MOLECULAR CHARACTERIZATION OF *CDU-B1*, A MAJOR LOCUS CONTROLLING CADMIUM ACCUMULATION IN DURUM WHEAT (*Triticum turgidum* L. *var durum*) GRAIN

A Thesis Submitted to the College of
Graduate Studies and Research
In Partial Fulfillment of the Requirements
For the Degree of Master of Science
In the Department of Plant Sciences
University of Saskatchewan
Saskatoon, Saskatchewan

By
Krystalee D. Wiebe

© Copyright Krystalee Wiebe, September 2012. All Rights Reserved

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements of a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised this thesis work or, in their absence, by the Head of the Department or the Dean of the College in which this thesis work was done. It is understood that any copying or publication or use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Plant Sciences
51 Campus Drive,
University of Saskatchewan,
Saskatoon, Saskatchewan, Canada,
S7N 5A8

A portion of this thesis was published previously in the journal Theoretical and Applied and Genetics:

Wiebe K, Pozniak CJ, Harris N, Faris JD, Clarke JM, Knox RE, Taylor G (2010) Targeted mapping of *Cdu1,* a major locus regulating grain cadmium concentration in durum wheat (*Triticum turgidum* L. var. *durum).* Theor Appl Genet 121:1047-1058.

Publication available at:
http://www.springerlink.com/content/h2n7670w16ku58p5/?MUD=MP
DOI: 10.1007/s00122-010-1370-1

ABSTRACT

A major gene controlling grain cadmium (Cd) concentration, designated as *Cdu-B1,* has been mapped to the long arm of chromosome 5B, but the genetic factor(s) conferring the low Cd phenotype are currently unknown. Genetic mapping of markers linked to *Cdu-B1* in a population of recombinant inbred substitution lines (RSLs) revealed that the gene(s) associated with variation in Cd concentration reside(s) in wheat deletion bin 5BL9 between fraction breakpoints 0.76 and 0.79, and linked to two candidate genes; *PCS2* (phytochelatin synthetase) and *Xwg644*, which codes for a known ABC (ATP-binding cassette) protein. Genetic mapping and quantitative trait locus (QTL) analysis of grain Cd concentration was performed in a doubled haploid (DH) population and revealed that these genes were not associated with *Cdu-B1*. Two expressed sequence markers (ESMs), and five sequence tagged site (STS) markers were identified that co-segregated with *Cdu-B1*, and explained >80% of the phenotypic variation in grain Cd concentration. A gene coding for a P_{1B}-ATPase, designated as *OsHMA3* (heavy metal associated), has recently been associated with phenotypic variation in grain Cd concentration in rice. Mapping of the orthologous gene to *OsHMA3* in the DH population revealed complete linkage with *Cdu-B1* and was designated as *HMA3-B1*. Fine mapping of *Cdu-B1* in >4000 F_2 plants localized *Cdu-B1* to a 0.14 cM interval containing *HMA3-B1*. Two bacterial artificial chromosomes (BACs) containing full-length coding sequence for *HMA3-B1* and *HMA3-A1* (homoeologous copy from the A genome) were identified and sequenced. Sequencing of *HMA3-B1* from high and low Cd accumulators of durum wheat revealed a 17 bp duplication in high accumulators that results in predicted pre-mature stop codon and thus, a severely truncated protein. Several DNA markers linked to *Cdu-B1,* including *HMA3-B1,* were successfully converted to high throughput markers and were evaluated for practical use in breeding programs. These markers were successful at classifying a collection of 96 genetically diverse cultivars and breeding lines into high and low Cd accumulators and will have broad application in breeding programs targeting selection for low grain Cd concentrations. Current results support *HMA3-B1* as a candidate gene responsible for phenotypic differences in grain Cd concentrations in durum wheat.

ACKNOWLEDGEMENTS

I would like to express sincere gratitude to my supervisor, Dr. Curtis Pozniak for his mentorship throughout my degree. Also, the feed back from my advisory committee members (Drs. K. Bett, G. Scoles, and B. Coulman) has been valuable in the progress and completion of this project.

Financial assistance provided by the Warburtons Award, the Roderick Alan McLean Postgraduate Scholarship, the John Wickhorst Memorial Scholarship, and the Harris and Lauretta Raymond Earl Parr Postgraduate Scholarship was greatly appreciated.

I would also like to acknowledge the people, past and present, working in the Durum Wheat Breeding program at the University of Saskatchewan. To the field crew for their hard work and help in processing samples and soil. To the lab crew for their technical support and expertise. To Ron MacLachlan for his bioinformatics expertise and support. It was a pleasure to complete my degree in the Plant Sciences Department at the University of Saskatchewan. I would like to thank all those in the department that contributed to my success by sharing their expertise and encouragement. I would also like to thank Dr. Neil Harris (University of Alberta) for Cd phenotyping of Chinese Spring and CS-DIC5B and Dr. Justin Faris for providing the CS/CS-DIC5B mapping population and several molecular markers.

Finally, I would especially like to thank my family and friends. Through their encouragement and support I was able to persevere.

TABLE OF CONTENTS

PERMISSION TO USE i
ABSTRACT ii
ACKNOWLEDGEMENTS iii
TABLE OF CONTENTS iv
LIST OF TABLES vi
LIST OF FIGURES vii
LIST OF APPENDICIES viii
LIST OF ABBREVIATIONS ix

1.0 INTRODUCTION
- 1.1 Background 1
- 1.2 Research and Hypothesis and Objectives 2

2.0 LITERATURE REVIEW
- 2.1 Durum Wheat 3
- 2.2 Cadmium Related Health Concerns and Trade Restrictions 4
- 2.3 Cadmium Uptake in Plants 5
  - 2.3.1 Cadmium and Soil Properties 5
  - 2.3.2 Physiological Bases for Grain Cadmium Accumulation in Field Crops 6
  - 2.3.3 Genetic and Molecular Bases of Grain Cadmium Accumulation 9
- 2.4 Breeding for Low Grain Cadmium Concentration 13
- 2.5 Fine Mapping and Positional Cloning; A Forward Genetic Approach 15
- 2.6 Wheat, Rice, and *Brachypodium* Colinearity 22

3.0 MATERIALS AND METHODS
- 3.1 Verification of LDN5D(5A) and LDN5D(5B) Substitution Lines 27
- 3.2 Genetic Mapping Population 27
- 3.3 Determination of Grain Cd Concentration Source in Hexaploid Wheat 28
- 3.4 ESM and Gene Specific Marker Development 29
- 3.5 Saturation Mapping Marker Analysis 29
- 3.6 Saturation Genetic Mapping and QTL Analysis 31
- 3.7 Colinearity With the Rice and *Brachypodium* Genomes 31

3.8 STS Marker Development and Analysis 31
3.9 *OsHMA3* Gene Specific Marker Development..32
3.10 Development of Markers for Fine Mapping 32
3.11 Development of F_2 Fine Mapping Populations 33
3.12 *HMA-B1* BAC Identification 35
3.13 *HMA* Sequencing 36
3.14 Marker Validation 38

4.0 RESULTS
4.1 Saturation Mapping of *Cdu-B1* in Durum Wheat 39
4.2 Fine Mapping of *Cdu-B1* in Durum Wheat 48
4.3 Sequencing and Mapping of *HMA3* Orthologues in Durum Wheat 55
4.4 Validation of Molecular Markers Linked to *Cdu-B1* in Durum Wheat 62

5.0 DISCUSSION
5.1 Precise Genetic Localization of *Cdu-B1* in Durum Wheat 65
5.2 Localization of Other Genes Influencing Expression of Grain Cadmium Concentration in Wheat 69
5.3 Reverse Genetic Studies...71
5.3.1 Phytochelatin Synthase and ABC Transporters 71
5.3.2 Association of *HMA3-B1* with *Cdu-B1* in durum wheat 72
5.3.2.1 Genetic Mapping of *OsHMA3* orthologues in durum wheat 72
5.3.2.2 Comparative analysis of *HMA3* in durum wheat 73
5.4 Marker Validation and Breeding for Low Grain Cd Concentration 76

6.0 CONCLUSIONS
6.1 General conclusions 79
6.2 Future Work 81

7.0 COPYRIGHT CLEARANCE...83

8.0 REFERENCES 86

9.0 APPENDICIES 106

LIST OF TABLES

Table 2.1 QTLs identified for Cd accumulation in plants..10
Table 2.2 Genes associated Cd accumulation in plants..11
Table 2.3 Postionally cloned wheat genes 21
Table 4.1 Least square means of grain Cd concentrations (ng g^{-1}) from two environments for three markers associated with *Cdu-B1* and *XCbf32* 47
Table 4.2 Marker designations for converted CAPS markers 49
Table 4.3 Haplotype of three recombinant F_2 plants from the Svevo/Brigade population 50
Table 4.4 Haplotype groupings of homozygous $F_{2:3}$ lines derived from recombinant F_2 lines from the Svevo/Brigade population 51
Table 4.5 Haplotype of 20 recombinant F_2 plants from the 8982-TL fine mapping population 53
Table 4.6 Haplotype groupings of homozygous $F_{2:3}$ lines derived from recombinant F_2 lines from the 8982-TL mapping population..54
Table 4.7 Molecular variant frequencies and chi sq (χ^2) tests for expected segregation ratios for two F_2 populations...55
Table 4.8 Haplotyping of Durum cultivars and breeding lines using CAPS markers, *ScOPC20* and *HMA3-B1* markers...63

LIST OF FIGURES

Figure 2.1 Cd content or concentration over time in durum wheat and rice....................7
Figure 2.2 Proposed model for Cd translocation in rice 8
Figure 2.3 Genetic linkage map of a portion of chromosome 5BL and QTL analysis of *Cdu-B1*...11
Figure 2.4 Example of a three parent cross designed to fix the low grain Cd phenotype in the F_1 generation...15
Figure 2.5 Process of positional cloning 16
Figure 2.6 Phylogeny of grass species 23
Figure 2.7 Wheat genome and its colinearity to rice 24
Figure 2.8 *Brachypodium* genome colinearity with rice and wheat...............................25
Figure 3.1 Outline of F_2 fine mapping population development, screening, and phenotyping of F_2 individuals 34
Figure 4.1 Polymorphism detected at ScOPC20 in CS and CS-DIC 5B 39
Figure 4.2 Physical map of a portion of chromosome 5BL showing the deletion bins of that chromosome and corresponding genetic map of CS/CS-DIC 5B 40
Figure 4.3 Effects of substituting the 5D chromosome into Langdon on grain Cd concentration..41
Figure 4.4 Polymorphic ESM markers associated with *Cdu-B1* in the W9262-260D3/Kofa population 41
Figure 4.5 Genetic map of a portion of chromosome 5BL in the W9262-260D3/Kofa population and *Cdu-B1* QTL analysis...42
Figure 4.6 Genetic map of a portion of chromosome 5AL in the W9262-260D3/Kofa population 43
Figure 4.7 Genetic map of the *Cdu-B1* region from W9262-260D3/Kofa (Tt5B_Cdu-B1) and its colinear physical region on rice chromosome three (R3), and *Brachypodium* chromosome 1 (Bd1) 44
Figure 4.8 Polymorphic STS markers associated with *Cdu-B1* in the W9262-260D3/Kofa population 45
Figure 4.9 Polymorphic gene specific markers in W9262-260D3/Kofa population designed from genes linked to $VrnA^m1$ in *T. monococcum*...............................46
Figure 4.10 CAPS markers developed from ESMs and STS marker 49
Figure 4.11 Comparison of W9262-260D3/Kofa genetic saturation map, Svevo/Brigade genetic fine map, and 8982-TL genetic fine map 50
Figure 4.12 Genetic position of *HMA3-B2* on a portion of chromosome 2B in the W9262-260D3/Kofa mapping population..56
Figure 4.13 *HMA3* primers localized to chromosome 5B in wheat...............................56
Figure 4.14 Localization of *HMA3* BACs to chromosome 5A and 5B..........................57
Figure 4.15 *HMA3* sequencing..58
Figure 4.16 Primer sets for sequencing *HMA3-B1* from Kofa and W9262-260D3........59
Figure 4.17 *HMA* dendrogram 61
Figure 4.18 *HMA3-B1* 17 bp duplication 61

LIST OF APPENDICIES

Appendix 1 ESM, STS, and gene specific markers polymorphic between W9262-260D3 and Kofa 106

Appendix 2 Gene products in colinear region of *Cdu-B1* on rice chromosome 3 and *Brachypodium* chromosome 1 108

Appendix 3 SNP identified to design CAPS markers 110

Appendix 4 Additional *Cdu-B1* mapping markers, CAPS markers and *HMA* sequencing markers 111

Appendix 5 Graphical genotype of *Cdu-B1* and markers associated with *Vrn-B1* in near isogenic parental lines, Kofa and W9262-260D3 111

Appendix 6 Alignment of *HMA3* genomic sequence and predicted CDS, and reported genomic and CDS sequence from rice and *Brachypodium* 112

Appendix 7 Alignment of HMA3-A1 and HMA3-B1 predicted protein sequence from durum wheat 118

Appendix 8 Origin and pedigree of global collection of durum breeding lines and cultivars 119

LIST OF ABBREVIATIONS

AAFC	Agriculture and Agri-Food Canada
ABC	ATP Binding Cassette
BAC	Bacterial Artificial Chromosome
BLAST	Basic Local Alignment Search Tool
CAPS	Cleavage Amplified Polymorphic Sequence
Cbf	C-repeat binding factor
Cd	Cadmium
CDS	Coding DNA Sequence
CE	Capillary Electrophoresis
CFIA	Canadian Food Inspection Agency
CS	Chinese Spring
CS-DIC 5B	CS *Triticum dicoccoides* 5B substitution
CSF-1	Cleavage Stimulation Factor Subunit 1
DH	Double Haploid
EA	Environmental Agency
ESM	Expressed Sequence Marker
EST	Expressed Sequence Tag
FAO	Food and Agriculture Organization
IARC	International Agency for Research on Cancer
INDEL	Insertion/Deletion
INRA-CNRGV	France Plant Genomic Resources Centre
HMA	Heavy Metal Associated
HPA	Health Protection Agency
HRM	High Resolution Melt
LDN	Langdon
LOD	Log of Odds
LSD	Least Significant Difference
LSM	Least Square Mean
MAS	Molecular Assisted Selection

MC	Mitochondrial Carrier Protein
MLM	Multiple Locus Model
MTK4	Putative Protein Kinase Tousled
Mya	Million years ago
NCBI	National Center for Biotechnology Information
NIL	Near Isogenic Lines
NRAMP	Natural Resistance-Associated Macrophage Protein
NRC	National Research Council
PC	Phytochelatin
PCR	Polymerase Chain Reaction
PCS	Phytochelatin Synthetase
PHY-C	Phytochrome-C
QTL	Quantitative Trait Locus
RAPD	Random Amplified Polymorphic DNA
RNAi	RNA interference
RSL	Recombinant Inbred Substitution Lines
SCAR	Sequence Characterized Amplified Region
SDR	Segregation Distortion Region
SEM	Standard Error of Means
SNP	Single Nucleotide Polymorphism
SRC	Saskatchewan Research Council
SSCP	Single Strand Conformation Polymorphism
SSR	Simple Sequence Repeat
STS	Sequence Tagged Site
Tsn	Toxin sensitivity gene
TILLING	Target Induced Local Lesions in Genomes
VIGS	Virus Induced Gene Silencing
χ^2	Chi Squared
ZIP	ZRT-and IRT-like proteins
Zn	Zinc

1.0 INTRODUCTION

1.1 Background

Canadian durum wheat (*Triticum turgidum* L. var *durum*) production accounts for approximately two-thirds of the world exports of durum wheat and 18% of all wheat hectares planted in Canada (Stats Canada 2009). In Canada, the majority of the durum wheat is produced in the semi-arid regions of the western provinces of Saskatchewan, Manitoba, and Alberta with most production (84%) concentrated in Saskatchewan (Canadian Food Inspection Agency (CFIA) 2006). Given the importance of durum exports to Canadian durum wheat producers, it is necessary that durum wheat cultivars meet international trade regulations.

In durum, a single major gene for grain cadmium (Cd) concentration has been reported on chromosome 5B (Knox et al. 2009) and has been designated *Cdu-B1*. The closest DNA marker is a SCAR (sequence characterized amplified region) marker ScOPC20 (Knox et al. 2009) that maps 4.6 cM from *Cdu-B1* (Penner et al. 1995). Physiologically, Cd enters the plant from the soil through the roots, and is then translocated to above ground tissue via the xylem, where it is then available for remobilization to the developing grain presumably via the phloem. Harris and Taylor (2004) identified restricted root-to-shoot translocation as the mechanism responsible for reduced grain Cd concentration in durum wheat. Therefore *Cdu-B1* may be associated with one or more processes that influence Cd sequestering in the roots, or that prevent xylem loading of Cd.

Several health concerns have been associated with exposure to Cd and since cereal grain is a large source of dietary intake, maximum allowable tolerances for grain Cd have been imposed on cereal grains marketed internationally. Currently, the Codex Alimentarius of the Food and Agriculture Organization (FAO) has set a maximum level of 200 ng g^{-1} for grain Cd concentration (CODEX STAN 193-1995 2009) and is proposing a maximum level of 150 ng g^{-1}. In response, the western Canadian variety registration system now imposes a limit of 100 ng g^{-1} for all newly registered western Canadian durum wheat cultivars. Many durum wheat cultivars accumulate Cd in grain to levels higher than 200 ng g^{-1}, but genetic variation for grain Cd accumulation exists in

durum (Clarke et al. 1997b), and breeding for low grain Cd concentration is a priority in Canadian programs and globally. Because of the new proposed limits, there is a continued need for research and breeding to further lower grain Cd concentrations in durum wheat. In order to develop effective breeding strategies, a better understanding of the gene(s) involved in determining grain Cd concentration and better DNA markers to aid in selection are first required.

1.2 Research Hypothesis and Objectives

The objective of this research is to utilize reverse genetic approaches to associate gene(s) with variation in grain Cd concentration of durum wheat. Based on current information, the hypothesis of this research is that the *Cdu-B1* locus contains a gene(s) of several genes already identified in plants that can sequester Cd in roots, and therefore limit Cd availability for subsequent transport to vegetative tissues and the grain. Specifically, one of three gene families are hypothesized to be associated with *Cdu-B1*: a) a gene coding for a heavy metal associated (HMA) transporter, b) an ATP-binding cassette (ABC)-like transporter or c) a phytochelatin synthetase (PCS), as these have been associated with Cd uptake and sequestration in other plants. In addition, a secondary objective was to utilize a forward genetics approach to identify putative genes that maybe associated with variation in grain Cd concentration in durum. The last objective was to develop and validate breeder friendly markers that could be used globally in durum wheat breeding programs targeting the low grain Cd phenotype.

2.0 LITERATURE REVIEW

2.1 Durum Wheat

Globally, durum wheat (*Triticum turgidum* L. *var durum*) is an economically important crop with 41.1 million tonnes produced worldwide in 2010 (Agriculture and Agri-Food Canada (AAFC) 2010). Durum wheat is used primarily for production of pasta products (pasta, spaghetti, and macaroni), but its use in non-pasta products (leavened and unleavened bread, and bulgur) is increasing, particularly in Mediterranean regions (Elias and Manthey 2005). Historically, durum wheat has been grown in the regions around the Mediterranean Sea, including North Africa, southern Europe, Syria, and Turkey as durum wheat is better suited to areas where annual precipitation is low (semi-arid climates) (Elias and Manthey 2005). In North America, good quality durum is also produced in the dry growing regions of western North Dakota and Montana in the US, and southern Saskatchewan, Manitoba, and Alberta in Canada (AAFC 2005). Over 80% of Canadian production occurs in Saskatchewan.

Canada is the second largest durum wheat producer with 5.4 million tonnes produced in 2010 (Stats Canada 2011) and is also a major exporter. Nearly 80% of Canadian produced durum is exported into high quality markets and comprises more than 60% of world durum wheat trade (Clarke 2005). Thus, breeding for pasta quality is a primary objective in Canadian durum breeding programs.

Genetically, durum wheat is an allotetraploid derived from the hybridization and polyploidization of two ancestral grass diploid species: *Triticum monococcum* carrying the A genome and most likely *Aegilops speltoides* (Kilian et al. 2007) carrying the B genome. Durum wheat is closely related to bread wheat (*Triticum aestivum* L.) which is an allohexaploid having three genomes, two of which are similar to durum wheat (A and B genomes), as well as a third “D” genome, derived from *Aegilops tauschii*. The hexaploid wheat genome is large, spanning approximately 17 Gb in comparison to rice (~465 Mb) and *Arabidopsis* (~100 Mb). It is estimated that approximately 80% of the wheat genome is comprised of repetitive DNA, with only 20% of the genome containing coding sequence (Li et al. 2004). It was initially hypothesized that genes in wheat reside in “gene rich regions/islands” interspersed between large repetitive regions (Erayman et

al. 2004). Recently, Rustenholz et al. (2010) showed a slightly increased gene frequency from centromeric to telomeric regions of the chromosome. In this study, 40% of genes were considered to be in gene islands while 60% were isolated genes but no large gene free or gene poor regions were identified. Distribution of isolated genes was uniform while a higher frequency of gene islands existed in telomeric regions (Rustenholz et al. 2010).

2.2 Cadmium Related Health Concerns and Trade Restrictions

The primary mechanism of human exposure to Cd is through the consumption of contaminated foods and water as well as the inhalation of polluted air and tobacco smoke. Contaminated foods have been shown to be a great risk for Cd exposure with an estimated daily intake of 800 – 25 000 ng day^{-1} for foods grown in non-contaminated environments. Cigarette smoking (one pack a day) can further increase exposure by up to 100 ng day^{-1} (Gallagher et al. 2010).

Cadmium (Cd) is a nephrotoxic pollutant (Suwazono et al. 2006) as the kidney is the main target for Cd introduced into the human body (U.S. Department of Health and Human Service 1999). The skeletal system can also sequester Cd, which can cause calcium loss leading to osteoporosis and osteomalacia (Kazantzis 2004). Changes to bone cells and their associated function have not been identified (Bodo et al. 2010). In addition, Cd has been classified as a human carcinogen by the International Agency on Cancer Research (IARC 1993) as Cd can bind to DNA causing strand breaks and chromosome aberrations (Beyersmann and Hechtenberg 1997), which could lead to cancer causing mutations. As well, Cd has also been found to inactivate the DNA mismatch repair (MMR) system in cells, which leads to genome instability and increased risk of several types of cancer (Jin et al. 2003). An association between Cd concentrations in the urine and breast cancer has also been established (Gallagher et al. 2010). Cadmium (Cd) has also been found to modify gene expression at a cellular level, including genes involved in stress response, apoptosis, and intracellular metabolic pathways (Luparello et al. 2011).

With the risk of Cd entering the food chain, there is a need to decrease the amount of Cd in food products. The Codex Alimentarius Commission (CODEX STAN

193-1995 2009) allows a max level of Cd at 200 ng g^{-1} and currently a guideline level of 150 ng g^{-1} is being proposed. Even crops grown in non-polluted soils can have grain Cd levels which can easily exceed these maximum allowable levels (Morghan 1993).

2.3 Cadmium Uptake in Plants

Cadmium (Cd) accumulation in crops is influenced by several factors, including soil Cd content, soil type, concentrations of other micronutrients, climate, agronomic management, and physiological mechanisms that are defined by the genotype of the plant (Grant et al. 1998).

2.3.1 Cadmium and Soil Properties

Cadmium (Cd) is a naturally occurring heavy metal that is found in most agricultural soils. Typically, Cd levels in most agricultural soils do not exceed 400 - 500 ng g^{-1} (Fleischer et al. 1974) and are therefore considered non-polluted (Health Protection Agency (HPA) 2009 quoting Reimann and Caritat 1998). However, higher levels of Cd content have been observed, particularly near mining activity (Fleischer et al. 1974). Cadmium (Cd) contents of agricultural soil in the UK can range from 100 - 1800 ng g^{-1} with a mean of 400 ng g^{-1} (Environmental Agency (EA) 2007) while Cd concentration in Canadian soils have been shown to range from 100 - 8100 ng g^{-1} (Frank et al. 1976). Cadmium (Cd) and other heavy metals are added to soils through industrial and agricultural activity such as the application of phosphate fertilizers. Worldwide approximately 15,000 tons of Cd is mined for products such as nickel-cadmium batteries each year (McMurray and Tainer 2003).

Cadmium (Cd) is water soluble, and enters the plant partially through passive (Kudo et al. 2011) but mostly through active transport systems (Grant et al. 1998) in the plasma membrane of root cells, similar to Zinc (Zn) (Hart et al. 1998). Soil properties can influence the availability of Cd for plant uptake. Lower soil pH results in an increase in available Cd (Del Castilho and Chardon 1995) as higher concentrations of H^+ compete with Cd at cation exchange sites in the soil. This results in more Cd being in solution and available for plant uptake (Reichman 2002). In addition to contributing

Cd to the soil, phosphate fertilizers can have an acidifying effect on soil pH (Lambert et al. 2007).

Cadmium (Cd) availability is also influenced by the relative concentrations of other cations in the soil. Zinc (Zn) is an essential micronutrient for plants and is thought to share common uptake, translocation and remobilization pathways with Cd (Hart et al. 1998). The presence of Cd in Zn-deficient soil has been found to reduce plant Zn uptake (Shute and Macfie 2006). In contrast, when Zn content is sufficient, Cd has no effect or increases Zn accumulation (Shute and Macfie 2006). Varying levels of soil-applied Zn can also show effects on Cd plant accumulation. When Zn was added to optimal levels for bread wheat production, grain Cd concentration decreased (Oliver et al. 1994) suggesting Zn may compete with Cd for uptake by the plant. In contrast, Hart et al. (2002) found wheat has a higher affinity for Cd than Zn. When soils with high concentrations of Cd and Zn were evaluated, Shute and Macfie (2006) observed decreased Zn and increased Cd in soybean. Though the interactions between Cd and Zn have been reported at an uptake level, Clarke et al. (2002) observed that *Cdu-B1* only influenced grain Cd concentration in durum wheat and did not influence concentrations of Zn or other micronutrients concentration in the grain. These results suggest that although Zn and Cd may compete for uptake, *Cdu-B1* is involved in Cd specific assimilation in the plant.

2.3.2 Physiological Bases for Grain Cadmium Accumulation in Field Crops

Once Cd has entered the plant, it can either be sequestered in root tissue, and/or be translocated to the above ground vegetative tissues via the xylem (Salt et al. 1995; Hart et al. 1998) and to the seed, most likely via the phloem (Popelka et al. 1996; Hart et al. 1998; Harris and Taylor 2001). Transpiration promotes Cd translocation in plants supporting the role of xylem-mediated transport of Cd from the roots to the leaves (Haag-Kerwer et al. 1999). However, several studies have shown that phenotypic differences in grain Cd concentrations are poorly related to whole-plant Cd accumulation (Farrell et al. 2005). Using near isogenic lines (NILs) of durum wheat, Harris and Taylor (2004) found that whole plant Cd concentrations were similar between high and low grain Cd lines (Fig. 2.1a). However, the concentration of Cd in

roots was higher in low Cd accumulating lines coupled with lower shoot Cd concentrations. The opposite was true in high accumulators (Fig. 2.1a). Similar results were found in Cd partitioning studies in rice, where in low grain accumulators of Cd, the majority of whole plant Cd was present in the roots (Fig. 2.1b; Ueno et al. 2009). These results suggest that Cd translocation from the roots to the shoots is the major physiological process associated with variation in Cd concentration of shoots and grain (Ueno et al. 2009; Uraguchi et al. 2009). In durum wheat lines with high grain Cd concentration, 40 to 50% of whole-plant Cd is translocated to the shoots after 14 d growth (Harris and Taylor 2004; Hart et al. 2006). Similarly, in rice, 49% of whole plant cadmium is translocated to shoots in high grain accumulators of rice (Ueno et al. 2011). Taken together, genotypic differences in grain Cd concentration in both rice and durum wheat is most associated with restricted root-to-shoot translocation. This would limit the pool of available Cd in vegetative tissues for subsequent remobilization during grain filling (Harris and Taylor 2004; Uraguchi et al. 2009). Given the physiological similarities in restricted root to shoot translocation of Cd between rice and durum wheat, it is plausible that a similar genetic mechanism is operating.

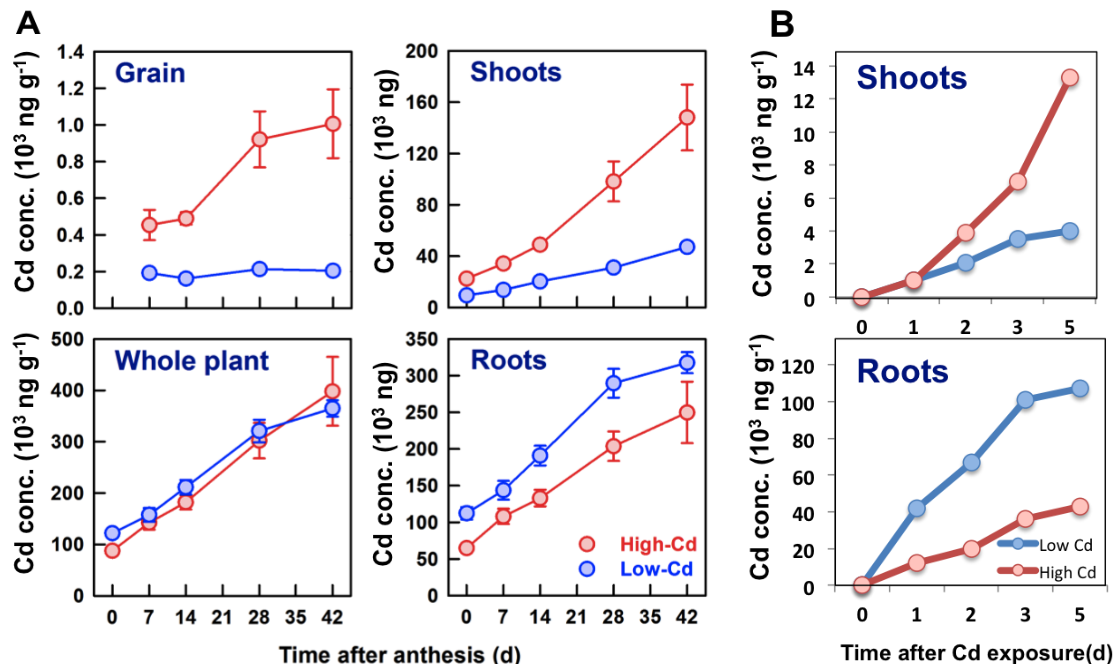

Fig. 2.1 Cd content or concentration over time in A) Durum wheat in the whole plant, grain, shoot and root tissue (Modified from Harris and Taylor 2004) and in B) Rice shoot and root tissue (Modified from Ueno et al. 2009).

The mechanism that limits Cd translocation to the shoots has not been identified in durum wheat. Cd translocation is likely to be complex and could include sequestration of Cd inside root cells, or reduced capacity for xylem loading or transport. Studies on the genetic inheritance of the low Cd trait have shown that the low Cd phenotype in most plants is controlled by few genes, and in durum is controlled by a single gene (Clarke et al. 1997b). It is doubtful that one or a few genes could account for genotypic differences in loading or transport of Cd. Rather, restricted Cd translocation to the shoots may be attributable to greater sequestration of Cd in the roots symplasm. In rice, a P_{1B}-ATPase transporter gene (*OsHMA3*) has been associated with limiting root to shoot translocation of Cd (Ueno et al. 2010). Miyadate et al. (2011) determined that the OsHMA3 coded protein localized to the tonoplast and functions to transport Cd into the vacuole. In hypersensitive yeast expressing a functional OsHMA3 protein, Cd tolerance was restored in the presence of Cd (Miyadate et al. 2011). These data support the hypothesis that rice plants with a functional allele of *OsHMA3* sequester Cd by transporting Cd into the vacuole, and thereby limiting the amount of Cd available for subsequent translocation to the shoots and grain (Fig 2.2). In contrast, rice plants with a

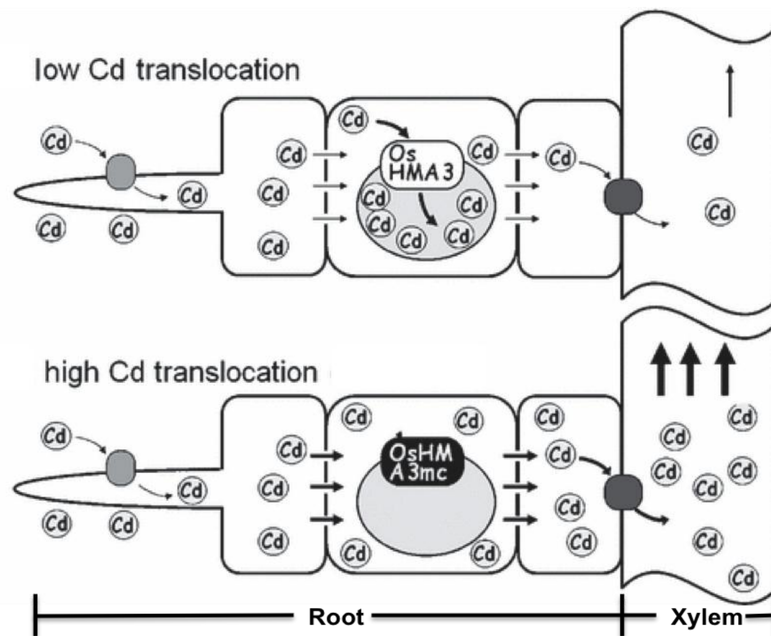

Fig. 2.2. Proposed model for Cd translocation in rice (modified from Miyadate et al. 2011). OsHMA3 is a functional HMA transporter while OsHMA3mc is a mutated non-functional HMA transporter.

non-functional OsHMA3 transporter protein cannot sequester Cd in the root cell vacuoles, and Cd is freely available for translocation to the shoots and grain (Fig. 2.2). To date, *HMA3*-related gene(s) have not been characterized in wheat, so it is not known if their coded protein(s) also function to restrict Cd to roots.

2.3.3 Genetic and Molecular Basis of Grain Cadmium Accumulation

Genetic variation for grain Cd concentration exists in many crop species including durum wheat (Penner et al. 1995), rice (*Oryza sativa* L.; Ishikawa et al. 2005), oat (*Avena sativa* L.; Tanhuanpää et al. 2007), sunflower (*Helianthus annuus*; Anderson and Hansen 1984), flax (*Linum usitatissimum*; Morghan 1993; Cieslinski et al. 1996), and soybean (*Glycine max* L. Merr; Kobori et al. 2010). In rice, Liu et al. (2006) found a range of 280 – 1840 ng g^{-1} in six rice cultivars. Tanhuanpää et al. (2007) found a range of 650 - 3490 ng g^{-1} in a mapping population of oat. Li et al. (1997) determined an average grain Cd concentration for nine hybrid sunflower cultivars to be 1510 ng g^{-1}, a range from 800 – 1550 ng g^{-1} in 14 flax cultivars, and a range of 110 – 340 ng g^{-1} for 30 durum wheat lines. Kobori et al. (2010) determined a range of 210 – 420 ng g^{-1} in three soybean cultivars. Li et al. (1997, quoting Hinesly et al. (1978)) determined a range of 50 - 1810 ng g^{-1} Cd in the grain of 20 corn inbred lines. Hexaploid wheat cultivars show little variation in grain Cd concentration (Zook et al. 1970) and are considered to be low in grain Cd concentration in comparison to high grain Cd accumulating durum wheat lines.

Given the complexity of Cd uptake, sequestration, and translocation, grain Cd accumulation can be regulated by multiple genes with combined effects (Tanhuanpää et al. 2007). However, in most cereals, Cd content is controlled by only one to three genes. In rice, three quantitative trait loci (QTL) on chromosomes 3, 6 and 8 have been identified (Ishikawa et al. 2005). Kashiwagi et al. (2009) also identified three QTL for Cd concentration in vegetative tissues of rice, two on chromosome 4 and on chromosome 11 (Table 2.1). More recently a single QTL on rice chromosome 7 was identified to control translocation of Cd from root to shoot (Ueno et al. 2009). Tezuka et al. (2009) and Ishikawa et al. (2010) also identified QTLs for grain Cd concentration in rice and in both studies chromosome 7 was identified. In maize (*Zea mays* L.), a

Table 2.1 QTLs identified for Cd accumulation in plants

Species	QTL Chromosome	Author
Oryza sativa L.	3, 6, 8	Ishikawa et al. 2005
	4, 11	Kashiwagi et al. 2009
	7	Ueno et al. 2009
	7	Tezuka et al. 2009
	7	Ishikawa et al. 2010
Thlaspi caerulescens	3	Deniau et al. 2006
Arabidopsis	3	Courbot et al. 2007
Zea mays L.	2	Soric et al. 2009
Avena sativa L.	unknown	Tanhuanpää et al. 2007
Triticum turgidum L.	5B	Knox et al. 2009
Glycine max L.	9	Jegadeesan et al. 2010
	9	Benitez et al. 2010

QTL for leaf Cd accumulation has recently been identified on chromosome 2 (Soric et al. 2009). In oat (*Avena sativa* L), a single QTL has been reported (Tanhuanpää et al. 2007). In soybean a major QTL has been identified on chromosome 9 in two independent studies (Jegadeesan et al. 2010; Benitez et al. 2010). In durum wheat, grain Cd concentration is controlled by a major gene designated as *Cdu-B1* and other minor genes (Knox et al. 2009). Recently, a major QTL associated with grain Cd concentration was reported on chromosome 5B of durum wheat (Fig. 2.3; Knox et al. 2009). In that study grain Cd concentration was also mapped as a Mendelian factor, and a single gene, designated as *Cdu-B1*. To date, no studies have been conducted to determine the number of genes controlling Cd concentration in flax or sunflower. *Thlaspi caerulescens* is a hyperaccumulator of Cd and a single QTL for Cd shoot concentration has been reported in that species (Deniau et al. 2006).

Several genes have been reported to be associated with phenotypic variation for Cd uptake, accumulation, and grain concentration (Table 2.2). In rice, a gene coding for a P_{1B}-ATPase was identified (Ueno et al. 2010) which is hypothesized to sequester Cd to root vacuoles where it is no longer available for transport to the shoots (see section 2.3.2). Genes associated with concentration of grain Cd in rice also include a novel gene *LCD* (low cadmium; Shimo et al. 2011) as well as *OsNRAMP1* (natural resistance-associated macrophage proteins) which is an iron transporter (Takahashi et al. 2011) and *OsZIP8* (ZRT-and IRT-like protein), which is a cadmium/zinc transporting

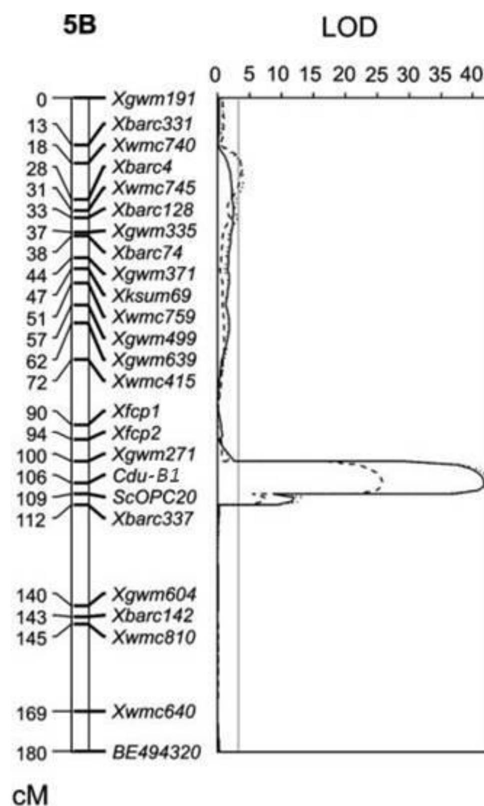

Fig. 2.3 Genetic linkage map of a portion of chromosome 5BL and QTL analysis of *Cdu-B1*. QTL analyses of cadmium concentration are presented as LOD scores for Swift Current 2000 (· · ·), Swift Current 2001 (—), and the combined analysis (- - -). Significance is declared for QTL to the right of the vertical line located at LOD 3.1. (modified from Knox et al. 2009).

Table 2.2 Genes associated with Cd accumulation in plants

Species	Gene	Author
Arabidopsis	P_{1B}-ATPase	Morel et al. 2009, Courbot et al. 2007
	ABC transporter	Wojas et al. 2009, Kim et al. 2007
	selenium binding protein	Dutilleul et al. 2008
	phytochelatins	Salt and Rauser 1995
Oryza sativa L.	P_{1B}-ATPase	Ueno et al. 2010
	LCD	Shimo et al. 2011
	OsNRAMP1	Takahashi et al. 2011
	OsZIP8	Ueno et al 2009, Ishimaru et al. 2005
Glycine max L.	ATPase	Jegadeesan et al. 2010
	P_{1B}-ATPase	Benitez et al. 2012

ATPase (Ueno et al. 2009; Ishimaru et al. 2005). Genes coding for ABC transporters (Wojas et al. 2009; Kim et al. 2007), P_{1B}-ATPases (Morel et al. 2009), and selenium-binding proteins (Dutilleul et al. 2008) have been associated with Cd accumulation and detoxification in *Arabidopsis,* as well as phytochelatins (PC) that bind Cd for subsequent sequestration in the vacuole (Salt and Rauser 1995). An ATPase was identified as a candidate gene for soybean grain Cd accumulation (Jegadeesan et al. 2010) as well as a P_{1B}-ATPases (Benitez et al. 2012).

Salt and Rauser (1995) found the vacuole to be an important organelle in the sequestration and detoxification of Cd in plants. The same authors suggested that membrane transporters facilitate this process by transporting Cd into vacuoles making Cd unavailable for transport/remobilization to the grain. Different families of transporter proteins have been found to be involved in Cd transport.

The ABC transporter family is a large ubiquitous and diverse superfamily of transmembrane bound proteins that facilitate the transportation, including lipids, hormones, and secondary metabolites (Verrier et al. 2008). Studies have shown that ABC transporters are involved in heavy metal sequestration in plants (Wojas et al. 2009, Kim et al. 2007). Yazaki et al. (2006) found that expression of a human ABC-MRP1 (multi drug resistance) protein expressed in tobacco can confer heavy metal tolerance and was thought to be localized to the vacuolar membrane, which allowed the sequestration of Cd. Transcript levels of ABC transporter *AtMRP3* showed a strong induction after the Cd treatment (Bovet et al. 2003), but currently, there is only anecdotal evidence that this transporter is involved in Cd transport. In durum, transcriptome profiling of high and low Cd accumulating NILs revealed higher expression of a gene coding for an ABC-MRP like transporter (Harris et al. 2007). An ABC-like transporter (*Xwg644*) have been identified on the homoeologous group five chromosomes in wheat near the vernalization loci (Yan et al. 2003) and has been sequenced from several cereals where a duplicate copy was identified in barley (Dubcovsky et al. 2001).

P_{1B}-ATPase proteins function by the binding and hydrolysis of ATP to create the energy for transport across a membrane against a concentration gradient (Williams and Mills 2005). In rice, Ueno et al. (2010) discovered a P_{1B}-ATPase (*OsHMA3*) as the

gene functioning to limit Cd translocation from root to shoots. As well, P_{1B}-ATPases have been associated in many species with grain Cd concentration including soybean (Benitez et al. 2012) and *Arabidopsis* (Morel et al. 2009). P_{1B}-ATPase in *Arabidopsis* (*AtHMA3*) has been shown to participate in vacuolar storage of Cd and other heavy metals (Morel et al. 2009). *OsHMA3* has been identified on chromosome 7 in rice, which shows colinearity to wheat group two chromosomes in wheat. To date, no reports of genes coding for P_{1B}-ATPases have been mapped in wheat.

Phytochelatins (PCs) are a class of small thiol (SH)-rich peptides that bind metals by thiolate coordination and are important cellular chelating agents. Several heavy metals can induce PC expression *in planta*, but Cd is the most effective metal inducer of PCs in *Arabidopsis* (Grill et al. 1985), *Brassica juncea* (Haag-Kerwer et al. 1999) and rice (Yan et al. 2000). PCs are thought to play a role in Cd detoxification by binding Cd to produce PC-Cd complexes, which can then be translocated into vacuoles (Salt and Rauser 1995; Vogeli-Large and Wagner 1996). In *Arapidopsis* the sequestration of PC-Cd complexes to the vacuole most likely occurs through ABC-MRP like transporter (Salt and Rauser 1995). In *Arabidopsis* and *Brassica juncea,* over-expression of *PCS* genes has been shown to increase Cd translocation from roots to the vegetative tissue (Heiss et al. 2003) contrary to the hypothesis of sequestration of Cd in the roots. Therefore, increased PC binding may allow increased efficiency of Cd translocation. *PCS2* has been identified on the group 5 chromosomes in wheat (Yan et al. 2003), and is tightly linked to *Xwg644*. Though some evidence may suggest PCs are not involved in Cd sequestration to the root, *PCS* is still a logical candidate gene due to its involvement in Cd binding and movement into the vacuole and due to the *PCS2* gene being mapped close to *Cdu-B1* in wheat.

2.4 Breeding for Low Grain Cadmium Concentration

Limiting the amount of Cd in durum grain to reduce potential risk to human health has been a priority of durum wheat breeders. For selection, a cost efficient and reliable method for assessing or predicting Cd concentrations in grain is required. Phenotypic selection for grain Cd concentration involves digestion of samples with nitric acid and assessing Cd using graphite furnace spectroscopy (Clarke et al. 2002).

Penner et al. (1995) determined the cost of commercial fees for phenotypically determining Cd content per sample to range from $15 to $23. Today, the cost of Cd analysis has increased to as high as $100/sample at commercial labs (i.e. Saskatchewan Research Council (SRC)). Due to high costs and low sample throughput, phenotypic selections were generally delayed to later generations on only the best breeding lines.

More recently, selection for grain Cd concentrations has been performed using marker assisted selection (MAS) with a SCAR marker (ScOPC20; Knox et al. 2009) developed from the sequence of a random amplified polymorphic DNA (RAPD) marker OPC-20 (Penner et al. 1995). With this marker, breeders can eliminate high grain Cd accumulating lines early in the breeding program at a reduced cost relative to phenotypic selection. ScOPC20 has been mapped approximately 5 cM away from the *Cdu-B1* (Penner et al, 1995) meaning molecular selection can still result in high Cd progeny as a result of recombination between *Cdu-B1* and ScOPC20. In addition, ScOPC20 is a dominant marker (Fig. 2.4a), and is linked in repulsion with *Cdu-B1* (Knox et al. 2009). Thus, ScOPC20 would not be an option for marker assisted backcrossing programs designed to introgress *Cdu-B1* into locally adapted material. However, this marker has been used effectively for gametic selection in the F_1 progeny resulting from three way crosses where one otherwise desirable parent is high in grain Cd. For example, breeders can make crosses where at least two of the parents are low Cd (i.e. High Cd/Low Cd//Low Cd) (Fig. 2.4b). As the third parent used is a low Cd type, Mendelian segregation would result in 50% of the F_1 progeny being homozygous for low grain Cd concentration and those individuals can be easily identified using ScOPC20 by the absence of a polymerase chain reaction (PCR) generated amplicon. However, because the marker is dominant, absence of an amplicon (homozygous low Cd line) could possibly be due to a failed PCR.

Indeed ScOPC20 is a better alternative to phenotypically testing grain samples for Cd concentration. However, a more closely linked, co-dominant marker for identification of *Cdu-B1* itself would improve selection efficiency and more accurately classify germplasm and breeding lines into high and low Cd accumulators. A more closely linked or perfect marker for *Cdu-B1* would reduce or eliminate selection error due to recombination. The development of a co-dominant marker will also reduce PCR

errors in that both the high and low grain Cd concentration molecular variants will be represented and the absence of a band would only mean that the PCR reaction failed.

