis revealed very little at first but yielded a number of potential candidates when investigated in greater depth. Two predictive markers ($p < 0.05$) were found within the GEM AT analyses and one SNP in AT analyses (**Table 5.2.2.a**).

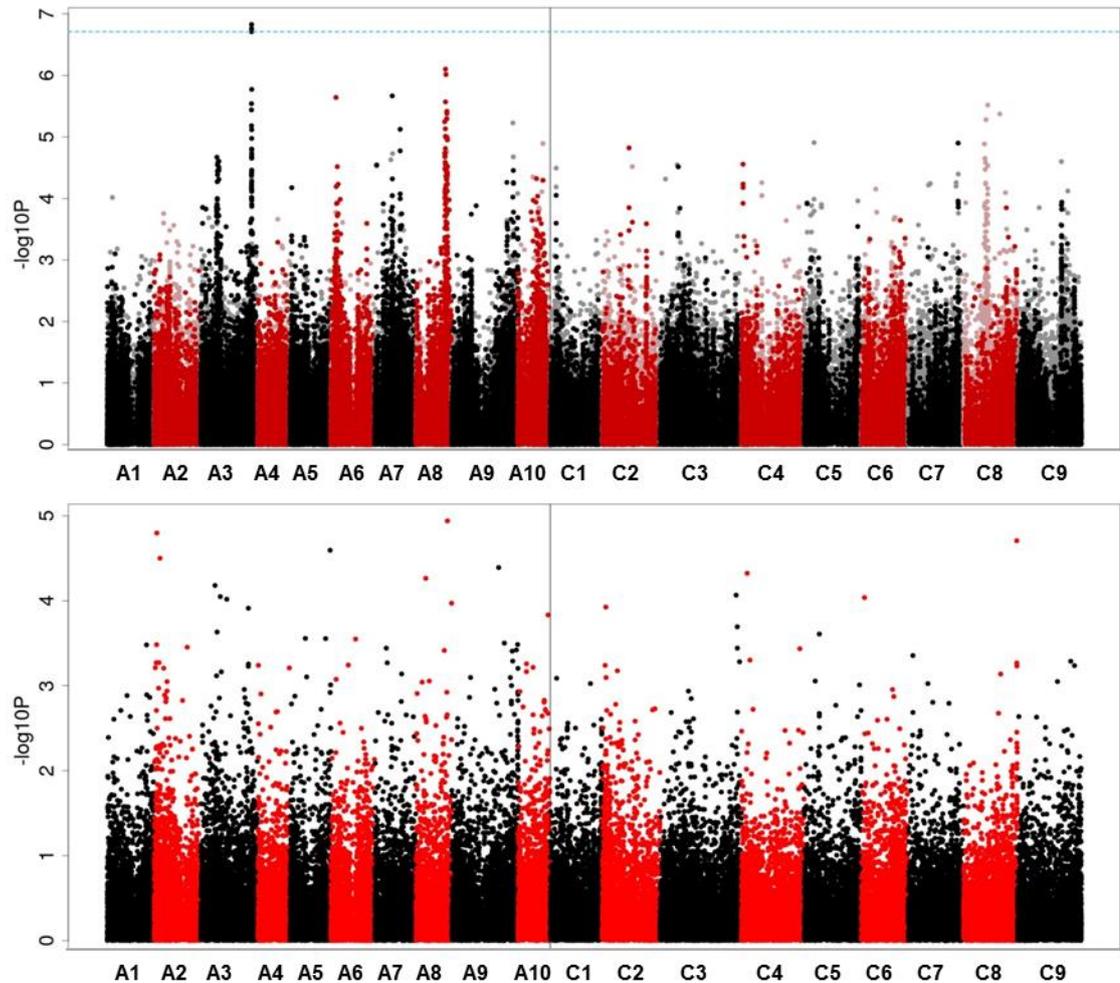


Figure 5.2.2.a Genome wide distribution of mapped markers associating with the Cd concentration in leaves (mg/kg DW) of all 383 accessions.

The average concentration of Cd in leaves was calculated from 5 separate plants from ICP-MS analysis for each of the 383 accessions. SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and leaf Cd concentration as the response variable. $-\log_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

Table 5.2.2.a Predictive capability of markers from AT analysis of Cd concentration in leaves.

For assessing the predictive capability of markers the highest SNP markers from discernible association peaks and the most highly associated GEMs were analysed. The marker type is given as either SNP or GEM, alongside their name and position, followed by their $-\log_{10}P$ value from the 274 AT analysis.

Finally, the correlation coefficient (R), significance (p) and sample size (n) are given for the predictions made on the 109 diversity panel. Markers that were significantly predictive are highlighted in bold.

Marker type	Marker	Position	AT 274 $-\log_{10}P$	R	p	n
GEM	Cab031859.1	A08_021311613_021314130	6.00	0.287	0.003	109
GEM	Bo4g024850.1	C04_004021498_004023941	6.00	0.271	0.004	109
GEM	Bo8g117680.1	C08_041396264_041397850	5.94	0.088	0.364	109
GEM	Bo6rg016920.1	C06_003381602_003382846	5.92	0.187	0.052	109
SNP	Cab011211.1:2640:G	A03_030093862_030089292	4.33	0.491	<0.001	109
SNP	Bo9g171620.1:85:T	C09_050592304_050593467	3.81	0.052	0.617	109
SNP	Bo9g135680.1:2288:C	C09_041790860_041794352	3.17	0.144	0.136	95

Within the predictive SNP peak on A3 (the second peak, passing the Bonferroni corrected significance and FDR thresholds) two well-known and well characterised Cd transporters were identified: Cab011213.1 and Cab011209.2, whose orthologues in *A. thaliana* are *heavy metal ATPase 2 (HMA2: AT4G30110.1)* and *heavy metal ATPase 3 (HMA3: AT4G30120.1)* respectively. HMA2 is a known Zn/Cd transporting ATPase important for the xylem loading of Zn (Wong et al., 2009), while it has been shown that HMA3 is important for the vacuolar storage of Cd (Morel et al., 2009). Given that these candidates were already well characterised in *A. thaliana* they were investigated no further in this research. However, within the first peak on A3 a CDS gene model, Cab002809.1, has an orthologous gene in *A. thaliana* described as an uncharacterised heavy metal transport/detoxification superfamily protein (AT2G36950.1), which was identified as a heavy metal associated isoprenylated plant protein (De Abreu-Neto et al., 2013). This had not previously been investigated with *A. thaliana* T-DNA insert mutants and was therefore taken forward for further analysis.

Within the remaining association regions in the SNPs, a number of previously identified genes with a role in the Cd response of the plant were identified. On the A7 association peak, a CDS gene model called Cab009060.1 has an orthologue in *A. thaliana* that is called *NPF6.2 (AT2G26690.1)*. It was previously found that this nitrate transporter displayed strong upregulation under Cd stress and a nitrate dependent Cd sensitive phenotype in an insertional mutant line (Li et al., 2010). Another previously recognised candidate was found on the A8 association peak; *Plant Cadmium Resistance 1 (PCR1: AT1G14880.1)* was within the correct localisation in *A. thaliana* despite not being present within the pan-transcriptome. When *PCR1*

was overexpressed in *A. thaliana*, plants had increased tolerance to Cd, while they displayed increased sensitivity when it was disrupted (Song, 2004). Furthermore, within the SNP peak on C9, two CDS gene models (Bo9g134860.1/Bo9g134880.1) were found whose orthologue in *A. thaliana* is described as the mitochondrial ABC transporter, *ATM3* (AT5G58270.1). As with *PCR1*, plants which overexpress *ATM3* have enhanced Cd tolerance, while those without it show increased sensitivity. Unlike *PCR1*, it has been suggested that this occurs due to the export of glutamine synthetase conjugated Cd (Kim et al., 2006a). Finally, the most significantly associated GEM marker on A2 is BnaA02g00090D whose orthologue is described as *Pyrophosphorylase 6* (*PPA6* : AT5G09650.1) and is known to be upregulated in *A. thaliana* in response to Cd treatment (Sarry et al., 2006).

Of the predictive GEMs, the first was the most highly associated GEM in AT results for leaf Cd concentration; Cab031859.1 on A8, whose orthologue in *A. thaliana* is described as a *methionine aminopeptidase* (AT1G13270). It is predicted to have a metal ion binding domain and metalloexopeptidase activity (TAIR, 2015). Other than this it is not well characterised within the literature. It would make for an excellent candidate to be assessed with *A. thaliana* insertional mutants considering its high association in AT outputs and the number of Cd specific candidates which had been observed alongside it in these analyses. However, due to time constraints a more in-depth analysis was not possible. The second predictive marker corresponded to a very well characterised gene; the major floral integrator *SOC1* (AT2G45660.1 orthologue of Bo4g024850.1) mentioned as part of the Cu concentration in leaf AT analysis. This will also be discussed in greater detail in the following chapter.

5.2.3 Associative transcriptomic outputs, predictions and candidates: Mn concentration in leaf material

In complete contrast to the Cd concentration in leaf AT analyses, the AT outputs for Mn concentrations in leaf material was the least successful (**Figure 5.2.3.a**). When investigated in detail, there were no clear associations within either the SNP or GEM AT analyses.

Interestingly, the SNP association outputs pass the Bonferroni corrected significance threshold and the 5% false discovery threshold; demonstrating that it is important to look for good association peaks along with high p values in GWAS analyses. Despite a lack clear of association peaks in AT analysis, three GEMs and one SNP were found to be predictive for leaf Mn concentration (**Table 5.2.3.a**).

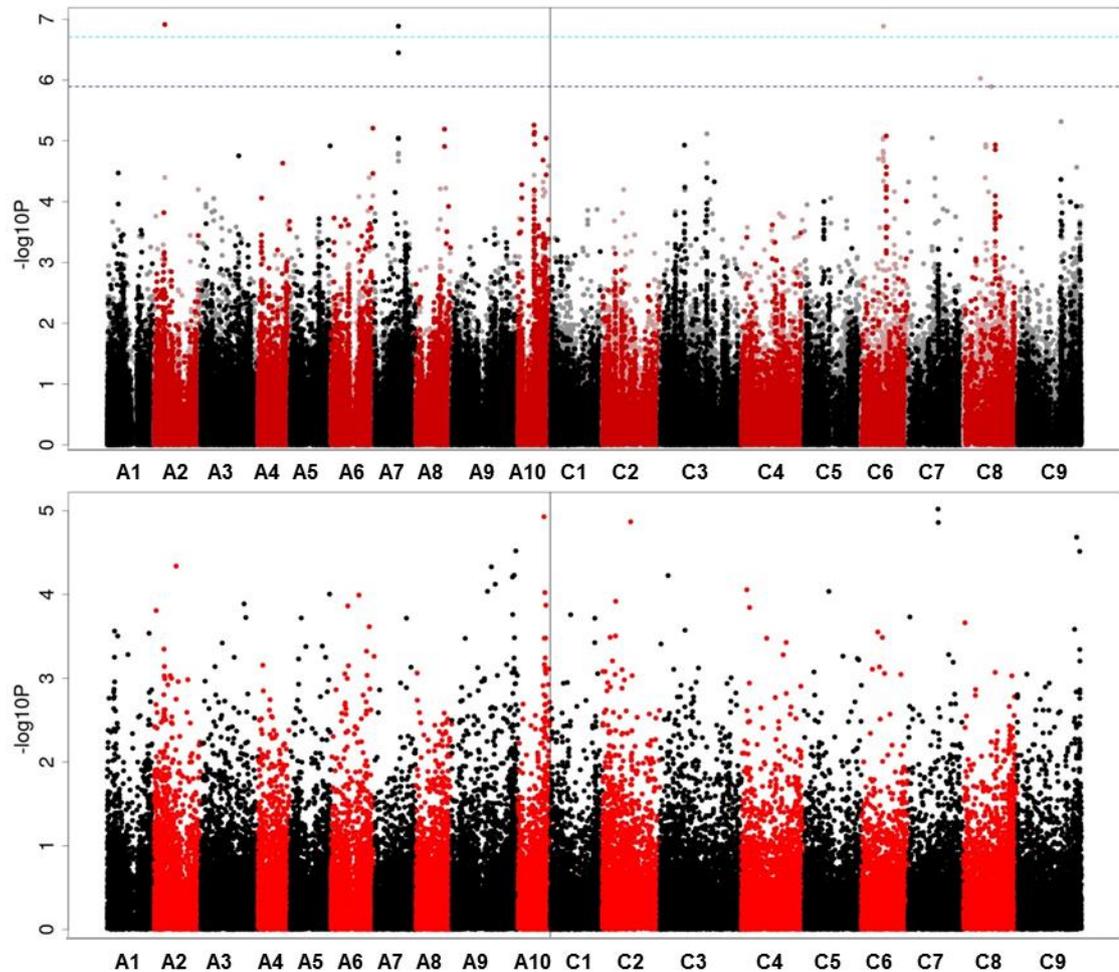


Figure 5.2.3.a Genome wide distribution of mapped markers associating with the Mn concentration in leaves (mg/kg DW) of all 383 accessions.

The average concentration of Mn in leaves was calculated from 5 separate plants from ICP-MS analysis for each of the 383 accessions. SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and leaf Mn concentration as the response variable. $-\text{Log}_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

Table 5.2.3.a Predictive capability of markers from AT analysis of Mn concentration in leaves.

For assessing the predictive capability of markers the highest SNP markers from discernible association peaks and the most highly associated GEMs were analysed. The marker type is given as either SNP or GEM, alongside their name and position, followed by their $-\log_{10}P$ value from the 274 AT analysis.

Finally, the correlation coefficient (R), significance (p) and sample size (n) are given for the predictions made on the 109 diversity panel. Markers that were significantly predictive are highlighted in bold.

Marker type	Marker	Position	AT 274 $-\log_{10}P$	R	p	n
GEM	BnaA09g49870D	A09_044670075_044672126	10.20	-0.004	0.971	109
GEM	Bo9g174920.1	C09_051742405_051744195	8.02	0.169	0.078	109
GEM	Cab006306.1	A06_019056334_019057817	7.52	0.179	0.063	109
GEM	Cab019794.2	A01_004090150_004088648	7.18	0.051	0.597	109
GEM	Bo2g023580.1	C02_006152073_006152889	6.91	0.105	0.277	109
GEM	Cab038054.1	A02_014437442_014439202	6.66	0.218	0.023	109
GEM	Bo2g083700.1	C02_023285173_023286819	6.65	0.260	0.006	109
GEM	Cab032604.2	A02_005555254_005558452	6.64	0.076	0.433	109
GEM	Bo9g181240.1	C09_053471789_053472822	6.45	0.275	0.004	109
GEM	Cab003267.1	A03_011912889_011915275	6.29	0.169	0.079	109
GEM	Bo3g021940.1	C03_007477013_007478491	6.25	0.033	0.735	109
GEM	Cab007391.2	A10_017777444_017787543	6.22	0.117	0.227	109
SNP	Bo5g039310.1:171:A	C05_012732381_012736308	5.54	-0.070	0.503	94
SNP	Bo5g039310.1:1230:T	C05_012732381_012736308	5.48	-0.092	0.347	106
SNP	Bo5g039310.1:749:G	C05_012732381_012736308	5.05	-0.067	0.493	106
SNP	Bo2g028890.1:408:A	C02_008692620_008693924	4.74	0.242	0.012	108
SNP	Cab040440.1:232:T	A07_000215898_000212523	4.63	0.004	0.972	103
SNP	BnaC03g44950D:228:C	C03_031706050_031707646	4.28	0.177	0.098	88
SNP	Bo2g028790.1:960:A	C02_008644153_008648863	3.94	0.191	0.065	94

All of the GEMs that were predictive for Mn concentration in leaves were unannotated. When a BLAST (Altschul et al., 1990) was performed to compare these GEMs to all other CDS gene models within the pan-transcriptome and plant species in the NCBI database (see 2.2.2), no CDS gene models annotated with *A. thaliana* orthologue information were found. Perhaps the reason why Mn leaf concentration assessment with AT has proven ineffective is because the mechanisms involved in Mn concentration in *B. napus* vary significantly from those of *A. thaliana*. However, this seems highly unlikely as previous work has linked an orthologue of the cation efflux facilitator transporter *MTP8* to Mn tolerance in *B. napus* QTL analysis (Raman et al., 2017). As observed in Cu and Cd leaf concentration AT analysis, *SOC1* and *FLC* were amongst the most highly associated GEMs. The predictive SNP marker Bo2g028890.1:408:A, corresponded to an orthologue in *A. thaliana* of a transport protein particle (TRAPP) component (AT5G58030.1). No evidence of a role for this TRAPP candidate in leaf Mn

concentration could be found within the literature (other than the general secretion of proteins containing Mn as a cofactor) and it was investigated no further.

5.2.4 Associative transcriptomic outputs, predictions and candidates: Zn concentration in leaf material

The SNP outputs for Zn concentration in leaves did not yield any association peaks, however the GEM association analysis seemed to show a potential deletion on A9 (**Figure 5.2.4.a**). Deletions within GEM association analysis are often observed as association peaks since all the genes within the region have their expression disrupted by the deletion. As previously observed in the Cu leaf concentration GEM association analysis, the Zn leaf concentration SNP Q-Q plots suggest that the model has been too stringent with the observed p values falling below those that were expected (**Figure 5.2.4.b**). Accordingly, no SNP markers were found that were predictive for Zn concentration within the leaves, whilst 5 GEMs displayed predictive capacity ($p < 0.05$, **Table 5.2.4.a**).

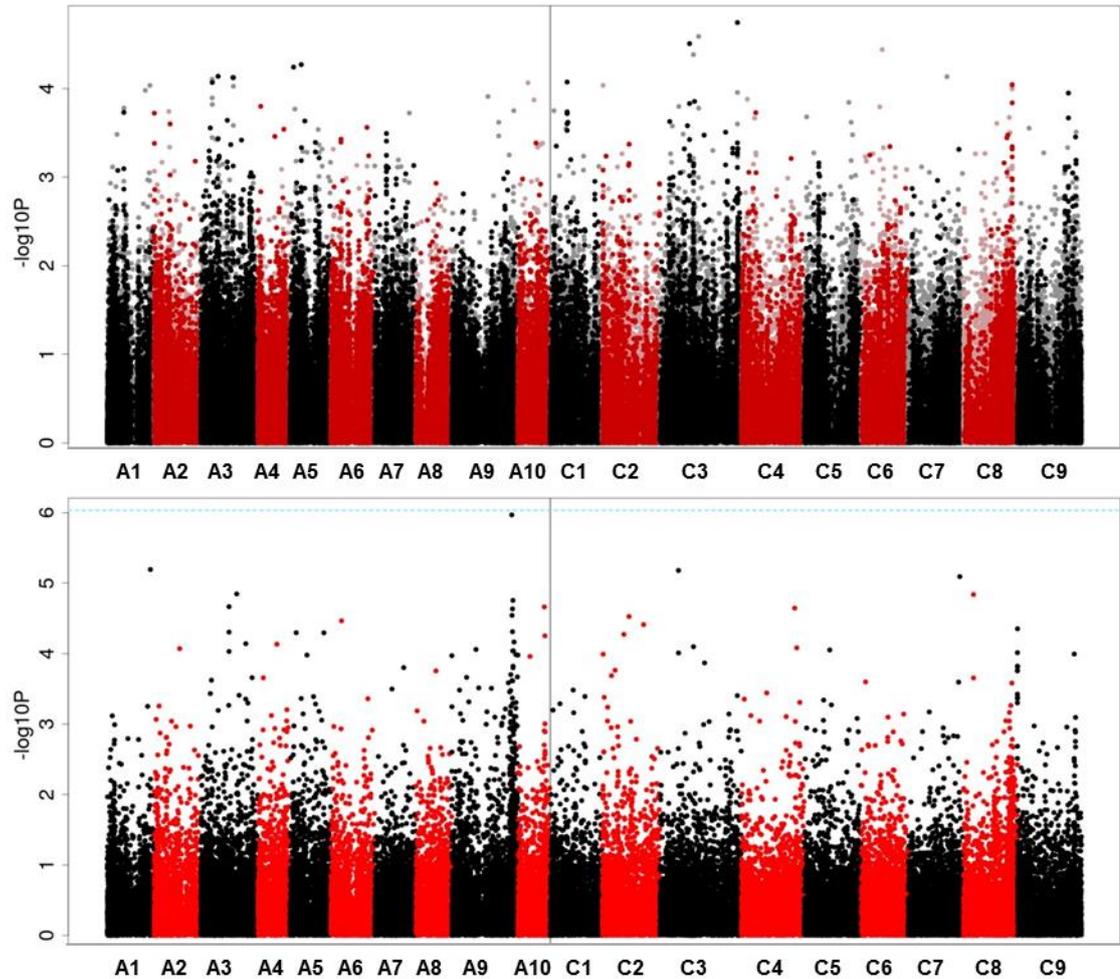


Figure 5.2.4.a Genome wide distribution of mapped markers associating with the Zn concentration in leaves (mg/kg DW) of all 383 accessions.

The average concentration of Zn in leaves was calculated from 5 separate plants from ICP-MS analysis for each of the 383 accessions. SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and leaf Zn concentration as the response variable. $-\text{Log}_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

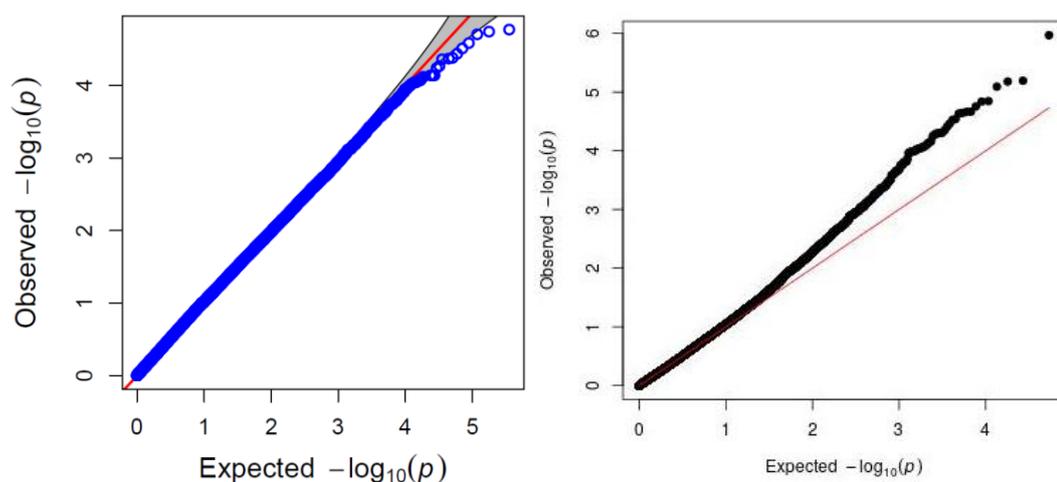


Figure 5.2.4.b Quantile-quantile plot of observed $-\log_{10}P$ values from AT SNP analysis (left) and AT GEM analysis (right) for Zn in leaf against expected $-\log_{10}P$ values.

The red line indicates a theoretical perfect fit of the expected $-\log_{10}P$ values, while the grey area indicates the 95% confidence interval under the null hypothesis of no association between the Zn concentration in the leaves and the SNPs/GEMs.

Table 5.2.4.a Predictive capability of markers from AT analysis of Zn concentration in leaves.

For assessing the predictive capability of markers the highest SNP markers from discernible association peaks and the most highly associated GEMs were analysed. The marker type is given as either SNP or GEM, alongside their name and position, followed by their $-\log_{10}P$ value from the 274 AT analysis. Finally, the correlation coefficient (R), significance (p) and sample size (n) are given for the predictions made on the 109 diversity panel. Markers that were significantly predictive are highlighted in bold.

Marker type	Marker	Position	AT 274 $-\log_{10}P$	R	p	n
GEM	Cab013955.2	A09_042635665_042636935	7.97	0.205	0.033	109
GEM	Bo2g056290.1	C02_016642943_016645398	6.76	0.150	0.120	109
GEM	Bo3g027950.1	C03_010615262_010619636	6.51	0.261	0.006	109
GEM	Cab007716.1	A10_017146559_017143933	6.39	0.179	0.063	109
GEM	Bo7g116050.1	C07_046288528_046290471	6.25	0.256	0.007	109
GEM	Bo8g032470.1	C08_010730007_010730725	6.20	0.214	0.025	109
GEM	Bo2g079140.1	C02_021942569_021944600	6.17	0.214	0.025	109
GEM	Cab013586.1	A09_040779359_040779808	6.15	0.085	0.377	109
GEM	Bo5g004680.1	C05_001368506_001369971	6.11	-0.008	0.935	109
GEM	Bo9g002180.1	C09_000123371_000125749	6.10	0.123	0.204	109
GEM	Bo8g108310.1	C08_038595797_038597360	6.07	0.108	0.263	109
GEM	Cab046623.1	A10_008626408_008628450	5.92	0.174	0.071	109
SNP	Cab033019.2:475:A	A07_010108135_010105975	3.86	0.109	0.309	89

Unfortunately, the GEM at the top of the GEM association peak on A9 was an unannotated marker. It most closely blasts to Cab031842.1, whose orthologue in *A. thaliana* AT1G13880.1 is an ELM2 domain-containing protein (S' 745, E: 0), which is a member of the Myb transcription factors (Marmioli et al., 2014). However, no direct link to Zn concentration in leaves could be established from the literature and this was pursued no further. There were no other candidates identified in this association region with a link to Zn leaf concentration. The top GEM on C8 however was Bo8g032470.1, orthologue of AT1G31260.1 which is described as a *zinc transporter precursor 10 (ZIP10)*. It has been described as having Zn²⁺ transport capacity (TAIR, 2015), and is grouped with a number of known Zn transporters in phylogenetic analysis of cation transport families within *A. thaliana* (*IRT1&2* and *ZIP 7 & 8*) (Maser, 2001). As of yet it is not thought to have been characterised within *A. thaliana* insert lines but due to time constraints was not assessed further as part of this study.

5.2.5 Associative Transcriptomic outputs, predictions and candidates: Mo concentration in leaf material

Given the distinct similarities observed between Mo and S concentrations following seed AT analysis (particularly in relation to the GLS) it was surprising that no relationship between the S and Mo AT outputs was observed in the leaves (cf. 5.2.6). In fact, when leaf S and Mo concentrations were correlated a significant positive correlation was observed ($R: 0.514$, $p < 0.001$, $n: 385$). This is in direct contrast to the relationship observed within the seed which has a significant negative correlation between the two ($R: -0.466$, $p < 0.001$, $n: 380$). A distinct association peak was observed on C1 in the SNP association analysis and a few less well defined SNP association peaks on A7, A10, C3 and C9 (**Figure 5.2.5.a**). The GEM association analysis showed no clear associations but did provide some predictive markers ($p < 0.05$, unlike the SNPs, **Table 5.2.5.a**).

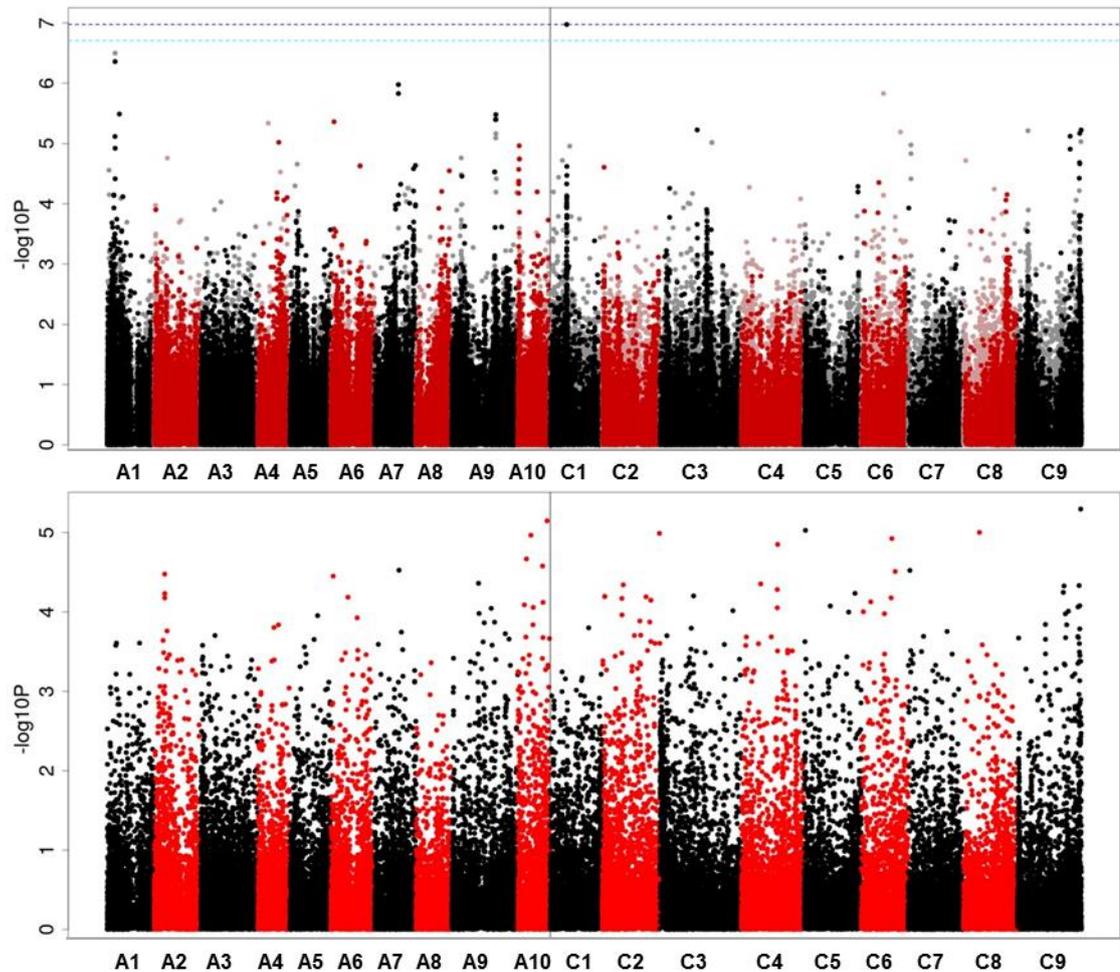


Figure 5.2.5.a Genome wide distribution of mapped markers associating with the Mo concentration in leaves (mg/kg DW) of all 383 accessions.

The average concentration of Mo in leaves was calculated from 5 separate plants from ICP-MS analysis for each of the 383 accessions. SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and leaf Mo concentration as the response variable. $-\text{Log}_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

Table 5.2.5.a Predictive capability of markers from AT analysis of Mo concentration in leaves.

For assessing the predictive capability of markers the highest SNP markers from discernible association peaks and the most highly associated GEMs were analysed. The marker type is given as either SNP or GEM, alongside their name and position, followed by their $-\log_{10}P$ value from the 274 AT analysis.

Finally, the correlation coefficient (R), significance (p) and sample size (n) are given for the predictions made on the 109 diversity panel. Markers that were significantly predictive are highlighted in bold.

Marker type	Marker	Position	AT 274 $-\log_{10}P$	R	p	n
GEM	Bo9g183340.1	C09_054156036_054157450	20.23	0.413	<0.001	109
GEM	Cab004528.1	A03_018750262_018751101	16.18	0.229	0.017	109

The association peak within the SNP AT output on C1 was very narrow (containing only ~69 CDS gene models), passing the Bonferroni and 5% false discovery thresholds without yielding any known Mo specific candidates. A large number of potential candidates with more general roles were observed (e.g. auxin response factors, cytokinin response factors, multiple transcription factors). An example of these more general candidates would be Bo1g037120.1 whose orthologue is AT4G23710.1, a vacuolar ATP synthase subunit G2. MoO_4^{2-} is known to be stored within the vacuole of Brassica species (Hale et al., 2001) and since V-type ATPases are involved in acidifying the vacuole (Finbow and Harrison, 1997) this would affect the availability of Mo to the rest of the leaf. However, the most likely of these general candidates is Bo1g037430.1, orthologue of *NLP7* (AT4G24020) which is involved in modulating nitrate sensing and metabolism (Castaings et al., 2009). Mo plays a central role in N metabolism as it is the Mo-molybdopterin (Mo-MPT) domain of Nitrate Reductase that is the nitrate reducing active site (Campbell, 1999). Interestingly when *NLP7* was knocked out in *A. thaliana* there was a three-fold reduction in nitrate reductase enzymatic activity (with no reduction in total N) (Castaings et al., 2009). It would make sense for a transcription factor involved in nitrate sensing and metabolism to be involved in regulating leaf Mo concentration due to its key role in nitrate assimilation. However, knocking out *NLP7* would not necessarily provide evidence of a direct role in Mo concentration within leaf, as when mutants were previously tested a wide range of N starvation symptoms were displayed which would affect multiple nutrients indirectly (Castaings et al., 2009). Rather it would be better to assess if the expression of any of the genes known to be involved in Mo transport/assimilation were disrupted in the *NLP7* knockout, thus providing a link between N and Mo regulation. Furthermore, the SNP association peak on A7 was shown to contain Cab020822.1 (orthologue of *Nitrate Reductase*: AT1G77760.1), while the orthologue of an ammonium transporter (*AMT1.5*: AT3G24290) was

found close to the association peak on C3 (Bo3g083120.1), adding credence to the assertion that NLP7 may be involved in the cross-regulation of N and Mo concentrations in the leaf.

The association peak on A10 was also very narrow and contained no obvious Mo candidates. Once again there was a number of general candidates which could affect Mo within leaves, for example Cab033364.1/ Cab029905.1. Its orthologue in *A. thaliana* is *NAC5* (AT1G02250.1) and is involved in xylem formation and would thus impact leaf Mo (Zhao et al., 2016). However, it would likely disrupt more than just Mo concentrations in leaves and was therefore not considered likely to be responsible for a Mo specific association peak. Another general candidate was Cab029913.1 (orthologue of *USUALLY MULTIPLE ACIDS MOVE IN AND OUT TRANSPORTERS 28/ UMAMIT28*: AT1G01070.1). This candidate had previously been tested with *A. thaliana* insertional mutants and displayed disrupted Mo within the leaves (PiiMs, (Baxter et al., 2007a)), however it was tested to confirm these results (as it seemed an unlikely candidate) and to expand the range of elements it was tested for (e.g. if there was any link to S concentrations given the positive association between leaf Mo and S).

Of the predictive GEMs one (Cab004528.1) was unannotated, whilst Bo9g183340.1 has an orthologue in *A. thaliana*, AT5G02820.1, described as “Spo11/DNA topoisomerase VI, subunit A protein”. After further investigation it was found to be *BRASSINOSTEROID INSENSITIVE 5 (BIN5)/ ROOT HAIRLESS 2 (RH2)*. It is involved in endoreduplication and when disrupted displays wide ranging phenotypes (e.g. stunting and hairless roots (Hartung et al., 2002; Sugimoto-Shirasu et al., 2002; Yin et al., 2002)). It would therefore not make a good candidate for analysis with insertional mutants; plants would likely not grow very well and it would be impossible to distinguish direct (i.e. a specific regulatory role in Mo concentration in leaves) or indirect effects (i.e. Mo disruption as an effect of its role in root hair development) on Mo concentration in leaves.

5.2.6 Associative Transcriptomic outputs, predictions and candidates: S concentration in leaf material

Once again, neither the SNP nor GEM AT analysis for S concentrations in leaf passed either the Bonferroni corrected significance or 5% FDR thresholds (**Figure 5.2.6.a**). Three association peaks were investigated within the SNP AT analysis; A6, A9 and C9. Once again, GEM AT analysis appeared to show close association to flowering time candidates (further discussed in 6). Two SNPs and all 10 GEMs analysed were predictive ($p < 0.05$) for S concentrations in leaf material (**Table 5.2.6.a**).

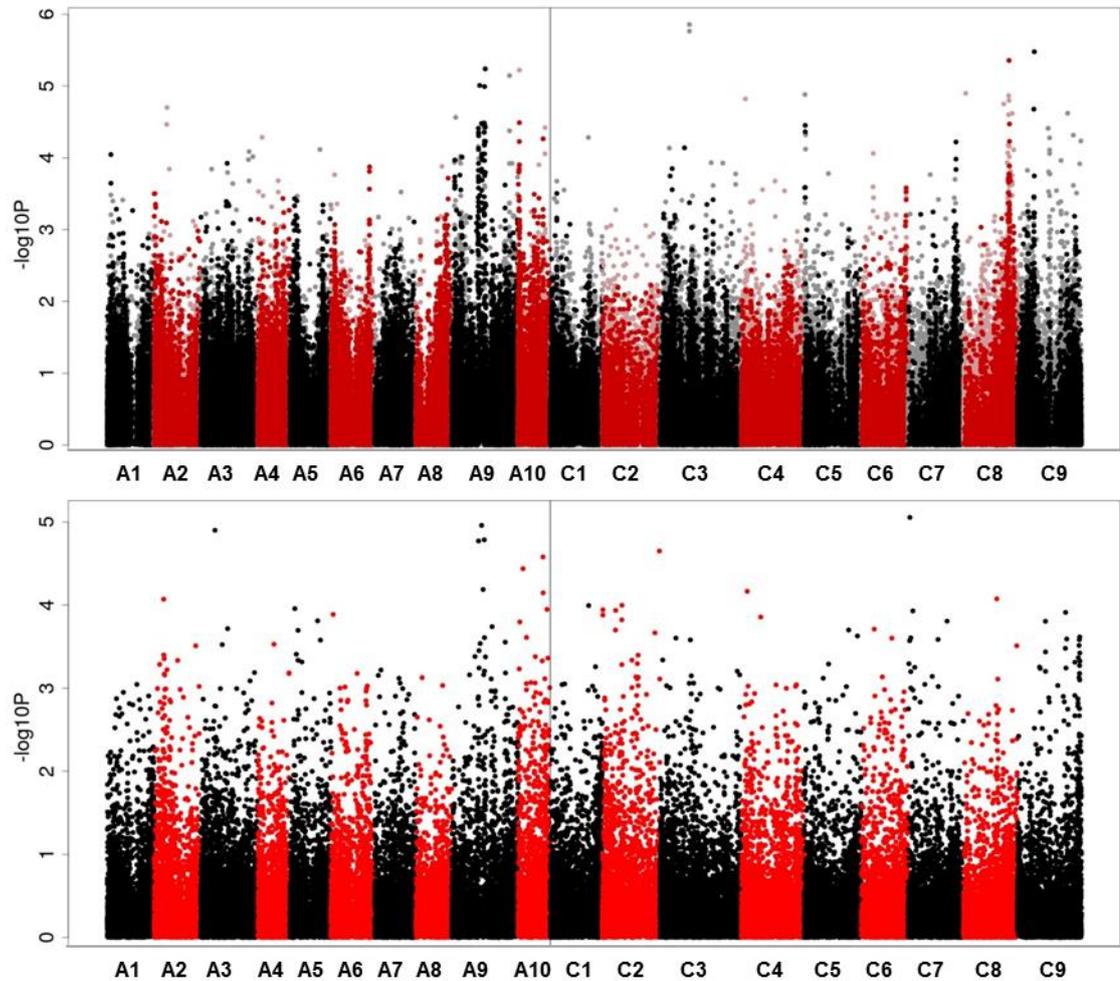


Figure 5.2.6.a Genome wide distribution of mapped markers associating with the S concentration in leaves (mg/kg DW) of all 383 accessions.

The average concentration of S in leaves was calculated from 5 separate plants from ICP-MS analysis for each of the 383 accessions. SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and leaf S concentration as the response variable. $-\text{Log}_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

Table 5.2.6.a Predictive capability of markers from AT analysis of S concentration in leaves.

For assessing the predictive capability of markers the highest SNP markers from discernible association peaks and the most highly associated GEMs were analysed. The marker type is given as either SNP or GEM, alongside their name and position, followed by their $-\log_{10}P$ value from the 274 AT analysis. Finally, the correlation coefficient (R), significance (p) and sample size (n) are given for the predictions made on the 109 diversity panel. Markers that were significantly predictive are highlighted in bold.

Marker type	Marker	Position	AT 274 $-\log_{10}P$	R	p	n
GEM	Cab002472.4	A03_007577717_007583343	20.32	0.402	<0.001	109
GEM	Cab045257.1	A09_023545564_023548145	19.85	0.413	<0.001	109
GEM	Cab007824.2	A10_016640393_016638608	18.64	0.415	<0.001	109
GEM	Cab005591.1	A09_021405072_021407365	17.91	0.496	<0.001	109
GEM	Bo4g024850.1	C04_004021498_004023941	17.61	0.427	<0.001	109
GEM	Bo7g005670.1	C07_001731985_001732598	17.22	0.550	<0.001	109
GEM	Cab003830.1	A03_015086160_015088422	17.20	0.378	<0.001	109
GEM	Cab005327.2	A09_018954048_018956064	17.18	0.494	<0.001	109
GEM	Cab017845.1	A09_022634254_022640222	17.05	0.415	<0.001	109
GEM	Cab025356.1	A05_002862036_002865272	15.96	0.416	<0.001	109
SNP	Cab045287.1:1751:A	A09_023798471_023802483	4.10	0.512	<0.001	108
SNP	BnaA06g34040D:219:T	A06_026508326_026506702	3.99	0.017	0.860	108
SNP	Cab039046.2:195:A	A08_017541264_017543216	3.98	-0.029	0.770	107
SNP	Cab035444.2:1693:A	A06_026149390_026151787	3.75	-0.088	0.365	109
SNP	Cab039021.1:766:C	A08_017418744_017420069	3.72	0.136	0.195	93
SNP	Cab039028.1:3366:C	A08_017443847_017452005	3.60	0.151	0.165	86
SNP	Cab005584.1:373:G	A09_021344941_021346222	3.01	0.550	<0.001	81

The SNP association peak on A6 yielded no S specific candidates. Instead, a number of other general candidates were found; including phosphate and nitrate regulatory candidates and root hair developmental candidates. One of the interesting candidates was Cab035423.3, whose orthologue in *A. thaliana* is the *EARLY FLOWERING MYB PROTEIN (EFM: AT2G03500)*, which is responsible for the flowering response to environmental cues. This is intriguing when compared to the GEM AT analysis, which seems to be highlighting many of the famous floral integrators (discussed further in 6). No candidates were taken forward from this region. Again the SNP association peak on A9 gave no S specific candidates and appeared to cross the centromere. This is in contrast to most of the other association peaks which occur towards the end of the chromosome where recombination is more likely to occur. The association peak on A9 covered an extremely large region of ~900 CDS models, making it incredibly difficult to analyse (perhaps indicating an association as the result of a large chromosomal rearrangement). Furthermore, these regions are very hard to analyse as it is likely that candidates related to the trait will be within the association region by chance. Two genes were

found which could play a role within S concentration in the leaves; Cab005432.1/ Cab005436.1 are orthologues of a Mo cofactor sulfurylase protein (AT1G30910.1) in *A. thaliana* and Cab045260.2 is an orthologue of *TAR3* (AT1G34040.1: described in TAIR as “Pyridoxal phosphate (PLP)-dependent transferases superfamily protein”) which has carbon-sulfur lyase activity. The C9 association peak was relatively narrow, and the most highly associated SNP Bo9g037180.1 has an orthologue in *A. thaliana* involved in GSL metabolism (*FMO GS-OX2*: AT1G62540.1). Specifically it is known to be involved in the conversion of methylthioalkyl glucosinolates to methylsulfinylalkyl glucosinolates (i.e. S oxygenation) (Hansen, Kliebenstein and Halkier, 2007). The predictive SNPs corresponded to: Cab045287.1, orthologue of AT1G33680.1 which is a KH domain containing protein, whilst Cab005584.1 was unannotated. In all, no S specific candidates were found from leaf S concentration SNP AT analysis which would be useful to test with an *A. thaliana* insertional mutant.

5.2.7 Summary of Associative Transcriptomic outputs, predictions and candidates

Very few candidates were taken forward from leaf AT analysis of individual element concentrations (**Table 5.2.7.a**). This was a result of the majority of the candidates already being previously tested in *A. thaliana* and/or due to a lack of being able to narrow down the candidates in broad association regions. The GEM data for many of the elements revealed a link to flowering time; therefore these candidates were investigated in more detail as part of the next chapter (see 6).

Table 5.2.7.a a list of the candidate genes taken forward for further study from leaf element concentration AT results.

Detailed are the original AT trait analysis the candidate was found for, its marker within the pan-transcriptome, AGI code, description in *A. thaliana*, line ordered from NASC and other potential element interactions are listed. Cu and Cd lines (*) were analysed as part of an undergraduate project.

Trait AT	Marker	AGI	Description	NASC ID	Interaction?
Cd leaf*	Cab002809.1	AT2G36950	Heavy metal transport/detoxification superfamily protein	SALK_151770C (Cd leaf HMA 1)	Cu
				SALK_069207C (Cd leaf HMA 2)	Cu
Cu leaf*	Bo2g052580.1	AT1G65840	polyamine oxidase 4	SALK_118752C (PAO4 1)	?
				SALK_062544C (PAO4 2)	?
Cu leaf*	Bo2g018010.1	AT5G19390	Rho GTPase activation protein (RhoGAP) with PH domain	SALK_088689C (RhoGAP1)	?
				SALK_083351C (RhoGAP2)	?
Cu leaf*	Bo2g052640.1	AT1G65930	cytosolic NADP+-dependent isocitrate dehydrogenase (CICDH)	SALK_056247C (CICDH 1)	?
				SALK_009094 (CICDH 2)	?
Cu leaf*	Bo1g039440.1	AT4G24930	thylakoid lumenal 17.9 kDa protein, chloroplast	SALK_099907C (Cu GEM)	?
Mo leaf	Cab029913.1	AT1G01070	nodulin MtN21 /EamA-like transporter family protein (UMAMIT28)	SALK_099741C (UMAMIT28 1)	?
				SALK_147481C (UMAMIT28 2)	?

The lack of candidates taken forward is to some extent a reflection of the difficulties of plant nutrient research. The ionome within plants is integral to plant growth, development and metabolism, so under conditions of nutrient sufficiency there are many different pathways and feedback mechanisms ensuring that the plant maintains a safe working range for all elements within the ionome. The cross-talk and interactions of these elements could be wide ranging, making the selection of specific candidates extremely difficult. In addition, plants are very good at coping with variable amounts of elements within their tissues and the broader environment (i.e. there is a range of acceptable concentrations). Therefore finding enough variation within

the panel that is specific to differences under nutrient sufficiency which could influence nutrient use efficiency is challenging. This is exemplified by the AT outputs of Cd concentrations in leaf. As a non-essential element and toxic 'heavy metal', there is likely to be strong selective pressure to minimise the biological effects of this element within tissues under all conditions (through specific detoxification pathways, such as sequestration or chelation). Selections against Cd within tissues may explain the strong association peaks observed within the Cd AT SNP outputs. The other (essential) elements analysed as part of this study are regulated in many different ways and at multiple levels of plant organisation and may therefore show weaker associations to specific mechanisms. Seed AT analyses may have given clear AT outputs comparatively as *B. napus* is predominantly grown for seeds. Therefore the seeds may have been under greater selective pressure during breeding (e.g. 4) leading to clearer associations.

5.3 Leaf candidate gene analysis

As discussed previously, very few leaf candidates were taken forward from leaf AT analysis due to many candidates already being well characterised previously or a lack of obvious candidates in broad association peaks (candidates listed in 5.2.7). As with seed analyses, the concentrations of some elements within the leaf ionome are known to correlate significantly (**Figure 5.2.7.a**), which highlights the interdependent nature of elements within the ionome. However, unlike the seeds, very few significant negative associations appear to be present within the leaf. In fact, very few correlations were observed between the seed and leaf datasets (reflected in correlation, ratio and discriminant analyses performed as part of Thomas et al., (2016)). The observation that the top GEMs in many of the leaf AT outputs were flowering time related emphasises the interactive nature of all elements within the ionome (as discussed as part of the next chapter, 6).

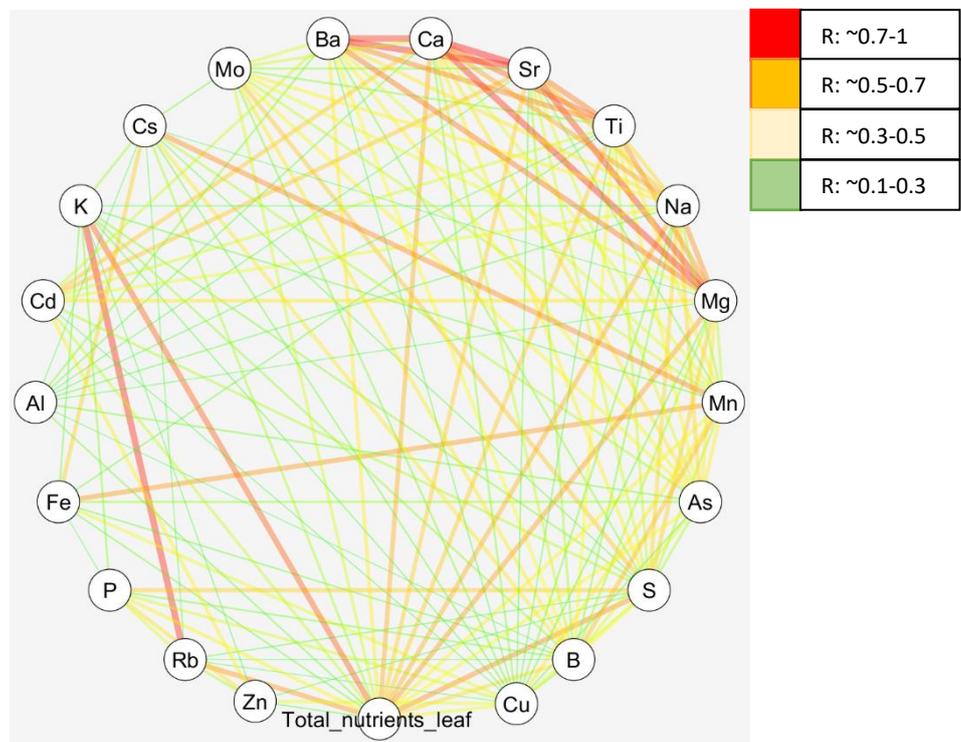


Figure 5.2.7.a The leaf 'interactome'.

Significant correlations ($p < 0.001$) for all element concentrations within the leaf are displayed, with the R value representing the lines connecting the element concentrations (thicker and darker lines are more highly correlated, key on the right hand side). Element concentrations which show a significant positive correlation ($p < 0.001$) are displayed, very few element concentrations showed significant negative ($p < 0.05$) correlations (Cd vs Cs: R -0.404; Cs vs Zn, R -0.191; Ba vs Rb, R -0.191). The diagram was generated with Cytoscape 3.2.1, using the R correlation coefficient as edges.

In this section, the investigation of the candidates taken forward from leaf AT analysis will be discussed. The first part will focus on Cu and Cd leaf concentration candidates which were originally part of an undergraduate project performed under the supervision of the author. As with the seed, these candidates were chosen from AT analysis performed as part of this research and taken forward into further analysis by the undergraduate student, Mr Jun Hee Jung (who designed primers, grew the plants and analysed seed and stem/pod materials with ICP-MS under the author's supervision). However, all analysis performed within leaves was generated as part of the current research; only the stem and seed analyses of the leaf candidates were performed by Mr. Jung (which were re-diluted and re-analysed for this study). The next section will focus on the Mo leaf concentration candidate that was carried forward, which was only analysed within leaf tissues as there was no evidence for a role outside of this tissue.

5.3.1 Leaf candidate gene analysis: Cu and Cd leaf concentration

Cu and Cd leaf concentration candidates were analysed simultaneously (**Figure 5.3.1.a**). Of the candidates which were found within leaf Cu concentration AT analysis, none varied significantly in comparison to the wildtype control for Cu concentration in leaves (Cu ANOVA: F 1.08, df 4, 35, p 0.38). Considering these candidates were chosen due to their high association and multiple occurrence in the Cu in leaf concentration AT SNP analysis (due to a lack of obvious testable candidates in a strong association peak) this is perhaps unsurprising. *POA4* was never successfully genotyped and was not taken further. *cicdh-2* could not be tested as it grew poorly and all the plants which survived to be genotyped were found to be wild type. This could have happened because it was essential for Cu concentration in plants, however since *cicdh-1* displayed no phenotype it is thought likely that an insert exists in this line elsewhere, causing the growth defective phenotype. However, the percentage recovery of Cu from leaves in this analysis was low (<75%), so perhaps Cu concentrations would vary if the recovery was better (e.g. maybe the Cu was bound within material more recalcitrant to digestion). On the other hand, Cd had an acceptable percentage recovery (~93%) and did show some significant differences. Of all the candidates carried forward only *Cd leaf hma-2* showed any significant differences within the leaves (the other insert line ordered in this gene, *Cd leaf hma-1*, only produced wild type plants). *Cd leaf hma-2* was analysed for both Cu and Cd concentration. This showed no significant variation in leaf Cu concentration but displayed a significant increase in leaf Cd concentration (**Figure 5.3.1.a**, Cd ANOVA: F 42.71, df 4, 35, p <0.001). However, Cd was not the only element to show a significant difference between the wildtype control and *Cd leaf hma-2*. When t-tests were performed between the wildtype control and all other elements successfully analysed as part of ICP-MS analysis a range of element concentrations varied (**Figure 5.3.1.b**).

