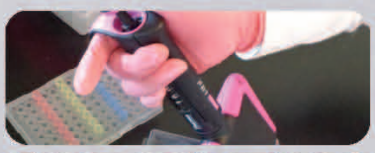
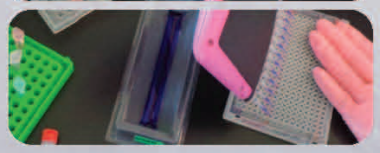
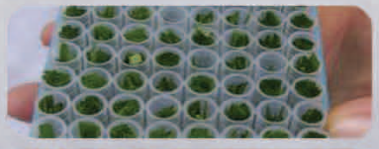
CIMMYT Wheat

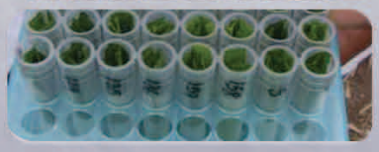
Molecular Genetics:

Laboratory Protocols and Applications to Wheat Breeding









CIMMYT MR
International Maize and Wheat Improvement Center

CIMMYT Wheat Molecular Genetics:

Laboratory Protocols and Applications to Wheat Breeding

Dreisigacker, S., Sehgal D., Reyes Jaimez A.E., Luna Garrido B.,
Muñoz Zavala S., Núñez Ríos C., Mollins J., Mall S. (Eds.)

Acknowledgements

The authors specially thank the Agri-Food System CRP WHEAT, the Bill and Melinda Gates Foundation (BMGF), the United States Agency for International Development (USAID), the Grain Research and Development Corporation (GRDC) for their support to molecular genetics initiatives at CIMMYT.

Correct citation: Dreisigacker, S., Sehgal D., Reyes Jaimez A.E., Luna Garrido B., Muñoz Zavala S., Núñez Ríos C., Mollins J., Mall S. (Eds.) (2016) *CIMMYT Wheat Molecular Genetics: Laboratory Protocols and Applications to Wheat Breeding*. Mexico, D.F.: CIMMYT.

AGROVOC Descriptors: Agricultural research; Agricultural engineering; Wheats; Plant breeding; Plant biotechnology; Breeding methods; Genotypes; Genetic markers; Laboratory equipment; Laboratory experimentation; Laboratory techniques.

AGRIS Category Codes: F30 Plant Genetics and Breeding

Additional Keywords: CIMMYT

Dewey Decimal Classification: 631.5233 DRE

ISBN: 978-607-8263-57-8

Design and layout: Miguel Mellado

Printed in México

INDEX

Appendices

Foreword

It is generally agreed that in order to meet future challenges in food production, multi-faceted breeding approaches are needed, including the use of current available genomics resources. Since more than three decades, molecular markers have acted as a versatile genomics tool for fast and unambiguous genetic analysis of plant species of both diploid and polyploid origin. With the enormous reductions in sequencing cost, the number of molecular markers, even for a more complex genome like wheat are abundant, economical, protocols are robust and high throughput. E.g. in wheat, fewer than 500 SNP markers were available in 2008, with the number increased to 1536 in 2010, 10,000 in 2011, 90,000 in 2012 and 820,000 in 2014 (Bevan and Uauy 2013). This is currently leading to the development of high-resolution genetic maps and the increased exploitation of genetic linkages between markers and important economic traits in bread, durum wheat and its wild relatives. QTL discovery and candidate gene identification will further be accelerated with the availability of the high quality reference sequence of the wheat genome (Choulet et al. 2014).

This manual describes the use of molecular markers in wheat breeding with emphasis on the status of marker-assisted selection (MAS) at CIMMYT. Together with decreasing marker assay costs and interconnected genotyping service facilities, the opportunity to apply MAS strategies is becoming accessible to more and more breeding programs. We have not attempted a comprehensive review of the literature related to the future potential of genomics resources in wheat improvement nor on the extensive availability of molecular biology techniques. In the context of wheat production challenges, this manual seeks to provide a practical guide and insights into the current use of molecular markers as a progressing selection tool in the hands of public program wheat breeders.

In Part 1 of the manual we describe various experimental protocols used in our laboratory for MAS, ranging from DNA extraction to polymerase chain reaction, gel and fluorescence detection methods, which are presented so as to be readily usable at the laboratory bench. These step-by-step protocols are intended to be concise and easy to follow. Suggestions to successfully apply the procedures are included, along with the recommended materials and suppliers. Some of the protocols described here are new; others are quite old. We have included the latter because, though they may be phased out in the future, they continue to be useful. Successive chapters deal with primer design protocols. The second part of the manual target marker deployment in the CIMMY breeding program. A number of chapters on QTL/gene identification approaches, how to optimize MAS strategies, how MAS is currently used at CIMMYT for major trait categories such as biotic stresses and quality traits are described and we share our experience on recently developed prediction methods using genome-wide markers to archive genetic gain for more complex traits. The chapters for the manual are written by a group of international scientists who are using molecular markers in wheat genome research and breeding at CIMMYT. Final appendices provide list of molecular markers currently used at CIMMYT, and the links to the useful websites and software packages with their characteristic features briefly described.

We encourage readers, especially those who have found the manual useful, to send us their comments. We also welcome any corrections and suggestions for improvement that may contribute to the success of future versions of this manual.

Please address your comments to: Wheat Molecular Breeding Laboratory
CIMMYT, Apdo. Postal 6-641
06600 Mexico, D.F., Mexico
Phone: +52 (55) 5804-2004
Email: sdreisigacker@cgiar.org

References:

Bevan M.W., C. Uauy. 2013. Genomics reveals new landscapes for crop improvement. Genome Biol, 14:206.

Choulet F., M. Caccamo, J. Wright, M. Alaux, H. Šimková, J. Šafář, P. Leroy, J. Dolezel, J. Rogers, K. Eversole, C. Feuillet. 2014. The wheat black jack: Advances towards sequencing the 21 chromosomes of bread wheat. In: R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources*, Springer Science, Netherlands, pp. 405-438.

Abbreviations/Acronyms

BME	β-Mercaptoethanol
BPB	Bromophenol blue
BSA	Bovine serum albumin
bp	base pairs
CTAB	Cetyltrimethyl/ammonium/bromide
cm	Centimeter
dATP	Deoxyadenosine 5’-triphosphate
dCTP	Deoxycytidine 5’-triphosphate
ddH_2O	double-distilled water
dGTP	Deoxyguanosine 5’-triphosphate
dH_2O	distilled water
DNA	Deoxyribose nucleic acid
dNTPs	Deoxynucleoside 5’-triphosphates
dTTP	Deoxythymidine triphosphate 5’-triphosphates
EDTA	Ethylenediaminetetraacetate
EtBr	Ethidium bromide
EtOH	Ethanol
g	gram(s)
h	hour(s)
HCl	Hydrochloride
HYB	hybridization
kb	kilobases
kcal	kilocalories
KOAc	Potassium acetate
LMP	low melting point
mA	milli Amperes
mBar	millibars
mg	microgram(s)= 10^{-6} gram
mg	milligram(s) = 10^{-3} gram
min	minute(s)
ml	milliliter(s) = 10^{-3} liter
mm	millimeter(s) = 10^{-3} meter
mM	milliMol
mol	
µl	microliter(s) = 10^{-6} liter
µM	micromolar
MW	molecular weight
ng	nanogram(s) = 10^{-9} gram
nM	nanomol
NaCl	Sodium chloride
NaOAc	Sodium acetate
NaOH	Sodium hydroxide
NH_4OAc	Ammonium acetate
OD	optical density unit
ODx	optical density unit at (260/280) nm
PCR	polymerase chain reaction
pH	potential of hydrogen
pmole	picomole
RNA	Ribonucleic acid
R×N	reaction(s)
rpm	rounds per minute
RT	room temperature
sec	second(s)
SGB	sample gel buffer
STE	Sodium Tris-EDTA (also TEN)
TAE	Tris-acetate EDTA (buffer)
TBE	Tris-borate EDTA
TE	Tris-EDTA (buffer)
TG	Tris-glycine buffer
T_m	melting temperature
TNE	Tris Sodium (Na) EDTA (buffer)
Tris	Tris (hydroxymethyl) amino-methane
Tris-HCL	Tris hydrochloride
TTE	Triton Tris-EDTA (buffer)
U	unit(s) of enzyme
UV	ultraviolet
V	volts
[FINAL]	FINAL concentration
[Stock]	stock concentration
°C	degree Celsius

Introduction

Susanne Dreisigacker

Since the early beginnings of plant domestication, plant breeding has been extremely successful in developing crops and varieties and contributed to the dramatic improvement of yield, nutritional qualities and other traits of value (Harlan, 1992). During the last century the success of plant breeding has mainly relied on the utilization of natural and mutant induced genetic variation and in the efficient selection of favorable genetic combinations by using suitable breeding methods. The evaluation and identification of genetic variants of interest as well as the selection methods used have largely been based on phenotypic evaluation.

Human population is rapidly growing and the production of high-quality food must increase with reduced inputs, an accomplishment that will be in particular challenging in the face of globally dynamic environmental changes and biotic threats (FAO, 2013). The wheat demand is predicted to increase by 40 percent by 2050 (Ray et al. 2013). Therefore, intensive investment in wheat research and more than traditional breeding approaches are required. Plant breeding must focus on traits with the greatest potential to increase yield. New strategies must be developed to increase and utilize the available genetic diversity in breeding germplasm and to accelerate breeding results.

Genomic tools provide breeders with a new set of opportunities for improving and accelerating genetic gains. While classical genetics (the fundamental association between genotype and phenotype first described by Mendel) revolutionized plant breeding at the beginning of the 20th century (Shull et al. 1909), plant genomics today has the potential to significantly contribute to crop improvement by the ability to dissect the molecular and genetic basis of traits and the characterization of whole genomes. The field of genomics and its applications are developing very quickly. Extensive knowledge from the analyses of genomes combined with traditional breeding methods can hopefully lead to more rapid and efficient plant breeding.

One of the main pillars of current plant genomics is the development of high-throughput DNA sequencing technologies also known as next generation sequencing (NGS), which provide i) large collections of SNP markers, ii) high-throughput genotyping technologies, iii) high density genetic maps and iv) transcriptome or whole genome sequences that can be incorporated into modern breeding. First developed for the human genome, SNP markers have been proven universal and are the most abundant forms of genetic variation among individuals within a species (Rafalski 2002). Although SNP markers are less polymorphic than e.g., SSRs markers because of their bi-allelic nature, they easily compensate this drawback by being abundant, ubiquitous and amenable to high and ultra-high-throughput automation. E.g. the latest array-based genotyping option in wheat contains 820K SNP markers (820K Axiom® Array probe set, http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/axiom_820K_search.php) in comparison to approximate 4 to 5000 wheat SSR markers published during the last three decades. Coupled with the genome sequence, NGS technologies open new ways of high-throughput genotyping for more effective genetic mapping and genome analysis (Poland et al. 2012). Genotyping-by-sequencing (GBS) increases sequencing throughput and allows sequencing of large numbers of samples using multiplexed sequencing, thus provides dense and unbiased marker coverage, high mapping accuracy and resolution for more comparable genome and genetic maps among mapping populations (Elshire et al. 2011). NGS data derived from targeted regions can also be used for variant detection in large datasets, identification of gene analogs and paralogs or SNP discovery in QTL regions (Clarke et al. 2014, Rife et al. 2015). SNP detection in the transcriptome of non-sequenced species provides an opportunity using NGS to additionally accelerate fine-mapping and cloning of genes having an impact on our ability to understand the fundamental biology of a trait and enabling wheat breeders to directly access the genetic variation in genes (Ramirez-Gonzalez et al. 2015). NGS technologies also supported the building of a first survey sequence in wheat that produced new estimates of the gene content, a better understanding of the phylogenic history of wheat and gene expression divergence among the three sub-genomes (Eversole et al. 2014).

An already known breeding and selection strategy that involves using genetic information is marker-assisted selection (MAS, Lande and Thompson, 1990). Marker-assisted selection is an indirect process where selection is carried out on the bases of a marker instead of the trait itself. The successful application of MAS especially relies on the tight association between the marker and the major gene or QTL responsible for the trait. The plant genomics enabled accelerated identification of markers tightly linked to target genomic regions will provide the ability to deploy MAS for a larger number of traits. The identification of intragenic marker (functional marker) can additionally reduce limitations due to recombination. The most useful application of MAS is the process of pyramiding genes via forward selection. The accumulation of genes from different sources conferring resistance against the same disease is an example and is one of the most widespread applications of gene pyramiding in wheat. Beside the growing number of tightly linked markers offered, the main benefit of the recent advantages in genomics is the accessibility to new genotyping platforms that allow to screen very large numbers of plants in a gene pyramiding program, which 1) is required to ensure with reasonable likelihood that the genotype combining favorable alleles is present, as the number of loci of interest is increasing, 2) can provide the desired genetic information in a more rapid, logistically easier and economically feasible manner.

Marker-assisted backcrossing facilitates the quick recovery of the recurrent parent genome. Foreground selection is being used extensively in wheat breeding making use of genomic resources, e.g., for the introgression of disease resistances genes observed in wild species. The application of efficient background selection strategies for gene introgression has still been somewhat limited by the high costs of marker analysis. Selection strategies combining single marker assays with high-throughput SNP assays have shown potential to greatly increase the efficiency and flexibility of marker-assisted background selection (Herzog et al. 2011). In some cases, the problem of recovering the genetic background of the recurrent parent arises because of linkage drag (the introgression of chromosome regions with deleterious effects tightly linked to the gene of interest). Dense genetic maps used in background selection can be an efficient way to break the genetic linkage drag. In summary, MAS can take benefit from new genomic technologies, speeding up the release of new varieties.

Current MAS strategies in the breeding programs fit for traits with high heritability and governed by a few major QTL that explain a large proportion of the phenotypic variability. However, the application of MAS for breeding traits with complex genetics based on the interaction of multiple QTL with minor and environmental effects has been inefficient. Examples of such complex traits are yield, drought tolerance or nitrogen and water use efficiency. Meuwissen et al. (2001) described a new methodology in plant breeding called genomic selection (GS) that is believed to solve problems related to MAS for complex traits. In GS high density marker coverage is needed to potentially have all QTL in linkage disequilibrium with at least one marker. The comprehensive information on all possible loci, haplotypes and marker effects across the entire genome is used to calculate the genomic estimated breeding values of a particular line in a breeding population.

The obvious advantages of GS over traditional MAS have been successfully proven in animal breeding (Hayes and Goddard, 2010) and initially in plant breeding including wheat (Crossa et al. 2014). The rapid evolution of NGS technologies are enabling generation and validation of millions of markers giving a cautious optimism for successful application of GS in plant breeding for complex traits.

In conclusion, considering current application levels and success in various crops, MAS deployment is getting increasingly attractive for the 21st century breeding and can be further advanced. Successful efforts of additionally incorporating large scale high-throughput genotyping are currently shifting the MAS theory from the transfer of larger-effect genes to predicting the performance of both phenotyped and unphenotyped individuals. A key challenge is the integration and interpretation of the massive amounts of data that is being generated and that needs to be addressed to reap the full potential of genomics.

References

Clarke, L. J., P. Czechowski, J. Soubrier, M. I. Stevens and A. Cooper. 2014. Modular tagging of amplicons using a single PCR for high-throughput sequencing. Molecular ecology resources, 14 (1): 117-121.

Crossa J., P. Pérez, J. Hickey, J. Burgueño, L. Ornella, J. Cerón-Rojas, X. Zhang, S. Dreisigacker, R. Babu, Y. Li, D. Bonnett and K. Mathews. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112: 48–60.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, , K. Kawamoto, E. S. Buckler and S. E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one, 6(5): e19379.

Eversole, K., C. Feuillet, K. F. Mayer and J. Rogers. 2014. Slicing the wheat genome. Science, 345(6194): 285-287.

Food and Agriculture Organisation (FAO) of the United Nations, Declaration of the World Summit on Food Security, Rome, 16–18 November 2009 (www.fao.org/wsfs/world-summit/en/).

Harlan, J.R. 1992. Crops and Man. American Society of Agronomy and Crop Science Society of America, Madison, WI.

Hayes, B. and M. Goddard. 2010. Genome-wide association and genomic selection in animal breeding. Genome, 53(11): 876-883.

Herzog, E. and M. Frisch. 2011. Selection strategies for marker-assisted backcrossing with high-throughput marker systems. Theoretical and applied genetics, 123(2): 251-260.

Lande, R. and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics, 124(3): 743-756.

Meuwissen, T.H.E., B.J. Hayes and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829.

Poland, J. A., P. J. Brown, M. E. Sorrells and J. L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PloS one, 7(2): e32253.

Rafalski, J. A. 2002. Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant science, 162(3): 329-333.

Ramirez-Gonzalez, R. H., V. Segovia, N. Bird, P. Fenwick, S. Holdgate, S. Berry and C. Uauy. 2015. RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. Plant biotechnology journal, 13(5): 613-624.

Ray D.K., N.D. Mueller, P.C. West, J.A. Foley. 2013. Yield trends are insufficient to double crop production by 2015. PLOS One 8 (6): e66428.

Rife, T. W., S. Wu, R. Bowden and J. A. Poland. 2015. Spiked GBS: a unified, open platform for single marker genotyping and whole-genome profiling. BMC genomics, 16(1): 248.

Shull G.H. 1909. A pure-line method in corn breeding. Am Breeders Assoc Rep 5: 51–59.

Part 1: Laboratory Protocols

PART 1 describes various experimental protocols used in our laboratory for MAS. These step-by-step protocols are intended to be concise and easy to follow. Suggestions to successfully apply the procedures are included, along with the recommended materials and suppliers. Some of the protocols described here are new; others are quite old. We have included the latter because, though they may be phased out in the future, they continue to be useful.

1. Plant Genomic DNA Isolation

1.1 DNA Extraction from Large Amounts of Lyophilized Tissue (15 ml tubes)

To extract DNA from large amounts of lyophilized tissue (300-400 mg), use 15 ml tubes and proceed as follows:

Lyophilization

1. Harvest 10-15 dry leaves from the greenhouse or field grown plants. It is preferable to use young leaves without necrotic areas or lesions, although older leaves which are not senescent may be used.

2. If the midrib is thick and tough, remove it. Cut or fold leaves into 10-15 cm sections and place in a waxed paper bag along with the tag identifying the sample. Place your paper bags in bigger plastic bags and then in an ice chest or other container with ice to keep samples cool (but do not allow them to freeze). Make sure samples do not get wet.

3. Place your plastic bags at -80°C until ready to be lyophilized, but minimum 4 h.

4. Transfer frozen leaf samples in the waxed paper bags to a lyophilizer (e.g., Labconco). Make sure the lyophilizer is down to the recommended temperature (usually the chamber is ≤ -50°C) and is pulling a good vacuum before loading the samples. Do not overload your lyophilizer: make sure the vacuum is always between 0.0 and 0.120 mBar. Samples should get dried during 72 h. Typically, fresh weight ≈ 10X dry weight.

5. Dried leaf samples may be stored in sealed plastic bags at room temperature for a few days or, preferably, at -20°C for several years.

Note: If samples are not fully dried before grinding, grinding will be inefficient and DNA yield will be poor. Alternatively, leaf samples can be tried with silica or at low temperature in an oven.

Grinding

1. Grind the dried leaf samples to a fine powder with a coffee miller (e.g., Braun KSM-2 Coffee Grinder) into a plastic scintillation vial or any other appropriate plastic container that can be closed airtight. The finer the grind, the greater the yield of DNA from a given amount of material.

2. Store ground samples tightly capped at -20°C. Samples are stable for several years.

Genomic DNA Isolation

(Based on method of Saghai-Maroof et al. 1984)

1. Weight 300-400 mg of ground, lyophilized tissue, into a 15 ml polypropylene centrifuge tube. DNA yields range from 50 to more than 100 µg DNA/100 mg dry tissue.

2. Add 9.0 ml of warm (65°C) CTAB extraction buffer to the 300-400 mg ground. It is best to distribute tissue along the sides of the tube before adding buffer, to avoid clumping of dry tissue in the bottom. Mix several times by gentle inversion.

3. Incubate for 60-90 min, with continuous gentle rocking in a 65°C oven.

4. Remove the tubes from the oven, wait 4-5 min for the tubes to cool down, and then add 4.5 ml chloroform/octanol (24:1). Rock gently to mix for 15 min.

5. Spin in a table-top centrifuge for 30 min at 3750 rpm at 4 °C to generate a yellow aqueous phase and a green organic phase.

6. Pour off top yellow aqueous layer into new 15 ml tubes. Add 4.5 ml chloroform/octanol (24:1) and rock gently for 15 min at RT.

7. Spin in a table-top centrifuge for 30 min at 3750 rpm at RT.

8. Pipette top aqueous layer into new 15 ml tubes containing 40 µl of 10 mg/ml RNase A. Mix by gentle inversion and incubate for 30 min at RT.

9. Add 6.0 ml of isopropanol (2-propanol 18L Baker 9084-18). Mix by very gentle inversion until you see the DNA strand.

10. Remove the precipitated DNA with a glass hook.

11. Place the hook with DNA in a 5 ml plastic tube containing 3-4 ml of WASH 1. Leave DNA on the hook in the tube for about 20 min.

12. Rinse DNA on the hook briefly in 3-4 ml of WASH 2 and transfer DNA to 2 ml microfuge tube containing 0.3-1.0 ml Tris-HCL (1 mM, pH 8); gently twirl hook until DNA slides off the hook. Cap the tube and rock gently overnight at room temperature to dissolve the DNA. Store samples at 4°C.

Note: Prepare glass hook by first sealing the end of a 23 cm glass transfer pipette by heating in a flame for a few seconds. Then gently heat the tip 1 cm while twirling the pipette. When soft, allow the tip to bend into a hook. Cool before use. Used hooks can be cleaned by washing in dH_2O and EtOH.

Note: DNA that is refrozen after being thawed begins to break after each freezing session, so freeze DNA only for long-term storage and preferably after all testing is finished. If DNA will be used for multiple projects with long periods of time between projects, it can be aliquoted into several tubes and frozen, so that each aliquot is thawed only once at the start of each project.

1.2 DNA Extraction from Small Amounts of Lyophilized Tissue (2 ml tubes)

To extract DNA from small amounts of lyophilized tissue (50 mg), use 2 ml tubes and proceed as follows:

Lyophilization

1. Harvest small dry leaves parts from greenhouse or field grown plants and place into 2ml tubes (3/4 of the tube).

2. Keep tubes of the leaves cool on ice until they can be frozen.

3. Place tubes at -80°C for at least 4 h.

4. Transfer the tubes into a lyophilizer (e.g., Labconco) for 24-72 h. Make sure the lyophilizer is down to the recommended temperature (usually the chamber is ≤ -50°C) and is pulling a good vacuum before loading the samples. Do not overload your lyophilizer: make sure the vacuum is always between 0.0 and 0.120 mBar. Samples should get dried during 72 h. Typically, fresh weight ≈ 10X dry weight. Lids of the tubes must be OPEN!

Note: If samples are not fully dried before grinding, grinding will be inefficient and DNA yield will be poor. Alternatively leaf samples can be tried with silica or at low temperature in an oven.

Grinding

1. Place 1-2 stainless steel balls (4 mm) into each tube and close securely. Place the tubes in a tissue grinder (e.g., GenoGrinder 2010, Zymo Research) and grind 2-3 min until the leaf tissue is ground to fine powder.

2. Leaf powder can be stored in the closed tubes, or DNA extraction can begin immediately in the same tubes.

Genomic DNA Isolation

1. Preheat the CTAB isolation buffer to 65°C. If the samples were in refrigeration or -20°C let them some minutes at RT.

2. Add 1 ml of CTAB isolation buffer at 65°C. Mix by gentle swirling to homogenize the tissue with the buffer.

3. Incubate at 65°C for 60-90 min with continuous gentle rocking.

4. Remove tubes from incubator, let them cool for 15 min.

5. Add 800 µl of chloroform:octanol (24:1). Mix for 15 min at RT with continuous rocking.

6. Centrifuge for 30 min at 3750 rpm at RT.

7. Remove 600 to 700 µl of the top yellow aqueous layer and place in a new tube with RNase with 10 µl of 10 mg/ml RNase A.

8. Mix with gentle inversion and incubate for 30 min at 37°C oven, or 1hour at RT.

9. Add 500 µl of isopropanol (2-propanol 18L Baker 9084-18). Previously cooled at -20°C and mix by inversion.

10. Incubate for 30 min or 1 h at -20°C.

Note: DNA extraction can be paused at this point and continued the following day.

11. Centrifuge tubes for 30 min at 3750 rpm at 4°C to precipitate DNA. Remove the supernatant.

12. Add 1 ml of 70% EtOH.

13. Centrifuge tubes for 20 min at 3750 rpm at RT.

14. Remove the EtOH by decantation.

15. Repeat steps 12 to 14.

16. Dry the DNA at RT overnight inside a fume hood.

17. Re-suspend in 150-200 µl Tris-HCL (1mM, pH 8). Store samples at 4°C.

Note: DNA that is refrozen after being thawed begins to break after each freezing session, so freeze DNA only for long-term storage and preferably after all testing is finished. If DNA will be used for multiple projects with long periods of time between projects, it can be aliquoted into several tubes and frozen, so that each aliquot is thawed only once at the start of each project.

1.3 Small-Scale DNA Extraction in 96-Well Plates

To extract DNA from large amounts of lyophilized tissue (10 mg), use 96-well plates and proceed as follows:

Tissue sampling

1. Small portions of dry leaf tissue are cut from each plant and placed in 8-strip 1.1 ml tubes, up to 12 strips located in 96-well racks. Fill the tubes only half.

2. Keep the tubes with the tissue cool until they can be frozen, but freeze as soon as possible. Store in a -80°C freezer for at least 3 h or use liquid nitrogen. Samples must not thaw before lyophilization.

3. Place trays with the tubes containing the frozen tissue into a lyophilizer (e.g., Labconco). Lids of the tubes must be OPEN! Be sure that the lyophilizer chamber is at -50°C at all times. Verify that it has reached the proper vacuum level after loading the samples, and that it maintains a vacuum level between 0.0 and 0.120 mBars. Fortunately, the small leaf size in each tube makes it hard to overload the machine. Approximately, 1 to 4 plates require 24 h and 5 to 15 plates require 48 h to dry.

4. Dried tissue may be stored in the tubes (with the lids now CLOSED) at room temperature for a few days, or can be stored for longer periods at -20°C. DNA extraction can be started in the same tubes.

Note: If samples are not fully dried before grinding, grinding will be inefficient and DNA yield will be poor. Alternatively leaf samples can be tried with silica or at low temperature in an oven.

Grinding

1. Place 1-2 stainless steel balls (4 mm) into each tube and close securely. Place the entire plate in a tissue grinder (e.g., GenoGrinder 2010, Zymo Research) and grind 2-3 min until the leaf tissue is ground to fine powder.

2. Leaf powder can be stored in the closed tubes, or DNA extraction can begin immediately in the same tubes.

DNA extraction

1. Preheat the CTAB isolation buffer to 65°C.

2. Add 400 µl of CTAB isolation buffer. Mix by gentle swirling to homogenize the tissue with the buffer.

3. Incubate the samples at 65°C for 90 min with continuous gentle rocking.

4. Remove tubes from the oven and allow them to cool for 10 min.

5. Add 300 µl of chloroform:octanol (24:1). Mix gently with continuous rocking for 15 min at RT.

6. Centrifuge at 3750 rpm for 30 min at 4°C to generate a yellow aqueous phase and a green organic phase.

7. Remove approximately 300 µl of the yellow aqueous phase and place in a new set of strips containing 15 µl RNAse A.

8. Mix with gentle inversion and incubate at 37°C for 30 min.

9. Add 280 µl of 100% isopropanol (2-propanol 18L Baker 9084-18). Mix very gently to precipitate the nucleic acid. Incubate samples at -20°C for 30 minutes to 1 h.

Note: DNA extraction can be paused at this point and continued the following day.

10. Centrifuge at 3750 rpm at 4°C for 30 min to form a pellet at the bottom of the tube. Discard the supernatant.

11. Add 400 µl of 70% EtOH. Wash the DNA pellet gently.

12. Centrifuge at 3750 rpm for 20 min at room temperature.

13. Discard EtOH by decantation.

14. Repeat steps 11 to 13.

15. Allow pellet to dry under a fume hood until ethanol evaporates completely.

Note: Any remaining alcohol smell indicates that the pellet is not completely dry.

16. Re-suspend the DNA pellet in 200 µl Tris-HCL (1mM, pH 8).

17.-Store samples at 4°C until use; if DNA will not be used for a long time, store at -20°C.

Note: DNA that is refrozen after being thawed begins to break after each freezing session, so freeze DNA only for long-term storage and preferably after all testing is finished. If DNA will be used for multiple projects with long periods of time between projects, it can be aliquoted into several tubes and frozen, so that each aliquot is thawed only once at the start of each project.

1.4 Automated DNA Extraction Using a BIOMEK FX^P Liquid Handling Station and the Sbeadex® Mini Plant Kit

The screening of large numbers of DNA samples requires a fast and cost effective methods of DNA extraction from plants. In addition, it is preferable for these methods to be automated and reliable for leaf and seed tissue.

We have therefore tested several commercial extraction kits, to test their outcome regards satisfactory DNA quality and quantity for multiple molecular biology techniques using leaves tissue and/or seed. We evaluated eight commercial DNA extraction kits:

NucleoSpin® 96 Plant II (Machery-Nagel)
ZR-96 Plant/Seed DNA kit™ (Zymo Research)
FastDNA™-96 Plant & Seed DNA kit (MP)
Kleargene Plant 96 DNA kit (LGC Genomics)
Wizard Magnetic 96 DNA Plant System (Promega)
NucleoMag 96 Plant (Machery-Nagel)
InviMag® Plant DNA Mini kit (Invitek)
Sbeadex® plant kit (LGC Genomics)

We tested their DNA quality and quantity and their ability to automate DNA extraction on a Biomek FX^P Liquid Handling Station. These included kits in a 96-well binding plate and magnetic bead formats. The performance characteristics of the two methods are summarized in Table 1.

Table 1. Current technologies for extracting DNA and their features.

Features	Magnetic beads	Silica membrane
Mechanism of action	Modified magnetic particles which bind nucleic acids via a classic polarity-based binding mechanism	Nucleic acids bound to silica particles for spin separation. Inserted to microtiter plate in 96 and 384-well formats
Grade of automation	Automated high throughput extraction	Manual extraction or (semi) automated high throughput extraction
Advantage	➢ No organic solvents or chaotropic salts in final wash buffer ➢ No drying of beads for evaporation of alcohols ➢ High flexibility of all extraction parameter	➢ Suitable for manual or (semi) automated extraction solutions 96 or 384 format. ➢ Flexible lysis solutions available

DNA yield and purity (A_{260}/A_{280} & A_{260}/A_{230}) of each kit were assessed by spectrophotometry (Nanodrop 8000), and 1% quality gel eletctrophoresis. The results are shown in Table 2.

Table 2. Comparison of results between the eight commercial kits.

Kit	Average concentration [ng/uL]*	260/280 ratio*	260/230 ratio*	Quality**
NucleoSpin® 96 Plant II	30	1.87	2.07	Good
ZR Plant/Seed DNA kit™	80	1.90	1.34	Good
FastDNA™-96 Plant & Seed DNA kit	52	1.85	0.25	Bad
Kleargene Plant 96 DNA kit	107	1.88	1.51	Good
Wizard Magnetic 96 DNA Plant System	49	1.69	0.10	Bad
NucleoMag 96 Plant	63	1.49	0.90	Good
InviMag® Plant DNA Mini kit	90	1.93	0.88	Good
Sbeadex® plant kit	141	1.79	1.43	Good

* Using Nanodrop 8000. **Agarose 1%.

The comparison of the eight commercial kits showed that the ZR-96 Plant/Seed DNA kit™, InviMag® Plant DNA Mini kit and Sbeadex® plant kit in our laboratory revealed similar DNA quality (Table 1, Figure 1); furthermore, the Sbeadex® plant kit revealed a higher DNA concentration and good purity, making it appropriate for high throughput extraction. Our results and additional reports from different sources showed that the magnetic bead-based method was preferable for avoiding organic extractions, centrifugation, or filtration steps that can result in clogging or cross contamination. Magnetic beads are also advantageous because they allow aggressive washes that can aid in the removal of particulate matter and potential PCR inhibitors as part of a fully automated system. Magnetic bead-based DNA extraction has been used to successfully isolate nucleic acids from a variety of sample types that contain PCR inhibitors.

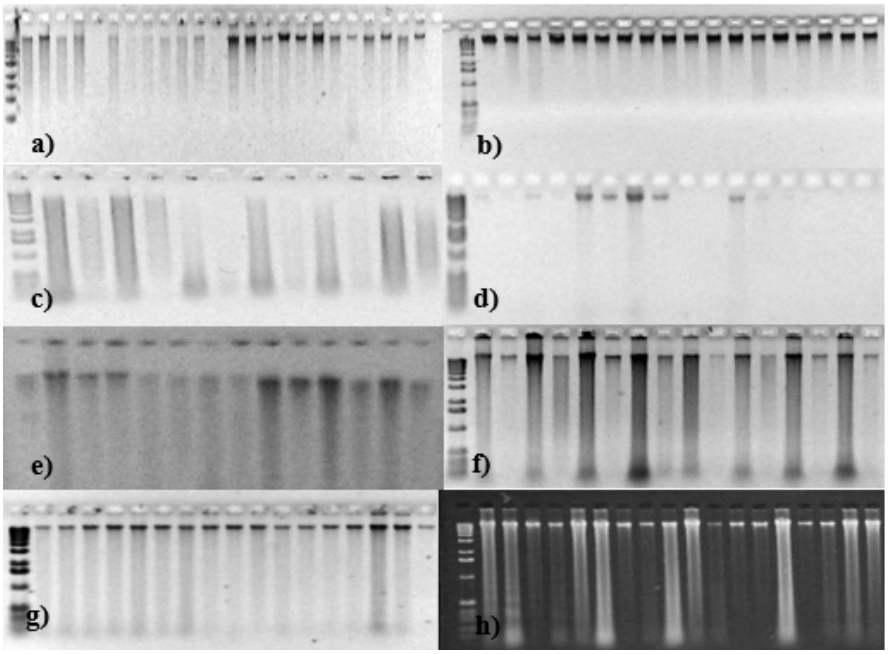


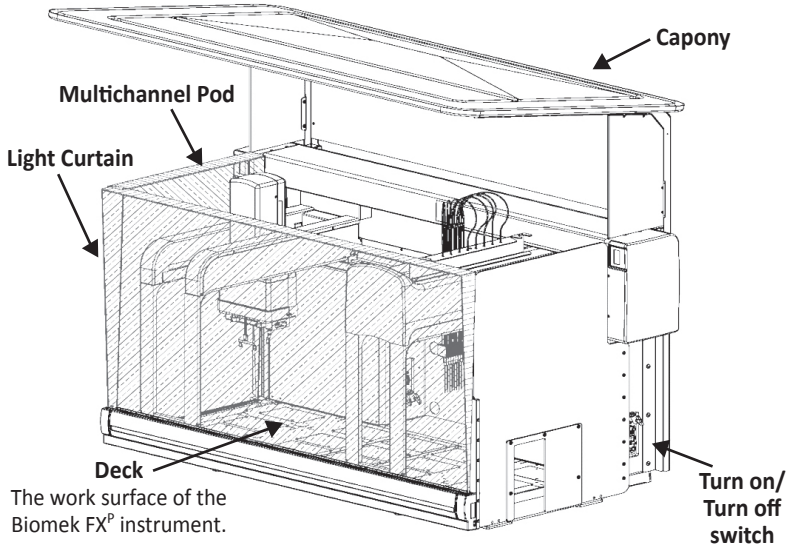
Figure 1. Agarose quality gel (1%) for the eight commercial DNA extraction kits. The first well belongs to a 1 kb marker a) NucleoSpin® 96 Plant II, b) ZR-96 Plant/Seed DNA kit™, c) FastDNA™-96 Plant & Seed DNA kit, d) Kleargene Plant 96 DNA kit, e) Wizard Magnetic 96 DNA Plant System, f) NucleoMag 96 Plant, g) InviMag® Plant DNA Mini kit, and h) Sbeadex® plant kit.

Based on the DNA yield and purity, ease of implementation and atomization and cost we have decided to further validate the Sbeadex kit in our laboratory. The Sbeadex magnetic particle protocol uses the cationic detergent CTAB, which builds complexes with proteins and carbohydrates. By centrifugation, cell residues like insoluble cell walls or and CTAB/protein/sugar complexes are spun down and separated from other cell components like DNA, which stays in solution. A novel two step binding mechanism via a patented adapter enables binding of nucleic acids more specifically and final washes with pure water to deliver nucleic acids in high yield, purity and quality.

Schematic process for DNA extraction on the Sbeadex LGC kits.

BIOMEK FX^P Liquid Handling Station

The CIMMYT laboratory accounts with a Beckman Biomek FX^P Liquid Handling Station for DNA extraction. The Biomek FX^P is a multiaxis liquid-handling instrument used in the laboratory, with a variety of operating components (Figure 2).



The work surface of the Biomek FX^P instrument.

Figure2. Biomek FX^P Main components.

Automated DNA extraction protocol for wheat leaf or seed tissue

The linkage of the liquid handling station and the commercial DNA extraction kits allows automating operations and provides cost and labor savings. During testing we also observed more constant and stable DNA quantities across 96-well plates, which is of advantage for MAS and reduce the need to normalize the DNA.

Using the liquid handling station combined with the Sbeadex® plant kit protocol, we extract two 96-well plates in 1.5 h, and the DNA is ready to use for any downstream application.

For the automated protocol we use the following materials:

a) Sbeadex® mini plant kit (960 samples), Cat. # 41610, LGC Genomics
b) DNA plates: Plates with ground plant materials (fine powder of dried leaf tissue or seed)
c) Sigma cell culture ddH_2O
d) Two Qiagen rack plates for holding the final DNA
e) AP96 Tips, P250 Nonsterile, Cat.# 727251
f) Reservoirs for all the components: lysis buffer, binding buffer, sbeadex, wash buffers, ddH_2O, elution buffer; all components are on the kit
g) Two deep-well plates 96/1000µl, Cat.# 30504208
h) Two magnetic plates
i) Two reservoirs for junk material

Instructions

1. Locate the power switch on the right side of the instrument, and flip it to the on position.

2. Launch the Biomek Software on the computer (Figure 3). Home All Axes. This automatic process initiates the pod and defines the home position from which all subsequent moves are determined. Go to Instrument menu→Home All Axes→OK.

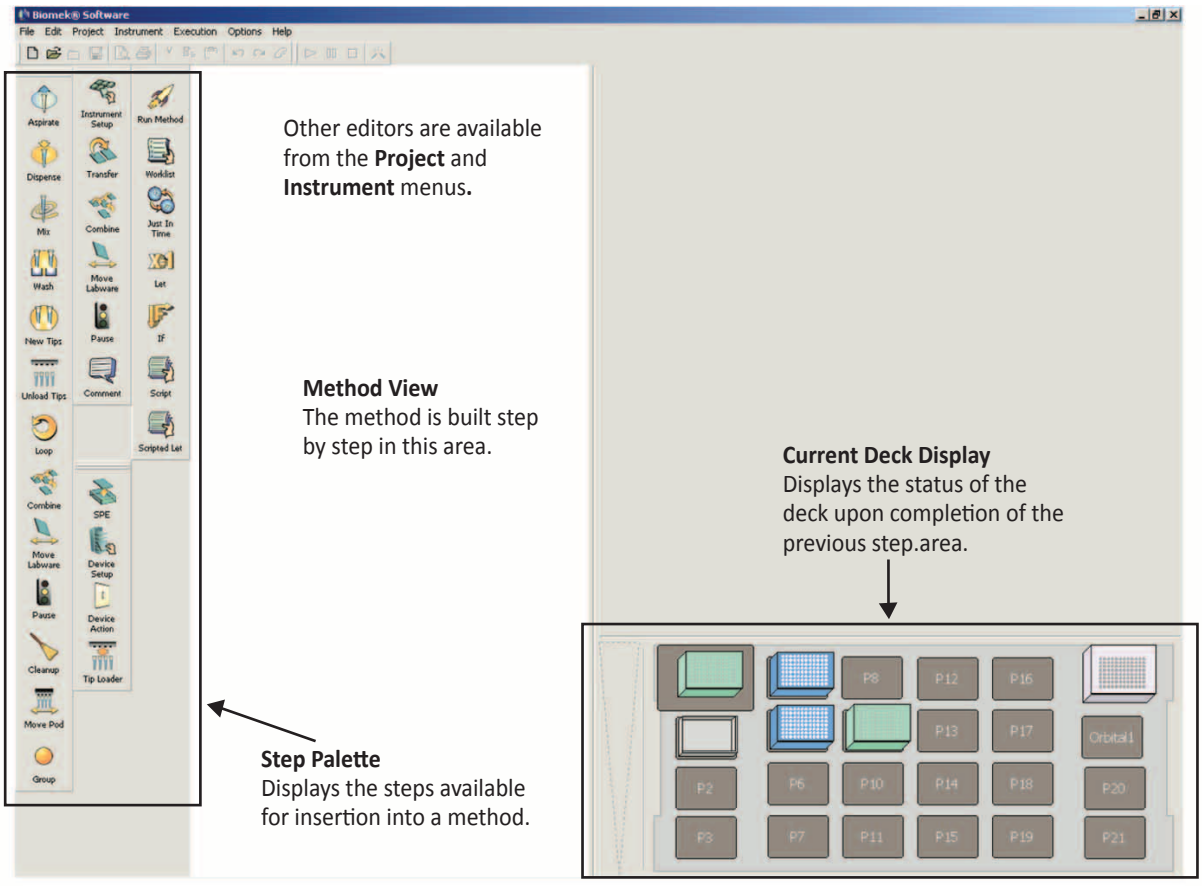


Figure 3. Biomek Software main editor.

3. Open the DNA extraction program 'LGC 2 extraction plates' in the Biomek Software and set up all the components according to the Deck display (Figure 4).

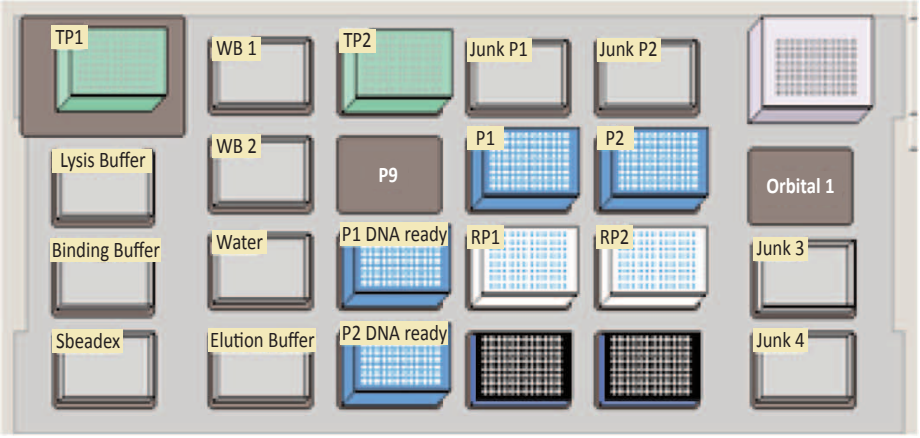


Figure 4. Diagram of the Biomek deck layout for DNA extraction.

4. Lysis: Press play to start the first step, the addition of 300 uL of lysis buffer. After the buffer is added into the DNA plates, a stop message will appear on the screen: remove the DNA plates from the deck.

5. Incubate the plates 30 min at 65°C on the orbital shaker.

6. While the plates are incubating, press → play to prepare the binding plate. During this step the robot will mix the binding buffer and the sbeadex magnetic particles to make the binding solution. A stop window will appear at the end of this step. Return the DNA plates to their original position and continue with the next step.

7. Binding DNA: During this step the lysate and the binding solution will be mixed; the DNA will be attached to the magnetic particles by electromagnetic forces.

8. Remove supernatant: During this step, the DNA plate is going to the deck place of the magnetic plate; all the magnetic particles go to the surface of the DNA plate, subsequently the supernatant is discarded.

9. Washing step 1: This step moves the DNA plate to another position, adds 400 µL of washing buffer 1; subsequently the plate is mixed with a shaker on the deck. Step 6 will be repeated.

10. Washing step 2: This step will repeat the previous step, but will add 400 µL of washing buffer 2 and 400 µL of ddH_2O.

11. Elution: After the washing steps, 100 µL of elution buffer or alternative 100 µL of Tris (pH 8.0) is added to the DNA plate. This process requires an incubation step outside the robot.

12. After incubation, 90 µl of eluted DNA is added into a new plate. This DNA is ready to use for downstream applications.

Plant DNA extraction: Solutions

CTAB extraction buffer[1]

STOCK	[FINAL]	1 R×N 10 ml	5 R×N 50 ml	10 R×N 100 ml	20 R×N 200 ml	50 R×N 500 ml	60 R×N 600 ml
dH_2O		6.5 ml	32.5 ml	65.0 ml	130 ml	325.0 ml	390.0 ml
1 M Tris-7.5	100 mM	1.0 ml	5.0 ml	10.0 ml	20.0 ml	50.0 ml	60.0 ml
5 M NaCl	700 mM	1.4 ml	7.0 ml	14.0 ml	28.0 ml	70.0 ml	84.0 ml
0.5 M EDTA-8.0	50 mM	1.0 ml	5.0 ml	10.0 ml	20.0 ml	50.0 ml	60.0 ml
CTAB[2]	1 %	0.1 g	0.5 g	1.0 g	2.0 g	5.0 g	6.0 g
14 M BME[3]	140 mM	0.1 ml	0.5 ml	1.0 ml	2.0 ml	5.0 ml	6.0 ml

[1] Use freshly made; warm buffer to 60-65°C before adding the CTAB and BME.
[2] CTAB (Sigma M7635).
[3] BME (Sigma M6250). Add BME just prior to use, under a fume hood.

CHLOROFORM: OCTANOL (24:1)

STOCK	100 ml	200 ml	300 ml	400 ml	500 ml
Chloroform[1]	96 ml	192 ml	288 ml	384 ml	480 ml
Octanol[2]	4 ml	8 ml	12 ml	16 ml	20 ml

[1] Cloroform 4 L (Baker 9180-03).
[2] 1-Octanol Octyl alcohol 500 ml (Baker 9085-01).

10 mg/ml RNAse A

Dissolve 100 mg of RNAse[1] in 10 ml of 10 mM Tris - 7.5, 15 mM NaCl. Heat in boiling water for 15 min and allow to cool slowly to room temperature (RT).
Dispense into 1 ml aliquots and store at -20°C.

[1] Ribonuclease A From Bovi-Ne Pancreas 500mg (Sigma R4875).

WASH 1: 76% EtOH, 0.2 M NaOAc

STOCK	100 ml	200 ml	300 ml	400 ml	500 ml
Absolute EtOH[1]	76 ml	152 ml	228 ml	304 ml	380 ml
2.5 M NaOAc	8 ml	16 ml	24 ml	32 ml	40 ml
dH_2O	16 ml	32 ml	48 ml	64 ml	80 ml

[1] Absolute Ethanol (Merck 100983).

WASH 2: 76% EtOH, 10 mM NH_4OAc

STOCK	100 ml	200 ml	300 ml	400 ml	500 ml
Absolute EtOH[1]	76 ml	152 ml	228 ml	304 ml	380 ml
1 M NH_4OAc[2]	1 ml	2 ml	3 ml	4 ml	5 ml
dH_2O	23 ml	46 ml	69 ml	92 ml	115 ml

[1] Absolute Ethanol (Merck 100983).
[2] NH_4OAc (Sigma A1542-500G).

1 M Tris - pH 7.5/ pH8.0

Dissolve 121 g Trizma-Base[1] in approx. 750 ml dH_2O. Add conc. HCl until desired pH is reached (75 ml HCl = pH 7.5, 49 ml HCl = pH 8.0). Bring solution to 1000 ml with dH_2O Autoclave.

[1] Trizma Base minimum 99.9% tritation 10k (Sigma T1503).

Tris- HCl 1 mM, pH 8.0

STOCK	1 L
Tris 1 M-pH 8.0	10 ml
dH_2O	990 ml

5 M NaCl

Dissolve 292.2 g NaCl[1] (MW=58.44) in dH_2O to a final volume of 1000 ml. Autoclave.

[1] NaCl (J.T. Baker, 3624-1OKG).

0.5 M EDTA- pH 8.0

Dissolve 186.12 g $Na_2EDTA \bullet 2H_2O$[1] (MW=372.24) in approx. 750 ml of dH_2O. Add NaOH pellets to bring pH to 8.0. After EDTA is in solution, bring to 1000 ml with dH_2O Autoclave.

[1] Ethylenediaminetetraacetic Acid Disodium Salt 1kg (Sigma E5134).

2.5 M NaOAc

Dissolve 20.5 g sodium acetate[1] (anhydrous, MW=82.03) in dH_2O to a final volume of 100 ml. Autoclave.

[1] Sodium Acetate Trihydrate 1kg (Sigma S-9513).

1 M NH_4OAc

Dissolve 7.71 g ammonium acetate[1] (MW=77.08) in dH_2O to a final volume of 100 ml. Filter sterilize.

[1] NH_4OAc (Sigma A1542-500G).

2. Quantification, Quality Control and purification of DNA

2.1 UV Quantification of DNA with a NanoDrop 8000 Spectrometer (Thermo Scientific, USA)

1. After re-suspension of your DNA in Tris (pH 8.0) read your samples at OD260 and OD280 to determine the quantity and the purity of your DNA.

2. For reading load first a reference sample of 2 µl Tris (pH 8.0) as blank, then load 2 µl of your sample.

Note: We use wipers (Kimwipes®EX-L, Cat. # 34155, Kimberly-Clark®) to clean every sample.

Note: (DNA concentration (µg/µl)= OD260 x 50 (dilution factor)x 50 µg/ml)/1000.

The ratio OD260/OD280 is determined to assess the purity of the sample. If this ratio is 1.8 - 2.0, the absorption is probably due to nucleic acids. A ratio of less than 1.8 indicates there may be proteins and/or other UV absorbers in the sample, in which case it is advisable to re-precipitate the DNA. A ratio higher than 2.0 indicates the samples may be contaminated with chloroform or phenol and should be re-precipitated with ethanol.

3. After UV quantification, adjust the concentration of each DNA sample to 0.3-0.5 µg/µl or a concentration of your choice with Tris (pH 8.0), and store at 4°C. Sample should be usable for up to six months. For long term storage, freezing temperature is more desirable.

Note: When analyzing many samples we usually use a dilution factor of 10 and proceed.

2.2 Molecular Weight Markers

Two types of molecular weight (MW) standards are routinely used. The Lambda/HindIII and PhiX174/ *Hae*III MW standards provide a useful reference for calculating molecular weights of large and small DNA fragments, respectively, after electrophoretic separation; the "internal MW standards" provide a means for normalizing fragment migration distances within each lane to facilitate comparisons between lanes on the same or different luminographs in fingerprinting studies.

PhiX174 DNA as a Molecular Weight Standard for Luminographs

Digestion of ØX174 DNA with *Hae*III:

STOCK	[Final] or amount	500 µl RXN
ddH_2O	----	326.5 µl
10X CutSmart Buffer	1X	50.0 µl
Spermidine 0.1 M	2.5 mM	12.5 µl
ØX174 RF1 DNA (0.5 µg/ µl)[1]	15 µg	100.0 µl
*Hae*III (10 000U/ml)[2]	2 U/ µg ADN	11.0 µl
		500.0 µl

[1] Verify the concentration of the commercial λ and make the necessary adjustments. FX 174 RF1. DNA 50mg. (Fermentas SD0031).

[2] *Hae*III 10,000 U/ml (BioLabs R0108S).

1. Calculate the number of tubes to prepare. One ØX174 tube yields 2000 µl of ladder at 25 ng/µl.

2. Allow to digest over 2 or 3 h in a thermal bath at 37 °C.

3. Check that digestion is complete by running about 50 ng on a 0.7% agarose gel. When it is complete, move to step 4 or 5 (Figure 1).

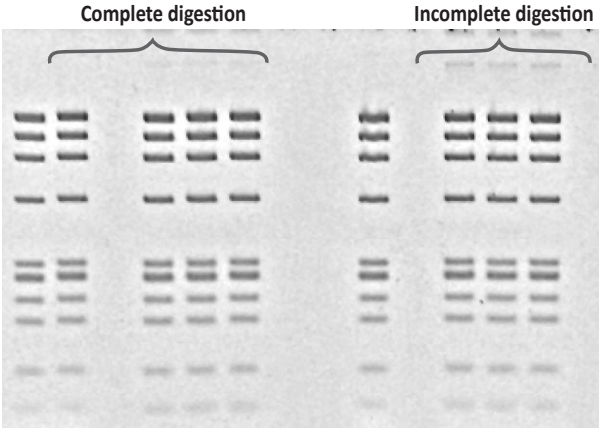


Figure 1. Fragment migration after digestion with *Hae*III.

4. If the digestion was not completed, add 2 µl of *Hae*III and digest again in a thermal bath at 37 °C for 30 min.

5. If you are going to use the digested DNA as a MW marker without end-labeling it, inactivate the enzyme by incubating at 80°C for 20 min.

6. For each ØX174 tube, add 1000 µl of TE and 500 µl of 5X loading buffer BPB solution to bring to a concentration of 25 ng/µl. Aliquot (200 µl per 0.5 ml tubes) and keep at 4°C or in the freezer.

Lambda (λ) DNA as a Molecular Weight Standard for Luminographs

Digestion of λ DNA with *Hind*III:

STOCK	[FINAL] or amount	50 µl RXN
ddH_2O		31.8 µl
10X Buffer	1X	5.0 µl
0.1 M Spermidine	2.5 mM	1.2 µl
λ DNA (0.45 µg/µl)[1]	5 µg	11.0 µl
*Hind*III (10 U/µl)[2]	2 U/µg DNA	1.0 µl

[1] Check the concentration of commercial λ and adjust quantities accordingly.
[2] Lambda DNA/*Hind*lll marker (Thermoscientific SM0101).

1. Allow to digest at 37°C for 2-3 h.

2. Check that digestion is complete by running about 50 ng on a 0.7% agarose gel. When it is complete, move to step 3 or 4.

3. Inactivate the enzyme by incubating at 65°C for 10 min. Then add 110 µl TE and 40 µl 5X SGB to bring to a concentration of 25 ng/µl. Aliquot and keep at 4°C or in the freezer.

Molecular Weight Markers: Solutions

0.1 M Spermidine

Dissolve 1 g spermidine (MW= 145.2, Sigma # S2626) in ddH_2O to a final volume of 69 ml. Filter sterilize and aliquot into 5 ml tubes. Store at -20°C.

TE-8: 10 mM Tris - 8.0, 1 mM EDTA - pH 8.0

	50 ml	100 ml	500 ml	1000 ml
1 M Tris- 8.0	0.5 ml	1.0 ml	5.0 ml	10.0 ml
0.5 M EDTA – 8.0	0.1 ml	0.2 ml	1.0 ml	2.0 ml
ddH_2O	to volume	to volume	to volume	to volume

2.3 DNA Quality Control of DNA via gel electrophoresis

This step is essential for checking that the isolated DNA is of high molecular weight. Native DNA should migrate as a tight band of molecular weight ≥ 40 Kb. However, degradation of part of the isolated DNA is inevitable, and the protocol below is designed to run the DNA under optimal conditions for ascertaining the relative amounts of degraded and high molecular weight DNA. The procedure also allows verifying the UV quantification performed above.

Note: If you have few DNA samples (say, less than 25), check all of them. Otherwise, check only10-20% of the samples, making sure that the selection is totally random.

1. Prepare a 50 ng/µl dilution of the selected samples

Note: See the section on gel electrophoresis for details about gel preparation, running conditions, and DNA visualization.

2. Load 100 ng of each diluted sample (2 µl DNA + 5 µl 1X sample gel buffer) in a 0.7% agarose gel. Include at least one lane per comb of uncut Lambda DNA (λ) as a molecular weight marker. Load 100 ng of this marker to check both quality and quantity of the sample DNA.

3. Run the gel at 70V for about 90 min and visualize the DNA.

2.4 DNA Purification for Sequencing

DNA purification is required for genomics applications ranging from cloning, sequencing and microarray analysis. A number of purification kits are offered that very in DNA treatment and the purification methods. The most popular methods are based on salt-precipitation and silica binding.

Qiagen's nucleic acid-purification kit

We have been using Qiagen's nucleic acid-purification kits that rely on the binding of DNA to a silica matrix, which is attached to a column. Inhibitors are washed away, and pure, intact DNA subsequently is eluted from the column at high yields.

E.g., we convert SSR/STS markers to SNP markers (KASP assays) for high-throughput genotyping by simple re-sequencing the marker bands from the genotypes having contrasting phenotypes.

Before purification: The DNA should be free of contaminants prior to sequencing to avoid mixed and unspecific sequence data. It is highly recommended to observe any amplification product you would like to sequence on an agarose gel to confirm the expected amplification product with the correct size. If you get a single clean product on an agarose gel e.g., among parental genotypes with contrasting phenotypes (Figure 1a) the use of a PCR purification kit is recommended. If the polymorphic product between the genotypes is accompanied with one or two more unspecific amplification products (Figure 1b), then a purification kit based on gel extraction is required, which allows to cut only the polymorphic band and avoids sequencing of the other irrelevant bands.

Purification: The Qiagen Gel extraction kit (cat. nos. 28704 and 28706) or PCR Purification kit (cat. nos. 28104 and 28106) can be used to remove all of the unwanted elements like primers, nucleotides, enzymes, mineral oil, salts, and other impurities from DNA samples. The kit instructions are easy to follow.

After purification: Purified samples can then be sequenced. We send our samples to Macrogen (http://dna.macrogen.com/eng/). Samples can be submitted online (choose the option "custom DNA sequencing/Standard-Seq" and read more information on how to send the samples). In addition to the purified DNA samples, supply primers at a concentration of 10 pmole/µl = 60 ng/µl in deionized water in volumes greater than 20 µl should be submitted and sent. The sequencing results are provided by Macrogen in a compressed zipped folder for sequence alignment in any program.

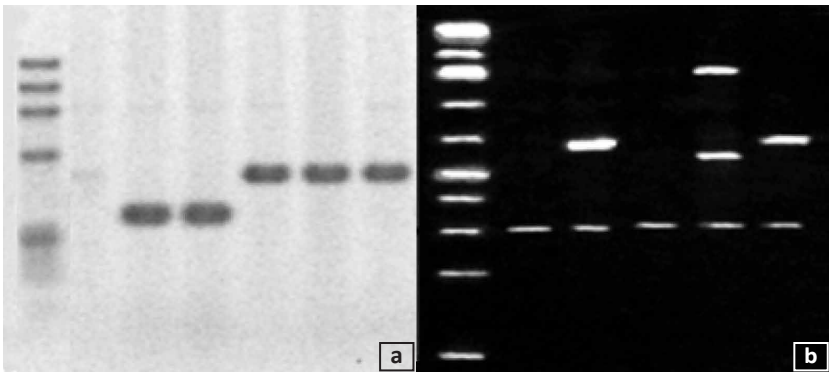


Figure 1. Amplification results (a) two single clear polymorphic bands between target genotypes and (b) polymorphic bands accompanied with unspecific bands.

3. PCR Protocol for STS/SSR Markers

Polymerase Chain Reaction is widely held as one of the most important inventions of the 20th century in molecular biology. Small amounts of the genetic material can be amplified to be able to identify polymorphisms in the DNA. PCR involves the following three steps: Denaturation, Annealing and Extension. First, the genetic material is denatured, converting the double stranded DNA molecules to single strands. The primers are then annealed to the complementary regions of the single stranded molecules. In the third step, they are extended by the action of the DNA polymerase. All these steps are temperature sensitive and the common choice of temperatures is 94°C, 60°C and 70°C respectively. Good primer design is essential for successful reactions.

Sequence-Tagged Sites (STS) are a relatively short, easily PCR-amplified sequence (200 to 500 bp) which can be specifically amplified by PCR and detected in the presence of all other genomic sequences and whose location in the genome is mapped. STS-based PCR produces a simple and reproducible pattern on agarose or polyacrylamide gel. In most cases STS markers are co-dominant, i.e., allow heterorozygotes to be distinguished from the two homozygotes.

Polymorphic loci present in nuclear DNA and organellar DNA that consist of repeating units of 1-10 base pairs, most typically, 2-3 bp in length, also called **Simple Sequence Repeats (SSR)**. SSRs are highly variable and evenly distributed throughout the genome. This type of repeated DNA is common in eukaryotes. These polymorphisms are identified by constructing PCR primers for the DNA flanking the microsatellite region. The flanking regions tend to be conserved within the species, although sometimes they may also be conserved in higher taxonomic levels.

Good sources of sequence information for both marker systems can be accessed via the Internet. For wheat, consult GrainGenes at http://wheat.pw.usda.gov.

The quality of the template DNA is less critical for STSs or SSRs. We get good results using DNA from large amounts of lyophilized, ground tissue, as well as DNA extracted from a small frozen leaf portion or fast DNA extraction kits.

Amplification

1. Prepare a bulk reaction mix containing all the reaction components listed below except DNA. We recommend assembling all reaction components on ice.

Using GoTaq®Flexi, Promega, USA, Cat. # M8295

STOCK or Amount	[FINAL]	10 µl RXN
ddH_2O[1]	-------	0.05 µl
5X Green or Colorless GoTaq®Flexi	1X	2.0 µl
25 mM $MgCl_2$[2]	1.5 mM	0.6 µl
dNTP Mix (2.5 mM each)	200 µM	0.8 µl
Primers F + R (1.0 µM each)[3]	0.25 µM each	2.5 µl
Go taq(® DNA Polymerase (5U/µl)	0.25 U	0.05 µl
DNA (10-50 ng/µl)	50-100 ng	4.0 µl

[1] *Sigma* Cell Culture Water, Cat. # W-3500.
[2] It is essential to determine optimal concentrations of $MgCl_2$ and Taq with each new lot of enzyme and DNA from species to be analyzed.
[3] Both forward and reverse primers are present in the same tube.

Using REDTaq, SIGMA-ALDRICH®, USA, Cat. # R2523

STOCK or Amount	[FINAL]	10 µl RXN
REDTaq®ReadyMix™ PCR Reaction Mix with $MgCl_2$	1X	4.0 µl
Primers F + R (1.0 µM each)	0.25 µM each	2.0 µl
DNA (10-50 ng/µl)	50-100 ng	4.0 µl

Note: Usually DNA primers are obtained from companies in lyophilized form. Primer can be ordered with a different purity grade (desalted or HPLC purified) and concentration. For AGE or PAGE desalted purity is sufficient. Primers have to be dissolved with ddH_2O or TRIS (pH 8.0) according to the concentration of each of the two primers.

Note: The optimum concentrations of various components are slightly different for each marker. If you need to prepare the bulk mix in advance, we suggest you include all components except the Taq polymerase and keep the mixture at either 4°C or -20°C until needed. The Taq enzyme would be added just before aliquoting the bulk mix.

2. Gently mix the bulk reaction and keep it on ice.

3. Add the DNA samples to each labeled PCR microtiter plate well.

4. Aliquot bulk mix into each labeled PCR microtiter plate well.

5. Gently mix the reaction. Collect all liquid to the bottom of the tube by a quick spin if necessary.

Note: Use separate tips for each DNA sample. Using multi-channel pipettes reduce pipetting errors. When using the same tips to aliquot the bulk of all samples take care not to touch the base of the plate wells containing the DNA.

6. Transfer the microtiter plate from ice to a PCR machine and begin thermocycling. Amplify using either of the following programs:

Standard PCR program

94°C for 2 min	
94°C for 1 min 50-68°C for 2 min[1] 72°C for 2 min	30 cycles
72°C for 5 min	1 cycle
Final temperature 15°C	

[1] Based on annealing temperature.

Touchdown (TD) PCR program

Example TD57 (annealing temperature can be varied)

94°C for 2 min	
94°C for 1 min 64°C for 1 min[1] 72°C for 1 min	7 cycles
94°C for 1 min 57°C for 1 min 72°C for 1 min	35 cycles
72°C for 5 min	1 cycle

[1] Annealing temperature decreases by 1°C each cycle to 57°C.

Quick PCR program

94°C for 30 sec	
94°C for 15-30 sec 50-68°C for 15-60 sec[1] 72°C for 1 min	30 cycles
72°C for 5 min	1 cycle
Final temperature 15°C	

[1] Based on annealing temperature.

Note: Each pair of primers has an optimal annealing temperature that should be determined from their sequences. Start testing new primers at 55°C annealing temperature. If satisfactory amplification does not occur, either reduce or increase the temperature by 4-5°C. The touchdown program may eliminate some unspecific bands compared to the standard program.

Note: For primers pairs resulting in amplification products of distinct sizes, a procedure called multiplexing allows the simultaneous amplification of two or more PCR based markers, provided they have similar annealing temperatures. We have used the procedure in duplexing (two primer pairs at a time). In some cases, combining two sets of primer pairs result in the preferential amplification of one of the two products. To improve the amplification of the other product, suggestions are to increase the amount of primers of the poorly amplified PCR based marker and/or decrease the amount of primers of the other PCR based marker, decrease the annealing temperature, and/or use a higher quality Taq polymerase. A second option is to amplify each PCR based marker separately and load the markers in the same gel system.

7. When using colorless buffer for PCR reactions, add 3-4 µl of loading dye “1X bromophenol blue (BPB)” to each tube and load on the desired gel system.

PCR protocol for STS/SSR markers: Solutions

dNTP mix (2.5 mM each of dCTP, dGTP, dATP, and dTTP)

We recommend using a deoxynucleoside triphosphate set, PCR grade (e.g.,Roche, Cat #. 1969064). Each set comes with four individual tubes containing dCTP, dGTP, dATP, and dTTP at 100 mM concentration. To mix, place 250 µl of each nucleotide in a 10 ml tube and add 9000 µl of sterile ddH_2O (Sigma, cat. W3500) to obtain a 2.5 mM concentration of each nucleotide.

Make 1 ml aliquots and label each tube with different color dots (red for dTTP, blue for dCTP, black for dATP, and green for dGTP) to indicate contents. Store at -20°C.

Note: For individual nucleotide solutions, mix 250 µl of each nucleotide separately with 2,250 µl sterile ddH_20. Make 200 µl aliquots and label. Store at -20°C.

4. Gel Electrophoresis

The choice of the gel electrophoresis system to be used, and of its various components, depends on the expected size of the amplification product(s), on the resolution required to clearly see the difference in size among the amplified products and, to a lesser extent, on the intensity of the amplified products. In our laboratory, we have tried horizontal agarose gels of different concentrations and various ratios of higher quality : normal quality agarose; small polyacrylamide vertical gels with different concentrations and ratios of acrylamide : bisacrylamide, stained with ethidium bromide, silver nitrate or GelRed. Below are the conditions we have been using for both agarose (AGE) and small nondenaturing polyacrylamide gel electrophoresis (PAGE).

Some general rules we follow:

- Use AGE if you expect clear difference in amplified products (> 20 bp).
- For genetic diversity/fingerprinting purposes, always use PAGE due to the required higher resolution.
- For mapping studies, start by screening parental lines for polymorphisms on agarose gels and re-run on polyacrylamide only those markers with small differences or low intensity that could not be clearly seen on agarose gels.

4.1 Agarose Gel Electrophoresis (AGE)

Factors you should consider when deciding on the type and size of agarose gels to be used:

- Agarose concentration, depending on the size of the amplified products and on the type of inheritance of the marker (dominant or co-dominant); typically we use 1.5% for larger fragments (200-3500 bp) and dominant inherited markers; 4% for smaller fragments or co-dominant markers with smaller differences in product size.
- Migration distance is an additional factor involved in the resolution of the differences in amplification product sizes. The larger the distance, the better the resolution (see point below on choice of electrophoresis tanks). We are using SFR agarose AMRESCO® (Code: J234) for all agarose concentrations.

We use 1X TBE buffer (both to prepare the gel and run it) rather than 1X TAE for better resolution. This buffer can be re-used once or twice since the running time is usually short. An alternative to re-using the buffer is to try using 0.5X TBE.

Electrophoresis Systems

We currently use smaller and bigger sizes of ExpressCast™ Horizontal Electrophoresis Systems (Gemini Scientific Inc. DBA Galileo Bioscience).

The small system with the dimensions 32 × 37.5 cm (Model 81-2325) hold four 50-tooth combs, which allow to electrophorese samples from up to two 96-well microtiter plates and to load samples using a multichannel pipette.

The large system with the dimensions 32 × 53 cm (Model 81-2340) hold 12 50-tooth combs, which allow to electrophorese samples from up to six 96-well microtiter plates and to load samples using a multichannel pipette.

Preparation of the agarose gel

1. Add agarose to proper amount of 1X TBE gel buffer (For small gels we use 200ml, for large gels 350 ml of 1X TBE).

2. Melt agarose in microwave oven, mixing carefully several times during heating. Make sure all the agarose is dissolved. To adjust evaporation during melting, add extra ddH_2O to maintain the desired concentration.

3. To eliminate very small bubbles created by much mixing, apply some vacuum to the flask (can be done by placing in a dessicator connected to the vacuum).

4. Level your gel tray and electrophoresis system.

5. Pour agarose right away into the gel tray and then insert combs. Allow to solidify for 20-30 min. You may want to cool it at 4°C for 15 min before loading your samples. We also often prepare such gels one day before and keep them covered with Saran Wrap in the cold.

6. Either load the samples in the "dry" gel using a multichannel pipette (dry loading) or place tray in rig with 1X TBE gel buffer (wet loading). For wet loading remove combs only when ready to load samples. Pour enough 1X TBE buffer into the gel rig to cover the gel by at least 0.5 cm.

7. Run samples into gel at 150/200 Volts at constant voltage. The running time depends on the agarose concentration and size of the amplified product. Approximate running time is between 1-2 h. A good indicator is to run the gel until the bromophenol blue (BPB) dye has migrated to just above the next set of wells.

8. Remove tray from rig and stain the gel with ethidium bromide solution for 10-30 min with gentle shaking.

CAUTION: Ethidium bromide is extremely mutagenic–wear a lab coat, protecting eye glases and double gloves when handling and use extra precaution.

9. Rinse gel in dH_2O for 5-10 min, slide gel onto a UV transilluminator, and photograph.

Non-toxic option: GelRed® STAINING PROTOCOL

GelRed™ Nucleic Acid Gel Stain, 10,000X in water. Catalog # 41003 Biotium

1. Dilute 50 ul of GelRed 10,000x in 500 ml of distilled water. Mix.
2. Put the gel into the dilution.
3. Mix gently for 30 min.
4. Take picture on UV transilluminator.

Note: Alternatively GelRed can be added to the agarose gel.

Note: Both ethidum bromide and GedRed are photosensitive and should be stored in dark.

Load 10 µl of each sample in an agarose gel prepared with 1X TBE gel buffer. Electrophorese in 1X TBE at 150 V (small tanks) and 250 V (large tanks), constant voltage, until the dye has migrated as required.

Note: If no dye has been included during PCR reaction add 2-3 µl blue dye (1x GB) before loading.

AGE: Solutions

10X TBE gel buffer: 0.9 M Tris-borate, 20 mM EDTA

STOCK	1 liter	2 liters	3 liters	4 liters	5 liters
Trizma Base[1](MW=121.10)	108.0 g	216.0 g	324.0 g	432.0 g	540.0 g
Boric acid[2](MW=61.83)	55.0 g	110.0 g	165.0 g	220.0 g	275.0 g
0.5 EDTA pH 8.0	40.0 ml	80.0 ml	120.0 ml	160.0 ml	200.0 ml

Ajust pH to 8.0 with glacial acetic acid or HCl (acetic acid for PAGE). A precipitate may form when stored for long periods of time.

[1] Trizma Base minimum 99.9% tritation 10 k (Sigma T1503).

[2] Boric Acid for Electrophoresis 1 K (Sigma B7901).

5X loading buffer BPB

STOCK	[FINAL]	50 ML	100 ML
1 M Tris-pH 8.0	50 mM	2.5 ml	5.0 ml
0.5 M EDTA-8.0	5 mM	0.5 ml	1.0 ml
Sucrose[1]	25%	12.5 g	25 g
Bromophenol blue[2]	2mg/ml	100 mg	200 mg
ddH_2O		up to 50 ml	up to 100 ml

Dilute to prepare a 1X dilution.

[1] Bromphenol Blue Sodium Salt 25g (Sigma B7021).

[2] Sucrose 10 kg (Sigma S-5390).

0.5 M EDTA-8.0

Dissolve 186.12 g Na2EDTA•2H20[1] (MW = 372.24) in approx. 750 ml of dH_2O. Add NaOH pellets to bring pH to 8.0. After EDTA is in solution, bring to 1000 ml with dH_2O. Autoclave.

[1] EDTA Ethylenediaminetetraacetic Acid Disodium Salt 1kg (Sigma E5134).

1 M Tris - pH 7.5, 8.0

Dissolve 121 g Trizma-Base[1] in approx. 750 ml dH_2O. Add concentrated HCl until desired pH is reached (75 ml HCl = pH 7.5, 49 ml HCl = pH 8.0). Bring solution to 1000 ml with dH_2O. Autoclave.

[1] Trizma Base minimum 99.9% tritation 10k (Sigma T1503).

Ethidium Bromide Solution[1]

Disolve 100 µl of 10 mg/ml ethidium bromide in 1000 ml dH_2O.

[1] Ethidium Bromide Solution (Sigma SLBB2512V).

GelRed® STAINING

Dilute 50 µl of GelRed[1] 10,000x in 500 ml of distilled water. Gently mix.

[1] GelRed™ Nucleic Acid Gel Stain, 10,000X in water, Catalog # 41003 Biotium.

4.2 Polyacrylamide Gel Electrophoresis (PAGE)

Polyacrylamide gel electrophoresis is used when higher band resolution is required for co-dominant markers.

In the laboratory we are currently only using mainly the system Model MGV-216-33 from CBS Scientific Co. with the dimensions 45 × 19 × 18 cm allow to electrophorese samples of one 96-well microtiter plates and to load samples using a multichannel pipette. We use combs with 63 wells of 1 mm width so that multi-channel pipettes fit to every alternate well. This is very convenient when a large number of samples have to be loaded.

Set up glass of plates

1. Clean glass plates before every use with 70% ethanol.

2. Assemble glass plates and sealers using clamps. Be sure the sealers are at the appropriate position between the two glass plates to avoid leaking. Two gels can be set in one apparatus.

Note: We use wipers (Kimwipes®EX-L, Cat.# 3425610, Kimberly-Clark®) to clean the glass plates.

Gel preparation (Non-denaturing gels)

We recommend using 12 to 14% of 29:1 acrylamide: bisacrylamide depending on the fragment size of the amplified marker products. Concentration may be reduced (e.g., to 10%) or increased (e.g., to 16%) for larger or smaller fragments, respectively. We purchase pre-mixed acrylamide/bisacrylamide from Sigma (Cat.# A2792). The stock can be stored at 4°C for a few months.

1. Place the plates with gels in the apparatus and add running buffer. One electrophoresis tank requires about 650 ml of 1X TG or 1X TBE.

2. Remove the combs and flush out the wells using a syringe. This is a critical step, especially for polymorphic bands that are close to each other. Otherwise, unpolymerized acrylamide solution will be polymerized at the bottom of the wells and will affect the migration of the fragments.

Note: TG buffer requires a longer time for running, but results in better band separation. The pH of TBE buffer should be adjusted with acetic acid so that the background of the gels is much reduced after silver staining. The same stock of TBE should be used to prepare both the gel and the running buffer.

PAGE Gel Preparation for the Model MGV-216-33

Concentration	10%			
	1 gel	2 gels	3 gels	4 gels
Acrylamide 40% (29:1)[1]	15 ml	25 ml	35 ml	45 ml
TG Buffer 5X	12 ml	20 ml	28 ml	36 ml
ddH_2O	33 ml	55 ml	77 ml	99 ml
APS 25%[2]	240 µl	400 µl	560 µl	720 µl
TEMED[3]	30 µl	50 µl	70 µl	90 µl

Concentration	12%			
	1 gel	2 gels	3 gels	4 gels
Acrylamide 40% (29:1)	18 ml	30 ml	42 ml	54 ml
TG Buffer 5X	12 ml	20 ml	28 ml	36 ml
ddH_2O	30 ml	50 ml	70 ml	90 ml
APS 25%	240 µl	400 µl	560 µl	720 µl
TEMED	30 µl	50 µl	70 µl	90 µl

Concentration	14%			
	1 gel	2 gels	3 gels	4 gels
Acrylamide 40% (29:1)	21 ml	35 ml	49 ml	63 ml
TG Buffer 5X	12 ml	20 ml	28 ml	36 ml
ddH_2O	27 ml	45 ml	63 ml	81 ml
APS 25%	240 µl	400 µl	560 µl	720 µl
TEMED	30 µl	50 µl	70 µl	90 µl

Concentration	16%			
	1 gel	2 gels	3 gels	4 gels
Acrylamide 40% (29:1)	24 ml	40 ml	56 ml	72 ml
TG Buffer 5X	12 ml	20 ml	28 ml	36 ml
ddH_2O	24 ml	40 ml	56 ml	72 ml
APS 25%	240 µl	400 µl	560 µl	720 µl
TEMED	30 µl	50 µl	70 µl	90 µl

[1] Acrylamide /Bis-Acrylamide 29:1 (Sigma A2792).

[2] APS: Ammonium Persulfate (Sigma A3678).

[3] TEMED: N,N,N',N'-Tetramethylethylenediamine (Sigma T7024).

NOTE: Polymerization is caused by both the APS and TEMED. Once you add those components, you should quickly pour the gel. The amount of APS added may be changed depending on ambient temperature and time required for polymerization.

CAUTION: Acrylamide, a potent neurotoxin, is absorbed through the skin. It should be handled in a fume hood–wear a lab coat, eye protection, mask, and gloves when handling powdered acrylamide and bisacrylamide, and use extra precaution. Wear a lab coat, gloves, and eye protection, when handling solutions containing these chemicals.

CAUTION: APS is a hazardous chemical–wear a lab coat, eye protection, and gloves when handling.

CAUTION: TEMED is highly flammable and corrosive–wear a lab coat, eye protection, and gloves when handling.

Sample loading

1. Add 2-4 µl of 1X loading buffer including bromophenol blue (BPB) and xylene-cyanole (XC) to each sample and load 6-10 µl of each sample using a micropipette. Use an appropriate MW marker in two wells; we use about 100 ng of the 100 bp ladder or Phi (ϕX174RF) plasmid digested with *Hae*III.

Electrophoresis

1. Run gels at constant 250-300V for 3-6 h, depending on the acrylamide concentration. Usually the BPB has run out of the gel and the XC has either just run out or is at the bottom of the gel (depending on acrylamide concentration).

2. Remove gels from plates. We cut one or more corners of the gels to identify the direction and number of each gel after silver staining.