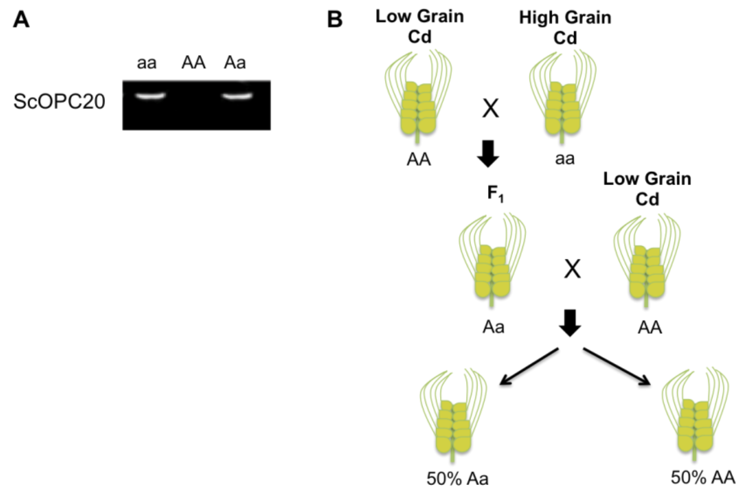

Fig. 2.4 Example of a three parent cross designed to fix the low grain Cd phenotype in the F_1 generation. A) Gel image of ScOPC20. B) Three parent cross used to identify plants homozygous for low grain Cd.

2.5 Fine Mapping and Positional Cloning; A Forward Genetic Approach

Positional or map-based cloning is a forward genetic approach to identify gene(s) controlling traits based on their precise genomic location, which is determined by phenotyping in well-defined large population. This is in contrast to a reverse genetics approach where a set of putative genes are identified (usually from related species) based on function and then tested for association with the phenotype in appropriate genetic populations. In a reverse genetics approach, if the hypothesized mechanism for the phenotype is incorrect, the gene will likely not be identified. A reverse genetic approach is generally used when there is strong evidence for a gene influencing a phenotype based on studies in model systems or other crop plants. For example, genes reported for variation in Cd accumulation in other species (Table 2.2) are reasonable candidate genes to test for association with *Cdu-B1*. In practice, reverse genetic approaches are generally used to determine the effect of a series of mutations of a phenotype and are used extensively for functional genomics studies in plants (Slade et al. 2005). In practice, both approaches are used in a combined effort to clone genes responsible for the phenotype.

Positional cloning involves five main steps: saturation mapping, fine mapping, chromosome walking, and identification and validation of candidate genes (Fig. 2.5). In the first step, a genomic location associated with phenotypic variation is saturated with molecular markers in a population segregating for that trait. This is generally achieved through classical genetic mapping experiments, usually in bi-parental mapping populations. In most cases, a small population (100-200 individuals) is used to estimate the general genomic location of the gene controlling the trait. To achieve map saturation, molecular markers co-segregating and closely flanking are required. Markers can include AFLP (amplified fragment length polymorphism), RFLP (restricted fragment length polymorphism), COS (conserved ortholog set), RAPD, CAPS (cleavage amplified polymorphic sequence), and SSR (simple sequence repeat). Additional markers can also be developed based on single nucleotide polymorphsims (SNPs) from within genes or ESTs (expressed sequence tag) known to be genetically linked to the

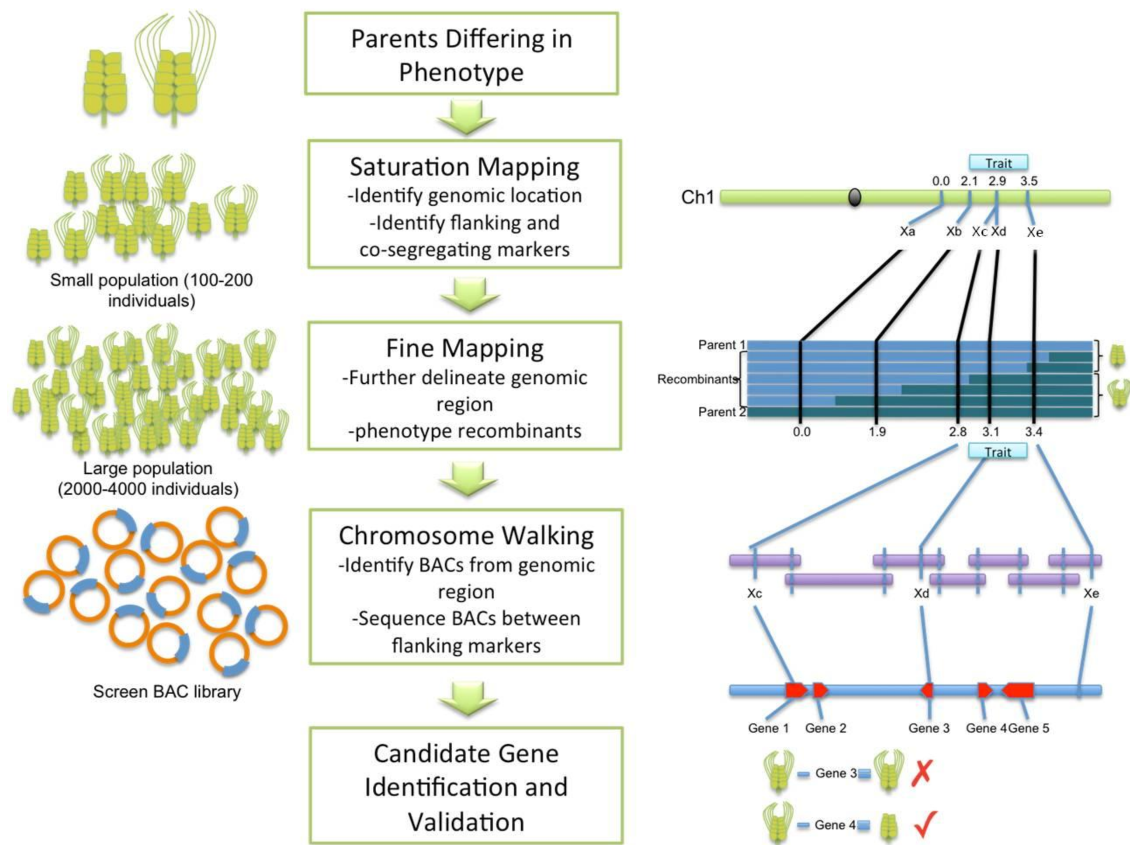

Fig. 2.5 Process of positional cloning.

trait or in colinear regions of a related species. For example, saturation of the *Tsn1* (toxin sensitivity gene) locus in wheat involved mapping of 28 EST-derived markers (ESMs) known to be physically linked to that locus to localize *Tsn1* to a 2.1 cM interval (Lu and Faris 2006). Ideally, to proceed with fine mapping, several co-segregating markers are required with closely linked (≤1 cM), flanking markers.

Once co-segregating markers are identified, the trait is first mendelized and mapped in a large F_2 population (2000-4000 individuals) to better delineate the genomic region (Fig. 2.5). For example, a total of 2707 F_2 plants were used to fine map the powdery mildew (*Pm6*) locus in hexaploid wheat (Qin et al. 2011). Large populations are favored to improve sampling of genetic recombination events, which in turn results in improved estimation of genetic distance and marker order (Ferreira et al 2006, Semagn et al. 2006).

The purpose of fine mapping is to identify the smallest possible genomic region associated with a trait. The physical size of the wheat genome is large, with the largest chromosome (3B) being over twice the size of the entire 370 Mb rice genome (Itoh et al. 2007). Furthermore, physical mapping of wheat chromosomes has revealed some chromosome segments which are high in gene density (Faris et al. 2000) and recombination frequencies are not consistent along chromosomes, with most cross-overs occurring in sub-telomeric regions of wheat chromosomes (Saintenac et al. 2009; Erayman et al. 2004). A major factor that would influence the success of fine mapping and ultimate gene identification is recombination frequency as higher recombination frequency results in a smaller genotypic to phenotypic distance ratio. With a high recombination frequency it is possible to detect a smaller physical region as mapping relies on recombinations between markers. Gene density can also affect fine mapping as high recombination frequencies are often associated with higher gene density (Faris et al. 2000). Therefore regions of high gene density should be regions of high recombination frequency and therefore ideal for fine mapping. Though some studies report higher recombination frequency in gene poor regions (Wei et al. 2002) therefore, gene density in not always a great predictor of recombination frequency (Mezard 2006). The particular combination of high gene density and low recombination frequency can make positional cloning of genes a daunting task due to the presence of many genes

over a large physical distance which cannot be resolved genetically through mapping. Recently, the *Tsn1* locus, located around 17 cM proximal to *Cdu-B1* (Knox et al. 2009) was cloned (Faris et al. 2010). Comparison with other high-density maps of 5B suggests that *Cdu-B1* resides in a gene-rich, recombination hot spot. Saturation mapping of *Tsn1* estimated the recombination frequency to be 400 kb/cM, an 11-fold increase in recombination compared to the genomic average (Faris et al. 2000), though further analysis of the locus determined a 3.18 Mb/cM physical to genomic distance (Faris et al. 2010).

Several loci may influence a phenotype (either through additive or epistatic effects), therefore it is generally accepted that lines near isogenic for the gene of interest are used as parents of the fine mapping population (Kim et al. 2010). This is to ensure only segregation occurs at the locus of interest and closely linked markers. In this large population, phenotyping of recombinants between flanking markers will position the phenotype in the fine map and better delineate the genetic distance between co-segregating markers and the trait. The markers closest genetically to the phenotype can now be used to screen a genomic, YAC (yeast artificial chromosome) or BAC (bacterial artificial chromosome) library. Several BAC libraries are available for hexaploid wheat (http://cnrgv.toulouse.inra.fr/Library), but only a single BAC library is available for durum wheat (Cenci et al. 2003). The library was generated from the cultivar Langdon, which carries a 30 cM insertion from *Triticum turgidum* ssp. *dicoccoides* on chromosome 6BS, and a gene for high grain protein content. The BAC library contains 516096 clones with an average insert size of 130kb (Cenci et al. 2003). The cultivar Langdon was chosen for the BAC library as disomic D genome substitution lines have been developed for this cultivar (Joppa and Williams, 1988) therefore allowing ease of genome assignment for BACs (Cenci et al. 2004).

For positional cloning experiments, BACs carrying co-segregating or flanking markers are first identified (Fig. 2.5). Often the 5’ and 3’ ends of these BACs are sequenced to allow development of additional markers to rescreen the BAC library to identify additional, overlapping BACs. In wheat it has been observed that BAC end sequence is often not sufficient for identifying unique sequence for developing additional marker (Stein et al. 2000). Generally, in wheat the whole BAC is sequenced

to identify unique sequence for additional markers, which would be useful in identifying overlapping BACs. This process of chromosome walking is repeated until BACs spanning the complete genetic interval are identified. Sequencing and annotation of the BACs in the minimum tiling path is then performed (either using Sanger or next generation sequencing). Software has been developed such as FGENESH (http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind), which can predict gene sequence from genomic sequences. These genes identified from this sequencing are all candidate genes. In wheat, several BACs have been sequenced and, on average, there is usually only one gene per BAC (Rustenholz et al, 2010; Devos et al. 2005). Some BACs from wheat are completely devoid of functional genes (Devos et al. 2008; Choulet et al. 2010) and as many as 10 genes have been identified on a single BAC (Choulet et al. 2010).

Validation of a candidate gene can be performed by RNAi (RNA interference), transgenics, VIGS (virus induced gene silencing), or TILLING (target induced local lesions in genomes). Huang et al. (2003) demonstrated the use of transgenics for candidate gene validation. Wheat lines susceptible to leaf rust were transformed with the candidate gene. These transformed susceptible lines then displayed a resistant reaction validating the candidate gene as causing resistance to leaf rust. Faris et al. (2010) employed TILLING for validating the *Tsn1* gene in wheat, by chemically mutagenizing a resistant line and screening for susceptibility. The candidate gene was then sequenced from multiple mutant lines and non-functional mutants were analyzed for phenotypic expression of resistance to *Stagonospora nodorum* blotch. Care must be taken when developing a TILLING population, as the line chosen for mutation must carry the functional gene. Mutations to genes are more likely to cause a functional protein to become non-functional, than a non-functional protein to become functional. Wang et al. (2011) employed RNAi to verify the *Opaque7* gene as the gene affecting storage protein synthesis in maize endosperm. RNAi employs dsRNA to target complementary mRNA for degradation and therefore gene silencing. Wang et al. (2011) targeted the candidate gene, *Opaque7*, through RNAi and saw a change in seed protein content compared to non-RNAi targeted lines, indicating that the *Opaque7* gene controls storage protein synthesis. Candidate genes can also be validated using allele

diversity studies, and expression analysis (Krattinger et al. 2009). In most cases a combination of above methods are used to validate candidate genes.

Major breakthroughs in sequencing and polymorphism detection technologies along with existing databases of developed markers and sequenced genomes has made positional cloning routine in some species including *Arabidopsis*, rice, *Brachypodium*, corn, and sorghum where the time and effort for positional cloning has decreased dramatically (Jander et al. 2002). However, positional cloning in plants with complex genomes with little sequence information such as hexaploid and durum wheat is more challenging. In wheat the greatest complexities are a) the polyploid nature of the wheat genome, b) the large size of the wheat genome (17 Mb in hexaploid wheat), and c) the repetitive nature of the wheat genome, where approximately 80 percent of DNA sequences in the wheat genome are repetitive partially from transposons (Li et al. 2004).

The polyploid nature of wheat can cause a challenge when positional cloning. This is because of similarity between homoeologous genes on homoeologous genomes and potential inability to easily determine one genome from another when identifying BACs. The size of the wheat genome is a detriment for positional cloning, as a larger library of BAC clones has to be developed and screened to sufficiently cover the whole genome. For example an *Arabidopsis* Ath-B-ITA library (http://cnrgv.toulouse.inra.fr/library/genomic_resource/Ath-B-ITA) required only 15360 BAC clones for 13X coverage of the genome while the durum wheat BAC library Ttu-B-LDN65 (http://cnrgv.toulouse.inra.fr/library/genomic_resource/Ttu-B-LDN65) is composed of 516096 BAC clones, representing only 5X coverage. A 9X coverage of the hexaploid wheat cultivar "Chinese Spring" (http://cnrgv.toulouse.inra.fr/library/genomic_resource/Tae-B-Chinese%20spring) contains 1147776 BAC clones. These large libraries increase the time and effort required to identify BACs carrying desirable genes. Also the repetitive nature of the wheat genome can further complicate BAC identification, as unique sequence is necessary for BAC identification. Entire BACs could be made up of repetitive DNA making it difficult for chromosome walking especially if the region to be spanned contains a large stretch of repetitive DNA.

Despite these complexities at least 11 wheat genes have been discovered through positional cloning in wheat (Table 2.3). These include genes for leaf rust resistance

Table 2.3 Postionally cloned wheat genes (Modified from Krattinger et al. 2009)

Gene Name	Trait	Gene Function	References
Gpc-B1	Grain protein content	NAC transcription factor controlling senescence and grain protein, zinc and iron content	Uauy et al. 2006
Q	Domestication	AP2 transcription factor influencing threshing character and spike phenotype	Faris et al. 2003
VRN1	Vernalization	AP1 like MADS-box transcription factor controlling flowering	Yan et al. 2003
VRN2	Vernalization	Dominant repressor of flowering that is down regulated by vernalization	Yan et al. 2004
VRN3	Vernalization	An orthologue of Arabidopsis flowering time (FT)	Yan et al. 2006
Ph1	Chromosome pairing	Multigene family, heterochromatin of both. A major chromosome pairing locus in polyploid wheat	Griffiths et al. 2006
Lr1	Leaf rust resistance	CC-NBS-LRR Leaf rust resistance	Qiu et al. 2007, Cloutier et al. 2007
Lr10	Leaf rust resistance	CC-NBS-LRR Leaf rust resistance	Stein et al. 2000, Wicker et al. 2001, Feuillet et al. 2003
Lr21	Leaf rust resistance	NBS-LRR Leaf rust resistance	Huang et al. 2003
Lr34	Leaf rust resistance	ABC transporter	Dakouri et al. 2010
Yr36	Stripe rust resistance	WKS1	Fu et al. 2009
Tsn1	Tan spot resistance	S/TPK -NBS-LRR ToxA sensitivity	Faris et al. 2010

NBS - nucleotide-binding site, LRR - leucine-rich repeat, CC-coiled-coil, S/TPK-serine/threonine protein kinase

(Qiu et al. 2007; Cloutier et al. 2007, Stein et al. 2000; Wicker et al. 2001; Feuillet et al. 2003; Huang et al. 2003; Dakouri et al. 2010), stripe rust resistance (Fu et al. 2009), resistance to tan spot (Faris et al. 2010), domestication (*Q*) (Faris et al. 2003), grain protein content (Uauy et al. 2006) and vernalization (Yan et al. 2003; Yan et al. 2004; Yan et al. 2006). Several of these genes are the targets of MAS in wheat breeding programs (Feuillet et al. 2003; Huang et al. 2003), and identification of allelic variation in these genes has allowed the development of useful DNA markers. Recent efforts to sequence the wheat genome (Feuillet and Eversole 2007) will increase the ability to identify wheat genes through positional cloning efforts.

2.6 Wheat, Rice, and *Brachypodium* Colinearity

Macro-colinearity has been described as conserved marker and/or gene order between species (Gale and Devos 1998). Extensive macro-colinearity has been observed throughout the grass (*Poaceae*) family. Conservation of gene sequence (micro-colinearity) has also been observed in the grass family. Colinearity is evident throughout the grass family but the extent of colinearity is based on evolution of the grass species. Species that diverged from each other later in evolution are expected to share greater colinearity than those that diverged earlier. Rice (*Oryza sativa*) and *Brachypodium* (Fig. 2.6) are the closest related species to wheat that have sequenced genomes and therefore are a good reference for studying the evolution of the wheat genome. Sorghum (*Sorghum bicolor*) (http://mips.helmholtz-muenchen.de/plant/sorghum/) and maize (*Zea mays*) (http://www.maizegdb.org/) also have sequenced genomes and although not closely related to wheat can still provide information about the wheat genome (Fig. 2.6).

The rice genome was sequenced because of its smaller size and agronomical importance and is considered one of the model species for the grass family (International Rice Genome Sequencing Project 2005). Several studies have examined the colinearity between the wheat and rice genomes using mapped ESTs and RFLPs (La Rota and Sorrells 2004, Salse et al. 2008) as genetic anchors and large portions of these genomes are colinear but large segments have been rearranged (Fig. 2.7). The long arm of the 5B chromosome of wheat shows colinearity to portions of rice chromosomes 6,

12, 9, and 3. This is in contrast to the wheat to group 3 chromosomes, which shows good macro-colinearity to rice chromosome 1 (LaRota and Sorrells 2004; Fig. 2.7).

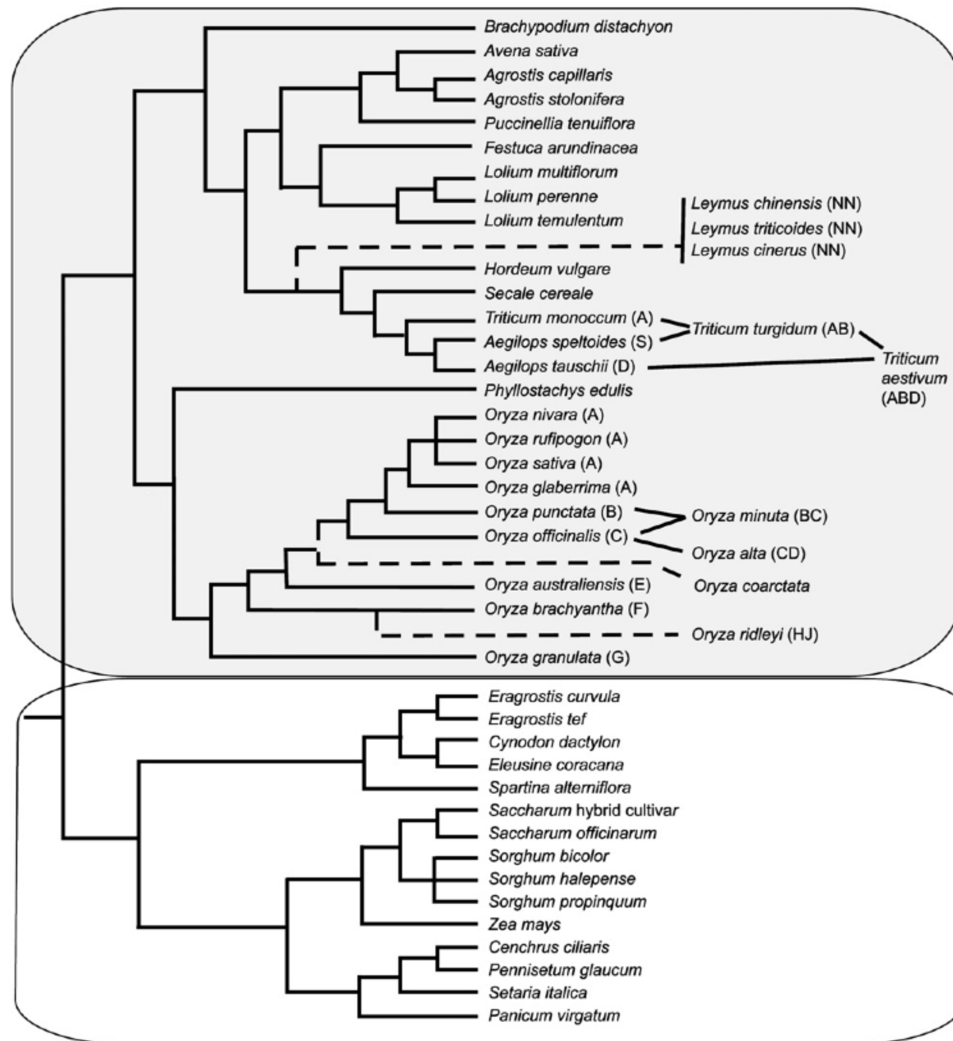

Fig. 2.6 Phylogeny of grass species (modified from Buell 2009). Genome designations are in brackets.

The *Brachypodium* genome has also been sequenced and it is phylogenetically closer to wheat than rice (International *Brachypodium* Initiative 2010) and colinearity relationships with rice and wheat have been assessed (Fig. 2.8). Current sequence information supports that *Brachypodium* is more closely related to wheat than rice (Bossolini et al. 2007; Huo et al. 2008; Kumar et al. 2009; Buell 2009; Fig. 2.8). Though the complete wheat genome has not been sequenced to make a complete

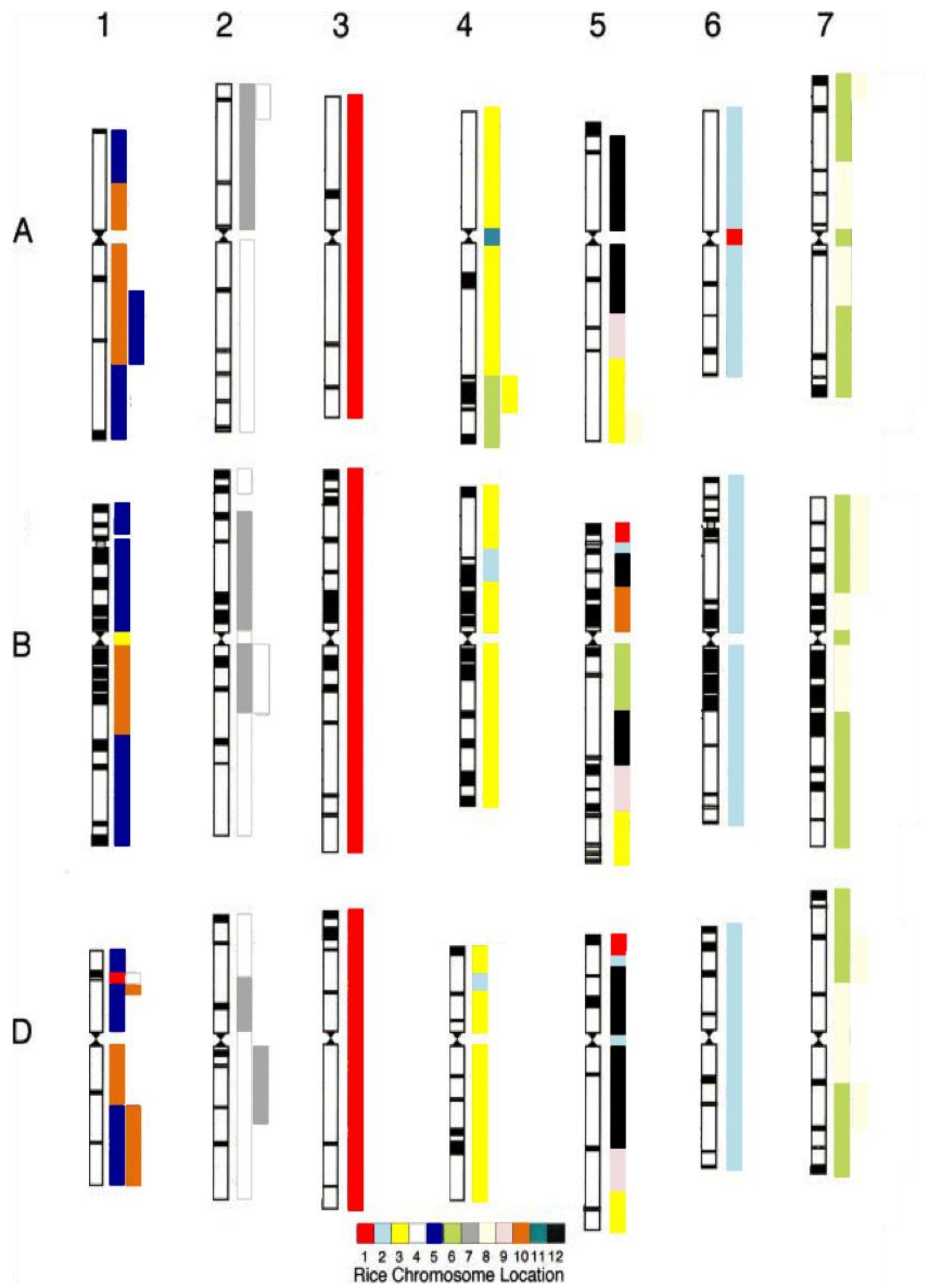

Fig. 2.7 Wheat genome and its colinearity to rice (modified from La Rota and Sorrells 2004)

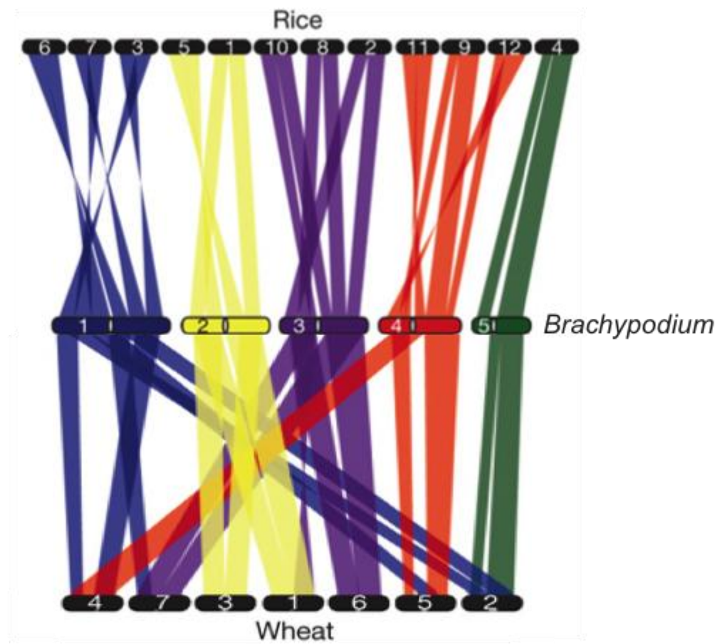

Fig. 2.8 *Brachypodium* genome colinearity with rice and wheat (modified from the International *Brachypodium* Initiative 2010)

comparison multiple loci have been compared between the three species as well as EST sequence. Both rice and *Brachypodium* genomes can provide gene information and putative locations for wheat scientists. Frequent interruptions in micro-colinearity are often observed at the level of individual genes in the form of deletions, translocations, and duplications (Bennetzen and Ma 2003; Feuillet and Keller 2002). Variability in the size of colinear regions between species is also not uncommon but this does not necessarily indicate a difference in gene content. The difference is usually due to differences in intragenic regions due to transposable elements (Bennetzen and Ma 2003). Wheat group 5 chromosomes show colinearity to *Brachypodium* chromosomes 1 and 4, which show colinearity to rice chromosomes 3, 7, 6, 9, 11 and 12 (Fig. 2.8) all of which, except chromosomes 7 and 11, show colinearity to the long arm of chromosome 5B (Fig 2.7).

Colinearity between species has been considered a useful tool in comparative genomics (Higgins et al. 2010), studying evolutionary relationship, and in gene discovery (Faris et al. 2008) including marker development (Qin et al. 2011). Positional cloning in wheat relies heavily on wheat/rice/*Brachypodium* colinearity. Many of the wheat genes cloned have benefited from sequenced colinear regions in rice and

Brachypodium. In most cases, high density genetic maps for wheat are achieved by developing markers from genes identified in colinear regions of these model plants (Yan et al. 2003; Fu et al. 2009). Also putative functions of sequenced wheat genes are identified through BLASTx (basic local alignment search tool) searches to identify orthologues in the rice and *Brachypodium* genomes (Faris et al. 2010).

Breakdowns in macro-colinearity and micro-colinearity are evident in all wheat positional cloning studies. For example, the *Tsn1* locus in wheat showed no colinearity (macro-colinearity) to *Brachypodium* but colinearity to rice aided in the identification of *Tsn1*, indicating that this region may not be conserved in *Brachypodium* as it is in wheat and rice (Lu and Faris 2006). At the wheat *Q* locus, wheat/rice micro-colinearity was more conserved than wheat/*Brachypodium* due to the presence of two non colinear genes identified in *Brachypodium* (Faris et al. 2008). However, when the sequences of the genes at the *Q* locus between the three species were analyzed, the wheat gene sequences were more closely related to *Brachypodium* than rice (Faris et al. 2008). For this particular locus, having more similar gene sequences agrees with the placement of *Brachypodium* closer to rice on the evolution scale. Also at this locus, the presence of two non-colinear *Brachypodium* genes identifies that micro-colinearity is no closer between wheat and *Brachypodium* than wheat and rice. Wicker et al. (2010), studying the "rate of gene movement" (the rate at which colinearity breaks down), observed that 20% of genes between species would be rearranged and no longer colinearity in 40 million years of evolution. With wheat and rice diverging 50 Mya (million years ago) and wheat and *Brachypodium* diverging around 35 – 40 Mya perfect colinearity cannot be expected and caution should be used when utilizing wheat/rice/*Brachypodium* colinearity in positional cloning. Regardless, rice and *Brachypodium* genomes are still valuable resources for marker development and positional cloning in wheat.

3.0 Materials and Methods

3.1 Verification of LDN5D(5A) and LDN5D(5B) Substitution Lines

As *Cdu-B1* has been localized to chromosome 5B, Langdon disomic substitution lines (Joppa and Williams 1988) LDN5D(5A) and LDN5D(5B) were used to validate 5B specific markers. In the Langdon disomic substitution lines, the Langdon 5A or 5B chromosome, respectively, have been substituted with the 5D chromosome from hexaploid wheat cultivar "Chinese Spring" (CS). These aneuploid lines were obtained from National Small Grains Collection (http://www.ars.usda.gov/main/docs.htm?docid=2884) and have limited quantities of seed. Because these are aneuploid lines, they were first verified to be missing the appropriate chromosome using simple sequence repeat (SSR) markers. SSR markers Xgwm291 and Xgwm293 were used to validate lines lacking the durum wheat 5A chromosome are 5A specific and SSR markers Xgwm371 and Xgwm408 were selected to verify the 5B substitution as these markers are known to be chromosome 5B specific (Somers et al. 2004). Primer sequences are available on the graingenes website (http://wheat.pw.usda.gov/GG2/index.shtml). The PCR reactions consisted of 50 mM KCl, 10 mM Tris-HCl, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.4 µM of forward and reverse SSR primers, 1.75 U of Taq DNA polymerase and 100 ng of genomic DNA. The total PCR volume was 25 µL. Temperature cycling was 94°C for 5 min followed by 32 cycles of 94°C for 30 s, 59°C for 30 s, 72°C for 1 min, then a final extension at 72°C for 10 min before cooling to 4°C. PCR samples were resolved by electrophoresis on 2.0-2.5% agarose gels at 140V for 1.5 hours in 1X TBE buffer and stained with ethidium bromide (0.5 µg/ml). The DNA banding patterns were visualized with UV light and recorded by a Canon Power Shot 7 digital camera and UVP imaging system. The presence/absence of these markers will determine whether the substitution has occurred.

3.2 Genetic Mapping Population

Genetic mapping was performed using 155 double haploid (DH) lines from the cross W9262-260D3/Kofa as this population was used previously to localize *Cdu-B1*

(Knox et al. 2009). Grain Cd concentration data was available from two environments for the DH population (Knox et al. 2009), and these data were used in the present study. A hexaploid wheat (*Triticum aestivum* L.) population consisting of 115 recombinant inbred substitution lines (RSLs) derived from a cross between Chinese Spring (CS) and a CS *Triticum dicoccoides* 5B substitution (CS-DIC 5B) described in Gill et al. (1996) was also used for genetic mapping. TA106, the *T. dicoccoides* source of 5B, was included in the molecular studies. Grain Cd from CS and CS-DIC 5B was performed by Dr. Neil Harris and has been published previously (Wiebe et al. 2010).

3.3 Determination of Grain Cd Concentration Source in Hexaploid Wheat

Durum wheat cultivars Langdon (high Cd), CDC Verona (low Cd), Kofa (high Cd), W9262-260D3 (low Cd), Commander (high Cd), and Strongfield (low Cd) as well a Langdon (LDN) disomic substitution lines LDN5A(5D) and LND5B(5D) (Joppa and Williams 1988) were used to determine if chromosome 5D contains genes that confer low grain Cd concentration in hexaploid wheat. These lines were planted and grown in soil collected from the Kernen Crop Research Farm near Saskatoon, Saskatchewan, which is known to contain sufficient Cd to differentiate high and low Cd accumulating durum wheat lines. Soils were air dried, mix thoroughly and distributed into 15 cm pots and placed into a growth chamber. Seeds of each line were pregerminated on Petri plates at 14°C for five days then one germinated seed was transplanted to each pot. Langdon substitution lines seed was of limited quantity and had to be verified as described in section 3.1. Therefore, only two replicates of each cultivar or line were planted. Growth conditions in the growth chamber were 16 h day, and 23/17°C day/night temperature. Plants were grown to maturity and grain from the entire plant was harvested for grain Cd concentration analysis. All grain Cd analysis was performed by Saskatoon Research Council (SRC) (Saskatoon, SK). The samples were digested with nitric acid and hydrogen peroxide using a Milestone ETHOS EZ microwave digester. The digest was analyzed on a Thermo X Series ll ICP-MS (Inductively Coupled Plasma-Mass Spectroscopy). Certified reference material, aquatic plant (*Lagarosiphon major*) from the European Joint Research Center, and matrix matched method blanks were carried through from the initial digest to the final analysis. The

phenotypic data was analyzed using Microsoft Excel (version 14.2.3). The mean and standard deviation of each mean is presented.

3.4 ESM and Gene Specific Marker Development

Markers were developed for ESTs previously localized to bin 5BL9 0.76-0.79. EST sequences were blasted against Michigan State University (MSU) rice genome annotation release 6.1 (Ouyang et al. 2007) using BLASTn (http://rice.plantbiology.msu.edu/blast.shtml). Rice genes with the best hit (e-values $<10^{-7}$ and ≥80% nucleotide identity for at least 60 bases) were then used as queries in BLASTn searches of *Triticum* sequences (National Center for Biotechnology Information (NCBI) BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi)). The *Triticum spp.* coding DNA sequence (CDS) and EST sequences were aligned with rice CDS and genomic sequences to determine putative intron/exon sites using AlignX (Vector NTI Advance 10.3; Invitrogen, Carlsbad, CA). If no significant hits were identified for *Triticum spp.* the rice genes were then BLASTn searched against the Chinese Spring 454 shotgun sequence (http://www.cerealsdb.uk.net/search_reads.htm) and these hits were used in the alignments. Primer pairs were designed from the wheat CDS or genomic sequences, and the target products included at least one intronic region. Overlapping primers were designed to ensure coverage of the majority of the CDS of each gene. A total of 120 ESM primer pairs were designed from 54 ESTs.

In addition, an earlier study (Knox et al. 2009) suggested that *Cdu-B1* resides in an area close to the major vernalization locus *vrn-B1*. The homoeologous locus $Vrn\text{-}A^{m}1$ has been sequenced, so a total of 56 primer pairs were designed for seven genes physically linked to that locus, including markers for genes *Xwg644* and *PCS2* using the same procedures as described above. All polymorphic ESM and gene-specific primer pairs are listed in Appendix 1.

3.5 Saturation Mapping Marker Analysis

A saturated map of 5BL has been reported previously for the CS/CS-DIC 5B population (Lu et al. 2006). As such, a dominant marker most closely associated with *Cdu-B1*, previously designated as ScOPC20 (Knox et al. 2009), was amplified in this

population using primers and PCR reaction conditions reported previously (Knox et al. 2009).

All ESM primers, and primers for the seven *Vrn-B1* associated genes were evaluated for polymorphisms first on genomic DNA from Kofa and W9262-260D3, and only those primers that produced polymorphic amplicons were assessed on the W9262-260D3/Kofa DH population and are reported here (Appendix 1). The PCR reactions consisted of 50 mM KCl, 10 mM Tris-HCl, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.1 μM of M13 sequence-modified forward ESM primer (M13 tag, CACGACGTTGTAAAACGAC, attached to 5' end of forward primer), 0.4 μM of reverse ESM primer, 0.152 μM of Universal dye-labeled M13 primer (Schuelke 2000), 1.75 U of Taq DNA polymerase and 100 ng of genomic DNA. The universal primer was labeled with either HEX, FAM, or NED fluorescent dyes. The total PCR volume was 25 μL. Temperature cycling was 94°C for 5 min followed by 3 cycles of 94°C for 30 s, 56°C for 45 s, 72°C for 45 s, 94°C for 30 s, 54°C for 45 s, 72°C for 45 s, 94°C for 30 s, 52°C for 45 s, 72°C for 45 s, 94°C for 30 s, 50°C for 45 s, 72°C for 45 s, then 32 cycles of 94°C for 30 s, 51°C for 45 s, 72°C for 45 s, then a final extension at 72°C for 10 min before cooling to 10°C. Primers were first assessed for polymorphisms using capillary electrophoresis (CE) (ABI3100xl; Applied Biosystems). For CE, 1 μL of diluted PCR product (diluted 1/20 or 1/10 in deionized water) was combined with 9.0 μL HiDi formamide (ABI, Foster City, CA), and 0.08 μL of 500(-250) ROX size standard. The samples were run on a 36 cm array, processed with Applied Biosystem Data Collection Software version 2.0, and genotyped using GeneMapper version 3.0. Monomorphic ESMs were further analyzed using single strand conformational polymorphism (SSCP). For SSCP analysis, 4 μL of the PCR product were mixed with 20 μL of loading buffer containing 95% formamide, 0.05% bromophenol blue, and 0.05% xylene cyanol. The samples were heated at 94°C for 5 min and then immediately placed on ice to allow single strand folding. The fragments were resolved on a 0.6X MDE gel (Lonza, Rockland, ME, USA) run at room temperature for 17 h (6 W) using 0.6X TBE buffer. The Bio-Rad Sequi-Gen GT System (38 cm X 50 cm) was used for electrophoresis. Gels were visualized by silver staining as described previously (Bassam and Gresshoff. 2007).

3.6 Saturation Genetic Mapping and QTL Analysis

Revised genetic linkage maps of the CS/CS-DIC 5B (Lu et al. 2006) and W9262-260D3/Kofa (Knox et al. 2009) populations were constructed using the Haldane mapping function of JoinMap 4.0 (van Ooijen and Voorrips 2004) at a minimum log of odds (LOD) score of 3.0. Only the W9262-260D3/Kofa population was used for QTL analysis using grain Cd concentration collected previously (Knox et al. 2009). QTL analysis was performed using a multiple locus model (MLM) in MapQTL Version 5.0 (van Ooijen 2004) and the significance threshold ($P<0.01$) of the LOD score was determined as described previously (van Ooijen 1999). For QTL analysis, the least square (LMS) means for each DH line was used and were estimated from data collected from two environments (Knox et al. 2009). The average QTL effects (one half the difference between parental marker class means) were estimated by MapQTL Version 5.0. All genetically mapped markers were designated with an X as per the recommended rules for gene symbolization in wheat.

3.7 Colinearity With the Rice and *Brachypodium* Genomes

For comparative analysis with the rice and *Brachypodium* genomes, the reported sequences of the ESMs linked to *Cdu-B1* were subjected to BLASTn searches of the rice and *Brachypodium* (http://www.brachybase.org/blast) genomes and parsed for sequences with e-values $<10^{-7}$ and ≥80% nucleotide identity for at least 60 bases. When several significant hits were found, only the best hit (lowest e-value) was reported. MapChart 2.1 (Voorips 2002) was used to view colinearity of the wheat genetic map, and the rice and *Brachypodium* physical maps. Colinear genes and functions are listed in Appendix 2.

3.8 STS Marker Development and Analysis

Once the *Cdu-B1* colinear regions were identified, markers were developed for rice and *Brachypodium* genes within the *Cdu-B1* colinear region in wheat using the same procedures described above (section 3.4). A total of 120 sequence tagged site (STS) primer pairs were designed based on 25 colinear genes. These markers were designed and analyzed as per the ESMs using rice and *Brachypodium* colinear gene

sequences (section 3.4). STS markers were analyzed as per ESMs (section 3.5). These markers were also assessed for polymorphism in the parents of our mapping population using high resolution melt (HRM) technology. For HRM analysis, all PCR reactions were performed using a BioRad CFX384 real time (RT) PCR machine. The PCR reactions included 1X Eva Green Supermix (BioRad), 1 μM forward and reverse primers, and 100 ng DNA. PCR temperature cycling was the same for all primers (98°C for 3 min followed by 40 cycles of 98°C for 10 s and 59°C for 10 s). Following PCR, amplicons were melted by ramping the temperature from 59°C to 98°C in 0.2°C increments which were held for 10 s after which fluorescence was measured. The resulting melting temperature curves were compared to determine polymorphisms. STS primer pairs polymorphic between W9262-260D3 and Kofa are listed in Appendix 1. Polymorphic markers were mapped in the W9262-260D3/Kofa population as described in section 3.6.

3.9 *OsHMA3* Gene Specific Marker Development

A P_{1B}-ATPase gene in rice (*OsHMA3, Os07g12900*) has been identified as controlling grain Cd concentration in rice (section 2.3.2), and seven primer pairs were designed from the closest orthologous wheat sequences (DQ490135 and AY829002) as well as the Chinese Spring shot gun 454 sequences (http://www.cerealsdb.uk.net/search_reads.htm). These primers were designed as described in section 3.4 and 5B and polymorphic primer pairs are listed in Appendix 1.

3.10 Development of Markers for Fine Mapping

Most *Cdu-B1* markers were mapped by using SSCP, which is not conducive to screening large numbers of F_2 individuals typical of fine mapping populations. Thus three co-segregating CAPS markers (*Xusw14*, *Xusw47*, *Xusw17*) and STS marker *Xusw15b* were converted to CAPS markers for the fine mapping experiments. For this, PCR amplicons of the appropriate size were cloned using the TOPO TA Cloning Kit (Invitrogen, Burlington, Ontario, Canada) following the manufacturer's instructions. Fragments from both high grain Cd concentration cultivars (Kofa and Commander) and low grain Cd concentration cultivars (W9262-260D3 and Strongfield) were cloned.

Recombinant plasmid DNA from clones carrying the inserted PCR product (as verified by PCR) were sequenced at McGill University (Genome Quebec Innovation Center in Montreal, Quebec, Canada). Sequences were aligned using AlignX (Vector NTI Advance 10.3; Invitrogen, Carlsbad, CA) and were analyzed for SNPs between high and low grain Cd accumulating cultivars (Appendix 3). Primers were redesigned to amplify B genome specific amplicons by placing a B genome specific SNP at the 3' end of the primer. Forward and reverse primers were selected to flank restriction sites differentiating alleles of high and low accumulators. Because ScOPC20 is a dominant (presence/absence) marker it could not be converted to a CAPS marker.

The resulting DNA fragment for each CAPS marker (Appendix 4) was PCR amplified using the following conditions. The PCR reaction consisted of 50 mM KCl, 10 mM Tris-HCl, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.25 pmol of forward CAPS primer, 0.25 pmol or reverse CAPS primer, 1.75 U of Taq DNA polymerase and 100 ng of genomic DNA. Temperature cycling was 94°C for 3 min followed by 35 cycles of 94°C for 30 sec, 57°C for 30 sec, 72°C for 1min, then a final extension at 72°C for 10 min before cooling to 10°C. PCR amplicons were resolved by electrophoresis on 2.0% (w/v) agarose gels at 140V for 1.5 hours in 1X TBE buffer and stained with ethidium bromide (0.5 μg/ml). The DNA banding patterns were visualized with UV light and recorded by a Canon Power Shot 7 digital camera and UVP imaging system.

All of the CAPS markers developed were first validated for their ability to amplify amplicons located on chromosome 5B by using the Langdon disomic D genome substitution lines (Joppa and Williams 1988). Markers were then scored on the W9262-260D3/Kofa mapping population to ensure the CAPS markers co-segregated with the ESM and STS markers from which they were developed.

3.11 Development of F_2 Fine Mapping Populations

Two F_2 populations were utilized for fine mapping. A smaller population consisting of 521 individuals was developed from the cross Svevo (high grain Cd) x Brigade (low Cd; Clarke et al. 2009). A larger F_2 population consisted of 3558 F_2 individuals derived from a cross between two near isogenic lines 8982-TL-L (low grain Cd concentration) and 8982-TL-H (high grain Cd concentration; Appendix 5; Clarke et

al. 1997a). The resulting F_1s from the latter cross were increased and were sampled and analyzed for markers linked to *Cdu-B1* for heterozygosity at the two leaf stage to ensure no F_1 plants were the result of self pollination. Only F_1s heterozygous at markers linked to *Cdu-B1* were grown to maturity and F_2 seeds harvested (Fig. 3.1).

For fine mapping, the F_2 plants were first germinated at 14°C in the dark until roots and shoots were visible. Germinated seedlings were then planted in sunshine mix (L22 germinating) in 8 by 16, 2.5 cm cell trays (96 individual seedlings per tray). Appropriate parents of each F_2 population were included in each tray as controls. Seedlings were allowed to grow for 7-10 days, until the seedling shoots were about 10 cm tall. DNA was extracted from seedlings as described previously (Eckstein et al 2004).