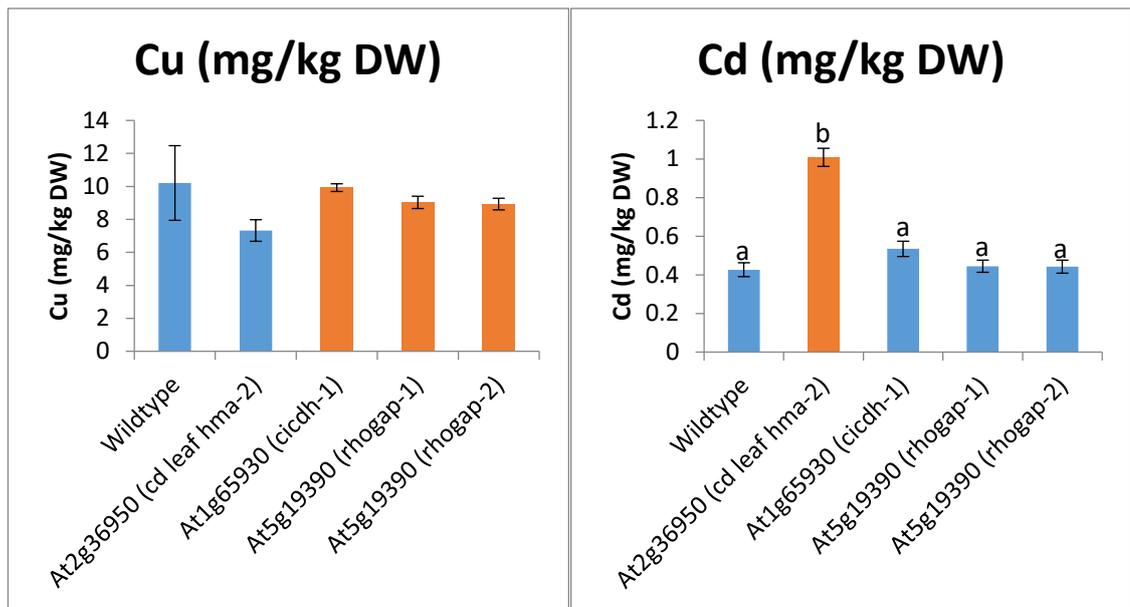


Figure 5.3.1.a Candidate gene analysis of Cu (left) and Cd (right) concentrations in the leaves of *A. thaliana* insert mutants as mg/kg DW of leaf tissue.

Wildtype: *A. thaliana* Col 0; At2g36950 = *hma-2* (SALK_069207C); At1g65930 = *cicdh-1* (SALK_056247C); At5g19390 = *rhogap-1* (SALK_088689C) and *rhogap-2* (SALK_083351C). The mean and standard error are shown for each line, n: 8 for all lines analysed, where n is the number of individual plants used in analyses. Highlighted in orange are the candidates which were picked out specifically from AT analysis of the element being portrayed. No significant differences were observed for Cu concentrations, but Cd concentrations displayed some differences (as indicated by differences in lettering). Cu ANOVA: F 1.08, df 4, 35, p 0.38; Cd ANOVA: F 42.71, df 4, 35, p <0.001.

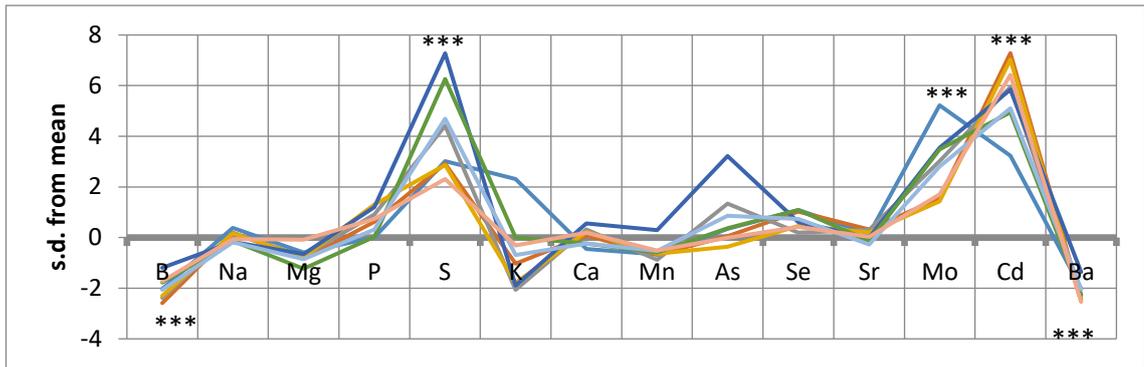


Figure 5.3.1.b Z-score graph detailing how leaf material (as a different coloured line) from the *Cd leaf hma-2* line (At2g36950/SALK_069207C) varied in comparison to the average wildtype control (n: 8 for both, where n is the number of individual plants sampled for analyses).

Only elements which had a recovery >85% and an average concentration greater than the LOD were analysed. The mean and standard deviation of the wildtype controls are used to calculate the number of standard deviations each mutant is away from the average of the wildtype control in each element, in accordance to methods previously outlined (Lahner et al., 2003) and in use by PiiMs (Baxter et al., 2007a). A t-test was used to determine whether the difference between wild type and mutant lines was significant for each element concentration ($p < 0.001$ ***), as detailed in Appendix 18.

Since *Cd hma-2* displayed a significant increase in leaf Cd concentration it was decided that the stem and leaf analyses would be re-diluted and analysed to assess whether the effect was localised to the leaves. No significant differences were observed for Cd concentrations in either stem or seed (**Figure 5.3.1.c**, Cd stem t-test: t 2.34, df approx. 3.2, $p = 0.096$; Cd seed t-test: t 1.95, df approx. 3.15, $p = 0.142$), although Cu concentration in seed was significantly reduced while the stem material showed no significant difference (**Figure 5.3.1.c**, Cu stem t-test: t 2.3, df 6, $p = 0.061$; Cu seed t-test: t 2.62, df 6, $p = 0.04$). The general pattern in *Cd hma-2* in stem and seeds showed a reduction in these elements and arguably the small sample size of $n = 4$ limited statistical power. Therefore, these tests should be repeated before any certain conclusion can be drawn about the activity of this gene. However, the pattern as it currently stands could imply that *Cd leaf HMA* increases the Cd concentration in leaf tissues at the expense of those in stem and seed. Interestingly, the concentration of other elements within the leaf may provide some explanation for this; a typical Cd toxicity response involving both cellular detoxification and sequestration. The concurrent increase in S concentration within the leaf tissues of these lines would suggest that toxicity pathways had been induced, generating the S rich chelating agents required to cope with increased Cd toxicity and thereby increasing the flux of S e.g. Yamaguchi et al., (2016). Once again, Mo concentration may be varying as a result of the chemical similarity between MoO_4^{2-} and SO_4^{2-} (Marschner, 1995b), while B concentration may vary as a consequence of the way the plants are sequestering the excess

Cd. To be specific, the only proven role for B in plants is as part of the cell wall within the rhamnogalacturonan II (RGII) polysaccharide domain of pectin (one of four pectin polysaccharide domains)(Kobayashi, Matoh and Azuma, 1996; Funakawa and Miwa, 2015). It is well known that the presence of Cd causes cell wall modifications; increasing the pectin content and production of low-methylesterified pectin (which can bind divalent cations, including Cd)(Loix et al., 2017). It is possible that the increased Cd concentration resulted in a greater requirement for B in the cell walls of roots (through increased pectin), reducing B available to the leaf. Further evidence may be seen in the reduction in Ba concentration within the leaf materials, another divalent cation *in planta*, which is also perhaps being trapped as a consequence of cell wall modifications within the roots. The other divalent cations (which could be accurately measured), including the essential nutrients Mg, Ca and Mn, may be less affected by this increased binding capacity because of their relatively higher concentration within tissues. However, this does not explain why Sr would not show the same pattern of reduction. Perhaps this is a consequence of the relative binding capacity of Ba^{2+} ; it has a higher atomic number in comparison to Sr^{2+} (as well as all other divalent cations under assessment, followed by Cd^{2+}) and therefore likely has a higher binding affinity (i.e. Ba^{2+} and Cd^{2+} have greater binding affinity in comparison to the other divalent cations and thus outcompete them). This would imply that *Cd HMA2* plays some role in increasing the uptake of Cd from the soil (or its translocation from the roots), which would be relatively easy to test by analysing the roots of the mutant plants for relative Cd concentrations.

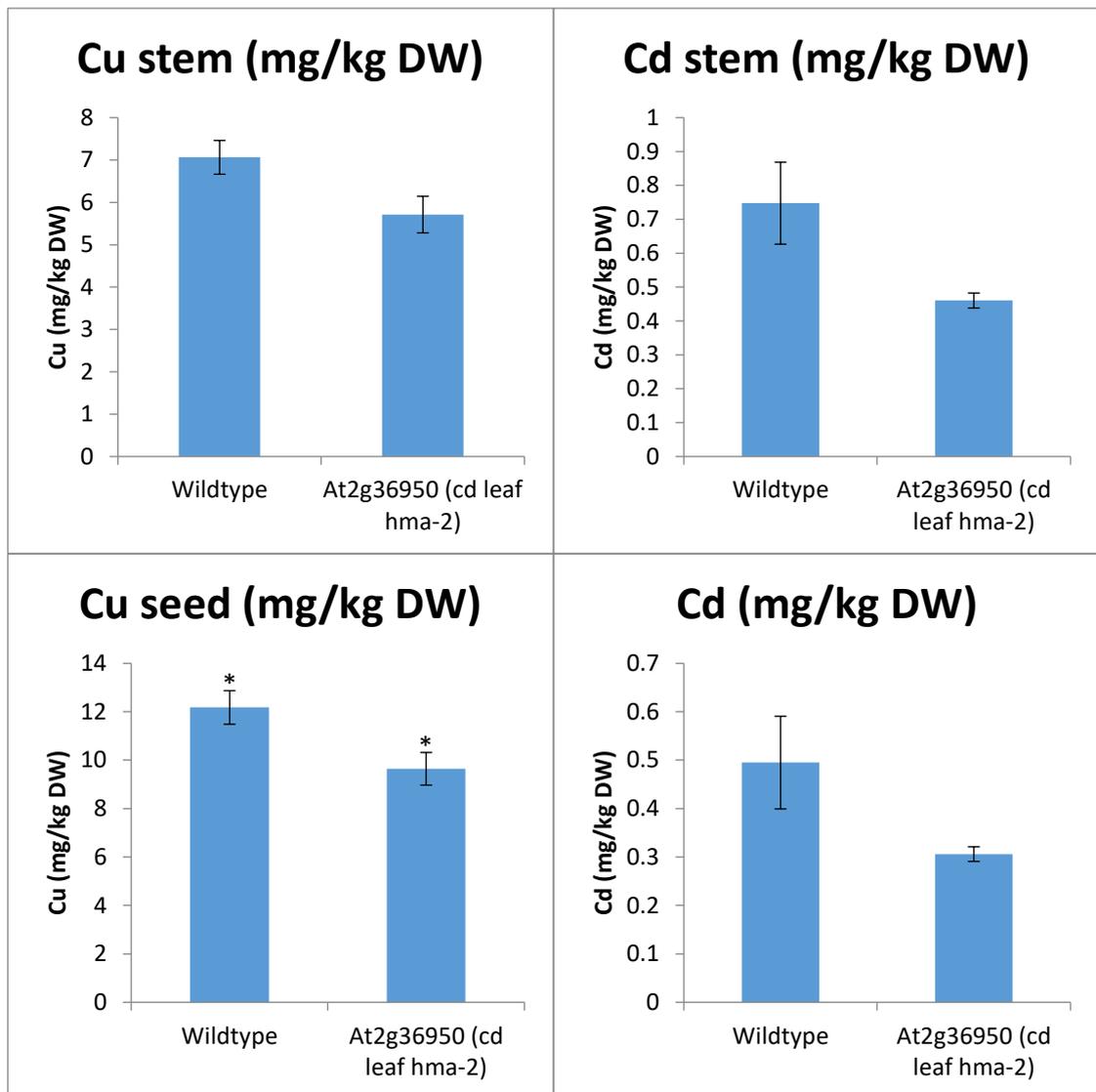


Figure 5.3.1.c Candidate gene analysis of Cu (left) and Cd (right) concentrations in the stem (top) and seeds (bottom) of *A. thaliana* insert mutants, as mg/kg DW of leaf tissue

Wildtype: *A. thaliana* Col 0; At2g36950 = *Cd leaf hma-2* (SALK_069207C). The mean and standard error are shown for each line, n = 4, where n is the number of individual plants sampled for analyses. No significant differences were observed for Cd concentrations, but Cu displayed some differences (* = p < 0.04). Cu stem t-test: t 2.3, df 6, p = 0.061; Cu seed t-test: t 2.62, df 6, p = 0.04; Cd stem t-test: t 2.34, df approx. 3.2, p = 0.096; Cd seed t-test: t 1.95, df approx. 3.15, p = 0.142

5.3.2 Leaf candidate gene analysis: Mo leaf concentration

As previously observed (according to the PiiMs database (Baxter et al., 2007a)), Mo was significantly disrupted in leaves of the *umamit 28* insert mutants (97% recovery from digestion, **Figure 5.3.2.a**). However, it was also found that S concentration was significantly higher ($p < 0.05$) in both insert mutant lines (95% recovery from digestion). Other elements showed inconsistency between the two lines e.g. *umamit28-1* had more B in comparison to the wildtype control and *umamit28-2*, but only Mo and S showed consistently high in comparison to the wildtype control for both lines investigated. The other element predicted to vary according to the PiiMs database was K (89% recovery), which in this experiment showed no significant difference to the wildtype control (although there were significant differences between the two lines). This could be due to the fact that different insert lines were analysed as part of this study, or it could be that when you assess multiple elements within one candidate you are statistically more likely to find an association by chance. Considering the candidate was picked out from Mo AT results and tested for S due to its relationship with Mo in *B. napus*, this adds credibility to the role of *UMAMIT 28* in leaf Mo and S.

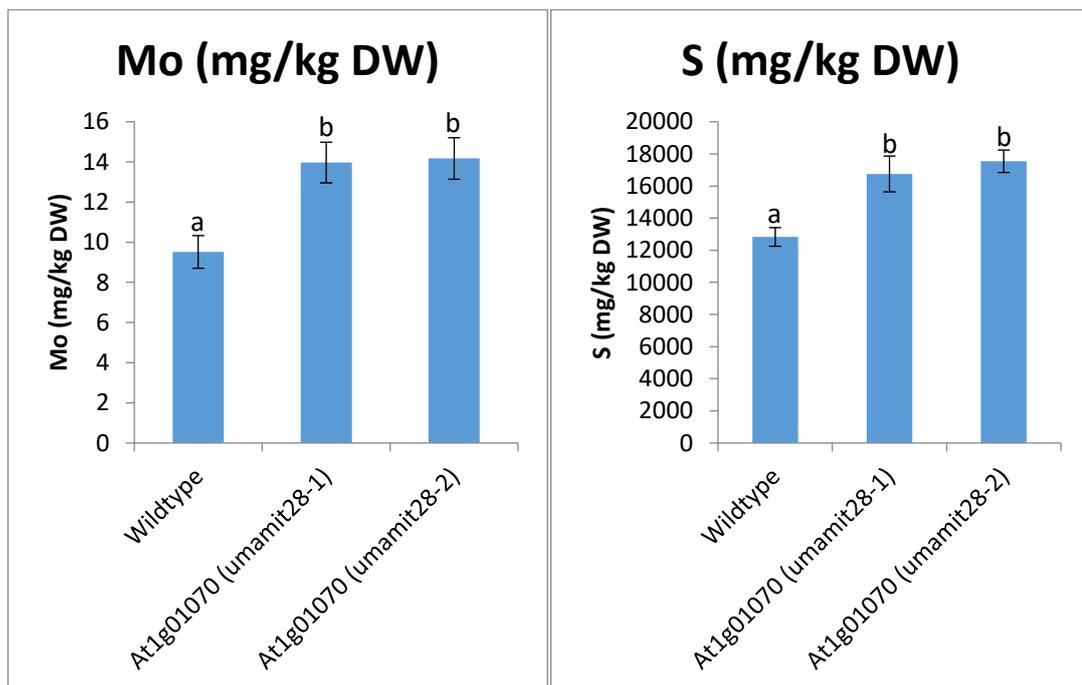


Figure 5.3.2.a Candidate gene analysis of Mo in the leaves of *A. thaliana* insert mutants, as mg/kg DW of leaf tissue

Wildtype: *A. thaliana* Col 0; At1g01070 = *umamit 28-1* (SALK_099741C) and *umamit 28-2* (SALK_147481C). The mean and standard error are shown for each line, Col $n = 8$, *umamit 28-1* $n = 12$, *umamit 28-2* $n = 8$, where n is the number of individual plants sampled for analyses. A significant difference between the wildtype control and the Mo and S concentrations was observed (Mo leaf ANOVA: $F 6.34$, $df 2, 25$, $p 0.006$; S leaf ANOVA: $F 6.28$, $df 2, 25$, $p 0.001$)

5.3.3 Leaf candidate gene analysis: summary, discussion and conclusions

Many of the candidates highlighted by AT analysis had previously been tested for a role in the element under investigation, resulting in very few candidates being taken forward into investigation with *A. thaliana* KO mutants. Furthermore, many of the GEMs were shared across multiple elements and were tested as part of the following chapter (see 6).

For Cu and Cd candidates, none of the candidates analysed for variation in leaf Cu concentration varied significantly from the wildtype control plants. It is thought this was primarily due to the way they were selected; to be specific, markers which were most commonly occurring in the top association peaks (*PAO4*, *CICDH* and *RhoGAP*), not necessarily that they had any indication of a major biological role relevant to leaf Cu concentration. These were analysed since the main candidate in the region of interest, *ATX1* (a known Cu chaperone), did not have a CDS gene model in the pan-transcriptome, although it should be within the syntenic block as indicated by the other CDS gene models and their *A. thaliana* orthologues. One of the Cu candidates (*cicdh-1*) displayed a stunted phenotype. However there were no T-DNA insertions within the candidate gene and therefore there must have been a spurious T-DNA insertion elsewhere. The *Cd hma-2* candidate (with Cu binding capacity according to (TAIR, 2015)), produced some interesting results. It displayed no significant variation in leaf Cu concentration ($p > 0.05$) but did display a significant increase in Cd concentration within the leaves ($p < 0.001$). A number of other element concentrations was also disrupted in the leaf, namely a significant decrease in B and Ba concentration ($p < 0.001$) and a significant increase in S, Mo and Cd concentrations ($p < 0.001$). This was thought to be a result of Cd toxicity mechanisms that cause increased S as the plant utilises S rich chelating agents and decreased B as more Cd is sequestered within the cell walls of roots. MoO_4^{2-} could be following as a structural analogue of SO_4^{2-} , while Ba may also be binding to the Cd modified cell walls. Other tissues (seed and stem) analysed showed no disruption or increase in Cd concentrations, making *Cd hma-2* a good potential candidate for phytoremedial purposes. The Mo candidate *umamit 28* showed a significant increase in both S and Mo concentration ($p < 0.01$). This contributes to previous research which observed a similar increase in leaf Mo concentration (according to the PiiMs database (Baxter et al., 2007a)), and adds weight to the link between S and Mo concentration in *B. napus*. Other element concentrations showed inconsistent results between the two mutant lines analysed as part of this study.

Overall, analysis of the limited leaf candidates was successful with two candidates being validated for a role within *A. thaliana*. Markers from this study could be used for MAS to breed for either increased S/Mo uptake/concentrations in the leaves for nutrient use efficiency, as

well as either increased (for phytoremedial purposes) or decreased (for human consumption) Cd concentrations.

5.4 Chapter summary and conclusions

In this chapter the genes controlling nutrient concentration within the leaves of *B. napus* have been explored. An AT approach was applied to a number of specific elements within the leaves (including Cu, Cd, Mn, Zn, Mo and S). Unlike seed analyses, only a very limited number of candidates was taken forward to be tested with *A. thaliana* T-DNA insert lines. This was a result of very few novel loci being identified which had not been previously tested in *A. thaliana*. Perhaps this is a consequence of there being many associated loci of small effect or that the traits under investigation were highly conserved across the diversity panel (see 7.1). Furthermore, perhaps the limitations of the pan-transcriptome coverage have limited the candidates identified, e.g. the lack of a CDS gene model for *ATX1*. Additionally, well known candidates were not tested further, such as *HMA 2+3* in Cd (see 5.2.2), but would make good targets for further testing with radiation/EMS TILLING or MAS. Furthermore, a common pattern was observed amongst most GEM AT outputs. Orthologues of the floral regulators *SOC1* and *FLC* were frequently the most highly associated GEMs. This became the focus of subsequent experiments, as detailed in the next chapter, 6. Therefore the overall number of candidates investigated for the leaves has been limited (alongside the previously discussed limitations of studying candidate genes in *A. thaliana* and experimental design limitations, such as lack of transcript quantifications of the T-DNA lines analysed, see 7.1).

Of the candidates taken forward for further analysis, two were successfully validated in *A. thaliana*. The Cd unidentified candidate had significantly ($p < 0.001$) higher concentrations of Cd within its leaves and showed disruption for a number of other elements (reduced B and Ba/ increased S, Mo and Cd ($p < 0.001$)). It was theorised that these elements may be disrupted as part of a Cd toxicity response; increased S (and Mo as an analogue) for the production of chelators to deal with the increase Cd within the plant; decreased B in the leaf as the root cell walls are modified to sequester the excess Cd and reduced Ba as a consequence of binding the same sites in the cell wall as Cd (but with higher affinity). However, this merely a theory and requires further validation. For example, other tissues such as the roots could be analysed with ICP-MS (e.g. for a change in B), whilst levels of GSH or phytochelatin could be measured with mass spectrometry and HPLC (e.g. Mendoza-Cózatl et al., 2008). The Mo candidate, *umamit 28* displayed increased Mo and S concentrations in leaves, corroborating previous research which found an increase in Mo (PiiMs, (Baxter et al., 2007a)) and a close relationship between Mo and S in *B. napus*. Once again, these candidates could be further investigated by exploiting the

natural variation within the diversity panel, potentially leading to MAS for phytoremedial purposes in the case of Cd HMA or improved S/Mo efficiency for *UMAMIT 28*.

6 Investigating the relationship between the wider leaf ionome and flowering

6.1 Introduction

Individual AT GEM analyses of leaf elements were commonly giving orthologues of *FLC* and *SOC1* as some of the most highly associated GEM AT results. When the trait data was further investigated, it was shown that not only were many of the element concentrations within the leaf ionome negatively correlated with *FLC* orthologue expression and positively correlated with *SOC1* orthologue expression, they showed an overall negative correlation to flowering time i.e. the longer the plants took to flower the less concentrated the elements within the leaves were (as measured in days taken to flower relative to the fastest flowering genotype, **Table 5.3.3.a**). There are many potential explanations for this observation. Firstly, there are both spring and winter (oilseed rape) OSR ecotypes within the RIPR diversity panel. This could create these associations if the plants were sampled prior to fulfilling the vernalisation requirement of all lines. Alternatively, the floral regulators could in some way be interacting with the nutrient status of the plants. To investigate this relationship further a number of experiments was designed. Firstly, an attempt was made to control variation for flowering by the analysis of a split data set; the data was split according to *FLC* expression and then by flowering time (as measured by a project student Mr Cándido José Martínez Ortuño, see 6.2.3). The intention was that by looking at winter and spring OSR ecotypes separately it would be possible to remove the association observed with the floral regulators. When this failed, values for the 'total leaf essential ionome' and the 'coefficient of variation' were calculated (see 6.2.2). These were analysed with AT and the association regions highlighted were analysed for common candidates with the flowering time (in days) AT analysis (using the flowering time data generated by Mr Martínez Ortuño). Candidates from these analyses were initially tested in *A. thaliana* insert lines before a wider leaf ionome timeline was conducted in *B. napus*.

Table 5.3.3.a Flowering time correlations (as days relative to the first flowering line, generated by a project student, Mr Martínez Ortuño) against the expression (as RPKM) of all CDS gene models whose orthologues in *A. thaliana* are annotated as *SOC1/FLC*. Leaf element concentrations, as mg/kg DW and summed, and 'yield', as g.

Yield is only indicative as any side stems outside of the pollination bags were removed, see (Thomas et al., 2016). R values are displayed to 3 decimal places, significant correlations ($p < 0.001$) are indicated in bold and underlined, while those for $p < 0.05$ are only in bold, $n = 297$. The table is coloured to highlight positive correlations (red) and negative correlations (blue). Relative differences in expression (average RPKM for each CDS gene model) are indicated in yellow for RPKM average values.

Trait	Flowering time (days)	A2_FLC_Cab0361 25.1	A3_FLC_Cab0024 72.4	A10_FLC_Cab00766 1.1	C2_FLC_BnaC02g00490D	C3_FLC_Bo3g00547 0.1	C3_FLC_Bo3g02425 0.1	C9_FLC_Bo9g17337 0.1	C9_FLC_Bo9g17340 0.1	A3_SOC1_Cab00326 7.1	A4_SOC1_Cab04575 4.1	A5_SOC1_Cab02535 6.1	C3_SOC1_Bo3g0388 80.1	C4_SOC1_Bo4g0248 50.1	C4_SOC1_Bo4g195720.1	Yield (g)
Na	<u>-0.470</u>	-0.096	<u>-0.470</u>	<u>-0.335</u>	<u>-0.472</u>	<u>-0.248</u>	<u>-0.424</u>	<u>-0.189</u>	-0.090	<u>0.382</u>	<u>0.267</u>	<u>0.370</u>	0.281	<u>0.392</u>	<u>0.214</u>	-0.095
Total	<u>-0.439</u>	<u>-0.140</u>	<u>-0.438</u>	<u>-0.382</u>	<u>-0.412</u>	<u>-0.219</u>	<u>-0.407</u>	<u>-0.242</u>	-0.082	<u>0.384</u>	<u>0.269</u>	<u>0.407</u>	0.290	<u>0.418</u>	<u>0.181</u>	-0.063
Mg	<u>-0.426</u>	-0.045	<u>-0.409</u>	<u>-0.213</u>	<u>-0.468</u>	<u>-0.153</u>	<u>-0.382</u>	<u>-0.114</u>	<u>-0.155</u>	<u>0.287</u>	<u>0.166</u>	<u>0.335</u>	<u>0.172</u>	<u>0.385</u>	<u>0.133</u>	-0.004
S	<u>-0.420</u>	<u>-0.224</u>	<u>-0.512</u>	<u>-0.428</u>	<u>-0.414</u>	<u>-0.303</u>	<u>-0.428</u>	<u>-0.313</u>	-0.015	<u>0.440</u>	<u>0.393</u>	<u>0.434</u>	0.328	<u>0.443</u>	<u>0.313</u>	-0.029
Ca	<u>-0.379</u>	-0.048	<u>-0.327</u>	<u>-0.157</u>	<u>-0.379</u>	<u>-0.116</u>	<u>-0.274</u>	-0.091	<u>-0.126</u>	<u>0.269</u>	0.067	<u>0.324</u>	<u>0.181</u>	<u>0.388</u>	0.052	0.012
Mo	<u>-0.367</u>	<u>-0.253</u>	<u>-0.455</u>	-0.317	<u>-0.339</u>	<u>-0.344</u>	<u>-0.390</u>	<u>-0.233</u>	0.002	<u>0.405</u>	<u>0.275</u>	<u>0.370</u>	0.285	<u>0.398</u>	<u>0.203</u>	-0.135
Cd	<u>-0.311</u>	-0.104	<u>-0.299</u>	-0.195	<u>-0.314</u>	<u>-0.173</u>	<u>-0.240</u>	-0.110	-0.087	<u>0.182</u>	0.020	<u>0.241</u>	0.072	<u>0.301</u>	0.094	<u>0.149</u>
Sr	<u>-0.299</u>	0.002	<u>-0.268</u>	-0.089	<u>-0.336</u>	-0.078	<u>-0.226</u>	-0.050	<u>-0.163</u>	<u>0.213</u>	0.010	<u>0.274</u>	<u>0.124</u>	<u>0.329</u>	0.014	0.017
Ti	<u>-0.297</u>	-0.105	<u>-0.263</u>	<u>-0.162</u>	<u>-0.300</u>	-0.100	<u>-0.212</u>	<u>-0.117</u>	-0.095	<u>0.196</u>	0.089	<u>0.283</u>	<u>0.175</u>	<u>0.302</u>	0.106	-0.009
As	<u>-0.286</u>	<u>-0.148</u>	<u>-0.238</u>	<u>-0.268</u>	<u>-0.263</u>	<u>-0.151</u>	<u>-0.220</u>	<u>-0.126</u>	-0.029	<u>0.317</u>	<u>0.179</u>	<u>0.354</u>	<u>0.254</u>	<u>0.382</u>	0.093	<u>-0.117</u>
Cu	<u>-0.228</u>	<u>-0.158</u>	<u>-0.312</u>	<u>-0.319</u>	-0.246	<u>-0.238</u>	<u>-0.236</u>	<u>-0.245</u>	0.104	<u>0.403</u>	<u>0.282</u>	<u>0.285</u>	<u>0.303</u>	<u>0.299</u>	<u>0.188</u>	0.024
Mn	<u>-0.216</u>	<u>-0.187</u>	<u>-0.221</u>	<u>-0.281</u>	<u>-0.139</u>	-0.085	<u>-0.130</u>	<u>-0.191</u>	0.055	<u>0.244</u>	<u>0.181</u>	<u>0.198</u>	<u>0.232</u>	<u>0.201</u>	<u>0.122</u>	-0.087
Ba	<u>-0.214</u>	-0.017	<u>-0.204</u>	-0.020	<u>-0.246</u>	-0.039	<u>-0.158</u>	0.009	<u>-0.140</u>	<u>0.217</u>	0.000	<u>0.257</u>	<u>0.125</u>	<u>0.303</u>	-0.021	-0.070
B	<u>-0.203</u>	-0.078	<u>-0.353</u>	<u>-0.237</u>	<u>-0.254</u>	<u>-0.163</u>	<u>-0.290</u>	<u>-0.182</u>	-0.054	<u>0.282</u>	<u>0.137</u>	<u>0.202</u>	<u>0.185</u>	<u>0.205</u>	0.102	<u>-0.132</u>
P	<u>-0.180</u>	-0.076	<u>-0.300</u>	<u>-0.339</u>	<u>-0.237</u>	<u>-0.200</u>	<u>-0.265</u>	<u>-0.214</u>	0.087	<u>0.221</u>	<u>0.241</u>	<u>0.159</u>	<u>0.157</u>	<u>0.163</u>	<u>0.155</u>	0.030
Al	<u>-0.161</u>	<u>-0.124</u>	-0.048	-0.101	<u>-0.123</u>	-0.013	-0.057	0.080	-0.001	<u>0.167</u>	0.101	<u>0.231</u>	0.105	0.233	0.014	<u>-0.129</u>
K	<u>-0.154</u>	-0.087	<u>-0.139</u>	<u>-0.227</u>	-0.096	-0.088	<u>-0.160</u>	<u>-0.146</u>	-0.013	<u>0.164</u>	<u>0.146</u>	<u>0.168</u>	<u>0.150</u>	<u>0.135</u>	0.072	-0.082
Rb	-0.110	-0.010	-0.091	<u>-0.211</u>	-0.066	-0.073	-0.108	-0.101	-0.009	0.058	0.060	0.049	0.066	0.032	0.007	-0.023
Cs	-0.107	-0.081	<u>-0.153</u>	<u>-0.167</u>	<u>-0.139</u>	-0.009	<u>-0.123</u>	-0.067	0.034	<u>0.114</u>	<u>0.138</u>	0.052	<u>0.141</u>	0.049	0.053	<u>-0.190</u>
Se	-0.095	-0.050	-0.099	-0.025	-0.015	0.029	-0.057	-0.064	-0.059	0.012	-0.026	0.041	-0.030	0.035	-0.039	-0.051
Zn	-0.048	-0.065	-0.090	<u>-0.169</u>	0.017	-0.080	-0.014	-0.095	0.063	0.049	-0.020	-0.007	0.007	0.012	-0.009	<u>0.128</u>
Fe	-0.017	<u>-0.124</u>	-0.014	<u>-0.121</u>	0.060	-0.005	0.024	-0.058	<u>0.149</u>	0.045	0.089	0.049	0.057	0.030	0.065	-0.041
Yield (g)	0.011	-0.001	0.092	0.000	0.106	0.077	0.085	-0.021	-0.031	-0.096	-0.086	-0.102	-0.053	-0.084	-0.071	NA
Flowering time (days)	NA	<u>0.255</u>	<u>0.574</u>	<u>0.494</u>	<u>0.432</u>	<u>0.318</u>	<u>0.548</u>	<u>0.256</u>	-0.021	<u>-0.501</u>	<u>-0.327</u>	<u>-0.438</u>	<u>-0.360</u>	<u>-0.470</u>	<u>-0.217</u>	0.011
Average RPKM		42.25	8.79	45.66	32.50	120.63	1.94	9.26	29.14	17.21	21.49	2.16	3.72	5.81	9.73	NA

6.2 Leaf ionome and flowering specific methods:

6.2.1 Controlling for spring and winter OSR ecotypes: splitting the diversity panel according to flowering time and *FLC* expression

Considering that an association between leaf nutrient status and flowering time may be a consequence of the presence of spring and winter ecotypes within the diversity panel (i.e. the leaves were at different developmental stages and could therefore have had different nutrient statuses) it was cogent to test if the relationship was maintained when the panel was split into spring and winter ecotypes. It was quickly observed that the ecotype information associated with the diversity panel seemed to display inconsistencies. For example spring ecotypes with high *FLC* expression and winter ecotypes with low *FLC*. Therefore in order to be certain that the panel was being correctly split, actual flowering time data was required. A suitable dataset had been previously produced as part of an Erasmus/Lifelong learning project by Mr Cándido José Martínez Ortuño (as detailed in 6.2.3). The data was then split in two ways: according to flowering time (as days to flowering) and based on A3 *FLC* expression (as RPKM). The A3 copy of *FLC* was used as it was consistently the most highly associated marker in AT analyses. Data was trimmed so that only accessions with both flowering time and leaf nutrient information were analysed, allowing for direct comparison. Furthermore, only lines which were designated as a spring or winter oilseed rape (OSR) ecotype were included (i.e. exotics, spring fodders, synthetic swede, swede and anything with no ecotype information were removed from analysis). The dataset was then ordered according to either *FLC* expression or flowering time in days, split into halves and then quartiles. Each subset was then assessed for correlations: between element concentrations, flowering time and the expression of *FLC* and *SOC1* orthologues.

6.2.2 ‘Total leaf essential ionome’ vs ‘Coefficient of Variation’

The ‘total leaf essential ionome’ was calculated as a sum for each accession of all the essential element concentrations from the RIPR dataset that were successfully analysed with ICP-MS in leaves (11 elements in total: B, Ca, Cu, Fe, K, Mg, Mn, Mo, P, S and Zn). This total was then analysed by AT and the associated regions of the genomes were assessed for the presence of candidate trait control genes. It was compared to the ‘total ionome’ (all 21 elements successfully analysed with ICP-MS; Al, As, B, Ba, Ca, Cd, Cs, Cu, Fe, K, Mg, Mn, Mo, Na, P, Rb, S, Se, Sr, Ti and Zn) but the results were almost exactly identical and therefore only the ‘essential total’ was taken forward (e.g. 6.3.1). To calculate the ‘coefficient of variation’ once again only the essential elements were included; these were individually ranked from 1- 383 (1 being the

line with the lowest concentration and 383 being the highest), the average rank and standard deviation across all elements for each accession was taken. The standard deviation was then divided by the average to give the 'coefficient of variation' for each accession; in this way it was possible to distinguish which lines had consistently higher element concentrations from those which had either consistently lower element concentrations or those which displayed a lot of variability. The purpose of this approach was to provide a clearer picture of overall concentrations in comparison to the 'total method' as it would be less skewed by the more abundant macronutrients.

6.2.3 *B. napus* flowering time

Flowering time was analysed as part of a separate project by a visiting student (Mr Cándido José Martínez Ortuño); its use within this study was for splitting the panel (see 6.2.1) and comparison to leaf element concentration data. Briefly, all the accessions from the whole RIPR diversity panel were sown (415 accessions, October 2014) with four replicate plants per accession each within a glasshouse. Once established, the plants were moved outside to vernalise over winter (21/11/14 to 13/2/15). Flowering was measured relative to the earliest accession to flower: as soon as yellow was observed on the flowering bud it was considered to have flowered and every plant which flowered thereafter was counted relative to the first (i.e. the first accessions were counted as day zero and everything following was the number of days relative to the first accession). The average flowering time for each of the four plants per accession was used in further analysis (for splitting the panel and in AT analysis). Plants which failed to flower were removed from analysis. In total this left flowering data for 368 accessions to be analysed with AT. This data was used within this study to: perform correlations against *FLC/SOC1* orthologue expression and leaf element concentration data (**Table 5.3.3.a**); split the panel in an attempt to break the association between leaf nutrients and flowering time (see 6.3.1); and to compare AT outputs (see 6.3.4) to those of 'total leaf essential elements' (see 6.3.2) and the 'coefficient of variation' (see 6.3.3).

6.2.4 Candidate gene analysis: growth condition variation in *A. thaliana*

Unlike the *A. thaliana* lines grown for seed and general leaf analysis, the plants grown for the leaf ionome flowering variation were grown under two conditions (although still within the same P24 pots). After a few weeks of growth at (12 hour day/ night, 16-20°C temperature range), half the plants were moved to short day conditions (8 hour day at 21°C and 16 hour night at 15°C) and half the plants moved to longer days (16 hour day, 16-20°C temperature range). This would allow the investigation of the potential flowering lines under both short and

long day conditions to investigate whether there was any link between flowering and the leaf ionome.

6.2.5 Leaf ionome timeline in *B. napus*

Plants for use in the leaf ionome timeline experiment were the same as those grown for the pod ionome comparisons. For details of how the plants were grown see (4.2.5.1).

To understand whether the link between flowering time and leaf nutrient concentration observed in AT analysis of the leaf ionome was the result of differences between spring and winter ecotypes or an adaptive mechanism to variable leaf nutrient concentration, a leaf ionome timeline experiment was designed. 5 trays of plants were grown so that leaf development stage could be compared between undamaged plants at different developmental stages, under vernalised and non-vernalised conditions. The second true leaves of plants from tray 1 were sampled once the plants had reached the 4th true leaf. Plants from trays 4 and 5 were moved into vernalisation conditions at this development stage (4th true leaf, see above). The 2nd and 4th trays were sampled for their 5th true leaf once the plants had reached the 7th true leaf, whilst the plants of the 3rd and 5th trays were sampled for their 7th true leaf when they reached the 9th true leaf. The leaves of the plants were dried and stored until they could be digested and analysed with ICP-MS (as described in 2.5).

6.3 Results

6.3.1 Controlling for spring and winter OSR ecotypes: splitting the diversity panel according to *FLC* expression and flowering time

None of the methods utilised for splitting the panel could remove the association between leaf nutrient concentrations and flowering time completely. This is unsurprising considering the number of CDS gene models included in analysis (8 for *FLC* and 6 for *SOC1*), the number of elements under analysis (21, i.e. the number of elements accurately analysed in leaves: Al, As, B, Ba, Ca, Cd, Cs, Cu, Fe, k, Mg, Mn, Mo, Na, P, Rb, S, Se, Sr, Ti and Zn), and the quantity of measurements for each trait under analysis (at least n: 63). This gave many instances where a significant association ($p < 0.001$) occurred between a CDS gene model and a trait (data not presented, cf. (Thomas et al., 2016) for full correlation tables comparing all elements within leaves, as well as in comparison to seed and various other trait datasets, alongside stepwise discriminant analysis). As such, analyses were limited to looking at associations between the various elements against the orthologues of *FLC* (A3 Cab002472.4) and *SOC1* (C4 Bo4g024850.1) which were the most highly associated with leaf nutrients, the most highly expressed copy of *FLC* (C3 Bo3g005470.1) and the flowering time in days. Furthermore, only analyses which were significant to $p < 0.001$ were considered. As the number of observations was generally large the chance of type 1 errors was increased, therefore a more stringent p value was used throughout these analyses alongside comparison to effect size ($p < 0.001$ generally reflected a higher effect size/Pearson's r correlation coefficient: small effect being +/-0.2, medium +/-0.5, large +/-0.8 (Nakagawa and Cuthill, 2007; Sullivan and Feinn, 2012)). To give a general overview of what was happening within the leaf ionome a number of summed totals was included (i.e. data was summed to give a general overview): total ionome (sum of all element concentrations in the ionome); essentials/non- essentials (sum of essentials (B, Ca, Cu, Fe, K, Mg, Mn, Mo, P, S, Zn)/non-essential (Al, As, Ba, Cd, Cs, Na, Rb, Se, Se, Sr, Ti) element concentrations included in the study); Mobile/immobile (sum of essential elements considered to be mobile (Cu, Fe, K, Mg, P and S)/ immobile (B, Ca, Mn, Mo, Zn) in *B. napus* (see 6.3.7 and Maillard et al., (2015)).

Analyses which were split based upon A3 *FLC* expression (**Table 6.3.1.a**) values were effective in removing any correlation between leaf nutrients and flowering time for both Q1 (the lowest quartile/hinge, i.e. the lowest *FLC* expression) and Q4 (the highest/upper quartile/hinge, i.e. the greatest *FLC* expression). However, there was a significant ($p < 0.001$) association between the 'sum total ionome' (and many of the individual elements) for *FLC* and flowering time in Q2. Further when *FLC* and *SOC1* expression were compared it was found that these were also

significantly associated with time to flowering, suggesting that the method of splitting the data had been ineffective for this group. Considering that Q2 plants had a low to intermediate *FLC* expression in comparison to the other groups, this would suggest that it is the difference between the spring ecotypes causing the association. This fits well with the AT results; population structure is controlled as part of AT analysis (with the use of a Q matrix) so it would be hypothesised that any associations observed were not a consequence of variation between different ecotypes. However, when this group was re-analysed with everything defined as a winter (n 22) ecotype removed, the correlation between C4 *SOC1* and flowering was maintained (R -0.697, $p < 0.001$, n 44), as was the association between flowering and 'total ionome' (R -0.522, $p < 0.001$, n 44), but the association between *FLC* and flowering was finally broken (R 0.131, $p = 0.398$, n 44). This is interesting as it implies that the link between flowering and leaf nutrient status may not be vernalisation dependent (since *FLC* is no longer correlating) but may be associated with the flowering pathway in general (as *SOC1* integrates signals from multiple pathways, (Lee and Lee, 2010)). None of the various *FLC* homologues as part of this reduced set showed a significant correlation to flowering time (as $p < 0.001$, although the A2 *FLC* was close, $r = 0.470$, $p = 0.0013$), whilst three *SOC1* homologues did (A3 $r = -0.554$, A5 $r = -0.582$ and C4 as above, $p < 0.001$ n 44). In addition, it was found that for Q3 a significant ($p < 0.001$) association between the A3 *FLC* expression and the 'total ionome', 'total essentials', 'total mobile' and K was maintained. Considering the apparent link between nutrients and flowering time in general it was therefore decided to try splitting the data based on flowering time.

Table 6.3.1.a Correlation table between *FLC/SOC1* expression, flowering time and leaf nutrient concentration when the diversity panel is split according to A3 *FLC* expression.

Correlation tables for flowering time data (as days, see 6.2.3), A3 *FLC*, C3 *FLC* and C4 *SOC1* RPKM expression values, yield (as g from 6 plants), all elements within the leaf ionome (as mg/kg) and a range of summed totals (total ionome: sum of all elements in the ionome; essentials/non- essentials: sum of essential/non-essential elements included in the study; Mobile/immobile essential (see); sum of elements considered to be mobile/ immobile in *B. napus* (see 6.3.7, (Maillard et al., 2015))). Data for the diversity panel was split based on *FLC* expression into quartiles/hinges from lowest (Q1) to highest (Q4) and tested for associations (Q1 top left, Q2 top right, Q3 bottom left and Q4 bottom right). R values are displayed to 3 decimal places, significant correlations ($p < 0.001$) are indicated in bold and underlined, while those for $p < 0.05$ are only in bold. The table is coloured to highlight positive correlations (red) and negative correlations (blue).

Q1 Trait (FLC, n:67)	A3_FLC_Cab002472.4	C3_FLC_Bo3g005470.1	C4_SOC1_Bo4g024850.1	Flowering time (days)
Na	-0.063	0.123	0.069	-0.164
Non-essentials	-0.068	0.121	0.076	-0.162
Al	-0.230	-0.082	0.273	-0.157
Se	0.067	0.289	0.089	-0.141
Mobile essentials	0.018	-0.117	0.163	-0.137
Rb	0.025	-0.139	-0.030	-0.126
K	0.086	-0.094	0.033	-0.125
Total Ionome	-0.071	-0.109	0.302	-0.094
Essentials	-0.068	-0.120	0.305	-0.085
Yield (g)	-0.255	-0.251	-0.013	-0.069
Fe	-0.020	0.003	-0.034	-0.031
Mg	-0.216	-0.102	0.305	-0.025
Cs	-0.075	0.071	-0.014	-0.021
P	0.118	0.166	-0.021	-0.019
Zn	-0.011	-0.045	0.026	-0.019
S	-0.082	-0.059	0.187	-0.014
Cd	-0.112	-0.037	0.270	0.000
Cu	-0.030	0.131	0.060	0.013
Mo	-0.051	-0.287	0.147	0.036
Mn	-0.056	0.075	0.007	0.073
Ca	-0.204	-0.017	0.353	0.112
Immobile essentials	-0.204	-0.016	0.352	0.113
Sr	-0.156	-0.014	0.306	0.117
Ti	-0.074	0.040	0.263	0.133
As	-0.233	-0.203	0.313	0.157
Ba	-0.146	-0.051	0.276	0.187
B	0.026	0.104	0.032	0.226
Flowering (days)	-0.039	-0.015	-0.113	NA

Q2 Trait (FLC, n:66)	A3_FLC_Cab002472.4	C3_FLC_Bo3g005470.1	C4_SOC1_Bo4g024850.1	Flowering time (days)
Total Ionome	-0.400	-0.060	0.361	-0.539
Essentials	-0.392	-0.041	0.361	-0.532
Cd	-0.184	-0.296	0.454	-0.483
Mg	-0.404	-0.126	0.295	-0.474
Immobile essentials	-0.277	-0.079	0.312	-0.471
Ca	-0.277	-0.078	0.311	-0.469
Non-essentials	-0.378	-0.346	0.210	-0.447
Na	-0.379	-0.351	0.207	-0.444
Mo	-0.258	-0.225	0.259	-0.441
As	-0.171	-0.120	0.303	-0.437
Mobile essentials	-0.337	-0.015	0.288	-0.419
Sr	-0.208	-0.001	0.243	-0.388
S	-0.337	-0.118	0.252	-0.375
Ti	-0.251	-0.046	0.280	-0.355
Mn	-0.122	-0.102	0.242	-0.343
Ba	-0.135	0.076	0.167	-0.295
B	-0.301	0.032	0.115	-0.245
K	-0.142	0.077	0.178	-0.226
P	-0.367	-0.172	0.060	-0.213
Zn	0.079	-0.075	0.062	-0.169
Fe	-0.095	-0.092	0.093	-0.161
Cu	-0.044	-0.095	0.087	-0.153
Yield (g)	-0.163	0.011	-0.013	-0.141
Rb	-0.032	0.087	0.037	-0.114
Al	0.082	0.186	0.043	-0.079
Cs	-0.118	0.156	-0.115	-0.031
Se	-0.012	-0.090	-0.035	0.103
Flowering (days)	0.449	0.351	-0.647	NA

Q3 Trait (FLC, n:66)	A3_FLC_Cab002472.4	C3_FLC_Bo3g005470.1	C4_SOC1_Bo4g024850.1	Flowering_time (days)
S	0.150	-0.128	0.273	-0.253
Mo	0.144	-0.133	0.173	-0.252
Ba	0.146	-0.005	-0.113	-0.244
Immobile essentials	0.148	-0.053	0.027	-0.234
Ca	0.150	-0.053	0.024	-0.232
Sr	0.186	-0.029	-0.078	-0.212
B	0.079	0.012	0.034	-0.209
Ti	0.302	-0.060	0.067	-0.196
Mg	0.102	0.010	0.116	-0.189
Total Ionome	0.410	-0.011	0.064	-0.187
Mn	-0.152	-0.014	0.219	-0.180
Essentials	0.422	-0.003	0.056	-0.180
Non-essentials	-0.070	-0.141	0.157	-0.172
Na	-0.076	-0.141	0.160	-0.169
Fe	0.008	0.018	0.202	-0.134
Mobile essentials	0.416	0.015	0.053	-0.121
Al	0.137	-0.075	-0.021	-0.111
Yield (g)	0.261	0.143	0.124	-0.098
Cd	0.179	-0.036	-0.065	-0.081
Cs	-0.237	0.038	0.279	-0.080
K	0.408	0.054	-0.061	-0.032
As	0.098	0.074	0.078	-0.028
Se	-0.069	-0.014	-0.060	-0.024
P	0.240	-0.071	0.237	-0.023
Rb	0.351	0.077	-0.096	-0.023
Zn	0.118	-0.039	-0.080	-0.003
Cu	0.088	-0.133	0.119	0.118
Flowering (days)	0.003	0.121	-0.265	NA

Q4 Trait (FLC, n:67)	A3_FLC_Cab002472.4	C3_FLC_Bo3g005470.1	C4_SOC1_Bo4g024850.1	Flowering_time (days)
Mg	0.092	0.150	0.134	-0.301
Non-essentials	0.127	0.137	0.080	-0.271
Na	0.127	0.135	0.082	-0.270
As	0.115	0.013	-0.073	-0.245
Ca	-0.014	0.174	0.034	-0.216
Immobile essentials	-0.014	0.176	0.032	-0.211
Mo	0.165	-0.100	0.025	-0.201
Sr	-0.034	0.144	-0.006	-0.174
Al	0.075	0.123	-0.071	-0.144
Cu	-0.077	-0.096	0.036	-0.136
Ba	-0.040	0.053	0.003	-0.120
Cd	-0.096	0.177	0.027	-0.104
Ti	-0.107	-0.072	-0.023	-0.037
Yield (g)	-0.027	0.138	0.120	-0.020
Cs	0.092	-0.043	-0.028	0.009
Total Ionome	0.119	0.128	0.024	0.015
Essentials	0.115	0.124	0.021	0.027
S	0.016	-0.126	0.168	0.043
Mobile essentials	0.137	0.092	0.015	0.092
Mn	-0.036	0.172	-0.064	0.143
P	0.023	-0.180	0.104	0.147
Rb	0.065	0.066	0.010	0.155
K	0.134	0.117	-0.060	0.162
Zn	-0.153	0.145	-0.020	0.165
Fe	-0.056	0.105	-0.166	0.169
Se	0.096	-0.015	-0.088	0.208
B	0.115	-0.105	-0.028	0.356
Flowering (days)	-0.059	-0.001	-0.129	NA

Splitting the panel based on flowering time data had a much greater success rate (**Table 6.3.1.b**). It removed the association between *FLC/SOC1* to flowering in all quartiles except Q1 (where there was a significant association with *C4 SOC1*). Within this quartile (made up of entirely spring and 2 winter OSR ecotypes according to accession information), there was a significant association ($p < 0.001$) between flowering time, essential elements and the total ionome. The more immobile elements seemed to show better associations to *SOC1* expression both in Q1 and Q4. However this could have been a confounding effect of the more abundant Ca. Nevertheless, the results of these associations further support the conclusion that the apparent association between flowering time and leaf nutrients is a consequence of the earlier flowering spring ecotypes. However the association with *FLC* may be a consequence of spring/winter ecotype differences i.e. there is an association between flowering and leaf nutrients independent of *FLC*, but this is amplified when *FLC* expression is also considered. Note, splitting based on *C4 SOC1* expression was also performed (data not presented, see Appendix 19) and supported the conclusion that earlier flowering genotypes were more readily associated with higher leaf nutrient concentrations (i.e. Q4 had significant, $p < 0.001$, associations with the total ionome etc.). Since none of the methods of splitting the data seemed to adequately remove the association between flowering time and leaf nutrient status, it was decided that AT analyses would be examined more closely to try to find any common candidates between leaf nutrients and flowering time.

Table 6.3.1.b Correlation table between *FLC/SOC1* expression, flowering time and leaf nutrient concentration when the diversity panel is split according to flowering time.