Silver staining

Trays are gently shaken throughout the steps. Wear gloves at all times and handle the gels gently because pressure and fingerprints produce staining artifacts. It is also important to use clean glassware and deionized distilled water because contaminants greatly reduce the sensitivity of silver staining.

1. Transfer each gel in 100 ml fixing solution, shake for 5 min and rinse in dH_2O.

2. Place each gel in 100 ml staining solution and shake for 10 min. The staining solution can be re-used many times.

3. Transfer each gel to 100 ml of developer 1X solution. When bands become visible, immediately transfer the gel to the stop 1X solution to stop further reaction. The gel can be kept in the stop solution until it is placed onto a light box with fluorescent lamps and photographed.

Note: Deionized-distilled water is recommended for all solutions involved in the staining process. Trays should be cleaned by wiping with soft wet paper towels to remove silver. If not cleaned, the surface of subsequent gels may become black because of the silver residue. The weaker the band intensity, the longer the developing time, resulting in a higher background. In this case, load more sample, or optimize PCR conditions to give better amplification.

Optional and non-toxic: Stain with GelRed

GelRed® STAINING PROTOCOL
GelRed™ Nucleic Acid Gel Stain, 10,000X in water.
Catalog Number: 41003 Biotium

1. Dilute 50 ul of GelRed 10,000x in 500 ml of distilled water. Gently mix.
2. Put the gel into the GelRed dilution
3. Incubate for 30 min.
4. Take a picture on UV lamp.-transilluminator.

PAGE: Solutions

5X loading buffer BPB and XC

STOCK	[FINAL]	50 ML	100 ML
1 M Tris-8.0	50 mM	2.5 ml	5.0 ml
0.5 M EDTA-8.0	5 mM	0.5 ml	1.0 ml
Sucrose[1]	25%	12.5 g	25 g
Bromophenol blue[2]	2 mg/ml	100 mg	200 mg
Xylene cyanole[3]	2 mg/ml	1000 mg	200 mg
ddH_2O		up to 50 ml	up to 100 ml

Dilute to prepare a 1X dilution with Sucrose (25%).
[1] Bromphenol Blue Sodium Salt 25g (Sigma B7021).
[2] Sucrose 10 kg (Sigma S5390).
[3] Xylene cyanole (Sigma X4126-10G).

10x TG Running Gel Buffer

	2 L	8 L	10 L
TrizmaBase[1] (MW=121.10)	60 g	240 g	300 g
Glycine[2] (MW=75.07)	288 g	1152 g	1440 g

[1] TrizmaBase (Sigma T1503).
[2] Glycine (Sigma G8898).

Ammonium persulfate (APS) 25%

STOCK	10 ml	20 ml	30 ml
Ammonium persulfate[1]	2.5 g	5.0 g	7.5 g

Dissolve in ddH_2O to the final volume. The stock can be stored at 4°C for up to a month.
[1] Ammonium Persulfate 100g (Sigma A3678).

Fixing solution

	2 L
Absolute EtOH[1]	200 ml
Acet Acid Glacial[2]	10 ml
dH_2O	1800 ml

[1] Absolute GR for analysis Ethanol (Merck 100983).
[2] Acet Acid Glacial 500 ml (J.T.Baker 9511-02).

Staining solution: 0.2% silver nitrate	
	2 L
$AgNO_3$ [1]	4 g
dH_2O	2000 ml

[1] Silver Nitrate ACS. Reagen 25g. (Sigma 209139) .Should be stored in dark.

CAUTION: Silver nitrate is an oxidizing corrosive-wear a labcoat, eye protection and gloves when handling.

Developer 10X Solution: 30% NAOH	
	2 L
Sodium Hydroxide, Pellets[1]	600 g
dH_2O	2000 ml

[1] NAOH Sodium Hydroxide, Pellets 2.5kg (Baker 3722-05).

CAUTION: Sodium Hydroxide is corrosive, wear a labcoat, eye protection and gloves, when handling.

Developer 1X Solution: 3% NAOH	
	2 L
Developer 10X	200 ml
36-38% Formaldehyde[1]	10 ml
dH_2O	1800 ml

[1] 37% Formaldehyde 1L (J.T.Baker 2106-02). Concentration of formaldehyde may vary depending on the company you purchase from. It should be added immediately before use.

CAUTION: Formaldehyde is a potential cancer hazard, a lachrymator, and combustible. It should be handled in a fume hood-wear a laboratory coat, eye protection, and gloves when handling and use extra precaution.

Stop 2X Solution	
	2 L
$Na_2EDTA \bullet 2H20$[1] (MW=372.24)	120 g
dH_2O	2000 ml

[1] EDTA Ethylenediaminetetraacetic Acid Disodium Salt 1kg, E5134.

GelRed® STAINING
Dilute 50 µl of GelRed1 10,000x in 500 ml of distilled water. Gently mix.

[1] GelRedTM Nucleic Acid Gel Stain, 10,000X in water, Catalog # 41003 Biotium.

5. Kompetitive Allele Specific PCR Genotyping System (KASP™)

The Kompetitive Allele Specific PCR genotyping system (KASPar) is a novel homogeneous fluorescent endpoint genotyping technology developed by LGC Genomics (www.LGCgenomics.com). KASPar offers a simple, cost effective and flexible way to determine both SNP genotypes and insertion/deletions (InDel) genotypes. Analysis can be carried out in 96, 384 and 1536 well plate formats.

The chemistry has the advantage to be easily implemented in-house. The assays have shown to work with a number of plate readers such as ABI7700, ABI7900, BMG PherastarPlus Tecan Saffire & PerkinElmer Envision, just make sure you select the correct filters and detectors to read the fluorescent dyes (FAM and HEX). Furthermore, LGC Genomics directly and through the Generation Challenge Program/Integrated Breeding Platform (https://www.Integratedbreeding .net/482/communities/genomics-crop-info/crop-information/gcp-kaspar-snpmarkers) offer genotyping services for several crops including wheat.

Information on wheat KASP probes can be found on various websites:

- LGCgenomics wheat panel (http://www.lgcgroup.com/wheat/#.VfMk3q10y70)
- CerealsDB KASP SNPs (http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/index NEW.php)
- Integrated Breeding Platform (https://www.integratedbreeding.net/482/communities/genomics-crop-info/crop-information/gcp-kaspar-snp-marker)

The KASP genotyping system is comprised of two components (see also Figure 1).

1. The **KASP Assay Mix** is specific to the SNP or InDel to be targeted and consists of two competitive, allele-specific forward primers and one common reverse primer. Each forward primer incorporates an additional tail sequence that corresponds with one of two universal FRET (fluorescent resonance energy transfer) cassettes present in the KASP Master mix.

2. The **KASP Master Mix** (supplied at 2X concentration) contains the two universal FRET cassettes (FAM and HEX), ROX™ passive reference dye, Taq polymerase, free nucleotides and $MgCl_2$ in an optimized buffer solution.

DNA Samples

DNA samples may be arrayed in any microliter PCR plate; In our laboratory typically 384 well plates are used. We recommend amounts of DNA between 2.5 and 3.5 µl of approximate 50 ng/µl for 384 well plates due to the large wheat genome. Genotyping should be carried out on at least 24 samples to ensure there are sufficient genotypes to develop cluster plots. It is also strongly recommended that at least three samples fluorescent labelled and clustering as FAM, HEX, and heterozygotes (HET) and two water samples (NTCs) are included per 384 well plate as controls. After arraying, the plates should be briefly centrifuged and samples can be dried in an oven for 45 min at 65°C. To dry the DNA of your samples in the plate is useful when performing large scale genotyping, as it allows many plates to be prepared in advance, without the concern of sample evaporation altering the reagent concentrations. The dried DNA samples are also stable at room temperature for at least three months if protected from moisture. Faster drying will occur if the oven is fan-assisted. A quick visual check is all that is required to ensure the samples are dry.

Dispensing the KASP genotyping reactions

In our laboratory we are working with a total volume of 5 µl per reaction for 384 well plates. 5 µl is the smallest volume we can currently run with our standard ABI thermal cyclers. The calculations to prepare the genotyping reaction bulk are carried using the customer KASP reaction volume calculator (Figure 2).

Depending of the total volume of the PCR reaction and if wet or dry DNA is used the addition of what needs to be considered (see Table 1). Volumes must be scaled-up depending on the number of reactions required. An optimal $MgCl_2$ concentration of 2.2 mM, in the reaction has been observed, however only a final concentration of is supplied in the 1.8 mM in the KASP Master Mix as some assays only work at this lower concentration (see troubleshooting/optimizing

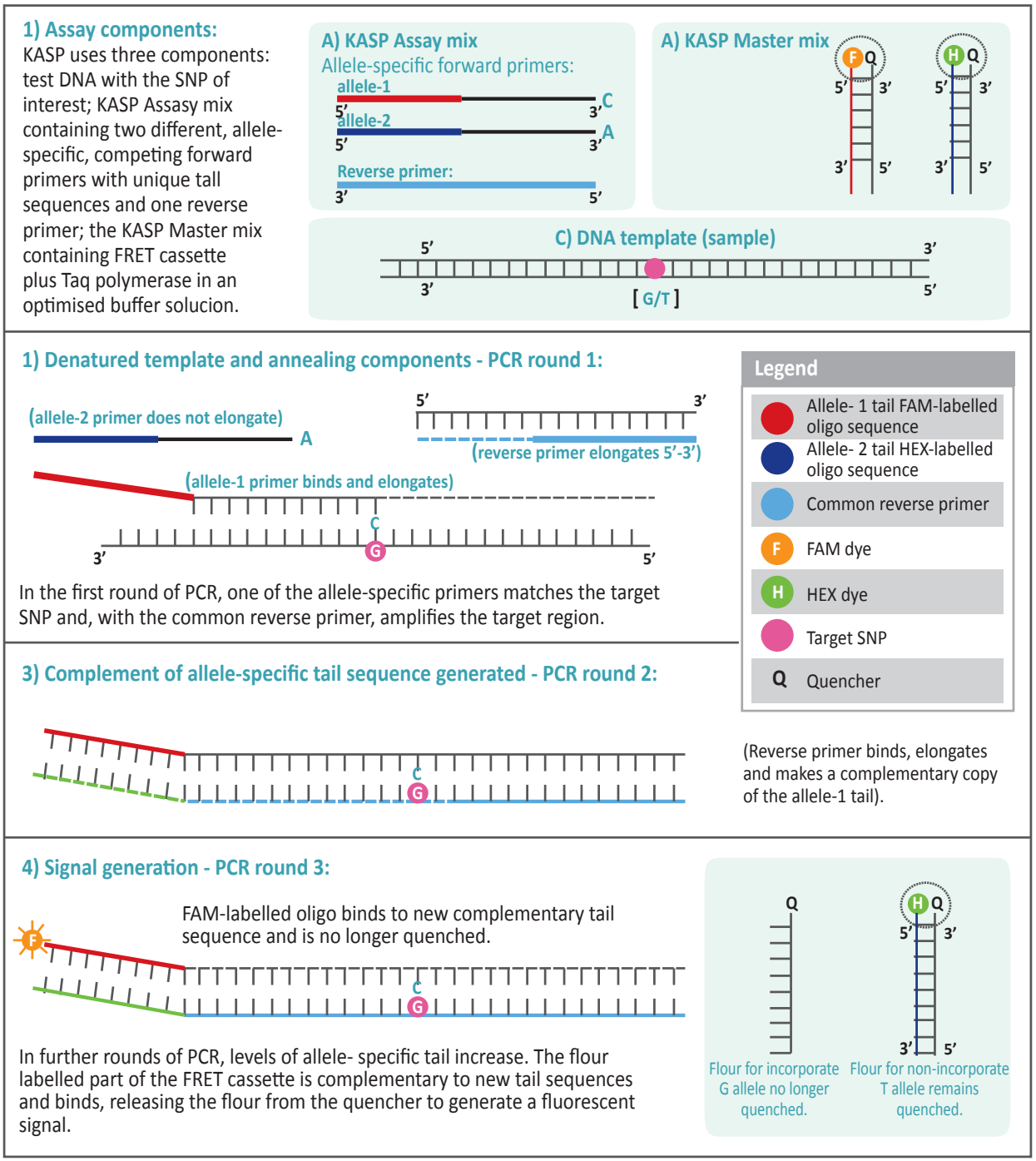


Figure 1. How KASPar works. Graph taken from: http://www.lgcgroup.com/LGCgroup/media/PDFs/products/genotyping/KASP-brochure.pdf?ext=.pdf.

guide). Therefore for most assays, the final $MgCl_2$ concentration must be adjusted to 2.2 mM before use (i.e. an increase of 0.4 mM, to increase the concentration from 1.8 mM to 2.2 mM). 100 µl of KASP Assay Mix is sufficient to carry out at least 1420 genotypes in 384 well plates). The KASP Assay and Master Mix can be safely stored 1 year at -20°C, or indefinitely at -80°C. We divide the KASP Master Mix into 15ml and 5ml aliquots and the KASP Assay Mix into 100 µl aliquots. Frequent freeze/thawing of both KASP Mixes may adversely affect performance; also we want to avoid contamination as much as possible. We use tubes used are light-protective and prepare aliquots.

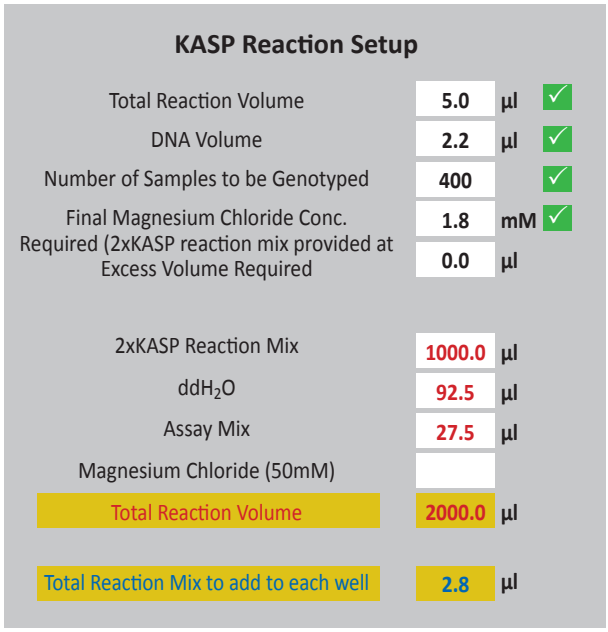


Figure 2. Customer KASP reaction volume calculator, it is an useful tool to calculate the µl of Master and Assay Mix that are necessary to carried a specific number of samples to be genotyped.

Table 1. Constituent reagent volumes for making KASP genotyping mix in 96-well reaction volume (10 µl final volume) or 384-well reaction volume (5 µl final volume).

Component	Wet DNA Method (µl)		Dry DNA Method (µl)	
DNA[1]	2.5	5	N/A	N/A
2X KASP Reaction Mix	2.5	5	2.5	5
Assay Mix	0.07	0.14	0.07	0.14
dd H_2O	N/A	N/A	2.5	5
Total Reaction Volume	5 µl	10 µl	5 µl	10 µl

[1] DNA samples diluted to final concentration of 30 ng/µl in the PCR.

[2] KASP 2XRxn Mix V4.0 25 ml (Kbioscience-uk KBS-1016-002).

Dispensing can be carried out robotically or manually with a suitable pipette, depending on plate type and sample number. Using dry DNA we dispense 5 µl of the final KASP Genotyping Mix to each sample. To seal our plates and avoid reaction losses during PCR, we use a rubber cover lid with a guaranteed, seal that can be reused many times.

Thermal cycling conditions

PCR cycling can be performed on any PCR thermal cycler, optimal results are generally obtained using a touchdown cycling program with two temperature steps detailed below. However, cycling conditions can be adapted as required.

Touchdown cycling program:

94°C for 15 minutes *Hot-start enzyme activation*

94°C for 20 seconds
65°C for 60 seconds[1]
72°C for 30 seconds
} 11 cycles

94°C for 20 seconds
57°C for 60 seconds
72°C for 30 seconds
} 26 cycles

72°C for 2 min
20°C final

[1] Touchdown over 65-57°C for 60 seconds 10 cycles (dropping 0.8°C per cycle)

In some cases we us a 2 step cycling program detailed below:

2 step cycling program:

94°C for 15 minutes *Hot-start enzyme activation*

94°C for 10 seconds
57°C for 5 seconds
72°C for 10 seconds
} 20 cycles

94°C for 10 seconds
57°C for 20 seconds
72°C for 40 seconds
} 24 cycles

Plate Reading

The plate should be read on a suitable fluorescent plate reader according to its specifications. In our laboratory we are using a BMG Pherastar Plus (Figure 3). Reading temperature should be 25°C or below (preferably ambient). Most FRET-capable plate readers (with the relevant filter sets) can be used in conjunction with KASP. Some plate readers can be set to read at a range of temperatures but

Figure 3. Pherastar plate reader from BMG Labtech Company.

elevated temperatures (above 40°C), will lead to poor/no data. Whilst using real time PCR machines, plates should be read at ambient temperature after the completion of the PCR run, rather than using the real time data to generate end point curves. KASP uses the fluorophores FAM (Fluorescein) and HEX (CAL Fluor Orange 560) for distinguishing genotypes. The relevant excitation and emission wave lengths are show in Table 2. ROX is also used to allow normalization of variations in signal caused by differences in well-to-well liquid volume by dividing FAM and HEX values by the passive reference value for a particular well (see Table 2). If using a plate reader optimized for use with the dye VIC (e.g. Applied Biosystems), no modification of settings will be necessary as the excitation and emission values for VIC and HEX are extremely close.

Table 2. Excitation and Emission values for the fluorescent dyes used in the KASP chemistry.

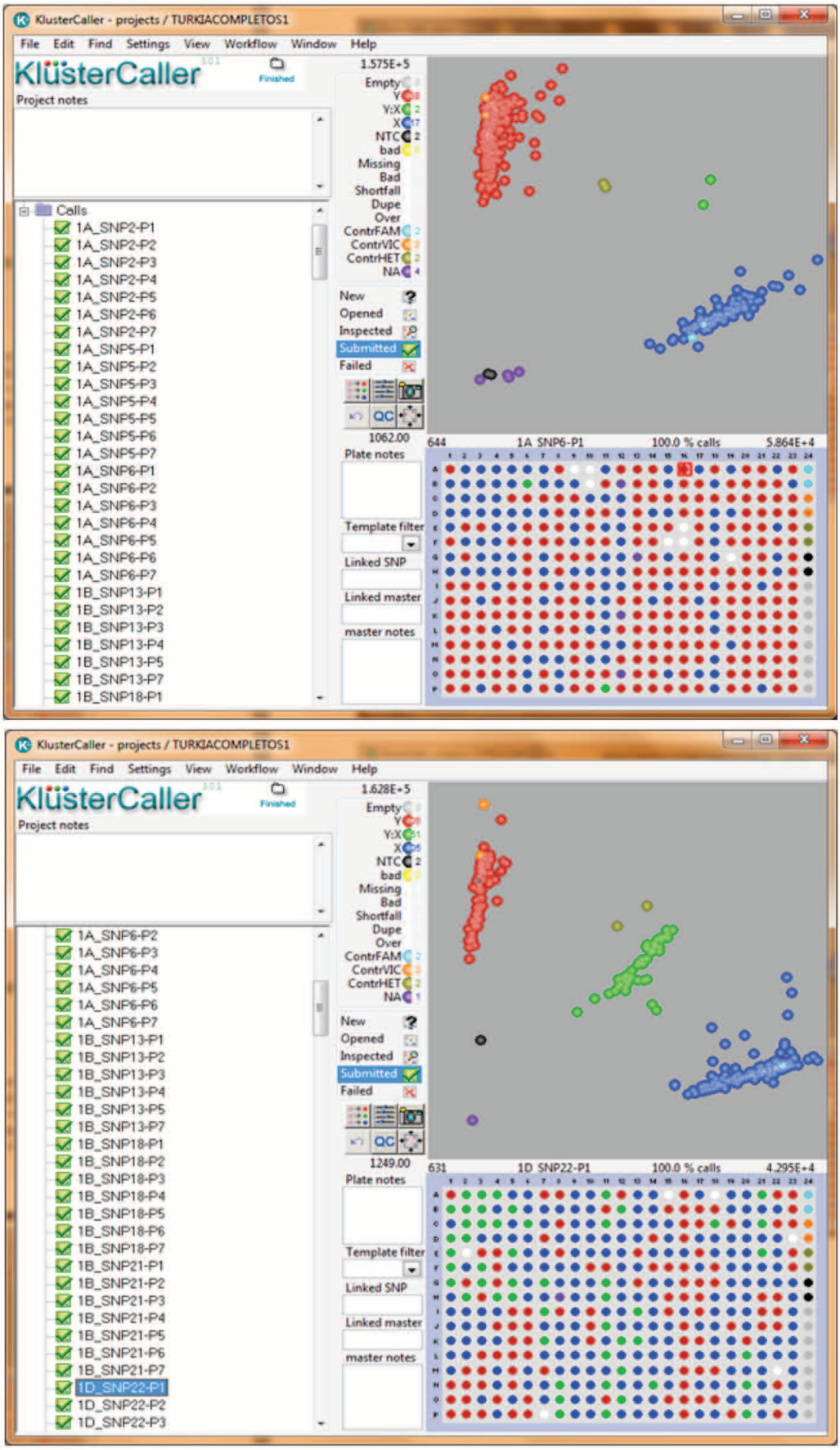
	Excitation (nm)	Emission (nm)
FAM (Fluorescein)	485	520
HEX (CAL Fluor Orange 560[1])	534	560
ROX (carboxy-X-rhodamine)	575	610

[1] The excitation and emission values for CAL Fluor Orange 560 are the same as those of VIC / JOE.

Geographical viewing of genotypic data (cluster plots)

Once the KASP reactions are complete, and the resulting fluorescence has been measured, the raw data must be interpreted to enable genotypes to be assigned to the DNA samples. The FAM and HEX fluorescence values are typically plotted on the X and Y axes of a Cartesian plot respectively. A sample that is homozygous for the allele reported by FAM will only generate FAM fluorescence during the KASP reaction. This data point will be plotted close to the X axis, representing high FAM signal and no HEX signal (blue data points in Figure 4). A sample that is homozygous for the allele reported by HEX will only generate HEX florescence and the data points will be plotted close to the Y axis (red data points in Figure 4). A sample that is heterozygous will contain both the allele reported by FAM and the allele reported by HEX. This sample will generate half as much FAM fluorescence and half as much HEX fluorescence as the samples that are homozygous for these alleles. This data point will be plotted in the center of the plot, representing half FAM signal and half HEX signal (green data points in Figure 4). To ensure the reliability of the results, a KASP reaction without any template DNA (Water) must be included as a negative control. This is typically referred to as a no template control or NTC. The NTC will not generate any fluorescence and the data point will therefore be plotted at the origin (black data points in Figure 4).

Figure 4. Examples of cluster plots of KASP SNPs in KlusterCaller.



All samples that have the same genotype will have generated similar levels of fluorescence and will therefore all appear close together on the plot. Based on the relative position of these clusters, it is possible to determine then genotype of all the data points. It is important that a sufficient number of individual samples are included in a KASP reaction plate to ensure that there are enough data points on the Cartesian plot to allow cluster analysis.

The fluorescence values can be plotted using Excel. A software package is additionally offered as a part of a full workflow management system (Kraken™) or a standalone version (KlusterCaller™). The inclusion of the passive reference dye (ROX) leads to tighter clustering and, as a result, more accurate calling of data.

Calling guidelines – Summary:

- Include a minimum of 24 data points on each plot to facilitate cluster analysis
- Check the scaling of the X and Y axes and, if required, rescale so that they are comparable. This will ensure that clusters are correctly centered and interpreted.
- Where possible, include positive, negative, heterozygotes and NTC controls for each KASP assay as this will facilitate data analysis.
- Do not view results for more than one KASP assay on an individual cluster plot – always view results assay by assay. Otherwise you might get confused.
- Do not view results from bread and durum wheat on one individual cluster plot, plots might be different for both species.
- Always check genotypes that are automatically assigned by the instrument software as they may need to be manually adjusted.
- The contamination of tips can be problem if you don't pipette carefully. You can identify any contamination easily when your NTC controls are plotting close to both X and Y axes.
- Always keep the same order when adding your controls. This helps in case your plate reader accounts for a technical some problem.

6. Primer Design

6.1 PCR Primer Design Guidelines

A good PCR primer design is essential to reveal successful amplification products of a particular target region. Important considerations are described below:

a) Primer Length:
It is generally accepted that the optimal length of PCR primers is 18-22 bp. This length is long enough for adequate specificity and short enough for primers to bind easily to the template at the annealing temperature.

b) Primer Melting Temperature:
Pairs of primers should have similar melting temperatures since annealing in a PCR occurs for both simultaneously. Primer Melting Temperature (T_m) by definition is the temperature at which one half of the DNA duplex will dissociate to become single stranded and indicates the duplex stability. Primers with melting temperatures in the range of 52-58°C generally produce the best results. Primers with melting temperatures above 65°C have a tendency for secondary annealing. The GC content of the sequence gives a fair indication of the primer T_m (see Wallace's rule below).

$$Tm = 4(G+C) + 2(A+T)°C$$

c) Primer Annealing Temperature:
The primer melting temperature is the estimate of the DNA-DNA hybrid stability and critical in determining the annealing temperature in a PCR reaction. Too high T_a will produce insufficient primer-template hybridization resulting in low PCR product yield. Too low T_a may fail to anneal and extend at all and may possibly lead to non-specific products caused by a high number of base pair mismatches.

d) GC Content:
The GC content (the number of G's and C's in the primer as a percentage of the total bases) of primer should be 40-60%.

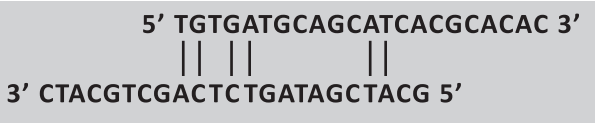
e) GC Clamp:
The presence of G or C bases within the last five bases from the 3' end of primers (GC clamp) helps promote specific binding at the 3' end due to the stronger bonding of G and C bases. More than 3 G's or C's should be avoided in the last 5 bases at the 3' end of the primer.

f) Primer Secondary Structures:
Presence of the primer secondary structures produced by intermolecular or intramolecular interactions can lead to poor or no yield of the product. They adversely affect primer template annealing and thus the amplification. They greatly reduce the availability of primers to the reaction.

i) Hairpins: It is formed by intramolecular interaction within the primer and should be avoided. Optimally a 3' end hairpin with a ΔG of -2 kcal/mol and an internal hairpin with a ΔG of -3 kcal/mol is tolerated generally.

ii) Self Dimer: A primer self-dimer is formed by intermolecular interactions between the two (same sense) primers, where the primer is homologous to itself. Generally a large amount of primers are used in PCR compared to the amount of target gene. When primers form intermolecular dimers much more readily than hybridizing to target DNA, they reduce the product yield. Optimally a 3' end self dimer with a ΔG of -5 kcal/mol and an internal self dimer with a ΔG of -6 kcal/mol is tolerated generally.

iii) Cross Dimer/Primer Dimer: Primer cross dimers are formed by intermolecular interaction between sense and antisense primers, where they are homologous. Optimally a 3' end cross dimer with a ΔG of -5 kcal/mol and an internal cross dimer with a ΔG of -6 kcal/mol is tolerated generally.



All these primer secondary structures can be checked while designing primers using OligoAnalyzer tool on IDT site (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/).

g) Repeats:
A repeat is a di-nucleotide occurring many times consecutively and should be avoided because they can misprime. For example: ATATATAT. A maximum number of di-nucleotide repeats acceptable in an oligo is four di-nucleotides.

h) Avoid Cross Homology:
To improve specificity of the primers it is necessary to avoid regions of homology. Primers designed for a sequence must not amplify other genes in the mixture. Commonly, primers are designed and then BLASTed to test the specificity.

6.2 SSR/STS Primer Design (with Batchprimer3)

1. Make a text file (.txt extension) or MS word file having the sequences of SSR clones in FASTA format. See below two examples of FASTA formats:

   >seq1
   agagattaggatcgatcgcgctctctctctctctcgatcgagatcgat
   ggccatcatcatcatcatcattgagatatagcgcgatatcgagagatctc
   agaatagatatcgcgctatagagagatcgagagagagtaga

   >seq2
   agagataggaatatgagatagcgggggggggggggcgctatacgcgctcg
   gagagagatctctctctctcttatagagatcgatcgactagctagatata
   agactcactcactcactcactcactcagcgcgat

2. Type **Batchprimer3** in Google search or go to the link http://probes.pw.usda.gov/batchprimer3/. The advantage of using **Batchprimer3** over other primer designing programs is that it can process for up to 500 sequences at a time and one can perform both SSR screening and primer designing at the same time.

3. On this site, you will see two servers for primer design, Primer design server 1 (Albany) and Primer design server 2 (Albany). Click Primer design server 1 (Albany).

4. Go to option choose primer type and click SSR screening and primers or Generic primers.

5. Now upload your FASTA sequence file (.txt file) or copy and paste your sequences in FASTA format in the blank space provided under Input Sequences

6. For SSR, markers go to option SSR Screening and click the SSR pattern types (di-, tri-, tetra-, penta- and hexa-nucleotide) you are looking for in your sequences or alternatively leave all options on. Then enter the minimum number of repeats you will allow or alternatively go ahead with default options which are 6 for dinucleotide, 4 for trinucleotide and 3 for each of the tetra-, Penta- and hexa-nucleotide motifs.

7. Go to General Settings for Generic Primers and tune the options of product size, primer size, primer Tm, primer GC etc. according to your need or alternatively leave them unchanged and go ahead with default settings.

8. Once you are satisfied with all primer designing parameters, go to the top of the site and click Pick Primers. You will get results from within few seconds to few minutes based on the number of sequences in the file.

9. For each sequence, you will get number of different types of SSR motifs and primer pairs designed for each of the SSR motif or different primer pairs for any generic sequence.

10. You can download entire result as a zip file or alternatively separately for SSR/STS screening and primers by choosing the options at the top of the output page.

6.3 CAPS Primer Design

1. Make a text file (.txt extension) or MS word file having two sequences in FASTA format or alternatively use the alignment file from CLUTSALW. The following 2 sequences in FASTA format can be used for practice:

>131017-11_C14_VrnA1c-1_VrnA1c.ab1
GGCGATTCGTTCGAACGGATTACTACTGCTTAGTAATATCCATTGTTGTT
TGTAATCTTGCTGAGAAAGCAACATTACCATCAGCTTAAGTGGTGAGTCA
GTCATAACCCCACCGTGTTCACTTCCCGAACTCCTTGGAAAAGAGACGAT
CACGTAACGCACGCGGTTGGTGTATTTTAATTGGGTTCAGTGTCAAGTTC
TCTAAAATCGGATATTATAAATTTTTAAGTCGCCACATAACCGCGGGCAC
GGCTTCCGAAAAGATTTAGCCCTGCAGGGGTGCACCAAGTAGTCCATTAT
AAATTACCACATGCATCGGATGGAACATCCTCACACCATGATAACACGAT
GCTTACAATAAGGAACCCCGGTGGACAAGCCACTCGTCAAAGGCAAAACT
AAACCAGCAAGACCACCCGGTGTGTCGTCACCCCGATAAGAGCCGCGCCT
ATTTTCTAGGGTTGCCTAACCCTTGGGATCCCTTGGACCACCTTACTATG
TGCATGTTTTCTTTTCACACGGGCATTTATCTGCTTTGGCATCAAAGCTT
TCATTTGAAAATTTGCTACTACCACCTATATTTGTACTGACAATACCTTT
GCATGGACCCACACATTAGGTTTTAAAATGGTTCTCACATTCGGGGGGCC
TTTAATTTAAAA

>131017-11_D14_VrnA1c-2_VrnA1c.ab1
AAGGGAAACGTTCGAAGGATCGCTACTGCTTAGTAAATATCCATTGTTGT
TTGTAATCTTGCTGAGAAAGCAACATTACCATCAGCTTAAGTGGTGAGTC
AGTCATAACCCCACCGTGTTCACTTCCCGAACTCCTTGGAAAAGAGACGA
TCACGTAACGCACGCGGTTGGTGTATTTTAATTGGGTTCAGTGTCAAGTT
CTCTAAAATCGGATATTATAAATTTTTAAGTCGCCACATAACCGCGGGCA
CGGCTTCCGAAAAGATTTAGCCCTGCAGGGGTGCACCAAGTAGTCCATTA
TAAATTACCACATGCATCGGATGGAACATCCTCACACCATGATAACACGA
TGCTTACAATAAGGAACCCCGGTGGACAAGCCACTCGTCAAAGGCAAAAC
TAAACCAGCAAGACCACCCGGTGTGTCGTCACCCCGATAAGAGCCGCGCC
TATTTTCTAGGGTTGCCTAACCCTTGGGATCCCTTGGACCACCTTACTAT
GTGCATGTTTTCTTTTCACACGGGCATTTATCTGCTTTGGCATCAAAGCT
TTCATTTGAAAATTTGCTACTACCACCTATATTTGTACTGACAATACCTT
TGCATGGACCCACACATTAGGTTTTAAAATGGTTCTCACATTCGTGTGGG
TCTTACTTAAACCGGCCC

2. Go to the Sol Genomics Network (SNG) webpage, CAPS Designer at http://solgenomics.net/tools/caps_designer/caps_input.pl

3. Either choose option of unaligned fast sequences and copy and paste your FASTA sequences in the blank space provided or clustal alignment option and upload your alignment file from CLUSTALW.

4. The second option of Find enzymes priced less than $65/1000u is optional; clicking it would display only inexpensive restriction enzymes that can be used for a CAPS marker.

5. Click Find CAPS

6. The output file will give you list of CAPS candidates along with expected band sizes and prices of the enzymes.

7. The results can be downloaded as a plain text file at the top of the site.

6.4 dCAPS Primer Design

1. Type dCAPS finder 2.0 in Google or go to site http://helix.wustl.edu/dcaps/

2. Paste the two haplotypes, with no gaps in sequence, into the blank boxes provided. Provide upto 20 bases nucleotides flanking the SNP. The following example can be taken for practice:

Wild Type TTAGCAGTAGCCTATTAACAGG**C**TCGATTGATACATTACGGTAC
Mutant TTAGCAGTAGCCTATTAACAGG**G**TCGATTGATACATTACGGTAC

3. In option How many mismatches in the primer, enter the number of mismatches allowed in your primer. If you enter “0” it will give you only enzyme sites and not the primers. In this case, you have to design a pair of primers to amplify the pertinent sequence containing the target restriction sequence separately using Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/).

4. The advice is to enter “1”. This will allow identification of potential dCAPS.

5. Click Submit.

6. The output will display the available enzymes sites and the possible primer sequences with one mismatch. The colored sequence indicates mismatches.

7. The output provides the primer sequence to introduce the restriction site. You need to design a reverse primer that will be used to amplify the region with the dCAPS primer output using Primer3Plus.

6.5 KASP Primer Design

1. Type Primer3Plus in Google or go to site http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/. You can use the following example:

   >Example
   TGGTCACCACCCTGAAGCGCGCCGTCAAGGTCGTCGGCACGCCGGCGTACCA
   TGAGATGGTCAAGAACT[G/C] CATGATACAGGATCTCTCCTGGAAGGTAAGTC
   GTCTCTGGTTCAGTATGCACTTCCTGGAACAACTAAGAGTGAAGGGC

2. In the blank space provided, paste a single sequence in FASTA format with the SNP clearly marked by brackets []. Remember to have atleast 50-100 bp high quality, clean and conserved sequence on either side of the SNP.

3. Go to advanced settings and put the minimum, optimum and maximum product size and product Tm without changing other default parameters. The suggestions for minimum, optimum and maximum product sizes are 20-50 bp, 50-100 bp and ~150 bp, respectively. For product Tm, the suggestions for minimum, optimum and maximum Tm are 56°C, 57°C and 58°C.

4. Click Pick Primers. The programme will return you with 5 primer pair options. You require at least one of the primers (forward or reverse), returned by the programme, to be overlapped with the SNP. Many of the times, you will get both primers far away from the target SNP. In such a case, move one of the primers (depending on whichever is closer to the target SNP) closer to the SNP so that its 3’ end overlaps with the SNP. For example, with the above example sequence you will get both primers (Forward primer is purple and Reverse primer is yellow) far away from the target SNP in the first instance. Hence, to go closer to the target first try to delete 10-20 bp of the sequence from left side and go to advanced setting with minimum, optimum and maximum product size settings of 20, 30 and 30, respectively. This way programme will be forced to select primers close to the target SNP. With this strategy you will see that the forward purple primer is just 5 nucleotides away from the target SNP.

5. However, the forward primer is still far from the target SNP. Now either move the forward primer 5 nucleotides in the 3’ end direction or simply add 5 nucleotides + 1 nucleotide base in the forward primer (as maximum primer size of 26 bases is allowed) originally returned by the programme. Now try these two kind of forward primer options in the programme. Go to advanced settings and again change minimum, optimum and maximum product size settings to 20, 30 and 30, respectively and click Pick primers. Either of these options might work.

6. If deleting sequences from left side does not work, try deleting sequence from the right side. This way programme will return reverse primer close to the target SNP. Then using the same methodology as described for forward primers in steps 4 and 5, you can design two reverse primers targeting the SNP.

7. For designing KASP primers for multiple sequences in one easy step, use the PolyMarker tool (http://polymarker.tgac.ac.uk/). PolyMarker has been specially constructed for designing primers from wheat genome. PolyMarker generates a multiple alignment between the target SNP sequence and the IWGSC chromosome survey sequences for each of the three wheat genomes. It generates KASP assays which are based on a three primer system. Two diagnostic primers incorporate the alternative varietal SNP at the 3' end, but are otherwise similar. The third common primer is preferentially selected to incorporate a genome-specific base at the 3' end, or a semi-specific base in the absence of an adequate genome specific position.

8. The input file must be uploaded as a CSV file (can be exported from Excel) with the following columns:
   a. Gene id: A unique identifier for the assay. It must be unique on each run.
   b. Target chromosome: In the form 1A, 2D, 7B, etc.
   c. Sequence: The sequence flanking the SNP. The SNP must be marked in the format [A/T] for a varietal SNP with alternative bases, A or T.

9. PolyMarker takes ~1 minute per marker assuming an input sequence of 200 bp (with the varietal SNP in the middle). Longer sequences can be used, but this will slow down the initial BLAST against the wheat survey sequence.

7. General Laboratory Procedures, Equipment Use, and Safety Considerations

7.1 Safety Procedures

A. Safety General Tour

Any new laboratory user must be introduced to the lab safety procedures and other guidelines in consultation with his/her supervisor and colleagues. The laboratory area which require special attention to safety procedures e.g., gel staining area should be visited. In that moment also all the personal protective equipment should be delivered (lab coat, googles and gloves). It is mandatory to respect any indication received in this safety general tour.

B. Chemicals

A number of chemicals used in any molecular biology laboratory are hazardous. All manufacturers of hazardous materials are required by law to supply the user with pertinent information on any hazards associated with their chemicals. This information is supplied in the form of Material Safety Data Sheets or MSDS (see also Table 1 and 2). This information contains the chemical name, CAS#, health hazard data, including first aid treatment, physical data, fire and explosion hazard data, reactivity data, spill or leak procedures, and any special precautions needed when handling this chemical. A file containing MSDS information on the hazardous substances should be kept in the laboratory for revision if required. This file should have to be updated annually due to their health hazard could change in consequence its classification. In addition, MSDS information can be accessed in the web. You are strongly urged to make use of this information prior to using a new chemical and certainly in the case of any accidental exposure or spill. The instructor/laboratory manager should be notified immediately in the case of an accident involving any potentially hazardous reagents.

The following chemicals are particularly noteworthy (see Table 1):

- Acrylamide - potential neurotoxin
- Chloroform – carcinogen
- Ethidium bromide – carcinogen

These chemicals are not harmful if used properly: always wear lab coat, googles, gloves and safety shoes when using potentially hazardous chemicals and never mouth-pipet them. If you accidentally splash any of these chemicals on your skin, immediately rinse the area thoroughly with water and inform the instructor, supervisor or any lab user to get the first aid treatment. Discard the waste in appropriate containers.

C. Ultraviolet Light

Exposure to ultraviolet light can cause acute eye irritation. Since the retina cannot detect UV light, you can have serious eye damage and not realize it until 30 min to 24 hours after exposure. Therefore, always wear appropriate eye protection when using UV lamps.

D. Electricity

The voltages used for some equipment but also electrophoresis are sufficient to cause electrocution. Cover the buffer reservoirs during electrophoresis. Always turn off the power supply and unplug the leads before removing a gel.

E. General Housekeeping

All common areas should be kept free of clutter and all dirty dishes, electrophoresis equipment, etc. should be dealt with appropriately. Since you have only a limited amount of space to call your own, it is to your advantage to keep your own area clean. Since you will use common facilities, all solutions and everything stored in an incubator, refrigerator, etc. must be labeled. In order to limit confusion, each person should use his initials or other unique designation for labeling plates, etc. Unlabeled material found in the refrigerators, incubators, or freezers may be destroyed. Always mark the backs of the plates with your initials, the date, and relevant experimental data, e.g. strain numbers.

Table 1. Health hazard, flammability and reactivity of reagents frequently used in the CIMMYT wheat laboratory.

Reagent	OSHA damage description	Warning word	NFPA Rating			Process in which it is used
			Health	Flammability	Reactivity	
ACETIC ACID	Harmful by skin absorption. Skin sensitizer. corrosive	Danger	3	2	0	Acrylamide
ACRYLAMIDE	Carcinogen. Teratogen. Mutagen. Reproductive hazard. irritant	Danger	2	1	0	Acrylamide
AMMONIUM PERSULFATE	Oxidizer. Irritant	Danger	2	0	1	Acrylamide
FORMALDEHYDE	Skin sensitizer. Carcinogen. Effect of target organ	Danger	3	2	0	Acrylamide
GLYCINE	No danger	Danger	1	0	0	Acrylamide
SODIUM HYDROXIDE	Corrosive	Danger	3	0	0	Acrylamide
SILVER NITRATE	Oxidizer. Carcinogen. Effect of target organ. Corrosive	Danger	3	0	2	Acrylamide
TEEMED	Toxic if ingested Corrosive	Danger	3	3	0	Acrylamide
BORIC ACID	Teratogen. Reproductive hazard	Danger	0	0	0	Electrophoresis
CHLOROFORM	Irritant. Carcinogen	Attention	3	0	0	DNA extraction
SODIUM CHLORIDE	Irritant	Attention	3	0	0	DNA extraction
CTAB	Corrosive. Causes burns	Danger	2	1	0	DNA extraction
ISOPROPANOL	Effect of target organ. Irritant	Danger	2	3	0	DNA extraction
MERCAPTOETHANOL	Skin sensitizer. Corrosive. Mutagen. Toxic	Danger	3	2	0	DNA extraction
OCTANOL	Effect of target organ. Irritant	Attention	2	2	0	DNA extraction
HYDROCHLORIC ACID	Toxic by inhalation. Harmful if swallowed. Corrosive	Danger	3	0	0	Solution preparation
ETHIDIUM BROMIDE	Very toxic by inhalation. Harmful if swallowed. Mutagen	Danger	4	0	0	DNA staining
EDTA	No danger	Attention	0	0	0	Various
TRIZMA BASE	Irritant	Attention	0	0	0	Various
ETHANOL	Effect of target organ. Irritant	Danger	2	3	0	Various

Code	meaning	Risk description					
		Health		Flammability		Reactivity	
4	**Very serious**	4	May cause death or major injury even with medical treatment	4	Volatile highly flammable gas or flammable liquids	4	Prepared to detonate or explode if exposed to fire
3	**Serious**	3	May cause serious injury even with medical treatment	3	Ignition can be started at normal temperatures	3	It can explode, but it requires a source of ignition under confinement
2	**Moderate**	2	May cause injury and requires immediate treatment	2	Ignition under moderate heating	2	Normally unstable, but won't explode. Violent chemical reaction
1	**Light**	1	May cause irritation if not treated	1	Moderate ignition after preheating	1	Normally stable. Unstable at high temperatures and under pressure. Reagent in water
0	**None**	0	No health risk	0	Nonflammable	0	Normally stable and unreactive with water

Table 2. Links to some relevant Material Safety Data Sheets.

Reagent	Link
Acetic Acid	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&productNumber=A6283&brand=SIAL&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsial%2Fa6283%3Flang%3Des
Acrylamide	http://www.sigmaaldrich.com/MSDS/MSDS/PleaseWaitMSDSPage.do?language=EN&country=MX&brand=SIGMA&productNumber=A9099&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsigma%2Fa9099%3Flang%3Des
Ammonium Persulfate	http://www.sigmaaldrich.com/MSDS/MSDS/PleaseWaitMSDSPage.do?language=EN&country=MX&brand=SIAL&productNumber=215589&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsial%2F215589%3Flang%3Des
Formaldehyde	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&productNumber=F8775&brand=SIGMA&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsigma%2Ff8775%3Flang%3Des
Glycine	http://www.sigmaaldrich.com/MSDS/MSDS/PleaseWaitMSDSPage.do?language=EN&country=MX&brand=SIGMA&productNumber=G8898&PageToGoToURL=%2Fsafety-center.html
Sodium Hydroxide	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&product Number=221465&brand=SIAL&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsial%2F221465%3Flang%3Des
Silver Nitrate	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&productNumber=209139&brand=SIAL&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsial%2F209139%3Flang%3Des
TEMED	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&productNumber=T9281&brand=SIGMA&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsigma%2Ft9281%3Flang%3Des
Boric Acid	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&product Number=B0394&brand=SIAL&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsial%2Fb0394%3Flang%3Des
Sodium Chloride	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&productNumber=S9888&brand=SIAL&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsial%2Fs9888%3Flang%3Des
Chloroform	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&product Number=288306&brand=SIAL&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsial%2F288306%3Flang%3Des
CTAB	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&productNumber=H9151&brand=SIGMA&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsigma%2Fh9151%3Flang%3Des
Isopropanol	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&productNumber=W292907&brand=ALDRICH&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Faldrich%2Fw292907%3Flang%3Des
Mercaptoethanol	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&productNumber=M6250&brand=ALDRICH&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Faldrich%2Fm6250%3Flang%3Des
Octanol	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&productNumber=W280100&brand=ALDRICH&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Faldrich%2Fw280100%3Flang%3Des
Hydrochloric Acid	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&product Number=339253&brand=SIAL&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsial%2F339253%3Flang%3Des
Ethidium Bromide	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&productNumber=E1510&brand=SIGMA&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsigma%2Fe1510%3Flang%3Des
EDTA	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&productNumber=E5134&brand=SIGMA&PageToGoToURL=%2Fsafety-center.html
Trizma Base	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&productNumber=T1503&brand=SIGMA&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsigma%2Ft1503%3Flang%3Des
Ethanol	http://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=MX&language=EN&productNumber=459836&brand=SIAL&PageToGoToURL=http%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsial%2F459836%3Flang%3Des

7.2 Disposal of Buffers and Chemicals

It is the clear responsibility of all research workers to ensure the safe and correct disposal of all wastes produced in the course of their work. Improper and irresponsible disposal of chemical wastes e.g. any uncontaminated, solidified agar or silver down drains is forbidden by law. Disposal of wastes should be discarded in a separate trash.

Ethidium bromide is a mutagenic substance that should be treated before disposal and should be handled only with double gloves. Ethidium bromide should be disposed of in laboratory controlled and labeled containers.

Acrylamide and Silver are also classified as hazardous waste and should be placed into suitable, leak-tight labelled containers.

Dirty glassware should be rinsed, all traces of agar or other substance that will not come clean in a dishwasher should be removed, all labels should be removed (if possible), and the glassware should be placed in the dirty dish bin. Bottle caps, stir bars and spatulas should not be placed in the bins but should be washed with hot soapy water, rinsed well with hot water, and rinsed three times with distilled water.

Different types of waste should never be mixed. Separate trash and labeled containers should be stored under suitable conditions in a waste room until collected by a specialized company.

7.3 Equipment

A. General Comments

It is to everyone's advantage to keep the equipment in good working condition. As a rule of thumb, don't use anything unless you have been instructed in the proper use. This is true not only for equipment in the laboratory but also departmental/institutional equipment. Report any malfunction to your manager or supplier. Keep also your basic equipment clean, e.g. rinse out all centrifuge rotors after use and in particular if anything spills. Maintain your equipment according to manufacturer's instructions, this helps to give your equipment maximum use. Try not to waste supplies - use only what you need for a better environment. If a supply is running low, plan accordingly; notify either the instructor/laboratory manager before the supply is completely exhausted. Occasionally, it is necessary to borrow a reagent or a piece of equipment from another laboratory. Apply the same rules.

B. Micropipettes

Most of the experiments you will conduct in the laboratory will depend on your ability to accurately measure volumes of solutions using micropipettes. The accuracy of your pipetting can only be as accurate as your pipettor and several steps should be taken to insure that your pipettes are accurate and are maintained in good working order. If they need to be recalibrated, do so. Do not drop it on the floor. If you suspect that something is wrong with your pipettor, first check the calibration to see if your suspicions were correct.

Part 2: Applications to Wheat Breeding

PART 2 of the manual targets marker deployment in the CIMMYT Global Wheat Program. Chapters on QTL/gene identification approaches, how to optimize MAS strategies, how MAS is currently used at CIMMYT for major trait categories such as biotic stresses and quality traits are described and we share our experience on recently developed prediction methods using genome-wide markers to archive genetic gain for more complex traits.

Chapter 1:
Overview of Bi-Parental QTL Mapping and Cloning Genes in the Context of Wheat Rust

Caixia Lan and Bhoja R. Banset
CIMMYT Int., Apdo. Postal 6-641,06600 Mexico, DF, Mexico

Quantitative trait locus (QTL) mapping

The regions within genomes that contain genes associated with a particular quantitative trait are known as quantitative trait loci (QTL). Breeders and geneticists have developed statistical methods to identify QTL by utilizing molecular markers. These methods have sought to answer basic questions concerning QTL (e.g. number, mode of action and size of action) and to map QTL on the genome to facilitate their manipulation for breeding purposes. In wheat, populations derived from single crosses of inbred lines have predominantly been used in QTL mapping experiments. In this chapter we describe the principles of bi-parental QTL mapping.

Development of bi-parental genetic populations

Several types of populations are currently being used in bi-parental QTL mapping studies. They include $F_{2:3}$ families, near-isogenic lines (NIL), doubled haploids (DH) and recombinant inbred lines (RIL). These populations differ in the method by which they are developed and the unique genetic structure they carry. $F_{2:3}$ families are also called tentative or temporary populations and their genetic constitution will change with recombination through further selfing or inbreeding (Xu 2010). Thus, $F_{2:3}$ families can only be used once for phenotyping and genotyping due to the variable genotypes in the following generation. This type of population is usually used to map the genes, which control qualitative traits such as seedling resistance genes to diseases in wheat. A NIL population consists of lines with nearly identical genetic backgrounds. It can be derived by continuous backcrossing of a hybrid to one of its parental lines so that lines only differ for a specific target trait or locus (Xu 2010). This kind of population is usually used for fine mapping of a targeted genetic locus and to estimates its genetic effect. The DH approach has several advantages that make it useful in genetic analysis in plants (Bal and Abak 2007; Ferrie 2007; Forster et al. 2007). The DH approach can specifically speed up the development of the genetic population and fix the trait of interest. However, this type of population presents more advantages in a diploid crop (such as maize) than in a polyploid crop (such as wheat) as only one recombination occurs during population construction. Recombinant inbred lines (RILs) are therefore more often used in wheat genetic research. They can be produced by various inbreeding procedures such as single seed descent (SSD) and single spike selection methods (Fehr 1987). Several quantitative traits loci have been mapped at CIMMYT using this type of population (Basnet et al. 2013; Lan et al. 2014; Rosewarne et al. 2012). As the RILs are genetically fixed, they can be easily phenotyped for multiple environments and years in replicated trials. Moreover, the RILs provide the better choice for genetic studies as they are developed by several generations of selfing which allows multiple meiotic divisions and more numbers of accumulated recombination events.

Molecular markers

DNA based molecular marker technologies have come a long way since the development of the first generation markers, such as restriction frangment length polymorphism (RFLPs) in the 1980s. In spite of many developed marker types, major technologies can broadly classified into the following three groups: 1) hybridization-based markers, such as RFLPs (Devos et al. 1992) and diversity arrays technology (DArT, Akbari et al. 2006), 2) PCR-based markers e.g., simple sequence repeats (SSR, Roeder et al. 1998), sequence tagged sited (STS), or single nucleotide polymorphisms (SNPs), and 3) sequencing-based markers such as genotyping-by-sequencing (GBS, Elshire et al. 2011). Some of the markers use combination of different techniques, e.g., amplified fragment length polymorphisms (AFLPs). RFLPs are mostly co-dominant and restricted to regions with low-copy sequences. The rust resistance genes

Lr10, Lr23, Lr27, Lr31, Lr34 were firstly mapped by this marker system (Nelson et al. 1997). Compared RFLPs, AFLPs markers give higher reproducibility and resolution at the whole genome level; however, the procedure of AFLP analysis is complex and costly (Mueller et al. 1999). The SSR marker system became the preferred system due to its co-dominance, accuracy, high repeatability, high levels of polymorphism, chromosome specificity, and ease of manipulation (Röder et al. 1998). More recently, high-throughput technologies, DArT, SNP and GBS, have become the major genotyping platforms. DArT was developed as a hybridisation-based system capable of generating whole-genome fingerprints by scoring presence versus absence of DNA fragments in genomic representations generated from samples of genomic DNA. However, all DArT markers are dominant. The chromosome information of this system is not very clear either and markers tend to form location clusters, although some markers have been located on the physical map (Wilkinson et al. 2012; http://www.cerealsdb.uk.net/CerealsDB/Documents/FORM_DArT_1A. php). DArT markers have often been combined with SSR in mapping QTL for traits such as biotic and abiotic stress tolerance in wheat. SNPs represent the smallest possible DNA polymorphism. Their abundance in any genome makes them ideally suited for the construction of high-resolution genetic maps, investigations of population evolutionary history and the discovery of marker-trait associations (Aranzana et al. 2005; Zhao et al. 2007). This capacity also helps in directly interrogating sequence variation and reducing genotyping errors compared to assays based on size discrimination (Li et al. 2014). However, the chromosome information of high-density SNP markers (1537 up to 830K SNP chips) is still somewhat limited in its use in wheat genetic analysis. Due to the accelerated development of sequencing technologies, the GBS approach has recently been applied to the construction of genetic maps of crops (Elshire et al. 2011; Poland et al. 2012). This allows direct analysis of genetic variation and reduction of the effect of ascertainment bias caused by the SNP discovery process. However, the lack of chromosome information has so far limited SNP application in genetic research. Thus, combining several well-established marker systems with a high-throughput method is currently the best option for genetic studies in wheat.

Traditional genetic analysis

Before molecular mapping of genes is carried out, it is important to understand the mode of inheritance of the trait(s) under study. For a monogenic and oligogenic trait such as rust resistance in wheat, gene action and gene quantity can be analyzed by progeny testing and Mendelian population segregation analysis. Singh and Rajaram (1992) demonstrated that the number of genes segregating in a population can be estimated using Mendelian segregation ratios for such quantitative traits as disease severities. The observed frequencies for the three homozygous resistant parental, homozygous susceptible parental and segregating type categories were tested against the expected frequencies for different numbers of additive genes using Chi-square (χ^2) analysis. In addition, the minimum number of genes in a RIL population in F_5 could also be estimated by using the quantitative approach described by the formula of Wright (1968):

$$n = (GR)^2/4.57(\sigma^2 g)$$

where n = the minimum number of genes, GR (genotypic range) = phenotype range × h (narrow-sense heritability, $h = \sigma^2 g / \sigma^2 g + \sigma^2 e$), $\sigma^2 g$ = genetic variance of F_5 RILs in the present population and 4.57 a correction factor for inbreeding at F_5.

QTL mapping

QTL mapping establishes linkages between marker loci and trait(s). There are two major types of QTL mapping methods based on the classification of individuals: marker-based analysis (MBA) and trait-based analysis (TBA) (Xu 2010). The MBA method locates chromosomal regions or QTL based on their linkage relationships to Mendelian marker loci (Thoday 1961). This method of linkage involves testing for phenotypic differences among marker genotypes (Soller and Beckmann 1990). Several wheat rust loci mapped by this method (Ren et al. 2012; Yang et al. 2013) are schematically demonstrated in Figure 1.

The TBA method examines marker allele frequencies in lines originating from a segregating population but selected for specific phenotypes (Stuber et al. 1980, 1982). Although variation in quantitative traits is continuous in a population, phenotype extremes can still be distinguished if the intermediate phenotypes are excluded. By selecting the extremes

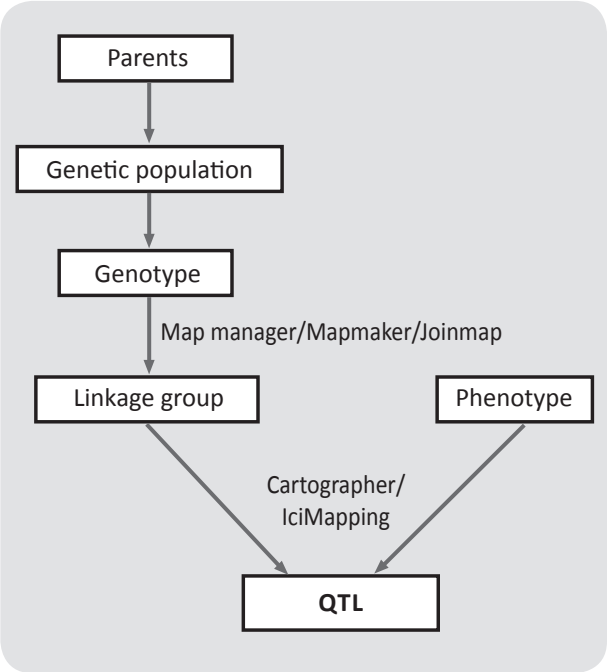


Figure 1. Strategy of QTL mapping for MBA.

of a trait phenotype, the difference of related allele frequencies will be maximized (Sun et al. 2010) and are thus considered to have an effect on the trait (Xu 2010). This method is good to be used in fine mapping single genes/QTL (Figure 2).

QTL mapping methods and related software

The QTL mapping methods can have significant impact on accuracy and efficiency of QTL discoveries. In the early 1990s, simple interval mapping (SIM) was wildly used in genetic analysis. This method detects the relationship between traits and markers by comparing the significant difference between phenotype means divided by single marker genotypes (Soller and Beckmann, 1990). However, SIM cannot determine the linkage between detected QTL and has low efficiency and easily provides false positive results. Composite interval mapping (CIM) was introduced by Zeng (1993) as improvement over SIM. It combines both multiple linear regression and SIM together to detect the QTL between several markers and different chromosome regions simultaneously. This method improved the sensitivity and accuracy of the QTL result by controlling genetic background effects. This method has been widely used in QTL mapping in particular because of its ability to do multiple traits analysis (Jiang and Zeng, 1995).

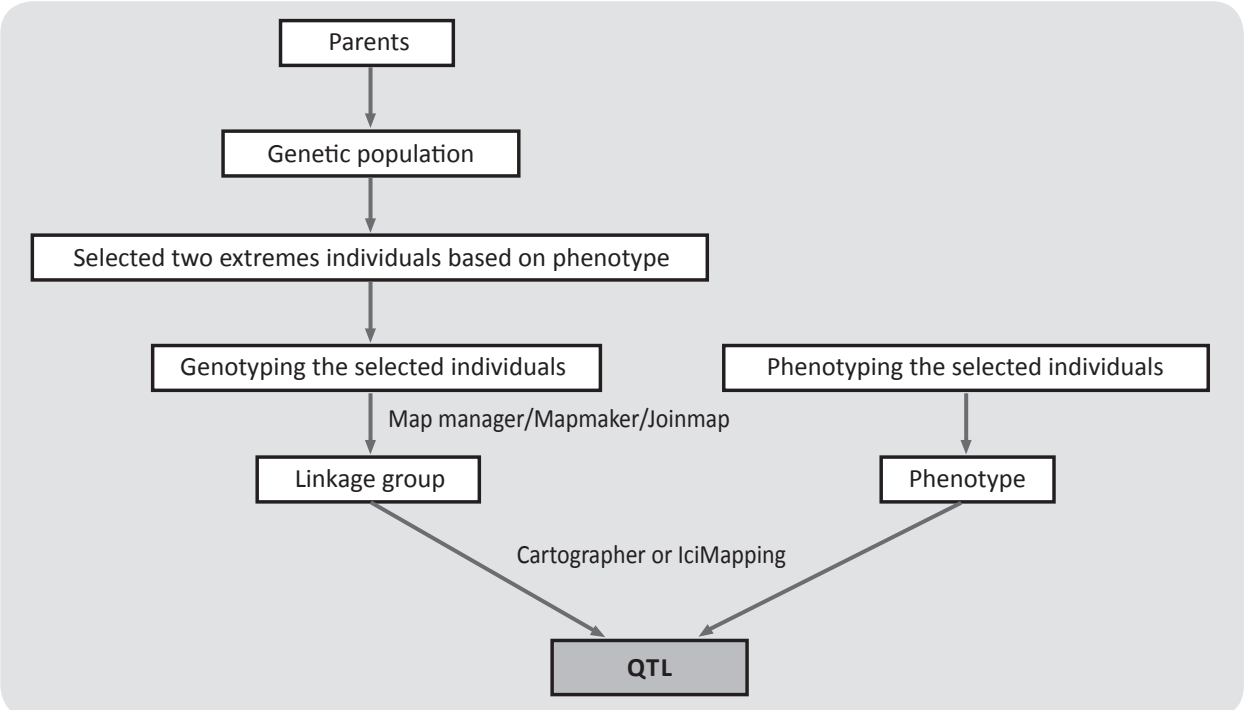


Figure 2. Strategy of QTL mapping for TBA.

CIM based on mixed model can detect additive, dominance and G×E interaction effects and also can detect QTL in multiple environments simultaneously, which markedly improves the efficiency and accuracy of the mapping result (Zhu et al. 1998). Recently, inclusive composite interval mapping (ICIM) has been reported by Li et al. (2007). The differences between this method and those previously mentioned are 1) it uses a selective analysis based on a stepwise regression model; 2) it adjusts the phenotype based on the selected model and 3) it carries out interval mapping based on a multiple regression equation. In recent years, ICIM has been a popular method in QTL mapping studies because of its very high efficiency in terms of processing time and higher accuracy of results, including less false positive results, better estimates of QTL positions and effects (Basnet et al. 2014, Lan et al. 2014, Rosewarne et al. 2012).

With advances in newer computer technologies, genetic-mapping theories and techniques, sophisticated software for genetic mapping and QTL analyses have also evolved rapidly. Some very popular non-commercial, genetic-mapping software includes: Mapmaker/QTL 1.1; Map Manager QTX (Manly et al. 2001); Windows QTL Cartographer (Wang et al. 2012); QTL IciMapping v3.3 (Li et al. 2007).

To conduct CIMMYT rust genetic research, we use both QTL Cartographer (CIM) and IciMapping (ICIM) for QTL analysis as well as for cross validation. The results from both methods are compared and confirmed with a logarithm of odds (LOD) threshold based on 1,000 permutations. Based on our QTL mapping experience, especially for disease resistance traits, there are several factors affecting the accuracy of QTL results:

- LOD threshold: although the general consensus of LOD 2.0-3.0 is used as a threshold for QTL detection, it is very important to understand the type of phenotypic data and using permutation (≥1,000). For small effect QTL, a higher threshold may reduce the power of QTL detection, resulting in false negatives, whereas lower thresholds can result in many false positives
- Mapping population size: bigger population size is required to detect multiple minor-effects QTL. In addition, the genetic population size will also affect the LOD threshold to detect the QTL, for example, the number of detected QTL will be significantly less if higher LOD thresholds are used in smaller genetic populations.
- QTL/gene effect: larger QTL or genes, have often been observed to overshadow the effect of smaller QTL in the population, which results in under-detection of true QTL. In this instance, the genetic population can be subdivided by excluding the large-effect QTL/gene in order to allow the small QTL to be detected (Basnet et al. 2013).
- Number and distribution of markers in the genome: genome coverage by markers is very important to detect larger numbers of QTL present in the population. The genetic distance between markers will also affect the accuracy of QTL result. A false QTL might be detected if the genetic distance between two markers is very far – for example, more than 30 cM based on our genetic population analysis. This QTL needs to be confirmed by the single marker and SIM analysis.

Thus, the accuracy of QTL will be affected by the size of the population, LOD score, genetic distance between markers and QTL mapping method. The results need to be further confirmed by single marker analysis for traits with low heritability and QTL with small effect.

The importance of phenotype for QTL mapping

The expression of minor genes can be affected by several factors. In the case of wheat rust, the first factor is environment, including temperature, light and moisture. It is very important to evaluate the population in different environments to identify stable QTLs. For example, one leaf rust APR QTL on 3BS had been identified in the RIL population from the cross Avocet-*YrA* × Francolin#1, it explained 30.4% of phenotype variance at the CIMMYT's Norman E. Borlaug Experimental Station (CENEB) wheat breeding station in Obregon in the 2008-09 season, but it was not identified in Obregon in the 2009-10 season (Lan et al. 2014). The second factor is the inoculation pressure. The method for inoculation and the inoculum amount had certain effects on expression of minor resistance genes. The reaction was significantly different when the inoculums were inoculated directly or indirectly (Zhang et al. 2009). The third factor is genetic background. The effects of the same APR QTL were different in various

segregating populations because interaction between a range of genes were present. For example, the leaf rust severity score was 5% when the single adult plant resistance gene *Lr34* was introgressed in the Avocet background, whereas it was 20% in the Lal Bahadur background. The final factor effecting the expression of minor genes is pathotypes. Non-race specific genes often confer consistent resistance to most *Pt* pathotypes, which means that the resistance gene presented broadly spectrum resistance to different pathotypes (McDonald and Linde 2002). Pathotypes have qualitative effects on the gene expression if a major gene is present in the population.

Mapping of rust resistance genes and QTL at CIMMYT

The rust group in the CIMMYT Global Wheat Program (GWP) started to map the rust resistance gene and QTL at the beginning of this century. With the cooperation of Australian scientist Evans Lagudah and his group at the Commonwealth Scientific and Industrial Research Organisation (CSIRO), we have mapped and cloned pleotropic adult plant resistance (PAPR) genes *Lr34/Yr18/Sr57/Pm38*, which confer durable and race non-specific resistance against the four bio-trophic fungal pathogens causing leaf rust, stripe rust, stem rust and powdery mildew (Krattinger et al. 2009). Its resistance has maintained effective and durable for at least 100 years. In addition, two PAPR genes, *Lr46/Yr29/Sr58/Pm39* and *Lr67/Yr46/Sr55/Pm46*, have also been mapped in CIMMYT germplasm 'Pavon 76' and 'Thatcher' derived line 'RL6077', respectively. These three genes have been widely used in the CIMMYT wheat breeding program as the partial basis for resistance against three rusts. So far, more than ten rust resistant genes have been officially designated by our group, such as *Lr61, Lr67/Yr46/Pm46, Lr68, Lr72, Yr54, Yr60, Sr2/Yr30, Sr55, Sr57, Sr58* as well as eight temporarily designated resistance genes, viz. *YrF, YrSuj, YrKK, SrND643, SrNini, SrSHA7/SrHaril, SrBlouk*. In addition, QTL mapping for adult plant resistance to leaf and stripe rust was performed in Avocet/Quaiu#3, Avocet/Francolin#1, Avocet/Pavon 76, Avocet/Pastor, Avocet/Atila, Avocet/Chapio, Avocet/Sujata, Avocet/Kenya kongoni and Avocet/Kundan populations. It was also performed to determine resistance to stem rust in Cacuke/Kenya Kudu, Cacuke/Kenya Nyangumi, Cacuke/Kingbird, and Cacuke/Kenya Swara populations. Altogether, more than 20 QTL have been mapped among these populations to determine resistance against the three rusts. Refer to chapter 5 "Marker-assisted selection for rust resistance in wheat" to learn more about how CIMMYT uses these genes in breeding

Cloning wheat genes

In order to clone wheat genes, accurate phenotyping and fine mapping is essential. SSR markers have often been used for primary molecular mapping, whereas SNP markers and comparative genomics approaches are used for fine mapping (Zhang et al. 2013). As a final step Fu et al. (2009) used chromosome walking combined with bacterial artificial chromosome (BAC) library screening to clone the resistance gene *Yr36* (Fu et al. 2009). So far, more than 10 wheat disease resistance genes have been cloned by this method, such as *Lr1, Lr10, Lr21, Lr34, Yr36, Sr33, Sr35, Pm3, Pm8* and *Pm21* (Cao et al. 2011; Cloutier et al. 2007; Feuillet et al. 2003; Fu et al. 2009; Huang et al. 2003; Hurni et al. 2013; Krattinger et al. 2009; Periyannan et al. 2013; Saintenac et al. 2013; Yahiaoui et al. 2004).

However, common wheat has a very large genome, and the genomic sequences are not yet assembled as thoroughly as those in rice and maize (Liu et al. 2012, Choulet et al 2014). Therefore, it is very difficult to clone genes by map-based cloning in common wheat. Alternatively, comparative genomics provides an efficient approach for the isolation of wheat genes based on orthologs descended from a common ancestor, which have often conserved functions and are expected to produce similar phenotypes across species (Devos 2005). The genomes of rice, maize and Brachypodium grasses have been sequenced and provide powerful tools for gene discovery in wheat (Matsumoto et al. 2005; Schnable et al. 2009; Vogel et al. 2010). This technology for in silico cloning was widely employed for the identification of interesting genes in wheat (Gill and Sanseau 2000; He et al. 2007, 2008, 2009; Su et al. 2011; Ma et al. 2012). Based on the large expressed sequence tags (EST) database (wheat.pw.usda.gov/wEST/), putative wheat gene sequences were obtained by aligning and jointing the orthologous genes with the same function in the related species. For example, the *Psy1* gene (GenBank accession U32636) has been cloned based on the cDNA sequence of maize, all wheat ESTs sharing high similarity with the reference gene were blasted and subjected to contig assembly (He et al. 2008). Recently, selective genotyping was used

in fine mapping and cloning of the *GCP-B1* gene in wheat (Trick et al. 2012). This cost-efficient method was used to genotype individuals from the extremes of the segregating population instead of the entire population. In the future, sequencing techniques combined with candidate gene association analysis and high-density wheat SNP arrays will be used for fine mapping and cloning genes in wheat.

References

Akbari, M., P. Wenzl, C. Vanessa, J. Carling, L. Xia, S. Yang, G. Uszynski, V. Mohler, A. Lehmensiek, H. Kuchel, M.J. Hayden, N. Howes, P. Sharp, B. Rathmell, P. Vaughan, E. Huttner and A. Kilian. 2006. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theoretical and Applied Genetics 113:1409–1420

Aranzana, M.J., S. Kim, K. Zhao, E. Bakker, M. Horton, K. Jakob, C. Lister, J. Molitor, C. Shindo, C. Tang, C. Toomajian, B. Traw, H. Zheng, J. Bergelson, C. Dean, P. Marjoram and M. Nordborg. 2005. Genomewide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. PLoS Genetics 1: e60

Bal, U. and K. Abak. 2007. Haploidy in tomato (*Lycopersicon esculenttum* Mill.): a critical review. Euphytica 158: 1-9

Basnet, B.R., R. P. Singh, A.M.H. Ibrahim, S.A. Herrera-Foessel, J. Huerta-Espino, C.X. Lan and J.C. Rudd. 2013. Characterization of *Yr54* and other genes associated with adult plant resistance to yellow rust and leaf rust in common wheat Quaiu 3. Molecular Breeding 33: (2): 385-399

Basnet, B.R., R.P. Singh, A.M.H. Ibrahim, S.A. Herrera-Foessel, J. Huerta-Espino, C.X. Lan, J.C. and Rudd. 2014. Characterization of *Yr54* and other genes associated with adult plant resistance to yellow rust and leaf rust in common wheat Quaiu 3. Molecular Breeding 33: 385-399

Cao, A., L. Xing, X. Wang, X. Yang, W. Wang, Y. Sun, C. Qian, J. Ni, Y. Chen, D. Liu, X. Wang and P. Chen. 2011. Serine/threonine kinase gene Stpk-V, a key member of powdery mildew resistance gene *Pm21*, confers powdery mildew resistance in wheat. Proceedings of the National Academy of Sciences United States of America 108: 7727-7732

Cloutier S., B.D. McCallum, C. Loutre, T.W. Banks, T. Wicker, C. Feuillet, B. Keller and M.C. Jordan. 2007. Leaf rust resistance gene Lr1, isolated from bread wheat (*Triticum aestivum* L.) is a member of the large psr567 gene family. Plant Molecular Biology 65: 93-106

Devos, K.M., M.D. Atkinson, C.N. Chinoy and M.D. Gale. 1992. RFLP-based genetic map of the homoeologous group 2 chromosomes of wheat and rye. Theoretical and Applied Genetics 72:820-828

Devos, K.M. 2005. Updating the 'crop circle'. Current Opinion in Plant Biology 8: 155-162

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler and S.E. Mitchel. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379

Fehr, W.R. 1987. Principles of Cultivar Development., Vol. 1. Theory and Techniques. London: Macmillan Publishing Company

Ferrie, A.M.R. 2007 Doubled haploid production in nutraceutical species: a review. Euphytica 158: 347-357

Feuillet, C., S. Travella, N. Stein, L. Albar, A. Nublat and B. Keller. 2003. Map-based isolation of the leaf rust disease resistance gene Lr10 from the hexaploid wheat (*Triticum aestivum* L.) genome. Proceedings of the National Academy of Sciences United States of America. 100: 15253-15258

Forster, B.P., E. Herberle-Bors, K.J. Kasha and A. Touraev. 2007. The resurgence of haploids in higher plants. Trends in Plant Science. 12: 368-375

Fu, D.L., C. Uauy, A. Distelfeld, A. Blechl, L. Epstein, X.M. Chen, H. Sela, T. Fahima, J. Dubcovsky. 2009. A kinase-start gene confers temperature dependent resistance to wheat stripe rust. Science 323: 1357-1360

Gill, R.W. and P. Sanseau. 2000. Rapid in silico cloning of genes using expressed sequence tags (ESTs). Biology Annual Review 5: 25-44

He, X.Y., Z.H. He, W. Ma, R. Appels and X.C. Xia. 2009. Allelic variants of phytoene synthase 1 (*Psy1*) genes in Chinese and CIMMYT wheat cultivars and development of functional markers for flour color. Molecular Breeding 23: 553-563

He, X.Y., Z.H. He, L.P. Zhang, D.J. Sun, C.F. Morris, E.P. Fuerst and X.C. Xia. 2007. Allelic variation of polyphenol oxidase (PPO) genes located on chromosomes 2A and 2D and development of functional markers for the PPO genes in common wheat. Theoretical and Applied Genetics 115: 47-58

He, X.Y., Y.L. Zhang, Z.H. He, X.P. Wu, Y.G. Xiao, C.X. Ma and X.C. Xia. 2008. Characterization of phytoene synthase 1 gene (*Psy1*) located on common wheat chromosome 7A and development of a functional marker. Theoretical and Applied Genetics 116: 213-221

Huang, L., S.A. Brooks, W. Li, J.P. Fellers, H.N. Trick and B.S. Gill. 2003. Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics 164: 655-664

Hurni, S., S. Brunner, G. Buchmann, G. Herren, T. Jordan, P. Krukowski, T. Wicker, N. Yahiaoui, R. Mago and B. Keller. 2013. Rye *Pm8* and wheat *Pm3* are orthologous genes and show evolutionary conservation of resistance function against powdery mildew. Plant Journal 76: 957-969

Krattinger S.G., E.S. Lagudah, W. Spielmeyer, R.P. Singh, J. Huerta-Espino, H. McFadden, E. Bossolini, L.L. Selter and B. Keller. 2009. A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science 323: 1360-1363

Jiang, C.J. and Z.B. Zeng. 1995. Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140: 1111-1127

Lan, C.X., G.M. Rosewarne, R.P. Singh, S. A. Herrera-Foessel, J. Huerta-Espino, B.R. Basnet, Y.L. Zhang and E.N. Yang. 2014. Genetic analysis of resistance to leaf rust and stripe rust in wheat cultivar 1 Francolin#1. Molecular Breeding 34:789-803

Li, H.H., G.Y. Ye and J.K. Wang. 2007. A modified algorithm for the improvement of composite interval mapping. Genetics 175: 361-374

Li, Z.F., C.X. Lan, Z.H. He, R.P. Singh, G.M. Rosewarne, X.M. Chen and X.C. Xia. 2014. Overview and application of QTL for adult plant resistance to leaf rust and powdery mildew in wheat. Crop Science 54: 1970-1925

Liu, Y.N., Z.H. He, R. Appels and X.C. Xia. 2012. Functional markers in wheat: current status and future prospects. Theoretical and Applied Genetics 125: 1-10

Ma, D.Y., J. Yan, Z.H. He, L. Wu and X.C. Xia. 2012. Characterization of a cell wall invertase gene *TaCwi-A1* on common wheat chromosome 2A and development of functional markers. Molecular Breeding 29: 43-52

Manly, K.F., R.H. Cudmore Jr and J.M. Meer. 2001. Map Manager QTX, cross-platform software for genetic mapping. Mammalian Genome 12: 930-932

Matsumoto, T., J.Z. Wu, H. Kanamori, Y. Katayose and M. Fujisawa. 2005. The map-based sequence of the rice genome. Nature 436: 793-800

McDonald, B.A. and C. Linde. 2002. Pathogen population genetics, evolutionary potential, and durable resistance. Annual Review of Phytopathology 40: 349-379

Mueller, U.G. and L.L. Wolfenbarger. 1999. AFLP genotyping and fingerprinting. Trends in Ecology and Evolution 14: 389-394

Nelson, J.C., R.P. Singh, J.E. Autrique and M.E. Sorrells. 1997. Mapping genes conferring and suppressing leaf rust resistance in wheat. Crop Science 37: 1928-1935

Periyannan, S., J. Moore, M. Ayliffe, U. Bansal, X. Wang, L. Huang, K. Deal, M. Luo, X. Kong, H. Bariana, R. Mago, R. McIntosh, P. Dodds, J. Dvorak and E. Lagudah. 2013. The gene *Sr33*, an ortholog of barley *Mla* genes, encodes resistance to wheat stem rust race Ug99. Science 341:786-768

Poland, J.A., P.J. Brown, M.E. Sorrells and J.L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253

Ren, Y., Z.F. Li, Z.H. He, L. Wu, B. Bai, C.X. Lan, C.F. Wang, G. Zhou, H.Z. Zhu and X.C. Xia. 2012. QTL mapping of adult-plant resistances to stripe rust and leaf rust in a Chinese wheat cultivar Bainong 64. Theoretical and Applied Genetics 125:1253-1262

Röder, M.S., V. Korzun, K. Wendehake, J. Plaschke, M.H. Tixier, P. Leroy and M.W. Ganal. 1998. A microsatellite map of wheat. Genetics 149:2007-2023

Rosewarne, G.M., R.P. Singh, J. Huerta-Espino, S.A. Herrera-Foessel, K.L. Forrest, M.J. Hayden and G.J. Rebetzke. 2012. Analysis of leaf and stripe rust severities reveals pathotype changes and multiple minor QTLs associated with resistance in Avocet × Pastor wheat population. Theoretical and Applied Genetics 124: 1283-1294

Saintenac, C., W. Zhang, A. Salcedo, M.N. Rouse, H.N. Trick, E. Akhunov and J. Dubcovsky. 2013. Identification of wheat gene *Sr35* that confers resistance to Ug99 stem rust race group. Science 341:783-786

Schnable, P.S., D. Ware, R.S. Fulton, J.C. Stein and F.S. Wei. 2009. The *B73* maize genome: complexity, diversity, and dynamics. Science 326:1112-1115

Singh, R.P. and S. Rajaram. 1992 Genetics of adult-plant resistance to leaf rust in ¨Frontana¨ and three CIMMYT wheats. Genome 35: 24-31

Soller, M. and J.S. Beckmann. 1990. Marker-based mapping of quantitative trait loci using replicated progenies. Theoretical and Applied Genetics 80: 205-208

Stuber, C.W. and R.H. Moll. 1982. Improvement of yield and ear number resulting from selection at allozyme loci in a maize population. Crop Science 22:737-740

Stuber, C.W., R.H. Moll, M.M. Goodman, H.E. Schaffer and B.S. Weir. 1980. Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays*). Genetics 95:225-336

Sun, Y.P., J.K. Wang, J.H. Crouch and Y.B. Xu. 2010. Efficiency of selective genotyping for genetic analysis of complex traits and potential applications in crop improvement. Molecular Breeding 26:493-511

Sun, Y.W., Z.H. He, W.J. Ma and X.C. Xia. 2011. Alternative splicing in the coding region of Ppo-A1 directly influences the polyphenol oxidase activity in common wheat (*Triticum aestivum* L.). Function Inter Genome 11:85-93

Thoday, J.M. 1961. Location of polygenes. Nature 191:368-370

Trick, M., N.M. Adamski, S.G. Mugford, C.C. Jiang, M. Freber and C. Uauy. 2012. Combining SNP discovery from next-generation sequencing data with bulk segregant analysis (BSA) to fine-map genes in polyploidy wheat. BMC Plant Biology 12:14

Vogel, J.P., D.F. Garvin and T.C. Mockler. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763-768

Wang, D.L., J. Zhu, Z.K. Li and A.H. Paterson. 1999. Mapping QTLs with epistatic effects and QTL×environment interactions by mixed linear model approaches. Theoretical and Applied Genetics 99:1255-1264.