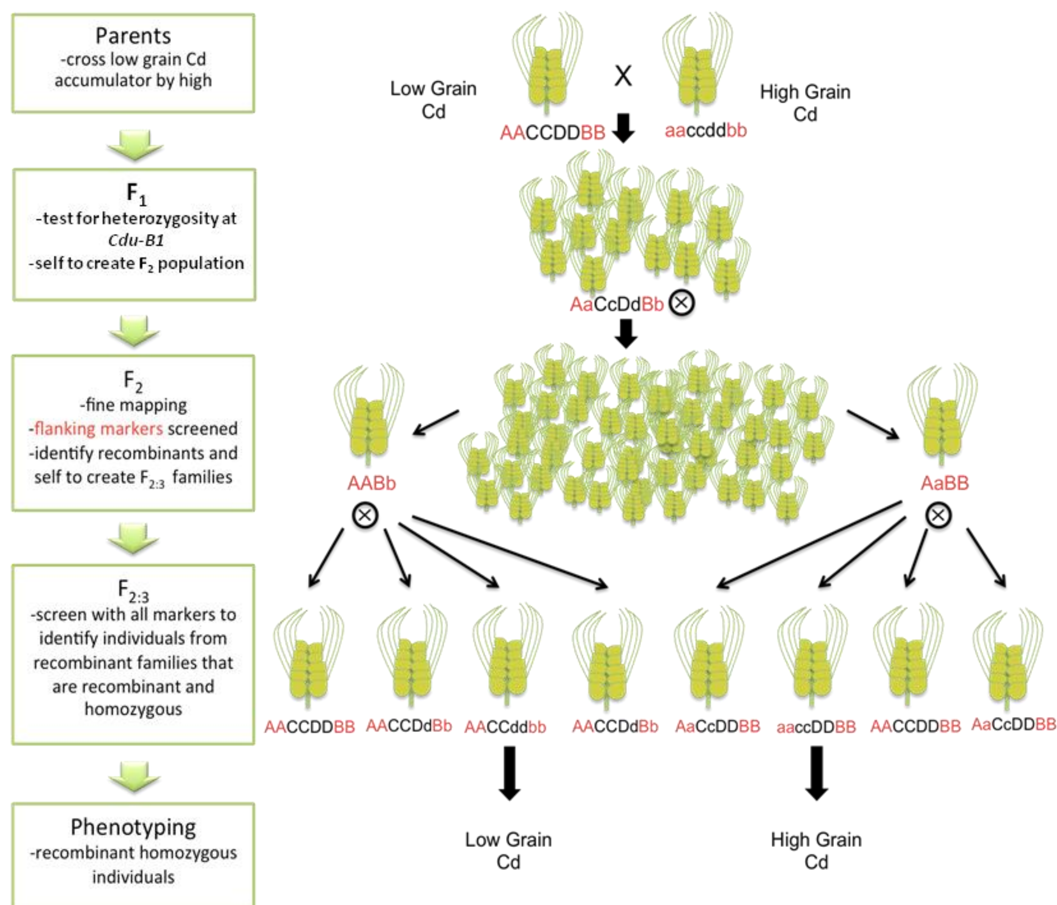

Fig. 3.1 Outline of F_2 fine mapping population development, screening, and phenotyping of F_2 individuals. Red alleles indicate flanking markers.

In total eleven markers were used for fine mapping (ScOPC20, *Xusw49, HMA3-B1*, *Xusw50*, *Xusw51*, *Xusw52*, *Xusw15b*, *Xusw17*, *Xusw47*, *Xusw53*, and *Xusw14).* The populations were first screened with three markers: ScOPC20 (section 3.5) and *Xusw14* (section 3.10) which flank *Cdu-B1*, and *Xusw47* (section 3.10), which co-segregates with *Cdu-B1.* The remaining eight markers that co-segregated with *Cdu-B1* in the W9262-260D3/Kofa population were scored only on F_2 plants showing recombination between at least two of the three original markers. Chi-square (χ^2) analysis of observed molecular variant frequencies was estimated using expected values for co-dominant (1:2:1) or dominant (3:1) segregation in an F_2 population.

$F_{2:3}$ progeny derived from only the recombinant F_2 individuals were grown as described for the F_2 plants, and all ten markers scored on 10-20 individual $F_{2:3}$ plants. Plants homozygous for recombinations were selected, and grown in Cd contaminated soil (section 3.3) and scored using all markers to identify $F_{2:3}$ families for phenotyping (Fig. 3.1). Parental material was grown and transplanted as checks. At maturity, grain was harvested from each selected $F_{2:3}$ plant individually, and analyzed for grain Cd content as described previously (section 3.3).

Genetic maps were constructed individually for each F_2 mapping population using the Haldane mapping function of JoinMap 4.0 (van Ooijen and Voorrips 2004) at a minimum LOD score of 3.0. ScOPC20 and several of the STS-HRM scored markers are dominant markers and therefore it was not possible to differentiate heterozygous and homozygous high grain Cd molecular variants. Therefore, the genotypes of individual recombinant $F_{2:3}$ families were used to determine the genetic state at the F_2 generation.

3.12 *HMA-B1* BAC Identification

The *Triticum turgidum* Ttu-B-LDN65 (Langdon) BAC library (Cenci et al 2003; 5X coverage) 2D pools, super pools and plate pools (2D pooling described at http://cnrgv.toulouse.inra.fr/en/services/dna_pool_production) were purchased from the French Plant Genomic Resource Centre (INRA-CNRGV). Positive plates were identified from super pools using primer pair TtHMA-n454-F3/R3 (F-5' ATGTCGTCGTTGATGAGCATG 3', and R-5' CCATTGTCCTCACGGCGATGT 3') and validated with plate pools. At the plate pool level, BAC plate pools were screened

to identify only 5B BACs. Langdon disomic subseries lines LDN5D(5A) and LDN5D(5B) were run as checks with the plate pools on SSCP (as described in Wiebe et al. 2010) using the TtHMA-n454-F3/R3 primer pair. One 5A and all 5B plates identified were also ordered from INRA-CNRGV and individual BACs were identified from plates using the same primer pair. Identified BACs were verified to localize to chromosome 5B or 5A using SSCP. The PCR reactions consisted of 50 mM KCl, 10 mM Tris-HCl, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.4 µM of forward and reverse primers, 1.75 U of Taq DNA polymerase and 100 ng of genomic DNA. PCR temperature cycling involves 95°C for 5 min followed by 35 cycles of 95°C for 30 s and 59°C for 30 s and 72°C followed by a final extension of 72°C for 10 min.

3.13 *HMA* Sequencing

For 454 BAC sequencing, BAC DNA was prepared by the method described by Poulsen (2004). Shotgun 454 sequence was obtained for BAC 532J14 using GS FLX Titanium chemistry with sequencing being performed at The Genome Quebec Innovation Center (Montreal, Quebec, Canada). 298B8 BAC 454 shotgun sequencing was performed using Roche 454 GS FLX Plus chemistry by Funomics (Saskatoon, Saskatchewan, Canada). Illumina GA paired-end sequence was also obtained for both BACs using a sequencing library size of 300 bp. BAC DNA for Illumina sequencing was prepared with the QIAGEN Large-construct kit according to the manufacturer's instructions (http://www.qiagen.com/products/plasmid/qiagenplasmidpurificationsystem/qiagenlargeconstructkit.aspx#Tabs=t2). Library preparation and sequencing was done at the National Research Council of Canada, Saskatoon, Sk. BAC end sequencing was performed using Sanger sequencing, with M13(-20) forward (GTAAAACGACGGCCAG) and M13 reverse (CAGGAAACAGCTATGAC) primers.

Hybrid assemblies incorporating both 454 and Illumina pair-end data were performed using Newbler version 2.6 (Roche). Illumina data was preprocessed as follows: reads matching the BAC cloning host sequence (*Escherichia coli* DH10B, NC_010473) or vector sequence (pIndigoBAC-5) were moved by bowtie (Langmead et al. 2009) and duplicate reads were removed using a Perl script obtained from Kevin Koh

(NRC, pers. comm.). Subsampling was done because of the high read depth obtained from the Illumina sequencing platform and increasing numbers of read pairs were sampled in increments of 10,000. A single 166,999 bp scaffold was obtained with 10,000 read pairs sampled for the 298B8 assembly (Appendix 6). BAC end sequencing confirmed the presence of both BAC ends. Similarly for 532J14, a single 124,137 bp scaffold was obtained with 20,000 read pairs sampled (Appendix 6); both the BAC-F and BAC-R ends were also found in the 532J14 scaffold, suggesting the assembly was complete.

Plant repeats were masked using RepeatMasker (Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-3.0, 1996-2010; http://www.repeatmasker.org) and the repeat masker specific libraries (RepBase Update 20110920) from the Genetic Information Research Institute (www.giri.org). The search program used by RepeatMasker was rmblastn version 2.2.23+ (http://www.repeatmasker.org/RMBlast.html). Coding regions were identified on repeat masked sequence by *ab initio* gene prediction using FGENESH (Softberry Inc.); cDNA sequence was used to guide the *ab initio* prediction when necessary. The predicted CDS of the *HMA3-B1* and *HMA3-A1* genes were aligned with similar genes from *Arabidopsis* (Baxter et al. 2003), *Brachypodium* (http://www.brachypodium.org/) and rice (Baxter et al. 2003) using ClustalX version 2.0.12 (Thompson et al. 1997). A bootstrap neighbor-join phylogenetic tree was generated using ClustalX. 1000 replicates were used for bootstrapping. The phylogenetic tree was viewed using Dendroscope (Huson et al, 2007).

The *HMA3* genes from durum cultivars Kofa (high grain Cd) and W9262-260D3 (low grain Cd) and Langdon BACs 298B8 (5B genome) and 532J14 (5A genome) were then sequenced from PCR amplicons. PCR amplicons of the approximate size were cloned using the TOPO TA Cloning Kit (Invitrogen, Burlington, Ontario, Canada) following the manufacturer's instructions and the plasmid DNA from clones carrying the inserted PCR product were sequenced at the National Research Council (NRC), Saskatoon, SK. Sequences were aligned using AlignX (Vector NTI Advance 10.3; Invitrogen, Carlsbad, CA). Primers for *HMA-B1* sequencing are listed in Appendix 3. The PCR reactions consisted of 50 mM KCl, 10 mM Tris-HCl, 1.5 mM $MgCl_2$, 0.2 mM

of each dNTP, 0.4 μM of forward and reverse primers, 1.75 U of Taq DNA polymerase and 100 ng of genomic DNA. PCR temperature cycling involves 95°C for 5 min followed by 35 cycles of 95°C for 30 s and 63°C for 30 s and 72°C followed by a final extension of 72°C for 10 min. Protein sequences were predicted for these genes using the ExPSAy translate tool (http://web.expasy.org/translate/) on the predicted CDS (Appendix 7) and functional domains were identified using the MyHits motif scan tool (http://myhits.isb-sib.ch/cgi-bin/motif_scan).

3.14 Marker Validation

A global collection of 96 durum wheat cultivars and breeding lines (Appendix 8) was used to validate five markers (ScOPC20, *Xusw47, Xusw17, Xusw15b,* and *HMA-B1)* for their potential for marker assisted selection (MAS) for the low grain Cd phenotype. This collection included breeding lines and cultivars from countries including Canada (25), Italy (17), USA (12), Australia (9), Spain (9), Argentina (5), Iran (4), France (3), Mexico (3), Morocco (3), Germany (2), New Zealand (2) and Russia (1), and has been described previously (Somers et al. 2007; Reimer et al. 2008).

Phenotypic data for grain Cd was collected from field trials conducted previously at Saskatoon, Saskatchewan over two years (2000 and 2001; Reimer et al. 2008). A single replicate of each line was grown in an alpha lattice design in both years (Reimer et al. 2008). Grain Cd concentration was determined using procedures described previously (Wiebe et al. 2010). Data analysis was performed using PROC MIXED of SAS (SAS Institute, Toronto, Canada) using each of the 2000 and 2001 environments as replications (random factor) and cultivars and breeding lines as a fixed factor. The least significant difference (LSD) was estimated using a significance level of p=0.05, and was used to classify the lines and cultivars discretely into high grain or low grain Cd accumulators.

4.0 RESULTS

4.1 Saturation Mapping of *Cdu-B1* in Durum Wheat

Cdu-B1 (Penner et al. 1995) was previously localized on chromosome 5BL (Knox et al. 2009). Using a DH population derived from the cross W9262-260D3 (low grain Cd)/Kofa (high grain Cd), Knox et al. (2009) localized *Cdu-B1* as a Mendelian factor approximately 3 cM distal to ScOPC20 (Knox et al. 2009; section 2.3.3, Fig. 2.3) and 12 cM distal to the *Tsn1* locus (markers *Xfcp1* and *Xfcp2*; Knox et al. 2009; section 2.3.3, Fig. 2.3). A 5BL map derived from the CS/CS-DIC 5B RSL hexaploid wheat population is available, and is saturated in the *Tsn1* region (Lu et al. 2006). Therefore, *Cdu-B1* was first localized in that population using *Tsn1* associated markers. ScOPC20 primers amplified the expected 394 bp fragment from CS-DIC 5B, but no fragment was amplified from CS (Fig. 4.1). *XBG608197* and *Xrz575* were mapped previously in the CS/CS-DIC 5B population (Lu et al. 2006), and ScOPC20 was found to co-segregate with these markers at a position 4.3 cM proximal to *Xwg644* (Fig. 4.2). In a previous study, grain Cd concentrations for CS and CS-DIC 5B was similar despite segregation at ScOPC20 (Wiebe et al. 2010), therefore it was not possible to localize *Cdu-B1* in this population. Regardless, the linkage of ScOPC20 with *Xwg644*, *XBG608197* and *Xrz575,* which were previously localized to deletion bin 5BL9 0.76-0.79 (Lu et al. 2006; Fig. 4.2), suggests *Cdu-B1* is also located within the same deletion bin.

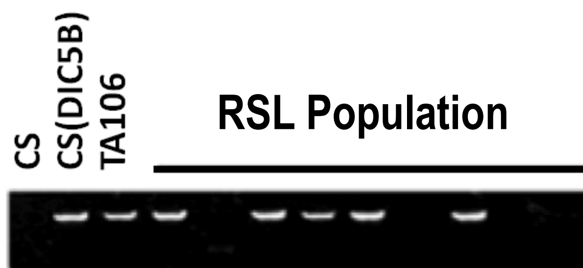

Fig. 4.1 Polymorphism detected at ScOPC20 in CS and CS-DIC 5B and a subsample of RSLs from the CS/CS-DIC 5B RSL.

It was not clear why no phenotypic variation in grain Cd concentration was observed in the CS/CS-DIC 5B RSL population, despite segregation at ScOPC20. There are at least two possibilities: a) there has been a recombination between

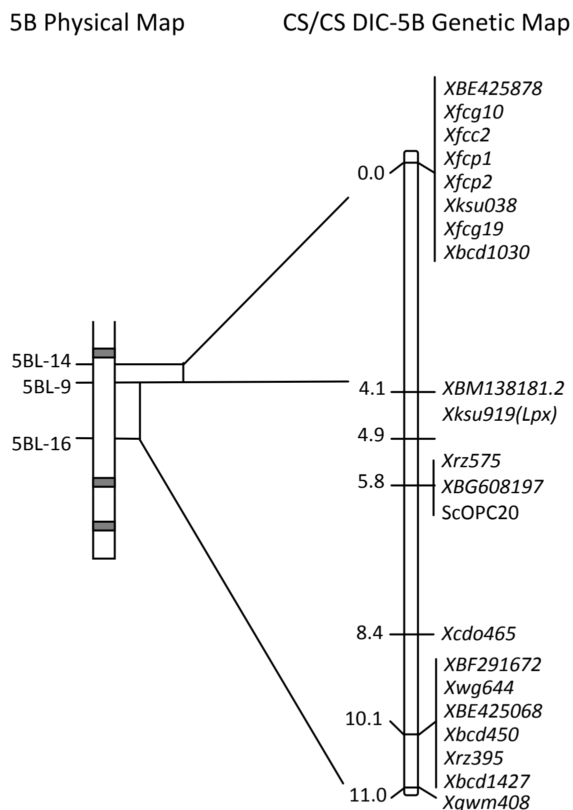

Fig. 4.2 Physical map of a portion of chromosome 5BL showing the deletion bins of that chromosome and corresponding genetic map of CS/CS-DIC 5B.

ScOPC20 and *Cdu-B1* in this population, or b) a gene homoeologous to *Cdu-B1* resides on chromosome 5D of hexaploid wheat. To test the latter hypothesis, grain Cd concentration of the Langdon-Chinese spring 5D disomic substitution lines LDN5D(5A) and LDN5D(5B) revealed that a substitution of chromosome 5D into Langdon resulted in a significant lowering of grain Cd concentration compared to Langdon (Fig. 4.3). These results confirm that a gene for reducing grain Cd is also present on chromosome 5D of hexaploid wheat.

Having established the likely physical location of *Cdu-B1*, 120 primer pairs were designed from the sequences of 54 wheat ESTs previously localized to bin 5BL9 0.76-0.79 (grain genes website http://wheat.pw.usda.gov/cgi-bin/westsql/bin_candidates.cgi?bin=5BL9-0.76-0.79). Twenty-five of these primers produced amplicons that were polymorphic between Kofa (high grain Cd) and W9262-260D3 (low grain Cd) and 13 of these (Appendix 1) were mapped to group five

chromosomes in the DH population. Marker *XBF474090.1* was polymorphic (Fig. 4.4) and co-segregated with *Cdu-B1* in the W9262-260D3/Kofa DH population (Fig. 4.5).

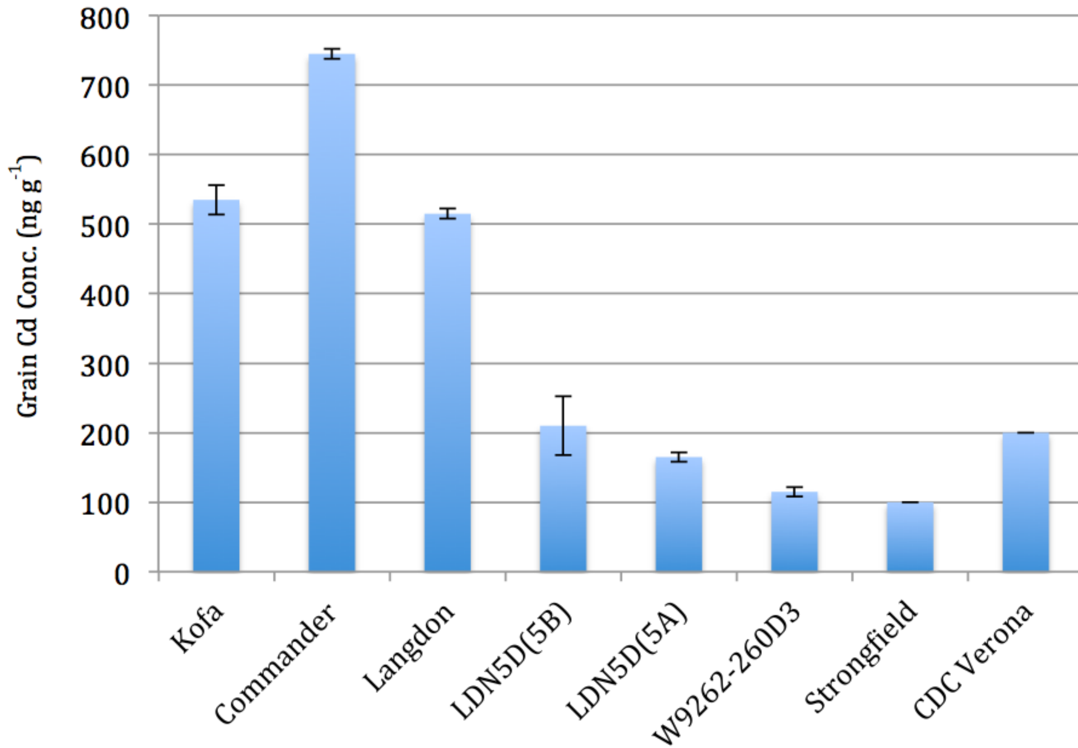

Fig. 4.3 Effects of substituting chromosome 5D into Langdon. Grain Cd concentration (ng g^{-1}) of known high grain Cd accumulators (Kofa, Commander, and Langdon) and known low grain Cd accumulators (W9262-260D3, Strongfield, and CDC Verona) and Langdon disomic substitution lines (LDN5D(5B), and LDN5D(5A)).

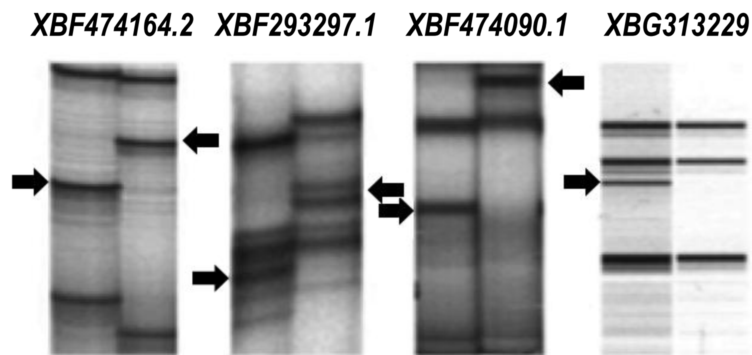

Fig. 4.4 Polymorphic ESMs associated with *Cdu-B1* in the W9262-260D3/Kofa population. For each marker, Kofa is lane 1 and W9262-260D3 is lane 2. All polymorphisms were detected with SSCP except *XBG313229* was detected with CE. Arrows indicate those polymorphic fragments that localized to 5B.

Primers designed from the wheat sequences of *XBF293297* and *XBF474164* each produced two polymorphic fragments (Fig. 4.4). *XBF293297.1* co-segregated with *Cdu-B1* and *XBF474164.2* mapped 0.2 cM distal to *Cdu-B1* (Fig. 4.5). *XBG313229* mapped 7 cM proximal to *Cdu-B1*. Primers for the ESMs *XBE604920*, *XBE426348*, *XBF474090.2, XBF145263, XBE494515*, *XBG262450*, *XBG274700, XBF474164.1,* and *XBF293297.2* were all polymorphic between Kofa and W9262-260D3, but all clustered on the distal region of chromosome 5AL when mapped in the DH population (Fig. 4.6).

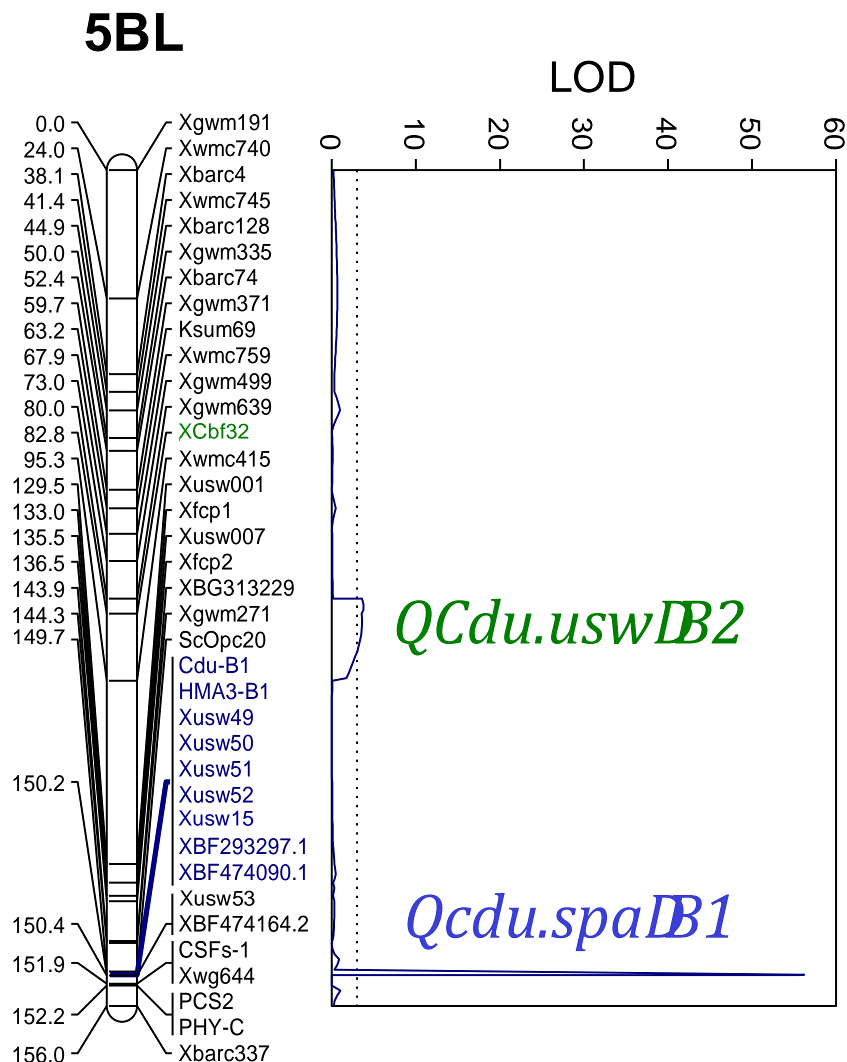

Fig. 4.5 Genetic map of a portion of chromosome 5BL in the W9262-260D3/Kofa population and *Cdu-B1* QTL analysis. Markers associated with *QCdu.spa-B1* are indicated by blue text. Marker in green text is associated with *QCdu.usw-B2*.

The sequences of *XBG608197*, *XBF293297.1*, *XBF474090.1*, and *XBF474164.2* were used to search against the available rice and *Brachypodium* genomes using BLASTx to identify the colinear regions (Fig. 4.7). All four ESM sequences showed similarity to genes localized to rice chromosome 3 and *Brachypodium* chromosome 1 except *XBF474090.1*, which localizes to *Brachypodium* chromosome 4. The durum wheat *Cdu-B1* region was colinear with a 286 Kbp region in rice, similar in size to the 282 Kbp region identified in *Brachypodium*. Eight *Brachypodium* genes were not present in the rice colinear region and 20 rice genes were absent in *Brachypodium*. Tandem repeats of the rice gene *Os03g53590* were identified in *Brachypodium*. The remaining twenty-nine genes showed near perfect gene order between rice and *Brachypodium* (Fig. 4.7). All genes from the *Cdu-B1* colinear regions in rice and *Brachypodium* and their predicted function are listed in Appendix 2. None of the genes have been associated with Cd uptake or sequestration in plants previously.

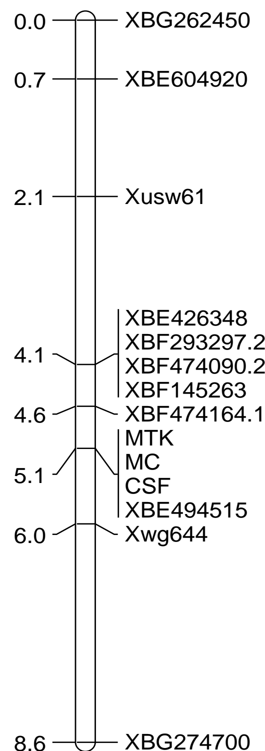

Fig. 4.6 Genetic map of a portion of chromosome 5AL in the W9262-260D3/Kofa population

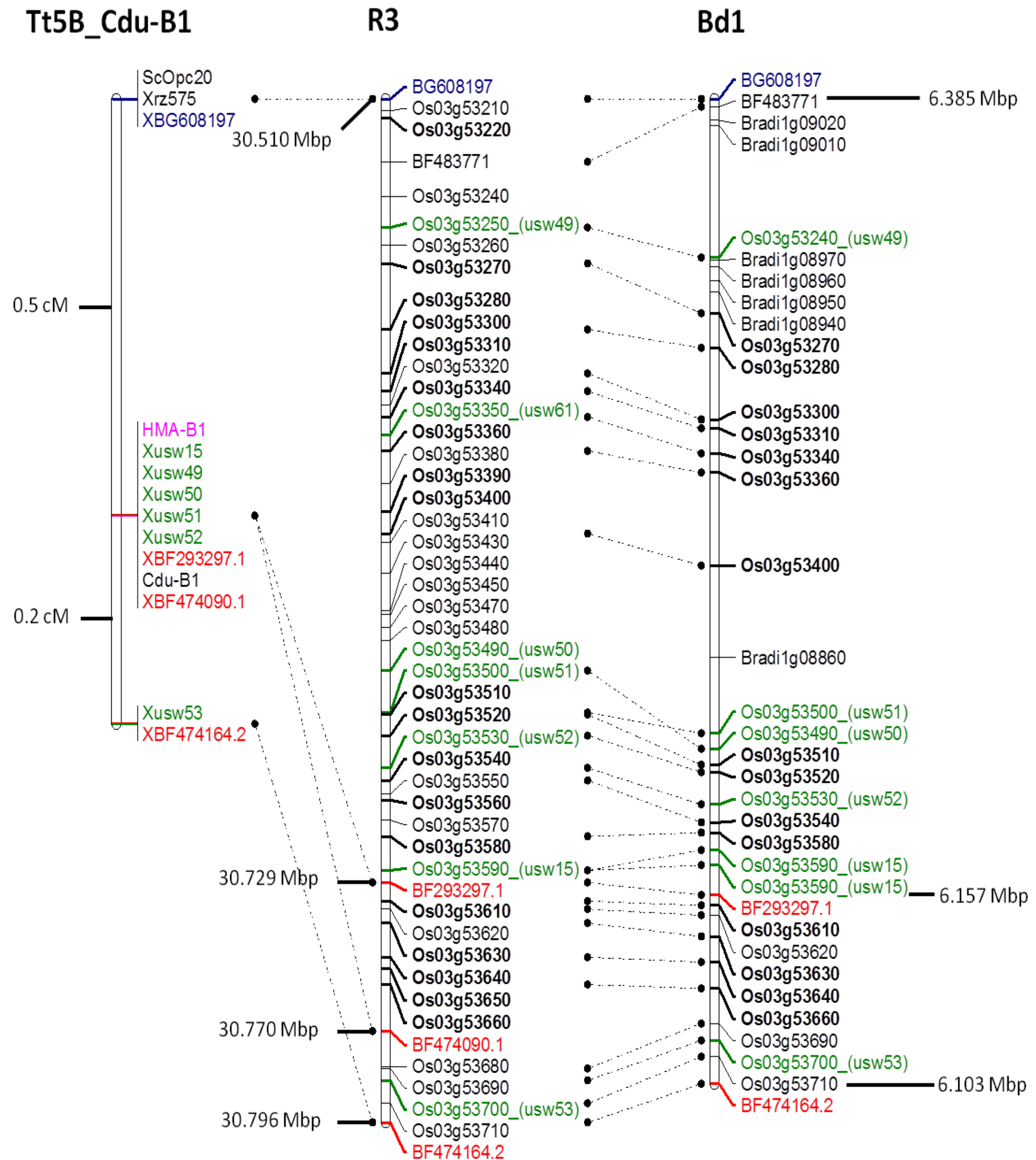

Fig. 4.7 Genetic map of the *Cdu-B1* region from W9262-260D3/Kofa (Tt5B_Cdu-B1) and its colinear physical region on rice chromosome three (R3), and *Brachypodium* chromosome 1 (Bd1). The positions of *Xrz575* and *XBG608197* (blue) relative to *Cdu-B1* were inferred from the CS/CS-DIC 5B genetic map (see Fig. 4.2). Red indicates ESMs that were used along with *XBG608197* to determine colinear regions in rice and *Brachypodium.* Green indicates those genes in which polymorphisms in W9262-260D3/Kofa were identified and mapped (STS markers). Bold indicates colinear genes, which were evaluated for polymorphisms in W9262-260D3/Kofa, but unable to map due to lack of polymorphisms.

To further saturate *Cdu-B1,* primers for 25 genes from the colinear region of rice were developed, and seven of these (*Os03g53590, Os03g53250, Os03g53490, Os03g53500, Os03g53530, Os03g53700,* and *Os03g53350*; Fig. 4.7) produced detectable polymorphisms between Kofa and W9262-260D3 (Fig. 4.8). These markers were designated as *Xusw15, Xusw49, Xusw50, Xusw51, Xusw52, Xusw53,* and *Xusw61*. These were added to the existing genetic map (Fig. 4.5). In the revised map, *Xusw15, Xusw49, Xusw50, Xusw51,* and *Xusw52* co-segregated with *Cdu-B1* and one (*Xusw53*) co-segregated with *XBF474164*.2 (Fig. 4.5). One mapped to chromosome 5A (*Xusw61*) (Fig. 4.6).

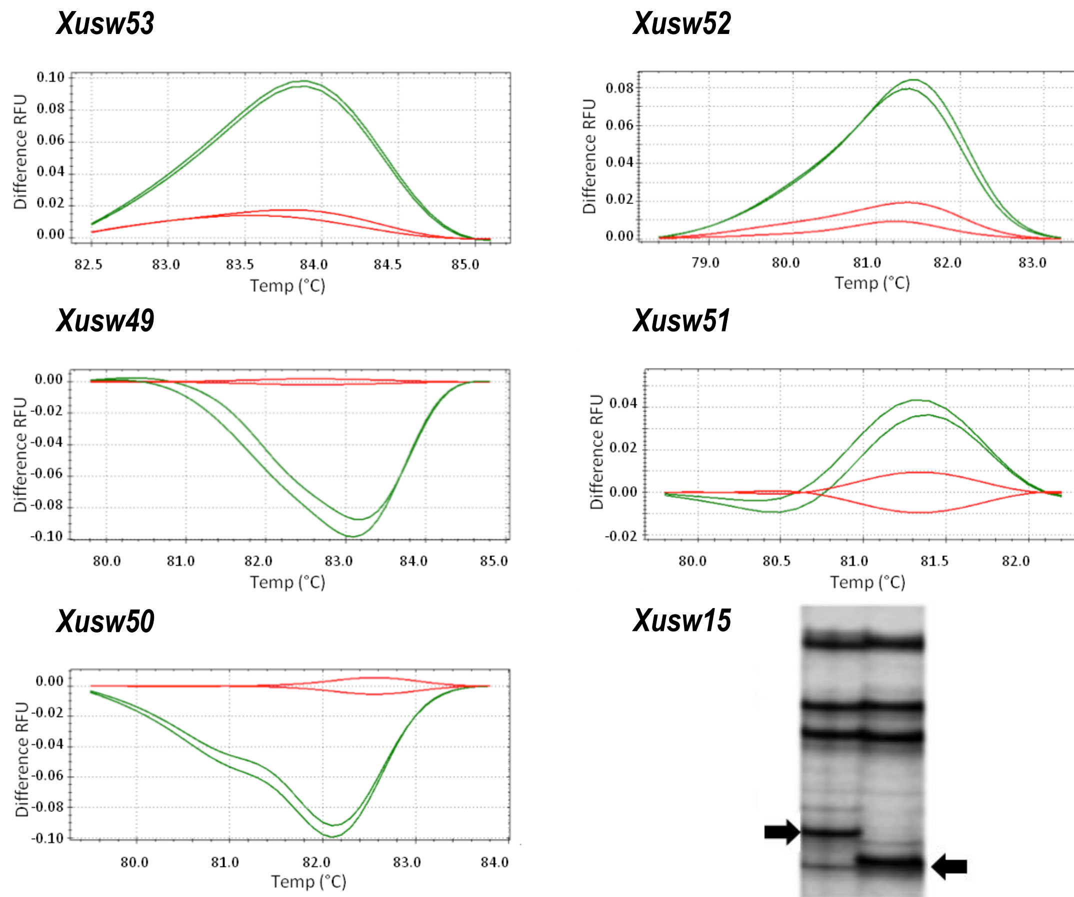

Fig. 4.8 Polymorphic STS markers associated with *Cdu-B1* in the W9262-260D3/Kofa population. Polymorphisms were detected with HRM (W9262-260D3 melt (green) in comparison to the Kofa melt (red)), except *Xusw15*, which was detected with SSCP. (Kofa is lane 1 and W9262-260D3 is lane 2). Arrows indicate those polymorphic fragments that localized to 5B.

To add additional markers to the *Cdu-B1* region, markers were identified from previously reported positional cloning studies in wheat that focused on chromosome 5BL. *Xwg644* mapped distal to ScOPC20 in the CS/CS-DIC 5B population (Fig. 4.2). That marker is linked to the *Vrn-A^{m}1* locus in *T. monococcum,* which has been sequenced (Yan et al. 2003). Thus, primers were designed for six genes physically linked to, and including, *Xwg644* in *T. monococcum*. The PCR products for two genes, *PHY-C* (phytochrome-C) and *PCS2* were polymorphic between W9262-260D3 and Kofa (Fig. 4.9), and mapped 1.8 cM distal to *Cdu-B1* in the DH mapping population (Fig. 4.5). The primers for *Xwg644* and *CSFs-1* (cleavage stimulation factor subunit 1) produced two polymorphic fragments each, and one from each gene (Fig. 4.9) mapped 1.5 cM distal to *Cdu-B1* (Fig. 4.5) while the second fragment mapped to chromosome 5AL (Fig. 4.6). Primers for *MTK4* (putative protein kinase tousled), *MC* (mitochondrial carrier protein) produced polymorphic amplicons, but these mapped to 5AL (Fig. 4.6).

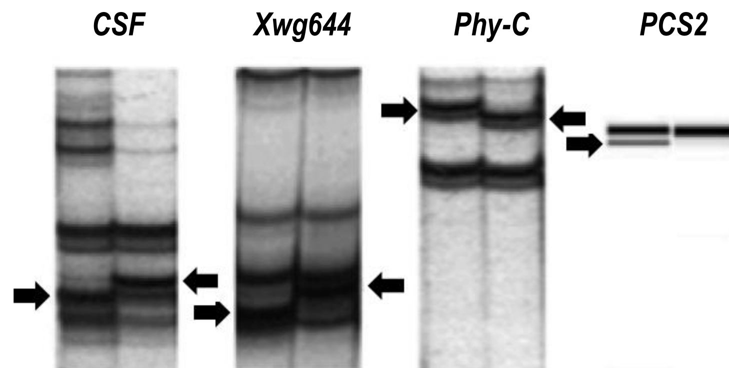

Fig. 4.9 Polymorphic gene specific markers in W9262-260D3/Kofa population designed from genes linked to VrnA^{m}1 in *T. monococcum*. Kofa is lane 1 and W9262-260D3 is lane 2. All polymorphism detected with SSCP except *PCS2* was detected with CE. Arrows indicate those polymorphic fragments that localized to 5B.

The revised genetic map of chromosome 5B from the W9262-260D3/Kofa DH population was used for QTL analysis of grain Cd concentration (Fig. 4.5) using data reported previously (Knox et al. 2009). A major QTL (LOD = 58) for grain Cd concentration centered at *XBF293297.1, XBF474090.1, Xusw15, Xusw49, Xusw50, Xusw51,* and *Xusw52* was identified and was flanked by ScOPC20 proximally and *XBF474164.2* and *Xusw53* distally (Fig. 4.5). This QTL, previously designated as *QCdu.spa-B1* (Knox et al. 2009), was reduced to a 0.7 cM interval (Fig. 4.5) and

explained 82% of the phenotypic variation in grain Cd concentration. W9262-260D3, the low Cd parent contributed the allele for low Cd with an additive effect of 47 ng g^{-1} (Table 4.1). Using MLM, a second minor QTL (LOD = 4.1) not previously reported by Knox et al. (2009) was also detected on 5BL around *XCbf32* (Fig. 4.5) and was designated as *QCdu.usw-B2*. The QTL effect was small relative to *QCdu.spa-B1,* and the low grain Cd parent W9262-260D3 contributed the allele for low grain Cd (Table

Table 4.1 Least square means (LSM) of grain Cd concentrations (ng g^{-1}) from two environments (Knox et al. 2009) for three markers associated with *Cdu-B1* and *XCbf32*

Molecular variants	**ScOPC20**	***HMA3-B1***	***XBF474164.2***	***XCbf32***
Least square means of genotypic groups				
Kofa (high)	160	157	160	121
W9262-260D3 (low)	71	67	72	101
Difference[a]	89**	90**	88**	20**
Effect of *XCbf32* in lines homozygous for low Cd uptake at *Cdu-B1*				
Kofa (high)	75	74	80	
W9262-260D3 (low)	63	58	63	
Difference[a]	12*	16**	17**	
Effect of *XCbf32* in lines homozygous for high Cd uptake at *Cdu-B1*				
Kofa (high)	166	165	165	
W9262-260D3 (low)	149	148	149	
Difference[a]	17**	17**	16*	

[a]Differences between genotypic classes were significant at * $P < 0.05$; ** significant at $P < 0.01$

4.1). The interactions between *XCbf32* and the three markers linked to *Cdu-B1* were not statistically significant, but compared to only *QCdu.spa-B1*, the combined effects of the W9262-260D3 molecular variants at *XCbf32* and *QCdu.spa-B1* reduced grain Cd by 17 ng g^{-1} (Table 4.1).

4.2 Fine Mapping of *Cdu-B1* in Durum Wheat

For ease of use in fine mapping, those markers that were mapped using SSCP (Fig. 4.4; Fig. 4.5; Fig. 4.8) were first converted to CAPS markers (Appendix 3 and 4), as SSCP is not conducive to high through-put genotyping of large populations typical of fine mapping populations. The PCR amplicons from *XBF474164.2*, *XBF474090.1*, *XBF293297.1*, and *Xusw15* were PCR cloned and sequenced from Kofa and W9262-260D3 to identify single nucleotide polymorphisms (SNPs) (Appendix 3). Appropriate restriction enzymes were identified to selectively cut either the Kofa or W9262-260D3 allele. Four new CAPS markers were developed (Fig. 4.10), and based on the recommended rules for gene symbolization in wheat, were designated as *Xusw14 (XBF474164.2), Xusw47 (XBF474090.1), Xusw17 (XBF273297.1),* and *Xusw15b (Xusw15)* (Table 4.2). The new CAPS markers genotyped the W9262-260D3/Kofa population the same as the SSCP markers from which they were developed (Fig. 4.11).

For fine mapping experiments, the flanking markers (ScOPC20, and *Xusw14*) were used first to identify recombinants within the *Cdu-B1* containing interval. *Xusw47* was also included because ScOPC20 is a dominant marker, which makes detection of recombination difficult, as it is impossible to identify heterozygotes with a dominant marker. Then, additional co-segregating markers were only scored on the recombinants.

In the Svevo/Brigade F_2 population, three recombinations were identified in the 521 individuals that were assayed (Table 4.3) and a genetic map was constructed (Fig. 4.11). Flanking markers (ScOPC20 and *Xusw14*) spanned an interval of 0.29 cM. Eight of the markers co-segregated and mapped 0.10 cM (one recombinant) distal to ScOPC20 (Fig. 4.11). Two recombinants (0.19 cM) separated the co-segregating markers from *Xusw14*.

The three recombinant plants identified were self-pollinated to generate $F_{2:3}$ families. A minimum of 16 individual plants were screened from each of the three

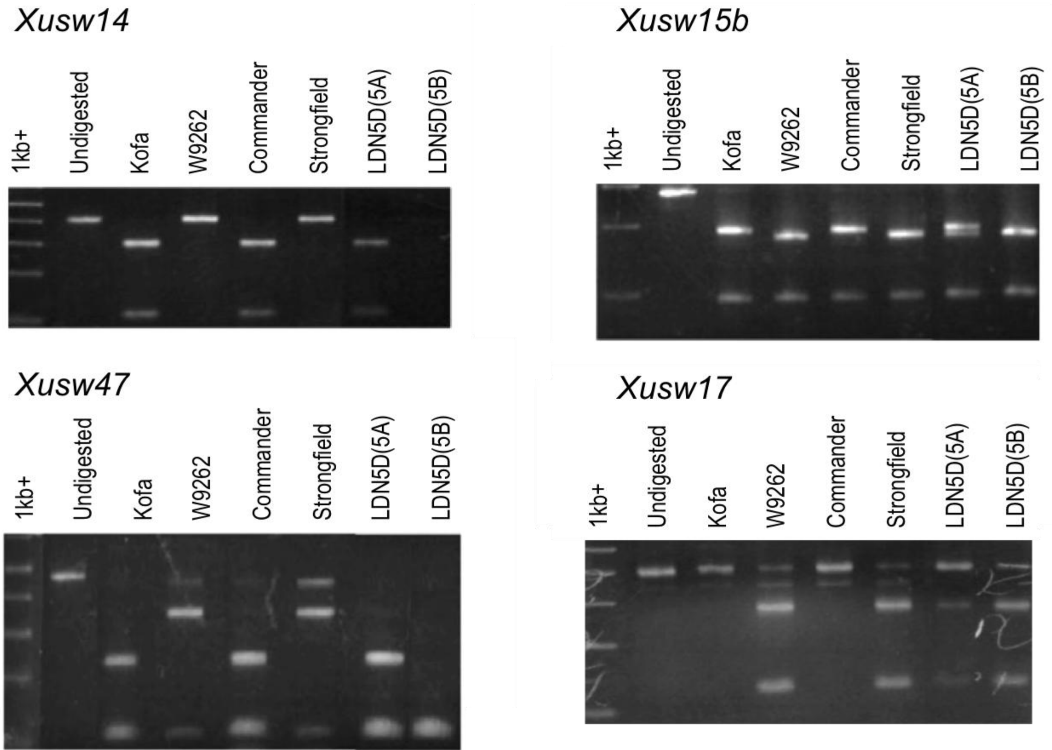

Fig. 4.10 CAPS markers developed from ESMs and STS marker. Validation of CAPS markers in Langdon group 5 disomic substitution lines and known high (Kofa, Commander) and low (W9262-260D3, Strongfield) grain Cd concentration cultivar.

Table 4.2 Marker designations for converted CAPS markers

ESM or STS marker	Converted CAPS marker
XBF474164.2	*Xusw14*
XBF474090.1	*Xusw47*
XBF293297.1	*Xusw17*
Xusw15	*Xusw15b*

families, and eight lines that were homozygous for the identified recombinations were identified and grouped into one of three haplotypes (Table 4.4). All eight lines, along with parents, were phenotyped for grain Cd concentration (Table 4.4). Brigade showed 48% less grain Cd than Svevo (Table 4.4) and would thus be classified as low grain Cd concentration because the high grain Cd concentration durum phenotype contains 45-60% more grain Cd when compared to low grain Cd concentration durum phenotype (Harris and Taylor 2004; Hart et al. 2006). Three lines with haplotype 3 carried the

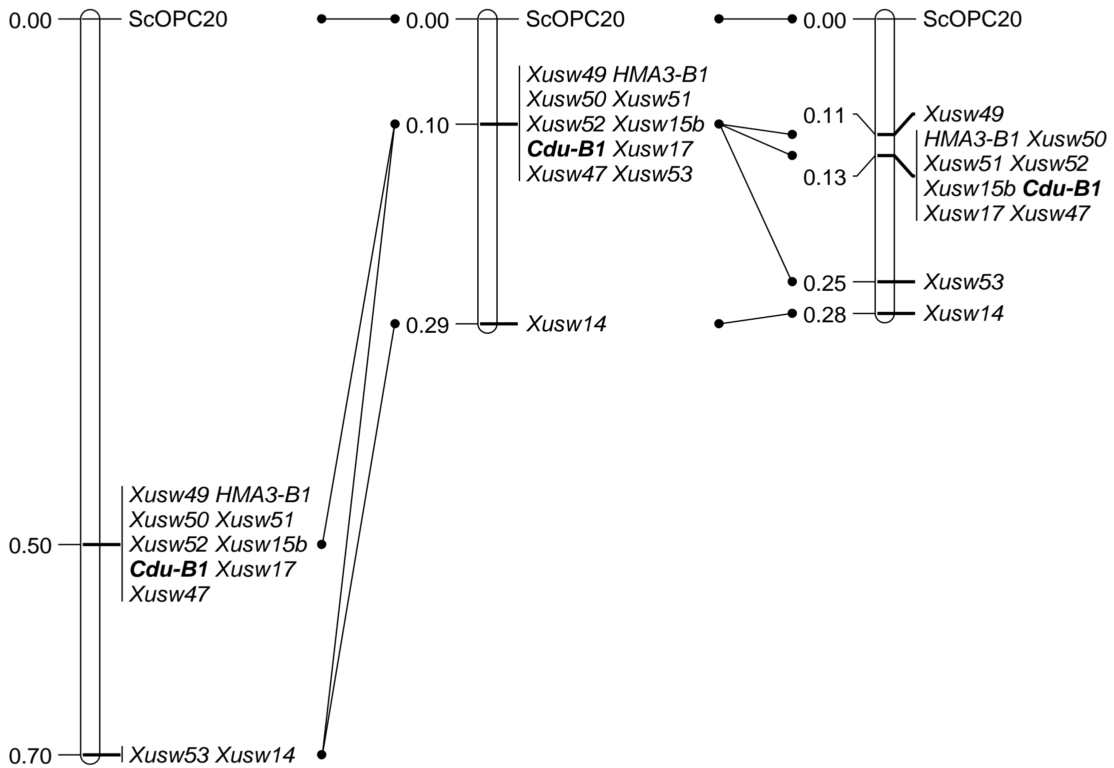

Fig. 4.11 Comparison of W9262-260D3/Kofa genetic saturation map, Svevo/Brigade genetic fine map, and 8982-TL genetic fine map.