Correlation tables for flowering time data (as days, see 6.2.3), A3 *FLC*, C3 *FLC* and C4 *SOC1* RPKM expression values, yield (as g from 6 plants), all elements within the leaf ionome (as mg/kg) and a range of summed totals (total ionome: sum of all elements in the ionome; essentials/non-essentials: sum of essential/non-essential elements included in the study; Mobile/immobile essential (see); sum of elements considered to be mobile/immobile in *B. napus* (see 6.3.7, (Maillard et al., 2015))). Data for the diversity panel was split based on flowering time data (as days to flowering) into quartiles/hinges from lowest (Q1) to highest (Q4) and tested for associations (Q1 top left, Q2 top right, Q3 bottom left and Q4 bottom right). R values are displayed to 3 decimal places, significant correlations ($p < 0.001$) are indicated in bold and underlined, while those for $p < 0.05$ are only in bold. The table is coloured to highlight positive correlations (red) and negative correlations (blue).

Q1 Trait (Flowering time)	A3_FLC_Cab002472.4	C3_FLC_Bo3g005470.1	C4_SOC1_Bo4g024850.1	Flowering_time (days)
Essentials	-0.1295	0.0164	0.3189	-0.4232
Total Ionome	-0.1341	0.0076	0.322	-0.4197
Mobile essentials	-0.0759	0.0333	0.172	-0.3445
K	-0.0331	0.0367	0.0263	-0.2765
Mg	-0.1187	-0.0599	0.3888	-0.2682
Immobile essentials	-0.1461	-0.0345	0.3929	-0.2611
Ca	-0.1461	-0.0349	0.3943	-0.2603
Cd	-0.1318	-0.0945	0.3908	-0.2377
Al	0.1105	0.0701	0.004	-0.2127
Sr	-0.1319	0.0212	0.3395	-0.1898
As	0.0923	-0.3051	0.2614	-0.167
Mn	-0.0381	-0.0018	0.0701	-0.1669
Rb	-0.0187	0.0106	-0.0702	-0.1655
Ba	-0.0706	0.0145	0.2306	-0.1627
Yield (g)	0.1792	-0.1058	0.0055	-0.1387
Ti	-0.1291	0.1	0.2506	-0.1339
Non-essentials	-0.1314	-0.1351	0.1896	-0.1281
Na	-0.1309	-0.1375	0.1851	-0.1225
Se	-0.2144	0.2318	0.1218	-0.1211
Cu	-0.1547	0.0268	0.1254	-0.1068
B	-0.1406	0.1558	-0.0639	-0.101
S	-0.0497	-0.0226	0.189	-0.0908
Mo	-0.164	-0.2423	0.1534	-0.0673
Fe	-0.0615	0.0642	-0.082	-0.0074
Cs	0.0285	0.1094	-0.1461	0.0098
Zn	-0.0196	0.0041	0.0586	0.0105
P	0.0073	0.2225	-0.0817	0.1172
Flowering (days)	0.196	0.2005	-0.5289	-

Q2 Trait (Flowering time)	A3_FLC_Cab002472.4	C3_FLC_Bo3g005470.1	C4_SOC1_Bo4g024850.1	Flowering_time (days)
P	-0.4716	-0.3953	0.3108	-0.4312
S	-0.5683	-0.376	0.2274	-0.3694
B	-0.37	-0.2101	0.102	-0.2916
Mobile essentials	-0.1683	-0.156	0.0031	-0.1921
Zn	-0.2479	-0.161	0.0515	-0.1883
Essentials	-0.163	-0.0802	-0.0301	-0.1839
Total Ionome	-0.1667	-0.0779	-0.0394	-0.1801
Mo	-0.2595	-0.343	0.1444	-0.1274
Cu	-0.2154	-0.2144	0.2796	-0.1256
Mn	-0.1219	0.0092	0.1002	-0.1113
Cs	-0.0798	-0.0511	0.151	-0.1048
Mg	-0.0699	0.0941	-0.0531	-0.0961
Rb	0.0514	-0.1135	-0.049	-0.0765
As	0.0412	0.1303	0.177	-0.0686
K	0.0423	-0.0583	-0.0782	-0.0309
Immobile essentials	-0.016	0.1748	-0.0884	-0.0127
Ca	-0.0134	0.1759	-0.0901	-0.0104
Non-essentials	-0.1234	0.0057	-0.1637	-0.0102
Na	-0.1256	0.0032	-0.1652	-0.01
Sr	0.0037	0.1992	-0.08	-0.0077
Al	0.0431	-0.0241	0.0961	0.0032
Fe	0.174	-0.0682	0.1521	0.0127
Cd	0.0532	0.129	-0.0968	0.0152
Ti	-0.0597	0.0528	0.0091	0.0208
Ba	0.0131	0.2157	-0.0508	0.0226
Yield (g)	0.0627	-0.0685	0.0277	0.0408
Se	-0.0014	0.0034	-0.1599	0.1479
Flowering (days)	0.384	0.2618	-0.2313	-

Q3 Trait (Flowering time)	A3_FLC_Cab002472.4	C3_FLC_Bo3g005470.1	C4_SOC1_Bo4g024850.1	Flowering time (days)
Al	-0.0083	-0.0289	0.2929	-0.1684
Ba	-0.0643	0.0392	-0.038	-0.1389
Non-essentials	-0.3903	-0.0392	0.1261	-0.1267
Na	-0.391	-0.0402	0.1247	-0.1247
Ca	-0.1049	0.075	-0.0232	-0.124
Immobile essentials	-0.1059	0.0755	-0.0216	-0.1238
Sr	-0.0883	0.0462	-0.0688	-0.1065
Cd	-0.333	-0.0801	0.102	-0.1056
S	-0.3868	-0.1454	0.3274	-0.1049
Mo	-0.3341	-0.118	0.2133	-0.0964
Zn	-0.1198	0.0902	0.0871	-0.0721
Total Ionome	-0.309	0.034	0.1688	-0.0612
Essentials	-0.2955	0.0366	0.1659	-0.056
B	-0.3645	-0.0547	0.3347	-0.0558
Yield (g)	0.0646	0.0919	0.101	-0.0526
P	-0.2739	-0.2115	0.2814	-0.0515
Mg	-0.288	-0.0054	0.1691	-0.051
Cu	-0.1101	-0.0707	-0.0033	-0.0385
Fe	-0.0053	0.1325	0.0456	-0.0371
Mobile essentials	-0.2952	0.0195	0.1879	-0.0275
Mn	-0.0216	0.0447	0.0197	-0.0106
Rb	-0.1783	0.0517	0.0144	0.0029
K	-0.1517	0.0896	0.0654	0.013
Ti	-0.0769	0.0271	0.0192	0.0256
Cs	-0.0987	0.0486	0.0948	0.0531
As	-0.0682	0.0694	0.027	0.0796
Se	0.0183	-0.0965	-0.1142	0.1604
Flowering (days)	0.1359	-0.0215	-0.1279	-

Q4 Trait (Flowering time)	A3_FLC_Cab002472.4	C3_FLC_Bo3g005470.1	C4_SOC1_Bo4g024850.1	Flowering time (days)
Al	-0.0456	0.2519	0.062	-0.2519
As	-0.3325	-0.0646	0.2532	-0.0985
Cd	-0.1642	-0.1898	0.2544	-0.0611
Zn	0.2007	-0.0044	-0.1479	-0.0466
Yield (g)	0.1238	0.3293	-0.1954	-0.0203
Ba	-0.3107	-0.2803	0.4672	-0.0107
Sr	-0.2873	-0.2784	0.4206	0.0011
Ca	-0.3223	-0.2514	0.5247	0.0203
Immobile essentials	-0.3214	-0.2491	0.5264	0.0234
Rb	0.1445	0.2081	-0.001	0.0782
Mg	-0.4074	-0.1824	0.3543	0.0789
Cs	-0.0712	0.1235	0.0932	0.0889
Non-essentials	-0.2052	-0.1636	0.1547	0.1086
K	0.0661	0.1431	0.1192	0.1106
Na	-0.202	-0.1642	0.1469	0.1119
Se	-0.2102	-0.0761	0.1006	0.1249
Essentials	-0.0964	0.0225	0.3098	0.1309
Mo	-0.3168	-0.1557	0.3783	0.133
Total Ionome	-0.1032	0.0154	0.3111	0.1332
P	0.1537	0.0747	-0.0613	0.1396
Fe	0.1778	0.0143	-0.1451	0.1423
Mobile essentials	-0.0177	0.0984	0.2043	0.1443
Ti	-0.1269	-0.3876	0.4794	0.1553
Cu	-0.246	-0.0971	0.1034	0.165
S	-0.1672	0.0126	0.3993	0.1837
Mn	-0.0702	0.0817	0.2627	0.1922
B	-0.0747	-0.0454	0.2985	0.2017
Flowering (days)	0.2017	-0.025	-0.0515	-

6.3.2 Associative transcriptomic outputs, predictions and candidates: 'Total leaf essential elements'

Unlike most of the individual element AT results, both the SNP and GEM analyses for 'total leaf essential elements' passed the Bonferroni corrected significance threshold and 5% false discovery rate (FDR, **Figure 6.3.2.a**). A significant association peak passing both the Bonferroni and FDR thresholds was found within the SNP AT analysis on A10 and one GEM (Cab002472.4 orthologue of *FLC*: AT5G10140) passed both the Bonferroni corrected significance threshold and the FDR. The predictive capability of markers was assessed with the 274 diversity panel (**Table 6.3.2.a**); the orthologues of *FLC* (Cab002472.4 and Bo3g024250.1) and *SOC1* (Bo4g024850.1) were predictive, as was an unannotated regulatory gene (Cab007824.2 orthologue of a RING/U box family protein: AT5G12310.1) and Bo3g032510.1 (orthologue of CASP-LIKE PROTEIN 4B: AT2G38480, discussed below). Five predictive SNP markers were found ($p < 0.05$), one of which was found on A10 within the same association peak observed in the 383 panel.

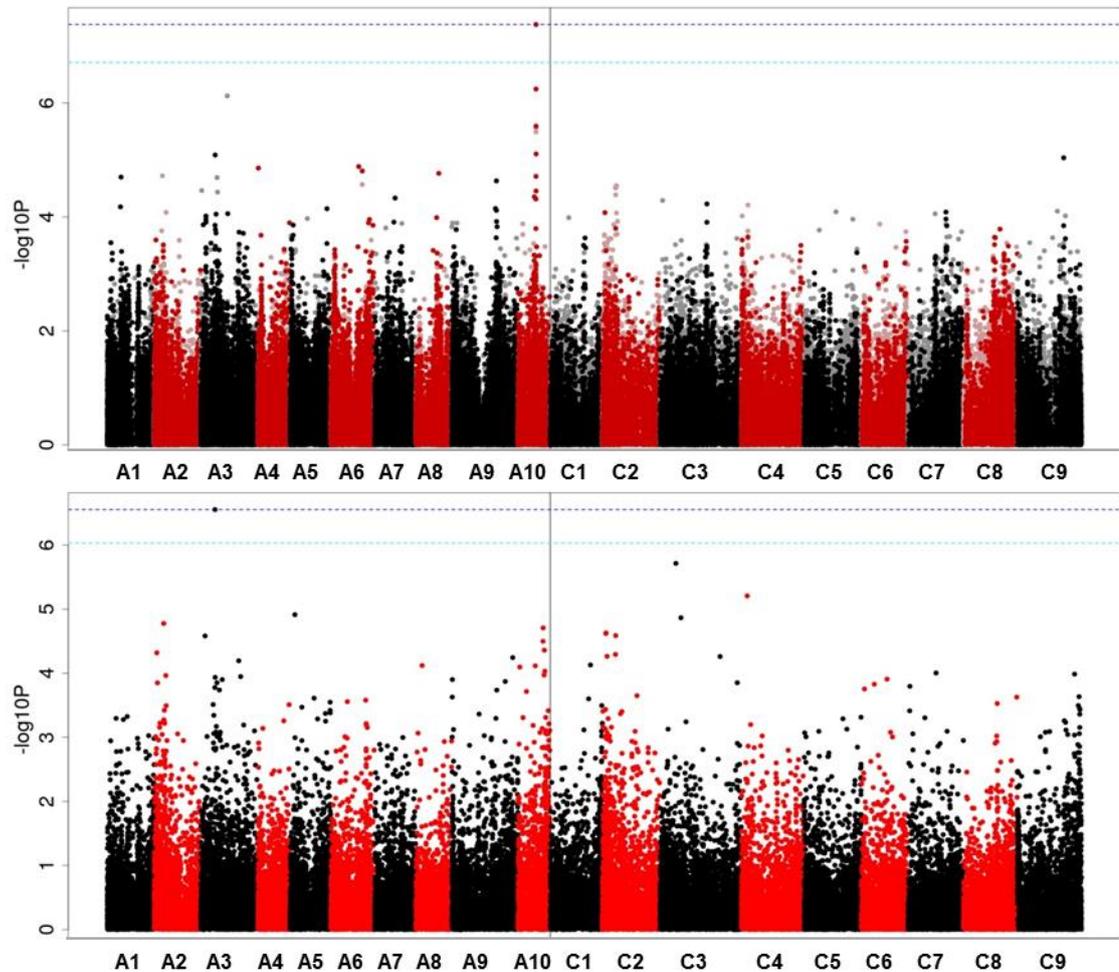


Figure 6.3.2.a Genome wide distribution of mapped markers associating with the ‘Total essential leaf ionome’ (mg/kg DW), calculated from the sum total of essential element concentrations measured with ICP-MS analysis for each of the 383 accessions.

SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and ‘total leaf essential elements’ as the response variable. $-\log_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

Table 6.3.2.a Predictive capability of markers from ‘total leaf essential ionome’ AT analysis.

For assessing the predictive capability of markers the highest SNP markers from discernible association peaks and the most highly associated GEMs were analysed. The marker type is given as either SNP or GEM, alongside their name and position, followed by their $-\log_{10}P$ value from the 274 AT analysis. Finally, the correlation coefficient (R), significance (p) and sample size (n) are given for the predictions made on the 109 diversity panel. Markers that were significantly predictive are highlighted in bold.

Marker type	Marker	Position	AT 274 $-\log_{10}P$	R	p	n
GEM	Cab002472.4	A03_007577717_007583343	6.34	0.400	<0.001	109
GEM	Cab007824.2	A10_016640393_016638608	5.18	0.363	<0.001	109
GEM	Bo3g024250.1	C03_009002289_009008999	4.98	0.439	<0.001	109
GEM	Bo3g032510.1	C03_012480257_012480710	4.82	0.338	<0.001	109
GEM	Bo4g024850.1	C04_004021498_004023941	4.45	0.444	<0.001	109
SNP	Cab017170.1:1314:T	A10_012578306_012574546	5.66	0.124	0.2	109
SNP	BnaA10g13370D:657:T	A10_012555821_012557320	5.61	0.255	0.008	109
SNP	Cab017170.1:1317:G	A10_012578306_012574546	5.54	0.124	0.2	109
SNP	Cab017372.1:381:A	A10_011480419_011479277	5.53	0.131	0.177	107
SNP	Bo3g085560.1:1067:C	C03_031724176_031725948	5.45	0.246	0.021	88
SNP	Cab002593.1:630:G	A03_008220680_008223280	5.30	0.246	0.01	109
SNP	Cab017167.1:227:C	A10_012611272_012606076	5.26	0.182	0.062	106
SNP	Cab006642.1:2565:G	A06_021380557_021384918	5.08	0.231	0.016	108
SNP	Bo2g009930.1:1786:G	C02_002200000_002203533	5.08	0.354	<0.001	98
SNP	Cab031715.1:1911:G	A07_016208430_016204916	5.00	0.147	0.138	103

The most highly associated GEM from AT analysis in the 383 diversity panel for the sum total of the essential element concentrations within the leaf gave Cab002472.4 (orthologue of *FLC*: AT5G10140). The second most highly associated GEM was another orthologue of *FLC* (Bo3g024250.1), followed by two orthologues of *SOC1* (AT2G45660.1), Bo4g024850.1 and Cab025356.1. Interestingly, the 5th most highly associated GEM from AT analysis was Bo3g032510.1, whose orthologue in *A. thaliana* is *CASP-LIKE PROTEIN 4B1* (*CASPL4B1*: AT2G38480). This gene is uncharacterised in *A. thaliana* and was therefore ordered as one of the candidates to be further tested. CASP proteins mediate the deposition of the casparian strip and it has been suggested that CASP-like proteins may play a role in membrane organisation or cell wall modification and could therefore affect multiple elements within the ionome (Roppolo et al., 2014). Following this was a GEM known as Cab032577.1. It is an orthologue of *target of early activation tagged 2* (*TOE2*: AT5G60120) which is an AP2- like gene known to have a role in the timing of flowering in *A. thaliana* (Aukerman and Sakai, 2003). Another homologue of *FLC* was the 9th GEM (BnaC02g00490D) and top marker on C2, as well as the 24th GEM (Cab007661.1). A further *SOC1* homologue was observed as the 33rd GEM (Cab003267.1). Yet another interesting candidate was observed as the 30th most highly

associated GEM; Cab001440.1 is an orthologue of *FRUITFUL* (*FUL*: AT5G60910), which has a known role in flowering and fruit development. *FUL* has been implicated in the control of meristem arrest, plant longevity and the control of the annual/perennial lifestyle in *A. thaliana* (Melzer et al., 2008; Balanzà et al., 2018). It has long been recognised within the literature that the coordinated arrest of meristem activity may be important for the correct allocation of resources to the developing seeds (Hensel et al., 1994), consistent with the observed correlation of leaf element concentration with flowering in this study (i.e. evidence of plant wide coordination of resource allocation between vegetative and reproductive development). What is more, *FUL* was shown to perform this function in part through its regulation of AP2-like genes (such as TOE1 and 3, however TOE2 which appears in the GEMs was not mentioned) (Balanzà et al., 2018) as well as having a regulatory role involving *FLC* and *SOC1* (Balanzà, Martínez-Fernández and Ferrándiz, 2014). This potentially highlights a similar or related pathway in *B. napus*, as within the pan-transcriptome the CDS gene models are being related to their orthologous genes in *A. thaliana* and may display variation in functionality. However, it was not possible to test this hypothesis as part of this study and it therefore remains mere speculation.

The SNP association peak on A10 was very narrow (27 CDS models). However, within and in the surrounding area of the association peak were a number of potential candidates which could influence total leaf nutrient concentration. None of the candidates in or around the association peak were specific flowering regulators as seen in the GEMs, however one of the candidates was known to delay flowering; Cab017149.1 is an orthologue of *Myb37* (AT5G23000.1) which when overexpressed in *A. thaliana* has been shown to delay flowering and to enhance the growth of plants, boosting seed production (Yu et al., 2016b). Furthermore like *FUL* it is also known to have a role in meristems, but instead of promoting meristem arrest, *Myb37* (alternatively known as *REGULATOR OF AXILLARY MERISTEMS1*) is involved in axillary meristem initiation and the vegetative to reproductive phase transition (Jeifetz et al., 2011; Keller, 2006). It is thought to be involved in the positive regulation of ABA signalling, but how it delays flowering and increases seed yield is unknown (Yu et al., 2016b). Being able to tightly regulate the balance between vegetative and reproductive development based on the concentration of available elements within the leaves is an attractive theory (e.g. Gol, Tomé and Von Korff, 2017; Vidal et al., 2014). Unfortunately, *Myb37* did not represent a good candidate for further investigation in this study, as previous analysis of T-DNA lines failed to produce any phenotype (thought to be a result of the functional redundancy between other members of the MYB family (Yu et al., 2016b)).

However, there was a number of other candidates in this region which could also play a role in leaf nutrient status. Firstly, close by were the orthologues of a number of well-known transporters; Cab017152.1 (orthologue of the nitrate transporter, *NRT2.3*: AT5G60780) and Cab017154.1 (orthologue of *NRT2.4*: AT5G60770) (Orsel, Krapp and Daniel-Vedele, 2002); Cab017138.1 (which is an orthologue of a predicted Na⁺/H⁺ antiporter, *CAX9*: AT5G22910) (TAIR, 2015) and BnaA10g13720D (orthologue of the magnesium transporter *MGT10*: AT5G22830) (Dynowski et al., 2008). Similarly, Cab017181.1 (an orthologue of *SIZ1*: AT5G60410) is also within this region; *SIZ1* is known for its role in P homeostasis and disruption leads to early flowering (Jin et al., 2008). These are all macronutrient transporters and it may be that the relatively large contribution of Mg, Ca and P to the sum total has skewed the AT SNP analysis. Another candidate could be Cab017179.2 which is an orthologue of *ARF4* (*AUXIN RESPONSE FACTOR 4*: AT5G60450) which has a role in leaf development in *A. thaliana* (specifically in the maintenance of leaf abaxial identity (Guan et al., 2017)) and could therefore affect the leaf ionome. However, it works redundantly with *ARF2* and *3* (Guan et al., 2017) so the conclusion that only this gene would display an association to total essential leaf nutrients is questionable.

6.3.3 Associative transcriptomic outputs, predictions and candidates: 'Leaf essential coefficient of variation'

Unlike AT analyses performed on 'total leaf essential elements', the analyses performed on the ranked 'coefficient of variation' data did not pass the Bonferroni corrected significance threshold but did pass the FDR (**Figure 6.3.3.a**). Nevertheless, AT SNP outputs gave considerably more association peaks; A1, A2, A9, A10, C5 and C8. GEM analysis once again gave orthologues of *FLC* as the most highly associated markers; however *SOC1* orthologues were no longer as highly ranked. Three of the GEMs tested were predictive to $p < 0.05$: Cab002472.4 (orthologue of *FLC*: AT5G10140.4), Bo2g002690.1 (no *A. thaliana* orthologue) and Bo9g181240.1 (no *A. thaliana* orthologue). One SNP was predictive to $p < 0.05$: Cab007376.2 (orthologue of AT5G08690.1, annotated as ATPsynthase alpha/beta family). Details are provided in **Table 6.3.3.a**.

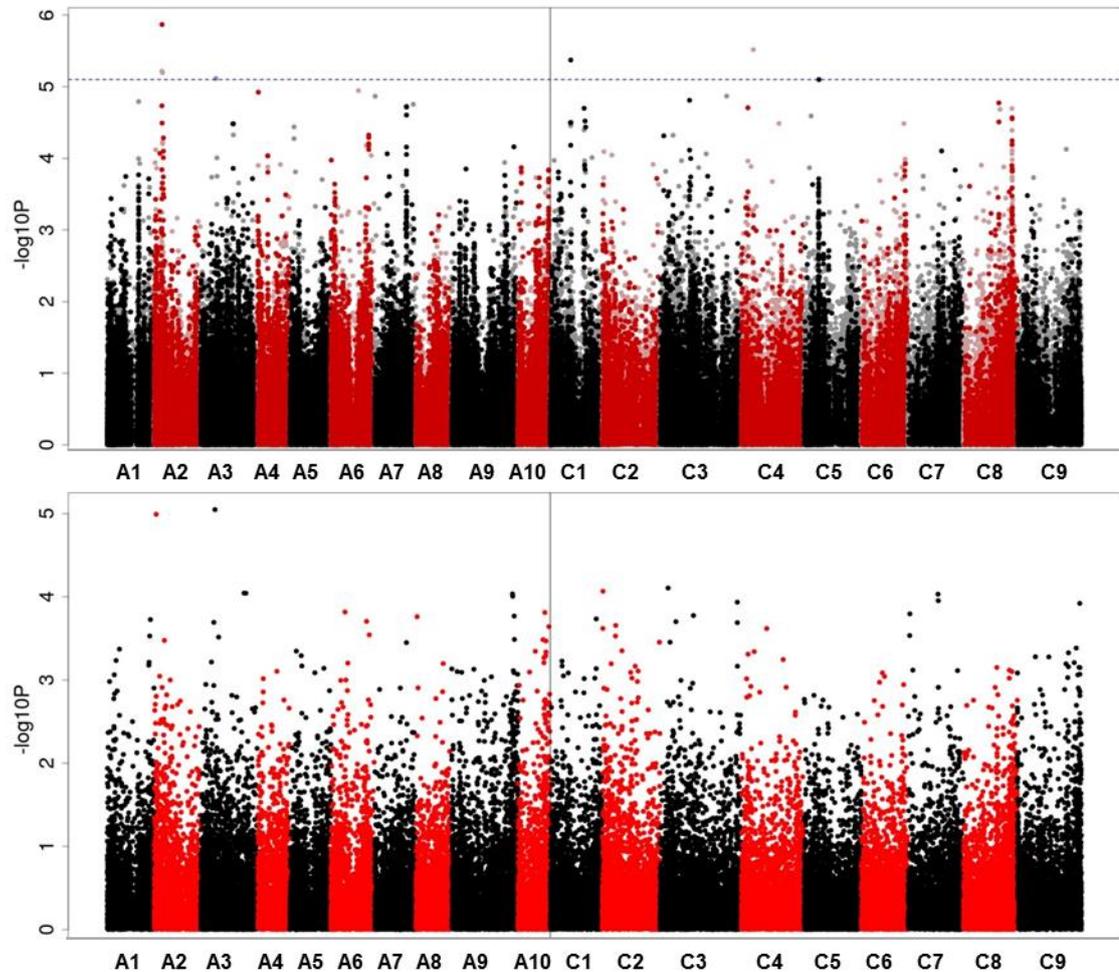


Figure 6.3.3.a Genome wide distribution of mapped markers associating with the ‘ranked leaf coefficient of variation’ for essential element concentrations.

Calculated by ranking data for each accession for all essential elements as 1-383 (1 being the accession with the lowest element concentration for the element under investigation and 383 being the highest), working out the average rank each accession displayed for all elements and dividing it by the standard deviation of ranks. It was theorised this would provide a way to distinguish the most efficient accessions from those that accumulate one or two macronutrients efficiently (i.e. those consistently ranked highly for multiple essential elements, cf ‘total essential elements’). SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and the ‘ranked correlation coefficient’ as the response variable. $-\log_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while

grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

Table 6.3.3.a Predictive capability of markers from AT analysis of ‘ranked leaf coefficient of variation’.

For assessing the predictive capability of markers the highest SNP markers from discernible association peaks and the most highly associated GEMs were analysed. The marker type is given as either SNP or GEM, alongside their name and position, followed by their $-\log_{10}P$ value from the 274 AT analysis. Finally, the correlation coefficient (R), significance (p) and sample size (n) are given for the predictions made on the 109 diversity panel. Markers that were significantly predictive are highlighted in bold.

Marker type	Marker	Position	AT 274 $-\log_{10}P$	R	p	n
GEM	Cab002472.4	A03_007577717_007583343	4.497060355	0.3788	<0.001	109
GEM	Bo2g002690.1	C02_000485370_000486065	4.013653756	0.3185	<0.001	109
GEM	Cab031056.1	A02_007296455_007298593	3.795204102	0.0392	0.6856	109
GEM	Bo9g181240.1	C09_053471789_053472822	3.701382062	0.2479	0.0093	109
GEM	Bo9g177260.1	C09_052892785_052894636	3.689934313	0.1084	0.2617	109
SNP	Cab032396.1:450:T	A02_004363480_004364717	5.556603452	0.0666	0.5337	109
SNP	Cab032382.1:1137:A	A02_004310030_004311540	4.769262946	0.0489	0.7099	109
SNP	Cab022984.1:234:G	A06_003072958_003073652	5.600623819	-0.18	0.0607	109
SNP	BnaA06g17070D:210:G	A06_011029175_011028017	5.574338653	-0.072	0.0641	109
SNP	Cab035375.2:1641:G	A06_025630994_025633924	5.057056486	0.178	0.455	109
SNP	Cab029652.2:2109:G	A10_002026337_002023360	3.740976841	0.308	0.2813	109
SNP	Cab007376.2:1365:G	A10_017704336_017706873	3.603783551	0.1041	0.0011	109
SNP	BnaA10g13370D:657:T	A10_012555821_012557320	3.415509523	0.0603	0.4914	109
SNP	Cab007263.1:528:G	A10_019369002_019366965	3.096844499	0.1425	0.1617	98
SNP	Bo6rg122160.1:546:T	C06_038967836_038969745	4.445014665	-0.036	0.6133	109

The narrow association peak on A1 yielded an obvious floral candidate; Cab009802.1 which is the orthologue of *VERNALIZATION INDEPENDENCE 5* (*VIP5*: AT1G61040.1). *VIP5* is closely related to a subunit of the yeast PAF1 complex (Oh et al., 2004); a transcription elongation factor which has a role in transcriptional regulation via the promotion of chromatin remodelling factors (e.g. methyltransferases) (Krogan et al., 2003). *In planta*, when *VIP* loss of function mutants are analysed they are known to flower earlier, displaying reduced methylation of *FLC* chromatin (specifically H3K4me3, an ‘active histone mark’ (Schneider et al., 2004)) which results in the silencing of *FLC* (He, Doyle and Amasino, 2004; Lu et al., 2017; Oh et al., 2004). However, why only *VIP5* would show an association and why this gene would occur in this analysis and not the specific flowering time analysis (see next section, 6.3.4) raises doubts as to its significance. Perhaps *VIP5* responds to nutrient status to help regulate the time

to flowering, which would be missed if flowering time were being analysed under sufficient nutrient and vernalisation conditions. Alternatively, perhaps the vernalisation requirement of plants within the diversity panel was not met. Thus due to the presence of spring and winter ecotypes in the diversity panel, when the coefficient of variation is analysed *VIP5* is observed as it is involved in how plants perceive cold treatment. The SNP associations in the coefficient of variation may be more variable than the 'total' as the skewing effect of macronutrients is absent. Other candidates in the association peak on A1 included; Cab004827.1/ Cab009793.1 which are orthologues of *GA20OX3*: AT5G07200, one of the five GA20 oxidases in *A. thaliana* involved in Gibberellic acid (GA) biosynthesis (which has a known role in regulating flowering through promotion of flowering time integrators e.g. *FLOWERING LOCUS T* and *TWIN SISTER OF FT*)(Galvao et al., 2012). However, why only one GA20 oxidase would show an association is uncertain.

Interestingly, on the A2 association peak an obvious floral homeotic gene was observed; Cab032392.2/ Cab047347.1 are orthologues of *PISTILLATA (PI)*: AT5G20240, which is known to play a role in floral organ patterning (which is dependent upon GA) (Yu et al., 2004; Plackett et al., 2017). Perhaps *GA20OX3* in the A1 association peak and *PI* in the A2 association peak indicates that early floral development is affected by leaf nutrient status. Alternatively, this could once again be a result of the skewing effects of spring and winter ecotypes within the diversity panel. No specific floral regulators was observed within the A2 association peak. A couple of nutrient specific candidates were found, however these were both specific to P, e.g. Cab032387.1 (orthologue of *SPX1*: AT5G20150) (Puga et al., 2014) and Cab032401.1 (*PHOSPHATE TRANSPORTER 4;5*: AT5G20380) (Guo et al., 2008)) and therefore not thought likely to play a role in multiple nutrients (cf. Flowering candidate gene analysis, 6.3.6).

The association peak on A9 had three obvious candidates. The first was Cab040282, which is an orthologue of *FRY1 (FIERY1)*: AT5G63980, which when T-DNA mutants were assessed in *A. thaliana* had delayed flowering under long and short day conditions (thought to act via the photomorphogenic signalling pathways)(Kim and Von Arnim, 2009). Interestingly *fry1* had a higher number of leaves at bolting in comparison to the control only under long day conditions, leading the authors of the previous study to suggest that late flowering under short day conditions in this mutant could be due to slower leaf initiation (Kim and Von Arnim, 2009). Furthermore, *FRY1* has been linked to lateral root initiation (Chen and Xiong, 2010; Hirsch et al., 2011). This may be consistent with the current observation of leaf nutrient concentration being tied to floral induction. However, it may also fit with the contrasting theory: that it is due to variation in developmental stages between the spring and winter ecotypes of the panel. The

link to *FLC* and other flowering time related candidates in the current study could have arisen as a consequence of the different vernalisation requirements of plants in the diversity panel. Similarly, *FRY1* is known to be involved in abiotic stress response, perhaps exposure of the spring ecotypes to cold conditions induced a stress response (Guzy-Wrobelska et al., 2013). Similar to *FRY1*, the other candidate observed in the A9 association peak was Cab040229, whose orthologue (*GRF8*: AT5G65430, annotated as *14-3-3 PROTEIN G-BOX FACTOR14 KAPPA*) has connections to the cold response through brassinosteroid signalling (which could link to the spring/winter ecotype association)(Gampala et al., 2007). The final candidate was Cab040266, whose orthologue in *A. thaliana* is *XND1* (*XYLEM NAC DOMAIN 1*: AT5G64530); it is involved in xylem differentiation and could therefore have a major impact on the leaf ionome (Zhao et al., 2008, 2017). However, once again for both *XND1* and *GRF8*, why these would show an association and no other genes within their relative pathways is unexplained.

The small association peak on A10 yielded one candidate: Cab007268 (orthologue of *EARLY FLOWERING 6*, *ELF6* : AT5G04240). *ELF6* T-DNA mutants have an early flowering phenotype under long and short day conditions (Noh et al., 2004). It is thought to act as part of the photoperiod pathway (Noh et al., 2004), modulating brassinosteroid regulated gene expression (Yu et al., 2008) and interacts directly with the *FT* transcription initiation region (Jeong et al., 2009). It has H3K27me3 demethylase activity, which is thought to play a pivotal role in preventing the inheritance of epigenetic vernalisation signals at the *FLC* locus (Crevillén et al., 2014). More recently, it has been suggested to work with SDG8 (a H3K36me3 methyltransferase) to coordinate transcriptional activation and de-repression of *FLC* through these chromatin modifications (*ELF6* removal of the silencing mark H3K27me3 and SDG8 addition of activation mark H3K36me3)(Yang, Howard and Dean, 2016). Once again, this may highlight the differences in *FLC* expression between spring and winter ecotypes. No nutrient specific candidates could be found.

The final SNP association peak analysed was on C5. This contained two candidates which could affect flowering time: Bo5g025070 (orthologue of *Receptor for Activated Kinase 1*, *RACK1A*: AT1G18080) and Bo5g025450 (orthologue of *ARP4*: AT1G18450). When *RACK1A* was investigated in *A. thaliana* T-DNA lines it was found to display multiple phenotypes, including late flowering, shorter primary roots (with decreased lateral roots) and reduced rosette leaf production (again linking flowering and leaf development)(Chen et al., 2006; Guo and Chen, 2008; Urano et al., 2015). *ARP4* RNA interference lines displayed early flowering, delayed senescence of flowers and partial sterility (T-DNA lines did not effectively control gene expression) (Kandasamy et al., 2005). It has been suggested that this acts as part of the

NuA4/SWR1-C chromatin remodelling complexes, leading to decreased *FLC* expression and early flowering through acetylation of nucleosomal histones (Bieluszewski et al., 2015). Once again, however, it is unknown why these candidates would occur only in this particular AT analysis and why only these specific genes were observed from their respective pathways.

Considering the large number of chromatin remodelling factors observed as part of the SNP association analysis (albeit from multiple pathways) it was hoped that the GEMs would continue to help elucidate this pattern. Interestingly, in these analyses *SOC1* CDS gene models were eliminated from the top GEM results (c.f. total nutrient and flowering time), with the most highly associated CDS gene model for *SOC1* ranked at 561st (when associations are scored from most closely associated to least, 1st being the most, 53890th being the lowest), whilst Cab002472.4 (orthologue of *FLC*) maintained its position as the most highly associated marker. This may reflect the important role that chromatin remodelling plays in the regulation of *FLC* expression (He, 2012). However, chromatin remodelling has also been implicated in the control of *SOC1* expression; interestingly in *A. thaliana* this is reportedly through a Nuclear Factor Y complex, which binds to the *SOC1* promoter and demethylates H3K27 via *RELATIVE OF EARLY FLOWERING 6 (REF6)* (Hou et al., 2014), which in turn is homologous to *ELF6* identified within SNP association analyses (although they have divergent roles in floral regulation) (Noh et al., 2004). Once again, it may be that the observations of the current analysis merely reflect the differences between spring and winter ecotypes (since splitting based on actual flowering time data was able to remove the association to *FLC*, 6.3.1). *FLC* represented the only obvious floral candidate within GEM analysis and no broad reaching leaf nutrient candidates were observed.

6.3.4 Associative transcriptomic outputs, predictions and candidates: Flowering time

Although the flowering time data was not generated as part of the current research, it was felt necessary to analyse it with AT to highlight differences between ‘total leaf essential elements’ (see 6.3.2) and Coefficient of Variation (see 6.3.3). Once again, neither the SNP nor the GEM AT analyses passed the Bonferroni or FDR thresholds (**Figure 6.3.4.a**), cf. total essential elements, 6.3.2. However, within the SNP association analysis a number of association peaks were observed; A2, A3, A6 and C1. These were not observed in either AT SNP analyses of ‘total essential elements’ or the ‘ranked coefficient of variation’. Furthermore, only one SNP marker could be found that was predictive ($p < 0.05$) although this was not found to be predictive for flowering time but rather for total essential elements and the coefficient of variation (**Table 6.3.4.a**). GEMs showed more similarity; once again Cab002472.4 was the most highly associated CDS gene model (*FLC*: AT5G10140.4) as well as the third highest (Bo3g024250.1), however Cab003267.1 (orthologue of *SOC1*: AT2G45660.1) was lower down in 12th and 17th rank (Bo4g024850.1). A candidate whose orthologue in *A. thaliana* is described as *E2F target gene 1* (*E2F*: AT2G40550.1) was repeatedly one of the most highly associated GEMs, in 2nd (BnaC09g21690D), 5th (BnaC07g31190D) and 7th (Bo2g158420.1) place. There is little description of function for *E2F* within the literature and since it had also been observed but overlooked in S leaf concentration AT analysis, it was selected for further investigation with *A. thaliana* T-DNA insert lines. As with the total leaf essential GEMs (see 6.3.2), Cab032577.1 (orthologue of *TOE2*: AT5G60120.2) was once again amongst the most highly associated GEMs and on A3 Cab001440.1 (orthologue of *FUL*: AT5G60910.1) was the fourth most highly associated GEM (after Cab002472.4 (orthologue of *FLC*), Cab003267.1 (orthologue of *SOC1*) and Cab016313.1 (unannotated in *A. thaliana*)). It is interesting to note that the leaf essential elements GEMs seem to give better flowering outputs than the flowering time data itself.

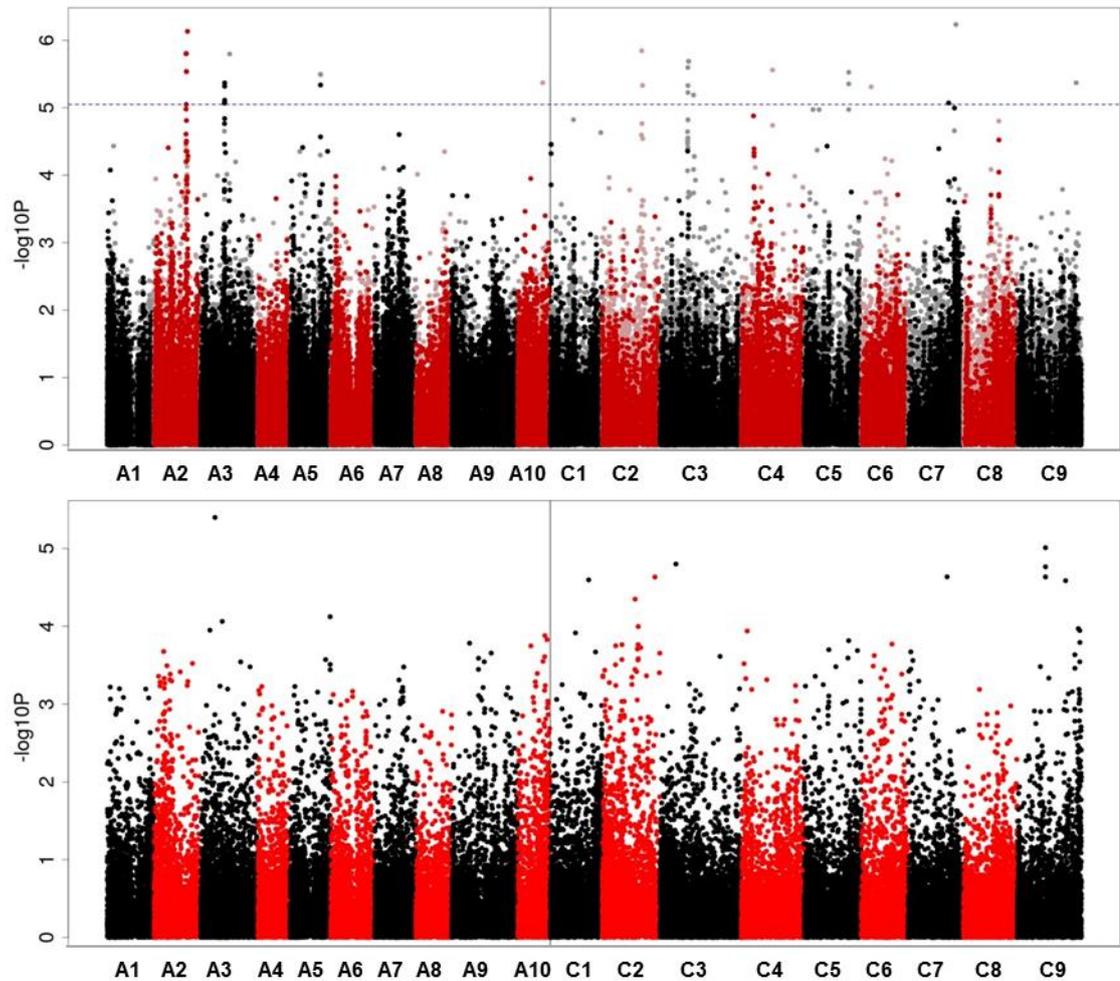


Figure 6.3.4.a Genome wide distribution of mapped markers associating with the days to flowering; calculated from the average of four plants.

SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014a). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and days to flowering as the response variable. $-\log_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

Table 6.3.4.a Predictive capability of markers from flowering time (as days) AT analysis.

For assessing the predictive capability of markers the highest SNP markers from discernible association peaks and the most highly associated GEMs were analysed. The marker type is given as either SNP or GEM, alongside their name and position, followed by their $-\log_{10}P$ value from the 274 AT analysis. Finally, the correlation coefficient (R), significance ($p < 0.001^{***}$, $p < 0.01^{**}$, $p < 0.05^*$) and sample size (n) are given for the predictions made on the 109 diversity panel for flowering, total essential elements and the ranked coefficient of variation (predictions made using flowering data and then tested for predictive capacity on the other trait datasets).

Marker type	Marker	Position	AT 274 - $\log_{10}P$	Flowering (R)	Total essential (R)	CV (R)	n
GEM	BnaC07g31190 D	C07_038545382_038544355.001	3.71	0.5955***	-0.3602***	0.2596**	98
GEM	BnaC09g21690 D	C09_022908323_022904991.001	3.81	-0.642***	0.3831***	-0.2583*	98
GEM	BnaC09g21700 D	C09_022910910_022912289	3.84	0.6035***	-0.3758***	0.2341*	98
GEM	BnaC09g37150 D	C09_044987850_044986184.001	3.67	0.5906***	-0.3957***	0.2782**	98
GEM	Bo1g108520.1	C01_033465858_033467740	3.55	0.6121***	-0.3987***	0.2555*	98
GEM	Bo2g158420.1	C02_049205683_049207550	3.51	0.6218***	-0.3402***	0.2175*	98
GEM	Bo3g024250.1	C03_009002289_009008999	4.26	-0.5633***	0.4437***	-0.3594***	98
GEM	Bo9g075990.1	C09_022909694_022910201	4.22	0.5513***	-0.3576***	0.2217**	98
GEM	Cab002472.4	A03_007577717_007583343	4.91	0.5731***	-0.4248***	0.4095***	98
SNP	Bo1g054960.1:1297:G	C01_015964461_015967118	5.21	0.158	-0.074	-0.013	98
SNP	Bo1g056630.1:1353:A	C01_016502800_016505838	3.35	0.193	-0.168	0.083	98
SNP	Bo1g056630.1:1578:G	C01_016502800_016505838	3.52	0.154	-0.08	0.054	98
SNP	Bo3g018870.1:987:A	C03_006367555_006370743	4.26	0.145	-0.05	0.067	98
SNP	Bo7g110810.1:687:C	C07_044028248_044031823	7.02	0.088	-0.015	0.041	98
SNP	Cab016457.1:1101:T	A03_005454287_005456021	4.48	0.124	-0.03	0.078	98
SNP	Cab016461.1:2295:T	A03_005469019_005473965	4.19	0.165	-0.234*	0.327**	94
SNP	Cab038703.2:1014:G	A02_022451577_022453059	4.97	0.182	0.056	0.061	96

The SNP association peak on A2 yielded no obvious floral regulators. Two senescence candidates were found which could be interesting from the perspective of flowering and leaf elements. The first was Cab038797.2, which is an orthologue of *senescence associated gene 12* (*SAG12*: AT5G45890), a cysteine protease only expressed in senescing tissues and regulated specifically by senescence pathways (Noh and Amasino, 1999). The second was Cab038798.1,

which is an orthologue of *AUTOPHAGY RELATED 7* (*ATG7*: AT5G45900) in *A. thaliana*. It is one of the rate limiting components of autophagosome formation and flux (Minina et al., 2018) and it is also hypersensitive to C and N deprivation (TAIR, (2015) and Ren, Liu and Gong, (2014)). As senescence is important for remobilising nutrients from old leaves during flowering and seed development, these candidates could represent potential links between flowering and leaf element status observed in this study. Other candidates in the region were Cab038779.1 (whose orthologue in *A. thaliana* is involved in defence and cold response, AT5G17890 (Yang et al., 2010)) and Cab038716.1 (orthologue of *CASPL-LIKE PROTEIN 1B1*: AT5G44550.1) cf. total essential elements *CASPL 4B*, 6.3.2.

As with the SNP association peak on A2, no flowering specific candidates could be found on the SNP A3/C3 association peak within the pan-transcriptome. One of the potential general candidates identified in this region includes Cab003487.1, which is an orthologue of *TUMOROUS SHOOT DEVELOPMENT 1* (*TSD1*: AT5G49720) and is involved in cellulose biosynthesis. It is therefore involved in shoot system development and produces abnormal flowers (Nicol et al., 1998). Another candidate which could affect flowering is Cab003507.1 (orthologue of *NRT 3.1*: AT5G50200.1, involved in high affinity nitrate transport) (Okamoto et al., 2006). *NRT3.1* is thought to act as part of a two component high affinity nitrate transporter system (involving *NRT2.1*: AT1G08090) (Yong, Kotur and Glass, 2010), therefore it is questionable why only this gene may show an association.

Once again, multiple potential candidates were found on the A6 SNP association peak; however, links to flowering were relatively tangential. The candidate closest to the most highly associated markers is Cab023002.1, its orthologue in *A. thaliana*, *YAB2* (*YABBY2*: AT1G08465), is involved in abaxial cell fate (Siegfried et al., 1999). It has been shown that *YAB2* is regulated through the same *SWI/SNF* chromatin remodelling complex (via *SWP73*) that is known to also regulate flowering time and floral development (i.e. another link between vegetative and floral development)(Sacharowski et al., 2015). However, once again, why only *YAB2* (and not *YAB3* or 5 etc.) would show an association is unknown. Another potential candidate would be Cab023090.1 (orthologue of *HYL1*: AT1G09700). *HYL1* is a dsRNA binding protein (Lu, 2000), which acts as part of the plant Microprocessor-Dicing complex alongside Dicer-like 1 (*DCL1*) to convert long pri-miRNA into miRNA (downstream of chromatin remodelling factor 2 (*CHR2*) and Serrate (*SE*))(Wang et al., 2018; Rogers and Chen, 2013). These miRNAs are small, non-coding RNAs which are involved in posttranscriptional regulation (Voinnet, 2009). As such, the *HYL1* T-DNA insert lines display a broad array of phenotypes including leaf hyponasty, delayed flowering and slower root growth (Lu, 2000; Han et al., 2004). Once again, however, why only

this specific aspect of the miRNA processing machinery would be associating with flowering time is unknown.

The final SNP association peak to be analysed was C1 and this gave a candidate which could affect both flowering and leaf nutrient status: Bo1g003130 (orthologue of *HB16*: AT4G40060). *HB16* thought to be involved in regulating leaf cell expansion and the photoperiodic control of flowering (potentially through blue light signalling (Wang et al., 2003)). This could therefore affect the vegetative to reproductive phase transition and may thus link leaf nutrient status to flowering time. This fits with the GEM data from total leaf nutrient analysis but fails to link to either the SNP or GEM analysis for flowering.

6.3.5 Summary of flowering time related associative transcriptomic outputs, predictions and candidates

From both AT analysis (of leaf nutrient concentrations and flowering time) and the background literature, there is no doubt that the vegetative to reproductive phase transition is closely related to plant nutrient status. The key question raised by the current research is whether the leaf nutrient concentration of plants under stress free conditions prior to the floral transition (as leaves were sampled for RNA before vernalisation and floral induction) is actively promoting/repressing the floral transition or if the association to flowering time is a mere consequence of developmental differences within and between spring/winter ecotypes (at the time of sampling for leaf nutrient analysis). Of all the flowering candidates highlighted by AT analysis only three were taken forward (summarised in **Table 6.3.5.a**). Analysing these flowering candidates in *A. thaliana* was not necessarily the most practical approach; the Columbia background for the T-DNA insert lines has minimal *FLC* expression as it does not require vernalisation (i.e. it is a 'spring' ecotype). Nevertheless, it was decided that for some of the many flowering genes discussed as part of AT analysis, these three would be analysed with *A. thaliana* insert mutants under long and short day conditions. This would allow a relatively quick analysis to complement work on the leaf ionome timeline (see 6.3.7).

Table 6.3.5.a A list of the flowering candidate genes taken forward for further study.

Details of the original AT trait analysis the candidate were found for, its marker within the pan-transcriptome, AGI code, description in *A. thaliana*, line ordered from NASC and other potential element interactions are listed.

Trait	Marker	AGI	Description	NASC name	Interaction?
S leaf/Flowering time/Total essential elements	Bo2g158420.1/ BnaC07g31190 D/BnaC09g2169 0D	AT2G40550.1	<i>E2F-target gene 1</i>	SALK_07 6472C	Flowering/Total leaf essential elements?
				SALK_05 5089C	
Cd leaf/ Cu leaf/Mn leaf/Mo leaf/S leaf/Flowering time/Total essential elements	Bo4g024850.1/ Cab025356.1/C ab003267.1	AT2G45660.1	<i>SOC1</i>	SALK_00 6054C	Flowering/Total leaf essential elements?
				SALK_13 8131C	
Total essential elements	Bo3g032510.1	AT2G38480.1	<i>CASP-LIKE PROTEIN 4B1</i>	SALK_02 2304C	Flowering/Total leaf essential elements?
				SALK_06 4085C	

6.3.6 Flowering candidate gene analysis

The *A. thaliana* T-DNA lines for the flowering candidates were grown under long (LD) and short day (SD) and compared to wildtype. For 'total essential elements' analysis only concentrations from elements that were essential, had greater than 85% recovery, and had an average greater than the LOD, were included (i.e. B, Mg, P, S, K, Ca, Mn, Cu, Zn and Mo were included, whilst Fe and Ni were excluded). One of the *soc1* lines failed all attempts at genotyping (SALK_138131C). For 'total essential elements' no significant difference could be found between any of the lines when the post hoc Bonferroni test was applied (ANOVA: F: 2.09, df 11,59, p 0.036, Bonferroni p<0.05). There does, however, appear to be a pattern within the data (**Figure 6.3.6.a**); LD plants appear to have slightly lower totals in comparison to their SD equivalents, but this difference is not large enough to be significant within the post hoc analyses. This could be a result of the initial growth conditions; all plants were grown initially under the same conditions before being moved to LD/SD. This was to allow enough leaf biomass to be produced for accurate weights and digestions for ICP-MS analysis. Perhaps plants were not moved in time, minimising the differences that LD and SD conditions would cause in the floral pathway. However, when individual element concentrations were assessed to ensure that one or two macronutrients were not masking a more general effect some significant differences were observed (see Appendix 20). Of all the individual elements, only P concentrations showed a consistent pattern between LD and SD (**Figure 6.3.6.b**). All insert lines showed significant differences between plants grown under LD and SD conditions. However only the *casp1* lines were significantly different to wildtype under LD conditions.

