MOLECULAR DIVERSITY, LINKAGE DISEQUILIBRIUM AND GENETIC MAPPING IN EAST AFRICA WHEAT

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNVERSITY OF MINNESOTA
BY

GODWIN KAMAU MACHARIA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

JAMES A. ANDERSON, ADVISER

April 2013

Acknowledgements

Foremost, I wish to thank my adviser, Dr Jim Anderson for providing me an opportunity to work with him in his program. His mentorship, encouragement and unconditional support are greatly appreciated.

I'm indebted to my research committee members, Dr Peter Morell, Dr Rex Bernado, Dr Brian Steffenson, Dr Yue Jin, and Dr Matt Rouse, who through careful review and insightful suggestions helped me accomplish my research objectives and prepare this dissertation.

Next, I would like to thank staff and faculty at the Agronomy and Plant genetics department for their invaluable support. Similar thanks are expressed to members of the wheat project, Brian Seda, Prabin Bajgain, Kathryn Turner, Emily Conley, Catherine Springer, Jen Flor, Susan Reynolds, Roger Caspers, Gary Linkert, Dr Xiaofei Zhang, Dr Liangliang Gao and all undergraduates who were members of the project during my tenure. They all provided friendship, assistance, and encouragement in many aspects of my research work.

I'm equally indebted to my collaborators Dr Shiaoman Chao, at USDA-ARS genotyping Lab in Fargo, North Dakota, Dr Ravi Singh, and Dr Javier R. Pena at CIMMYT, Mexico, Dr Senjay Simsek at the Wheat Quality and Carbohydrate research Lab at the North Dakota State University in Fargo, and Dr Jim Kolmer at USDA-ARS Cereal Disease Lab in St Paul, Minnesota. All dedicated their time and resources to help my project succeed.

I express my sincere appreciation to my employer Kenya Agricultural Research Instiute, through Dr Ephraim A. Mukisira for granting me study leave.

I would wish to specially recognize Monsanto Company through Monsanto Beachell Borlaug International Scholars Program (MBBISP) for not only providing me a fellowship to pursue my studies, but also facilitating my participation in several career growth meetings. Last but not least, many thanks are expressed to the MBBISP judges for having found merit in my research proposal and fast-tracking my desire to learn the best of plant science.

Dedication

I dedicate this dissertation to my parents, Johnson and Sifira who early in life helped me appreciate the values of hardwork and patience, to my siblings and friends who through lots of encouragement helped me keep my eyes trained on the ultimate prize, and most importantly to my wife, Patricia, and daughters Alicia and Rosemary who despite the many days that separated us while I was abroad working on this dissertation, remained strong! To them I say…behold, the prize is finally home!

Table of Contents

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Appendices

Chapter 1

SNP-based molecular characterization of population structure, diversity and genomic footprints of selection in East Africa bread wheat

Wheat productivity in the East African region has perennially suffered losses due to diseases. The recent emergence of the highly virulent race TTKS of *Puccinia graminis* f.sp *tritici* followed by a quick succession of lineal races has resulted in efforts to determine the genetic diversity of local germplasm under the hypothesis that patterns of host genotype variation over time may be contributory to the observed perennial vulnerability to rust diseases and slow genetic gain in the region. A panel of 300 bread wheat lines, 90% of which are historical and modern East African cultivars, breeder germplasm and landraces were assembled. A few vintage North American lines also were included. Here, upstream studies of the population structure, genetic diversity and signatures of selection in this panel based on a genome-wide set of 6488 informative single nucleotide polymorphism (SNPs) markers are discussed. Both 'discriminant analysis of principal components' and 'structure' revealed population subclustering largely consistent with line breeding era. SNP diversity was also found to vary among the identified subpopulations with the oldest East African cultivars and a few landraces (*i.e*. subpopulation East Africa3) being the most diverse. In similar analyses, the D- genome depicted the lowest diversity compared to either the A- or B-genome. Though 'Bayescan' and 'Lositan' methods were able to detect SNPs at elevated F_{st}, perhaps signifying past selection, there was minimal overlap between results obtained from the two methods.

INTRODUCTION

Wheat (*Triticum* spp.) is one of the several crops domesticated in the 'Fertile Crescent' (Brown et al. 2009). Cultivation of diploid (genome $A^m A^m$, *2n*=14) einkorn wheat (*Triticum monococcum* spp.*monococcum*) and tetraploid (genomes AABB, *2n*=28) emmer wheat (*Triticum turgidum* spp. *dicoccoides*) about 10000 years ago (Tanno and Willcox 2006) contributed to the 'Neolithic Revolution' (Shewry 2009), a highlight in the evolution of human societies marked by a transition from hunting and gathering of food to agrarian lifestyles (Dubcovsky and Dvorak 2007). Presently, bread wheat (*Triticum aestivum* spp. *aestivum* L.) an allohexaploid species (genomes AABBDD, *2n*=42) that originated around 8000 years ago following hybridization of emmer wheat with goat grass (*Aegilops tauschii*, genome DD) (McFadden and Sears 1946), accounts for over 95% of all cultivated wheat (Dubcovsky and Dvorak 2007). This species is counted among the most important cereal crops not only in global production but also in its ecological range of cultivation, diversity and the extent to which it has become embedded in the culture and religion of diverse societies. Wheat, mostly the hexaploid and tetraploid type, is extensively grown on 17% of all crop area, in the temperate, Mediterranean-type and subtropical parts of both hemispheres, from 67^0 N in Norway, Finland and Russia to 45^0 S in Argentina (Peng et al. 2011).

Wheat research and production through introduction, hybridization and selection has been underway in Ethiopia (Hailu et al. 1991) and Kenya (Dixon, 1960) for nearly a century. Past achievements have led to the development of cultivars highly adaptable to the East African highlands with most commercial production practiced at altitudes above 1500m. Diseases, especially rusts have reduced wheat productivity in the region ever

since the crop was first grown commercially in 1906 (Lathbury 1947, Evans et al. 1969, Makanda and Oehmke 1993, Dubin and Brennan 2009, Wanyera et al. 2009). Devastating historical and current epidemics such as race Ug99, a highly virulent race of *Puccinia graminis* Pers. f. sp. *tritici* Eriks the causal agent of wheat stem rust that was first detected in the region in 1999 (Pretorius et al. 2000), has reduced countries in the region to perennial net wheat grain importers. This is in the backdrop of increased consumption needs, estimated at more than 150 percent of regional production (FAOSTAT 2009). Additionally, wheat yield potential has faced the severe challenges of drought (Hailu et al. 1991, Kinyua et al. 2000) and attacks by insect pests, particularly the Russian wheat aphid (*Diuraphis noxia*) (Malinga 2007).

Assessment of genetic structure, relationships, and diversity within a given set of germplasm is useful in plant breeding for various reasons that include: (i) assisting in the selection of parental combinations for developing progenies with maximum genetic variability for genetic mapping or further selection (Barrett et al. 1998) (ii) determining the level of genetic variation when defining core subsets selected for specific traits (Mohammadi et al. 2003) and (iii) estimating possible loss of genetic diversity during conservation or selection programs (Reif et al. 2003).

It is widely acknowledged that the success of a plant breeding program depends on having adequate diversity in the germplasm (Johal et al. 2008). Genetic diversity is the amount of genetic variability in individuals of a variety, population or species (Brown 1983). It is the foundation for genetic improvement (Phillips 2010, Kilian 2012) from which plant breeders create novel plant genome combinations and select crop varieties more suited to the needs of diverse agricultural systems (Glazmann et al. 2010). This

underscores the pivotal role plant diversity plays in contributing to world food security (Abdurakhmonov and Abdukarimov 2008). Efficient utilization of genetic diversity within breeding programs and that locked in genebanks (Tanksley and McCouch, 1997) for crop improvement can best be achieved when its pattern in plant germplasm and its temporal and spatial distribution are clearly characterized.

In the course of agricultural development, early domesticates were gradually replaced first by landraces and traditional varieties, and later by genetically less-diverse modern cultivars. These changes reflect demographic and selective events often characterized by small initial population sizes and intense human selection for agronomic traits with an attendant decline in the diversity of breeding germplasm (Tanksley and McCouch, 1997, Hoisington et al. 1999, Nevo 2005). Using a set of microsatellite loci, Thuillet et al. (2005) characterized such bottlenecking. They estimate wild *dicoccoides* to have had an average effective population size (N_e) of 32,500 which corresponds to the number of breeding individuals in an ideal Wright-Fisher population (Haudry et al. 2007). This number was estimated at only about half (N_e = 12000) in emmer (*dicoccum*), a fifth (N_e = 6,000) in old cultivated *durum* landraces and a twenty fifth (N_e = 1,300) in most recently improved varieties. Furthermore, modern crop breeders have preferentially utilized 'advanced cycle breeding', a method in which two elite inbred lines are crossed to form a base population from which new inbreds are developed (Allard 1960 p. 115). While this favored closed pedigree structure has proven successful to achieve gains in selection and to conserve favorable gene complexes (Condon et al. 2009), the challenge is that over cycles of selection, the genetic diversity within breeding populations is reduced, and presumably impedes future genetic gains (Munoz-Amatriain et al. 2010).

For example, Lu and Bernado (2001) suggested that the reduction in genetic distances among elite lines within Iowa Stiff Stock Synthetic and non-Stiff Stock maize heterotic groups could limit future gains from selection. Additionally, genetic uniformity arising from the recycling of related elite material can lead to genetic vulnerability to disease and insect pests (Committee on Genetic Vulnerability, USA 1972).

In bread wheat, initial bottlenecking that occurred during the polyploid speciation event (Dubcovsky and Dvorak 2007) is suggested to have contributed to a decline in genetic diversity within modern inbred elite cultivars (Haudry et al. 2007). However, though experiencing diversity bottlenecks, the adaptive genetic complexes that evolved in its progenitor wild emmer at the site of domestication still provide modern wheat with its remarkable adaptive ability to diverse environments and end uses (Nevo et al. 2002). Wheat thus compensates for bottlenecks by capturing part of the genetic diversity of its progenitors and by generating new diversity at a relatively fast pace (Dubcovsky and Dvorak 2007). Various molecular analyses for wheat diversity have been carried out in the past (*e.g*. Donini et al. 2000, Roussel et al. 2004, Fu 2006, Huang et al. 2007, White et al. 2008). A recent meta-analysis (Van De Woouw et al. 2010) combining results of these and other studies didn't find universal support for a significant gradual reduction in wheat diversity. However, examination of specific regional germplasm (*e.g*. Fu and Somers 2009) revealed some evidence for a negative impact of plant breeding on diversity.