Table 4.3 Haplotype of three recombinant F_2 plants from the Svevo/Brigade population (a = high Cd molecular variant, b = low Cd molecular variant, h = heterozygous).

F_2 Plant ID	***ScOPC20***	***Xusw49***	***Xusw50***	***Xusw51***	***Xusw52***	***Xusw15b***	***Xusw17***	***Xusw47***	***Xusw53***	***Xusw14***
D09.42-F2-037	a	a	a	a	a	a	a	a	a	h
D09.42-F2-115	b	h	h	h	h	h	h	h	h	h
D09.42-F2-384	h	h	h	h	h	h	h	h	h	a
Svevo	a	a	a	a	a	a	a	a	a	a
Brigade	b	b	b	b	b	b	b	b	b	b

Table 4.4 Haplotype groupings of homozygous $F_{2:3}$ lines derived from recombinant F_2 lines from the Svevo/Brigade population. The mean grain Cd concentration (ng g^{-1}), standard error of the means (SEM), and classification of each haplotype into high or low grain Cd concentration. The number of $F_{2:3}$ lines classified into each haplotype group are presented. a = high Cd molecular variant, b = low Cd molecular variant.

No. of homozygous $F_{2:3}$ Families	ScOPC20	*Xusw49*	*HMA3-B1*	*Xusw50*	*Xusw51*	*Xusw52*	*Xusw15b*	*Xusw17*	*Xusw47*	*Xusw53*	*Xusw14*	Grain Cd (ng g^{-1})	SEM (ng g^{-1})	Classification
Haplotype 1 (n=3)	a	**a**	**a**	**a**	**a**	**a**	**a**	**a**	**a**	**a**	b	630	15	**High**
Haplotype 2 (n=2)	b	**a**	**a**	**a**	**a**	**a**	**a**	**a**	**a**	**a**	a	665	75	**High**
Haplotype 3 (n=3)	b	**b**	**b**	**b**	**b**	**b**	**b**	**b**	**b**	**b**	a	300	15	**Low**
Svevo	a	**a**	**a**	**a**	**a**	**a**	**a**	**a**	**a**	**a**	a	500	21	**High**
Brigade	b	**b**	**b**	**b**	**b**	**b**	**b**	**b**	**b**	**b**	b	260	30	**Low**

Svevo (high) molecular variant at *Xusw14* and all three expressed low grain Cd like Brigade (Table 4.4). Haplotype 2 showed the Brigade (low) molecular variant at ScOPC20, but expressed high grain Cd (Table 4.4). Taken together, it was concluded that the *Cdu-B1* locus was flanked by ScOPC20 and *Xusw14*, and was associated with the eight co-segregating markers.

The results from the Svevo/Brigade F_2 population did not improve the genetic resolution at *Cdu-B1*, so a second, larger F_2 population was developed. For the second population (herein designated as 8982-TL population), two near isogenic lines (8982-TL-Low/8982-TL-High (Clarke et al. 1997a) for *Cdu-B1* were crossed to produce and an F_2 population of 3558 F_2 plants, which was used for mapping. *Cdu-B1* and *Vrn-B1* markers were screened on parents of the 8982-TL F_2 population. Results showed that the 8982-TL population is not segregating for *Vrn-B1* and therefore confirmed *Cdu-B1* is not associated with genes *PCS2* and *Xwg644* (Appendix 5). In total, 20 F_2 plants were identified with recombinations between ScOPC20 and *Xusw14* (Table 4.5). The remaining eight co-segregating markers were scored on the population, and the resulting data generated a final map which spanned 0.28 cM (Fig. 4.11). No recombination was detected between *Xusw50*, *Xusw51*, *Xusw52*, *Xusw15b*, *Xusw17*, and *Xusw47,* but a single recombination was detected between this cluster of markers and *Xusw49* (Table 4.5). Two recombinants were identified with between *Xusw53* and *Xusw14* (Table 4.5).

The 20 recombinant F_2 plants were self-pollinated, and individuals from the resulting $F_{2:3}$ families were screened to identify 20 lines homozygous for recombination and these were classified into one of seven haplotypes (Table 4.6). Four $F_{2:3}$ individuals homozygous for the low grain Cd concentration molecular variant (haplotype 2) at *Xusw53* and *Xusw14* expressed high grain Cd concentration like 8982-TL-H, the high Cd parent (Table 4.6). Lines with haplotype 1 expressed high grain Cd, but the low molecular variant at *Xusw14* (Table 4.6). Taken together, these data show that *Xusw53* flanks *Cdu-B1* distally. Two $F_{2:3}$ lines homozygous for the high molecular variant at ScOPC20 and *Xusw49* expressed grain Cd concentration similar to 8982-TL-L, the low Cd parent. Of all 20 $F_{2:3}$ individuals screened, only Haplotypes 3-6 expressed low grain Cd (Table 4.6). Taken together, *Cdu-B1* maps 0.02 cM distal to *Xusw49*, and 0.12 cM proximal to *Xusw53* in the 8982-TL population (Fig. 4.11).

Table 4.5 Haplotype of 20 recombinant F_2 plants from the 8982-TL fine mapping population (a = high Cd molecular variant, b = low Cd molecular variant, h = heterozygous).

F_2 Plant ID	*ScOPC20*	*Xusw49*	*Xusw50*	*Xusw51*	*Xusw52*	*Xusw15b*	*Xusw17*	*Xusw47*	*Xusw53*	*Xusw14*
8982-TL-L/H-F2-0874	a	a	a	a	a	a	a	a	a	h
8982-TL-L/H-F2-1153	a	a	a	a	a	a	a	a	h	h
8982-TL-L/H-F2-2306	a	a	a	a	a	a	a	a	h	h
8982-TL-L/H-F2-3006	a	a	a	a	a	a	a	a	h	h
8982-TL-L/H-F2-1451	b	b	b	b	b	b	b	b	h	h
8982-TL-L/H-F2-2332	h	b	b	b	b	b	b	b	b	b
8982-TL-L/H-F2-3365	h	b	b	b	b	b	b	b	b	b
8982-TL-L/H-F2-3397	h	b	b	b	b	b	b	b	b	b
8982-TL-L/H-F2-3860	h	b	b	b	b	b	b	b	b	b
8982-TL-L/H-F2-3823	h	h	b	b	b	b	b	b	b	b
8982-TL-L/H-F2-2124	h	h	h	h	h	h	h	h	h	a
8982-TL-L/H-F2-0565	h	h	h	h	h	h	h	h	a	a
8982-TL-L/H-F2-0774	h	h	h	h	h	h	h	h	a	a
8982-TL-L/H-F2-2063	h	h	h	h	h	h	h	h	a	a
8982-TL-L/H-F2-1912	h	h	h	h	h	h	h	h	b	b
8982-TL-L/H-F2-2619	h	h	h	h	h	h	h	h	b	b
8982-TL-L/H-F2-0633	b	h	h	h	h	h	h	h	h	h
8982-TL-L/H-F2-2394	b	h	h	h	h	h	h	h	h	h
8982-TL-L/H-F2-2466	b	h	h	h	h	h	h	h	h	h
8982-TL-L/H-F2-3975	b	h	h	h	h	h	h	h	h	h
8982-TL-L	b	b	b	b	b	b	b	b	b	b
8982-TL-H	a	a	a	a	a	a	a	a	a	a

Observed molecular variant frequencies in both F_2 fine populations differed from expected frequencies based on Chi squared (χ^2) tests (Table 4.7), indicating segregation distortion was present in both populations. Expected ratios for co-dominant markers in an F_2 population would be 1a:2h:1b (where "a" was the high Cd parent molecular variant; "b" was the low Cd parent molecular variant) and 3a:1b for dominant markers. In the Svevo/Brigade population a greater number of "b" molecular variants (molecular variant of the low Cd parent) were observed for all markers screened (Chi squared P value <0.0001) (Table 4.7). In contrast, segregation distortion was less pronounced but favored the "a" molecular variant (Chi square P value <0.05) in the 8982-TL population (Table 4.7).

Table 4.6 Haplotype groupings of homozygous $F_{2:3}$ lines derived from recombinant F_2 lines from the 8982-TL mapping population. The mean grain Cd concentration (ng g^{-1}), standard error of the means (SEM), and classification of each haplotype into high or low grain Cd concentration. The number of $F_{2:3}$ lines classified into each haplotype group are presented. a = high Cd molecular variant, b = low Cd molecular variant.

No. of homozygous $F_{2:3}$ Families	ScOPC20	*Xusw49*	*HMA3-B1*	*Xusw50*	*Xusw51*	*Xusw52*	*Xusw15b*	*Xusw17*	*Xusw47*	*Xusw53*	*Xusw14*	Grain Cd (ng g^{-1})	SEM (ng g^{-1})	Classification
Haplotype 1 (n=1)	a	a	**a**	**a**	**a**	**a**	**a**	**a**	**a**	a	b	680	-	**High**
Haplotype 2 (n=4)	a	a	**a**	**a**	**a**	**a**	**a**	**a**	**a**	b	b	623	34	**High**
Haplotype 3 (n=2)	a	a	**b**	**b**	**b**	**b**	**b**	**b**	**b**	b	b	220	10	**Low**
Haplotype 4 (n=4)	a	b	**b**	**b**	**b**	**b**	**b**	**b**	**b**	b	b	213	21	**Low**
Haplotype 5 (n=1)	b	b	**b**	**b**	**b**	**b**	**b**	**b**	**b**	b	a	230	-	**Low**
Haplotype 6 (n=4)	b	b	**b**	**b**	**b**	**b**	**b**	**b**	**b**	a	a	195	23	**Low**
Haplotype 7 (n=4)	b	a	**a**	**a**	**a**	**a**	**a**	**a**	**a**	a	a	595	89	**High**
8982-TL-H	a	a	**a**	**a**	**a**	**a**	**a**	**a**	**a**	a	a	690	98	**High**
8982-TL-L	b	b	**b**	**b**	**b**	**b**	**b**	**b**	**b**	b	b	198	23	**Low**

Table 4.7 Molecular variant frequencies and chi sq (χ^2) tests for expected segregation ratios for two F_2 populations. For each marker "a" represents the molecular variant for from the high Cd parent, and "b" represents for the low Cd parent. Heterozygous F_2 plants were scored as "h".

Locus	Ratio Tested[a]	Svevo x Brigade					8982-TL				
		a	h	b	χ^2	χ^2 P-value[b]	a	h	b	χ^2	χ^2 P-value[b]
ScOPC20	3:-:1	259	-	262	175	0.0001	2734	-	830	5.57	0.05
Xusw49	1:2:1	83	177	261	175.16	0.0001	954	1778	826	9.21	0.01
Xusw50	1:2:1	83	177	261	175.16	0.0001	954	1777	827	9.07	0.05
Xusw51	1:2:1	83	177	261	175.16	0.0001	954	1777	827	9.07	0.05
Xusw52	1:2:1	83	177	261	175.16	0.0001	954	1777	827	9.07	0.05
Xusw15b	1:2:1	83	177	261	175.16	0.0001	955	1777	827	9.21	0.01
Xusw17	1:2:1	83	177	261	175.16	0.0001	955	1777	827	9.21	0.01
Xusw47	1:2:1	83	177	261	175.16	0.0001	955	1780	828	9.06	0.05
Xusw53	1:2:1	83	177	261	175.16	0.0001	955	1779	829	8.92	0.05
Xusw14	1:2:1	83	177	261	175.16	0.0001	955	1779	830	8.78	0.05

[a]Expected ratios tested were a:h (heterozygous):b. For ScOPC20, it was not possible to identify heterozygotes as this marker is dominant in nature. Thus the expected ratio is a:-:b.
[b]Probability that deviations from expected ratio are due to chance alone.

4.3 Sequencing and Mapping of *HMA3* Orthologues in Durum Wheat

A recent report has implicated *OsHMA3* as a gene that is responsible for limiting root to shoot translocation of Cd, and reduced grain Cd concentration in rice grain (section 2.3.2). Several primers were designed from the sequence of *OsHMA3* to amplify the orthologous sequence in durum wheat, and one pair (TtHMA3-B2-F/R) produced polymorphic fragments between Kofa and W9262-260D3. The polymorphic fragment mapped to chromosomes 2B and was designated as *HMA3-B2* (Fig. 4.12). *HMA3-B2* was not significantly associated with grain Cd concentration in the W9262-260D3/Kofa mapping population.

Given that large HMA-protein families exist in most plants (Baxter et al. 2003), the *OsHMA3* CDS sequence were used to search against the current 5X shotgun sequence assembly of the cultivar "Chinese Spring" using BLASTn. In total 23 contigs were identified (e value < e^{-5}) and from these, 13 new primers were designed. Primer

amplicons for these were first screened in the LDN5D(5B) disomic substitution line to determine if any of the amplicons localized to chromosome 5B. Three amplicons were identified on chromosome 5B. (Fig. 4.13). However, none of these primers produced polymorphic fragments between Kofa and W9262-260D3 when assayed using SSCP (Fig. 4.13).

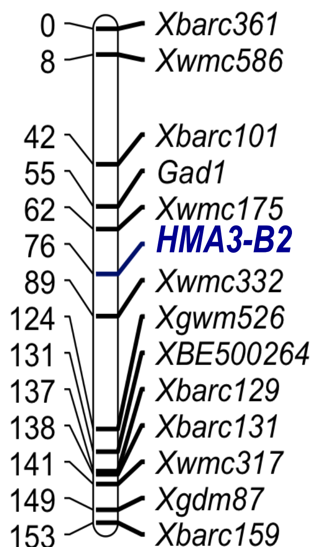

Fig. 4.12 Genetic position of *HMA3-B2* on a portion of chromosome 2B in the W9262-260D3/Kofa mapping population.

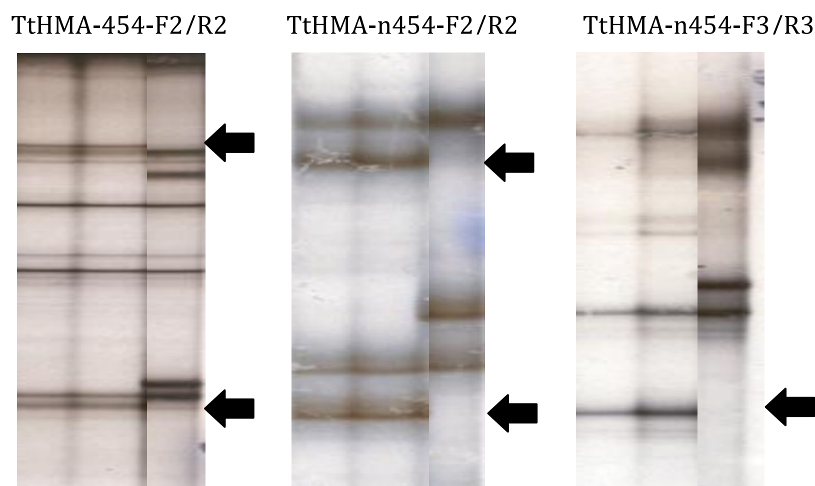

Fig. 4.13 *HMA3* primers localized to chromosome 5B in wheat. For all primers lane 1 is Kofa, lane 2 is W9262-260D3, and lane 3 is LDN5D(5B). Arrows indicate those amplicons that localize to 5B as they are absent in the substitution line. The additional amplicons present in the substitution line not present in Kofa or W9262-260D3 localize to chromosome 5D.

Given the association of *OsHMA3* in rice with variation in grain Cd concentration and the physical mapping of an orthologous gene to chromosome 5B of wheat, mapping of this gene in durum became a high priority. To obtain full length genomic DNA of *HMA3* from durum wheat, a BAC library derived from the cultivar Langdon (Cenci et al. 2003) was screened with primers TtHMA-n454-F3/R3. In total, six BACs were identified, five of which were mapped to chromosome 5B by comparing single strand conformational polymorphisms between individual BACs and LDN5D(5B) and LDN5D(5A) (Fig. 4.14). A single BAC localized to chromosome 5A (Fig. 4.14). One BAC was selected from chromosome 5A (532J14) and one from 5B (298B8) for sequencing using a combination of Roche 454 Titanium and Illumina platforms. Hybrid assemblies were generated for both BACs and both were assembled in their entirety into single scaffolds. 298B8 and 532J14 assembled into 166,999 bp and 124,137 bp respectively. Sequences of these BACs have been submitted to NCBI under the accession numbers JX454959 (298B8) and JX454960 (532J14).

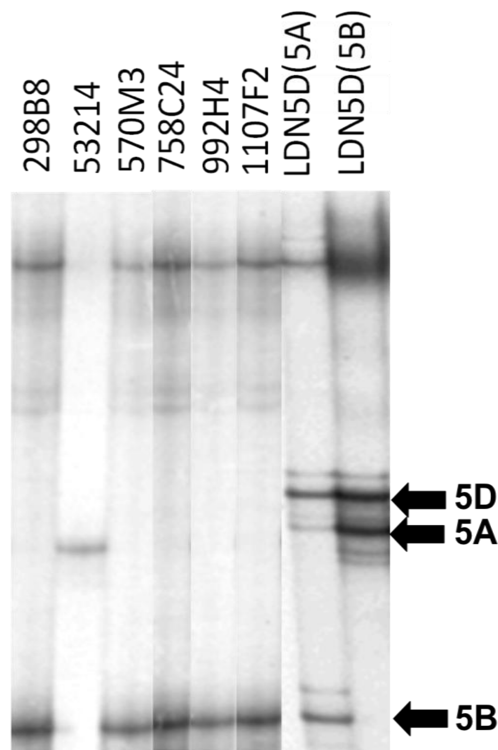

Fig. 4.14 Localization of *HMA3* BACs to chromosome 5A and 5B using SSCP.

FGENESH predicted full-length genomic sequences of *HMA3* from both 5A and 5B-derived BACs (Appendix 6), and these were designated as *HMA3-A1* and *HMA3-B1*, respectively. The genomic sequence of *HMA3-A1* was 2905 bp in length, coding for six predicted exons and was smaller than the 2978 bp sequence of *HMA3-B1*. The genomic sequences were 92% similar, but the Langdon *HMA-B1* could be easily differentiated by a 17 bp duplication in the first exon (Fig. 4.15a) and a 33 bp insertion/deletion (INDEL) in the last predicted exon (Fig. 4.15b). Predicted *HMA3-A1* and *HMA3-B1* coding sequencing showed 80% and 79% identity respectively with rice *OsHMA3* and 85% and 84% identity with *Brachypodium* orthologue *BdHMA3*.

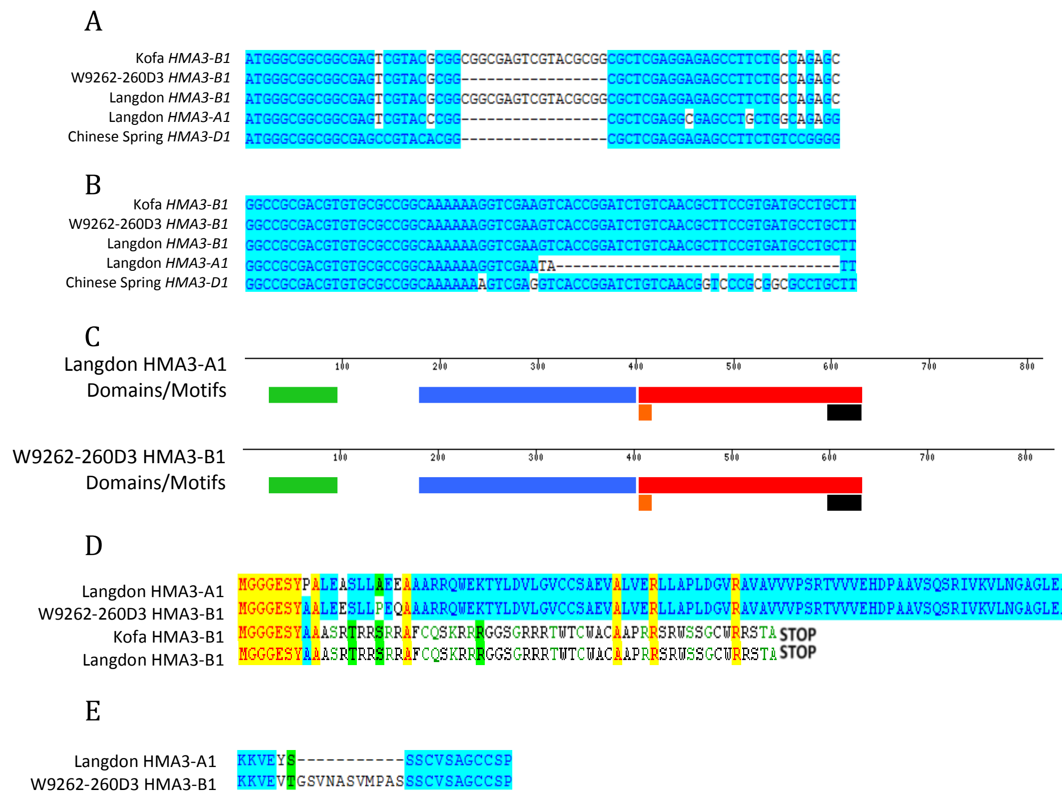

Fig. 4.15 *HMA3* sequencing a) Nucleotide sequence from the first 70 bp of wheat *HMA-A1* and *HMA3-B1* including 17 bp duplication b) nucleotide sequence of 70 bp from 3' end including 33bp INDEL c) conserved protein domains/motifs identified for HMA3-B1 from W9262-260D3 and HMA3-A1 from Langdon aligned against amino acid sequences (green – heavy metal associated domain; blue – ATPase domain; red – hydrolase domain; orange – P-ATPase; black – COF 2). d) Protein sequence from the first 90 amino acids of wheat HMA3-A1 and HMA3-B1 proteins including premature stop codon due to 17bp duplication. e) Protein sequence from the last 28 amino acids of wheat HMA3-A1 and HMA3-B1 including 11 amino acid deletion.

Partial sequence of *HMA3-D1,* the *D* genome homoeologue was obtained by using BLASTx against the chromosome 5DL survey sequence (http://urgi.versailles.inra.fr/Species/Wheat/Sequence-Repository) derived from the cultivar Chinese Spring. Two contigs were identified that, together, represent nearly a full-length sequence (Appendix 6). A small portion of the 3' sequence is missing (around 38 bp) as well as 224 bp after the 2463 bp. At the 5' end of the sequence, *HMA3-D1* lacked the 17bp duplication present in *HMA3-B1* present in Langdon, similar to *HMA3-A1*. In contrast, at the 3' end of the *HMA3-D1* sequence contained the 33 bp sequence absent in *HMA3-A1*, but present in the Langdon sequence of *HMA3-B1* (Fig. 4.15b).

To identify allelic variation at *HMA3-B1*, six primer sets were designed from the full length Langdon sequence of *HMA3-B1* to amplify overlapping, contiguous fragments to generate full length genomic sequence from Kofa and W9262-260D3 (Fig. 4.16). Alignment of *HMA3-B1* sequences revealed that the W9262-260D3 sequence was 98% similar to the *HMA3-B1* from BAC 298B8 (Appendix 6). In contrast, the sequence from Kofa was identical to the 298B8 sequence, and when compared to W9262-260D3, both Kofa and Langdon alleles contained a 17 bp duplication 27 bp from the start codon (Fig. 4.15a). Predicted protein sequence (Appendix 7) from the Kofa and Langdon *HMA3-B1* sequence revealed the 17 bp duplication results in a premature stop codon, and thus a severely truncated, non-functional protein.

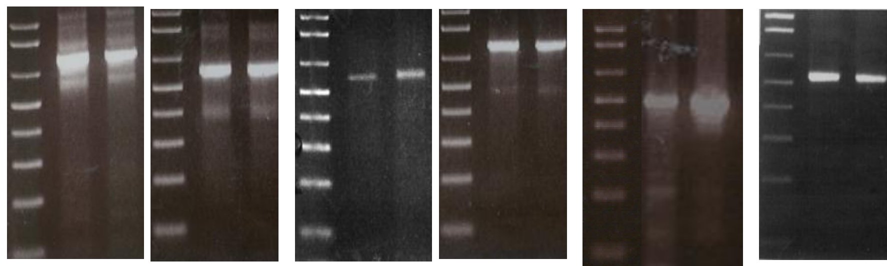

Fig. 4.16 Primer sets for sequencing *HMA3-B1* from Kofa (lane 1) and W9262-260D3 (lane 2). Primer sets from left to right HMA3-F1/R1, HMA3-F2/R2, HMA3-F3/R3, HMA3-F4/R4, HMA3-F5/R5, and HMA3-F6/R6. 1kb+ ladder to the left of each primer sets.

Predicted protein sequences for *HMA3-A1* and *HMA3-B1* genes were scanned for conserved protein domains and motifs (http://myhits.isb-sib.ch/cgi-bin/motif_scan). The Langdon HMA3-A1 and the W9262-260D3 HMA3-B1 predicted proteins contained three predicted conserved domains including a heavy metal associated domain, an ATPase domain, and a hydrolase domain, and two motifs including a P-ATPase motif, and COF 2 predicted hydrolase motif (Fig. 4.15c). No conserved protein domains were predicted from the Langdon and Kofa HMA3-B1 sequence as the predicted protein sequence contained a premature stop codon (Fig. 4.15d) and therefore a severely truncated protein. The Langdon HMA3-A1 predicted protein did not show any conserved domains or motifs in the C terminus, nor did the W9262-260D3 HMA3-B1 predicted protein. However, the Langdon HMA3-A1 predicted protein showed an eleven amino acid deletion when compared to HMA3-B1 from W9262-260D3 (Fig 4.15e).

Multiple *HMA* genes have been sequenced and characterized from rice, *Brachypodium*, and *Arabidopsis*. To determine the similarity between *HMA3-A1* and *HMA3-B1* and other *HMA* genes, a sequence similarity dendrogram was constructed using only CDS (Fig. 4.17). The *HMA* genes grouped into six clades (Fig. 4.17; Baxter et al. 2003; Williams and Mills 2005) and *HMA3-A1* and *HMA3-B1* belong to clade 2 which is characterized by a HMA domain with CCxx motif (in place of the CxxC motif which is thought to be involved in metal binding) in the N terminus and multiple CC motifs in the C terminus.

To determine if the 17 bp duplication identified between Kofa and W9262-260D3 was associated with variation in grain Cd concentration, primers flanking the 17 bp duplication were designed and evaluated for polymorphism. The primer HMA3-B1-F/R amplified the expected 297 bp fragment from Kofa, and a 280 bp fragment was amplified from W9262-260D3 (Fig. 4.18). The fragment was mapped in the W9262-260D3/Kofa mapping population (Fig. 4.18) and co-segregated with *Cdu-B1* (Fig. 4.5; Fig. 4.7; Fig. 4.11). The effect of the W9262-260D3 *HMA3-B1* allele was to reduced grain Cd by 90 ng g^{-1} in the DH mapping population (Table 4.1). The primer was also tested in the parents of the two fine mapping populations, and Svevo and 8982-TL-H (both high Cd accumulators) produced a banding pattern like Kofa, and the two low

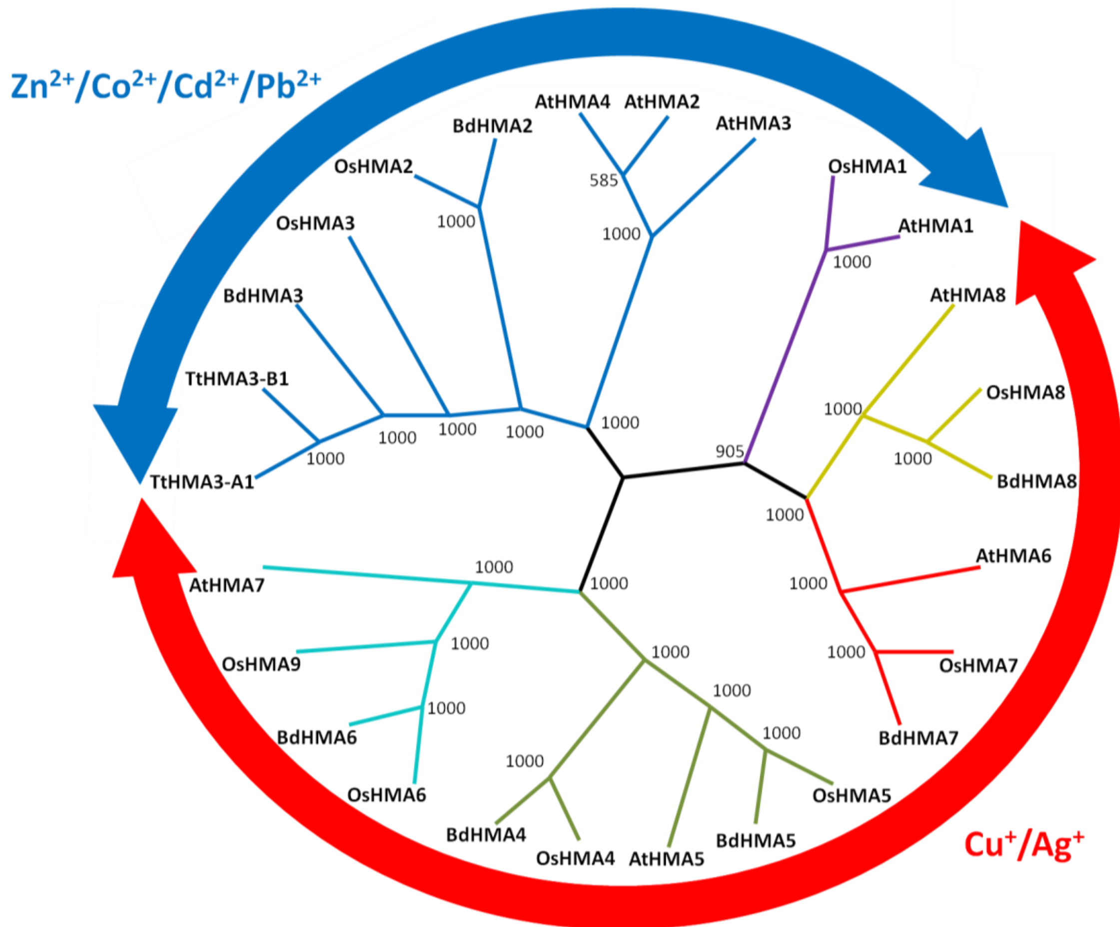

Fig. 4.17 *HMA* dendrogram. Different color lines represent different *HMA* groups (Baxter et al. 2003) (purple = group 1, blue = group 2, green = group 3, aqua = group 4, red = group 5, and orange = group6). Arrows indicate putative function of *HMA* (blue = $Zn^{2+}/Co^{2+}/Cd^{2+}/Pb^{2+}$ metal transporters, and red = Cu^{+}/Ag^{+} metal transporters). 1000 replicates used for bootstrapping.

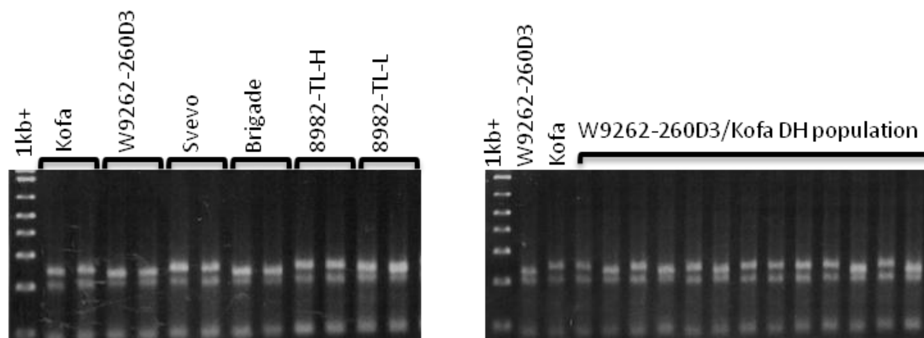

Fig. 4.18 *HMA3-B1* 17 bp duplication in parents of populations as well as a portion of the W9262-260D3/Kofa DH population.

accumulating parents Brigade and 8982-TL produced amplicons identical in size to W9262-260D3 (Fig. 4.18). HMA3-B1-F/R was scored on the recombinant $F_{2:3}$ families (from the two fine mapping populations (Tables 4.5 and 4.7)), and in both cases, co-segregated with *Cdu-B1* (Fig. 4.11).

HMA3-B1 co-segregated with several genes that were colinear on rice chromosome 3 and *Brachypodium* chromosome 1 (Fig 4.7). However, no sequences similar to *HMA3-B1* were identified in these colinear regions. *HMA3-B1* was most similar to *OsHMA3* and *BdHMA3*, which localize to chromosomes 7 and 1, respectively. These results suggest a break in colinearity between wheat-rice-*Brachypodium* in the *Cdu-B1* interval.

4.4. Validation of Molecular Markers Linked to *Cdu-B1* in Durum Wheat

A diverse set of 96 durum wheat cultivars and breeding lines was used to determine if the 17 bp duplication identified from sequencing experiments of *HMA3-B1* was associated with variation in grain Cd. Grain Cd concentration was collected for all 96 lines over two years. Averaged over both years, grain Cd concentrations ranged from 59 ng g^{-1} to 296 ng g^{-1} (Table 4.8). The majority of cultivars and breeding lines could be classified into either high (>136 ng g^{-1}) or low (<121 ng g^{-1}) grain Cd accumulators. The majority of accessions (66.7%) expressed high grain Cd with an average concentration of 203 ng g^{-1}. In contrast, the average concentration of low accumulators over two years of testing was 85 ng g^{-1}. Consistent with previous results (Knox et al. 2009), Kofa expressed high grain Cd concentration (276 ng g^{-1}) and was similar to Kronos (296 ng g^{-1}) and Westbred 881 (254 ng g^{-1}). Commander (292 ng g^{-1}), and AC Navigator (Clarke et al. 2001) (296 ng g^{-1}) are derived from Westbred 881 and both expressed high grain Cd. CDC Verona (90 ng g^{-1}), Strongfield (88 ng g^{-1}) and Napoleon (109 ng g^{-1}) are low Cd accumulating varieties, and all three expressed grain Cd concentrations 167 ng g^{-1} less than Kofa. Within each of the high and low groupings, variation in grain Cd was still evident. Borli (61 ng g^{-1}), Camacho (63 ng g^{-1}), and Buck Topacio (78 ng g^{-1}) expressed significantly lower ($P<0.05$) grain Cd than Strongfield and CDC Verona. Langdon(DIC-6B) (Joppa and Cantrell 1990) (264 ng g^{-1}), which carries the high protein gene *Gpc-B1* (Chee et al. 2001) was

Table 4.8 Haplotyping of Durum cultivars and breeding lines using CAPS markers, *ScOPC20* and *HMA3-B1* marker. Cd concentration in ng g^{-1}. Molecular variant "a" is "Kofa-like" and "high Cd", "b" is "W9262-260D3-like" and "low Cd, and "H" is heterogenous.

High Grain Cd Accumulators

Cultivar	Origin	Cd Concentration	Classification	*ScOPC20*	*Xusw17*	*Xusw47*	*Xusw15b*	*HMA-B1*
Avonlea	CDN	214	H	H	a	a	a	a
AC Pathfinder	CDN	223	H	a	a	a	a	a
Morse	CDN	187	H	a	a	a	a	a
AC Navigator	CDN	296	H	a	a	a	a	a
Kyle	CDN	215	H	a	a	a	a	a
Commander	CDN	292	H	a	a	a	a	a
D24-1773	CDN	212	H	a	a	a	a	a
DT513	CDN	181	H	a	a	a	a	a
DT532	CDN	193	H	a	a	a	a	a
DT536	CDN	165	H	a	a	a	a	a
44616	Iran	216	H	a	a	a	a	a
BICumenary	Argentina	249	H	a	a	a	a	a
Bquilaco	Argentina	187	H	a	a	a	a	a
Bvalverde	Argentina	213	H	a	a	a	a	a
BuckAmbar	Argentina	226	H	a	a	a	a	a
920334	Australia	240	H	a	a	a	a	a
940030	Australia	242	H	a	a	a	a	a
940435	Australia	267	H	a	a	a	a	a
950090	Australia	183	H	a	a	a	a	a
950844	Australia	193	H	a	a	a	a	a
940955	Australia	211	H	a	a	a	a	a
Carioca	France	203	H	a	a	a	a	a
Simeto	Italy	171	H	a	a	a	a	a
Tetradur	France	257	H	a	a	a	a	a
Wollaroi	Australia	229	H	a	a	a	a	a
Durabon	Germany	165	H	a	a	a	a	a
Durafit	Germany	141	H	a	a	a	a	a
Colosseo	Italy	187	H	a	a	a	a	a
Duilio	Italy	161	H	a	a	a	a	a
Grazia	Italy	205	H	a	a	a	a	a
Fortore	Italy	181	H	a	a	a	a	a
Lesina	Italy	194	H	a	a	a	a	a
Varano	Italy	173	H	a	a	a	a	a

Cultivar	Origin	Cd Concentration	Classification	*ScOPC20*	*Xusw17*	*Xusw47*	*Xusw15b*	*HMA-B1*
Nacori97	Mexico	225	H	a	a	a	a	a
DHTON1	Morrocco	219	H	a	a	a	a	a
Gidara17a	Morroco	170	H	a	a	a	a	a
D940098	ND-USDA	181	H	a	a	a	a	a
D941038	ND-USDA	223	H	a	a	a	a	a
D95580	ND-USDA	205	H	a	a	a	a	a
Plaza	ND-USDA	248	H	a	a	a	a	a
Durex	U.S.	214	H	a	a	a	a	a
Langdon	U.S.	159	H	a	a	a	a	a
Langdon(DIC-6B)	U.S.	264	H	a	a	a	a	a
Westbred881	U.S.	254	H	a	a	a	a	a
Kofa	U.S.	276	H	a	a	a	a	a
Kronos	U.S.	296	H	a	a	a	a	a
Arrivato	New Zealand	217	H	a	a	a	a	a
K-39099	Russia	136	H	a	a	a	a	a
44721	Iran	208	H	a	a	a	a	a
Ciccio	Italy	201	H	a	a	a	a	a
CRDW17	Iran	161	H	a	a	a	a	a
D-73-15	Iran	182	H	a	a	a	a	a
Medda	Italy	123	H	a	a	a	a	a
Mongibello	Italy	242	H	a	a	a	a	a
Parsifal	Italy	156	H	a	a	a	a	a
Svevo	Italy	159	H	H	H	H	H	H
Tresor	Italy	193	H	a	a	a	a	a
Agridur	Spain	176	H	a	a	a	a	a
Ariesol	Spain	268	H	a	a	a	a	a
Mexa	Spain	263	H	a	a	a	a	a
Vitron	Spain	191	H	a	a	a	a	a
Ocotillo	Arizona	217	H	a	a	a	a	a
Average		**203**						
Minimum		**136**						
Maximum		**296**						

Low Grain Cd Accumulators

Cultivar	Origin	Cd Concentration	Classification	*ScOPC20*	*Xusw17*	*Xusw47*	*Xusw15b*	*HMA-B1*
9661-AF1D	CDN	60	L	b	b	b	b	b
9661-CA5E	CDN	89	L	b	b	b	b	b
CDC Verona	CDN	90	L	b	b	b	b	b
DT691	CDN	90	L	b	b	b	b	b
DT695	CDN	98	L	b	b	b	b	b
DT696	CDN	121	L	H	H	H	H	H
DT704	CDN	111	L	b	b	b	b	b
DT705	CDN	103	L	b	b	b	b	b
DT707	CDN	82	L	b	b	b	b	b
DT709	CDN	109	L	b	b	b	b	b
DT710	CDN	95	L	b	b	b	b	b
DT711	CDN	108	L	b	b	b	b	b
Napoleon	CDN	109	L	b	b	b	b	b
Strongfield	CDN	88	L	b	b	b	b	b
Altar-Aos	Spain	65	L	b	b	b	b	b
Arcobelano	Spain	85	L	b	b	b	b	b
950329	Australia	124	L	a	H	H	H	H
Tamario	Australia	121	H	b	b	b	b	b

Cultivar	Origin	Cd Concentration	Classification	*ScOPC20*	*Xusw17*	*Xusw47*	*Xusw15b*	*HMA-B1*
Borli	Spain	61	L	b	b	b	b	b
BuckTopacio	Argentina	78	L	b	b	b	b	b
Camacho	Spain	63	L	b	b	b	b	b
CFR5001	New Zealand	86	L	b	b	b	b	b
Bronte	Italy	80	L	b	b	b	b	b
Demetra	Italy	90	L	a	b	b	b	b
Flavio	Italy	59	L	b	b	b	b	b
Gallereta	Spain	86	L	b	b	b	b	b
Gianna	Italy	69	L	b	b	b	b	b
Green27	Mexico	98	L	b	b	b	b	b
Green34	Mexico	85	L	b	b	b	b	b
Iride	Italy	73	L	b	b	b	b	b
Marjak	Morrocco	82	L	a	b	b	H	b
RABD93.40	France	72	L	b	b	b	b	b
Average		**85**						
Minimum		**59**						
Maximum		**124**						
LSD		**42**						

classified as high grain Cd; higher than its isogenic parent Langdon (159 ng g^{-1}). Molecular variation at *Cdu-B1* was assessed in the 96 cultivars using five DNA markers (Table 4.8). ScOPC20 correctly classified all high accumulating cultivars. However, the cultivars 950329, Demetra, and Marjak all carried the high Cd molecular variant at ScOPC20, but all three expressed low grain Cd concentrations (Table 4.8). Molecular variants at *Xusw17*, *Xusw47* and *Xusw15b*, which co-segregated with *Cdu-B1* in both fine mapping populations (Fig. 4.11), agreed with the phenotypic classification into high and low Cd accumulators (Table 4.8), except for three (four in *Xusw15b*) lines. For all five markers, Svevo, DT696 and 950329 were classified as heterogeneous (containing both high and low molecular variants). DT696 and 950329 did express higher grain Cd concentrations than the average of the low Cd grouping. Svevo grain Cd concentration was 156 ng g^{-1}, and was statistically lower than the majority of high accumulators (Table 4.8). The 17 bp duplication in the Kofa and Langdon allele of *HMA3-B1* was present in all high Cd accumulators and was absent in all low lines (except heterogeneous lines) (Table 4.8), confirming the association of this duplication with variation in grain Cd concentration.

5.0 DISCUSSION

5.1 Precise Genetic Localization of *Cdu-B1* in Durum Wheat, a Forward Genetic Study

Large genetic variation for grain Cd concentration exists in durum wheat (Table 4.8) and breeding for low concentration of grain Cd is a target of breeding programs globally. To date, the genetic factor(s) associated with phenotypic variation in grain Cd concentration have not been identified in durum wheat. Gene discovery in wheat is slow because of the large genome size, repetitive nature of the genome, and the polyploid nature of the wheat genome (Li et al. 2004). As well, a reference genome sequence for wheat is not available as a source to generate polymorphic DNA markers that are required for high density and fine mapping experiments. Currently, an international effort to sequence the wheat genome is ongoing (http://www.wheatgenome.org), but it is not likely to be available for several years. However, several genes have been cloned in wheat using colinearity between wheat, rice and *Brachypodium* to develop DNA markers for the high density and fine mapping experiments required for positional cloning (section 2.5). A similar approach was used in this project to develop a high density map of *Cdu-B1* in durum wheat.

A first step in isolating the genetic factor(s) associated with *Cdu-B1* was to develop a dense genetic map of *Cdu-B1*. *Cdu-B1* was mapped previously as a Mendelian factor to chromosome 5BL near ScOPC20 (Knox et al. 2009). Here, ScOPC20 was localized in the CS/CS-DIC 5B population as dense genetic map exists for that population (Lu et al. 2006). Results from this study showed that ScOPC20 localizes to deletion bin 5BL9 0.76-0.79, suggesting that *Cdu-B1* also resides in the same bin. As such, available ESTs previously mapped to this bin were converted to ESMs and mapped relative to *Cdu-B1* (Appendix 1). Two ESMs (*XBF293297.1* and *XBF474090.1*) were identified to be associated with *Cdu-B1* (Fig. 4.5). Using these mapped ESMs, the region of *Cdu-B1* was found to be colinear with a 286 Kbp region of rice chromosome 3 and a 282 Kbp region of *Brachypodium* chromosome 1. However, no obvious genes that code for known metal transporters or plant metal chelators were identified in either rice or *Brachypodium* in these regions (Appendix 2). There were,

however, several breaks in microcolinearity observed between rice and *Brachypodium*. Indeed the *Xwg644* locus consists of two tandem genes coding for independent half-sized ABC transporters (Ramakrishna et al. 2002), but only a single copy exists in rice and *Brachypodium*. Also, the ESM marker derived from *XBF474090.1* localized to chromosome four in *Brachypodium.* Eight genes were identified in the *Brachypodium* colinear region not identified in the rice colinear region and conversely 20 rice genes were absent in *Brachypodium* (Fig. 4.7). However, it is not reasonable to expect perfect colinearity between wheat, rice, and *Brachypodium* because multiple breaks in microcolinearity due to inversions, deletions, duplications and other rearrangements have been reported (Bennetzen 2000; Feuillet and Keller 2002; Li and Gill 2002; Sorrells et al. 2003; Francki et al. 2004; Lagudah et al. 2006; Lu and Faris 2006; Valárik et al. 2006; Bossolini et al. 2007; Faris et al. 2008).

Despite the break in colinearity, the rice/*Brachypodium* colinear region was used to develop an additional five STS markers (*Xusw15*, *Xusw49, Xusw50, Xusw51,* and *Xusw52)* that were also associated with *Cdu-B1*. Quantitative trait locus (QTL) analysis using the W9262-260D3/Kofa DH population and data collected previously (Knox et al. 2009) confirmed that these seven markers were strongly associated with grain Cd concentration, explaining greater than 80% of the observed phenotypic variation (Table 4.1). The additional markers reduced the *Cdu-B1* interval to 0.7 cM (Fig. 4.5) and represents a significant step towards positional cloning of *Cdu-B1* and the development of DNA markers to select for reduced grain Cd concentration (section 5.4 below).

Two fine mapping populations were developed and were used to localize *Cdu-B1* to an interval with a genetic distance of less than 0.29 cM on chromosome 5BL of durum wheat (Fig. 4.11). The high density of markers reported here, coupled with the reduced genetic distance could allow for positional cloning of *Cdu-B1*. However, the success of positional cloning is not only a function of marker density, but the physical: genetic distance ratio, where a small physical distance per cM is desirable to minimize the extent of chromosome walking. Recombination frequency near *Tsn1*, which maps approximately 7 cM proximal to *Cdu-B1* (Knox et al. 2009) was initially estimated at 400 kb/cM (Faris et al. 2000). Higher resolution mapping in later studies revealed a physical to genomic distance closer to 3.18 Mb/cM (Faris et al. 2010). In the 8982-TL

population, *Cdu-B1* spanned 0.14 cM (Fig. 4.11) and if the *Tsn1* locus physical to genomic distance is indicative of the physical to genomic distance at *Cdu-B1*, the physical distance between flanking genetic markers would be estimated at approx. 446 Kb. The average size of BACs from the Langdon durum wheat library is 130 Kb (Cenci et al. 2003), and thus ten BACs would span that physical distance (depending on the extent of BAC overlap). Although the physical to genomic distance at *Tsn1* is a reasonable estimate for the *Cdu-B1* interval (given its proximity to *Cdu-B1*), recombination frequencies are variable throughout the genome and even genetically linked loci can have large physical to genetic distance ratios. Generally, large physical to genetic distances are the result of recombination suppression. This was observed at the powdery mildew resistance locus in barley (*Mla*). This study showed *Mla* to repress recombination with a physical to genetic distance of 5.6 Mb/cM while just proximal the physical to genetic distance ratio was calculated at 140 kb/cM (Wei et al 2002). Physical to genetic distance also vary in wheat: approximately 1.25 Mb/cM for *Gpc-B1* (Distelfeld et al. 2006), 0.33 Mb/cM for the domestication locus (Q) (Faris et al. 2003), 1.5 Mb/cM for *VRN2* (Yan et al. 2004), and around 2.24 Mb/cM for *Yr36* (Fu et al. 2009), with a genome average calculated at 4.4 Mb/cM (Faris and Gill 2002).