Wang, S., C.J. Basten and Z.B. Zeng. 2012. Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)

Wilkinson, P.A., M.O. Winfield, G.L.A. Barker, A.M. Allen, A. Burridge, J.A. Coghill, A. Burridge and K.J. Edwards. 2012. CerealsDB 2.0: an integrated resource for plant breeders and scientists. BMC Plant Biology 13:219

Wright, S. 1968. Evolution and the genetics of populations. Vol. 1. Genetic and biometric foundations. The University of Chicago Press, Chicago

Wu, W.R., X.M. Li and H.R. Lu. 1996. Quantitative trait loci analysis with composite interval mapping based on least squares estimation. Fujang Agricultural University Journal 25:394-399 (in Chinese).

Xu, Y.B. 2010. Molecular plant breeding. CABI Head Office, Nosworthy way Wallingford Oxfordshird OX10 8DE, UK, ISBN 978-1-84593-392-0

Yahiaoui, N., P. Srichumpa, R. Dudler and B. Keller. 2004. Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. Plant Journal 37:528-538

Yang, E.N., G.M. Rosewarne, S.A. Herrera-Foessel, J. Huerta-Espino, Z.X. Tang, C.F. Sun, Z.L. Ren and R.P. Singh. 2013. QTL analysis of the spring wheat ''Chapio'' identifies stable stripe rust resistance despite inter-continental genotype×environment interactions. Theoretical and Applied Genetics 126:1721-1732

Zeng, Z.B. 1993. Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait Proceedings of the National Academy of Sciences of USA 90: 10972-10976

Zhang, L., Z. Li, M. Lillemo, X.C. Xia, D. Liu, W. Yang, J. Luo and H. Wang. 2009. QTL mapping for adult-plant resistance to leaf rust in CIMMYT wheat cultivar Saar. Science Agriculture Sin 42:388-397

Zhang, X., D. Han, Q. Zeng, Y. Duan, F. Yuan, J. Shi, Q. Wang, J. Wu, L. Huang and Z.S. Kang. 2013. Fine mapping of wheat stripe rust resistance gene *Yr26* based on collinearity of wheat with Brachypodium distachyon and rice. PLoS One 8:e57885

Zhao, K., M.J. Aranzana, S. Kim, C. Lister, C. Shindo, C. Tang, C. Toomajian, H. Zheng, C. Dean, P. Marjoram and M. Nordborg. 2007. An Arabidopsis example of association mapping in structured samples. PLoS Genetics 3:e4

Zhu, J. 1998. Mixed linear model in mapping complex quanatitative traits. National workshop of plant breeding. 11-20

Chapter 2:
Association Mapping for Dissecting Complex Traits

Deepmala Seghal, Sivakumar Sukumaran, Susanne Dreisigacker
CIMMYT Int., Apdo. Postal 6-641,06600 Mexico, DF, Mexico

Introduction

The objective of genetic mapping is to identify simply inherited markers in close proximity to genetic factors affecting quantitative traits (Quantitative trait loci, or QTL). Genetic mapping can be done in two ways: (1) using experimental populations (bi-parental mapping populations), called "QTL-mapping" or "gene tagging," (see Chapter 1); and (2) using diverse lines from natural populations or germplasm collections, called "genome-wide association studies," "association mapping (AM)" or "linkage disequilibrium (LD) mapping". The traditional QTL mapping approach suffers from a number of limitations. First, allelic variation in each cross is restricted because typically only two parents are used to create a QTL mapping population. Second, since often segregating or double haploid populations are used, the number of recombination events per chromosome is small. Third, a typical QTL identified from a cross consisting of a few hundred offspring can span anywhere between a few to tens of centiMorgan encompassing several megabases. Such large genome regions contain, typically, hundreds if not thousands of genes, making the process of identifying the causal gene in a QTL region a tedious and quite time-consuming task through map-based cloning (Price 2006).

Association mapping has emerged as a more efficient way of determining the genetic basis of complex traits where a large population is surveyed to determine marker-trait associations using linkage disequilibrium. This approach has many major advantages over conventional QTL mapping. First, a larger and more representative genepool can be surveyed. Second, it bypasses the expense and time of developing mapping populations and enables the simultaneous mapping of many traits in one set of genotypes. Third, a much finer mapping resolution can be achieved, resulting in small confidence intervals of the detected loci compared to classical mapping, where the identified loci need to be fine-mapped. Finally, it has the potential not only to identify and map QTLs but also to identify the causal polymorphism within a gene that is responsible for the difference in two alternative phenotypes (Yu et al. 2013). However, AM is prone to the identification of false positives, especially if the experimental design is not rigorously controlled. For example, population structure has long been known to induce many false positives and accounting for population structure has become one of the main issues when implementing AM in plants (Breseghello and Sorrells 2006). Also, with an increasing number of genetic markers used, the problem of separating true from false positive marker associations becomes relevant and highlights the need for independent validation of identified associations. With these caveats in mind, AM nevertheless shows great promise for helping us understand the genetic basis of polygenic traits of agronomic importance.

The performance of AM includes the following general steps: (1) selection of a group of individuals from a natural population or germplasm collection with wide coverage of genetic diversity; (2) measuring the phenotypic characteristics (yield, quality, tolerance, resistance etc.) in the population, preferably, in different environments and multiple replication/trial design; (3) genotyping the mapping population individuals with molecular markers; (4) quantification of the extent of LD using molecular marker data; (5) assessment of the population structure (the level of genetic differentiation among groups within a sampled population individuals) and kinship (coefficient of relatedness between pairs of each individuals within a sample); and (6) based on information gained through quantification of LD and population structure, correlation of phenotypic and genotypic data with the application of an appropriate statistical approach that reveals "marker tags" positioned within close proximity of targeted trait of interest. In this chapter we have provided an overview of important considerations while selecting an AM panel, issues related to genotyping and population structure, and some important programs and tips to analyze the data data.

Selection of AM panels

Various population types can be used for AM; gene bank collections, elite breeding materials and/or specialized populations (e.g., nested association mapping, or NAM, and multiparent advanced generation inter-cross, or MAGIC, populations). In the case of gene bank material, core collections are expected to represent most of the genetic variability with a manageable number of accessions, and thus are suitable for AM studies. In addition, the process of selecting a minimum sample size with maximum variation has a normalizing effect that is expected to reduce population structure and decrease LD thus creating a situation favorable for AM. On the other hand, large numbers of rare alleles might be captured in a core collection whose effects are difficult to identify with AM. Core collections are useful for mapping qualitative traits, such as disease resistance or quality characteristics. Their broad genetic variability makes them often unsuitable for analysis of quantitative traits because accessions are usually unadapted to growing conditions, resulting in poor precision of trait measurement.

In plant breeding programs, a large body of phenotypic data is accumulated for elite lines from replicated field experiments over locations and years, thereby saving time on developing a panel. The use of those data for AM requires statistical models accounting for covariances introduced both by experimental design (years, locations, replicates) and polygenic effects. Moreover, those data are often unbalanced because new lines are included in field trials each year, while other lines are discarded. Population structure and higher LD can be prominent in elite material because it is common for closely related lines to be admitted to advanced trials. However, if pedigrees are known, the relationships among the lines can be determined and used to control for polygenic effects. Although AM in elite lines may not offer much improved resolution compared with QTL analysis in bi-parental mapping populations, there are at least two important advantages: a substantially higher level of polymorphism and detection of favorable alleles directly in the target population. At CIMMYT, panels of elite lines that will form international wheat screening nurseries and yield trails are being used for AM for yield and yield components under drought and heat stress, quality traits and resistance to various diseases. Elite lines might be more desirable materials for mapping low heritability traits, as the material is genetically more stable and are well adapted to normal growing conditions.

To increase the power and mapping resolution of marker-trait associations, some specialized populations have been constructed utilizing a joint strength of QTL mapping and AM. For example, NAM populations and MAGIC populations have been developed in wheat and other crops (Cavanagh et al 2008; Kover et al. 2009; McMullen et al. 2009; Huang et al. 2012). NAM populations are developed by crossing a set of diverse lines (up to 25) to one reference line. F_1's of each cross are then selfed to develop recombinant inbred lines for each population. MAGIC populations are created by several generation of intercrossing among multiple founder lines, for example four or eight lines. Multiple founders similar to a NAM population capture more allelic diversity than bi-parental mapping populations whereas the multiple cycles of intercrossing give greater opportunity of recombination and hence greater precision of QTL mapping. However, it should be kept in mind that generatin g such specialized populations requires and a lot of effort, time and investment.

Genotyping

Choice of markers for genome-wide study

Amplified fragment length polymorphism (AFLP), diversity array technology (DArT), simple sequence repeats (SSR) and single nucleotide polymorphisms (SNP) are the most commonly used markers in genome wide association studies (Honsdorf et al. 2010; Adhikari et al. 2012; Gupta et al. 2014; Upadhyaya et al. 2013; Crossa et al. 2007). AFLP and DArT markers are easily obtained in almost any organism, even for those lacking any kind of genomics data. Similarly, the highly polymorphic, multiallelic and co-dominant nature of SSR markers, combined with the availability of semi-automated detection methods, have made them highly suitable for AM in many crops including wheat (Peng et al. 2009; Yao et al. 2009; Liu et al. 2010; Zhang et al. 2011; Reif et al. 2011; Dodig et al. 2012). However, AFLP and DArT markers are exclusively dominant (i.e. heterozygous genotypes cannot be distinguished from homozygous genotypes) which introduces a number of problems when estimating population structure or using them directly in mapping (Ritland 2005). Moreover, AFLP, DArT and SSR markers are today seen as rather time-

consuming and cost-intensive technologies, and the overall number of markers (up to 2000) that can be revealed by all three technologies is low.

The development of next-generation sequencing (NGS) technologies has allowed for unprecedented genotyping capabilities, even for large complex polyploid genomes such as wheat (Poland et al. 2012). The current NGS technologies are capable of analyzing anywhere from hundreds of thousands to tens of millions of DNA molecules in parallel compared with hundreds at a time which is the maximum throughput of most traditional sequencing instruments. Next generation sequencing technologies allow the rapid identification of a large numbers of genetic markers, mainly SNPs (Imelfor et al. 2009). SNPs are biallelic and the information content per marker is therefore much lower than in SSR markers. This, however, is compensated for by a higher genome density and a wide distribution across the genome. SNP markers are therefore rapidly becoming the marker of choice for most AM studies. SNP markers are also amenable to high-throughput genotyping in multiplex or microarray format. SNP marker platforms have been established in wheat (Akhunov et al. 2009; Wang et al. 2014) and genotyping of wheat association panels with up to 90,000 SNP markers is now achievable in wheat (Wang et al. 2014). The potential of SNP markers in determining marker-trait associations has just begun in wheat and initial studies have shown good results (Lopez et al. 2015).

An important consideration that has been receiving an increasing interest in using SNPs for AM considers how the selection of SNPs can potentially bias the results. For instance, SNP discovery panels are often small, which means that low frequency mutations often go undetected and SNPs occurring at intermediate frequencies are frequently sampled. Therefore, the ascertainment bias introduced in the SNP selection process has important consequences for the inferences drawn; for AM the most detrimental effect is an over-sampling of mutations at intermediate frequencies which results in lower levels of LD than if SNPs were selected completely at random. Hence, including as many diverse genotypes (possibly from diverse origins, landraces, etc.) as possible in the SNP discovery panel would be an ideal way to accumulate SNPs with both low and intermediate frequencies. Another important factor to be aware of in AM is the genotyping error rate. While state-of-the-art SNP scoring methods are usually quite robust, the rate of genotyping errors especially in a highly polyploid crop such as wheat can vary substantially between different SNPs. Even low error rates (around 3-5%) can affect estimates of LD (Akey et al. 2001) and consequently AM.

With further developments in NGS technologies, sequencing today has extended to entire populations, rather than to a few parental individuals, thus enabling the simultaneous genome-wide detection and scoring of hundreds of thousands of markers (Elshire et al. 2011). This new approach, called "genotyping-by-sequencing" (GBS), uses data directly from the populations being genotyped, thus removing ascertainment bias towards a particular population. GBS is a highly cost-effective technology producing up to a million SNPs per genotype with a cost as low as 20-40 USD. However, one of the unique features associated with GBS is the generation of highly incomplete datasets (Fu 2014), sometimes with up to 90% missing observations per line (Elshire et al. 2011; Fu and Peterson 2011). Such data either needs to be discarded or imputed before it can be used for any genetic analysis (Fu 2014). Many methods are now available for imputation; regression-based methods such as random forest (RF, Stekhoven and Bühlmann 2011) and principal component analysis (PCA)-based tools (Stacklies et al. 2007). The potential of GBS approach for studying marker-trait associations in wheat for yield and yield components under contrasting water regimes and heat stress is underway at CIMMYT.

Candidate gene-based association study

Candidate gene AM is aimed at linking phenotypic variation with polymorphic sites in candidate genes to identify causative polymorphisms. A candidate gene-based association study is more hypothesis-driven than a genome-wide study. The construction of molecular linkage maps based on genes (for example, expressed sequence tags [ESTs, EST-SSRs]) is one way of identifying the candidate genes underlying QTL, instead of time-consuming fine mapping (Sehgal et al. 2012). Also, the choice of candidate gene(s) can be based on relevant information obtained from genetic, biochemical, physiological or expression studies in both model and non-model plant species (Sehgal and Yadav 2009). Standard neutrality tests

applied to DNA sequence variation data can also be used to select candidate genes or amino acid sites that are putatively under selection for AM. This is one of the effective alternative strategies in AM that allows reducing the total amount of marker genotyping in a lower number of individuals and increases the power and precision of the trait-marker correlations. However, it is important to remember that a candidate gene approach is limited by the choice of candidate genes that are identified (and probably explain a small percent of the phenotypic variance) and hence always runs the risk of missing out on identifying causal mutations located in non-identified genes. In wheat, a candidate gene-based association study identified marker-trait associations for drought tolerance traits (Edae et al. 2013). In this study, known drought stress-induced genes in ABA-dependent (*ERA1*) and ABA-independent (*DREB1A, 1-FEH*) pathways were used as candidates genes (Edae et al. 2013). In an another candidate gene-based AM in wheat, an SNF-1 type serine-threonine protein kinase *TaSnRK2.8* showed association with plant height, flag leaf width and water-soluble carbohydrates under drought conditions (Zhang et al. 2013). This gene was selected based on previous evidence (Zhang et al. 2010) of its role in enhancing tolerance to drought, salt and low temperature.

Confounding effects of population structure

One of the main hurdles for using AM to dissect the genetic architecture of complex traits in plants is the risk of incurring false positives due to population structure (Pritchard et al. 2000). The problem of population structure arises because any phenotypic trait that is also correlated with the underlying population structure at neutral loci will show an inflated number of positive associations resulting in Type I errors. Among many methods developed to deal with this problem, the 'genomic control' (GC) method (Devlin and Roeder 1999) estimates association using a large number of putative neutral markers or markers not thought to be involved in controlling the trait of interest. The distribution of the test statistic of interest is then calculated from these associations and a critical value corresponding to the desired Type I error rate is chosen from this distribution. Another method that is commonly used is structured associations (SA) (Pritchard et al. 2000). SA first searches a population for closely related clusters/subdivisions using a Bayesian approach, and then uses the clustering matrices (Q) in AM (by a logistic regression) to correct for false associations. Population structure and shared co-ancestry coefficients between individuals of subdivisions of a population can be effectively estimated with the STRUCTURE program (Pritchard et al. 2000) using several models for linked and unlinked markers.

Principal component analysis (PCA) was recently suggested as a fast and effective way to diagnose population structure (Zhu and Yu 2009). The PCA method summarizes variation observed across all markers into a number of underlying component variables and these components, typically the first few, can then be used to replace Q to adjust for population structure. The PCA method makes it computationally feasible to handle a large number of markers (tens of thousands) and correct for subtle population stratification. There are many programs that can be used to calculate PCA such as DARwin (Perrier and Jacquemoud-Collet, 2006) and EIGENSTRAT (Price et al. 2006).

However, incorporating only population structure information in the analysis is not good enough itself when highly structured population with some degree of related individuals are used in the AM. A mixed linear model (MLM) that combines both population structure information (*Q*-matrix or PCA) and level of pairwise relatedness coefficients (kinship-matrix) should be used in the analysis. While the *Q*-matrix explains the structure between groups in a population, the kinship-matrix accounts additionally for the within group structure. Although computationally intensive, the MLM approach is very effective in removing the confounding effects of the population in AM (Yu et al. 2006).

Estimates of LD

The terms linkage and LD are often confused. Linkage refers to the correlated inheritance of loci through the physical connection on a chromosome, whereas LD refers to the correlation between alleles in a population (Flint-Garcia et al. 2003) but not necessarily on the same chromosome. As a starting point for AM, it is important to gain knowledge of the patterns of LD for genomic regions of the "target" organisms and the specificity of the extent of LD

among different populations or groups to design and conduct unbiased association mapping. The two most commonly used statistics to measure LD are r^2 (square of the correlation coefficient) and D' (disequilibrium coefficient). The statistics r^2 and D' reflect different aspects of LD and perform differently under various conditions. The r^2 is affected by both mutation and recombination while D' is affected by more mutational histories.

There are many freely available softwares such as GOLD (Abecasis and Cookson 2000), TASSEL (www.maizegenetics.net) or Powermarker (Liu and Muse 2005) to depict the structure and pattern of LD. One can estimate an average genome-wide decay of LD by plotting LD values (r^2 values) obtained from a data set covering an entire genome (i.e., with more or less evenly spaced markers across all chromosomes in a genome) against the genetic or physical distance between markers. When such a LD decay plot is generated, the usual practice is to determine the distance where LD values (r^2) decrease below 0.1 or half strength of D' (D' = 0.5) based on the curve of the nonlinear logarithmic trend line. This gives a rough estimate of the extent of LD for association studies, but for more accurate estimates, highly significant threshold LD values ($r^2 \geq 0.2$) are also used as a cutoff point. The decrease of the LD within the genetic distance indicates that the portion of LD is conserved with linkage and proportional to recombination (Gupta et al. 2005). The decay of LD over physical/genetic distance in a population determines the density of marker coverage needed to perform an association analysis. If LD decays rapidly, then a higher marker density is required to capture markers located close enough to functional sites (Flint-Garcia et al. 2003; Gaut and Long, 2003). In wheat, depending on the populations used, LD decay have been reported to vary from 0.5 to 40cM (Chao et al. 2007, 2010; Tommasini et al. 2007; Crossa et al. 2007; Somers at al. 2007; Yao et al. 2009; Dreisigacker et al 2012).

Association with raw data or BLUPs or residuals

In general, raw data can be used directly in association analysis provided it is available for all entries and for all replicates in different locations/years. For cases where phenotypes are not evaluated for all individuals and replicates due to large sample size, BLUPs from a mixed model may be substituted as the dependent variable. In such cases, the association analysis using BLUPs can be performed with many fewer observations and require much less time.

More recently, researchers have also used residuals instead of raw data. The rationale is that after removing all the effects except the marker, including the polygenic genetic variance captured by the BLUPs, the signal due to marker association is still contained in the residuals. Signals from the markers will be removed only to the extent that it is correlated with the other effects. The residuals approach performs as well as the approach using raw phenotype directly for low heritability traits (Aulchenko et al. 2007). Because the association test using residuals is performed without including the polygenic random effect, tests of individual markers run quickly. The mixed model equations with thousands of individuals only need to be solved once for any particular phenotype. After that, the millions of association tests for individual markers can then be performed using simple t-tests or F-tests of the marker classes.

Association analysis programs

Public, freely available software suitable for association analysis using mixed models in plants include TASSEL and EMMA/R. Both analyze moderately large datasets in a reasonable amount of time but only allow a single effect (samples or taxa) to be fit as a random effect. All other effects are treated as fixed. EMMA relies on the R for data management and visualization whereas TASSEL handles those functions itself. Several commercial software packages available for association studies include ASREML, JMP Genomics, SAS and GenStat. ASREML and JMP Genomics are specifically engineered for genetic analysis and can handle more complex models, whereas general purpose packages such as SAS Proc Mixed and GenStat can perform association analysis but require more expertise and programming on the part of the user.

From the user's perspective, clearly freely available software such as TASSEL plays an important role in scientific investigation. Another advantage of using TASSEL is that both GUI (graphical user interface) and CLI (command line interface) versions exist. In the GUI, the plug-ins are invoked by clicking buttons on the interface. With the CLI, the plug-ins are used in a

predetermined pipeline that passes the output from one step to the input of another. Hence, scientists can use these versions depending on their expertise and consistent results are achieved independent of the interface. In the latest version of TASSEL (TASSEL 5.0), a compressed MLM method is available for computing large datasets with up to 500,000 markers.

Significance threshold

A threshold is set to declare significant associations. Either of the two statistical methods – False Discovery Rate (FDR) or Bonferroni correction – can be used to correct for multiple comparisons. The correction is needed whenever one would like to test multiple hypotheses simultaneously. FDR controls the expected proportion of false positives among significant results by determining a threshold from the observed p-value distribution in the data, whereas Bonferroni corrections control the chance of any false positives (Benjamini and Hochberg 1995). Given the aims of the study, one may consider a high FDR for some projects (e.g. investigating the genetic architecture of a trait) and a low FDR for others (e.g. identifying candidate loci for follow-up studies).

Validation of association results

Validation of AM results is required before marker information is incorporated in selection decisions, or before larger efforts are invested into identification of causal factors and gene cloning. The most straightforward way is to compare the AM results with previous results published for the trait; for example using bi-parental populations. If markers in close proximity (within 10 cM) to previously reported QTLs/genes are identified, the result will not only be validated but also increase the confidence to pursue the new genomic target identified for the trait. Secondly, results can be validated in different populations. This is more reliable as the probability of observing false positives becomes small if significant associations are confirmed in two or more validation populations. Third, if association studies point to alleles with opposite effects on a trait of interest, one can generate multiple F_2 populations from parents that harbor contrasting alleles and determine whether differences in phenotype co-segregate with the locus in question. Once markers tightly linked to the target trait are validated, they provide several magnitudes of return on investment through increased speed and cost efficiency of breeding programs.

References

Abecasis, G.R. and W.O. Cookson. 2000. GOLD—graphical overview of linkage disequilibrium. Bioinformatics 16:182–183

Adhikari, T.B., S. Gurung, J.M. Hansen, E.W. Jackson and J.M. Bonman. 2012. Association mapping of quantitative trait loci in spring wheat landraces conferring resistance to bacterial leaf streak and spot blotch. The plant genome 5: NO. 1

Akey, J.M., K. Zhang, M. Xiong, P. Doris and L. Jin. 2001. The effect that genotyping errors have on the robustness of common linkage disequilibrium measures. American Journal of Human Genetics 68: 1447–56

Akhunov, E., C. Nicolet, and J. Dvorak. 2009. "Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay," Theoretical and Applied Genetics119: 507–517

Aulchenko, Y.S., D.J. de Koning and C. Haley. 2007. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics177:577–85

Benjamini, Y. and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society 57:289–300

Breseghello, F., P.L. Finney, C. Gaines, L. Andrews, J. Tanaka, G. Penner and M.E. Sorrells. 2005. Genetic loci related to kernel quality differences between a soft and a hard wheat cultivar. Crop Science 45: 1685-1695

Breseghello, F. and M.E. Sorrells. 2005. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 2005, 172:1165-1177

Cavanagh, C.R., S. Chao, S. Wang, B.E. Huang, S. Stephen, S. Kiani, K. Forrest, C. Saintenac, G.L. Brown-Guedira, A. Akhunova, D. See, G. Bai, M. Pumphrey, L. Tomar, D. Wong, S. Kong, M. Reynolds, M.L. da Silva, H. Bockelman, L. Talbert, J.A. Anderson, S. Dreisigacker, S. Baenziger, A. Carter, V. Korzun, P.L. Morrell, J. Dubcovsky, M.K. Morell, M.E. Sorrells, M.J. Hayden and E. Akhunov. 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proceedings of the National Academy of Sciences United States of America, 110: 8057–8062

Chao, S., J. Dubcovsky, J. Dvorak, M.C. Luo, S.P. Baenziger, R. Matnyazov, D.R. Clark, L.E. Talbert, J.A. Anderson, S. Dreisigacker, K. Glover, J. Chen, K. Campbell, P.L. Bruckner, J.C. Rudd, S. Haley, B.F. Carver, S. Perry, M.E. Sorrells, and E.D. Akhunov. 2010. Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). BMC Genomics 11: 727

Chao, S., W. Zhang, J. Dubcovsky, and M.E. Sorrells. 2007. Evaluation of genetic diversity and genome-wide linkage disequilibrium Association mappingong U.S wheat (*Triticum aestivum* L.) germplasm representing different market classes. Crop Science 47: 1018-1030

Clark, R.M., E. Linton, J. Messing and J.F. Doebley. 2004. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. Proceedings of the National Academy of Sciences United States of America, 101:700-707

Crossa, J., J. Burgueno, S. Dreisickacker, M. Vargas, S.A. Herrera-Foessel, M. Lillemo, R.P. Singh, R. Trethowan, M. Warburton, J. Franco, M. Reynolds, J.H. Crouch and R. Ortiz. 2007. Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 177: 1889-1913

Devlin, B. and K. Roeder. 1999. Genomic control for association studies. Biometrics 55:997–1004

Dreisigacker, S., H. Shewayrga, J. Crossa, V.N. Arief, I.H. DeLacy, R.P. Singh, M.J. Dieters, and H.J. Braun. 2012. Genetic structures of the CIMMYT international yield trial targeted to irrigated environments Molecular Breeding 29: 529-541

Edae, E.A., P.F. Byrne, H. Manmathan, S.D. Haley, M. Moragues, M.S. Lopes and M.P. Reynolds. 2013. Association Mapping and Nucleotide Sequence Variation in Five Drought Tolerance Candidate Genes in Spring Wheat. The Plant Genome doi: 10.3835/plantgenome2013.04.0010

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler and S.E. Mitchell 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379

Flint-Garcia, S., J.M. Thornsberry and E.S. Buckler IV. 2003. Structure of linkage disequilibrium in plants. Annual Review of Plant Biology 54: 357-374

Fu, Y.B. 2014. Genetic Diversity Analysis of Highly Incomplete SNP Genotype Data with Imputations: An Empirical Assessment. Genes Genomes Genetics 4: 891-900

Fu, Y. B. and G. W. Peterson. 2011. Genetic diversity analysis with 454 pyrosequencing and genomic reduction confirmed the eastern and western division in the cultivated barley gene pool. Plant Genetics 4: 226–237

Gaut, B.S. and A.D. Long. 2003. The lowdown on linkage disequilibrium. Plant Cell 15 (7): 1502-1506

Gupta, S., K. Kumari, M. Muthamilarasan, S.K. Parida, and M. Prasad. 2014. Population structure and association mapping of yield contributing agronomic traits in foxtail millet. Plant Cell Reports 33 (6): 881-893

Honsdorf, N., H.C. Becker and W. Ecke.2010. Association mapping for phenological, morphological, and quality traits in canola quality winter rapeseed (*Brassica napus L.*). Genome. 53(11):899-907

Huang, B.E., A.W. George, K.L. Forrest, A. Kilian, M.J. Hayden, M.K. and C.R. Cavanagh. 2012. A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotechnology Journal 10 (7): 826-829

Imelfort M, C. Duran , J. Batley and D. Edwards . 2009. Discovering genetic polymorphisms in next-generation sequencing data. Plant Biotechnol J 7:312-17

Kover, P.X., W. Valdar, J. Trakalo, N. Scarcelli, I.M. Ehrenreich, M.D. Purugganan, C. Durrant and A. Mott. 2009. A multiparent advanced generation inter-cross to fine-map quantitative traits in Arabidopsis thaliana. PLoS Genetics 5:e1000551

Liu, L., L. Wang, J. Yao, Y. Zheng and C. Zhao. 2010. Association mapping of six agronomic traits on chromosome 4A of wheat (*Triticum aestivum* L.). Molecular Plant Breeding 1: 1-10

Liu, K. and S.V. Muse. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. Bioinformatics 21(9): 2128-2129

Lopes, M.S., S. Dreisigacker, R.J. Peña, S. Sukumaran and M.P. Reynolds. 2015. Genetic characterization of the wheat association mapping initiative (WAMI) panel for dissection of complex traits in spring wheat. Theoretical and Applied Genetics 128(3):453-64

McMullen, M.D., S. Kresovich, H.S. Villeda, P. Bradbury, H. Li, Q. Sun, S. Flint-Garcia, J. Thornsberry, C. Acharya, C. Bottoms, P. Brown, C. Browne, M. Eller, K. Guill, C. Harjes, D. Kroon, N. Lepak, S.E. Mitchell, B. Peterson, G. Pressoir, S. Romero, M. Oropeza Rosas, S. Salvo, H. Yates, M. Hanson, E. Jones, S. Smith, J.C. Glaubitz, M. Goodman, D. Ware, J.B. Holland and E.S. Buckler. 2009. Genetic properties of the maize nested association mapping population. Science 325:737–740

Palaisa KA, M. Morgante, M. Williams and A. Rafalski: Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 2003, 15:1795-1806

Peng, J.H., Y. Bai, S.D. Haley and N.L.V. Lapitan. 2009. Microsatellite based molecular diversity of bread wheat germplasm and association mapping of wheat resistance to the Russian wheat aphid. Genetica 135: 95-122

Perrier, X. and J.P. Jacquemoud-Collet. 2006. DARwin software http://darwin.cirad.fr.

Poland, J.A., P.J. Brown, M.E. Sorrells and J.L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253. doi:10.1371/journal.pone.0032253

Price, A.H. 2006. Believe it or not, QTLs are accurate! Trends in Plant Science 11:213-6.

Price, A.L., N.J. Patterson and R.M. Plenge. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38:904–909

Pritchard, J.K., M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–59

Rasheed, A., X. Xia, F. Ogbonnaya, T. Mahmood, Z. Zongwen, A. Mujeed-Kazi and Z. He. 2014. Genome-wide association for grain morphology in synthetic hexaploid wheats using digital imaging analysis. BMC Plant Biology 14:128

Reif, J.C., M. Gowda, H.P. Maurer, C.F.H. Longin, V. Korzun, E. Ebmeyer, R. Bothe, C. Pietsch and T. Würschum. 2011. Association mapping for quality traits in soft winter wheat. Theoretical and Applied Genetics 122: 961-970

Ritland, K. 2005. Multilocus estimation of pairwise relatedness with dominant markers. Molecular Ecology 14:3157–65

Sehgal, D., V. Rajaram, V. Amstead, V. Vadez, C.T. Hash and R.S. Yadav. 2012. Integration of gene-based markers in a pearl millet genetic map for identification of candidate genes for drought tolerance. BMC Plant Biology 2012, DOI, 10.1186/1471-2229-12-9

Sehgal D. and R.S. Yadav. 2009. Molecular marker based approaches for drought tolerance. In: Mohan JS, Brar DS eds. Molecular techniques in crop improvement, 2009, 2nd edition, pp 207-230

Sehgal D., E. Autrique, R. Sing and Dreisigacker. 2015. Genome-wide association analysis of grain yield across multiple environments in a large elite bread wheat population using a genotyping-by-sequencing approach (under review)

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

Stekhoven, D. J. and P. Bühlmann, 2011. MissForest—nonparametric missing value imputation for mixed-type data. Bioinformatics 28: 112–118

Stacklies, W., H. Redestig, M. Scholz, D. Walther and J. Selbig, 2007. pcaMethods—a bioconductor package providing PCA methods for incomplete data. Bioinformatics 23: 1164.

Tommasini, L., T. Schnurbusch, D. Fossati, F. Mascher and B. Keller. 2007. Association mapping of *Stagonospora nodorum* blotch resistance in modern European winter wheat varieties. Theoretical and Applied Genetics 115: 697-708

Upadhyaya, H.D., Y.H.Wang, C.L.L.Gowda and S. Sharma. 2013. Association mapping of maturity and plant height using SNP markers with the sorghum mini core collection Theoretical and Applied Genetics 126:2003–2015

Visscher, P.M. 2008. Sizing up human height variation. Nature Genetics 40:489–490

Wang, S., D. Wong, K. Forrest, A. Allen, S. Chao, B.E. Huang, M. Maccaferri, S. Salvi, S.G. Milner, L. Cattivelli, A.M. Mastrangelo, A. Whan, S. Stephen, G. Barker, R. Wieseke, J. Plieske, M. Lillemo, D. Mather, D., R. Appels, R. Dolferus, G. Brown-Guedira, A. Korol1, A.R. Akhunova, C. Feuillet, J. Salse, M. Morgante, C. Pozniak, M-C. Luo, J. Dvorak, M. Morell, J. Dubcovsky, M. Ganal, R. Tuberosa, C. Lawley, I. Mikoulitch, C. Cavanagh, K.J. Edwards, M. Hayden and E. Akhunov. 2014. Characterization of polyploid wheat genomic diversity using a high density 90000 single nucleotide polymorphism array. Plant Biotechnology Journal 12, 787-796

Weigel, D. and M. Nordborg. 2005. Natural variation in Arabidopsis. How do we find the causal genes? Plant Physiology 138:567-568

Yao, J., L. Wang, L. Liu, C. Zhao and Y. Zheng. 2009. Association mapping of agronomic traits on chromosome 2A of wheat. Genetica 137: 67-75

Yu, J., G. Pressoir and W.H. Briggs .2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38:203–8

Yu, X., G. Bai, S. Liu, N. Luo, Y. Wang, D.S. Richmond, P.M. Pijut, S.A. Jackson, J. Yu and Y. Jiang. 2013. Association of candidate genes with drought tolerance traits in diverse perennial ryegrass accessions. Journal of Experimental Botany 64: 1537-1551

Zegeye, H., A. Rasheed, F. Makdis, A. Badebo and F.C. Ogbonnaya. 2014. Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. PLoS One 9: e 105593

Zhang, H., X. Mao, C. Wang and R. Jing. 2010. Overexpression of a common wheat gene *TaSnRK2.8* enhances tolerance to drought, salt and low temperature in Arabidopsis. PLoS One 5: e16041

Zhang, H., X. Mao, J. Zhang, X. Chang and R. Jing. 2013. Single-nucleotide polymorphisms and association analysis of drought-resistance gene *TaSnRK2.8* in common wheat. Plant Physiology and Biochemistry 70:174-181

Zhang, J., C. Hao, Q. Ren, X. Chang, G. Liu and R. Jing. 2011. Association mapping of dynamic developmental plant height in common wheat. Planta 234: 891-902

Zhu, C. and J. Yu. 2009. Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. Genetics 182 (3): 875-888

Chapter 3:
Optimizing Marker-Assisted Selection (MAS) Strategies

Susanne Dreisigacker and David G. Bonnett
CIMMYT Int., Apdo. Postal 6-641,06600 Mexico, DF, Mexico

Introduction

Marker-assisted selection allows for the selection of QTL or genes that control traits of interest. In wheat, the number of markers known to be associated with QTL/genes for major economic traits has been growing during the last decade and marker discovery will be further be accelerated with the availability of a high quality reference sequence of the wheat genome (Choulet et al. 2014). Together with decreasing marker assay costs and interconnected genotyping service facilities, the opportunity to apply MAS strategies is becoming accessible to more and more breeding programs.

Marker-assisted selection can supplement conventional breeding to increase genetic gain. Efficient MAS strategies can substantially cut down population sizes, allow selection for a maximum number of loci and thus reduce the time and cost needed to recover a desirable genotype. However, several factors need to be considered in choosing the trait and strategy for which MAS is appropriate. Depending on the populations and the trait selected for, empirical comparisons of MAS and phenotypic selection for increasing genetic gain from selection revealed different results. In some studies, MAS is reported to achieve higher selection gains than phenotypic selection (Abalo et al. 2009; Kuchel et al. 2007; Miedaner et al. 2009). Other studies considered the two methods equally effective (Moreau et al. 2004). In a third group of studies phenotypic selection proved to me more efficient than MAS (Davis et al. 2006; Wilde et al. 2007). Most studies conclude that MAS is most appropriate when the target traits 1) have low heritability, 2) are difficult and cost-prohibitive to measure, or 3) require desired pyramiding of genes. Every breeding program has its own set of breeding objectives and its own way to measure a trait; therefore the choice of traits for MAS and to be combined with phenotypic selection is individual for each breeding program and might vary between programs. The careful and efficient integration of marker and phenotypic selection in very individual program is crucial to maximize overall gains.

Traits that have been targeted for MAS in wheat include: (1) disease/pest resistance (e.g., rust, Fusarium head blight, wheat nematodes and aphids); (2) quality traits (e.g., grain hardiness, grain color, grain texture and gluten strength); (3) traits linked to development and growths (e.g., height, phenology); and (4) abiotic stresses (e.g. drought, heat). In subsequent chapters we will describe in more detail how we use MAS for some of these traits in the CIMMYT Global Wheat Program. This chapter aim to outline principles that should be useful in designing a breeding strategy and critical factors to consider integrating MAS with currently available technology.

Marker-assisted selection strategies

In every breeding program, modern varieties are combinations of alleles for yield, grain quality and tolerance to biotic and abiotic stresses that have been assembled over multiple cycles of crossing and selection. A cross made with the aim of producing a variety will have parents with many alleles in common controlling these characters and simple or topcrosses will be made. If parents have a lower co-ancestry and differ for a greater number of alleles, genetic variation of the progenies will increase, but it will be difficult to produce a line suitable for release as a variety from a simple bi-parental cross (Longin et al. 2014). In these cases, or where one parent contributes only a small number of desirable attributes and the other contribute many more, one or more backcrosses may be necessary to recover a commercially viable line. Effective use of markers can make a large difference to the probability of obtaining a desirable genotype in elite or wide-cross populations or estimate the population size needed to have a reasonable probability of recovering it.

Marker-assisted allele enrichment strategies in early generations

In the commonly used breeding methods for self-pollinating crops, selecting desirable plants begins in early generations for traits of higher heritability. For traits of low heritability, selection is often postponed until the lines become more homozygous in later

generations (F_5 or F_6). Selection of superior plants involves visual assessment for agronomic traits or resistance to stresses, as well as laboratory tests for quality or other traits. To improve early generation selection, markers should decrease the number of plants retained due to their early generation performance, and at the same time they should ensure a high probability of retaining superior lines for selection in later generations (Eathington et al. 1997). If markers are cheaper, more heritable, easier or more accurate to select for than the target trait, they should be used in earlier stages of a breeding program to increase the frequency of the alleles favorable for the trait under selection.

The principle of F_2-enrichment introduced by Bonnett et al. 2005 is applied at CIMMYT after parental material is first characterized with markers for known genes to identify those parents with favorable alleles, which are then selectively combined in crosses. The concept of F_2-enrichment is illustrated in Figure 1. In the F_2 of a bi-parental cross, at every polymorphic locus, three-fourths of F_2 individuals will carry at least one copy of the preferred A allele. Both AA and Aa individuals will produce the preferred AA homozygous progeny and should be retained in the population. Individuals with the aa genotype cannot produce the AA progeny and should be culled from the population. Culling aa and retaining both AA and Aa increases the frequency of the A allele from one in two to two in three, and thus enriches the frequency of the A allele in the population. If no further selection was applied and the population was progressed to homozygosity by inbreeding or production of double haploids (DHs) from selected F_2s, the frequency of AA genotypes in the final population would be two in three and the frequency of aa only one in three. Alternatively to retaining all carriers of the desirable allele (AA, Aa) only the AA homozygous individuals could be retained in the population. The locus would not further segregate in the progeny; however, as the frequency of the AA individuals is only one in four, one-half less of F_2 individuals would be overall retained.

The advantage of F_2-enrichment becomes even more apparent with greater numbers of polymorphic loci (B, C, D, etc.); the breeding target being to pyramid or combine several genes into a single genotype. The difference in the frequencies and the population sizes needed to recover only homozygotes vs. carriers of all desirable alleles becomes larger (see Table 1). E.g., in a population segregating at two loci, A and B, the frequency of the preferred AABB homozygote is much smaller at just one in 16 than the frequency of A-B- carriers (AABB, AABb, AaBB, AaBb) at nine in 16. With no polymorphic loci, the frequency of homozygotes in F_2 is (one in four)^n, the frequency of carriers is (three in four)^n. F_2-enrichment will increase the frequency of desirable homozygotes to (two of three)^n in inbred or DH lines produced from the selected F_2s that carry at least one copy of the target allele at all loci. At CIMMYT A-B- carriers are usually advanced via bulk breeding. In cases when three loci are combined, several bulks are sometimes advanced, e.g., a bulk that includes the A-B-carriers of all three loci but also a bulk that includes the A-B-carriers of two of the three loci.

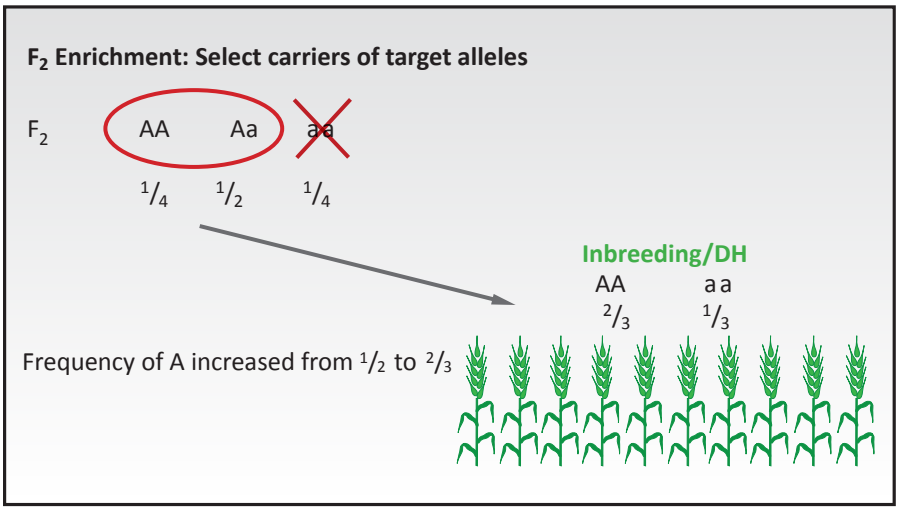


Figure 1. Strategy of QTL mapping for MBA.

Table 1. Population sizes required for enrichment (enrich) vs. fixation (fix) of target alleles in biparental F_2 populations and to obtain at least one target homozygous genotype in later generation enriched (enrich) and non-enriched (rand) populations for different numbers of segregating loci (Bonnett et al. 2005).

	Pop. required for fix vs. enrich (p = 0.05)		Population size required to obtain a target homozygote at all loci in rand and enrich populations (p = 0.05)									
Gen:	F_2		$F_{2:3}$		$F_{3:4}$		$F_{4:5}$		$F_{5:6}$		DH	
Loci:	Fix	Enrich	Rand	Enrich	Rand	Enrich	Rand	Enrich	Rand	Enrich	Rand	Enrich
1	11	3	7	5	6	4	5	3	5	3	5	3
2	47	4	20	11	15	8	13	6	12	6	11	6
3	191	6	56	23	35	14	28	11	25	10	23	9
4	766	8	151	47	81	25	61	18	53	16	47	14
5	3067	11	403	95	186	43	131	30	111	26	95	22
6	12270	16	1076	191	426	75	281	49	231	40	191	33
7	49081	21	2872	382	975	129	601	79	478	63	382	50
8	196327	29	7659	766	2231	222	1284	128	988	98	766	76
9	785312	39	20427	1533	5100	382	2741	205	2040	152	1533	114
10	3141252	52	54473	3067	11660	656	5848	329	4213	236	3067	172

Each selected F_2 will need to produce several progeny to make up the required number of lines in subsequent generations. Each selected F_2 should contribute equal numbers of progeny to the subsequent population in order to avoid changes in allele frequencies due to genetic drift. Table 1 shows the population sizes needed to use F_2 enrichment in a bi-parental cross in the F_2 generation and in later generation populations derived from the selected F_2s. For comparison it also shows the population sizes needed to recover homozygotes in different generations when enrichment has not been applied.

In certain cases backcross (BC_1F_1) or top cross (TCF_1) populations are made to combine genes of interest. If markers are going to be used in BC_1F_1 or TCF_1 populations, the desired alleles or allele combinations are of lower frequency. For example, desirable alleles coming from the non-recurrent or donor line will have a frequency of one in four in BC_1F_1 or TCF_1 populations and one-half of the population will lack the allele. Selection among BC_1F_1 or TCF_1 populations will increase the frequency of target alleles from donors from one in four to one in two and ensure all selected individuals carry one copy of all target allele. If followed by F_2-enrichment, the frequency of donor alleles is increased from one in four to two in three. Table 2 shows the frequencies of carriers and homozygotes for target alleles at single loci with a range of common initial allelic frequencies in different generations. This table can be used to calculate frequencies of carriers or homozygotes that can be selected in a desired generation.

In populations with differing frequencies of target alleles at different polymorphic loci, the frequency of an individual with a particular genotype across all loci can be calculated by multiplying the individual frequencies at each locus. For example, in a bi-parental population in which F_2-enrichment has been applied for target alleles at 6 loci, the frequency of a genotype homozygous at all loci in the F_4 generation is $0.583^6 = 0.060$. In a similar backcross populations in which target alleles at 4 loci coming from the recurrent parent and 2 from the donor with enrichment applied in the BC_1F_1 for donor alleles and in F_2 for donor and recurrent parent alleles, the frequency of an individual in a DH population developed following F_2-enrichment would be 0.67^2 (donor alleles) × 0.857^4 (recurrent parent alleles) in both BC_1F_1 (increasing frequency at each locus from one in four to one in two) and subsequent enrichment in F_2 increasing the frequency of these donor alleles from one in two to two of three. Enrichment of the recurrent parent alleles in BC_1F_2 increases their frequency from three in four to seven of eight. In spite of the relatively high frequency of homozygotes for the recurrent parent alleles in a backcross, enrichment still requires smaller

Table 2. Frequencies of homozygotes (homo) and carriers of a target allele (A) for different allele frequencies and levels of inbreeding (Bonnett et al. 2005).

Allelic frequency	1/4 (e.g. non-recurrent parent allele in BC_1)		1/2 (e.g. biparental cross)		3/4 (e.g. recurrent parent allele in BC_1)		2/3 (e.g. following F_2 enrichment of biparental cross)		7/8 (e.g. following F_2 enrichment of recurrent parent allele in BC_1)	
Gen	Homo (AA)	Carrier (A-)	Homo (AA)	Carrier (A-)	Homo (AA)	Carrier (A-)	Homo (AA)	Carrier (A-)	Homo (AA)	Carrier (A-)
F_2	0.125	0.375	0.25	0.75	0.625	0.875	0.333	1	0.714	1
F_3	0.188	0.313	0.375	0.625	0.688	0.813	0.5	0.833	0.786	0.929
F_4	0.219	0.281	0.438	0.563	0.719	0.781	0.583	0.75	0.821	0.893
F_5	0.234	0.266	0.469	0.531	0.734	0.766	0.625	0.708	0.839	0.875
F_6	0.242	0.258	0.484	0.516	0.742	0.758	0.646	0.688	0.848	0.866
F_7	0.246	0.254	0.492	0.508	0.746	0.754	0.656	0.677	0.853	0.862
F_8	0.248	0.252	0.496	0.504	0.748	0.752	0.661	0.672	0.855	0.859
F_9	0.249	0.251	0.498	0.502	0.749	0.751	0.664	0.669	0.856	0.858
F_{10}	0.25	0.25	0.499	0.501	0.75	0.75	0.665	0.668	0.857	0.858
DH	0.25	0.25	0.5	0.5	0.75	0.75	0.667	0.667	0.857	0.857

population sizes than selection of homozygotes. For more information on application of allele enrichment refer to the publication of Bonnett et al. (2005). On high priority materials, at CIMMYT we apply marker assays additionally at the F_4 or F_5 generation to ensure a high frequency of advanced progeny containing the alleles of interest.

Balancing early and late generation, marker and phenotypic selection

In reality, markers or efficient phenotypic screens will rarely be available for alleles at all important loci segregating in a cross and it will not be possible to enrich frequencies of these alleles in early generations. Early generation selection strategies must therefore be designed to retain important allelic variation until later stages of the breeding process to select for more complex traits like yield that require homogeneous lines, large seed quantities and expensive phenotypic screens to achieve acceptable heritability.

Estimating the number of important polymorphic loci or deciding on a certain number of inbred lines to be retained for phenotypic selection can be implemented to optimize overall required population sizes. Required population sizes to recover an individual with a target genotype are inversely related to the frequency of those individuals. A formula for calculating population size for any frequency and desired level of confidence of recovery was given by Hanson (1959):

$$N = \frac{log_n (x)}{log_n (1\text{-}G)'}$$

where *N* is the population size, x the specified probability of failure and *G* the genotypic frequency. A useful rule of thumb is to multiply the inverse of the frequency by 3 to achieve a commonly desired 95% probability of recovery. For example with a frequency of one in 16, the population size needed for 95% probability of recovering the target genotype is 16 × 3 = 48. In other words: Population size = (1/frequency of target genotype) × 3. This formula applies regardless of whether the target genotype is homozygous or heterozygous. Often, the number of important loci contributing variation to important traits in a cross will not be known and partial enrichment is applied by estimating the number of important polymorphic loci or deciding on a certain number of inbred lines to retain for phenotypic selection. E.g., measures can be translated to a partial enrichment strategy where e.g. six loci are taken into account for enrichment plus additional four important polymorphic loci for which markers are not available.

If greater numbers of important loci are present in the population, it may not be possible to recover an individual with the most favorable allele at all loci but it may still be possible to make useful progress. Although partial F_2-enrichment requires larger population sizes than if markers were available for alleles at all important loci, substantial reductions in population sizes can still be achieved compared with not enriching the frequency of any alleles. See Passioura et al. 2007.

Use of modified pedigree or bulk breeding methodologies

After enrichment in F_2, inbreeding by whatever system will ultimately produce the same frequency of target homozygotes as would be produced through SSD, providing selection for other traits does not affect frequencies of the target alleles through linkage or pleiotropy and population sizes remain large enough to avoid changes in allele frequencies due to drift. Bulk breeding methodologies as used at CIMMYT may be a very efficient means of progressing populations to homozygosity while selecting for other traits and provided this selection does not cause changes in the frequencies of 'target' alleles following the enrichment steps due to linkage, pleiotropy, or genetic drift, the expected frequencies of target genotypes should be similar to those predicted.

Marker-assisted backcrossing (MABC)

As outlined above, if parents have a low co-ancestry, genetic variation of the progeny will increase, but it will be difficult to produce a line suitable for release as a variety from a simple bi-parental cross. Various generations of backcrossing are used in some cases to transfer a desired trait from a rather unadapted donor plant into an elite genotype (recurrent parent) until most of the genes stemming from the donor are eliminated (Becker 1993).

Markers can be used in the context of MABC to either control the target gene (foreground selection) or to accelerate the reconstruction of the recurrent parent genotype (background selection). According to Tanksley et al. (1989), in traditional backcross breeding the reconstruction of the recurrent parent genotype requires more than six generations, while this may be reduced to only three generations in MABC. Similarly, Hospital et al. (1992) and Ribaut and Hoisington (1998) concluded that employing molecular markers with known map position can speed up the recovery of the recurrent parent genome by about two to three generations. These findings are confirmed by results of Frisch et al. (1999), who showed in a computer simulation that MAS can reconstruct a level of recurrent parent genome in BC_3 which would only be reached in BC_7 without the use of markers. Prigge et al. (2008) compared simulated and experimental data of a MABC program in rice and revealed good agreement.

The effectiveness of MABC depends on the availability of closely linked markers/flanking markers for the target loci, the size of the population, the number of backcrosses, and the position and number of markers for background selection. A straight-forward way to accomplish MABC is the two-stage selection strategy. In BC_1F_1 populations, individual plants heterozygote at the target loci are first identified reducing the population size for further screening (foreground selection). For the background selection step, individuals with the fewest number of background markers from the donor parent are then selected. The upper limit of the number of background markers is defined by the number and length of the chromosomes. In rice and sugar beets, 50 to 60 background markers resulted in efficient selection response (Frisch and Melchinger et al. 2005; Prigge et al. 2008). Markers should be evenly distributed to reflect all proportions of the genome. In subsequent backcross generations, selection is carried out to the same scheme, but only those markers are analyzed which have not been fixed for the recurrent parent in the preceding generation.

In BC_1F_1 populations, MABC would be more efficient for larger populations. Larger population sizes in earlier generations are also of advantage for more quantitative traits. However larger populations also increase the number of marker data points required and hence the cost. In comparison to BC_1F_1 populations, the number of markers that needs to be analyzed in later backcross generations is lower. In a two stage selection strategy, increasing the population size with the number of backcross generations reduces the number of marker loci and cost with comparable percentages of recovery of the recurrent parent genome (Frisch et al. 1999). Prigge et al. (2009) additionally showed in a computer simulation that the approach of increasing population sizes in

advanced backcross generations can be refined by additionally increasing marker densities sequentially across backcross generations.

Two additional selection steps can follow the two-stage selection strategy. As a third step, after preselecting the individuals with the target loci, individuals can be analyzed for the two markers flanking the target locus. Individuals with one or two flanking markers fixed for the recurrent parent's allele are retained and then analyzed for the remaining markers. In some publications this three-stage selection strategy is also called "recombinant selection". As a fourth step, individuals with the maximum number of markers fixed on the chromosome of the target locus can be selected before analyzing all other remaining markers. These two steps provide an option which significantly reduces the number of data points required in comparison to the two-stage selection strategy.

Factors to take into considerations when applying MAS

Imperfect linkage between markers and target alleles

Due to the increased marker availability recent genetic maps are dense; however, markers are mostly not perfectly linked with the target allele, which reflects the accuracy of MAS. For example, if the genetic distance between the marker and the target allele is 5 cM, on average five recombinants occur in a set of 100 progenies. In such cases flanking markers can be very useful to decrease the probably of recombinants between target alleles and markers. If two flanking markers with a genetic distance of 5 cM to the target allele are applied, on average only one recombinant occur in a set of 100 progenies. If imperfect markers are used in F_2-enrichment, the change in allele frequency will be slightly less than if markers were perfect. In spite of a slight reduction in efficiency, use of imperfect markers still increases allele frequencies and is very worthwhile. For more details on using imperfect markers, refer to Wang et al. (2007).

Dominant vs co-dominant markers

Markers can be dominant or co-dominant, the latter being able to distinguish heterozygote and homozygote carriers of the target allele. Due to improved marker technologies most of the more recent developed SNP markers for relevant genes in wheat are co-dominant. The advantages of co-dominant markers in F_2-enrichment are that they allow a more direct assessment of the frequencies of target alleles that they remove the need for progeny testing of selected later generation individuals (e.g., F_5 or F_6) to recover homozygotes. When dominant markers are used and progeny testing is not done, some selected individuals will be heterozygous for some of the target alleles. However, because the frequency of heterozygous individuals is halved with each generation of inbreeding, only relatively small numbers of selected F_6 individuals would be heterozygous at any of the target loci. In MABC the advantages of co-dominant markers is more evident. For the background selection step in MABC, loci homozygote for the recurrent parent can be identified.

Linkage between two alleles

With greater numbers of markers available for selection, it is inevitable at some point that a cross will involve target alleles that are linked. If the alleles are linked in coupling they will behave more like a single gene and required population sizes will be smaller than if they were unlinked. If they are linked in repulsion and a crossover between the loci is necessary to bring the target alleles together on the same chromosome, required population sizes will be considerably larger. If target m alleles are linked in repulsion it will usually be best to first recover a recombinant with the target alleles in coupling and then focus on combining the other alleles (Wang et al. 2007). For example, the wheat stem rust gene *Sr2* and the fusarium head blight gene *Fhb1* are linked in repulsion on chromosome 3BS (Anderson et al. 2007).

Polymorphism and genetic backgrounds

Ideally, a marker should be highly polymorphic in breeding materials and discriminate between different genotypes. In some cases, the target polymorphism of a marker is only specific in certain donors and therefore not diagnostic in all genetic backgrounds. These markers cannot be used for the screening of unknown sets of germplasm of a breeding program. They can be used to follow a target allele in segregating populations including the parental line known to carry the target allele and the

marker showing polymorphism between the carrier and non-carrier of the allele.

In QTL mapping experiments, parents that represent the extreme ends of a trait phenotype are chosen. The effect of the QTL might therefore be less significant when used for introgression into an elite breeding line. In other cases, the effect of a locus may differ in different genetic backgrounds due to the interaction with other loci (epistasis) (Holland et al. 2001).

Environmental effects

While the effect of a QTL appears to be consistent across environments, the magnitude of the effect may vary. The extent of the QTL × environment interaction is often unknown because the mapping studies have been limited to only a few years or locations (Wang et al. 2007).

Genome structure

Markers can be identified and developed using populations where parents do not represent adapted germplasm, such as diploid or tetraploid wheat species. In such cases, the known polymorphism can be of little practical value despite it is transferred to wheat though interspecific hybridization.

Many key traits for wheat improvement present in alien segments have been transferred to wheat. The alien segments however are often large and can carry undesired characters in addition to the favorable trait, such as the high grain protein content gene *Gpc-B1* transferred from *Triticum turgidum ssp. dicoccoides*, which is negatively correlated with grain yield (Uauy et al. 2006). Recombination within these alien segments is very low and advanced approaches that reduce the large linkage blocks are needed.

Cost and logistic of MAS

Typically breeding programs grow hundreds of populations and many thousands of individual plants. Given the extent and the complexity of selection required in breeding programs one can easily appreciate the usefulness of new tools that may assist breeders in plant selection. The scale of the breeding programs, however, also underlines the challenges of incorporating MAS. A close relationship between breeders and molecular biologists support the level of integration of MAS. For example, it is vital that the robustness and reliability of the markers available for genes or alleles of interest are evaluated before considering their routine application. Lack of confidence in published information is cited as one of the reasons that limit the use of markers in practical plant breeding (Kuchel et al. 2003). Leaf tissue collected in the field has also to be brought to the laboratory in time to provide the marker data to the breeders prior to selection or harvest.

Despite the recent shift to SNP based platforms, e.g. KASP in wheat ((http://www.cerealsdb.uk.net/) at least at CIMMYT, the cost of marker assays remains the rate limiting factor for the adoption of MAS. Taking advantage of that present specialized genotype service providers can evade the requirements of large capital investments for the acquisition of equipment and the regular labor expenses, drastic reductions in assay costs are, however, difficult to achieve unless very large numbers of marker assays are deployed. For service providers and genotyping platforms, the cost per marker assay is associated with the sample volume. With its current sample volume CIMMYT is reaching a minimum of 0.2 USD cost per SNP assay, while one-tenth of the cost would be desired. Next generation sequencing is likely to provide future technologies that can currently combine single marker assays with a number of background marker for a broad use in forward and background selection at very low cost that will make MAS further attractive.

References

Abalo, G., P. Tongoona, J. Derera and R. Edema. 2009. A comparative analysis of conventional and marker-assisted selection methods in breeding maize streak virus resistance in maize. Crop Science 49: 509-520

Anderson, J.A., S. Chao and S. Liu. 2007. Molecular breeding using a major QTL for Fusarium head blight resistance in wheat. Crop Science 47 (Supplement_3), S-112

Becker, H., 1993. Pflanzenzüchtung. Ulmerverlag, 327 S.

Bonnett, D.G., G.J. Rebetzke and W. Spielmeyer. 2005. Strategies for efficient implementation of molecular markers in wheat breeding. Molecular Breeding 15: 75-85

Choulet, F., M. Caccamo, J. Wright, M. Alaux, H. Šimková, J. Šafář, P. Leroy, J. Dolezel, J. Rogers, K. Eversole and C. Feuillet. 2014. The wheat black jack: Advances towards sequencing the 21 chromosomes of bread wheat. In: R. Tuberosa et al. (eds.), Genomics of Plant Genetic Resources, Springer Science +Business Media Dordrecht 2014

Davies, J., W. Berzonsky and G. Leach. 2006. A comparison of marker-assisted and phenotypic selection for high grain protein content in spring wheat. Euphytica 152: 117-134

Eathington, S.R., T.M. Crosbie, M.D. Edwards, R.S. Reiter and J.K. Bull. 2007. Molecular markers in a commercial breeding program. Crop Science 47, Supplement 3: S154

Frisch, M. and A.E. Melchinger. 2005. Selection theory for marker-assisted backcrossing. Genetics 170:909-917

Frisch, M., M. Bohn and A.E. Melchinger.1999. Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. Crop Science 39:967–975

Hanson, W.D. 1959. Minimum family sizes for the planning of genetic experiments. Agronomy Journal 51:711-715

Holland, J. 2001. Epistasis and plant breeding. Plant Breeding Reviews. 21: 27-92.

Hospital, F., C. Chevalet and P Mulsant. 1992. Using markers in gene introgression breeding programs. Genetics 132:1199–1210

Kuchel, H., R. Fox, J. Reinheimer, L. Mosionek, N. Willey, H. Bariana and S. Jefferies. 2007. The successful application of a marker-assisted wheat breeding strategy. Molecular Breeding 20:295–308

Longin, C.F. and J.C. Reif. 2014. Redesigning the exploitation of wheat genetic resources. Trends in Plant Science 19, (10): 631–636

Miedaner, T., F. Wilde, V. Korzun, E. Ebmeyer, M. Schmolke, L. Hartl and C.C. Schoen. 2009. Marker selection for fusarium head blight resistance based on quantitative trait loci (*qtl*) from two European sources compared to phenotypic selection in winter wheat. Euphytica 166:219-227

Moreau, L., A. Charcosset and A. Gallais. 2004. Experimental evaluation of several cycles of marker-assisted selection in maize. Euphytica 137: 111-118

Prigge, V., H.P. Maurer, D.J. Mackhill, A.E. Melchinger and M. Frisch. 2008. Comparison of the observed with the simulated distribution of the parental genome contribution in two marker-assisted backcross programs in rice. Theoretical and Applied Genetics 116:739-744

Prigge, V., A.E. Melchinger, B.S. Dhillon and M. Frisch. 2009. Efficiency Gain of marker-assisted backcrossing by sequentially increasing marker density over generations. Theoretical and Applied Genetics 119:23-32

Ribaut, J-M. and D. Hoisington. 1998. Marker-assisted Selection: new Tools and Strategies. Trends in Plant Science 3:236–239

Tanksley, S.D., N.D. Young, A.H. Patterson and M.W. Bonierbale. 1989. RFLP mapping in plant breeding. New tools for an old science. Bio/Technology 7:257-263

Uauy, C., J.C. Brevis and J. Dubcovsky. 2006. The high grain protein content gene *Gpc-B1* accelerates senescence and has pleiotropic effects on protein content in wheat. Journal of Experimental Botany 57(11):2785-2794

Wang, J, S.C. Chapman, D.G. Bonnett, G.J. Rebetzke and J. Crouch. 2007. Application of population genetic theory and simulation models to efficiently pyramid multiple gene via marker-assisted selection. Crop Science 47:582-590

Wilde, F., V. Korzun, E. Ebmeyer, H. Geiger and T. Miedaner. 2007. Comparison of phenotypic and marker-based selection for fusarium head blight resistance and don content in spring wheat. Molecular Breeding 19: 357-370

Chapter 4:
Marker-Assisted Selection for Quality Traits

Carlos Guzmán[1], Susanne Dreisigacker[1], Juan B. Alvarez[2], Roberto Javier Peña[1]

[1] CIMMYT Int., Apdo. Postal 6-641,06600 Mexico, DF, Mexico

[2] Departamento de Genética, E.T.S.I.A.M., Universidad de Córdoba, CeiA3, 14071, Córdoba, Spain

Introduction

The biggest challenge for wheat breeders is to enhance concomitantly grain yield, biotic/abiotic stress resistance, and grain quality for diverse end products. New wheat varieties should meet specific grain quality requirements to satisfy the increasing demand for processed conventional and novel wheat-based foods. Grain quality is a variable concept, and its meaning depends on the type of flour to be produced (whole meal flour, refined flour, semolina, etc.), end product to manufacture (bread, biscuit, pasta, etc.), the process used to produce it (handmade, semi- mechanized, mechanized, etc.), and the consumer's preferences. In defining quality for any given end use, processing performance and end-product properties have to be considered.

Processing quality and end product properties are determined by a set of complex traits, the most important being the endosperm texture or grain hardness, the content and composition of storage proteins (mainly glutenins), the composition of starch and non-starch polysaccharides, and, for some specific products, the color of the flour/semolina. The high variability in grain quality traits existing in wheat has led to the creation of thousands of varieties possessing many different grain composition combinations, allowing wheat to be used to manufacture many different types of food products.

Although the main quality traits are influenced by the environment and cropping practices, their expression is mainly controlled by qualitative genes and their allelic variations. The good association between genotype and phenotype for main grain quality parameters has made the use of these parameters possible to estimate the presence/absence of quality-related loci. In addition, this genotypic-phenotypic association has contributed significantly to the development and validation of several functional or allele-specific markers, derived from polymorphic sites within the genes that are directly associated with phenotypic variations. This kind of markers are developed from single nucleotide polymorphism (SNPs) or insertion/deletions (InDels) between different alleles, requiring the gene sequences of functional motifs associated with plant phenotypes. Until now, functional markers for almost all important high-molecular-weight (HMWGs) and low-molecular-weight (LMWGs) glutenin subunits associated with the gluten properties are available, as well as for genes related to flour/semolina color (phytoene synthase, lipoxygenase and polyphenol oxidase), grain hardness (puroindolines), protein content, and starch properties (waxy genes) (Liu et al. 2012). Due to the complexity of analyzing some quantitative and qualitative traits using conventional non molecular tools, these molecular markers have received great attention and are being implemented in some breeding programs. Thus, MAS is becoming a reality in breeding for main grain quality traits. MAS is now considered a useful tool, efficiently complementing traditional selection, increasing the understanding of phenotypic characteristics and their genetic control to design better crossing schemes, and therefore breeding strategies. Molecular markers can be used to characterize and select parental genotypes possessing specific genes/alleles to perform crosses conferring desirable quality traits, as well as to select for desirable quality traits in segregating or early advanced stages. This is particularly useful when fixing quality traits under simple genetic control is desirable at early stages of breeding, or to select for traits difficult to measure using phenotypic assays requiring large amount of grain sample.

Grain hardness

Grain hardness, or endosperm texture, can be considered the most important single factor determining the general end use (bread, biscuit, pasta) of a wheat cultivar (Morris 2002): hard wheat is for bread while soft wheat is for biscuits, and the very hard, vitreous grain of durum wheat, is suitable for pasta. Actually, common (hexaploid) wheat is generally marketed according to grain hardness class, as soft or hard, while durum (tetraploid wheat) is

marketed as a different class compared to common wheat. Grain hardness is important in the flour milling process. Flour of different extraction rates are obtained from common wheat flour milling while semolina (coarse flour particles) is obtained from durum wheat only. The importance of grain hardness resides in its influence on the level of damaged starch resulting during flour milling; the harder the grain, the higher the level of damaged starch in the flour, and the higher the water hydration (water absorption) capacity of the flour (Posner 2000). While a low level of damaged starch (and low flour water absorption) is required for the manufacture of biscuits, the manufacture of different types of flat and leavened breads requires different levels of water absorption capacity.

Grain hardness is controlled in common wheat by *Pina-D1* and *Pinb-D1*, two intronless small genes (444 bp of coding region) located at the short arm of chromosome 5D (Morris 2002). These genes codified for two proteins named puroindolines (PINA and PINB), associated with the membrane surrounding starch granules, and that have direct role in the definition of grain hardness (see Morris and Bhave 2008 for a complete review). When wild forms of both proteins are present the grain texture is always soft. However if one of the proteins is missing or has a modified amino acid sequence (the change of just one amino acid is in several cases enough) the texture will be hard or semi-hard. In the case of the tetraploid durum wheat, the texture is very hard due to the lack of both genes/proteins because of the absence of D genome, and the complete deletion of these genes in the A and B genome during the synthesis event of the tetraploid wheat (Chantret et al. 2005).

Up to now, four and 14 different alleles leading to hard texture have been identified for *Pina-D1* and *Pinb-D1*, respectively (McIntosh et al. 2014). All these alleles, except *Pina-D1b*, are characterized by the presence of one SNP in the coding region that either changes the ORF leading to the appearance of a premature stop codon or leading to the change of one amino acid in the protein sequence that modifies the functionality of the protein. Simple STS or CAPS molecular markers have been designed for all these mutations and are of public domain.

The most common *Pin* alleles causing hard texture are *Pina-D1b* and *Pinb-D1b*. The first one, predominant in CIMMYT germplasm (Lillemo et al. 2006), is characterized by the almost complete deletion of the *Pina-D1* gene. This is easily detectable in a PCR with any kind of primers that are designed to amplify this gene, as the ones presented in Gautier et al. (1994), which show an amplicon of 349 bp when the genotype carry any *Pina-D1* allele except *Pina-D1b*, which does not show any amplification product. It is important to remark that the presence of PCR product using these primers does not mean that the genotype is carrying the wild allele *Pina-D1a*: it could carry any other *Pina-D1* allele as the l, m or n that only have one SNP in their sequence, but that also lead to hard texture. The other common allele, *Pinb-D1b*, is characterized by one SNP mutation that changes Gly46 → Ser, which apparently is sufficient to disrupt the softening effect of PINB. STS primers have been designed to detect this mutation, with one of the primers annealing in the area of the SNP.

Although grain hardness is easily measured in the laboratory (by near infrared spectroscopy, or NIRS; single kernel characterization system, or SKCS; or other more traditional methods such as Particle Size Index) and shows only a small environmental influence, the molecular markers to determine *Pin-D1* genes are useful in the breeding process, because the different allele combinations for both genes are associated with different hardness levels (Martin et al. 2001; Takata et al. 2010), which have different effects on end-use quality (Eagles et al. 2006). For example, *Pina-D1b* is associated with a harder texture than *Pinb-D1b* and in some cases with higher flour yield. Therefore, knowing the *Pin* genotype can help the breeder to develop wheat with a specific and desirable texture for a specific end product. Besides, with the use of the markers, the selection of offspring homozygous for *Pin* alleles can be started at segregation stages, when the available seed is not enough to measure hardness by non-molecular means.

At CIMMYT, common wheat parental lines are analyzed with respect to *Pina-D1* and *Pinb-D1* with developed SNP markers (Appendix 2). Currently the predominance of *Pina-D1b* in CIMMYT germplasm is very high (more than 90%); therefore, the markers

are being used to identify lines carrying *Pinb-D1b* and are used to try to enhance variability for both loci. In addition, SNP markers for other less frequently found alleles that could lead to a different texture, such as *Pinb-D1c, Pinb-D1d* or *Pina-D1m*, are being developed to screen for them in different wheat collections and introduced in our germplasm. The specific effect of each allele will be tested soon in a study with near-isogenic lines (NILs).

Gluten composition

When wheat flour is mixed with water to form dough, the storage proteins of the grain are aggregated in gluten, a visco-elastic protein network. Gluten confers the dough its viscoelastic properties (elasticity or strength and extensibility), which are the main functional properties defining the processing and end-use quality of any wheat variety. The processing of most wheat-based products requires gluten strength and extensibility in a greater or lesser extent. Gluten elasticity or strength requirements depend on the processing conditions (higher strength is required in mechanized production than in hand-made production) and the end product to be manufactured (bread-making and noodle-making require medium to strong gluten; pasta-making requires medium to strong and inextensible gluten; while biscuit-making requires a weak gluten type).

Gluten is composed of a large number of proteins, mainly glutenins and gliadins. Glutenins contribute more to gluten strength while gliadins contribute to extensibility and viscosity. It is not possible to understand dough/gluten viscoelastic properties studying both kinds of proteins independently but coexisting together in the intricate gluten protein network and in the complex dough system. Among the glutenins there are high-molecular-weight glutenin subunits (HMWGs), codified by the *Glu-A1*, *Glu-B1* and *Glu-D1* loci (located at the long arm of chromosomes 1A, 1B and 1D, respectively), and low-molecular-weight glutenin subunits (LMWGs), codified by the *Glu-A3, Glu-B3* and *Glu-D3* loci (located at the short arms of chromosomes 1A, 1B and 1D, respectively). The HMWGs loci are formed by two linked genes that codified two subunits (x + y), composing a single allele. The alleles are named usually with two numbers, each one identifying one subunit. Different alleles for each of the glutenin genes have been detected and classified mainly by SDS-PAGE protein electrophoresis. This system, when the protein extraction process is done properly and the separation conditions are optimum in the gels, allows the identification of the six loci aforementioned (*Glu-A1, Glu-B1, Glu-D1, Glu-A3, Glu-B3* and *GluD-3*) and classification of the alleles with only two SDS-PAGE gels. Most of these alleles have been associated with high or poor quality; the clearest example is the association of the *Glu-D1d* allele (subunits 5+10) with higher elasticity and extensibility than that conferred by the allele *Glu-D1a* (subunits 2+12) (Payne et al. 1987).

The six *Glu-1* and *Glu-3* loci already have been characterized at a molecular level and molecular markers are available for almost all the alleles (see Gale et al. 2005 and Rasheed et al. 2013 for a review of markers available). Most of them are STS type, although at CIMMYT several SNPs markers have been developed and validated, as the one for *Glu-D1d* (subunits 5 +10) (Appendix 1 and 2). The use of the DNA molecular markers is recommended when the facilities and experience to perform grain protein extraction and electrophoretic separation are limited. Up to now, SDS-PAGE electrophoresis is more time-efficient and has a lower cost than the use of several individual markers, as low cost multiplex assays for all observed alleles still do not exist. However, SDS-PAGE also requires highly experienced staff to perform identification of the LMW alleles. At CIMMYT, parental lines are analyzed for the six glutenin loci by SDS-PAGE electrophoresis. The lines that carry SDS-PAGE subunit bands are difficult to differentiate, mainly due to band overlapping. These are analyzed by the specific molecular markers. Those are usually the lines carrying the subunit 7 overexpressed (Bx-7^{OE}), which has been shown to increase the concentration of this subunit, which is difficult to detect in an SDS-PAGE gel. Other alleles difficult to identify by SDS-PAGE are *Glu-A3f* and e, as well as *Glu-B3f* and g. In those cases, the power resolution of the molecular markers is very useful to have a concluding result.

Starch properties

Starch (the major component of the wheat grain) is composed mainly of amylose and amylopectin. The ratio of both macromolecules has a significant impact on the main functional properties of starch (pasting and gelation) and dough characteristics such as

viscosity, extensibility and expansion, particularly at the oven stage in bread-making, when hydration of macromolecules change, mainly due to denaturation. In common wheat, the amount of amylose and amylopectin is 25-28% and 75-72%, respectively. Variations in this ratio will lead to positive or negative changes in the water absorption of the dough and in the end product quality (firmness, texture, freshness retention). For noodles, a lower amylose content is desirable to get better volume, firmness and texture while for the bread crumb a high amylose content is preferable to achieve uniform structure and soft texture. In regard to nutrition and health, a high amylose content is related to high resistant starch concentration, which acts as pseudo-fiber in the human intestine during digestion and which is associated with a healthier diet due to its lower glycemic index and because it increases satiation with less ingestion of food.

The amylose/amylopectin ratio is controlled by the enzymes responsible for their synthesis (Morell et al. 2001); the environment has little effect on this trait. The amylopectin synthesis is complex and is carried out by different starch synthases (SGP-1, SGP-2 and SGP-3), branching and debranching enzymes of the grain. The case of the amylose is simpler and it is exclusively synthetized by the Granule-Bound Starch Synthase (GBSS) I, commonly named waxy protein. In common wheat, three different waxy proteins are present and controlled by the three *Wx* loci (*Wx-A1*, *Wx-B1* and *Wx-D1*) located at chromosomes 7AS, 4AL and 7DS, respectively (Yamamori et al. 1994). These proteins have shown polymorphism, denoting the existence of null alleles (absence of the protein) that lead to the reduction of the amylose content, especially when Wx-B1 protein is lacking because its major proportion in relation to the other waxy proteins (Wx-A1 and Wx-D1). The combination, using classical breeding, in the same wheat of the three null waxy proteins generated is called waxy wheat, which has 0% of amylose content and has properties very different to standard wheat (Nakamura et al. 1995).