Many of the earlier studies reviewed by Van De Woouw et al. (2010) utilized either one or a combination of different marker systems in quantifying wheat diversity including restriction fragment length polymorphisms (RFLPs); random amplified

polymorphic DNAs (RAPDs), and amplified fragment length polymorphisms (AFLPs). Importantly, the time-consuming multi-step protocols for some of these markers such as the hybridization-based RFLPs, coupled with low levels of genetic polymorphism in wheat (Bryan et al. 1999) may have paved way for the heavier utilization of microsatellites or simple sequence repeats (SSRs) in recent wheat diversity studies. SSRs are preferred because of their reproducibility, multiallelic nature, relative abundance and good genome coverage (Varshney et al. 2005). However it has been suggested that the high mutation rate of microsatellites might have allowed some recovery of diversity since domestication (Vigouroux et al. 2002) and may underestimate the consequences of bottlenecks for nucleotide diversity (Haudry et al. 2007). Further, microsatellites are prone to size homoplasy which occurs when different copies of a locus are identical in state although not identical by descent (Estoup et al. 2002), with the identity generated through mutation. Homoplasy violates a basic assumption in the analysis of genetic markers: that variants of similar phenotype (*e.g*. base pair size) are assumed to derive from a common ancestor (Sanderson and Hufford, 1996). Another challenge with this marker system is the need for estimates of fragment lengths which has limited the standardization of polymorphism data across different laboratories and equipment (Chao et al. 2009). Although single nucleotide polymorphisms (SNPs) are generally biallelic and have a lower information content than SSRs (Van Inghelandt 2010), attention has been drawn to their utility in plant diversity studies since they provide an abundant source of DNA polymorphisms (Zhu et al. 2003), the ease to automate SNP allele calls (Akhunov et al. 2009, Semagn et al. 2010) that enable high throughput genotyping, and the virtual lack of homoplasy in SNPs (Stucki et al. 2012). However, not many wheat

diversity studies have involved use of SNP markers to date. Large-scale SNPs discovery and development in a polyploid is a daunting task because of the presence of homoeologous and paralogous genes (Saintenac et al. 2011). Despite these challenges, a number of SNP platforms have recently been developed to perform high-density wheat genotyping (Allen et al. 2011). Chao et al. (2009) analyzed SNP marker polymorphism among 20 US elite wheat cultivars and advanced breeding lines representing seven market classes based on 359 SNP-detection primers. The authors demonstrated a marked reduction in the diversity of the D genome, measured as polymorphism information content (PIC), relative to A- and B- genomes. A collaborative project involving US and Australian scientists have recently designed a 9000 SNP iselect beadchip assay (http://wheatgenomics.plantpath.ksu.edu/snp) with over 80% of the discovered SNPs mapped. This resource promises great utility in wheat diversity studies and similar research designed to address fundamental questions of the relative roles of adaptive and non-adaptive processes in shaping patterns of genomic variation (Crisci et al. 2012).

To improve crop plants, it is essential to sort available diversity for alleles and polymorphisms that are beneficial (Buckler and Thornsberry 2002). Approaches to dissect diversity include the conventional 'top-down' (*e.g.* quantitative trait loci (QTL) and linkage disequilibrium (LD) mapping), that begins with trait and proceeds to identify candidate genes or the relatively new 'bottom-up' methods that start with detecting genes with signatures of adaptation using population genetics based methods to identify the traits affected (Ross-Ibarra et al. 2007). Owing to the idea that selection has a localized effect on molecular variation in populations (Nielsen 2005), bottom-up approaches that detect footprints of selection are based on pin-pointing genomic regions with patterns of

variation different from the genome-wide average (Wright et al. 2005, Horton et al. 2012). Using genome-wide scans based on the analysis of LD patterns and genetic differentiation between populations, targets of selection have been detected in several natural (Wang et al. 2012) and cultivated crop populations including maize (Vigouroux et al. 2002, Hufford et al. 2012), sorghum (Hamblin et al. 2006, Bouchet et al. 2012), pearl millet (Mariac et al. 2011) and tomato (Sim et al. 2012). Analysis of genetic variation on the basis of genetic scans for targets of selection in bread wheat have been limited in scope possibly due to lack of high-density markers for diverse populations (Cavanagh et al. in press). However, this analysis has been conducted by Chao et al. (2010) using a genome-wide set of 1536 SNPs. The existence of known sets of genes that control vernalization, photoperiod response, early maturity (William et al. 2008) and possibly other unknown chromosomal regions (Law et al. 1998) in plant germplasm when successfully flagged, might provide insights about linked molecular markers.

The objectives of this study were to (i) detect patterns of population structure within historical and contemporary bread wheat genotypes adapted to East Africa, (ii) assess the level of SNP polymorphism across East Africa wheat lines (iii) and evaluate the potential use of wheat 9000 iselect SNP coverage to detect genomic regions providing signals of selection in bread wheat.

MATERIALS AND METHODS

Plant materials

Two hundred four wheat cultivars and advanced breeding lines released or developed in East Africa from 1920 - 2010 were assembled for this study. This panel consists of 71 Ethiopian and 133 Kenyan genotypes respectively. Several databases including 'Genetic Resources Information System for Wheat and Triticale' (http://wheatpedigree.net) and 'Wheat pedigree and Identified Alleles of Genes online' (genbank.vurv.cz/wheat/pedigree/) were interrogated to provide pedigree information and cultivar release dates. Supplementary information was requested from Ethiopian and Kenyan breeders. To increase the scope of the study, 42 'founder lines' and 51 North American (mainly Minnesota) vintage cultivars were included. Founder lines were identified as those parents frequent in the pedigrees of the East Africa germplasm. The complete list of the 297 lines and pedigrees is provided as Appendix I. Seeds of most lines were sourced from Ethiopia and Kenya with a significant contribution from the USDA-ARS National Small Grains Collection.

DNA extraction

Plants were grown in the greenhouse and DNA was isolated from 12-14 day old seedlings. About 20g of leaf tissue was harvested in 2ml microtubes on ice from one plant of each line. Tissue was lyophilized and then ground to fine powder using 4mm glass beads by shaking on a GenoGrinder®. Subsequently, 600 µl of extraction buffer (0.1M Tris-HCL pH 7.5; 0.05 EDTA pH 8.0; 1.25% SDS) was added to each tube, shaken thoroughly, and incubated in a water bath at 65^0C for 30 minutes with an intermittent occasional gentle shaking. Tubes were placed on ice to cool for about 5

minutes before adding 300 µl of 6M ammonium acetate to each tube. After shaking to facilitate mixing, the tubes were left to stand on ice for 15 minutes. A 720 µl supernatant was recovered from each tube into new 1.5 ml microtubes following 5 minutes of centrifugation at 13,000 rpm. To each tube, 460 µl of isopropanol was added and mixed thoroughly before allowing the DNA to precipitate on ice. Samples were centrifuged for 10 minutes at 13,000 rpm to pellet DNA which was then washed in 1ml of 70% ethanol and re-suspended in 200 µl of TE buffer. After spinning down the tubes for 5 minutes at 13000 rpm, 190 µl supernatant was transferred to clean tubes and diluted using distilled water.

SNP genotyping

SNP genotyping assays were performed at the USDA-ARS genotyping laboratory in Fargo, North Dakota using the wheat 9,000 iselect SNP assay. This was done on the Illumina BeadStation and iscan instruments according to standard Illumina GoldenGate assay protocols (Shen et al. 2005). Subsequently, genotype calling was carried out using Illumina's BeadStudio software v.3. In this procedure, the software's default clustering algorithm was initially used to identify SNPs with three unambiguous allele clusters and to automatically classify those according to their matching genotypic states. To reduce genotyping error, likely due to peculiarities of clustering patterns in polyploid wheat (Chao et al. 2010), manual curation was applied for assays that produced compressed SNP allele clusters (Akhunov et al. 2009), which the allele calling algorithm couldn't decisively classify. Homozygous genotypic classes were hence manually defined leaving heterozygous genotypes uncalled.

Data Analysis

The consensus integrated genetic map for 7,500 SNPs was kindly provided by the team that designed the 9000 iselect assay through Dr. Eduard Akhunov from Kansas State University. This map is largely based on the MAGIC population (Huang et al. 2012), which provided most of the mapped markers plus markers mapped in several biparental populations. In total the present assays generated 6,488 polymorphic SNPs with < 10% missing data per SNP to be used to compute several diversity statistics including, polymorphic information content (PIC), minor allele frequency (MAF), expected heterozygosity/gene diversity (H_e) and allelic richness (A_R). The same set of SNPs was utilized to generate several analyses of molecular variance (AMOVA), pairwise F_{st} and for F_{st} outlier detection to infer putative selection based on genomic scan methods outlined below. Population structure was evaluated using 538 SNPs (225 for the A genome, 216 for the B genome and 97 for the D genome (Appendix II). Only markers located approximately 4 cM apart, or more were selected for this analysis to reduce the effect of frequency correlation between linked alleles (Chao et al. 2010).

Population structure

Discriminant analysis of principal components (DAPC) was used to elucidate discrete genetic clusters for the inclusive set of 297 lines. This analysis was performed using the Adegenet package version 1.3.5 (Jombart, 2010) implemented in R (R Development Core Team, 2011). DAPC is a multivariate method that identifies clusters of genetically similar individuals by using principal components analysis. Subsequently, a discriminant analysis is applied which maximizes the among-population variation and minimizes variation within pre-defined groups (Jombart 2012).

Additionally, population structure was inferred using the model-based Bayesian clustering method of Pritchard et al. (2000) as implemented in Structure version 2.3.4 (pritch.bsd.uchicago.edu/structure.html). A total of 10 iterations of Gibbs sampler were run for an admixture model with both correlated and non-correlated allele frequencies (Falush et al. 2003). Run length was set to 10^6 Markov chain Monte Carlo replicates following a burn-in period of 10^5. Five independent replicates were implemented in each of the iterations. To estimate the correct number of K (number of subpopulations) in the sample, two *ad hoc* methods were used. Firstly, the second order rate of change of the likelihood (ΔK) was plotted for each successive K as proposed by Evanno et al. (2005). The true value of K is inferred when ΔK shows a clear peak. Next, the rate of change of the log likelihood for each successive K *i.e.* L (K), was plotted for each K as proposed by Pritchard et al. (2000). When K is approaching a true value, L (K) plateaus or continues to increase slightly and has high variance between runs (Rosenberg et al. 2001). To summarize Structure analyses, graphical visuals of assignment of each wheat genotype to the respective assumed clusters were developed using package RSvg device version 0.6.4.1 (Luciani 2009) implemented in R (R Development Core Team, 2011). In this step which was also extensively informed by breeding histories and origins of individual lines, those with probability of membership $\geq 51\%$ were assigned to the same group.

Genetic distance and relative kinship

Matrices of Nei's genetic distance (Nei 1972) and relative kinship were calculated between each pair of lines respectively using package Adegenet version 1.3.5 (Jombart, 2010) of R and TASSEL version 3.0 (Bradbury et al. 2007) across all 6488 SNPs. A dendrogram was constructed to reveal groups of closely related lines using the neighbor

joining algorithm of the R package. The resulting tree was then formatted and visualized using TreeDyn (Chevenet et al. 2006). Results from kinship analysis were summarized and a bar plot drawn to represent percentage of pairs of lines corresponding to various kinship coefficient classes.

Genetic diversity

PIC which refers to the relative value of a marker with respect to the polymorphism it exhibits, a concept first described by Botstein et al. (1980) was calculated for each locus. Thus; $PIC = 1 - \sum_{i}^{n} p_i^2$, where p_i is the frequency of the i^{th} allele (Anderson et al. 1993). Estimates were made using PowerMarker software (Liu and Muse 2005) for sets of SNPs grouped by (i) chromosome (ii) genome (iii) subpopulations inferred from DAPC, and (iv) subpopulations inferred from Structure. PowerMarker was also used to calculate expected heterozygosity/gene diversity (H_e), the probability that two randomly chosen alleles from the population are different (Weir 1996) and minor allele frequency (MAF), the frequency at which the less common allele occurs in a population. Allelic richness (A_R), a count of the number of the number of alleles observed in a locus in a population (Kalinowski 2004) was determined with the rarefaction method available in HP-RARE version 1.0 (Kalinowski 2005) in order to obtain estimation based on equivalent sample size. Rarefaction was not used in calculating H_e because this is far less sensitive to sample size (Petit et al. 2008).