Despite the large number of F_2 lines screened in the 8982-TL population, it was not possible to identify recombinations between *Cdu-B1* and the six co-segregating markers. Perhaps if additional F_2 plants were screened, recombinations could have been identified that would have allowed more precise localization of *Cdu-B1*. Larger populations allow for greater sampling of recombination and are required to precisely map DNA markers particularly for those that are very closely linked. Radiation mapping has also been used in mapping studies and is of particular interest, as it does not rely on recombination. Using this approach in wheat, Kalavacharla et al. (2006) estimated resolution at about 199 kb/break.

The population size used in this study is larger than most that have been used in fine mapping studies of wheat. For fine mapping of *Gpc-B1* a total of 935 gametes were used (Distelfeld et al. 2006), 2719 F_2 plants were used to fine map *Tsn1* (Lu and Faris 2006) and 3095 F_2 plants *VRN1* (Yan et al. 2003). The lack of recombination between markers could indicate that the markers are physically very close on chromosome 5BL

or alternatively, supports that recombination around the *Cdu-B1* locus is suppressed. Recombination suppression is common near centromeres where DNA is tightly packed into heterochromatic chromatin, which restricts chismata formation and chromatid crossovers (Fan et al. 2011). However, *Cdu-B1* is distal from the centromere, thus it is not likely that centromere related suppression is acting in the populations used here. Structural rearrangements along the chromosome could also cause recombination suppression such as an inversion or an alien introgression, as observed in Neu et al. (2002). An introgression would reduce homologous chromosome pairing during meiosis, which would restrict chismata formation and crossing over. Several introgressions have been identified in the wheat genome and have been a useful strategy to transfer disease resistance from wild relatives to cultivated wheat. For example, resistance to the stem rust race TTKS (*Ug99;* Niu et al. 2011) and multiple effective leaf and stripe rust resistance genes *Yr17, Lr37,* and *Sr38* (Seah et al. 2001) in wheat are derived from alien introgressions. Perhaps *Cdu-B1* also resides on an introgressed segment into 5BL. However, this is less likely because nearly all markers developed at *Cdu-B1* co-segregated in a co-dominant fashion (Fig. 4.4). Generally, markers derived from introgressed segments lack an alternate allele, and thus segregate in a dominant fashion. For example, DNA markers associated with VPM (*Yr17, Lr37,* and *Sr38*) are dominant markers (Helguera et al. 2003).

Both fine mapping populations revealed segregation distortion around the *Cdu-B1* locus, which could support reduced recombination within the *Cdu-B1* region. In the 8982-TL population, F_1 plants were screened for heterozygosity; therefore, segregation distortion was not the result of heterogeneous parents or a selfed female plant (i.e., homozygous, F_1 plants). However, during the development of the Svevo/Brigade population, the F_1s were not tested for heterozygosity, and during the population screen, it was noted that Svevo occasionally scored as having the Brigade molecular variant. Thus for this population, it cannot be ruled out that the segregation distortion observed was the result of heterogeneity within the variety Svevo. Interestingly, segregation distortion occurred in the favor of the male parental molecular variant in both populations (Table 4.7). Segregation distortion on wheat chromosome 5B has been observed in previous studies (Faris and Friesen 2009; Kumar et al. 2007; Faris et al.

1998). Faris and Friesen (2009) observed segregation distortion to be caused by the preferential transmission of male gametes (Faris and Friesen 2009). Kumar et al. (2007) observed three segregation distortion regions (SDR) on chromosome 5B in wheat, two regions that showed preferential transmission of male gametes on 5BL (SRD1 and SRD3) while the third showed a fertilization preference for male gametes on 5BS (SRD2). SRD1 has been mapped to a large portion of 5BL spanning three deletion bins including 5BL9 0.76-0.79 (Kumar et al. 2007) where *Cdu-B1* is located. Kumar et al. (2007) also found evidence of suppressed recombination at SRD2.

5.2 Localization of Other Genes Influencing Expression of Grain Cadmium Concentration in Wheat

In a previous study, only a single QTL on 5BL was associated with phenotypic variation in grain Cd concentration (Knox et al. 2009). In that study, transgressive segregation for grain Cd concentration was observed, suggesting that additional minor genes influence grain Cd concentration, supporting an earlier hypothesis that other minor genes influence grain Cd concentration in durum wheat (Clarke et al. 1997b). With the improved genetic map of 5BL reported in this thesis, a second QTL designated as *QCdu.usw-B2* was identified at the marker *XCbf32* (Fig 4.5). Relative to *Cdu-B1*, the effect of *QCdu.usw-B2* was small, but DH lines carrying the W9262-260D3 molecular variant at *XCbf32* consistently expressed lower Cd content than lines carrying the Kofa molecular variant (Table 4.1). The primers for *XCbf32* are known to amplify a portion of *CbfIIId-12* (EU194246; Campoli et al. 2009), a gene coding for a C-repeat binding factor (Cbf) (Campoli et al. 2009). The *Cbfs* are known transcription factors involved in activation of abiotic stress responsive genes in plants and have been associated with enhanced tolerance to cold (Knox et al. 2008; Campoli et al. 2009) and drought responses (Haake et al. 2002). In rye, Cbf expression patterns are dependent on the allelic state at Vrn1 (Campoli et al. 2009) and this has also been shown in wheat (Badawi et al. 2007), and barley (Stockinger et al. 2007). Recently, RT-PCR analysis of several *Cbf* genes revealed transient expression induced by copper stress in hexaploid wheat, suggesting that *Cbfs* may enhance copper tolerance in wheat (Szira et al. 2008). It is possible that the *Cbf* genes have a pleiotropic effect on Cd concentration, possibly

by regulating transpiration rates. Higher transpiration rates have been associated with elevated concentrations of several metals and ions in plants, likely the result of increased movement to sink tissue. Indeed, overexpression of an *Arabidopsis Cbf* gene has been shown to improve water use efficiency and reduce transpiration in rice (Karaba et al. 2007). As well, higher expression of *Cbfs* has been associated with reduced transpiration in wheat. Alternatively, it cannot be ruled out that linked genes near *XCbf32* are influencing grain Cd concentrations.

ScOPC20 was mapped in the CS/CS-DIC 5B population, but in a previous study (Wiebe et al. 2010), no variation in grain Cd concentration and shoot-to-root partitioning was detected between the parents of this mapping population, and both parents were classified as low Cd accumulators. Thus it was not possible to localize *Cdu-B1* in this population. The lack of phenotypic segregation between the parents could suggest one of at least two possibilities. First, a genetic recombination between ScOPC20 and *Cdu-B1* may have occurred, such that *Cdu-B1* itself is not segregating in this population or secondly, additional genes in the D genome lower grain Cd regardless of the allelic state at *Cdu-B1*. The co-segregating markers associated with *Cdu-B1* in the W9262-260D3/Kofa and two fine mapping populations were assessed for polymorphism in the CS/CS-DIC 5B population, and both CS and CS-DIC-5B showed molecular variants identical to Kofa, a high Cd accumulator. These data would suggest that both CS and CS-DIC 5B should express a high Cd phenotype. It is thus more probable that additional genes, perhaps from the hexaploid wheat D genome, maybe compensating for *Cdu-B1* in this population. This is not an unreasonable hypothesis, as most genes contain orthologous copies that are present on each of the homeologous chromosomes of wheat (Mochida et al. 2003). To test this hypothesis, we evaluated Cd in Langdon, a high Cd accumulator. Substitution of Langdon chromosome 5B with chromosome 5D of CS resulted in >50% reduction on Cd concentration (Fig. 4.3). It could be possible that the loss of chromosome 5B (and thus *Cdu-B1*) resulted in the low Cd phenotype. However, substitution of chromosome 5A with CS 5D produced a similar effect as grain Cd concentrations were also reduced by >50% relative to Langdon (Fig. 4.3). Since the Langdon *Cdu-B1* allele is still present in this latter substitution line (as the Langdon 5B chromosome is still present), it is clear that chromosome 5D also contains gene(s) that

can compensate for the high Cd allele at *Cdu-B1* in this population. This may explain in part why hexaploid wheat cultivars show little variation in grain Cd concentration and are considered to be low in grain Cd concentration compared to durum wheat (Zook et al. 1970).

5.3 Reverse Genetic Studies

5.3.1. Phytochelatin Synthase and ABC Transporters

A hypothesis of this thesis was that low Cd is the result of a functional transporter or chelator that transports, or aids in the transport of, Cd to root organelles, thus preventing subsequent translocation to shoots for remobilization to the grain (section 2.3.2, Fig. 2.2). Sequestration of Cd into chemical complexes or physical compartments, such as the vacuole, could occur in root tissues thereby reducing its availability for loading into xylem and phloem. Recent studies have shown the potential of several ABC transporters to sequester Cd in plants by transporting Cd conjugates (glutathione or phytochelatin) into the vacuole (Song et al. 2003; Klein et al. 2006; Wojas et al. 2009). Many higher plants synthesize PCs in response to Cd treatment and bind Cd to form Cd–PC complexes which can be transported across the tonoplast (Salt and Rauser 1995) into the vacuole for sequestration (Vogeli-Lange and Wagner 1990). Therefore, it is possible that transport of Cd–PC complexes via one (or more) ABC transporters into the vacuole of root cells might limit Cd translocation to the shoot for subsequent remobilization to the grain (Stolt et al. 2003). In this study, it was hypothesized that *Xwg644* (a known ABC transporter) and/or *PCS2* would be associated with *Cdu-B1*, as these genes both map to a similar region of chromosome 5AL (Yan et al. 2003). Although logical candidates, both genes were ruled out because they mapped distal to the *QCdu.spa-B1* QTL and neither was associated with *Cdu-B1* (Fig. 4.5). Also, NIL parents for the fine mapping population 8982-TL were segregating for *Cdu-B1* markers but fixed for *Vrn-B1* markers including *PCS2* and *Xwg644* (Appendix 5). This supports the work of Hart et al. (2006) who reported that PC synthesis was not a limiting factor in the differential storage of Cd in roots of high and low Cd accumulating NILs. However, the possibility that other ABC-like transporter or metal chelator genes

exist in the *Cdu-B1* region cannot be ruled out as large ABC transporter families exist in most plants.

5.3.2 Association of *HMA3-B1* with *Cdu-B1* in durum wheat

5.3.2.1 Genetic Mapping of *OsHMA3* orthologues in durum wheat

A major QTL associated with grain and shoot Cd concentration on rice chromosome 7 has been identified (Tezuka et al. 2009; Ueno et al. 2009) and much like *Cdu-B1*, explained a large proportion of the phenotypic variation and low Cd concentration was a dominant trait. In this region of chromosome 7, several putative metal transporter-encoding genes, including *OsZIP8*, cadmium/zinc transporting ATPase (*OsHMA3*) and *OsNramp1* exist. Recently, Takahashi et al. (2009) reported that *OsNramp1* had the capacity to transport both Cd and Fe (iron). Some ZIP proteins have been implicated in heavy-metal uptake in rice (Ishimaru et al. 2005) and an *Arabidopsis* homologue of *OsHMA3* has been shown to transport Cd from the cytosol to vacuoles (Morel et al. 2009). More recently *OsHMA3,* was identified as the gene limiting Cd accumulation in rice grain (Ueno et al. 2010) and in its functional state, was responsible for sequestration of Cd to roots and thus, low Cd phenotype (section 2.3.2, Fig. 2.2). The low Cd phenotype is dominant (Clarke et al. 1997b), and thus the presence of a functional transporter or chelator that sequesters Cd in roots would result in a low Cd phenotype. Therefore, a similar mechanism may be operating in durum wheat as is in rice.

Primers designed from *OsHMA3* of rice mapped to chromosome 2B in the W9262-260D3/Kofa population (Fig. 4.12) and were not associated with phenotypic variation for grain Cd concentration. Wheat chromosome 2B shows colinearity to chromosome 7 in rice (La Rota and Sorrells 2004) where *OsHMA3* has been localized. However, design of additional primers based on the current Chinese Spring 5X sequence was an effective strategy to design several additional primers, which allowed mapping of *HMA3-B1* to the *Cdu-B1* interval in durum wheat. These results may indicate duplication of the gene on chromosome 2B and movement of one duplicated copy to chromosome 5B during the evolution of durum wheat as a homolog is still evident in the donor site (chromosome 2B) (Wicker et al. 2010), which is colinear to rice chromosome

7 (section 2.6, Fig. 2.7) where *OsHMA3* is located. Sequencing of the *HMA-B2* from chromosome 2B would determine whether *HMA-B2* is in fact orthologous to *OsHMA3*. Little work to date has been done on identifying and characterizing *HMA* genes and proteins in wheat. Seven, nine, and eight *HMA* genes have been identified in *Brachypodium*, rice, and *Arabidopsis*, respectively. Due to durum wheat containing two genomes there is also the potential for 16 to 18 P_{1B}-ATPases assuming that each genome would contain at least one of the P_{1B}-ATPases found in rice and *Brachypodium*.

Mapping of *HMA3-B1*, the durum wheat orthologue of *OsHMA3*, revealed complete linkage with *Cdu-B1* in all mapping populations reported here as well as the 96 genetically diverse durum cultivars and breeding lines. Sequence analysis of *HMA3-B1* revealed high grain Cd lines (Langdon and Kofa) to have a 17 bp duplication resulting in a premature stop codon and a non-functioning HMA-B1 protein (Fig.4.15a; d). Marker analysis also revealed all high Cd lines in the genetically diverse durum breeding lines and cultivars are carrying this 17 bp duplication. The effect of the W9262-260D3 allele at *HMA3-B1* was to reduce grain Cd by >50% in the W9262-260D3/Kofa mapping population and the two fine mapping populations. *OsHMA3* is a P_{1B}-ATPase (Ueno et al 2010), which is an ancient family of P-ATPases which physiologically are involved in maintaining homeostasis and transporting transition metals across membranes against their electrochemical gradient (Williams and Mills 2005). A functioning *OsHMA3* gene confers low Cd accumulation in rice and much like *HMA3-B1*, accounts for >70% of phenotypic variation (Ueno et al. 2010). This is similar to the *Cdu-B1* locus in durum wheat, indicating that the same mechanism may be functioning in durum wheat.

5.3.2.2 Comparative analysis of *HMA3* in durum wheat

The duplication in the non-functional allele is likely the result of either an unequal crossover during meiosis or DNA slippage during DNA replication. Both of these processes have resulted in duplications and novel allelic variation in the high molecular weight glutenins (Liu et al. 2007) and disease resistant gene analogues (Luo et al. 2011). This would imply that the low Cd allele at *HMA3-B1* is the wild type allele and the functional allele of *HMA3-B1* would be expected to result in grain Cd

concentration segregating in a dominant fashion, which is the case in durum wheat (Clarke et al. 1997b).

Because durum wheat is a tetraploid, a functional allele coded by the A genome homeologous copy could function to sequester Cd in root vacuole cells, and thus confer a low Cd phenotype. Analysis of *HMA3-A1* sequence from the high accumulating cultivar Langdon revealed a 33 bp deletion in exon 6 when compared to the putative functional allele *HMA3-B1* derived from W9262-260D3. The resulting deletion would translate into a deletion of 11 amino acids from the C terminus. (Fig. 4.15e). These 11 amino acids were not associated with any functional protein domains, so it is not clear if this deletion would impact on Cd transport or impact on protein folding and or integration into the vacuole membrane. Indeed most HMA proteins have eight transmembrane domains (Williams and Mills 2005) with the C and N terminus in the cytosol of the cell which are thought to be involved in metal binding (Williams and Mills 2005) therefore changes to the C and N terminus may influencing binding and therefore transportation of metals. Based on these data, it does appear the *HMA3-A1* allele from Langdon is also non-functional, and would thus not confer the low Cd phenotype, even in the presence of a non-functional *HMA3-B1* allele. Further functional studies would be required to determine if this is a reasonable hypothesis. In addition, sequence analysis of *HMA3-A1* in diverse genetic backgrounds (such as the 96 cultivars and breeding lines used in this thesis to test for association of *HMA3-B1* (section 4.4)) may reveal allelic variation in *HMA3-A1* that could be used to test for association with further reduced grain Cd concentration. Indeed several durum wheat cultivars (9661-AF1D, Borli, and Flavio) studied here expressed grain Cd concentrations >25 ng g-1 less than Strongfield and CDC Verona (Table 4.8), both of which express low grain Cd concentration and carry a putative functional allele at *HMA3-B1*.

To determine the relationship of *HMA3-A1* and *HMA3-B1* to other related *HMA* genes from other plant species, a dendrogram of these sequences was generated (Fig 4.17). Durum wheat *HMA3* sequences show more similarity to *Brachypodium* than rice, which was also expected due to the close relation of Durum wheat to *Brachypodium* than rice (section 2.6, Fig. 2.6). The dendrogram of *HMA* CDS shows similar results to Baxter et al. (2003) except *HMA* CDS identified from *Brachypodium* were included

(Fig. 4.17). HMA proteins were classified into two groups based on the substrate they were predicted to transport, either Cu^{+}/Ag^{+} or $Zn^{2+}/Co^{2+}/Cd^{2+}/Pb^{2+}$. *OsHMA3* and *BdHMA3* show the closest sequence similarities to *HMA3-A1* and *HMA3-B1* and were classified as $Zn^{2+}/Co^{2+}/Cd^{2+}/Pb^{2+}$ transporters. According to Baxter et al. (2003), at least two other *HMA* sequences in rice (*OsHMA2* and *OsHMA1*) have conserved domains, which suggest that they could transport $Zn^{2+}/Co^{2+}/Cd^{2+}/Pb^{2+}$ but neither of these have been associated with phenotypic variation in grain Cd concentration in rice. Potential reasons for this could be a cellular location of the translated protein other than the tonoplast, a non-Cd substrate specificity, the translated proteins are nonfunctional, or the populations used for mapping were not segregating for these genes. For example, diversity of cellular location is identified in the four HMA *Arabidopsis* proteins that are identified to transport $Zn^{2+}/Co^{2+}/Cd^{2+}/Pb^{2+}$ with evidence to support cellular locations of the tonoplast, the plasma membrane and the cholorplastic envelope (Williams and Mills 2005).

Substitution of chromosome 5D from Chinese Spring for chromosome 5A and 5B in the cultivar Langdon resulted in a >50% reduction in grain Cd concentration (Fig 4.3). This could suggest that a homoeologous copy of *HMA3-B1* (designated here as *HMA3-D1*) exists and codes for a functional transporter that could compensate for the non-functional *HMA3-B1* allele present in Langdon. This is a reasonable hypothesis given that several genes in wheat contain orthologous copies on each of the homoeologous chromosomes. Partial genomic sequence of *HMA3-D1* from CS was obtained from the long arm of chromosome 5D from the current wheat survey sequence (Appendix 6). Although the complete coding sequence was not obtained for *HMA3-D1*, two lines of evidence support that it could code for a functional protein. First, *HMA3-D1* revealed the absence of the 5' duplication, similar to the W9262-260D3 *HMA3-B1* (Fig. 4.15a; b). In addition, the 3' 33 bp deletion noted in the Langdon *HMA3-A1* sequence was absent in the CS sequence (Similar to *HMA3-B1* from W9262-260D3). Taken together, the major sequence difference between the functioning transporters *HMA3-B1*/*HMA3-D1* and *HMA3-A1* is the presence of a 33 bp deletion in *HMA3-A1*. This could support that *HMA3-A1* does not function to transport Cd to the vacuoles in

roots due to the 33 bp deletion at the 3' end of that sequence, but more work is required to clarify if this is the case.

5.4 Marker Validation and Breeding for Low Grain Cd Concentration

Breeding is the most reliable approach to reduce grain Cd concentrations in durum wheat. The SCAR marker ScOPC20 has been successfully utilized in MAS programs to develop durum wheat cultivars with reduced grain Cd concentration (Clarke et al. 2006; Pozniak et al. 2009; Grant et al. 2008). However, this marker is a dominant marker that is linked in repulsion to the low Cd phenotype. The three CAPS markers associated with *Cdu-B1* reported here were all validated in a large global collection of cultivars and breeding lines as well as the candidate gene *HMA3-B1*. In the validation population *HMA3-B1*, *Xusw47*, *Xusw17* and *Xusw15b* were accurate in correctly classifying lines into high and low grain Cd concentration with the exception of heterogeneous lines. In comparison to ScOPC20 which misclassified three low Cd lines as high as well as classifying three lines as heterogeneous. This suggests that these markers (*HMA3-B1*, *Xusw47*, *Xusw17* and *Xusw15b*) would have broad application in a range of germplasm and breeding programs globally. In addition, the co-dominant nature of these markers allows for identification of heterozygous individuals, which will be useful for backcrossing the low Cd allele into otherwise agronomically acceptable cultivars. Given the large proportion of durum wheat cultivars evaluated in this study that are high in grain Cd, utilization of these marker to improve grain Cd levels should be a priority. Given that *HMA3-B1* is a candidate gene responsible for phenotypic differences in grain Cd concentration in durum wheat, the marker developed for this locus should be preferred by durum wheat breeders because if it is the gene, no recombination is possible and therefore no error.

In this study, five accessions in the validation population were heterogeneous at at least one marker linked to *Cdu-B1* (Table 4.8). The effect of heterogeneity on grain Cd would be a function of the relative proportions of the high and low alleles present in the variety. Of the lines classified as high accumulators, Svevo (159 ng g^{-1}) was heterogeneous at all markers and had 44 ng g^{-1} less Cd than the average of the high accumulators (203 ng g^{-1}), but still significantly more (73 ng g^{-1}) than the average low

accumulating varieties (85 ng g^{-1}). Given that the low phenotype is dominant, it is likely that Svevo carries a higher frequency of the high accumulating allele. Breeding lines DT696 (121 ng g^{-1}) and 950329 (120 ng g^{-1}) are classified as low grain Cd concentration and are heterogeneous at all or nearly all markers linked to *Cdu-B1*. In these two breeding lines the high allele may be at a higher frequency than the grain Cd concentration indicates due to the dominant nature of the low accumulating allele masking the effect of the high allele. These breeding lines were scored as low accumulators but were some of the highest scoring low Cd concentration lines; this may be due to their heterogeneous nature. High grain Cd concentration cultivar Avonlea (214 ng g^{-1}) carries one heterogeneous marker (ScOPC20) but seems to have no effect on the grain Cd concentration classification. While low grain Cd concentration cultivar Marjak (82 ng g^{-1}) also carries one heterogeneous marker (*Xusw15b*) it also seems to have no effect of grain Cd concentration classification.

Interestingly, less than half the Canadian (11 of 25) and Spanish (4 of 9) breeding lines and cultivars phenotyped are classified as high grain Cd concentration while around three quarters of the Italian (12 of 17) and Australian (7 of 9) breeding lines and cultivar tested were classified as high (Fig. 4.8). All the lines tested from the United States (12) were classified as high grain Cd concentration (Fig. 4.8). However, these results could be due to sampling error or the result of breeding activity and may not be a true representation of the genetic diversity of these agroecological regions. To accurately depict the variation from these regions more research is necessary.

The *Gpc-B1* gene derived from *T. dicoccoides* is a NAC transcription factor (*NAM-B1*) that is known to increase grain protein concentration, by increased nutrient remobilization to the grain from the leaves and accelerated senescence, with little effect on grain yield (Uauy et al. 2006). In addition, recombinant substitution lines carrying the *T. dicoccoides Gpc-B1* allele accumulate more zinc, iron, and manganese (Distelfeld et al. 2006). The chromosome substitution line LDN(DIC-6B) carries *Gpc-B1*, and in this study, expressed 66% more grain Cd than its isogenic parent Langdon (Table 4.8) and supports the hypothesis *Gpc-B1* is involved in more efficient remobilization of metals, from leaves to the grains (Distelfeld et al. 2006). However, it cannot be ruled out that gene(s) other than *Gpc-B1* are involved, as LDN(DIC-6B) carries a complete *T.*

dicoccoides 5B chromosome substitution. However, if *Gpc-B1* does elevate grain Cd, even in lines carrying *Cdu-B1,* this would limit its use in durum wheat breeding programs targeting reduced grain Cd concentrations. Evaluating *Gpc-B1* in near isogenic lines with and without *Cdu-B1* will be required to resolve this hypothesis. Fortunately *Cdu-B1* has no pleiotropic effects on zinc, iron, or manganese concentrations in durum wheat (Clarke et al. 2002).

In this study, *Cdu-B1* was tightly linked to the *Vrn-B1* locus (markers *CSFs*-1, *Xwg644*, *PCS2, PHY-C*; Fig. 4.5), which controls vernalization response in wheat (Iwaki et al. 2002). A recent study in hexaploid wheat by Ferenc Bálint et al. (2009) also reported a QTL for copper tolerance that was associated with *vrn-A1* on 5AL. However, the reported map was not well saturated, so it is was difficult to ascertain if *vrn-A1* per se was associated or if linked genes were responsible for the observed variation in copper tolerance. The *Cdu-B1*/*Vrn-B1* linkage could have implications for breeding low Cd durum wheat cultivars, because the presence of vernalization genes in spring wheat lines can influence flowering time, and thus yield (Iqbal et al. 2007). The parents of the mapping population used here have not been assayed for vernalization requirement, but the markers reported here for *Cdu-B1* could be used effectively to break any undesirable relationships between the low Cd phenotype and any vernalization response associated with *Vrn-B1*.

Hexaploid wheat lines and cultivars are known for having low grain Cd accumulation with little variation of grain Cd concentrations. Studying the effects of the 5D chromosome substituted into the Langdon background for either 5A or 5B chromosome shows a significant decrease in grain Cd (Fig. 4.3). Only when the 5D chromosome was present grain Cd concentration was low. Therefore in Langdon it is possible that neither HMA3-A1 nor HMA3-B1 proteins are functioning to sequester Cd to the vacuole in root cells. If *HMA3-A1* codes for a non-functional P_{1B}-ATPAse it may be of interest to breeders to identify lines containing an allele of *HMA3-A1* that codes for a functional protein. This allele could then be combined with the functional allele of *HMA3-B1* already present in several Canadian durum wheat cultivars to further reduce grain Cd in durum wheat. This should be a high priority for breeders as CODEX is proposing to lower the limits of grain Cd in cereal grain products (section 2.2).

6.0 CONCLUSIONS

6.1 General Conclusions

The main objective of this project was to utilize forward (fine mapping) and reverse (candidate gene) genetic approaches to better understand the molecular basis of Cd accumulation in durum wheat grain. For the reverse genetic studies, it was hypothesized that one of three gene families would be associated with *Cdu-B1* that could function by restricting Cd to the roots of low Cd lines of durum wheat; a) a gene coding for a heavy metal associated (HMA) transporter, b) an ATP-binding cassette (ABC)-like transporter or c) a phytochelatin synthetase (PCS), as these have been associated with Cd uptake and sequestration in other plants. For the forward genetics approach, colinearity was established with rice and *Brachypodium* to identify putative genes that maybe associated with variation in grain Cd concentration in durum. The last objective was to develop and validate breeder friendly markers that could be used globally in durum wheat breeding programs targeting the low grain Cd phenotype.

From the work reported here, the following major conclusions were made:

1) In durum wheat, the *Cdu-B1* locus on chromosome 5B was fine mapped to a genetic interval of 0.14 cM. Fine mapping reduced the genetic interval of *Cdu-B1* from 0.7 cM to 0.14 cM using 3558 F_2 individuals. Seven markers that co-segregate with *Cdu-B1* were identified.

2) The *Cdu-B1* locus in durum wheat is colinear with a 286 Kbp region in rice and a 282 Kbp region identified in *Brachypodium*. No obvious genes associated with Cd uptake, sequestration, or translocation were identified in these colinear regions that could be the target of study in durum wheat

3) *PCS2* and *Xwg644* genes were hypothesized to be associated with *Cdu-B1* as both have been mapped previously to the general vicinity of *Cdu-B1*, and both have been associated with Cd assimilation in other plant species. However,

these two genes could be ruled out as candidates as both genes map distal to *Cdu-B1* in the W9262-260D3/Kofa population and neither gene was segregating in the two near isogenic lines used to develop the 8982-TL population.

4) In rice, *OsHMA3* gene has been shown to restrict Cd to the root vacuoles and thus limits translocation of Cd to above ground tissues for subsequent translocation to the rice grain. The orthologous gene, *HMA3-B1* was characterized from durum wheat and current evidence suggests it is a strong candidate gene for *Cdu-B1*. Several lines of evidence to support this are:

   - *HMA3-B1* co-segregated with *Cdu-B1* in all durum wheat populations studied in this thesis, and was able to discretely classify a set of 96 genetically diverse cultivars and breeding lines into high and low Cd accumulators.

   - Sequencing revealed high grain Cd lines possess a 17 bp duplication that results in a premature stop codon and a severely truncated protein;

   - A non-functional Cd transporter in high accumulating Cd lines and a functional Cd transporter in low grain Cd lines fits the model proposed in rice by Myiadate et al. (2011) where a functional transporter would sequester Cd to the vacuole in roots and therefore would not be available for transport to the shoots and grain.

5) Grain Cd analysis of Langdon disomic lines revealed that the substitution of the hexaploid wheat 5D chromosome into the high grain Cd line, Langdon, for either chromosome 5A or 5B results in a low grain Cd phenotype

6) Analysis of *HMA3-A1* from chromosome 5AL revealed a 33 bp deletion which could result in a non-functional protein. However, detailed functional analysis was not conducted to determine if this hypothesis is correct.

7) Several DNA markers were developed to assist breeding of low grain Cd concentration in durum wheat. In total, seven markers that co-segregated with *Cdu-B1* were identified. Four of these (including *HMA3-B1*) were converted to high throughput co-dominant markers for use in breeding programs. Selection

using *HMA3-B1* is recommended, because if this is the gene that regulates Cd concentration in durum wheat grain, the DNA marker would be considered a "perfect" marker.

6.2 Future Work

To complete the forward genetics approach, a positional cloning strategy (section 2.5) could be pursued. This would require chromosome walking across the *Cdu-B1* region, BAC sequencing, and sequence assembly and annotation. However, because a strong candidate gene (*HMA3-B1*) has been identified through the reverse genetic studies, positional cloning would be less of a priority in favour of functional analysis of HMA3-B1 protein in durum wheat.

For functional analysis of *HMA3-B1,* several strategies could be pursued. These could include a combination of Target Induced Local Lesion in Genomes (TILLING) studies, transgenic experiments, expression studies, and yeast complementary studies. TILLING studies would include generation of a TILLING population using one of several chemical mutagens (Pozniak and Hucl 2004) and screening for mutations within *HMA3-B1* that would result in a non-functional protein. Choosing an appropriate parent to generate the TILLING population is critical, and in this case, would require TILLING in a durum wheat cultivar that expresses low grain Cd (and thus a hypothesized, functional Cd transporter). Knock-out mutations in a low grain Cd cultivar should confer high grain Cd concentration if *HMA3-B1* is indeed the casual gene. However, TILLING can be complicated in a polyploid genome due to homoeologous copies of genes, but as seen in this project, Langdon disomic substitution lines are effective in assigning gene amplicons to the individual wheat chromosomes. Another limitation of TILLING is a large population is necessary for screening for mutations. However, Uauy et al. (2009) has utilized a four-fold pooling strategy in tetraploid and hexaploid wheat mutation populations that reduces reactions required for mutation screening.

Transgenic studies could also be employed. In this case, transforming a functional *HMA3-B1* gene (like that from W9262-260D3) into a high grain Cd accumulator should confer the low phenotype. Because the HMA3-B1 protein is hypothesized to transport Cd to root cell vacuoles, selection of an appropriate promoter

that results in a high level of *HMA3-B1* transcript expression in the roots would be required. Several root specific promoters exist for functional studies, and have been used effectively for functional studies of genes related to silicon (Montpetit et al. 2012) and boron tolerance in wheat (Sutton 2012).

Yeast complementation studies could also be used to determine cellular location and affinity of the functional HMA3-B1 protein for Cd. A similar approach was described by Miyadate et al. (2011) and Morel et al. (2009) who identified OsHMA3 and AtHMA3 proteins respectively to be localized to the tonoplast. In yeast complementary studies, *HMA3-B1* could be transformed into Cd sensitive yeast to determine if the durum functional allele of *HMA3-B1* restores Cd tolerance. A similar approach was described by Miydate et al. (2011) and Gravot et al. (2004) who showed Cd tolerance restored by OsHMA3 and AtHMA3 proteins respectively. To fully verify the *HMA3-B1* gene a combination of all of the above strategies would likely be useful.

In allopolyploid genomes, orthologous copies are evident on each of the homoeologous chromosomes, and may perform a similar function. *HMA3-A1* and *HMA3-D1* genes have been identified in durum and hexaploid wheat and further work would be required to verify their functionality. *HMA3-A1* from this thesis was hypothesized not to function due to a 33 bp deletion in the cultivar Langdon. An allelic survey could be performed on the 96 genetically diverse lines and cultivars from this thesis to determine if a functional *HMA3-A1* allele is present in any lines. That work would then lay the foundation for functional analysis of the A genome copy, and to determine if combining functional alleles at both *HMA3-A1* and *HMA3-B1* further reduces grain Cd concentration in durum wheat.

7.0 COPYRIGHT CLEARANCE

As per the requirements of the College of Graduate Studies at the University of Saskatchewan, the following lists the copyright clearance and license number for figures presented in this thesis that have been modified from works reported previously in the published literature or from appropriate web sites.

Figure 2.1a (pg 7)
Modified from Harris and Taylor 2004
Open access journal

Figure 2.1b (pg 7)
Modified from Ueno et al. 2009
Licensee: Krystalee D Wiebe
License Date: Sep 7, 2012
License Number: 2983690670483
Publication: Plant and Cell Physiology
Title: Identification of a Novel Major Quantitative Trait Locus Controlling Distribution of Cd Between Roots and Shoots in Rice
Type Of Use: Thesis/Dissertation
Order Reference: Figure 2

Figure 2.2 (pg 8)
Modified from Miyadate et al. 2011
Licensee: Krystalee D Wiebe
License Date: Sep 6, 2012
License Number: 2983150627198
Publication: New Phytologist
Title: OsHMA3, a P1B-type of ATPase affects root-to-shoot cadmium translocation in rice by mediating efflux into vacuoles
Type Of Use: Dissertation/Thesis

Figure 2.3 (pg 11)
Modified from Knox et al. 2009
Licensee: Krystalee D Wiebe
License Date: Sep 6, 2012
License Number: 2983141282818
Publication: Genome
Title: Chromosomal location of the cadmium uptake gene (*Cdu1*) in durum wheat
Type Of Use: Thesis/Dissertation
Order Reference: Figure 2

Table 2.3 (pg 11)
Modified from Krattinger et al. 2009
Licensee: Krystalee D Wiebe
License Date: Sep 6, 2012
License Number: 2983150145507
Publication: Springer eBook
Title: Map-Based Cloning of Genes in *Triticeae* (Wheat and Barley)
Type Of Use: Thesis/Dissertation
Order Reference: Table 12.1

Figure 2.6 (pg 23)
Modified from Buell 2009
Permission Status: Granted
Permission type: Republish or display content
Type of use: Republish in a dissertation
Requested use: Dissertation
Organization status: Non-profit 501(c)(3)
Circulation/ Distribution: 5000
Type of content: Figure/ diagram/ table
Description of requested content: *Poaceae* genomics: going from unattainable to becoming a model clade for comparative plant genomics
Page range(s): pg 114

Figure 2.7 (pg 24)
Modified from La Rota and Sorrells 2004
Licensee: Krystalee D Wiebe
License Date: Sep 6, 2012
License Number: 2983150295618
Publication: Functional & Integrative Genomics
Title: Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between rice and wheat
Type Of Use: Thesis/Dissertation
Order Reference: Figure 2

Figure 2.8 (pg 25)
Modified from The International *Brachypodium* Initiative 2010
Licensee: Krystalee D Wiebe
License Date: Sep 6, 2012
License Number: 2983140991597
Publication: Nature
Title: Genome sequencing and analysis of the model grass *Brachypodium distachyon*
Type Of Use: reuse in a thesis/dissertation

Appendix 8 (pg 119)

Modified from Reimer et al. 2008

Licensee: Krystalee D Wiebe
License Date: Sep 7, 2012
License Number: 2983670655839
Publication: Genome
Title: Association mapping of yellow pigment in an elite collection of durum wheat cultivars and breeding lines
Type Of Use: Thesis/Dissertation
Order Reference: Supplementary Table 1

8.0 REFERENCES

AAFC (2010) Wheat: Situation and Outlook. Market Outlook Report. Agriculture Agri-Food Canada. April 1, 2010. Vol. 2 No. 1

AAFC (2005) Durum wheat: 2005-2006 situation and outlook. Bi-weekly Bulletin. Agriculture Agri-Food Canada. December 16, 2005. Vol. 18. No. 20

Anderson A, Hanson HN (1984) High cadmium and nickel contents in sunflower kernels. Z Lebensm Unters Forsch 179:399-440

Badawi M, Danyluk J, Boucho B, Houde M, Sarhan F (2007) The *CBF* gene family in hexaploid wheat and its relationship to the phylogenetic complex of cereal *CBF*s. Mol Genet Genomics 277:533-554

Bassam BJ, Gresshoff PM (2007) Silver staining DNA in polyacrylamide gels. Nat Protoc 2:2649- 2654

Baxter I, Tchieu J, Sussman MR, Boutry M, Palmgren MG, Gribskov M, Harper JF, Axelsen KB (2003) Genomic comparison of P-type ATPase ion pumps in *Arabidopsis* and rice. Plant Physiol 132:618-628

Benitez ER, Hajika M, Yamada T, Takahashi K, Oki N, Yamada N, Nakamura T, Kanamaru K (2010) A major QTL controlling seed cadmium accumulation in soybean. Crop Sci 50:1728-1734

Benitez ER, Makita Hajika M, Takahashi R (2012) Single-base substitution in P_{1B}-ATPase gene is associated with a major QTL for seed cadmium concentration in soybean. J Hered 103:278-286

Bennetzen JL (2000) Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell 12:1021-1029

Bennetzen JL, Ma J (2003) The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr Opin Plant Biol 6:128-133

Beyersmann D, Hechtenberg S (1997) Cadmium, gene regulation, and cellular signalling in mammalian cells. Toxicol Appl Pharmacol 144:247-261

Bodo M, Balloni S, Lumare E, Bacci M, Calvitti M, Dell'Omo M, Murgia N, Marinucci L (2010) Effects of sub-toxic cadmium concentrations on bone gene expression program: Results of an *in vitro* study. Toxicol In Vitro 24:1670-1680

Bossolini E, Wicker T, Knobel PA, Keller B (2007) Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation. Plant J 49:704-717

Bovet L, Feller U, Martinoia E (2003) Possible involvement of plant ABC transporters in cadmium detoxification: a cDNA sub-microarray approach. Environ Int 31:263-7

Buell CR (2009) *Poaceae* genomes: going from unattainable to becoming a model clade for comparative plant genomics. Plant Physiol 149:111-116

Campoli C, Matus-Cádiz MA, Pozniak C, Cattivelli L, Fowler DB (2009) Comparative expression of *Cbf* genes in the *Triticeae* under different acclimation induction temperatures. Mol Genet 282:141-152

Cenci A, Chantret N, Kong X, Gu Y, Anderson OD, Fahima T, Distelfeld A, Dubcovsky J. (2003) Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. *durum*). Theor Appl Genet 107:931-939

Cenci A, Somma S, Chantret N, Dubcovsky J, Blanco A (2004) PCR identification of durum wheat BAC clones containing genes coding for carotenoid biosynthesis enzymes and their chromosome localization. Genome 47:911-917

CFIA (2006) The Biology of *Triticum turgidum* ssp. *durum* (Durum Wheat). Canadian Food Inspection Agency. July 6, 2006. Biology Document BIO2006-07 http://www.inspection.gc.ca/english/plaveg/bio/dir/dir0607e.pdf

Chee PW, Elias EM, Anderson JA, Kianian SF (2001) Evaluation of a high grain protein QTL from *Triticum turgidum* L. var. *dicoccoides* in an adapted durum wheat background. Crop Sci 41:295-301

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Le Paslier M-C, Magdelenat G, Gonthier C, Couloux A, Budak H, Breen J, Pumphrey M, Liu S, Kong X, Jia J, Gut M, Brunel D, Anderson JA, Gill BS, Appels R, Keller B, Feuillet C (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. The Plant Cell 22:1686-1701

Cieśliński G, Van Rees KCJ, Huang PM, Kozak LM, Rostad HPW, Knott DR (1996) Cadmium uptake and bioaccumulation in selected cultivars of durum wheat and flax as affected by soil type. Plant and Soil 182:115-124

Clarke JM, Knox RE, DePauw RM, Clarke FR, Fernandez MR, McCaig TN, Singh AK (2009) Brigade durum wheat. Can J Plant Sci 89:505-509

Clarke JM, Leisle D, DePauw RM, Thiessen LL (1997a) Registration of five pairs of durum wheat genetic stocks near-isogenic for cadmium concentration. Crop Sci 37:297

Clarke JM, Leisle D, Kopytko GL (1997b) Inheritance of cadmium concentration in five durum wheat crosses. Crop Sci 37:1722-1726

Clarke JM, McCaig TN**,** DePauw RM**,** Knox RE**,** Clarke FR**,** Fernandez MR, Ames N (2006) Registration of 'Strongfield' Durum Wheat. Crop Sci 46:2306-2307

Clarke JM, McLeod JG, DePauw RM, Marchylo BA, McCaig TN, Knox RE, Fernandez MR (2001) Registration of 'AC Navigator' Durum wheat. Crop Sci 41: 270-271

Clarke JM, Norvell WA, Clarke FR, Buckley WT (2002) Concentration of cadmium and other elements in the grain of near-isogenic durum lines. Can J Plant Sci 82:27-33

Clarke JM (2005) Durum wheat improvement in Canada. In: Royo C, Nachit MM, Di Fonzo N, Araus JL, Pfeiffer WH, Slafer GA (eds.) Durum wheat breeding. Current approaches and future strategies. Vol.2, pp 921-938

Cloutier S, McCallum BD, Loutre C, Banks TW, Wicker T, Feuillet C, Keller B, Jordan MC (2007) Leaf rust resistance gene *Lr1*, isolated from bread wheat is a member of the large psr567 gene family. Plant Mol Biol 65:93-106

CODEX STAN 193-1995 (2009) Codex General Standard for Contaminants and Toxins in Foods and Feed. Revision 5, 2009. http://www.codexalimentarius.net/download/standards/17/CXS_193e.pdf.

Courbot M, Willems G, Motte P, Arvidsson S, Roosens N, Saumitou- Laprade P, Verbruggen N. (2007) A major quantitative trait locus for cadmium tolerance in *Arabidopsis halleri* colocalizes with *HMA4*, a gene encoding a heavy metal ATPase1. Plant Physiol 144:1052-1065

Dakouri A, McCallum BD, Walichnowski AZ, Cloutier S (2010) Fine-mapping of the leaf rust *Lr34* locus in *Triticum aestivum* (L.) and characterization of large germplasm collections support the ABC transporter as essential for gene function. Theor Appl Genet 121:373-384

Del Castilho P, Chardon WJ (1995) Uptake of soil cadmium by three field crops and its prediction by a pH-dependent Freundlich sorption model. Plant Soil 171:263-266

Deniau AX, Pieper B, Ten Bookum WM, Lindhout P, Aarts MGM, Schat H (2006) QTL analysis of cadmium and zinc accumulation in the heavy metal hyperaccumulator *Thlaspi caerulescens*. Theor Appl Genet 113:907-920

Devos KM, Costa de Oliveira A, Xu X, Estill JC, Estep M, Jogi A, Morales M, Pinheiro J, San Miguel P, Bennetzen JL (2008) Structure and organization of the wheat genome – the number of genes in the hexaploid wheat genome. In: Appels R, Esatwood R, Lagudah E, Langridge P, Mackay M, McIntyre L, Sharp P (eds) Proc 11th Int Wheat Genet Symp. Sydney, Sydney University Press, eO24, pp1-5

Devos KM, Ma J, Pontaroli AC, Pratt LH, Bennetzen JL (2005) Analysis and mapping of randomly chosen bacterial artificial chromosome clones from hexaploid bread wheat. Proc Natl Acad Sci U S A 102:19243-19248

Distelfeld A, Cakmak I , Peleg Z , Ozturk L, Yazici AM , Budak H , Saranga Y, Fahima T (2006) Multiple QTL-effects of wheat *Gpc-B1* locus on grain protein and micronutrient concentrations. Physiologia Plantarum 129:635-653

Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan L, Shiloff BA, Bennetzen JL (2001) Comparative sequence analysis of collinear barley and rice bacterial artificial chromosomes. Plant Physiol 125:1342-1353

Dutilleul C, Jourdain A, Bourguignon J, Hugouvieux V (2008) The *Arabidopsis* putative selenium-binding protein family: Expression study and characterization of *SBP1* as a potential new player in cadmium detoxification processes. Plant Physiol 147:239-251

Eckstein PE, Hay D, Rossnagel BG, Scoles GJ (2004) A simple and effective method for molecular marker-assisted selection. In: Proceedings of 9th International Barley Genetics Symposium, Brno, Czech Republic, 20–26 June, 2004, pp 259-262

Elias EM, Manthey FA (2005) End products: present and future uses. In: Durum wheat breeding current approaches and future strategies. C. Royo, M.M. Nachit, N. DiFonzo, J.L. Araus, W.H. Pfeiffer, and G.A. Slafer eds. Food Products Press. New York, pp 63-86

EA (2007) UK Soil and Herbage Pollutant Survey, Report Number 7. EA. Bristol, UK.