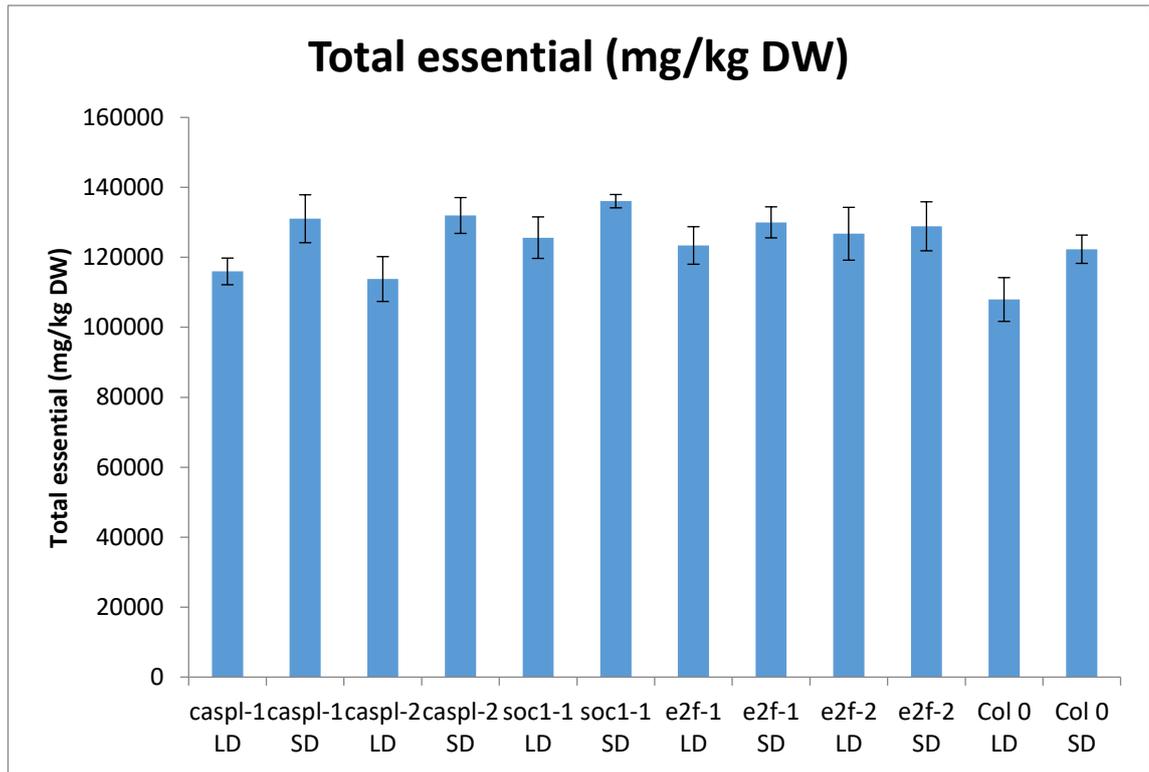


Figure 6.3.6.a Candidate gene analysis of flowering candidates against ‘total essential elements’ in leaf tissues (as mg/kg DW) of *A. thaliana* insert mutants.

Wildtype: *A. thaliana* Col 0; *caspl-1* (At2g38480/SALK_022304C), *caspl-2*(At2g38480/SALK_064085C), *soc1-1* (At2g45660 /SALK_006054C), *e2f-1*(At2g40550/SALK_076472C) and *e2f-2*

(At2g40550/SALK_055089C). Lines were grown under either long day (LD) or short day (SD) conditions.

The mean and standard error are displayed for each line, n = 6 for all lines except the Col 0 SD which had 5 (where n=1 is an individual plant). No significant differences were observed between any of the lines or growth conditions (ANOVA: F: 2.09, df 11, 59, p 0.036, Bonferroni p<0.05).

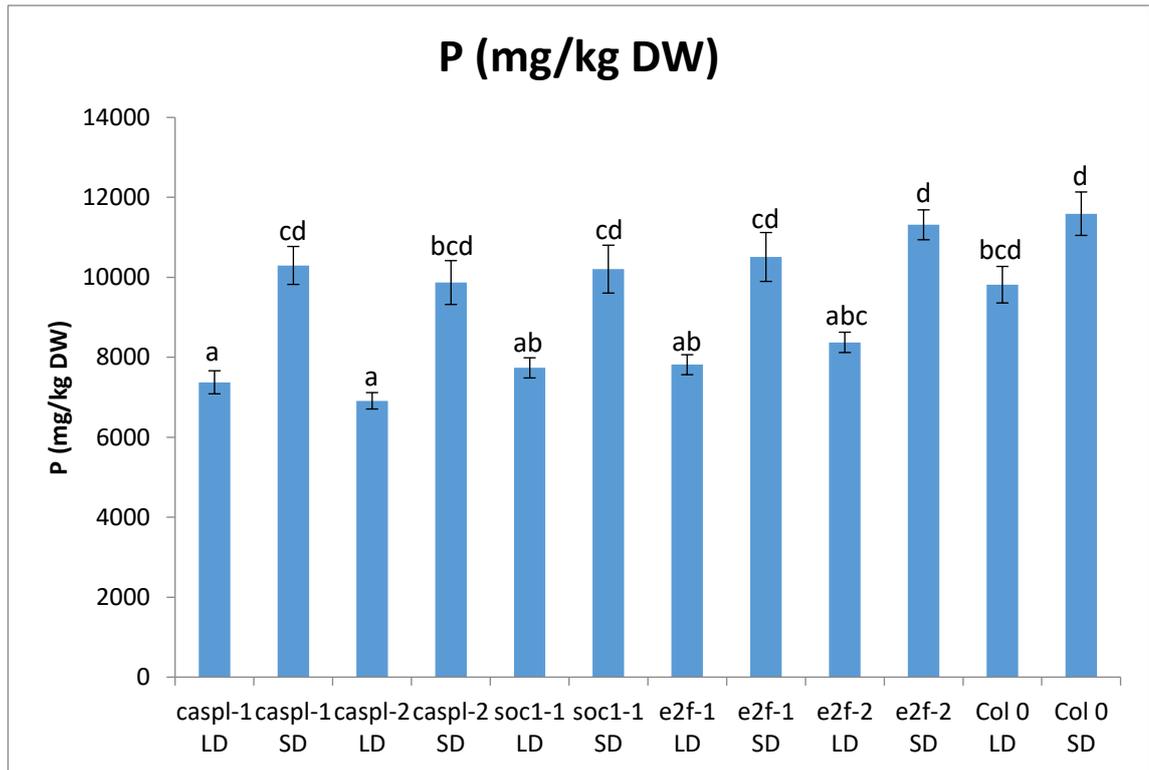


Figure 6.3.6.b Candidate gene analysis of flowering candidates against P in leaf tissues (as mg/kg DW) of *A. thaliana* insert mutants

Wildtype: *A. thaliana* Col 0; *caspl-1* (At2g38480/SALK_022304C), *caspl-2*(At2g38480/SALK_064085C), *soc1-1* (At2g45660 /SALK_006054C), *e2f-1*(At2g40550/SALK_076472C) and *e2f-2*

(At2g40550/SALK_055089C). Lines were grown under either long day (LD) or short day (SD) conditions.

The mean and standard error are displayed for each line, n = 6 for all lines except the Col 0 SD which had 5 (where n= 1 is an individual plant). Significant differences were observed between all lines to their counterparts grown under different day length conditions except the wildtype control. However only the *caspl* lines under LD showed a significant difference to the wildtype control under the same growth conditions. Significant differences are indicated where letters are not shared. ANOVA: F 14.01, df 11, 59, p<0.001, Bonferroni p<0.05.

6.3.7 Leaf ionome timeline

The first aspect of the leaf ionome timeline which needed to be considered was whether a significant difference was observed in the development of spring and winter ecotypes under the different vernalisation conditions. This was done simply by recording what day each leaf was sampled for each developmental time-point (**Figure 6.3.7.a**). There was no significant difference between spring and winter ecotypes when sampled for their 2nd true leaf (prior to any vernalisation treatment), but once plants were sampled for the 5th true leaf significant differences were observed. At the 5th true leaf under non-vernalised conditions there was a significant difference in when plants achieved this development stage, with winter ecotypes taking significantly longer than the spring ecotypes. However, when the same accessions were compared for plants which had been vernalised, there was no significant difference at this developmental time-point. The vernalised plants took longer than their non-vernalised counterparts in general due to developmental delays caused by vernalisation treatment (i.e. cold conditions slowed the development of spring and winter ecotypes alike). By the 7th true leaf sampling, the same relationship between spring and winter ecotypes was observed under non-vernalised conditions. However, the winter ecotypes which had been vernalised were taking longer than their spring counterparts to reach the next developmental stage at the 7th sampling point.

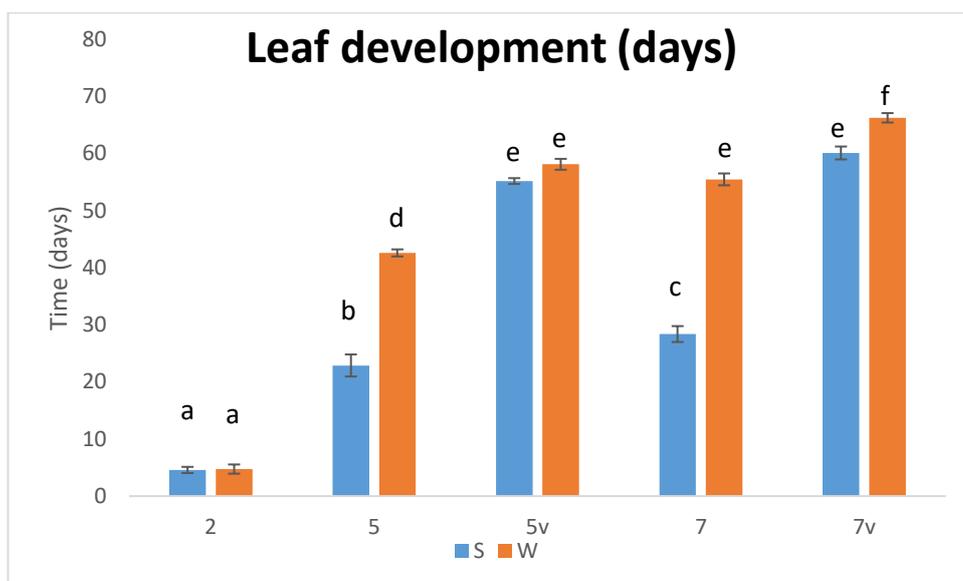


Figure 6.3.7.a leaf development timeline (as days from sowing).

The time it took from when plants were sown (day 0) to when they were sampled for a leaf at different developmental stages was recorded for winter (W) and spring (S) ecotypes: 2 (number of days before the 2nd true leaf was sampled), 5 (number of days before the 5th true leaf was sampled), 5V (number of days before the 5th true leaf was sampled from plants which had been vernalised), 7 (number of days before the 7th true leaf was sampled) and 7V (number of days before the 7th true leaf was sampled from plants which had been vernalised). Leaves were sampled from 5 separate trays for the same 6 winter and spring accessions (four replicate plants per accession, each sampled for one leaf). ANOVA: F 477.97, df 9, 50, $p < 0.001$, Bonferroni $p < 0.05$, significantly different means are indicated where letters are not shared.

After confirming that the lines were displaying developmental differences between spring and winter ecotypes under the different vernalisation treatments, the leaf ionome could be assessed under these conditions (**Figure 6.3.7.b**). Once again, only the essential elements which passed 85% recovery and had an average greater than the LOD were included (i.e. B, Mg, P, S, K, Ca, Mn, Zn and Mo included, Cu, Fe and Ni excluded). At the 2nd true leaf there was no significant difference between the essential elements of the ionome of spring and winter ecotypes. However, by the time the 5th true leaf was sampled there was a significant difference between the spring and winter ecotypes, with the spring ecotypes having significantly higher elemental concentrations in comparison to the winter ecotypes under non-vernalised conditions. This is in contrast to the vernalised plants which displayed no significant difference in leaf ionome at this time point or the next (7th). By the 7th true leaf however the difference

between the spring and winter ecotypes of the non-vernalised plants was not maintained. It is thought this is because by this point in time the plants were all becoming stressed by the cramped growth conditions; vernalised plants got potted on into bigger pots to be included as part of the high S seed (HS) and low S seed (LS)/pod/stem experiment, see 4.2.5.1. There was no significant difference between HS and LS lines in this experiment within time points for total essential elements (ANOVA: F 29.18, df 9,50, $p < 0.001$, Bonferroni $p < 0.05$, shared letter is no significant difference: A) 5 (H and LS), 7 (H and LS); B) 2 (H and LS), 5v (H and LS) and 7v (H and LS)).

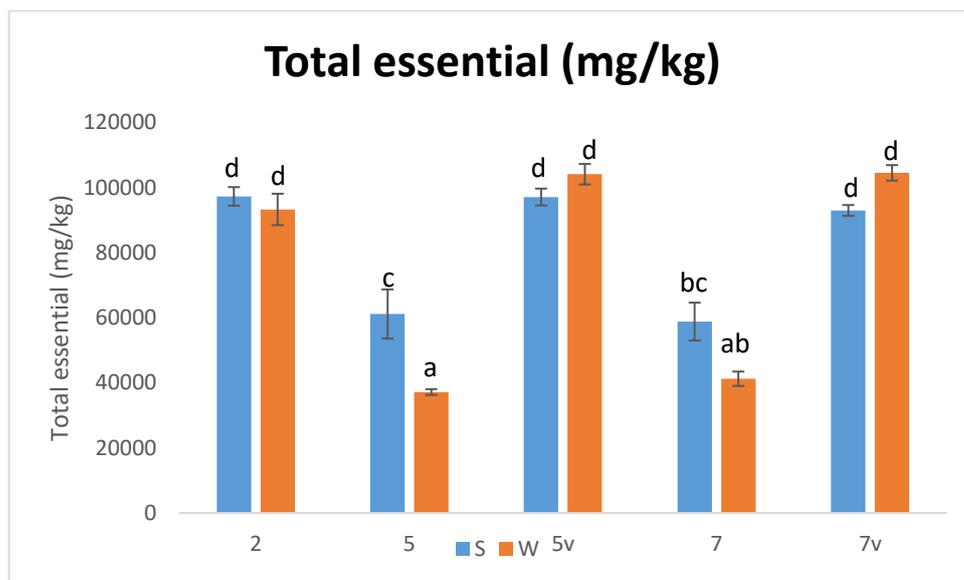


Figure 6.3.7.b Total essential elements (mg/kg) within leaves at different developmental and vernalisation states.

The difference in leaf ionome between spring (S) and winter (W) ecotypes was measured at: the 2nd true leaf (2) prior to vernalisation; the fifth true leaf with (5V) and without vernalisation (5); and the 7th true leaf with (7V) and without vernalisation (7). Leaves were sampled from 5 separate trays for the same 6 winter and spring accessions (four replicate plants per accession, each sampled for one leaf). ANOVA: F: 46.65, df 9,50, $p < 0.001$, Bonferroni $p < 0.05$, significantly different means are indicated where letters are not shared.

Individual essential element concentrations were investigated to see if the patterns observed in the 'total' were a consequence of one or two elements, or if there was a difference between elements known to be more mobile within the plants, i.e. whether they would be able to remobilise the elements from the leaves during the senescence associated with the floral transition (

Table 6.3.7.a). For this analysis, elements were split based upon previous research where the Apparent Nutrient Remobilisation (ANR%, expressed as the maximum % of nutrient content

during leaf senescence, 0% as no remobilisation to -100%, full remobilisation) was calculated for a number of species, including *B. napus* (Maillard et al., 2015). It was important to find data specific to *B. napus* as the mobility of nutrients is species dependent (Fernández and Brown, 2013). For the purposes of this study, an ANR of 0% classed the nutrient as immobile (i.e. B, Ca, Mn, Mo, Zn), while anything below 0% was classed as mobile (i.e. K, Mg, P and S) (Maillard et al., 2015). The element concentrations in each of these groups was then summed to give an indication as to what was happening in general within the groups (i.e. immobile and mobile elements generally). From this analysis a difference was found between how the winter and spring ecotypes appear to be moving the mobile nutrients around; with there being higher concentrations of the more mobile nutrients in the leaves of the older (5/7 leaf) treatments under the non-vernalised conditions within the spring ecotypes (Table 6.3.7.a).

Table 6.3.7.a ANOVA results for individual elements, as well as those from elements which have been grouped/summed according to their Apparent Nutrient Remobilisation (ANR%) in *B. napus* from (Maillard et al., 2015) across vernalisation and development stages.

Elements are coloured according their ANR: red – immobile (B, Ca, Mn, Mo, Zn); blue – mobile (K, Mg, P, S). The F statistic (F), degrees of Freedom (df) and p value (p) are given, alongside the differences highlighted with post hoc Bonferroni analysis ($p < 0.05$) between the different time points (2, 5, 7), vernalisation treatments (indicated by a v) and ecotype (Spring –S and Winter – W). Where the means are significantly different the letters are not shared, for ease of viewing some differences are highlighted in yellow (where there is a significant difference between S and W ecotypes within a treatment). Lettering starts from the smallest mean in the group under analysis, i.e. the group/s with the lowest average will be lettered ‘a’ and the remaining means labelled in ascending order.

ANOVA	F	df	p	Bonferroni ($p < 0.05$)									
				2_S	2_W	5_S	5_W	5v_S	5v_W	7_S	7_W	7v_S	7v_W
B	12.68	9,50	<0.001	bc	c	ab	a	bc	bc	ab	a	bc	c
Ca	4.61	9,50	<0.001	bc	c	abc	ab	abc	abc	abc	a	ab	abc
Mn	26	9,50	<0.001	cd	d	b	a	b	b	b	a	bc	b
Mo	7.71	9,50	<0.001	c	abc	abc	a	ab	a	bc	a	ab	ab
Zn	35.71	9,50	<0.001	b	b	a	a	b	b	a	a	b	b
K	61.8	9,50	<0.001	bc	b	a	a	cde	de	a	a	bcd	e
Mg	23.75	9,50	<0.001	b	b	b	a	a	b	a	a	b	b
P	30.27	9,50	<0.001	def	cde	bc	a	efg	fg	cd	ab	g	efg
S	21.5	9,50	<0.001	c	c	b	a	b	b	ab	a	b	b
Immobile	4.99	9,50	<0.001	bc	c	abc	ab	abc	abc	abc	a	ab	abc
Mobile	59.32	9,50	<0.001	c	c	b	a	c	c	b	ab	c	c
TOTAL	45.65	9,50	<0.001	d	d	c	a	d	d	bc	ab	d	d

Both the results of the 'total' and 'mobile' elements fit with what was observed as part of the wider RIPR panel: plants which were closer to flowering had a higher concentration of elements within their leaves (i.e. there was a negative correlation between time to flowering and leaf nutrient status, see **Table 5.3.3.a**). It may be that there is an accumulation of the elements which can be remobilised more easily prior to flowering within the leaves, acting as a reservoir of elements. The question is whether this reservoir plays any role in actively promoting flowering. It is well known that nutrient deficient plants will flower early (i.e. stress induced flowering), however the current question is whether there is a link between leaf nutrient status under sufficient conditions and the timing of flowering. Interestingly, previous research on nitrate in *A. thaliana* suggests that flowering time is modulated by nitrate in a mechanism that works in tandem with the autonomous, gibberellin and photoperiod pathways, whilst also being repressed by *FLC* (Marín et al., 2011). This is consistent with the candidates observed as part of AT analysis; *SOC1*, *TOE2* (*AP2 like*) and *FUL* are all part of the age-related flowering pathway (Wang, 2014; Wu et al., 2009; Wang, Czech and Weigel, 2009; Fornara and Coupland, 2009), working independently and in parallel to the other flowering pathways, while also being capable of repression via *FLC* (Marín et al., 2011). Perhaps this interaction occurs via *SOC1* which was the only single knock out mutant in the study of nitrate not to flower earlier with low nitrate (glutamine feeding was used for amino acid biosynthesis to prevent confounding growth effects). Alternatively, this line may have had no response to nitrate because it was not actually controlling *SOC1* expression, since its flowering time was not significantly different to the wildtype control (Marín et al., 2011). Combined with that of another study, which found carbohydrate status of the plant was involved in regulating flowering in part by its effects on the age pathway at the shoot apical meristem and in part through direct interaction with *FT* in the leaves (Wahl et al., 2013) and the argument that ionic status may be involved in regulating the floral transition through the age-related pathway becomes more convincing. Nevertheless, the question remains as to whether for the leaf ionome this response is an artefact of leaf age/development (i.e. as the plant ages how it allocates nutrients varies) or a direct consequence of a signal from the ionome. It has been argued that this response will be somewhat independent of age, supporting plant environmental adaptability and increased fitness (Guilbaud et al., 2015), which is in agreement with the difference between spring ecotypes relative to the winter ecotypes under varying vernalisation conditions as seen in the current study (i.e. spring ecotypes accumulate relatively more nutrients in comparison to the winter ecotypes only when vernalisation has not occurred, responding to the environment). Still, much further investigation is required,

involving mutants of differing growth rates under high/low nutrient conditions and varying photoperiod (instead of vernalisation).

6.4 Chapter summary and conclusions

Within this chapter the apparent association between flowering time and leaf nutrient status has been explored. To begin the panel was split in various ways in order to remove variation which could be occurring as a consequence of vernalisation/developmental differences between the leaves of spring and winter ecotypes. When *FLC* expression was used to split the panel it failed to remove the association within the quartile which had low to intermediate expression of *FLC* (Q2). This implied that this method had not successfully controlled for variation in flowering time and therefore relative leaf development. However, the panel was successfully split when actual flowering time data (as days to flower) was used. This effectively removed the association between *FLC* and flowering (implying that at least some of the differences observed were a consequence of the winter and spring ecotypes) but did not remove the association with *SOC1* in the quartile with highest *SOC1* expression/abundance of spring ecotypes. This supported the conclusion that the association between leaf nutrient status and flowering time was a consequence of the earlier flowering spring ecotypes. Yet, as none of the methods of splitting the diversity panel could completely remove the association between flowering time and leaf nutrients it was decided the relationship would be investigated further.

After splitting the data failed to remove the association between flowering time and leaf nutrients it was decided that AT analysis would be performed with an emphasis on finding common candidates. Three sets of data were analysed: a sum total of essential elements; the coefficient of variation for essential elements and flowering time (measured in days by Mr Martínez Ortuño). Coefficient of variation analyses failed to clarify the observations of either flowering time or sum total essential element analyses although it did remove the association to *SOC1*, suggesting perhaps that only a subset of elements are responsible for the associations since this technique better controls for the effects of one or two macronutrients. Conversely, GEM AT analysis of both total essential elements and flowering time had distinct similarities; *FLC*, *SOC1*, *TOE2* and *FUL* were observed in both, perhaps indicating a link to the vegetative-reproductive phase transition through the control/arrest of meristem activity. Indeed, the SNP analysis of total essential elements provided another meristem candidate, *Myb37*, which is involved in promoting axillary meristem activity (Keller, 2006; Jeifetz et al., 2011; Yu et al., 2016b). This implies a link between leaf nutrient status and reproductive development prior to floral initiation; however it could also be linking to 'age' related pathways, which *SOC1*, *TOE2* and *FUL* are downstream targets of (Wang, 2014; Wu et al., 2009; Wang, Czech and Weigel, 2009), with *FLC* potentially varying as a consequence of the confounding effect of spring and

winter ecotypes. Other main targets in the age related flowering pathway are floral meristem identity genes, i.e. *FT* (AT1G65480), *AP1* (AT1G69120) and *LFY* (AT5G61850) and therefore may be missing from the associations as they are primarily expressed within the meristem and not the leaf (in contrast to *SOC1*, *TOE2* and *FUL* which are all expressed within the leaves during vegetative development) (Fornara and Coupland, 2009; Wang, Czech and Weigel, 2009; Wang, 2014). *FT* is known to be expressed in leaf vasculature (Takada and Goto, 2003), however at the time of the RNA sampling of the diversity panel this must have been minimal as no RPKM values for *FT* were recorded/passed the minimum threshold (i.e. RPKM values below 0.4 were excluded from AT analysis (Havlickova et al., 2018)). The upstream regulators of these genes are microRNAs (specifically miR156 and miR172) which have no expression data within the RIPR diversity panel and one of many SPL transcription factors (which may have shown little association due to functional redundancy (Preston and Hileman, 2013)). As plants were all sampled for ICP-MS analysis at the same time, this could mean that developmentally the leaves of some of the spring ecotypes may be 'older' leading to the association between leaf nutrient concentrations, flowering time and the floral regulators respectively. In addition, the age related pathways have also been shown to regulate lateral root development (linking nutrient acquisition, leaf development and flowering time) (Yu et al., 2015).

To explore whether developmental differences were responsible for the association between leaf nutrient concentrations and flowering time, two further experiments were designed. One would utilise *A. thaliana* T-DNA insert lines in three floral candidates grown under long and short day conditions from the AT analysis described above, whilst the second would utilise natural variation to explore the association in *B. napus* (with the use of vernalised/non-vernalised, spring/winter ecotypes). The *A. thaliana* experiments failed to reveal any significant difference between the T-DNA lines and wildtype control plants under the varying conditions. Only P showed any significant relationships; the casparian strip like protein (CASPL) was significantly different from the wildtype controls under long and short day conditions. This could reflect the important role that root architecture plays in the acquisition of P (Peret et al., 2014) or could merely be an effect of analysing so many elements (i.e. making it more likely to observe a significant association).

The ionome timeline in *B. napus* showed significant differences between spring and winter ecotypes under different vernalisation conditions. From analyses of the time it took plants to reach each sampling/leaf development time point it appeared that the experiment had successfully controlled for developmental differences in the leaves of spring and winter ecotypes; specifically the sampling date for the 5th leaf was significantly different between

spring and winter ecotypes only for non-vernalised plants. These differences were mirrored at the 5th leaf stage in the ionomic results; there was only a significant difference between the 'total' essential elements between spring and winter ecotypes under non-vernalised conditions. Furthermore, when investigated in greater depth for multiple element concentrations, it appeared that the difference between the two ecotypes was created as a consequence of the more mobile elements. It appears from analysis that the spring ecotypes had a greater accumulation of the more mobile elements in leaf tissues. This supports the hypothesis that plants should only flower at optimum nutrient conditions and perhaps even link the growth rate rules of previous research with the age related flowering pathway theorised from GEM analysis (Guilbaud et al., 2015). Conceivably, as vegetative growth becomes limited (e.g. by the availability of nutrients or genetic control of leaf initiation through aging), there could be a decrease in investment of the more mobile nutrients in new vegetative growth, leading to a greater accumulation within the older leaves. Since growth rate is reduced, proportionally more of the leaves 'age'/are older, promoting the signalling of flowering i.e. detection of growth rate as a product of the proportion of older/mature leaves to younger/developing leaves. This could be investigated further with grafting experiments of plants with different growth rates; measuring the effect on time to flowering and the leaf ionome. In spite of this result, this experiment is inconclusive in highlighting whether leaf nutrient status may actively be playing a role in floral induction but does provide a potential link between established pathways and the leaf ionome that could be investigated further.

7 Conclusions and General discussions

7.1 General discussion

The current research has focused on investigating whether variation in micronutrient concentration in the seeds and leaves of *B. napus* is a consequence of underlying genetic loci. The project exploited a form of GWAS known as Associative Transcriptomics (AT), relating element concentration data from the seeds and leaves of *B. napus* (for Cu, Mn, Zn, Cd, Mo and S) to functional genotypes in the form of SNPs and Gene Expression Markers (GEMs). As with all forms of GWAS, a large and genetically diverse panel of plants was required to perform the trait-marker association analyses. This data was provided as part of the wider research consortium associated with the project (i.e. the Renewable Industrial Products from Rapeseed/RIPR consortium (BBSRC, 2014)) as described in the Introduction (see 1.2.7), Methods (see 2.1.1), Thomas et al., (2016) and Havlickova et al., (2018). The aim of the current research was that if genetic loci did seem to be responsible for variation in the concentration of nutrients in *B. napus* under nutrient sufficient conditions, these loci/associated markers could be targeted in breeding efforts to improve nutrient use efficiency in *B. napus* without perturbing the rest of the ionome. However, it is likely that this limited the capacity of the AT approach to detect marker-trait associations across the diversity panel (as a range of acceptable concentrations exists within and between plants that is not necessarily subject to genetic control). With the benefit of hindsight, it could be argued that looking at a smaller number of nutrients assessed under a range of conditions (e.g. deficiency, sufficiency and excess/toxicity) would have yielded more definitive results.

Furthermore, given that plants were grown in soil in 5L pots it is possible that individual plants were exposed to different concentrations or availabilities of nutrients (as soil is a heterogeneous matrix, variation in the volume of soil per pot etc.). However, it is unlikely that any nutrients were limiting in such a large volume of soil relative to the growth stage of the plants sampled. It would have been easier to ensure that plants were all exposed to the same range and availabilities of nutrients if they had been grown hydroponically. However, this would have been too expensive, time consuming and impractical for the volume of plants under analysis (2160 experimental units). Nevertheless, it is interesting to note that experimental design accounted for ~33%/24% (leaf/seed) of variation in the variation component analysis performed as part of Thomas et al., (2016) for the micronutrients (B, Cu, Fe, Mn, Mo, Zn) and ~21%/38% (leaf/seed) for the macronutrients (Ca, K, Mg, P, S) (Thomas et al., 2016). This may indicate that the concentration of micronutrients in the leaves and

macronutrients in the seeds was more susceptible to variation in experimental design. However, there is no significant difference in % variation of micronutrient concentrations between the seeds or leaves that is attributable to experimental design (t-test, t 0.54, df 7, p 0.60). On the other hand, the effect of experimental design was significantly higher for macronutrients in the seeds in comparison to the leaves (t-test, t -2.60, df 6, p <0.05). Overall, there was no significant difference between the % variation assigned to experimental design between the macronutrients and the micronutrients (t-test, t -0.25, df 16, p 0.81). This is important to note as it could be argued that the micronutrients would be more susceptible to slight variations in soil or experimental design. Since they are required in much smaller amounts, it could be expected that any change could have a disproportionate effect. Alternatively, the significantly large effect of experimental design on the macronutrients in the seeds relative to the leaves could be a consequence of differences between ecotypes in the diversity panel and the effect of breeding for low GSL (discussed further in 4 and 7.2).

The effects of underlying population structure may be responsible for a number of issues in the current research. For example, the plants were sampled for RNA at the 2nd true leaf prior to vernalisation, whilst sampling for leaf element concentrations for ICP-MS analysis occurred at the 6th-8th true leaf stage after vernalisation. As the diversity panel contained a mixture of spring and winter OSR (oilseed rape) ecotypes, if the plants were not sufficiently vernalised when they were sampled it could mean the plants were sampled at different developmental time points (despite being sampled at the same time). If true, this effect could easily create the association between flowering time and leaf nutrient status (discussed in greater detail in 6). Nonetheless, when the temperature timelines are examined (Figure 7.1.a) it would appear that the plants were likely subjected to the correct temperature and photoperiod to achieve vernalisation (one month of average temperatures between 6 and 9°C (Tommev and Evans, 1991)). Furthermore, the use of both Q and K matrices in the AT analysis should account for spurious associations caused by underlying population structure (Price et al., 2010). Nevertheless, it could be further argued that the leaf expression data used for AT analysis was not representative of the leaves or seeds at the time they were sampled for ICP-MS analysis. It is possible that this has limited the discovery of candidate genes which are only expressed at certain developmental stages. However, the AT technique has previously been successfully used in such instances (e.g. with seed (Harper et al., 2012; Lu et al., 2014; Havlickova et al., 2018) and stem data (Wood et al., 2017)) as a consequence of variation in surrounding genes in linkage disequilibrium with the causative gene generating an association peak.

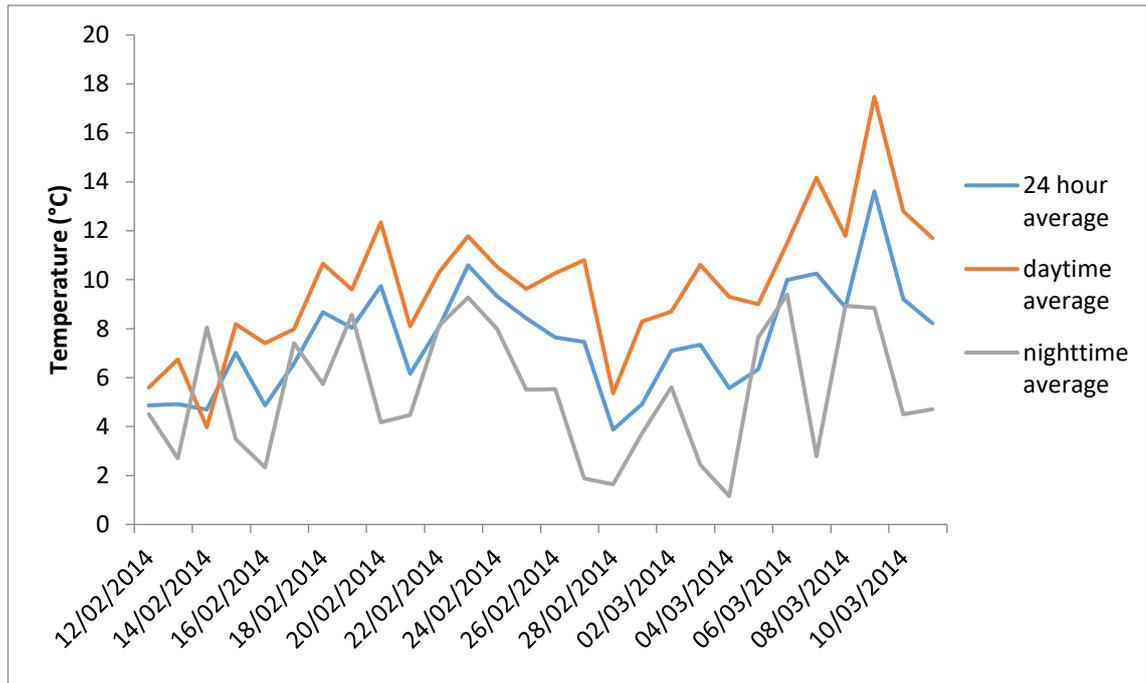


Figure 7.1.a temperature timeline for one of the polytunnels used during the growth of the RIPR panel.

The average temperature is displayed per day in one of three ways (48 temperature measurements per day): the 24 hour average (an average of all temperature recordings within a 24 hour period measured from midnight-midnight, from 48 measurements), daytime average (an average of all temperature recordings between 6am and 6pm, from 24 measurements) and night-time average (an average of all temperature recordings between 6pm and 6 am, from 24 measurements). A one sample t-test was performed to assess if the daytime average was significantly different from the highest acceptable vernalisation temperature (9°C (Tommev and Evans, 1991)) and it was not significantly different (t 1.53, df 27, p 0.132).

The limited number of candidate genes taken forward from AT analysis for the leaves reflects the difficulty of both plant ionomic research and GWAS. Complex traits often have many associated SNPs which explain only a fraction of trait heritability in GWAS analyses (Korte and Farlow, 2013). Perhaps this was why so few of the AT analyses passed the 5% false discovery rate or Bonferroni corrected significance threshold of 0.05 (i.e. many markers with minor effect on the trait). Lowering the false discovery thresholds could have yielded more candidates, however it was considered more important to maintain consistency across all traits analysed as part of the RIPR consortium and as such the threshold was maintained at 5%. Another issue is that the candidate gene approach used in the current research has limited capacity to detect novel candidates as it relies on orthologue information. One of the main advantages of the GWAS approach is the capacity to narrow down the association regions

under analysis (in comparison to QTL) and to potentially identify the candidate gene directly (1.2.7). In the current research there were some instances where a GEM marker with trait-predictive capability could represent a novel candidate gene (e.g. in Mn leaf GEM analyses, see 5.2.3) however the lack of orthologue information in the pan-transcriptome prevented any form of validation in the current analyses with *A. thaliana*. Nevertheless, it is rare within GWAS analyses that the marker most highly associated with the trait is a causative allele. An alternate explanation for the lack of candidates from leaf analyses would be that the candidate genes important for variation in leaf element concentrations did not vary as they are highly conserved across the diversity panel (i.e. they are essential and therefore do not vary). However, of the candidates and markers identified in the current study, many were found to be significantly predictive in an independent panel. Indeed at least one predictive marker was found for all element concentrations investigated in each tissue. This implies some level of biological relevance, most likely through marker LD with causative loci, despite the apparent lack of candidate genes/markers passing false discovery thresholds. Furthermore, it is likely that the way markers were assessed for trait predictive capacity was limited by splitting the panel in two (i.e. making it hard and sometimes impossible to assess trait predictability or rare alleles). A better approach would have been to use the “leave one out” methodology of Harper et al., (2016).

The key difference between AT and GWAS is the use of transcribed sequence data instead of genomic data for the marker-trait association analysis. AT was initially developed as a means to avoid the difficulties of genome sequence order and SNP discovery in polyploids without a reference genome (Harper et al., 2012). Subsequently a draft genome sequence has become available in *B. napus* (Chalhoub et al., 2014), however the updated AT pan-transcriptome has been shown to contain a greater number of CDS gene models (He et al., 2015). Furthermore, the updated AT pan-transcriptome has been shown to have a higher SNP density than the commercially available 60K Brassica Infinium® SNP array (Xu et al., 2015; Li et al., 2014; Havlickova et al., 2018). Nevertheless, the AT technique is limited to only transcribed sequences unlike other GWAS studies. Therefore, any variation in non-coding sequences associated with the trait under investigation will only be detected if they have an effect on gene expression and/or are in LD with surrounding SNPs in CDS gene models. The use of GEM analysis in AT partially mitigates such issues as transcript abundance data is available for ~46% of genes present in the pan-transcriptome (Havlickova et al., 2018). It has also previously been shown that whole genome re-sequencing can identify a larger number of SNPs than AT (Huang et al., 2013). However there are further benefits to the gene expression data used in AT analysis of polyploids (Havlickova et al., 2018), with the measurement of transcript abundance

enabling discrimination between homoeologous genes in polyploids (Higgins et al., 2012). In the current research, the GEMs played an important role in candidate gene analysis and exploring the potential relationship between S, GSL and the seed ionome (e.g. via WGCNA analysis, see 4.3.1). This work reiterated the likelihood of the low GSL phenotype being a result of a homoeologous exchange as all the GEMs showed similarly disrupted expression patterns (see 4 and 7.2).

As with all forms of GWAS, candidate genes identified with AT need to be functionally verified. In the current research this validation was performed with *A. thaliana* T-DNA insertional mutants. If the T-DNA insert mutant displayed a significant difference in the concentration of the element under investigation relative to a wildtype control it was considered validated. However, there are a number of issues with such an approach. Firstly, *A. thaliana* and *B. napus* diverged ~20 Mya (Yang et al., 1999), therefore the candidates identified in *B. napus* may have functionally diverged from their *A. thaliana* orthologues. As such, candidates may have been rejected in *A. thaliana* which could have a different functional role in *B. napus*. Other issues within the current research were the availability of appropriate T-DNA lines. This was a multifaceted issue. For example, some candidate genes were already well characterised within the literature and Ionomics Hub (Baxter et al., 2007b). These candidate genes were usually not subjected to further validation, however when some candidate genes were re-analysed in *A. thaliana* (e.g. *UMAMIT 28* or even *Myb28/HAG1*) inconsistencies were observed (e.g. which elements vary in *UMAMIT 28* or the multi-element phenotype of *Myb28/HAG1* (see 3.3)). This emphasises the need for the repetition of results in general (both within the current research and wider academic literature) but could also be an artefact of the experimental design used in this study. For example, the growth of candidates on nutrient rich soil in comparison to hydroponics for *Myb28/HAG1* (Martinez-Ballesta et al., 2015), analysis of all candidates at only one developmental stage or pooling of seed samples. In other instances, the *A. thaliana* approach failed as all the plants produced from a T-DNA line were genotyped as wildtype. This could be a consequence of the candidate gene being essential for plant survival/reproduction, however this would likely have been an issue no matter which experimental system was used for validation (except RNAi, see 1.2.8). Another issue with the current research was the lack of transcript quantification of the T-DNA lines (e.g. qRT-PCR) or measurement of other important phenotypes (e.g. plant size). Once again this may in part be able to explain the differences observed between the current research and the literature/Ionomics hub if the T-DNA lines assessed were not truly 'knock-outs' but 'knock-downs'. Another explanation could be that the *A. thaliana* wildtype background was not truly representative of each of the T-DNA lines under analysis. As an additional check it would have been useful to compare the plants which were

genotyped as wildtype from each of the T-DNA lines to the independent wildtype control and the plants genotyped as containing the T-DNA insertion. If there was no difference between the plants known to have the T-DNA insertion and the plants which had been genotyped as wildtype (but a significant difference was observed relative to the independent wildtype control) it would suggest either that the method of genotyping was inaccurate or that another mutation in the T-DNA line was responsible for the phenotype. Similarly, if the independent wildtype line and the genotyped wildtype from the T-DNA seed stock produced the same phenotype it would confirm that it was the known T-DNA insertion responsible for any observed phenotype in the T-DNA line.

Despite many of the aforementioned limitations of the techniques used throughout the current research, a number of promising candidate genes for further investigation were successfully identified. Three candidate genes were already well characterised in the literature: Bo3g053000.1, which is an orthologue of *Heavy metal ATPase 5* in *A. thaliana* (*HMA5*: AT1G63440.1) was found in seed Cu AT analysis; Cab011213.1 and Cab011209.2, whose orthologues in *A. thaliana* are *heavy metal ATPase 2* (*HMA2*: AT4G30110.1) and *heavy metal ATPase 3* (*HMA3*: AT4G30120.1) were found in leaf Cd analysis. *HMA5* is a known Cu transporter (Andres-Colas et al., 2006), *HMA 2* is a known Zn/Cd transporting ATPase (Wong et al., 2009) and *HMA 3* is known to be important for the vacuolar storage of Cd (Morel et al., 2009). As these genes had already been validated in the literature they were investigated no further as part of the current investigation. Time permitting, it would have been interesting to see if the phenotype of *A. thaliana* T-DNA lines could be complemented with the *B. napus* orthologue (i.e. transform the *A. thaliana* T-DNA lines with the *B. napus* orthologue to see if the disrupted phenotype could be rescued). Of the novel candidates identified in the current investigation, *Cd leaf HMA-2* (Cab002809.1/ AT2G36950) would warrant similar investigation. Given the significant increase in Cd within leaf tissues and the apparent Cd toxicity response observed in the other element concentrations in the T-DNA line, this gene would represent a suitable target for improving Cd accumulation in *B. napus* for phytoremedial purposes. An obvious next step for some of the candidate genes with inconsistent results in *A. thaliana* would be to assess the expression of the candidate gene in the T-DNA line (e.g. with qRT-PCR). This would be particularly useful for seed *Per1L- 1* and *2* where it was theorised that alternative splicing may have been responsible for the contrasting results obtained. Similarly, given the inconsistency of some of the results with the literature/ionomics hub, a repetition of the current experiments (with transcript quantification) would be ideal for *Myb28/HAG1* and *UMAMIT 28* T-DNA lines.

7.2 Understanding micronutrient concentration in *B. napus*

Over the course of the current investigation our understanding of micronutrient concentration in the *B. napus* tissues under investigation has expanded. One of the quickest and easiest ways to determine differences in nutrient allocation between tissues is to look at them as a ratio (**Figure 7.2.a**). When the ratios of all the elements under investigation were compared some interesting patterns were observed in line with the evidence generated as part of the current investigation. If the ratio for seed (mg/kg DW): leaf (mg/kg DW) is >1 it indicates a net increase in concentration of that element in the seed, while a ratio <1 indicates higher concentrations within the leaves in comparison to the seeds (potentially indicating some level of retention). Across the elements analysed the most striking example of this would be the Cd seed: leaf ratio. Among all of the crop types compared, none display an increase in concentration in Cd in the seeds relative to the leaves. This fits with previous research which has shown that vacuolar sequestration and cell wall immobilisation of Cd in the leaves is an important defence mechanism against the biologically toxic Cd (Carrier, Barylá and Havaux, 2003). Furthermore, it emphasises why AT analyses of Cd concentrations in leaves gave clear association peaks, as the candidates identified in this analysis were involved in Cd translocation/sequestration. Interestingly, the only micronutrient analysed which had a seed: leaf ratio of <1 was Mn. This is indicative of the integral role that Mn plays in photosynthesis (see 1.2.5.2 and Barber, 2009), leading to its immobility with respect to transport into the seeds. Of the other micronutrients analysed (Mo, Cu and Zn) all showed a seed: leaf ratio of >1 , indicating a higher net concentration in the seeds.

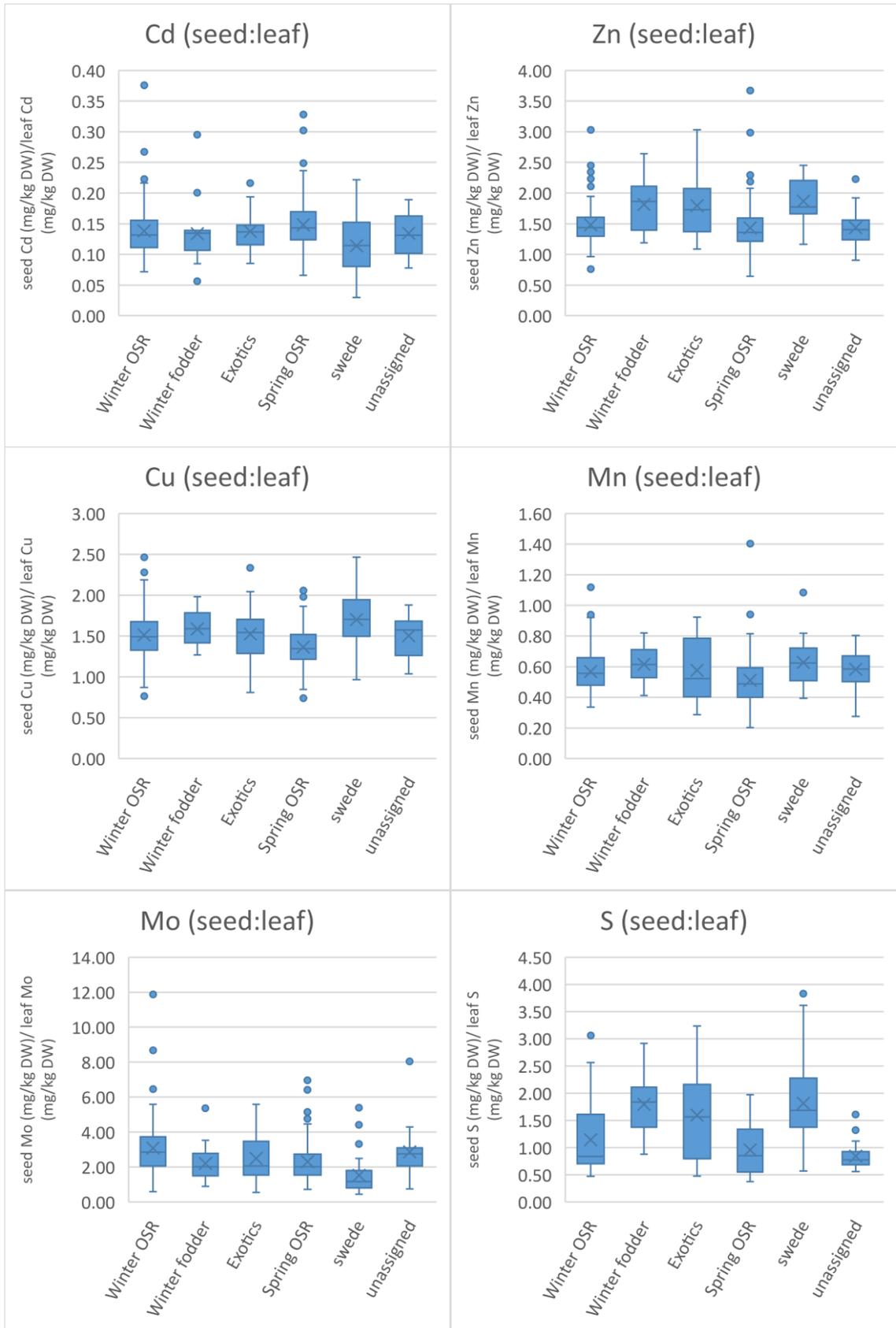


Figure 7.2.a seed: leaf ratios of all elements under investigation across the RIPR diversity panel

Left to right and top to bottom in order: Cd, Zn, Cu, Mn, Mo and S. Seed: leaf ratios calculated by [seed element concentration (as mg/kg DW)/ leaf element concentration (as mg/kg DW)]. Data is presented as

box plots with the mean displayed as an X, the interquartile range with median drawn as the box, whiskers are the 95% confidence interval and dots are outlying values. Data calculated per ecotype from the RIPR leaf and seed concentration datasets. Winter OSR (n 159), winter fodder (n 15), exotics (n 26), spring OSR (n 127), swede (n 30) and unassigned (n 21).

Mo and S displayed some variability in ratios between crop types (**Figure 7.2.a** and Thomas et al., (2016)). The winter fodder and swede ecotypes generally had higher concentrations of S within their seeds relative to the spring or winter OSR ecotypes, while the reverse was observed in Mo (higher concentrations in seeds within the OSR ecotypes). It is thought that this is a consequence of breeding for low GSL lines in the OSR ecotypes, with low GSL lines having low S concentrations in their seeds. This is consistent with the seed AT results for S and Mo (see 3 and 4). It also adds weight to the results of the senescence investigation, which showed that low GSL lines under low N conditions had higher remaining S concentration in senesced leaves (see 4.3.3). One of the explanations afforded to the link between S and Mo is that sulfate and molybdate are chemical analogues (Marschner, 1995b). The low GSL phenotype could affect S translocation to the seed, indirectly affecting the Mo concentration. The same pattern can be observed in the ratio of $([S]: [Mo] \text{ seed}) / ([S]: [Mo] \text{ leaf})$ (**Figure 7.2.b**). In the context of this study this would imply net accumulation of Mo relative to S in the spring and winter OSR ecotypes as the ratio is generally <1 . This is different to the swede ecotypes, where there appears to be a higher concentration of S in the seeds relative to Mo.

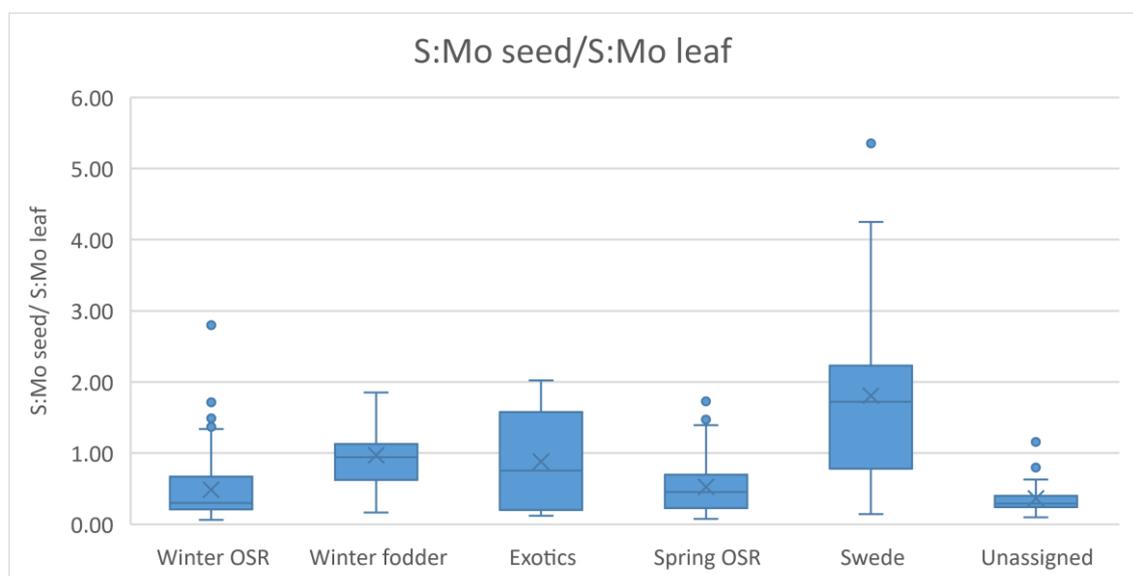


Figure 7.2.b ratio of S and Mo concentrations across seed and leaf tissues in the RIPR diversity panel
The ratio was calculated as $([S] \text{ seed} / [Mo] \text{ seed}) / ([S] \text{ leaf} / [Mo] \text{ leaf})$. Data is presented as box plots with the mean displayed as an X, the interquartile range with median drawn as the box, whiskers are the

95% confidence interval and dots are outlying values. Data calculated per ecotype from the RIPR leaf and seed concentration datasets. Winter OSR (n 159), winter fodder (n 15), exotics (n 26), spring OSR (n 127), swede (n 30) and unassigned (n 21).

However, the effect of GSL on the wider seed ionome stretches further than simply disrupted Mo concentrations in the seeds. When AT was used to analyse S concentrations in seed, the SNP and GEM markers assessed were able to predict the concentration of a number of other elements (including S, Mo, B, Mg, Zn and Sr). There are two potential explanations for this result. The first is that breeding for the low GSL phenotype indirectly selected for the concentration of the various elements indirectly. For example, within the A9 association peak in S seed concentration analyses there is a potential Mg candidate, Cab040264.1, whose orthologue in *A. thaliana* is a Mg transporter (*MGT9*: AT5G64560.1). Within *A. thaliana*, *MGT9* has been shown to be important for male fertility and to be expressed within both the dry seeds and the vasculature of plants (Chen et al., 2009). It is possible that candidate genes like this were selected for randomly throughout the breeding process for low GSL and have caused the association between GSL markers and element concentrations in seed by chance. However the presence of this candidate gene and others could also be due to chance as the associated region on A9 is extremely large (~870 CDS gene models).