The partitioning of SNP variation was conducted using analysis of molecular variance (AMOVA) as implemented in Arlequin version 3.5 (Excoffier and Lischer 2010). Several comparisons were made based on $F_{st,}$ the population differentiation statistic (Weir and Cockerham 1984): (i) within and among release era for 204 East

Africa cultivars (landraces and advanced lines excluded). Three *ad hoc* eras were defined for this analysis- early century (before 1940), mid-century (1941-1980), late century (1981-2010); (ii) within and between release region for 255 cultivars. Two geographic regions were defined- East Africa, North America; (iii) within and among subpopulation inferred from DAPC; (iv) within and among subpopulations inferred from Structure; (v) within and among genomes *i.e*. A-genome, B-genome, D-genome.

Detection of outlier SNP loci

Wright's fixation index, F_{st} (Weir and Cockerham 1984) provides a measure of population differentiation by estimating the correlation of alleles within a subpopulation relative to alleles in the entire population. Markers that present higher than expected F_{st} values under neutral assumptions are candidates for positive selection where different populations have fixed different alleles. Markers that present lower than expected F_{st} values under neutral assumptions are candidates for balancing selection where diversity (heterozygosity) tend to be conserved in populations (Bouchet et al. 2012). Two outlier detection frameworks- Bayescan (Foll and Gaggiotti 2008) and FDIST2 (Beaumont and Nichols 1996) as implemented in Lositan (Antao et al. 2008) were used to calculate single-locus F_{st} for 5947 mapped SNP loci along each chromosome. These tests were applied on the East African cultivars grouped to reflect 3 *ad hoc* breeding eras (before 1940, 1940-1980, after 1980). The same applications were used to assess whether any of the SNP loci exhibited outlier behavior assuming an infinite island model (Bazin et al. 2010). The basic postulate of the infinite island model is that subpopulation allele frequencies are correlated through a common migrant genepool from which they differ in varying degrees. Here, a common migrant genepool was roughly assumed to be 'the rest

of world wheat', from which individual introductions have been sourced to East Africa through the respective breeding eras. In Lositan, expected heterozygosity (H_e) and F_{st} were estimated through five coalescent simulations for each pairwise comparison. A simulation run comprised of 100,000 iterations conducted using the neutral mean, forced mean, and an assumed infinite allele mutation model (Antao et al. 2008). Outliers were detected by comparing the simulated H_e and F_{st} distribution to the observed distribution of the two parameters at a 99% confidence interval. Markers that presented higher than 99% confidence interval were considered candidates for positive selection. It has been suggested that a major weakness of the Lositan method is that all markers including those that will be candidates for selection are used to construct confidence intervals for neutral marker F_{st} distribution (Bouchet et al. 2012). Besides, Lositan is expected to have a higher type I error rate than Bayescan (Narum and Hess 2011), the alternative bayesian method that implements the multinomial-Dirichet model (Balding 2003, Foll and Gaggiotti 2008) to detect loci under selection. The Bayescan program takes into account uncertainty in allele frequency estimates in populations of small size, providing a more efficient estimate of genetic differentiation. In Bayescan, the posterior probability of any of the SNP loci being under selection was estimated through a reversible-jump Markov chain Monte Carlo approach by implementing 20 pilot runs of 5000 iterations, an additional burn-in of 50,000 iterations followed by 100,000 iterations with a sample size of 5,000 and a thinning interval of 20. A false discovery rate of 0.05, which corresponded to a log10 (Bayes Factor) equal or greater than 1, was used as the threshold to accept signatures of selection. On the Jeffrey's scale (Jeffry 1961), a log10 (Bayes Factor)

greater than 1 is considered to match a posterior probability indicative of strong evidence for selection (Foll, 2012).

RESULTS

SNP genotyping and polymorphism

A total of 2,673,000 genotyping data points were generated on interrogating the inclusive set of 297 wheat lines with the 9000 wheat iselect SNP assay. After eliminating SNPs that could not be decisively clustered, 6488 SNPs were used, representing 1,771,308 data points. Among these, 97.8% were polymorphic with 825,312 coded as AA and 908,200 as BB. Only a small proportion (0.003% = 5,721) showed residual heterozygosity (coded as Aa and Bb) or no amplification (2.1% = 37796). These latter categories were treated as missing data points in subsequent analyses. Of the polymorphic SNPs, 5962 (92%) are in genetically mapped positions.

The mean MAF values (Table 1) ranged from 0.13 in chromosome 4D to 0.33 in chromosome 1D. Across genomes, mean MAF was 0.28, 0.29, and 0.24 for the A-, B-, and D-genome respectively. In addition, there was evidence for significant differences in the distribution of MAF classes among the genomes ($\chi^2_{127} > \chi^2_{34,\ P=0.005}$; Figure 1). Both genomes A and B had distributions skewed towards medium MAF (> 20%) compared to intermediate frequencies (5-20%) observed for the D genome.

Population structure

DAPC analysis identified 7 clusters within the data (Figure 2, Appendix I) after retaining 15 principal components that explained over 99% of the observed variation. The number of principal components retained was chosen to optimize 'a-score' (Jombart 2012), the difference between the proportion of successful reassignment of the analysis (observed discrimination) and values obtained using random groups (random discrimination). These clusters arbitrarily designated as East Africa1, East Africa2, East

Africa3, North America, Mixed Identity1, Mixed Identity2 and Tetraploid consisted of 67, 63, 83, 57, 8, 12 and 7 lines respectively. This broad cluster naming criterion was informed by an in-depth study of the breeding history and origin of individual lines that grouped together. The East Africa1 subpopulation (cluster 3, Figure 2) had many of the present commercially popular cultivars and those released in Ethiopia and Kenya after 1977. East Africa2 subpopulation (cluster 6, Figure 2) mostly spanned materials cultivated in the region from 1960-1977 while lines released before 1960 or those recognized as important breeder lines then and landraces pooled into the East Africa3 subpopulation (cluster 7, Figure 2). The North America subpopulation (cluster 4, Figure 2) was clearly enriched of Minnesota old germplasm as well as other regional lines. The accord of individuals labeled Mixed Identity1 (cluster 2, Figure 2) and Mixed Identity2 (cluster 5, Figure2) was not readily distinguishable since they didn't follow an evident pattern. DAPC was most successful in isolating the seven lines labeled Tetraploid which are actually durum wheat landraces misclassified as bread wheat among the global sample. Results presented elsewhere (Chapter 4, this thesis), involving profiles of high molecular weight glutenin protein in the grain of all 297 individual lines supports that the 7 misclassified lines entirely lack the *GluD1x* and *GluD1y* phenotype. Alleles at *GluD1x* and *GluD1y* are encoded by loci in the D genome (Gianibelli et al. 2001) which is absent in *durum* wheat.

The model-based approach of inferring population sub-structuring was equally insightful. The comparison of admixture models assuming correlated and non-correlated allele frequencies revealed that both of the approaches were correspondingly appropriate for inferring the likely number of clusters within the studied data set. In either case, using

the ΔK method of Evanno et al. (2005), maximal ΔK occurred at K=2 (Figure 3, plot A) with the next largest peak at K=4. At K=2, the wheat lines were broadly clustered into North America vs. all other. Given that the East African lines have a long breeding history and are expected to include diverse lines falling into subpopulations subjected to different selection regimes over time and with differentiated allelic frequencies, the K value with the maximal likelihood (K=4) over that with maximal ΔK (K=2) was deemed most favorable. Consistently, the log likelihood method of Pritchard et al. (2000) seems to predict more than two clusters. Although L (K) is not yet plateaued at K=4, the value for this statistic continues to increase only slightly at K=4 (Figure 3, plot B), diagnostic of the true number of clusters. By assuming 4 clusters and supported by information of the origin, pedigree and breeding histories of the sampled lines, the recognized broad categories (Figure 4, Appendix I) were: East Africa1 (modern East African commercial and cultivars released after 1977); East Africa2 (East African lines released 1960 - 1977); East Africa3 (East African lines released before 1960, landraces and lines of mixed identity); North America (vintage North American cultivars).

Genetic distance and relative kinship

The neighbor joining tree generated from Nei's genetic distance matrix grouped the 297 lines into 5 major groups and 8 subgroups. These are: group 1 (East African commercial and cultivars released after 1977); group 2 (East African lines released 1960 - 1977); group 3 (some of the East African lines released before 1960); group 4 and group 5 (some of the East African lines released before 1960, landraces, some of the North American lines). When the model-based population partitioning Structure result was superimposed on the generated neighbor joining clustering data (Figure 5), moderate

concordance was observed. While the North American lines did not separate distinctly as in model-based structuring, instead tending to group with the East African most vintage material, the East African commercial and East African lines released 1960 - 1977 were clearly differentiated. Besides, the neighbor joining tree suggested that very few of the lines released earlier in the century were related to those released in 1960 - 1977 and consistently, even fewer to modern East Africa commercial cultivars. This outcome reflects the insight that beginning in the mid 1950's Kenya breeders started utilization of 'International Spring Wheat Nurseries' (reviewed in Dixon, 1969), largely abandoning earlier germplasm.

Relative kinship coefficients ranged from 0, indicative of no relationship to 1 suggesting very close relationship (Figure 6), with an overall mean of 0.24. Most of the lines (83%) had coefficients ranging from 0.11 to 0.40.

Patterns of genetic diversity

There was unevenness in the number of SNP markers tested for each genome (Table 1). Only 418 (7% of the mapped markers) markers tested were in the D genome, 2807 (47%) in the A genome, and 2737 (46%) in the B genome for a total of 11,902 alleles. The fewer number of markers in the D genome reflects the lower number of polymorphic D genome-SNPs found in the discovery panel.

The chromosome with the highest estimate of PIC was 6B. This chromosome was also amongst those with the highest *He*, including 2A and 7B, suggesting that they were the most diverse. The lowest values for both PIC and *He* were recorded in chromosome 4D. Across homoelogous groups and genomes, group 1 chromosomes and genome B scored highest for PIC and H_e on average, the lowest values being observed in group 4

chromosomes and genome D. A summary of mean MAF, PIC, *He* and A_R was also made for the various subpopulations inferred in both DAPC and Structure analyses (Table 2a). All subpopulations had distributions skewed towards common alleles as suggested by the medium values of MAF (>20%) observed across the board. An exception was the Tetraploid subpopulation made of 7 *durum* wheat individuals that was inclined towards low MAF (5 - 20%). In both DAPC and Structure subpopulations, those generally enhanced with lines released in East Africa before 1960 and landraces, designated as East Africa3 (Table 2a) had the highest PIC, H_e and A_R hinting to higher molecular diversity. A comparison of each of the other subpopulations was made with East Africa3 to reveal discrepancies in diversity (Table 2b). In general, lines designated as Mixed Identity2 were observed to be as diverse as East Africa3. The Mixed Identity1 and North America subgroups appear to be the least diverse relative to East Africa3. On the other hand, East Africa1 subpopulation, which as earlier indicated had generally many of the modern East Africa commercial and cultivars released after 1977 appear to share much of the allelic diversity present in the East Africa3 subpopulation.