Erayman M, Sandhu D, Sidhu D, Dilbirligi M, Baenziger1 PS, Gill KS (2004) Demarcating the gene rich regions of the wheat genome. Nucleic Acids Res 32:3546-3565

Fan C, Walling JG, Zhang J, Hirsch CD, Jiang J, Winga RA (2011) Conservation and purifying selection of transcribed genes located in a rice centromere. The Plant Cell 23:2821-2830

Faris JD, Fellers JP, Brooks SA, Gill BS (2003) A bacterial artificial chromosome contig spanning the major domestication locus *Q* in wheat and identification of a candidate gene. Genetics 164: 311–321

Faris JD, Friesen TL (2009) Reevaluation of a tetraploid wheat population indicates that the *Tsn1*–ToxA interaction is the only factor governing *Stagonospora nodorum* blotch susceptibility. Phytopathology 99:906-912

Faris JD, Gill BS (2002) Genomic targeting and high-resolution mapping of the domestication gene *Q* in wheat. Genome 45:706-718

Faris JD, Haen KM, Gill BS (2000) Saturation mapping of a gene-rich recombination hot spot region in wheat. Genetics 154:823-835

Faris JD, Laddomada B, Gill BS (1998) Molecular mapping of segregation distortion loci in *Aegilops tauschii*. Genetics 149:319-327

Faris JD, Zhang Z, Fellers JP, Gill BS (2008) Micro-colinearity between rice, *Brachypodium*, and *Triticum monococcum* at the wheat domestication locus *Q*. Funct Integr Genomics 8:149-164

Faris JD, Zhang Z, Lu H, Lu S, Reddy L, Cloutier S, Fellers JP, Meinhardt SW, Rasmussen JB, Xu SS, Oliver RP, Simons KJ, Friesen TL (2010) A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. Proc Natl Acad Sci U S A 107:13544-13549

Ferreira A, da Silva MF, da Costa e Silva L, Cruz CD (2006). Estimating the effects of population size and type on the accuracy of genetic maps. Genet Mol Biol 29: 182-192

Farrell RE, McArthur DFE, Van Rees KCJ (2005) Net Cd^{2+} flux at the root surface of durum wheat (*Triticum turgidum* L. var. *durum*) cultivars in relation to cultivar differences in Cd accumulation. Can J Plant Sci 85:103-107

Ferenc Bálint A, Szira F, Röder MS, Galiba G Börner A (2009) Mapping of loci affecting copper tolerance in wheat—The possible impact of the vernalization gene *Vrn-A1*. Environ Exp Bot 65:369-375

Feuillet C, Eversole K (2007) Physical mapping of the wheat genome: a coordinated effort to lay the foundation for genome sequencing and develop tools for breeders. Isr J Plant Sci 55:307-313

Feuillet C, Keller B (2002) Comparative genomics in the grass family: Molecular characterization of grass genome structure and evolution. Ann Bot 89:3-10

Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B (2003) Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc Natl Acad Sci USA 100:15253-15258

Fleischer M, Sarofim AF, Fassett DW, Hammond P, Schacklette HT, Nisbet IC, Epstein S (1974) Environmental impact of cadmium: a review by the panel on hazardous trace substances. Environ Health Perspect 7:263-323

Francki M, Carter M, Ryan K, Hunter A, Bellgard M, Appels R (2004) Comparative organization of wheat homoeologous group 3S and 7L using wheat–rice synteny and identification of potential markers for genes controlling xanthophyll content in wheat. Funct Integr Genomics 4:118-130

Frank R, Ishida K, Suda P (1976) Metals in Agricultural Soils of Ontario. Can J Soil Sci 56:181-196

Fu D, Uauy C, Distelfeld A, Blechl A, Epstein, L, Chen X, Sela, H, Fahima T, Dubcovsky J (2009) A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. Science 323:1357-136

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95:1971–1974

Gallagher CM, Chen JJ, Kovach JS (2010) Environmental cadmium and breast cancer risk. Aging (Albany NY) 2:804-814

Gill KS, Gill BS, Endo TR, Boiko EV (1996) Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. Genetics 143:1001-1012

Grant CA, Buckley WT, Bailey LD, Selles F (1998) Cadmium accumulation in crops. Can J Plant Sci 78:1-17

Grant CA, Clarke JM, Duguid S, Chaney RL (2008) Selection and breeding of plant cultivars to minimize cadmium accumulation. Sci of the Total Environ 390:301-310

Gravot A, Lieutaud A, Verret F, Auroy P, Vavasseur A, Richaud P (2004) AtHMA3, a plant P-1B-ATPase, functions as a Cd/Pb transporter in yeast. FEBS Lett 561:22-28

Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G (2006) Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. Nature 439:749-752

Grill E, Winnacker E-L, Zenk MH (1985) Phytochelatins: the principal heavy-metal complexing peptides of higher plants. Science 230:674-676

Haag-Kerwer A, Schaèfer HJ, Heiss S, Walter C, Rausch T (1999) Cadmium exposure in *Brassica juncea* causes a decline in transpiration rate and leaf expansion without effect on photosynthesis. J of Exp Bot 50:1827-1835

Haake V, Cook D, Riechmann JL, Pineda O, Thomashow MF, Zhang JZ (2002) Transcription factor CBF4 is a regulator of drought adaptation in *Arabidopsis*. Plant Physiol 130:639-648

Harris NS, Bryman MJ, Pozniak CJ, Taylor GJ (2007) Transcriptome comparison of high and low cadmium accumulating near isogenic lines of durum wheat. Plant Canada 2007, Saskatoon, SK, Canada, 10-14 June, 2007

Harris NS, Taylor GJ (2001) Remobilization of cadmium in maturing shoots of near-isogenic lines of durum wheat that differ in grain cadmium accumulation. J of Exp Bot 52:1473-1481

Harris NS, Taylor GJ (2004) Cadmium uptake and translocation in seedlings of near isogenic lines of durum wheat that differ in grain cadmium accumulation. BMC Plant Biol 4:4

Hart JJ, Welch RM, Norvell WA, Kochian LV (2002) Transport interactions between cadmium and zinc in roots of bread and durum wheat seedlings. Physiologia Plantarum 116:73-78

Hart JJ, Welch RM, Norvell WA, Kochian LV (2006) Characterization of cadmium uptake, translocation and storage in near-isogenic lines of durum wheat that differ in grain cadmium concentration. New Phytol 172:261-271

Hart JJ, Welch RM, Norvell WA, Sullivan LA, Kochian LV (1998) Characterization of cadmium binding, uptake, and translocation in intact seedlings of bread and durum wheat cultivars. Plant Physiol 116:1413–1420

HPA (2009) Contaminated Land Information Sheet, Cadmium Version 2 CHAPD HQ & CHAPD Cardiff

Heiss S, Wachtre A, Bogs J, Cobbet C, Rausch T (2003) Phytochelatin synthase (PCS) protein is induced in *Brassica juncea* leaves after prolonged Cd exposure. J Exp Bot 54:1833-1839

Helguera M, Khan IA, Kolmer J, Lijavetzky D, Zhong-qi L, Dubcovsky J (2003) PCR assays for the *Lr37-Yr17-Sr38* cluster of rust resistance genes and their use to develop isogenic hard red spring wheat lines. Crop Sci 43:1839-1847

Higgins JA, Bailey PC, Laurie DA (2010) Comparative Genomics of Flowering Time Pathways Using *Brachypodium distachyon* as a Model for the Temperate Grasses. PLoS ONE 5(4): e10065 .doi:10.1371/journal.pone.0010065

Hinesly TD, Alexander DE, Ziegler EL, Barrett GL (1978) Zinc and Cd accumulation by corn inbreds grown on sludge amended soil. Agron J 70: 425–428

Huang L, Brooks SA, Li W, Fellers JP, Trick HN, Gill BS (2003) Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics 164: 655–664

Huo N, Lazo GR, Vogel JP, You FM, Ma Y, Hayden DM, Coleman-Derr D, Hill TA, Dvorak J, Anderson OD, Luo, M-C, Gu YQ (2008) The nuclear genome of *Brachypodium distachyon*: Analysis of BAC end sequences. Funct Integr Genomics 8:135-147

Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R (2007) Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics 8:460

IARC (1993) Cadmium and cadmium compounds. In Beryllium, Cadmium, Mercury and Exposures in the Glass Manufacturing Industry. IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans, vol. 58. Lyon, France: International Agency for Research on Cancer, pp 119-239

The International *Brachypodium* Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763-768

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793-800

Iqbal M, Navabi A, Yang R-C, Salmon DF, Spaner D (2007) Molecular characterization of vernalization response genes in Canadian spring wheat. Genome 50:511-516

Ishikawa S, Abe T, Kuramata M, Yamaguchi M, Ando T, Yamamoto T, Yano M (2010) A major quantitative trait locus for increasing cadmium specific concentration in rice grain is located on the short arm of chromosome 7. J. Exp. Bot. 61: 923–934.

Ishikawa S, Ae N, Yano M (2005) Chromosomal regions with quantitative trait loci controlling cadmium concentration in brown rice (*Oryza satiza*). New Phytol 168:345-350

Ishimaru Y, Suzuki M, Kobayashi T, Takahashi M, Nakanishi H, Mori S, Nishizawa NK (2005) *OsZIP4*, a novel zinc-regulated zinc transporter in rice. J of Exp Bot 56: 3207–3214

Itoh T, Tanaka T, Barrero RA, Yamasaki C, Fujii Y, Hilton PB, Antonio BA, Aono H, Apweiler R, Bruskiewich R, et al (2007) Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. Genome Res 17:175-83

Iwaki K, Nishida J, Yanagisawa T, Yoshida H, Kato K (2002) Genetic analysis of *Vrn-B1* for vernalization requirement by using linked dCAPS markers in bread wheat (*Triticum aestivum* L.). Theor Appl Genet 104:571-576

Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002) *Arabidopsis* map-based cloning in the post-genome era. Plant Physiology 129:440-450

Jegadeesan S, Yu K, Poysa V, Gawalko E, Morrison MJ, Shi C, Cober E (2010) Mapping and validation of simple sequence repeat markers linked to a major gene controlling seed cadmium accumulation in soybean [*Glycine max* (L.) Merr.]. Theor Appl Genet 121:283-294

Jin YH, Clark AB, Selbos RJ, Al-Refai H, Taylor JA, Kunkel TA, Resnick MA, Gordenin DA (2003) Cadmium is a mutagen that acts by inhibiting mismatch repair. Nature Genet 34: 326-329

Joppa LR, Cantrell RG (1990) Chromosomal location of genes for grain protein content in wild tetraploid wheat. Crop Sci 30:1059–1064

Joppa LR, Williams ND (1988) Langdon durum disomic substitution lines and aneuploid analysis in tetraploid wheat. Genome 30: 222–228

Kalavacharla V, Hossain K, Gu Y, Riera-Lizarazu O, Vales MI, Bhamidimarri S, Gonzalez- Hernandez JL, Maan SS, Kianian SF (2006) High-resolution radiation hybrid map of wheat chromosome 1D. Genetics 173:1089-1099

Karaba A, Dixit S, Greco R, Aharoni A, Trijatmiko KR, Marsch-Martinez N, Krishnan A, Nataraja KN, Udayakumar M, Pereira A (2007) Improvement of water use efficiency in rice by expression of *HARDY*, an *Arabidopsis* drought and salt tolerance gene. Proc Natl Acad Sci USA 104:15270-15275

Kashiwagi T, Shindoh K, Hirotsu N, Ishimaru K (2009) Evidence for separate translocation pathways in determining cadmium accumulation in grain and aerial plant parts in rice. BMC Plant Biol 9:8

Kazantzis G (2004) Cadmium, osteoporosis and calcium metabolism. Biometals 17:493–498

Kilian B, Özkan H, Deusch O, Effgen S, Brandolini A, Kohl J, Martin W, Salamini F. (2007) Independent wheat B and G genome origins in outcrossing *Aegilops* progenitor haplotypes. Mol Biol Evol 24: 217–227

Kim DY, Bovet L, Maeshima M, Martinoia E, Lee Y (2007) The ABC transporter AtPDR8 is a cadmium extrusion pump conferring heavy metal resistance. Plant J 50:207-218

Kim DH, Kim GH, Van K, Kim MY, Lee SH (2010) Fine mapping of a resistance gene to bacterial leaf pustule in soybean. Theor Appl Genet 120:1443-1450

Klein M, Burla B, Martinoia E (2006) The multidrug resistance-associated protein (MRP/ABCC) subfamily of ATP-binding cassette transporters in plants. FEBS Lett 580:1112-1122

Knox AK, Li C, Vágújfalvi A, Galiba G, Stockinger E, Dubcovsky J (2008) Identification of candidate *CBF* genes for the frost tolerance locus $Fr\text{-}A^{m}2$ in *Triticum monococcum*. Plant Mol Biol 67:257-270

Knox RE, Pozniak CJ, Clarke FR, Clarke JM, Houshmand S, Singh AK (2009) Chromosomal location of the cadmium uptake gene (*Cdu1*) in durum wheat. Genome 52:741-747

Kobori G, Okazaki M, Motobayashi T, Matsukawa K (2010) Differences in cadmium uptake and accumulation in three soybean cultivars. Pedologist 214-221

Krattinger SG, Wicker T, Keller B (2009) Map-based cloning of genes in *Triticeae* (wheat and barley). In: Feuillet C, Muehlbauer GJ (eds) Genetics and Genomics of the *Triticeae*. Springer Science+Business Media, LLC2009, pp 337-357

Kudo H, Kudo K, Amboc H, Uemuraa M, Kawai S (2011) Cadmium sorption to plasma membrane isolated from barley roots is impeded by copper association onto membranes. Plant Sci 180:300-305

Kumar S, Gill BS, Faris JD (2007) Identification and characterization of segregation distortion loci along chromosome 5B in tetraploid wheat. Mol Genet Genomics 278:187-196

Kumar S, Mohan A, Balyan HS, Gupta PK (2009) Orthology between genomes of *Brachypodium*, wheat and rice. BMC Res Notes 2:93

Lagudah ES, McFadden H, Singh RP, Huerta-Espino J, Bariana HS, Spielmeyer W (2006) Molecular genetic characterization of the *Lr34/Yr18* slow rusting resistance gene region in wheat. Theor Appl Genet 114:21-30

Lambert R, Grant C, Sauve S (2007) Cadmium and zinc in soil solution extracts following the application of phosphate fertilizers. Sci of the Total Environ 378:293-305

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

La Rota M, Sorrells ME (2004) Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between wheat and rice. Funct Integr Genomics 4:34-46

Liu J, Qian M, Cai G, Yang J, Zhu Q (2006) Uptake and translocation of Cd in differentrice cultivars and the relation with Cd accumulation in rice grain. J Haz Mat 143:443-447

Liu S, Zhao S, Chen F, Xia G (2007) Generation of novel high quality HMW-GS genes in two introgression lines of *Triticum aestivum/Agropyron elongatum*. BMC Evol Biol 7:76-84

Li W, Gill BS (2002) The colinearity of the *Sh2*/*A1* orthologous region in rice sorghum and maize is interrupted and accompanied by genome expansion in the *Triticeae*. Genetics 160:1153-1162

Li W, Zhang P, Fellers JP, Friebe B, Gill BS (2004) Sequence composition, organization, and evolution of the core *Triticeae* genome. Plant J 40:500-511

Li YM, Chaney RL, Schneiter AA, Miller JF, Elias EM, Hammond JJ (1997) Screening for low grain cadmium phenotypes in sunflower, durum wheat and flax. Euphytica 94:23–30

Luo S, Peng J, Li K, Wang M, Kuang H (2011) Contrasting Evolutionary Patterns of the *Rp1* Resistance Gene Family in Different Species of *Poaceae*. Mol Biol Evol 28:313-325

Lu H, Faris JD (2006) Macro- and microcolinearity between the genomic region of wheat chromosome 5B containing the *Tsn1* gene and the rice genome. Funct Integr Genomics 6:90-105

Lu H-J, Fellers JP, Friesen TL, Meinhardt SW, Faris JD (2006) Genomic analysis and marker development for the *Tsn1* locus in wheat using bin-mapped ESTs and flanking BAC contigs. Theor Appl Genet 112: 1132-1142

Luparello C, Sirchia R, Longo A (2011) Cadmium as a transcriptional modulator in human cells. Crit Rev Toxicol 41:75-82

McMurray CT, Tainer JA (2003) Cancer, cadmium and genome integrity. Nature Genet 34:239-241

Mezard C (2006) Meiotic recombination hotspots in plants. Biochem Soc Trans 34:531-534

Miyadate H, Adachi S, Hiraizumi A, Tezuka K, Nakazawa N, Kawamoto T, Katou K, Komada I, Sakurai K, Takahishi H et al. (2011) OsHMA3, a P1B-type of ATPase affects root-to-shoot cadmium translocation in rice by mediating efflux into vacuoles. New Phytol 189:190-199

Mochida K, Yamazaki Y, Ogihara Y (2003) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. Mol Gen Genomics 270:371-377

Montpetit J, Vivancos J, Mitani-Ueno N, Yamaji N, Rémus-Borel W, Belzile F, Ma JF, Bélanger RR (2012) Cloning, functional characterization and heterologous expression of *TaLsi1*, a wheat silicon transporter gene. Plant Mol Biol 79:35-46

Morghan JT (1993) Accumulation of cadmium and selected elements in flax seed grown on a calcareous soil. Plant Soil 150: 61-68

Morel M, Crouzet J, Gravot A, Auroy P, Leonhardt N, Vavasseur A, Richaud P (2009) AtHMA3, a P1B-ATPase Allowing Cd/Zn/Co/Pb Vacuolar Storage in *Arabidopsis*. Plant Physiol 149:894-904

Neu C, Stein N, Keller B (2002) Genetic mapping of the *Lr20 - Pm1* resistance locus reveals suppressed recombination on chromosome arm 7AL in hexaploid wheat. Genome 45:737-744

Niu Z, Klindworth DL, Friesen TL, Chao S, Jin Y, Cai X, Xu SS (2011) Targeted introgression of a wheat stem rust resistance gene by DNA marker- assisted chromosome engineering. Genet advanced access 10.1534/genetics.110.123588

Oliver DP, Hannam R, Tiller KG, Wilhelm NS, Merry RH, Cozens GD (1994) The effects of zinc fertilization on cadmium concentration in wheat grain. J Environ Qual 23:705-711

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR (2007) The TIGR rice genome annotation resource: improvements and new features. Nucleic Acids Res 35:883-887

Penner GA, Clarke J, Bezte LJ, and Leisle D (1995) Identification of RAPD markers linked to a gene governing cadmium uptake in durum wheat. Genome 38:543-547

Popelka JC, Schubert S, Schulz R, Hansen AP (1996) Cadmium uptake and translocation during reproductive development of peanut (*Arachis hypogaea* L.). Angew Bot 70: 140–143

Poulsen TS (2004) Purification of BAC DNA. Methods Mol Biol 255:91-100

Pozniak CJ, Fox SL, Knott DR (2009) CDC Verona durum wheat. Can J Plant Sci 89:321-324

Pozniak CJ, Hucl PJ (2004) Genetic analysis of imidazolinone resistance in mutation-derived lines of common wheat. Crop Sci 44:23–30

Qin B, Cao A, Wang H, Chen T, You FM, Liu Y, Ji J, Liu D, Chen P, Wang X (2011) Collinearity-based marker mining for the fine mapping of *Pm6*, a powdery mildew resistance gene in wheat. Theor Appl Genet 123:207-218

Qiu J-W, Schürch AC, Yahiaoui N, Dong LL, Fan HJ, Zhang ZJ, Keller B, Ling H-Q (2007) Physical mapping and identification of a candidate for the leaf rust resistance gene *Lr1* of wheat. Theor Appl Genet 115:159-168

Ramakrishna W, Dubcovsky J, Park Y-J, Busso C, Emberton J, SanMiguel P, Bennetzen JL (2002) Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. Genetics 162:1389-1400

Reichman SM (2002) The responses of plants to metal toxicity: A review focusing on copper, manganese and zinc. The Australian Minerals and Energy Environment Foundation, Melbourne, Australia. Occasional Paper 14:1-54

Reimann C, Caritat P (1998) Chemical elements in the environments. Springer, Heidelberg

Reimer S, Pozniak CJ, Clarke FR, Clarke JM, Somers DJ, Knox RE, Singh AK (2008) Association mapping of the yellow pigment in an elite collection of durum wheat cultivars and breeding lines. Genome 51:1016-1025

Rustenholz C, Hedley PE, Morris J, Choulet F, Feuillet C, Waugh R, Paux E (2010) Specific patterns of gene space organization revealed in wheat by using the combination of barley and wheat genomic resources. BMC Genomics 2010 11:714-726

Saintenac C, Falque M, Martin OC, Paux E, Feuillet C, Sourdille P (2009) Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat (*Triticum aestivum* L.). Genetics 181:393-403

Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. Plant Cell 20:11-24

Salt DE, Prince RC, Pickering IJ, Raskin I (1995) Mechanisms of cadmium mobility and accumulation in Indian mustard. Plant Physiol 109: 1427–1433

Salt DE, Rauser WE (1995) MgATP-dependent transport of phytochelatins across the tonoplast of oat roots. Plant Physiol 107:1293-1301

Schuelke, M (2000) An economic method for the fluorescent labeling of PCR fragments. A poor man’s approach to genotyping for research and high-throughput diagnostics. Nat Biotechnol 16:233-235

Seah S, Bariana M, Jahier J, Sivasithamparam K, Lagudah ES (2001) The introgressed segment carrying rust resistance genes *Yr17*, *Lr37* and *Sr38* in wheat can be assayed by a cloned disease resistance gene-like sequence. Theor Appl Genet 102:600-605

Semagn K, Bjørnstad Å, Ndjiondjop MN (2006) Principles, requirements and prospects of genetic mapping in plants. African J Biotechnol 5:2569-2587

Shimo H, Ishimaru Y, An G, Yamakawa T, Nakanishi H, Nishizawa NK (2011) *Low cadmium* (*LCD*), a novel gene related to cadmium tolerance and accumulation in rice. J Exp Bot 62:5727-5734

Shute T, Macfie SM (2006) Cadmium and zinc accumulation in soybean: A threat to food safety? Sci of the Total Environ 37:63-73

Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat Biotechnol 23:75-81

Somers DJ, Banks T, DePauw, R., Fox S, Clarke J, Pozniak C, and McCartney C (2007) Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. Genome 50:557-567

Somers DJ, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (Triticum *aestivum* L.). Theor Appl Genet 109:1105-1114

Song WY, Sohn EJ, Martinoia E, Lee YJ, Yang YY, Jasinski M, Forestier C, Hwang I, Lee Y (2003) Engineering tolerance and accumulation of lead and cadmium in transgenic plants. Nat Biotechnol 21:914-919

Soric R, Loncaric Z, Kovacevic V, Brkic I, Simic D (2009) A major gene for leaf cadmium accumulation in maize (*Zea mays* L.). UC Davis: The Proceedings of the International Plant Nutrition Colloquim XVI.

Sorrells ME, LaRota M , Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin, Mahmoud A, Ma X, Gustafson PJ, et al (2003) Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 13:1818-1827

Stats Canada (2009) Market Outlook Report Wheat Outlook for 2009-10. Statistics Canada. November 6, 2009. ISSN 1920-20082X Vol. 1 No. 4 http://publications.gc.ca/collections/collection_2011/agr/A27-32-1-4-eng.pdf

Stats Canada (2011) Ceraeals and Oilseed Review January 2011. Statistics Canada. Feburary 9, 2011. 22-007-X, Vol. 34, No. 1 http://www.statcan.gc.ca/pub/22-007-x/22-007-x2011001-eng.pdf

Stein N, Feuillet C, Wicker T, Schlagenhauf E, Keller B (2000) Subgenome chromosome walking in wheat: a 450-kb physical contig in *Triticum monococcum* L. spans the Lr10 resistance locus in hexaploid wheat (*Triticum aestivum* L.). Proc Natl Acad Sci USA 97:13436-13441

Stockinger EJ, Skinner JS, Gardner KG, Francia E, Pecchioni N (2007) Expression levels of barley *Cbf* genes at the *Frost resistance-H2* locus are dependent upon alleles at *Fr-H1* and *Fr-H2*. Plant J 51:308-321

Stolt JP, Sneller FEC, Bryngelsson T, Lunborg T, Schat H (2003) Phytochelatin and cadmium accumulation in wheat. Environ Exp Bot 49:21-28

Sutton T (2012) Cloning the Chromosome 7BL Boron Toxicity Tolerance Locus *Bo1* and Analysis of Allelic Variation in Bread and Durum Wheat. International Plant and Animal Genome Conference. San Diego, CA. Jan 14-18, 2012

Suwazono Y, Sand S, Vahter M, Filipsson AF, Skerfving S, Lidfeldt J, Akesson A (2006) Benchmark dose for cadmium-induced renal effects in humans. Environ Health Perspect 114:1072-1076

Szira F, Ferenc Bálint A, Vágújfalvi A, Galiba G (2008) Are *Cbf* genes involved in copper tolerance? Acta Biologica Szegediensis 52:205-207

Takahashi M, Nozoye T, Kitajima N, Fukuda N, Hokura A, Terada Y, et al. (2009) In vivo analysis of metal distribution and expression of metal transporters in rice seed during germination process by microarray and X-ray fluorescence imaging of Fe, Zn, Mn, and Cu. Plant Soil 325:39–51

Takahashi R, Ishimaru Y, Senoura T, Shimo H, Ishikawa S, Arao T, Nakanishi H, Nishizawa NK (2011) The OsNRAMP1 iron transporter is involved in Cd accumulation in rice. J Exp Bot advanced access doi:10.1093/jxb/err136

Tanhuanpää P, Kalendar R, Schulman AH, Kiviharju E (2007) A major gene for grain cadmium accumulation in oat (*Avena sativa* L.). Genome 50:588-594

Tezuka K, Miyadate H, Katou K, Kodama I, Matsumoto S, Kawamoto T, Masaki S, Satoh H, Yamaguchi M, Sakurai K, et al. (2009) A single recessive gene controls cadmium translocation in the cadmium hyperaccumulating rice cultivar Cho-Ko-Koku. Theor Appl Genet 120: 1175-1182

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876-4882

Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J (2006) A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science 314:5803-1301

Ueno D, Koyama E, Kono I, Ando T, Yano M, Ma JF (2009) Identification of a novel major quantitative trait locus controlling distribution of Cd between roots and shoots in rice. Plant Cell Physiol 50:2223-2233

Ueno D, Koyama E, Yamaji N, Ma JF (2011) Physiological, genetic, and molecular characterization of a high-Cd-accumulating rice cultivar, Jarjan. J Exp Bot 62: 2265–2272

Ueno D, Yamaji N, Kono I, Huang CF, Ando T, Yano M, Ma JF (2010) Gene limiting cadmium accumulation in rice. Proc Natl Acad Sci USA 107: 16500-16505

Uraguchi S, Mori S, Kuramata M, Kawasaki A, Arao T, Ishikawa S (2009) Root-to-shoot Cd translocation via the xylem is the major process determining shoot and grain cadmium accumulation in rice. J Exp Bot 60:2677-2688

U.S. Department of Health and Human Service, Public Health Service Agency for Toxic Substances and Disease Registry (1999) Toxicology profile for cadmium http://labourandemployment.gov.in/idmis/tox_profile/cadmium.pdf

Valárik M, Linkiewicz A, Dubcovsky J (2006) A microcolinearity study at the earliness per se gene *Eps-A^{m}1* region reveals an ancient duplication that preceded the wheat–rice divergence. Theor Appl Genet 112:945-957

van Ooijen JW, Voorips RE (2004) JoinMap Version 4.0, Software for the calculation of genetic linkage maps. Kyazma BV, Wageningen.

van Ooijen JW (1999) LOD significance thresholds for QTL analysis in experimental populations of diploid species. Heredity 83:613-624

van Ooijen JW (2004) MapQTL 5, Software for the mapping of quantitative trait loci in experimental populations. Kyazma BV, Wageningen

Verrier PJ, Bird D, Burla B, Dassa E, Forestier C, Geisler M, Klein M, Kolukisaoglu U, Lee Y, Martinoia E, Murphy A, Rea PA, Samuels L, Schulz B, Spalding EJ, Yazaki K, Theodoulou FL (2008) Plant ABC proteins—A unified nomenclature and updated inventory. Trends Plant Sci. 13:151-159

Vögeli-Lange R, Wagner GJ (1990) Subcellular localization of cadmium and cadmium-binding peptides in tobacco leaves. Implication of a transport function for cadmium-binding peptides. Plant Physiol 92:1086-1093

Vögeli-Lange R, Wagner GJ (1996) Relationship between cadmium, glutathione and cadmium-binding peptides (phytochelatins) in leaves of intact tobacco seedlings. Plant Sci 114:11-18

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93: 77–78

Wang G, Sun X, Wang G, Wang F, Gao Q, Sun X, Tang Y, Chang C, Lai J, Zhu L, Xu Z, Song R (2011) *Opaque7* encodes an acyl activating enzyme-like protein that affects storage protein synthesis in maize endosperm. Genet advanced access 10.1534/genetics.111.133967

Wei F S, Wing R A, Wise R P (2002) Genome Dynamics and evolution of the *Mla* (Powdery Mildew) resistance locus in barley. Plant Cell 14:1903-1917

Wicker T, Buchmann JP, Keller B (2010) Patching gaps in plant genomes results in gene movement and erosion of colinearity. Genome Res 20:1229-1237

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E et al. (2001) Analysis of a contigous 211kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. The Plant J 26:307-316

Wiebe K, Pozniak CJ, Harris N, Faris JD, Clarke JM, Knox RE, Taylor G (2010) Targeted mapping of *Cdu1,* a major locus regulating grain cadmium concentration in durum wheat (*Triticum turgidum* L. var. *durum).* Theor Appl Genet 121:1047-1058

Williams LE, Mills RF (2005) P_{1B}-ATPases—an ancient family of transition metal pumps with diverse functions in plants. Trends Plant Sci 10:491-502

Wojas S, Hennig J, Plaza S, Geisler M, Siemianowski O, Sklodowska A, Ruszczynska A, Bulska E, Antosiewicz DM (2009) Ectopic expression of *Arabidopsis* ABC transporter MRP7 modifies cadmium root-to-shoot transport and accumulation. Environ Pollut 157:2781-2789

Yan L, Fu D, Blech A, Tranquilli G, Bonafed M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J (2006) The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. Proc Natl Acad Sci USA 103:19581-19586

Yan L, Loukoianov A, Blecht A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J (2004) The wheat *VRN2* gene is a flowering repressor down-regulated by vernalisation. Science 303:1640-1644

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of wheat vernalization gene *VRN1*. Proc Natl Acad Sci USA 100:6263-6268

Yan SH, Tsay CC, Chen YR (2000) Isolation and characterization of phytochelatin synthase in rice seedlings. Proc Natl Counc ROC (B) 24:202-207

Yazaki K, Yamanaka N, Masuno T, Konagai S, Shitan N, Kaneko S, Ueda K, Sato F (2006) Heterologous expression of a mammalian ABC transporter in plant and its application to phytoremediation. Plant Mol Biol 61:491-503

Zook EG, Greene FE, and Morris ER (1970) Nutrient composition of selected wheats and products. VI. Distribution of manganese, copper, nickel, zinc, magnesium, lead, tin, cadmium, chromium, and selenium as determined by atomic absorption spectroscopy and colorimetry. Cereal Chem 47:720-731

.0 Appendices

Appendix 1 ESM, STS, and gene specific markers polymorphic between W9262-260D3 and Kofa

Marker name	Detection method[a]	Map location	Primer sequence[b] (3'-5')	Rice gene
ESMs				
XBE425993 [c]	SSCP	5A	AAGACATCCTGAACCTGGTGTA GTCCCAGTCGAACTTGTTCAT	Os03g55070
XBE426348 [c]	SSCP	5A	CTATAAGATGAACCGGGGTTTT TACGCTACCTATGAAGTACTTGGAC	Os03g53800
XBE604920 [c]	Agarose	5A	TCCCCTACATGCTGCTCTAC CAACATCGACTTCATTATTGGAC	Os03g52860
XBF474090.1	SSCP	5B	GAGGCCATGGACCCCAACTTT GGACAGGAGAACCTGAAGGAT	Os03g53670
XBF474090.2 [c]	SSCP	5A	GTAGATTATTGGCAACAAGACAAGT GCGTAAGAAATATATCACGCTAGTT	Os03g53670
XBF474164.1/.2 [c]	SSCP	5A/5B	AGACTTTCTCGTCCCGATACTT CAACATATGTCTGGCCTACTACTCT	Os03g53720
XBG262450 [c]	SSCP	5A	GATAATTTCAGAACAATGCCATTAC AAGAGTAGCCAATCTGTAGTTGATG	Os03g51020
XBG274700 [c]	SSCP	5A	CAGAAGACAGTGAAGAACCAAAAC AACTCTCAAGTCACTCATCTCAATC	Os03g55950
XBG313229 [c]	CE	5B	CTTGCTGTCCTCGAGAAGTTT ATAGTATCCCATCAATTGTAAGCTG	Os03g58470
XBG607162	SSCP	5B	ATGCATACAAGGACCGCTAC AATCACACCCTTGCGAATAAT	Os03g63140
XBF293297.2/.1	SSCP	5A/5B	TGGCCGCGCCCTTCTTCTCCA TTGTCCTGCGGCTTCACCATC	Os03g53600
XBF145263	SSCP	5A	ACGTGGACGACTACTTGGAGT CAGGTCATAAGCTTGGCGTGC	Os03g53700
STS markers				
Xusw15	SSCP	5B	ACCAGCAGGACATTGGGAACA GAACCTTGGACGATTGCTAAC	Os03g53590
Xusw49	HRM	5B	CACCGAGCTGTCCTAATGAAG CTGCAGAAGTACTCTGGATCC	Os03g53250
Xusw50	HRM	5B	TTCAGTGATAACTTACACCAG AGCTTCTTGCGTTCTTCCATC	Os03g53490

Appendix 1 (con't).

Marker name	Detection method[a]	Map location	Primer sequence[b] (3'-5')	Rice gene
STS markers (con't)				
Xusw51	HRM	5B	ATGGTTGGCTGTAGAACAAGG CTCACGCCGTGAGAACGTTAC	Os03g53500
Xusw52	HRM	5B	TTCATTGTCAGATGATTCTGG CTTCCAGATCTTCACAAGCTT	Os03g53530
Xusw53	HRM	5B	GATGAACCGCATATCCTTCCT CTCATTGTCACAAGCAATCAT	Os03g53700
Xusw61	HRM	5A	TGGAGACGGTGGCCGCCGGTG GGGCATGAACGCCATCACTA	Os03g53350
Genes associated with Vrn-B1				
Xwg644	SSCP	5A	GACTTGTTCAGTCATCTCATA GCAGCTTGTGTCTGATGTGAA	Os03g54790
Xwg644	SSCP	5B	GCTCTTAAGCAGGCTTTCTGA CTGTAAGGCTGTATAAGATGA	Os03g54790
MC	SSCP	5A	AGTCGGTGTTCAAGCAACAGG GCGATCAATCTTCTAACTACC	Os03g54760
CSFs	SSCP	5B	TCGGCACCAATGCCGTGGATT AGAACTTAATGGATGTGTCCC	Os03g54770
CSFs	SSCP	5A	CCAGTAGCTCATCTCTATGAT ACTCGTAGCTTCTACAGATCC	Os03g54770
PCS2	CE	5B	TCAACTACCAGCAGTTCCGAC GTAGGCCTGCCAACAAGAGCA	Os03g54750
PHY-C	SSCP	5B	ACTGGAAGCAGGCTATCCTGG AACATAGTCGCCTTGTATCCG	Os03g54084
MTK4	SSCP	5A	CGTGGTGGAACAGGACGAGGG CATCATTCCCAGGTAGAACAC	Os03g53880
P1B-ATPases gene (HMA)				
HMA3-B2	SSCP	2B	AGTGATGCTTACTGGCGATAG ACAAGCATGTGCCAACATCAG	Os07g12900
HMA3-B1	CE	5B	TTCTTGCTGTTCATCCGCCTG AATACGGGACTGCGAGACGGC	Os07g12900

[a]SSCP=single strand conformational polymorphism gel electrophoresis; CE=capillary electrophoresis; HRM=high resolution melt

[b]M13 (CACGACGTTGTAAAACGAC) tag attached to 5' end of forward primer (listed first for each primer pair).

[c] – at least one of markers primers was designed by Dr. Justin Faris

Appendix 2 Gene products in colinear region of *Cdu1* on rice chromosome 3 and *Brachypodium* chromosome 1.

gene name	gene product name	gene name	gene product name
BF293297	HTH DNA-binding protein, putative, expressed	Os03g53430	retrotransposon protein, putative, unclassified
BF474090	YT521-B-like family domain containing protein, expressed	Os03g53440	hypothetical protein
BF474164	SRPK4, putative, expressed	Os03g53450	hypothetical protein
BF483771	bifunctional 3-phosphoadenosine 5-phosphosulfate synthetase, putative, expressed	Os03g53470	hypothetical protein
BG608197	OsCML4 - Calmodulin-related calcium sensor protein, expressed	Os03g53480	hypothetical protein
Bradi1g08860.1	Core histone H2A/H2B/H3/H4 domain containing protein, putative, expressed	Os03g53490	PPR repeat containing protein, expressed
Bradi1g08940.1	tetratricopeptide repeat domain containing protein, expressed	Os03g53500	helicase conserved C-terminal domain containing protein, expressed
Bradi1g08950.1	ubiquitin carboxyl-terminal hydrolase, family 1, putative	Os03g53510	WD domain, G-beta repeat domain containing protein, expressed
Bradi1g08960.1	acanthoscurrin-1 precursor, putative, expressed	Os03g53520	expressed protein
Bradi1g08970.1	conserved hypothetical protein	Os03g53530	WD domain, G-beta repeat domain containing protein, expressed
Bradi1g09000.1	lectin-like receptor kinase 7, putative, expressed	Os03g53540	expressed protein
Bradi1g09010.1	expressed protein	Os03g53550	retrotransposon protein, putative, unclassified, expressed
Bradi1g09020.1	hypothetical protein	Os03g53560	retrotransposon protein, putative, Ty1-copia subclass
Os03g53210	expressed protein	Os03g53570	retrotransposon protein, putative, unclassified
Os03g53220	U5 small nuclear ribonucleoprotein 200 kDa helicase, putative	Os03g53580	expressed protein
Os03g53240	hypothetical protein	Os03g53590	expressed protein
Os03g53250	expressed protein	Os03g53590	expressed protein
Os03g53260	conserved hypothetical protein	Os03g53610	late embryogenesis abundant protein D-34, putative

Appendix 2 (con't)

gene name	gene product name	gene name	gene product name
Os03g53270	stem-specific protein TSJT1, putative, expressed	Os03g53620	late embryogenesis abundant protein D-34, putative, expressed
Os03g53280	WD domain containing protein, putative, expressed	Os03g53630	PHD finger family protein, putative, expressed
Os03g53300	expressed protein	Os03g53640	expressed protein
Os03g53310	emp24/gp25L/p24 family protein, putative, expressed	Os03g53650	cysteine synthase, putative, expressed
Os03g53320	hypothetical protein	Os03g53660	Myosin head domain containing protein, expressed
Os03g53340	HSF-type DNA-binding domain containing protein, expressed	Os03g53680	hypothetical protein
Os03g53350	anthocyanin 3-O-beta-glucosyltransferase, putative	Os03g53690	oxidoreductase, short chain dehydrogenase/reductase family domain containing protein, expressed
Os03g53360	transferase family protein, putative, expressed	Os03g53700	PHD-finger domain containing protein, putative, expressed
Os03g53380	hypothetical protein	Os03g53710	aldose 1-epimerase, putative, expressed
Os03g53390	expressed protein	Os03g55350	OsSub31 - Putative Subtilisin homologue, expressed
Os03g53400	transmembrane BAX inhibitor motif-containing protein, putative, expressed	Os03g56160	lectin-like receptor kinase 7, putative, expressed
Os03g53410	protein kinase domain containing protein, expressed		

Appendix 3 SNP identified to design CAPS markers

usw14

BsoBI (CYCGRG)

usw15b

TaqI (TCGA)

usw47

Hpy188I (TCNGA)

usw16

SbfI (CCTGCAGG)

Appendix 4 Additional *Cdu-B1* mapping markers and *HMA* sequencing markers, primer sequences, restriction enzymes, detection method, and ESM, STS marker, or Rice gene from which the new marker was developed

Additional *Cdu-B1* Mapping Markers

CAPS Marker	Forward Primer sequence	Reverse Primer Sequence	Restriction Enzyme	ESM or STS marker developed from
Xusw14	TACAGCCGCTCAGTTGCTC	CAACATATGTCTGGCCTACTACTCT	BsoB1	*XBF474164*
Xusw17	TCCACCCCCTTCCATCCCTAT	TTGTCCTGCGGCTTCACCATC	Sbf1	*XBF293297*
Xusw15b	TATGTGTTGTGATTTGCTGAG	GAACCTTGGACGATTGCTAAC	Taq1	*Xusw15*
Xusw16	GCTAGGACTTGATTCATTGAT	GGACAGGAGAACCTGAAGGAT	Hpy188I &MboII	*XBF474090*
Xusw47	GCTAGGACTTGATTCATTGAT	AGTGATCTAAACGTTCTTATA	Hpy188I	*XBF474090*
Other Marker	**Forward Primer sequence**	**Reverse Primer Sequence**	**Detection Method**	**Marker developed from**
ScOPC20	ACTTCGCCACTCCAGATGTACT	ACTTCGCCACCATGGTCACA	Agarose	--

Additional Markers for HMA3 Cloning for Sequencing

HMA Marker	Forward Primer sequence	Reverse Primer Sequence	Rice gene developed from
HMA3-F1/R1	ATGGGCGGCGGCGAGTCGTAC	GTGGTGAAGAGGAAGACGATG	Os07g12900
HMA3-F2/R2	GACATCAACATCCTCATGCTT	CCATTGTCCTCACGGCGATGT	Os07g12900
HMA3-F3/R3	ACATCGCCGTGAGGACAATGG	TTTGCTCTCGATGCTTGAGAT	Os07g12900
HMA3-F4/R4	ATCTCAAGCATCGAGAGCAAA	TGAGGATGTCGCTGGACATGA	Os07g12900
HMA3-F5/R5	TCATGTCCAGCGACATCCTCA	GCCGACACGCAGCTCGATGAA	Os07g12900
HMA3-F6/R6	CGTGCTCAACAGCATGCTGCT	AAGATCGAACGGCCATTCTTC	Os07g12900

Appendix 5 Graphical genotype of *Cdu-B1* and markers associated with *Vrn-B1* in near isogenic parental lines, Kofa and W9262-260D3

	***Cdu-B1* locus markers**					***Vrn-B1* locus markers**			
Durum Line	ScOPc20	*Xusw17*	*Xusw47*	*Xusw15b*	*Xusw14*	CSFs-1	PCS2	*Xwg644*	PHY-C
Kofa	D	A	A	A	A	A	A	A	A
W9262-260D3	B	B	B	B	B	B	B	B	B
8982-TL-H	D	A	A	A	A	A	A	A	A
8982-TL-L	B	B	B	B	B	A	A	A	A