The second potential explanation for the association between S, the GSL and wider seed ionome could be that the candidate gene thought to be responsible for the low GSL phenotype, *Myb28/HAG1*, is somehow responsible for causing variation in the concentration of multiple elements in the seeds. This hypothesis is supported by the results observed in the *A. thaliana* T-DNA *myb28/hag1* line in the current investigation, which appeared to show disruption in multiple element concentrations within the seeds and leaves (see 3.3.2). *Myb28/HAG1* could be causing variation in the concentration of nutrients in a number of ways. One possibility is that *Myb28/HAG1* is in some way responsible for regulating the movement of sulfate. Perhaps the low GSL phenotype/ low *Myb28/HAG1* expression is perceived as S deficiency, upregulating sulfate transporters which are capable of transporting other elements, such as Mo. Alternately, perhaps low expression of the major aliphatic GLS regulator *Myb28/HAG1* causes a subsequent upregulation of aromatic GSL biosynthesis which requires a Mo containing enzyme (causing an increase in Mo concentration within the seeds)(Ibdah et al., 2009). However, neither of these explanations would explain the variation in other elements (e.g. B, Mg or Zn). It was therefore suggested that a reduction in GSL could be associated with variation in nutrient concentrations in the seeds through their role as plant defence

compounds. If the plant is subject to increased biotic stress it could affect plant wide nutrient concentrations (e.g. increased root or leaf damage may lead to variation in plant wide nutrient allocation).

However, considering pod, stem and seed ionome experiments (see 4.3.4) did not show any variation in nutrient concentrations other than S across tissues and Mo in green seeds (in high and low S/GSL lines), it appears more likely that variation in nutrient concentrations in the seeds of different *B. napus* lines across the RIPR diversity panel is a consequence of indirect selection through breeding. Nevertheless, it is still likely that S and Mo are varying in seeds as a direct consequence of the low GSL phenotype. It will be important to understand the exact mechanism if this information is going to be of use for crop breeding. For example, from senescence experiments it appears that low GSL lines retain higher S concentrations in senescing leaves. As it has been suggested that the low GSL lines perform poorly under S limitation (Bloem, Haneklaus and Schnug, 2007), it is important to understand the broader effects of *Myb28/HAG1* on plant-wide S concentrations. Given that Mo appears to vary in seeds but not the pods or stems of low GSL lines, it could be that the movement of S into the seeds has been perturbed (if they share a transporter) or that there is an increase in demand for Mo in the production of aromatic GSL. Since S is required for aromatic GSL production the latter explanation seems unlikely. It is questionable whether this could then be applied to the whole plant, i.e. that plant wide movement of S is perturbed as a direct consequence of *Myb28/HAG1* regulation of S transporters. If the variation in seed is a consequence of a shared transporter, a general increase in Mo concentrations in all tissues, not just seeds, would be expected. Alternatively, it could be that aromatic GSL' account for such a small proportion of seed GSL that demand for Mo is increased without any obvious effect on S concentrations. Another explanation would be that there is tissue specific expression of certain S transporters and only those expressed in seed are capable of transporting both S and Mo.

To try and understand whether and why low GSL lines might be performing poorly under limited S a range of further experiments could be undertaken. Firstly, it would be useful to re-characterise the lines used in the pod/seed/stem experiment across the spectrum of S containing compounds (i.e. free sulfate, GSL, proteins and lipids). It would be of particular interest to compare how the low and high S lines would respond to high and low S fertilisation. For example, if reduced *Myb28/HAG1* expression is perceived as a high S signal (i.e. when there are enough GSL, *Myb28/HAG1* is down regulated) perhaps there is increased allocation of S into less mobile pools. Under limiting S this would result in lower S being available for remobilisation, potentially reducing the S concentrations in seeds alongside GSL. However it

has been shown that *Myb28/HAG1* is downregulated by S deficiency (Aarabi et al., 2016). It has been hypothesised that binding between Myb28/HAG1 and Sulfur Deficiency Induced 1 (SD1) is important in this response. Perhaps the absence of an upregulation in *SD1* etc. combined with already low expression of *Myb28/HAG1* is perceived by the plant as a high GSL signal. Along similar lines, as *Myb28/HAG1* expression has been linked to regulation of genes involved in primary S metabolism, it could be that S assimilation throughout the plant is down regulated in low GSL lines resulting in greater storage of sulfate in vacuoles and reduced translocation. It would be revealing to look at plant wide gene expression of primary S metabolism genes in the low and high S lines under different S concentrations. It would also be interesting to compare the performance of the *myb28/hag1* T-DNA line under varying herbivore burden. This could be achieved by re-growing lines in sterilised and non-sterilised soil in separate growth chambers. Elucidating the effects of GSL regulation would enable targeted breeding for higher GSL leaves for human consumption in leafy crop types, improving their nutritional value. On the other hand, increasing GSL in leaves without increasing them in the seeds of OSR would promote pest resistance without limiting the use of seed meal as an animal feed.

7.3 Flowering time and leaf element concentration

Individual GEM AT analyses for leaf element concentration commonly highlighted orthologues of FLC and SOC1 as some of the most highly associated GEM results. This was initially thought to be a consequence of lines within the diversity panel being insufficiently vernalised. However, temperature timelines (**Figure 7.1.a**), evidence from splitting the diversity panel, flowering AT results and a leaf ionome timeline have all added weight to the hypothesis that flowering time and leaf nutrient concentration may be linked prior to floral induction (see 6 and **Figure 7.3.a**). There could be a number of reasons why flowering time, leaf element concentrations and some of the major floral regulators (*FLC* and *SOC1*) seem to correlate prior to floral induction. The key question to be solved is whether leaf nutrient concentration is actively promoting or repressing flowering (i.e. if leaf nutrient concentration causes variation in flowering time) or if the association is an artefact of developmental differences between the various ecotypes in the diversity panel (i.e. if time to flowering is causing variation in the concentration of elements in the leaf ionome) (**Figure 7.3.a**).

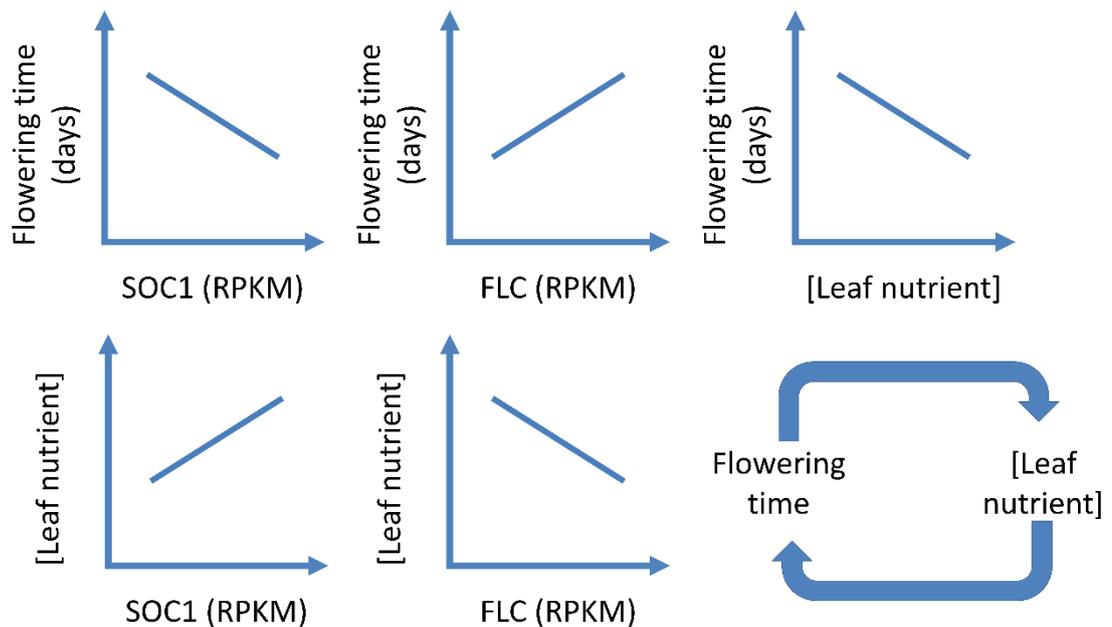


Figure 7.3.a Schematic representation of the relationship between leaf nutrient concentration and flowering time

Representations of the correlations between flowering time (as days to flowering), leaf nutrient concentration (as mg/kg DW), FLC and SOC1 expression (as RPKM) across the RIPR diversity panel, alongside proposed causal model of leaf nutrient concentrations influencing flowering time and vice versa.

There is already evidence that plants may modify their flowering time in response to nutrient deficiency (e.g. Tienderen et al., 1996; Zhang and Lechowicz, 1994; Pigliucci and Schlichting, 1998; Kolář and Seňková, 2008). However, whether nutrient concentrations under nutrient sufficient conditions can affect flowering time is yet to be determined. Feasibly both of these hypotheses could be responsible via various flowering pathways. As all the plants in this study had their leaves sampled at the same time for ICP-MS analysis it could be that developmental differences between the various ecotypes caused the association to leaf nutrient concentration (i.e. plants at different developmental stages may have different leaf element profiles). However, this should have been mitigated (at least in part) by the sampling method (i.e. a minimum of three fully expanded leaves taken per plant at approximately the 6th-8th true leaf stage (Thomas et al., 2016)). Alternately, the genes coordinating flowering time may directly affect the nutrient concentration in leaves to ensure the availability of nutrients throughout the plant during the vegetative to reproductive phase transition. On the other hand, perhaps leaf element concentrations directly interact with plant development, causing the association with flowering. Specifically, the leaf nutrient concentration may affect flowering time precisely because it is vital to plant development and the success of flowering. Marín et al., (2011) have already provided evidence that nitrate may play a role in modulating plant flowering time in a mechanism that works in tandem with the autonomous, gibberellin and photoperiod pathways but is still repressed by *FLC*. Further research has linked expression of *FLC*, *LFY* and *AP1* to low nitrate conditions (Kant, Peng and Rothstein, 2011), whilst other workers have linked both the photoperiod and gibberellic acid pathways to low nitrate responses (Liu et al., 2013). Other authors have also suggested that the effect of nutrients on flowering is part of a stress induced flowering response (Takeno, 2016). It is tempting to apply these observations to the results of the current investigation.

However, the link between flowering and nutrient status has historically been very hard to deconstruct. This is because it is difficult to separate any primary effects (e.g. leaf nutrient concentrations directly affecting flowering time) from secondary effects (e.g. nutrient availability causing disruption in vegetative growth leading to changes in flowering time) of nutrient concentrations on flowering time (Marín et al., 2011). To assess if leaf nutrient concentration is directly affecting flowering time, it would be necessary to repeat the leaf ionome timeline at low and sufficient nutrient concentrations. The leaf ionome of the spring and vernalised winter ecotypes should be able to respond to the availability of nutrients and perhaps modify flowering time. The winter ecotypes that are not vernalised would be unable to modify flowering time and thus it would be possible to assess if/how the ionome responded. Additionally, analysis of *FLC*, *SOC1*, *TOE2* and *FUL* expression (i.e. genes highlighted

in GEM AT analysis) across the developmental time points, in both the leaves and floral meristem as they develop, would provide further insight into the contribution of leaf nutrient concentration to flowering time. Comparison between the vernalised and non-vernalised winter ecotypes should highlight any difference in the response of both the leaf ionome and gene expression as a result of floral regulators, whilst the analysis of the spring ecotypes would confirm that any effects observed were not vernalisation dependent. Analysis of the effect of leaf nutrients on flowering time would be useful to confirm the effect in spring ecotypes (e.g. delayed flowering at lower nutrients) but would not be useful for the vernalised/non-vernalised winter ecotypes as the non-vernalised accessions would not flower.

7.4 Overall conclusions

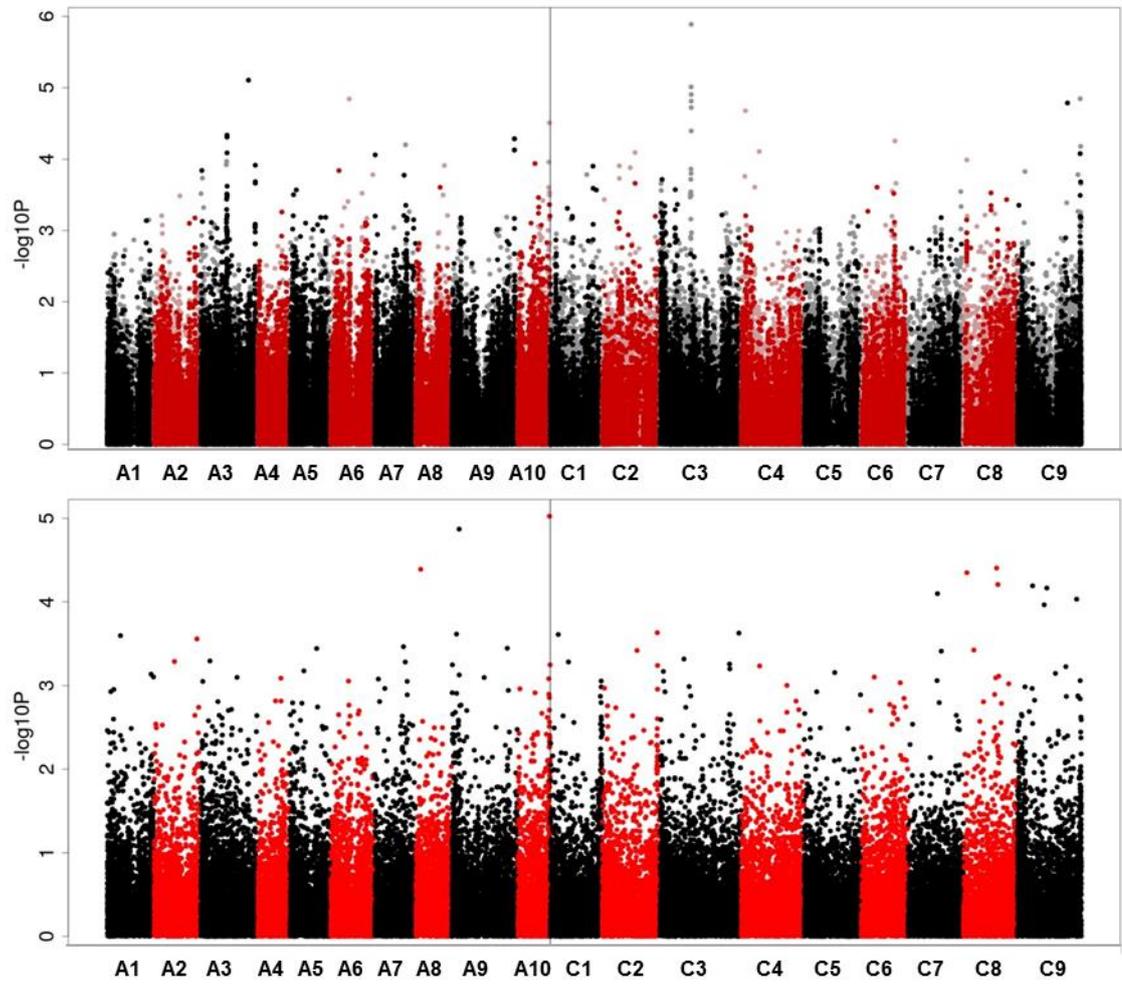
All in all, the current research has successfully tested whether underlying variation in the leaves and seeds of *B. napus* is a consequence of genetic loci. The AT approach, in spite of its limitations, was able to identify a number of candidate genes to be taken forward to further analysis with *A. thaliana* T-DNA lines. Some of the candidate genes analysed in *A. thaliana* displayed disrupted element concentrations relative to the wildtype control. Further testing of these candidate genes is necessary to confirm their biological role in *B. napus*, e.g. with TILLING. Additional research into the relationship between GSLs and the seed ionome revealed how breeding for one trait may disrupt others. This research requires further investigation, perhaps by analysing the expression of primary S metabolism genes in high and low S lines under different S concentrations. A potential link between the concentrations of elements within the leaves of *B. napus* and flowering time was also observed. Whether leaf nutrient concentrations can cause variation in flowering time under nutrient sufficiency or vice versa requires further investigation. This could be achieved with further analysis of the leaf ionome timeline accessions under varying fertilisation regimes. The significantly associated markers highlighted from AT analysis within this study could also be carried forward into MAS and further testing in *B. napus*. This work represents the first step towards breeding *B. napus* with improved micronutrient use efficiency.

Appendices

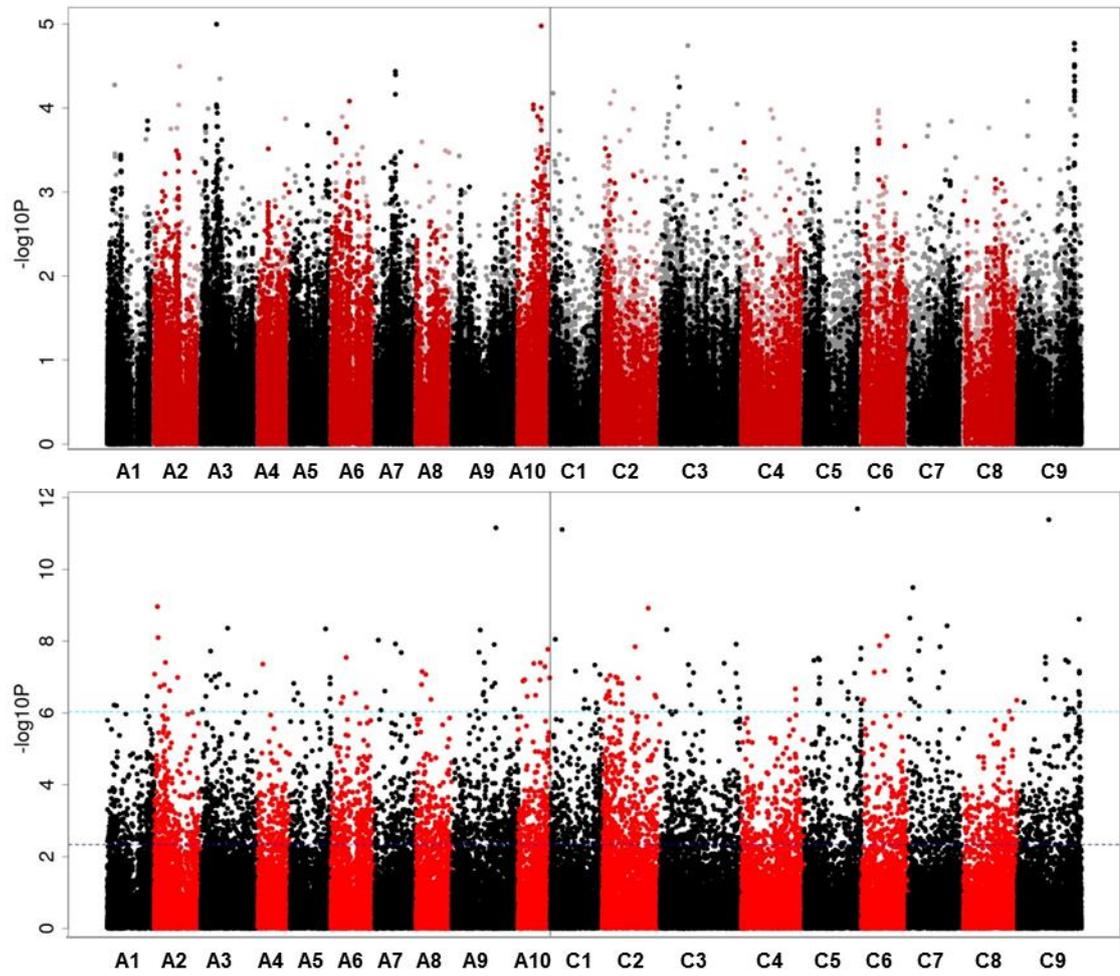
Associative Transcriptomic outputs on the 274 diversity panel

Genome wide distribution of mapped markers for various traits (as indicated above each Manhattan plot) on the 274 diversity panel. SNP associations (top) were calculated with the R script GAPIT (Lipka et al., 2012), using a compressed linear mixed model capable of accounting for population structure and relatedness with a Q matrix inferred by PSIKO (Popescu et al., 2014). GEM associations (bottom) were calculated with the R script Regress, performing fixed effect linear modelling with the Q matrix and RPKM data as explanatory variables and the various element concentrations as the response variable. $-\log_{10}P$ values from the SNP and GEM association analysis were plotted against the pseudomolecules (representing the 19 *B. napus* chromosomes) based on the CDS gene model order (labelled on the X axis from chromosome A1-C9). For the SNP analysis, black and dark red points represent simple SNPs and hemi-SNPs that have been linkage mapped to a genome, while grey and light red points represent hemi-SNPs which have not been linkage mapped but assigned to the genome of the CDS gene model they were called from. The two type 1 error tests are portrayed as dashed lines when associations pass these thresholds; the Bonferroni corrected significance threshold of 0.05 as light blue and the 5% false discovery rate (FDR) as dark blue.

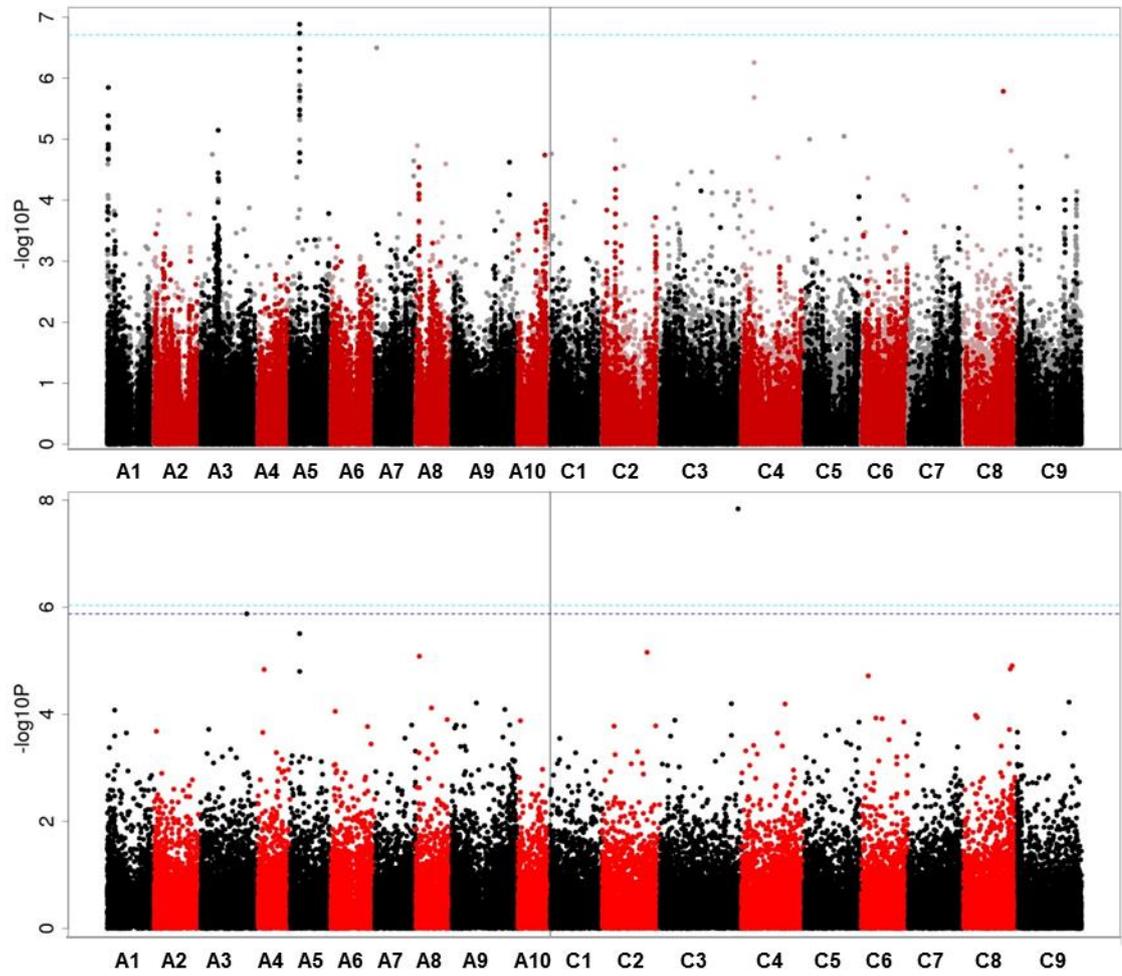
Appendix 1 AT output for Cu (mg/kg DW) seed on the 274 diversity panel



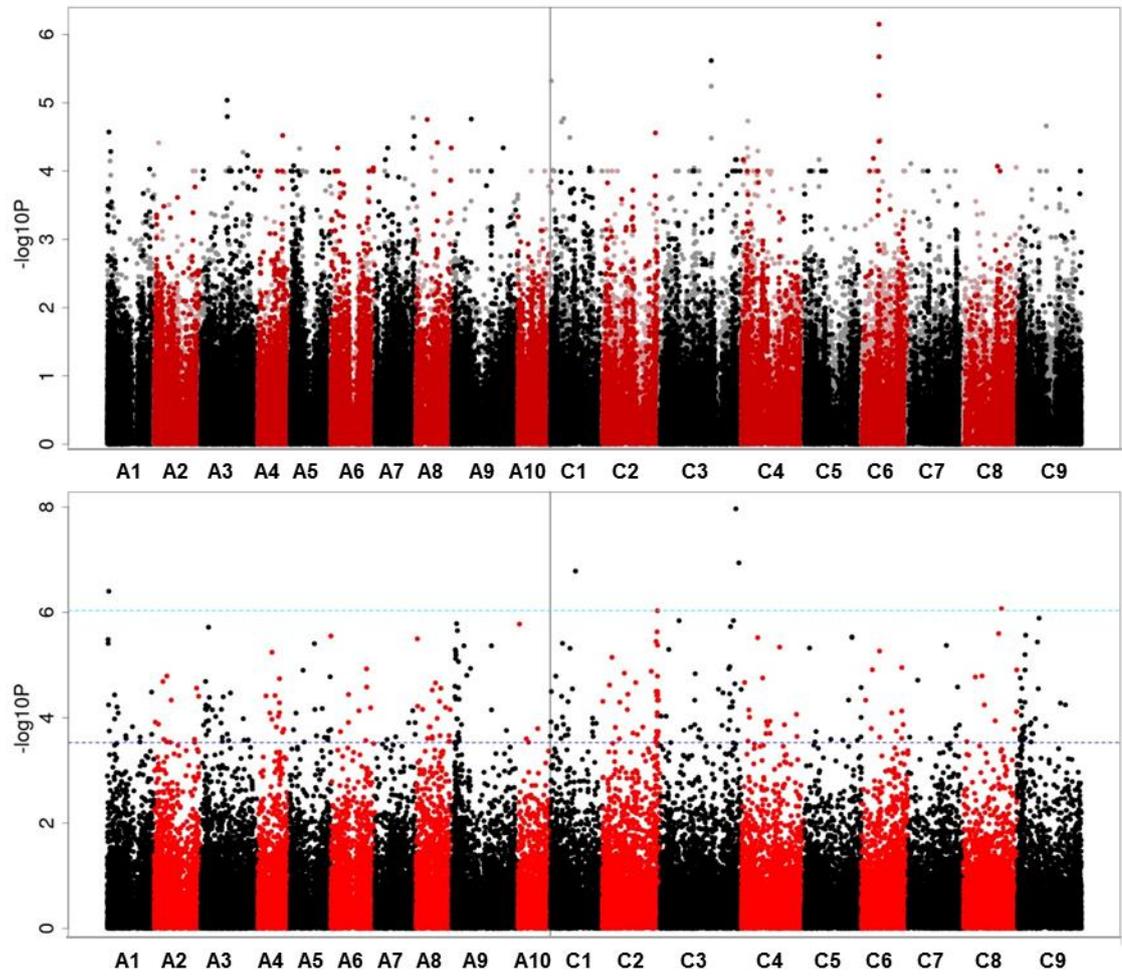
Appendix 2 AT output for Cd (mg/kg DW) seed on the 274 diversity panel



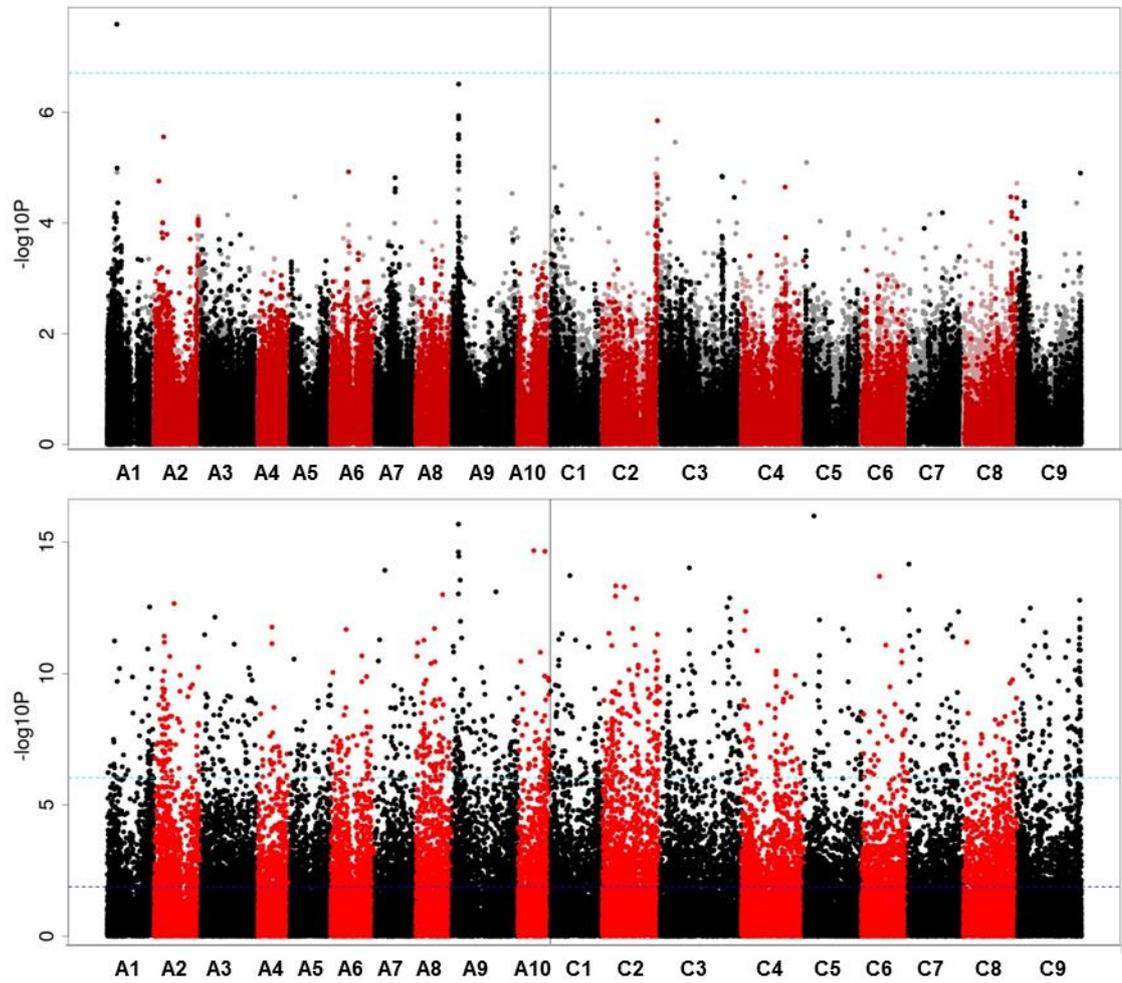
Appendix 3 AT output for Mn (mg/kg DW) seed on the 274 diversity panel



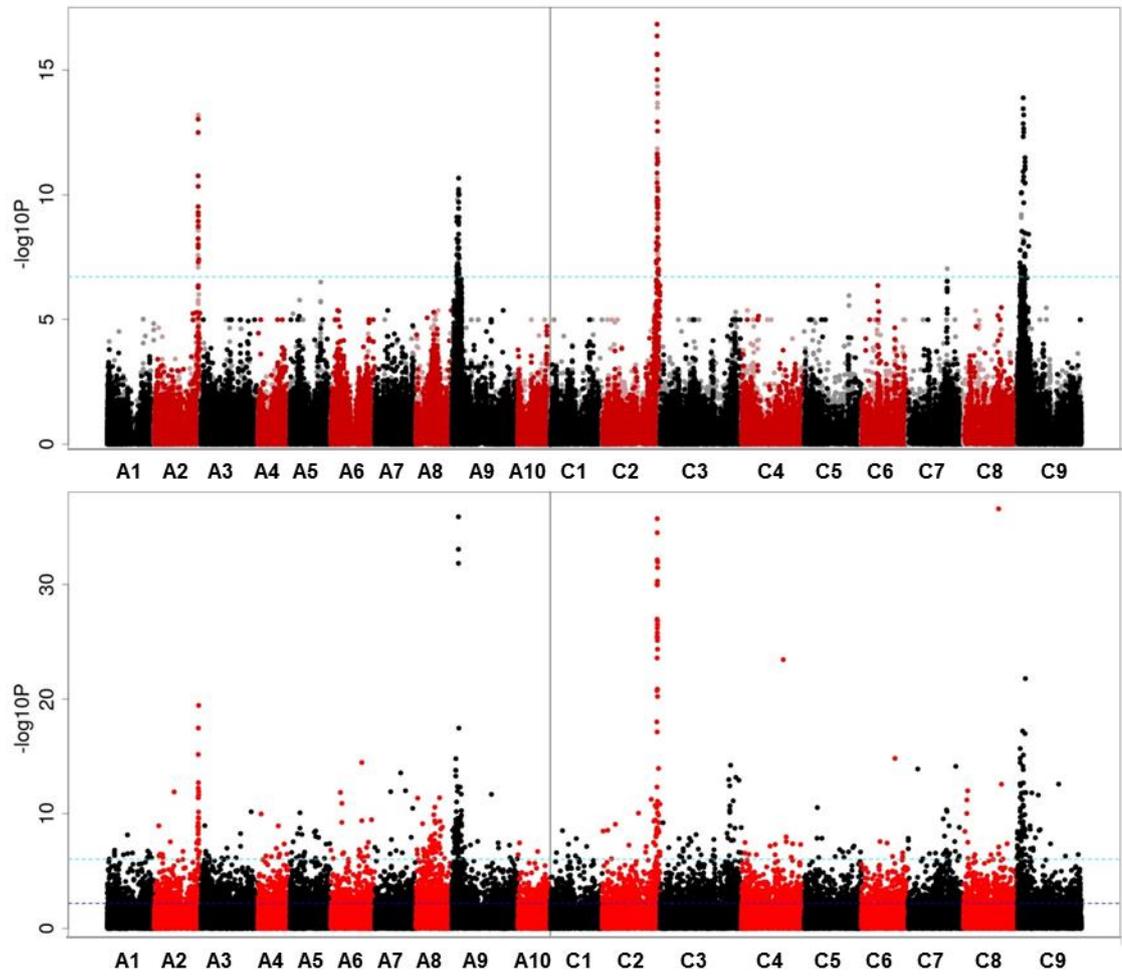
Appendix 4 AT output for Zn (mg/kg DW) seed on the 274 diversity panel



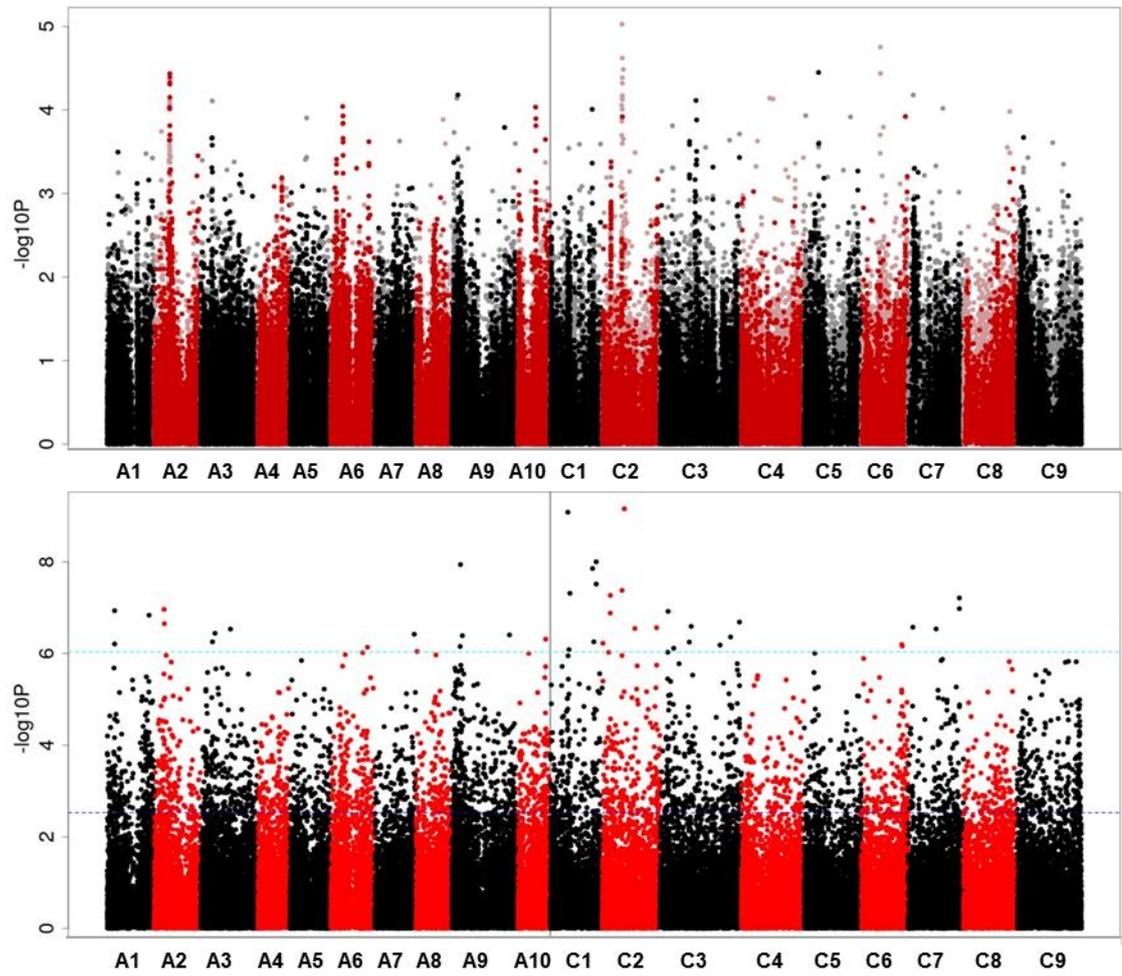
Appendix 5 AT output for Mo (mg/kg DW) seed on the 274 diversity panel



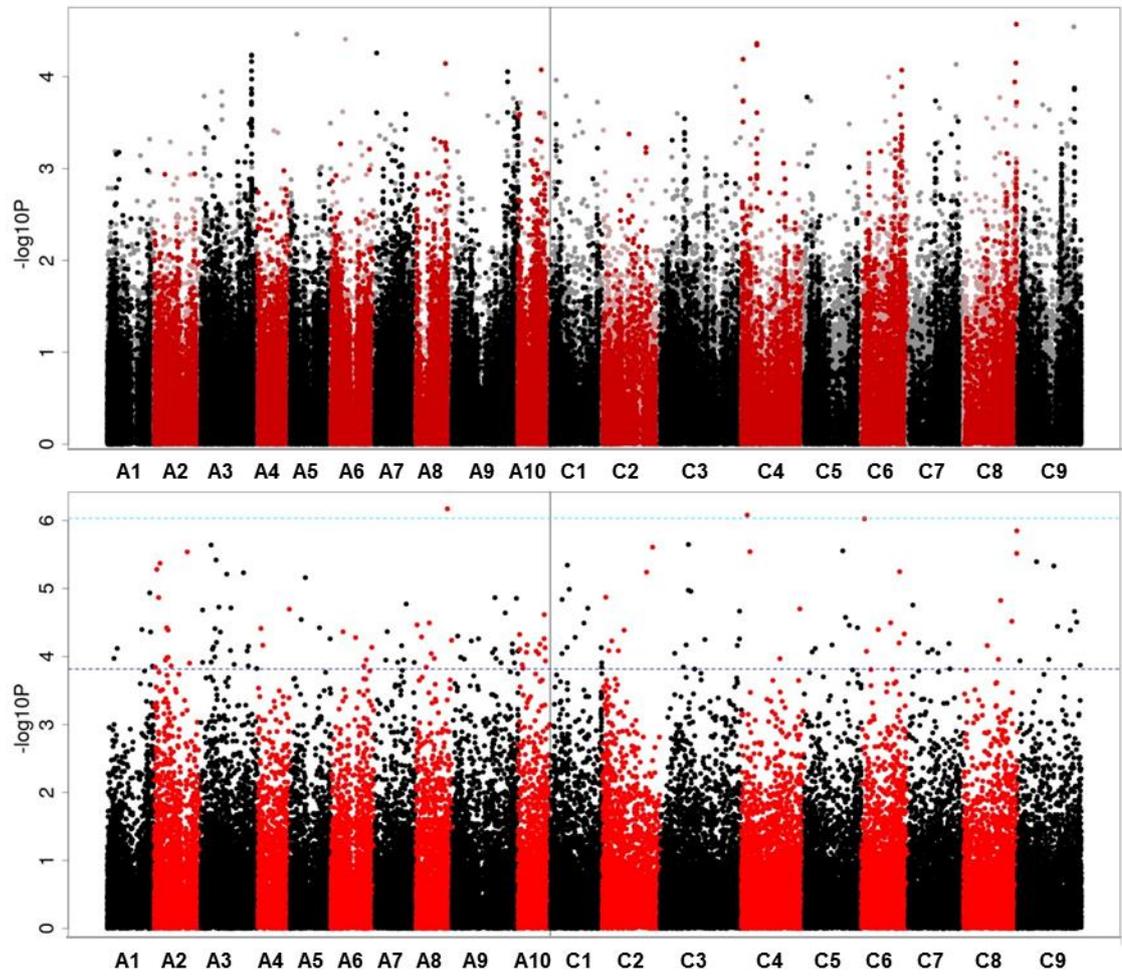
Appendix 6 AT output for S (mg/kg DW) seed on the 274 diversity panel



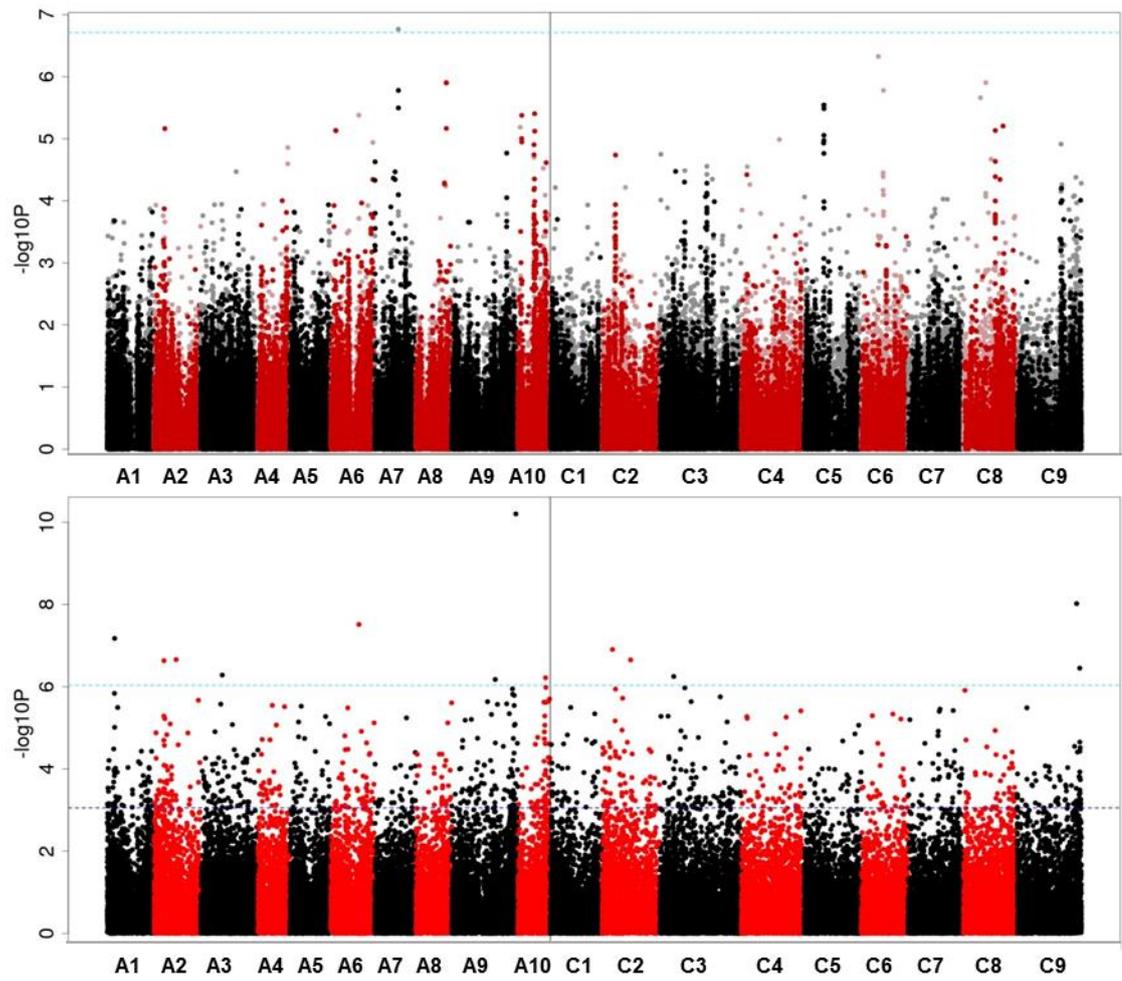
Appendix 7 AT output for Cu (mg/kg DW) leaf on the 274 diversity panel



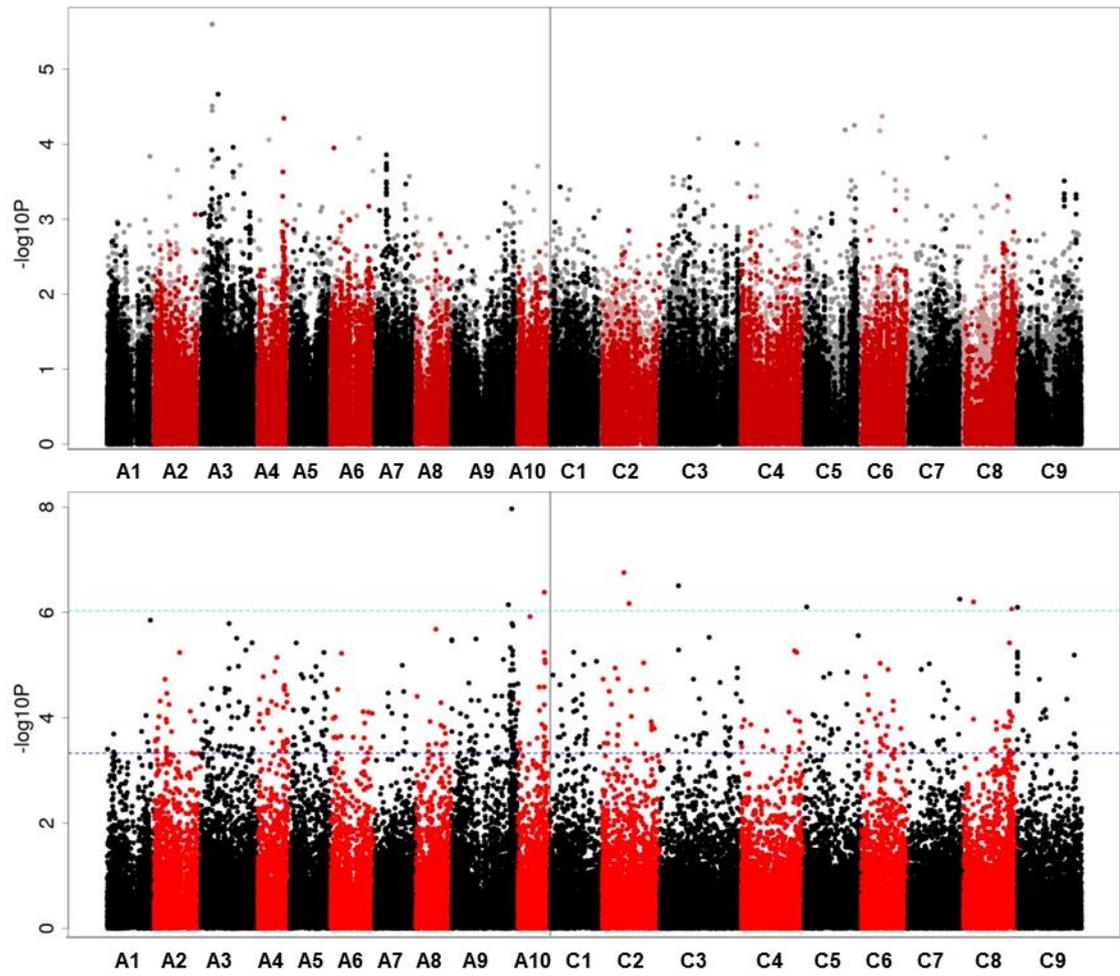
Appendix 8 AT output for Cd (mg/kg DW) leaf on the 274 diversity panel



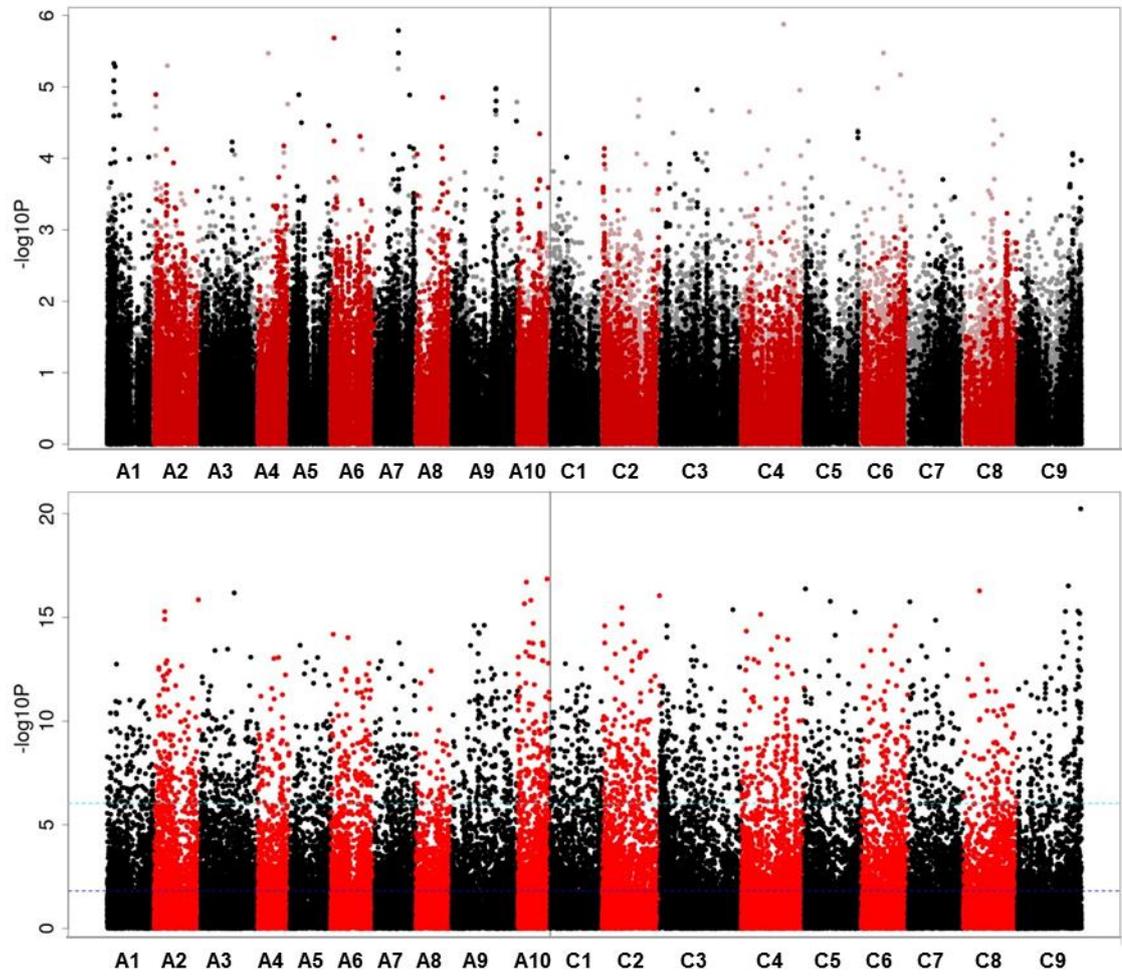
Appendix 9 AT output for Mn (mg/kg DW) leaf on the 274 diversity panel