For all group comparisons made in the analyses of molecular variance (AMOVA), it was evident that lines within a given comparison were more genetically differentiated than when contrasted across (Table 3). This is reflected by the higher proportion of total variance for the 'within' analyses. For example, when AMOVA was conducted for 204 East African cultivars released in three *ad hoc* pre-defined breeding eras (before 1940, 1941-1980, 1981-2010), 94.2% of the total variation could be ascribed to genetic differences within an era relative to only 5.8% due to genetic differences of lines across era. Additionally, the coefficient of differentiation, F_{st} for the 'across/among' analyses in

all contrasts was markedly low, ranging from 0.003 when AMOVA was conducted across genome (A, B, D) to 0.157 when done across DAPC clusters. The proportion of total variation explained notwithstanding, AMOVA results also indicated that all variance components were significant (P <0.001) at any level of group comparison.

SNP outlier detection

Genome scans based on both Bayescan and Lositan detected SNP loci at elevated F_{st} values. Across all 21 wheat chromosomes, 55 and 102 markers with F_{st} values equal to or greater than 0.3 were detected by both methods respectively (Figure 7). Amongst the East African cultivars considered across breeding regimes (before 1940, 1941-1980, 1981-2010), Bayescan led to the detection of 43 candidate loci under positive selection (Figure 8a). For the same analysis Lositan identified 48 and 75 loci under positive and balancing selections respectively (Figure 8b). Oddly, there was minimal overlap of loci detected by the two approaches with only 4 loci, all on chromosome 1A, observed in both (Table 4).

DISCUSSION

This study has illuminated several important concepts that have direct implications to plant germplasm handling procedures including parental selection, creation of breeding populations, conservation of vintage lines and varietal protection. A comprehensive sample of mainly East African wheat cultivars released over a span of several decades was chosen for the various assays with the understanding that despite past success, this region still lags behind its targeted production needs for the crop. From a different perspective and disregarding that East Africa is a historically recognized 'hotspot' for major wheat rusts epidemics exemplified by *Yr9* stripe rust virulence which emerged in Ethiopia in 1986 (Nazari, 2010) and recently stem rust race *Ug99*, the region has in the past stood out as a major source of wheat lines with unique resistance genes. For instance, the cultivars Kenya-117A, Kenya-58 and Kenya-farmer that were released in 1937, 1939 and 1954 respectively, were found to be resistant to the damaging North American race 15B of stem rust in the 1950s and were extensively used in developing breeding populations in the Canadian and some US wheat programs.

Achieving wheat yield stability and closing the local yield gap will necessarily come from contributions made in selection of superior genotypes suited to the East African environment and an introgression of useful alleles into new cultivars. An outlook of the temporal trends in the deployment of wheat diversity in the region might inform a better understanding of the recurrent challenge of new rust races, given the indication that genetically narrow germplasm for instance, have the potential to invoke intense directional selective pressure to pathogen populations favoring novel races.

The present study has reiterated the utility of SNP- based markers in characterizing population structure and quantifying genetic diversity and relatedness in

wheat. Although not with clear-cut results, use of SNP markers to detect signatures of selection, as a bottom-up approach to understand the nature of adaptive alleles has been utilized.

SNP polymorphism

SNP allele distributions for A- and B- genomes were observed to be skewed towards medium MAF relative to the D-genome. Likely, this could be a reflection of the smaller number of SNPs sampled in the latter. Additionally, the observed distribution can be attributed to the intentional bias introduced in the procedures used for SNP discovery. In particular, development of the assay favored common alleles by selecting SNPs that were present in at least two individuals in the discovery panel, and that were shared between accessions of different geographic origin (Cavanagh et al. in press).

Population structure and genetic relatedness

In their overlapping results, the two multivariate methods DAPC and Structure were successful in identifying sub-populations whose distinction is consistent with East Africa's wheat breeding history. Dixon (1960) is particularly instructive. The author directs that the origin of wheat in Kenya traces from introduction of Australian varieties at the beginning of the 20th century subsequently augmented with a few Egyptian, Italian and Canadian founder lines decades later. Breeding from 1930 - 1950 was largely based on crossing within this collection of germplasm with relatively few additions from abroad. In the same time period, wheat improvement in Ethiopia mainly involved evaluation of indigenous wheat and a few introductions from Europe and Kenya (Hailu et al. 1991). In this study, germplasm together handled by East Africa breeders prior to 1960 is labeled East Africa3. This apparently unique genepool that shows only marginal

relatedness with subsequent sub-populations in the genetic distance results (Figure 5) was correctly inferred by all methods. Earlier breeders acknowledged the genetic narrowness of this initial set which only offered limited resistance to groups of rust races that had arisen since 1948 (Dixon 1960). Beginning in the mid 1950's there was a major shift to a new genepool comprising of cultivars derived from the International Spring Wheat Nursery initiated in 1950 by B. B. Bayles and R. A. Rodenhiser of USDA-ARS (United States Department of Agriculture-Agricultural Research Services) and Near East Nurseries organized by the Food and Agriculture Organization (F.A.O.) . This study successfully identified this group of individuals, labeled East Africa2. Members of this cluster are only related to a few individuals of the initial genepool, perhaps those that had maintained useful levels of resistance to the rust diseases. The third unique sub-population is East Africa1, consisting of most of the modern East Africa cultivars and those released in the late 1970s. Coincident with this period, CIMMYT, the main source of cultivars in this genepool had embarked in developing wheat lines having the 1RS translocation from rye (*Secale cereale*). The hallmark of this effort was a release of a series of lines harboring that translocation- most notable the 'Veery wheats', which includes Glennson, Ures, Genaro and Seri (Martin et al. 1990). Veery lines are popular under drought or heat stressed conditions and on average produce a 7% grain yield advantage (Rajaram et al. 1983). Equally significant, the 1RS translocation, whose frequency went up to ~70% at one stage in CIMMYT's spring wheat germplasm is associated with resistance to all three wheat rusts (stem rust, leaf rust and stripe rust) and powdery mildew (Singh et al. 2008). The universal susceptibility of the Kenyan commercial cultivars to race TTKSK (*Ug99*) of *Puccinia graminis* fsp. *tritici* (Njau et al.

2009) underscores the close relatedness of this germplasm since the race is virulent to gene *Sr31* on the 1RS translocation (Pretorius et al. 2000). To enhance insights from this study, 51 vintage North American lines were included in all assays. This sub-population was well identified in the DAPC and Structure results. It is intuitive that selection in the North American genepool must have favored combinations of alleles different from those in the East African wheat cultivars to help adapt to uniquely different growth environments. Interestingly, individuals in this sub-population seemed to be generally more closely related to the oldest members of the East Africa programs (Figure 5). One possible interpretation of this outcome is that as pointed above, some of the vintage Kenyan germplasm were intensively used to generate crosses by North American breeders in the 1950s to subdue local stem rust races including race 15B.

Overall, population structure heralded in the present wheat panel necessitates that this must be correctly modeled and sufficiently controlled in experiments that undertake to identify potentially useful alleles in this resource through genome-wide association mapping. Association mapping has recently been advocated as the method of choice for identification of loci involved in the inheritance of complex traits in plants (Zhu et al. 2008). However, the presence of population structure in the panel used in mapping can result in “spurious associations”, where putative associations are found between a phenotype and markers that are in reality not linked to any causative loci (Pritchard et al. 2000).

The ability to identify breeding populations with the highest mean performance has been underscored as an overarching breeding objective (Bernado 2003). It is widely acknowledged that groups of closely related individuals tend to bring redundancy in a

breeding program (Semagn et al. 2012), and their use as breeding populations may indicate a less than optimal allocation of resources. By selecting parents that are genetically similar, a breeder restricts the amount of variation that will be evaluated in the offspring (McCouch 2004). On the other hand, by crossing genetically divergent parents, the range of phenotypic variation will be much more extensive. The relatedness information generated in this study will encourage breeders' innovativeness in planning crosses that can most optimally exploit the variation in the East African wheat resources.

Genetic diversity

Lower diversity in the D genome than in either A and B as reflected by the lower number of polymorphisms tested for this genome in the SNP assay, and the consistent lower values of PIC and H_e is agreeable to previously published reports. Comparisons based on genome-wide estimates of average nucleotide polymorphism (θ_w) and nucleotide diversity (θ_π) revealed that the values of these measures were significantly lower in genome D, being less than half those in either A or B genomes (Akhunov et al. 2010). In their study utilizing 253 pre-selected SNPs genotyped on a set of 20 US wheat cultivars, Chao et al. (2009) found the ratio of the average PIC value for the A and B genomes to that of D genome (PIC_{AB}/PIC_D) to be 1.7, a near two-fold difference. In this study, the ratio is at least one-fold higher for PIC and also for comparisons involving H_e and A_R. A strong population bottleneck during the last polyploidization event from which hexaploid bread wheat originated from tetraploid emmer wheat (AABB) and diploid *A. tauschii* (DD) is suggested to have had a dramatic impact on the level of genetic diversity in the D genome (Akhunov et al. 2010, Dvorak et al. 1998), given that it involved only a small number of *A. tauschii* accessions (Dvorak et al. 1988). It is also indicated that since

the tetraploid progenitor and hexaploid wheat are inter-fertile and readily hybridize to give rise to fertile pentaploids (Kihara 1982), this has allowed for continued gene flow between the two species which is further encouraged by their sympatric nature (Akhunov et al. 2010). Such cannot be said of the diploid progenitor, *A. tauschii* and hexaploid wheat which share limited geographic distribution and hence little gene flow between them.

In addition to being an artifact of the enrichment bias for common SNPs in the ascertainment scheme applied during SNP discovery, the shift in the MAF spectrum towards a higher proportion of lower frequency alleles observed in the D genome (Figure 1) could also be a result of selection or demography. As a result of genetic drift, an additional diversity bottleneck involving a reduction in effective population size following intensive selection in modern breeding may have had a stronger effect on the D genome than on the A and B genomes (Chao et al. 2009). A higher loss of allelic variants, consistent with a lower mean H_e observed for the D genome (Table 1) in this study relative to either A or B genomes, may have been more dramatic on this genome during the bottleneck imposed by selection from landraces to modern improved cultivars. However, at the population level, while the East Africa3, a sub-population deemed to be richer in landraces and older cultivars had the highest diversity (Table 2a), in relative terms that of modern and most recently released cultivars as exemplified by East Africa1 subpopulation is at least 0.9 times as high. Still this number might be an underestimate considering the bias towards common SNPs included in the genotyping assay.

The mean PIC value of 0.28 found for the inclusive wheat panel, which compares closely with 0.25 when the populations are sub-structured (Table 2a) suggest that on

average, up to a third of the transcriptome-derived SNPs in the 9000 iselect assay are expected to be polymorphic in any two East African wheat lines. Furthermore, the probability that any two randomly chosen alleles in the complete population or in the inferred subpopulations are different as indicated by mean $H_{e,}$ is high (0.31 - 0.35) symptomatic of significant overall diversity. In a project intended at discovering and validating SNPs, Somers et al. (2003) recorded a similarly high mean PIC of 0.27 using a diverse set of 12 wheat genotypes including a synthetic derived mapping parent. However in their study, Chao et al. (2009) observed a lower value of 0.18 which they credited to less diversity in the study population.

Taken together, our results suggest a high genetic diversity in the wheat lines under study despite evidence for population sub-structuring. Validity for this observation is further provided by the large proportion of genetic variance within sub-populations from the AMOVA tests.