The three *Wx* loci have been well characterized at the molecular level (Murai et al. 1999). The coding region of these genes is composed of 11 exons and 10 introns of a total size between 2,781 and 2,862 bp. The sequences of the null alleles known as *Wx-A1b*, *Wx-B1b* and *Wx-D1b* already have been described and different molecular markers have been validated to screen for these mutations (Liu et al. 2005; McLauchlan et al. 2001; Nakamura et al. 2002), especially for the *Wx-B1b* allele that has greater impact than the others (Saito et al. 2009). All of them are STS markers. At CIMMYT, there is an increasing interest in determining the variation in starch properties of modern germplasm, as well as diverse genetic resources, such as wheat landraces of different origins. Although starch composition is not considered a major factor in defining processing quality, parental lines have been analyzed for the presence of different null alleles, finding significant presence of the *Wx-B1b* allele (11% of the lines) but not for the others. The use of these markers, especially when breeding, is oriented to improve very specific products like biscuits, flat unleavened breads, or noodles, with very specific quality requirements, is useful because colorimetric assays to determine amylose content or electrophoretic separation or waxy proteins are time consuming processes and interpretation of the results is not always easy.

Flour and semolina color

Flour color is an important trait in the assessment of flour quality, especially in noodles and other related products. The enzyme polyphenol-oxidase (PPO) has been found to be involved in undesirable time-dependent browning of noodles, flat breads, and steam bread. The PPO activity, although it has been shown to be largely dependent on the environment, is variable among different genotypes. Six loci, two per each genome (*Ppo-1* and *Ppo-2*), have been characterized at a molecular level and alleles associated with high or low PPO activity are already available.

Semolina yellow color has become an important quality trait for durum wheat end products. The yellow color is due to the presence of yellow carotenoid pigment. Genetic variation and high heritability has been reported for this trait, although environmental effects and the processing conditions will also influence the final result in the end product. The final yellow pigment concentrations in the end product are affected by the carotenoids synthesis in the grain, catalyzed by phytoene synthase (PSY), but also by carotenoid degradation during grain and semolina storage, and pasta processing. The latter is a consequence of the activity of oxidative enzymes

denoting the lypoxygenases family (LOXs). Several alleles and their corresponding molecular markers have been described for the genes controlling PSY in durum wheat, *Psy-A1* and *Psy-B1*, some of them associated with higher yellow pigment concentration (Ficco et al. 2014). On the other hand, three LOX isoforms have been described, each one controlled by a *Lpx-1* locus; the LOX-1 is the one with major role in oxidation of carotenoid pigments during pasta processing because of its higher concentration. The c allele of one of the copies of the *Lpx-1* located on chromosome 4B (*Lpx-B1.1*), containing a large central deletion that probably leads to the production of a non-functional enzyme, is associated to low LOX activity and therefore to a small decrease in yellow pigment during processing. PPO activity, as above mentioned in flour from common wheat, sometimes appears in durum wheat, causing an undesirable brown color.

At CIMMYT durum wheat recombinant lines carrying the Lr19+Yp genes from *Agropyron elongatum* have been actively introgressed into various CIMMYT elite durum wheat lines via F_2 enrichment. "Yp" is designated by two loci for increased yellow pigment closely mapped distal to Lr19 on the distal region of chromosome arm 7EL (Ceoloni et al. 2000). For MAS the marker published by Zhang et al. (2008) was applied (Appendix 1). While the tight linkage results are unfavorable in common wheat, the single transfer produced beneficial effects for both the leaf rust resistant trait and the yellow color for semolina and pasta products. The validation and implementation of markers for LOX and PPO activity have been initiated and are very interesting because those are physiological-biochemical traits that cannot be easily evaluated based on phenotype. The use of these markers would allow the selection of wheat progenies in the early generations and would greatly improve selection efficiency for color.

References

Ceoloni C., P. Forte, Y.M. Cia, M. Nenno, A. Bitti, P. De Vita, M.G. and D'Egidio. 2000. Chromosomally engineered durum wheat: the potential of alien gene introgressions affecting disease resistance and quality. In: Royo C, Nachit MM, Di Fonzo N, Araus JL (*eds*) Durum wheat improvement in the Mediterranean region: new challenges, vol 40. International centre for advanced Mediterranean agronomic studies, Zaragoza, Spain, pp 363–371

Chantret N., J. Salse, F. Sabot, S. Rahman and A. Bellec, et al. 2005. Molecular basis of evolutionary events that shaped the Hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). Plant Cell 17: 1033-1045

Eagles H., K. Cane and R. Eastwood. 2006. Contributions of glutenin and puroindoline genes to grain quality traits in southern Australian wheat breeding programs. Crop Pasture Science 57: 179-186

Ficco D.B.M., A.M. Mastrangelo, D. Trono, G.M. Borelli, P. De Vita, C. Fares, R. Beleggia, C. Platani and R. Papa. 2014. The colours of durum wheat: a review. Crop Pasture Science 65: 1-15

Gale K.R. 2005. Diagnostic DNA markers for quality traits in wheat. Journal of Cereal Science 41: 181-192

Gautier M.F., M.E. Aleman, A. Guirao, D. Marion and P. Joudrier. 1994. *Triticum aestivum* puroindolines, two basic cystine-rich seed proteins: cDNA sequence analysis and developmental gene expression. Plant Molecular Biology 25: 43-57

Lillemo M., F. Chen, X. Xia, M. William, R.J. Peña, R. Trethowan and Z. He. 2006. Puroindoline grain hardness alleles in CIMMYT bread wheat germplasm. Journal of Cereal Science 44: 86-92

Liu L., Z. He, J. Yan, Y. Zhang, X. Xia and R.J. Peña. 2005. Allelic variation at the Glu-1 and Glu-3 loci, presence of the 1B.1R translocation, and their effects on mixographic properties in Chinese bread wheats. Euphytica 142: 197-204

Liu Y., Z. He, R. Appels and X. Xia. 2012. Functional markers in wheat: current status and future prospects. Theoretical and Applied Genetics 125: 1-10

Martin J., R. Frohberg, C. Morris. 2001. Milling and bread baking traits associated with puroindoline sequence type in hard red spring wheat. Crop Science 41: 228-234

McIntosh R.A., Y. Yamazaki, J. Dubcovsky, W.J. Rogers, C. Morris, R. Appels and X.C. Xia. 2014. Catalogue of gene symbols for wheat. [http://www.shigen.nig.ac.jp/wheat/komugi/genes/macgene/2013/GeneSymbol.pdf] Date of last successful access 23 May 2014

McLauchlan A., F.C. Ogbonnaya, B. Hollingsworth, M. Carter, K.R. Gale, R.J. Henry, T.A. Holton, M.K. Morell, L.R. Rampling, P.J. Sharp, M.R. Shariflou, M.G.K. Jones and R. Appels. 2001. Development of robust PCR-based DNA markers for each homoeo-allele of granule-bound starch synthase and their application in wheat breeding programs. Australian Journal of Agricultural Research 52: 1409-1416

Morell M.K., S. Rahman, A. Regina, R. Appels and Z. Li. 2001. Wheat starch biosynthesis. Euphytica 119: 55–58

Morris C.F. 2002. Puroindolines: the molecular genetic basis of wheat grain hardness. Plant Molecular Biology 48: 633-647

Morris C.F. and M. Bhave. 2008. Reconciliation of D-genome puroindoline allele designations with current DNA sequence data. Journal of Cereal Science 48: 277-287

Murai J., T. Taira and D. Ohta. 1999. Isolation and characterization of the three Waxy genes encoding the granule-bound starch synthase in hexaploid wheat. Gene 234: 71-9

Nakamura T., M. Yamamori and H. Hirano. 1995. Production of waxy (*amylose-free*) wheats. Molecular General Genetics 248: 253-9

Nakamura T., P. Vrinten, M. Saito and M. Konda. 2002. Rapid classification of partial waxy wheats using PCR-based markers. Genome 45: 1150-6

Payne P.I., M.A. Nightingale, A.F. Krattiger and L.M. Holt. 1987. The relationship between HMW glutenin composition and the bread-making quality of British-grown wheat varieties. Journal of the Science of Food and Agriculture 40: 51-65

Posner E.S. 2000. Wheat. In: Kulp K, Ponte JGJ, editors. Handbook of cereal science and technology. New York, NY, USA: Marcel Dekker, Inc. pp 1–29

Rasheed A., T. Mahmood, A. Gul-Kazi and A. Mujeeb-Kazi. 2013. An overview of omics for wheat grain quality improvement. In: Hakeem KR, Ahmad P, Ozturk M (*eds*) Crop Improvement. Springer US, pp 307-344

Saito M., P. Vrinten, G. Ishikawa, R. Graybosch and T. Nakamura. 2009. A novel codominant marker for selection of the null *Wx-B1* allele in wheat breeding programs. Molecular Breeding 23: 209-217

Takata K., T.M. Ikeda, M. Yanaka, H. Matsunaka, M. Seki, N. Ishikawa and H. Yamauchi. 2010. Comparison of five puroindoline alleles on grain hardness and flour properties using near isogenic wheat lines. Breeding Science 60: 228-232

Yamamori M., T. Nakamura, T.R. Endo and T. Nagamine. 1994. Waxy protein deficiency and chromosomal location of coding genes in common wheat. Theoretical and Applied Genetics 89: 179-184

Zhang W. and J. Dubcovsky. 2008. Association between allelic variation at the Phytoene synthase 1 gene and yellow pigment content in the wheat grain. Theoretical and Applied Genetics 116: 635-45

Chapter 5:
Marker Assistant Selection for Rust Resistance in Wheat

Caixia Lan, Bhoja R. Basnet and Susanne Dreisigacker
CIMMYT Int., Apdo. Postal 6-641,06600 Mexico, DF, Mexico

Wheat leaf brown rust, stripe yellow rust, and stem black rust caused by *Puccinia triticina Pt, P. striiformis* f. sp. *tritici Pst*, and *P. graminis* f. sp. *tritici Pgt*, respectively, are the most widely distributed wheat *Triticum aestivum* L. diseases in the world. They can cause yield losses of up to 40% of the wheat production area in countries such as Bangladesh, China, India, Mexico and Pakistan Dubin and Brennan 2009, and more than 20 stripe rust epidemics have been documented worldwide Wellings 2011. Stem rust has historically been a big threat to wheat production Saari and Prescott 1985. A new *Pgt* race, TTKSK commonly referred to as Ug99, detected in Uganda in 1998, had virulence to most of the widely deployed specific resistance genes and was seen as threat to global food security Li et al. 2014. There are many ways to manage these diseases; however, development and using resistant cultivars is the most efficient method to control them.

Rust resistance genes in wheat

Two types of rust resistance genes are often defined in wheat. Race-specific resistance genes usually confer protection throughout the growth cycle and therefore resistance conferred by them is also called all-stage resistance Chen 2013. These resistance genes cause various degrees of hypersensitive reactions in the host if the pathogen possesses corresponding avirulence genes Flor 1942. In contrast, race non-specific minor genes that confer adult plant resistance APR are usually present together with other similar genes and therefore associated with quantitative inheritance Das et al. 1992; Johnson and Law 1973. Most cultivars with multiple genes for APR are susceptible at the seedling stage but later display resistance to a number of races Bjarko and Line 1988. For the last decade, identification of new sources of rust resistance and molecular marker based gene discovery has gained high priority in the wheat scientific community. As a result, several new sources of resistance have been identified in global wheat germplasm including CIMMYT germplasm by QTL mapping and association mapping approaches. Several race-specific and race non-specific genes have been tagged with molecular markers. As one of the important members in the global partnership of the Borlaug Global Rust Initiative www.globalrust.org, CIMMYT's global wheat program is playing an active role to fulfill different objectives of the Durable Rust Resistance in Wheat DRRW project. Close collaboration with CSIRO, Australia through the Grains Research & Development Corporation GRDC-funded projects identifying and molecular mapping of novel resistance genes to all three rusts and developing elite wheat germplasm with durable rust resistance is a long-term objective.

Available molecular markers of rusts resistance gene in CIMMYT

Although many advanced wheat lines/cultivars have been bred by traditional breeding methods, MAS provides opportunities for enhancing the response from selection, due to high precision and low-cost using available molecular markers of rust resistance genes at both seedling and adult plant stages. Table 1, Table 2 and Table 3 provide an overview of available markers for seedling resistance genes, APR genes with pleiotropic effect, and APR genes, respectively, present in CIMMYT wheat germplasm. Additional information regarding the most actively used markers can be seen in Table 4. Single gene resistance can usually be selected phenotypically in the greenhouse. For quantitative disease resistance, MAS can be very useful for pyramiding more small effect individual QTL in elite wheat lines. In addition, with more and more newly identified pleotropic resistance genes/QTL, conferring resistance to multiple taxa of pathogens, provides additional perspectives for MAS in wheat rusts.

Marker assistant selection in wheat breeding program in CIMMYT

During the last two to three years, CIMMYT has started to routinely evaluate its bread wheat screening nurseries e.g. IWSN with markers associated to rust resistance genes. In addition, early generation MAS is applied on a project basis in the bread and durum wheat programs. Table 5 shows the results of a F_2-

enrichment strategy for Ug99 stem rust resistance. The in total 305 lines in the 31st Semi-Arid Wheat Yield Screening Nursery SAWYSN are listed according the environment they have been selected using conventional selection or MAS. The lines within each environment and selection scheme were subsequently classified according their disease rating. Forty lines within the SAWYSN were derived from

Table 1. Available molecular markers for seedling resistance genes to rusts in the CIMMYT wheat breeding program.

Genes	Flanking markers	Marker type	Resistance source	Reference
Lr16	*gwm210, wmc661*	SSR	Francolin#1	Lan et al. 2014
Lr19	*wMAS000032, PSY-E, wmc221*	SNP, STS, SSR	Agatha, Misr#1	Zhang and Dubkovsky 2008, Dreisigacker pers. comm.
Lr21	*D14*	STS		Talbert et al. 1994
Lr42	*cfd15, wmc432*	SSR	Quaiu#3	Basnet et al. 2014a
Lr47	*PS10R/ PS10L, PS10R/PS10L2, CIMwMAS0055*	STS, SNP		Helguera et al. 2000
Lr51	*S30-13L/AGA7-759R*	STS		Helguera et al. 2005
Yr17	*VENTRIUp/LN2, URIC/LN2, CIMwMAS004*	STS, SNP	Milan	Helguera et al. 2003
Yr15	*Yr15-R5, Yr15-R8*	SNP	Blanca Grande 515	Ramirez-Gonzalez et al. 2015
Yr24/26	*We173*	STS, SSR	Chuanmai 42	Wang et al. 2008
Yr41	*gwm410, gwm374*	SSR	Chuannong 19	Luo et al. 2008
Yr43	*wgp110, wgp103, barc139*	RGA, SSR	ID0377S	Cheng and Chen 2010
Yr44	*pWB5/N1R1,wgp100, gwm501*	RGA, SSR	Zak	Cheng and Chen 2010
Yr50	*gwm540, barc1096, wmc47, wmc310*	SSR	CH223	Liu et al. 2013
Yr60	*wmc776*	SSR	Lal Bahadur	Herrera-Foessel, pers comm
YrF	*gwm374, wmc474*	SSR	Francolin#1	Lan et al. 2014
YrSuj/Yr67	*cfa2040, wmc526*	SSR	Sujata, C591	Lan CX, pers comm
Sr1BL.1RS	*SCM9, wMAS000011*	STS, SNP	Kavkaz, Veery 'S'	Weng et al. 2007
Sr13	*barc104, dupw167, CD926040, BE471213*	SSR	Kofa, Kronos	Simons et al. 2011
Sr22	*wmc633, cfa2123*	SSR	Sr22Tb, Steinwedel	Olson et al. 2010
Sr23	*gwm210*	SSR	AC Domain	McCartney et al. 2005
Sr25	*wMAS000032, wmc221*	SSR	Agatha, Misr#1	Zhang and Dubkovsky 2008
Sr26	*Sr26#43, BE518379*	STS	WA1	Mago et al. 2005
Sr32	*csSr32#1, csSr32#2*	STS		Mago et al. 2013
Sr33	*barc152, cfd15, BE405778, BE499711*	SSR	RL5288	Sambasivam et al. 2008; Periyannan et al. 2013
Sr35	*cfa2170, cfa2076, wmc169, wmc559*	SSR	G2919	Zhang et al. 2010
Sr38	*CIMwMAS0004, Ventriup/LN2*	SNP, STS	VPM1	Helguera et al. 2003;
Sr42	*barc183, FSD_RSA*	SSR	Norin 40	Ghazvini et al. 2012
Sr-6DS	*gpw5182, cfd49*	SSR	Niini, Coni, Blouk	Lopez-Vera et al. 2014
SrND643	*gwm350, Xwmc219*	SSR	ND643	Basnet et al. 2014b
SrHuw234	*wmc332*	SSR	Huwa	Lopez-Vera et al. 2014
SrYanac	*barc200*	SSR	Yaye	Lopez-Vera et al. 2014

Table 2. Available molecular markers for pleotropic adult plant resistance PAPR genes to rusts in the CIMMYT wheat breeding program.

Genes	Flanking markers	Marker type	Resistance source	Reference
Lr34/Yr18/Pm38/Sr57	*wMAS000003, wMAS000004*	STS, SNP	Parula	Lagudah et al. 2009, http://www.cerealsdb.uk.net/
Lr46/Yr29/Pm39/Sr58	*csLv46, csLV46G22*	CAPS	Pavon 76, Kenya Kongoni	Lagudah ES pers comm
Lr67/Yr46/Pm46/Sr55	*csSNP856*	SNP	RL6077	Forrest et al. 2014
Sr2/Yr30	*csSr2, wMAS000005*	CAPS	Pavon76	Mago et al. 2011, http://www.cerealsdb.uk.net/
Sr2/Yr30/Lr27	*gwm533*	SSR	Quaiu#3	Spielmayr et al. 2003

Table 3. Available molecular markers for adult plant resistance APR genes to rusts in the CIMMYT wheat breeding program.

Genes	Flanking markers	Marker type	Resistance source	Reference
Lr68	*CSGS, cs7BLNLRR,CIMwMAS0056*	CAPS , SNP	Parula	Herrera-Foessel et al. 2012, Dreisigacker pers. comm.
Yr54	*gwm301*	SSR	Quaiu#3	Basnet et al. 2014
Yr36	*Gpc-B1, wMAS000017*	STS, SNP	Glupro	Uauy et al. 2005, http://www.cerealsdb.uk.net/
Yr39	*wgp36,wgp45,gwm18, gwm11*	RGA, SSR	Alpowa	Lin and Chen 2007
Yr52	*barc182, wgp5258*	RGA, SSR	PI 183527	Ren et al. 2012
Yr59	*wgp5175, barc32, barc182*	RGA, SSR	PI178759,PI 660061	Chen XM pers comm
Sr56	*Sun209, Sun320*	SSR	Arina	Bansal et al. 2014

Table 4. Additional information of the most actively used markers of APR genes to rusts.

Genes	Flanking markers	Forward (STS) or HEX (SNP) primer 5'-3'	Reverse (STS) or FAM (SNP) primer 5'-3'	Reverse (SNP) common primer	Expected product	Reference
Lr34/Yr18/ Pm38/Sr57	*cssfr1*	TGATGAAACCAGTTTTTTTTCTA	GCCATTTAACATAATCATGA TGGA		517 bp	Lagudah et al. 2009
	cssfr2	TGATGAAACCAGTTTTTTTTCTA	ATGCCATTTAACATAATCATGAA		523 bp	Lagudah et al. 2009
	cssfr5	TGAGGCACTCTTTCCTGTA CAAAG	ATTCAATGAGCAATGGTTATC		751 bp	Lagudah et al. 2009
	wMAS000003	GGTATGCCATTTAACATAATC ATGAA	GGTATGCCATTTAACATAATC ATGAT	TACTATATGGGAGCAT TATTTTTTTCC		http://www.cerealsdb.uk.net/
	wMAS000004	TGTAATGTATCGTGAGAGATTT GCAG	ATTGTAATGTATCGTGAGAGAT TTGCAT	GATCATTATCTGACCTGT GCGAATGAATA		http://www.cerealsdb.uk.net/
Lr67/Yr46/ Pm46/Sr55	*csSNP723*	S1: GCTGCCCAGAGACGCTT GAGC	S2: GCTGCCCAGAGACGCTT GAGT	GTAGCTCCTCCCTG CGATG		Forrest et al. 2014
	csSNP754	S1: AAATTATAGCAACTAGAATA CCTGCATCAA	S2: AAATTATAGCAACTAGAAT ACCTGCATCAT	GTGCACTAAAGAATTGC ACATGTGCATAA		Forrest et al. 2014
	csSNP856	S1: GCTACTACTATTGGTAGCCTG	S2: GCTACTACTATTGGTAGCCTA	CCAGTAGCTTATGGCACTC AAA		Forrest et al. 2014
	csSNP275	S1: ATAAGAGTATGGTTCTCTG GCGACT	S2: ATAAGAGTATGGTTCTCTGG CGACA	TTTGGTCTCTATAACCCG CCAGGAT		Forrest et al. 2014
Sr2/Yr30	*csSr2*	CAAGGGTTGCTAGGATTGG AAAAC	AGATAACTCTTATGATCTTACAT TTTTCTG		172 bp	Mago et al. 2011
Sr2/Yr30/ Lr27	*gwm533*	AAGGCGAATCAAACGGAATA	GTTGCTTTAGGGGAAAAGCC		120 bp	Mago et al. 2011
Lr68	*CSGS*	AAGATTGTTCACAGATCCATGTCA	GAGTATTCCGGCTCAAAAAGG		385 bp	Herrera-Foessel et al. 2012
Sr2/Yr30	*wMAS000005*	GTGCGAGACATCCAACACTCAC	GTGCGAGACATCCAACACTCAT	CTCAAATGGTCGAG CACAAGCTCTA		http://www.cerealsdb.uk.net/
	cs7BLNLRR	GAAGGAGTGCTTCCTCCACTG	CTTGGTTCTCCTGTTCTTCCC		738 bp	Herrera-Foessel et al. 2012
	CIMwMAS0056	CGTGTCTTGGACCTGAGCAAT	CGTGTCTTGGACCTGAGCAAC	TGACCTGAGTCCCGTCAAGA		Dreisigacker pers. comm.
Yr54	*gwm301*	GAGGAGTAAGACACATGCCC	GTGGCTGGAGATTCAGGTTC		?	Basnet et al. 2014
Yr36	*wMAS000017*	CAAGAGGGGAGAGACATGTT ACTTA	CAAGAGGGGAGAGACATGTT ACTTT	GATTATGGGAGTAGGT TGGTGAGATAAAA		http://www.cerealsdb.uk.net/
Yr39	*wgp36*	GAYGTNAARCCIGARAA	GCATTGGAACAAGGTGAA		830 bp	Lin and Chen 2007
	wgp45	CCGTTGGACAGGAAGGAG	GCATTGGAACAAGGTGAA		940 bp	Lin and Chen 2007
Yr52	*barc182*	CCATGGCCAACAGCTCAAGG TCTC	CGCAAAACCGCATCAGGGAAG CA CCAAT		75 bp	Ren et al. 2012
	wgp5258	GGCAAGACCACATTA	CCATATGTCATCAATGAG		650 bp	Ren et al. 2012
Yr59	*wgp5175*	GGAGGCTTAGGGAAG	TGGTAGGTCCTTGTA		450 bp	Zhou et al. 2014
	barc32	GCGTGAATCCGGAAACCCAA TCTGTG	TGGAGAACCTTCGCATTGTGTC ATTA		165 bp	Zhou et al. 2014
Sr56	*Sun209*	CTGTAAGGTTCTTTCGGATTGG	CATGGTCTTCGACGACTTAGTG		448 bp	Bansal et al. 2014
	Sun320	TAGCAAACGCAACAATTTGG	CATCAGTTTCTACGGCAGCA		179 bp	Bansal et al. 2014

the MAS program in Mexico with the main target of pyramiding two to three markers linked to the genes *Sr25, Sr1A1R, Sr24, Sr26,* and *SrCad.* Within these 40 lines the percentage of lines with R, R-MR, and MR ratings was highest, and about two-fold larger than the percentage of lines with the same ratings derived from selections in Mexico not using MAS. A higher percentage of lines with R, R-MR and MR ratings was derived from selection in Kenya in comparison to selection in Mexico using conventional selection due to the selection pressure that is obtained in Kenya.

In durum wheat the two resistance genes *Lr19/Sr25* and *Lr47* were introgressed in different genetic backgrounds via marker assisted backcrossing MABC. Furthermore, *Lr19/Sr25* and *Sr22*, both genes tightly linked on the long arm of chromosome 7A, were combined via the F_2-enrichment strategy.

Strategy of pyramiding rusts resistance genes in CIMMYT

Recently, the bread wheat improvement program has also started targeted development of rusts-resistant germplasm via early generation MAS or MABC see Chapter 3 and Figure 1. The newly developed elite lines, which will carry multiple resistance genes, can be effectively used as source parents for durable resistance in the future. For gene introgression and pyramiding, there are several candidate race-specific genes, which are effective against the majority of the three rust races including Ug99, already available in CIMMYT germplasm Table 1. Similarly, some important genes currently not present in CIMMYT germplasm, *Yr36, Lr67/Yr46/Sr55, Sr33* and *Sr35* for example, will be introduced into CIMMYT germplasm using diagnostic and/or gene-based markers. Similarly, race non-specific slow-rusting APR genes, such as *Lr34/Yr18/Sr57/Pm38, Lr46/Yr29/Sr58/Pm39, Lr67/*

Table 5. Disease rating for stem rust in the 31st Semi-Arid Wheat Yield Screening Nursery. Lines were classified according the selection environment and using conventional selection or MAS.

		Disease rating for stem rust							
Selection environment		**Total**	**R**	**R-MR**	**MR**	**MR-MS**	**MS**	**MSS**	**S**
Mexico - no MAS	No	159	25	52	30	27	22	3	0
	%		15.7	48.4	67.3	84.3	98.1	100.0	
Mexico - MAS	No	40	23	13	4	0	0	0	0
	%		57.5	90.0	100.0				
Kenya	No	106	44	46	10	5	1	0	0
	%		41.5	84.9	94.3	99.1	100.0		
ALL	No	305	92	111	44	32	23	3	0
	%		30.2	66.6	81.0	91.5	99.0	100.0	

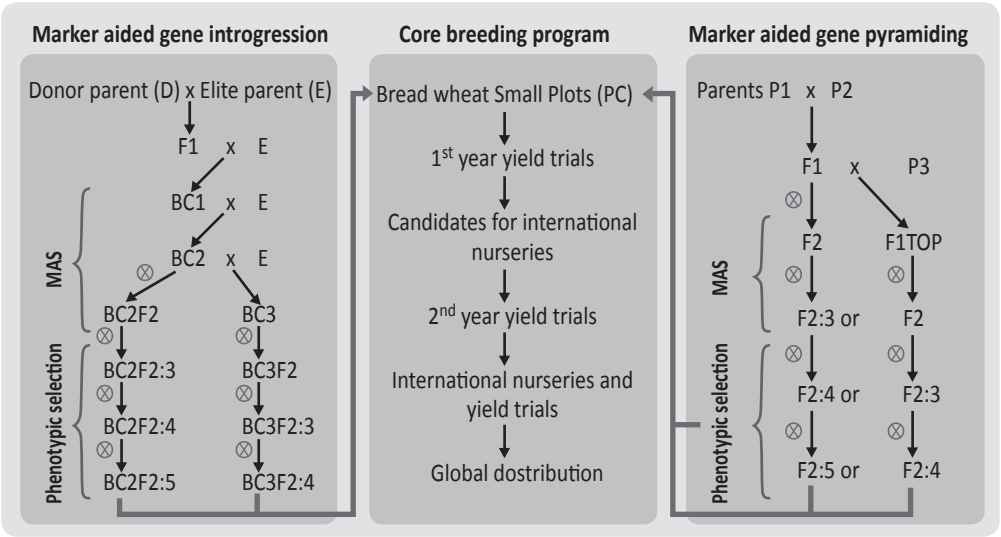


Figure 1. Proposed schematic diagram of marker assistant selection MAS in wheat rust resistance gene introgression and pyramiding approaches, and their integration into bread wheat breeding program at CIMMYT.

Yr46/Sr55/Pm46, and *Lr27/Yr30/Sr2* have been the major foundation of durable resistance in CIMMYT's germplasm. These prominent slow-rusting APR genes Table 2 can be effectively used to combine with each other as well as with the race-specific genes Table 1 via MAS approaches. Enhancing and strengthening the resistance diversity in CIMMYT wheat germplasm with multiple gene pyramiding will provide a better foundation for future breeding tools and approaches, including genome-wide selection, to reach the goal of developing durably resistant high-yielding wheat varieties for resource-poor farmers.

References

Bansal, U., Bariana, H., Wong, D., Randhawa, M., Wicker, T., Hayden, M., Keller, B. 2014 Molecular mapping of an adult plant stem rust resistance gene *Sr56* in winter wheat cultivar Arina. Theoretical and Applied Genetics 127:1441-1448

Basnet, B.R., Singh, R.P., Ibrahim, A.M.H., Herrera-Foessel, S.A., Huerta-Espino, J., Lan, C.X., Rudd, J.C. 2014a. Characterization of *Yr54* and other genes associated with adult plant resistance to yellow rust and leaf rust in common wheat Quaiu 3. Molecular Breeding 33: 385-399

Basnet, B.R., Singh, S., Lopez-Vera, E.E., Huerta-Espino, J., Bhavani, S., Jin, Y., Rouse, M.N., Singh, R.P. 2014b. Molecular mapping and validation of *SrND643*: A new wheat gene for resistance to stem rust race Ug99. Phytopathology (*in press*)

Bhavani, S., Singh, R.P., Argillier, O., Huerta-Espino, J., Singh, S., Njau, P., Brun, S., Lacam, S., Desmouceaux, N. 2011. Mapping durable adult plant stem rust resistance to the race Ug99 group in six CIMMYT wheats. In: McIntosh RA (ed) BGRI 2011 Technical Workshop. Borlaug Global Rust Initiative, St Paul, MN, USA, pp. 43–53

Cheng, P., Chen, XM. 2010. Molecular mapping of a gene for stripe rust resistance in spring wheat cultivar IDO377s. Theoretical and Applied Genetics 121:195-204

Dubin, H.J., Brennan, J.P. 2009. Fighting a 'Shifty Enemy', the international collaboration to contain wheat rusts. Combating stem and leaf rust of wheat: Historical perspective, impacts, and lessons learned. Washington D.C.: International Food Policy Research Institute. Chapter 2, pp 19-24

Forrest, K., Pujol, V., Bulli, P., Pumphrey, M., Wellings, C., Herrera-Foessel, S., Huerta-Espino, J., Singh, R., Lagudah, E., Hayden, M., Spielmeyer, W.. 2014. Development of a SNP marker assay for the *Lr67* gene of wheat using a genotyping by sequencing approach. Molecular Breeding 34:2109-2118

Ghazvini, H., Hiebert, C.W., Zegeye, T., Liu, S., Dilawari, M., Tsilo, T., Anderson, J.A., Rouse, M.N., Jin, Y., Fetch, T. 2012. Inheritance of resistance to Ug99 stem rust in wheat cultivar Norin 40 and genetic mapping of *Sr42*. Theoretical and Applied Genetics 125:817-824

Helguera, M., Khan, I.A., Dubcovsky, J. 2000. Development of PCR markers for the wheat leaf rust resistance gene *Lr47*. Theoretical and Applied Genetics 100:1137-1143

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

Helguera, M., Vanzetti, L., Soria, M., Khan, I.A., Kolmer, J., Dubcovsky, J. 2005. PCR markers for *Triticum speltoides* leaf rust resistance gene *Lr51* and their use to develop isogenic hard red spring wheat lines. Crop Science 45:728-734

Herrera-Foessel, S.A., Singh, R.P., Huerta-Espino, J., Rosewarne, G.M., Periyannan, S.K., Viccars, L., Calvo-Salazar, V., Lan, C.X., Lagudah, E.S. 2012. *Lr68*: a new gene conferring slow rusting resistance to leaf rust in wheat. *Theoretical and Applied Genetics* 124:1475-1486

Huang, L, Brooks, SA, Li, WL, Fellers, JP, Trick, HN, Gill, BS. 2003. Map-based cloning of leaf rust resistance gene Lr21 from the large and ploypoid genome of breed wheat. Genetics 164:655-664

Lagudah, E.S., Krattinger, S.G., Herrera-Foessel, S., Singh, R.P., Huerta-Espino, J., Spielmeyer, W., Brown-Guedira, G., Selter, L.L., Keller, B. 2009. Gene-specific markers for the wheat gene *Lr34/Yr18/Pm28* which confers resistance to multiple fungal pathogens. Theoretical and Applied Genetics 119:889-898

Lagudah, E.S., McFadden, H., Singh, R.P., Huerta-Espino, J., Bariana, H.S., Spielmeyer, W. 2006. Molecular genetic characterization of the *Lr34/Yr18* slow rusting resistance gene region in wheat. Theoretical and Applied Genetics 114:21-30

Lan, C.X., Rosewarne, G.M., Singh, R.P., Herrera-Foessel, S.A., Huerta-Espino, J., Basnet, B.R., Zhang, Y.L., Yang, E.N. 2014. Genetic analysis of resistance to leaf rust and stripe rust in wheat cultivar Francolin#1. Molecular Breeding 34:789-803

Li, Z.F., Lan, C.X., He, Z.H., Singh, R.P., Rosewarne, G.M., Chen, X.M., Xia, X.C. 2014. Overview and application of QTL for adult plant resistance to leaf rust and powdery mildew in wheat. Crop Science 54:1907-1925

Lin, F., Chen, X.M. 2007. Genetics and molecular mapping of genes for race-specific all-stage resistance and non-race-specific high-temperature adult-plant resistance to stripe rust in spring wheat cultivar Alpowa. Theoretical and Applied Genetics 114:1277-1287

Liu, J., Chang, Z.J., Zhang, X.J., Yang, Z.J., Li, X., Jia, J.Q., Zhan, H.X., Guo, H.J., Wang, J.M. 2013. Putative *Thinopyrum intermedium*-derived stripe rust resistance gene *Yr50* maps on wheat chromosome arm 4BL. Theoretical and Applied Genetics 126:265-274

Lopez-Vera, E.E., Nelson, S., Singh, R.P., Basnet, B.R., Haley, S.D., Bhavani, S., Huerta-Espino, J., Xoconostle-Cazares, B., Ruiz-Medrano, R., Rouse, M.N., Singh, S.. 2013. Resistance to stem rust Ug99 in six bread wheat cultivars maps to chromosome 6DS. Theoretical and Applied Genetics 127: 231-239

López-Vera E.E., Singh, S., Bhavani, S., Basnet, B.R., Singh, R.P., Huerta-Espino, J., Rouse, M.N., Xoconostle-Cázares, B., Ruiz-Medrano, R. 2014. Identification of genes on wheat chromosome 2B conferring resistance to stem rust fungus race Ug99. In: Plant and Animal Genome XXII Conference, 2014, Jan 14-15, San Diego, California, USA

Luo, P.G., Hu, X.Y., Ren, Z.L., Zhang, H.Y., Shu, K., Yang, Z.J. 2008. Allelic analysis of stripe rust resistance gene on wheat chromosome 2B. Genome 51:922-927

Mago, R., Verlin, D., Zhang, P., Bansal, U., Bariana, H., Jin, Y., Ellis, J., Hoxha, S., Dundas, I. 2013. Development of wheat-*Aegilops speltoides* recombinants and simple PCR-based markers for *Sr32* and new stem rust gene on the 2S#1 chromosome. Theoretical and Applied Genetics 126:2943-2955

Mago, R., Bariana, H.S., Dundas, I.S., Spielmeyer, W., Lawrence, G.J., Pryor, A.J., Ellis J.G. 2005. Development of PCR markers for the selection of wheat stem rust resistance genes *Sr24* and *Sr26* in diverse wheat germplasm. Theoretical and Applied Genetics 111:496-504

Mago R., Brown-Guedira, G., Dreisigacker, S., Breen, J., Jin, Y., Singh, R., Appels, R., Lagudah, E.S., Ellis, J., Spielmeyer, W. 2011. An accurate DNA marker assay for stem rust resistance gene *Sr2* in wheat. Theoretical and Applied Genetics 122:735-744

McCartney, C.A., Somers, D.J., McCallum, B.D., Thomas, J., Humphreys, D.G., Menzies, J.G., Brown, P.D. 2005. Microsatellite tagging of the leaf rust resistance gene *Lr16* on wheat chromosome 2BSc. Molecular Breeding 15: 329-337

Njau, P.N., Bhavani, S., Huerta-Espino, J., Keller, B., Singh, R.P. 2012. Identification of QTL associated with durable adult plant resistance to stem rust race Ug99 in wheat cultivar "Pavon 76." Euphytica 190:33-44

Olson, E.L., Brown-Guedira, G., Marshall, D., Stack, E., Bowden, R.L., Jin, Y., Rouse, M., Pumphrey, M.O. 2010. Development of wheat lines having a small introgressed segment carrying stem rust resistance gene *Sr22*. Crop Science 50:1823-1830

Periyannan, S., Moore, J., Ayliffe, M., Bansal, U., Wang, X., Huang, L., Deal, K., Luo, M., Kong, X., Bariana, H., Mago, R., McIntosh, R., Dodds, P., Dvorak, J., Lagudah, E. 2013. The gene *Sr33*, an ortholog of barley *Mla* genes, encodes resistance to wheat stem rust race Ug99. Science 341:786-788

Ramirez-Gonzalez, R. H., Segovia, V., Bird, N., Fenwick, P., Holdgate, S., Berry, S., Jack, P., Caccamo, M., Uauy, C. 2015. RNA-- bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. Plant Biotechnology Journal, 135:613-624

Ren, R.S., Wang, M.N., Chen, X.M., Zhang, Z.J. 2012. Characterization and molecular mapping of *Yr52* for high-temperature adult-plant resistance to stripe rust in spring wheat germplasm PI 183527. Theoretical and Applied Genetics 125:847-857

Saari, E.E., Prescott, J.M. 1985. World distribution in relation to economic losses. In: Roelfs A. P. and W. R. Bushnell eds. The Cereal Rusts Vol. II: Diseases, distribution, epidemiology and control, Academic Press, Orlando, Fl. pp:259-298

Sambasivam, P.K., Bansal, U.K., Hayden, M.J., Dvorak, J., Lagudah, E.S., Bariana, H.S. Identification of markers linked with stem rust resistance genes *Sr33* and *Sr45*. 2008. In: Appels, R., Eastwood, R., Lagudah, E., Langridge, P., Michael, M.L. (ed) Proceedings of the 11th International Wheat Genetics Symposium. Brisbane, Australia, 24-29 August 2008

Simons, K., Abate, Z., Chao, S., Zhang, W., Rouse, M., Jin, Y., Elias, E., Dubcovsky, J. 2011. Genetic mapping of stem rust resistance gene *Sr13* in tetraploid wheat *Triticum turgidum* ssp. *durum* L. Theoretical and Applied Genetics 122:649-658

Spielmeyer, W., Sharp, P.J., Lagudah, E.S. 2003. Identification and validation of markers linked to broad-spectrum stem rust resistance gene *Sr2* in wheat *Triticum aestivum* L. Crop Science 43:333-336

Uauy, C., Brevis, J.C., Chen, X.M., Khan, I., Jackson, L., Chicaiza, O., Distelfeld, A., Fahima, T., Dubcovsky, J. 2005. High-temperature adult-plant HTAP stripe rust resistance gene *Yr36* from *Triticum turgidum* ssp. *Dicoccoides* is closely linked to the grain protein content locus *Gpc-B1*. *Theoretical and Applied Genetics* 112:97-105

Wang, C., Zhang ,Y., Han, D., Kang, Z., Li, G., Cao, A., Chen, P. 2008. SSR and STS markers for wheat stripe rust resistance gene *Yr26*. Euphytica, 1593:359-366

Wellings, C.R. 2011. Global status of stripe rust: a review of historical and current threats. Euphytica 179:129-141

Weng, Y., Azhaguvel, P., Devkota, R.N., Rudd, J.C. 2007. PCR-based markers for detection of different sources of 1AL.1RS and 1BL.1RS wheat–rye translocations in wheat background. *Plant Breeding* 126:482-484

Zhang, W., Dubcovsky, J. 2008. Association between allelic variation at the Phytoene synthase 1 gene and yellow pigment content in the wheat grain. Theoretical and Applied Genetics 116:635-645

Zhang, W., Olson, E., Saintenac, C., Rouse, M., Abate, Z., Jin, Y., Akhunov, E., Pumphrey, M., Dubcovsky, J. 2010. Genetic maps of stem rust resistance gene *Sr35* in diploid and hexaploid wheat. Crop Science 50:2464-2474

Zhou, X.L., Wang, M.N., Chen, X.M., Lu, Y., Kang, Z.S., Jing, J.X. 2014. Identification of *Yr59* conferring high-temperature adult-plant resistance to stripe rust in wheat germplasm PI 178759. Theoretical and Applied Genetics 127: 935-945

Chapter 6:
Marker-Assisted Selection for Spike and Foliar Blight Diseases

Xinyao He, and Pawan K. Singh
CIMMYT Int., Apdo. Postal 6-641,06600 Mexico, DF, Mexico

MAS for Fusarium head blight (FHB)

FHB is a major wheat disease globally, with major epidemic regions being North America, Europe, East Asia and the South Cone of South America. *Fusarium graminearum* (teleom. *Gibberella zeae*) is the most important causal agent worldwide. Besides yield reduction, FHB produces a set of mycotoxins, particularly deoxynivalenol (DON), which is harmful to both human and livestock. In most developed countries, legally enforceable DON limits in wheat grain and food products have been set, reflecting concerns for food safety.

Host resistance is the most important component in the disease management system, although other measures like fungicide and cultural practices should also be considered to achieve a satisfactory control (Gilbert and Haber 2013). There are three major difficulties for breeding FHB resistance varieties: 1) the multigenic control of host resistance and a lack of functional markers; 2) limited resistance sources in adapted elite germplasm; and 3) multiple resistance components. Numerous host resistance mechanisms have been proposed, each having its own evaluation methods. The most famous resistance components are Type I for initial infection and Type II for disease spread in spike tissues (Liu et al. 2009).

FHB is a quantitatively inherited disease, making the application of MAS in this disease more difficult than in qualitatively inherited traits. Until now, more than 100 published studies have been performed to identify FHB resistance QTLs, which have been mapped to all the 21 wheat chromosomes, with various phenotypic effects (Buerstmayr et al. 2009; Liu et al. 2009). Until now, five QTLs have been fine mapped and designated (Table 1, Appendix 1), but none has been cloned and therefore no functional marker is available for FHB.

Fhb1 identified in Sumai 3 is the most well-studied FHB resistance gene. After its fine mapping, a co-dominant STS marker, *umn10*, was developed based on a polymorphic site near the candidate gene region. Recently, two SNP markers, *Xsnp3BS-8* and *Xsnp3BS-11*, were developed to facilitate high-throughput genotyping (Bernardo et al. 2012). It is noteworthy, however, that this gene is usually in repulsive phase with the famous stem rust resistance gene *Sr2*, compromising its application in breeding practices; this situation is being changed with the availability of lines in which the two genes are in coupling phase (Thapa et al. 2013).

In addition to these FHB resistance QTLs, dwarfing genes like *Rht-B1b*, *Rht-D1b* and *Rht8* have also been proven to be associated with FHB resistance, based on either pleiotropy, tight linkage or disease escape (Buerstmayr et al. 2009). Although all the three aforementioned dwarfing genes reduce Type I resistance, *Rht-B1b* and *Rht8* confer less FHB susceptibility compared with *Rht-D1*, and there was evidence showing that *Rht-B1b* is able to confer Type

Table 1. Information on the five nominated FHB resistance genes.

FHB genes	Source of resistance	Chromosome	Resistance Type	Flanking markers	References
Fhb1	Sumai 3	3BS	II	*gwm533* and *gwm493*	Cuthbert et al. 2006
Fhb2	Sumai 3	6BS	II	*gwm133* and *gwm644*	Cuthbert et al. 2007
Fhb3	*Leymus racemosus*	T7AL·7Lr#1S	II	*BE586744-STS, BE404728-STS*, and *BE586111-STS*	Qi et al. 2008
Fhb4	Wangshuibai	4B	I	*barc20* and *wmc349*	Xue et al. 2010
Fhb5	Wangshuibai	5A	I	*barc56* and *barc100*	Xue et al. 2011

II resistance. Therefore, it is recommended to use *Rht-B1b* and *Rht8* in breeding practices to combine reduced plant stature with acceptable FHB resistance (Gilbert and Haber 2013).

Despite the extensive efforts relating to identification of resistance QTLs, limited progress has been obtained regarding the utilization of those QTLs in MAS, primarily due to the lack of functional markers. Nevertheless, there were examples where MAS was employed to develop FHB-resistant cultivars, such as the newly registered Canadian wheat cultivar 'Cardale' (Gilbert and Haber 2013). The utilization of MAS in FHB resistance breeding at CIMMYT dates back to 2008, when a collaborative project with USDA-ARS Small Grains Genotyping Center, Fargo, was initiated (Duveiller et al. 2008). And nowadays the 'haplotyping' system at CIMMYT comprises 17 markers for 10 validated QTLs on seven chromosomes (Table 2, Appendix 1), which has facilitated the genotypic characterization of numerous CIMMYT elite lines, including the 13th and 14th FHBSN (He et al. 2013a; He et al. 2013b). However, this system is being upgraded to incorporate markers in closer linkage with several of the QTLs, and it is predictable that the SSR and STS markers will be replaced by high throughput SNP markers in the near future.

MAS for Septoria tritici blotch (STB)

STB is a foliar blight disease that reduces yields up to 60% under conducive environmental conditions, with Europe, North America, South America, Australia and the Central West Asia and North Africa (CWANA) region being the major epidemic regions (Raman and Milgate 2012). This disease is caused by the ascomycete fungal agent *Zymoseptoria tritici* (anamorph: *Septoria tritici*).

The host resistance to STB is reported to be both qualitative and quantitative. Although gene-for-gene interactions exist between a certain resistance genes and the corresponding pathogen isolates, the resistance conferred by each gene is weak and cannot provide sufficient protection to wheat as those in rusts and powdery mildew (Goodwin 2012). Like in other diseases, breakdown of STB resistance genes has been observed, e.g. resistance of the wheat cultivar 'Gene' was defeated only five years after its release, implying its resistance nature of 'race-specific' (Cowger et al. 2000). Thus it is recommended to pyramid both qualitative and quantitative resistance genes in breeding materials to achieve durable resistance (Raman and Milgate 2012).

The first STB resistance gene, *Stb1*, was discovered in 1966 and designated in 1985, followed by *Stb2* and *Stb3* in 1985, and *Stb4* in 1994 (Goodwin 2012). But it is *Stb5* that was first mapped on a genetic map (Arraiano et al. 2001). Soon after this landmark work, 12 more resistance genes were reported during the 2000s as reviewed by Goodwin (2012). In the last few years, *Stb16*, *Stb17* and *Stb18* were identified and mapped (Ghaffary et al. 2011; Ghaffary et al. 2012), and the chromosomal localization of *Stb2* and *Stb3* were adjusted from 3BS to 1BS and 6DS to 7AS, respectively (Liu et al. 2013). In addition to major resistance genes, several quantitative loci with minor effects (Simon et al. 2012; Kelm et al. 2013; Risser et al. 2012; Kosellek et al. 2013) have been identified. At CIMMYT, two new resistant QTL were recently mapped on chromosome 5AL and 3BS in two populations under field conditions in Mexico (Dreisigacker et al. 2015).

All mapped *Stb* genes but *Stb15* have at least one linked SSR marker, which enables the application of MAS (Ghaffary et al. 2011; Ghaffary et al. 2012; Goodwin 2012). At CIMMYT, the haplotyping work on STB began in the late 2000s

Table 2. FHB markers used in CIMMYT's haplotyping system.

Source of resistance	Chromosome	Resistance Type	Flanking marker(s)	Locus
Sumai 3	3BS	II	*umn10, Xsnp3BS-11, Xsnp3BS-8*	*Fhb1*
	5AS	I	*barc186* and *barc180*	*Fhb5*
	6BS	II	*gwm133* and *wmc179*	*Fhb2*
Frontana	3A	I	*dupw227*	
	5AS	I	*barc197* and *gwm129*	*Fhb5*
Wuhan 1	2DL	II	*wmc144* and *wmc245*	
	4BS	II	*wmc238* and *gwn149*	*Fhb4*
CJ 9306	2DL	II	*gwm157* and *gwm539*	
T. dicoccoides	3A	II	*gwm2*	
	7A	II	*barc121* and *wmc488*	

* This table was modified from He et al. (2013b)

on elite breeding materials and parents of mapping populations, and recently it became a routine procedure to characterize genotypically CIMMYT's International Septoria Observation Nurseries (ISEPTON) with linked SSRs (Table 3, Appendix 1). Similar to the FHB markers, the markers for STB are also being constantly updated to incorporate new findings.

Table 3. STB markers used in CIMMYT's haplotyping system.

Gene locus	Linked markers	Chromosome	Source of resistance
Stb1	*barc74*	5BL	Bulgaria 88
Stb2	*wmc230*	1BS	Veranapolis
Stb3	*gdm132*	7ASS	Israel 493
Stb4	*gwm111*	7DS	Tadinia
Stb5	*gwm44*	7DS	CS/Synthetic 6x
Stb6	*gwm369*	3AS	Flame
Stb7	*wmc313*	4AL	Estanzuela Federal, Kavkaz-K4500 L.6.A.4
Stb8	*gwm577, gwm146*	7BL	Synthetic W7984, Opata m85
Stb9	*wmc317*	2B	Courtot
Stb10	*wms848*	1D	Kavkaz-K4500 L.6.A.4
Stb11	*barc008*	1BS	TE 9111
Stb12	*wmc219*	4AL	Estanzuela Federal, Kavkaz-K4500 L.6.A.4
Stb13	*wmc396*	7B	Salamouni
Stb14	*wmc500, wmc623*	3B	Salamouni
Stb16	*wmc494*	6B	SH M3, Opata M85
Stb17	*hbg247*	5A	Kavkaz-K4500 L.6.A.4
Stb18	*gpw3087, gpw5176*	6DS	Balance

MAS for tan spot (TS) and Stagonospora nodorum blotch (SNB)

In addition to STB, TS and SNB are two more foliar blight diseases with global importance. TS is caused by *Pyrenophora tritici-repentis* (Died.) Drechs. [anamorph *Drechslera tritici-repentis* (Died.) Shoem.] and the major epidemic regions include North America, the Southern Cone region of South America, Australia, Europe and Central Asia, leading to yield reduction of up to 50% (Singh et al. 2012). SNB is caused by *Parastagonospora nodorum* (Berk.) Quaedvlieg, Verkley & Crous (anamorph *Stagonospora nodorum* (Berk.) E. Castell. & Germano), with its major epidemic regions being Australia, Europe and North America (Oliver et al. 2012); this disease usually causes lower yield losses compared to STB and TS, but the values of 31% and even around 40% have been reported (Eyal et al. 1987; Bhathal et al. 2003). Under the field natural infection conditions, TS and SNB showed very similar symptoms to STB, and thus in practice these three diseases are usually scored together as leaf blotch disease complex (Lu and Lillemo 2014).

For both TS and SNB, both qualitative and quantitative host resistance have been reported (Singh et al. 2012; Francki 2013). For TS, host genes corresponding to host-specific toxins Ptr ToxA, Ptr ToxB and Ptr ToxC have been located on chromosomes 5BL, 2BS and 1AS, respectively (Singh et al. 2012); for SNB, host genes *Snn4, Snn1, Snn2, Snn5, Snn3* and *Tsn1* have been respectively located on 1AS, 1BS, 2DS, 4BL, 5BS and 5BL (Lu and Lillemo 2014). In addition, numerous resistance QTLs for both seedling and adult plant resistance have been reported for the two diseases, as reviewed by Singh et al. (2012) for TS and by Francki (2013) for SNB. Despite the identification of molecular markers linked to those genes/QTLs, their application in breeding practices has been limited, with the only exception being *Tsn1*. This gene was mapped and designated by Faris et al. (1996) and Stock et al. (1996) in their respective genetic studies for wheat resistance to TS. Afterwards, the corresponding fungal effector gene of *Tsn1, Ptr ToxA*, was found to be transferred from *P. nodorum*, the causal agent of SNB, accounting for the fact that Tsn1 in wheat is responsible for the sensitivity to both TS and SNB (Friesen et al. 2006). The gene has been cloned (Faris et al. 2010) and a SNP marker *wMAS000020* and SSR markers *fcp1* and *fcp394* are available and used for MAS at CIMMYT (Appendix 1).

References

Arraiano, L.S., A.J. Worland, C. Ellerbrook and J.K.M. Brown. 2001. Chromosomal location of a gene for resistance to *septoria tritici* blotch (*Mycosphaerella graminicola*) in the hexaploid wheat 'Synthetic 6x'. Theoretical and Applied Genetics, 103, 758-764

Bernardo, A.N., H.X. Ma, D.D. Zhang and G.H. Bai. 2012. Single nucleotide polymorphism in wheat chromosome region harboring *Fhb1* for *Fusarium* head blight resistance. Molecular Breeding, 29, 477-488

Bhathal, J. S., R. Loughman and J. Speijers. 2003. Yield reduction in wheat in relation to leaf disease from yellow (*tan*) spot and *septoria nodorum* blotch. European Journal of Plant Pathology, 109, 435-443

Buerstmayr, H., T. Ban and J.A. Anderson. 2009. QTL mapping and marker-assisted selection for *Fusarium* head blight resistance in wheat: a review. Plant Breeding, 128, 1-26

Cowger, C., M.E. Hoffer and C.C. Mundt. 2000. Specific adaptation by *Mycosphaerella graminicola* to a resistant wheat cultivar. Plant Pathology, 49, 445-451

Cuthbert, P.A., D.J. Somers and A. Brulé-Babel. 2007. Mapping of Fhb2 on chromosome 6BS: a gene controlling *Fusarium* head blight field resistance in bread wheat (*Triticum aestivum L.*). Theoretical and Applied Genetics, 114, 429-437

Cuthbert, P.A., D.J. Somers, J. Thomas, S. Cloutier and A. Brulé-Babel, A. 2006. Fine mapping *Fhb1*, a major gene controlling *Fusarium* head blight resistance in bread wheat (*Triticum aestivum L.*). Theoretical and Applied Genetics, 112, 1465-1472

Duveiller, E., M. Mezzalama, J. Murakami, J. Lewis and T. Ban. 2008. Global *Fusarium* networking. Cereal Research Communications, 36, 11-19

Dreisigacker, S., X. Wang, B.A. Martinez Cisneros, R. Jing, P.K. Singh. 2015. Adult plant resistance to *Septoria tritici blotch* in hexaploid spring wheat Theoretical and Applied Genetics, 128, 2317-2329

Eyal, Z., A.L. Scharen, J.M. Prescott, and M.V. Ginkel, The Septoria diseases of wheat: Concepts and methods of disease management, 1987 (Mexico, D.F.: CIMMYT).

Faris, J.D., J.A. Anderson, L.J. Francl and J.G. Jordahl. 1996. Chromosomal location of a gene conditioning insensitivity in wheat to a necrosis-inducing culture filtrate from *Pyrenophora tritici-repentis*. Phytopathology, 86, 459-463

Faris, J.D., Z. Zhang, H. Lu, S. Lu, L. Reddy, S. Cloutier, J.P. Fellers, S.W. Meinhardt, J.B. Rasmussen, S.S. Xu, R.P. Oliver, K.J. Simons and T.L. Friesen. 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

Francki, M.G. 2013. Improving *Stagonospora nodorum* Resistance in Wheat: A Review. Crop Science, 53, 355-365

Friesen, T.L., E.H. Stukenbrock, Z. Liu, S. Meinhardt, H. Ling, J.D. Faris, J.B. Rasmussen, P S. Solomon, B.A. McDonald and R.P. Oliver. 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. Nature Genetics, 38, 953-956

Ghaffary, S.M., O. Robert, V. Laurent, P. Lonnet, E. Margale, T.A. van der Lee, R.G. Visser and G.H. Kema. 2011. Genetic analysis of resistance to septoria tritici blotch in the French winter wheat cultivars Balance and Apache. Theoretical and Applied Genetics, 123, 741-754

Ghaffary, S.M.T., J.D. Faris, T.L. Friesen, R.G.F. Visser, T.A.J. van der Lee, O. Robert and G.H.J. Kema. 2012. New broad-spectrum resistance to septoria tritici blotch derived from synthetic hexaploid wheat. Theoretical and Applied Genetics, 124, 125-142

Gilbert, J. and S. Haber. 2013. Overview of some recent research developments in *Fusarium* head blight of wheat. Canadian Journal of Plant Pathology, 35, 149-174

Goodwin, S.B. 2012. Resistance in Wheat to Septoria Diseases Caused by *Mycosphaerella graminicola* (*Septoria tritici*) and Phaeosphaeria (*Stagonospora*) *nodorum*. In Sharma, I. (Ed.). Disease resistance in wheat (pp. 151-159). Wallingford, Oxfordshire, UK; Cambridge, MA: CABI

He, X., P. K. Singh, E. Duveiller, S. Dreisigacker and R.P. Singh. 2013a. Development and characterization of International Maize and Wheat Improvement Center (CIMMYT) germplasm for *Fusarium* head blight resistance. In Alconada, T. M. & Chulze, S. N. (Eds.). *Fusarium* Head Blight in Latin America (pp. 241-262).

He, X., P.K. Singh, E. Duveiller, N. Schlang, S. Dreisigacker and R.P. Singh. 2013b. Identification and characterization of international *Fusarium* head blight screening nurseries of wheat at CIMMYT, Mexico. European Journal of Plant Pathology, 136, 123-134

Liu, S.Y., M.D. Hall, C.A. Griffey and A.L. McKendry. 2009. Meta-Analysis of QTL Associated with *Fusarium* Head Blight Resistance in Wheat. Crop Science, 49, 1955-1968

Liu, Y., L. Zhang, I.A.Thompson, S.B. Goodwin and H.W. Ohm. 2013. Molecular mapping re-locates the *Stb2* gene for resistance to Septoria tritici blotch derived from cultivar Veranopolis on wheat chromosome *1BS*. Euphytica, 190, 145-156

Lu, Q. and M. Lillemo. 2014. Molecular mapping of adult plant resistance to *Parastagonospora nodorum* leaf blotch in bread wheat lines 'Shanghai-3/Catbird' and 'Naxos'. Theoretical and Applied Genetics, 127, 2635-2644

Oliver, R. P., T.L. Friesen, J.D. Faris and P.S. Solomon. 2012. *Stagonospora nodorum*: From pathology to genomics and host resistance, Annual Review of Phytopathology (Vol. 50, pp. 23-43).

Qi, L. L., M.O. Pumphrey, B. Friebe, P.D. Chen and B.S. Gill. 2008. Molecular cytogenetic characterization of alien introgressions with gene *Fhb3* for resistance to *Fusarium* head blight disease of wheat. Theoretical and Applied Genetics, 117, 1155-1166

Raman, H. and A. Milgate. 2012. Molecular breeding for Septoria tritici blotch resistance in wheat. Cereal Research Communications, 40, 451-466

Singh, P.K., E. Duveiller and R.P. Singh. 2012. Resistance Breeding for Tan Spot (*Pyrenophora tritici-repentis*) of Wheat. In Sharma, I. (Ed.). Disease resistance in wheat (pp. 136-150). Wallingford, Oxfordshire, UK; Cambridge, MA: CABI

Stock, W.S., A.L. Brûlé-Babel and G.A. Penner. 1996. A gene for resistance to a necrosis-inducing isolate of *Pyrenophora triticirepentis* located on 5BL of *Triticum aestivum* cv. Chinese Spring. Genome, 39, 598-604

Thapa, R., K. Wise, S. Goodwin, G. Brown-Guedira and H. Ohm. 2013. Determining the order of resistance genes *Qsng-3BS, Fhb1* and *Sr2* and combining them in coupling on wheat chromosome 3BS. In. Proceedings of the 12th International Wheat Genetics Symposium (pp. 84). Pacifico Yokohama, Japan

Xue, S., G.Q. Li, H.Y. Jia, F. Xu, F. Lin, M.Z. Tang, Y. Wang, X. An, H.B. Xu, L.X. Zhang, Z.X. Kong and Z.Q. Ma. 2010. Fine mapping Fhb4, a major QTL conditioning resistance to *Fusarium* infection in bread wheat (*Triticum aestivum L.*). Theoretical and Applied Genetics, 121, 147-156

Xue, S., F. Xu, M. Tang, Y. Zhou, G. Li, X. An, F. Lin, H. Xu, H. Jia, L. Zhang, Z. Kong and Z. Ma. 2011. Precise mapping Fhb5, a major QTL conditioning resistance to *Fusarium* infection in bread wheat (*Triticum aestivum L.*). Theoretical and Applied Genetics, 123, 1055-1063

Chapter 7:
Molecular Marker Strategy to Enhance Grain Zinc Concentration in Bread Wheat

Govindan Velu and Yuanfeng Hao
CIMMYT Int., Apdo. Postal 6-641,06600 Mexico, DF, Mexico

Micronutrient malnutrition, resulting from diets primarily deficient in zinc (Zn) and iron (Fe) has been widely recognized as a major health problem affecting almost two billion people worldwide, especially in countries with a high consumption of cereals (Black et al. 2013). Zinc is an essential trace element for all organisms and its role has been thoroughly reviewed in both plant and human health (Cakmak, 2000; Graham et al. 2012). About 17% of the world's population is at a risk of micronutrient deficiency due to inadequate Zn intake (Wessells & Brown, 2012), and annually more than 100,000 deaths of children under age five are attributed to Zn deficiency (Black et al. 2013).

The biofortification approach of improving the nutritional quality of staple food crops through breeding offers a cost-effective and sustainable solution to the global malnutrition problems. Wheat is the second most produced cereal crop after Rice and constitutes about 28% of dietary energy and 20% protein to consumers in many parts of the world (Braun et al. 2010). Improving the nutritional levels of wheat is therefore of paramount importance. The wheat biofortification breeding program at the International Maize and Wheat Improvement Center (CIMMYT) is leading the global partnership to develop and disseminate competitive wheat varieties with high grain Zn and other essential agronomic features to target regions in South Asia and beyond. The emphasis of this breeding program is to introduce novel sources of genetic diversity from wild species and landraces, into the adapted wheat background which resulted in the development of widely adapted, durable rust and foliar disease resistant, high Zn wheat varieties (Velu et al. 2014). The high zinc wheat varieties with 20-40% superior in grain Zn concentration over the baseline commercial cultivars are being adapted by small-holder farmers in target countries (Velu et al. 2015).

Identification of molecular markers linked to nutritional traits would be of great interest as nutritional elements are rather difficult and are cost expensive to phenotype. QTL mapping is a highly useful tool for the discovery of markers to use in breeding programs, especially in the post genomics era genotyping costs getting cheaper and cheaper and application of molecular markers in wheat breeding have been increased. One way to implement molecular markers in breeding programs is by the identification of linkage between DNA markers and the loci that control the traits of interest, to make selections in segregating progenies based on those marker-trait associations. The marker-assistant selection (MAS) procedures can greatly facilitate the breeding programs by identifying genomic regions associated with higher grain Zn concentrations. To date, twenty-one QTL for increased grain zinc content (GZnC) have been reported on ten chromosomes in diploid (*T. monococcum* and *T. boeoticum*), tetraploid (*T. dicoccoides* and *T. durum*), and hexaploid wheat (*T. aestivum*) (Xu et al. 2011; Velu et al. 2014).

We initiated the marker discovery for GZnC in wheat at CIMMYT using a recombinant inbred line (RIL) population from the cross between PBW343 and Kenya Swara. Two novel QTL of large effect were stably detected for increasing GZnC on chromosomes 2Bc (centromeric region) and 3AL (long arm). The two QTL individually explained about 10 to 15% of the total phenotypic variation (Hao et al., 2014). The 2Bc QTL from PBW343 has pleiotropic effect and can increase thousand-kernel weight at significant level. The flanking markers associated for these QTLs (Table 2) are being converted into breeder-friendly KASPar (SNP) marker to be able to use in early generation MAS or MABC program.

Another QTL mapping study conducted in a cross Seri M82 x Synthetic Hexaploid Wheat revealed two major QTL for GZnC on chromosome 4B and 6B, interestingly the 4B QTL appears to have pleiotropic effects for GFeC (Crespo et al., 2015). The QTL on 4B is fully linked with the marker TP81797 and the QTL on chromosome 6B is also rather close to marker TP29689 (0.9 cM).

Another mapping study involving RILs derived from a cross between PICUS/3/..* FRANCOLIN (low Zn parent) and REH/HARE//2*BCN/3/CROC_1/AE.SQUARROSA (213)//…/HUITES/7/MUTUS (high Zn parent) showed three major QTL on chromosomes 1B, 5B, and 6A, respectively. The 1B QTL from the high Zn parent explained about 15% of phenotypic variation, and the 5B and 6A QTL from the low Zn parent contribute about 11% and 8% of total variation, respectively. Major QTL contributed from both parents explain the trangressive segregation pattern observed in the RIL population. The three QTL were closely linked to markers *wPt-10518*, *wPt-8163* and *4990410*, respectively (Table 3). Both the 1B and 5B QTL should represent novel loci for increasing GZnC based on our literature review (Table 1).

Table 1 Summary of QTL detected for GZnC in diploid, tetraploid and hexaploid wheat and source cultivars for GZnC from our mapping studies and a literature review.

GZnC	1	2	3	4	5	6	7
A	*T. monococcum* ID-362	*T. durum* Langdon, *T. dicoccoides G18-16*	*T. aestivum* Kenya Swara	*T. aestivum* Hanxuan10	*T. aestivum* Hanxuan10, *T. dicoccoides* G18-16, *T. aestivum* Xiaoyan 54, *T. monococcum* ID-362	*T. spelta* H+26 (PI348449), *T. aestivum* BV2010-13	*T. aestivum* Lumai 14, *T. dicoccoides* G18-16, *T. boeoticum* pau5088, *T. monococcum* pau14087, *T. aestivum* RAC875-2
B	*T. aestivum* BV2010-48	*T. aestivum* PBW343, *T. aestivum* HUW234		*T. aestivum* Jin 411, *T. aestivum* RAC875-2	*T. aestivum* BV2010-13	*T. dicoccoides* G18-16, *T. aestivum* Cascades, *T. dicoccoides* LDN (DIC6B)	*T. dicoccoides* G18-16
D			*T. aestivum* RAC875-2, *T. aestivum* BV2010-48	*T. aestivum* Lumai14			

Note: The QTL detected in diploid, tetraploid, and hexaploid wheat were colored in red, brown and blue, respectively, followed by variety or accession names as source; the boxes in grey shading are QTL we have identified in PBW343/Kenya Swara population; the boxes in pink represent chromosomal location of QTL detected in the Picus…Francolin x Croc/Ae. squarrosa population.

Table 2. Position and QTL effect associated with high grain zinc content (GZnC) in the PBW343 × Kenya Swara RIL population.

Environ.	QTL name	Marker interval	Marker position	Peak LOD	Peak position (cM)	R^2 (%)[b]	Additive effect[b]
GZnC-2012	*QGzncpk.cimmyt-1BS*	*wPt-8622*	68.3	5.6***	68.3	10	-2.16
	QGzncpk.cimmyt-2Bc	*wPt-6174–wPt-2430*	37.6–41.9	4.9**	38.6	9	1.96
	QGzncpk.cimmyt-3AL	*wPt-0286*	79.0	7.9***	79.0	15	-2.43
	QGzncpk.cimmyt-4AS	*wPt-7191–wPt-8007*	24.0–26.3	3.4*	26.1	7	-1.70
	QGzncpk.cimmyt-5BL	*Xcfp393–tPt-3144*	220.8–225.4	3.4*	221.9	8	-1.49
GZnC-2013	*QGzncpk.cimmyt-1BS*	*wPt-8622*	68.3	3.1	68.3	5	-1.65
	QGzncpk.cimmyt-1BL	*Xwmc44–wPt2861*	245.8–249.7	3.0	248.9	5	1.46
	QGzncpk.cimmyt-2Bc	*wPt-6174–wPt-2430*	37.6–41.9	6.6***	37.7	11	2.52
	QGzncpk.cimmyt-2D	*wPt-6847–tPt-6105*	8.9–49.0	5.2***	22.5	26	3.35
	QGzncpk.cimmyt-3AL	*wPt-0286–Xwmc222*	79.0–87.3	4.7**	80.0	11	-2.34
	QGzncpk.cimmyt-6AL	*wPt-667817–tPt-6278*	100.6–103.0	4.0**	101.7	7	-2.16
	QGzncpk.cimmyt-7DL	*wPt-671530*	142.9	3.4	142.9	5	1.36
GZnC-Mean	*QGzncpk.cimmyt-1BS*	*wPt-3103–wPt-8622*	68.0–68.3	7.0***	68.1	11	-2.47
	QGzncpk.cimmyt-2Bc	*wPt-6174*	37.6	6.6***	37.6	10	2.09
	QGzncpk.cimmyt-3AL	*wPt-0286–Xwmc222*	79.0–87.3	9.0***	79.1	15	-2.56
	QGzncpk.cimmyt-4AS	*wPt-7191–wPt-8007*	24.0–26.3	3.1	26.2	5	-1.52
	QGzncpk.cimmyt-6AL	*wPt-667817–tPt-6278*	100.6–103.0	3.0	101.7	6	-1.68

Table 3 Position and effect of QTL associated with high grain zinc concentration in a RIL from PICUS/3/..*FRANCOLIN and REH/HARE//2*BCN/3/ CROC_1/AE.SQUARROSA(213)//…/HUITES/7/MUTUS.

QTL name	Interval	Range	Peak position (cM)	Peak LOD	R2 (%)a	Additive effectb
QGzncpr.cmt-1Bc	*4663991–wPt-10518*	48.8–49.0	48.9	8.8**	15	-2.02
QGzncpr.cmt-3D	*1007328*	12.4	12.4	3.4	6	-1.23
QGzncpr.cmt-5B	*wPt-8163–1139328*	84.3–84.5	84.4	6.5**	11	1.75
QGzncpr.cmt-6A	*4990410*	109.1	109.1	4.7**	8	1.36

**Significant at the 0.01 probability level

Marker Assisted Selection

During the 2014-15 crop season as a proof-of-concept strategy we started applying marker-assisted backcrossing using selected RILs that showed significantly enhanced GZnC than either of the parental lines from PBW 343 x Kenya Swara populations to transfer QTL of interest. Selected RILs high in GZnC and adapted parents were crossed and F_1's were backcrossed to the adapted parent. Marker assisted selection was begun with BC_1 plants to select plants with favorable GZnC alleles. DNA samples from individual BC_1 plants were genotyped, and PCR based probes for these QTL would be used to identify plants which have the favorable donor alleles before the pollination. The plants which are positive for the donor allele would then be backcrossed again to the recurrent parent. The resultant BC_2 families are being advanced through conventional selection schemes of shuttle breeding between Ciudad Obregon and Toluca valley in Mexico. Agronomically superior plants will be selected which might have the favorable alleles for GZnC. This strategy will serve to move the desirable alleles quickly and more precisely into the adapted background.

References

Black, R.E., Victora, C.G., Walker, S.P., Bhutta, Z.A., Christian, P., de Onis, M., Ezzati, M., Grantham-McGregor, S., Katz, J., Martorell, R., Uauy, R. 2013. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet 382:427-451

Braun, H.J., Atlin, G., Payne, T. 2010. Multi-location testing as a tool to identify plant response to global climate change. In: MP Reynold (ed) Climate Change and Crop Production, Oxford, UK: CABI press pp 115–138

Cakmak, I. 2000. Possible roles of zinc in protecting plant cells from damage by reactive oxygen species. New Phytologist 146:185-205.

Crespo-Hererra, L.A., Velu, G., Singh, R.P. 2015. QTL mapping reveals pleiotropic effect for grain iron and zinc concentrations in wheat. Annals of Applied Biology (in press)

Graham R, Knez M, Welch R (2012) How much nutritional iron deficiency in humans globally is due to an underlying zinc deficiency? Adv AgronAdv Agron 115:1-40

Hao, Y., Velu, G., Pena, R.J., Singh, S., Singh, R.P. 2014. Genetic loci associated with high grain zinc concentration and pleiotropic effect on kernel weight in wheat (Triticum aestivum L.). Molecular Breeding 34:1893-1902

Velu G, Ortiz-Monasterio I, Cakmak I, Hao Y, Singh RP (2014) Biofortification strategies to increase grain zinc and iron concentrations in wheat. J Cereal Sci. 59: 365-372

Velu, G., Singh, R.P., Arun, B., Mishra, V.K., Tiwari, C., Joshi, A., Cherian, B., Virk, P., Pfeiffer, W.H. 2015. Reaching out to Farmers with High Zinc Wheat Varieties through Public-Private Partnerships – An Experience from Eastern-Gangetic Plains of India. Advances in Food Technology and Nutritional Sciences 1(3): 73-75

Wessells, K.R., Brown, K.H. 2012. Estimating the global prevalence of zinc deficiency: results based on zinc availability in national food supplies and the prevalence of stunting. PLoS ONE 7: e50568

Xu, Y., An, D., Li, H., Xu, H. 2011. Breeding wheat for enhanced micronutrients. Canadian Journal of Plant Science 91:231-237

Chapter 8:
Marker-Assisted Selection for Grain Yield and Developmental Traits

Susanne Dreisigacker, Sivakumar Sukumaran and Matthew P. Reynolds
CIMMYT Int., Apdo. Postal 6-641,06600 Mexico, DF, Mexico

Introduction

Grain yield is the most important trait plant breeders are interested in. It is reflecting the culmination of all the processes of vegetative and reproductive growth and development, and their interactions with the edaphic and aerial environments. The global average of grain yield in wheat is ~3.3t/ha (FAO, 2013). However, to meet the predicted global demand, grain yield will need to increase with expected gains of ~2% annually, a cumulative increase of 50% in ~20 years (Lopes et al. 2012).

To be able to predict the future land use and food supply, various studies have reported on the genetic gain for grain yield increase during the last decades and across diverse geographical regions, especially in the light of climate change, application of agronomic practices and agricultural policy. The reported studies identified contrasting results. A first group of studies showed a slowing yield growth during the last decades (Finger et al. 2010; Ray et al. 2013), a second group observed a yield plateau (Slafer and Peltonen-Sainio, 2001; Lobell and Field, 2007; Graybosch and Peterson, 2010; Fischer and Edmeades, 2010) with no statistically significant trend over a time period, and a third group of studies revealed a recent genetic gain for yield increase (Peltonen-Sainio et al. 2009; Mackay et al. 2011; Zheng et al. 2011). Genetic gains in yield of CIMMYT spring wheat in favorable environments averaged 0.6% annually between 1995 and 2010, based on data from hundreds of testing sites worldwide mainly through conventional breeding (Sharma et al. 2012). In Mexico, grain yield progress was significantly linear and about 0.7% annually in a set of CIMMYT historical lines representing 30 years of breeding (Lopes et al. 2012). Genetic progress for yield assessed globally in the semi-arid wheat yield trials of CIMMYT increased at approximately 1% annually over a 17-year period expressed as a percentage of the long-term check cultivar Dharwar Dry (Manés et al. 2012).

Most of the improvements in grain yield have arisen through incremental genetic advances. For example, wheat varieties with reduced plant height were introduced to the global wheat industry during the "Green Revolution." These varieties substantially improved grain yield through increased harvest index and straw strength (Borlaug 1968). In Australia, the production of early maturing, photoperiod-insensitive varieties allowed a significant expansion away from the fringes of the environmentally favorable eastern coast (Kuchel et al. 2007), leading to increased wheat production. Stably expressed genes leading to higher grain yield are therefore important targets of wheat breeding. In this chapter we summarize published information and our own work related to marker-assisted selection (MAS) for yield and related development traits.

Grain yield

Yield is a highly quantitative trait that is influenced by environmental and genetic interactions at all stages of the plant's growth. Yield is usually broken down into three components; number of spikes per area, grain number per spike and thousand grain weight (TGW). These yield components are sequentially fixed during the growth cycle, vary in terms of their heritability, and are not always positively correlated with yield. TGW usually shows stable heritability (Kuchel et al. 2007) and can be further broken down into individual components including physical parameters (grain length, width, thickness) and grain filling characteristics, which are also under independent genetic control. In the past decade, there have been significant advances in the understanding of the genetic control of grain size, shape and grain filling parameters in the diploid crop species, especially in rice (Ikeda et al. 2013). Several genes with relatively large effects have been identified through map-based cloning and support the independent genetic control of grain length, width and grain filling parameters. In wheat, there is still a limited understanding of grain weight genetic control. Many studies have identified quantitative trait loci (QTL) for TGW, grain size and shape (Gross et al. 2003; Kumar et al. 2006; Breseghello and Sorrells 2007; Tsilo et al. 2010; Rustgi et al. 2013; Sun et al. 2009; Zhang et al. 2010); but no gene has yet been cloned. Many of the observed QTL

are mapped to relatively wide genomic regions and have not been validated and fine-mapped and have therefore limited impact in breeding.

For some of the genes associated with TGW and grain shape cloned in rice, orthologs have been identified in wheat via comparative genetics. These genes play different roles in various stages of grain development and include: 1) sucrose synthase genes (*TaSus1* and *2*) which are correlated to dry matter accumulation (Hou et al. 2014); 2) cell wall invertase genes (*TaCwi-2A, -4A, -4B* and *-5D*) related to sink tissue development and carbon partitioning (Ma et al. 2012; Jiang et al. 2015); 3) *TaGW2* (*TaGw2-6A, 6B, 6D*), a orthologous gene to the rice gene *OsGW2* and associated with kernel width and weight by controlling endosperm cell number in both the cell division and late grain-filling phases (Su et al. 2011); 4) a cytokinin oxidase/dehydrogenase gene (*TaCKX6-D1*) that plays a principal role in controlling cytokinin levels and affects grain weight in wheat (Zhang et al. 2012); 5) *TaSAP1*, a member of the stress association protein (SAP) gene family in wheat associated with grain weight, number of grains per spike, spike length, and peduncle length in multiple environments (Chang et al. 2013); and 6) *TaGS-D1* and TaGASR7, two genes mainly related to grain length (Zhang et al. 2015; Dong et al. 2014).

The exact effect of these genes on TGW, grain size or shape in wheat is still not well understood to date. E.g. several studies have examined the role of *TaGW2* on grain size parameters and contradictory results have been reported. Two studies have described a SNP upstream of the putative start codon as significantly associated with wider grains and increased TGW in Chinese germplasm (Su et al. 2011; Zhang et al. 2013). However, each study found the positive association with the opposite SNP, and a negative association between *TaGW2* expression levels and grain width. Yang et al. (2012) identified a *TaGW2* frame-shift mutation in a large-grain variety, and associated this mutant allele with increased grain width and TGW in a large $F_{2:3}$ population. However, down-regulation of *TaGW2* through RNA interference (RNAi) resulted in decreased grain size and TGW in wheat, suggesting that *TaGW2* is a positive regulator of grain size.