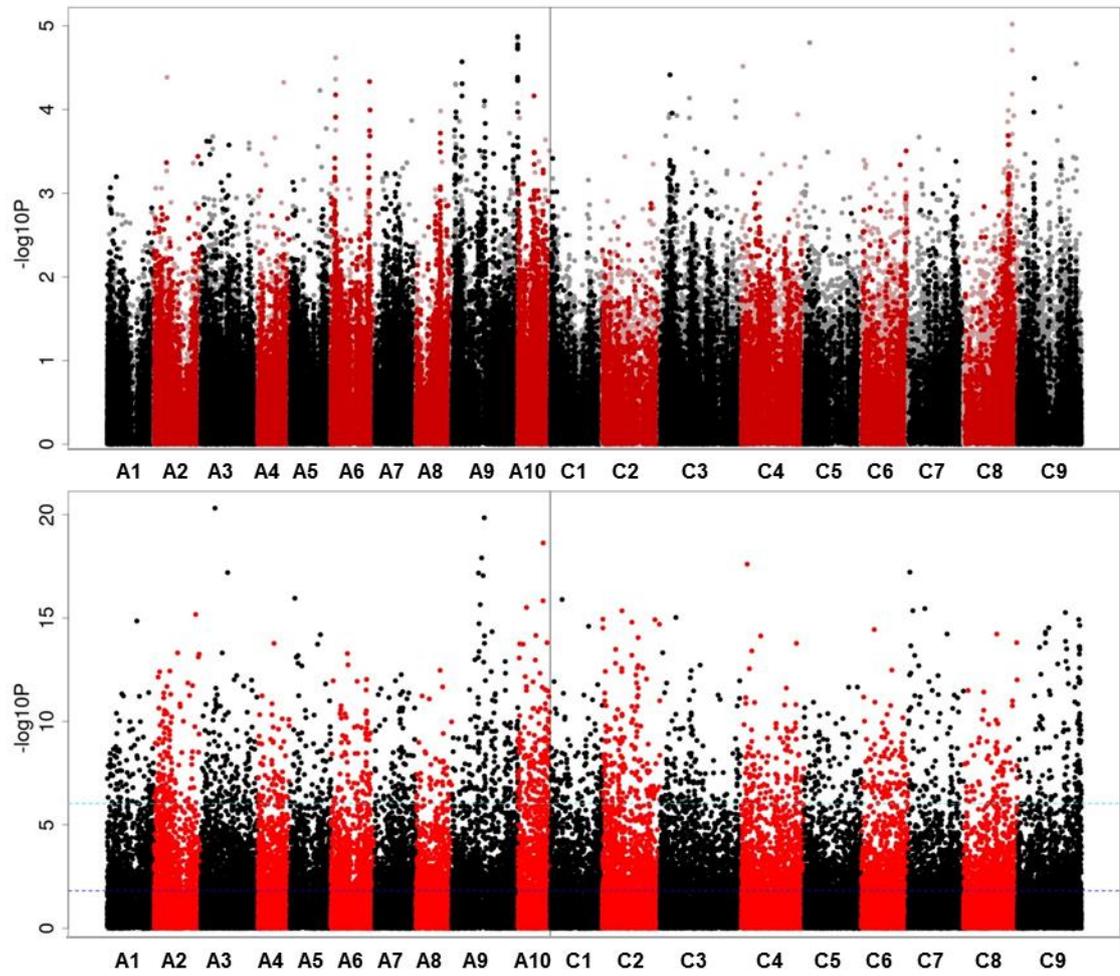
Appendix 10 AT output for Zn (mg/kg DW) leaf on the 274 diversity panel



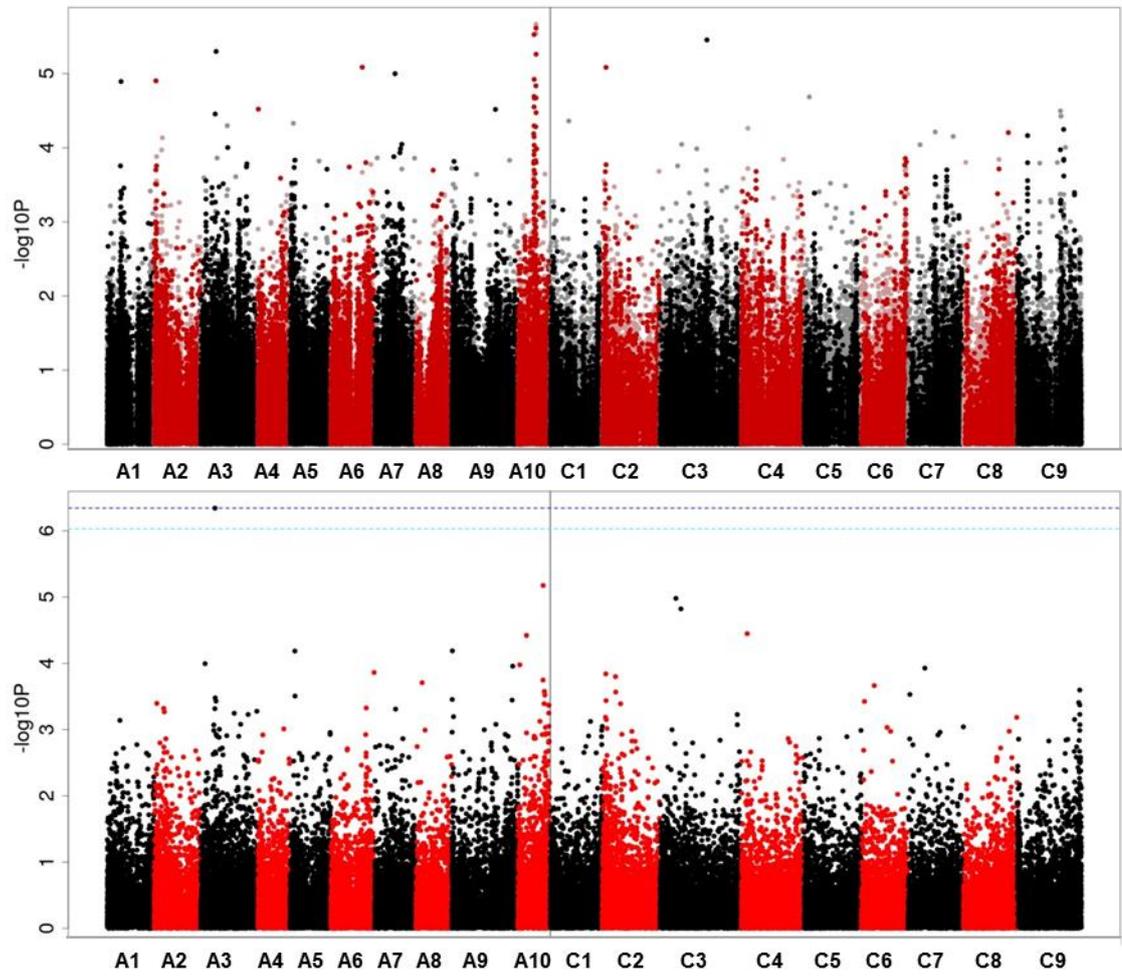
Appendix 11 AT output for Mo (mg/kg DW) leaf on the 274 diversity panel



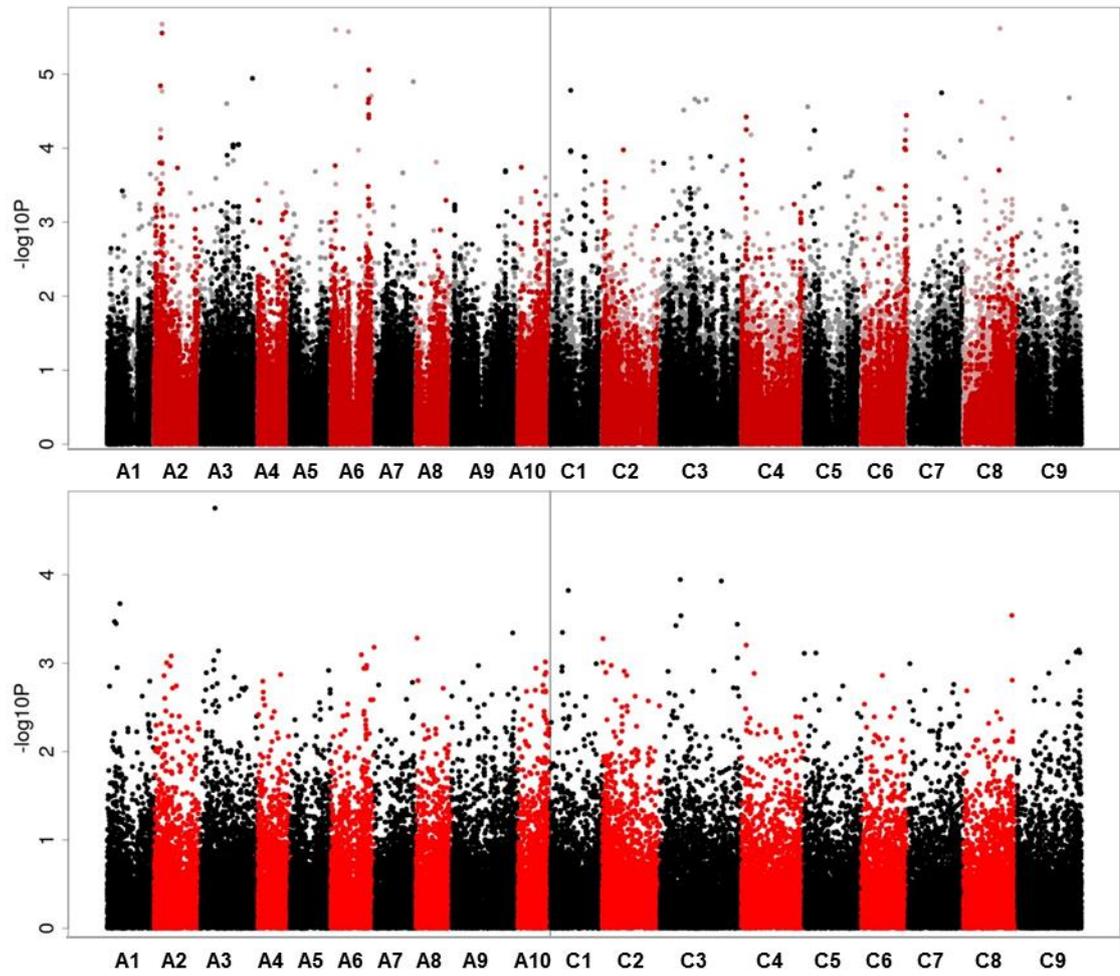
Appendix 12 AT output for S (mg/kg DW) leaf on the 274 diversity panel



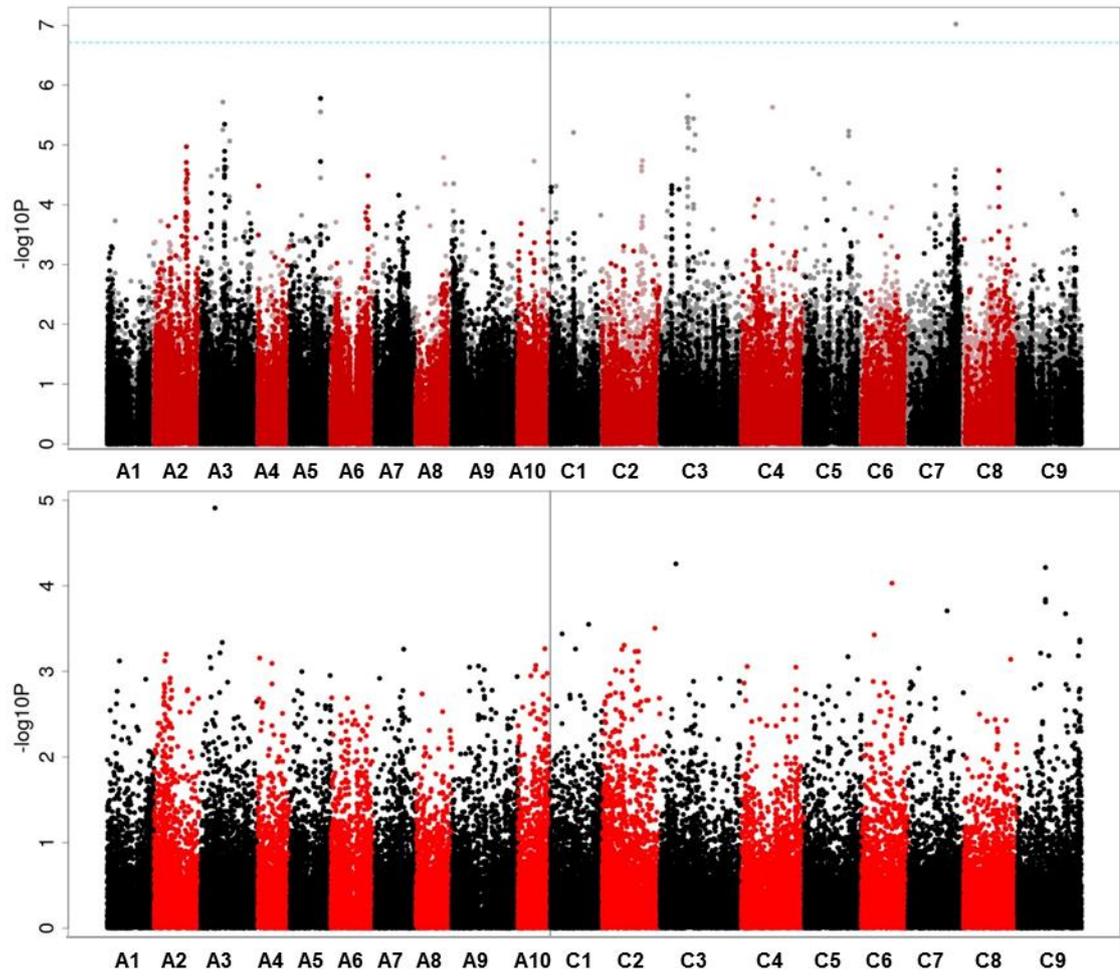
Appendix 13 AT output for Total leaf essential elements (mg/kg DW) on the 274 diversity panel



Appendix 14 AT output for Leaf essential coefficient of variation on the 274 diversity panel



Appendix 15 AT output for Flowering time (days, data produced by Mr Cándido José Martínez Ortuño) on the 274 diversity panel



T-Test outputs for Myb28/HAG1 (AT5G61420/SALK_136312C) in comparison to the wildtype control

Appendix 16 T-Test outputs for Myb28/HAG1 (AT5G61420/SALK_136312C) in comparison to the wildtype control in seeds (as mg/kg DW for each element).

The T-test value (t), probability (p) and degrees of freedom (df) are given for each element analysed. N: 6 for each group

Myb28/HAG1 vs control	B	Mg	P	S	K	Ca	Mn	Co	Cu	Zn	As	Mo	Cd	Pb
t	1.06	-3.69	-5.35	25.81	-5.58	-0.14	1.92	4.09	3.57	0.48	20.72	11.43	-9.18	-3.37
p	0.314	0.004	<0.001	<0.001	<0.001	0.892	0.084	0.002	0.005	0.639	<0.001	<0.001	<0.001	<0.001
df	10	10	10	10	10	10	10	10	10	10	10	10	10	10

Appendix 17 T-Test outputs for Myb28/HAG1 (AT5G61420/SALK_136312C) in comparison to the wildtype control in leaves (as mg/kg DW for each element).

The T-test value (t), probability (p) and degrees of freedom (df) are given for each element analysed.

Where there was evidence of unequal variance between the two means, they were estimated individually for each and an amended df calculated. N: 8 for each group

Myb28/HAG1 vs control	B	Na	Mg	P	S	K	Ca	Mn	As	Se	Sr	Mo	Cd	Ba
t	3.440	0.100	1.050	1.670	2.290	1.790	1.380	0.620	2.640	3.490	1.760	6.990	7.490	4.760
p	0.004	0.922	0.310	0.117	0.038	0.105	0.200	0.547	0.019	0.004	0.102	<0.001	<0.001	<0.001
df	14	7.49	14	14	14	9.62	9.01	14	14	14	13	14	14	14

T-Test outputs for Cd leaf HMA2 (AT2G36950/SALK_069207C) in comparison to the wildtype control

Appendix 18 T-test outputs for Cd leaf HMA2 (AT2G36950/SALK_069207C) in comparison to the wildtype control in leaves (as mg/kg DW for each element).

The T-test value (t), probability (p) and degrees of freedom (df) are given for each element analysed.

Where there was evidence of unequal variance between the two means, they were estimated individually for each and an amended df calculated. N: 8 for each group

Cd Leaf HMA 2 vs control	B	Na	Mg	P	S	K	Ca	Mn	As	Se	Sr	Mo	Cd	Ba
t	-5.19	0.13	-1.94	1.62	5.82	-1.1	0.15	-1.4	1.36	1.87	0.2	4.94	9.85	-5.74
p	<0.001	0.902	0.086	0.127	<0.001	0.288	0.886	0.195	0.196	0.096	0.848	<0.001	<0.001	<0.001
df	9.65	7.62	8.43	14	14	14	8.57	8.67	14	8.59	7.62	14	14	9.04

SOC1 correlation tables between flowering (days), FLC/SOC1 expression (days) and leaf nutrients (as mg/kg DW) when split

Appendix 19 *SOC1* correlation tables between flowering (as days), *FLC/SOC1* expression (as RPKM) and leaf nutrients (as mg/kg DW) when split correlation tables for flowering time data (as days, see 1.4.2.2), A3 FLC, C3 FLC and C4 SOC1 RPKM expression values, yield (as g from 6 plants), all elements within the leaf ionome (as mg/kg DW) and a range of summed totals (total ionome: sum of all elements in the ionome; essentials/non- essentials: sum of essential/non-essential elements included in the study; Mobile/immobile essential (see 6.3.7); sum of elements considered to be mobile/ immobile in *B. napus* (see 6.3.7, (Maillard et al., 2015))). Data for the diversity panel was split based on *SOC1* expression into quartiles/hinges from lowest (Q1) to highest (Q4) and tested for associations (Q1 top left, Q2 top right, Q3 bottom left and Q4 bottom right). R values are displayed to 3 decimal places, significant correlations ($p < 0.001$) are indicated in bold and underlined, whilst those for $p < 0.05$ are only in bold. The table is coloured to highlight positive correlations (red) and negative correlations (blue).

Q1 SOC1 (n:66)	A3_FLC_Cab0024 72.4	C3_FLC_Bo3g0054 70.1	C4_SOC1_Bo4g0248 50.1	Flowering time (days)
Ba	-0.2689	0.0082	-0.0805	-0.3346
Sr	-0.3373	-0.0264	-0.0833	-0.3287
Immobile essentials	-0.3212	-0.052	-0.1116	-0.3195
Ca	-0.3231	-0.0516	-0.1115	-0.3193
Mg	-0.4614	-0.0354	-0.0505	-0.3191
Non-essentials	-0.4119	-0.0955	-0.0455	-0.3137
Na	-0.4122	-0.0972	-0.0453	-0.31
Ti	-0.1776	-0.1934	-0.1229	-0.2687
S	-0.3137	-0.0722	0.0963	-0.228
Al	0.0061	0.0705	0.1138	-0.1916
Cd	-0.1783	0.0487	0.0023	-0.1888
Mo	-0.2199	-0.2585	-0.0408	-0.1624
Mn	-0.0039	-0.0348	-0.0365	-0.1536
Total ionome	-0.0555	0.0604	-0.1555	-0.1409
Essentials	-0.0325	0.0676	-0.1567	-0.1259
Fe	0.1974	-0.132	-0.0555	-0.1005
B	0.0024	-0.0209	-0.0876	-0.0981
Yield (g)	-0.0948	-0.1148	0.2178	-0.0919
As	0.1927	0.1511	0.001	-0.0513
Mobile essentials	0.0677	0.0933	-0.1409	-0.0384
Cs	0.0088	-0.0358	-0.1882	-0.0325
P	-0.0905	-0.0598	-0.0362	0.0597
Rb	0.19	0.1426	-0.1126	0.0623
Zn	-0.0517	-0.0437	-0.0307	0.0809
K	0.2854	0.1364	-0.1571	0.0904
Se	-0.2223	-0.0672	-0.0199	0.1259
Cu	0.044	-0.1475	0.0396	0.2163
Flowering (days)	0.261	-0.0115	-0.2481	NA

Q2 SOC1 (n:67)	A3_FLC_Cab0024 72.4	C3_FLC_Bo3g0054 70.1	C4_SOC1_Bo4g0248 50.1	Flowering time (days)
Non-essentials	-0.4713	-0.0158	-0.1636	-0.314
Na	-0.4745	-0.02	-0.1637	-0.3125
Mg	-0.3279	0.0453	-0.0537	-0.2908
Cu	-0.3212	-0.1757	-0.0775	-0.2858
Ca	-0.1762	0.1212	-0.1426	-0.2534
Immobile essentials	-0.1765	0.1252	-0.1436	-0.2511
Mo	-0.5099	0.0162	-0.0114	-0.2328
S	-0.4995	-0.0675	0.1634	-0.2068
Sr	-0.1065	0.0648	-0.171	-0.1998
Total ionome	-0.2674	0.1135	0.0257	-0.1779
Essentials	-0.2501	0.1169	0.0344	-0.1664
Cd	-0.1238	-0.2118	-0.1232	-0.164
Se	-0.0456	-0.0247	-0.0992	-0.1578
Ba	-0.0776	0.1438	-0.1921	-0.1377
Al	0.1342	0.2262	0.0872	-0.1262
Mobile essentials	-0.232	0.0964	0.0792	-0.1166
P	-0.2851	-0.2144	0.1674	-0.1059
As	-0.2083	0.0689	0.0049	-0.0994
Zn	0.0542	-0.0454	-0.1321	-0.0871
Ti	-0.0363	0.2679	-0.1303	-0.0272
Cs	-0.1453	0.2728	0.0683	-0.0152
K	-0.0663	0.1524	0.0594	-0.0141
Rb	-0.0031	0.0797	0.1073	0.0036
B	-0.4368	-0.0023	0.0509	0.022
Mn	-0.0259	0.3161	-0.0856	0.0999
Yield (g)	0.1082	0.0559	0.2679	0.1182
Fe	0.1331	0.2844	-0.042	0.178
Flowering (days)	0.5075	0.1258	0.1722	NA

Q3 SOC1 (n:66)	A3_FLC_Cab0024	C3_FLC_Bo3g0054	C4_SOC1_Bo4g0248	Flowering time
----------------	----------------	-----------------	------------------	----------------

Q4 SOC1 (n:67)	A3_FLC_Cab0024	C3_FLC_Bo3g0054	C4_SOC1_Bo4g0248	Flowering time
----------------	----------------	-----------------	------------------	----------------

	72.4	70.1	50.1	(days)
Mo	-0.255	-0.3265	0.0599	-0.4174
Mg	-0.2277	-0.0595	-0.0112	-0.3699
Total ionome	-0.3841	-0.1372	0.0679	-0.3661
S	-0.3252	-0.2151	-0.0745	-0.3578
Essentials	-0.3781	-0.1322	0.0714	-0.3557
Non-essentials	-0.2914	-0.1517	-0.0252	-0.3549
Na	-0.2908	-0.1541	-0.0252	-0.3541
Mobile essentials	-0.3558	-0.149	0.0822	-0.2981
B	-0.375	-0.1239	0.0799	-0.2697
P	-0.2891	-0.2906	0.1397	-0.2452
Immobile essentials	-0.1525	0.0135	-0.0119	-0.2431
Ca	-0.1498	0.0153	-0.0134	-0.2416
Sr	-0.1156	0.0596	-0.0197	-0.1944
Ba	-0.0525	0.0543	-0.0077	-0.1484
As	0.0464	-0.0748	-0.0934	-0.1474
Mn	-0.1442	-0.1272	0.1021	-0.1197
K	-0.229	-0.0611	0.1101	-0.1037
Rb	-0.2507	-0.0281	0.113	-0.0994
Cd	-0.2155	-0.0069	0.0262	-0.0973
Cu	-0.1413	-0.0971	0.2046	-0.0961
Ti	-0.1422	-0.0421	0.0239	-0.0687
Cs	-0.0605	-0.0396	0.0333	-0.0554
Al	0.0182	0.0498	-0.0563	-0.0301
Zn	-0.2032	-0.0057	0.0798	-0.0128
Yield (g)	0.109	0.1521	-0.0542	0.0143
Se	-0.0902	-0.0433	0.0536	0.0613
Fe	-0.0295	-0.1028	0.1492	0.0676
Flowering (days)	0.4537	0.3438	-0.0159	NA

	72.4	70.1	50.1	(days)
Total ionome	-0.4078	-0.1543	0.368	-0.4615
Essentials	-0.3994	-0.1417	0.3662	-0.448
Non-essentials	-0.3387	-0.2598	0.2239	-0.4391
Na	-0.3363	-0.2599	0.2179	-0.4382
Mobile essentials	-0.3255	-0.1135	0.2092	-0.3822
Cd	-0.3027	-0.1375	0.4118	-0.3599
Mg	-0.2689	-0.1468	0.413	-0.3557
Mo	-0.4262	-0.3173	0.1895	-0.3555
Immobile essentials	-0.2563	-0.0951	0.4249	-0.2511
Ca	-0.2549	-0.0956	0.426	-0.2499
K	-0.1554	-0.0643	0.0716	-0.2447
As	-0.239	-0.2779	0.3211	-0.226
S	-0.3824	-0.1365	0.1339	-0.2046
Sr	-0.2452	-0.0696	0.3875	-0.1948
Rb	-0.0702	-0.1293	-0.0056	-0.1916
Mn	-0.1743	-0.0093	0.0898	-0.1759
Ti	-0.2284	-0.0501	0.2841	-0.1587
Cu	-0.2598	0.1442	-0.0786	-0.1568
Zn	-0.0929	0.0128	0.0701	-0.1255
Al	-0.1387	-0.0795	0.221	-0.1243
Ba	-0.1985	-0.1017	0.3222	-0.091
Se	-0.1234	0.1836	-0.0028	-0.0893
B	-0.2584	0.0067	0.092	-0.0685
Fe	-0.0703	0.1448	-0.1209	-0.0485
Cs	-0.0736	0.0827	-0.1256	-0.0187
Yield (g)	0.2998	0.2953	-0.1697	-0.0045
P	-0.0732	0.2182	-0.0986	0.0333
Flowering (days)	0.4156	0.176	-0.3902	NA

ANOVA outputs for flowering T-DNA lines for individual elements in comparison to a wildtype control

Appendix 20 ANOVA results for individual elements (in mg/kg DW) across flowering T-DNA lines (control: *A. thaliana* Col; CAPL 1 (AT2G38480/SALK_022304C), CAPL 2(AT2G38480/SALK_064085C), SOC1 (AT2G45660 /SALK_006054C), E2F 1(AT2G40550/SALK_076472C) and E2F 2 (AT2G40550/SALK_055089C)), under long (LD) and short (SD) day conditions. The F statistic (F), degrees of Freedom (df) and p value (p) are given, alongside the differences highlighted with post hoc Bonferroni analysis ($p < 0.05$). Where the means are significantly different the letters are not shared, for ease of viewing some differences are highlighted in yellow (where there is a significant difference between LD and SD within a T-DNA line). Lettering starts from the smallest mean in the group under analysis, i.e. the group/s with the lowest average will be lettered 'a' and the remaining means labelled in ascending order.

ANOVA	F	df	p	Bonferroni (P<0.05)												
				CAPL 1 LD	CAPL 1 SD	CAPL 2 LD	CAPL 2 SD	SOC 1 LD	SOC1 SD	E2F 1 LD	E2F 1 SD	E2F 2 LD	E2F 2 SD	Col LD	Col SD	
B	3.48	11, 59	<0.001	abc	abc	c	abc	abc	ab	bc	abc	abc	abc	abc	a	
Ca	0.82	11, 59	0.617													
Cd	42.32	11, 59	<0.001	b	b	b	b	b	b	b	a	a	b	a	a	
Cu	3.71	11, 59	<0.001	a	ab	a	ab	a	ab	a	a	a	a	a	ab	b
K	1.51	11, 59	0.153													
Mg	1.88	11, 59	0.061													
Mn	2.3	11, 59	0.02	a	a	a	a	a	a	a	a	a	a	a	a	
Mo	3.48	11, 59	<0.001	b	a	ab	ab	ab	a	ab	ab	ab	ab	ab	ab	
Na	1.03	11, 59	0.437													
P	14.01	11, 59	<0.001	a	cd	a	bcd	ab	cd	ab	cd	abc	d	bcd	d	
S	2.19	11, 59	0.027	a	a	a	a	a	a	a	a	a	a	a	a	
Zn	3.85	11, 59	<0.001	ab	ab	ab	ab	b	a	b	ab	ab	a	b	ab	
sum	2.09	11, 59	0.036	a	a	a	a	a	a	a	a	a	a	a	a	

Abbreviations

Abbreviation	Full terminology
AAS	Atomic Absorption Spectroscopy
ABC	ATP Binding Cassette transporter
AGI	Arabidopsis Genome Initiative
Al	Aluminium
AO	Aldehyde Oxidase
APS	Adenosine 5'-phosphosulfate
AT	Associative Transcriptomics
ATPase	Adenosinetriphosphatase
B	Boron
BASS	Sodium Bile Acid Symporter family
BBSRC	Biotechnology and Biological Sciences Research Council
BLAST	Basic Local Alignment Search Tool
BOR	Boron transporter
BP	Border Primer
BSA	Bulk Segregant Analysis
Ca	Calcium
CaCO ₃	Calcium carbonate
CAM	Crassulacean Acid Metabolism
CAX	Cation exchanger transporter
Cd	Cadmium
CDF	Cation Diffusion Facilitator
cDNA	Complementary DNA
CDS	Coding DNA Sequence
Cl	Chlorine
CNX	Co-factor for Nitrate reductase and Xanthine dehydrogenase
Co	Cobalt
CO ₂	Carbon dioxide
COPT	Copper Transporter
CRM	Certified Reference Material
Cu	Copper
df	Degrees of Freedom
DNA	Deoxyribonucleic acid
dsRNA	Double Stranded RNA
E2FL	E2F like protein
ECA	ER-type Ca ²⁺ ATPase
ER	Endoplasmic Reticulum
EST	Expressed Sequence Tag
EXAFS	Extended X-ray Absorption Fine Structure
FDR	False Discovery Rate
Fe	Iron
FLC	Flowering Locus C
FRET	Forster Resonance Energy Transfer
FST	Flanking Sequence Tag
GAPIT	Genomic Association and Prediction Integrated Tool

GEM	Gene Expression Marker
GFP	Green fluorescent protein
GM	Genetic Modification
GSH	Glutathione
GSL	Glucosinolate
GTR	Glucosinolate Transporter
GUS	Beta-glucuronidase
GWAS	Genome Wide Association Study
H ₂ S	Hydrogen sulfide
HAG	High Aliphatic Glucosinolate
Hg	Mercury
HIPP	Heavy metal Isoprenylated Plant Protein
HMA	Heavy Metal ATPase
HPLC	High Performance Liquid Chromatography
ICP-MS	Inductively Coupled Plasma Mass Spectrometry
ICP-O/AES	Inductively Coupled Plasma Optical/Atomic Emission Spectroscopy
IHP	Interhomeolog polymorphism
IRT	Iron Regulated Transporter
K	Potassium
LA-ICP-MS	Laser Ablation ICP-MS
LN/HN	Low Nitrogen / High Nitrogen
LOD	Limit Of Detection
LP	Left genomic Primer
LS/HS	Low Sulfur / High Sulfur
LZRT	LIV-1 Zinc Transporter subfamily
MAGIC	Multiparent Advanced Generation Inter-Cross
MAS	Marker Assisted Selection
Mg	Magnesium
MIR	Mid-Infrared spectroscopy
miRNA	MicroRNA
MLM	Mixed Linear Model
Mn	Manganese
Mo	Molybdenum
MoO ₄ ²⁻	Molybdate
MOT	Molybdenum transporter
MTP	Metal Tolerance Protein
Myb	Myeloblastosis family of transcription factors
N ₂	Atmospheric nitrogen
Na	Sodium
NAA	Neutron Activation Analysis
NaCl	Sodium Chloride
Na-EDTA	Sodium Ethylenediaminetetraacetic acid
NASC	National Arabidopsis Stock Centre
NCBI	National Centre for Biotechnology Information
NH ₃	Ammonia
Ni	Nickel

NIR	Near-Infrared spectroscopy
NRAMP	Natural Resistance-Associated Macrophage Protein (transporter)
NUE	Nutrient Use Efficiency
OAS	O-Acetylserine
OASTL	OAS thiollyase
P	Phosphorous
PAPS	3'-phosphoadenosine 5'-phosphosulfate
PCR	Polymerase Chain Reaction
PCRs	Plant Cadmium Resistance
PiiMs	Purdue Ionomics Information Management System
PSII	Photosystem 2
PSIKO	Population Structure Inference using Kernel-PCA and Optimisation
QTL	Quantitative Trait Loci
R	Correlation Coefficient
RIL	Recombinant Inbred Line
RIPR	Renewable Industrial Products from Rapeseed
RNA	Ribonucleic acid
RNA-Seq	Ribonucleic acid Sequencing
ROS	Reactive Oxygen Species
RP	Right genomic Primer
RPKM	Reads Per Kilobase per Million aligned reads
RSA	Root System Architecture
RT-PCR	Reverse transcription polymerase chain reaction
S	Sulfur
S/W	Spring / Winter
S ²⁻	Sulfide
SAT	Serine Acetyl-transferase
SBP	Selenium Binding Protein
Se	Selenium
Si	Silicon
SIGnAL	Salk Institute Genomic Analysis Laboratory
SIMS	Secondary Ion Mass Spectrometry
SIZ1	Small Ubiquitin-like Modifier (SUMO) E3 ligase
SNP	Single Nucleotide Polymorphism
SO ₂	Sulfur dioxide
SO ₃ ²⁻	Sulfite
SO ₄ ²⁻	Sulfate
SOC1	Supressor of Constans
SOD	Superoxide dismutase
SPL/SPB	SQUAMOSA Promoter Binding protein-Like
Sr	Strontium
SULTR	Sulfate transporter
S-XRF	Synchrotron X-ray fluorescence
TAIR	The Arabidopsis Information Resource
T-DNA	Transposon DNA
TEM-EDX	Transmission Electron Microscopy Energy-Dispersive X-ray spectroscopy

TILLING	Targeted Induced Local Lesions in Genomes
Tris-HCl	Tris-hydrochloride
UMAMIT	Usually Multiple Acids Move In and out Transporter
UPL	Ubiquitin Protein Ligase
UV	Ultraviolet
Vis	Visual
VIT	Vacuolar Iron Transporter
WGCNA	Weighted Gene Co-expression Network Analysis
WT	wild type
XANES	X-ray Absorption Near Edge Structure
XAS	X-ray Absorption Spectrometry
XRF	X-ray Fluorescence Spectrometry
YSL	Yellow Stripe Like transporter
ZFN	Zinc finger nucleases
ZIFL	Zinc Induced Facilitator Like
ZIP	Zinc transporter precursor
Zn	Zinc
ZnOH ⁺	Zinc hydroxide
ZTR	Zinc transporter

References

- Aarabi, F. et al. (2016). Sulfur deficiency-induced repressor proteins optimize glucosinolate biosynthesis in plants. *Science Advances*, 2 (10), pp.e1601087–e1601087. [Online]. Available at: doi:10.1126/sciadv.1601087.
- Abdel-Ghany, S. E. and Pilon, M. (2008). MicroRNA-mediated systemic down-regulation of copper protein expression in response to low copper availability in Arabidopsis. *Journal of Biological Chemistry*, 283 (23), pp.15932–15945. [Online]. Available at: doi:10.1074/jbc.M801406200.
- De Abreu-Neto, J. B. et al. (2013). Heavy metal-associated isoprenylated plant protein (HIPP): Characterization of a family of proteins exclusive to plants. *FEBS Journal*, 280 (7), pp.1604–1616. [Online]. Available at: doi:10.1111/febs.12159.
- Abreu, I. A. and Cabelli, D. E. (2010). Superoxide dismutases-a review of the metal-associated mechanistic variations. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1804 (2), pp.263–274. [Online]. Available at: doi:10.1016/j.bbapap.2009.11.005.
- Adrees, M. et al. (2015). The effect of excess copper on growth and physiology of important food crops: a review. *Environmental Science and Pollution Research*, 22 (11), pp.8148–8162. [Online]. Available at: doi:10.1007/s11356-015-4496-5.
- Alcock, T. D. et al. (2018). Species-Wide Variation in Shoot Nitrate Concentration, and Genetic Loci Controlling Nitrate, Phosphorus and Potassium Accumulation in Brassica napus L. *Frontiers in Plant Science*, 9 (October). [Online]. Available at: doi:10.3389/fpls.2018.01487.
- Alexander, D. H., Novembre, J. and Lange, K. (2009). Supplementary Material for Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome research*, 19 (9), pp.1655–1664. [Online]. Available at: doi:10.1101/gr.094052.109.
- Ali, B. et al. (2014). Hydrogen sulfide alleviates cadmium-induced morpho-physiological and ultrastructural changes in Brassica napus. *Ecotoxicology and Environmental Safety*, 110, pp.197–207. [Online]. Available at: doi:10.1016/j.ecoenv.2014.08.027.
- Ali, H., Khan, E. and Sajad, M. A. (2013). Phytoremediation of heavy metals-Concepts and applications. *Chemosphere*, 91 (7), pp.869–881. [Online]. Available at: doi:10.1016/j.chemosphere.2013.01.075.
- Allen, M. D. et al. (2007). Manganese deficiency in Chlamydomonas results in loss of

photosystem II and MnSOD function, sensitivity to peroxides, and secondary phosphorus and iron deficiency. *Plant physiology*, 143 (1), pp.263–277. [Online]. Available at: doi:10.1104/pp.106.088609 [Accessed 4 November 2014].

Allender, C. J. and King, G. J. (2010). Origins of the amphiploid species *Brassica napus* L. investigated by chloroplast and nuclear molecular markers. *BMC Plant Biology*, 10. [Online]. Available at: doi:10.1186/1471-2229-10-54.

Alloway (Eds), B. (2008). *Micronutrient deficiencies in global crop production*. Alloway, B. J. (Ed). Springer. [Online]. Available at: doi:10.1007/978-1-4020-6860-7 [Accessed 28 October 2014].

Alloway, B. J. (2009). Soil factors associated with zinc deficiency in crops and humans. *Environmental Geochemistry and Health*, 31 (5), pp.537–548. [Online]. Available at: doi:10.1007/s10653-009-9255-4.

Alloway, B. J. (2013). *Heavy metals in Soils*. Third Edit. Alloway, B. J. (Ed). London: Springer.

Alomari, D. Z. et al. (2017). Genome-Wide Association Study of Calcium Accumulation in Grains of European Wheat Cultivars. *Frontiers in Plant Science*, 8 (October), pp.1–11. [Online]. Available at: doi:10.3389/fpls.2017.01797.

Altschul, S. F. et al. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215 (3), pp.403–410. [Online]. Available at: doi:10.1016/S0022-2836(05)80360-2.

Andersen, T. G. et al. (2013). Integration of Biosynthesis and Long-Distance Transport Establish Organ-Specific Glucosinolate Profiles in Vegetative Arabidopsis. *The Plant Cell*, 25 (8), pp.3133–3145. [Online]. Available at: doi:10.1105/tpc.113.110890.

Andersen, T. G. and Halkier, B. A. (2014). Upon bolting the GTR1 and GTR2 transporters mediate transport of glucosinolates to the inflorescence rather than roots. *Plant signaling & behavior*, 9, pp.9–10. [Online]. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24481282>.

Andres-Colas, N. et al. (2006). The Arabidopsis heavy metal P-type ATPase HMA5 interacts with metallochaperones and functions in copper detoxification of roots. *Plant Journal*, 45 (2), pp.225–236. [Online]. Available at: doi:10.1111/j.1365-313X.2005.02601.x.

Antosiewicz, D. M. and Hennig, J. (2004). Overexpression of LCT1 in tobacco enhances the protective action of calcium against cadmium toxicity. *Environmental Pollution*, 129 (2), pp.237–245. [Online]. Available at: doi:10.1016/j.envpol.2003.10.025.

- Araki, R. et al. (2018). SPL7 locally regulates copper - homeostasis - related genes in Arabidopsis. *Journal of Plant Physiology*, 224–225 (March), pp.137–143. [Online]. Available at: doi:10.1016/j.jplph.2018.03.014.
- Arnon, D. and Stout, P. (1939). Molybdenum as an essential element for higher plants. *Plant physiology*, 14, pp.599–602. [Online]. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC437771/> [Accessed 23 November 2014].
- Assuncao, A. G. L. et al. (2010). Arabidopsis thaliana transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. *Proceedings of the National Academy of Sciences*, 107 (22), pp.10296–10301. [Online]. Available at: doi:10.1073/pnas.1004788107.
- Atwell, S. et al. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature*, 465 (7298), pp.627–631. [Online]. Available at: doi:10.1038/nature08800.
- Aukerman, M. J. and Sakai, H. (2003). Regulation of Flowering Time and Floral Organ Identity by a MicroRNA and Its APETALA2 -Like Target Genes. *The Plant cell*, 15 (November), pp.2730–2741. [Online]. Available at: doi:10.1105/tpc.016238.pression.
- Auld, D. S. (2001). Zinc coordination sphere in biochemical zinc sites. *BioMetals*, 14 (3–4), pp.271–313. [Online]. Available at: doi:10.1023/A:1012976615056.
- Balanzà, V. et al. (2018). Genetic control of meristem arrest and life span in Arabidopsis by a FRUITFULL-APETALA2 pathway. *Nature Communications*, 9 (1). [Online]. Available at: doi:10.1038/s41467-018-03067-5.
- Balanzà, V., Martínez-Fernández, I. and Ferrándiz, C. (2014). Sequential action of FRUITFULL as a modulator of the activity of the floral regulators SVP and SOC1. *Journal of Experimental Botany*, 65 (4), pp.1193–1203. [Online]. Available at: doi:10.1093/jxb/ert482.
- Balík, J. et al. (2006). The fluctuation of molybdenum content in oilseed rape plants after the application of nitrogen and sulphur fertilizers. *Plant, Soil and Environment*, 52 (7), pp.301–307.
- Bancroft, I. et al. (2011). Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. *Nature biotechnology*, 29 (8), pp.762–766. [Online]. Available at: doi:10.1038/nbt.1926 [Accessed 1 September 2014].
- Bancroft, I. et al. (2015). Collinearity analysis of Brassica A and C genomes based on an updated inferred unigene order. *Data in Brief*, 3, pp.51–55. [Online]. Available at:

doi:10.1016/j.dib.2015.01.004.

Bao, F., Azhakanandam, S. and Franks, R. G. (2010). SEUSS and SEUSS-LIKE transcriptional adaptors regulate floral and embryonic development in Arabidopsis. *Plant physiology*, 152 (2), pp.821–836. [Online]. Available at: doi:10.1104/pp.109.146183.

Barber, J. (2009). Photosynthetic energy conversion: natural and artificial. *Chemical Society reviews*, 38, pp.185–196. [Online]. Available at: doi:10.1039/b802262n [Accessed 22 March 2014].

Barbosa, J. T. P. et al. (2015). Microwave-assisted diluted acid digestion for trace elements analysis of edible soybean products. *Food Chemistry*, 175, pp.212–217. [Online]. Available at: doi:10.1016/j.foodchem.2014.11.092.

Baxter, I. et al. (2007a). Purdue Ionomics Information Management System. An Integrated Functional Genomics Platform. *Plant Physiology*, 143 (2), pp.600–611. [Online]. Available at: doi:10.1104/pp.106.092528.

Baxter, I. et al. (2007b). Purdue Ionomics Information Management System. An Integrated Functional Genomics Platform. *Plant Physiology*, 143 (2), pp.600–611. [Online]. Available at: doi:10.1104/pp.106.092528.

Baxter, I. et al. (2008). Variation in molybdenum content across broadly distributed populations of Arabidopsis thaliana is controlled by a mitochondrial molybdenum transporter (MOT1). *PLoS genetics*, 4 (2), p.e1000004. [Online]. Available at: doi:10.1371/journal.pgen.1000004 [Accessed 5 November 2014].

Baxter, I. et al. (2009). Root suberin forms an extracellular barrier that affects water relations and mineral nutrition in Arabidopsis. *PLoS Genetics*, 5 (5). [Online]. Available at: doi:10.1371/journal.pgen.1000492.

Baxter, I. et al. (2010). A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1;1. *PLoS Genetics*, 6 (11). [Online]. Available at: doi:10.1371/journal.pgen.1001193.

Baxter, I. (2015). Should we treat the ionome as a combination of individual elements, or should we be deriving novel combined traits? *Journal of Experimental Botany*, 66 (8), pp.2127–2131. [Online]. Available at: doi:10.1093/jxb/erv040.

Bazakos, C. et al. (2017). New Strategies and Tools in Quantitative Genetics: How to Go from

the Phenotype to the Genotype. *Annu. Rev. Plant Biol*, 68, pp.435–455. [Online]. Available at: doi:10.1146/annurev-arplant-042916.

BBSRC. (2014). *Award details: BBSRC Renewable Industrial Products from Rapeseed (RIPR) Programme*. [Online]. Available at: <http://www.bbsrc.ac.uk/pa/grants/AwardDetails.aspx?FundingReference=BB%2FL002124%2F1> [Accessed 14 November 2016].

Beekweelder, J. et al. (2008). The impact of the absence of aliphatic glucosinolates on insect herbivory in Arabidopsis. *PLoS ONE*, 3 (4). [Online]. Available at: doi:10.1371/journal.pone.0002068.

Belouchrani, A. S. et al. (2016). Phytoremediation of soil contaminated with Zn using Canola (*Brassica napus* L). *Ecological Engineering*, 95, pp.43–49. [Online]. Available at: doi:10.1016/j.ecoleng.2016.06.064.

Bernal, M. et al. (2012). Transcriptome Sequencing Identifies SPL7-Regulated Copper Acquisition Genes FRO4/FRO5 and the Copper Dependence of Iron Homeostasis in Arabidopsis. *the Plant Cell Online*, 24 (2), pp.738–761. [Online]. Available at: doi:10.1105/tpc.111.090431.

Bieluszewski, T. et al. (2015). AtEAF1 is a potential platform protein for Arabidopsis NuA4 acetyltransferase complex. *BMC Plant Biology*, 15 (1), pp.1–15. [Online]. Available at: doi:10.1186/s12870-015-0461-1.

Billard, V. et al. (2014). Copper-Deficiency in Brassica napus Induces Copper Remobilization, Molybdenum Accumulation and Modification of the Expression of Chloroplastic Proteins. *PLoS ONE*, 9 (10), p.e109889. [Online]. Available at: doi:10.1371/journal.pone.0109889 [Accessed 5 November 2014].

Billard, V. et al. (2015). Zn deficiency in Brassica napus induces Mo and Mn accumulation associated with chloroplast proteins variation without Zn remobilization. *Plant Physiology and Biochemistry*, 86, pp.66–71. [Online]. Available at: doi:10.1016/j.plaphy.2014.11.005.

Billard, V. et al. (2016). Mg deficiency affects leaf Mg remobilization and the proteome in Brassica napus. *Plant Physiology and Biochemistry*, 107, pp.337–343. [Online]. Available at: doi:10.1016/j.plaphy.2016.06.025.

Bittner, F. (2014). Molybdenum metabolism in plants and crosstalk to iron. *Frontiers in plant science*, 5, pp.1–6. [Online]. Available at: doi:10.3389/fpls.2014.00028 [Accessed 17 November

2014].

Bloem, E., Haneklaus, S. and Schnug, E. (2005). Relation between total sulphur analysed by ICP-AES and glucosinolates in oilseed rape and Indian mustard seeds. *Landbauforschung Völkenrode*, 55 (4), pp.205–210.

Bloem, E., Haneklaus, S. and Schnug, E. (2007). Changes in the sulphur and glucosinolate content of developing pods and seeds of oilseed rape (*Brassica napus* L .) in relation to different cultivars. *Landbauforschung Völkenrode*, 4 (57), pp.297–306.

Bones, a M. and Rossiter, J. T. (1996). The myrosinase-glucosinolate system, its organisation and biochemistry. *Physiologia Plantarum*, 97, pp.194–208. [Online]. Available at: doi:10.1034/j.1399-3054.1996.970128.x.

Bortesi, L. and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Advances*, 33 (1), pp.41–52. [Online]. Available at: doi:10.1016/j.biotechadv.2014.12.006.

Bourke, P. M. et al. (2018). Tools for Genetic Studies in Experimental Populations of Polyploids. *Frontiers in Plant Science*, 9 (April), pp.1–17. [Online]. Available at: doi:10.3389/fpls.2018.00513.

Bowler, C. and Slooten, L. (1991). Manganese superoxide dismutase can reduce cellular damage mediated by oxygen radicals in transgenic plants. *The EMBO journal*, 10 (7), pp.1723–1732. [Online]. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/pmc452843/> [Accessed 25 November 2014].

Bozdog, G. O. et al. (2014). Characterization of a cDNA from *Beta maritima* that confers nickel tolerance in yeast. *Gene*, 538 (2), pp.251–257. [Online]. Available at: doi:10.1016/j.gene.2014.01.052.

Bradbury, P. J. et al. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23 (19), pp.2633–2635. [Online]. Available at: doi:10.1093/bioinformatics/btm308.

Brennan, R. F. and Bolland, M. D. A. (2011). Foliar spray experiments identify no boron deficiency but molybdenum and manganese deficiency for canola grain production on acidified sandy gravel soils in Southwestern Australia. *Journal of Plant Nutrition*, 34 (8), pp.1186–1197. [Online]. Available at: doi:10.1080/01904167.2011.558161.

- Brennan, R. F. and Bolland, M. D. A. (2015). Manganese Deficiency in Canola Induced by Liming to Ameliorate Soil Acidity. *Journal of Plant Nutrition*, 38 (6), pp.886–903. [Online]. Available at: doi:10.1080/01904167.2014.964362.
- Briat, J.-F. et al. (2015). Integration of P, S, Fe, and Zn nutrition signals in *Arabidopsis thaliana*: potential involvement of PHOSPHATE STARVATION RESPONSE 1 (PHR1). *Frontiers in Plant Science*, 06 (April), pp.1–16. [Online]. Available at: doi:10.3389/fpls.2015.00290.
- Bristow, A. W. and Garwood, E. A. (1984). Deposition of sulphur from the atmosphere and the sulphur balance in four soils under grass. *The Journal of Agricultural Science*, 103 (2), pp.463–468. [Online]. Available at: doi:10.1017/S0021859600047444.
- Broadley, M. R. et al. (2007). Zinc in plants: Tansley review. *New Phytologist*, 173 (4), pp.677–702. [Online]. Available at: doi:10.1111/j.1469-8137.2007.01996.x.
- Brunel-Muguet, S. et al. (2015). SuMoToRI, an Ecophysiological Model to Predict Growth and Sulfur Allocation and Partitioning in Oilseed Rape (*Brassica napus* L.) Until the Onset of Pod Formation. *Frontiers in plant science*, 6 (November), p.993. [Online]. Available at: doi:10.3389/fpls.2015.00993.
- Buhtz, A. et al. (2008). Identification and characterization of small RNAs from the phloem of *Brassica napus*. *Plant Journal*, 53 (5), pp.739–749. [Online]. Available at: doi:10.1111/j.1365-313X.2007.03368.x.
- Burkhead, J. L. et al. (2009). Copper homeostasis. *New Phytologist*, 182 (4), pp.799–816. [Online]. Available at: doi:10.1111/j.1469-8137.2009.02846.x.
- Cakmak, I. (2002). Plant nutrition research : Priorities to meet human needs for food in sustainable ways. *Plant and Soil*, 247, pp.3–24. [Online]. Available at: doi:10.1023/a:1021194511492.
- Campbell, D. et al. (2017). Growth and metal uptake of canola and sunflower along a thickness gradient of organic-rich covers over metal mine tailings. *Ecological Engineering*, 109 (June), pp.133–139. [Online]. Available at: doi:10.1016/j.ecoleng.2017.08.019.
- Campbell, W. H. (1999). NITRATE REDUCTASE STRUCTURE, FUNCTION AND REGULATION: Bridging the Gap between Biochemistry and Physiology. *Annual review of plant physiology and plant molecular biology*, 50, pp.277–303. [Online]. Available at: doi:10.1146/annurev.arplant.50.1.277.