Signatures of selection

F_{st} scan for signatures of selection detect loci at different stages of selection (Horton et al. 2012) and is a function of the selected alleles in a population. While initial attempts to find correspondence of the detected alleles to known adaptive loci (*e.g*. *Ppd-B1*, *Rht-B1* and *Vrn1*) did not find outright hits, future work will involve application of the complimentary and non-parametric approach *i.e.* extent of pairwise sharing (Lewontin and Krakauer, 1973) to seek consensus with methods used presently. It is probable that the identified high F_{st} SNPs recorded in this study could have been selected during improvement and represent a wide array of loci across different traits.

CONCLUSION

This study has provided critical insights about an unprecedentedly large collection of the East Africa's bread wheat. It is envisaged that the knowledge so generated will enhance local breeders' ability to organize and select germplasm used in the crossing blocks for cultivar development. Information that vintage lines are a degree of magnitude richer in allelic diversity than the present cultivars necessitates that useful alleles are mapped and further characterized in a bid to utilize them for breeding. Methods of genome-wide association mapping and/or genomic selection modeling are predictably applicable in the East African breeding programs if additional genotyping and comprehensive phenotyping of important traits is implemented in the assembled wheat panel. Lastly, the procedure of identifying signatures of selection illustrated in this study, as a bottom-up approach to unraveling plant variation may become increasingly valuable as methods of genotyping by sequencing become routine in wheat.

Table 1 Number of markers and alleles, minor allele frequency (MAF), polymorphism information content (PIC) and expected heterozygosity (H_e) averaged across 5962 mapped SNP loci in an East African enriched set of 297 wheat lines.

			Mean		
	No. of marker	**No. of alleles**	**MAF**	**PIC**	H_e
Chromosome					
1A	510	1015	0.29	0.30	0.38
1B	358	716	0.27	0.29	0.36
1D	91	182	0.33	0.30	0.29
2A	325	650	0.31	0.30	0.39
2B	625	1250	0.28	0.30	0.37
2D	98	196	0.22	0.25	0.31
3A	390	780	0.26	0.28	0.35
3B	402	800	0.29	0.29	0.37
3D	33	66	0.25	0.27	0.34
4A	360	720	0.27	0.29	0.36
4B	159	318	0.27	0.29	0.37
4D	28	56	0.13	0.18	0.21
5A	408	812	0.28	0.30	0.38
5B	511	1017	0.30	0.30	0.38
5D	73	146	0.20	0.24	0.29
6A	417	834	0.26	0.28	0.36
6B	403	802	0.30	0.31	0.39
6D	52	104	0.27	0.27	0.34
7A	397	794	0.27	0.29	0.36
7B	279	558	0.31	0.30	0.39
7D	43	86	0.28	0.27	0.35
Homoelogous group					
1	959	1913	0.30	0.30	0.38
2	1048	2096	0.27	0.28	0.36
3	825	1646	0.27	0.28	0.35
4	547	1094	0.22	0.25	0.31
5	992	1975	0.26	0.28	0.35
6	872	1740	0.28	0.29	0.36
7	719	1438	0.28	0.29	0.37
Genome					
A	2807	5605	0.28	0.29	0.37
B	2737	5461	0.29	0.30	0.38
D	418	836	0.24	0.26	0.32
Total	5962	11902			
Overall mean		1.997	0.27	0.28	0.35

Table 2a Mean minor allele frequency (MAF), polymorphism information content (PIC), expected heterozygosity (H_e), and allelic richness (A_R) estimated for 6488 SNP markers across DAPC and Structure inferred clusters.

Sub-population	Size	Sub-population description	Mean (across loci)			
			MAF	**PIC**	H_e	A_R
DAPC						
East Africa1	67	E. African modern and after 1977	0.23	0.25	0.31	1.31
East Africa2	63	E. African 1960-1977	0.21	0.23	0.29	1.29
East Africa3	83	E. African before 1960, landraces	0.26	0.27	0.34	1.34
North America	57	N. American vintage	0.21	0.23	0.29	1.29
Mixed Identity1	8	No absolute description	0.21	0.22	0.27	1.30
Mixed Identy2	12	No absolute description	0.26	0.27	0.33	1.35
Tetraploid	7	Durum	0.13	0.14	0.17	1.22
Combined (*Durum excluded*)			0.23	0.25	0.31	1.31
Structure						
East Africa1	68	E. African modern and after 1977	0.25	0.26	0.33	1.40
East Africa2	47	E. African 1960-1977	0.24	0.25	0.32	1.39
East Africa3	131	E. African before 1960, landraces	0.27	0.28	0.36	1.43
North America	51	N. American vintage	0.21	0.24	0.29	1.39
Combined (*Durum excluded*)			0.24	0.25	0.33	1.40

Table 2b Comparison of PIC, H_e and A_R values amongst inferred subpopulations relative to East Africa3 group, estimated to be most diverse in Table 2a.

Sub-population[a]	Ratio of subpopulation relative to East Africa3		
	PIC	H_e	A_R
DAPC			
East Africa1	0.92	0.92	0.98
East Africa2	0.85	0.85	0.96
North America	0.85	0.85	0.96
Mixed Identity1	0.81	0.64	0.97
Mixed Identy2	1.00	0.97	1.10
Structure			
East Africa1	0.96	0.91	0.97
East Africa2	0.92	0.88	0.97
North America	0.88	0.81	0.97

[a] Subpopulation annotation same as in Table 2a.

Table 3 Results of analysis of molecular variance (AMOVA) components estimated based on 6488 SNP markers for 297 wheat lines for 5 considerations.

Comparison[a]	**Group /source**	**Variance component**	**Variation (%)**	**F_{st}**
Breeding era				
	Across era	67***	5.8	0.058
	Within era	1077***	94.2	
Release region				
	Across region	78***	6.7	0.067
	Within region	1090***	93.3	
DAPC inferred clusters				
	Among clusters	183***	15.7	0.157
	Within clusters	986***	84.3	
Structure inferred clusters				
	Among clusters	77***	5.5	0.053
	Within clusters	1325***	94.5	
Genome				
	Across genome	0.2***	0.2	0.003
	Within genomes	64.4***	99.8	

***Significant at P< 0.001

[a] Three breeding eras (before 1940, 1941-1980, 1981-2010); two release regions (East Africa, North America); seven DAPC inferred clusters (East Africa1, East Africa2, East Africa3, Mixed Identity1, Mixed Identity2, North America, Tetraploid); four Structure inferred clusters (East Africa1, East Africa2, East Africa3, and North America) and three genomes (A, B, D).

Table 4 SNP markers detected (FDR=0.01) as candidates of positive selection in both Bayescan and LOSITAN scans. Methods were implemented on East African cultivars grouped based on breeding era.

Candidate SNP	Chromosome	MAF	Bayescan F_{st}[b]	Lositan F_{st}[c]
wsnp_Ex_c4605_8239915	1A	0.18	0.30 (0.45)	0.41 (0.99)
wsnp_Ex_c4605_8240189	1A	0.37	0.31 (0.50)	0.41 (0.99)
wsnp_Ex_rep_c105244_89727666	1A	0.20	0.32 (0.56)	0.42 (0.99)
wsnp_Ex_rep_c105244_89727820	1A	0.20	0.32 (0.57)	0.40 (0.99)

[a] MAF, minor allele frequency

[b] The F_{st} obtained with Bayescan, the log 10(q value) is provided in parenthesis.

[c] The F_{st} obtained with Lositan, *P*value corresponding to *P*(Simul $F_{st} <$ sample F_{st}) is provided in parenthesis.

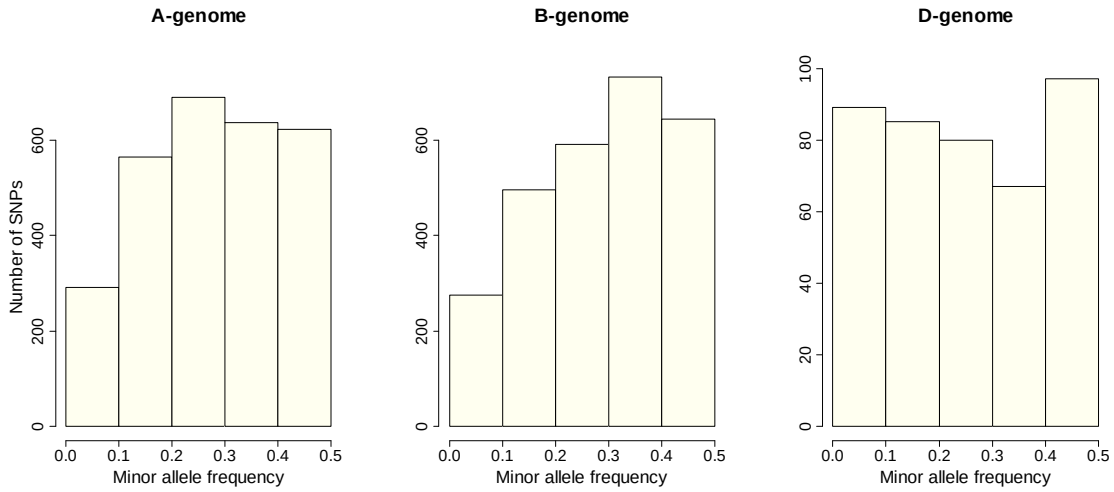


Figure 1 Distribution of minor allele classes across the A, B and D genomes of 297 wheat lines for 5962 SNP markers.

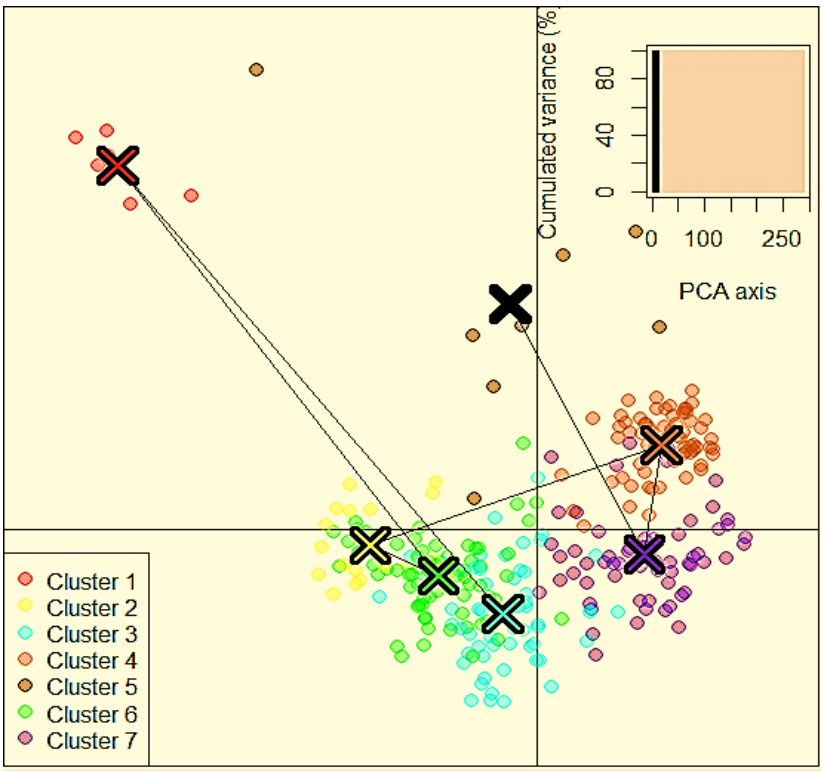
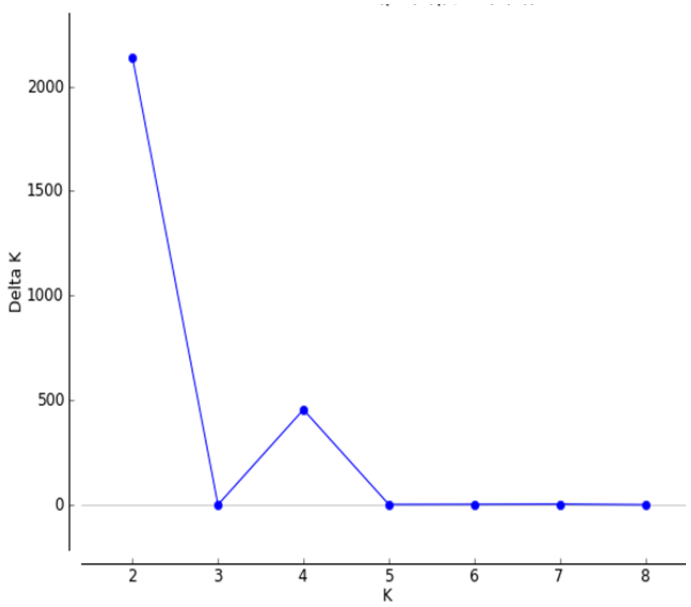