At CIMMYT we have utilized molecular markers related to some of the published genes in four different germplasm sets (Table 1).

Frequencies of the published favorable alleles varied between the data sets and the most frequent favorable allele was the allele for *TaCWi-2A*. Initial analyses evaluating the effect of each of the alleles on TGW were inconclusive and only positively consistent for *TaCWi-2A* over all germplasm sets and in different environments (data not shown).

Simmonds et al. (2014) recently positioned the *TaGW2-6A* gene within the 6A QTL interval of the population Rialto × Spark. This QTL on chromosome 6A has been consistent in different populations and showed significant effects over seasons and environments (Snape et al. 2007). The QTL has been introgressed into CIMMYT germplasm via Spring × Winter wheat crosses using UK cultivars such as 'Premio' or 'Mercato.' The QTL has also been observed in a recent CIMMYT parent 'PFAU/SERI.1B//AMAD/3/WAXWING' which shows good yield stability. The QTL has most likely been introgressed via the line 'PFAU' which also has European lines in its pedigree. The effect of this QTL in CIMMYT germplasm has not yet been determined. Overall, therefore, it is too early to conclude if and how these genes affects TGW, grain size, and shape and additional experiments are required.

Table 1. Allele frequencies of the favorable allele of diverse published genes related to TKW in four different datasets.

Germplasm	Number of lines	*TaSus1*	*Tasus2-2B*	*TaCWi-2A*	*TaGW2-6A*
		% favorable allele			
Mexican Landraces	33	0.58	0.09	0.85	0.2
Turkish Landraces	153	1	0.37	0.86	0.01
Historical set of CIMMYT lines	54	0.39	0.00	0.83	0.13
Current CIMMYT elite lines	112	0.48	0.00	0.91	0.12

Another strategy to increase wheat yield is by reducing lodging. Lodging is a persistent phenomenon in wheat that reduces harvestable yield by up to 80% as well as reduces grain quality. Lodging is the permanent displacement of cereal stems from the vertical position and results from either plastic failure of the stem base (stem lodging) or failure of the anchorage system (root lodging). Lodging is difficult to measure on a phenotypic basis. Therefore MAS could be an important tool in breeding for lodging resistance. Plant height is strongly correlated with lodging and it is hard to identify other parameters associated. However, reduction in plant height may reduce the capacity of plant photosynthesis, thus reducing yield. So other mechanisms that reduce the lodging resistance – structural carbohydrates – needed to be studied. So far only one gene COMT gene, *TaCM* in wheat related to lignin strength (lignin is a structural component of cell wall that imparts strength and lodging resistance) has been studied as a source of lodging resistance H4564 (Ma 2009). Verma et al. (2005) observed several traits such as shoot and root traits and various components to yield were important for lodging and identified several QTL for each of the traits measured and indicated that they are controlled by several genes. It is likely that some of the lodging traits, e.g. the spread of the root plate, will not be found within elite germplasm and therefore wide crosses with novel germplasm might be required to achieve the target traits. Keller et al. (1999) also identified more than one QTL in a wheat × spelt (*Triticum spelta*, an ancient grain) population. Further work should therefore be carried out to better understand the genetic control of the traits and to investigate whether reliable genetic markers can be identified that work across a range of genotypes and environments and have a sufficiently large effect to be useful.

The number of tillers established in wheat far exceeds the number that end up being grain-bearing at maturity. Improving the economy in tillering has therefore also been proposed to improve wheat yields. A tiller inhibition mutant (*tin*) identified in a wheat from Israel is potentially useful for breeding varieties with a greater economy of tillering. However, the mutant has the pleiotropic effect of stunting under long day and low temperatures in some genetic backgrounds which has limited its use (Kebrom et al. 2013). Underlying genetic variation for tillering was also observed by QTL. In spring wheat, QTL with significant effect on tiller number were found to be located on chromosomes 6A, 1D and 3AL (Li et al. 2002; Shah et al. 1999). However, genotype-specific management practices may need to be developed to ensure the changed crop architecture associated with tin results in high yields.

More recently there has been an increasing focus on morpho-physiological dissection of grain yield, with the aim of improving grain production, particularly under abiotic stresses such as drought and heat. Dissection traits include e.g. early vigor, canopy temperature, Normalized Differential Vegetative Index or abscissic acid (ABA)-independent and ABA-dependent such as *Dreb1* or *Era1*. Research in this area is summarized in Chapter 8.

Plant phenology

Variation in expression to phenology is the most essential physiological adaptation of wheat to its cropping system. Archiving the appropriate plant phenology permits wheat varieties: 1) to fit into the time frame of the cropping cycle, 2) to avoid extreme weather events (e.g. frost, drought); and 3) to optimize the use of resources to maximize yield. The manipulation of plant phenology is therefore a common target. The genetic determination of plant phenology has demonstrated that it is a complex character which exhibits a continuous variation and is controlled by many genes scattered over the whole genome (Snape et al. 1996). In wheat the genetic bases of flowering time has been well studied and related genes have been classified according to whether they respond to vernalization or to photoperiod or to earliness *per se* during the pre-anthesis developmental phases. Stelmakh (1998) estimated that the vernalization gene system accounts for about 70-75%, the photoperiod gene system for about 20-25% and the earliness *per se* for about 5% of the genetic variability in the flowering time of bread wheat. Cane et al. (2013) estimated the effects of frequent alleles of *Vrn-1, Ppd-D1* and *Ppd-B1* genes on flowering time, which accounted for 53% of the genotypic variance of the trait. Diagnostic markers to identify the allelic variation of the genes have been developed (for a recent review see Kamran et al. 2014 and (Appendix 1) and can be used in breeding programs. A recent study by Chen et al. (2014) has shown that Pytochrome C and circadian

clock output genes play an additional role in long-day induced flowering in wheat. Pytochrome C operates by light activation upstream of the vernalization and photoperiod pathways and on the regulation of the circadian clock.

Vernalization requirement

Vernalization is the acquisition or acceleration of a plant's ability to flower by exposure to cold (Chouard, 1960). According to the vernalization requirements wheat is classified as being either winter or spring wheat. Spring wheat is insensitive or partly sensitive to vernalization, but winter wheat has a considerable vernalization requirement. Genetic differences are caused by allelic variation at *Vrn-1, Vrn-2, Vrn-3* and *Vrn-4* loci (Distelfeld et al. 2009). Spring and facultative wheat is manifested by the presence of one or more dominant alleles at *Vrn-1* which confer the insensitivity or partial sensitivity to vernalization. Winter wheat possesses dominant alleles at *Vrn-2*, a floral repressor which is considered to delay flowering until the plants are vernalized and recessive alleles at the other three loci (Trevaskis et al. 2007). Fu et al. (2005) sequenced the *Vrn-1* genes located on the homologues chromosomes 5 from diverse wheat accessions. Several *Vrn-1* alleles result from insertion and deletions in the promoter and intron-1 regions of the gene (Yan et al. 2004a, Yan et al. 2004b, Fu et al. 2005, Diaz et al. 2012). The role in altering vernalization response of each allele differs. While some alleles have large effects on the growth habit, others are silent mutations and therefore unlikely to have any direct role in the vernalization response. The *Vrn-3* genes mapped on the homologues chromosomes 7 promote the transcription of *Vrn-1*, thereby accelerating flowering time further. Genetic variation has been observed in *Vrn-B3* and *Vrn-D3* (Yan et al. 2006, Chen et al. 2010). Rather limited information is available for *Vrn-4*. So far only one allele have been described which was designated *Vrn-D4* and assigned to chromosome 5D (Kato et al. 2003, Yoshida et al. 2010).

At CIMMYT molecular markers linked to the *Vrn-1* and *Vrn-3* genes are utilized to evaluate CIMMYT advanced wheat lines (Table 2, Appendix 1). Summarizing the results of more than 1000 lines, the most common allele in CIMMYT wheat is the dominant spring allele *Vrn-D1a* (99%), followed by *Vrn-B1a* (70%). Stelmakh (1993) and Eagles et al. (2011) evaluated the genetic effect of the three *Vrn-1* genes and suggested a higher effect on heading date of *Vrn-D1* or *-A1* compared with *Vrn-B1*. The Japanese cultivar 'Akakomugi' is thought to be the donor parent of the *Vrn-D1a* allele (Stelmakh, 1990) which was later transferred into early Green Revolution cultivars like 'Lerma Rojo' and 'Sonora 64.' These two cultivars are thought to be the potential source of the *Vrn-D1a* allele in South and Southeast Asian wheat (Stelmakh, 1990; Van Beem et al. 2005). Stelmakh (1993) also concluded that the highest yield was predicted for varieties containing *Vrn-D1a*. *Vrn-A1a* is almost absent in CIMMYT wheat. Different recessive winter *vrn-A1* alleles (V and W) have been identified which are distinguished by a C/T SNP in the fourth exon of the gene and are also associated with copy number variation (Zhu et al. 2014). The 'W' allele is present in 80% of CIMMYT wheat lines. The allele is characterized by a higher copy number variation, a greater vernalization requirement and increased frost tolerance (Zhu et al. 2014). The allele was previously observed in CIMMYT 'Veery' lines and derivatives such as 'Attila' and 'Babax' (Eagles et al. 2011) and might have some adaptive advantage or is linked to another favorable allele. For *Vrn-3* genes, no variation was observed for *Vrn-B3* in CIMMYT wheat. However, the published allele of *Vrn-D3* from the cultivar 'Jagger' is present in 60% of the CIMMYT lines evaluated. The *Vrn-D3* allele further promotes development and according to Chen et al. (2010) maximize effects at physiological maturity. The effect of *Vrn-D3* on heading and maturity date could not be confirmed yet. Overall the most common haplotype at *Vrn-1* and *Vrn-3* in CIMMYT wheat is *vrn-A1W, Vrn-B1a, Vrn-D1a, Vrn-D3*. As compared to hexaploid wheat the major elite durum wheat gene pools show no major vernalization requirements (spring wheat) and functionally variant alleles are present at main loci for the photoperiod-sensitive response (Clarke et al. 1998).

Photoperiod sensitivity genes

Photoperiod-sensitive wheat is stimulated to flower only in long days and flowering is delayed under short days provided that any requirement for vernalization is met. In spring wheat, photoperiod- sensitive types cannot be grown as an overwinter crop in tropical or low latitude areas, since the day length requirement would not be satisfied in a short enough time-frame to produce a commercially viable crop (Worland and Snape, 2001). Photoperiod-insensitive wheat flowers

independently of day length and can be grown to maturity in long or short day environments. This is of particular advantage in warmer and dry climates as early flowering varieties are able to fill their grains prior to the onset of high temperatures and drought stress occurring late in the season (Worland and Snape, 2001). To date, three such genes have been identified, including Ppd-*A1, Ppd-B1* and *Ppd-D1* located on chromosomes 2A, 2B and 2D, respectively. The primary influence of the genes is on ear growth and spikelet growth (Scrath et al. 1985). A novel photoperiod response gene designated as *Ppd-B2* has been mapped on wheat chromosome 7BS (Khlestkina et al. 2009). This gene accelerates flowering only under long photoperiods in contrast to the *Ppd-1* genes that induce earlier flowering irrespective of day length. *Ppd-D1* is the photoperiod- insensitive locus with the largest effect followed by *Ppd-B1* and *Ppd-A1* (Worland, 1996). Photoperiod insensitivity are induced by In/Dels in the 5' upstream region of the gene pseudo-response regulator genes; they do not exist in the photoperiod-sensitive alleles (Beales et al. 2007, Wilhelm et al. 2009, Nishida et al. 2013). Furthermore Diaz et al. (2012) showed that for *Ppd-B1*, alleles conferring altered flowering time had an increased copy number of the gene and altered gene expression.

Ppd-D1a is predominant in CIMMYT wheat germplasm (Table 2). Across the same 1041 recent advanced lines described above, 95% of the lines carry the *Ppd-D1a* allele. Since its beginning by Norman Borlaug and his colleagues, the CIMMYT wheat program is based in Mexico and shuttles germplasm between two contrasting environments. This shuttle breeding exposes wheat materials to diverse photoperiod and temperatures, to a range of important diseases and led to the selection of photoperiod-insensitive lines. The *Ppd-A1a* alleles first described in durum wheat is present in 3 to 5% CIMMYT bread wheat germplasm. The allele was transferred from durum wheat via synthetic hexaploid wheat derivative that have been incorporated with increasing number into the bread wheat breeding programs (Dreisigacker et al. 2008).

Table 2. Allele frequency of reported *Vrn* and *Ppd* alleles in two sets of germplasm forming the international bread wheat screening nursery (IBWSN).

Trial	M45IWBSN			M46IBWSN		
Gene/Allele	No	No of lines	%	No	No of lines	%
Vrn-A1a	706	4	0.6	317	0	0.0
vrn-A1	706	702	99.4	317	317	100.0
Vrn-A1v	704	168	23.9	317	63	19.9
Vrn-A1w	704	536	76.1	317	254	80.1
Vrn-B1a	720	452	62.8	298	252	84.6
Vrn-B1b	720	20	2.8	298	16	5.4
vrn-B1	720	248	34.4	298	30	10.1
Vrn-D1a	706	702	99.4	319	319	100.0
vrn-D1	706	4	0.6	319	0	0.0
Vrn-D3a	698	452	64.8	321	187	58.3
vrn-D3	698	246	35.2	321	134	41.7
Ppd-A1a (GS105)	714	10	1.4	319	5	1.6
Ppd-A1b	713	5	0.7	298	2	0.7
Ppd-A1(Null)	701	145	20.7	318	133	41.8
Ppd-A1	714	554	77.6	319	181	56.7
Ppd-B1a (4x)	709	3	0.4	319	2	0.6
Ppd-B1a (3x)	720	497	69.0	308	207	67.2
Ppd-B1b (1x)	719	213	29.6	308	88	28.6
Ppd-B1b (1x)	720	6	0.8	308	10	3.2
Ppd-D1a	716	641	89.5	319	289	90.6
Ppd-D1b	705	76	10.8	319	30	9.4
Ppd-D1(null)	714	1	0.14	317	1	0.3

The *Ppd-B1* alleles show the largest variation in CIMMYT wheat. Most of the reported alleles were observed, e.g. the four copy number variant initially identified in 'Chinese Spring.' The most frequent allele is the three copy number variant first characterized in the Green Revolution line 'Sonora 64.'

In CIMMYT durum wheat segregation functionally variant alleles are present at the main loci *Ppd-A1* and *Ppd-B1* for the photoperiod-sensitive response. From the two insensitive alleles 'GS105' and 'GS100,' the former has been found to be more frequent. Molecular markers used to evaluate the CIMMYT germplasm for the major photoperiod alleles are described in Appendix 1.

Earliness

Earliness *per se* (*Eps*) genes are those that regulate flowering time independently of vernalization and photoperiod, and are important for the fine-tuning of flowering time and for the wide adaptation of wheat to different environments. Among the contributing factors influencing time to flowering, *Eps* has been least investigated. *Eps* loci have already been identified in wheat (Hoogendoorn 1985) and meta-QTL analysis of heading time in bread wheat revealed that numerous QTL co-located in chromosomal regions known to carry *Eps* loci (Hanocq et al. 2007; Griffiths et al. 2009). However, most of these *Eps* loci remained molecularly undefined, and only the *Eps-A^m1* locus in einkorn wheat has been fine mapped and phenotypically characterized (Faricelli et al. 2010). Gawronski and Schnurbusch (2012) recently fine mapped *Eps-A^m3*, a second gene derived from einkorn wheat.

Under combined vernalization and photoperiod treatments we identified an *Eps* QTL on chromosome 1DL using genome wide association mapping in the WAMI population genotyped with the 90K Wheat Illumina SNP array (Sukumaran personal communication). Subsequent BLAST searches indicated that the QTL region with sequence similarity identity higher than 96% contained the *Mot1* and *ELF3* genes that were candidates for *Eps* from earlier studies in einkorn wheat, so they are likely orthologues of *Eps-A^m1*. A recent study using four independent pairs of NILs derived from a cross between Spark and Rialto winter wheat varieties identified the same region on 1DL for *Eps* in wheat suggesting that MAS of *Eps* effects is getting feasible (Zikhali et al. 2014).

Plant height

Plant height is an important agronomic trait in cereal crops. It not only determines plant architecture but also contributes a lot to grain yield. The *Rht-B1b* (*Rht1*) and *Rht-D1b* (*Rht2*) semi-dwarfing genes were introduced into commercial wheat cultivars from the Japanese variety Norin10 in the 1960s as part of wheat improvement programs in the USA and at CIMMYT and led to the first Green Revolution wheat varieties. A reduction in plant height improved stem strength and thus lodging resistance and Harvest Index, the partitioning of assimilates to the developing grain (Borlaug, 1968). The large increases in yield that followed the introduction of these dwarfing genes led to widespread adoption of the dwarfing genes throughout the world (Gale et al. 1985). Perfect STS markers were developed in wheat for these genes (Ellis et al. 2002). *Rht1* and *Rht2* encode proteins involved in gibberellin signal transduction, but also have pleiotropic effects on plant growth, causing reductions in coleoptile length and seedling leaf area. These genes reduce the leaf elongation rate and coleoptile length. A number of alternative plant height genes have been observed that might be more suitable final plant height without compromising early plant growth (Ellis et al. 2004). Examples are *Rht4, Rht5* or *Rht8* that do not reduce the leaf elongation rate, coleoptile length and do not affect early growth. *Rht1* is predominant in CIMMYT germplasm due to the introgression of this gene during the Green Revolution. However, efforts are underway to incorporate some of the alternative alleles e.g. *Rht4, Rht5* or *Rht13* into CIMMYT germplasm. SNP KASP (Kompetitive Allele Specific PCR, www.lgcgenomics.com) assays were designed based on the by Ellis et al. (2002) developed markers and are routinely used to evaluated the CIMMYT elite germplasm. For the alternative drawing genes the markers reported in Ellis et al. (2005) are utilized (Appendix 1).

References

Beales J., A.Turner, S. Griffiths, J.W. Snape and D.A. Laurie. 2007. A Pseudo-Response Regulator is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum L.*). Theoretical and Applied Genetics 115:721–733

Borlaug, N. E. 1968.Wheat breeding and its impact on world food supply (pp. 1-36). Australian Academy of Science. Breseghello F, Sorrells ME (2007) QTL analysis of kernel size and shape in two hexaploid wheat mapping populations. Field Crop Research 101(2): 172–179

Cane K., H.A. Eagles, D.A .Laurie, B. Trevaskis, N. Vallance, R.F. Eastwood, N.N. Gororo, H. Kuchel and P.J. Martin. 2013. *Ppd-B1* and *Ppd-D1* and their effects in southern Australian wheat. Crop Pasture Science 64: 100-114

Chen A., C. Li, W. Hu, M.Y. Lau, H. Lin, N.C. Rockwell, S.S. Martin, J.A. Jernstedt, J.C. Lagarias and J. Dubcovsky. 2014. PHYTOCHROME C plays a major role in the acceleration of wheat flowering under long-day photoperiod. Proceedings of the National Academy of Sciences, USA 111: 10037–10044

Chen Y.H., B.F. Carver, S.W. Wang, S.H. Cao and L.L.Yan. 2010. Genetic regulation of developmental phases in winter wheat. Molecular Breeding 26: 573–582

Chang, J, J, Zhang, X. Mao, A. Li, J. Jia and R. Jing. 2013. Polymorphism of *TaSAP1-A1* and its association with agronomic traits in wheat. Planta 237(6): 1495-1508

Chouard P. 1960. Vernalization and its relation to dormancy. Annu Rev Plant Physiol 11: 191–237

Clarke JM, R.M. DePauw, L.L. Thiessen. 1998. Registrationof wheat genetic stocks near-isogenic for photoperiod sensitivity. Crop Sci 38: 882-882

Distelfeld A, C. Li and J. Dubcovsky. 2009. Regulation of flowering in temperate cereals. Curr Opin Plant Biol 12: 1-7

Díaz A, M. Zikhali, A.S. Turner, P. Isaac and D.A. Laurie. 2012. Copy number variation affecting the Photoperiod-B1 and Vernalization-A1 genes is associated with altered flowering time in wheat (*Triticum aestivum*). PLoSONE 7: e33234

Dong L, F. Wang, T. Liu, Z. Dong, A. Li, R. Jing, L. Mao, Y. Li, X. Liu, K. Zhang and D. Wang. 2014. Natural variation of affects grain length in common wheat under multiple cultivation conditions. Mol Breed 34 (3): 937-947

Dreisigacker S, M. Kishii, J. Lage and M. Warburton. 2008 . Use of synthetic hexaploid wheat to increase diversity for CIMMYT bread wheat improvement. Australian Journal of Agricultural Research 59: 413-420

Eagles H.A., K. Cane and B. Trevaskis. 2011. Veery wheats carry an allele of *Vrn-A1* that has implications for freezing tolerance in winter wheats. Plant Breed 130: 413-418

Ellis M.H., W. Spielmeyer, K. Gale, G. Rebetzke and R. Richards. 2002. Perfect markers for the *Rht-B1b* and *Rht-D1b* dwarfing genes in wheat. Theor Appl Genet 105 (6-7): 1038-1042

Ellis MH, G.J. Rebetzke, P. Chandler, D. Bonnett, W. Spielmeyer and R.A. Richards. 2004. The effect of different height reducing genes on the early growth of wheat Functional Plant Biology 31 (6): 583-589

Ellis MH, G.J. Rebetzke, F. Azanza, R.A. Richards and W. Spielmeyer. 2005. Molecular mapping of gibberellin-responsive dwarfing genes in bread wheat. Theor Appl Genet 111: 423–430

Faricelli ME, M. Valarik and J. Dubcovsky. 2010. Control of flowering time and spike development in cereals: the earliness per se *Eps-1* region in wheat, rice, and *Brachypodium*. Functional & Integrative Genomics 10: 293-306

FAO. 2013. Fisheries Statistics and Information. Fisheries and Aquaculture Department, Food and Agriculture Organization of the UN, Rome, Italy. Http://www.fao.org/fishery [accessed February 10 2015]

Finger R. 2010. Evidence of slowing yield growth – the example of Swiss cereal yields. Food Policy 35: 175-182

Fischer R.A. and G.O. Edmeades. 2010. Breeding and cereal yield progress. Crop Sci 50: S85–S98

Fu D, Szucs P, Yan L, Helguera M, Skinner JS, Zitzewitz J, Hayes PM, Dubcovsky J (2005) Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. Mol Gen Genom 273: 54–65

Gale M.D., S. Youssefian and G.E. Russell. 1985. Dwarfing genes in wheat. In: Progress in plant breeding, ed Russell GE, (Butterworths, London, UK), pp 1–35

Gawronski P. and T. Schnurbusch. 2012. High-density mapping of the earliness per se-3A^m (Eps-3A^m) locus in diploid einkorn wheat and its relation to the syntenic regions in rice and *Brachypodium* distachyon L.

Graybosch, R.A. and C.J. Peterson. 2010. Genetic improvement in winter wheat yields in the great plains of North America, 1959–2008. Crop Sci 50: 1882–1890

Griffiths S, J. Simmonds, M. Leverington, Y. Wang, L. Fish, L. Sayers, L. Alibert, S. Orford, L. Wingen, L. Herry, S. Faure, D. Laurie, L. Bilham and J. Snape. 2009. Meta-QTL analysis of the genetic control of ear emergence in elite European,winter wheat germplasm. Theor Appl Genet 119: 383–395

Groos C, N. Robert, E. Bervas and G. Charmet. 2003. Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. Theor Appl Genet 106 (6): 1032-1040

Hanocq E., A. Laperche, O. Jaminon, A.L. Laine and J. Le Gouis. 2007. Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis. Theor Appl Genet 114: 569–584

Hoogendoorn, J. 1985. A reciprocal *F1* monosomic analysis of the genetic control of time of ear emergence, number of leaves and number of spikelets in wheat (*Triticum aestivum* L.). Euphytica 34: 545–558

Hou J., Q.Y. Jiang, C.Y. Hao, Y.Q. Wang, H.N. Zhang and X.Y. Zhang. 2014. Global selection on sucrose synthase haplotypes during a century of wheat breeding. Plant Physiol 164: 1918–1929

Ikeda M., K. Miura, K. Aya, H. Kitano and M. Matsuoka. 2013. Genes offering the potential for designing yield-related traits in rice. Curr Opin Plant Biol 16(2): 213–220

Jiang Y., Q. Jiang, C. Hao, J. Hou, L. Wang, H. Zhang, S. Zhang, X. Chen and X. Zhang. 2015. A yield-associated gene *TaCWI*, in wheat: its function, selection and evolution in global breeding revealed by haplotype analysis." Theor Appl Genet 128 (1): 131-143

Kato K, M. Yamashita, K. Ishimoto, H. Yoshino and M. Fujita. 2003. Genetic analysis of two genes for vernalization response, the former *Vrn2* and *Vrn4*, by using PCR based molecular markers. In: Pogna NE, Romano M, Pogna EA, Galterio G (eds) Proceedings of the 10th international wheat genetics symposium. Instituto Sperimentale per la Cerealicoltura, Paestum (3): 971–973

Kamran A., M. Iqbal and D. Spaner. 2014. Flowering time in wheat (*Triticum aestivum* L.): a key factor for global adaptability. Euphytica 197: 1–26

Kebrom T.H. and R.A. Richards. 2013. Physiological perspectives of reduced tillering and stunting in the tiller inhibition (tin) mutant of wheat. Functional Plant Biology 40: 977-985

Keller M., C.H. Karutz, J.E. Schmid, P. Stamp, M. Winzeler, B. Keller and M.M. Messmer. 1999. Quantitative trait loci for lodging resistance in a segregating wheat × spelt population Theor Appl Genet 98:1171-1182

Khlestkina E.K., A. Giura, M.S. Roder, A. Borner. 2009. A new gene controlling the flowering response to photoperiod in wheat. Euphytica 165: 579–585

Kuchel H., K. J. Williams, P. Langridge, H.A. Eagles and S.P.Jefferies. 2007. Genetic dissection of grain yield in bread wheat. I. QTL analysis. Theor Appl Genet 115: 1029–1041

Kumar N., P.L. Kuwal, A. Gaur, A.K. Tyagi, J.P. Khurana, P. Khurana, H.S. Balyan and P.K. Gupta. 2006. QTL analyses for grain weight in common wheat. Euphytica 151: 135-144

Li WL, J.C. Nelson, C.Y. Chu , L.H. Shi, S.H. Huang and D.J. Liu. 2002. Chromosomal locations and genetic relationships of tiller and spike characters in wheat. Euphytica 125:357–366

Lobell D.B. and C.B. Field. 2007. Global scale climate-crop yield relationships and the impacts of recent warming. Environ Res Lett 2: 1-7

Lopes M.S. M.P. Reynolds, Y. Manes, R.P. Singh, J. Crossa and H.J. Braun. 2012. Genetic Yield Gains and Changes in Associated Traits of CIMMYT Spring Bread Wheat in a "Historic" Set Representing 30 Years of Breeding. Crop Sci 52: 1123

Ma Q.H. 2009. The expression of caffeic acid *3-O-methyltransferase* in two wheat genotypes differing in lodging resistance. J Exp Bot 60:2763–2771. doi: 10.1093/jxb/erp132

Ma D.Y., J. Yan, Z.H. He, L. Wu and X.C. Xia. 2012. Characterization of a cell wall invertase gene *TaCwi-A1* on common wheat chromosome 2A and development of functional markers. Mol Breeding 29: 43–52

Mackay I, A. Horwell, J. Garner, J. White, J. McKee and H. Philpott. 2011. Reanalyzes of the historical series of UK variety trials to quantify the contributions of genetic and environmental factors to trends and variability in yield over time. Theor Appl Genet 122: 225-238

Manés Y., H.F. Gomez, L. Puhl, M.P. Reynolds, H.J. Braun and R.M. Trethowan. 2012. Genetic yield gains of the CIMMYT international semi-arid wheat yield trials from 1994 to 2010. Crop Sci 52: 1543–1552

Nishida H., T. Yoshida, K. Kawakami, M. Fujita, B. Long, Y. Akashi, D.A. Laurie and K. Kato. 2013. Structural variation in the 5' upstream region of photoperiod-insensitive alleles *Ppd-A1a* and *Ppd-B1a* identified in hexaploid wheat (*Triticum aestivum L.*), and their effect on heading time. Mol. Breed. 31: 27–37

Peltonen-Sainio P., L. Jauhiainen and I.P. Laurila. 2009. Cereal yield trends in northern European conditions: Changes in yield potential and its realization. Field Crops Res 110: 85–90

Ray D.K., N.D. Mueller, P.C. West and J.A. Foley. 2013. Yield trends are insufficient to double global crop production by 2050. PLoS ONE 8: e66428

Rustgi S., M.N. Shafqat, N. Kumar, P.S. Baenziger, M.L. Ali, I. Dweikat, B.T. Campbell and K.S. Gill. 2013. Genetic dissection of yield and its component traits using high-density composite map of wheat chromosome *3A:* bridging gaps between *QTLs* and underlying genes. PLoS ONE 8 (7): e70526

Scarth R., E.J.M. Kirby and C.N. Law.1985. Effects of the photoperiod genes *Ppd1* and *Ppd2* on growth and development of the shoot apex in wheat. Annals of botany 55 (3): 351-359

Shah M.M., K.S. Gill, P.S. Baenziger, Y. Yen, S.M. Kaeppler and H.M. Ariyarathne. 1999. Molecular mapping of loci for agronomic traits on chromosome 3A of bread wheat. Crop Sci 39:1728–1732

Sharma R.C., J. Crossa, G. Velu, J. Huerta-Espino, M. Vargas, T.S. Payne and R.P. Singh. 2012. Genetic gains for grain yield in CIMMYT spring bread wheat across international environments. Crop Sci. 52: 1522-1533

Simmonds J., P. Scott, M. Leverington-Waite, A.S. Turner, J. Brinton, V. Korzun, J. Snape and C. Uauy. 2014. Identification and independent validation of a stable yield and thousand grain weight *QTL* on chromosome *6A* of hexapoid wheat (*Triticum aestivum L.*). BMC Plant Biology 14:191

Slafer, G.A. and P. Peltonen-Sainio P. 2001. Yield trends of temperate cereals in high latitude countries from 1940 to 1998. Agric Food Sci Finl 10: 121–131

Snape J.W., Quarrie SA, Laurie DA (1996) Comparative mapping and its use for the genetic analysis of agronomic characters in wheat. Euphytica 89: 27–31

Snape J.W., M.J. Foulkes, J. Simmonds, M. Leverington, L.J. Fish, Y. Wang and M. Ciavarrella. 2007. Dissecting gene × environmental effects of wheat via *QLT* and physiological analysis. Euphytica 154:401-408

Stelmakh A.F. 1990. Geographic distribution of *Vrn* genes in landraces and improved varieties of spring bread wheat. Euphytica 45: 113–118

Stelmakh A.F. 1993. Genetic effect of *Vrn* genes on heading date and agronomic traits in bread wheat. Euphytica 65: 53–60

Stelmakh A.F. 1998. Genetic systems regulating flowering response in wheat. Euphytica 100:359–369

Su Z, C. Hao, L. Wang, Y. Dong and X. Zhang. 2011. Identification and development of a functional marker of *TaGW2* associated with grain weight in bread wheat (*Triticum aestivum L.*). Theor Appl Genet 122(1): 211–223

Sun X-Y., K.Wu, Y. Zhao, F-M. Kong, G-Z. Han, H-M. Jiang, X-J. Huang, R-J. Li, H-G. Wang, S-S Li. 2009. QTL analysis of kernel shape and weight using recombinant inbred lines in wheat. Euphytica 165(3): 615–624

Trevaskis B., M.N. Hemming, E.S. Dennis and W.J. Peacock. 2007. The molecular basis of vernalization-induced flowering in cereals. Trends Plant Sci 12: 351–357

Tsilo TJ, G.A. Hareland, S. Simsek, S. Chao and J.A. Anderson. 2010. Genome mapping of kernel characteristics in hard red spring wheat breeding lines. Theor Appl Genet 121: 717–730

Van Beem J., V. Mohler, R. Lukman, M. Van Ginkel, M. William, J. Crossa and A.J. Worland. 2005. Analysis of genetic factors influencing the developmental rate of globally important CIMMYT wheat cultivars. Crop Sci 45: 2113–2119

Verma V., A.J. Worland, E.J. Sayers, L. Fish, D.S. Caligari, J. Snape. 2005. Identification and characterization of quantitative trait loci related to lodging resistance and associated traits in bread wheat. Plant Breeding 124: 234-241

Wilhelm, Edward P., A.S. Turner and D. A. Laurie. 2009. Photoperiod insensitive *Ppd-A1a* mutations in tetraploid wheat (*Triticum durum Desf.*). Theor Appl Genet 118 (2): 285-294

Worland T., J.W. Snape. 2001. Genetic basis of worldwide wheat varietal improvement. In: Bonjean AP, Angus WJ (eds) The world wheat book: a history of wheat breeding. Intercept Ltd., London, pp 61–67

Worland A.J. 1996. The influence of flowering time genes on environmental adaptability in European wheats. Euphytica 89: 49-57

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

Yan L., M. Helgura, K. Kato, S. Fukuyama, J. Sherman and J. Dubcovsky. 2004b. Allelic variation at the *Vrn-1* promoter region in polyploid wheat. Theor Appl Genet 109: 1677-1686

Yan L., D. Fu, C. Li, A. Blechl, G. Tranquilli, M. Bonafede, A. Sanchez, M. Valarik, S. Yasuda, J. Dubkovsky. 2006. The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. Proc Natl Acad Sci USA 103: 19581-19586

Yang Z., Z. Bai, X. Li, P. Wang, Q. Wu, L. Yang, L. Li and X. Li. 2012. SNP identification and allelic-specific PCR markers development for *TaGW2*, a gene linked to wheat kernel weight. Theor Appl Genet 125 (5): 1057–1068

Yoshida T., H. Nishida, J. Zhu, R. Nitcher, A. Distelfeld, Y. Akashi, K. Kato and J. Dubcovsky. 2010. *Vrn-D4* is a vernalization gene located on the centromeric region of chromosome 5D in hexaploid wheat. Theor Appl Genet 120: 543-552

Zhang D., B. Wang, J. Zhao, X. Zhao, L. Zhang, D. Liu, L. Dong, D. Wang, L. Mao and A. Li. 2015. Divergence in homoeolog expression of the grain length-associated gene GASR7 during wheat allo-hexaploidization. The Crop Journal 3 (1): 1-9

Zhang L., Y-L. Zhao, L-F. Gao, G-Y. Zhao, R-H. Zhou, B-S. Zhang and J-Z. Jia. 2012. TaCKX6-D1, the ortholog of rice OsCKX2, is associated with grain weight in hexaploid wheat. New Phytol 195: 574-584

Zhang L-Y, D-C. Liu, X-L. Guo, W-L. Yang, J-Z. Sun, D-W. Wang and A. Zhang. 2010 Genomic distribution of quantitative trait loci for yield and yield-related traits in common wheat. J Integr Plant Biol 52 (11): 996-1007

Zhang X, J. Chen, C. Shi, J. Chen, F. Zheng and J. Tian. 2013. Function of *TaGW2-6A* and its effect on grain weight in wheat (*Triticum aestivum L.*). Euphytica 192: 347–357

Zheng, T.C., X.K. Zhang, G.H. Yin, L.N. Wang, Y.L. Han, L. Chen, F. Huang, J.W. Tang, X.C. Xia, Z.H. He. 2011. Genetic gains in grain yield, net photosynthesis and stomatal conductance achieved in Henan province of China between 1981 and 2008. Field Crops Res 122: 225-233

Zhu J., S. Pearce, A. Burke, D.R. See, D.Z. Skinner, J. Dubcovsky and K. Garland-Campbell. 2014. Copy number and haplotype variation at the *VRN-A1* and central *FR-A2* loci are associated with frost tolerance in hexaploid wheat. Theor Appl Genet 127: 1183–1197

Zikhali M., M. Leverington-Waite, L. Fish, J. Simmonds, S. Orford, L.U. Wingen, R. Goram, N. Gosman, A. Bentley and S. Griffiths. 2014. Validation of a *1DL* earliness per se (eps) flowering QTL in bread wheat (*Triticum aestivum*). Mol Breed 34: 1023–1033

Chapter 9:
Marker-Assisted Selection for Abiotic Stresses in Wheat

Sivakumar Sukumaran, Susanne Dreisigacker and Matthew P. Reynolds
CIMMYT Int., Apdo. Postal 6-641,06600 Mexico, DF, Mexico

Introduction

Abiotic stresses affect plant development, productivity and grain quality in wheat. Research on plant responses to abiotic stresses and their impact on crop production continue to be a major focus in breeding; especially in the current scenario of climate change, climate-resilient wheat is a necessity. Among the four main abiotic stresses – drought, heat, salinity and metal toxicity – drought is the single largest threat to food security. Genetic studies to combat and adapt wheat to abiotic stresses through genomic-based approaches can reduce the time and cost of varietal development. However, given the complex genetic architecture of abiotic stress traits, their large genotype-by-environment (G × E) interaction, and the difficulty of phenotyping – especially for drought and heat tolerance – gene discovery and their application in MAS has so far been limited. Several initiatives to study these complex traits were attempted. Here, we address some of these studies.

Drought and heat tolerance

Plant responses to drought and heat stress are complex depending on the genotypes, environments and the G × E interaction. In addition, the difficulty to identify QTLs for traits under drought and heat are: (1) the availability of mapping population with controlled height and phenology to avoid confounding effect of major genes; and (2) the phenotyping procedure in a time frame in large populations that will avoid confounding masking effect of major genes on minor genes (Reynolds and Tuberosa, 2008).

Drought tolerance phenotyping can be realized indirectly by measurements of morpho-physiological traits, mainly: water use (WU); water use efficiency (WUE); carbon partitioning to grain; carbon isotope discrimination to determine transpiration efficiency; canopy temperature (CT); green leaf area; stay green; water soluble carbohydrates (WSC); above ground biomass; grain yield and root parameters; root biomass; rooting depth; and root development under drought conditions (Foulkes et al. 2007, Rashid et al. 1999). At present, these traits are followed in CIMMYT to perform trait-based crosses to combine the high value alleles. Parents with contrasting desired physiological traits are selected from available evaluated germplasm and crosses are made with recurrent parents or elite lines. The developed lines are then tested under different environmental conditions.

At the candidate gene level, among the most common gene networks and pathways related to drought tolerance, Abscissic acid (ABA)-dependent and ABA-independent pathways are the most studied. Abscissic acid is a plant growth regulator and stress hormone, which induces leaf stomata closure to reduce water loss through transpiration and decreases the photosynthetic rate in order to improve the WUE of plants. A major QTL affecting drought-induced ABA accumulation was located on chromosome 5A in wheat and examples of source genotypes are the cultivars 'Ciano 67' and 'SQ1' (Quarrie et al. 1994). Furthermore, many families of transcription factors have been demonstrated to play a role in stress responses in plants. *bZIP, DREB, WRKY, bHLH, MYB* and *NAC* transcription factors represent the major groups of regulatory genes of which some members are found to be involved in wheat stress tolerance. A very limited number of markers are developed for these genes. Wei et al. (2009) developed a functional marker for the *Dreb-B1* gene involved in abiotic stress tolerance, Chang et al. (2013) identified linked markers to the gene thylakoid bound ascorbate peroxidase *TaSAP-A1*. Ascorbate peroxidase enzymes play a key role in detoxifying the reactive oxygen species that can cause damage to the cells (Caverzan et al. 2012).

Further international efforts to detect QTLs for grain yield under drought stress conditions have been made with some success. Fleury et al. (2010) summarized more than 20 QTL for drought in wheat. To give an example: a QTL on chromosome 3BL was detected under heat, drought, and high yield potential conditions that explained up to 22% of the variance for grain yield and canopy temperature (Bennett et al. 2012). The same QTL on chromosome 3B was also

associated with grain yield in the studies (Bonneau et al. 2013, Sukumaran et al. 2015).

Heat stress mostly occurs in combination with drought and the combined effect of drought and heat is more severe than any one of the stresses. Higher temperature above 30°C at grain filling period is detrimental to wheat crop yield. For heat tolerance studies, several traits have shown promises *viz*. light interception traits, rapid ground cover, canopy structure, radiation use efficiency, stay green, photosynthesis and reduced photorespiration, photo protective metabolites, wax, membrane thermostability, spike fertility, water soluble carbohydrate, starch synthesis and plant signaling (Cossani and Reynolds, 2012).

While confronting high temperature stress and alleviation from damage of cellular protein structure is essential for survival in stressed conditions, plants trigger a novel class of protein called heat shock proteins (HSPs). These HSPs serve as molecular chaperones to maintain conformational protein functions as well as cellular protein refolding, thereby protecting plants under heat stress conditions (Wang et al. 2014). So far only one attempt has been made to identify single nucleotide polymorphisms (SNPs) that differentiate heat-tolerant and heat-susceptible genotypes of wheat analyzing the heat shock protein HSP16.9 as the target gene. DNA fragments covering a partial sequence of wheat HSP16.9, were amplified from the heat-tolerant genotype 'K7903' and heat-susceptible genotype'RAJ4014,' and subsequently analyzed for the presence of SNPs. One SNP was found between these genotypes and the analysis of the corresponding amino acid sequence showed that the base transition (A/G) positioned at 31 amino acid resulted in a missense mutation from *aspartic acid to asparagine* residue (Garg et al. 2012). Allele-specific primers based on SNP explained 29% and 24% phenotypic variation for grain weight and thousand grain weights, respectively.

Despite the importance of heat tolerance only a few studies have focused to identify QTL via genome wide scans. Yang et al (2002) found QTL linked to grain filling duration on the short arms of chromosomes1B and 5A. Vijayalakshmi et al. (2010) reported QTL with significant effects on grain yield, grain weight, grain filling, stay-green and senescence-associated traits on 2A, 3A, 4A, 6A, 6B and 7A under post-anthesis high temperature stress in wheat. Heat susceptibility index (HSI) calculated from agronomic traits associated with heat stress tolerance and QTL were mapped on chromosomes 1A, 2A, 2B, and 3B for HSI calculated from the kernel characteristics under stress conditions applied during early grain filling stage that explained up to 31% of the variation in the traits (Mason et al. 2010). QTL were mapped for heat stress using the Fischer susceptibility index on chromosomes 1B, 5B, and 7B that explained up to 44% of the variation in the traits (Mohammadi et al. 2010). QTL mapping for terminal heat stress has identified QTL in chromosome 2B, 7B, and 7D that co-localized for kernel weight, grain filling duration, and canopy temperature difference. During flowering, higher temperature can cause pollen abortion and subsequently low yield. QTLs for heat stress tolerance was identified in a cross on cultivar NW1014 (heat-tolerant) and HUW468 (heat-susceptible) using HSI of grain weight, grain fill duration, grain yield, and canopy temperature depression on chromosomes 2B, 7B, and 7D. These explained up to 20% of the phenotypic variation for the traits (Paliwal et al. 2012). Mondal et al. (2015) identified QTL for leaf wax content located on chromosomes 1B and 5A with the 5A QTL region showing localization with QTL for leaf and spike temperature depression, indicating a genetic link between these traits. Composite interval mapping by the study of Talukder et al. (2014) identified five QTL regions significantly associated with response to heat stress. Associations were identified for plasma membrane damage on chromosomes 7A, 2B and 1D, SPAD chlorophyll content on 6A, 7A, 1B and 1D and thylakoid membrane damage on 6A, 7A and 1D. The variability explained by these QTL ranged from 11.9 to 30.6% for thylakoid membrane damage, 11.4 to 30.8% for SPAD chlorophyll content, and 10.5 to 33.5% for plasma membrane damage.

The plant developmental genes in wheat for vernalization and photoperiod (*Vrn* and *Ppd*, respectively) are related to the performance of the lines under drought and heat stress (Bogard, et al. 2014). Therefore, care must be taken to avoid these effects in gene discovery when developing mapping populations. At CIMMYT, to study the genetic basis of drought and heat tolerance and to make strategic crosses for trait integration and line development, mapping populations are available with restricted

phenology and plant height (Sukumaran et al. 2013). The Seri × Babax recombinant inbred line (RIL) population was the first of this type of population developed with a phenology range of 10 days. Using this Seri × Babax population QTL were identified for drought and heat tolerance on chromosome 4A that co-localized with grain yield. Common QTLs for drought and heat tolerance were identified on 1B, 2B, 3B, 4A, and 7A using canopy temperature measurements (Pinto et al. 2010). Canopy temperature measurements are surrogates for estimating stomatal conductance (Rebetzke et al. 2013).

The Wheat Association Mapping Initiative (WAMI) population was subsequently created and consists of 287 lines selected from a series of CIMMYT international nurseries that has a phenology range of 14 days. The CIMMYT WAMI population was studied through genome-wide association study (GWAS) at the population level and by candidate gene approach (Edae et al. 2014, Edae et al. 2013, Lopes et al. 2015, Sukumaran et al. 2015). Edae et al. (2013) performed a candidate gene study and confirmed the effects of the dehydration responsive element binding 1A (*DREB1A*) gene on NDVI, heading date, biomass, and spikelet number, the enhanced response to abscissic acid (*ERA1-A* and *ERA1-B*) genes on harvest index, flag leaf senescence, and flag leaf width, and the fructan 1-exohydrolase (*1-FEH-A* and *1-FEH-B*) genes associated with grain yield and thousand kernel weight, respectively. A consistent QTL in chromosome 2DS for grain yield and yield components under contrasting moisture conditions in the U.S. and Ethiopia was additionally identified through GWAS (Edae et al. 2014). On chromosomes 5A and 6A, loci for grain yield, thousand kernel weight, grain number, and canopy temperature were detected in the WAMI grown at yield potential conditions in Mexico (Sukumaran et al. 2015).

A recently developed RIL population at CIMMYT – Synthetic × Weebil – has a phenology range of three days that was phenotyped under drought and heat conditions. Identifying QTL for heat and drought tolerance in these populations will be more independent of the confounding effects of phenology. This population was phenotyped though 90K Illumina Bead chip array (Wang et al. 2014) and the research is in progress to detect QTL. The´detected QTL in the phenology controlled populations is recent and the validation of the identified QTL is only underway. Based on the current knowledge obtained on the underlying mechanisms and architecture of heat and drought tolerance, the applicability of MAS for both traits has to been questioned to its inability to capture small effect QTL. A promising approach, termed genomic selection, attempts to avoid this deficiency by capturing both large and small-effect QTL with dense genome-wide molecular marker coverage to predict complex trait values (Meuwissen et al. 2001).

Among the physiological traits the stay-green phenotype has also shown proven utility to improve yields under abiotic stress. Stay green governs long grain-filling period and higher yields. In wheat, stay green QTL were mapped to 2A, 3A, 3B, 6A, 6B, and 7A for the traits chlorophyll content, chlorophyll fluorescence, and green leaf area duration (Vijayalakshmi et al. 2010). The stability of thylakoid membrane proteins and antioxidant competence contribute to the drought stress resistance in a wheat stay green mutant *tasg1*. Genes *TaLhcb4* and *TaLhcb6* were higher expressed in the mutant than the wild type plants (Tian et al. 2013). QTL for stay green were identified on chromosomes 1A, 3BS, and 7DS derived from the parent Chirya 3; a synthetic wheat developed by CIMMYT (Kumar et al. 2010). Stay green studies in CIMMYT on the Seri × Babax population focus on measurement of the traits using NDVI and SPAD meter (Lopes and Reynolds, 2012).

Breeding for root traits is an order of magnitude more difficult than for most above-ground traits. Root characteristics are important in drought tolerance. A deep and thick root system is associated positively with grain yield under drought conditions. Twenty nine QTLs were identified for the root traits seminal root number, total root length, maximum root length; root area, root surface area, and seminal root angle (Liu et al. 2013). The drought-tolerant variety Seri M82 had a compact root system compared to the susceptible variety Hartog (Manschadi et al. 2006). Coleoptile length is another feature that is important in drought tolerance and QTLs were identified on chromosomes 1B, 3D, 4DS, 4DL, 5AS, and 5B (Yu and Bai, 2010). Spielmeyer et al. (2007) identified a QTL on chromosome 6A in bread wheat associated with longer coleoptiles, greater seeding vigor and final plant height.

To breed for abiotic stress, an important source of drought and heat tolerance is the 1B.1R rye translocation in spring wheat. This 1B.1R translocation delayed heading and maturity by 7 and 5 days in a study using Seri 82 genotype (Singh et al. 1998). Positive performance is reported for translocations of chromosome 1 of rye in bread wheat. 1RS translocations in 'Pavon' delayed maturity, reduced plant height, and increased root biomass. It also increased grain yield and grain weight under well-watered conditions. The 1RS translocations were tolerant to moisture stress (Ehdaie et al. 2003). Even though the rye translocations are associated with high grain yield, stress and disease tolerance they are also associated with poor bread-making quality: sticky dough and weak gluten. Recently, a region within the 1RS.1BL translocation affecting grain yield and canopy water status was mapped that was not associated with low bread-making quality (Howell et al. 2014). 1BL.1RS translocation is widely reported in Indian spring wheat genotypes. The association of the 1B.1R region with high yield is associated with high root biomass. A QTL mapping study identified the terminal 15% of the rye 1RS arm carries genes for greater rooting ability (Sharma et al. 2009). Apart from the studies to find tolerance to drought and heat stress on breeding lines, integrating novel traits from wild wheat could improve drought and heat stress in wheat (Placido et al. 2013).

Salinity

Excess dissolved salts in the soil can inhibit plant growth. This can occur due to natural processes (e.g. drought) or due to human activities (e.g. application of poor irrigation water). Of the 17% irrigated crop land in the world, 8% of the land area is affected by salinity and in arid and semi-arid regions, it is up to 25%, which creates significant consequences to world food production. Salinity effects the seedling germination (Fogle and Munns, 1973), grain yield and quality (Turki et al. 2012) and yield components (Nia et al. 2012). Bread wheat is a moderately salt-tolerant crop whereas durum wheat is less salt-tolerant (Maas and Hoffman, 1977, Munns et al. 2006).

Main mechanisms of salt tolerance involve Na+ exclusion from the transpiration stream, sequestration of Na+ and Cl− in the vacuoles of root and leaf cells, and other processes that promote fast growth despite the osmotic stress of the salt outside the roots. Bread wheat has a low rate of Na+ accumulation and enhanced K+/Na+ discrimination, a character controlled by a locus (*Kna1*) on chromosome 4D (Dubcovsky et al. 1996). Two loci, *Nax1* and *Nax2* on chromosome 2A and 5A, ,respectively, controlling Na+ accumulation has been found in the durum genotype 'Line 149' (Shaw et al. 2012). Both genes are not present in modern durum wheat and appear to originate from a wheat relative, *Triticum monococcum* (C68-101) that was crossed with a durum wheat to transfer rust resistance genes, and this cross inadvertently transferred the *Nax* genes into Line 149 as well (James et al. 2006). These genes were therefore named *TmHKT7* (*TmHKT1;4-A2*) and *TmHKT8* (*TmHKT1;5-A*) to recognize their origin in *Triticum monococcum*. Linked molecular markers have been developed for these two genes, but have not been applied in CIMMYT wheat breeding yet. Salinity-responsive bread wheat gene *TaAOC1*, which encodes an allene oxide cyclase involved in the α-linolenic acid metabolism pathway, was constitutively expressed in both bread wheat and Arabidopsis (*Arabidopsis thaliana*) (Zhao et al. 2014). QTL mapping studies have identified 47 QTLs mapping to all wheat chromosomes except 1B, 1D, 4B, 5D and 7D for salinity tolerance (Ma et al. 2006, Xu et al. 2013).

Metal toxicity

Wheat is susceptible to excess amounts of aluminum (Al), boron (B), cadmium (Cd) and copper (Cu). Out of these, under low pH, Al is the most prevalent and most toxic to wheat plants (Delhaize and Ryan, 1995, Kochian et al. 2005). Al tolerance is polygenic and controlled by at least three genes. *TaALMT1* is a dominant gene which encodes a malate transporter on chromosome 4D constitutively expressed on root apices. Raman et al. (2006) developed SSR markers, *ALMT1-SSR3a* and *ALMT1-SSR3b* and a CAPS marker from the repetitive InDels and substitution region of the *TaALMT1* gene which can be used in MAS, but has not yet been applied in CIMMYT (Raman et al. 2006, Raman et al. 2005). A minor QTL on 3BL was detected by (Zhou et al. 2007) in recombinant inbred lines (RILs) using 'Atlas 66' as tolerant parent (Ma et al. 2005, Tang et al. 2002). Similarly Al tolerance QTL on chromosome 3BL was also contributed by 'Chinese Spring' (Navakode et al. 2009). QTLs on chromosome 3BL and 2A apart from the major effect gene on chromosome 4D, collectively explained 80%

of the phenotypic variation (Cai et al. 2008, Dai et al. 2013).

Boron toxicity occurs when plants are grown in alkaline or volcanic soils. Boron has the narrowest range between deficient and toxic soil solution concentration of all plant nutrients. Boron toxicity in wheat can cause poor root growth, low above ground biomass, low seed set and sterility, and low grain yield (Pallotta et al. 2014). *TaBot1L* (*Bo1*), and *Bo4* are the two major effect QTLs for boron tolerance in wheat. The utilization of *Bo1* on the long arm of chromosome 7BL has been a long-term priority for marker-assisted selection in wheat breeding programs in Australia. At CIMMYT the STS marker *AWW5L7* (Schnurbusch et al. 2010, Scoles et al. 2008) is used. Sources of resistance were e.g. the Australian line 'Gladius.' *Bo4* is located on chromosome 4AL and was recently placed with the marker interval *Xabg390-4A–XksuG10-4A* (Pallotta et al. 2014).

Cadmium is a toxic metal that is naturally present in trace quantities in almost all soils, but is a non-essential plant element. Cadmium in soil is readily absorbed by roots and transported in plants (Grant et al. 1998). Cadmium-contaminated foods are the dominant source of human exposure to environmental Cd (Satarug et al. 2011, Satarug and Moore, 2004), with cereals and vegetables contributing the majority of dietary Cd (McLaughlin and Singh, 1999). Among cereals, some durum wheat (*Triticum turgidum* L. var durum) cultivars have the genetic potential to accumulate Cd in grain to levels that exceed the Codex standard (Grant et al. 2008). Marker-assisted selection is preferred for selecting breeding lines expressing low Cd because measuring grain Cd is laborious and expensive relative to PCR-based screening. Genetic studies identified a gene *Cdu1* in chromosome 5BL responsible for accumulation of Cd in the grains of durum (Penner et al. 1995). We have tested the marker *ScOpc20* and *usw47* (Knox et al. 2009, Wiebe et al. 2010) for evaluation of CIMMYT durum wheat germplasm.

Copper toxicity can cause yield losses by reduced fertility, by affecting germination and root and shoot elongation. Cu is also reported to inhibit root growth (Fargašová, 2001), damages to chromosome, nuclei, and cell membranes (Jiang et al. 2001), and chlorosis (Eleftheriou and Karataglis, 1989, Ganeva et al. 2003). Genomic regions associated with copper tolerance has been found in chromosomes 1AL, 2DS, 3DS, 4AL, 5AL, 5DL, 5BL, and 7DS (Bálint et al. 2007). Gene for copper tolerance *cbf1* may be pleotropically affected by *Vrn-1* loci (Bálint et al. 2009). Cu toxicity is sometimes overestimated in acidic soils with the presence of Aluminum and Manganese.

References

Bálint, A.F., M.S. Röder, R. Hell, G. Galiba and A. Börner. 2007. Mapping of *QTLs* affecting copper tolerance and the Cu, Fe, Mn and Zn contents in the shoots of wheat seedlings. Biologia Plantarum 51: 129-134

Bálint, A.F., F. Szira, M.S. Röder, G. Galiba and A. Börner. 2009. Mapping of loci affecting copper tolerance in wheat—The possible impact of the vernalization gene *Vrn-A1*. Environmental and Experimental Botany 65: 369-375

Bennett, D., M. Reynolds, D. Mullan, A. Izanloo, H. Kuchel, P. Langridge, et al. 2012. Detection of two major grain yield *QTL* in bread wheat (*Triticum aestivum L.*) under heat, drought and high yield potential environments. Theoretical and Applied Genetics 125: 1473-1485

Bogard, M., C. Ravel, E. Paux, J. Bordes, F. Balfourier, S. Chapman, et al. 2014. Predictions of heading date in bread wheat *(Triticum aestivum L.*) using *QTL*-based parameters of an ecophysiological model. Journal of experimental botany: eru328

Bonneau, J., J. Taylor, B. Parent, D. Bennett, M. Reynolds, C. Feuillet, et al. 2013. Multi-environment analysis and improved mapping of a yield-related *QTL* on chromosome *3B* of wheat. Theoretical and Applied Genetics 126: 747-761

Cai, S., G.-H. Bai and D. Zhang. 2008. Quantitative trait loci for aluminum resistance in Chinese wheat landrace *FSW*. Theoretical and Applied Genetics 117: 49-56

Caverzan, A., G. Passaia, S.B. Rosa, C.W. Ribeiro, F. Lazzarotto and M. Margis-Pinheiro. 2012. Plant responses to stresses: role of ascorbate peroxidase in the antioxidant protection. Genetics and molecular biology 35: 1011-1019

Chang, J., J. Zhang, X. Mao, A. Li, J. Jia and R. Jing. 2013. Polymorphism of *TaSAP1-A1* and its association with agronomic traits in wheat. Planta 237: 1495-1508

Cossani, C.M. and M.P. Reynolds. 2012. Physiological traits for improving heat tolerance in wheat. Plant physiology 160: 1710-1718

Dai, J., G. Bai, D. Zhang and D. Hong. 2013. Validation of quantitative trait loci for aluminum tolerance in Chinese wheat landrace *FSW*. Euphytica 192: 171-179

Delhaize, E. and P.R. Ryan. 1995. Aluminum toxicity and tolerance in plants. Plant physiology 107: 315.

Dubcovsky, J., G. Santa Maria, E. Epstein, M.-C. Luo and J. Dvořák. 1996. Mapping of the *K+/Na+* discrimination locus *Kna1* in wheat. Theoretical and Applied Genetics 92: 448-454

Edae, E.A., P.F. Byrne, S.D. Haley, M.S. Lopes and M.P. Reynolds. 2014. Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. Theoretical and Applied Genetics 127: 791-807

Edae, E.A., P.F. Byrne, H. Manmathan, S.D. Haley, M. Moragues, M.S. Lopes, et al. 2013. Association mapping and nucleotide sequence variation in five drought tolerance candidate genes in spring wheat. The Plant Genome 6: 2

Ehdaie, B., R. Whitkus and J. Waines. 2003. Root biomass, water-use efficiency, and performance of wheat–rye translocations of chromosomes 1 and 2 in spring bread wheat *'Pavon'*. Crop Science 43: 710-717

Eleftheriou, A. and S. Karataglis. 1989. Ultrastructural and Morphological Characteristics of Cultivated Wheat Growing on Copper-Polluted Fields. Botanica acta 102: 134-140

Fargašová, A. 2001. Phytotoxic effects of Cd, Zn, Pb, Cu and Fe on Sinapis alba L. seedlings and their accumulation in roots and shoots. Biologia Plantarum 44: 471-473

Fleury, D., S. Jefferies, H. Kuchel and P. Langridge. 2010. Genetic and genomic tools to improve drought tolerance in wheat. Journal of Experimental Botany 61: 3211-3222.

Fogle, V. and D. Munns. 1973. Effect of salinity on the time course of wheat seedling growth. Plant physiology 51: 987.

Foulkes, M., R. Sylvester-Bradley, R. Weightman and J. Snape. 2007. Identifying physiological traits associated with improved drought resistance in winter wheat. Field Crops Research 103: 11-24

Ganeva, G., S. Landjeva and M. Merakchijska. 2003. Effects of chromosome substitutions on copper toxicity tolerance in wheat seedlings. Biologia Plantarum 47: 621-623

Garg, D., S. Sareen, S. Dalal, R. Tiwari and R. Singh. 2012. Heat shock protein based SNP marker for terminal heat stress in wheat (*Triticum aestivum L.*). Australian Journal of Crop Science 6: 1516-1521

Grant, C., W. Buckley, L. Bailey and F. Selles. 1998. Cadmium accumulation in crops. Canadian Journal of Plant Science 78: 1-17

Grant, C., J. Clarke, S. Duguid and R. Chaney. 2008. Selection and breeding of plant cultivars to minimize cadmium accumulation. Science of the Total Environment 390: 301-310

Howell, T., I. Hale, L. Jankuloski, M. Bonafede, M. Gilbert and J. Dubcovsky. 2014. Mapping a region within the *1RS*. 1BL translocation in common wheat affecting grain yield and canopy water status. Theoretical and Applied Genetics 127: 2695-2709

James, R.A., R.J. Davenport and R. Munns. 2006. Physiological characterization of two genes for *Na+* exclusion in durum wheat, Nax1 and Nax2. Plant Physiology 142: 1537-1547

Jiang, W., D. Liu and X. Liu. 2001. Effects of copper on root growth, cell division, and nucleolus of Zea mays. Biologia Plantarum 44: 105-109

Knox, R., G. Scoles, C. Pozniak, F. Clarke, J. Clarke, S. Houshmand, et al. 2009. Chromosomal location of the cadmium uptake gene (*Cdu1*) in durum wheat. Genome 52: 741-747

Kochian, L.V., M.A. Pineros and O.A. Hoekenga. 2005. The physiology, genetics and molecular biology of plant aluminum resistance and toxicity. Root Physiology: From Gene to Function. Springer.pp. 175-195

Kumar, U., A.K. Joshi, M. Kumari, R. Paliwal and S. Kumar. 2010. Identification of *QTLs* for stay green trait in wheat (*Triticum aestivum L.*) in the *'Chirya 3'×'Sonalika'* population. Euphytica.174: 437-445

Liu, X., R. Li, X. Chang and R. Jing. 2013. Mapping *QTLs* for seedling root traits in a doubled haploid wheat population under different water regimes. Euphytica 189: 51-66

Lopes, M.S., S. Dreisigacker, R.J. Peña, S. Sukumaran and M.P. Reynolds. 2015. Genetic characterization of the wheat association mapping initiative (*WAMI*) panel for dissection of complex traits in spring wheat. Theoretical and Applied Genetics 128: 453-464

Lopes, M.S. and M.P. Reynolds. 2012. Stay-green in spring wheat can be determined by spectral reflectance measurements (normalized difference vegetation index) independently from phenology. Journal of experimental botany. 63(10) 3789-3798

Ma, H.-X., G.-H. Bai, B. Carver and L.-L. Zhou. 2005. Molecular mapping of a quantitative trait locus for aluminum tolerance in wheat cultivar Atlas 66. Theoretical and Applied Genetics 112: 51-57

Ma, L., E. Zhou, N. Huo, R. Zhou, G. Wang and J. Jia. 2006. Genetic analysis of salt tolerance in a recombinant inbred population of wheat (*Triticum aestivum L.*). Euphytica 153: 109-117

Maas, E. and G. Hoffman. 1977. Crop salt tolerance\-current assessment. Journal of the irrigation and drainage division 103: 115-134

Manschadi, A.M., J. Christopher and G.L. Hammer. 2006. The role of root architectural traits in adaptation of wheat to water-limited environments. Functional Plant Biology 33: 823-837

Mason, R.E., S. Mondal, F.W. Beecher, A. Pacheco, B. Jampala, A.M. Ibrahim, et al. 2010. *QTL* associated with heat susceptibility index in wheat (*Triticum aestivum L.*) under short-term reproductive stage heat stress. Euphytica 174: 423-436

McLaughlin, M.J. and B.R. Singh. 1999. Cadmium in soils and plants Springer Netherlands. pp 1-9

Meuwissen, T.H., B.J. Hayes and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-1829

Mohammadi, V., A. Zali and M. Bihamta. 2010. Mapping *QTLs* for heat tolerance in wheat. Journal of Agricultural Science and Technology 10: 261-267

Mondal, S., R.E. Mason, T. Huggins and D.B. Hays. 2015. QTL on wheat (*Triticum aestivum L.*) chromosomes 1B, 3D and 5A are associated with constitutive production of leaf cuticular wax and may contribute to lower leaf temperatures under heat stress. Euphytica 201: 123-130

Munns, R., R.A. James and A. Läuchli. 2006. Approaches to increasing the salt tolerance of wheat and other cereals. Journal of Experimental Botany 57: 1025-1043

Navakode, S., A. Weidner, U. Lohwasser, M. Röder and A. Börner. 2009. Molecular mapping of quantitative trait loci (*QTLs*) controlling aluminium tolerance in bread wheat. Euphytica 166: 283-290

Nia, S.H., M.J. Zarea, F. Rejali and A. Varma. 2012. Yield and yield components of wheat as affected by salinity and inoculation with *Azospirillum* strains from saline or non-saline soil. Journal of the Saudi Society of Agricultural Sciences 11: 113-121

Paliwal, R., M.S. Röder, U. Kumar, J. Srivastava and A.K. Joshi. 2012. *QTL* mapping of terminal heat tolerance in hexaploid wheat (*T. aestivum L.*). Theoretical and Applied Genetics 125: 561-575

Pallotta, M., T. Schnurbusch, J. Hayes, A. Hay, U. Baumann, J. Paull, et al. 2014. Molecular basis of adaptation to high soil boron in wheat landraces and elite cultivars. Nature. 514: 88-91

Penner, G.A., L.J. Bezte, D. Leisle and J. Clarke. 1995. Identification of *RAPD* markers linked to a gene governing cadmium uptake in durum wheat. Genome 38: 543-547

Pinto, R.S., M.P. Reynolds, K.L. Mathews, C.L. McIntyre, J.-J. Olivares-Villegas and S.C. Chapman. 2010. Heat and drought adaptive *QTL* in a wheat population designed to minimize confounding agronomic effects. Theoretical and Applied Genetics 121: 1001-1021

Placido, D.F., M.T. Campbell, J.J. Folsom, X. Cui, G.R. Kruger, P.S. Baenziger, et al. 2013. Introgression of novel traits from a wild wheat relative improves drought adaptation in wheat. Plant physiology 161: 1806-1819

Quarrie, S., M. Gulli, C. Calestani, A. Steed and N. Marmiroli. 1994. Location of a gene regulating drought-induced abscisic acid production on the long arm of chromosome *5A* of wheat. Theoretical and Applied Genetics 89: 794-800

Raman, H., R. Raman, R. Wood and P. Martin. 2006. Repetitive indel markers within the *ALMT1* gene conditioning aluminium tolerance in wheat (*Triticum aestivum L.*). Molecular Breeding 18: 171-183

Raman, H., K. Zhang, M. Cakir, R. Appels, D.F. Garvin, L.G. Maron, et al. 2005. Molecular characterization and mapping of *ALMT1*, the aluminium-tolerance gene of bread wheat (*Triticum aestivum L.*). Genome 48: 781-791

Rashid, A., J. Stark, A. Tanveer and T. Mustafa. 1999. Use of canopy temperature measurements as a screening tool for drought tolerance in spring wheat. Journal of agronomy and crop science 182: 231-238

Rebetzke, G.J., A.R. Rattey, G.D. Farquhar, R.A. Richards and A.T.G. Condon. 2013. Genomic regions for canopy temperature and their genetic association with stomatal conductance and grain yield in wheat. Functional Plant Biology 40: 14-33

Reynolds, M. and R. Tuberosa. 2008. Translational research impacting on crop productivity in drought-prone environments. Current opinion in plant biology 11: 171-179

Satarug, S., S.H. Garrett, M.A. Sens and D.A. Sens. 2011. Cadmium, environmental exposure, and health outcomes. Ciência & Saúde Coletiva 16: 2587-2602.

Satarug, S. and M.R. Moore. 2004. Adverse health effects of chronic exposure to low-level cadmium in foodstuffs and cigarette smoke. Environmental health perspectives 112: 1099.

Schnurbusch, T., J. Hayes, M. Pallotta, P. Langridge and T. Sutton. 2010. Identification of boron (B) tolerance genes in barley and wheat. Tagungsband der 60. Jahrestagung der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs 24.-26. November 2009, Raumberg-Gumpenstein: 39.

Scoles, G., T. Schnurbusch, P. Langridge and T. Sutton. 2008. The *Bo1*-specific PCR marker *AWW5L7* is predictive of boron tolerance status in a range of exotic durum and bread wheats. Genome 51: 963-971

Sharma, S., P.R. Bhat, B. Ehdaie, T.J. Close, A.J. Lukaszewski and J.G. Waines. 2009. Integrated genetic map and genetic analysis of a region associated with root traits on the short arm of rye chromosome 1 in bread wheat. Theoretical and Applied Genetics 119: 783-793

Shaw, L.M., A.S. Turner and D.A. Laurie. 2012. The impact of photoperiod insensitive *Ppd-1a* mutations on the photoperiod pathway across the three genomes of hexaploid wheat (*Triticum aestivum*). The Plant Journal 71: 71-84

Singh, R.P., J. Huerta-Espino, S. Rajaram and J. Crossa. 1998. Agronomic effects from chromosome translocations 7DL. 7Ag and 1BL. 1RS in spring wheat. Crop Science 38: 27-33

Spielmeyer, W., J. Hyles, P. Joaquim, F. Azanza, D. Bonnett, M. Ellis, et al. 2007. A *QTL* on chromosome 6A in bread wheat (*Triticum aestivum*) is associated with longer coleoptiles, greater seedling vigour and final plant height. Theoretical and Applied Genetics 115: 59-66

Sukumaran, S., S. Dreisigacker, M. Lopes, P. Chavez and M.P. Reynolds. 2015. Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. Theoretical and Applied Genetics 128: 353-363

Sukumaran, S., M. Ellis, M. Lopes, S. Pinto, D. Bonnett, S. Dreisigacker, D. Mullan, and M.P. Reynolds. 2013. Wheat mapping populations available at CIMMYT for yield potential research. International Workshop of the Wheat Yield Consortium, 3. Proceedings; CENEB, CIMMYT. Cd. Obregon, Sonora (Mexico); 5-7 Mar 2013, CIMMYT.

Talukder, S.K., P.V. Prasad, T. Todd, M.A. Babar, J. Poland, R. Bowden, et al. 2015. Effect of cytoplasmic diversity on post anthesis heat tolerance in wheat. Euphytica: 1-12

Tang, Y., D. Garvin, L. Kochian, M. Sorrells and B. Carver. 2002. Physiological genetics of aluminum tolerance in the wheat cultivar Atlas 66. Crop science 42: 1541-1546

Tian, F., J. Gong, J. Zhang, M. Zhang, G. Wang, A. Li, et al. 2013. Enhanced stability of thylakoid membrane proteins and antioxidant competence contribute to drought stress resistance in the tasg1 wheat stay-green mutant. Journal of experimental botany 64: 1509-1520

Turki, N., M. Harrabi and K. Okuno. 2012. Effect of Salinity on Grain Yield and Quality of Wheat and Genetic Relationships among Durum and Common Wheat. Journal of Arid Land Studies: 22-21

Vijayalakshmi, K., A.K. Fritz, G.M. Paulsen, G. Bai, S. Pandravada and B.S. Gill. 2010. Modeling and mapping *QTL* for senescence-related traits in winter wheat under high temperature. Molecular Breeding 26: 163-175

Wang, S., D. Wong, K. Forrest, A. Allen, S. Chao, B.E. Huang, et al. 2014. Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. Plant Biotechnol Journal 12: 787-796

Wei, B., R. Jing, C. Wang, J. Chen, X. Mao, X. Chang, et al. 2009. Dreb1 genes in wheat (*Triticum aestivum L.*): development of functional markers and gene mapping based on *SNPs*. Molecular Breeding 23: 13-22

Wiebe, K., N.S. Harris, J.D. Faris, J.M. Clarke, R.E. Knox, G.J. Taylor, et al. 2010. Targeted mapping of *Cdu1*, a major locus regulating grain cadmium concentration in durum wheat (*Triticum turgidum L. var durum*). Theoretical and Applied Genetics 121: 1047-1058. -1.

Xu, Y., S. Li, L. Li, X. Zhang, H. Xu and D. An. 2013. Mapping *QTLs* for salt tolerance with additive, epistatic and *QTL*× treatment interaction effects at seedling stage in wheat. Plant Breeding 132: 276-283

Yang, J., R.G. Sears, B.S. Gill and G.M. Paulsen. 2002. Quantitative and molecular characterization of heat tolerance in hexaploid wheat. Euphytica.126: 275-282

Yu, J.-B. and G.-H. Bai. 2010. Mapping quantitative trait loci for long coleoptile in Chinese wheat landrace Wangshuibai. Crop science 50: 43-50

Zhao, Y., W. Dong, N. Zhang, X. Ai, M. Wang, Z. Huang, et al. 2014. A wheat allene oxide cyclase gene enhances salinity tolerance via jasmonate signaling. Plant physiology 164: 1068-1076

Zhou, L.-L., G.-H. Bai, H.-X. Ma and B.F. Carver. 2007. Quantitative trait loci for aluminum resistance in wheat. Molecular Breeding 19: 153-161

Chapter 10:
Marker Use in Wide Crosses

Masahiro Kishii
CIMMYT Int., Apdo. Postal 6-641,06600 Mexico, DF, Mexico

Wheat wide crossing utilizes the diversity of wheat wild relatives. Wheat wide crossing can be seen as a special area of wheat breeding where the crossability of wheat with wild relatives, the meiotic chromosome pairing between wheat and the wild relative, and necrosis become important and often problematic. Molecular markers can aid in facilitating some of the wide crossing constraints. There are four important categories of markers used in wheat wide crossing: 1) markers detecting the segments or genes in the wild relatives; 2) crossability-related genes; 3) meiotic chromosomal pairing related genes; and 4) necrosis related genes. The markers used in the CIMMYT wide crosses group are listed in Table 1.

Markers for detecting the chromosome segments/genes from wild relatives

The list of resistance genes discovered in wild relatives and already transferred into diverse wheat backgrounds can be found in the "CATALOGUE OF GENE SYMBOLS FOR WHEAT" (McIbtosh et al. 2013; http://wheat.pw.usda.gov/GG2/Triticum/wgc/2013/) that has been updated elsewhere (http://www.shigen.nig.ac.jp/wheat/ komugi/genes/download.jsp#mg2012). Wild relative-derived genes have been also reviewed in intervals in international journals (e.g., Tyrka and Chekowski, 2004) and many genes identified for diseases resistance such as wheat rust and their linked markers are reported at the web site "MAS WHEAT" (http://maswheat.ucdavis.edu/Index.htm).

The CIMMYT wide crossing group is currently working on the transfer of Ug99 stem rust resistance genes, namely *Sr26, Sr32, Sr37, Sr39,* and *Sr40*, into CIMMYT elite wheat varieties. For the genes *Sr26, Sr32,* and *Sr39*, perfect markers that detect the wild relative segments, including the gene itself, are listed in "MAS WHEAT" (Figure 1). For *Sr37* and *Sr40*, only linked markers are available (Wu et al., 2009), keeping in mind that the genes may be lost during breeding.

Markers for crossability-related genes

The homeologous genes *Kr1* (5BL), *Kr2* (5AL), and *Kr3* (5DL, Fedak and Jui 1982) are well known to be related to crossability. The dominant genes have shown to inhibit pollen-tube growth in rye, barley or *Aegilops*, which reduce the ability to successfully produce interspecific hybrids. In wheat, *Kr1* has the highest inhibition effect and *Kr1* and *Kr2* are additive. The inhibition effect therefore varies with different genotypes. *Kr1Kr2* is considered to have less than 10% crossability, *Kr1kr2* between 10 and 25% crossability, *kr1Kr2* between 25 and 50%, and *kr1kr2* more than 50% crossability (Lein 1943). More recently, the gene

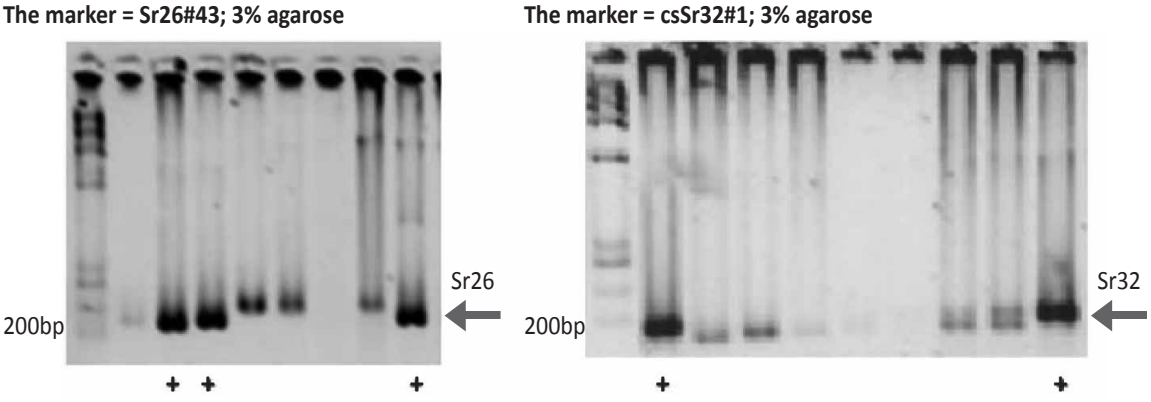


Figure 1. Screening of Sr26 and Sr32 positive plants.

Table 1. The list of molecular markers.

Target genes	Name of markers	Sequences	Comments	References
Sr26	Sr26#43-F Sr26#43-R	AATCGTCCACATTGGCTICT CGCAACAAAATCATGCACTA	Perfect marker	Mago et al. 2005
Sr32	caSr32#1-F caSr32#1-R	GGTTTGGTGGCAACTCAGGT CATAAGCCAAAGAGGCACCA	Perfect marker	Mago et al. 2013
Sr39	Sr39#50s-F Sr39#50s-R	CCAATGAGGAGATCAAAACAACC CTAGCAAGGACCAAGCAATCTTG	Perfect marker (Not verified at CIMMYT)	Mago et al. 2009
Sr40	wmc344-F wmc344-R	ATTTCAGTCTAATTAGCGTTGG AACAAAGAACATAATTAACCCC	0.7 cM to *Sr40* (Not verified at CIMMYT)	Wu et al. 2009
Sr40	wmc474-F wmc474-R	ATGCTATTAAACTAGCATGTGTCG AGTGGAAACATCATTCCTGGTA	~2.5 cM proximal to *Sr40* (Not verified at CIMMYT)	Wu et al. 2009
Sr40	wmc477-F wmc477-R	CGTCGAAAACCGTACACTCTCC GCGAAACAGAATAGCCCTGATG	~3.5 cM proximal to *Sr40* (Not verified at CIMMYT)	Wu et al. 2009
Ph1b, Ne1, Kr1	gwm213-F gwm213-R	TGCCTGGCTCGTTCTATCTC CTAGCTTAGCACTGTCGCCC	Perfect maker for *ph1b*	Bertin et al. 2009
Ph1b, Ne1, Kr1	gwm371-F gwm371-R	GACCAAGATATTCAAACTGGCC AGCTCAGCTTGCTTGGTACC	Perfect maker for *ph1b*	Bertin et al. 2009
Skr	cfb341-F cfb341-R	TAATTAGGGCCTGCTTCTGCT TTCCTTCATCCAAAGAGACTGG	Perfect marker	Alfares et al. 2009
Ne1	barc0074-F barc0074-R	GCGCTTGCCCCTTCAGGCGAG CGCGGGAGAACCACCAGTGACAGAGC	2.0 cM proximal to *Ne1*	Chu et al. 2009
Ne1	barc0216-F barc0216-R	TGACGACCCAATCCATAGACA GGTGATTATTCGTGAGTTCCCTGTG	8.3 cM distal to *Ne1*	Chu et al. 2009
Ne2	barc0055-F barc0055-R	GCGGTCAACACACTCCACTCCTCTCTC CGCTGCTCCCATTGCTCGCCGTTA	3.2 cM proximal to *Ne2*	Chu et al. 2009
Ne2	gwm148-F gwm148-R	GTGAGGCAGCAAGAGAGAAA CAAAGCTTGACTCAGACCAAA	6.7 cM distal to *Ne2*	Chu et al. 2009

Skr (5BS) has additionally been recognized to prevent the effect of *Kr1* in crosses between wheat and rye, therefore, promoting crossability (Lamoureux et al. 2002).