Appendix 6 Alignment of *HMA3* genomic sequence and predicted CDS sequence and reported genomic and CDS sequences from rice and *Brachypodium*. Nucleotides color coded based on similarity: yellow – identical, blue – conserved, white – non-similar.

```
Section 1
                             (1) 1        10        20        30        40        50        60             76
OsHMA3_genomic               (1) ATGGCCGGAAAGGATGAGGCGGAAGGGCTCGAGGCGAGGCTGCTGCTGCCGCCTGAGGCGGCGGCGGAGGAGC
OsHMA3_CDS                   (1) ATGGCCGGAAAGGATGAGGCGGAAGGGCTCGAGGCGAGGCTGCTGCTGCCGCCTGAGGCGGCGGCGGAGGAGC
BdHMA3_genomic               (1) -------------------------------------------------------ATGAG-TGGCGGCG-AGC-TC
BdHMA3_CDS                   (1) -------------------------------------------------------ATGAG-TGGCGGCG-AGC-TC
Langdon_HMA3-A1_genomic      (1) -------------------------------------------------------ATGGG-CGGCGGCG-AG--TC
Langdon_HMA3-A1_CDS          (1) -------------------------------------------------------ATGGG-CGGCGGCG-AG--TC
Langdon_TtHMA3-B1_genomic    (1) -------------------------------------------------------ATGGG-CGGCGGCG-AG--TC
Langdon_TtHMA3-B1_CDS        (1) -------------------------------------------------------ATGGG-CGGCGGCG-AG--TC
Partial_CS_HMA3-D1_genomic   (1) -------------------------------------------------------ATGGG-CGGCGGCG-AGC--C

Section 2
                            (77) 77          90        100       110       120       130       140         152
OsHMA3_genomic              (77) CGACGAGGT---------------GTGGCGGCGGCGA--CGGCGGCG---------G-CGGCG-GGAGGAAGCGGA
OsHMA3_CDS                  (77) CGACGAGGT---------------GTGGCGGCGGCGA--CGGCGGCG---------G-CGGCG-GGAGGAAGCGGA
BdHMA3_genomic              (19) CGTCGGGGAGCG---------------CTTGAGGAGAGCCTGCTGCT---------GACGGGGAGGACGCAGTGGG
BdHMA3_CDS                  (19) CGTCGGGGAGCG---------------CTTGAGGAGAGCCTGCTGCT---------GACGGGGAGGACGCAGTGGG
Langdon_HMA3-A1_genomic     (18) GTACCCGGCG-----------------CTCGAGGCGAGCCTGCTGGCAGAGGAAGCGGCGGCGAGGAGGCAGTGGG
Langdon_HMA3-A1_CDS         (18) GTACCCGGCG-----------------CTCGAGGCGAGCCTGCTGGCAGAGGAAGCGGCGGCGAGGAGGCAGTGGG
Langdon_TtHMA3-B1_genomic   (18) GTACGCGGCGGCGAGTCGTACGCGGCGCTCGAGGAGAGCCTTCTGCCAGAGCAAGCGGCGGCGAGGAGGCAGTGGG
Langdon_TtHMA3-B1_CDS       (18) GTACGCGGCGGCGAGTCGTACGCGGCGCTCGAGGAGAGCCTTCTGCCAGAGCAAGCGGCGGCGAGGAGGCAGTGGG
Partial_CS_HMA3-D1_genomic  (18) GTACACGGCG-----------------CTCGAGGAGAGCCTTCTGTCCGGGGAAGCGGCGGCGAGGAGGCAGTGGG

Section 3
                           (153) 153    160       170       180       190       200       210            228
OsHMA3_genomic             (125) AGAAGACGTACCTTGATGTCCTTGGCGTGTGCTGCTCCGCGGAGGTCGCGCTGGTGGAGCGGCTGCTCGCGCCGCT
OsHMA3_CDS                 (125) AGAAGACGTACCTTGATGTCCTTGGCGTGTGCTGCTCCGCGGAGGTCGCGCTGGTGGAGCGGCTGCTCGCGCCGCT
BdHMA3_genomic              (71) ACAAGACATACCTGGACGTGCTGGGCGTGTGCTGCTCGGCGGAGGTCGCGCTCGTGGAGAGGCTCCTCTCGCCGCT
BdHMA3_CDS                  (71) ACAAGACATACCTGGACGTGCTGGGCGTGTGCTGCTCGGCGGAGGTCGCGCTCGTGGAGAGGCTCCTCTCGCCGCT
Langdon_HMA3-A1_genomic     (77) AGAAGACGTACCTGGACGTGCTGGGCGTGTGCTGCTCCGCGGAGGTCGCGCTCGTCGAGCGGCTGCTGGCGCCGCT
Langdon_HMA3-A1_CDS         (77) AGAAGACGTACCTGGACGTGCTGGGCGTGTGCTGCTCCGCGGAGGTCGCGCTCGTCGAGCGGCTGCTGGCGCCGCT
Langdon_TtHMA3-B1_genomic   (94) AGAAGACGTACCTGGACGTGCTGGGCGTGTGCTGCTCCGCGGAGGTCGCGCTGGTCGAGCGGCTGCTGGCGCCGCT
Langdon_TtHMA3-B1_CDS       (94) AGAAGACGTACCTGGACGTGCTGGGCGTGTGCTGCTCCGCGGAGGTCGCGCTGGTCGAGCGGCTGCTGGCGCCGCT
Partial_CS_HMA3-D1_genomic  (77) AGAAGACGTACCTGGACGTGCTGGGCGTGTGCTGCTCGGCGGAGGTCGCGCTCGTCGAGCGGCTGCTGGCGCCGCT

Section 4
                           (229) 229        240       250       260       270       280       290         304
OsHMA3_genomic             (201) CGACGGCGTCCGCGTGGTGTCCGTCGTCGTGGCGTCCCGCACCGTCGTCGTCGAGCACGACCCCGCCGCCGCCCCG
OsHMA3_CDS                 (201) CGACGGCGTCCGCGTGGTGTCCGTCGTCGTGGCGTCCCGCACCGTCGTCGTCGAGCACGACCCCGCCGCCGCCCCG
BdHMA3_genomic             (147) CGCCGGCGTGAGGACGGTGTCCGTCGTCGTCCCCTCCAGGACCGTTGTCGTCCAGCACGACCCCGCCGCCGTCTCC
BdHMA3_CDS                 (147) CGCCGGCGTGAGGACGGTGTCCGTCGTCGTCCCCTCCAGGACCGTTGTCGTCCAGCACGACCCCGCCGCCGTCTCC
Langdon_HMA3-A1_genomic    (153) CGACGGCGTGAGGGCGGTGGCCGTCGTCGTCCCCTCCCGCACCGTCGTCGTCGAGCACGACCCCGCCGCCGTCTCG
Langdon_HMA3-A1_CDS        (153) CGACGGCGTGAGGGCGGTGGCCGTCGTCGTCCCCTCCCGCACCGTCGTCGTCGAGCACGACCCCGCCGCCGTCTCG
Langdon_TtHMA3-B1_genomic  (170) CGACGGCGTGAGGGCGGTGTCCGTCGTCGTCCCCTCCCGCACCGTCGTCGTCGAGCACGACCCCGCCGCCGTCTCG
Langdon_TtHMA3-B1_CDS      (170) CGACGGCGTGAGGGCGGTGTCCGTCGTCGTCCCCTCCCGCACCGTCGTCGTCGAGCACGACCCCGCCGCCGTCTCG
Partial_CS_HMA3-D1_genomic (153) CGACGGCGTGAGGGCGGTGTCCGTCGTCGTCCCCTCCCGCACCGTCGTCGTCGAGCACGACCCCGCCGCCGTCTCT

Section 5
                           (305) 305  310        320       330       340       350       360       370     380
OsHMA3_genomic             (277) GAGTCCGCCATCGGTAAGCCAC------------CGCGCGCGCGCGCGCCATTAATCCT---------CCACGCGA
OsHMA3_CDS                 (277) GAGTCCGCCATCG---------------------------------------------------------------
BdHMA3_genomic             (223) CAGTCCCACATCGGTACTAAAA------------CTC---------CCTCTTTTACTTT---------CAAGCCTA
BdHMA3_CDS                 (223) CAGTCCCACATCG---------------------------------------------------------------
Langdon_HMA3-A1_genomic    (229) CAGTCCCGTATTGGTAACCAATCA----ATCCGTCCCCTCTCCTGACCTTACTTACCCTGATCTGAACCCAAAATC
Langdon_HMA3-A1_CDS        (229) CAGTCCCGTATTG---------------------------------------------------------------
Langdon_TtHMA3-B1_genomic  (246) CAGTCCCGTATCGGTACTAACCAATAAATCCATCCCCTCTCCTGATCTTTACTTGCTCTGATCTGAGCCCAGAATC
Langdon_TtHMA3-B1_CDS      (246) CAGTCCCGTAT-----------------------------------------------------------------
Partial_CS_HMA3-D1_genomic (229) CAGTCCCGTATTGGTAACCAATCAA--------TCCGTCCCCTCTCCTGATCTTACTCTAATCTGAACCCAAAATC

Section 6
                           (381) 381      390       400       410       420       430       440            456
OsHMA3_genomic             (332) TTCTCGCGCGAGATTCTCACTGAATCGAGTCGTCGTCGTCTTTGTCTTGGAGTG-----------------CAGTG
OsHMA3_CDS                 (290) --------------------------------------------------------------------------TG
BdHMA3_genomic             (269) AACCAACAAAATTCTAACGTTGCTGTGAACTGGCG-------CTTTGCGCCGGCCGG--------------CAGTG
BdHMA3_CDS                 (236) --------------------------------------------------------------------------TG
Langdon_HMA3-A1_genomic    (301) TCCATGAAATAAAGCTTAATTACAAGTCGATCTCAACGTGTTTGTTGCCTTGTCCGGACTCCGAAACGCACCAGTG
Langdon_HMA3-A1_CDS        (242) --------------------------------------------------------------------------TG
Langdon_TtHMA3-B1_genomic  (322) TCCATGAAAGAAAGCTTAATTACAAGTCGATCACAACGTGTTTGTTGCCTTGTCCGGACTCCGGAACGCATCAGTG
Langdon_TtHMA3-B1_CDS      (257) -------------------------------------------------------------------------GTG
Partial_CS_HMA3-D1_genomic (297) TCCATGAAAGAAAGC----TTACAAGTCGATCTCAACGTCTTTGTTGCCTTGTCCGGACTCCGGAACGCATCAGTC

Section 7
                           (457) 457         470       480       490       500       510       520         532
OsHMA3_genomic             (394) AAGGCGCTGAACAAGGCCGGCCTCGAGGCGTCGGTGCGAGCCTACGGCAGCAGCGGCGTGGTCAGCCGGTGGCCGA
OsHMA3_CDS                 (292) AAGGCGCTGAACAAGGCCGGCCTCGAGGCGTCGGTGCGAGCCTACGGCAGCAGCGGCGTGGTCAGCCGGTGGCCGA
BdHMA3_genomic             (324) AAGGTTCTGAACGGGGCAGGCCTGGAGGCGTCGGTGCGAGCCTATGGGAGCAGCGGGATCATCAGCCGGTGGCCCA
BdHMA3_CDS                 (238) AAGGTTCTGAACGGGGCAGGCCTGGAGGCGTCGGTGCGAGCCTATGGGAGCAGCGGGATCATCAGCCGGTGGCCCA
Langdon_HMA3-A1_genomic    (377) AAGGTCCTCAACGGGGCGGGCCTGGAAGCCTCGGTGCGAGCGTACGGCAGCAGCGGGTTCATCGGCCGACGGCCCA
Langdon_HMA3-A1_CDS        (244) AAGGTCCTCAACGGGGCGGGCCTGGAAGCCTCGGTGCGAGCGTACGGCAGCAGCGGGTTCATCGGCCGACGGCCCA
Langdon_TtHMA3-B1_genomic  (398) AAGGTCCTCAACGGGGCGGGCCTGGAAGCCTCGGTGCGAGCCTACGGCAGCAGCGGGGTCATCGGCCGATGGCCCA
Langdon_TtHMA3-B1_CDS      (260) AAGGTCCTCAACGGGGCGGGCCTGGAAGCCTCGGTGCGAGCCTACGGCAGCAGCGGGGTCATCGGCCGATGGCCCA
Partial_CS_HMA3-D1_genomic (369) AAGGTCCTGAACGGGGCGGGCCTGGAGGCCTCGGTGCGAGCCTACGGCAGCAGCGGGGTCATCGGCCGATGGCCCA

Section 8
                           (533) 533         540       550       560       570       580       590         608
OsHMA3_genomic             (470) GCCCGTACATCGTCGCCTCCGGCGTCCTCCTCACGGCGTCCTTCTTCGAGTGGCTCTTCCCT---CCCCTGCA---
OsHMA3_CDS                 (368) GCCCGTACATCGTCGCCTCCGGCGTCCTCCTCACGGCGTCCTTCTTCGAGTGGCTCTTCCCT---CCCCTGCA---
BdHMA3_genomic             (400) GCCCGTACATCCTCGCCTCCGGCGTCCTCCTCCTGGCGTCCTCCCTGAGCTGGCTCCTCCCTGATCCCCTGCATCA
BdHMA3_CDS                 (314) GCCCGTACATCCTCGCCTCCGGCGTCCTCCTCCTGGCGTCCTCCCTGAGCTGGCTCCTCCCTGATCCCCTGCATCA
Langdon_HMA3-A1_genomic    (453) GCCCGTACATCGTCGCCTGCGGCGCCCTCCTCCTCGCGTCCTCCTTCAGGTGGCTCCTGCTT---CCCCTGCA---
Langdon_HMA3-A1_CDS        (320) GCCCGTACATCGTCGCCTGCGGCGCCCTCCTCCTCGCGTCCTCCTTCAGGTGGCTCCTGCTT---CCCCTGCA---
Langdon_TtHMA3-B1_genomic  (474) GCCCGTACATCGTCGCCTGCGGCGTCCTCCTCCTCGCGTCCTCCTTCAGGTGGCTCCTGCCT---CCCCTGCA---
Langdon_TtHMA3-B1_CDS      (336) GCCCGTACATCGTCGCCTGCGGCGTCCTCCTCCTCGCGTCCTCCTTCAGGTGGCTCCTGCCT---CCCCTGCA---
Partial_CS_HMA3-D1_genomic (445) GCCCGTACATCGTCGCCTGCGGCGCCCTCCTCCTCGCGTCTTCCTTCAGGTGGCTCCTGCCT---CCCCTGCA---

Section 9
                           (609) 609         620       630       640       650       660       670         684
OsHMA3_genomic             (540) ------GTGCCTGGCCGTGGCGGCCGTCGTCGCCGGCGCGCCGCCGATGGTGCGCCGTGGGTTCGCCGCGGCGAGC
OsHMA3_CDS                 (438) ------GTGCCTGGCCGTGGCGGCCGTCGTCGCCGGCGCGCCGCCGATGGTGCGCCGTGGGTTCGCCGCGGCGAGC
BdHMA3_genomic             (476) TCTGCAGTGGCTGGCCCTGGCCGCGGCCTTCGTCGGCGCTCCCCCGATGCTGCTCCGGGGCTTCGCGGCCGCCAGC
BdHMA3_CDS                 (390) TCTGCAGTGGCTGGCCCTGGCCGCGGCCTTCGTCGGCGCTCCCCCGATGCTGCTCCGGGGCTTCGCGGCCGCCAGC
Langdon_HMA3-A1_genomic    (523) ------GTGGCTGGCCCTGGGGGCGGCCTGCGCCGGCGCTCCCCCGATGGTCCTCCGAGGGCTCGCCGCCGCCAGC
Langdon_HMA3-A1_CDS        (390) ------GTGGCTGGCCCTGGGGGCGGCCTGCGCCGGCGCTCCCCCGATGGTCCTCCGAGGGCTCGCCGCCGCCAGC
Langdon_TtHMA3-B1_genomic  (544) ------GTGGCTGGCCCTGGCGGCGGCCTGCGCCGGCGCTCCTCCGATGCTTCTCCGTGGCATCGCCGCCGCCAGC
Langdon_TtHMA3-B1_CDS      (406) ------GTGGCTGGCCCTGGCGGCGGCCTGCGCCGGCGCTCCTCCGATGCTTCTCCGTGGCATCGCCGCCGCCAGC
Partial_CS_HMA3-D1_genomic (515) ------GTGGCTGGCCCTGGGGGCGGCCTGCGCCGGCGCTCCCCCGATGCTTCTCCGAGGGCTCGCCGCCGCCAGC
```

```
                                                                                                   Section 10
                            (685) 685  690        700       710       720       730       740       750        760
           OsHMA3_genomic   (610) CGGCTGTCGCTCGACATCAACGTCCTCATGCTCATCGCCGTCGCCGGCGCGCTCTGCCTCGGCGACTACACGGAGG
              OsHMA3_CDS    (508) CGGCTGTCGCTCGACATCAACGTCCTCATGCTCATCGCCGTCGCCGGCGCGCTCTGCCTCGGCGACTACACGGAGG
           BdHMA3_genomic   (552) AGGCTCACGCTGGACATCAACGTCCTGATGCTCATCGCCGTCGCCGGCGCCGCCGCTCTCGGGGACTACACTGAGG
              BdHMA3_CDS    (466) AGGCTCACGCTGGACATCAACGTCCTGATGCTCATCGCCGTCGCCGGCGCCGCCGCTCTCGGGGACTACACTGAGG
   Langdon_HMA3-A1_genomic  (593) AGGCTCGCGCTGGACATCAACATCCTCATGCTTATCGCCGTCGCCGGTGCCGTCGCCCTCAAGGACTACACGGAGG
      Langdon_HMA3-A1_CDS   (460) AGGCTCGCGCTGGACATCAACATCCTCATGCTTATCGCCGTCGCCGGTGCCGTCGCCCTCAAGGACTACACGGAGG
 Langdon_TtHMA3-B1_genomic  (614) AGGCTGACGCTGGACATCAACATCCTCATGCTTATCGCCGTTGCCGGTGCCGTCGCCCTCAAGGACTACACGGAGG
    Langdon_TtHMA3-B1_CDS   (476) AGGCTGACGCTGGACATCAACATCCTCATGCTTATCGCCGTTGCCGGTGCCGTCGCCCTCAAGGACTACACGGAGG
 Partial_CS_HMA3-D1_genomic (585) AGGCTGACGCTGGACATCAACATCCTCATGCTTATCGCCGTTGCCGGTGCGGTTGCGCTCAAGGACTACGCGGAGG
                                                                                                   Section 11
                            (761) 761        770       780       790       800       810       820             836
           OsHMA3_genomic   (686) CCGGCGCCATCGTCTTCCTCTTCACCACCGCAGAATGGCTCGAGACGCTGGCCTGCACAAAGGTAGGA---GTACA
              OsHMA3_CDS    (584) CCGGCGCCATCGTCTTCCTCTTCACCACCGCAGAATGGCTCGAGACGCTGGCCTGCACAAAGG------------
           BdHMA3_genomic   (628) CCGGCGCCATCGTCTTCCTCTTCACCACCGCCGAATGGCTCGAGACCCTGGCTTGCACCAAGGTAATC---AAGAA
              BdHMA3_CDS    (542) CCGGCGCCATCGTCTTCCTCTTCACCACCGCCGAATGGCTCGAGACCCTGGCTTGCACCAAGG------------
   Langdon_HMA3-A1_genomic  (669) CAGGCGTCATCGTCTTCCTCTTCACCACCGCAGAGTGGCTCGAGACCCTGGCCTGCACCAAGGTAATACAATTACA
      Langdon_HMA3-A1_CDS   (536) CAGGCGTCATCGTCTTCCTCTTCACCACCGCAGAGTGGCTCGAGACCCTGGCCTGCACCAAGG------------
 Langdon_TtHMA3-B1_genomic  (690) CAGGCGTCATCGTCTTCCTCTTCACCACTGCGGAGTGGCTCGAGACCCTGGCCTGCACCAAGGTTACCTACAACAA
    Langdon_TtHMA3-B1_CDS   (552) CAGGCGTCATCGTCTTCCTCTTCACCACTGCGGAGTGGCTCGAGACCCTGGCCTGCACCAAGG------------
 Partial_CS_HMA3-D1_genomic (661) CAGGCGTCATCGTCTTCCTCTTCACCACTGCAGAGTGGCTCGAGACCCTGGCCTGCACCAAGGTACTACAATTACA
                                                                                                   Section 12
                            (837) 837         850       860       870       880       890       900        912
           OsHMA3_genomic   (759) TGTCAATGT-CCTCTTCACCAAATCCATGGCGATCTTGGCACGAACCCGGACGAACACTCACGTTGCGAGGT---G
              OsHMA3_CDS    (647) ----------------------------------------------------------------------------
           BdHMA3_genomic   (701) CGAGCCTCTTCCAGAACTCCAGGTCTTCAGAAACTTTTGATTGGACGGAATCCTATGTTGACCTTAATTTGC---G
              BdHMA3_CDS    (605) ----------------------------------------------------------------------------
   Langdon_HMA3-A1_genomic  (745) GTGTGGATCCTAAATTACTTTCCATGATCGAATCGAAGGAATCAATATGAATTTTCCTAACACAAAGTGTGATGTG
      Langdon_HMA3-A1_CDS   (599) ----------------------------------------------------------------------------
 Langdon_TtHMA3-B1_genomic  (766) TTACAGTATGAATCCTCAGTTACTTTCCATGATCGATGGTATCAATATGAATTTTCCTAACACAAAGTGTGA---T
    Langdon_TtHMA3-B1_CDS   (615) ----------------------------------------------------------------------------
 Partial_CS_HMA3-D1_genomic (737) GTGTGAAG--AATCCTCAGTTACTTTCCGTGATCGATGGAATGAATGTGAATTTTCCTAACACAAAGTGTGATGTG
                                                                                                   Section 13
                            (913) 913    920       930       940       950       960       970             988
           OsHMA3_genomic   (831) TG--CAGGCGAGCGCCGGGATGTCGTCGTTGATGGCATGCTGCCGGTGAAGGCGGTCATCGCGACGACGGGCGAG
              OsHMA3_CDS    (647) --------CGAGCGCCGGGATGTCGTCGTTGATGGCATGCTGCCGGTGAAGGCGGTCATCGCGACGACGGGCGAG
           BdHMA3_genomic   (774) TGGTCAGGCTAGCGCCGGCATGTCGTCGTTGATGAGCATGATCCCGCCAAAGGCAGTGCTCGCGGAGACGGGTCAG
              BdHMA3_CDS    (605) --------CTAGCGCCGGCATGTCGTCGTTGATGAGCATGATCCCGCCAAAGGCAGTGCTCGCGGAGACGGGTCAG
   Langdon_HMA3-A1_genomic  (821) GTGTCAGGCCAGCGCCGGGATGTCGTCGTTGATGAGCATGATCCCGCCGAAGGCAGTCCTCGCCGAGACGGGCGAG
      Langdon_HMA3-A1_CDS   (599) --------CCAGCGCCGGGATGTCGTCGTTGATGAGCATGATCCCGCCGAAGGCAGTCCTCGCCGAGACGGGCGAG
 Langdon_TtHMA3-B1_genomic  (839) GTGTCAGGCCAGCGCCGGGATGTCGTCGTTGATGAGCATGATCCCGCCGAAGGCAGTCCTCGCCGAGACGGGCGAT
    Langdon_TtHMA3-B1_CDS   (615) --------CCAGCGCCGGGATGTCGTCGTTGATGAGCATGATCCCGCCGAAGGCAGTCCTCGCCGAGACGGGCGAT
 Partial_CS_HMA3-D1_genomic (811) GTGTCAGGCCAGCGCCGGGATGTCGTCGTTGATGAGCATGATCCCGCCGAAGGCAGTCCTCGCGGAGACGGGCGAG
                                                                                                   Section 14
                            (989) 989        1000      1010      1020      1030      1040      1050        1064
           OsHMA3_genomic   (905) GTTGTCAGCGTGCGCGACGTCCGCGTGGGCGACGTCGTCGCGGTCAGGGCCGGCGAGATCGTCCCCGTCGACGGCG
              OsHMA3_CDS    (715) GTTGTCAGCGTGCGCGACGTCCGCGTGGGCGACGTCGTCGCGGTCAGGGCCGGCGAGATCGTCCCCGTCGACGGCG
           BdHMA3_genomic   (850) GTGGTCAACGTGAGCGACATCGGTGTAGGCGCCGTCGTCGCCGTCAGGGCCGGCGAGGTGGTCCCGGTCGACGGCG
              BdHMA3_CDS    (673) GTGGTCAACGTGAGCGACATCGGTGTAGGCGCCGTCGTCGCCGTCAGGGCCGGCGAGGTGGTCCCGGTCGACGGCG
   Langdon_HMA3-A1_genomic  (897) GTCGTCAACGTTCGCGACATCGGTGTGGGCGTCGTCATCGCGGTCAGAGCAGGGGAGATGGTGCCGGTGGACGGCG
      Langdon_HMA3-A1_CDS   (667) GTCGTCAACGTTCGCGACATCGGTGTGGGCGTCGTCATCGCGGTCAGAGCAGGGGAGATGGTGCCGGTGGACGGCG
 Langdon_TtHMA3-B1_genomic  (915) GTCGTCAACGTGCGCGACATCGGTGTGGGCGCCGTCATCGCGGTCAGAGCAGGGGAGATGGTGCCGGTGGACGGCA
    Langdon_TtHMA3-B1_CDS   (683) GTCGTCAACGTGCGCGACATCGGTGTGGGCGCCGTCATCGCGGTCAGAGCAGGGGAGATGGTGCCGGTGGACGGCA
 Partial_CS_HMA3-D1_genomic (887) GTCGTCAACGTGCGCGACATAGGTGTGGGCGCCGTCATCGCGGTCAGAGCCGGGGAGATGGTGCCGGTGGATGGCG
                                                                                                   Section 15
                           (1065) 1065 1070      1080      1090      1100      1110      1120      1130      1140
           OsHMA3_genomic   (981) TGGTGGTCGACGGCCAGAGCGAGGTCGACGAGAGGAGCCTCACCGGCGAGTCCTTCCCGGTGCCGAAGCAGCCGCA
              OsHMA3_CDS    (791) TGGTGGTCGACGGCCAGAGCGAGGTCGACGAGAGGAGCCTCACCGGCGAGTCCTTCCCGGTGCCGAAGCAGCCGCA
           BdHMA3_genomic   (926) TGGTGGTGGACGGGCAGAGCGAGGTCGACGAGAGGAGCCTCACCGGGGAGTCCTACCCGGTGCCCAAGCAGCCGCA
              BdHMA3_CDS    (749) TGGTGGTGGACGGGCAGAGCGAGGTCGACGAGAGGAGCCTCACCGGGGAGTCCTACCCGGTGCCCAAGCAGCCGCA
   Langdon_HMA3-A1_genomic  (973) TGGTGGTCGACGGGCAGAGTGAGGTCGACGAGAGGAGCCTCACCGGCGAGTCGTACCCGGTGCCCAAGCAACCGCA
      Langdon_HMA3-A1_CDS   (743) TGGTGGTCGACGGGCAGAGTGAGGTCGACGAGAGGAGCCTCACCGGCGAGTCGTACCCGGTGCCCAAGCAACCGCA
 Langdon_TtHMA3-B1_genomic  (991) TGGTGGTCGACGGGCAGAGTGAGGTCGACGAGAGGAGCCTCACCGGCGAGTCGTACCCGGTGCCCAAGCAGCCGCA
    Langdon_TtHMA3-B1_CDS   (759) TGGTGGTCGACGGGCAGAGTGAGGTCGACGAGAGGAGCCTCACCGGCGAGTCGTACCCGGTGCCCAAGCAGCCGCA
 Partial_CS_HMA3-D1_genomic (963) TGGTGGTCGACGGGCAGAGNNNNNNNNNNNNNNNNNNNNNNNNNNGCGAGTCGTACCCGGTGCCCAAGCAACCGCA
                                                                                                   Section 16
                           (1141) 1141       1150      1160      1170      1180      1190      1200        1216
           OsHMA3_genomic  (1057) CTCCGAGGTCTGGGCCGGCACAATGAACTTTGACGGTAAACAAACAAACAATCAAAAAAATCTCTCATCTCTTAGT
              OsHMA3_CDS    (867) CTCCGAGGTCTGGGCCGGCACAATGAACTTTGACGGT---------------------------------------
           BdHMA3_genomic  (1002) GTCCGAGGTCTGGGCCGGCACCCTCAATTTGGATGGTAATCT--CAAAG-----------------------TGGA
              BdHMA3_CDS    (825) GTCCGAGGTCTGGGCCGGCACCCTCAATTTGGATGGC---------------------------------------
   Langdon_HMA3-A1_genomic (1049) GTCCGAGGTCTGGGCCGGCACGCTCAACTTGGACGGTACCAA----------------------------------
      Langdon_HMA3-A1_CDS   (819) GTCCGAGGTCTGGGCCGGCACGCTCAACTTGGACGGT---------------------------------------
 Langdon_TtHMA3-B1_genomic (1067) CTCCGAGGTCTGGGCCGGCACGCTCAACTTGGACGGTACAAA----AAC---------------------------
    Langdon_TtHMA3-B1_CDS   (835) CTCCGAGGTCTGGGCCGGCACGCTCAACTTGGACGGT---------------------------------------
 Partial_CS_HMA3-D1_genomic(1039) GTCCGAGGTCTGGGCCGGCACGCTCAACTTGGACGGTACTAA----------------------------------
                                                                                                   Section 17
                           (1217) 1217        1230      1240      1250      1260      1270      1280        1292
           OsHMA3_genomic  (1133) TTCAGAGTTCCCAAATGTACGGAACTCTGAAACGTCCGTGTCGTTGAATTGTCCGTCTTTTCTCATCAGCTTAAGT
              OsHMA3_CDS    (904) ----------------------------------------------------------------------------
           BdHMA3_genomic  (1053) TTCCCAGACCACAGTAGTACT--CC---------------------------------------------------
              BdHMA3_CDS    (862) ----------------------------------------------------------------------------
   Langdon_HMA3-A1_genomic (1091) ----------------------------------------------------------------------------
      Langdon_HMA3-A1_CDS   (856) ----------------------------------------------------------------------------
 Langdon_TtHMA3-B1_genomic (1112) ------------------------------------------------------------T---------------
    Langdon_TtHMA3-B1_CDS   (872) ----------------------------------------------------------------------------
 Partial_CS_HMA3-D1_genomic(1081) ------------AAC--T----------------------------------------------------------
                                                                                                   Section 18
                           (1293) 1293   1300      1310      1320      1330      1340      1350              1368
           OsHMA3_genomic  (1209) TCTGGTAAACTGATTGACGCATCCAACTCAATAGTCACAATAGCACACATTTGTTCATGTCATGATCAATGGGATT
              OsHMA3_CDS    (904) ----------------------------------------------------------------------------
           BdHMA3_genomic  (1076) ----GTAACATTTTGCACACAAAATTCTCAAGGCCAAGGCAACCATGTTCTTGATCTGGTCCCTTTC---------
              BdHMA3_CDS    (862) ----------------------------------------------------------------------------
   Langdon_HMA3-A1_genomic (1091) ------AACTGGATTGTTTCTTCTCCTTGAAGGCAGCGACGTTCTTGCGGTGATCTGAGCCTACTGTCA-------
      Langdon_HMA3-A1_CDS   (856) ----------------------------------------------------------------------------
 Langdon_TtHMA3-B1_genomic (1113) ---GGCCGGCTTGTTTCTTCTCCTTCTTGAAGGCAGCGACGTTCTTGCGGTGATCTGAGCTACTGTCA--------
    Langdon_TtHMA3-B1_CDS   (872) ----------------------------------------------------------------------------
 Partial_CS_HMA3-D1_genomic(1085) -------GCAATGTTCCTTCTCTTTCTTGAAGGCAGCGACATTCTCGTGGTGATCTGAGCCTACTGTGA-------
```

```
Section 19
(1369) 1369 1380 1390 1400 1410 1420 1430 1444
OsHMA3_genomic (1285)            GAGAGGCTGTCACGTTTCAGGTTACATTGCTGTGAGAACTACGGCTCTCGCCGAGAACTCGACGGTGGCGAAGATG
OsHMA3_CDS (904)                 ----------------------TACATTGCTGTGAGAACTACGGCTCTCGCCGAGAACTCGACGGTGGCGAAGATG
BdHMA3_genomic (1139)            --------------TTGCAGGCTACATTGCCGTGAGGACAAGTGCTCTTGCGGAGAACTCGACGGTGGCCAGGATG
BdHMA3_CDS (862)                 ----------------------TACATTGCCGTGAGGACAAGTGCTCTTGCGGAGAACTCGACGGTGGCCAGGATG
Langdon_HMA3-A1_genomic (1154)   ----------------TGCAGGTTACATCGCCGTGAGAACAATGGCTCTCGCCGAGAACTCCACGGTGGCCAAGATG
Langdon_HMA3-A1_CDS (856)        ----------------------TACATCGCCGTGAGAACAATGGCTCTCGCCGAGAACTCCACGGTGGCCAAGATG
Langdon_TtHMA3-B1_genomic (1178) --------------T-GCAGGTTACATCGCCGTGAGGACAATGGCTCTCGCCGAGAACTCCACGGTGGCCAAGATG
Langdon_TtHMA3-B1_CDS (872)      ----------------------TACATCGCCGTGAGGACAATGGCTCTCGCCGAGAACTCCACGGTGGCCAAGATG
Partial_CS_HMA3-D1_genomic (1147) ----------------TGCAGGCTACATCGCCGTGAGGACAATGGCCCTCGCCGAGAACTCCACGGTGGCCAAGATG

Section 20
(1445) 1445 1450 1460 1470 1480 1490 1500 1510 1520
OsHMA3_genomic (1361)            GAGAGGCTGGTGGAGGCGGCGCAGAACAGCAGGTCGAAGACGCAGCGGCTGATCGATTCGTGCGCAAAGTACTACA
OsHMA3_CDS (958)                 GAGAGGCTGGTGGAGGCGGCGCAGAACAGCAGGTCGAAGACGCAGCGGCTGATCGATTCGTGCGCAAAGTACTACA
BdHMA3_genomic (1201)            GAGAAGCTGGTGGAAGAGGCCCAGAACAGCAGGTCCAAGACGCAGCGGCTTATCGATTCCTGCGCGAAATACTACA
BdHMA3_CDS (916)                 GAGAAGCTGGTGGAAGAGGCCCAGAACAGCAGGTCCAAGACGCAGCGGCTTATCGATTCCTGCGCGAAATACTACA
Langdon_HMA3-A1_genomic (1215)   GAGAGGCTGGTGGAGGAGGCCCAGCAGAGCAAGTCCAAGACGCAGCGGCTGATCGACTCCTGCGCCAAGTACTACA
Langdon_HMA3-A1_CDS (910)        GAGAGGCTGGTGGAGGAGGCCCAGCAGAGCAAGTCCAAGACGCAGCGGCTGATCGACTCCTGCGCCAAGTACTACA
Langdon_TtHMA3-B1_genomic (1239) GAGAGGCTGGTGGAGGAGGCCCAGCAGAGCAAGTCCAGGACGCAGCGGCTGATCGACTCCTGCGCCAAGTACTACA
Langdon_TtHMA3-B1_CDS (926)      GAGAGGCTGGTGGAGGAGGCCCAGCAGAGCAAGTCCAGGACGCAGCGGCTGATCGACTCCTGCGCCAAGTACTACA
Partial_CS_HMA3-D1_genomic (1208) GAGAGGCTGGTGGAGGAGGCCCAGCAGAGCAAGTCCAAGACGCAGCGGCTGATCGACTCCTGCGCCAGGTACTACA

Section 21
(1521) 1521 1530 1540 1550 1560 1570 1580 1596
OsHMA3_genomic (1437)            CGCCGGGTGCGCAATCGGCATCGCCTAAACTCTCAACATCTCCAATTTTACAACACAAAATTCATCAGTAACAAAG
OsHMA3_CDS (1034)                CGCCGG----------------------------------------------------------------------
BdHMA3_genomic (1277)            CGCCAG----------------------------------------------------------------------
BdHMA3_CDS (992)                 CGCCAG----------------------------------------------------------------------
Langdon_HMA3-A1_genomic (1291)   CGCCGG----------------------------------------------------------------------
Langdon_HMA3-A1_CDS (986)        CGCCGG----------------------------------------------------------------------
Langdon_TtHMA3-B1_genomic (1315) CGCCGG----------------------------------------------------------------------
Langdon_TtHMA3-B1_CDS (1002)     CGCCGG----------------------------------------------------------------------
Partial_CS_HMA3-D1_genomic (1284) CGCCGG----------------------------------------------------------------------

Section 22
(1597) 1597 1610 1620 1630 1640 1650 1660 1672
OsHMA3_genomic (1513)            ACATTTCTGTTCTTGGTGCGTTGCAGCCGTGGTGGTTGTTGCAGCAGGAGTGGCCCTGATCCCGGCTCTGCTCGGA
OsHMA3_CDS (1040)                --------------------------CCGTGGTGGTTGTTGCAGCAGGAGTGGCCCTGATCCCGGCTCTGCTCGGA
BdHMA3_genomic (1283)            --------------------------CCGTGGTGGTCATGGGGGCAGGGGTGGCGCTGCCGCCGCTGCTGCTGGGA
BdHMA3_CDS (998)                 --------------------------CCGTGGTGGTCATGGGGGCAGGGGTGGCGCTGCCGCCGCTGCTGCTGGGA
Langdon_HMA3-A1_genomic (1297)   --------------------------CCGTGGTGGTTCTCGGAGCAGGGGTGGCGCTGCTGCCGCCGCTGCTGGGG
Langdon_HMA3-A1_CDS (992)        --------------------------CCGTGGTGGTTCTCGGAGCAGGGGTGGCGCTGCTGCCGCCGCTGCTGGGG
Langdon_TtHMA3-B1_genomic (1321) --------------------------CCGTGGTGGTTCTCGGAGCAGGGGTGGCGCTGCTGCCGCCGCTGCTGGGG
Langdon_TtHMA3-B1_CDS (1008)     --------------------------CCGTGGTGGTTCTCGGAGCAGGGGTGGCGCTGCTGCCGCCGCTGCTGGGG
Partial_CS_HMA3-D1_genomic (1290) --------------------------CCGTGGTGGTTCTCGGAGCAGGGGTGGCGCTGCTGCCGCCGCTGCTGGGG

Section 23
(1673) 1673 1680 1690 1700 1710 1720 1730 1748
OsHMA3_genomic (1589)            GCAGATGGCCTTGAGCAATGGTGGAAGCTGGCTCTGGTGATGCTTGTGAGCGCGTGCCCCTGCGCATTAGTGCTGT
OsHMA3_CDS (1090)                GCAGATGGCCTTGAGCAATGGTGGAAGCTGGCTCTGGTGATGCTTGTGAGCGCGTGCCCCTGCGCATTAGTGCTGT
BdHMA3_genomic (1333)            GCAGGGGACCCGGAGCGGTGGTTCAGGCTGTCCCTGGTGCTGCTGGTGAGCGCGTGCCCGTGCGCGCTGGTGCTGT
BdHMA3_CDS (1048)                GCAGGGGACCCGGAGCGGTGGTTCAGGCTGTCCCTGGTGCTGCTGGTGAGCGCGTGCCCGTGCGCGCTGGTGCTGT
Langdon_HMA3-A1_genomic (1347)   GCGCGCGACGCGGAGCGGTGGTTCAGGCTGGCGCTGGTGCTGCTGGTGAGCGCGTGCCCGTGCGCGCTGGTGCTGT
Langdon_HMA3-A1_CDS (1042)       GCGCGCGACGCGGAGCGGTGGTTCAGGCTGGCGCTGGTGCTGCTGGTGAGCGCGTGCCCGTGCGCGCTGGTGCTGT
Langdon_TtHMA3-B1_genomic (1371) GCGCGCGACGCGGAGCGGTGGTTCCGGCTGGCGCTGGTGCTGCTGGTGAGCGCGTGCCCGTGCGCTCTGGTGCTGT
Langdon_TtHMA3-B1_CDS (1058)     GCGCGCGACGCGGAGCGGTGGTTCCGGCTGGCGCTGGTGCTGCTGGTGAGCGCGTGCCCGTGCGCTCTGGTGCTGT
Partial_CS_HMA3-D1_genomic (1340) GCGCGCGACGCGGAGCGGTGGTTCAGGCTGGCGCTGGTGCTGCTGGTGAGCGCGTGCCCGTGCGCGCTGGTGCTGT

Section 24
(1749) 1749 1760 1770 1780 1790 1800 1810 1824
OsHMA3_genomic (1665)            CGACACCGGTGGCATCCTTCTGCGCAATGCTGCGCGCTGCGAGGATGGGGATCTTCATCAAGGGTGGAGATGTTCT
OsHMA3_CDS (1166)                CGACACCGGTGGCATCCTTCTGCGCAATGCTGCGCGCTGCGAGGATGGGGATCTTCATCAAGGGTGGAGATGTTCT
BdHMA3_genomic (1409)            CGACGCCCGTGGCCACCTTCTGCGCGCTCCTCACGGCGGCCAGGATGGGTGTCCTAATCAAGGGAGGGGACGTCCT
BdHMA3_CDS (1124)                CGACGCCCGTGGCCACCTTCTGCGCGCTCCTCACGGCGGCCAGGATGGGTGTCCTAATCAAGGGAGGGGACGTCCT
Langdon_HMA3-A1_genomic (1423)   CGACGCCCGTCGCGACGTTCTGCGCGCTCCTGACGGCGGCGCGGATGGGGGTCCTCGTCAAGGGAGGGGACGTCCT
Langdon_HMA3-A1_CDS (1118)       CGACGCCCGTCGCGACGTTCTGCGCGCTCCTGACGGCGGCGCGGATGGGGGTCCTCGTCAAGGGAGGGGACGTCCT
Langdon_TtHMA3-B1_genomic (1447) CGACGCCGGTCGCGACGTTCTGCGCGCTGCTGACGGCGGCGAGGATGGGGGTCCTCGTCAAGGGAGGGGACGTCCT
Langdon_TtHMA3-B1_CDS (1134)     CGACGCCGGTCGCGACGTTCTGCGCGCTGCTGACGGCGGCGAGGATGGGGGTCCTCGTCAAGGGAGGGGACGTCCT
Partial_CS_HMA3-D1_genomic (1416) CGACGCCCGTCGCGACGTTCTGCGCGCTCCTGACGGCGGCGAGGATGGGGGTCCTCGTCAAGGGAGGGGACGTCCT

Section 25
(1825) 1825 1830 1840 1850 1860 1870 1880 1890 1900
OsHMA3_genomic (1741)            TGAATCACTTGGGGAGATCAGGGCCGTCGCGTTCGACAAGACCGGGACGATCACCAGAGGAGAGTTCAGCATCGAC
OsHMA3_CDS (1242)                TGAATCACTTGGGGAGATCAGGGCCGTCGCGTTCGACAAGACCGGGACGATCACCAGAGGAGAGTTCAGCATCGAC
BdHMA3_genomic (1485)            CGAGTCCCTGGGGGAGATCACGGCCGTCGCCTTCGACAAGACCGGCACCATCACCAGAGGCGAGTTCAGCATCGCC
BdHMA3_CDS (1200)                CGAGTCCCTGGGGGAGATCACGGCCGTCGCCTTCGACAAGACCGGCACCATCACCAGAGGCGAGTTCAGCATCGCC
Langdon_HMA3-A1_genomic (1499)   CGAGTCGCTGGGCGAGATCAGGGCCGTGGCGTTCGACAAGACCGGCACCATCACCAGAGGGGAGTTCACCGTCGAC
Langdon_HMA3-A1_CDS (1194)       CGAGTCGCTGGGCGAGATCAGGGCCGTGGCGTTCGACAAGACCGGCACCATCACCAGAGGGGAGTTCACCGTCGAC
Langdon_TtHMA3-B1_genomic (1523) CGAGTCGCTTGGCGAGATCAAGGCCGTGGCGTTCGACAAGACCGGCACCATCACCAGAGGGGAGTTCACCGTCGAC
Langdon_TtHMA3-B1_CDS (1210)     CGAGTCGCTTGGCGAGATCAAGGCCGTGGCGTTCGACAAGACCGGCACCATCACCAGAGGGGAGTTCACCGTCGAC
Partial_CS_HMA3-D1_genomic (1492) CGAGTCGCTGGGCGAGATCAGGGCCGTGGCGTTCGACAAGACCGGCACCATCACCAGAGGGGAGTTCAGCGTCGAC

Section 26
(1901) 1901 1910 1920 1930 1940 1950 1960 1976
OsHMA3_genomic (1817)            TCGTTCCATCTGGTTGGGGATCACAAGGTTGAGATGGATCATCTTCTTTACTGGTAAGACTGAGAAAGAAGATTTC
OsHMA3_CDS (1318)                TCGTTCCATCTGGTTGGGGATCACAAGGTTGAGATGGATCATCTTCTTTACTGG----------------------
BdHMA3_genomic (1561)            GCGTTCCATGTGGTCGCGGA---CAAGGTTCAGATCACCCACCTTCTTTACTGGTAA--------AT-ATG--TTC
BdHMA3_CDS (1276)                GCGTTCCATGTGGTCGCGGA---CAAGGTTCAGATCACCCACCTTCTTTACTGG----------------------
Langdon_HMA3-A1_genomic (1575)   ATGTTCGATGTGGTCGAACA---GAAGGTTCAGATGAGCCATCTTCTTTACTGGTATG---------------TTC
Langdon_HMA3-A1_CDS (1270)       ATGTTCGATGTGGTCGAACA---GAAGGTTCAGATGAGCCATCTTCTTTACTGG----------------------
Langdon_TtHMA3-B1_genomic (1599) ATGTTCGACGTGGTTGGACA---CAAGGTTCAGATGAGCCATCTTCTTTACTGGTATG---------------TGT
Langdon_TtHMA3-B1_CDS (1286)     ATGTTCGACGTGGTTGGACA---CAAGGTTCAGATGAGCCATCTTCTTTACTGG----------------------
Partial_CS_HMA3-D1_genomic (1568) ATGTTCGATGTGGTTGGGCA---CAAGGTTCAGATGAGCTATCTTCTTTACTGGTACGTGTG----------TCCA

Section 27
(1977) 1977 1990 2000 2010 2020 2030 2040 2052
OsHMA3_genomic (1893)            ATTTCAGCTCTTGGAATGCAGCTGCTACACTGAACAAATTTGACAGATTTTCAACTGTTGTCGGTTTACGACTTGT
OsHMA3_CDS (1372)                ----------------------------------------------------------------------------
BdHMA3_genomic (1623)            ACTTCCACTTT--CAC-----------CCCTGA-------------------AACTCATGT---------------
BdHMA3_CDS (1327)                ----------------------------------------------------------------------------
Langdon_HMA3-A1_genomic (1633)   ACCTC-----------------------------------------------------------------------
Langdon_HMA3-A1_CDS (1321)       ----------------------------------------------------------------------------
Langdon_TtHMA3-B1_genomic (1657) ACGTCCTCTGGGCCGGCATTGAAATTTCACTGATGA----------------AAA---------------------
Langdon_TtHMA3-B1_CDS (1337)     ----------------------------------------------------------------------------
Partial_CS_HMA3-D1_genomic (1633) CTGTCTATTGCCACACCTCTTGGGCGGCATTGAAC-----------------------------------------
```

```
                                                                                                    Section 28
                         (2053) 2053    2060      2070      2080      2090      2100      2110            2128
          OsHMA3_genomic (1969) CGAAAAATCCTTAGTACTTGTGATGGTTTGAATTGACTTGAACAGTGAAATGCTTGGCACAAATTATGTCTTTGGT
             OsHMA3_CDS (1372) ----------------------------------------------------------------------------
          BdHMA3_genomic (1652) -------------------------GTGTGAAGTGAAATGATGA---------------------------TTTCT
             BdHMA3_CDS (1327) ----------------------------------------------------------------------------
   Langdon_HMA3-A1_genomic (1638) --------------------------TGAGCATTGAAATTTTCACCGATGAAAACTCGTGTTGAAATGGAGCTTC
      Langdon_HMA3-A1_CDS (1321) ----------------------------------------------------------------------------
 Langdon_TtHMA3-B1_genomic (1696) --------------------------TATCCAGTGCTGAAATGAG---AG---------------------CTTCT
    Langdon_TtHMA3-B1_CDS (1337) ----------------------------------------------------------------------------
Partial_CS_HMA3-D1_genomic (1668) --------------------------TTTTCACTGATGAAAACTCGTGCTGAAATGAG---AG--------CTTCT
                                                                                                    Section 29
                         (2129) 2129      2140      2150      2160      2170      2180      2190          2204
          OsHMA3_genomic (2045) TATTGCAGGATTGCAAGCATTGAGAGCAAGTCAAGCCACCCAATGGCAGCTGCACTTGTGGAGTATGCTCAATCCA
             OsHMA3_CDS (1372) ---------ATTGCAAGCATTGAGAGCAAGTCAAGCCACCCAATGGCAGCTGCACTTGTGGAGTATGCTCAATCCA
          BdHMA3_genomic (1676) TGTTGCAGGGTGTCAAGCATCGAGAGCAAATCGAGCCATCCAATGGCGGCTGCACTTGTGGAGTACGCGCGGTCCA
             BdHMA3_CDS (1327) ---------GTGTCAAGCATCGAGAGCAAATCGAGCCATCCAATGGCGGCTGCACTTGTGGAGTACGCGCGGTCCA
   Langdon_HMA3-A1_genomic (1688) TTTGGCAGGATCTCAAGCATCGAGAGCAAATCCAGCCACCCAATGGCGGCTGCGCTCGTGGAGCACGCGCAGTCGA
      Langdon_HMA3-A1_CDS (1321) ---------ATCTCAAGCATCGAGAGCAAATCCAGCCACCCAATGGCGGCTGCGCTCGTGGAGCACGCGCAGTCGA
 Langdon_TtHMA3-B1_genomic (1722) TCTCGCAGGATCTCAAGCATCGAGAGCAAATCTAGCCACCCAATGGCGGCTGCGCTCGTGGAGCACGCGCGGTCGA
    Langdon_TtHMA3-B1_CDS (1337) ---------ATCTCAAGCATCGAGAGCAAATCTAGCCACCCAATGGCGGCTGCGCTCGTGGAGCACGCGCGGTCGA
Partial_CS_HMA3-D1_genomic (1707) TCTCGCAGGATCTCAAGCATCGAGAGCAAATCCAGCCACCCAATGGCGGCTGCGCTCGTGGAGTACGCGCAGTCGA
                                                                                                    Section 30
                         (2205) 2205 2210      2220      2230      2240      2250      2260      2270      2280
          OsHMA3_genomic (2121) AATCCATCCAACCAAACCCGGAAAATGTGGGCGATTTTCGGATATATCCCGGGGAGGGGATCTATGGAGAGATCCA
             OsHMA3_CDS (1439) AATCCATCCAACCAAACCCGGAAAATGTGGGCGATTTTCGGATATATCCCGGGGAGGGGATCTATGGAGAGATCCA
          BdHMA3_genomic (1752) AATCTATCCAGCCAAAGCCGGAGAATGTCGCCGAGTTTCGCGTGATTCCCGGAGAGGGCGTCTATGGAGAGATCCA
             BdHMA3_CDS (1394) AATCTATCCAGCCAAAGCCGGAGAATGTCGCCGAGTTTCGCGTGATTCCCGGAGAGGGCGTCTATGGAGAGATCCA
   Langdon_HMA3-A1_genomic (1764) AATCCATCCAGCCGAAACCGGAATGCGTCGCCGAGTTCCGCATCCTTCCCGGCGAGGGCGTCTACGGAGAGATCGA
      Langdon_HMA3-A1_CDS (1388) AATCCATCCAGCCGAAACCGGAATGCGTCGCCGAGTTCCGCATCCTTCCCGGCGAGGGCGTCTACGGAGAGATCGA
 Langdon_TtHMA3-B1_genomic (1798) AATCCATCGAGCCGAAACCGGAATGCGTCGCCGAGTTCCGCGTCCTTCCCGGCGAGGGCATATATGGAGAGATCGA
    Langdon_TtHMA3-B1_CDS (1404) AATCCATCGAGCCGAAACCGGAATGCGTCGCCGAGTTCCGCGTCCTTCCCGGCGAGGGCATATATGGAGAGATCGA
Partial_CS_HMA3-D1_genomic (1783) AATCCATCGAGCCGAA-CCGGAATGCGTCGCC-AGTTCCGCATCCTTCCCGGCGAGGGCATCTATGGAGAGATCGA
                                                                                                    Section 31
                         (2281) 2281      2290      2300      2310      2320      2330      2340            2356
          OsHMA3_genomic (2197) TGGAAAGCACATCTACATTGGGAACAGAAGGACAT---TGGCAAGAGCGTCAT-----CACCTCAGTCAAGTGAGC
             OsHMA3_CDS (1515) TGGAAAGCACATCTACATTGGGAACAGAAGGACAT---TGGCAAGAGCGTCAT-----CACCTCAGTCAA------
          BdHMA3_genomic (1828) TGGAAGGCACATCTACATCGGAAACAAGAGGGCGGCCCTGGCAAGGGATCGTG---TCACACAGGTGAA-----C
             BdHMA3_CDS (1470) TGGAAGGCACATCTACATCGGAAACAAGAGGGCGGCCCTGGCAAGGGATCGTG---TCACACAG-----------
   Langdon_HMA3-A1_genomic (1840) CGGGAAGCGCATCTACGTCGGGAACAAGAGGGTCT---TGGCGAGGGGATCCTCCTGTCAGACAGGTGAG------
      Langdon_HMA3-A1_CDS (1464) CGGGAAGCGCATCTACGTCGGGAACAAGAGGGTCT---TGGCGAGGGGATCCTCCTGTCAGACAG-----------
 Langdon_TtHMA3-B1_genomic (1874) CGGGATGCGCATCTACGTCGGGAACAAGAGGGTCT---TGGCGAGGGGATCATCCTGTCAGACAGGTGA-------
    Langdon_TtHMA3-B1_CDS (1480) CGGGATGCGCATCTACGTCGGGAACAAGAGGGTCT---TGGCGAGGGGATCATCCTGTCAGACAG-----------
Partial_CS_HMA3-D1_genomic (1857) CGGGAAACGCATCTACGTCGGGAACAAGAGGGTCT---TGGCGAGGGGATCCTCCTGTCAGACAGGTGAG------
                                                                                                    Section 32
                         (2357) 2357           2370      2380      2390      2400      2410      2420      2432
          OsHMA3_genomic (2265) AAAACACATCATTTGTAGTAGTAGTACAGTTCCTGACAATTAAGCTTCAGAATTAGCTGATCAGCTGAAGCTGAGC
             OsHMA3_CDS (1577) ----------------------------------------------------------------------------
          BdHMA3_genomic (1896) A---C----------TTGCTGCTTGAAAAATGATG----TTTTGTTTCAGAAGAAACTGACTCAACTGAACTCAAC
             BdHMA3_CDS (1532) ----------------------------------------------------------------------------
   Langdon_HMA3-A1_genomic (1907) --------------CTGAGAGCAGAGCAAACATACA----TTTGGTGCTTTGCTTGGTCAATGAAGTTTGAGAAGA
      Langdon_HMA3-A1_CDS (1526) ----------------------------------------------------------------------------
 Langdon_TtHMA3-B1_genomic (1940) -------------GCTGAGAGCAGAGCAAACATGCA----TTTGGTACGTGCTTGACCGATGAAGTTTGAGAAAGA
    Langdon_TtHMA3-B1_CDS (1542) ----------------------------------------------------------------------------
Partial_CS_HMA3-D1_genomic (1924) --------------CTGAGAGCAGAGCAAACATACA--------TTCTGTGCTTGATCAATGAAGTTTGAGAAGGA
                                                                                                    Section 33
                         (2433) 2433    2440      2450      2460      2470      2480      2490            2508
          OsHMA3_genomic (2341) TGACTGAACTGAAATCTGTGACGTTTACAGCTCAAGAAATGGGTGA-GATGATC--AAGGGCGTGTCGATCGGCTA
             OsHMA3_CDS (1577) ----------------------------CTCAAGAAATGGGTGA-GATGATC--AAGGGCGTGTCGATCGGCTA
          BdHMA3_genomic (1955) T--TTGTGTC---------GACGTTGACAGTTCCAGAGAAAAAGGACGGTGATCTGAAAGGCGTGTCCGTGGGATA
             BdHMA3_CDS (1532) -----------------------------TTCCAGAGAAAAAGGACGGTGATCTGAAAGGCGTGTCCGTGGGATA
   Langdon_HMA3-A1_genomic (1965) AAATTTTCTT------------GCTTGCAGTTCCAGAAAGAATGAATGGT---CTGAAGGGCGTCTCGATCGGCTA
      Langdon_HMA3-A1_CDS (1526) -----------------------------TTCCAGAAAGAATGAATGGT---CTGAAGGGCGTCTCGATCGGCTA
 Langdon_TtHMA3-B1_genomic (1999) AAACTTTCTT------------GCTTGCAGTTCCAGAAAGAATGAACGGT---CTGAAAGGCGTCTCCATGGGCTA
    Langdon_TtHMA3-B1_CDS (1542) -----------------------------TTCCAGAAAGAATGAACGGT---CTGAAAGGCGTCTCCATGGGCTA
Partial_CS_HMA3-D1_genomic (1978) AAACTTTCTTA------------CTTGCAGTGCCAGAAAGAATGAATGGC---CTGAAAGGCGTCTCGATCGGCTA
                                                                                                    Section 34
                         (2509) 2509      2520      2530      2540      2550      2560      2570          2584
          OsHMA3_genomic (2414) CGTGATCTGCGACGGCGAGCTCGCCGGCGTCTTCTCGCTCTCCGACGACTGCCGGACCGGCGCGGCGGAGGCGATC
             OsHMA3_CDS (1620) CGTGATCTGCGACGGCGAGCTCGCCGGCGTCTTCTCGCTCTCCGACGACTGCCGGACCGGCGCGGCGGAGGCGATC
          BdHMA3_genomic (2020) CGTGATCTGCGACGGGGACCTCGTCGGGGCGTTCTCCCTCTCGGACGATTGCCGGACCGGAGCGGCAGAAGCGATT
             BdHMA3_CDS (1578) CGTGATCTGCGACGGGGACCTCGTCGGGGCGTTCTCCCTCTCGGACGATTGCCGGACCGGAGCGGCAGAAGCGATT
   Langdon_HMA3-A1_genomic (2026) CGTGATCTGCGACGGGGACCTCGTCGGGGTGTTCTCGCTCTCCGACGACTGCCGGACCGGCGCGGCCGAAGCGATT
      Langdon_HMA3-A1_CDS (1569) CGTGATCTGCGACGGGGACCTCGTCGGGGTGTTCTCGCTCTCCGACGACTGCCGGACCGGCGCGGCCGAAGCGATT
 Langdon_TtHMA3-B1_genomic (2060) CGTGATCTGCGACGGGGACCTCGTCGGAGTGTTTTCCCTCTCAGACGACTGCCGGACCGGCGCGGCCGAGGCGATT
    Langdon_TtHMA3-B1_CDS (1585) CGTGATCTGCGACGGGGACCTCGTCGGAGTGTTTTCCCTCTCAGACGACTGCCGGACCGGCGCGGCCGAGGCGATT
Partial_CS_HMA3-D1_genomic (2039) CGTGATCTGCGACGGGGACCTCGTCGGGGTGTTCTCGTTCTCCGACGACTGCCGGACCGGCGCGGCCGAAGCGATT
                                                                                                    Section 35
                         (2585) 2585 2590      2600      2610      2620      2630      2640      2650      2660
          OsHMA3_genomic (2490) CGGGAGCTTGGATCGCTGGGCATCAAGTCGGTTATGCTGACCGGGGACAGCAGCGCGGCGGCGACGCACGCGCAGG
             OsHMA3_CDS (1696) CGGGAGCTTGGATCGCTGGGCATCAAGTCGGTTATGCTGACCGGGGACAGCAGCGCGGCGGCGACGCACGCGCAGG
          BdHMA3_genomic (2096) CAGGAACTAACAGCCATGGGCATCAAATCCGTAATGCTCACGGGAGACAGCGCGGAAGCCGCGGCGCACGCGCAGG
             BdHMA3_CDS (1654) CAGGAACTAACAGCCATGGGCATCAAATCCGTAATGCTCACGGGAGACAGCGCGGAAGCCGCGGCGCACGCGCAGG
   Langdon_HMA3-A1_genomic (2102) CGAGAGCTGGCGTCCATGGGCATCAGCTCAGTACTGCTCACGGGGGACAGCGCGGAGGCGGCGGTGCACGCGCAGG
      Langdon_HMA3-A1_CDS (1645) CGAGAGCTGGCGTCCATGGGCATCAGCTCAGTACTGCTCACGGGGGACAGCGCGGAGGCGGCGGTGCACGCGCAGG
 Langdon_TtHMA3-B1_genomic (2136) CGAGAGCTGGCGTCCATGGGCATCAGCTCAGTACTGCTCACGGGGGACATCGCGGAGGCGGCCATGCACGCGCAGG
    Langdon_TtHMA3-B1_CDS (1661) CGAGAGCTGGCGTCCATGGGCATCAGCTCAGTACTGCTCACGGGGGACATCGCGGAGGCGGCCATGCACGCGCAGG
Partial_CS_HMA3-D1_genomic (2115) CGAGAGCTGGCGTCCATGGGCATCAGCTCAGTACTGCTCACGGGGGACAGCGCGGAGGCGGCCGTGCACGCGCAGC
                                                                                                    Section 36
                         (2661) 2661    2670      2680      2690      2700      2710      2720            2736
          OsHMA3_genomic (2566) GCCAGCTCGGCGGCGT---CATGGAGGAGCTCCACTCCGAGCTCCTCCCGGAGGACAAGGTCCGGCTCGTCAGCGG
             OsHMA3_CDS (1772) GCCAGCTCGGCGGCGT---CATGGAGGAGCTCCACTCCGAGCTCCTCCCGGAGGACAAGGTCCGGCTCGTCAGCGG
          BdHMA3_genomic (2172) AGCAGCTCGGGGGCGC---GCTCGAGGAGCTCCACTGGGATCTCTTCCCGGAGGACAAGGTGCGGCTCGTCGCCGC
             BdHMA3_CDS (1730) AGCAGCTCGGGGGCGC---GCTCGAGGAGCTCCACTGGGATCTCTTCCCGGAGGACAAGGTGCGGCTCGTCGCCGC
   Langdon_HMA3-A1_genomic (2178) AGCGGCTGGGCGGCGC---CCTCGAGGAGCTCCACTCGGAGCTCTTCCCGGAGGACAAGGTCCGGCTGGTGGGCGC
      Langdon_HMA3-A1_CDS (1721) AGCGGCTGGGCGGCGC---CCTCGAGGAGCTCCACTCGGAGCTCTTCCCGGAGGACAAGGTCCGGCTGGTGGGCGC
 Langdon_TtHMA3-B1_genomic (2212) AGCAGCTGGGCGGCGCCTTGTTGGAGGAGGTCCACTCGGAGCTCTTCCCGGAGGACAAGGTCCGGCTGGTGGGCGC
    Langdon_TtHMA3-B1_CDS (1737) AGCAGCTGGGCGGCGCCTTGTTGGAGGAGGTCCACTCGGAGCTCTTCCCGGAGGACAAGGTCCGGCTGGTGGGCGC
Partial_CS_HMA3-D1_genomic (2191) AGCAGCTGGGCGGCGC---CTTGGAGGAGCTCCACTCGGAGCTCTTCCCGGAGGACAAGGTCCGGCTGGTGGGCGC
```

```
                                                                                    Section 37
                            (2737) 2737          2750      2760      2770      2780      2790      2800       2812
           OsHMA3_genomic (2639) CCTCAAGGCGAGGTTCGGCCCGACGATGATGGTCGGGGACGGGATGAACGACGCGGCGGCGCTGGCGGCGGCGGAC
               OsHMA3_CDS (1845) CCTCAAGGCGAGGTTCGGCCCGACGATGATGGTCGGGGACGGGATGAACGACGCGGCGGCGCTGGCGGCGGCGGAC
           BdHMA3_genomic (2245) CCTGAAGGCGAGGGCGGGGCCGACGATGATGGTCGGCGACGGGATGAACGACGCCCCGGCGCTGGCCATGGCGGAC
               BdHMA3_CDS (1803) CCTGAAGGCGAGGGCGGGGCCGACGATGATGGTCGGCGACGGGATGAACGACGCCCCGGCGCTGGCCATGGCGGAC
  Langdon_HMA3-A1_genomic (2251) GCTGAAGGCGAGGGCTGGGCCGACGATGATGGTCGGCGACGGCATGAACGACGCCCCGGCGCTGGCGACGGCGGAC
      Langdon_HMA3-A1_CDS (1794) GCTGAAGGCGAGGGCTGGGCCGACGATGATGGTCGGCGACGGCATGAACGACGCCCCGGCGCTGGCGACGGCGGAC
Langdon_TtHMA3-B1_genomic (2288) GCTGAAGGCGAGGGCTGGGCCGACGATGATGGTCGGCGACGGCATGAACGACGCCCCGGCGCTGGCAACGGCGGAC
    Langdon_TtHMA3-B1_CDS (1813) GCTGAAGGCGAGGGCTGGGCCGACGATGATGGTCGGCGACGGCATGAACGACGCCCCGGCGCTGGCAACGGCGGAC
Partial_CS_HMA3-D1_genomic (2264) GCTGAAGGCGAGGGCTGGGCCGACGATGATGGTCGGCGACGGCATGAACGACGCCCCGGCACTGGCGACGGCGGAC
                                                                                    Section 38
                            (2813) 2813      2820       2830      2840      2850      2860      2870            2888
           OsHMA3_genomic (2715) GTGGGCGTGTCCATGGGCATCTCCGGGTCGGCGGCGGCGATGGAGACCAGCCACGCGACGCTCATGTCGAGCGACG
               OsHMA3_CDS (1921) GTGGGCGTGTCCATGGGCATCTCCGGGTCGGCGGCGGCGATGGAGACCAGCCACGCGACGCTCATGTCGAGCGACG
           BdHMA3_genomic (2321) GTGGGCGTGTCCATGGGCGTCTCGGGCTCCGCGGCCGCCATGGAGACCAGCCACGCGACGCTCATGTCCAGCGACG
               BdHMA3_CDS (1879) GTGGGCGTGTCCATGGGCGTCTCGGGCTCCGCGGCCGCCATGGAGACCAGCCACGCGACGCTCATGTCCAGCGACG
  Langdon_HMA3-A1_genomic (2327) GTGGGCGTCTCCATGGGCATCTCCGGCTCGGCCGCGGCCATGGAGACCAGCCACGCGACGCTCATGTCCAGCGACA
      Langdon_HMA3-A1_CDS (1870) GTGGGCGTCTCCATGGGCATCTCCGGCTCGGCCGCGGCCATGGAGACCAGCCACGCGACGCTCATGTCCAGCGACA
Langdon_TtHMA3-B1_genomic (2364) GTGGGCGTCTCCATGGGCATCTCCGGCTCGGCCGCGGCCATGGAGACCAGCCACGCGACGCTCATGTCCAGCGACA
    Langdon_TtHMA3-B1_CDS (1889) GTGGGCGTCTCCATGGGCATCTCCGGCTCGGCCGCGGCCATGGAGACCAGCCACGCGACGCTCATGTCCAGCGACA
Partial_CS_HMA3-D1_genomic (2340) GTGGGCGTCTCCATGGGCATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
                                                                                    Section 39
                            (2889) 2889          2900      2910      2920      2930      2940      2950       2964
           OsHMA3_genomic (2791) TGCTCAGGGTCCCCGAGGCCGTCAGGCTCGGCCGGTGCGCCCGCCGGACCATCGCCGTCAACGTGGCCGGCTCGGT
               OsHMA3_CDS (1997) TGCTCAGGGTCCCCGAGGCCGTCAGGCTCGGCCGGTGCGCCCGCCGGACCATCGCCGTCAACGTGGCCGGCTCGGT
           BdHMA3_genomic (2397) TGCTCAGGGTGCCCGAGGCCGTCAGGCTCGGGCGGCGCGCGCGCCGGACCATCGCCGTGAATCTGGTGGCCTCGGT
               BdHMA3_CDS (1955) TGCTCAGGGTGCCCGAGGCCGTCAGGCTCGGGCGGCGCGCGCGCCGGACCATCGCCGTGAATCTGGTGGCCTCGGT
  Langdon_HMA3-A1_genomic (2403) TCCTCAGGGTCCCCGAGGCCGTGAGGCTCGGCAGGCGCGCCCGCCGGACCATCGCCGTCAACATGGTGTCCTCGGT
      Langdon_HMA3-A1_CDS (1946) TCCTCAGGGTCCCCGAGGCCGTGAGGCTCGGCAGGCGCGCCCGCCGGACCATCGCCGTCAACATGGTGTCCTCGGT
Langdon_TtHMA3-B1_genomic (2440) TCCTCAGGGTCCCGGAGGCCGTGAGGCTCGGCAGGCGCGCGCGCCGGACCATCGCCGTCAACATGGCGTCCTCGGT
    Langdon_TtHMA3-B1_CDS (1965) TCCTCAGGGTCCCGGAGGCCGTGAGGCTCGGCAGGCGCGCGCGCCGGACCATCGCCGTCAACATGGCGTCCTCGGT
Partial_CS_HMA3-D1_genomic (2416) NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
                                                                                    Section 40
                            (2965) 2965  2970        2980      2990      3000      3010      3020      3030      3040
           OsHMA3_genomic (2867) GGCCGTGAAGGCCGCCGTGCTCGCGCTGGCCGCGGCGTGGCGCCCCGTGCTCTGGGCCGCCGTGCTCGCCGACGTC
               OsHMA3_CDS (2073) GGCCGTGAAGGCCGCCGTGCTCGCGCTGGCCGCGGCGTGGCGCCCCGTGCTCTGGGCCGCCGTGCTCGCCGACGTC
           BdHMA3_genomic (2473) GGGGATCAAGGCCGCCGTGTTAGTGCTCGCGGTCGCCTGGACCCCGGCGCTCTGGGCGGCGGTGCTCGCTGATGTC
               BdHMA3_CDS (2031) GGGGATCAAGGCCGCCGTGTTAGTGCTCGCGGTCGCCTGGACCCCGGCGCTCTGGGCGGCGGTGCTCGCTGATGTC
  Langdon_HMA3-A1_genomic (2479) CGCCGCCAAGGCCGCCGTCCTCGCGCTCGCGGTCGCCTGGCGCCCGGTGCTGTGGGCGGCGGTCCTCGCCGACGTG
      Langdon_HMA3-A1_CDS (2022) CGCCGCCAAGGCCGCCGTCCTCGCGCTCGCGGTCGCCTGGCGCCCGGTGCTGTGGGCGGCGGTCCTCGCCGACGTG
Langdon_TtHMA3-B1_genomic (2516) CGCCGCCAAGGCCGCCGTCCTCGCGCTCGCGGTCGCCTGGCGCCCGGTGCTGTGGGCGGCGGTCCTCGCCGACGTG
    Langdon_TtHMA3-B1_CDS (2041) CGCCGCCAAGGCCGCCGTCCTCGCGCTCGCGGTCGCCTGGCGCCCGGTGCTGTGGGCGGCGGTCCTCGCCGACGTG
Partial_CS_HMA3-D1_genomic (2492) NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
                                                                                    Section 41
                            (3041) 3041          3050      3060      3070      3080      3090      3100       3116
           OsHMA3_genomic (2943) GGGACGTGCCTCCTCGTCGTGCTCAACAGCATGACGCTGCTCAGGGAGGAGTGGAAGGGCGGCGCC---------A
               OsHMA3_CDS (2149) GGGACGTGCCTCCTCGTCGTGCTCAACAGCATGACGCTGCTCAGGGAGGAGTGGAAGGGCGGCGCC---------A
           BdHMA3_genomic (2549) GGGACGTGCCTGCTCGTCGTGCTCAATAGCATGCGGCTGCTGGGGGAGGGGAACAGAGGGGGAGGTGGGAAGAAGA
               BdHMA3_CDS (2107) GGGACGTGCCTGCTCGTCGTGCTCAATAGCATGCGGCTGCTGGGGGAGGGGAACAGAGGGGGAGGTGGGAAGAAGA
  Langdon_HMA3-A1_genomic (2555) GGGACGTGCCTGCTCGTCGTGCTCAACAGCATGCTGCTGCTGGGGGAGGGGCGCAGACGCCGCGG-----------
      Langdon_HMA3-A1_CDS (2098) GGGACGTGCCTGCTCGTCGTGCTCAACAGCATGCTGCTGCTGGGGGAGGGGCGCAGACGCCGCGG-----------
Langdon_TtHMA3-B1_genomic (2592) GGGACGTGCCTGCTCGTCGTGCTCAACAGCATGCTGCTGCTGGGGGAGCGGCGCGGACGCCGCGG-----------
    Langdon_TtHMA3-B1_CDS (2117) GGGACGTGCCTGCTCGTCGTGCTCAACAGCATGCTGCTGCTGGGGGAGCGGCGCGGACGCCGCGG-----------
Partial_CS_HMA3-D1_genomic (2568) NNNNNNNNNNNNNNNNNGTGCTCAACAGCATGCTGCTGCTGGGGGAGGGGCGCGGGCGCCGCGG-----------
                                                                                    Section 42
                            (3117) 3117          3130      3140      3150      3160      3170      3180       3192
           OsHMA3_genomic (3010) AGGAGGACGGCGCGTGCCGCGCCACGGCGAGGTCGCTGGTGATGA---GGTCCCAGCT------CGCCGCCGACTC
               OsHMA3_CDS (2216) AGGAGGACGGCGCGTGCCGCGCCACGGCGAGGTCGCTGGTGATGA---GGTCCCAGCT------CGCCGCCGACTC
           BdHMA3_genomic (2625) AGAAGGAGGAGGCTTGCCGTGCCACGGCGAGATCGCTGGAGATGAGAAGGGCTCAGCTGCTGCTCAACGCCGATGA
               BdHMA3_CDS (2183) AGAAGGAGGAGGCTTGCCGTGCCACGGCGAGATCGCTGGAGATGAGAAGGGCTCAGCTGCTGCTCAACGCCGATGA
  Langdon_HMA3-A1_genomic (2620) -GAAGGAGGAGGCGTGCCGCGCCACCGCCAGGTCGCTGGAGATGAGAAGGTCTCAGCT------CGCCGCCGTTTC
      Langdon_HMA3-A1_CDS (2163) -GAAGGAGGAGGCGTGCCGCGCCACCGCCAGGTCGCTGGAGATGAGAAGGTCTCAGCT------CGCCGCCGTTTC
Langdon_TtHMA3-B1_genomic (2657) -GAAGGAGGACGCGTGCCGCGCCACGGCCAGGTCGCTGGAGATGAGAAGGTCTCAGCT------CGCCGCTGTCTC
    Langdon_TtHMA3-B1_CDS (2182) -GAAGGAGGACGCGTGCCGCGCCACGGCCAGGTCGCTGGAGATGAGAAGGTCTCAGCT------CGCCGCTGTCTC
Partial_CS_HMA3-D1_genomic (2633) -GAAGGAGGAGGCGTGCCGCGCCACGGCCAGGTCGCTGGAGATGAGAAGGTCTCAGCT------CGCCGCCGTTTC
                                                                                    Section 43
                            (3193) 3193      3200       3210      3220      3230      3240      3250            3268
           OsHMA3_genomic (3077) CCAAGCACCCAACGCCGC---AGACGCTG-----GCGCCGCC-GGCCGTGAGCAAACGAACGGC---TGCCGTTGC
               OsHMA3_CDS (2283) CCAAGCACCCAACGCCGC---AGACGCTG-----GCGCCGCC-GGCCGTGAGCAAACGAACGGC---TGCCGTTGC
           BdHMA3_genomic (2701) TGCGTCCAACGCTGCTAGTACTACGGTGTTGTTGAGAGACCAGGTGGTGGCGGTGGAGAGAGAAATGCCATTGC
               BdHMA3_CDS (2259) TGCGTCCAACGCTGCTAGTACTACGGTGTTGTTGAGAGACCAGGTGGTGGCGGTGGAGAGAGAAATGCCATTGC
  Langdon_HMA3-A1_genomic (2689) ACCGGA---CGCTGCCACTAAAAGCGTTG-----GAAAGACG-GGCGGCGACGCACCGAAAGGC---TGCCATTGC
      Langdon_HMA3-A1_CDS (2232) ACCGGA---CGCTGCCACTAAAAGCGTTG-----GAAAGACG-GGCGGCGACGCACCGAAAGGC---TGCCATTGC
Langdon_TtHMA3-B1_genomic (2726) ATCGGATTCCGCTGCCACTAAAAGCGTTG-----GAAAGACG-GGCGGCGACGCATCGAAAGGC---TGCCATTGC
    Langdon_TtHMA3-B1_CDS (2251) ATCGGATTCCGCTGCCACTAAAAGCGTTG-----GAAAGACG-GGCGGCGACGCATCGAAAGGC---TGCCATTGC
Partial_CS_HMA3-D1_genomic (2702) ACCGGA---CGCTGCCGCTAAAAGCGTTG-----GAAAGACG-GGCGGCGACGCATCGAAAGGC---TGCCATTGC
                                                                                    Section 44
                            (3269) 3269          3280      3290      3300      3310      3320      3330       3344
           OsHMA3_genomic (3141) TGTCCGAAGCCGGGCATGTCGCCCGAGCACTCGGTTGTCATCGACATC---CGAGCCGACGGCGAGCGTCAAGAAG
               OsHMA3_CDS (2347) TGTCCGAAGCCGGGCATGTCGCCCGAGCACTCGGTTGTCATCGACATC---CGAGCCGACGGCGAGCGTCAAGAAG
           BdHMA3_genomic (2777) TGTCACAGGGAAAACGAGTCTCCTGAGAGTTCGGTTGCCATCGACATTATTCGAGTCGACGAGCGTCCGGAGGAAG
               BdHMA3_CDS (2335) TGTCACAGGGAAAACGAGTCTCCTGAGAGTTCGGTTGCCATCGACATTATTCGAGTCGACGAGCGTCCGGAGGAAG
  Langdon_HMA3-A1_genomic (2753) TGCCACAAGCCAAGCAGGTCTCCTGAGCACTCGGTTGCCATCGACGTA---CGAGTCGACGAGCAGCGTGA---AG
      Langdon_HMA3-A1_CDS (2296) TGCCACAAGCCAAGCAGGTCTCCTGAGCACTCGGTTGCCATCGACGTA---CGAGTCGACGAGCAGCGTGA---AG
Langdon_TtHMA3-B1_genomic (2793) TGCCACAAGCCAAGCAGGTCCCCTGAGCACTCGGTTGCCATCGACGTA---CGGGCCGACGAGCAGCGTGA---AG
    Langdon_TtHMA3-B1_CDS (2318) TGCCACAAGCCAAGCAGGTCCCCTGAGCACTCGGTTGCCATCGACGTA---CGGGCCGACGAGCAGCGTGA---AG
Partial_CS_HMA3-D1_genomic (2766) TGCCACAAGCCAAGCAAATCTCCTGAGCACTCGGTCGCCATCGACGTA---CGAGCCGGCGAGCAGCGTGA---AA
                                                                                    Section 45
                            (3345) 3345  3350        3360      3370      3380      3390      3400      3410      3420
           OsHMA3_genomic (3214) AGCGGCCGGCCGAGGCGGCCGTCGTTGCCAAATGTTGCGGCGGAGGCGGCGGCGAGGGCATCCGCTGCGGAGCCTC
               OsHMA3_CDS (2420) AGCGGCCGGCCGAGGCGGCCGTCGTTGCCAAATGTTGCGGCGGAGGCGGCGGCGAGGGCATCCGCTGCGGAGCCTC
           BdHMA3_genomic (2853) TGAGGAAGATTGCGGCGTGTGCGCCTGCGAAAGG---CGGAGCGGACGCCTGATGTTGCGGC---------AGCAG
               BdHMA3_CDS (2411) TGAGGAAGATTGCGGCGTGTGCGCCTGCGAAAGG---CGGAGCGGACGCCTGA-----------------------
  Langdon_HMA3-A1_genomic (2823) GGCCGACGGCCGCGACGTGTGCGCCGGCAAAAAA---GGTCGAAT--A---------------------------
      Langdon_HMA3-A1_CDS (2366) GGCCGACGGCCGCGACGTGTGCGCCGGCAAAAAA---GGTCGAAT--A---------------------------
Langdon_TtHMA3-B1_genomic (2863) GGCCGACGGCCGCGACGTGTGCGCCGGCAAAAAA---GGTCGAAGTCACCGGATCTGTCAAC---------GCTTC
    Langdon_TtHMA3-B1_CDS (2388) GGCCGACGGCCGCGACGTGTGCGCCGGCAAAAAA---GGTCGAAGTCACCGGATCTGTCAAC---------GCTTC
Partial_CS_HMA3-D1_genomic (2836) GCCCGACGGCCGCGACGTGTGCGCCGGCAAAAAA---AGTCGAGGTCACCGGATCTGTCAAC---------GGTCC
```

```
                                                                              Section 46
                         (3421) 3421     3430      3440      3450      3460      3470      3480      3496
            OsHMA3_genomic (3290) CAAGAAGCCAACTGCAACGGTCGTTGTTGCCAAATGTTGCGGGGGCGGCGGCGGCGGCGAGGGCACCAGGTGCGGC
               OsHMA3_CDS (2496) CAAGAAGCCAACTGCAACGGTCGTTGTTGCCAAATGTTGCGGGGGCGGCGGCGGCGGCGAGGGCACCAGGTGCGGC
            BdHMA3_genomic (2917) CAGCAAGGCTACTGGATGCGGCACTTTGGTCAGGCCTATTTCATCGACCCGCG-CATCATCGCTACGATGTT-GTT
               BdHMA3_CDS (2461) ----------------------------------------------------------------------------
    Langdon_HMA3-A1_genomic (2866) ------------TTCATCGAGCTGCGTGTCGGCAGGATGCTGCTCCCCGTGA------------------------
       Langdon_HMA3-A1_CDS (2409) ------------TTCATCGAGCTGCGTGTCGGCAGGATGCTGCTCCCCGTGA------------------------
  Langdon_TtHMA3-B1_genomic (2927) CGTGATGCCTGCTTCATCGAGCTGCGTGTCGGCAGGATGCTGCTCCCCGTGA------------------------
     Langdon_TtHMA3-B1_CDS (2452) CGTGATGCCTGCTTCATCGAGCTGCGTGTCGGCAGGATGCTGCTCCCCGTGA------------------------
Partial_CS_HMA3-D1_genomic (2900) CGCGGCGCCTGCTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN------------------------
                                                                              Section 47
                         (3497) 3497         3510      3520      3530      3540      3550      3560      3572
            OsHMA3_genomic (3366) GCGTCCAAGAATCCGGCAACGGCGGCCGTCGTTGCCAAATGTTGCAGCGGAGGCGGCGGCGAGGGCATCGGCTGCG
               OsHMA3_CDS (2572) GCGTCCAAGAATCCGGCAACGGCGGCCGTCGTTGCCAAATGTTGCAGCGGAGGCGGCGGCGAGGGCATCGGCTGCG
            BdHMA3_genomic (2991) CCGGCCAAATAGACAGCTAAGAA----ATTGTACACCAAAATGGCGCCCGTTGATG-GTAGGGTGAAAGTGAT---
               BdHMA3_CDS (2461) ----------------------------------------------------------------------------
    Langdon_HMA3-A1_genomic (2906) ----------------------------------------------------------------------------
       Langdon_HMA3-A1_CDS (2449) ----------------------------------------------------------------------------
  Langdon_TtHMA3-B1_genomic (2979) ----------------------------------------------------------------------------
     Langdon_TtHMA3-B1_CDS (2504) ----------------------------------------------------------------------------
Partial_CS_HMA3-D1_genomic (2952) ----------------------------------------------------------------------------
                                                                              Section 48
                         (3573) 3573   3580       3590      3600      3610      3620      3630      3648
            OsHMA3_genomic (3442) GAGCCTCCAAGAAGCCAAC-TGCAACGGCCGTTGTTGCCAAATGTTGCGG-CGGCGGCGGCGAGGGCACCAGGTGC
               OsHMA3_CDS (2648) GAGCCTCCAAGAAGCCAAC-TGCAACGGCCGTTGTTGCCAAATGTTGCGG-CGGCGGCGGCGAGGGCACCAGGTGC
            BdHMA3_genomic (3059) -AGCCCCCCGATAGACATCCTTCTAGTGTTTAAAATGATGTATTTTTCAGACAGACATTTTTTAGTCATAAATTCA
               BdHMA3_CDS (2461) ----------------------------------------------------------------------------
    Langdon_HMA3-A1_genomic (2906) ----------------------------------------------------------------------------
       Langdon_HMA3-A1_CDS (2449) ----------------------------------------------------------------------------
  Langdon_TtHMA3-B1_genomic (2979) ----------------------------------------------------------------------------
     Langdon_TtHMA3-B1_CDS (2504) ----------------------------------------------------------------------------
Partial_CS_HMA3-D1_genomic (2952) ----------------------------------------------------------------------------
                                                                              Section 49
                         (3649) 3649    3660      3670      3680      3690      3700      3710      3724
            OsHMA3_genomic (3516) GCCGCGTCCAAGAAGCCGGCAACGGCGGCCGTCGTTGCCAAATGTTGCGGTGGCGATGGCGGCGAGGGCACCGGAT
               OsHMA3_CDS (2722) GCCGCGTCCAAGAAGCCGGCAACGGCGGCCGTCGTTGCCAAATGTTGCGGTGGCGATGGCGGCGAGGGCACCGGAT
            BdHMA3_genomic (3134) TACACA-CAAAT----------------------------------------------------------------
               BdHMA3_CDS (2461) ----------------------------------------------------------------------------
    Langdon_HMA3-A1_genomic (2906) ----------------------------------------------------------------------------
       Langdon_HMA3-A1_CDS (2449) ----------------------------------------------------------------------------
  Langdon_TtHMA3-B1_genomic (2979) ----------------------------------------------------------------------------
     Langdon_TtHMA3-B1_CDS (2504) ----------------------------------------------------------------------------
Partial_CS_HMA3-D1_genomic (2952) ----------------------------------------------------------------------------
                                                                              Section 50
                         (3725) 3725 3730      3740      3750      3760      3770      3780      3790      3800
            OsHMA3_genomic (3592) GTGGTGCCTCCAAGAGGTCGCCGCCGGCTGAGGGAAGCTGCAGCGGCGGTGAAGGCGGTACCAATGGTGTTGGTCG
               OsHMA3_CDS (2798) GTGGTGCCTCCAAGAGGTCGCCGCCGGCTGAGGGAAGCTGCAGCGGCGGTGAAGGCGGTACCAATGGTGTTGGTCG
            BdHMA3_genomic (3145) ----------------------------------------------------------------------------
               BdHMA3_CDS (2461) ----------------------------------------------------------------------------
    Langdon_HMA3-A1_genomic (2906) ----------------------------------------------------------------------------
       Langdon_HMA3-A1_CDS (2449) ----------------------------------------------------------------------------
  Langdon_TtHMA3-B1_genomic (2979) ----------------------------------------------------------------------------
     Langdon_TtHMA3-B1_CDS (2504) ----------------------------------------------------------------------------
Partial_CS_HMA3-D1_genomic (2952) ----------------------------------------------------------------------------
                                                                              Section 51
                         (3801) 3801     3810      3820      3830      3840      3850      3860      3876
            OsHMA3_genomic (3668) TTGCTGCACGAGCGTGAAGAGGCCAACCTGTTGCGACATGGGAGCGGCGGAGGTGTCCGATTCTTCGCCGGAGACG
               OsHMA3_CDS (2874) TTGCTGCACGAGCGTGAAGAGGCCAACCTGTTGCGACATGGGAGCGGCGGAGGTGTCCGATTCTTCGCCGGAGACG
            BdHMA3_genomic (3145) ----------------------------------------------------------------------------
               BdHMA3_CDS (2461) ----------------------------------------------------------------------------
    Langdon_HMA3-A1_genomic (2906) ----------------------------------------------------------------------------
       Langdon_HMA3-A1_CDS (2449) ----------------------------------------------------------------------------
  Langdon_TtHMA3-B1_genomic (2979) ----------------------------------------------------------------------------
     Langdon_TtHMA3-B1_CDS (2504) ----------------------------------------------------------------------------
Partial_CS_HMA3-D1_genomic (2952) ----------------------------------------------------------------------------
                                                                              Section 52
                         (3877) 3877         3890      3900      3910      3920      3930      3942
            OsHMA3_genomic (3744) GCGAAAGACTGCAGAAATGGAGGTGTTGCGCGAAGACGATGAACTCCGGTGAAGTGAAAGGATGA
               OsHMA3_CDS (2950) GCGAAAGACTGCAGAAATGGAGGTGTTGCGCGAAGACGATGAACTCCGGTGAAGTGAAAGGATGA
            BdHMA3_genomic (3145) -----------------------------------------------------------------
               BdHMA3_CDS (2461) -----------------------------------------------------------------
    Langdon_HMA3-A1_genomic (2906) -----------------------------------------------------------------
       Langdon_HMA3-A1_CDS (2449) -----------------------------------------------------------------
  Langdon_TtHMA3-B1_genomic (2979) -----------------------------------------------------------------
     Langdon_TtHMA3-B1_CDS (2504) -----------------------------------------------------------------
Partial_CS_HMA3-D1_genomic (2952) -----------------------------------------------------------------
```