Figure 2 Discriminant Analysis of Principal Components (DAPC) scatterplot for 297 wheat lines based on 538 SNP markers showing 7 inferred clusters: Cluster 1 (Durum), Cluster 2 (Mixed Identity1), Cluster 3 (East Africa3), Cluster 4 (North America), Cluster 5 (Mixed Identity2), Cluster 6 (East Africa2), Cluster 7 (East Africa3). The center of each cluster is indicated by a cross. A minimum joining spanning tree clarifies spatial proximities between populations. On the bottom left is the legend and top right an insert of the first fifteen principal axes (> 99% of total variation).

A



B

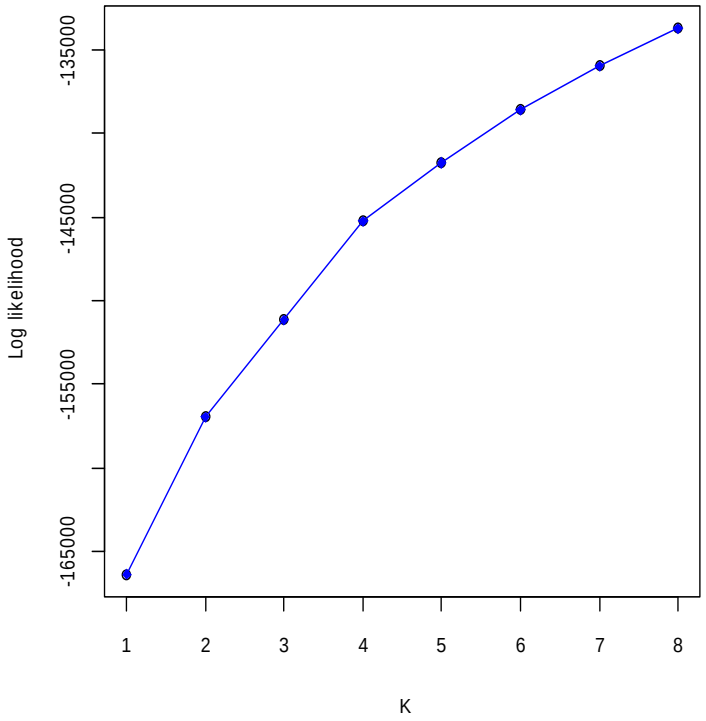
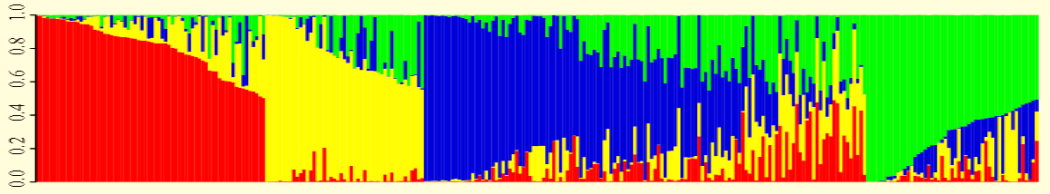


Figure 3 Plots of (A) delta K (Evanno et al. 2005) and (B) the log likelihood (Pritchard et al. 2000) of data averaged over 5 replicate runs for each K used to infer correct number of clusters.

A



B

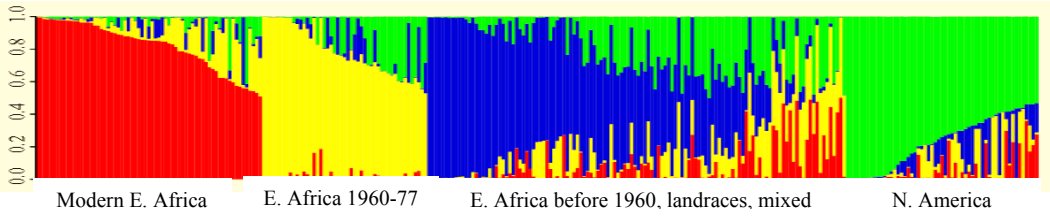


Figure 4 Structure clustering of 297 wheat lines based on 538 SNP markers. Analyses assuming admixed model with independent allele frequencies (Plot A) and with correlated allele frequencies (Plot B). The genotype of each line on the figure is represented by a colored line where each color reflects membership of a cultivar in one of the K= 4 clusters: red (modern East African commercial, and cultivars released after 1977); yellow (East African cultivars released 1960 – 1977); blue (East African cultivars released before 1960, landraces and lines of mixed identity); green (North American old cultivars).

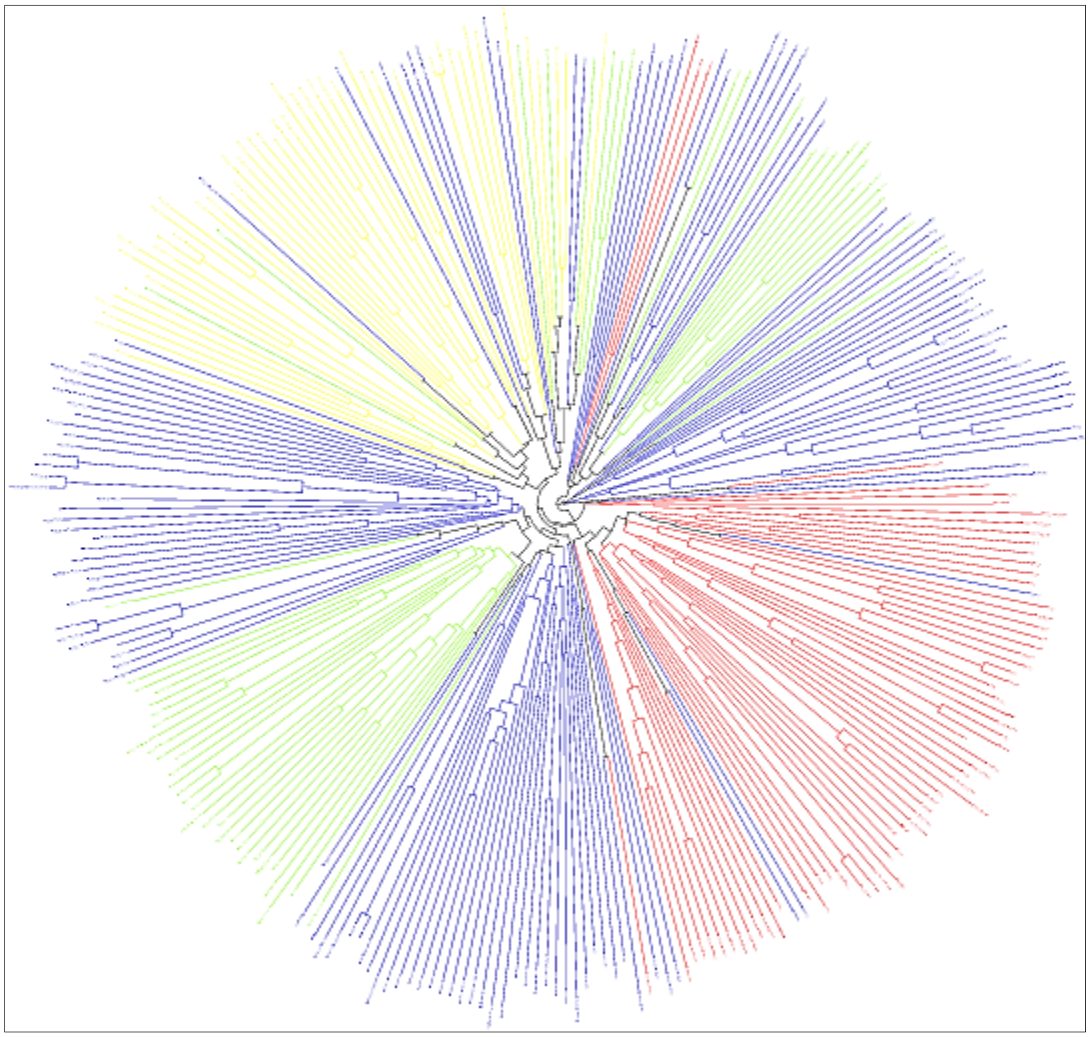


Figure 5 Neighbor joining tree for 297 wheat lines based on Nei's genetic distance calculated from 6488 SNP markers. Subgroups are colored to match clusters inferred in Structure analysis: red (modern East African commercial and cultivars released after 1977); yellow (East African cultivars released 1960-1977); blue (East African cultivars released before 1960, landraces and lines of Mixed identity); green (North America old cultivars).