Kr1 (5BL): According to Bertin et al. (2009), *Kr1* locates between SSR markers *gwm213* and *gwm371*. Both, *gwm213* and *gwm371* have shown fairly good polymorphism among CIMMYT wheat lines (Figure 2), but it may be necessary to use neighboring markers in case marker are monomorphic in the varieties of interest. Since some of best CIMMYT durum and bread wheat varieties don't have good crossability with *Aegilops* species, we are transferring *kr1* (high crossable allele) from a high crossable variety, Chinese Spring, to the CIMMYT best varieties using *gwm213* and *gwm371*.

SKr (5BS): According to Alfares et al. 2009, flanking markers for *SKr* are *cfd341* (less than 0.1cM), *cfd306* (less than 0.1cM), and *gwm234* (0.1cM).

The upper band of marker *cfd341* is associated with a high crossable genotype (Figure 3). Our primary results indicate that most bread and durum wheat varieties used as recurrent parents in the wide crossing group have the upper (high crossable) band, which, however, does not coincide with our phenotypic observation of high variation in their crossability with *Aegilops* species. It is therefore possible that the *SKr* gene does not have the same effect in *Aegilops* than in rye or additional genes are involved in the determination of crossability in CIMMYT varieties. Further research on the crossability effect of *SKr* for crosses with *Aegilops* species is therefore necessary.

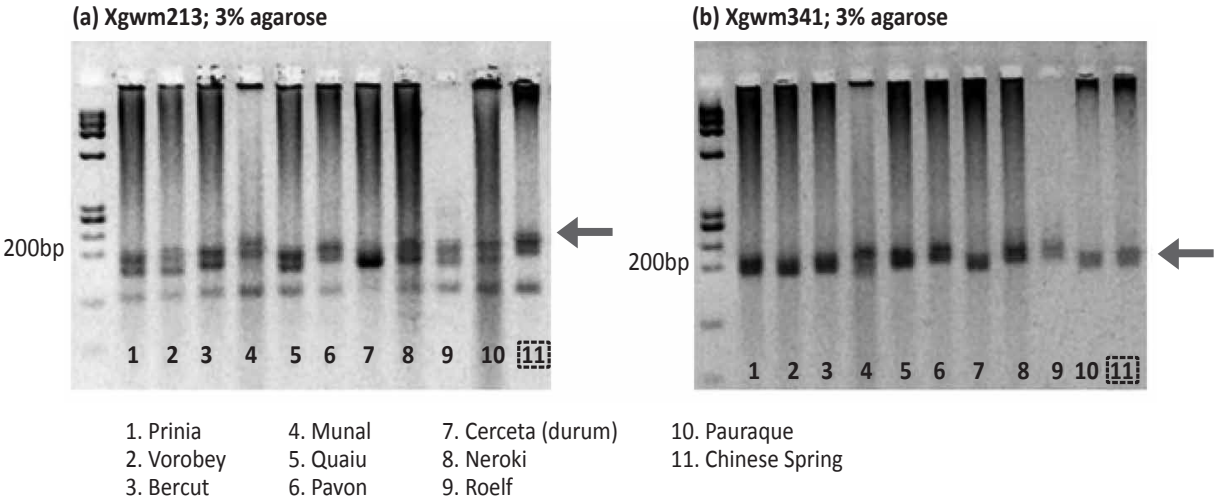


Figure 2. Polymorphism of *Kr1* linked markers among CIMMYT elite varieties and Chinese Spring.

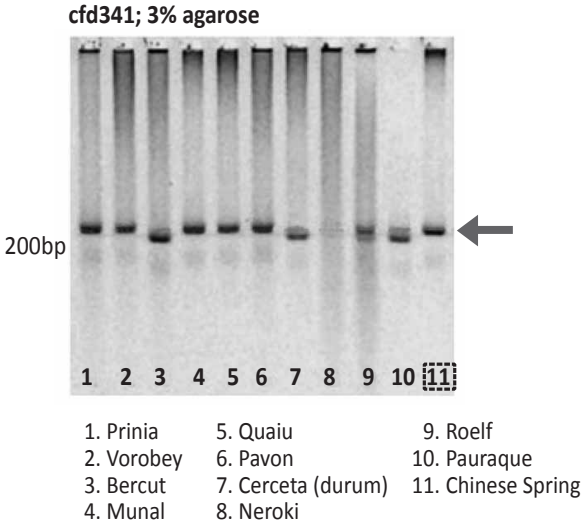


Figure 3. Polymorphism of *Skr* gene markers among CIMMYT elite varieties and Chinese Spring.

Markers for meiotic chromosomal pairing related genes

Ph1, Ph2 and two *PhI* genes from *Aegilops speltoides* fall in this marker category. While *Ph1* and *Ph2* inhibit the chromosomal pairing, *PhI* genes have been considered to inhibit the effect of *Ph1* (Dover and Riley 1977). The *Ph1* gene is the most effective gene for inhibiting chromosomal pairing, so that mutants of the gene have been routinely used for inducing meiotic pairing between wheat and alien species around the world.

ph1b: *gwm213* and *gwm371* are two SSRs markers in the *Ph1* gene region.

The *Ph1* mutant includes a Mbp deletion in the corresponding gene region (Sears 1977; Gill et al. 1993; Fig. 4a), so that any marker within this deletion can be used for ph1b detection. From our experience, the marker *gwm213* is the best marker to screen for the mutant as the marker distinguishes ph1b homozygous plants from PCR false negatives (Figure 4b). Since *ph1b* is a deletion, the marker is dominant and can only determine the homozygous *ph1bph1b* genotype but not heterogeneous *Ph1ph1b* genotype (Figure 4c). It is recommended to use markers outside or flanking the deletion when heterozygote genotype wants to be detected.

Markers for necrosis-related genes

Many types of necrosis have been recognized in wheat, but the *Ne1-Ne2* necrosis is the best known. Necrosis in expressed when both *Ne1* and *Ne2* dominant alleles genes build the genotype (Figure 5a). This type of necrosis is problematic in wheat wide crossing, particularly when synthetic hexaploid wheat lines are developed (a cross between durum wheat and *Aegilops tauschii*, Figure 5b), as most durum wheat lines (and especially CIMMYT lines) have the dominant allele of *Ne1*. This means that almost all synthetic hexaploid wheat lines have *Ne1*. It has also been reported that about 89% of wheat lines in Central America possess the dominant allele of *Ne2* (Pukhalskiy et al. 2000), which is why CIMMYT

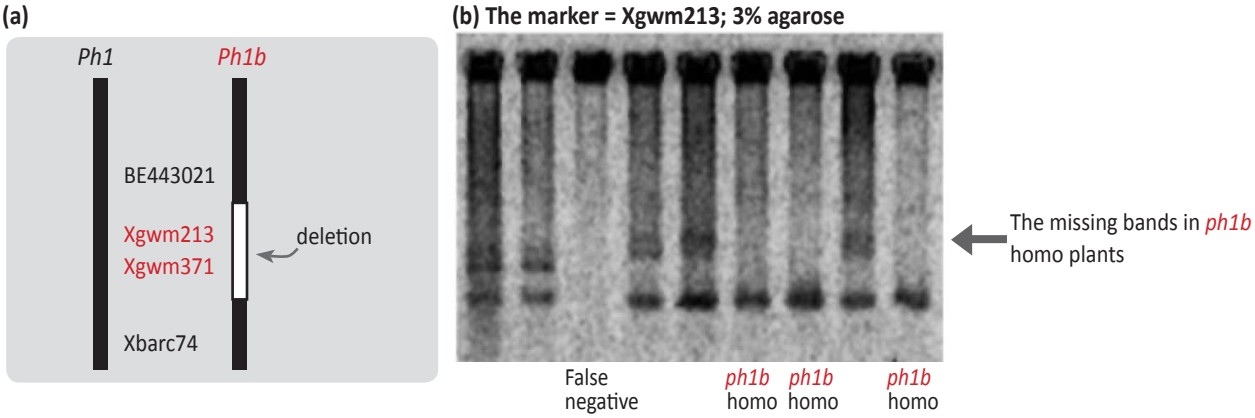


Figure 4. Deletion mutant of ph1b (a), screening of ph1b homo plants with Xgwm213 (b), and presumable PCR band patterns using markers inside and outside the deletion (c).

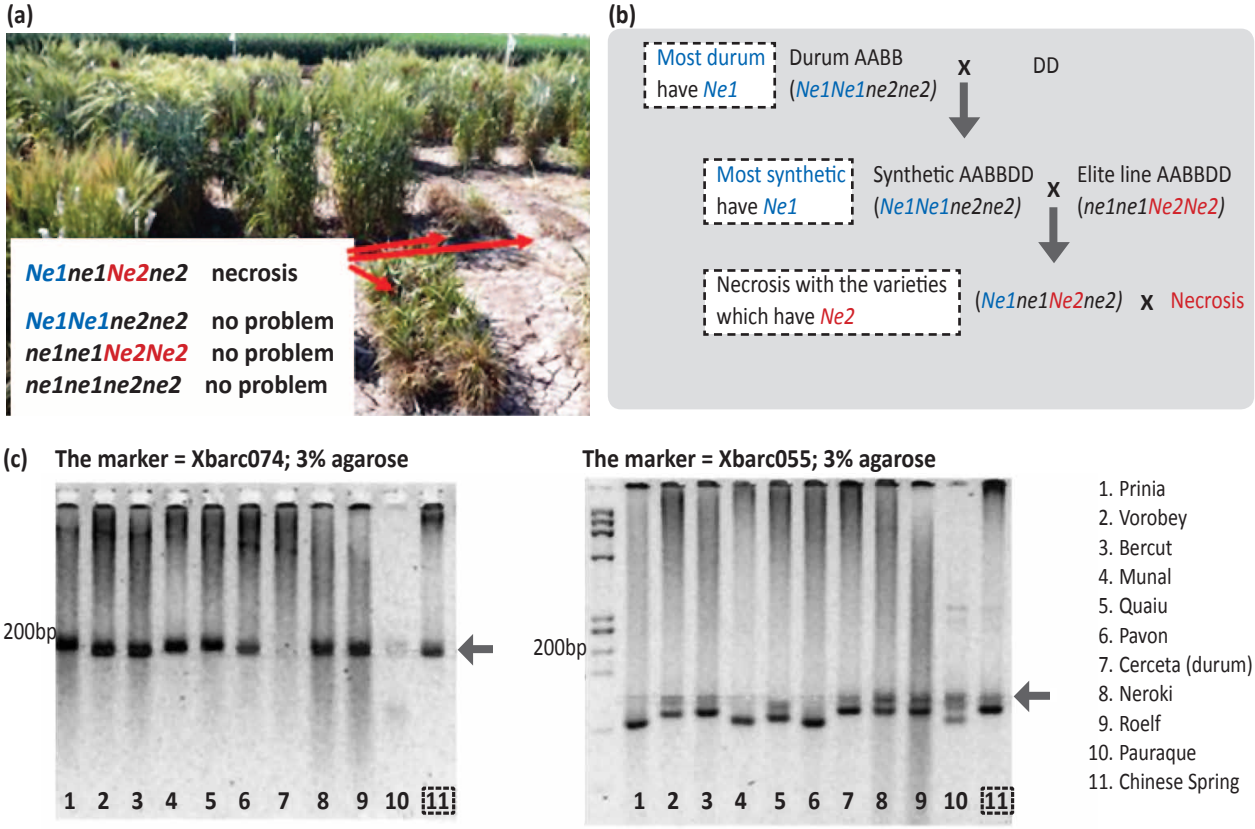


Figure 5. *Ne1* and *Ne2* necrosis (a), the inherent *Ne1* and *Ne2* necrosis problem in synthetic wheat lines (b), and polymorphism of a *Ne1* linked marker (Xbarc074) and *Ne2* linked marker (Xbarc055) (c).

wheat breeders have not been able to bring all possible synthetic hexaploid wheat lines into their breeding program. It is therefore a target at CIMMYT to eliminate the dominant *Ne1* gene from durum varieties for synthetic hexaploid wheat line production.

Ne1: Chu et al. 2009 localized the gene between markers *barc216* (8.3 cM distal) and *barc074* (2.0 cM proximal).

The wide crossing group has been working on transferring the recessive *ne1* allele (non-necrotic alley) into CIMMYT elite durum varieties. Since *barc074* and *barc216* have not shown much polymorphism among CIMMYT elite varieties, other markers in the vicinity such as *gwm213* (Fig. 2) can be also employed for selections.

Ne2: Chu et al. 2009 was localized this gene between markers *gwm148* (6.7 cM distal) and *barc055* (3.2 cM proximal).

Additionally to *ne1*, the wide crosses group has made the effort to transfer the recessive allele *ne2* (non-necrotic alley) to a limited number of CIMMYT elite bread wheat varieties as new synthetic hexaploid wheat lines might be developed in the future. Testing the two linked markers to *ne2*, we have seen that the marker barc055 was more polymorph among CIMMYT varieties, while *gwm148* was almost monomorph (Figure 5c).

References

Alfares, W., A. Bouguennec, F. Balfourier, G. Gay, H. Bergès, S. Vautrin, P. Sourdille, M. Bernard and C. Feuillet. 2009. Fine mapping and marker development for the crossability gene *SKr* on chromosome 5BS of hexaploid wheat (*Triticum aestivum L.*). Genetics 183: 469-481

Bertin, I., L. Fish, T.N. Foote, E. Knight, J. Snape and G. Moore. 2009. Development of consistently crossable wheat genotypes for alien wheat gene transfer through fine-mapping of the *Kr1* locus. Theor. Appl. Genet. 119: 1371-81

Chu, C.G., J.D. Faris, T.L. Friesen and S.S. Xu. 2006. Molecular mapping of hybrid necrosis genes *Ne1* and *Ne2* in hexaploid wheat using microsatellite markers. Theor Appl Genet. 112: 1374-1381

Dover, G. and R. Riley. 1977. Inferences from genetical evidence on the course of meiotic chromosome pairing in plants. Philos. Trans. R. Soc. Lond. Ser. B 277: 313–326

Fedak, G., P. and Y. Jui. 1982. Chromosomes of Chinese Spring wheat carrying genes for crossability with Betzes barley. Can. J. Genet. Cytol. 24: 227–233

Gill, K.S., B.S. Gill, T.R. Endo and Y. Mukai. 1993. Fine physical mapping of *Ph1*, a chromosome pairing regulator gene in polyploid wheat. Genetics 134: 1231–1236

Lamoureux, D., C. Boeuf, F. Regad, O. Garsmeur and G. Charmet. 2002. Comparative mapping of the wheat *5B* short chromosome arm distal region with rice, relative to a crossability locus. Theor. Appl. Genet. 105: 759–765

Lein, A. 1943. Die genetische Grundlage der Kreuzbarkeit zwischen Weizen und Roggen. Z Indukt Abstamm und Vererb Lehre 81: 28–61

Mago, R., H.S. Bariana, I.S. Dundas, W. Spielmeyer, G.J. Lawrence, A.J. Pryor and J.G. Ellis. 2005. Development of *PCR* markers for the selection of wheat stem rust resistance genes *Sr24* and *Sr26* in diverse wheat germplasm. Theor. Appl. Genet. 111: 496–504

Mago, R., D. Verlin, P. Zhang, U. Bansal, H. Bariana, Y. Jin, J. Ellis, S. Hoxha and I. Dundas I. 2013. Development of wheat–Aegilops speltoides recombinants and simple *PCR*-based markers for *Sr32* and a new stem rust resistance gene on the *2S#1* chromosome. Theor. Appl. Genet. 126: 2943-2955

Mago, R., P. Zhang, H.S. Bariana, D.C. Verlin, U.K. Bansal, J.G. Ellis and I.S. Dundas. 2009. Development of wheat lines carrying stem rust resistance gene *Sr39* with reduced Aegilops speltoides chromatin and simple *PCR* markers for marker-assisted selection. Theor. Appl. Genet. 124: 65-70

McIntosh, R.A., J. Dubcovsky, W.J. Rogers, C. Morris, R. Appels and X.C. Xia. 2013. Catalogue of gene symbols for wheat. Proc The 12th International Wheat Genetics Symposium. Yokohama, Japan. 8-14 September (http://wheat.pw.usda.gov/GG2/Triticum/wgc/2013/).

Pukhalskiy, V.A., S.P. Martynov and T.V. Dobrotvorskaya. 2000. Analysis of geographical and breeding-related distribution of hybrid necrosis genes in bread wheat (*Triticum aestivum L.*). Euphytica 114: 233–240

Sears, E. R. 1977. An induced mutation with homoeologous pairing in wheat. Can. J. Genet. Cytol. 19: 585–593

Tyrka, M. and J. Chekowski. 2004. Enhancing the resistance of triticale by using genes from wheat and rye. J. Appl. Genet. 45: 283–295

Wu, S., M. Pumphrey and G. Bai. 2009. Molecular Mapping of Stem-Rust-Resistance Gene *Sr40* in Wheat. Crop Science 49: 1681-1686

Chapter 11:
Genomic Predictions in CIMMYT Wheat

Susanne Dreisigacker[1], David Bonnett[2], Paulino Perez-Rodriguez[4], Gustavo de los Campos[3], Marco Lopez-Cruz[3], Jose Crossa[1]

[1] CIMMYT Int., Apdo. Postal 6-641,06600 Mexico, DF, Mexico
[2] Bayer Crop Science, Fargo, USA
[3] Colegio de Postgraduados, Montecillo, Mexico
[4] Michigan State University, East Lansing, Michigan

Introduction

Molecular markers in a breeding program are largely considered as a supplemental tool for variety development. They are primarily used as an indirect selection tool, specifically when traits have low heritability, are difficult and cost-prohibitive to measure or require the desired pyramiding of genes. However, especially for more complex traits, the lack of large-scale validation and refinement of larger-effect QTL, the fact that selection can be based only on a subset of markers that most likely capture only a smaller portion of the total genetic variation, the high cost of genotyping individual loci and of overcoming many practical, logistic, and genetic constraints limit the implementation of marker-assisted selection (MAS) methods in plant breeding programs for this traits (Xu and Crouch 2008).

With the development of modern genotyping and sequencing methods, MAS theory has recently shifted from the transfer of larger-effect genes to the use of genome-wide markers to predict the performance of both phenotyped and unphenotyped individuals (genomic selection, or GS). Using genome-wide markers, every trait locus is likely to be in linkage disequilibrium (LD) with a minimum of one marker locus in the entire target population. In GS, a training population related to the breeding germplasm is genotyped with genome-wide markers and phenotyped in target environments. These data are used to derive a prediction equation that can then be applied to genotypes of unphenotyped individuals to calculate their genomically estimated breeding values (GEBVs) that can be used to inform selection decisions (Meuwissen et al. 2001; Bernardo and Yu 2007; de los Campos et al. 2009). Genetically estimated breeding values open up several new routes for increasing genetic improvement rates in plant breeding programs. They offer opportunities to: (1) increase the selection efficiency of preliminary and multi-location yi eld trials; and (2) shorten the breeding cycle by repeated early generation selection, thus increasing the genetic gain per unit of time.

Several GS studies, many of which were developed or used data generated by CIMMYT's wheat breeding program (see Crossa et al. 2014, for a recent review of the topic), have shown that GS can achieve reasonably high prediction accuracy. This has raised expectations about the prospects of implementing GS in wheat breeding programs. However, implementing GS in breeding problems also presents important challenges. In this chapter, we describe the results and the important lessons learned when testing genomic prediction in CIMMYT's Global Wheat Program, from the initial assessment of the prediction ability of different models using pedigree and marker information to the present day, when methods for implementing GS in practical wheat breeding programs are being studied and investigated.

Prediction results using CIMMYT wheat data sets

Varying levels of genomic prediction accuracy have been obtained in plants. For inbreeding species, their relatively small effective population size (N_e) is the main advantage for genomic selection. The smaller the N_e, the smaller the number of independent chromosome segments in the genome that allow for higher GEBV accuracies (Lin et al. 2014). Most studies report prediction accuracy as the correlation between GEBVs and phenotypes. In CIMMYT wheat data sets, prediction accuracies for grain yield, for example, ranged from 0.3 (Poland et al. 2012) to 0.7 (Perez-Rodriguez et al. 2012); for days to heading, prediction accuracies between 0.4 (Poland et al. 2012) and 0.7 (Wimmer et al. 2013) were reported. Prediction accuracy depends on the prediction problem assessed and on several other factors, such as trait heritability and the genetic architecture of

the trait, the relationship between the individuals to be predicted and those used to train the models for prediction, sample size, number of markers and marker platform, choice of statistical model, and genotype × environment interaction (GE).

Trait heritability

Heritability allows comparing the relative importance of genes and environments to the variation of traits within and across populations. Heritability depends on the genetic properties of a trait, the range of typical environments in the studied population, as well as various interactions between genes and environmental factors. For traits with low heritability, genes contribute little to individual phenotypic differences; for highly heritable traits, genes are the main reason for individual differences. Genomic selection performs differently in traits with distinct genetic properties. In Table 1, we compared seven well-established genome-wide selection methods for 16 traits with heritabilities ranging from 0.18 to 0.93. For all methods, the ability to predict phenotypes was linearly correlated with trait heritability. For example, prediction accuracy was high for the quality traits of grain hardiness and flour sedimentation, which have high heritabilities, whereas prediction accuracy was moderate for the agronomic traits of plant height and grain yield, which have low to moderate heritabilities. Correlations ranged between 0.5 and 0.6 (Table 1). This correlation was expected, given that, as described above, traits with lower heritability have phenotypes less reflective of their genetic content, and are consequently less predictable through genomic selection. Genomic selection is mainly seen as a way to achieve more genetic gains in traits with lower heritability; however, phenotyping for some higher heritable traits can be very expensive (e.g., some quality traits) and genomic predictions of unphenotyped individuals can therefore be of great advantage. The heritability of the trait is also crucial because, with greater heritabilities, fewer genotyping and phenotyping records are required. Reduced heritability will lead to a decrease in accuracy of predicting the breeding value but this can be compensated for by using a larger number of observations to estimate the marker effects. Phenotypes with higher heritability can be created by averaging the phenotypic performance of varieties across replicate plots, thereby reducing the environmental variance of average performance.

Table 1. Cross-validated accuracy (r_{cv}) and trait heritability (h^2) for 15 traits, using seven prediction models with a design matrix based on genotyping by sequencing markers in the CIMMYT C29 Semi-Arid Wheat Screening Nurseries (SAWSN).

Trait	h^2	*BL	PBL	BRR	PBRR	RKHS	PRKHS	BayesB
Stem rust	0.75	0.68	0.73	0.67	0.73	0.68	0.72	0.68
Leaf rust	0.60	0.54	0.58	0.54	0.57	0.55	0.57	0.55
Yellow rust	0.18	0.64	0.70	0.64	0.70	0.65	0.69	0.63
Spot blotch	0.78	0.63	0.68	0.63	0.68	0.62	0.67	0.63
Fusarium head blight	-	0.50	0.55	0.51	0.54	0.52	0.55	0.47
Grain hardiness	0.88	0.72	0.79	0.72	0.79	0.70	0.79	0.68
Grain protein	0.57	0.55	0.58	0.55	0.58	0.54	0.56	0.53
Flour protein	0.69	0.58	0.63	0.58	0.63	0.59	0.62	0.57
Flour sedimentation	0.83	0.76	0.77	0.76	0.77	0.76	0.78	0.75
Bread loaf volume	0.68	0.62	0.74	0.62	0.74	0.62	0.73	0.56
Test weight	0.69	0.49	0.54	0.49	0.54	0.48	0.51	0.52
Alveograph ratio	0.47	0.41	0.52	0.41	0.52	0.41	0.52	0.42
Plant height	0.31	0.40	0.46	0.39	0.44	0.42	0.48	0.41
Days to heading	0.84	0.64	0.67	0.64	0.67	0.65	0.67	0.63
Thousand kernel weight	0.93	0.71	0.73	0.71	0.73	0.70	0.73	0.72
Grain yield	0.48	0.57	0.60	0.57	0.57	0.61	0.65	0.63
Cor (r_{cv}, h^2)		0.62	0.58	0.62	0.59	0.56	0.54	0.60

* BL, Bayesian LASSO; PBL, Bayesian LASSO with pedigree; BRR, Bayesian ridge regression; PBRR, Bayesian ridge regression with pedigree; RKHS, reproducing kernel Hilbert space; PRKHS, reproducing kernel Hilbert space with pedigree.

Training population size and the relationship between the individuals in training and selection populations

One approach to implementing GS is to select individuals with the highest genetic merit in the early generations of a breeding cycle (e.g., selecting F_2 individuals). The selected individuals can be intercrossed and the resulting progenies can be selected again a number of times before extracting inbred lines, also called rapid cycling GS. In a simulation study, we explored the relative importance of the relatedness between training and selection populations, sample size and marker density for the accuracy of genomic prediction in an early generation selection approach (Hickey et al. 2014). For simulation, several biparental populations, each having 550 F_2-individuals, were created that were related to each other in different ways: biparental populations that have one parent in common (BP-P); that have one grandparent in common (BP-G); or that are unrelated (BP-U). The accuracy of selection was evaluated on 50 unphenotyped F2-individuals from a single biparental population using the correlation between the GEBVs and the true breeding values. For the phenotypes, a polygenic trait was simulated with 0.5 heritability, and a range of 50 to 10,000 single nucleotide polymorphism (SNP) markers was tested.

The phenotypes and genotypes that were used to train the prediction equation were either generated inside the single biparental population or inside the other biparental populations (BP-P, BP-G, or BP-U) that were simulated to have relationships with the given single biparental population. Between 1 to 40 populations and 50 to 500 F_2-individuals per population were used to train the prediction equations. Figure 1 shows the accuracy of prediction inside a single biparental population. The accuracy of the breeding values increased as the size of the training population increased. Training with up to 50 phenotypes gave accuracies between 0.2 and 0.6, while training with 100 or more phenotypes gave accuracies of 0.8 or higher.

The results when differently related populations were used for genomic prediction are displayed in Figure 2. Prediction accuracies decrease with decreasing relatedness to the given single biparental population (BP-P -> BP-U). Using information from unrelated populations generally gave low accuracies unless very large numbers of phenotypes (more populations and more individuals per population) were used. This means there is a tradeoff between relationship and population size that affects prediction accuracy. When using information from close relatives, the marker associations are due to the linkage between markers and QTLs, whereas when using information from distant relatives, marker associations are due to linkage disequilibrium. Closer relatives share longer chromosome segments or haplotypes; therefore, a training population with close relatives will have a smaller number of independent haplotypes and a larger sample size per haplotype, leading to more precise predications. Distant relatives share shorter haplotypes and a training set will have a large number of independent haplotypes with different allele, different allele frequencies, linkage phases and background effects (including epistasis, the interaction between two or more genes controlling a single genotype) leading to less accurate predictions.

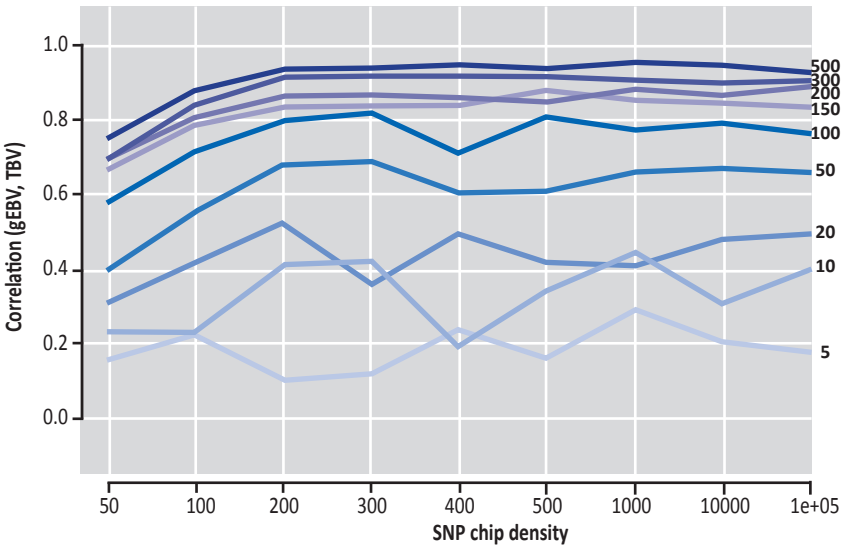


Figure 1. Accuracy of breeding values inside a given biparental population when training in the same population with different numbers of markers and F_2s (from Hickey et al. 2014).

We observed the same trend in prediction accuracies in the study by Ornella et al. (2012). Actual stem rust data from five recombinant inbred line (RIL) populations (PBW343/Juchi, PBW343/Pavon76, PBW343/Muu, PBW343/Kingbird and PBW343/K-Nyangumi) were used for genomic prediction. All populations were derived from crosses between resistant parents (Juchi, Pavon76, Muu, Kingbird and K-Nyangumi) and PBW343, a moderately susceptible parent. The populations were evaluated for reaction to stem rust at different locations. Stem rust resistance is known to be affected by major genes, along with several slow-rusting genes with small additive effects (Singh et al. 2011). The sample sizes of the five populations were between 92 and 176 molecularly characterized using Diversity Arrays Technology markers (http://www.diversityarrays.com/). As depicted in Figure 3, there are five clearly distinct but related half-sib populations, except lines in the PBW343/Juchi population, which do not seem to be closely related. Results in Table 2 indicate that prediction of stem rust data in each individual population using stem rust data from the other four populations gave relatively high correlations, except for population PBW343/Juchi, which does not have lines that are closely related among themselves or to lines in other populations. It is expected that population structure could become the main driving force for increasing the prediction accuracy in meta-populations, while prediction between populations significantly decreases. Several recent studies have furthermore illustrated the importance of a close relationship between the training and selection populations (Windhausen et al. 2012; Guo et al.

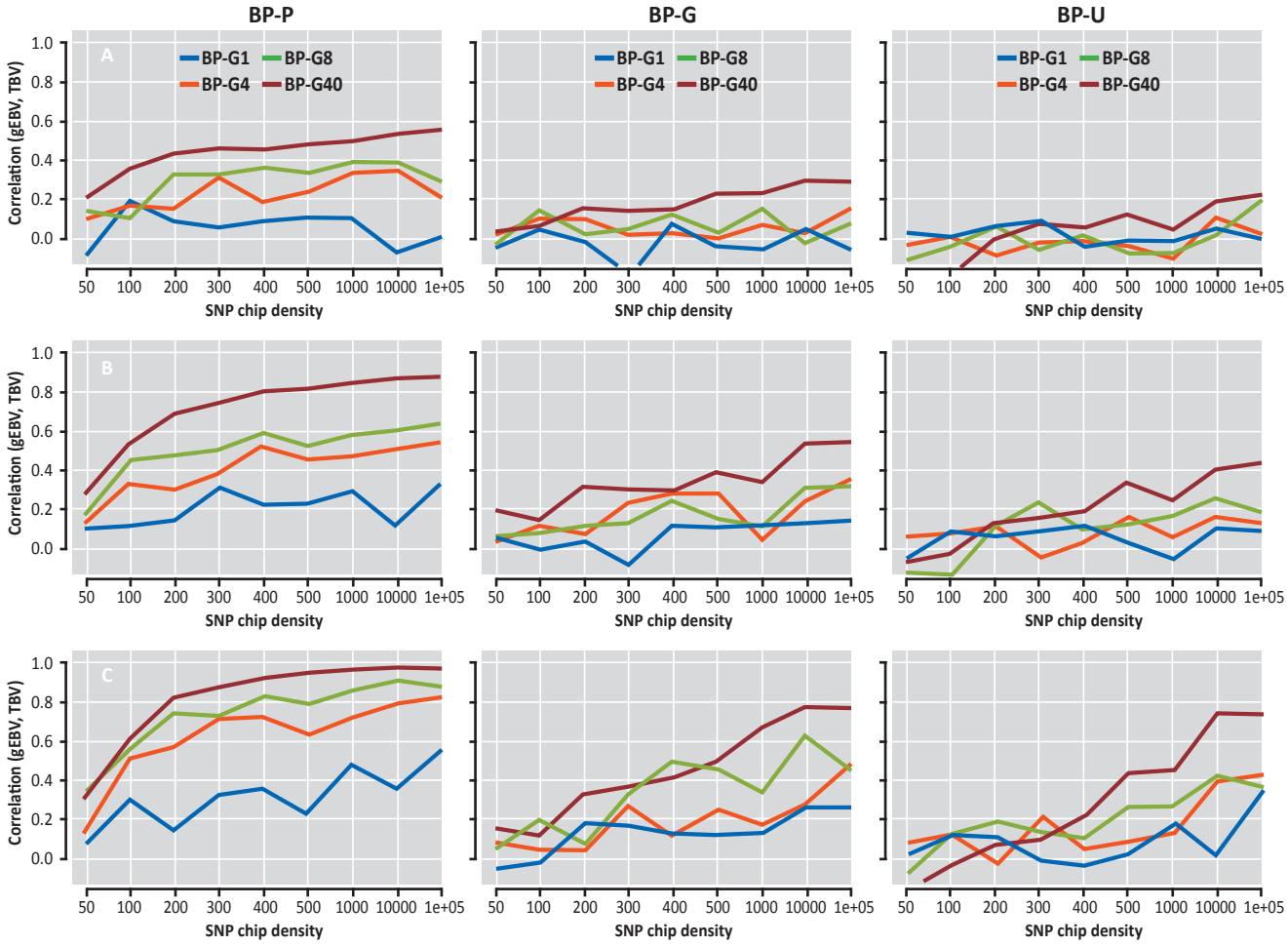


Figure 2. Accuracy of breeding values inside a given biparental population when training in 1, 4, 8 or 40 biparental populations with one parent in common (BP-P1 to BP-P40), one grandparent in common (BP-G1 to BP-G40) or with no pedigree relationship (BP-U1 to BP-U40). A = 5 F_2s recorded in each population, B = 50 F_2s recorded in each population, C = 500 F_2s recorded in each population, TBV: true breeding value, gEBV: genetically estimated breeding value (from Hickey et al. 2014).

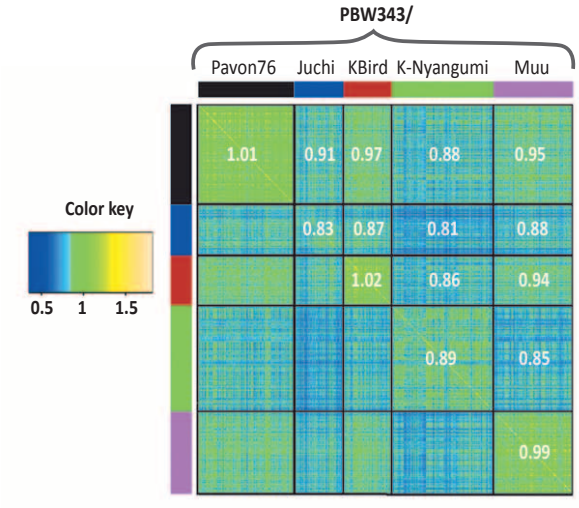


Figure 3. Heat map of the genomic relationship matrix G of five wheat populations. The numbers indicate average values of the corresponding elements of G within and between populations (from Ornella et al., 2012).

Table 2. Pair-wise correlations between observed and predicted stem rust values of GBLUP, trained within each population and in one population and evaluated in the other population for five populations (adapted from Ornella et al., 2012).

	Training set				
Testing set	PBW343/ Juchi	PBW343/ Kingbird	PBW343/ Knyangumi	PBW343/ Muu	Within population
PBW343/Juchi	-				0.41
PBW343/Kingbird	0.53	-			0.79
PBW343/KNyangumi	0.14	0.30	-		0.52
PBW343/Muu	0.18	0.30	0.33	-	0.59
PBW343/Pavon76	0.37	0.51	0.22	0.33	0.59

2014). It's therefore a priority for plant breeders to develop ad hoc training populations for each breeding population, and to update the training population to avoid the divergence of training and breeding population. Also it has to be kept in mind that if a training population does not segregate for a trait it is not possible to improve it with selection.

Marker density

Genotypic data can still be expensive, especially when low-cost genotyping approaches such as genotyping by sequencing or genotyping strategies involving low-density SNP chips are not used. In the simulation study conducted by Hickey et al. (2014), marker density was considered as one important factor in genomic prediction. The simulation results in Figure 2 show that the marker density required to obtain accurate genomic predictions depends on the degree of relatedness between the training and selection populations. With close relatives (e.g., BP-P), accurate predictions could be obtained with 200 markers. Increasing the marker density up to 10,000 markers did not improve prediction accuracy. A small number of markers is sufficient because the shared haplotypes and linkage blocks are large. When using distant relatives (e.g., BP-U in Fig. 2), more markers are required because of the lower linkage disequilibrium between markers and QTLs. Low marker density in distant relatives can cause artificial overestimates of linkage disequilibrium, when coupled with near homozygosity of late filial generations and decrease prediction accuracy. Similar results were also found by Solberg et al. (2008) and Meuwissen et al. (2009).

Relationship between environments

Multi-environment trials are widely used by plant breeders to evaluate the relative performance of genotypes across environments. Multi-environment trials are conducted because GE introduces uncertainty into the measure of genotypic performance and complicates the selection of superior genotypes. Thus accounting for GE has always been a concern when analyzing agronomic data and many different methods have been proposed for analyzing multi-environmental trials conducted by breeders.

Genome-wide markers provide a new tool that can be used in multi-environment trial analyses. Genome-wide markers do not change the fact of GE; however, their use could enable better selection decisions. Burgueño et al. (2012) and Jarquin et al. (2013) used genome-wide markers and additional environmental factors for multi-environment trial analyses and showed that when gene and environmental interaction terms were introduced in the prediction equation, prediction accuracy increased, suggesting that the proportion of variance accounted for by the prediction model was higher.

In Burgueño et al. (2012), the prediction accuracies showed the same pattern as the genetic correlations between environments. The interaction term allows

borrowing information between environments, and for environments that are positively correlated; this increases prediction accuracy.

The study conducted by Ornella et al. (2012) reached the same conclusion that higher prediction accuracy in correlated environments could be achieved. Actual stem rust data of the five RIL populations in this study were collected during the main and off-season in Njoro, Kenya, while the yellow rust data on the same populations were collected in Njoro, Kenya, and Toluca, Mexico. The positive correlation between environments (for stem rust, the same environment but different seasons) clearly favored prediction accuracy when data from one environment were used as the training population for the other environment (Figure 4).

Modeling selected markers as fixed effects

A recent simulation study (Bernardo 2013) found that modeling a large-effect locus as fixed to be advantageous when trait heritability was greater than 0.5 and the proportion of genetic variance explained by the locus was greater than 0.25. Rutkoski et al. (2014) confirmed these results using GS as a potential tool to select for adult-plant stem rust resistance. In a set of CIMMYT advanced lines that were tested for adult-plant stem rust resistance across environments and years, markers linked to the stem rust gene *Sr2* were applied and its results included as fixed effects in the prediction model that was more accurate than using genome-wide markers only. Overall, the levels of prediction accuracy found in this study indicate that GS can be effectively applied to improve stem rust APR in this germplasm.

Implementation of GS

Genomic selection has a great number of uses in a breeding program. Similar to MAS strategies, the introduction of GS is flexible and may vary for each breeding program, depending on the target traits and breeding scheme. The greatest potential use of GS is at points in the breeding program where selection using traditional methods is too expensive, time-consuming or not biologically or logistically possible. Two main applications of GS are being studied in CIMMYT's global wheat breeding program: 1) to predict the genotypic value of individuals for potential release as cultivars; and 2) to predict the breeding value of candidates in rapid-cycle populations.

Predicting the genotypic value of individuals for potential release

The breeding methodology applied by the CIMMYT wheat breeding program includes modified bulk selection. After population advancement with selection for more heritable traits via shuttle breeding, inbred lines are extracted and tested in preliminary yield trials (PYTs) to identify superior entries which will then be evaluated in the following year in more extensive multi-environment yield trials and/or used as parents to begin another breeding cycle. CIMMYT PYTs usually include up to 10,000 genotypes, of which approximately 1,000 are selected and evaluated in five to six different environments with two to three replications in the subsequent cycle.

CIMMYT PYTs are carried out in replicated yield trials in small plots and a single environment. GS could be useful to predict the GEBVs based on a large training population that includes previous breeding germplasm and amend the selection of lines. There is also a trade-off between the number of genotypes that enter the multi-environment trials and the number of plots per entry. A larger number of plots per entry allow a more accurate estimate of the performance of each genotype across environments, whereas a larger number of entries enhances the germplasm pool from which selections are made. If the number of plots is fixed, a larger number of entries can only be tested if they are divided across environments. Consequently, not all entries would be present in all environments but the average genotypic performance across all environments could be determined using genomic prediction. Initial results testing this approach using diverse models that incorporate pedigree, marker, environment and interaction terms into the prediction equation revealed relatively high prediction accuracies: an average 0.6 when 20% of the entries were present in only one of five environments (unpublished data). These results indicate that not all entries have to be evaluated in all environments and that more entries could be tested. This approach can be optimized further by maximizing the relationship between the training and testing populations and by varying the number of genotypes and environments to be predicted. Several sister lines are usually present in each CIMMYT multi-environment trial. Dividing the sister lines across environments would additionally increase the relatedness between the training and testing populations and allow higher prediction accuracies.

The best performing lines in multi-environment trials are selected to form CIMMYT's international screening nurseries and yield trials, which are distributed globally via the International Wheat Improvement Network. Although trait heritabilities are high at this stage of the breeding program, genomic prediction could be useful for boosting the selection of lines to be included in each yield trial specific to a target mega environment and that could potentially be released by national programs.

Predicting the breeding value of candidates in rapid-cycle populations

In a rapid cycle GS breeding scheme, segregating populations can be genotyped at the seedling stage, then selected based on GEBVs derived from a related training population. The resulting F_2-candidates can be used to extract inbred lines or intercrossed. Applying GS rapid cycling in early generations (e.g., F_2) is a high risk but high turnover approach. In conventional breeding, early generation intercrossing is not practiced due to highly heterozygous and heterogeneous progenies and the unfeasibility of selecting for complex traits based on a single plant. Although genomic prediction accuracy may not be high, shortening the cycle time or generation interval is expected to increase genetic gains.

In a proof-of-concept experiment, CIMMYT has initiated a rapid cycle GS scheme in wheat with grain yield as the target trait. Genomic prediction was applied in 40 F_2-populations, using historical data as a training population with two cycles of subsequent intercrossing within and between populations. In each generation, inbred lines were extracted based on genomic and conventional selection. Initial results in two seasons replicated trials of lines derived from single plants with a range of GEBVs have shown that higher GEBV individuals produce higher yielding derivative lines (Bonnett et al. unpublished). A yield increase of 7% in derivatives of plants with the highest 13% of GEBVs compared to the lowest 13% of GEBVs was found following just one cycle of GS. This represents a 30% realized gain from selection and is an encouraging initial result. A larger number of derivatives from simple cross F_2 derived bulks and first and second cycle intercrosses compared to conventionally selected cohorts from the same initial crosses are currently being tested to extend the evaluation of rapid cycle GS. These evaluations give the first indications of genetics gains from GS for a highly complex trait in an actual wheat breeding program.

In a second experiment, individuals from $F_{2:4}$ derived populations will be used as a training population to derive GEBVs within and between populations with close and distinct relationships.

Prediction algorithms

A large number of prediction models have been developed or adopted from other fields to handle the high-dimensional marker datasets that are typical of GS. The various types of models respond differently because they vary in their assumptions when treating the variance of complex traits. In GS, the number of predictors (p) is usually far greater than the number of individuals (n). In such cases, estimates of ordinary least-squares have poor prediction ability because marker effects are treated as fixed effects, which leads to multicolinearity and overfitting among predictors, thereby making the model unfeasible (for a review, see Lorenz et al. 2011, or de los Campos et al., 2013). To further improve genome-wide predictions in wheat, the CIMMYT Biometrics Unit has developed and recommended various prediction algorithms, primarily for low heritable traits. The R (R Core Team, 2013) package "Bayesian Generalized Linear Regression (BGLR)" (de los Campos et al. 2013) is applied for most CIMMYT implementations.

Models using combined pedigree and marker information

CIMMYT and many other breeding institutions have been using pedigree notations for many years, based on what is commonly known as the Purdy method (Purdy et al., 1968). These notations give an explicit description of the crossing of a line. Previously, selection in plant breeding was usually based on estimates of breeding values obtained using pedigree-based mixed models (Crossa et al. 2006). Information from relatives is integrated through the coefficient of parentage (COP) in the form of the additive relationship matrix A, which is twice the COP. However, pedigree-based models cannot account for Mendelian segregation. Molecular markers allow tracing Mendelian segregation across positions in the genome and are therefore expected to increase the accuracy of breeding value estimates and thus of the genetic progress attainable when

these predictions are used for selection. Studies by Pérez-Rodriguez et al. (2012) and Burgueño et al. (2012) on prediction ability using two CIMMYT wheat data sets confirmed this assumption and showed that molecular markers increased prediction ability over the pedigree-derived model in all environments of both data sets. Considering markers and pedigree together in the two datasets consistently increased the prediction ability of all models compared with models that used markers or pedigree only. This also can be seen in Table 1, where three different methods (BL: Bayesian Lasso; BRR: Bayesian ridge regression; and RKHS: Reproducing Kernel Hilbert Space) with and without pedigree information were used on 16 different traits. For all traits, models using markers and pedigree combined showed higher prediction ability than modes using markers only.

Models accounting for epistasis

There is concrete proof that the agglomeration of multiple gene × gene interactions (epistasis) having small effects and acting in small epistatic networks is important for explaining the heritability of complex traits such as grain yield (McKinney and Pajewski 2012). Evidence from studies on complex traits conducted at CIMMYT shows that models that allow for non-linear components consistently predicted the individuals in the validation set better than linear models (Crossa et al. 2010; Gonzalez-Camacho et al. 2012; Pérez -Rodriguez et al. 2012). Simulated data by Gonzalez-Camacho et al. (2012) indicate that the Reproducing Kernel Hilbert Space (RKHS) regression approach and the Radial Basis Function Neutral Network (RBFNN) model captured epistatic effects; however, adding redundant predictors (e.g., the interaction between markers) can adversely affect the prediction accuracy of the non-linear regression models.

Models accounting for GE

Genotype × environment interaction produces environments that are structured into related and unrelated subsets that may increase or decrease prediction accuracy. In plants, prediction accuracy has been estimated by evaluating training and validation populations in single environments or in a subset of similar environments (e.g., within irrigated or drought environments). Burgueño et al. (2012) studied the effect of GE modeling on genomic prediction ability when using pedigree and marker information. They used a multi-environment version of the genomic best linear unbiased predictor (G-BLUP), where GE was modeled using genetic cor-relations between environments, and found that the multi-environment G-BLUP had higher prediction accuracy than the single-trait G-BLUP so benefit from borrowing information from correlated environments and from using pedigree and genetic marker information. Jarquin et al. (2013) proposed a model that also allows incorporating main and interaction effects of markers, as well as environmental covariances such as climatic records or soil characteristics.

Conclusions

The large scale implementation of GS in breeding programs will shift efforts from evaluating the whole plant to evaluating marker effects. We think that initial GS implementation should not significantly affect the way plant field trials are conducted in each breeding program as there are still many unanswered questions regarding how to achieve the optimal balance between genotyping and phenotyping and the best use of marker effect evaluations to maximize the overall genetic gain for single or multiple target traits in a particular breeding program. Significant challenges also remain with respect to the successful implementation of GS. The cost of genotyping large numbers of SNPs is still an impediment, although technologies such as Genotyping-by-Sequencing (GBS) are reducing these costs significantly. Collecting large meaningful reference populations can also be costly due to extensive phenotyping. Furthermore, logistical limitations such as DNA extraction turn-around time, SNP genotyping and biometric analysis have to be carefully considered. Improved databasing and bi-informatics pipelines will be needed to support rapid analyses. These challenges warrant further research and a progressive increase in implementation of GS in areas where studies inside and outside CIMMYT indicate is the greatest opportunities to accelerate genetic gain and transform plant breeding.

References

Bernardo R. and J. Yu. 2007. Prospects for genome-wide selection for quantitative traits in maize. Crop Science 47:1082–1090

Bernardo R. 2013. Genomewide selection when major genes are known. Crop Science. 2014. 54:68-75.

Burgueño J., G. de los Campos, K. Weigel and J. Crossa. 2012. Genomic prediction of breeding values when modeling genotype × envi¬ronment interaction using pedigree and dense molecular markers. Crop Science 52:707

Crossa J., J. Burgueño, P.L. Cornelius, G. McLaren, R. Trethowan and A. Krishnamachari. 2006. Modeling geno¬type × environment interaction using additive genetic covari¬ances of relatives for predicting breeding values of wheat geno¬types. Crop Science 46:1722.

Crossa J., G. de los Campos, P. Perez, D. Gianola, J. Burgueño, J.L. Araus, D. Makumbi, R.P. Singh, S. Dreisigacker, J. Yan, V. Arief, M. Banzinger and H-J. Braun. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186(2): 713–724

Crossa J., P. Pérez, J. Hickey, J. Burgueño, L. Ornella, J. Cerón-Rojas, X. Zhang, S. Dreisigacker, R. Babu, Y. Li, D. Bonnett and K. Mathews. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112, 48–60

de los Campos G, H. Naya, D. Gianola, J. Crossa, A. Legarra, E. Manfredi, K. Weigel and J.M. Cotes. 2009. Predicting quan¬titative traits with regression models for dense molecular markers and pedigree. Genetics 182:375–385

de los Campos G, P. Pérez, A.I. Vazquez and J. Crossa. 2013. Genome-enabled prediction using the *BLR* (Bayesian Linear Regression) R-package. Genome-Wide Association Studies and Genomic Prediction. New York: Springer. pp. 299–320

González-Camacho J.M., G. de los Campos, P. Perez, D. Gianola, J.E. Cairns, G. Mahuku, R. Babu and J. Crossa. 2012. Genome-enabled prediction of genetic values using radial basis function. Theoretical and Applied Genetics 125: 759–771

Guo Z., D.M. Tucker, C.J. Basten, H. Gandhi, E. Ersoz, B. Guo, Z. Xu, D. Wang and G. Gilles. 2014. The impact of population structure on genomic prediction in stratified populations. Theoretical and Applied Genetics 127:749–762

Hickey J.M., S. Dreisigacker, J. Crossa, S. Hearne, R. Babu, B.M. Prasanna, M. Grondona, A. Zambelli, V.S. Windhausen, K. Mathews and G. Gorjanc. 2014. Evaluation of Genomic Selection Training Population Designs and Genotyping Strategies in Plant Breeding Programs Using Simulation. Crop Science 54:1476–1488

Jarquín D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, F. Piraux, L. Guerreiro, P. Pérez and M. Calus. 2013. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theoretical and Applied Genetics. 1-13

Lin Z., B.J. Hayes and H.D. Daetwyler. 2014. Genomic selection in crops, trees and forages: a review. Crop and Pasture Science

Lorenz A.J., S.M. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, K.P. Smith, M.E. Sorrells and J.L. Jannink. 2011. Genomic selection in plant breeding: knowledge and prospects. In 'Advances in Agronomy'. Vol. 110. (Ed. DL Sparks). Elsevier Academic Press Inc.: San Diego, CA, USA. pp. 77–123

McKinney B.A. and N.M. Pajewski. 2012. Six degrees of epistasis: statistical network models for *GWAS*. Frontiers in Genetics 2: 109

Meuwissen, T.H.E. B.J. Hayes and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Meuwissen T.H.E. 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense *SNP* genotyping. Genetics Selection Evolution 41:35.

Ornella L., S. Singh, P. Pérez, J. Burgueño, R. Singh, E. Tapia, S. Bhavani, S. Dreisigacker, H.J. Braun, K. Mathews and J. Crossa. 2012. Genomic prediction of genetic values for resistance to wheat rusts. The Plant Genome 5:136–148

Perez-Rodriguez P., D. Gianola , J.M. González-Camacho, J. Crossa, Y. Manes and S. Dreisigacker. 2012. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. G3 – Genes, Genomes, Genetics 2:1595–1605

Poland J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, S. Dreisigacker, J. Crossa, H. Sanchez-Villeda, M. Sorrells and J-L. Jannink. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. The Plant Genome 5 (3): 103-113

Purdy L.H., W.Q. Loegering, C.F. Konzak, C.J. Peterson and R.E. Allan. 1968. A Proposed Standard Method for Illustrating Pedigrees of Small Grain Varieties. Crop Sciene 8:405-406

R Development Core Team. 2010. R: A Language and environment for statistical computing. Available at http://www.r-project.org/

Rutkoski J.E., J.A. Poland, R.P. Singh, J. Huerta-Espino, S. Bhavani, H. Barbier, M.R. Rouse, J-L. Jannink and M.E. Sorrells. 2014. Genomic Selection for Quantitative Adult Plant Stem Rust Resistance In Wheat doi: 10.3835/plantgenome2014.02.0006

Singh R.P., D.P. Hodson, J. Huerta-Espino, Y. Jin, S. Bhavani, P. Njau, S. Herrera-Foessel, P.K. Singh, S. Singh and V. Govindan. 2011. The emergence of *Ug99* races of the stem rust fungus is a threat to world wheat production. Annual Review of Phytopathology 49: 465–481

Solberg T.R., A.K. Sonesson, J.A. Woolliams and T.H.E. Meuwissen. 2008. Genomic selection using different marker types and densities. Journal of Animal Science 86:2447–54

Wimmer, V., C. Lehermeier, T. Albrecht, H.J. Auinger, Y. Wang and C.C. Schön. 2013. Genome-wide prediction of traits with different genetic architecture through efficient variable selection. Genetics, 195(2):573-587

Windhausen V.S., G.N. Atlin, J. Crossa, J.M. Hickey, J. Crossa, J-L. Jannink, M. Sorrells, B. Raman, J. Cairns, A. Terekegne, K. Semagn, Y. Beyene, P. Grudloyma, F. Technow, C. Riedelsheimer and A.E. Melchinger. 2012. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3* - Genes Genomes Genetics 2:1427–1436

Xu Y. and J.H. Crouch. 2008. Marker-assisted selection in plant breeding: From publications to practice. Crop Science 48:391–407

Appendix 1. List of gene-specific STS markers currently in use at CIMMYT
(see a more complete list in http://repository.cimmyt.org/dvn)

Trait	Locus/Gene	Species	Allele	Marker	Chr.	Source
Leaf rust	*Lr3*	DW	*Lr3+*	*mwg798*	6B	Herrera-Foessil et al. 2007
Leaf rust	*Lr14*	DW	*Lr14a+*	*gwm146*	7BL	Herrera-Foessil et al. 2008
Leaf rust	*Lr14*	DW	*Lr14a+*	*gwm344*	7BL	Herrera-Foessil et al. 2008
Leaf rust	*Lr42*	*Ae. tauschii*	*Lr42+*	*wmc432*	1D	Sun et al. 2010, Basnet et al. 2014
Leaf rust	*Lr47*	*Ae. speltoides*	*Lr47+*	*PS10*	7A	Helguera et al. 2000
Leaf rust	*Lr53*	*Tr. dicoccoides*	*Lr53+*	*cfd1*	6BS	Dadkhodaie et al. 2011
Leaf rust	*Lr72*	DW	*Lr72+*	*wmc606*	7BS	Herrera-Foessil et al. 2014
Stem rust	*SrAes1t*	*Ae. speltoides*	*SrAes1t+*	*csSrAes1t*	2D	Mago et al. 2013
Stem rust	*SrCad/Sr42*	BW	*SrCad/Sr42+*	*FSD,RSA*	6DS	Laroche et al. 2000
Stem rust	*Sr22*	*Tr. monococcum*	*Sr22+*	*cfa2123*	7AL	Khan et al. 2005
Stem rust	*Sr22*	*Tr. monococcum*	*Sr22+*	*csKP81*	7AL	
Stem rust	*Sr24*	*Ae. Elongatum*	*Sr24+*	*Sr24 #12*	3DL	Mago et al. 2005
Stem rust	*Sr26*	*Ae. Elongatum*	*Sr26+*	*Sr26#43*	6AL	Mago et al. 2005
Stem rust	*Sr26*	*Ae. Elongatum*	*Sr26+*	*BE518379*	6AL	Liu et al. 2009
Stem rust	*Sr32*	*Ae. speltoides*	*Sr32+*	*csSr32#1*	2D	Mago et al. 2013
Stem rust	*Sr36*	*Tr. Timopheevi*	*Sr36+*	*stm773-2*	2BL	Tsilo et al. 2007, Bariana et al. 2001
Stem rust	Sr39/*Lr35*	*Ae. speltoides*	*Sr39+*	*Sr39#22r*	2B	Mago et al. 2005
Stem rust	*Sr39/Lr35*	*Ae. speltoides*	*Sr39+*	*Sr39#50s*	2B	Mago et al. 2005
Stem rust	*Sr50 (formerly SrR)*	rye	*Sr50+*	*IB-267*	1DS	Mago et al. 2002
Stem rust	*SrND643*	BW	*SrND643+*	*gwm350*	4AL	Basnet et al. 2014
Stem rust	*SrND643*	BW	*SrND643+*	*wmc219*	4AL	Basnet et al. 2014
Yellow rust	*Yr24/Yr26*	*T. ae.H. villosa 6VS/6AL*	*Yr24/Yr26+*	*We173*	6AL	Wang et al. 2008
Yellow rust	*Yr39*	BW	*Yr39+*	*wgp36*	7BL	Lin et al. 2007
Yellow rust	*Yr39*	BW	*Yr39+*	*wgp45*	7BL	Lin et al. 2007
Yellow rust	*Yr41*	BW	*Yr41+*	*gwm410*	2BS	Luo et al. 2008
Yellow rust	*Yr41*	BW	*Yr41+*	*gwm374*	2BS	Luo et al. 2008
Yellow rust	*Yr51*	BW	*Yr51+*	*sun104*	4AL	Randhawa et al. 2013
Yellow rust	*Yr52*	BW	*Yr52+*	*barc182*	7BL	Ren et al. 2012
Yellow rust	*Yr52*	BW	*Yr52+*	*wgp5258*	7BL	Ren et al. 2012
Yellow rust	*Yr57*	BW	*Yr57+*	*gwm389*	3BS	Randhawa et al. 2015
Yellow rust	*Yr59*	BW	*Yr59+*	*wgp5175*	7B	Zhou et al. 2014
Yellow rust	*Yr59*	BW	*Yr59+*	*barc32*	7B	Zhou et al. 2014
Yellow rust	*Yr60*	BW	*Yr60+*	*wmc776*	4AL	Herrera-Foessel et al. 2014
Yellow rust	*YrF*	BW	*YrF+*	*gwm374*	2BS	Lan et al. 2014
Yellow rust	*YrF*	BW	*YrF+*	*wmc474*	2BS	Lan et al. 2014
Combined resistance	*Sr2/Yr30*	BW	*Sr2/Yr30+*	*csSr2 digestion with BspH1*	3BS	Spielmayr et al. 2011
Combined resistance	*Lr16/Sr23*	BW	*Lr16/Sr23+*	*gwm210*	2B	McCartney et al. 2005
Combined resistance	*Lr19/Sr25/Psy-E1*	*Thinopyrum ponticum*	*Lr19/Sr25/Psy-E1+*	*wmc221*	7D	Isaac et al. 2004
Combined resistance	*Lr19/Sr25/Psy-E1*	*Thinopyrum ponticum*	*Lr19/Sr25/Psy-E1+*	*Psy1_F3/Xr1, Psy1-EF2/ER4*	7A	Zhang and Dubcovsky, 2008
Combined resistance	*Lr19/Sr25/Psy-E1*	*Thinopyrum ponticum*	*Lr19/Sr25/Psy-E1+*	*Psy1-DF2/R3, Psy1-EF2/ER4*	7D	Zhang and Dubcovsky, 2008
Combined resistance	*Lr34/Yr18/Sr57/Sb1/ Pm38/Ltn1*	BW	*Lr34/Yr18/Sr57/Sb1/ Pm38/Ltn1+*	*cslv34+cssfr1*	7DS	Lagudah et al. 2006/2009
Combined resistance	*Lr37/Sr38/Yr17 (VPM)*	BW	*VPM+*	*Ventriup +Ln*	2AS	Helguera et al. 2003
Combined resistance	*Lr46/Yr29/Sr58/Pm39*	BW	*Lr46/Yr29/Sr58/Pm39+*	*csLV46 digestion with Taq1*	1BL	Lagudah personal comm.
Combined resistance	*Lr67/Yr46*	*BW*	*Lr67/Yr46+*	*csLV67*	4DL	Moore et al. 2015

Trait	Locus/Gene	Species	Allele	Marker	Chr.	Source
Combined resistance	*Lr68*	BW	*Lr68+*	*cs7BLNLRR digestion with HaeIII*	7BL	Herrera-Foessil et al. 2012
Fusarium (spread)	*Fhb1*	BW	*Fhb1+*	*Umn10*	3BS	Liu et al. 2008
Fusarium (spread, severity)	*Fhb2*	BW	*Fhb2+*	*gwm133*	6BS	Cuthbert et al. 2007
Fusarium (spread, severity)	*Fhb2*	BW	*Fhb2+*	*gwm644*	6BS	Cuthbert et al. 2007
Fusarium (spread)	*Qfhs.ifa-5A (Sumai)*	BW	*Qfhs.ifa-5A+*	*barc186*	5AS	Buerstmayr et al. 2003
Fusarium (spread)	*Qfhs.ifa-5A (Sumai)*	BW	*Qfhs.ifa-5A+*	*barc180*	5AS	Buerstmayr et al. 2003
Fusarium (severity)	*QTL_3A (Frontana)*	BW	*QTL_3A+*	*dupw227*	3A	Steiner et al. 2004
Fusarium (severity)	*QTL_3A (Frontana)*	BW	*QTL_3A+*	*gwm2*	3A	Steiner et al. 2004
Fusarium (spread)	*QTL_3A (DIC FA-15-3)*	*Tr. dicoccoides*	*QTL_3A+*	*gwm2*	3A	Otto et al. 2002, Chen et al. 2007
Fusarium (severity)	*QTL_5A (Frontana)*	BW	*QTL_5A+*	*barc197*	5A	Steiner et al. 2004
Fusarium (severity)	*QTL_5A (Frontana)*	BW	*QTL_5A+*	*gwm129*	5A	Steiner et al. 2004
Fusarium (spread)	*QTL_2D (Wuhan1)*	BW	*QTL_2D+*	*wmc144*	2DL	Somers et al. 2003
Fusarium (spread)	*QTL_2D (Wuhan1)*	BW	*QTL_2D+*	*wmc245*	2DL	Somers et al. 2003
Fusarium (severity)	*Fhb4, QTL_4B (Wuhan1)*	BW	*Fhb4+*	*wmc238*	4BS	Somers et al. 2003
Fusarium (severity)	*Fhb4, QTL_4B (Wuhan1)*	BW	*Fhb4+*	*gwm149*	4BS	Somers et al. 2003, Xue et al. 2011
Fusarium (spread)	*Fhb4 Qfhi.nau-4B (WSB)*	BW	*Fhb4+*	*hbg226*	4BS	Xue et al. 2010
Fusarium (spread, T1)	*Fhb5, Qfhi.nau-5A (WSB)*	BW	*Fhb5+*	*gwm415*	5A	Xue et al. 2011
Fusarium (spread, T1)	*Fhb5, Qfhi.nau-5A (WSB)*	BW	*Fh5b+*	*gwm304*	5A	Xue et al. 2011
Fusarium (spread, DON)	*QTL_2D (CJ9306)*	BW	*QTL_2D+*	*gwm157*	2D	Jiang et al. 2007
Fusarium (spread, DON)	*QTL_2D (CJ9306)*	BW	*QTL_2D+*	*gwm539*	2D	Jiang et al. 2007
Fusarium (spread)	*QTL_2D (Gamenya)*	BW	*QTL_2D+*	*gwm261*	2D	Handa et al 2008
Fusarium (spread)	*QTL_7A (DIC PI478742)*	*Tr. dicoccoides*	*QTL_7A+*	*barc121*	7A	Kumar et al. 2007
Fusarium (spread)	*QTL_7A (DIC PI478742)*	*Tr. dicoccoides*	*QTL_7A+*	*wmc488*	7A	Kumar et al. 2007
Septoria tritici blotch	*stb1*	BW	*stb1+*	*barc74*	5BL	Adhikari et al. 2004
Septoria tritici blotch	*stb1*	BW	*stb1+*	*gwm335*	5BL	Adhikari et al. 2004
Septoria tritici blotch	*stb2*	BW	*stb2+*	*barc008*	1BS	Liu et al. 2013
Septoria tritici blotch	*stb2*	BW	*stb2+*	*wmc230*	1BS	Liu et al. 2013
Septoria tritici blotch	*stb2*	BW	*stb2+*	*wmc406*	1BS	Liu et al. 2013
Septoria tritici blotch	*stb3*	BW	*stb3+*	*wmc83*	7A	Adhikari et al. 2004, Goodwin 2011
Septoria tritici blotch	*stb4*	BW	*stb4+*	*gwm111*	7DS	Adhikari et al. 2004
Septoria tritici blotch	*stb5*	BW	*stb5+*	*gwm44*	7DS	Arraiano et al. 2001, Simon et al. 2007
Septoria tritici blotch	*stb6*	BW	*stb6+*	*gwm369*	3AS	Brading et al. 2001, Chartrain et al. 2005
Septoria tritici blotch	*stb7/stb12*	BW	*stb7/stb12+*	*wmc313*	4AL	McCartney et al. 2003, Chartrain et al 2005
Septoria tritici blotch	*stb7/stb12*	BW	*stb7/stb12+*	*wmc219*	4AL	McCartney et al. 2003, Chartrain et al 2005
Septoria tritici blotch	*stb8*	BW	*stb8+*	*gwm146*	7BL	Adhikari et al. 2003
Septoria tritici blotch	*stb8*	BW	*stb8+*	*gwm577*	7BL	Adhikari et al. 2003
Septoria tritici blotch	*stb9*	BW	*stb9+*	*wmc317*	6AS	Chartrain et al. 2009
Septoria tritici blotch	*stb10*	BW	*stb10+*	*gwm848*	1D	Chartrain et al. 2005
Septoria tritici blotch	*stb11*	BW	*stb11+*	*barc008*	1BS	Chartrain et al. 2005
Septoria tritici blotch	*stb11*	BW	*stb11+*	*barc137*	1BS	Chartrain et al. 2005
Septoria tritici blotch	*stb13*	BW	*stb13+*	*wmc396*	7B	McCartney 2002, Cowling et al. 2007
Septoria tritici blotch	*stb14*	BW	*stb14+*	*wmc623*	3B	McCartney 2002, Brule Babel 2007
Septoria tritici blotch	*stb14*	BW	*stb14+*	*wmc500*	3B	McCartney 2002
Septoria tritici blotch	*stb16*	SHW	*stb16+*	*wmc494*	3D	Ghaffary et al. 2012
Septoria tritici blotch	*stb17*	SHW	*stb17+*	*hbg247*	5A	Ghaffary et al. 2012
Septoria tritici blotch	*stb18*	BW	*stb18+*	*gpw5176*	6DS	Ghaffary et al. 2011
Septoria tritici blotch	*stb18*	BW	*stb18+*	*gpw3087*	6DS	Ghaffary et al. 2011
Ceral cyst nemathode	*Cre1*	BW	*Cre1+*	*Cre1M19*	2BL	Ogbonnaya et al. 2001
Ceral cyst nemathode	*Cre3*	BW	*Cre3+*	*Cre3SP*	2DL	Martin et al. 2004
Ceral cyst nemathode	*Cre8*	BW	*Cre8+*	*wri15*	6BL	Jayatilake et al. 2014

Trait	Locus/Gene	Species	Allele	Marker	Chr.	Source
Root lesion nematode	*Rlnn1*	BW	*Rlnn1+*	*uat0002*	7BL	Diane Mather, personal communication
Crown rot	*2.49 1DL QTL*	BW	*QTL (2.49)+*	*wmc429*	1DL	Collard et al. 2005
Crown rot	*2.49 1DL QTL*	BW	*QTL (2.49)+*	*wmc216*	1DL	Collard et al. 2005
BYDV	*Bdv-2 (TC14)*	*Thinopyrum intermedium*	*Bdv-2 (TC14)+*	*BYAgi*	7DL	Stoutjesdijk 2001
Hessian Fly	*H25*	rye	*H25+*	*gwm610*	4A	http://maswheat.ucdavis.edu/protocols/H25/index.htm
Grain protein content (DIC allele)	*Gpc-B1*	BW	*high-GPC*	*uhw89*	6BS	Distelfeld et al. 2006
Grain texture Hardness (Ha-locus)	*Pina-D1*	BW	*Pina-D1a/b*	*PinaD1*	5D	Gautier et al. 1994
Grain texture Hardness (Ha-locus)	*Pinb-D1*	BW	*Pinb-D1a (wild type)*	*PB5/SR*	5D	Gautier et al. 1994
Grain texture Hardness (Ha-locus)	*Pinb-D1*	BW	*Pinb-D1b (mutant)*	*PB5/HR*	5D	Gautier et al. 1994
Gluten strength	*Glu-A1*	BW	*Ax1, Ax2*, Ax-null*	*UMN19*	1AL	Liu et al. 2008
Gluten strength	*Glu-B1*	BW	*7OE*	*TaBAC1215C06-F517/R964*	1B	Raja et al. 2008
Gluten strength	*Glu-B1*	BW	*7OE*	*TaBAC1215C06-F24671/R25515*	1B	Raja et al. 2008
Gluten strength	*Glu-D1*	BW	*Glu-D1d (x5+y10)*	*Dx*	1DL	Ishikawa et al. 2007
Gluten strength	*Glu-D1*	BW	*Dx2, Dx5*	*UMN25*	1DL	Liu et al. 2008
Gluten strength	*Glu-D1*	BW	*Dy10, Dy12*	*UMN26*	1DL	Liu et al. 2008
Gluten strength	*Glu-A3*	BW	*GluA3a*	*LA1F,SA1R*	1AS	Wang et al. 2010
Gluten strength	*Glu-A3*	BW	*GluA3b*	*LA3F,SA2R*	1AS	Wang et al. 2010
Gluten strength	*Glu-A3*	BW	*GluA3ac*	*LA1F,SA3R*	1AS	Wang et al. 2010
Gluten strength	*Glu-A3*	*BW*	*GluA3d*	*LA3F,SA4R*	1AS	Wang et al. 2010
Gluten strength	*Glu-A3*	BW	*GluA3e*	*LA1F,SA5R*	1AS	Wang et al. 2010
Gluten strength	*Glu-A3*	BW	*GluA3f*	*LA1F,SA6R*	1AS	Wang et al. 2010
Gluten strength	*Glu-A3*	BW	*GluA3g*	*LA1F,SA7R*	1AS	Wang et al. 2010
Gluten strength	*Glu-B3*	BW	*Glu-B3a*	*SB1*	1BS	Wang et al. 2009
Gluten strength	*Glu-B3*	BW	*Glu-B3b*	*SB2*	1BS	Wang et al. 2009
Gluten strength	*Glu-B3*	BW	*Glu-B3c*	*SB3*	1BS	Wang et al. 2009
Gluten strength	*Glu-B3*	BW	*Glu-B3d*	*SB4*	1BS	Wang et al. 2009
Gluten strength	*Glu-B3*	BW	*Glu-B3e*	*SB5*	1BS	Wang et al. 2009
Gluten strength	*Glu-B3*	BW	*Glu-B3fg*	*SB6*	1BS	Wang et al. 2009
Gluten strength	*Glu-B3*	BW	*Glu-B3g*	*SB7*	1BS	Wang et al. 2009
Gluten strength	*Glu-B3*	BW	*Glu-B3h*	*SB8*	1BS	Wang et al. 2009
Gluten strength	*Glu-B3*	BW	*Glu-B3i*	*SB9*	1BS	Wang et al. 2009
Gluten strength	*Glu-B3*	BW	*Glu-B3bef*	*SB10*	1BS	Wang et al. 2009
Starch properties	*Wx-A1*	BW	*Wx-A1a/b*	*MAG264*	7A	Liu et al. 2005
Starch properties	*GBSS – Null/Wx-B1*	BW	*Wx-B1a/b*	*GBSS*	4A	McLauchlan et al. 2001
Starch properties	*GBSS –Wx-B1*	BW	*Wx-B1a*	*GBSS*	4A	Saito 2009
Starch properties	*GBSS – Null*	BW	*Wx-B1b*	*GBSS*	4A	Saito 2009
Starch properties	*Wx-D1a/b*	BW	*Wx-D1a/b*	*MAG269*	7D	Liu et al. 2005
bread making	*wmb*	BW	*wbm+*	*NWP*	7	Furtado et al. 2015
Lipoxygenase activity	*Lox-B1*	DW	*Lox-B1.1+*	*LOXA*	4B	Carrera et al. 2007
Cadmium concentration	*Cdu1-B1*	DW	*Cdu1+*	*usw47 digestion with Hpy188*	5BL	Wiebe et al. 2010
Cadmium concentration	*Cdu1-B1*	DW	*Cdu1+*	*HMA3-B1*	5BL	Wiebe et al. 2014 (phd thesis)
Height	*Rht-B1*	BW	*Rht-B1b, Rht1*	*BF/MR1*	4B	Ellis et al. 2002
Height	*Rht-B1*	BW	*Rht-B1a, rht1*	*BF/WR1*	4B	Ellis et al. 2002
Height	*Rht-D1*	BW	*Rht-D1b, Rht2*	*DF/MR2*	4D	Ellis et al. 2002
Height	*Rht-D1*	BW	*Rht-D1a, rht2*	*DF2/WR2*	4D	Ellis et al. 2002

Trait	Locus/Gene	Species	Allele	Marker	Chr.	Source
Pairing homologue	*Ph1*	BW	*ph1b*	*Qu9-5*	5BL	Qu et al. 1998
Pairing homologue	*Ph1*	BW	*ph1b*	*wms213*	5BL	
Crossability	*Skr*	BW	*Skr+*	*cfb341*	5BS	Alfares et al. 2009
Necrosis	*Ne2*	*BW*	*Ne2+*	*barc55*	2BS	Chu et al. 2006
Tiller inhibition	*tin1*	BW	*tin1+*	*gwm136*	1A	Spielmayr et al. 2004
Photoperiod	*Ppd-A1*	DW	*Ppd-A1a/b*	*Ag5del (2A_Ins1)*	2A	Bentley et al. 2011
Photoperiod	*Ppd-B1*	BW	*Ppd-B1a*	*gwm148*	2B	Hanocq et al. 2007
Photoperiod	*Ppd-B1*	BW	*Ppd-B1a*	*gwm257*	2B	Hanocq et al. 2007
Photoperiod	*Ppd-B1*	BW	*Ppd-B1a (CS)*	*219H05F2,97J10R2*	2B	Diaz et al. 2012
Photoperiod	*Ppd-B1*	BW	*Ppd-B1a (CS)*	*PpdB1_F25,PpdB1_R70*	2B	Diaz et al. 2012
Photoperiod	*Ppd-B1*	BW	*Ppd-B1a (Sonora 64)*	*PpdB1_F3/PpdB1_R36*	2B	Diaz et al. 2012
Photoperiod	*Ppd-B1*	BW	*Ppd-B1a (1 and 2 copies)*	*CNV10*	2B	Diaz et al. 2012
Photoperiod	*Ppd-B1*	BW	*Ppd-B1a (1 and 2 copies)*	*CNV13*	2B	Cane et al. 2013
Photoperiod	*Ppd-D1*	BW	*Ppd-D1a/b*	*Ppd-D1*	2D	Baeles et al. 2007
Vernalization	*Vrn-A1 (promoter)*	BW	*Vrn-A1a, Vrn-A1c, Vrn-A1b, vrn-A1*	VRN-A	5A	Yan et al. 2004
Vernalization	*Vrn-A1 (Intron1)*	BW	*Vrn-A1c*	*Intr1/A/F2, R3*	5A	Fu et al. 2005
Vernalization	Vrn-*A1 (Intron1)*	BW	*vrn-A1*	*Intr1/C/F, R*	5A	Fu et al. 2005
Vernalization	*Vrn-A1 (intron1)*	DW	*Vrn-A1 (Langdon)*	*Ex1/C/F, Intr1/A/R3*	5A	Yan et al. 2004
Vernalization	*Vrn-B1 (Intron 1)*	BW	*Vrn-B1a, Vrn-B1b, vrn-B1*	*Intro 1/B/F, R3, R4*	5B	Fu et al. 2005
Vernalization	*Vrn-B1 (Intron 1)*	BW	*Vrn-B1c*	*Ex1/B/F3, Intro 1/B/R3*	5B	Milec et al. 2012
Vernalization	*Vrn-D1 (Intron 1)*	BW	*Vrn-D1, vrn-D1*	*Intro 1/D/F. R3, R4*	5D	Fu et al. 2005
Vernalization	*Vrn-B3*	BW	*Vrn-B3*	*VRN4-B-INS-F_VRN4-B-INS-R*	7BS	Yan et al. 2006
Vernalization	*Vrn-B3*	BW	*vrn-B3*	*VRN4-B-NONINS-F_VRN4-B-NONINS-R*	7BS	Yan et al. 2006
wheat–rye translocation	*T1AL.1RS and T1BL.1RS*	rye	*T1AL.1RS and T1BL.1RS*	*SCM009*	1R	Weng et al. 2007
Sucrose Synthase	*TaSus1-7B*	BW	*TaSus1, Hap-T, Hap-C*	*Sus1-7B-2548f, 3671r digestion with Sph1*	7B	Hou et al. 2014
Sucrose Synthase	*TaSus1-7A*	BW	*TaSus1, Hap-1 to 5*	*Sus1-7A-539f,1720r digestion with TaqαI*	7A	Hou et al. 2014
Sucrose Synthase	*TaSus1-7A*	BW	*TaSus1, Hap-1 to 5*	*Sus1-7A-2636f, 3696r digestion with ApaL1*	7A	Hou et al. 2014
Sucrose Synthase	*TaSus1-7A*	BW	*TaSus1, Hap-1 to 5*	*Sus1-7A-1149f, 1620r*	7A	Hou et al. 2014
Sucrose Synthase	*TaSus2-2A*	BW	*TaSus2, Hap-G, Hap-A*	*Sus2-2AP-302f, Sus-2A214r with Asc1*	2A	Hou et al. 2014
Sucrose Synthase	*TaSus2-2B*	BW	*TaSu2, Hap-H, Hap-L*	*Sus2-227, Sus2-589L2*	2B	Jiang et al. 2011
Thousand kernel weight	*TaCWi-2A*	BW	*TaCWi-A1a*	*CWI21, CWI22*	2A	Ma et al. 2012
Thousand kernel weight	*TaCKX6*	BW	*TaCKX-D1a*	*C19*	3D	Zhang et al. 2012
Thousand kernel weight	*TaGW2-6A*	BW	*TaGW2-6A, Hap-A, Hap-G*	*Hap-6A-P2 digestion with Taq1*	6A	Su et al. 2011
Thousand kernel weight	*TaGW2-6B*	BW	*TaGW2-6B, Hap-1 to 4*	*TaGW2-6B-CAPS digestion with BstNI*	6B	Quin et al. 2014
Thousand kernel weight	*TaGW2-6B*	BW	*TaGW2-6B, Hap-1 to 4*	*TaGW2-6B-ACAS1*	6B	Quin et al. 2014
Thousand kernel weight	*TaGW2-6B*	BW	*TaGW2-6B, Hap-1 to 4*	*TaGW2-6B-ACAS2*	6B	Quin et al. 2014
Thousand kernel weight	*TaGW2-6B*	BW	*TaGW2-6B, Hap-1 to 4*	*TaGW2-6B-dCAPS digestion with Hpy166II*	6B	Quin et al. 2014
Thousand kernel weight	*TaGS*	BW	*TaGS-D1a*	*GS7D*	7D	Zhang et al. 2014
Bo tolerance	*Bo1*	BW	*Bo1+*	*AWW5L7*	7BL	Schnurbusch et al. 2007