- Campos, A. C. A. L. et al. (2017). Natural variation in *Arabidopsis thaliana* reveals shoot ionome, biomass, and gene expression changes as biomarkers for zinc deficiency tolerance. *Journal of Experimental Botany*, 68 (13), pp.3643–3656. [Online]. Available at: doi:10.1093/jxb/erx191.
- Carrier, P., Baryla, A. and Havaux, M. (2003). Cadmium distribution and microlocalization in oilseed rape (*Brassica napus*) after long-term growth on cadmium-contaminated soil. *Planta*, 216 (6), pp.939–950. [Online]. Available at: doi:10.1007/s00425-002-0947-6.
- Carrió-Seguí, A. et al. (2015). Defective copper transport in the *copt5* mutant affects cadmium tolerance. *Plant and Cell Physiology*, 56 (3), pp.442–454. [Online]. Available at: doi:10.1093/pcp/pcu180.
- Castaigns, L. et al. (2009). The nodule inception-like protein 7 modulates nitrate sensing and metabolism in *Arabidopsis*. *Plant Journal*, 57 (3), pp.426–435. [Online]. Available at: doi:10.1111/j.1365-3113X.2008.03695.x.
- Chalhoub, B. et al. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, 345 (6199), pp.950–953. [Online]. Available at: doi:10.1126/science.1253435.
- Chandra, N. and Pandey, N. (2016). Role of sulfur nutrition in plant and seed metabolism of *Glycine max* L. *Journal of Plant Nutrition*, 39 (8), pp.1103–1111. [Online]. Available at: doi:10.1080/01904167.2016.1143495.
- Chao, D.-Y. et al. (2011). Sphingolipids in the Root Play an Important Role in Regulating the Leaf Ionome in *Arabidopsis thaliana*. *The Plant Cell*, 23 (3), pp.1061–1081. [Online]. Available at: doi:10.1105/tpc.110.079095.
- Chen, C.-C. et al. (2011). *Arabidopsis* SUMO E3 Ligase SIZ1 Is Involved in Excess Copper Tolerance. *Plant Physiology*, 156 (4), pp.2225–2234. [Online]. Available at: doi:10.1104/pp.111.178996.
- Chen, H. and Xiong, L. (2010). The bifunctional abiotic stress signalling regulator and endogenous RNA silencing suppressor FIERY1 is required for lateral root formation. *Plant, Cell and Environment*, 33 (12), pp.2180–2190. [Online]. Available at: doi:10.1111/j.1365-3040.2010.02218.x.
- Chen, J. et al. (2009). Magnesium transporter AtMGT9 is essential for pollen development in *Arabidopsis*. *Cell research*, 19 (7), pp.887–898. [Online]. Available at: doi:10.1038/cr.2009.58

[Accessed 6 November 2014].

Chen, J. G. et al. (2006). RACK1 mediates multiple hormone responsiveness and developmental processes in Arabidopsis. *Journal of Experimental Botany*, 57 (11), pp.2697–2708. [Online]. Available at: doi:10.1093/jxb/erl035.

Chen, L. et al. (2018). Genome-Wide Association Study of Cadmium Accumulation at the Seedling Stage in Rapeseed (*Brassica napus* L.). *Frontiers in plant science*, 9 (April), pp.1–15. [Online]. Available at: doi:10.3389/fpls.2018.00375.

Choppala, G. et al. (2014). Cellular Mechanisms in Higher Plants Governing Tolerance to Cadmium Toxicity. *Critical Reviews in Plant Sciences*, 33 (5), pp.374–391. [Online]. Available at: doi:10.1080/07352689.2014.903747.

Christeller, J. T. (1981). The effects of bivalent cations on ribulose biphosphate carboxylase/oxygenase. *The Biochemical journal*, 1 (193), pp.839–844.

Christensen, T. (1984). Cadmium soil sorption at low concentrations. I.Effect of time, cadmium load, pH, and calcium. *Water, Air, & Soil Pollution*, 21, pp.105–114.

Clemens, S. (2006). Toxic metal accumulation, responses to exposure and mechanisms of tolerance in plants. *Biochimie*, 88 (11), pp.1707–1719. [Online]. Available at: doi:10.1016/j.biochi.2006.07.003 [Accessed 1 August 2014].

Clemens, S. et al. (2013). Plant science: The key to preventing slow cadmium poisoning. *Trends in Plant Science*, 18 (2), pp.92–99. [Online]. Available at: doi:10.1016/j.tplants.2012.08.003.

Connolly, E. L. (2002). Expression of the IRT1 Metal Transporter Is Controlled by Metals at the Levels of Transcript and Protein Accumulation. *the Plant Cell Online*, 14 (6), pp.1347–1357. [Online]. Available at: doi:10.1105/tpc.001263.

Coolong, T. W. and Randle, W. M. (2003). Zinc concentration in hydroponic solution culture influences zinc and sulfur accumulation in *Brassica rapa* L. *Journal of Plant Nutrition*, 26 (5), pp.949–959. [Online]. Available at: doi:10.1081/PLN-120020067.

Crevillén, P. et al. (2014). Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state. *Nature*, 515 (7528), pp.587–590. [Online]. Available at: doi:10.1038/nature13722.

Curie, C. et al. (2009). Metal movement within the plant: Contribution of nicotianamine and yellow stripe 1-like transporters. *Annals of Botany*, 103 (1), pp.1–11. [Online]. Available at:

doi:10.1093/aob/mcn207.

Curtin, D. and Syers, J. K. (1990). Extractability and adsorption of sulphate in soils. *Journal of Soil Science*, 41 (2), pp.305–312. [Online]. Available at: doi:10.1111/j.1365-2389.1990.tb00065.x.

Danku, J. et al. (2013). Large-scale plant ionomics. In: *Plant Mineral Nutrients: Methods and Protocols*. New York: Humana Press. pp.255–276.

Deinlein, U. et al. (2012). Elevated Nicotianamine Levels in Arabidopsis halleri Roots Play a Key Role in Zinc Hyperaccumulation. *the Plant Cell Online*, 24 (2), pp.708–723. [Online]. Available at: doi:10.1105/tpc.111.095000.

Delhaize, E. et al. (2003). Genes encoding proteins of the cation diffusion facilitator family that confer manganese tolerance. *The Plant Cell*, 15 (May), pp.1131–1142. [Online]. Available at: doi:10.1105/tpc.009134.al. [Accessed 20 November 2013].

Delhaize, E. et al. (2007). A role for the AtMTP11 gene of Arabidopsis in manganese transport and tolerance. *The Plant journal : for cell and molecular biology*, 51 (2), pp.198–210. [Online]. Available at: doi:10.1111/j.1365-313X.2007.03138.x [Accessed 22 April 2014].

Deng, F. et al. (2013). A Member of the Heavy Metal P-Type ATPase OsHMA5 Is Involved in Xylem Loading of Copper in Rice. *Plant Physiology*, 163 (3), pp.1353–1362. [Online]. Available at: doi:10.1104/pp.113.226225.

Desbrosses-Fonrouge, A. G. et al. (2005). Arabidopsis thaliana MTP1 is a Zn transporter in the vacuolar membrane which mediates Zn detoxification and drives leaf Zn accumulation. *FEBS Letters*, 579 (19), pp.4165–4174. [Online]. Available at: doi:10.1016/j.febslet.2005.06.046.

Devlin, B. and Roeder. (1999). Genomic control for association studies. *Biometrics*, 55 (December), pp.997–1004.

Djingova, R. et al. (2013). Multielement analytical spectroscopy in plant ionomics research. *Applied Spectroscopy Reviews*, 48 (5), pp.384–424. [Online]. Available at: doi:10.1080/05704928.2012.703153.

Doebley, J., Stec, A. and Hubbard, L. (1997). The evolution of apical dominance in maize. *Nature*, 386, pp.485–488.

Dresbøll, D. B. and Thorup-Kristensen, K. (2014). Will breeding for nitrogen use efficient crops lead to nitrogen use efficient cropping systems?: A simulation study of G×E×M interactions.

Euphytica, 199 (1–2), pp.97–117. [Online]. Available at: doi:10.1007/s10681-014-1199-9.

Droux, M. et al. (1998). Interactions between serine acetyltransferase and O-acetylserine (thiol) lyase in higher plants. Structural and kinetic properties of the free and bound enzymes. *European Journal of Biochemistry*, 255 (1), pp.235–245. [Online]. Available at: doi:10.1046/j.1432-1327.1998.2550235.x.

Dubousset, L., Etienne, P. and Avice, J. C. (2010). Is the remobilization of S and N reserves for seed filling of winter oilseed rape modulated by sulphate restrictions occurring at different growth stages? *Journal of experimental botany*, 61 (15), pp.4313–4324. [Online]. Available at: doi:10.1093/jxb/erq233.

Dynowski, M. et al. (2008). Plant plasma membrane water channels conduct the signalling molecule H₂O₂. *Biochemical Journal*, 414 (1), pp.53–61. [Online]. Available at: doi:10.1042/BJ20080287.

Ebbs, S. D. and Kochian, L. V. (1997). Toxicity of Zinc and Copper to Brassica Species: Implications for Phytoremediation. *Journal of Environment Quality*, 26 (3), p.776. [Online]. Available at: doi:10.2134/jeq1997.00472425002600030026x.

ECJ. (2018). *Organisms obtained by mutagenesis are GMOs and are, in principle, subject to the obligations laid down by the GMO Directive*. pp.1–2.

Eggert, K. and von Wirén, N. (2016). The role of boron nutrition in seed vigour of oilseed rape (*Brassica napus* L.). *Plant and Soil*, 402 (1–2), pp.63–76. [Online]. Available at: doi:10.1007/s11104-015-2765-1.

El-Jaoual, T. and Cox, D. A. (1998). Manganese toxicity in plants. *Journal of Plant Nutrition*, 21 (2), pp.353–386. [Online]. Available at: doi:10.1080/01904169809365409 [Accessed 22 April 2014].

Engel, E. et al. (2002). Flavor-active compounds potentially implicated in cooked cauliflower acceptance. *Journal of Agricultural and Food Chemistry*, 50 (22), pp.6459–6467. [Online]. Available at: doi:10.1021/jf025579u.

Englbrecht, C. C., Schoof, H. and Böhm, S. (2004). Conservation, diversification and expansion of C2H2 zinc finger proteins in the *Arabidopsis thaliana* genome. *BMC Genomics*, 5, pp.1–17. [Online]. Available at: doi:10.1186/1471-2164-5-39.

Erisman, J. W. et al. (2008). How a century of ammonia synthesis changed the world. *Nature*

Geoscience, 1 (10), pp.636–639. [Online]. Available at: doi:10.1038/ngeo325 [Accessed 19 November 2014].

Esteban, E., Deza, M. J. and Zornoza, P. (2013). Kinetics of mercury uptake by oilseed rape and white lupin: influence of Mn and Cu. *Acta Physiologiae Plantarum*, 35 (7), pp.2339–2344. [Online]. Available at: doi:10.1007/s11738-013-1253-6 [Accessed 24 October 2014].

Evanno, G., Regnaut, S. and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14 (8), pp.2611–2620. [Online]. Available at: doi:10.1111/j.1365-294X.2005.02553.x.

Fabro, G. et al. (2008). Genome-Wide Expression Profiling Arabidopsis at the Stage of Golovinomyces cichoracearum Haustorium Formation. *Plant Physiology*, 146 (3), pp.1421–1439. [Online]. Available at: doi:10.1104/pp.107.111286.

Fageria, N. K., Baligar, V. C. and Li, Y. C. (2008). The Role of Nutrient Efficient Plants in Improving Crop Yields in the Twenty First Century. *Journal of Plant Nutrition*, 31 (6), pp.1121–1157. [Online]. Available at: doi:10.1080/01904160802116068 [Accessed 14 October 2014].

Falk, K. L., Tokuhisa, J. G. and Gershenzon, J. (2007). The effect of sulfur nutrition on plant glucosinolate content: Physiology and molecular mechanisms. *Plant Biology*, 9 (5), pp.573–581. [Online]. Available at: doi:10.1055/s-2007-965431.

Feldmann, K. A. and Marks, D. M. (1987). Agrobacterium-mediated transformation of germinating seeds of Arabidopsis thaliana: A non-tissue culture approach. *MGG Molecular & General Genetics*, 208 (1–2), pp.1–9. [Online]. Available at: doi:10.1007/BF00330414.

Fernández, V. and Brown, P. H. (2013). From plant surface to plant metabolism: the uncertain fate of foliar-applied nutrients. *Frontiers in Plant Science*, 4 (July), pp.1–5. [Online]. Available at: doi:10.3389/fpls.2013.00289.

Fernando, D. R. et al. (2010). Characterization of foliar manganese (Mn) in Mn (hyper)accumulators using X-ray absorption spectroscopy. *New Phytologist*, 188 (4), pp.1014–1027. [Online]. Available at: doi:10.1111/j.1469-8137.2010.03431.x.

Fernando, D. R. et al. (2012). Plant homeostasis of foliar manganese sinks: specific variation in hyperaccumulators. *Planta*, 236 (5), pp.1459–1470. [Online]. Available at: doi:10.1007/s00425-012-1699-6 [Accessed 5 October 2013].

Fernando, D. R. and Lynch, J. P. (2015). Manganese phytotoxicity: New light on an old problem.

Annals of Botany, 116 (3), pp.313–319. [Online]. Available at: doi:10.1093/aob/mcv111.

Finbow, M. E. and Harrison, M. A. (1997). The vacuolar H⁺-ATPase: a universal proton pump of eukaryotes. *The Biochemical journal*, 712 (Pt 3), pp.697–712. [Online]. Available at: doi:10.1042/bj3240697.

Fitzgerald, T. L., Kazan, K. and Manners, J. M. (2012). The Application of Reverse Genetics to Polyploid Plant Species. *Critical Reviews in Plant Sciences*, 31 (2), pp.181–200. [Online]. Available at: doi:10.1080/07352689.2011.635538.

Fitzpatrick, K. L., Tyerman, S. D. and Kaiser, B. N. (2008). Molybdate transport through the plant sulfate transporter SHST1. *FEBS Letters*, 582 (10), pp.1508–1513. [Online]. Available at: doi:10.1016/j.febslet.2008.03.045.

Fornara, F. and Coupland, G. (2009). Plant Phase Transitions Make a SPLash. *Cell*, 138 (4), pp.625–627. [Online]. Available at: doi:10.1016/j.cell.2009.08.011.

Führs, H. et al. (2010). Physiological and proteomic characterization of manganese sensitivity and tolerance in rice (*Oryza sativa*) in comparison with barley (*Hordeum vulgare*). *Annals of Botany*, 105 (7), pp.1129–1140. [Online]. Available at: doi:10.1093/aob/mcq046 [Accessed 6 October 2013].

Funakawa, H. and Miwa, K. (2015). Synthesis of borate cross-linked rhamnogalacturonan II. *Frontiers in Plant Science*, 6 (April), pp.1–8. [Online]. Available at: doi:10.3389/fpls.2015.00223.

Gallego, S. M. et al. (2012). Unravelling cadmium toxicity and tolerance in plants: Insight into regulatory mechanisms. *Environmental and Experimental Botany*, 83, pp.33–46. [Online]. Available at: doi:10.1016/j.envexpbot.2012.04.006 [Accessed 29 October 2014].

Galvao, V. C. et al. (2012). Spatial control of flowering by DELLA proteins in *Arabidopsis thaliana*. *Development*, 139 (21), pp.4072–4082. [Online]. Available at: doi:10.1242/dev.080879.

Gampala, S. S. et al. (2007). An Essential Role for 14-3-3 Proteins in Brassinosteroid Signal Transduction in *Arabidopsis*. *Developmental Cell*, 13 (2), pp.177–189. [Online]. Available at: doi:10.1016/j.devcel.2007.06.009.

Gasber, a et al. (2011). Identification of an *Arabidopsis* solute carrier critical for intracellular transport and inter-organ allocation of molybdate. *Plant biology*, 13 (5), pp.710–718. [Online]. Available at: doi:10.1111/j.1438-8677.2011.00448.x [Accessed 5 November 2014].

- Gigolashvili, T. and Kopriva, S. (2014). Transporters in plant sulfur metabolism. *Frontiers in plant science*, 5 (September), p.442. [Online]. Available at: doi:10.3389/fpls.2014.00442.
- Gill, S. S. and Tuteja, N. (2010). Polyamines and abiotic stress tolerance in plants. *Plant Signaling & Behavior*, 5 (1), pp.26–33. [Online]. Available at: doi:10.4161/psb.5.1.10291.
- Gisbert, C. et al. (2006). Tolerance and accumulation of heavy metals by Brassicaceae species grown in contaminated soils from Mediterranean regions of Spain. *Environmental and Experimental Botany*, 56 (1), pp.19–27. [Online]. Available at: doi:10.1016/j.envexpbot.2004.12.002.
- Glover, N. M., Redestig, H. and Dessimoz, C. (2016). Homoeologs: What Are They and How Do We Infer Them? *Trends in Plant Science*, 21 (7), pp.609–621. [Online]. Available at: doi:10.1016/j.tplants.2016.02.005.
- Gol, L., Tomé, F. and Von Korff, M. (2017). Floral transitions in wheat and barley: Interactions between photoperiod, abiotic stresses, and nutrient status. *Journal of Experimental Botany*, 68 (7), pp.1399–1410. [Online]. Available at: doi:10.1093/jxb/erx055.
- Gong, J.-M. et al. (2004). Microarray-based rapid cloning of an ion accumulation deletion mutant in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 101 (43), pp.15404–15409. [Online]. Available at: doi:10.1073/pnas.0404780101.
- Goulding, K., Jarvis, S. and Whitmore, A. (2008). Optimizing nutrient management for farm systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363 (1491), pp.667–680. [Online]. Available at: doi:10.1098/rstb.2007.2177.
- Gourley, C. J. P., Allan, D. L. and Russelle, M. P. (1994). Plant nutrient efficiency: A comparison of definitions and suggested improvement. *Plant and Soil*, 158, pp.29–37. [Online]. Available at: doi:10.1007/BF00007914.
- Griffiths, D. W., Birch, A. N. E. and Hillman, J. R. (1998). Antinutritional compounds in the Brassicaceae: Analysis, biosynthesis, chemistry and dietary effects. *Journal of Horticultural Science and Biotechnology*, 73 (1), pp.1–18. [Online]. Available at: doi:10.1080/14620316.1998.11510937.
- Grispen, V. M. J., Nelissen, H. J. M. and Verkleij, J. A. C. (2006). Phytoextraction with *Brassica napus* L.: A tool for sustainable management of heavy metal contaminated soils. *Environmental Pollution*, 144 (1), pp.77–83. [Online]. Available at:

doi:10.1016/j.envpol.2006.01.007.

Grotz, N. et al. (1998). Identification of a family of zinc transporter genes from Arabidopsis that respond to zinc deficiency. *Proceedings of the National Academy of Sciences*, 95 (12), pp.7220–7224. [Online]. Available at: doi:10.1073/pnas.95.12.7220.

Guan, C. et al. (2017). Spatial Auxin Signaling Controls Leaf Flattening in Arabidopsis. *Current Biology*, 27 (19), p.2940–2950.e4. [Online]. Available at: doi:10.1016/j.cub.2017.08.042.

Guerinot, M. L. and Salt, D. E. (2001). Fortified foods and phytoremediation. Two sides of the same coin. *Plant physiology*, 125 (1), pp.164–167. [Online]. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1539353&tool=pmcentrez&rendertype=abstract>.

Guilbaud, C. S. E. et al. (2015). Is ‘peak N’ key to understanding the timing of flowering in annual plants ? *New Phytologist*, 205, pp.918–927. [Online]. Available at: doi:10.1111/nph.13095.

Guo, B. et al. (2008). Functional analysis of the Arabidopsis PHT4 family of intracellular phosphate transporters. *New Phytologist*, 177 (4), pp.889–898. [Online]. Available at: doi:10.1111/j.1469-8137.2007.02331.x.

Guo, J. and Chen, J. G. (2008). RACK1 genes regulate plant development with unequal genetic redundancy in Arabidopsis. *BMC Plant Biology*, 8, pp.1–11. [Online]. Available at: doi:10.1186/1471-2229-8-108.

Guo, W.-J., Meenam, M. and Goldsbrough, P. B. (2008). Examining the Specific Contributions of Individual Arabidopsis Metallothioneins to Copper Distribution and Metal Tolerance. *Plant Physiology*, 146 (4), pp.1697–1706. [Online]. Available at: doi:10.1104/pp.108.115782.

Guo, W. J., Bundithya, W. and Goldsbrough, P. B. (2003). Characterization of the Arabidopsis metallothionein gene family: Tissue-specific expression and induction during senescence and in response to copper. *New Phytologist*, 159 (2), pp.369–381. [Online]. Available at: doi:10.1046/j.1469-8137.2003.00813.x.

Gupta, U. C. and Gupta, S. C. (1998). Trace element toxicity relationships to crop production and livestock and human health: implications for management. *Communications in Soil Science and Plant Analysis*, 29 (11–14), pp.1491–1522. [Online]. Available at: doi:10.1080/00103629809370045 [Accessed 30 April 2014].

- Guzy-Wrobelska, J. et al. (2013). Vernalization and photoperiod-related changes in the DNA methylation state in winter and spring rapeseed. *Acta Physiologiae Plantarum*, 35 (3), pp.817–827. [Online]. Available at: doi:10.1007/s11738-012-1126-4.
- Habiba, U. et al. (2014). EDTA enhanced plant growth, antioxidant defense system, and phytoextraction of copper by *Brassica napus* L. *Environmental Science and Pollution Research*, 22 (2), pp.1534–1544. [Online]. Available at: doi:10.1007/s11356-014-3431-5.
- Hale, K. L. et al. (2001). Molybdenum sequestration in *Brassica* species. A role for anthocyanins? *Plant physiology*, 126 (4), pp.1391–1402. [Online]. Available at: doi:10.1104/pp.126.4.1391.
- Halkier, B. A. and Gershenzon, J. (2006). Biology and Biochemistry of Glucosinolates. *Annual Review of Plant Biology*, 57, pp.303–333. [Online]. Available at: doi:10.1146/annurev.arplant.57.032905.105228.
- Hall, A. J., Savin, R. and Slafer, G. a. (2014). Is time to flowering in wheat and barley influenced by nitrogen?: A critical appraisal of recent published reports. *European Journal of Agronomy*, 54, pp.40–46. [Online]. Available at: doi:10.1016/j.eja.2013.11.006.
- Hall, J. L. and Williams, L. E. (2003). Transition metal transporters in plants. *Journal of Experimental Botany*, 54 (393), pp.2601–2613. [Online]. Available at: doi:10.1093/jxb/erg303 [Accessed 9 April 2014].
- Hamlin, R. and Barker, A. (2006). Phytoextraction potential of indian mustard at various levels of zinc exposure. *Journal of Plant Nutrition*, 29 (7), pp.1257–1272. [Online]. Available at: doi:10.1080/01904160600767526.
- Hamlin, R. L. and Barker, A. V. (2008). Nutritional alleviation of zinc-induced iron deficiency in Indian mustard and the effects on zinc phytoextraction. *Journal of Plant Nutrition*, 31 (12), pp.2196–2213. [Online]. Available at: doi:10.1080/01904160802463858.
- Han, M.-H. et al. (2004). The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proceedings of the National Academy of Sciences*, 101 (4), pp.1093–1098. [Online]. Available at: doi:10.1073/pnas.0307969100.
- Hanikenne, M. et al. (2008). Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of HMA4. *Nature*, 453 (7193), pp.391–395. [Online]. Available at: doi:10.1038/nature06877.

- Hannon, G. J. (2002). RNA interference. *Nature*, 418 (6894), pp.244–251. [Online]. Available at: doi:10.1038/418244a [doi]\r418244a [pii].
- Hänsch, R. and Mendel, R. R. (2009). Physiological functions of mineral micronutrients (Cu, Zn, Mn, Fe, Ni, Mo, B, Cl). *Current opinion in plant biology*, 12 (3), pp.259–266. [Online]. Available at: doi:10.1016/j.pbi.2009.05.006 [Accessed 6 October 2014].
- Hansen, B. G., Kliebenstein, D. J. and Halkier, B. A. (2007). Identification of a flavin-monooxygenase as the S-oxygenating enzyme in aliphatic glucosinolate biosynthesis in Arabidopsis. *Plant Journal*, 50 (5), pp.902–910. [Online]. Available at: doi:10.1111/j.1365-313X.2007.03101.x.
- Harper, A. L. et al. (2012). Associative transcriptomics of traits in the polyploid crop species Brassica napus. *Nature biotechnology*, 30 (8), pp.798–802. [Online]. Available at: doi:10.1038/nbt.2302 [Accessed 10 July 2014].
- Harper, A. L. et al. (2016). Molecular markers for tolerance of European ash (*Fraxinus excelsior*) to dieback disease identified using Associative Transcriptomics. *Scientific Reports*, 6 (August 2015), pp.1–7. [Online]. Available at: doi:10.1038/srep19335.
- Hartung, F. et al. (2002). An archaeobacterial topoisomerase homolog not present in other eukaryotes is indispensable for cell proliferation of plants. *Current Biology*, 12 (20), pp.1787–1791. [Online]. Available at: doi:10.1016/S0960-9822(02)01218-6.
- Havlickova, L. et al. (2018). Validation of an updated Associative Transcriptomics platform for the polyploid crop species Brassica napus by dissection of the genetic architecture of erucic acid and tocopherol isoform variation in seeds. *The Plant Journal*, pp.181–192. [Online]. Available at: doi:10.1111/tpj.13767.
- Haydon, M. J. et al. (2012). Vacuolar Nicotianamine Has Critical and Distinct Roles under Iron Deficiency and for Zinc Sequestration in Arabidopsis. *the Plant Cell Online*, 24 (2), pp.724–737. [Online]. Available at: doi:10.1105/tpc.111.095042.
- He, Y. (2012). Chromatin regulation of flowering. *Trends in Plant Science*, 17 (9), pp.556–562. [Online]. Available at: doi:10.1016/j.tplants.2012.05.001.
- He, Y., Doyle, M. R. and Amasino, R. M. (2004). PAF1-complex-mediated histone methylation of FLOWERING LOCUS C chromatin is required for the vernalization-responsive, winter-annual habit in Arabidopsis. *Genes and Development*, 18 (22), pp.2774–2784. [Online]. Available at: doi:10.1101/gad.1244504.

- He, Z. et al. (2015). Construction of Brassica A and C genome-based ordered pan-transcriptomes for use in rapeseed genomic research. *Data in brief*, 4, pp.357–362. [Online]. Available at: doi:10.1016/j.dib.2015.06.016.
- He, Z. et al. (2016). Extensive homoeologous genome exchanges in allopolyploid crops revealed by mRNAseq-based visualization. *Plant Biotechnology Journal*, pp.1–11. [Online]. Available at: doi:10.1111/pbi.12657.
- Hensel, L. L. et al. (1994). The fate of inflorescence meristems is controlled by developing fruits in Arabidopsis. *Plant physiology*, 106 (3), pp.863–876. [Online]. Available at: doi:10.1104/pp.106.3.863.
- Higgins, J. et al. (2012). Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species Brassica napus. *BMC Genomics*, 13 (1), p.247. [Online]. Available at: doi:10.1186/1471-2164-13-247.
- Hindu, V. et al. (2018). Identification and validation of genomic regions influencing kernel zinc and iron in maize. *Theoretical and Applied Genetics*, 131 (7), pp.1–15. [Online]. Available at: doi:10.1007/s00122-018-3089-3.
- Hinsinger, P. et al. (2009). Rhizosphere : biophysics , biogeochemistry and ecological relevance. *Plant and Soil*, 321, pp.117–152. [Online]. Available at: doi:10.1007/s11104-008-9885-9.
- Hirsch, J. et al. (2011). A novel fry1 allele reveals the existence of a mutant phenotype unrelated to 5'→3' exoribonuclease (XRN) activities in Arabidopsis thaliana roots. *PLoS ONE*, 6 (2). [Online]. Available at: doi:10.1371/journal.pone.0016724.
- Hirschi, K. D. et al. (2000). Expression of Arabidopsis CAX2 in Tobacco. Altered Metal Accumulation and Increased Manganese Tolerance. *Plant physiology*, 124 (1), pp.125–133. [Online]. Available at: doi:10.1104/pp.124.1.125.
- Hou, X. et al. (2014). Nuclear factor Y-mediated H3K27me3 demethylation of the SOC1 locus orchestrates flowering responses of Arabidopsis. *Nature Communications*, 5, pp.1–14. [Online]. Available at: doi:10.1038/ncomms5601.
- Houtz, R. L., Nable, R. O. and Cheniae, G. M. (1988). Evidence for Effects on the in Vivo Activity of Ribulose-Bisphosphate Carboxylase/Oxygenase during Development of Mn Toxicity in Tobacco. *Plant physiology*, 86, pp.1143–1149. [Online]. Available at: doi:10.1104/pp.86.4.1143.
- Huang, S. et al. (2013). Identification of genome-wide single nucleotide polymorphisms in

- allopolyploid crop *Brassica napus*. *BMC Genomics*, 14 (1). [Online]. Available at: doi:10.1186/1471-2164-14-717.
- Huang, X. et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, 42 (11), pp.961–967. [Online]. Available at: doi:10.1038/ng.695.
- Huang, X. et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature*, 490 (7421), pp.497–501. [Online]. Available at: doi:10.1038/nature11532.
- Hugouvieux, V. et al. (2009). Arabidopsis Putative Selenium-Binding Protein1 Expression Is Tightly Linked to Cellular Sulfur Demand and Can Reduce Sensitivity to Stresses Requiring Glutathione for Tolerance. *Plant Physiology*, 151 (2), pp.768–781. [Online]. Available at: doi:10.1104/pp.109.144808.
- Hussain, D. et al. (2004). P-Type ATPase Heavy Metal Transporters with Roles in Essential Zinc Homeostasis in Arabidopsis. *The Plant Cell*, 16 (5), pp.1327–1339. [Online]. Available at: doi:10.1105/tpc.020487.
- Husted, S. et al. (2011). Review: The role of atomic spectrometry in plant science. *Journal of Analytical Atomic Spectrometry*, 26 (1), pp.52–79. [Online]. Available at: doi:10.1039/c0ja00058b.
- Hwang, J. U. et al. (2008). A Tip-Localized RhoGAP Controls Cell Polarity by Globally Inhibiting Rho GTPase at the Cell Apex. *Current Biology*, 18 (24), pp.1907–1916. [Online]. Available at: doi:10.1016/j.cub.2008.11.057.
- Ibdah, M. et al. (2009). An aldehyde oxidase in developing seeds of Arabidopsis converts benzaldehyde to benzoic Acid. *Plant physiology*, 150 (1), pp.416–423. [Online]. Available at: doi:10.1104/pp.109.135848.
- Inaba, R. and Nishio, T. (2002). Phylogenetic analysis of Brassiceae based on the nucleotide sequences of the S-locus related gene, SLR1. *Theoretical and Applied Genetics*, 105 (8), pp.1159–1165. [Online]. Available at: doi:10.1007/s00122-002-0968-3 [Accessed 16 December 2018].
- Inaba, T. et al. (2005). Estimation of cumulative cadmium intake causing Itai-itai disease. *Toxicology Letters*, 159 (2), pp.192–201. [Online]. Available at: doi:10.1016/j.toxlet.2005.05.011.
- Ivanova, E. M., Kholodova, V. P. and Kuznetsov, V. V. (2010). Biological effects of high copper

and zinc concentrations and their interaction in rapeseed plants. *Russian Journal of Plant Physiology*, 57 (6), pp.806–814. [Online]. Available at: doi:10.1134/S1021443710060099.

Jeifetz, D. et al. (2011). CaBLIND regulates axillary meristem initiation and transition to flowering in pepper. *Planta*, 234 (6), pp.1227–1236. [Online]. Available at: doi:10.1007/s00425-011-1479-8.

Jeong, J. H. et al. (2009). Repression of FLOWERING LOCUS T chromatin by functionally redundant histone H3 lysine 4 demethylases in Arabidopsis. *PLoS ONE*, 4 (11). [Online]. Available at: doi:10.1371/journal.pone.0008033.

Jiang, W. Z. (2006). Mn use efficiency in different wheat cultivars. *Environmental and Experimental Botany*, 57 (1–2), pp.41–50. [Online]. Available at: doi:10.1016/j.envexpbot.2005.04.008 [Accessed 25 November 2014].

Jin, J. B. et al. (2008). The SUMO E3 ligase, AtSIZ1, regulates flowering by controlling a salicylic acid-mediated floral promotion pathway and through affects on FLC chromatin structure. *Plant Journal*, 53 (3), pp.530–540. [Online]. Available at: doi:10.1111/j.1365-313X.2007.03359.x.

Jones, D. L. et al. (2013). REVIEW: Nutrient stripping: the global disparity between food security and soil nutrient stocks. Kardol, P. (Ed). *Journal of Applied Ecology*, 50 (4), pp.851–862. [Online]. Available at: doi:10.1111/1365-2664.12089 [Accessed 17 November 2014].

Jones, M. G. et al. (2004). Biosynthesis of the flavour precursors of onion and garlic. *Journal of Experimental Botany*, 55 (404), pp.1903–1918. [Online]. Available at: doi:10.1093/jxb/erh138.

Jordan, D. B. and Ogren, W. L. (1981). A Sensitive Assay Procedure for Simultaneous Determination of Ribulose-1,5-bisphosphate Carboxylase and Oxygenase Activities. *Plant physiology*, 67 (2), pp.237–245. [Online]. Available at: doi:10.1104/pp.67.2.237.

Jørgensen, M. E., Nour-Eldin, H. H. and Halkier, B. A. (2015). Transport of defense compounds from source to sink: lessons learned from glucosinolates. *Trends in Plant Science*, 20 (8), pp.508–514. [Online]. Available at: doi:10.1016/j.tplants.2015.04.006.

Josefsson, E. (1971). Studies of the Biochemical Background to Differences in Glucosinolate Content in Brassica napus L. II. Administration of Some Sulphur-35 and Carbon-14 Compounds and Localization of Metabolic Blocks. *Physiologia Plantarum*, 24 (2), pp.161–175. [Online]. Available at: doi:10.1111/j.1399-3054.1971.tb03473.x.

Josefsson, E. (1973). Studies of the Biochemical Background to Differences in Glucosinolate

Content in *Brassica ruipus* L. III. Further Studies to Localize Metabolic Blocks. *Plant Physiology*, 29, pp.28–32.

Kaiser, B. N. et al. (2005). The role of molybdenum in agricultural plant production. *Annals of Botany*, 96 (5), pp.745–754. [Online]. Available at: doi:10.1093/aob/mci226 [Accessed 29 September 2014].

Kamada-Nobusada, T. et al. (2008). A putative peroxisomal polyamine oxidase, AtPAO4, is involved in polyamine catabolism in *Arabidopsis thaliana*. *Plant and Cell Physiology*, 49 (9), pp.1272–1282. [Online]. Available at: doi:10.1093/pcp/pcn114.

Kamiya, T. et al. (2015). The MYB36 transcription factor orchestrates Casparian strip formation. *Proceedings of the National Academy of Sciences of the United States of America*, 112 (33), pp.10533–10538. [Online]. Available at: doi:10.1073/pnas.1507691112.

Kandasamy, M. K. et al. (2005). Silencing the nuclear actin-related protein AtARP4 in *Arabidopsis* has multiple effects on plant development, including early flowering and delayed floral senescence. *Plant Journal*, 41 (6), pp.845–858. [Online]. Available at: doi:10.1111/j.1365-313X.2005.02345.x.

Kant, S., Peng, M. and Rothstein, S. J. (2011). Genetic regulation by NLA and microRNA827 for maintaining nitrate-dependent phosphate homeostasis in *Arabidopsis*. *PLoS Genetics*, 7 (3). [Online]. Available at: doi:10.1371/journal.pgen.1002021.

Kataoka, T. et al. (2004). Vacuolar sulfate transporters are essential determinants controlling internal distribution of sulfate in *Arabidopsis*. *The Plant cell*, 16 (10), pp.2693–2704. [Online]. Available at: doi:10.1105/tpc.104.023960.

Kawachi, M. et al. (2009). A Mutant Strain *Arabidopsis thaliana* that Lacks Vacuolar Membrane Zinc Transporter MTP1 Revealed the Latent Tolerance to Excessive Zinc. *Plant and Cell Physiology*, 50 (6), pp.1156–1170. [Online]. Available at: doi:10.1093/pcp/pcp076.

Kawashima, C. G. et al. (2009). Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. *Plant Journal*, 57 (2), pp.313–321. [Online]. Available at: doi:10.1111/j.1365-313X.2008.03690.x.

Kawashima, C. G. et al. (2011). Interplay of SLIM1 and miR395 in the regulation of sulfate assimilation in *Arabidopsis*. *Plant Journal*, 66 (5), pp.863–876. [Online]. Available at: doi:10.1111/j.1365-313X.2011.04547.x.

- Keller, T. (2006). Arabidopsis REGULATOR OF AXILLARY MERISTEMS1 Controls a Leaf Axil Stem Cell Niche and Modulates Vegetative Development. *the Plant Cell Online*, 18 (3), pp.598–611. [Online]. Available at: doi:10.1105/tpc.105.038588.
- Kellermeier, F. et al. (2014). Analysis of the Root System Architecture of Arabidopsis Provides a Quantitative Readout of Crosstalk between Nutritional Signals. *The Plant cell*, 26 (4), pp.1480–1496. [Online]. Available at: doi:10.1105/tpc.113.122101.
- Khoshgoftarmanesh, A. et al. (2010). Micronutrient-efficient genotypes for crop yield and nutritional quality in sustainable agriculture. A review. *Agronomy for Sustainable Development*, 30, pp.83–107. [Online]. Available at: doi:10.1051/agro/2009017 [Accessed 14 October 2014].
- Khurana, N., Singh, M. V. and Chatterjee, C. (2006). Copper stress alters physiology and deteriorates seed quality of rapeseed. *Journal of Plant Nutrition*, 29 (1), pp.93–101. [Online]. Available at: doi:10.1080/01904160500416489.
- Kim, B. H. and Von Arnim, A. G. (2009). FIERY1 regulates light-mediated repression of cell elongation and flowering time via its 3'(2'),5'-bisphosphate nucleotidase activity. *Plant Journal*, 58 (2), pp.208–219. [Online]. Available at: doi:10.1111/j.1365-313X.2008.03770.x.
- Kim, D.-Y. et al. (2006a). AtATM3 is involved in heavy metal resistance in Arabidopsis. *Plant Physiology*, 140 (March), pp.922–932. [Online]. Available at: doi:10.1104/pp.105.074146.922.
- Kim, S. A. et al. (2006b). Localization of Iron in Arabidopsis. *Science*, 314 (November), pp.1295–1299. [Online]. Available at: doi:10.1126/science.1132563.
- Klug, A. (1999). Zinc finger peptides for the regulation of gene expression. *Journal of molecular biology*, 293 (2), pp.215–218. [Online]. Available at: doi:10.1006/jmbi.1999.3007.
- Kluska, K., Adamczyk, J. and Krężel, A. (2018). Metal binding properties, stability and reactivity of zinc fingers. *Coordination Chemistry Reviews*, 367, pp.18–64. [Online]. Available at: doi:10.1016/j.ccr.2018.04.009.
- Kneip, C. et al. (2007). Nitrogen fixation in eukaryotes - New models for symbiosis. *BMC Evolutionary Biology*, 7, pp.1–12. [Online]. Available at: doi:10.1186/1471-2148-7-55.
- Kobayashi, M., Matoh, T. and Azuma, J. (1996). Two Chains of Rhamnogalacturonan II Are Cross-Linked by Borate-Diol Ester Bonds in Higher Plant Cell Walls. *Plant physiology*, 110 (3), pp.1017–1020. [Online]. Available at: doi:10.1104/pp.110.3.1017.
- Kobayashi, Y. et al. (2008). Amino Acid Polymorphisms in Strictly Conserved Domains of a P-

- Type ATPase HMA5 Are Involved in the Mechanism of Copper Tolerance Variation in Arabidopsis. *Plant Physiology*, 148 (2), pp.969–980. [Online]. Available at: doi:10.1104/pp.108.119933.
- Koike, S. et al. (2004). OsYSL2 is a rice metal-nicotianamine transporter that is regulated by iron and expressed in the phloem. *The Plant journal : for cell and molecular biology*, 39 (3), pp.415–424. [Online]. Available at: doi:10.1111/j.1365-313X.2004.02146.x [Accessed 25 November 2014].
- Kolář, J. and Seňková, J. (2008). Reduction of mineral nutrient availability accelerates flowering of Arabidopsis thaliana. *Journal of Plant Physiology*, 165 (15), pp.1601–1609. [Online]. Available at: doi:10.1016/j.jplph.2007.11.010.
- Kopriva, S. (2015). Plant sulfur nutrition: From sachs to big data. *Plant Signaling and Behavior*, 10 (9), pp.1–5. [Online]. Available at: doi:10.1080/15592324.2015.1055436.
- Koprivova, A. et al. (2014). Dissection of the control of anion homeostasis by associative transcriptomics in Brassica napus. *Plant physiology*, 166 (1), pp.442–450. [Online]. Available at: doi:10.1104/pp.114.239947 [Accessed 28 October 2014].
- Koprivova, A. and Kopriva, S. (2014). Molecular mechanisms of regulation of sulfate assimilation: first steps on a long road. *Frontiers in Plant Science*, 5 (October), pp.1–11. [Online]. Available at: doi:10.3389/fpls.2014.00589.
- Koprivova, A. and Kopriva, S. (2016). Hormonal control of sulfate uptake and assimilation. *Plant Molecular Biology*, 91 (6), pp.617–627. [Online]. Available at: doi:10.1007/s11103-016-0438-y.
- Korenkov, V. et al. (2009). Root-selective expression of AtCAX4 and AtCAX2 results in reduced lamina cadmium in field-grown Nicotiana tabacum L. *Plant Biotechnology Journal*, 7 (3), pp.219–226. [Online]. Available at: doi:10.1111/j.1467-7652.2008.00390.x.
- Korshunova, Y. et al. (1999). The IRT1 protein from *Arabidopsis thaliana* is a metal transporter with a broad substrate range. *Plant Molecular Biology*, 40 (1), pp.37–44. [Online]. Available at: doi:10.1023/A:1026438615520.
- Korte, A. and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, 9 (1), p.1. [Online]. Available at: doi:10.1186/1746-4811-9-29.
- Kover, P. X. et al. (2009). A multiparent advanced generation inter-cross to fine-map

quantitative traits in *Arabidopsis thaliana*. *PLoS Genetics*, 5 (7). [Online]. Available at: doi:10.1371/journal.pgen.1000551.

Kozhevnikova, A. D. et al. (2014). Histidine-mediated xylem loading of zinc is a species-wide character in *Noccaea caerulescens*. *New Phytologist*, 203 (2), pp.508–519. [Online]. Available at: doi:10.1111/nph.12816.

Krishnakumar, V. et al. (2015). Araport: The Arabidopsis Information Portal. *Nucleic Acids Research*, 43 (D1), pp.D1003–D1009. [Online]. Available at: doi:10.1093/nar/gku1200.

Krogan, N. J. et al. (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: Linking transcriptional elongation to histone methylation. *Molecular Cell*, 11 (3), pp.721–729. [Online]. Available at: doi:10.1016/S1097-2765(03)00091-1.

Krzyszowska, M. (2011). The cell wall in plant cell response to trace metals: Polysaccharide remodeling and its role in defense strategy. *Acta Physiologiae Plantarum*, 33 (1), pp.35–51. [Online]. Available at: doi:10.1007/s11738-010-0581-z.

Kuper, J. et al. (2004). Structure of the molybdopterin-bound Cnx1G domain links molybdenum and copper metabolism. *Nature*, 430, pp.803–806. [Online]. Available at: doi:10.1038/nature02681 [Accessed 5 November 2014].

Kupper, H. et al. (2000). Cellular compartmentation of cadmium and zinc in relation to other elements in the hyperaccumulator *Arabidopsis halleri*. *Planta*, 212, pp.75–84. [Online]. Available at: doi:10.1007/s004250000366.

Küpper, H. and Andresen, E. (2016). Mechanisms of metal toxicity in plants. *Metallomics*, 8 (3), pp.269–285. [Online]. Available at: doi:10.1039/c5mt00244c.

Lahner, B. et al. (2003). Genomic scale profiling of nutrient and trace elements in *Arabidopsis thaliana*. *Nature biotechnology*, 21 (10), pp.1215–1221. [Online]. Available at: doi:10.1038/nbt865 [Accessed 29 October 2014].

Lal, R. (2009). Soil degradation as a reason for inadequate human nutrition. *Food Security*, 1 (1), pp.45–57. [Online]. Available at: doi:10.1007/s12571-009-0009-z.

Lamesch, P. et al. (2012). The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Research*, 40 (D1), pp.1202–1210. [Online]. Available at: doi:10.1093/nar/gkr1090.

Langfelder, P. and Horvath, S. (2008). WGCNA: an R package for weighted correlation network

analysis. *BMC bioinformatics*, 9, p.559. [Online]. Available at: doi:10.1186/1471-2105-9-559.

Larsson, E. H., Bornman, J. F. and Asp, H. (1998). Influence of UV-B radiation and Cd 2+ on chlorophyll fluorescence, growth and nutrient content in Brassica napus. *Journal of Experimental Botany*, 323 (49), pp.1031–1039. [Online]. Available at: doi:10.1093/jxb/49.323.1031.

Lasat, M. M. (2002). Phytoextraction of toxic metals: a review of biological mechanisms. *Journal of environmental quality*, 31 (1), pp.109–120. [Online]. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11837415>.

Lasswell, J. et al. (2000). Cloning and characterization of IAR1, a gene required for auxin conjugate sensitivity in *Arabidopsis*. *Plant Cell*, 12 (12), pp.2395–2408. [Online]. Available at: doi:10.1105/tpc.12.12.2395.

Lee, H. K. et al. (2017). The relationship between SO₂ exposure and plant physiology: A mini review. *Horticulture, Environment, and Biotechnology*, 58 (6), pp.523–529. [Online]. Available at: doi:10.1007/s13580-017-0053-0.

Lee, J. E. et al. (2014). SEUSS and SEUSS-LIKE 2 coordinate auxin distribution and KNOX1 activity during embryogenesis. *The Plant journal : for cell and molecular biology*, pp.122–135. [Online]. Available at: doi:10.1111/tpj.12625.

Lee, J. and Lee, I. (2010). Regulation and function of SOC1, a flowering pathway integrator. *Journal of Experimental Botany*, 61 (9), pp.2247–2254. [Online]. Available at: doi:10.1093/jxb/erq098.

Lequeux, H. et al. (2010). Response to copper excess in *Arabidopsis thaliana*: Impact on the root system architecture, hormone distribution, lignin accumulation and mineral profile. *Plant Physiology and Biochemistry*, 48 (8), pp.673–682. [Online]. Available at: doi:10.1016/j.plaphy.2010.05.005.

Li, F. et al. (2014). Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). *DNA Research*, 21 (4), pp.355–367. [Online]. Available at: doi:10.1093/dnares/dsu002.

Li, H., Ruan, J. and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping. *Genome research*, 18 (11), pp.1851–1858. [Online]. Available at: doi:10.1101/gr.078212.108.

- Li, J.-Y. et al. (2010). The Arabidopsis nitrate transporter NRT1.8 functions in nitrate removal from the xylem sap and mediates cadmium tolerance. *The Plant cell*, 22 (5), pp.1633–1646. [Online]. Available at: doi:10.1105/tpc.110.075242.
- Li, N. et al. (2018). Genome-wide analysis and expression profiling of the HMA gene family in Brassica napus under cd stress. *Plant and Soil*, 426 (1–2), pp.365–381. [Online]. Available at: doi:10.1007/s11104-018-3637-2.
- Lipka, A. E. et al. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, 28 (18), pp.2397–2399. [Online]. Available at: doi:10.1093/bioinformatics/bts444.
- Liu, S. et al. (2016). A genome-wide association study reveals novel elite allelic variations in seed oil content of Brassica napus. *Theoretical and Applied Genetics*, 129 (6), pp.1203–1215. [Online]. Available at: doi:10.1007/s00122-016-2697-z.
- Liu, T. et al. (2013). Nitrate or NaCl regulates floral induction in Arabidopsis thaliana. *Biologia*, 68 (2). [Online]. Available at: doi:10.2478/s11756-013-0004-x.
- Liu, X. F. et al. (1997). Negative control of heavy metal uptake by the Saccharomyces cerevisiae BSD2 gene. *Journal of Biological Chemistry*, 272 (18), pp.11763–11769. [Online]. Available at: doi:10.1074/jbc.272.18.11763.
- Llamas, A. et al. (2006). The mechanism of nucleotide-assisted molybdenum insertion into molybdopterin: A novel route toward metal cofactor assembly. *Journal of Biological Chemistry*, 281 (27), pp.18343–18350. [Online]. Available at: doi:10.1074/jbc.M601415200.
- Loix, C. et al. (2017). Reciprocal Interactions between Cadmium-Induced Cell Wall Responses and Oxidative Stress in Plants. *Frontiers in Plant Science*, 8 (October), pp.1–19. [Online]. Available at: doi:10.3389/fpls.2017.01867.
- Lu, C. (2000). A Mutation in the Arabidopsis HYL1 Gene Encoding a dsRNA Binding Protein Affects Responses to Abscisic Acid, Auxin, and Cytokinin. *the Plant Cell Online*, 12 (12), pp.2351–2366. [Online]. Available at: doi:10.1105/tpc.12.12.2351.
- Lu, C. et al. (2017). Phosphorylation of SPT5 by CDKD;2 Is Required for VIP5 Recruitment and Normal Flowering in Arabidopsis thaliana. *The Plant Cell*, 29 (2), pp.277–291. [Online]. Available at: doi:10.1105/tpc.16.00568.
- Lu, G. et al. (2014). Associative Transcriptomics Study Dissects the Genetic Architecture of Seed

Glucosinolate Content in *Brassica napus*. *DNA research : an international journal for rapid publication of reports on genes and genomes*, 21 (6), pp.613–625. [Online]. Available at: doi:10.1093/dnares/dsu024.

Lynch, J. P. and St.Clair, S. B. (2004). Mineral stress: The missing link in understanding how global climate change will affect plants in real world soils. *Field Crops Research*, 90 (1), pp.101–115. [Online]. Available at: doi:10.1016/j.fcr.2004.07.008.

Lysak, M. a. et al. (2005). Chromosome triplication found across the tribe Brassiceae. *Genome Research*, 15 (4), pp.516–525. [Online]. Available at: doi:10.1101/gr.3531105.

Lysak, M. A. et al. (2007). Ancestral Chromosomal Blocks Are Triplicated in Brassiceae Species with Varying Chromosome Number and Genome Size. *Plant Physiology*, 145 (2), pp.402–410. [Online]. Available at: doi:10.1104/pp.107.104380.

van Maarschalkerweerd, M. and Husted, S. (2015). Recent developments in fast spectroscopy for plant mineral analysis. *Frontiers in Plant Science*, 6 (March), pp.1–14. [Online]. Available at: doi:10.3389/fpls.2015.00169.

Maathuis, F. J. M. (2009). Physiological functions of mineral macronutrients. *Current opinion in plant biology*, 12 (3), pp.250–258. [Online]. Available at: doi:10.1016/j.pbi.2009.04.003 [Accessed 16 October 2014].

Maillard, A. et al. (2015). Leaf mineral nutrient remobilization during leaf senescence and modulation by nutrient deficiency. *Frontiers in Plant Science*, 6 (May), pp.1–15. [Online]. Available at: doi:10.3389/fpls.2015.00317.

Maillard, A. et al. (2016a). Non-Specific root transport of nutrient gives access to an early nutritional indicator: The case of sulfate and molybdate. *PLoS ONE*, 11 (11), pp.1–20. [Online]. Available at: doi:10.1371/journal.pone.0166910.

Maillard, A. et al. (2016b). Nutrient deficiencies in *Brassica napus* modify the ionic composition of plant tissues: a focus on cross-talk between molybdenum and other nutrients. *Journal of Experimental Botany*, 67 (19), p.erw322. [Online]. Available at: doi:10.1093/jxb/erw322.

Maret, W. (2005). Zinc coordination environments in proteins determine zinc functions. *Journal of trace elements in medicine and biology*, 19, pp.7–12. [Online]. Available at: doi:10.1016/j.jinorgbio.2011.11.018.