Appendix 7 Alignment of HMA3-A1 and HMA3-B1 predicted protein sequences from durum wheat. Amin acids color coded based on similarity: yellow – identical, blue – conserved, green – similar, white – non-similar

```
                                                                                                    Section 1
                          (1) 1        10        20        30        40        50        60        70      80
        Langdon HMA3-A1   (1) MGGGESYPALEASLLAEEAAARRQWEKTYLDVLGVCCSAEVALVERLLAPLDGVRAVAVVVPSRTVVVEHDPAAVSQSRI
   W9262-260D3_HMA3-B1    (1) MGGGESYAALEESLLPEQAAARRQWEKTYLDVLGVCCSAEVALVERLLAPLDGVRAVAVVVPSRTVVVEHDPAAVSQSRI
   Langdon/Kofa_HMA3-B1   (1) MGGGESYAAASRTRRSRRAFCQSKRRRGGSGRRRTWTCWACAAPRRSRWSSGCWRRSTA---------------------
                                                                                                    Section 2
                         (81) 81       90        100       110       120       130       140       150     160
        Langdon HMA3-A1  (81) VKVLNGAGLEASVRAYGSSGFIGRRPSPYIVACGALLLASSFRWLLLPLQWLALGAACAGAPPMVLRGLAAASRLALDIN
   W9262-260D3_HMA3-B1   (81) VKVLNGAGLEASVRAYGSSGFIGRRPSPYIVACGALLLASSFRWLLLPLQWLALGAACAGAPPMVLRGLAAASRLALDIN
   Langdon/Kofa_HMA3-B1  (60) --------------------------------------------------------------------------------
                                                                                                    Section 3
                        (161) 161      170       180       190       200       210       220       230     240
        Langdon HMA3-A1 (161) ILMLIAVAGAVALKDYTEAGVIVFLFTTAEWLETLACTKASAGMSSLMSMIPPKAVLAETGEVVNVRDIGVGVVIAVRAG
   W9262-260D3_HMA3-B1  (161) ILMLIAVAGAVALKDYTEAGVIVFLFTTAEWLETLACTKASAGMSSLMSMIPPKAVLAETGEVVNVRDIGVGVVIAVRAG
   Langdon/Kofa_HMA3-B1  (60) --------------------------------------------------------------------------------
                                                                                                    Section 4
                        (241) 241      250       260       270       280       290       300       310     320
        Langdon HMA3-A1 (241) EMVPVDGVVVDGQSEVDERSLTGESYPVPKQPQSEVWAGTLNLDGYIAVRTMALAENSTVAKMERLVEEAQQSKSKTQRL
   W9262-260D3_HMA3-B1  (241) EMVPVDGVVVDGQSEVDERSLTGESYPVPKQPQSEVWAGTLNLDGYIAVRTMALAENSTVAKMERLVEEAQQSKSRTQRL
   Langdon/Kofa_HMA3-B1  (60) --------------------------------------------------------------------------------
                                                                                                    Section 5
                        (321) 321      330       340       350       360       370       380       390     400
        Langdon HMA3-A1 (321) IDSCAKYYTPAVVVLGAGVALLPPLLGARDAERWFRLALVLLVSACPCALVLSTPVATFCALLTAARMGVLVKGGDVLES
   W9262-260D3_HMA3-B1  (321) IDSCAKYYTPAVVVLGAGVALLPPLLGARDAERWFRLALVLLVSACPCALVLSTPVATFCALLTAARMGVLVKGGDVLES
   Langdon/Kofa_HMA3-B1  (60) --------------------------------------------------------------------------------
                                                                                                    Section 6
                        (401) 401      410       420       430       440       450       460       470     480
        Langdon HMA3-A1 (401) LGEIRAVAFDKTGTITRGEFTVDMFDVVEQKVQMSHLLYWISSIESKSSHPMAAALVEHAQSKSIQPKPECVAEFRILPG
   W9262-260D3_HMA3-B1  (401) LGEIKAVAFDKTGTITRGEFTVDMFDVVGHKVQMSHLLYWISSIESKSSHPMAAALVEHARSKSIEPKPECVAEFRVLPG
   Langdon/Kofa_HMA3-B1  (60) --------------------------------------------------------------------------------
                                                                                                    Section 7
                        (481) 481      490       500       510       520       530       540       550     560
        Langdon HMA3-A1 (481) EGVYGEIDGKRIYVGNKRVLARGSSCQTVPERMNGLKGVSIGYVICDGDLVGVFSLSDDCRTGAAEAIRELASMGISSVL
   W9262-260D3_HMA3-B1  (481) EGIYGEIDGMRIYVGNKRVLARGSSCQTVPERMNGLKGVSMGYVICDGDLVGVFSLSDDCRTGAAEAIRELASMGISSVL
   Langdon/Kofa_HMA3-B1  (60) --------------------------------------------------------------------------------
                                                                                                    Section 8
                        (561) 561      570       580       590       600       610       620       630     640
        Langdon HMA3-A1 (561) LTGDSAEAAVHAQERLGGAL-EELHSELFPEDKVRLVGALKARAGPTMMVGDGMNDAPALATADVGVSMGISGSAAAMET
   W9262-260D3_HMA3-B1  (561) LTGDIAEAAMHAQEQLGGALLEEVHSELFPEDKVRLVGALKARAGPTMMVGDGMNDAPALATADVGVSMGISGSAAAMET
   Langdon/Kofa_HMA3-B1  (60) --------------------------------------------------------------------------------
                                                                                                    Section 9
                        (641) 641      650       660       670       680       690       700       710     720
        Langdon HMA3-A1 (640) SHATLMSSDILRVPEAVRLGRRARRTIAVNMVSSVAAKAAVLALAVAWRPVLWAAVLADVGTCLLVVLNSMLLLGEGRRR
   W9262-260D3_HMA3-B1  (641) SHATLMSSDILRVPEAVRLGRRARRTIAVNMASSVAAKAAVLALAVAWRPVLWAAVLADVGTCLLVVLNSMLLLGERRGR
   Langdon/Kofa_HMA3-B1  (60) --------------------------------------------------------------------------------
                                                                                                    Section 10
                        (721) 721      730       740       750       760       770       780       790     800
        Langdon HMA3-A1 (720) RGKEEACRATARSLEMRRSQLAAVSPD-AATKSVGKTGGDAPKGCHCCHKPSRSPEHSVAIDVRVDEQREGPTAATCAPA
   W9262-260D3_HMA3-B1  (721) RGKEDACRATARSLEMRRSQLAAVSSDSAATKSVGKTGGDASKGCHCCHKPSRSPEHSVAIDVRADEQREGPTAATCAPA
   Langdon/Kofa_HMA3-B1  (60) --------------------------------------------------------------------------------
                                                                                                    Section 11
                        (801) 801      810               828
        Langdon HMA3-A1 (799) KKVEYS-----------SSCVSAGCCSP
   W9262-260D3_HMA3-B1  (801) KKVEVTGSVNASVMPASSSCVSAGCCSP
   Langdon/Kofa_HMA3-B1  (60) ----------------------------
```

Appendix 8 Origin and pedigree of global collection of durum wheat breeding lines and cultivars (Modified from Reimer et al. 2008)

Accession	Origin	Pedigree
Bonaerance Inta Cumenay	Argentina	Unknown
Bonaerance Quilaco	Argentina	MAGH72//GS/AA///RABI//D21563/AA
Bonaerance Valverde	Argentina	GIORGIO//CAPELLI/YUMA
Buck Ambar	Argentina	Unknown
Buck Topacio	Argentina	Unknown
Ocotillo	Arizona	Unknown
920334	Australia	69850/ 86014
940030	Australia	Unknown
940435	Australia	Unknown
950329	Australia	Unknown
950090	Australia	Unknown
950844	Australia	Unknown
940955	Australia	Unknown
Tamaroi	Australia	RUFF/FLAMINGO-DW//MEXICALI-75///SHEARWATER/56113/TAM-1-B-17/KAMILAROI/56112/WELLS/56111//GUILLEMOT
Wollaroi	Australia	TAM-1-B-17/(SIB)KAMILAROI//ROKEL(S)/(SIB)KAMILAROI
AC Morse	Canada	RL 7196/DT 610
AC Napoleon	Canada	VIC/DT384//DT 471
9661-AF1D	Canada	W9262-260D3/ARUBA//DT 662
9661-CA5E	Canada	W9262-260D3/ARUBA//DT 662
AC Avonlea	Canada	8267-AD2A/DT 61
AC Melita	Canada	MEDORA/LLOYD
AC Navigator	Canada	KYLE/WESTBRED 881
AC Pathfinder	Canada	WESTBRED 881/DT 367
DT691	Canada	DT618/ 8667-D216C//DT 637
DT695	Canada	DT 471/2*KYLE
DT696	Canada	DT618/DT 637//KYLE
Kyle	Canada	6962-92-8-5/ 6965-494-
Commander	Canada	W9260-BK03/AC NAVIGATOR//AC PATHFINDER
DT704	Canada	AC AVONLEA/DT 665

Accession	Origin	Pedigree
DT705	Canada	AC AVONLEA/DT 665
DT707	Canada	AC AVONLEA/DT 665
DT709	Canada	DT 674/DT 665
DT710	Canada	DT618/GREEN 27
DT711	Canada	WESTBRED 881/W9260-BK03
Strongfield	Canada	AC AVONLEA/DT 665
D24-1773	Canada	DT 520/D94078
DT513	Canada	DT 625/DT 612
DT532	Canada	D92269/D92413
DT536	Canada	D94350/D93108
CDC Verona	Canada	D95253/D95116
Carioca	France	CID 479402
RABD 93.40	France	Unknown
Tetradur	France	EDMORE//CAPDUR/REGAL
Durabon	Germany	SIGNADUR/EDM//P 4312.86
Durafit	Germany	Unknown
44616	Iran	Unknown
44721	Iran	Unknown
CRDW17	Iran	Unknown
D-73-15	Iran	Unknown
Simeto	Italy	CAPEITI/VALNOVA
Colosseo	Italy	CRESO/MEXA
Duilio	Italy	CAPPELLI//ANHINGA/FLAMINGO
Grazia	Italy	ISWRN-21/VALSELVA
Fortore	Italy	CAPEITI 8/VALFORTE
Flavio	Italy	LATINO/CAPPELLI
Lesina	Italy	Unknown
Varano	Italy	CAPEITI 8/CRESO//CRESO///VALFORTE/TRINAKRIA
Bronte	Italy	BERILLO/LATINO
Ciccio	Italy	APPULO/VALNOVA//VALFORTE/PATRIZIO
Demetra	Italy	MESSAPIA/GIOIA
Gianna	Italy	Unknown

Accession	Origin	Pedigree
Iride	Italy	ALTAR 84/ARES-SIB
Medda	Italy	TRINAKRIA/VALFORTE
Mongibello	Italy	TRINAKRIA/VALFORTE
Parsifal	Italy	INRA92-1/D81028
Svevo	Italy	SELEZIONE CIMMYT/WB881
Tresor	Italy	AMBER-DURUM/S-22-80
Green27	Mexico	STERNA- DW 2/GRAVELOTE
Green34	Mexico	STERNA- DW 2/GRAVELOTE
Nacori97a	Mexico	ALTAR 84/CMH82A.1062//CD58230- ?
Gidara 17a	Morocco	Unknown
Marjak	Morocco	Unknown
DHTON1	Morocco	Unknown
D940027	ND-USDA	D88104/D88207
D940098	ND-USDA	D88450/D87436
D941038	ND-USDA	D86117/D88289
D95580	ND-USDA	BELZER/D88058//D88276
Plaza	ND-USDA	PLENTY/D8291
Arrivato	New Zealand	Unknown
CFR5001	New Zealand	Unknown
K-39099	Russia	LV-URAZOVSKII R-N,VORONEZHSKAYA OBL
Agridur	Spain	EDMORE//CIMMYT 303/CHANDUR
Altar-Aos	Spain	Altar/Aos
Arcobelano	Spain	CHEN/ALTAR 84
Ariesol	Spain	Unknown
Borli	Spain	Unknown
Camacho	Spain	Unknown
Gallareta	Spain	RUFF/FLAMINGO-DW//MEXICALI-75/3/SHEARWATER/4/?
Mexa	Spain	GDOVZ469///JO 1//61.130/LDS
Vitron	Spain	TURCHIA-77///JORI-SIB/ANHINGA-SIB//FLAMINGO-SIB
Durex	U.S.	AZ-MFSRS-86
Langdon	U.S.	LDN240/KHAPLI//LANGDON 308///MINDUM*3/VERNAL/4/VERNAL EMMER/3*MINDUM
Langdon(DIC-6B)	U.S.	LANGDON/*Triticum dicoccoides* (disomic chromosome substitution line)
Westbred881	U.S.	WARD/WLS//CNDO/WCA///MEXI/WB1000

Accession	Origin	Pedigree
Kofa	U.S.	DICOCCUM ALPHA
Kronos	U.S.	APB MSFRS POP SEL (D03-21)