Appendix 2. List of gene-specific SNP markers currently in use at CIMMYT

(see a more complete list in http://repository.cimmyt.org/dvn)

Trait	Gene	SNP id	Index	Marker Name	FAM primer	VIC primer
Plant height	*Rht-B1*	wMAS000001	CIMwMAS0028	*Rht-B1_SNP*	CCCATGGCCATCTCSAGCTG	CCCATGGCCATCTCSAGCTA
Plant height	*Rht-D1*	wMAS000002	CIMwMAS0029	*Rht-D1_SNP*	CATGGCCATCTCGAGCTRCTC	CATGGCCATCTCGAGCTRCTA
Rust	*Lr34*		CIMwMAS0001	*Lr34_TCCIND*	GGTATGCCATTTAACATAATCATGAA	GGTATGCCATTTAACATAATCATGAT
Rust	*Sr2*	wMAS000005	CIMwMAS0003	*Sr2_ger9 3p*	GTGCGAGACATCCAACACTCAC	GTGCGAGACATCCAACACTCAT
Rust	*Lr37/Yr17/Sr38*		CIMwMAS0004	*VPM_SNP*	CGCCGTTCCGAAYACGAGA	CGCCGTTCCGAAYACGAGG
Fusarium	*Fhb1*	wMAS000008	CIMwMAS0008	*snp3BS-8*	CACATGCATTTGCAAGGTTGTTATCC	CACATGCATTTGCAAGGTTGTTATCG
Fusarium	*Fhb1*	wMAS000009	CIMwMAS0007	*UMN10_SNP*	GAATTACTCATTTTTAGATTTGTCTACATACA	GAATTACTCATTTTTAGATTTGTCTACATACG
Quality	*GluD1*	wMAS000014	CIMwMAS0013	*Glu-D1d_SNP*	ATAGTATGAAACCTGCTGCGGAG	ATAGTATGAAACCTGCTGCGGAC
Quality	*Gpc-B1*	wMAS000017	CIMwMAS0024	*GCP_DUP*	CAAGAGGGGAGAGACATGTTACTTA	CAAGAGGGGAGAGACATGTTACTTT
Quality	*Pinb-D1*	wMAS000018	CIMwMAS0025	*Pinb-D1_INS*	CTCATGCTCACAGCCGCC	CCTCATGCTCACAGCCGCT
Quality	*Pina-D1*	wMAS000019	CIMwMAS0026	*Pina-D1_INS*	AACTGCCAACAACTTCGCTA	TTGTCTAGTACCCCGCTCTG
Tan spot	*Tsn1*	wMAS000020	CIMwMAS0009	*Tsn1*		CTATTCGTAATCGTGCCTTCCGG
Quality	*TaSus2*	wMAS000021	CIMwMAS0027	*TaSus2-2B_SNP*	GCGGTGTCCTTGAGCTTCTCA	GCGGTGTCCTTGAGCTTCTCG
Rust	*Yr36*	wMAS000022		*WKS*	CGATGCTTCTCTCAGAACGA	TTCGATGCTTCTTGTAGAACACA
Photoperiod	*Ppd-D1*	wMAS000024	CIMwMAS0040	*TaPpdDD001*	CAAGGAAGTATGAGCAGCGGTT	AAGAGGAAACATGTTGGGGTCC
Photoperiod	*Ppd-D1*	wMAS000025	CIMwMAS0041	*TaPpdDI001*	TGACTTATACACCCGGACGGAG	GAACATGACACACAACCAACGC
Photoperiod	*Ppd-D1*	wMAS000026	CIMwMAS0042	*TaPpdDD002*	CGAGCAGCTCCCGACG	GGGCGAGCAGCTCCAAC
Photoperiod	*Ppd-B1*	wMAS000027	CIMwMAS0035	*TaPpdBJ001*	CCGTTTTCGCGGCCTT	GACGTTATGAACGCTTGGCA
Photoperiod	*Ppd-B1*	wMAS000028	CIMwMAS0036	*TaPpdBJ003*		CGTGAAGAGCTAGCGATGAACA
Photoperiod	*Ppd-A1*	wMAS000029	CIMwMAS0033	*Cdex5-6ID*	CATTAGTTTCTTTTGGTTTCTGGCA	CAATCAGATCAGCAGCTCGAAC
Photoperiod	*Ppd-A1*	wMAS000030	CIMwMAS0031	*GS100-1027IND*	CCAGTATCTTTAGATGCACCATGC	GCCGGCGGCTAAAAGG
Photoperiod	*Ppd-A1*	wMAS000031	CIMwMAS0032	*GS105-1117IND*	GGGGACCAAATACCGCTCG	CGTTTGGTGGTGGACGGG
Vernalization	*Vrn-A1*	wMAS000033	CIMwMAS0043	*Vrn-A1_9K0001*	AGAGTTTTCCAAAAAGATAGATCAATGTAAAT	GAGTTTTCCAAAAAGATAGATCAATGTAAAC
Vernalization	Vrn-A1	wMAS000034	CIMwMAS0045	*Vrn1_new*	CAACTCCTTGAGATTCAAAGATTCAAG	GCAACTCCTTGAGATTCAAAGATTCAAA
Vernalization	*Vrn-A1*	wMAS000035	CIMwMAS0046	*Vrn-A1b-Marq*	GTTTTGGCCTGGCCATCCTCC	GTTTTGGCCTGGCCATCCTCA
Vernalization	*Vrn-D1*	wMAS000039	CIMwMAS0048	*Vrn-D1-D1a_A*	ATCATTCGAATTGCTAGCTCCGG	ATCATTCGAATTGCTAGCTCCGC
Soil born diseases	*Cre8*		CIMwMAS0010	*Cre8_SNP*	TTGATTAGGATCAGGGCATTG	TGCTTTGATTAGGATCAGGGCATTC
Quality	*Glu-A1*		CIMwMAS0011	*Glu-Ax1/x2*_SNP*	AAGTGTAACTTCTCCGCAACG	ACCTAAGTGTAACTTCTCCGCAACA
Quality	*Glu-A1*		CIMwMAS0012	*Glu-Ax2_IND*	ATTCTTGTTGTCCTTGTCCTGGCT	CTTGTTGTCCTTGTCCTGGCC
Vernalization	*Vrn-A1*		CIMwMAS0047	*Exon7_C/T_Vrn-A1*	gagtttgatcttgctgcgccG	ctgagtttgatcttgctgcgccA
Soil born diseases	*Rln1*		CIMwMAS0061	*Rlnn1_SNP*	GTCAGGAGAAAAGCAGCCATT	GTCAGGAGAAAAGCAGCCATA
Rust	*Lr14a*		CIMwMAS0054	*ubw14*	CTACACTAGTACTACTTTGAGACAATTTTT	ACACTAGTACTACTTTGAGACAATTTTA
Rust	*Lr47*		CIMwMAS0055	*Lr47-1*	GCAGCCTGGTAAGTTATCTGAC	GCAGCCTGGTAAGTTATCTGAG
Rust	*Lr68*		CIMwMAS0056	*Lr68-2*	CGTGTCTTGGACCTGAGCAAT	CGTGTCTTGGACCTGAGCAAC
Dreb	*Dreb*			*Dreb-B1*	CCTGCGCACTTTCTTCTTCCTGT	CTGCGCACTTTCTTCTTCCTGG
Rust	*Lr67*			*csSNP856 (Lr67)*	GCTACTACTATTGGTAGCCTG	GCTACTACTATTGGTAGCCTA
Rust	*Yr15*			*Yr15-R5*	agtcaacttggattacactgaagtT	agtcaacttggattacactgaagtC
Rust	*Yr15*			*Yr15-R8*	cagatccccggttctctcaaG	cagatccccggttctctcaaA

Common primer	FAM allele	FAM	VIC allele	VIC	Source	Reference
TCGGGTACAAGGTGCGGGCG	C	wildtype	T	Rht-B1a	CerealdB	Ellis et al (2002)
CGGGTACAAGGTGCGCGCC	G	wildtype	T	Rht-D1a	CerealdB	Ellis et al (2002)
TACTATATGGGAGCATTATTTTTTTCC	Ins	Lr34-	Del	Lr34+	CIMMYT-SD	Lagudah et al (2009)
CTCAAATGGTCGAGCACAAGCTCTA	G	Sr2-	A	Sr2 (Hope type)	CerealdB	Mago et al (2011)
CCCTGGCTTGCACCTTCGACAA	T	Lr37/Yr17/Sr38+	C	Lr37/Yr17/Sr38-	CIMMYT-SD	Helguera et al (2003)
CAAAGCAGCCTTAGGTCAATAGTTTGAAA	C	Fhb1-	G	Fhb1+ (Sumai3 type)	CerealdB	Bernardo et al (2011)
GAAGTTCATGCCACGCATATGCTAGTA	A	Fhb1-	G	Fhb1+	CerealdB	Liu et al (2008)
TACTAAAAAGGTATTACCCAAGTGTAACTT	C	2+12 or others	G	5+10	CerealdB	Ishikawa and Nakamura (2007)
GATTATGGGAGTAGGTTGGTGAGATAAAA	A	nromal GPC	T	increased GPC	CerealdB	Distelfeld et al (2006)
GTCACCTGGCCCACAAAATG	C	Pinb-D1a	T	Pinb-D1b	CerealdB	Giroux and Morris (1997)
ATGAAGGCCCTCTTCCTCATAGG	A	Pina-D1a	G	Pina-D1b	CerealdB	Giroux and Morris (1998)
CTGCCCTTCACTTAGCCTGTCAC	Null	Tsn1+	G	Tsn1-	CerealdB	Helguera et al. (2003)
ACTGCTGAGTACAATGCCGCGATCCCA	T	Hap-L	C	Hap-H	CerealdB	Jiang et al (2011)
GATTGGTTCTTGACGTATGTTTT	WKS2	Yr36+	WKS1	Retrotrasposon	CerealdB	Fu et al (2009)
GCCTCCCACTACACTGGGC	wildtype (C)	sensitive	Del (T)	insensitive	CerealdB	Beales et al (2007)
TGTTAATTAATTTGTACTGGCTCGGTA	Ins (G)	sensitive (Mercia type)	wildtype (C)	insensitive	CerealdB	Beales et al (2007)
GGTCTCCAATCAAGGCGGT	wildtype (G)	sensitive	Del (C)	loss of function (null)	CerealdB	Beales et al (2007)
GGGTTCGTCGGGAGCTGT	wildtype (T)	-	Ins (A)	truncated copy	CerealdB	Beales et al (2007)
TGGGCACGTTAACACACCTTT	null	-	A	insensitive	CerealdB	Beales et al (2007)
CCTGAAGTCAGAGATATGCAGCAAC	Ins		wildtype	sensitive (null)	CerealdB	Beales et al (2007)
CTATACAATGCTAAAGTCGCACAT	wildtype (C)	sensitive	Del	insensitive	CerealdB	Beales et al (2007)
GAAACAGAGGGGTGGTTTGAAAT	wildtype	sensitive	Del	insensitive	CerealdB	Beales et al (2007)
GTTAGTAGTGATGGTCCAATAATGCCAAA	A	vrn-A1	G	Vrn-A1	CerealdB	90K assay
CATCCTGCATCTGCAGGCATCTC	C	short vrn	T	long vrn (2147 type)	CerealdB	Chen et al (2009)
TATCAGGTGGTTGGGTGAGGACGT	C	Vrn-A1b	A	Vrn-A1a	CerealdB	Yan et al (2004)
GCCTGAACGCCTAGCCTGTGTA	G	winter	C	Vrn-D1a	CerealdB	Fu et al (2005)
GAGAGATTATGTTATATTCGTCCAACGGTT	C	susceptible	G	resistant	Jayatilake et al (2014)	Jayatilake et al (2014)
CGAAGAAGCTTGGCCTGGATAGTAT	G	Glu-Ax1 or Ax2*	A	Ax-null	CIMMYT-SD	Liu et al (2008)
GGTTTCATACTATCCAGGCCAAGCTT	INS	Glu-Ax1 or Ax-null	DEL	Glu-Ax2*	CIMMYT-SD	Liu et al (2008)
cttccccacagctcgtggagaa	C	early flower (Claire type)	T	late flower (Hereward type)	Diaz et al (2012)	Diaz et al (2012)
GGTTTGCAATCTTACAAYGACAAGGTAA	T	Rlnn1-	A	Rlnn1+	CIMMYT-SD	Sharp et al. (2001
AACAAACTCCAGTGTAAACACCACAGTTT	T	Lr14a-	A	Lr14a+	Teracciano et al (2013)	Teracciano et al (2013)
GCCTGGATTCAAGAGAACAT	G	Lr47-	C	Lr47+	CIMMYT-SD	Huelguera et al (2000)
TGACCTGAGTCCCGTCAAGA	T	Lr68+	C	Lr68-	CIMMYT-SD	Herrera-Fossel et al (2012)
TTTCACCTTGTGATATGGATTGCCTTGAT	A	TaDREB-B1a	C	TaDREB-B1b	CIMMYT-SD	Wei et al (2009)
CCAGTAGCTTATGGCACTCAAA		Susceptible		Resistent	Forrest et al 2014	Forrest et al (2014)
agatatcacactgaacatactgatgaG	A	Susceptible	G	Resistent	Ramirez-Gonzalez et al 2014	Ramirez-Gonzalez et al (2014)
cccccaaatgatcgagaata	C	Resistent	T	Susceptible	Ramirez-Gonzalez et al 2014	Ramirez-Gonzalez et al (2014)

Appendix 3. Examples STS and KASP markers

3.1 Examples of amplification products of gene-specific STS/SSR marker

SSR marker linked to *Lr16/Sr23*

Marker: *gwm210*

Marker Type: SSR

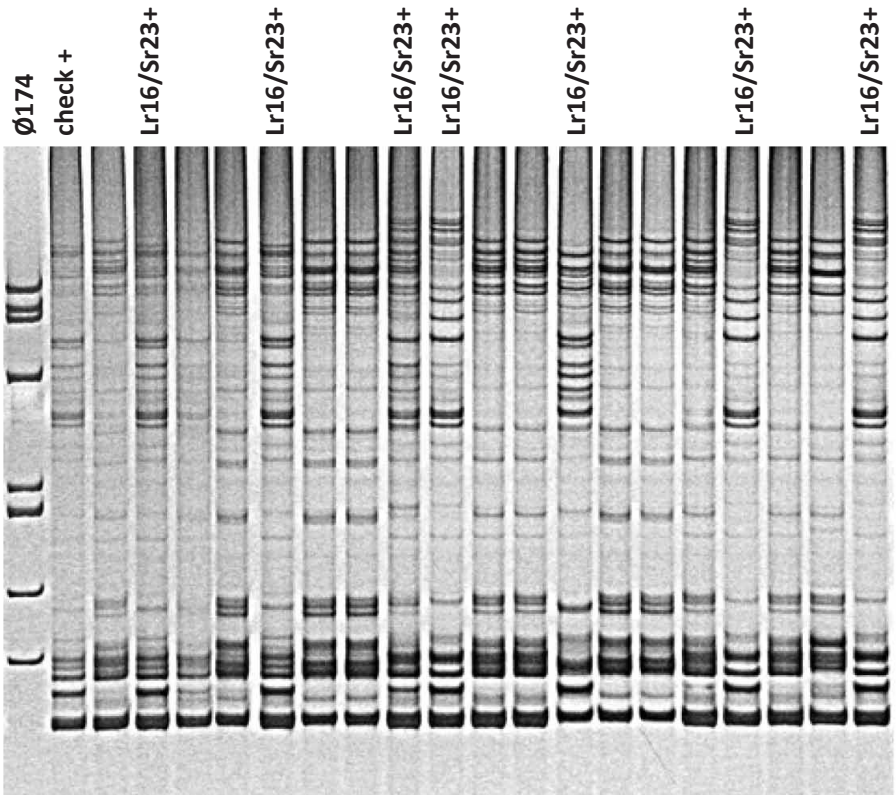
Gene: *Lr16/Sr23*

Product: *Lr16/Sr23*+ =~170bp, many bands

Inheritance: dominant

Medium: Acrylamide

Reference: McCartney et al. 2005



STS marker linked to *Sr26*

Marker: *Sr26#43 + BE518379*

Marker Type: STS

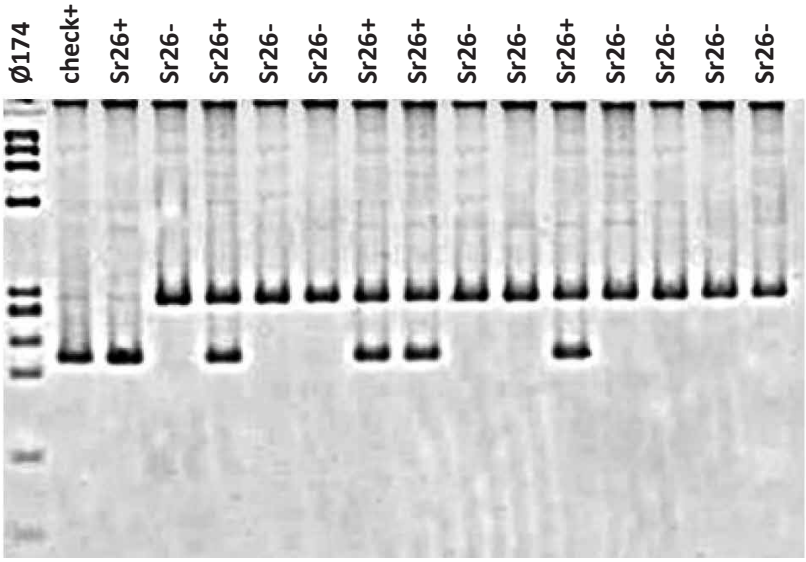
Locus/Gene: *Sr26*

Product: *Sr26+: 207* bp
Sr26- : 303 bp

Inheritance: co-dominant

Medium: Agarose

Reference: Mago et al. 2005, Liu et al. 2009



SSR marker linked to *Lr14a* in durum wheat

Marker: *wms146*

Marker Type: SSR

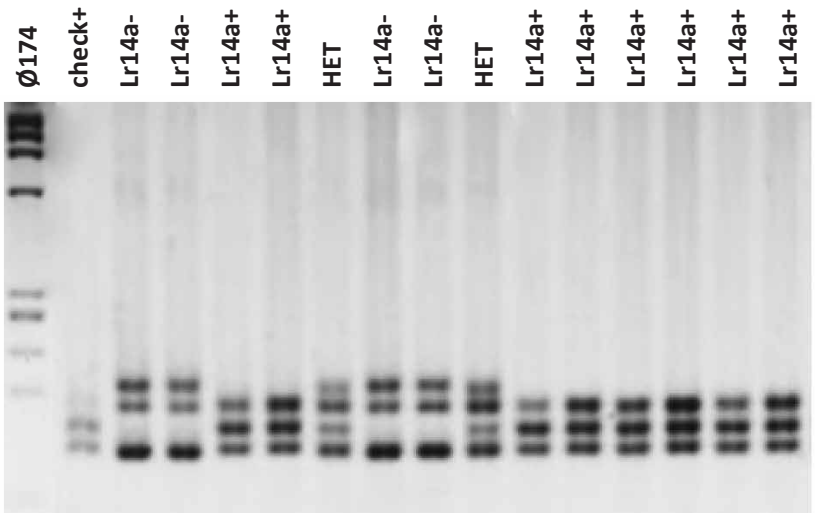
Locus/Gene: *Lr14a*

Product: *Lr14a+ = ~190/210* bp

Inheritance: co-dominant

Medium: Agarose

Reference: Herrera-Foessil et al. 2008



STS marker linked to *Waxy-B1*

Marker: *GBSS*

Marker Type: GBSS

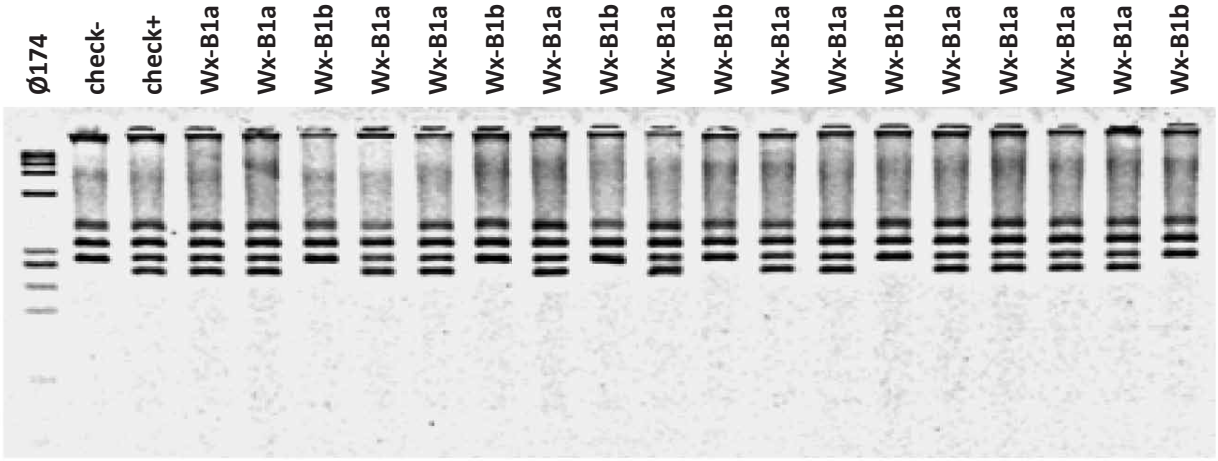
Locus/Gene: *Wx-B1*

Product: *Wx-B1a (wildtype) = 320* bp
Wx-B1b (null) = no band

Inheritance: dominant

Medium: Agarose

Reference: McLauchlan et al. 2001



STS marker linked to *Glu-B3d*

Marker: *SB4*

Marker Type: STS

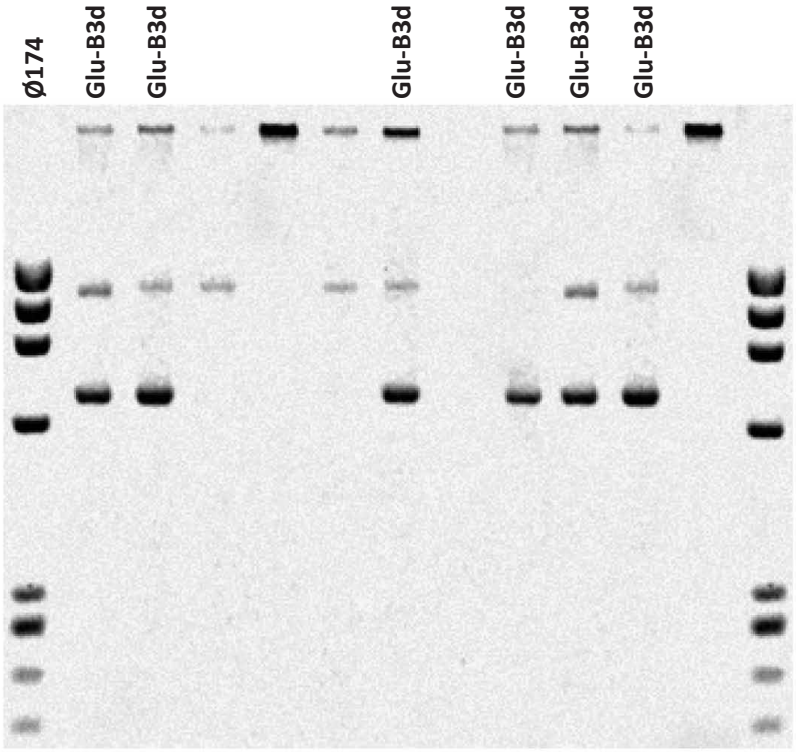
Locus/Gene: *GluB3*

Product: *: Glu-B3d = 662* bp

Inheritance: dominant

Medium: Agarose

Reference: Wang et al. 2009



STS marker linked to *Vrn-B1*

Marker: *Intro1/B/F + Intro1/B/R3 + Intro1/B/R4 + Ex1/B/F3*

Marker Type: STS

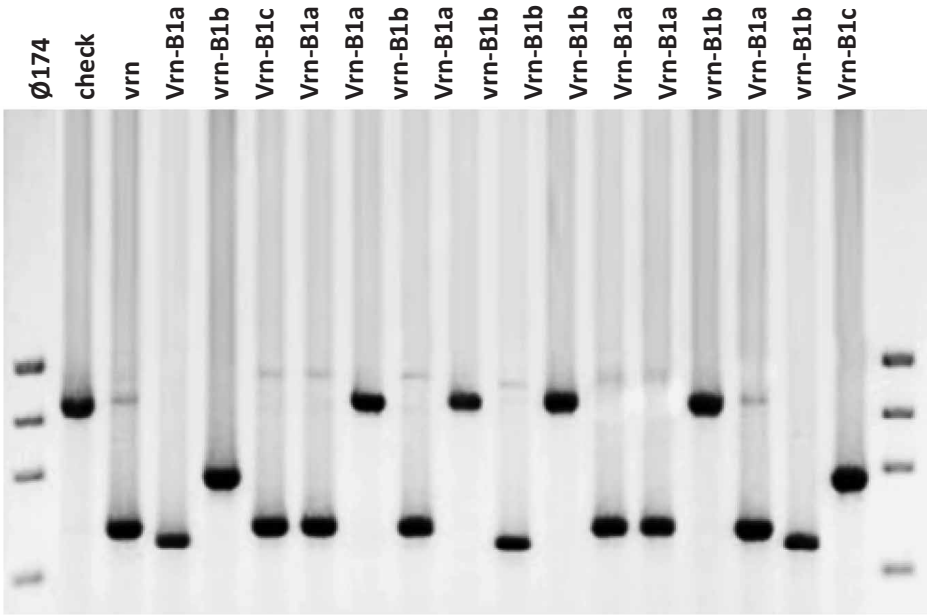
Locus/Gene: *Vrn-B1*

Product: *Vrn-B1a = 709* bp,
Vrn-B1 = 1149 bp,
Vrn-B1b = 673 bp,
Vrn-B1c = 849 bp

Inheritance: co-dominant

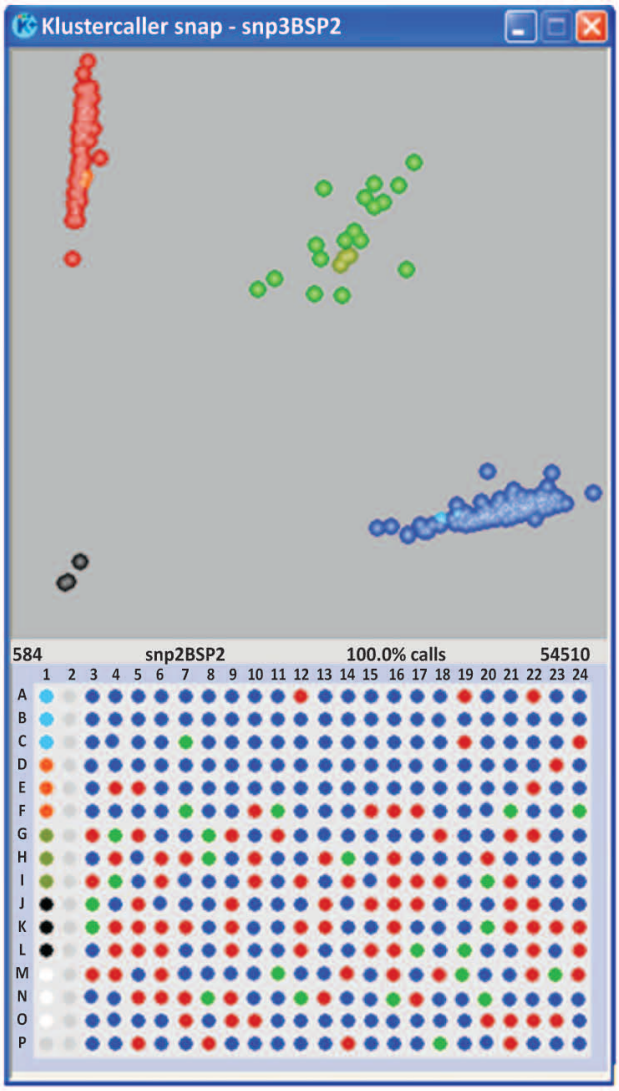
Medium: Agarose

Reference: Milec et al. 2012



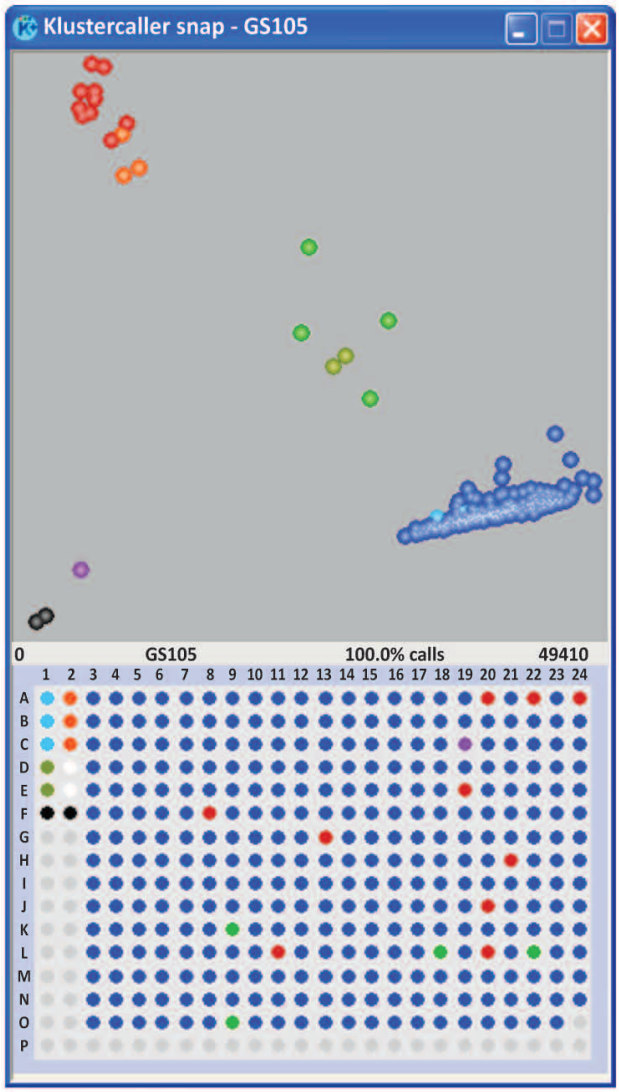
3.2 Examples of SNP visualization in Klustercaller

Trait-based KASP assay for *Fhb1*
Trait: Fusarium (FHB)
Gen: *Fhb1*
Name SNP: *snp3BS-8* (see Appendix 2)
Material: F_7 bread wheat, 384-well plate
PCR Program: snp (td) (see KASP assay protocol)
Allele FAM (blue): Susceptible to FHB
Allele VIC (red): Resistant to Fusarium + Sumai3 type
Allele Heterozygote (green)
Reference: Bernardo et al. (2011)



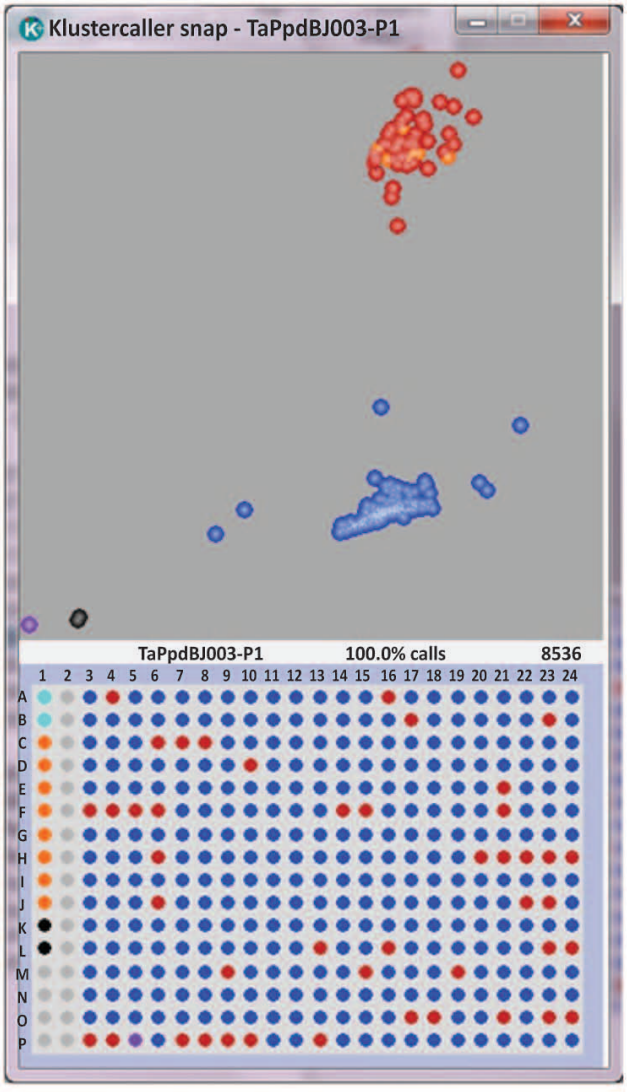
snp3BS-8 is a SNP used to detect the Fusarium head blight gene *Fhb1*, here in F_7 population of bread wheat. In the scatterplot lines in blue lack *Fhb1*. Lines colored in red have *Fhb1*. Lines colored in green are heterozygotes, and black dots represent NTC (non-template control).

Trait-based KASP assay for the gene *Ppd-A1*
Trait: Photoperiod
Gen: *Ppd-A1*
Name SNP: *GS105-1117ID* (see Appendix 2)
Material: Bread wheat core collection
PCR Program: snp (20-24) (see KASP assay protocol)
Allele FAM (blue): Sensitive to Photoperiod
Allele VIC (red): Insensitive to Photoperiod
Allele Heterozygote (green)
Reference: Beales et al. (2007)



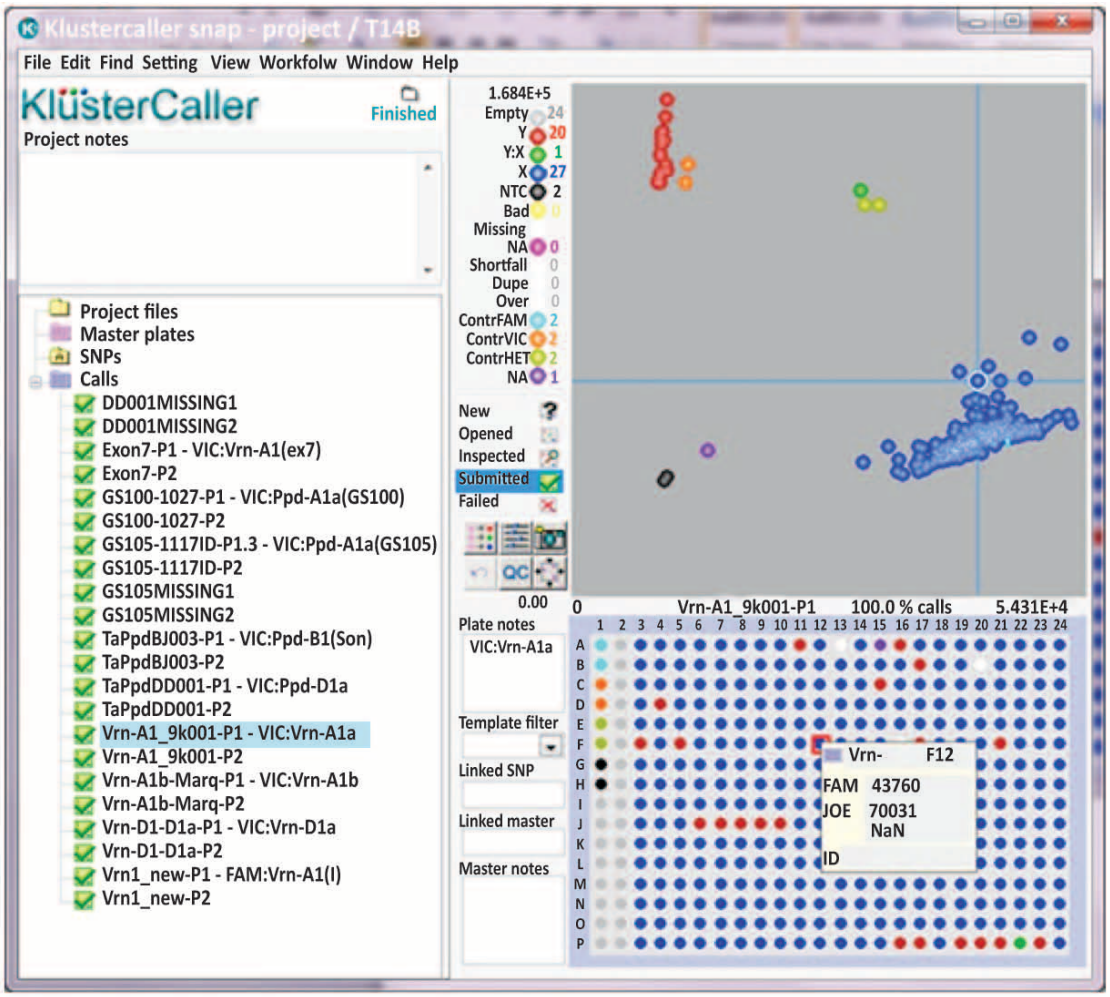
GS105-1117ID is a KASP assay used in a bread wheat core collection to detect a large deletion in the gene *Ppd-A1*. In the scatterplot lines colored in blue have the sensitive photoperiod allele, lines colored in red have the insensitive photoperiod allele, lines colored green are heterozygotes, lines in colored purple did not amplify and black dots represent NTC (non-template control).

Trait-based KASP assay for the gene *Ppd-B1*
Trait: Photoperiod
Gen: *Ppd-B1*
Name SNP: *TaPpdBJ003* (see Appendix 2)
Material: Bread wheat core collection
PCR Program: snp (td) (see KASP assay protocol)
Allele FAM (blue): Sensitive to Photoperiod
Allele VIC (red): Insensitive to Photoperiod
Allele Heterozygote (green)
Reference: Beales et al. (2007)



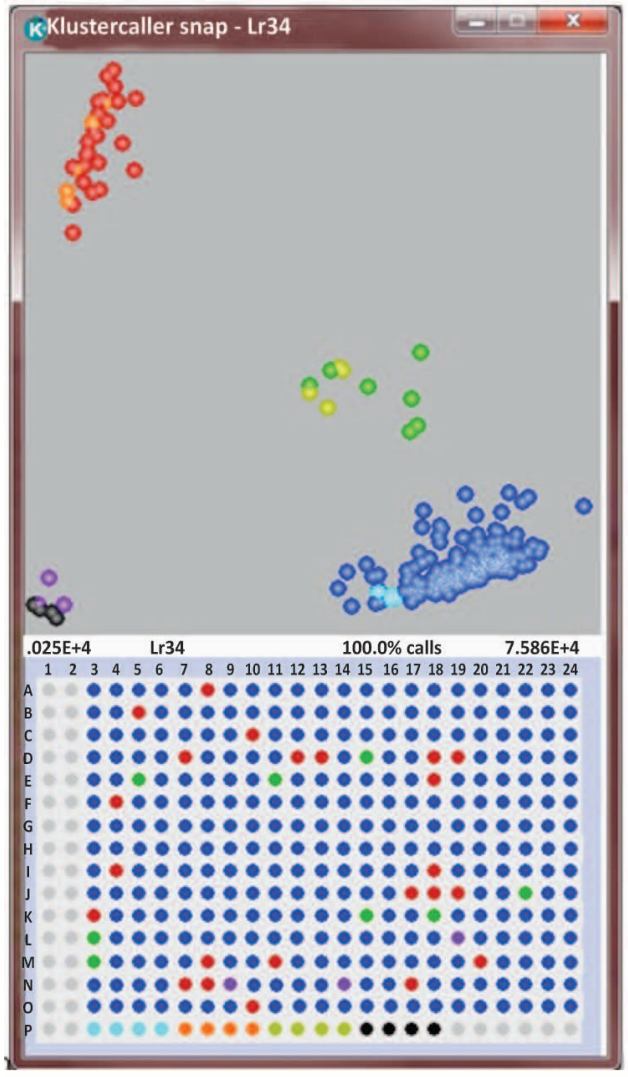
TaPpdBJ003 is a dominant KASP assay used here in a bread wheat core collection. In the scatterplot lines colored in blue have the sensitive photoperiod allele, lines colored red have the insensitive photoperiod allele, lines colored in purple did not amplify and black dots represent NTC (non-template control).

Trait-based KASP assay for the gene *Vrn-A1*
Trait: Vernalization
Gen: *Vrn-A1*
Name SNP: *Vrn-A1_9K0001* (see Appendix 2)
Material: Bread wheat core collection
PCR Program: snp (td) (see KASP assay protocol)
Allele FAM (blue): others
Allele VIC (red): *Vrn-A1a*
Allele Heterozygote (green)



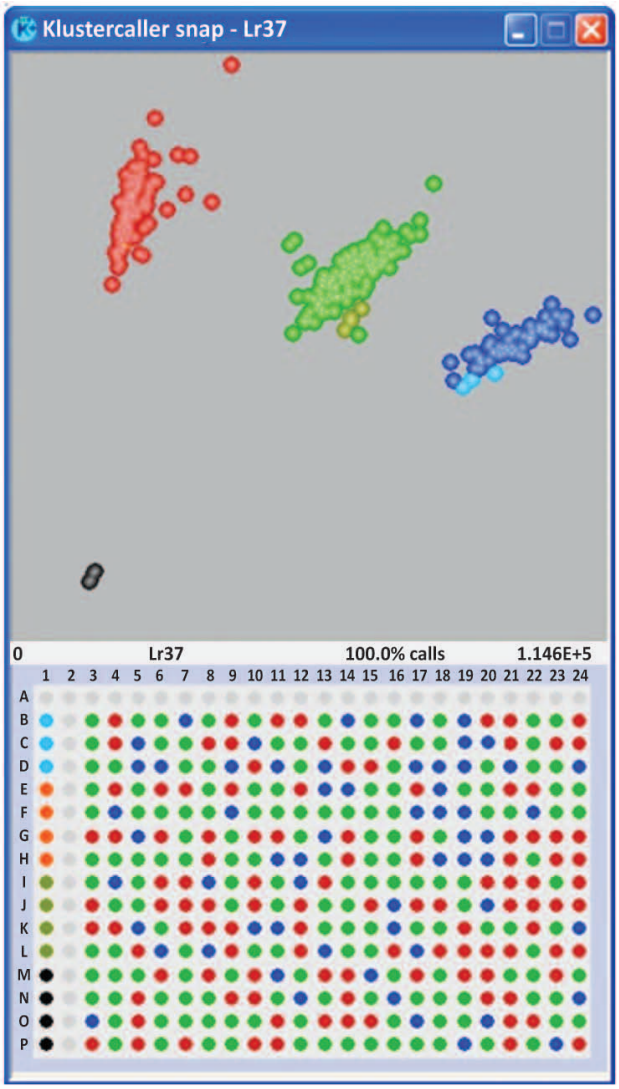
Vrn-A1_9K0001 is a KASP assay used bread wheat core collection. In the scatterplot lines colored in red have the *Vrn-A1a* spring allele, lines colored in blue have the *vrn-A1* winter allele or an alternative spring allele, lines colored in green are heterozygotes, lines colored in purple did not amplify and black dots represent NTC (non-template control). The blue cross indicates a selected line (in well F12) in the plate.

Trait-based KASP assay for the gene *Lr34*
Trait: Rust
Gen: *Lr34*
Name SNP: *Lr34_TCCIND* (see Appendix 2)
Material: Bread wheat core collection
PCR Program: snp (td) (see KASP assay protocol)
Allele FAM (blue): Susceptible
Allele VIC (red): Resistant
Allele Heterozygote (green)



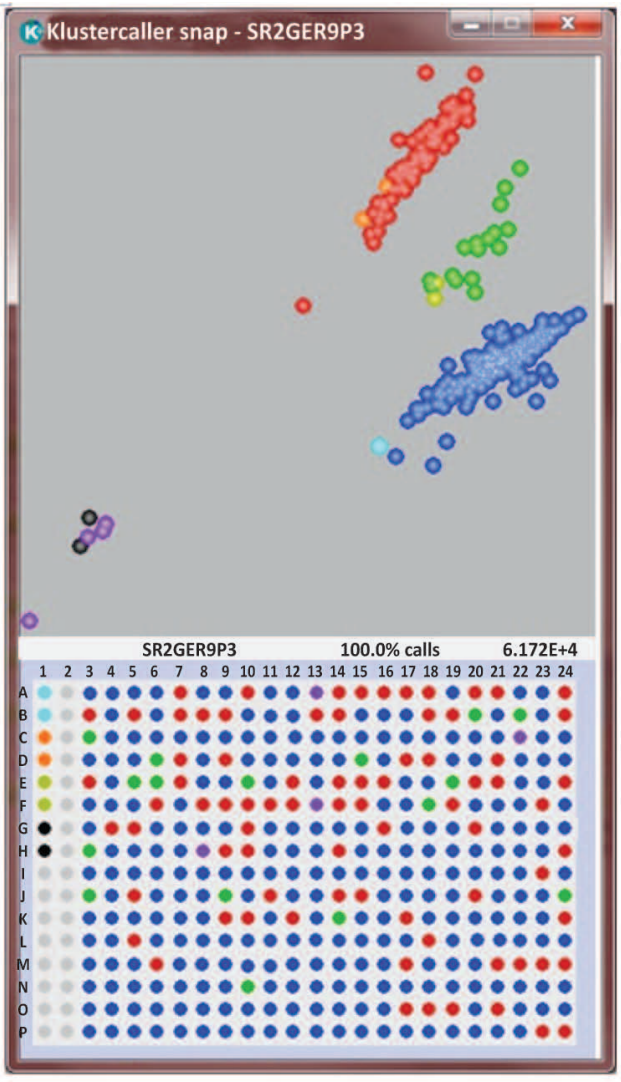
Lr34_TCCIND is a KASP assay used here ina bread wheat core collection. In the scatterplot lines colored in red carry the Lr34 resistance allele, lines colored in blue carry the susceptible allele, lines colored in green are heterozygotes, lines colored in purple did not amplify and black dots represent NTC (non-template control).

Trait-based KASP assay for the gene *Lr37*
Trait: Rust
Gen: *Lr37/Yr17/Sr38*
Name SNP: *VPM_SNP* (see Appendix 2)
Material: Bread wheat core collection
PCR Program: snp (td td) (see KASP assay protocol)
Allele FAM (blue): Resistant
Allele VIC (red): Susceptible
Allele Heterozygote (green)



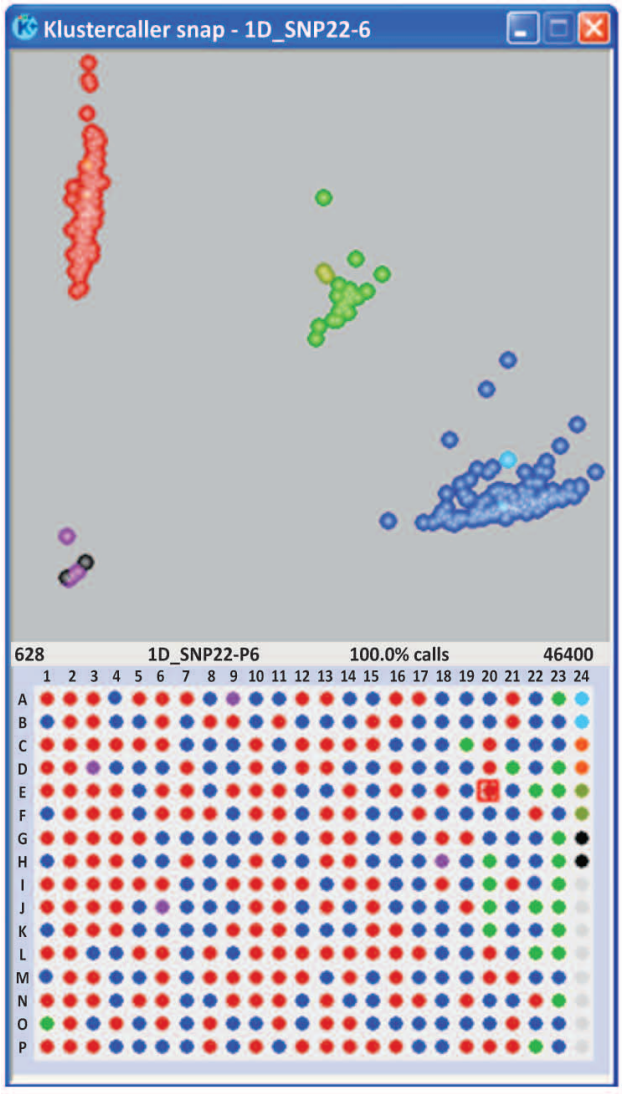
VPM_SNP is a KASP assay used here in a bread wheat core collection. In the scatterplot lines colored in red do not carry the VPM translocation, lines colored in blue carry the VPM translocation, lines colored in green are heterozygotes, lines colored in purple did not amplify and black dots represent NTC (non-template control).

Trait-based KASP assay for the gene *Sr2*
Trait: Rust
Gen: *Sr2*
Name SNP: *Sr2_ger9 3p* (see Appendix 2)
Material: Bread wheat core collection
PCR Program: snp (td) (see KASP assay protocol)
Allele FAM (blue): Susceptible to Sr2
Allele VIC (red): Resistance to Sr2 (Hope type)
Allele Heterozygote (green)



Sr2_ger9 3p is a KASP assay used here in a bread wheat core collection. In the scatterplot lines colored in red carry the Sr2+ allele, lines colored in blue carry the Sr2- allele, lines colored in green are heterozygotes, lines colored in purple are nulls and black dots represent NTC (non-template control).

Example of a neutral KASP assay in chromosome 1D
SNP ID: *IWA3753* (see Appendix 5)
Name SNP: *wsnp_Ex_c41048_47969948*
Material: Turkish Landraces
PCR Program: snp (td) (see KASP assay protocol)
Allele FAM (blue): A
Allele VIC (red): G
Allele Heterozygote (green)



wsnp_Ex_c41048_47969948 is a KASP assay developed from the 90K iselect Illumina What SNP chip used in this example on a set of Turkish landrace material. In the scatterplot lines colored in blue have the FAM- A allele, lines colored in red have the HEX- G allele, lines colored in green are heterozygotes, lines colored in purple did not amplify and black dots represent NTC (non-template control).

Appendix 4: Set of across the genome distributed and highly polymorphic SSR markers

Nr	Primer Name	Chr	Primer Prefix	Primer Sequence	Primer Prefix	Primer Sequence
1	cfa2129	1A	F (Forward)	5′ GTTGCACGACCTACAAAGCA 3′	R (Reverse)	5′ ATCGCTCACTCACTATCGGG 3′
2	cfa2219	1A	F	5′ TCTGCCGAGTCACTTCATTG 3′	R	5′ GACAAGGCCAGTCCAAAAGA 3′
3	cfd15	1A	F	5′ CTCCCGTATTGAGCAGGAAG 3′	R	5′ GGCAGGTGTGGTGATGATCT 3′
4	gwm135	1A	F	5′ ACACTGTCAACCTGGCAATG 3′	R	5′ TGTCAACATCGTTTTGAAAAGG 3′
5	gwm164	1A	F	5′ ACATTTCTCCCCCATCGTC 3′	R	5′ TTGTAAACAAATCGCATGCG 3′
6	gwm357	1A	F	5′ AGGCTGCAGCTCTTCTTCAG 3′	R	5′ TATGGTCAAAGTTGGACCTCG 3′
7	gwm497	1A	F	5′ CCGAAAGTTGGGTGATATAC 3′	R	5′ GTAGTGAAGACAAGGGCATT 3′
8	wmc24	1A	F	5′ GTGAGCAATTTTGATTATACTG 3′	R	5′ TACCCTGATGCTGTAATATGTG 3′
9	wmc59	1A	F	5′ TCATTCGTTGCAGATACACCAC 3′	R	5′ TCAATGCCCTTGTTTCTGACCT 3′
10	wmc183	1A	F	5′ CAGAAACGGCTCAACTTAACAA 3′	R	5′ TCTGATCTCGTGATCAGAATAG 3′
11	wmc278	1A	F	5′ AAACGATAGTAAAATTACCTCGGAT 3′	R	5′ TCAAAAAATAGCAACTTGAAGACAT 3′
12	wmc312	1A	F	5′ TGTGCCCGCTGGTGCGAAG 3′	R	5′ CCGACGCAGGTGAGCGAAG 3′
13	wmc336	1A	F	5′ GTCTTACCCCGCGATCTGC 3′	R	5′ GCGGCCTGAGCTTCTTGAG 3′
14	wmc469	1A	F	5′ AGGtGGCtGCCAACG 3′	R	5′ CAAttttAtCAGAtGCCCGA 3′
15	wmc716	1A	F	5′ CATTTATGTGCACGCCGAAG 3′	R	5′ CCATAAGCATCGTCACCCTG 3′
16	wmc818	1A	F	5′ TGAAGGGTGCGTGTGGTC 3′	R	5′ GCGTCGATTTTAATTTGATGATGG 3′
17	barc17	1AL	F	5′ GCGCAACATATTCAGCTCAACA 3′	R	5′ TCCACATCTCGTCCCTCATAGTTTG 3′
18	barc83	1AL	F	5′ AAGCAAGGAACGAGCAAGAGCAGTAG 3′	R	5′ TGGATTTACGACGACGATGAAGATGA 3′
19	barc158	1AL	F	5′ TGTGTGGGAAGAAACTGAGTCATC 3′	R	5′ AGGAATACCAAAAGAAGCAAACCAAC 3′
20	barc148	1AS	F	5′ GCGCAACCACAATGTATGCT 3′	R	5′ GGGGTGTTTTCCTATTTCTT 3′
21	cfd20	1B	F	5′ TGATGGGAAGGTAATGGGAG 3′	R	5′ ATCCAGTTCTCGTCCAAAGC 3′
22	cfd48	1B	F	5′ ATGGTTGATGGTGGGTGTTT 3′	R	5′ ATGTATCGATGAAGGGCCAA 3′
23	gwm11	1B	F	5′ GGATAGTCAGACAATTCTTGTG 3′	R	5′ GTGAATTGTGTCTTGTATGCTTCC 3′
24	gwm124	1B	F	5′ ACTGTTCGGTGCAATTTGAG 3′	R	5′ GCCATGGCTATCACCCAG 3′
25	gwm264	1B	F	5′ GAGAAACATGCCGAACAACA 3′	R	5′ GCATGCATGAGAATAGGAACTG 3′
26	gwm273	1B	F	5′ AGCAGTGAGGAAGGGGATC 3′	R	5′ ATTGGACGGACAGATGCTTT 3′
27	gwm274	1B	F	5′ AACTTGCAAAACTGTTCTGA 3′	R	5′ TATTTGAAGCGGTTTGATTT 3′
28	gwm413	1B	F	5′ GATCGTCTCGTCCTTGGCA 3′	R	5′ TGCTTGTCTAGATTGCTTGGG 3′
29	gwm498	1B	F	5′ GGTGGTATGGACTATGGACACT 3′	R	5′ TTTGCATGGAGGCACATACT 3′
30	wmc44	1B	F	5′ GGTCTTCTGGGCTTTGATCCTG 3′	R	5′ TGTTGCTAGGGACCCGTAGTGG 3′
31	wmc128	1B	F	5′ CGGACAGCTACTGCTCTCCTTA 3′	R	5′ CTGTTGCTTGCTCTGCACCCTT 3′
32	wmc134	1B	F	5′ CCAAGCTGTCTGACTGCCATAG 3′	R	5′ AGTATAGACCTCTGGCTCACGG 3′
33	wmc216	1B	F	5′ ACGTATCCAGACACTGTGGTAA 3′	R	5′ TAATGGTGGATCCATGATAGCC 3′
34	wmc367	1B	F	5′ CTGACGTTGATGGGCCACTATT 3′	R	5′ GTGGTGGAAGAGGAAGGAGAGG 3′
35	wmc416	1B	F	5′ AGCCCTTTCTACCGTGTTTCTT 3′	R	5′ TATGGTCGATGGACTGTCCCTA 3′
36	wmc419	1B	F	5′ GTTTCGGATAAAACCGGAGTGC 3′	R	5′ ACTACTTGTGGGTTATCACCAGCC 3′
37	wmc626	1B	F	5′ AGCCCATAAACATCCAACACGG 3′	R	5′ AGGTGGGCTTGGTTACGCTCTC 3′
38	wmc694	1B	F	5′ ATTTGCCCTTGTGAGCCGTT 3′	R	5′ GACCTGGGTGGGACCCATTA 3′
39	wmc830	1B	F	5′ ACCTTTTCCTGCATCGGCT 3′	R	5′ CTCCGCTCGTGTCCAACTATC 3′
40	barc80	1BL	F	5′ GCGAATTAGCATCTGCATCTGTTTGAG 3′	R	5′ CGGTCAACCAACTACTGCACAAC 3′
41	barc137	1BL	F	5′ GGCCCATTTCCCACTTTCCA 3′	R	5′ CCAGCCCCTCTACACATTTT 3′
42	barc174	1BL	F	5′ TGGCATTTTTCTAGCACCAATACAT 3′	R	5′ GCGAACTGGACCAGCCTTCTATCTGTTC 3′
43	barc181	1BL	F	5′ CGCTGGAGGGGGTAAGTCATCAC 3′	R	5′ CGCAAATCAAGAACACGGGAGAAAGAA 3′
44	barc8	1BS	F	5′ GCGGGAATCATGCATAGGAAAACAGAA 3′	R	5′ GCGGGGGCGAAACATACACATAAAAACA 3′
45	GDM33	1D	F	5′ GGCTCAATTCAACCGTTCTT 3′	R	5′ TACGTTCTGGTGGCTGCTC 3′
46	GDM126	1D	F	5′ TCCATCATATCCGTAGCACA 3′	R	5′ CGTGGTTGATTTCAGGAGGT 3′
47	cfa2147	1D	F	5′ TCATCCCCTACATAACCCGA 3′	R	5′ ATCGTGCACCAAGCAATACA 3′
48	cfd59	1D	F	5′ TCACCTGGAAAATGGTCACA 3′	R	5′ AAGAAGGCTAGGGTTCAGGC 3′
49	cfd61	1D	F	5′ ATTCAAATGCAACGCAAACA 3′	R	5′ GTTAGCCAAGGACCCCTTTC 3′
50	cfd63	1D	F	5′ TCCTGAGGATGTTGAGGACC 3′	R	5′ GAGAGAGGCGAAACATGGAC 3′
51	cfd65	1D	F	5′ AGACGATGAGAAGGAAGCCA 3′	R	5′ CCTCCCTTGTTTTTGGGATT 3′
52	cfd72	1D	F	5′ CTCCTTGGAATCTCACCGAA 3′	R	5′ TCCTTGGGAATATGCCTCCT 3′
53	cfd92	1D	F	5′ CTTGTTGATCTCCTTCCCCA 3′	R	5′ TTCTCTCATGACGGCAACAC 3′

Nr	Primer Name	Chr	Primer Prefix	Primer Sequence	Primer Prefix	Primer Sequence
54	gwm232	1D	F	5′ ATCTCAACGGCAAGCCG 3′	R	5′ CTGATGCAAGCAATCCACC 3′
55	gwm337	1D	F	5′ CCTCTTCCTCCCTCACTTAGC 3′	R	5′ TGCTAACTGGCCTTTGCC 3′
56	gwm642	1D	F	5′ ACGGCGAGAAGGTGCTC 3′	R	5′ CATGAAAGGCAAGTTCGTCA 3′
57	wmc339	1D	F	5′ CCGCTCGCCTTCTTCCAG 3′	R	5′ TCCGGAACATGCCGATAC 3′
58	wmc429	1D	F	5′ CGTAAAGATTTTCATTTGGCG 3′	R	5′ AACGGCAGCTTGAAAACATAG 3′
59	wmc432	1D	F	5′ ATgAcAccAgATcTAgcAc 3′	R	5′ AATATTggcATgATTAcAcA 3′
60	barc62	1DL	F	5′ TTGCCTGAGACATACATACACCTAA 3′	R	5′ GCCAGAACAGAATGAGTGCT 3′
61	barc66	1DL	F	5′ CGCGATCGATCTCCCGGTTTGCT 3′	R	5′ GGGAAGAGGACCAAGGCCACTA 3′
62	barc119	1DL	F	5′ CACCCGATGATGAAAAT 3′	R	5′ GATGGCACAAGAAATGAT 3′
63	barc149	1DS	F	5′ ATTCACTTGCCCCTTTTAAACTCT 3′	R	5′ GAGCCGTAGGAAGGACATCTAGTG 3′
64	barc5	2A	F	5′ GCGCCTGGACCGGTTTTCTATTTT 3′	R	5′ GCGTTGGGAATTCCTGAACATTTT 3′
65	cfd2	2A	F	5′ GGTTGCAGTTTCCACCTTGT 3′	R	5′ CATCTATTGCCAAAATCGCA 3′
66	gwm95	2A	F	5′ AATGCAAAGTGAAAAACCCG 3′	R	5′ GATCAAACACACACCCCTCC 3′
67	gwm294	2A	F	5′ GCAGAGTGATCAATGCCAGA 3′	R	5′ GGATTGGAGTTAAGAGAGAACCG 3′
68	gwm312	2A	F	5′ ACATGCATGCCTACCTAATGG 3′	R	5′ ATCGCATGATGCACGTAGAG 3′
69	gwm372	2A	F	5′ AATAGAGCCCTGGGACTGGG 3′	R	5′ GAAGGACGACATTCCACCTG 3′
70	gwm558	2A	F	5′ GGGATTGCATATGAGACAACG 3′	R	5′ TGCCATGGTTGTAGTAGCCA 3′
71	wmc177	2A	F	5′ AGGGCTCTCTTTAATTCTTGCT 3′	R	5′ GGTCTATCGTAATCCACCTGTA 3′
72	wmc181	2A	F	5′ TCCTTGACCCCTTGCACTAACT 3′	R	5′ ATGGTTGGGAGCACTAGCTTGG 3′
73	wmc296	2A	F	5′ GAATCTCATCTTCCCTTGCCAC 3′	R	5′ ATGGAGGGGTATAAAGACAGCG 3′
74	wmc382	2A	F	5′ CATGAATGGAGGCACTGAAACA 3′	R	5′ CCTTCCGGTCGACGCAAC 3′
75	wmc407	2A	F	5′ GGTAATTCTAGGCTGACATATGCTC 3′	R	5′ CATATTTCCAAATCCCCAACTC 3′
76	wmc522	2A	F	5′ AAAAATCTCACGAGTCGGGC 3′	R	5′ CCCGAGCAGGAGCTACAAAT 3′
77	wmc658	2A	F	5′ CTCATCGTCCTCCTCCACTTTG 3′	R	5′ GCCATCCGTTGACTTGAGGTTA 3′
78	wmc667	2A	F	5′ GAGGAGAGGAAAAGGCAGGCTA 3′	R	5′ AACTCTTGCGTGTCTCAAACCG 3′
79	barc98	2B	F	5′ CCGTCCTATTCGCAAACCAGATT 3′	R	5′ GCGGATATGTTCTCTAACTCAAGCAATG 3′
80	barc167	2B	F	5′ AAAGGCCCATCAACATGCAAGTACC 3′	R	5′ CGCAGTATTCTTAGTCCCTCAT 3′
81	cfd73	2B	F	5′ GATAGATCAATGTGGGCCGT 3′	R	5′ AACTGTTCTGCCATCTGAGC 3′
82	gwm148	2B	F	5′ CAAAGCTTGACTCAGACCAAA 3′	R	5′ GTGAGGCAGCAAGAGAGAAA 3′
83	gwm257	2B	F	5′ AGAGTGCATGGTGGGACG 3′	R	5′ CCAAGACGATGCTGAAGTCA 3′
84	gwm388	2B	F	5′ CACCGCGTCAACTACTTAAGC 3′	R	5′ CTACAATTCGAAGGAGAGGGG 3′
85	gwm429	2B	F	5′ TTGTACATTAAGTTCCCATTA 3′	R	5′ TTTAAGGACCTACATGACAC 3′
86	gwm630	2B	F	5′ CGAAAGTAACAGCGCAGTGA 3′	R	5′ GTGCCTGTGCCATCGTC 3′
87	wmc25	2B	F	5′ TCTGGCCAGGATCAATATTACT 3′	R	5′ TAAGATACATAGATCCAACACC 3′
88	wmc149	2B	F	5′ ACAGACTTGGTTGGTGCCGAGC 3′	R	5′ ATGGGCGGGGGTGTAGAGTTTG 3′
89	wmc154	2B	F	5′ ATGCTCGTCAGTGTCATGTTTG 3′	R	5′ AAACGGAACCTACCTCACTCTT 3′
90	wmc332	2B	F	5′ CATTTACAAAGCGCATGAAGCC 3′	R	5′ GAAAACTTTGGGAACAAGAGCA 3′
91	wmc477	2B	F	5′ CGTCGAAAACCGTACACTCTCC 3′	R	5′ GCGAAACAGAATAGCCCTGATG 3′
92	wmc764	2B	F	5′ CCTCGAACCTGAAGCTCTGA 3′	R	5′ TTCGCAAGGACTCCGTAACA 3′
93	barc101	2BL	F	5′ GCTCCTCTCACGATCACGCAAAG 3′	R	5′ GCGAGTCGATCACACTATGAGCCAATG 3′
94	barc13	2BS	F	5′ GCAGGAACAACCACGCCATCTTAC 3′	R	5′ GCGTCGCAATTTGAAGAAAATCATC 3′
95	barc18	2BS	F	5′ CGCTTCCCATAACGCCGATAGTAA 3′	R	5′ CGCCCGCATCATGAGCAATTCTATCC 3′
96	cfa2262	2D	F	5′ ACAATGTGGAGATGGCACAA 3′	R	5′ TACCAGCTGCACTTCCATTG 3′
97	cfd36	2D	F	5′ GCAAAGTGTAGCCGAGGAAG 3′	R	5′ TTAGAGTTTTGCAGCGCCTT 3′
98	cfd43	2D	F	5′ AACAAAAGTCGGTGCAGTCC 3′	R	5′ CCAAAAACATGGTTAAAGGGG 3′
99	cfd56	2D	F	5′ TTGCATAATTACTTGCCCTCC 3′	R	5′ CTGGTCCAACTTCCATCCAT 3′
100	cfd116	2D	F	5′ TTTGCCCATTACAACAAGCA 3′	R	5′ CAAGCAGCACCTCATGACAG 3′
101	cfd168	2D	F	5′ CTTCGCAAATCGAGGATGAT 3′	R	5′ TTCACGCCCAGTATTAAGGC 3′
102	cfd175	2D	F	5′ TGTCGGGGACACTCTCTCTT 3′	R	5′ ACCAATGGGATGCTTCTTTG 3′
103	gwm102	2D	F	5′ TCTCCCATCCAACGCCTC 3′	R	5′ TGT TGG TGG CTT GAC TAT TG 3′
104	gwm157	2D	F	5′ GAGTGAACACACGAGGCTTG 3′	R	5′ GTCGTCGCGGTAAGCTTG 3′
105	gwm301	2D	F	5′ GAGGAGTAAGACACATGCCC 3′	R	5′ GTGGCTGGAGATTCAGGTTC 3′
106	gwm311	2D	F	5′ CTACGTGCACCACCATTTTG 3′	R	5′ TCACGTGGAAGACGCTCC 3′
107	gwm484	2D	F	5′ ACATCGCTCTTCACAAACCC 3′	R	5′ AGTTCCGGTCATGGCTAGG 3′
108	gwm539	2D	F	5′ CTGCTCTAAGATTCATGCAACC 3′	R	5′ GAGGCTTGTGCCCTCTGTAG 3′
109	wmc18	2D	F	5′ CTGGGGCTTGGATCACGTCATT 3′	R	5′ AGCCATGGACATGGTGTCCTTC 3′
110	wmc112	2D	F	5′ TGAGTTGTGGGGTCTTGTTTGG 3′	R	5′ TGAAGGAGGGCACATATCGTTG 3′

Nr	Primer Name	Chr	Primer Prefix	Primer Sequence	Primer Prefix	Primer Sequence
111	wmc144	2D	F	5′ GGACACCAATCCAACATGAACA 3′	R	5′ AAGGATAGTTGGGTGGTGCTGA 3′
112	wmc167	2D	F	5′ AGTGGTAATGAGGTGAAAGAAG 3′	R	5′ TCGGTCGTATATGCATGTAAAG 3′
113	wmc503	2D	F	5′ GCAATAGTTCCCGCAAGAAAAG 3′	R	5′ ATCAACTACCTCCAGATCCCGT 3′
114	wmc601	2D	F	5′ ACAGAGGCATATGCAAAGGAGG 3′	R	5′ CTTGTCTCTTTATCGAGGGTGG 3′
115	barc159	2DL	F	5′ CGCAATTTATTATCGGTTTTAGGAA 3′	R	5′ CGCCCGATAGTTTTTCTAATTTCTGA 3′
116	barc228	2DL	F	5′ CCCTCCTCTCTTTAGCCATCC 3′	R	5′ GCACGTACTATTCGCCTTCACTTA 3′
117	barc124	2DS	F	5′ TGCACCCCTTCCAAATCT 3′	R	5′ TGCGAGTCGTGTGGTTGT 3′
118	barc54	3A	F	5′ GCGAACAGGAGGACAGAGGGCACGAGAG 3′	R	5′ GCGCTTTCCCACGTTCCATGTTTCT 3′
119	cfa2193	3A	F	5′ ACATGTGATGTGCGGTCATT 3′	R	5′ TCCTCAGAACCCCATTCTTG 3′
120	gwm5	3A	F	5′ AGAAAGGGCCAGGCTAGTAGT 3′	R	5′ GCCAGCTACCTCGATACAACTC 3′
121	gwm369	3A	F	5′ ACCGTGGGTGTTGTGAGC 3′	R	5′ CTGCAGGCCATGATGATG 3′
122	gwm666	3A	F	5′ GCACCCACATCTTCGACC 3′	R	5′ TGCTGCTGGTCTCTGTGC 3′
123	wmc11	3A	F	5′ TTGTGATCCTGGTTGTGTTGTGA 3′	R	5′ CACCCAGCCGTTATATATGTTGA 3′
124	wmc153	3A	F	5′ ATGAGGACTCGAAGCTTGGC 3′	R	5′ CTGAGCTTTTGCGCGTTGAC 3′
125	wmc264	3A	F	5′ CTCCATCTATTGAGCGAAGGTT 3′	R	5′ CAAGATGAAGCTCATGCAAGTG 3′
126	wmc532	3A	F	5′ GATACATCAAGATCGTGCCAAA 3′	R	5′ GGGAGAAATCATTAACGAAGGG 3′
127	wmc559	3A	F	5′ ACACCACGAATGATGTGCCA 3′	R	5′ ACGACGCCATGTATGCAGAA 3′
128	wmc594	3A	F	5′ CCCCTCACTGCCG 3′	R	5′ ATATCCATATAGTACTCGCAC 3′
129	barc45	3AS	F	5′ CCCAGATGCAATGAAACCACAAT 3′	R	5′ GCGTAGAACTGAAGCGTAAAATTA 3′
130	cfa2226	3B	F	5′ GGAGAAAAGCAAACAGCGAC 3′	R	5′ CAGTAGCATCTTCCATGGCG 3′
131	gwm108	3B	F	5′ CGACAATGGGGTCTTAGCAT 3′	R	5′ TGCACACTTAAATTACATCCGC 3′
132	gwm299	3B	F	5′ ACTACTTAGGCCTCCCGCC 3′	R	5′ TGACCCACTTGCAATTCATC 3′
133	gwm340	3B	F	5′ ACGAGGCAAGAACACACATG 3′	R	5′ GCAATCTTTTTTCTGACCACG 3′
134	gwm389	3B	F	5′ ATCATGTCGATCTCCTTGACG 3′	R	5′ TGCCATGCACATTAGCAGAT 3′
135	gwm493	3B	F	5′ GGAACATCATTTCTGGACTTTG 3′	R	5′ TTCCCATAACTAAAACCGCG 3′
136	wmc43	3B	F	5′ TAGCTCAACCACCACCCTACTG 3′	R	5′ ACTTCAACATCCAAACTGACCG 3′
137	wmc231	3B	F	5′ CATGGCGAGGAGCTCGGTGGTC 3′	R	5′ GTGGAGCACAGGCGGAGCAAGG 3′
138	wmc307	3B	F	5′ GTTTGAAGACCAAGCTCCTCCT 3′	R	5′ ACCATAACCTCTCAAGAACCCA 3′
139	wmc418	3B	F	5′ AGAGCAGCAAGTTGTGTAGCCA 3′	R	5′ TGAAGCTATTGCCAGCACGAG 3′
140	wmc471	3B	F	5′ GGCAATAATAGTGCAAGGAATG 3′	R	5′ GCCGATAATGGGCAATATAAGT 3′
141	wmc625	3B	F	5′ CACAGACCTCAACCTCTTCTT 3′	R	5′ AGTACTGTTCACAGCAGACGA 3′
142	wmc632	3B	F	5′ GTTTGATTGGTCGTTCCTGGTC 3′	R	5′ AACAGCGAATGGAGGGCTTTAG 3′
143	wmc687	3B	F	5′ AGGACGCCTGAATCCGAG 3′	R	5′ GGGAGCGTAGGAGGACTAACA 3′
144	wmc777	3B	F	5′ GCCATCAAGCGGATCAACT 3′	R	5′ GTAGCGCCCTGTTTCACCTC 3′
145	wmc808	3B	F	5′ TGAACCATCATCGGAGCTTG 3′	R	5′ TTTTAGCCGAAGTCAAACATTGC 3′
146	barc77	3BL	F	5′ GCGTATTCTCCCTCGTTTCCAAGTCTG 3′	R	5′ GTGGGAATTTCTTGGGAGTCTGTA 3′
147	barc84	3BL	F	5′ CGCATAACCGTTGGGAAGACATCTG 3′	R	5′ GGTGCAACTAGAACGTACTTCCAGTC 3′
148	barc164	3BL	F	5′ TGCAAACTAATCACCAGCGTAA 3′	R	5′ CGCTTTCTAAAACTGTTCGGGATTTCTAA 3′
149	barc75	3BS	F	5′ AGGGTTACAGTTTGCTCTTTTAC 3′	R	5′ CCCGACGACCTATCTATACTTCTCTA 3′
150	barc87	3BS	F	5′ GCTCACCGGGCATTGGGATCA 3′	R	5′ GCGATGACGAGATAAAGGTGGAGAAC 3′
151	barc147	3BS	F	5′ GCGCCATTTATTCATGTTCCTCAT 3′	R	5′ CCGCTTCACATGCAATCCGTTGAT 3′
152	cfa2076	3D	F	5′ CGAAAAACCATGATCGACAG 3′	R	5′ ACCTGTCCAGCTAGCCTCCA 3′
153	cfd35	3D	F	5′ GGGATGACACATAACGGACA 3′	R	5′ ATCAGCGGCGCTATAGTACG 3′
154	cfd223	3D	F	5′ AAGAGCTACAATGACCAGCAGA 3′	R	5′ GCAGTGTATGTCAGGAGAAGCA 3′
155	gwm161	3D	F	5′ GATCGAGTGATGGCAGATGG 3′	R	5′ TGTGAATTACTTGGACGTGG 3′
156	gwm383	3D	F	5′ ACGCCAGTTGATCCGTAAAC 3′	R	5′ GACATCAATAACCGTGGATGG 3′
157	gwm456	3D	F	5′ TCTGAACATTACACAACCCTGA 3′	R	5′ TGCTCTCTCTGAACCTGAAGC 3′
158	wmc492	3D	F	5′ AGGATCAGAATAGTGCTACCC 3′	R	5′ ATCCCGTGATCAGAATAGTGT 3′
159	wmc533	3D	F	5′ AATTGGATCGGCAGTTGGAG 3′	R	5′ AGCAAGCAGAGCATTGCGTT 3′
160	wmc549	3D	F	5′ TTGTCACACACGCACTCCC 3′	R	5′ GTCCTTCCCTCGTTCATCCT 3′
161	wmc552	3D	F	5′ ACTAAGGAGTGTGAGGGCTGTG 3′	R	5′ CTCTCGCGCTATAAAAGAAGGA 3′
162	wmc631	3D	F	5′ TTGCTCGCCCACCTTCTACC 3′	R	5′ GGAAACCATGCGCTTCACAC 3′
163	wmc656	3D	F	5′ AAGTAGGCGAGCGTTGT 3′	R	5′ TTTCCCTGGCGAGATG 3′
164	barc71	3DL	F	5′ GCGCTTGTTCCTCACCTGCTCATA 3′	R	5′ GCGTATATTCTCTCGTCTTCTTGTTGGTT 3′
165	cfd31	4A	F	5′ GCACCAACCTTGATAGGGAA 3′	R	5′ GTGCCTGATGATTTTACCCG 3′
166	gwm160	4A	F	5′ CTGCAGGAAAAAAAGTACACCC 3′	R	5′ TTCAATTCAGTCTTGGCTTGG 3′
167	gwm610	4A	F	5′ AATGGCCAAAGGTTATGAAGG 3′	R	5′ CTGCCTTCTCCATGGTTTGT 3′