- Maret, W. (2012). New perspectives of zinc coordination environments in proteins. *Journal of Inorganic Biochemistry*, 111, pp.110–116. [Online]. Available at: doi:10.1016/j.jinorgbio.2011.11.018.
- Marín, I. C. et al. (2011). Nitrate regulates floral induction in Arabidopsis, acting independently of light, gibberellin and autonomous pathways. *Planta*, 233 (3), pp.539–552. [Online]. Available at: doi:10.1007/s00425-010-1316-5.
- Marmioli, M. et al. (2014). Genome-wide approach in Arabidopsis thaliana to assess the toxicity of cadmium sulfide quantum dots. *Environmental Science and Technology*, 48 (10), pp.5902–5909. [Online]. Available at: doi:10.1021/es404958r.
- Marschner, H. (1995a). Functions of Mineral Nutrients: Macronutrients 8. In: *Mineral Nutrition of Higher Plants*. Second Edi. Academic Press Inc. pp.229–312. [Online]. Available at: doi:10.1016/B978-0-08-057187-4.50014-X.
- Marschner, H. (1995b). Functions of Mineral Nutrients: Micronutrients 9. In: *Mineral Nutrition of Higher Plants (Second Edition)*. Academic Press Inc. pp.313–404. [Online]. Available at: doi:http://dx.doi.org/10.1016/B978-012473542-2/50011-0.
- Marschner, H. (1995c). Introduction, Definition, and Classification of Mineral Nutrients 1. In: *Mineral Nutrition of Higher Plants*. Second Edi. Academic Press Inc. pp.3–5. [Online]. Available at: doi:https://doi.org/10.1016/B978-012473542-2/50003-1.
- Martinez-Ballesta, M. et al. (2015). The impact of the absence of aliphatic glucosinolates on water transport under salt stress in Arabidopsis thaliana. *Frontiers in Plant Science*, 6, pp.1–12. [Online]. Available at: doi:10.1371/journal.pone.0002068.
- Maruyama-Nakashita, A. et al. (2006). Arabidopsis SLIM1 is a central transcriptional regulator of plant sulfur response and metabolism. *The Plant cell*, 18 (11), pp.3235–3251. [Online]. Available at: doi:10.1105/tpc.106.046458.
- Maser, P. (2001). Phylogenetic Relationships within Cation Transporter Families of Arabidopsis. *Plant Physiology*, 126 (4), pp.1646–1667. [Online]. Available at: doi:10.1104/pp.126.4.1646.
- McCallum, C. M. et al. (2000a). Targeted screening for induced mutations. *Nature Biotechnology*, 18 (4), pp.455–457. [Online]. Available at: doi:10.1038/74542.
- McCallum, C. M. et al. (2000b). Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant physiology*, 123 (2), pp.439–442. [Online]. Available at:

doi:10.1104/pp.123.2.439.

Mcvittie, B. et al. (2011). Mapping of the locus associated with tolerance to high manganese in rapeseed (*Brassica napus* L .) Mapping of the locus associated with tolerance to high manganese in rapeseed (*Brassica napus* L .). In: *17th Australian Research Assembly on Brassicas (ARAB)*. (August). 2011. pp.12–16.

Melzer, S. et al. (2008). Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. *Nature Genetics*, 40 (12), pp.1489–1492. [Online]. Available at: doi:10.1038/ng.253.

Mendel, R. R. and Schwarz, G. (2011). Molybdenum cofactor biosynthesis in plants and humans. *Coordination Chemistry Reviews*, 255 (9–10), pp.1145–1158. [Online]. Available at: doi:10.1016/j.ccr.2011.01.054.

Mendoza-Cózatl, D. G. et al. (2008). Identification of high levels of phytochelatin, glutathione and cadmium in the phloem sap of *Brassica napus*. A role for thiol-peptides in the long-distance transport of cadmium and the effect of cadmium on iron translocation. *The Plant Journal*, 54 (2), pp.249–259. [Online]. Available at: doi:10.1111/j.1365-3113X.2008.03410.x.

Meng, J. G. et al. (2017). Genome-wide identification of Cd-responsive NRAMP transporter genes and analyzing expression of NRAMP 1 mediated by miR167 in *Brassica napus*. *BioMetals*, 30 (6), pp.917–931. [Online]. Available at: doi:10.1007/s10534-017-0057-3.

Mhamdi, A. et al. (2010). Cytosolic NADP-dependent isocitrate dehydrogenase contributes to redox homeostasis and the regulation of pathogen responses in *Arabidopsis* leaves. *Plant, Cell and Environment*, 33 (7), pp.1112–1123. [Online]. Available at: doi:10.1111/j.1365-3040.2010.02133.x.

Milazzo, M. F. et al. (2013). *Brassica* biodiesels: Past, present and future. *Renewable and Sustainable Energy Reviews*, 18, pp.350–389. [Online]. Available at: doi:10.1016/j.rser.2012.09.033.

Mills, R. F. et al. (2005). The plant P1B-type ATPase AtHMA4 transports Zn and Cd and plays a role in detoxification of transition metals supplied at elevated levels. *FEBS Letters*, 579 (3), pp.783–791. [Online]. Available at: doi:10.1016/j.febslet.2004.12.040.

Milner, M. et al. (2013). Transport properties of members of the ZIP family in plants and their role in Zn and Mn homeostasis. *Journal of Experimental Botany*, 64 (1), pp.369–381. [Online]. Available at: doi:10.1093/jxb/err313.

Minina, E. A. et al. (2018). Transcriptional stimulation of rate-limiting components of the autophagic pathway improves plant fitness. *Journal of Experimental Botany*, 69 (6), pp.1415–1432. [Online]. Available at: doi:10.1093/jxb/ery010.

Mishra, S., Mishra, A. and Küpper, H. (2017). Protein Biochemistry and Expression Regulation of Cadmium/Zinc Pumping ATPases in the Hyperaccumulator Plants *Arabidopsis halleri* and *Noccaea caerulescens*. *Frontiers in Plant Science*, 8 (May), pp.1–13. [Online]. Available at: doi:10.3389/fpls.2017.00835.

Morel, M. et al. (2009). AtHMA3, a P1B-ATPase allowing Cd/Zn/Co/Pb vacuolar storage in *Arabidopsis*. *Plant physiology*, 149 (2), pp.894–904. [Online]. Available at: doi:10.1104/pp.108.130294.

Moriyama, E. N. et al. (2006). Mining the *Arabidopsis thaliana* genome for highly-divergent seven transmembrane receptors. *Genome biology*, 7 (10), p.R96. [Online]. Available at: doi:10.1186/gb-2006-7-10-r96.

Moroni, J. S., Scott, B. J. and Wratten, N. (2003). Differential tolerance of high manganese among rapeseed genotypes. *Plant and Soil*, 253 (2), pp.507–519. [Online]. Available at: doi:10.1023/A:1024899215845.

van de Mortel, J. E. et al. (2006). Large Expression Differences in Genes for Iron and Zinc Homeostasis, Stress Response, and Lignin Biosynthesis Distinguish Roots of *Arabidopsis thaliana* and the Related Metal Hyperaccumulator *Thlaspi caerulescens*. *Plant Physiology*, 142 (3), pp.1127–1147. [Online]. Available at: doi:10.1104/pp.106.082073.

Mourato, M. P. et al. (2015). Effect of heavy metals in plants of the genus *Brassica*. *International Journal of Molecular Sciences*, 16 (8), pp.17975–17998. [Online]. Available at: doi:10.3390/ijms160817975.

Mugford, S. G. et al. (2011). Control of sulfur partitioning between primary and secondary metabolism. *Plant Journal*, 65 (1), pp.96–105. [Online]. Available at: doi:10.1111/j.1365-313X.2010.04410.x.

Mwamba, T. M. et al. (2016). Interactive effects of cadmium and copper on metal accumulation, oxidative stress, and mineral composition in *Brassica napus*. *International Journal of Environmental Science and Technology*, 13 (9), pp.2163–2174. [Online]. Available at: doi:10.1007/s13762-016-1040-1.

Nagaharu U. (1935). Genome analysis in *Brassica* with special reference to the experimental

formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany*, 7, pp.389–452.

Nakagawa, S. and Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews*, 82 (4), pp.591–605. [Online]. Available at: doi:10.1111/j.1469-185X.2007.00027.x.

Nakamura, M. et al. (2009). Increased thiol biosynthesis of transgenic poplar expressing a wheat O-acetylserine(thiol) lyase enhances resistance to hydrogen sulfide and sulfur dioxide toxicity. *Plant Cell Reports*, 28 (2), pp.313–323. [Online]. Available at: doi:10.1007/s00299-008-0635-5.

NC-IUBMB. (2018). *Enzyme Nomenclature*. [Online]. Available at: <http://www.sbcs.qmul.ac.uk/iubmb/enzyme/> [Accessed 13 July 2018].

Nicol, F. et al. (1998). A plasma membrane-bound putative endo-1, 4- β -d-glucanase is required for normal wall assembly and cell elongation in *Arabidopsis*. *The EMBO journal*, 17 (19), pp.5563–5576.

Noctor, G. et al. (2012). Glutathione in plants: An integrated overview. *Plant, Cell and Environment*, 35 (2), pp.454–484. [Online]. Available at: doi:10.1111/j.1365-3040.2011.02400.x.

Noh, B. et al. (2004). Divergent Roles of a Pair of Homologous Jumonji / Zinc-Finger – Class Transcription Factor Proteins in the Regulation of Arabidopsis Flowering Time. *The Plant cell*, 16 (October), pp.2601–2613. [Online]. Available at: doi:10.1105/tpc.104.025353.2000.

Noh, Y. S. and Amasino, R. M. (1999). Identification of a promoter region responsible for the senescence-specific expression of SAG12. *Plant Molecular Biology*, 41 (2), pp.181–194. [Online]. Available at: doi:10.1023/A:1006342412688.

Nordberg, G. F. (2009). Historical perspectives on cadmium toxicology. *Toxicology and Applied Pharmacology*, 238 (3), pp.192–200. [Online]. Available at: doi:10.1016/j.taap.2009.03.015.

Nouairi, I. et al. (2006). Comparative study of cadmium effects on membrane lipid composition of *Brassica juncea* and *Brassica napus* leaves. *Plant Science*, 170 (3), pp.511–519. [Online]. Available at: doi:10.1016/j.plantsci.2005.10.003.

Nour-Eldin, H. H. et al. (2012). NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature*, 488 (7412), pp.531–534. [Online].

Available at: doi:10.1038/nature11285.

Nour-Eldin, H. H. et al. (2017). Reduction of antinutritional glucosinolates in Brassica oilseeds by mutation of genes encoding transporters. *Nature Biotechnology*, 35 (4), pp.377–382. [Online]. Available at: doi:10.1038/nbt.3823.

O'Halloran, T. V. and Culotta, V. C. (2000). Metallochaperones, an intracellular shuttle service for metal ions. *Journal of Biological Chemistry*, 275 (33), pp.25057–25060. [Online]. Available at: doi:10.1074/jbc.R000006200.

O'Malley, R. C., Barragan, C. C. and Ecker, J. R. (2015). A user's guide to the Arabidopsis T-DNA insertion mutant collections. *Methods in molecular biology (Clifton, N.J.)*, 1284, pp.323–342. [Online]. Available at: doi:10.1007/978-1-4939-2444-8_16.

Oh, S. et al. (2004). A Mechanism Related to the Yeast Transcriptional Regulator Paf1c Is Required for Expression of the Arabidopsis FLC / MAF MADS Box Gene Family. *The Plant Cell*, 16 (November), pp.2940–2953. [Online]. Available at: doi:10.1105/tpc.104.026062.1.

Ohnishi, J. et al. (1990). Involvement of na in active uptake of pyruvate in mesophyll chloroplasts of some c(4) plants : na/pyruvate cotransport. *Plant Physiology*, 94 (3), pp.950–959. [Online]. Available at: doi:10.1104/pp.94.3.950.

Okamoto, M. et al. (2006). High-Affinity Nitrate Transport in Roots of Arabidopsis Depends on Expression of the NAR2-Like Gene AtNRT3.1. *Plant Physiology*, 140 (3), pp.1036–1046. [Online]. Available at: doi:10.1104/pp.105.074385.

Oomen, R. J. F. J. et al. (2009). Functional characterization of NRAMP3 and NRAMP4 from the metal hyperaccumulator *Thlaspi caerulescens*. *New Phytologist*, 181 (3), pp.637–650. [Online]. Available at: doi:10.1111/j.1469-8137.2008.02694.x.

Orsel, M., Krapp, A. and Daniel-Vedele, F. (2002). Analysis of the NRT2 nitrate transporter family in Arabidopsis. Structure and gene expression. *Plant physiology*, 129 (2), pp.886–896. [Online]. Available at: doi:10.1104/pp.005280.

Page, V. and Feller, U. (2015). Heavy Metals in Crop Plants: Transport and Redistribution Processes on the Whole Plant Level. *Agronomy*, 5 (3), pp.447–463. [Online]. Available at: doi:10.3390/agronomy5030447.

Pasricha, N. S. and Randhawa, N. S. (1972). Interaction effect of sulphur and molybdenum on the uptake and utilization of these elements by raya (*Brassica juncea* L.). *Plant and Soil*, 37 (1),

pp.215–220. [Online]. Available at: doi:10.1007/BF01578498.

Pedas, P. et al. (2008). Manganese efficiency in barley: identification and characterization of the metal ion transporter HvIRT1. *Plant physiology*, 148 (1), pp.455–466. [Online]. Available at: doi:10.1104/pp.108.118851 [Accessed 22 April 2014].

Peiter, E. et al. (2007). A secretory pathway-localized cation diffusion facilitator confers plant manganese tolerance. *Proceedings of the National Academy of Sciences of the United States of America*, 104 (20), pp.8532–8537. [Online]. Available at: doi:10.1073/pnas.0609507104.

Peñarrubia, L. et al. (2015). Temporal aspects of copper homeostasis and its crosstalk with hormones. *Frontiers in Plant Science*, 6 (April), pp.1–18. [Online]. Available at: doi:10.3389/fpls.2015.00255.

Peret, B. et al. (2014). Root Architecture Responses: In Search of Phosphate. *Plant Physiology*, 166 (4), pp.1713–1723. [Online]. Available at: doi:10.1104/pp.114.244541.

Pflieger, S., Lefebvre, V. and Causse, M. (2001). The candidate gene approach in plant genetics: A review. *Molecular Breeding*, 7 (4), pp.275–291. [Online]. Available at: doi:10.1023/A:1011605013259.

Phan-Thien, K. Y., Wright, G. C. and Lee, N. A. (2012). Inductively coupled plasma-mass spectrometry (ICP-MS) and -optical emission spectroscopy (ICP-OES) for determination of essential minerals in closed acid digestates of peanuts (*Arachis hypogaea* L.). *Food Chemistry*, 134 (1), pp.453–460. [Online]. Available at: doi:10.1016/j.foodchem.2012.02.095.

Pigliucci, M. and Schlichting, C. D. (1998). Reaction norms of Arabidopsis. V. Flowering time controls phenotypic architecture in response to nutrient stress. *Journal of Evolutionary Biology*, 11 (3), pp.285–301. [Online]. Available at: doi:10.1007/s000360050089.

Pilon-Smits, E. A. et al. (2009). Physiological functions of beneficial elements. *Current Opinion in Plant Biology*, 12 (3), pp.267–274. [Online]. Available at: doi:10.1016/j.pbi.2009.04.009.

Pittman, J. K. et al. (2004). Functional and regulatory analysis of the Arabidopsis thaliana CAX2 cation transporter. *Plant molecular biology*, 56 (6), pp.959–971. [Online]. Available at: doi:10.1007/s11103-004-6446-3.

Pittman, J. K. (2005). Managing the manganese: molecular mechanisms of manganese transport and homeostasis. *The New phytologist*, 167 (3), pp.733–742. [Online]. Available at: doi:10.1111/j.1469-8137.2005.01453.x [Accessed 6 October 2013].

- Plackett, A. R. G. et al. (2017). The early inflorescence of *Arabidopsis thaliana* demonstrates positional effects in floral organ growth and meristem patterning. *Plant Reproduction*, 31 (2), pp.1–21. [Online]. Available at: doi:10.1007/s00497-017-0320-3.
- Podar, D., Ramsey, M. H. and Hutchings, M. J. (2004). Effect of cadmium, zinc and substrate heterogeneity on yield, shoot metal concentration and metal uptake by *Brassica juncea* : Implications for human health risk assesment and phytoremediation. *New Phytologist*, 163 (2), pp.313–324. [Online]. Available at: doi:10.1111/j.1469-8137.2004.1122.x.
- Popescu, A.-A. et al. (2014a). A Novel and Fast Approach for Population Structure Inference Using Kernel-PCA and Optimization. *Genetics*, 198 (December), pp.1421–1431. [Online]. Available at: doi:10.1534/genetics.114.171314.
- Popescu, A. A. et al. (2014b). A Novel and Fast Approach for Population Structure Inference Using Kernel-PCA and Optimization. *Genetics*, 198 (December), pp.1421–1431. [Online]. Available at: doi:10.1534/genetics.114.171314.
- Preston, J. C. and Hileman, L. C. (2013). Functional Evolution in the Plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) Gene Family. *Frontiers in Plant Science*, 4 (April), pp.1–13. [Online]. Available at: doi:10.3389/fpls.2013.00080.
- Price, A. L. et al. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38 (8), pp.904–909. [Online]. Available at: doi:10.1038/ng1847.
- Price, A. L. et al. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, 11 (7), pp.459–463. [Online]. Available at: doi:10.1016/j.dcn.2011.01.002.The.
- Printz, B. et al. (2016). Copper Trafficking in Plants and Its Implication on Cell Wall Dynamics. *Frontiers in Plant Science*, 7 (May), pp.1–16. [Online]. Available at: doi:10.3389/fpls.2016.00601.
- Pritchard, J. K., Stephens, M. and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, pp.945–959. [Online]. Available at: doi:10.1111/j.1471-8286.2007.01758.x.
- Puga, M. I. et al. (2014). SPX1 is a phosphate-dependent inhibitor of PHOSPHATE STARVATION RESPONSE 1 in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 111 (41), pp.14947–14952. [Online]. Available at: doi:10.1073/pnas.1404654111.

- Puig, S. et al. (2007a). Copper and iron homeostasis in Arabidopsis: Responses to metal deficiencies, interactions and biotechnological applications. *Plant, Cell and Environment*, 30 (3), pp.271–290. [Online]. Available at: doi:10.1111/j.1365-3040.2007.01642.x.
- Puig, S. et al. (2007b). Higher plants possess two different types of ATX1-like copper chaperones. *Biochemical and Biophysical Research Communications*, 354 (2), pp.385–390. [Online]. Available at: doi:10.1016/j.bbrc.2006.12.215.
- Purakayastha, T. J. et al. (2008). Phytoextraction of zinc, copper, nickel and lead from a contaminated soil by different species of Brassica. *International Journal of Phytoremediation*, 10 (1), pp.61–72. [Online]. Available at: doi:10.1080/15226510701827077.
- Queralt, I. et al. (2005). Quantitative determination of essential and trace element content of medicinal plants and their infusions by XRF and ICF techniques. *X-Ray Spectrometry*, 34 (3), pp.213–217. [Online]. Available at: doi:10.1002/xrs.795.
- Raman, H. et al. (2017). A Major Locus for Manganese Tolerance Maps on Chromosome A09 in a Doubled Haploid Population of Brassica napus L. *Frontiers in Plant Science*, 8 (December), pp.1–14. [Online]. Available at: doi:10.3389/fpls.2017.01952.
- Rampey, R. A. et al. (2013). Compensatory Mutations in Predicted Metal Transporters Modulate Auxin Conjugate Responsiveness in Arabidopsis. *G3 & Genes/Genomes/Genetics*, 3 (1), pp.131–141. [Online]. Available at: doi:10.1534/g3.112.004655.
- Ren, C., Liu, J. and Gong, Q. (2014). Functions of autophagy in plant carbon and nitrogen metabolism. *Frontiers in Plant Science*, 5 (June), pp.1–5. [Online]. Available at: doi:10.3389/fpls.2014.00301.
- Ricachenevsky, F. K. et al. (2015). Got to hide your Zn away: Molecular control of Zn accumulation and biotechnological applications. *Plant Science*, 236, pp.1–17. [Online]. Available at: doi:10.1016/j.plantsci.2015.03.009.
- Rietra, R. P. J. J. et al. (2017). Effects of Nutrient Antagonism and Synergism on Yield and Fertilizer Use Efficiency. *Communications in Soil Science and Plant Analysis*, 48 (16), pp.1895–1920. [Online]. Available at: doi:10.1080/00103624.2017.1407429.
- Rizwan, M. et al. (2018). Cadmium phytoremediation potential of Brassica crop species: A review. *Science of the Total Environment*, 631–632, pp.1175–1191. [Online]. Available at: doi:10.1016/j.scitotenv.2018.03.104.

- Robinson, B. et al. (2003). Phytoextraction: an assessment of biogeochemical and economic viability. *Plant and Soil*, 249 (1), pp.117–125. [Online]. Available at: doi:10.1023/A:1022586524971 [Accessed 12 November 2013].
- Rogers, K. and Chen, X. (2013). Biogenesis, Turnover, and Mode of Action of Plant MicroRNAs. *The Plant Cell*, 25 (7), pp.2383–2399. [Online]. Available at: doi:10.1105/tpc.113.113159.
- Roppolo, D. et al. (2014). Functional and evolutionary analysis of the CASPARIAN STRIP MEMBRANE DOMAIN PROTEIN family. *Plant physiology*, 165 (August), pp.1709–1722. [Online]. Available at: doi:10.1104/pp.114.239137.
- Rouached, H. et al. (2008). Differential Regulation of the Expression of Two High-Affinity Sulfate Transporters, SULTR1.1 and SULTR1.2, in Arabidopsis. *Plant Physiology*, 147 (2), pp.897–911. [Online]. Available at: doi:10.1104/pp.108.118612.
- Ruiz, J. M. and Blumwald, E. (2002). Salinity-induced glutathione synthesis in Brassica napus. *Planta*, 214 (6), pp.965–969. [Online]. Available at: doi:10.1007/s00425-002-0748-y.
- Rus, A. et al. (2006). Natural variants of AtHKT1 enhance Na⁺ accumulation in two wild populations of Arabidopsis. *PLoS Genetics*, 2 (12), pp.1964–1973. [Online]. Available at: doi:10.1371/journal.pgen.0020210.
- Sacharowski, S. P. et al. (2015). SWP73 Subunits of Arabidopsis SWI/SNF Chromatin Remodeling Complexes Play Distinct Roles in Leaf and Flower Development. *The Plant Cell*, 27 (7), pp.1889–1906. [Online]. Available at: doi:10.1105/tpc.15.00233.
- Salt, D. E. et al. (1995). Mechanisms of Cadmium Mobility and Accumulation in Indian Mustard. *Plant Physiology*, 109 (4), pp.1427–1433. [Online]. Available at: doi:10.1104/pp.109.4.1427.
- Salt, D. E., Baxter, I. and Lahner, B. (2008). Ionomics and the study of the plant ionome. *Annual review of plant biology*, 59, pp.709–733. [Online]. Available at: doi:10.1146/annurev.arplant.59.032607.092942 [Accessed 13 December 2014].
- Sancenon, V. et al. (2003). Identification of a copper transporter family in Arabidopsis thaliana. *Plant Molecular Biology*, 51 (4), pp.577–587. [Online]. Available at: doi:10.1023/A:1022345507112.
- Sarry, J. E. et al. (2006). The early responses of Arabidopsis thaliana cells to cadmium exposure explored by protein and metabolite profiling analyses. *Proteomics*, 6 (7), pp.2180–2198. [Online]. Available at: doi:10.1002/pmic.200500543.

- Sasaki, A. et al. (2012). Nramp5 Is a Major Transporter Responsible for Manganese and Cadmium Uptake in Rice. *The Plant Cell*, 24 (5), pp.2155–2167. [Online]. Available at: doi:10.1105/tpc.112.096925.
- Schiavon, M. et al. (2012). Selenate and molybdate alter sulfate transport and assimilation in Brassica juncea L. Czern.: Implications for phytoremediation. *Environmental and Experimental Botany*, 75, pp.41–51. [Online]. Available at: doi:10.1016/j.envexpbot.2011.08.016.
- Schneider, R. et al. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nature Cell Biology*, 6 (1), pp.73–77. [Online]. Available at: doi:10.1038/ncb1076.
- Schneider, T. et al. (2013). A proteomics approach to investigate the process of Zn hyperaccumulation in *Noccaea caerulescens* (J and C. Presl) F.K. Meyer. *Plant Journal*, 73 (1), pp.131–142. [Online]. Available at: doi:10.1111/tpj.12022.
- Scholl, R. L., May, S. T. and Ware, D. H. (2000). Seed and molecular resources for Arabidopsis. *Plant physiology*, 124 (4), pp.1477–1480. [Online]. Available at: doi:10.1104/pp.124.4.1477.
- Schützendübel, A. and Polle, A. (2002). Plant responses to abiotic stresses: Heavy metal-induced oxidative stress and protection by mycorrhization. *Journal of Experimental Botany*, 53 (372), pp.1351–1365. [Online]. Available at: doi:10.1093/jxb/53.372.1351.
- Schwartzman, M. S. et al. (2018). Adaptation to high zinc depends on distinct mechanisms in metallicolous populations of *Arabidopsis halleri*. *New Phytologist*, 218 (1), pp.269–282. [Online]. Available at: doi:10.1111/nph.14949.
- Schwarz, G. and Mendel, R. R. (2006). Molybdenum cofactor biosynthesis and molybdenum enzymes. *Annual review of plant biology*, 57, pp.623–647. [Online]. Available at: doi:10.1146/annurev.arplant.57.032905.105437 [Accessed 29 September 2014].
- Selvam, A. and Wong, J. W. C. (2008). Phytochelatin synthesis and cadmium uptake of *Brassica napus*. *Environmental Technology*, 29 (7), pp.765–773. [Online]. Available at: doi:10.1080/09593330801987079.
- Sequera-Mutiozabal, M. I. et al. (2016). Global Metabolic Profiling of Arabidopsis Polyamine Oxidase 4 (AtPAO4) Loss-of-Function Mutants Exhibiting Delayed Dark-Induced Senescence. *Frontiers in Plant Science*, 7 (February), pp.1–13. [Online]. Available at: doi:10.3389/fpls.2016.00173.
- Shahzad, Z. et al. (2010). The five AHMTP1 zinc transporters undergo different evolutionary

fates towards adaptive evolution to zinc tolerance in *Arabidopsis halleri*. *PLoS Genetics*, 6 (4). [Online]. Available at: doi:10.1371/journal.pgen.1000911.

Shimajima, M. (2011). Biosynthesis and functions of the plant sulfolipid. *Progress in Lipid Research*, 50 (3), pp.234–239. [Online]. Available at: doi:10.1016/j.plipres.2011.02.003.

Shin, L.-J. et al. (2013). IRT1 DEGRADATION FACTOR1, a RING E3 Ubiquitin Ligase, Regulates the Degradation of IRON-REGULATED TRANSPORTER1 in *Arabidopsis*. *The Plant Cell*, 25 (8), pp.3039–3051. [Online]. Available at: doi:10.1105/tpc.113.115212.

Shin, L.-J., Lo, J.-C. and Yeh, K.-C. (2012). Copper Chaperone Antioxidant Protein1 Is Essential for Copper Homeostasis. *Plant Physiology*, 159 (3), pp.1099–1110. [Online]. Available at: doi:10.1104/pp.112.195974.

Shrestha, A. et al. (2018). Genome-wide association study to identify candidate loci and genes for Mn toxicity tolerance in rice. *PLoS ONE*, 13 (2), pp.1–15. [Online]. Available at: doi:10.1371/journal.pone.0192116.

Siegfried, K. R. et al. (1999). Members of the YABBY gene family specify abaxial cell fate in *Arabidopsis*. *Development (Cambridge, England)*, 126 (18), pp.4117–4128. [Online]. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10457020>.

SIGNAL. *Salk Institute Genomic Analysis Laboratory: T-DNA Primer Design Tool*. [Online]. Available at: <http://signal.salk.edu/tdnaprimers.2.html> [Accessed 7 February 2016].

Socha, A. L. and Guerinot, M. Lou. (2014). Mn-euvering manganese: the role of transporter gene family members in manganese uptake and mobilization in plants. *Frontiers in plant science*, 5, pp.1–16. [Online]. Available at: doi:10.3389/fpls.2014.00106 [Accessed 24 October 2014].

Sollars, E. S. A. et al. (2017). Genome sequence and genetic diversity of European ash trees. *Nature*, 541 (7636), pp.212–216. [Online]. Available at: doi:10.1038/nature20786.

Song, W.-Y. (2004). A Novel Family of Cys-Rich Membrane Proteins Mediates Cadmium Resistance in *Arabidopsis*. *Plant Physiology*, 135 (2), pp.1027–1039. [Online]. Available at: doi:10.1104/pp.103.037739.

Song, W.-Y. et al. (2010). *Arabidopsis* PCR2 Is a Zinc Exporter Involved in Both Zinc Extrusion and Long-Distance Zinc Transport. *the Plant Cell Online*, 22 (7), pp.2237–2252. [Online]. Available at: doi:10.1105/tpc.109.070185.

Soriano, M. A. and Fereres, E. (2003). Use of crops for in situ phytoremediation of polluted soils following a toxic flood from a mine spill. *Plant and Soil*, 256 (2), pp.253–264. [Online]. Available at: doi:10.1023/A:1026155423727.

South, P. F. et al. (2017). Bile Acid Sodium Symporter BASS6 Can Transport Glycolate and Is Involved in Photorespiratory Metabolism in *Arabidopsis thaliana*. *The Plant Cell*, 29 (4), pp.808–823. [Online]. Available at: doi:10.1105/tpc.16.00775.

Stockle, D. et al. (2016). Putative RopGAPs impact division plane selection and interact with kinesin-12 POK1. *Nature Plants*, 2 (September), pp.1–6. [Online]. Available at: doi:10.1038/nplants.2016.120.

Sugimoto-Shirasu, K. et al. (2002). DNA topoisomerase VI is essential for endoreduplication in *Arabidopsis*. *Current Biology*, 12 (20), pp.1782–1786. [Online]. Available at: doi:10.1016/S0960-9822(02)01198-3.

Sullivan, G. M. and Feinn, R. (2012). Using Effect Size—or Why the *P* Value Is Not Enough. *Journal of Graduate Medical Education*, 4 (3), pp.279–282. [Online]. Available at: doi:10.4300/JGME-D-12-00156.1.

TAIR. (2015). *The Arabidopsis Information Resource: GBrowse*. [Online]. Available at: <http://tairm17.tacc.utexas.edu/cgi-bin/gb2/gbrowse/arabidopsis/>.

Takada, S. and Goto, K. (2003). Terminal flower2, an Arabidopsis homolog of heterochromatin protein1, counteracts the activation of flowering locus T by CONSTANS in the vascular tissues of leaves to regulate flowering time. *The Plant Cell*, 15 (12), pp.2856–2865. [Online]. Available at: doi:10.1105/tpc.016345.

Takahashi, H. et al. (2000). The roles of three functional sulphate transporters involved in uptake and translocation of sulphate in *Arabidopsis thaliana*. *Plant Journal*, 23 (2), pp.171–182. [Online]. Available at: doi:10.1046/j.1365-3113X.2000.00768.x.

Takahashi, H. et al. (2011). Sulfur Assimilation in Photosynthetic Organisms: Molecular Functions and Regulations of Transporters and Assimilatory Enzymes. *Annual Review of Plant Biology*, 62 (1), pp.157–184. [Online]. Available at: doi:10.1146/annurev-arplant-042110-103921.

Takahashi, H. et al. (2012). Evolutionary Relationships and Functional Diversity of Plant Sulfate Transporters. *Frontiers in Plant Science*, 2 (January), pp.1–9. [Online]. Available at: doi:10.3389/fpls.2011.00119.

- Takeo, K. (2016). Stress-induced flowering: The third category of flowering response. *Journal of Experimental Botany*, 67 (17), pp.4925–4934. [Online]. Available at: doi:10.1093/jxb/erw272.
- Tapken, W. et al. (2015). The Clp protease system is required for copper ion-dependent turnover of the PAA2/HMA8 copper transporter in chloroplasts. *New Phytologist*, 205 (2), pp.511–517. [Online]. Available at: doi:10.1111/nph.13093.
- Taylor, K. M. and Nicholson, R. I. (2003). The LZT proteins; The LIV-1 subfamily of zinc transporters. *Biochimica et Biophysica Acta - Biomembranes*, 1611 (1–2), pp.16–30. [Online]. Available at: doi:10.1016/S0005-2736(03)00048-8.
- Tejada-Jiménez, M. (2007). A high-affinity molybdate transporter in eukaryotes. *Proceedings of the National Academy of Sciences*, 104, pp.20126–20130. [Online]. Available at: <http://www.pnas.org/content/104/50/20126.short> [Accessed 6 November 2014].
- ThermoFisher. *ThermoFisher Scientific Oligoperfect (TM) designer*. [Online]. Available at: <https://tools.thermofisher.com/content.cfm?pageid=9716> [Accessed 7 October 2017].
- Thomas, C. L. et al. (2016). Root morphology and seed and leaf ionomic traits in a Brassica napus L. diversity panel show wide phenotypic variation and are characteristic of crop habit. *BMC Plant Biology*, 16 (1), p.214. [Online]. Available at: doi:10.1186/s12870-016-0902-5.
- Thomine, S. et al. (2000). Cadmium and iron transport by members of a plant metal transporter family in Arabidopsis with homology to Nramp genes. *Proceedings of the National Academy of Sciences*, 97 (9), pp.4991–4996. [Online]. Available at: doi:10.1073/pnas.97.9.4991.
- Thomine, S. et al. (2003). AtNRAMP3, a multispecific vacuolar metal transporter involved in plant responses to iron deficiency. *Plant Journal*, 34 (5), pp.685–695. [Online]. Available at: doi:10.1046/j.1365-313X.2003.01760.x.
- Tian, H. et al. (2010). Arabidopsis NPCC6/NaKR1 Is a Phloem Mobile Metal Binding Protein Necessary for Phloem Function and Root Meristem Maintenance. *The Plant Cell*, 22 (12), pp.3963–3979. [Online]. Available at: doi:10.1105/tpc.110.080010.
- Tienderen, P. H. Van et al. (1996). Pleiotropic effects of flowering time genes in the annual crucifer Arabidopsis thaliana (Brassicaceae). *American Journal of Botany*, 83 (2), pp.169–174. [Online]. Available at: doi:10.2307/2445934.

- Tomatsu, H. et al. (2007). An *Arabidopsis thaliana* high-affinity molybdate transporter required for efficient uptake of molybdate from soil. *Proceedings of the National Academy of Sciences of the United States of America*, 104 (47), pp.18807–18812. [Online]. Available at: doi:10.1073/pnas.0706373104.
- Tommey, A. M. and Evans, E. J. (1991). Temperature and daylength control of flower initiation in winter oilseed rape (*Brassica napus* L.). *Annals of Applied Biology*, 118 (1), pp.201–208. [Online]. Available at: doi:10.1111/j.1744-7348.1991.tb06098.x.
- Toroser, D., Griffiths, H. and Thomas, D. R. (1995). Glucosinolate biosynthesis in oilseed rape (*Brassica napus* L.): studies with ³⁵SO₄²⁻ and glucosinolate precursors using oilseed rape pods and seeds. *Journal of Experimental Botany*, 46(288) (288), pp.787–794.
- Trick, M. et al. (2009a). A newly-developed community microarray resource for transcriptome profiling in Brassica species enables the confirmation of Brassica-specific expressed sequences. *BMC plant biology*, 9, p.50. [Online]. Available at: doi:10.1186/1471-2229-9-50 [Accessed 10 October 2014].
- Trick, M. et al. (2009b). Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. *Plant biotechnology journal*, 7 (4), pp.334–346. [Online]. Available at: doi:10.1111/j.1467-7652.2008.00396.x [Accessed 8 October 2014].
- UN-DESA. (2017). World Population Prospects The 2017 Revision Key Findings and Advance Tables. *World Population Prospects The 2017*, pp.1–46. [Online]. Available at: doi:10.1017/CBO9781107415324.004.
- Urano, D. et al. (2015). *Arabidopsis* Receptor of Activated C Kinase1 Phosphorylation by WITH NO LYSINE8 KINASE. *Plant Physiology*, 167 (2), pp.507–516. [Online]. Available at: doi:10.1104/pp.114.247460.
- USDA. (2018). Oilseeds: World Markets and Trade. *Office of Global Analysis*. [Online]. Available at: doi:10.2337/dc11-S062.
- Verbruggen, N., Hermans, C. and Schat, H. (2009a). Mechanisms to cope with arsenic or cadmium excess in plants. *Current opinion in plant biology*, 12 (3), pp.364–372. [Online]. Available at: doi:10.1016/j.pbi.2009.05.001 [Accessed 29 October 2014].
- Verbruggen, N., Hermans, C. and Schat, H. (2009b). Molecular mechanisms of metal hyperaccumulation in plants. *The New phytologist*, 181 (4), pp.759–776. [Online]. Available at:

doi:10.1111/j.1469-8137.2008.02748.x.

Vert, G. and Grotz, N. (2002). IRT1, an Arabidopsis transporter essential for iron uptake from the soil and for plant growth. *The Plant Cell*, 14, pp.1223–1233. [Online]. Available at: doi:10.1105/tpc.001388.Nongrasses [Accessed 25 November 2014].

Vidal, E. A. et al. (2014). Nitrogen control of developmental phase transitions in Arabidopsis thaliana. *Journal of experimental botany*, 65 (19), pp.5611–5618. [Online]. Available at: doi:10.1093/jxb/eru326.

Voinnet, O. (2009). Origin, Biogenesis, and Activity of Plant MicroRNAs. *Cell*, 136 (4), pp.669–687. [Online]. Available at: doi:10.1016/j.cell.2009.01.046.

Vreugdenhil, D. et al. (2004). Natural variation and QTL analysis for cationic mineral content in seeds of *Arabidopsis thaliana*. *Plant, Cell & Environment*, 27 (7), pp.828–839. [Online]. Available at: doi:10.1111/j.1365-3040.2004.01189.x.

Wahl, V. et al. (2013). Regulation of Flowering by Trehalose-6-Phosphate Signaling in Arabidopsis thaliana. *Science*, 339 (February), pp.704–708.

Wang, J. W. (2014). Regulation of flowering time by the miR156-mediated age pathway. *Journal of Experimental Botany*, 65 (17), pp.4723–4730. [Online]. Available at: doi:10.1093/jxb/eru246.

Wang, J. W., Czech, B. and Weigel, D. (2009). miR156-Regulated SPL Transcription Factors Define an Endogenous Flowering Pathway in Arabidopsis thaliana. *Cell*, 138 (4), pp.738–749. [Online]. Available at: doi:10.1016/j.cell.2009.06.014.

Wang, N. et al. (2008). A functional genomics resource for Brassica napus: Development of an EMS mutagenized population and discovery of FAE1 point mutations by TILLING. *New Phytologist*, 180 (4), pp.751–765. [Online]. Available at: doi:10.1111/j.1469-8137.2008.02619.x.

Wang, Y. et al. (2003). The Arabidopsis homeobox gene, ATHB16, regulates leaf development and the sensitivity to photoperiod in Arabidopsis. *Developmental Biology*, 264 (1), pp.228–239. [Online]. Available at: doi:10.1016/j.ydbio.2003.07.017.

Wang, Z. et al. (2018). SWI2/SNF2 ATPase CHR2 remodels pri-miRNAs via Serrate to impede miRNA production. *Nature*, 557. [Online]. Available at: doi:10.1038/s41586-018-0135-x.

Waters, B. M. et al. (2006). Mutations in Arabidopsis yellow stripe-like1 and yellow stripe-like3

reveal their roles in metal ion homeostasis and loading of metal ions in seeds. *Plant Physiology*, 141 (August), pp.1446–1458. [Online]. Available at: doi:10.1104/pp.106.082586.

Weber, M., Trampczynska, A. and Clemens, S. (2006). Comparative transcriptome analysis of toxic metal responses in *Arabidopsis thaliana* and the Cd²⁺-hypertolerant facultative metallophyte *Arabidopsis halleri*. *Plant, Cell and Environment*, 29 (5), pp.950–963. [Online]. Available at: doi:10.1111/j.1365-3040.2005.01479.x.

White, P. et al. (2018). Limits to the Biofortification of Leafy Brassicas with Zinc. *Agriculture*, 8 (3), p.32. [Online]. Available at: doi:10.3390/agriculture8030032.

White, P. J. et al. (2013). Root traits for infertile soils. *Frontiers in Plant Science*, 4 (June), pp.1–7. [Online]. Available at: doi:10.3389/fpls.2013.00193.

White, P. J. and Broadley, M. R. (2009). Biofortification of crops with seven mineral elements often lacking in. *New Phytologist*, 182 (1), pp.49–84. [Online]. Available at: doi:10.1111/j.1469-8137.2008.02738.x.

Wildner, G. F. and Henkel, J. (1979). The effect of divalent metal ions on the activity of Mg⁺⁺ depleted ribulose-1,5-bisphosphate oxygenase. *Planta*, 146 (2), pp.223–228. [Online]. Available at: doi:10.1007/BF00388236.

Wintz, H. et al. (2003). Expression Profiles of *Arabidopsis thaliana* in Mineral Deficiencies Reveal Novel Transporters Involved in Metal Homeostasis. *Journal of Biological Chemistry*, 278 (48), pp.47644–47653. [Online]. Available at: doi:10.1074/jbc.M309338200.

Wirtz, M. and Hell, R. (2006). Functional analysis of the cysteine synthase protein complex from plants: Structural, biochemical and regulatory properties. *Journal of Plant Physiology*, 163 (3), pp.273–286. [Online]. Available at: doi:10.1016/j.jplph.2005.11.013.

Wong, C. K. E. et al. (2009). Functional analysis of the heavy metal binding domains of the Zn/Cd-transporting ATPase, HMA2, in *Arabidopsis thaliana*. *New Phytologist*, 181 (1), pp.79–88. [Online]. Available at: doi:10.1111/j.1469-8137.2008.02637.x.

Wood, I. P. et al. (2017). Carbohydrate microarrays and their use for the identification of molecular markers for plant cell wall composition. *Proceedings of the National Academy of Sciences*, p.201619033. [Online]. Available at: doi:10.1073/pnas.1619033114.

Wu, G. et al. (2009). The Sequential Action of miR156 and miR172 Regulates Developmental Timing in *Arabidopsis*. *Cell*, 138 (4), pp.750–759. [Online]. Available at:

doi:10.1016/j.cell.2009.06.031.

Wu, Z. et al. (2002). An endoplasmic reticulum-bound Ca²⁺/Mn²⁺ pump, ECA1, supports plant growth and confers tolerance to Mn²⁺ stress. *Plant Physiology*, 130, pp.128–137. [Online]. Available at: doi:10.1104/pp.004440.the [Accessed 1 May 2014].

Xu, L. et al. (2015). Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (*Brassica napus* L.). *DNA Research*, 23 (1), pp.43–52. [Online]. Available at: doi:10.1093/dnares/dsv035.

Yamaguchi, C. et al. (2016). Effects of cadmium treatment on the uptake and translocation of sulfate in *Arabidopsis thaliana*. *Plant and Cell Physiology*, 57 (11), pp.2353–2366. [Online]. Available at: doi:10.1093/pcp/pcw156.

Yamasaki, H. et al. (2007). Regulation of copper homeostasis by micro-RNA in *Arabidopsis*. *Journal of Biological Chemistry*, 282 (22), pp.16369–16378. [Online]. Available at: doi:10.1074/jbc.M700138200.

Yamasaki, H. et al. (2009). SQUAMOSA Promoter Binding Protein-Like7 Is a Central Regulator for Copper Homeostasis in *Arabidopsis*. *the Plant Cell Online*, 21 (1), pp.347–361. [Online]. Available at: doi:10.1105/tpc.108.060137.

Yang, H. et al. (2010). A mutant CHS3 protein with TIR-NB-LRR-LIM domains modulates growth, cell death and freezing tolerance in a temperature-dependent manner in *Arabidopsis*. *Plant Journal*, 63 (2), pp.283–296. [Online]. Available at: doi:10.1111/j.1365-3113X.2010.04241.x.

Yang, H., Howard, M. and Dean, C. (2016). Physical coupling of activation and derepression activities to maintain an active transcriptional state at *FLC*. *Proceedings of the National Academy of Sciences*, 113 (33), pp.9369–9374. [Online]. Available at: doi:10.1073/pnas.1605733113.

Yang, Y. et al. (2018). Precise editing of *CLAVATA* genes in *Brassica napus* L. regulates multilocular silique development. *Plant Biotechnology Journal*, 16 (7), pp.1322–1335. [Online]. Available at: doi:10.1111/pbi.12872.

Yang, Y. W. et al. (1999). Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. *Journal of Molecular Evolution*, 48 (5), pp.597–604. [Online]. Available at: doi:10.1007/PL00006502.

Yatusevich, R. et al. (2010). Genes of primary sulfate assimilation are part of the glucosinolate

biosynthetic network in *Arabidopsis thaliana*. *Plant Journal*, 62 (1), pp.1–11. [Online]. Available at: doi:10.1111/j.1365-313X.2009.04118.x.

Yin, Y. et al. (2002). A crucial role for the putative *Arabidopsis* topoisomerase VI in plant growth and development. *Proceedings of the National Academy of Sciences of the United States of America*, 99 (15), pp.10191–10196. [Online]. Available at: doi:10.1073/pnas.152337599.

Yong, Z., Kotur, Z. and Glass, A. D. M. (2010). Characterization of an intact two-component high-affinity nitrate transporter from *Arabidopsis* roots. *Plant Journal*, 63 (5), pp.739–748. [Online]. Available at: doi:10.1111/j.1365-313X.2010.04278.x.

Yoshimoto, N. et al. (2007). Posttranscriptional Regulation of High-Affinity Sulfate Transporters in *Arabidopsis* by Sulfur Nutrition. *Plant Physiology*, 145 (2), pp.378–388. [Online]. Available at: doi:10.1104/pp.107.105742.

Yoshimoto, N. et al. (2018). *Mediates Re-Distribution of Sulfur from Source to Sink Organs in Arabidopsis 1*. D (April 2003), pp.1511–1517. [Online]. Available at: doi:10.1104/pp.014712.plasmamembranes.

Yu, H. et al. (2004). Floral homeotic genes are targets of gibberellin signaling in flower development. *Proceedings of the National Academy of Sciences*, 101 (20), pp.7827–7832. [Online]. Available at: doi:10.1073/pnas.0402377101.

Yu, L. H. et al. (2016a). Overexpression of *Arabidopsis* NLP7 improves plant growth under both nitrogen-limiting and-sufficient conditions by enhancing nitrogen and carbon assimilation. *Scientific Reports*, 6 (June), pp.1–13. [Online]. Available at: doi:10.1038/srep27795.

Yu, N. et al. (2015). The role of miR156/SPLs modules in *Arabidopsis* lateral root development. *Plant Journal*, 83 (4), pp.673–685. [Online]. Available at: doi:10.1111/tpj.12919.

Yu, R. et al. (2012). Accumulation and translocation of heavy metals in the canola (*Brassica napus* L.)-soil system in Yangtze River Delta, China. *Plant and Soil*, 353 (1–2), pp.33–45. [Online]. Available at: doi:10.1007/s11104-011-1006-5.

Yu, X. et al. (2008). Modulation of brassinosteroid-regulated gene expression by jumonji domain-containing proteins ELF6 and REF6 in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 105 (21), pp.7618–7623. [Online]. Available at: doi:10.1073/pnas.0802254105.

- Yu, Y. T. et al. (2016b). Overexpression of the MYB37 transcription factor enhances abscisic acid sensitivity, and improves both drought tolerance and seed productivity in *Arabidopsis thaliana*. *Plant Molecular Biology*, 90 (3), pp.267–279. [Online]. Available at: doi:10.1007/s11103-015-0411-1.
- Yuan, M. et al. (2011). Molecular and functional analyses of COPT/Ctr-type copper transporter-like gene family in rice. *BMC Plant Biology*, 11 (1), p.69. [Online]. Available at: doi:10.1186/1471-2229-11-69.
- Zaheer, I. E. et al. (2015). Citric acid assisted phytoremediation of copper by *Brassica napus* L. *Ecotoxicology and Environmental Safety*, 120, pp.310–317. [Online]. Available at: doi:10.1016/j.ecoenv.2015.06.020.
- Zhan, E. et al. (2018). OTS1-dependent deSUMOylation increases tolerance to high copper levels in *Arabidopsis*. *Journal of Integrative Plant Biology*, 60 (4), pp.310–322. [Online]. Available at: doi:10.1111/jipb.12618.
- Zhang, I. and Lechowicz, M. (1994). Correlation Between Time of Flowering and Phenotypic Plasticity in *Arabidopsis thaliana*. *American Journal of Botany*, 81 (10), pp.1336–1342.
- Zhang, L. W. et al. (2013). MiR395 is involved in detoxification of cadmium in *Brassica napus*. *Journal of Hazardous Materials*, 250–251, pp.204–211. [Online]. Available at: doi:10.1016/j.jhazmat.2013.01.053.
- Zhang, X. D. et al. (2018). Identification of Cd-responsive RNA helicase genes and expression of a putative BnRH 24 mediated by miR158 in canola (*Brassica napus*). *Ecotoxicology and Environmental Safety*, 157 (December 2017), pp.159–168. [Online]. Available at: doi:10.1016/j.ecoenv.2018.03.081.
- Zhang, X. D., Zhao, K. X. and Yang, Z. M. (2018). Identification of genomic ATP binding cassette (ABC) transporter genes and Cd-responsive ABCs in *Brassica napus*. *Gene*, 664 (April), pp.139–151. [Online]. Available at: doi:10.1016/j.gene.2018.04.060.
- Zhang, Y. et al. (1992). A major inducer of anticarcinogenic protective enzymes from broccoli: Isolation and elucidation of structure (chemoprotection/enzyme induction/isothiocyanates/sulforaphane/quinone reductase). *Medical Sciences*, 89 (March), pp.2399–2403. [Online]. Available at: doi:10.1073/pnas.89.6.2399.
- Zhang, Y. et al. (2012). Vacuolar membrane transporters OsVIT1 and OsVIT2 modulate iron translocation between flag leaves and seeds in rice. *Plant Journal*, 72 (3), pp.400–410. [Online].

Available at: doi:10.1111/j.1365-313X.2012.05088.x.

Zhang, Z. et al. (2010a). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42 (4), pp.355–360. [Online]. Available at: doi:10.1038/ng.546.

Zhang, Z. et al. (2010b). Mixed linear model approach adapted for genome wide association studies. *Nature Genetics*, 42, pp.355–360.

Zhao, C. et al. (2008). XND1, a member of the NAC domain family in *Arabidopsis thaliana*, negatively regulates lignocellulose synthesis and programmed cell death in xylem. *Plant Journal*, 53 (3), pp.425–436. [Online]. Available at: doi:10.1111/j.1365-313X.2007.03350.x.

Zhao, C. et al. (2017). XYLEM NAC DOMAIN1, an angiosperm NAC transcription factor, inhibits xylem differentiation through conserved motifs that interact with RETINOBLASTOMA-RELATED. *New Phytologist*, 1. [Online]. Available at: doi:10.1111/nph.14704.

Zhao, F. J. et al. (1993). Sulphur uptake and distribution in double and single low varieties of oilseed rape (*Brassica napus* L.). *Plant and Soil*, 150 (1), pp.69–76. [Online]. Available at: doi:10.1007/BF00779177.

Zhao, F. J. et al. (1997). Nitrogen to sulphur ratio in rapeseed and in rapeseed protein and its use in diagnosing sulphur deficiency. *Journal of Plant Nutrition*, 20 (4–5), pp.549–558. [Online]. Available at: doi:10.1080/01904169709365273.

Zhao, F. J. et al. (2014). Imaging element distribution and speciation in plant cells. *Trends in Plant Science*, 19 (3), pp.183–192. [Online]. Available at: doi:10.1016/j.tplants.2013.12.001.

Zhao, J. et al. (2016). ANAC005 is a membrane-associated transcription factor and regulates vascular development in *Arabidopsis*. *Journal of Integrative Plant Biology*, 58 (5), pp.442–451. [Online]. Available at: doi:10.1111/jipb.12379.

Zheng, L. et al. (2012). YSL16 Is a Phloem-Localized Transporter of the Copper-Nicotianamine Complex That Is Responsible for Copper Distribution in Rice. *The Plant Cell*, 24 (9), pp.3767–3782. [Online]. Available at: doi:10.1105/tpc.112.103820.

Zhou, Y. et al. (2018). Comparative transcriptomics provides novel insights into the mechanisms of selenium tolerance in the hyperaccumulator plant *Cardamine hupingshanensis*. *Scientific Reports*, 8 (1), pp.1–17. [Online]. Available at: doi:10.1038/s41598-018-21268-2.

Zlobin, I. E. et al. (2015). *Brassica napus* responses to short-term excessive copper treatment with decrease of photosynthetic pigments, differential expression of heavy metal homeostasis

genes including activation of gene NRAMP4 involved in photosystem II stabilization.

Photosynthesis Research, 125 (1–2), pp.141–150. [Online]. Available at: doi:10.1007/s11120-014-0054-0.

Zlobin, I. E., Kartashov, A. V. and Shpakovski, G. V. (2017). Different roles of glutathione in copper and zinc chelation in *Brassica napus* roots. *Plant Physiology and Biochemistry*, 118, pp.333–341. [Online]. Available at: doi:10.1016/j.plaphy.2017.06.029.