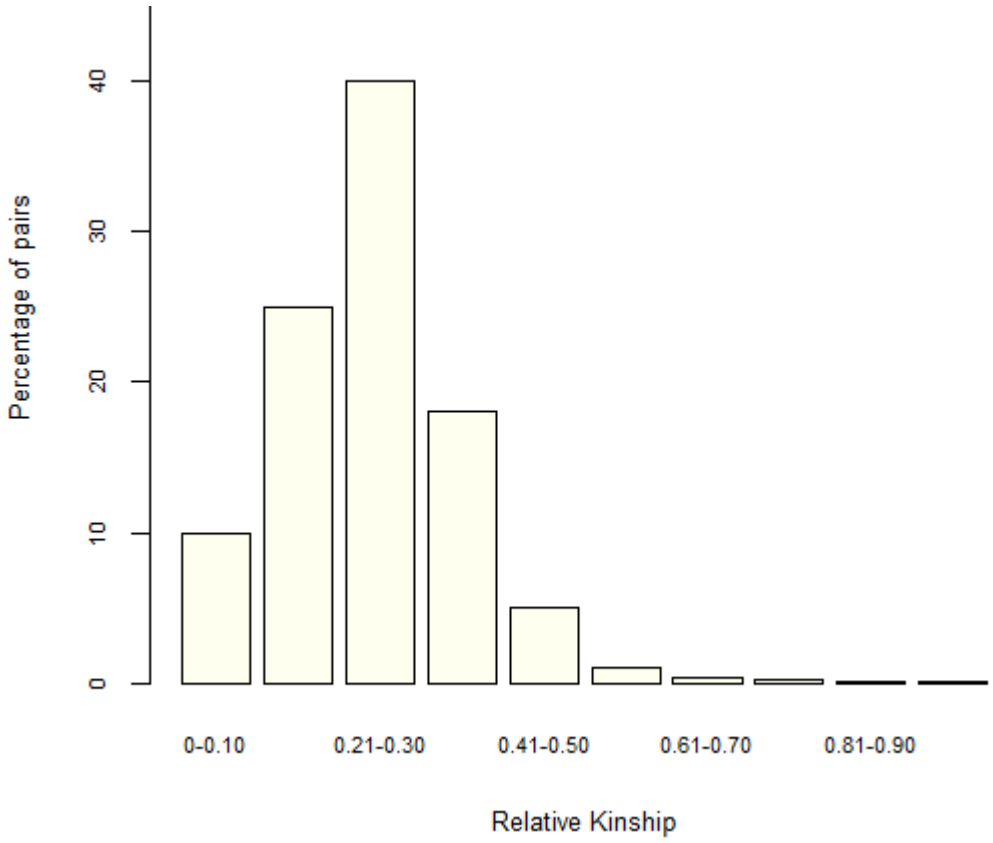


Figure 6 Distribution of pairwise relative kinship estimates among 297 wheat lines. Values are from TASSEL based on 6488 SNP markers.

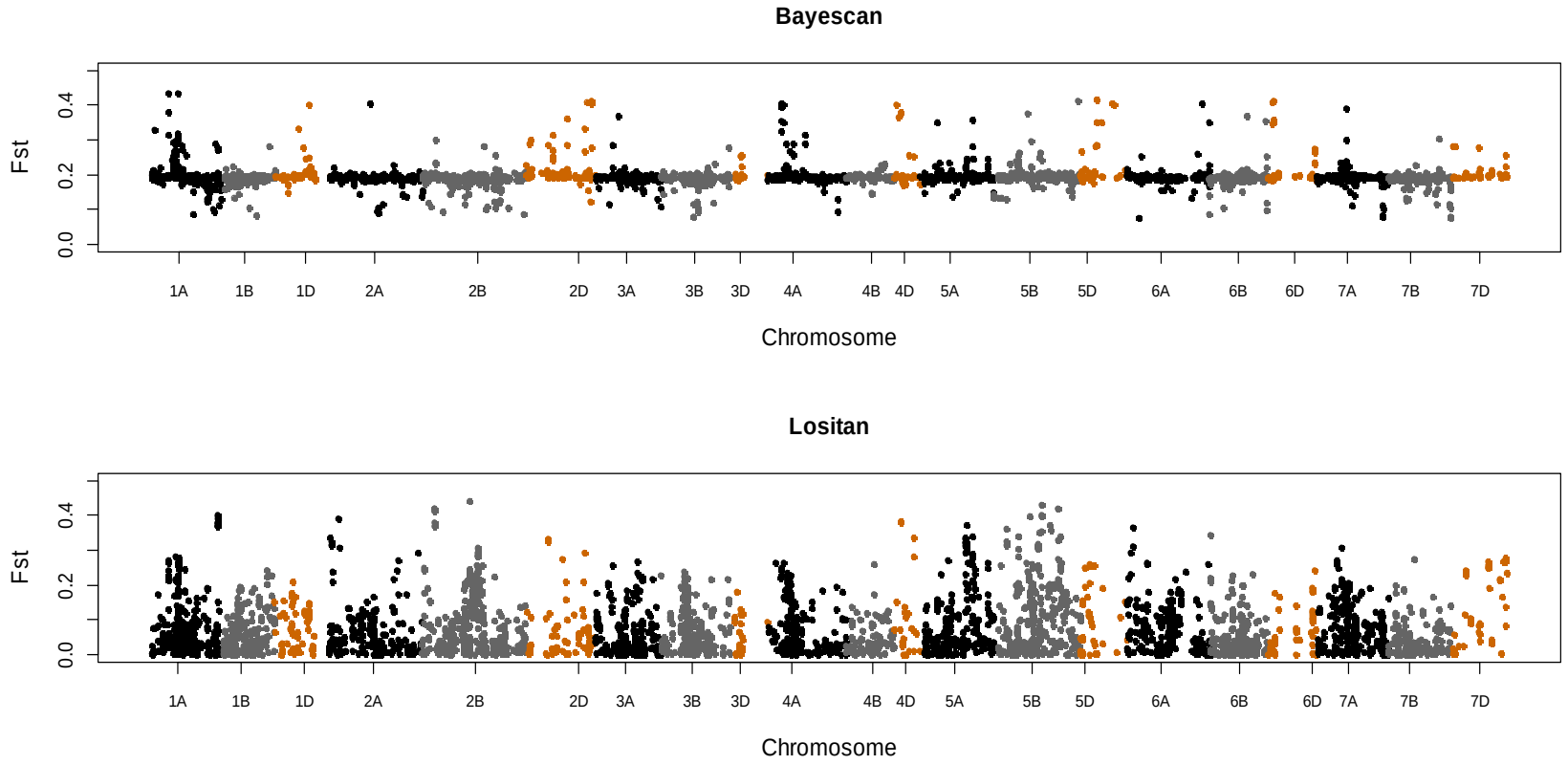


Figure 7 Manhattan plots of the F_{st} values for 5962 SNP loci along each of 21 chromosomes of wheat estimated through Bayescan and Lositan methods. Several chromosomes are associated with a few SNPs at elevated F_{st} (*e.g.* above 0.3), most evident for Lositan scan.

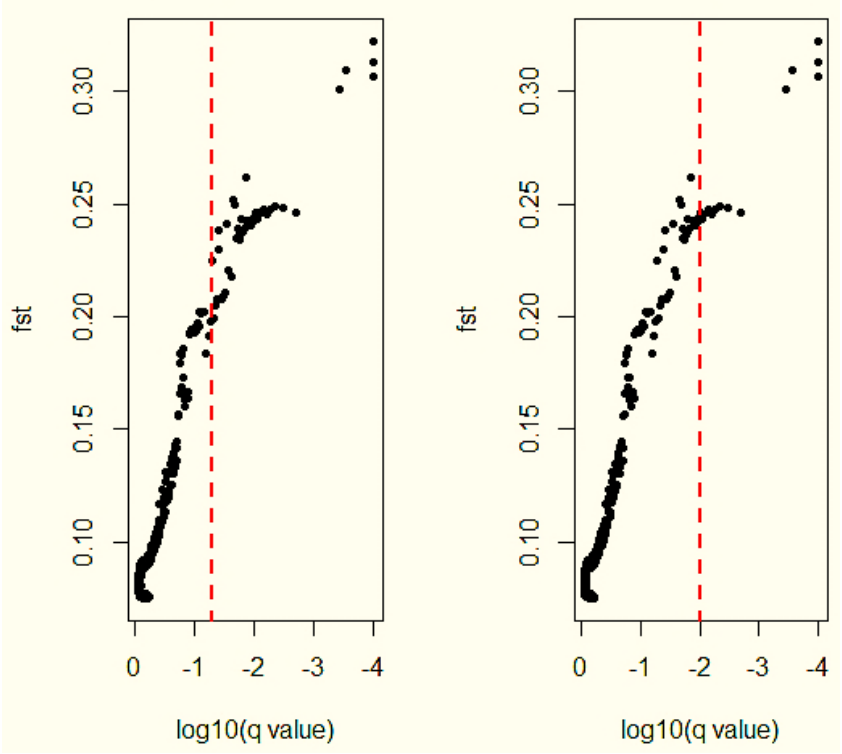


Figure 8a Genomic scan to detect loci under selection by Bayescan method amongst East African cultivars grouped based on three *ad hoc* release eras (before 1940, between 1940-1980, after 1980). Each point corresponds to a SNP locus. F_{st} is plotted against posterior odds i.e. log10 (q value). Evidence for putative loci under selection is indicated by the red hashed line. Left and right plots are decisive thresholds at FDR=0.05 and FDR=0.01 respectively.

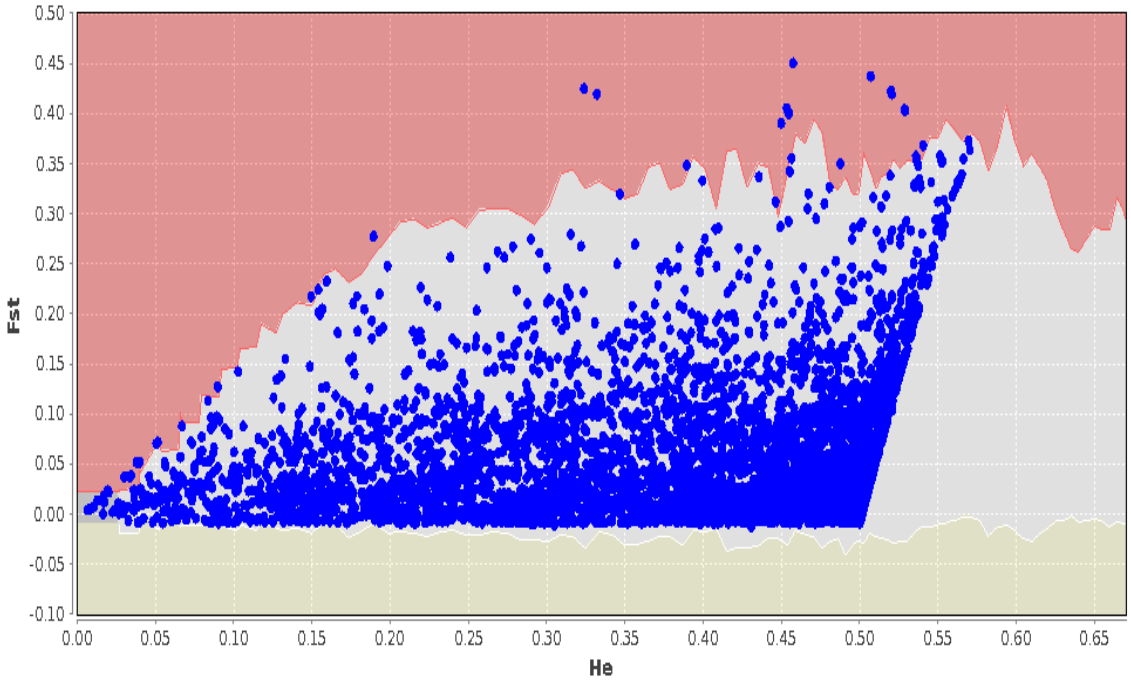


Figure 8b Genomic scan to detect loci under selection by FDIST2 approach implemented in Lositan amongst East African cultivars grouped based on three *ad hoc* release eras (before 1940, between 1940-1980, after 1980). F_{st} is plotted against expected heterozygosity (H_e) where each point corresponds to a SNP locus. FDR=0.01 is used as the decisive threshold to infer selection. Dots in salmon and green backgrounds represent candidate loci for positive and balancing selection respectively. Loci in the grey background are considered neutral.