Nr	Primer Name	Chr	Primer Prefix	Primer Sequence	Primer Prefix	Primer Sequence
168	wmc48	4A	F	5′ GAGGGTTCTGAAATGTTTTGCC 3′	R	5′ ACGTGCTAGGGAGGTATCTTGC 3′
169	wmc96	4A	F	5′ TAGCAGCCATGCTTAGCATCAA 3′	R	5′ GTTTCAGTCTTTCACGAACACG 3′
170	wmc219	4A	F	5′ TGCTAGTTTGTCATCCGGGCGA 3′	R	5′ CAATCCCGTTCTACAAGTTCCA 3′
171	wmc232	4A	F	5′ GAGATTTGTTCATTTCATCTTCGCA 3′	R	5′ TATATTAAAGGTTAGAGGTAGTCAG 3′
172	wmc283	4A	F	5′ CGTTGGCTGGGTTATATCATCT 3′	R	5′ GACCCGCGTGTAAGTGATAGGA 3′
173	wmc313	4A	F	5′ GCAGTCTAATTATCTGCTGGCG 3′	R	5′ GGGTCCTTGTCTACTCATGTCT 3′
174	wmc468	4A	F	5′ AGCTGGGTTAATAACAGAGGAT 3′	R	5′ CACATAACTGTCCACTCCTTTC 3′
175	wmc491	4A	F	5′ GGTAAAACTTCGTGTCCCTTGC 3′	R	5′ TAGTTGCGAGTCGGTAGTCTGC 3′
176	wmc650	4A	F	5′ AAAGCAAGAGCAGACTGGC 3′	R	5′ GCACATCAGTAACGCATCTC 3′
177	wmc680	4A	F	5′ TGAGTGTTCAGGCCGCACTATG 3′	R	5′ ATCCTTGTTCAGGAATCCCCGT 3′
178	wmc698	4A	F	5′ GTGAAGGGAGAGCTAGCAA 3′	R	5′ ACAGTTGGCCCAGCTAGTA 3′
179	wmc707	4A	F	5′ GCTAGCTGACACTTTTCCTTTG 3′	R	5′ TCAGTTTCCCACTCACTTCTTT 3′
180	barc170	4AL	F	5′ CGCTTGACTTTGAATGGCTGAACA 3′	R	5′ CGCCCACTTTTTACCTAATCCTTTTGAA 3′
181	barc184	4AL	F	5′ TTCGGTGATATCTTTTCCCCTTGA 3′	R	5′ CCGAGTTGACTGTGTGGGCTTGCTG 3′
182	barc206	4AS	F	5′ GCTTTGCCAGGTGAGCACTCT 3′	R	5′ TGGCCGGGTATTTGAGTTGGAGTTT 3′
183	barc20	4B	F	5′ GCGATCCACACTTTGCCTCTTTTACA 3′	R	5′ GCGATGTCGGTTTTCAGCCTTTT 3′
184	gwm6	4B	F	5′ AGCCTTATCATGACCCTACCTT 3′	R	5′ CGTATCACCTCCTAGCTAAACTAG 3′
185	gwm149	4B	F	5′ CATTGTTTTCTGCCTCTAGCC 3′	R	5′ CTAGCATCGAACCTGAACAAG 3′
186	gwm192	4B	F	5′ CGTTGTCTAATCTTGCCTTGC 3′	R	5′ GGTTTTCTTTCAGATTGCGC 3′
187	gwm251	4B	F	5′ CAACTGGTTGCTACACAAGCA 3′	R	5′ GGGATGTCTGTTCCATCTTAG 3′
188	gwm368	4B	F	5′ AATAAAACCATGAGCTCACTTGC 3′	R	5′ CCATTTCACCTAATGCCTGC 3′
189	wmc47	4B	F	5′ GAAACAGGGTTAACCATGCCAA 3′	R	5′ ATGGTGCTGCCAACAACATACA 3′
190	wmc89	4B	F	5′ ATGTCCACGTGCTAGGGAGGTA 3′	R	5′ TTGCCTCCCAAGACGAAATAAC 3′
191	wmc125	4B	F	5′ ATACCACCATGCATGTGGAAGT 3′	R	5′ ACCGCTTGTCATTTCCTTCTGT 3′
192	wmc238	4B	F	5′ TCTTCCTGCTTACCCAAACACA 3′	R	5′ TACTGGGGGATCGTGGATGACA 3′
193	wmc349	4B	F	5′ ACACACACTCGATCGCAC 3′	R	5′ GCAGTTGATCATCAAAACACA 3′
194	wmc413	4B	F	5′ CACTGGAAACATCTCTTCAACT 3′	R	5′ ACAGGAAAGGATGATGTTCTCT 3′
195	wmc710	4B	F	5′ GTAAGAAGGCAGCACGTATGAA 3′	R	5′ TAAGCATTCCCAATCACTCTCA 3′
196	cfd71	4D	F	5′ CAATAAGTAGGCCGGGACAA 3′	R	5′ TGTGCCAGTTGAGTTTGCTC 3′
197	cfd84	4D	F	5′ GTTGCCTCGGTGTCGTTTAT 3′	R	5′ TCCTCGAGGTCCAAAACATC 3′
198	wmc285	4D	F	5′ TGTGGTTGTATTTGCGGTATGG 3′	R	5′ TTGTGGTGCTGAGTTAGCTTGT 3′
199	wmc331	4D	F	5′ CCTGTTGCATACTTGACCTTTTT 3′	R	5′ GGAGTTCAATCTTTCATCACCAT 3′
200	wmc457	4D	F	5′ CTTCCATGAATCAAAGCAGCAC 3′	R	5′ CATCCATGGCAGAAACAATAGC 3′
201	wmc622	4D	F	5′ CAGGAAGAAGAGCTCCGAGAAA 3′	R	5′ CTTGCTAACCCGCGCC 3′
202	cfa2104	5A	F	5′ CCTGGCAGAGAAAGTGAAGG 3′	R	5′ AGTCGCCGTTGTATAGTGCC 3′
203	cfa2141	5A	F	5′ GAATGGAAGGCGGACATAGA 3′	R	5′ GCCTCCACAACAGCCATAAT 3′
204	cfa2163	5A	F	5′ TTGATCCTTGATGGGAGGAG 3′	R	5′ CATCATTGTGTTTACGTTCTTTCA 3′
205	cfa2190	5A	F	5′ CAGTCTGCAATCCACTTTGC 3′	R	5′ AAAAGGAAACTAAAGCGATGGA 3′
206	cfa2250	5A	F	5′ AGCCATAGATGGCCCTACCT 3′	R	5′ CACTCAATGGCAGGTCCTTT 3′
207	gwm126	5A	F	5′ CACACGCTCCACCATGAC 3′	R	5′ GTTGAGTTGATGCGGGAGG 3′
208	gwm156	5A	F	5′ CAATGCAGGCCCTCCTAAC 3′	R	5′ CCAACCGTGCTATTAGTCATTC 3′
209	gwm291	5A	F	5′ AATGGTATCTATTCCGACCCG 3′	R	5′ CATCCCTACGCCACTCTGC 3′
210	gwm293	5A	F	5′ TACTGGTTCACATTGGTGCG 3′	R	5′ TCGCCATCACTCGTTCAAG 3′
211	gwm304	5A	F	5′ AGGAAACAGAAATATCGCGG 3′	R	5′ AGGACTGTGGGGAATGAATG 3′
212	gwm595	5A	F	5′ GCATAGCATCGCATATGCAT 3′	R	5′ GCCACGCTTGGACAAGATAT 3′
213	gwm617	5A	F	5′ CTCCGATGGATTACTCGCAC 3′	R	5′ GATCTTGGCGCTGAGAGAGA 3′
214	wmc110	5A	F	5′ GCAGATGAGTTGAGTTGGATTG 3′	R	5′ GTACTTGGAAACTGTGTTTGGG 3′
215	wmc415	5A	F	5′ AATTCGATACCTCTCACTCACG 3′	R	5′ TCAACTGCTACAACCTAGACCC 3′
216	wmc445	5A	F	5′ AGAATAGGTTCTTGGGCCAGTC 3′	R	5′ GAGATGATCTCCTCCATCAGCA 3′
217	wmc524	5A	F	5′ TAGTCCACCGGACGGAAAGTAT 3′	R	5′ GTACCACCGATTGATGCTTGAG 3′
218	wmc713	5A	F	5′ ACATAGCATCCCATACTGAGAGAGG 3′	R	5′ ATGCGGGGAATAGAGACACAC 3′
219	wmc727	5A	F	5′ CATAATCAGGACAGCCGCAC 3′	R	5′ TAGTGGCCTGATGTATCTAGTTGG 3′
220	wmc805	5A	F	5′ GATGCTGCTGCACCAAACTC 3′	R	5′ GCCTTTTCCATGCCACACT 3′
221	barc151	5AL	F	5′ TGAGGAAAATGTCTCTATAGCATCC 3′	R	5′ CGCATAAACACCTTCGCTCTTCCACTC 3′
222	barc165	5AL	F	5′ GCGTAGAGCGGCTGTTAGTGTCAAATTA 3′	R	5′ GCGTTATCTCAAGTTTTGTAGCAGA 3′
223	gwm371	5B	F	5′ AGCTCAGCTTGCTTGGTACC 3′	R	5′ GACCAAGATATTCAAACTGGCC 3′
224	gwm408	5B	F	5′ GTATAATTCGTTCACAGCACGC 3′	R	5′ TCGATTTATTTGGGCCACTG 3′

Nr	Primer Name	Chr	Primer Prefix	Primer Sequence	Primer Prefix	Primer Sequence
225	gwm443	5B	F	5′ CCATGATTTATAAATTCCACC 3′	R	5′ GGGTCTTCATCCGGAACTCT 3′
226	gwm554	5B	F	5′ GCAACCACCAAGCACAAAGT 3′	R	5′ TGCCCACAACGGAACTTG 3′
227	gwm604	5B	F	5′ ATCTTTTGAACCAAATGTG 3′	R	5′ TATATAGTTCAATATGACCCG 3′
228	wmc73	5B	F	5′ TTGTGCACCGCACTTACGTCTC 3′	R	5′ ACACCCGGTCTCCGATCCTTAG 3′
229	wmc99	5B	F	5′ ATTACAATTGCTTCAGTGAGTG 3′	R	5′ TCATGATCATTGTTATAACGGT 3′
230	wmc160	5B	F	5′ CATGGCTCCAAGATACAAAAAG 3′	R	5′ AGGCCTGGATTCATGATAGATA 3′
231	wmc289	5B	F	5′ CATATGCATGCTATGCTGGCTA 3′	R	5′ AGCCTTTCAAATCCATCCACTG 3′
232	wmc376	5B	F	5′ TCTCAACCACCGACTTGTAA 3′	R	5′ ACATGTAATTGGGGACACTG 3′
233	wmc508	5B	F	5′ AGCCCTTGAGTTGGTCTCATTT 3′	R	5′ GAGCAGAGCTCCACTCACATTT 3′
234	wmc537	5B	F	5′ TCTTCTGTACATTGAACAACGA 3′	R	5′ ATGCAGAACCGTGATAGGAT 3′
235	wmc740	5B	F	5′ CTTGGTTGCAGACGGGG 3′	R	5′ GCTGGGTGCAATGCAGATAG 3′
236	barc4	5BL	F	5′ GCGTGTTTGTGTCTGCGTTCTA 3′	R	5′ CACCACACATGCCACCTTCTTT 3′
237	barc59	5BL	F	5′ GCGTTGGCTAATCATCGTTCCTTC 3′	R	5′ AGCACCCTACCCAGCGTCAGTCAAT 3′
238	barc232	5BL	F	5′ CGCATCCAACCATCCCCACCCAACA 3′	R	5′ CGCAGTAGATCCACCACCCCGCCAGA 3′
239	cfa2185	5D	F	5′ TTCTTCAGTTGTTTTGGGGG 3′	R	5′ TTTGGTCGACAAGCAAATCA 3′
240	cfd7	5D	F	5′ AGCTACCAGCCTAGCAGCAG 3′	R	5′ TCAGACACGTCTCCTGACAAA 3′
241	cfd8	5D	F	5′ ACCACCGTCATGTCACTGAG 3′	R	5′ GTGAAGACGACAAGACGCAA 3′
242	cfd10	5D	F	5′ CGTTCTATGACGTGTCATGCT 3′	R	5′ TCCATTTTCAAAAACACCCTG 3′
243	cfd18	5D	F	5′ CATCCAACAGCACCAAGAGA 3′	R	5′ GCTACTACTATTTCATTGCGACCA 3′
244	cfd29	5D	F	5′ GGTTGTCAGGCAGGATATTTG 3′	R	5′ TATTGATAGATCAGGGCGCA 3′
245	cfd40	5D	F	5′ GCGACAAGTAATTCAGAACGG 3′	R	5′ CGCTTCGGTAAAGTTTTTGC 3′
246	cfd57	5D	F	5′ ATCGCCGTTAACATAGGCAG 3′	R	5′ TCACTGCTGTATTTGCTCCG 3′
247	cfd86	5D	F	5′ TTAATGAGCGTCAGTACTCCC 3′	R	5′ GCAACCATGTTTAAGCCGAT 3′
248	cfd102	5D	F	5′ TTGTGGAAGGGTTTGATGAAG 3′	R	5′ TGCAGGACCAAACATAGCTG 3′
249	cfd183	5D	F	5′ ACTTGCACTTGCTATACTTACGAA 3′	R	5′ GTGTGTCGGTGTGTGGAAAG 3′
250	cfd189	5D	F	5′ GCTAAAGCCACATAGGACGG 3′	R	5′ GCACAAGATTTTGCAAGGCT 3′
251	gwm182	5D	F	5′ TGATGTAGTGAGCCCATAGGC 3′	R	5′ TTGCACACAGCCAAATAAGG 3′
252	gwm190	5D	F	5′ GTGCCACGTGGTACCTTTG 3′	R	5′ GTGCTTGCTGAGCTATGAGTC 3′
253	gwm271	5D	F	5′ AGCTGCTAGCTTTTGGGACA 3′	R	5′ CAAGATCGTGGAGCCAGC 3′
254	gwm272	5D	F	5′ GTTCAAAACAAATTAAAAGGCCC 3′	R	5′ TGCTCTTTGGCGAATATATGG 3′
255	gwm292	5D	F	5′ CCACCGAGCCGATAATGTAC 3′	R	5′ TCACCGTGGTCACCGAC 3′
256	gwm335	5D	F	5′ CGGTCCAAGTGCTACCTTTC 3′	R	5′ CGTACTCCACTCCACACGG 3′
257	wmc233	5D	F	5′ GACGTCAAGAATCTTCGTCGGA 3′	R	5′ ATCTGCTGAGCAGATCGTGGTT 3′
258	wmc357	5D	F	5′ TAGTGGGTGACCGGTCAAGA 3′	R	5′ TGGACGGATTTGGTCATTTC 3′
259	wmc765	5D	F	5′ GGGATCAGACTGGGACTGGAG 3′	R	5′ GGGTTGGCTTGGCAGAGAA 3′
260	barc143	5DL	F	5′ TTGTGCCAAATCAAGAACAT 3′	R	5′ GGTTGGGCTAGGATGAAAAT 3′
261	barc3	6A	F	5′ TTCCCTGTGTCTTTCTAATTTTTTTT 3′	R	5′ GCGAACTCCCGAACATTTTTAT 3′
262	cfd190	6A	F	5′ CAATCAGAAGCGCCATTGTT 3′	R	5′ CCCTGATGTTTTCTTTTTCTCC 3′
263	gwm334	6A	F	5′ AACATGTGTTTTTAGCTATC 3′	R	5′ AATTTCAAAAAGGAGAGAGA 3′
264	gwm356	6A	F	5′ AGCGTTCTTGGGAATTAGAGA 3′	R	5′ CCAATCAGCCTGCAACAAC 3′
265	gwm494	6A	F	5′ ATTGAACAGGAAGACATCAGGG 3′	R	5′ TTCCTGGAGCTGTCTGGC 3′
266	wmc201	6A	F	5′ CATGCTCTTTCACTTGGGTTCG 3′	R	5′ GCGCTTGCAGGAATTCAACACT 3′
267	wmc254	6A	F	5′ AGTAATCTGGTCCTCTCTTCTTCT 3′	R	5′ AGGTAATCTCCGAGTGCACTTCAT 3′
268	wmc256	6A	F	5′ CCAAATCTTCGAACAAGAACCC 3′	R	5′ ACCGATCGATGGTGTATACTGA 3′
269	wmc417	6A	F	5′ GTTCTTTTAGTTGCGACTGAGG 3′	R	5′ CGATGTATGCCGTATGAATGTT 3′
270	wmc553	6A	F	5′ CGGAGCATGCAGCTAGTAA 3′	R	5′ CGCCTGCAGAATTCAACAC 3′
271	cfd13	6B	F	5′ CCACTAACCAAGCTGCCATT 3′	R	5′ TTTTTGGCATTGATCTGCTG 3′
272	gwm193	6B	F	5′ AATTGTGTTGATGATTTGGGG 3′	R	5′ CTTTGTGCACCTCTCTCTCC 3′
273	gwm518	6B	F	5′ AATCACAACAAGGCGTGACA 3′	R	5′ CAGGGTGGTGCATGCAT 3′
274	gwm705	6B	F	5′ ACC ATA AAA TAT GAG CTA AGG 3′	R	5′ TCC TAC AAG GTG AAG TAA AA 3′
275	wmc487	6B	F	5′ CAAATTTGGCCACCATTTTACA 3′	R	5′ CGGTTCAATCCTTGGATTTACA 3′
276	wmc494	6B	F	5′ GGATCGAGTCTCAAGTCTACAA 3′	R	5′ AGAAGGAACAAGCAACATCATA 3′
277	barc24	6BL	F	5′ CGCCTCTTATGGACCAGCCTAT 3′	R	5′ GCGGTGAGCCATCGGGTTACAAAG 3′
278	barc67	6D	F	5′ GCGGCATTTACATTTCAGATAGA 3′	R	5′ TGTGCCTGATTGTAGTAACGTATGTA 3′
279	barc96	6D	F	5′ AAGCCTTGTTGTTCCGTATTATT 3′	R	5′ GCGGTTTATATTTTGTGGTTGAGCATTTT 3′
280	cfd5	6D	F	5′ TGCCCTGTCCACAGTGAAG 3′	R	5′ TTGCCAGTTCCAAGGAGAAT 3′
281	cfd49	6D	F	5′ TGAGTTCTTCTGGTGAGGCA 3′	R	5′ GAATCGGTTCACAAGGGAAA 3′

Nr	Primer Name	Chr	Primer Prefix	Primer Sequence	Primer Prefix	Primer Sequence
282	cfd60	6D	F	5′ TGACCGGCATTCAGTATCAA 3′	R	5′ TGGTCACTTTGATGAGCAGG 3′
283	cfd75	6D	F	5′ GCATAAACTTGGACCCTGGA 3′	R	5′ GCTAAGCCACGCTACCACTC 3′
284	cfd132	6D	F	5′ CAAATGCTAATCCCCGCC 3′	R	5′ TGTAAACAAGGTCGCAGGTG 3′
285	cfd188	6D	F	5′ AATGGCTTCACTGTTTGCCT 3′	R	5′ AAATGGTCCCAGCATTCAAG 3′
286	gwm469	6D	F	5′ CAACTCAGTGCTCACACAACG	R	5′ CGATAACCACTCATCCACACC
287	barc146	6DS	F	5′ AAGGCGATGCTGCAGCTAAT 3′	R	5′ GGCAATATGGAAACTGGAGAGAAAT 3′
288	barc173	6DS	F	5′ GGGGATCCTTCAACAATAACA 3′	R	5′ GCGAGATGGCATTTTTAAATAAAGAGAC 3′
289	cfa2019	7A	F	5′ GACGAGCTAACTGCAGACCC 3′	R	5′ CTCAATCCTGATGCGGAGAT 3′
290	cfa2028	7A	F	5′ TGGGTATGAAAGGCTGAAGG 3′	R	5′ ATCGCGACTATTCAACGCTT 3′
291	cfa2040	7A	F	5′ TCAAATGATTTCAGGTAACCACTA 3′	R	5′ TTCCTGATCCCACCAAACAT 3′
292	cfa2049	7A	F	5′ TAATTTGATTGGGTCGGAGC 3′	R	5′ CGTGTCGATGGTCTCCTTG 3′
293	cfa2257	7A	F	5′ GATACAATAGGTGCCTCCGC 3′	R	5′ CCATTATGTAAATGCTTCTGTTTGA 3′
294	cfd6	7A	F	5′ ACTCTCCCCCTCGTTGCTAT 3′	R	5′ ATTTAAGGGAGACATCGGGC 3′
295	gwm276	7A	F	5′ AATTTCACTGCATACACAAG 3′	R	5′ ATTTGCCTGAAGAAAATATT 3′
296	gwm471	7A	F	5′ CGGCCCTATCATGGCTG 3′	R	5′ GCTTGCAAGTTCCATTTTGC 3′
297	wmc9	7A	F	5′ AACTAGTCAAATAGTCGTGTCCG 3′	R	5′ GTCAAGTCATCTGACTTAACCCG 3′
298	wmc83	7A	F	5′ TGGAGGAAACACAATGGATGCC 3′	R	5′ GAGTATCGCCGACGAAAGGGAA 3′
299	wmc593	7A	F	5′ GGGGAGAAGCAGCAGGG 3′	R	5′ CGCGCGGTTGCCGGTGG 3′
300	wmc603	7A	F	5′ ACAAACGGTGACAATGCAAGGA 3′	R	5′ CGCCTCTCTCGTAAGCCTCAAC 3′
301	wmc646	7A	F	5′ GGAGTAAATGGAGACGGGGAC 3′	R	5′ GCCAGTGTGATGCATGTGAC 3′
302	wmc790	7A	F	5′ AATTAAGATAGACCGTCCATATCATCCA 3′	R	5′ CGACAACGTACGCGCC 3′
303	wmc809	7A	F	5′ CAGGTCGTAGTTGGTACCCTGAA 3′	R	5′ TGAACACGGCTGGATGTGA 3′
304	barc121	7AL	F	5′ ACTGATCAGCAATGTCAACTGAA 3′	R	5′ CCGGTGTCTTTCCTAACGCTATG 3′
305	barc108	7AS	F	5′ GCGGGTCGTTTCCTGGAAATTCATCTAA 3′	R	5′ GCGAAATGATTGGCGTTACACCTGTTG 3′
306	barc127	7AS	F	5′ TGCATGCACTGTCCTTTGTATT 3′	R	5′ AAGATGCGGGCTGTTTTCTA 3′
307	gwm146	7B	F	5′ CCAAAAAAACTGCCTGCATG 3′	R	5′ CTCTGGCATTGCTCCTTGG 3′
308	gwm333	7B	F	5′ GCCCGGTCATGTAAAACG 3′	R	5′ TTTCAGTTTGCGTTAAGCTTTG 3′
309	gwm400	7B	F	5′ GTGCTGCCACCACTTGC 3′	R	5′ TGTAGGCACTGCTTGGGAG 3′
310	gwm537	7B	F	5′ ACATAATGCTTCCTGTGCACC 3′	R	5′ GCCACTTTTGTGTCGTTCCT 3′
311	wmc10	7B	F	5′ GATCCGTTCTGAGGTGAGTT 3′	R	5′ GGCAGCACCCTCTATTGTCT 3′
312	wmc76	7B	F	5′ CTTCAGAGCCTCTTTCTCTACA 3′	R	5′ CTGCTTCACTTGCTGATCTTTG 3′
313	wmc311	7B	F	5′ GGGCCTGCATTTCTCCTTTCTT 3′	R	5′ CTGAACTTGCTAGACGTTCCGA 3′
314	wmc396	7B	F	5′ TGCACTGTTTTACCTTCACGGA 3′	R	5′ CAAAGCAAGAACCAGAGCCACT 3′
315	wmc475	7B	F	5′ AACACATTTTCTGTCTTTCGCC 3′	R	5′ TGTAGTTATGCCCAACCTTTCC 3′
316	wmc517	7B	F	5′ ATCCTGACGTTACACGCACC 3′	R	5′ ACCTGGAACACCACGACAAA 3′
317	wmc606	7B	F	5′ CCGATGAACAGACTCGACAAGG 3′	R	5′ GGCTTCGGCCAGTAGTACAGGA 3′
318	wmc696	7B	F	5′ ACCCGAGAGAGATTAGGGCTTG 3′	R	5′ CACTCGCAGCCTCTCTTCTACC 3′
319	wmc758	7B	F	5′ TAGGGGAGGCGACGGAG 3′	R	5′ GTTGCTGGAGAGTGGATTGC 3′
320	cfd14	7D	F	5′ CCACCGGCCAGAGTAGTATT 3′	R	5′ TCCTGGTCTAACAACGAGAAGA 3′
321	cfd21	7D	F	5′ CCTCCATGTAGGCGGAAATA 3′	R	5′ TGTGTCCCATTCACTAACCG 3′
322	cfd69	7D	F	5′ AAATACCTTGAATTGTGAGCTGC 3′	R	5′ TCTGTTCATCCCCAAAGTCC 3′
323	gwm44	7D	F	5′ ACTGGCATCCACTGAGCTG 3′	R	5′ GTTGAGCTTTTCAGTTCGGC 3′
324	gwm111	7D	F	5′ ACC TGA TCA GAT CCC ACT CG 3′	R	5′ ACCTGATCAGATCCCACTCG 3′
325	gwm428	7D	F	5′ CGAGGCAGCGAGGATTT 3′	R	5′ TTCTCCACTAGCCCCGC 3′
326	gwm635	7D	F	5′ CAGCCTTAGCCTTGGCG 3′	R	5′ TTCCTCACTGTAAGGGCGTT 3′
327	wmc438	7D	F	5′ GACCGTTGGGCTGTATAGCATT 3′	R	5′ CTCTGACAGTGGTGGAGCTTGA 3′
328	wmc463	7D	F	5′ GATTGTATAGTCGGTTACCCCT 3′	R	5′ ATTAGTGCCCTCCATAATTGTG 3′
329	wmc506	7D	F	5′ CACTTCCTCAACATGCCAGA 3′	R	5′ CTTTCAATGTGGAAGGCGAC 3′
330	wmc634	7D	F	5′ AGCGAGGAGGATGCATCTTATT 3′	R	5′ GACATACACATGATGGACACGG 3′
331	wmc824	7D	F	5′ CCGATGAACTTAAAAGTACCACCTG 3′	R	5′ CATGGATTGACACGATTGGC 3′
332	barc76	7DL	F	5′ ATTCGTTGCTGCCACTTGCTG 3′	R	5′ GCGCGACACGGAGTAAGGACACC 3′
333	barc111	7DL	F	5′ GCGGTCACCAGTAGTTCAACA 3′	R	5′ GCGTATCCCATTGCTCTTCTTCACTAAC 3′
334	barc172	7DL	F	5′ GCGAAATGTGATGGGGTTTATCTA 3′	R	5′ GCGATTTGATTTAACTTTAGCAGTGAG 3′

Appendix 5. Set of across the genome distributed and highly polymorphic KASP assays

No	SNP ID	Name SNP	Chr	Pos*	FAM allele	VIC allele	FAM Primer	VIC Primer	Common Primer
1	IWB34801	IAAV3919	1A	13.7	A	G	tctaatccagagacgacGcT	tctaatccagagacgacGcC	CCCacccaccAcatttgttG
2	IWB28577	Excalibur_c71158_54	1A	27.1	G	A	cctttgaaTgcttctctgtaAggT	cctttgaaTgcttctctgtaAggC	gccattgtggacggacgG
3	IWB30651	Excalibur_rep_c113950_132	1A	35.2	A	G	gcactcaacttttctgtgttctA	gcactcaacttttctgtgttctG	acggtcaggggagttcaaga
4	IWB59016	RAC875_c54245_88	1A	43.3	A	G	gcagatcatcgtacatgtcgaT	gcagatcatcgtacatgtcgaC	caagtaccgggcttacctcA
5	IWB73417	Tdurum_contig77289_179	1A	56.8	A	G	aagtacctgaattctgtGtggT	aagtacctgaattctgtGtggC	gacaCatgaCcagcaTcagA
6	IWB59114	RAC875_c5544_4156	1A	70.1	A	G	actcaaggcaaggatgttgaatA	actcaaggcaaggatgttgaatG	actttatacgcttgCtttgtggT
7	IWB35897	IACX3047	1A	80.8	A	G	tttcagaagctcagaaatactgctA	tttcagaagctcagaaatactgctG	ttcccattgatacaaggaaactacG
8	IWB8167	BS00039378_51	1A	94.5	T	C	aagggcatgtctaagctcgA	aagggcatgtctaagctcgG	acgtcaaatatggtcatacaaaccc
9	IWA1081	wsnp_CAP7_c3472_1623955	1A	102.3	A	G	catttggcctgtTgtttgcT	catttggcctgtTgtttgcC	tgTttgtttacagcaggtgaaaT
10	IWB60380	RAC875_c7563_273	1A	118.8	A	G	aggtgcaactgtgtatcttcttaT	aggtgcaactgtgtatcttcttaC	cagcatccaccgaattatatcag
11	IWB8918	BS00062759_51	1A	144.4	T	C	ggtatccaaatgttcctagcacA	ggtatccaaatgttcctagcacG	gacaaaccagggTacGgaAT
12	IWB10518	BS00070991_51	1A	149.8	A	G	gcccaggaagcgaagtacA	gcccaggaagcgaagtacG	cacctcaacttcggcctcaa
13	IWB7109	BS00022482_51	1B	9.7	T	C	tccaccatactgctacaccaT	tccaccatactgctacaccaC	TtctcagtagccgcccatT
14	IWB44529	Kukri_c36151_170	1B	24.5	A	G	cagcagtctGcctccctT	cagcagtctGcctccctC	gatactattacagtaccgtaggacc
15	IWB43758	Kukri_c29655_194	1B	41.5	T	C	ccagcaggatggGtgttgaT	ccagcaggatggGtgttgaC	ccgtacctccagccatcC
16	IWB7324	BS00022920_51	1B	75.1	T	C	acaatgtcgcttaggcatacA	acaatgtcgcttaggcatacG	gtctcggttcagcttttcct
17	IWB6709	BS00021680_51	1B	96.2	A	G	tctgtttgatttggatcagatctA	tctgtttgatttggatcagatctG	acgacgctggtcaatgctaa
18	IWB1342.1	BobWhite_c20073_443	1B	148.4	A	G	aactttattacgtttctgaagccA	aactttattacgtttctgaagccG	gcctatttgtgcgccggc
19	IWB65272	RFL_Contig785_535	1B	164.6	A	G	aggagacgCaggAgccaT	aggagacgCaggAgccaC	ctTGcattgccggctcAaC
20	IWB12221	BS00104270_51	1B	173.6	A	G	gagagggtccCgatggcT	gagagggtccCgatggcC	gcAaccatgttcgtcaagatC
21	IWA3753	wsnp_Ex_c41048_47969948	1D	8.5	A	G	GgCttaTcaTcttcGatctgcA	GgCttaTcaTcttcGatctgcG	aaagacCTgtaacaagtccaGaA
22	IWB1009	BobWhite_c1715_887	1D	21.8	A	C	cgggtggatgtccggaatA	cgggtggatgtccggaatC	tccaagccttcgcaactctt
23	IWB14612	CAP8_c2401_433	1D	45.4	A	G	ccacAgtccacaCaaagcataT	ccacAgtccacaCaaagcataC	tgaggtctcgtgttaataactgC
24	IWB9925	BS00066976_51	1D	78.4	T	C	atccggcgctccctcaaT	atccggcgctccctcaaC	ttccggggatcgccgatg
25	IWB34905	IAAV4656	1D	99.6	T	C	gtgtgcattcTagtcccacA	gtgtgcattcTagtcccacC	ggaaaaccccagtacgcaTG
26	IWB11807	BS00093390_51	1D	115.6	T	C	ggcaaatgcacgacaaaattT	ggcaaatgcacgacaaaattC	tcccttgctcCccaaagtcT
27	IWB56360	RAC875_c29598_336	1D	129.7	A	G	tcAggtccacggttgacA	tcAggtccacggttgacG	caagaaaggttttcgAgaggtG
28	IWB69918	Tdurum_contig29915_167	1D	179.5	A	C	acaattattctccagttgcagtttT	acaattattctccagttgcagtttG	tcttggatgcgcttgcgA
29	IWB20060	Ex_c19516_3687	2A	6.0	A	C	ttcacaTgtaacagaggctacA	ttcacaTgtaacagaggctacC	GcggtgttggattttcAccT
30	IWB65498	TA002095-0637	2A	26.0	T	C	cccttagcaatcggatcAacA	cccttagcaatcggatcAacG	ccactatttcgtgacccgataC
31	IWB22049	Excalibur_c12177_285	2A	47.2	A	G	gccagcagcgctaatattA	gccagcagcgctaatattG	ttactcttcccttaagcagcttAT
32	IWA2433	wsnp_Ex_c19556_28530231	2A	73.9	T	C	accatgttaggctgagtggA	accatgttaggctgagtggG	ttcctggtcatgcacaacC
33	IWB65859	TA004602-1630	2A	94.4	A	G	ccttcaccagcttccggatA	ccttcaccagcttccggatG	gctcttcattctGcattatgAtgtT
34	IWB61745	RAC875_rep_c107961_348	2A	130.4	T	C	cccatacctatactgtgaccgA	cccatacctatactgtgaccgG	taaggcatcacgacggcaaa
35	IWB34772	IAAV3791	2A	141.4	A	G	cttggtcgtctcGgtaacaT	cttggtcgtctcGgtaacaC	gCgcaatcagatcagtggaT
36	IWB14668	CAP8_c3129_381	2A	154.8	A	G	tgcgtcattgatagtgttagtgaA	tgcgtcattgatagtgttagtgaG	gtaaaCtactcaagacaaacagacC
37	IWA5161	wsnp_Ex_rep_c108004_91402649	2A	167.9	T	C	tggCatcatatttttgcaggaaaT	tggCatcatatttttgcaggaaaC	aaccacatTcaggatcattttCC
38	IWB62322	RAC875_rep_c115433_378	2B	20.6	T	C	ccatCtcaagtggaccaaaCA	ccatCtcaagtggaccaaaCG	agacaccatggaaccttgtcttC
39	IWB57369	RAC875_c38003_164	2B	35.1	T	C	gcacatataccctggatgctcT	gcacatataccctggatgctcC	cccatccatgtagcaaagcg
40	IWB64873	RFL_Contig5031_526	2B	52.1	A	C	gctcaacatcaaactgcagcT	gctcaacatcaaactgcagcG	tgggatttgggcatcaaaacA
41	IWB34673	IAAV3165	2B	72.0	A	G	tgtttgtgttgtatGgtgctTT	tgtttgtgttgtatGgtgctTC	ttattaccAtcggtccaagattcT
42	IWA5811	wsnp_JD_c1236_1789566	2B	95.8	A	G	accagatacatggattggagactA	accagatacatggattggagactG	agacgcacaacttttgcaaC
43	IWA8406	wsnp_RFL_Contig2914_2757372	2B	119.6	A	G	agtgacAAagcaaacagtgaaT	agtgacAAagcaaacagtgaaC	gggtCcttctgggatcatgttcA
44	IWB2628	BobWhite_c33464_133	2B	140.5	A	G	ttttgtgagtgtcagaaatgcA	ttttgtgagtgtcagaaatgcG	gcttatcgggtaactatgtttgG
45	IWB11333	BS00083998_51	2B	161.4	A	G	cacggaaccagactgGcA	cacggaaccagactgGcG	gaacccgttctcagcGaaT
46	IWB24255	Excalibur_c25043_357	2B	181.9	A	G	tcaggtgcctcatccacA	tcaggtgcctcatccacG	cagtgagtttggctagtaaattttg
47	IWB25618	Excalibur_c35611_446	2D	11.2	A	G	aagaataagttcataattgtcgccA	aagaataagttcataattgtcgccG	tccggcttgaagaggagG
48	IWB55568	RAC875_c23815_545	2D	98.6	A	G	gtgTttggacaaatgcTcctcgtaT	gtgTttggacaaatgcTcctcgtaC	agaagctgaaAaaggtgctggagA
49	IWB41521	Kukri_c16094_406	2D	103.3	T	C	aacgcCgtcttcttggcA	aacgcCgtcttcttggcG	gagcagcgcattgtcgcC
50	IWB7261	BS00022798_51	3A	23.6	T	C	caccttctcTgcCacctcA	caccttctcTgcCacctcG	tTgacggtgaccatggcatT

No	SNP ID	Name SNP	Chr	Pos*	FAM allele	VIC allele	FAM Primer	VIC Primer	Common Primer
51	IWB65471	TA001885-0568	3A	35.5	A	G	ggagtGcagggTcgacaA	ggagtGcagggTcgacaG	tccgaatctccgagttcttagT
52	IWB24284	Excalibur_c25239_283	3A	60.3	A	G	aatggcccatgcattGacgaT	aatggcccatgcattGacgaC	gatacatggttcctgcTgggacA
53	IWB48828	Kukri_rep_c102953_304	3A	83.3	A	G	gcagggatcccattcaccaT	gcagggatcccattcaccaC	ccgatggtttgtggtctatccA
54	IWB60376	RAC875_c75448_80	3A	97.8	T	C	cagagcagatcgtccccA	cagagcagatcgtccccG	gtggtttggtttggtttggC
55	IWB54593.1	RAC875_c18058_1070	3A	123.0	A	G	gcctgacacccttgctcaaT	gcctgacacccttgctcaaC	gctgatgcaacagaggcaac
56	IWB58806	RAC875_c52195_324	3A	146.9	A	G	aaaccttgtccacTaacaaagatA	aaaccttgtccacTaacaaagatG	cctttcaaggctccagtgag
57	IWB34789	IAAV3851	3A	169.9	T	G	atgtatcgttatgtcatgcacttT	atgtatcgttatgtcatgcacttG	gccggagcaatgctgtagat
58	IWA3177	wsnp_Ex_c28679_37784735	3A	185.4	A	C	caaaaatgatcgatGgcAccaT	caaaaatgatcgatGgcAccaG	tggaaaacccataccttcaaatgT
59	IWB72604	Tdurum_contig57914_1156	3B	5.8	A	G	gGtccttgatgtctttagaacaaA	gGtccttgatgtctttagaacaaG	tcaGtTACCCaAcagcaGAG
60	IWB9677	BS00065934_51	3B	45.9	A	G	ccttctcgcttagcatcctT	ccttctcgcttagcatcctC	ggcgagctgacactgaatct
61	IWB28234	Excalibur_c63730_660	3B	65.6	A	G	agtatgagaacatcagagggacA	agtatgagaacatcagagggacG	gcacAggtttaagttgtactccT
62	IWB25787	Excalibur_c37115_306	3B	78.2	T	C	tggagatcttgcaaagtttggT	tggagatcttgcaaagtttggC	acctgaacaagctgatcttcG
63	IWB71098	Tdurum_contig42366_944	3B	91.5	T	C	ggatactgctgttcatgtactcA	ggatactgctgttcatgtactcG	ggacagactcctgaagcatgt
64	IWB22558	Excalibur_c15095_852	3B	125.4	A	G	agaagtgtgctggctatgagA	agaagtgtgctggctatgagG	ctgaattcagccactcacgc
65	IWB50139	Kukri_rep_c71747_150	3B	144.4	T	C	gacgaatagtgaatttggaccatT	gacgaatagtgaatttggaccatC	gtgagagaacatgttgaccaca
66	IWB52937	RAC875_c101793_136	3D	36.4	A	G	ttgtatgtataatcatgctgcgaaT	ttgtatgtataatcatgctgcgaaC	accaacgaaccatctttgca
67	IWB7702.1	BS00028997_51	3D	97.7	A	G	agtgatcatccactgccatctT	agtgatcatccactgccatctC	agaaagcatctaacacatggct
68	IWB44015	Kukri_c31733_290	3D	107.9	A	G	cctattgttgtccaaatccgtGatA	cctattgttgtccaaatccgtGatG	tgccatacatgtttagcagggaC
69	IWB12285	BS00106545_51	4A	11.6	A	G	ccGttcgagcacgtgacT	ccGttcgagcacgtgacC	caatctgatcatggttgctgG
70	IWB72664	Tdurum_contig59603_74	4A	26.5	A	G	taGgggttgCgtttcaCcA	taGgggttgCgtttcaCcG	tgcAcAcaaGgcatgtatATgtA
71	IWA2533	wsnp_Ex_c20386_29451037	4A	40.3	T	C	ccgtgtaaggattcacaaatacaT	ccgtgtaaggattcacaaatacaC	tGCAGgatgtGttgtgttcA
72	IWB12180	BS00101512_51	4A	108.7	A	G	gcatcaatgaacctcgcatgT	gcatcaatgaacctcgcatgC	tgaatgcagattgctggagga
73	IWB34807	IAAV3960	4A	123.0	A	G	aaaccgctttctggaagagT	aaaccgctttctggaagagC	cgCCgccGctaatttacaA
74	IWB64511	RFL_Contig3841_2595	4A	137.0	A	G	gctttagctttagcAccaGaT	gctttagctttagcAccaGaC	gtGttcTgGagcttgcagA
75	IWB66655	Tdurum_contig10654_704	4A	147.2	T	C	gatcAaacacctccccAccaaT	gatcAaacacctccccAccaaC	cTtGtattCCTGgCCGTTCTttGT
76	IWB65772	TA004020-0357	4A	164.1	A	G	gcaggctaccatttgtgcA	gcaggctaccatttgtgcG	cctgtagcaatgaaacggcg
77	IWB931	BobWhite_c1656_186	4B	24.7	T	C	tttgctgtccagatcctcatataT	tttgctgtccagatcctcatataC	tgcaaattgttattagttgctccC
78	IWA3290	wsnp_Ex_c30695_39579408	4B	44.6	T	C	CAGGTTTCTGATACAATTTCGCTGCA	AGGTTTCTGATACAATTTCGCTGCG	CGTGCTCTGCTGCTGTGAATAGATT
79	IWB52254	Ra_c5508_706	4B	80.9	T	C	gcaaccgcatctcaagctaT	gcaaccgcatctcaagctaC	aattggtatcagtagccatcga
80	IWB54964	RAC875_c202_474	4B	95.6	A	G	tgtagaagaagattccccggaT	tgtagaagaagattccccggaC	aagaCGcactcctcacTacG
81	IWB3246	BobWhite_c4256_213	4B	104.1	A	G	tgtgtcctaactcgctcaatagatT	tgtgtcctaactcgctcaatagatC	tgccacccgaaatgtcaact
82	IWB58397	RAC875_c48025_483	4B	114.9	T	C	TAgaccgTcctggactcgT	TAgaccgTcctggactcgC	gctagatggaccaatccctC
83	IWB12054	BS00099053_51	4D	21.3	T	C	GgaAtccaaTggAcatcactacatT	GgaAtccaaTggAcatcactacatC	TcaGTccctccttgtcgtcT
84	IWB61486	RAC875_rep_c105718_304	4D	69.2	T	C	gctgcattagggttctattgctaT	gctgcattagggttctattgctaC	catgtggactctgcaccctt
85	IWB8050	BS00036421_51	4D	74.0	T	C	tcaccagctcccctcccT	tcaccagctcccctcccC	gatgcttgTtcacctaaatccC
86	IWB30317	Excalibur_rep_c108030_260	4D	83.8	T	C	cccatgAtCagCgacaCgA	cccatgAtCagCgacaCgG	GttgGtacatGacgtgaaggC
87	IWB10207	BS00068013_51	4D	92.8	A	G	tgttcattcatgaagtgtgagtcT	tgttcattcatgaagtgtgagtcC	gctTatataacacaaaaacctcacG
88	IWB28897	Excalibur_c79009_131	4D	105.1	A	G	tgaggaaggatcttatcaagctcA	tgaggaaggatcttatcaagctcG	acggtcgacatttcacagct
89	IWA3566	wsnp_Ex_c3620_6612231	5A	15.9	T	C	AATGTTTCTCCAAGAGACGTTGAGTTT	GTTTCTCCAAGAGACGTTGAGTTC	GCTTGAAAGGGTCCTCCTTGATGTA
90	IWB21456	Ex_c95453_1499	5A	26.5	A	G	atcgacaattacatcAaatgactgA	atcgacaattacatcAaatgactgG	aataacgtggctatcagtggT
91	IWA8154	wsnp_Ra_rep_c69221_66574148	5A	42.0	A	C	ATATGCCGCAGGCTTAGGTAACT	ATGCCGCAGGCTTAGGTAACG	AAGAAACTCTGCTGYGCTATAGCTGAT
92	IWA300	wsnp_BE497820A_Ta_2_2	5A	50.4	T	C	cCAcAatgacaaaaGgagccT	cCAcAatgacaaaaGgagccC	aatttcagattctgtttgctttgaC
93	IWB4836	BobWhite_rep_c50013_65	5A	127.6	T	C	ggaaggtcgtctgcgtcaT	ggaaggtcgtctgcgtcaC	gcgatcgagaaggagcttga
94	IWB69078	Tdurum_contig25432_1218	5B	19.7	T	C	cttgaaggttcaatctgtggaatT	cttgaaggttcaatctgtggaatC	CTCtCTGCTCCCTtTCtTtGtA
95	IWA4378	wsnp_Ex_c5915_10378807	5B	38.5	T	C	ATTATGAAATGACATCGGGCCCTCA	ATGAAATGACATCGGGCCCTCG	GGCAAGGTCGCCACGGACATTT
96	IWB36619	Jagger_c505_141	5B	49.0	A	G	tctattgtgtgatgtgatgctcaT	tctattgtgtgatgtgatgctcaC	ggcatctttcaagcatatcctagT
97	IWB45559	Kukri_c45713_151	5B	61.9	A	G	gctgcctcctcgtcaattgT	gctgcctcctcgtcaattgC	atacaagggctccagggagg
98	IWA1780	wsnp_Ex_c13485_21225504	5B	97.3	T	C	AGACGCAGAGCATACTGGGGT	GACGCAGAGCATACTGGGGC	GGACAAGCAAATGGACCTGAGTTCTT
99	IWB69931	Tdurum_contig29967_456	5B	120.9	T	G	ctaatcagccgccagggA	ctaatcagccgccagggC	cttgcaagttgtttaattagcacat
100	IWB9238	BS00064272_51	5B	139.4	A	G	tcaaatgggcaagtaggaagtaT	tcaaatgggcaagtaggaagtaC	cgttgatcacttctctttgcca
101	IWA7903	wsnp_Ra_c38873_46699852	5B	154.1	T	C	AAGATGCCGCTTGTTCTCAGTGAT	GATGCCGCTTGTTCTCAGTGAC	CCGGCTAAATATCAACCAACATACTAGTA
102	IWB1968	BobWhite_c26082_239	5B	170.5	A	G	aattcacgaggggcacaT	aattcacgaggggcacaC	gtgtctgatgatatcgccactT
103	IWB74021	Tdurum_contig97942_51	5B	180.7	A	G	cataaccatcgccgacGtcaaA	cataaccatcgccgacGtcaaG	aggcatacaatcagtttcaagagtC
104	IWB61615	RAC875_rep_c106589_784	5B	212.4	T	C	aaatctagatggttgtgagcctaT	aaatctagatggttgtgagcctaC	tggccaagactacAtataagcttT

No	SNP ID	Name SNP	Chr	Pos*	FAM allele	VIC allele	FAM Primer	VIC Primer	Common Primer
105	IWB29164	Excalibur_c86388_413	5B	217.6	A	G	tgcgcttcaccaagtaatcA	tgcgcttcaccaagtaatcG	attgtcctgattggggattgt
106	IWB11256	BS00082423_51	5D	0.0	T	C	gttccGttgtacaGTtcctcA	gttccGttgtacaGTtcctcG	acttaaccatggcattgtgttT
107	IWA3429	wsnp_Ex_c33327_41834973	5D	58.9	T	C	GTCTACATCAGAATGGTGTTGAGCA	CTACATCAGAATGGTGTTGAGCG	CCTTCCAGTGTGCTATTGCTGATGTT
108	IWB67782	Tdurum_contig12926_687	5D	70.1	A	G	tGctcaAaggaaggtttaaaggaT	tGctcaAaggaaggtttaaaggaC	AAGtTTGcCTcgGTaAgctG
109	IWA3015	wsnp_Ex_c2598_4832869	5D	96.2	A	G	gcaaaaagagCgccgTtttT	gcaaaaagagCgccgTtttC	ggttcctttcctttctctgtGtT
110	IWA1681	wsnp_Ex_c1278_2449191	5D	116.6	T	C	ctccgtcccataatAtgagacaT	ctccgtcccataatAtgagacaC	ctgctatcattcatcGatatggttT
111	IWA4087	wsnp_Ex_c508_1008029	5D	125.7	A	G	CATTCAAGCGGGCTGACCAACT	CAAGCGGGCTGACCAACC	CCCGTGCCTCGCCTTGGCAA
112	IWB49479	Kukri_rep_c110911_477	5D	152.5	A	G	gactctacagagctcttgtatgTaT	gactctacagagctcttgtatgTaC	acctccttacaagcagctattG
113	IWB54292	RAC875_c16419_585	5D	204.6	T	G	tcataaatgattgcttgcggaaaT	tcataaatgattgcttgcggaaaG	ctcgcatgtaggccttgcat
114	IWB23519	Excalibur_c20597_509	6A	17.0	A	G	AgtcggtaaagcggcGtT	AgtcggtaaagcggcGtC	cctcCgtcagctcctggtaC
115	IWB1550	BobWhite_c22086_444	6A	25.5	A	G	gtgaaaggagctgttcacagT	gtgaaaggagctgttcacagC	tggaggggatgtctgtccT
116	IWB36506	Jagger_c2853_75	6A	63.7	T	G	gtactgTctcagaagtagcttaagT	gtactgTctcagaagtagcttaagG	cctagctctctgggactcatC
117	IWB10758	BS00074752_51	6A	82.4	T	G	tccattaccctgcctacGA	tccattaccctgcctacGC	gcccttcactgagccggT
118	IWB34395	IAAV151	6A	94.9	T	C	gtagaggagacGaggctgT	gtagaggagacGaggctgC	GgaagcagtcaaaattttccttcG
119	IWB11953	BS00096240_51	6A	130.7	T	C	accTcgtggaacccgAcA	accTcgtggaacccgAcG	cctaCtatgtctgatgaGcgG
120	IWB20412	Ex_c28973_935	6A	140.7	A	G	ccatatggagaacttcagtgttacT	ccatatggagaacttcagtgttacC	aggtggacataTgtagtaagctaaT
121	IWB11445	BS00086173_51	6A	159.6	T	C	cgtggtcaagaagttgtgcaT	cgtggtcaagaagttgtgcaC	agcgcgtgatcatggagT
122	IWB71432	Tdurum_contig43538_1306	6B	5.0	A	G	ggaatcgggataatctttccctgA	ggaatcgggataatctttccctgG	tgcatatcgagggacggga
123	IWB60019	RAC875_c66376_395	6B	36.7	A	G	caagctttcagggtagacgT	caagctttcagggtagacgC	gggaggagatactgcggaag
124	IWB60404	RAC875_c76124_264	6B	53.8	T	C	gggctgaacagttatacatgatggT	gggctgaacagttatacatgatggC	ttcagtaagttcatttatctgccCG
125	IWA6770	wsnp_Ku_c24981_34948114	6B	67.2	A	G	TGACTGATGCATTGGGCAATCAA	GACTGATGCATTGGGCAATCAG	CCATAAGAAAATGTCCTTTGAAGCAGCAA
126	IWB46771	Kukri_c58961_76	6B	83.0	A	G	agtagcgtgccaaggccA	agtagcgtgccaaggccG	cgacatcacctacatgggca
127	IWB39172	Ku_c32100_105	6B	91.5	A	G	tggataagcgttgaagacatgtA	tggataagcgttgaagacatgtG	gcacgtaactgcaagtgtcg
128	IWB45581	Kukri_c45876_157	6B	120.6	A	G	cagtcatgggccctcacA	cagtcatgggccctcacG	tagacgctctgctctggaca
129	IWB65608	TA002853-0110-w	6D	17.0	A	G	ttttcagttgaacttgctgaagT	ttttcagttgaacttgctgaagC	cttgtcggtggtgctcgT
130	IWB72406	Tdurum_contig54917_597	6D	39.9	A	G	catgctcgaccagtccctT	catgctcgaccagtccctC	cactgcttggtctcgttCaG
131	IWB55904	RAC875_c26177_632	6D	54.3	T	C	agtaaatatggggagctaacacA	agtaaatatggggagctaacacG	aagtactacggccccggt
132	IWB11946	BS00095826_51	7A	33.2	A	G	accgtgctacccaccacT	accgtgctacccaccacC	TgaTgagttgcgGtcgtcG
133	IWB44621	Kukri_c36885_58	7A	45.4	A	G	catgtttcggttgattagctcattA	catgtttcggttgattagctcattG	tgttgAggatGttggAataGgT
134	IWB43661	Kukri_c29154_977	7A	52.4	A	G	gggacggggttcctcttcA	gggacggggttcctcttcG	ccctgatagtgggcaagtga
135	IWB58667	RAC875_c50665_202	7A	78.0	T	C	gtagtcgaaaacCcgatgatagT	gtagtcgaaaacCcgatgatagC	tccatgtctcatccatcatcaaG
136	IWB7554	BS00024786_51	7A	110.2	T	C	tgctcCccattcAgtccT	tgctcCccattcAgtccC	tctatgtaggacggattagaagttG
137	IWA5645	wsnp_Ex_rep_c69838_68799256	7A	127.8	T	C	TTTTCTGATTCTGGGCTAGGATTTCA	TCTGATTCTGGGCTAGGATTTCG	CCTGCCTGCATTGCTCTTTACCTTA
138	IWB9062	BS00063458_51	7A	136.4	A	G	tccacTtggacggcagcA	tccacTtggacggcagcG	gtcctccaacacgtagtacG
139	IWA7046	wsnp_Ku_c42539_50247426	7A	152.3	A	G	tgcgcaagatgaaggaacaA	tgcgcaagatgaaggaacaG	attcagcgggccttcagttt
140	IWA2724	wsnp_Ex_c22547_31738007	7A	185.9	T	C	ccttgggagctatgtcGgaA	ccttgggagctatgtcGgaG	tgaaagtgaaaatactattgcTggT
141	IWB30143	Excalibur_rep_c105674_315	7A	212.7	T	C	cctcgtcgtcaccatcaaT	cctcgtcgtcaccatcaaC	aGgaagaaGatgatgaggtaggAT
142	IWB74437	tplb0031i24_1212	7A	221.4	T	G	ttttgctgctgttattcacctT	ttttgctgctgttattcacctG	cccagaactgaaacatctccGaT
143	IWB27367	Excalibur_c53111_215	7B	3.3	A	G	tcCccaGacatcataagttcatT	tcCccaGacatcataagttcatC	aatgcctttcaacacagccA
144	IWA2568.1	wsnp_Ex_c2103_3947695	7B	50.4	A	G	agccaggcaaCgagaacA	agccaggcaaCgagaacG	gcaCtggtggagtAgactgC
145	IWB50943	Ra_c11468_305	7B	68.3	T	C	gctatctcctgcaaaagaatggtA	gctatctcctgcaaaagaatggtG	gcattgttcGatgggcgTaT
146	IWB41479	Kukri_c15912_1189	7B	112.4	A	G	cctcagcaaatggaagtgtacA	cctcagcaaatggaagtgtacG	ccattgacaagaactccgttgT
147	IWB34893	IAAV4582	7B	145.3	T	C	tctagcttctgataaagtgtgagT	tctagcttctgataaagtgtgagC	gctctgagggaccttcaGA
148	IWB9811	BS00066456_51	7B	163.9	A	G	gttgcaatgcttgtggccaA	gttgcaatgcttgtggccaG	gccccagacctgCgttataa
149	IWB3312	BobWhite_c43557_103	7B	171.1	A	G	gCaacaaattaagcttTtcCccA	gCaacaaattaagcttTtcCccG	caagtggaaacagtgcaggc
150	IWB35592	IAAV9104	7D	22.8	A	C	gggatgactgaccatggatgttaA	gggatgactgaccatggatgttaC	ctacaagtGcctgcCtagagaataT
151	IWB18914	D_GCE8AKX02ILA1U_88	7D	88.6	T	G	ggcatgagCagagaCcTGttA	ggcatgagCagagaCcTGttC	gccagGgcatgcttgacT
152	IWB23802	Excalibur_c22419_460	7D	93.0	T	C	ccaaagttgttcttcatcagatcA	ccaaagttgttcttcatcagatcG	tgatgacagggcatgccG
153	IWB13142	CAP11_rep_c8279_82	7D	112.4	A	G	cagcagttgccattggagttatA	cagcagttgccattggagttatG	tggtcgGaaggtattggtga
154	IWB58968	RAC875_c53629_483	7D	124.5	A	G	gccaacaaagaaccgaacTgT	gccaacaaagaaccgaacTgC	cagtacgatgctgtccgC
155	IWB34836	IAAV4133	7D	135.4	T	G	tctgcTtttggacatcaCgA	tctgcTtttggacatcaCgC	gcatgccAgTtcgtgagttG

* Map positions see Wang et al. 2014

Appendix 6: Useful software packages

1. Software packages for diversity and population genetic structure analysis

ARLEQUIN (http://cmpg.unibe.ch/software/arlequin35/Arl35Downloads.html)
Performs a wide variety of tests of inter and intra genetic diversity including e.g., hierarchical analysis of variance. It has import feature for GENEPOP files. However, it might produce spurious results if there are too many missing data.

CurlyWhirly (https://ics.hutton.ac.uk/curlywhirly/download-curlywhirly/)
User-friendly freely available software for visualization of 3D ordination data. It loads any kind of numerical data and any number of variables and highlights data points according to category to selectively view data from one or more categories. Also displays data labels moving the mouse over any data dot.

DARWin (http://darwin.cirad.fr/product.php)
Free package for diversity analysis. It generates trees and PCA plots like the NTSYS program. However, it contains more options for generating distance matrices using different coefficients and clustering methods including Ward's method, Single linkage, UPGMA, NJ etc.

Flapjack (https://ics.hutton.ac.uk/flapjack/download-flapjack/)
User-friendly freely available software for the visualization of maps, genotypes, traits, and QTL. It can handle data sets approaching 1 billion genotypes in size. Other important features are: individual alleles colored by state, frequency or similarity to a given standard line, sorting by genotype similarity to other lines, or by trait scores, map based information e.g. QTL positions can be aligned against graphical genotypes.

GENEPOP (http://genepop.curtin.edu.au/)
Performs exact tests for deviation from Hardy-Weinberg, linkage disequilibrium, population differentiation and isolation by distance (DOS). DNASP: performs population genetic analyses of sequence data, including tests for selection

GENAlEx (http://biology-assets.anu.edu.au/GenAlEx/Download.html)
Excel Add-In for the analysis of genetic data. It is particularly useful for dominant data such as RAPD and AFLP data.

NTSYS (commercial package)
A basic commercial package for diversity analysis. It constructs dendrograms/trees using distance matrices and includes UPGMA and neighbor joining (NJ) methods. It also performs principal component analysis (PCA) and Mantel test (to check correlation between different data sets).

POPGENE (https://www.ualberta.ca/~fyeh/popgene_download.html)
Performs analyses genetic variation among and within populations using co-dominant and dominant markers, and quantitative data.

Powermarker (http://statgen.ncsu.edu/powermarker/)
Handles a variety of marker data, either haplotypic or diplotypic. It can calculate a wide variety of diversity parameters such as allele number, gene diversity, inbreeding coefficient, estimation of allelic, genotypic and haplotypic frequency, Hardy-Weinberg disequilibrium and linkage disequilibrium

STRUCTURE (http://pritchardlab.stanford.edu/structure_software/release_versions/v2.3.4/html/structure.html)
A software that uses a Bayesian clustering method to identify population structure and assigns individuals to predefined (K) populations.

2. Software packages for QTL mapping

ICImapping (https://www.integratedbreeding.net/386/breeding-services/more-software-tools/icimapping)
A user-friendly freely available public software capable of building high-density linkage maps and mapping QTLs in bi-parental populations. Eight functionalities are integrated in this software package: (1) BIN: binning of redundant markers; (2) MAP: construction of linkage maps in biparental populations; (3) CMP: consensus map construction from multiple linkage maps sharing common markers; (4) SDL: mapping of segregation distortion loci; (5) BIP: mapping of additive, dominant, and digenic epistasis genes; (6) MET: QTL-by-environment interaction analysis; (7) CSL: mapping of additive and digenic epistasis genes with chromosome segment substitution lines; and (8) NAM: QTL mapping in NAM populations.

MapMaker/QTL (ftp://genome.wi.mit.edu/pub/mapmaker3/)
A user-friendly, freely distributed software program, and runs on almost all platforms. It analyzes F_2 or backcross data using standard interval mapping.

MQTL (request to Nick Tinker at nick.tinker@agr.gc.ca)
A computer program for composite interval mapping in multiple environments. It can also perform simple interval mapping. Currently, MQTL is restricted to the analysis of data from homozygous progeny (double haploids, or recombinant inbred lines). Progeny types with more than two marker classes (e.g., F_2) are not handled.

MapQTL (http://www.cpro.dlo.nl/cbw/)
A licensed software program. It performs Kruskal-Wallis test (single marker analysis), composite interval mapping and multiple interval mapping on almost all kind of mapping populations.

PLABQTL (http://www.uni-hohenheim.de/~ipspwww/soft.html)
A freely distributed computer program for composite interval mapping and simple interval mapping of QTL. Its main purpose is to localize and characterize QTL in mapping populations derived from a bi-parental cross by selfing or production of double haploids. Currently, this program is the easiest software for composite interval mapping.

QTL Cartographer (http://statgen.mcsu.edu/qtlcart/cartographer.html)
A QTL software written for either UNIX, Macintosh, or Windows. It performs single-marker regression, interval mapping, and composite interval mapping. It permits analysis from F_2 or backcross populations. It displays map positions of QTL using the GNUPLOT software.

Qgene (http://www.qgene.org/qgene/download.php)
A QTL mapping and marker-aided breeding package written for Macintosh. It has a user-friendly graphical interface and produces graphical outputs. QTL mapping is conducted by either single-marker regression or interval regression.

SAS (commercial package)
A general statistical analysis software. It can detect QTL by identifying associations between marker genotype and quantitative trait phenotype by single marker analysis approach such as ANOVA, t-test, GLM or REG.

3. Software packages for association mapping

ASREML, JMP Genomics, SAS and GenStat (commercial packages)
Commercial software packages. ASREML and JMP Genomics are specifically engineered for genetic analysis and can handle more complex models, whereas general purpose packages such as SAS Proc Mixed and GenStat can perform association analysis but require more expertise and programming on the part of the user. Also, a large data set takes too much time.

EMMA (http://mouse.cs.ucla.edu/emma/)
Freely available software suitable for association analysis using mixed models. It is very useful for handling large datasets. However, it requires familiarity with R as all the functions for data management and visualization are command based.

TASSEL (http://www.maizegenetics.net/#!tassel/c17q9)
Freely available software, can handle both general linear model- (GLM) and mixed linear model (MLM)-based association analysis. Visualization functions are efficient and user friendly. In the latest version of TASSEL (TASSEL 5.0), compressed MLM method is available for computing large datasets with upto 500,000 markers.

RRBLUP (use install packages in R)
Freely available R-based package which can perform association analysis using mixed models. It is basically designed for genomic prediction.

Appendix 7: Useful weblinks

CerealsDB (http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/indexNEW.php)
A site created by the University of Bristol which provides a range of information to study wheat genome. Principally the site provides information about SNP markers; e.g., primer sequences, haplotypes of UK and some CIMMYT varieties. The information regarding SNP markers is divided by platform: SNPs (KASP), SNPs (Axiom®), SNPs (iSelect) and SNPs (Taqman®).

- **KASP:** Includes a list of KASP assays linked to agronomic traits useful for MAS: http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/kasp_download.php?URL=Traits include e.g. disease resistance, quality traits, traits related to plant development such as vernalization requirement, photoperiod and plant height.In addition, > 7000 validated KASP assays across the genome are given, of which > 5000 are mapped.
- **AXIOM® 820K and 35K SNP Arrays:** The Affymetrix code, Bristol SNP code and probe sequence containing the SNP ambiguity code for '820,000' and '35,000' wheat SNP array can be downloaded from the following link http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/axiom_download.php
  The '35K' array is a more manageable, cheaper and highly efficient system for screening large numbers of lines. This array is available in 96 and 384 plate-based formats and consists of a subset of the 820K array with probes selected on the basis of a combination of their map location, their PIC value and their usefulness in a range of germplasm.
- **TaqMan® probes:** In partnership with Life Technologies, a collection of 4,800 TaqMan® SNP Assays are available, of which 4,735 are mapped. These assays come as single tube, pre-formulated primer and probe sets which can be used right out of the box requiring no optimization or laborious setup.

Other portals in CerealsDB:
- **Wheat EST database:** contains a collection of over 25,000 well annotated ESTs.
- **BLAST Wheat genomic sequence:** contains a number of BLAST pages, e.g., the sequence data of wheat variety Chinese Spring (5x genome coverage). The draft assembly of the gene-rich regions of the genome, or the raw sequence reads can be searched.
- **Genomics Facility:** This link takes you to the webpage of Genomics facility at University of Bristol. It is a core service laboratory equipped with various services.

GRAMENE (http://www.gramene.org/)
A curated, open-source, integrated data resource for comparative functional genomics in crops and model plant species. Currently, it hosts annotated whole genomes of over two dozen plant species and partial assemblies for almost a dozen wild rice species in the Ensembl browser, genetic and physical maps with genes, ESTs and QTLs locations, genetic diversity data sets, structure-function analysis of proteins, plant pathways databases (BioCyc and Plant Reactome platforms), and descriptions of phenotypic traits and mutations. Comparative maps of rice, maize, sorghum, barley, wheat and oat are anchored by a set of curated correspondences.

Important portals in GRAMENE:
- **Genome Browser:** Browse gene annotations & diversity data
- **BLAST:** Align DNA & protein sequences
- **Plant Reactome:** Browse metabolic & regulatory pathways
- **Pathways databases:** BioCyc based cellular metabolic networks for 10 plant species
- **Gramene Mart:** Customized data queries
- **Bulk downloads**
- **ARCHIVE:** Markers, Proteins and Ontology databases, QTLs, Comparative Maps
- **SSRIT (SSR Identification Tool):** Searches for SSRs in submitted sequences

GrainGenes (http://wheat.pw.usda.gov/GG3/)

Graingenes includes access to the microsatellites, STS and DArT markers, maps, mapped genes, gene sequences and QTLs of various cereal crops and model species. Like CerealsDB, it also provides a platform for comparative mapping among grasses. Besides, this is a very good web resource for accessing general information about traits, pathology and also published articles in various sections such as Reviews, Monographs, Journals and Newsletters, Reports and Proceedings, Mapping Publications, Agronomy Publications and Pathology Publications.

Important portals in GrainGenes:

- **Browse GrainGenes:** Browse GrainGenes data assembled into 29 classes; Allele, Assembly, Author, Colleague, Collection, Gene, Gene Class, Gene Product, Germplasm, Image, Journal, Keyword, Library, Locus, Map, Map Data, Marker, Pathology, Polymorphism, Probe, Protein, QTL, Rearrangement, Reference, Sequence, Species, Trait, Trait Study, Two Point Data.
- **Quick Queries:** Allows to retrieve marker/map/gene/QTLs/germplasm information using various options.
- **Advanced Queries:** contains several advanced modes for searching the MySQL database.
- **BLAST:** Allows you to Blast Mapped Wheat ESTs, Contigs containing Mapped Wheat ESTs, Poaceae EST-SSRs, TREP Triticeae Repeats, Barley 1 Gene Chip, Consensus Sequences from Wheat SNPs Assembly Dec 02 and All GrainGenes Sequences.
- **CMap:** Draws and displays comparative maps of different cereal species.
- **Gbrowse:** Graphical display of genes, gene-markers, Bins, Binlocus.

IWGSC (http://www.wheatgenome.org/)

The site contains a variety of tools and sequence resources available to the entire wheat scientific community. On behalf of the IWGSC, a central repository for access to physical map data and sequences has been established by the URGI at the following link http://wheat-urgi.versailles.inra.fr/Seq-Repository/.

Important portals in IWGSC:

- **BLAST:** Browse the *Triticum aestivum* survey sequence assemblies, the 3B reference sequence (whole chromosome and CDS only), 454 assemblies and other wheat species assemblies.
- **Genes and annotations:** contains downloadable data on Gene models, Genome Zipper, POPSEQ and 1AS sequence model.
- **Physical maps:** Display the physical maps using the Physical map viewer tool.
- **Transcriptome:** Search transcriptome data from diploid (*T. urartu*) and tetraploid wheat (*T. turgidum* ssp. *durum* cultivar Kronos).
- **RNA-Seq:** contains downloadable RNA-Seq data from two libraries: i) non-oriented library (TruSeq, Illumina) sequenced on Illumina HiSeq2000 2x100bp (PE) for 15 different conditions corresponding to five wheat organs (root, leaf, stem, spike, grain) at three developmental stages each in duplicates and (ii) oriented library (ScriptSeq, Epicentre) sequenced on Illumina HiSeq2000 1x100bp (SE) for five conditions corresponding to five wheat organs (root, leaf, stem, spike, grain) without duplicates.
- **Variations:** contains downloadable VCF files containing sequence data of 62 diverse wheat lines re-sequenced using the whole exome capture (WEC) and genotyping-by-sequencing (GBS) approaches. It also has a link to wheat hapmap.

MASWheat (http://maswheat.ucdavis.edu/)

This site contains abundant information on gene-based markers available for MAS in wheat, but also educational, outreach materials and laboratory methods. Under the link “MAS protocols” gene-based STS and SNP marker information has been divided by various categories: e.g., marker protocol related to genes for Fungi resistance (rust, powdery mildew, fusarium head blight, and septoria blotch), virus and insect resistance, quality genes, and abiotic stress tolerance.

Integrated Breeding Platform (https://www.integratedbreeding.net/)

This platform provides a suite of interconnected software specifically designed to help breeders anywhere from emerging national programmes to well-established companies manage their day-to-day activities through all

phases of their breeding programmes. From straightforward phenotyping to complex genotyping, it provides all the tools needed to conduct modern breeding in one comprehensive package. The IBP also provides training opportunities, responsive technical support and community space for meaningful exchanges with peers and other experts.

The important resources available on this page to optimise your plant breeding programme are:

- downloadable, comprehensive software tools: the **Breeding Management System (BMS)** and more **tools from our partners**;
- a network of accessible and reliable **breeding service providers**;
- a **resource library** with products and information for over 10 crops, including **diagnostic markers** and **trait dictionaries**;
- **training material and activities** for an optimal use of our technology as well as for integrating good breeding practises;
- support through **peer communities** and dedicated **technical assistance**.

dbSNP (http://www.ncbi.nlm.nih.gov/SNP/)

The Single Nucleotide Polymorphism database (dbSNP) is a public-domain archive for a broad collection of simple genetic polymorphisms. This collection of polymorphisms includes single-base nucleotide substitutions (also known as single nucleotide polymorphisms or SNPs), small-scale multi-base deletions or insertions (also called deletion insertion polymorphisms or DIPs), and retroposable element insertions and microsatellite repeat variations (also called short tandem repeats or STRs). The dbSNP has been designed to support submissions and research into a broad range of biological problems including physical mapping, functional analysis, pharmacogenomics, association studies, and evolutionary studies. It accepts submissions for variations in any species and from any part of a genome.

The SNP database can be queried from the dbSNP homepage, by using Entrez SNP, or by using the links to the six basic dbSNP search options as mentioned below.

1. Search by IDs (Single record query: Accession, ID or Cluster).
2. Submission information (Submission property query: method, paper, submitter, latest data).
3. Batch (Batch query: retrieve upto 20,000 records of interest at a time).
4. Locus information (Locus query: retrieve lists of variations in known gene regions or mRNA transcripts).
5. Free form (Easy form queries).
6. Between markers (Positional query: query the database for variations bounded by STS markers).

***e!*EnsemblPlants** (http://plants.ensembl.org/index.html)

EnsemblPlants is developed in coordination with other plant genomics and bioinformatics groups. It harbours genomic sequence, sequences of genes and transcripts and protein model predictions of about 39 plant species. The bread wheat genome in Ensembl Plants is the Chromosome Survey Sequence (CSS) for *Triticum aestivum* cv. Chinese Spring, combined with the reference sequence of chromosome 3B, both generated by the International Wheat Genome Sequencing Consortium. The CSS assemblies have been further refined into chromosomal pseudomolecules using POPSEQ data. The important tools available on this site are as under:

- Variant effect predictor (VEP): VEP tools allows to analyse your own variants and predict the functional consequences of known and unknown variants.
- BLAST/BLAT: allows to search genomes in e!EnsemblPlants for your DNA or protein sequence.
- BioMart: a data-mining tool to export custom datasets from Ensembl.
- Assembly converter: allows to map your data's coordinates to the current assembly.

Further, downloadable data on the FTP site are available as:

- FASTA formatted sequences of masked and unmasked genomic sequences associated with the assembly (contigs, chromosomes etc.), coding sequences (CDS), cDNA sequences and protein sequences.

- Annotated sequences annotated by the automated Ensembl genome annotation pipeline. Each nucleotide sequence record in a flat file represents a 1Mb slice of the genome sequence. Flat files are broken into chunks of 1000 sequence records for easier downloading.
- MySQL: All Ensembl MySQL databases are available in text format as are the SQL table definition files. These can be imported into any SQL database for a local installation of a mirror site.
- GTF: contains gene sets for each species. These files include annotations of both coding and non-coding genes.

Wheat genotyping service providers

Kansas State University (KSU)

Wheat 9K and 90K iSelect assays have been developed for low-cost high-throughput genotyping of wheat germplasm. The design of the wheat 9K and 90K iSelect assay and a consensus genetic map are available for download at http://wheatgenomics.plantpath.ksu.edu/

- Total number of markers on 90K iSelect array: **81,587**
- Markers that have been mapped: **43,999**

KSU also provides genotyping service by GBS markers. The facility accepts a minimum of 48 ready-to-load samples in 96- or 384-well ABI-compatible plates.

To submit samples:

- Fill out the **sample submission form**, and print.
- For billing purposes, we require an Interdepartmental Requisition Form (IDR) or Purchase Order (PO) number.
- Bring the sample submission form and IDR/PO to 3304 Throckmorton Hall along with your sample plate.
- Please send us your plate record by e-mail, **dnaseq@ k-state.edu**.
- Raw data will be e-mailed to you after the run.

TraitGenetics (http://www.traitgenetics.com/en/index.php?option=com_content&task=view&id=20&Itemid=46)

Provides efficient genotyping service for SSR and SNP markers in the range of hundred thousand SSRs to millions SNP data points/week. Beside its experience in marker development and analysis, TraitGenetics has also experience in a variety of fields that are associated with the use of molecular markers for specific applications such as:

- Variety identification and variety infringement
- Germplasm identification and protection
- Genetic distance analysis
- Marker-trait associations
- Marker-assisted selection
- Marker-assisted backcrossing
- Identification of markers within specific chromosomal regions
- Marker development from BAC clones
- Candidate gene analysis in specific germplasm
- Identification of mutations in specific genes
- Adaptation of markers to specific applications
- Generation of high density genetic maps
- Map-based cloning projects
- Consultations regarding optimal marker use in breeding applications

DArT Pty. Ltd (http://www.diversityarrays.com/dart-application-dartseq)

Provides a range of products and services including genotyping service for high throughput DArT and DArTSeq markers for a range of crops. Other important service includes development of DArT arrays for new species. For

DArT assays, 500 - 1000 ng of restriction enzyme grade DNA, resuspended in aqueous solution such as TE at a concentration of 50 - 100 ng/microliter should be submitted.

LGCGenomics (http://www.lgcgroup.com/services/#.VXXXjVXbKM8)
Important services provided by LGC include DNA sequencing, Genotyping services (SNP markers), SNP marker design, whole genome amplification, DNA/RNA extraction and next generation sequencing among others.

Bristol Genomics Facility (http://www.bristol.ac.uk/biology/research/transcriptomics/)

Next generation sequencing
The Bristol Genomics Facility offers a range of services for next generation sequencing and library preparation for certain applications. The sequencing service has been operating for more than 3 years using Illumina sequencing-by-synthesis and now includes Ion Torrent sequencing on the Life Technologies Ion PGM™ and Ion Proton™ sequencers. All platforms utilize a massively parallel sequencing approach to generate billions of high quality bases per run and are suitable for a broad range of applications including *De novo* sequencing, Re-sequencing, Stranded RNA-seq, RNA-seq, small RNA sequencing, Targete re-sequencing, ChIP-seq, Methylation-seq and MeDIP-seq.

Affymetrix Microarray
The facility currently operates an Affymetrix GeneChip® Array platform consisting of the 4-colour 7G upgrade enabling to process the highest resolution 5µm arrays available from Affymetrix. Soon the new Affymetrix GeneTitan® system will be offered for both Axiom genotyping and gene expression arrays.

Capillary sequencing and fragment analysis
The Bristol Genomics Facility also houses an Applied Biosystems 3500 Genetic Analyser, an 8-capillary system for DNA sequencing and fragment analysis applications.
Sample submission:
- Completed sequencing reactions should be provided dried down following ethanol precipitation.
- Completed fragment analysis reactions should be provided pre-diluted (1:10 is usually sufficient) in 8-strip tubes or a 96-well plate.

Real-time qPCR
The Bristol Genomics Facility houses a 96-well LightCycler® LC480 Real-Time PCR system for medium- and high-throughput applications in gene expression and genotyping analysis. They offer a 'ready-to-go' service for internal researchers at Bristol University and external researchers in which the investigator can book time on the instrument to run pre-prepared plates if familiar with the instrument software, or can provide the facility with plates to run and deliver the data via email, typically within 48 hours.

The instrument is most commonly used for the following applications:
- Gene Detection
  - Absolute quantification
  - Relative quantification
- Genetic Variation
  - Detection unknown variants
  - Detecting known variants

Illumina Genotyping Service-Qiagen (https://www.qiagen.com/us/shop/assay-technologies/illumina-genotyping-service/)
The Illumina Genotyping Service provides a complete service for genomic analysis (whole genome SNP genotyping, focused panel genotyping, or copy number variation analysis) that delivers robust and reproducible results. The service accepts and processes a variety of biological samples and then leverages cutting-edge tools for either pathway-focused or genome-wide analysis yielding superior results for scientists in academic, government, and industrial settings. QIAGEN is an Illumina-certified service provider.

ISBN: 978-607-8263-57-8

CIMMYT MR
International Maize and Wheat Improvement Center