CHAPTER 2

Patterns of linkage disequilibrium and a genome-wide SNP scan for QTL associated with rust resistance in East Africa bread wheat

Genome-wide association (GWA), which uses linkage disequilibrium to locate causal genes or genomic regions, has recently gained popularity as a means to study the architecture of complex traits. The power and resolution of genome-wide association mapping (GWAM) depends on the extent of LD, the non-random association between alleles at two or more loci. LD is a function of true physical linkages between loci or haplotypes but could also be an attribute of many genetic and population-specific demographic factors that either create novel LD or dissipate existing LD blocks. Profiling patterns of LD and estimating its decay over genetic distance in populations designated for GWAM provides invaluable information including hints on the marker density required to achieve the desired objectives in GWAM. In this study, a genome-wide set of 6487 informative single nucleotide polymorphism (SNPs) markers were used to derive insights on patterns of LD in a diverse panel of 300 spring wheat lines, mostly historical and commercially popular East African germplasm. Utilizing the mixed model regression approach, 16, 28, and 20 SNP markers significantly associated with resistance to adult plant resistance to stem rust, leaf rust and stripe rust respectively under diverse environments were identified.

INTRODUCTION

Rust diseases are among the major constraints to wheat performance and productivity worldwide. Epidemics of stem rust (*Puccinia graminis* Pers. f. sp. *tritici*), leaf rust (*Puccinia triticina*) and stripe rust (*Puccinia striiformis* Westend. f. sp. *tritici* Erikss) are recurrent events that cause significant grain yield losses and reduced grain quality (Samborski 1985, Roelfs et al. 1992, Line and Chen 1995). Control of these fungal pathogens to achieve stable wheat production has been a challenge to crop scientists for over a century (Singh 2012). Apart from directly sequestering assimilates for spore production, rust pathogens succeed in reducing crop yield by limiting photosynthesis due to diminished photosynthetic area (Roelfs et al. 1992, Rosewarne et al. 2012). Stem rust is deemed to be the most destructive of the three rust species since a crop that appears healthy three weeks before harvest can be devastated by explosive buildup of the disease if sufficient inoculum arrives from a heavily infected crop, even from a distant region (Leonard and Szabo, 2005). In favorable conditions, leaf rust can cause yield losses of up to 40% (Knott 1989, Zhao et al. 2008) while depending upon cultivar, earliness of the initial infection, rate of disease development and duration of disease, yield losses caused by stripe rust range from 10 to 70% (Chen 2005, Afzal et al. 2007).

Disease resistance is an important trait in both public and commercial plant breeding programs (Miedaner and Korzun, 2012) and breeding effectively involves critical resource allocation decisions. Considerable effort and resources are placed in screening breeding populations for rust resistance and incorporating it in finished cultivars. Detection of highly virulent races of wheat rusts as exemplified by stem rust

race TTKSK (*Ug99*) that was first reported in East Africa in 1998/99 (Pretorius et al. 2000), notwithstanding the variants within its lineage (Singh et al. 2011) underscores renewed and sustained need to breed wheat for resistance. The devastating outbreaks of stripe rust in 2010 that occurred throughout the major wheat growing areas in the Central, West Asia and North Africa region (Nazari et al. 2010), and the extremely dynamic South American leaf rust populations (German et al. 2007) equally prioritizes effort to identify and deploy rust resistance genes globally to mitigate against imminent yield losses.

Inheritance of wheat rusts can either be qualitative or quantitative. Qualitative resistance is recognized as being race-specific or *R*-gene resistance and is controlled by single or a few major genes (Kilpatrick 1975, Zhao et al. 2008) that show discontinuous distribution of infection types (Miedaner and Korzun, 2012). The term seedling resistance is often used to refer to this type of host response given that it is readily characterized in greenhouse seedling tests, and remains highly effective throughout the entire growth cycle of the crop (Ren et al. 2012). The *R*-genes follow the "gene-for-gene" relationship first proposed by Flor in 1956 (Singh, 2012). Host resistance requires the simultaneous interaction of the resistance allele in the host and the corresponding avirulence allele in the pathogen. This model that indicates the race- specific nature of *R*-gene resistance has been hypothesized to be based on an elicitor- receptor recognition mechanism (McDonald and Linde 2002). Past research through most of the past century has led to identification of numerous rust *R*-genes in wheat. At least 60 unique stem rust, 68 leaf rust and 49 stripe rust *R*-genes are catalogued (McIntosh et al. 2011). Additional rust resistance genes have been reported but have not been named. The catalogued *R*-genes are unique as indicated by different chromosomal locations, responses to disease races,

and wheat genotypes or wild species (Chen, 2005). Unlike qualitative resistance, quantitatively inherited resistance is non race-specific and shows continuous distribution of disease severity (Miedaner and Korzun, 2012). It has been termed "partial resistance'' (Parlevliet, 1975), "slow rusting resistance" (Singh 2012) and "adult plant resistance" (McIntosh 1992, Singh and Rajaram, 1993). This type of resistance is often characterized by lower frequencies of infections, lower uredinial size and less urediniospore production (Caldwell 1968, Chen and Line 1995, Liang et al. 2006) and in the cases of stem rust and leaf rust, absence of hypersensitivity (Singh 2012). Accordingly, this resistance describes genes that confer moderate to inadequate resistance under high disease pressure and rather than completely halt fungal infections, only slow the infection process in the adult plants but not seedlings (Rosewarne et al. 2012). In the present study the term adult plant resistance, abbreviated as APR, is preferred to describe quantitative resistance. To date, only a few well characterized, non-hypersensitive APR genes including *Lr34/Yr18*, *Lr46/Yr29*, *Lr67/Yr46*, *Lr68* and *Sr2/Yr30* are known to be effective in the field. A notable characteristic of these genes is that they often have pleiotropic effects on multiple rust diseases (Rosewarne et al. 2012, Singh 2012). Several recent studies (*e.g*. Crossa et al. 2007, Bhavani et al. 2011, Yu et al. 2011, and Yu et al. 2012) have identified and reported novel genomic regions associated with APR to wheat rusts.

Plant breeders desire resistance that is durable. Durable resistance is that which remains effective for a long period when applied on a large scale in a region that is undergoing regular epidemics of the pathogen (Johnson 1978). Although the multigenic APR is more difficult to evaluate, it is believed to be more durable than *R*-gene resistance (McIntosh 1992, Rutkoski et al. 2011). This assertion is supported by the sustained long

and wide spread deployment of known APR rust genes *e.g*. *Sr2* and *Lr34* (Ginkel and Rajaram 1993, McIntosh et al. 1995, Mago et al. 2011) compared to the ephemeral nature of *R*-genes that impart intense directional selection to rust pathogen populations leading to emergence of overwhelming virulent mutants and a typical "boom and bust" cycles of host resistance (McDonald and Linde 2002, Parlevliet 2002, Singh et al. 2004).

Selection of APR can be made effective if molecular markers linked to underlying genomic regions (i.e. quantitative trait loci or QTL) are detected in wheat populations and subsequently used in marker assisted selection (Miedaner and Korzun, 2012) and/or in genomic selection (Rutkoski et al. 2011). Detection of neutrally inherited markers in close proximity to the genetic causatives or genes controlling economically important complex quantitative traits, including APR is achieved through genetic mapping. The classical approach called QTL-mapping implements linkage analysis using experimental populations (also termed 'biparental' mapping populations) such as F_2, backcross (BC), double haploid (DH), recombinant inbred line (RIL) and near isogenic line (NIL). All these populations are derived from the genetic hybridization of two parental genotypes with an alternative trait of interest (Abdurakhmonov and Abdukarimov 2008). Rather than planned crosses, association mapping as a recent alternative approach (Soto-Cerda and Cloutier 2012) utilizes a diverse set of individuals. In plants, this method relies on historic patterns of recombination that have occurred within a collection of varieties, landraces, or breeders' lines (Vinod 2011). Though the terms linkage disequilibrium (LD) and association mapping have often been used interchangeably, Gupta et al. 2005 provides a distinction: association mapping, which essentially is one use of LD, refers to significant statistical association of a molecular marker with a phenotypic trait. LD refers

to non-random association between two markers or two genes/QTLs or between a gene/QTL and a marker locus. The utility of association mapping in hexaploid wheat was first demonstrated by Breseghello and Sorrells (2006), who successfully detected simple sequence repeat (SSR) markers linked with various grain milling quality traits. Subsequently, at least a dozen more studies have been implemented to dissect agronomic traits including those involved in adaptation, and to characterize many disease and insect resistances QTL in wheat (reviewed in Sajjad et al. 2012).

Association mapping has commonly been implemented following two approaches: (i) whole- genome scan (Rafalski 2002, Kraakman et al. 2004, 2006), which is also called genome-wide association study (GWAS) using high-density markers such as single nuclueotide polymorphisms (SNPs) and diversity array technology (DArT) that provide a reasonable genome coverage to identify genomic regions throughout the genome associated with the trait of interest; or (ii) candidate gene approach (Thornsbery et al. 2001, Szalma et al. 2005) which is more hypothesis- driven than GWAS (Vinod 2011), relies on prior knowledge about the biology (*e.g.* genetic, biochemical or physiological) of the trait of interest and genes underlying a QTL interval directly and tests the effects of genetic variants of the gene on the trait. The choice between GWAS and candidate gene approach is mostly dictated by the extent of LD in the organism under study since the extent of LD determines not only the mapping resolution that can be achieved, but also the number of markers that are needed for adequate coverage of the genome in GWAS (Vinod 2011).

The concept of LD, also referred to as "gametic phase disequilibrium" or "gametic disequilibrium" (Jannick and Walsh 2002, Gupta et al. 2005) was first described

by Jennings in 1917, and its quantification (*D*) developed by Lewtonin in 1964 (Abdurakhmonov and Abdukarimov, 2008). The LD measure *D* is the difference between the observed gametic frequencies of haplotypes and the expected gametic haplotype frequencies under linkage equilibrium ($D = P_{AB} - P_A P_B = P_{AB}\, P_{ab} - P_{Ab}\, P_{Ab}$) (Oraguzie et al. 2007). Other measures developed to quantify LD are D', r^2, D^2, D^*, *F*, *X* (2) and δ (Flint-Garcia et al. 2003, Gupta et al. 2005, Oraguzie et al. 2007). The r^2 statistic, the square of the correlation coefficient between two loci whose values range from 0 to 1, is the most appropriate LD quantification measure needed for association mapping (Abdalla et al. 2003, Gupta et al. 2005, Oraguzie et al. 2007) since it has more reliable sampling properties than D' in case of low allele frequencies (Oraguzie et al. 2007).

An "LD decay plot" which has the r^2 values plotted against genetic distance (cM) or weighted (bp) distance (Abdurakhmonov and Abdukarimov, 2008) provides a useful tool to estimate the portion of LD that is conserved with linkage and proportional to recombination (Stich et al. 2005, Gupta et al. 2005). LD conserved with linkage is very useful for association mapping but often there is significant LD between pairs of loci located far away from each other or even in different chromosomes that might cause false-positive associations between a marker (s) and a trait (Abdurakhmonov and Abdukarimov, 2008, Soto-Cerda and Cloutier, 2012).

Many genetic and demographic factors including new mutation, mating system, genetic isolation, population structure, relatedness (kinship), small founder population size or genetic drift, admixture, selection, epistasis, genomic rearrangements and gene conversion influence haplotypic LD blocks (Gupta et al. 2005, Abdurakhmonov and Abdukarimov 2008, Vinod 2011) . For instance, considering that the power of association

mapping can be strongly reduced as a result of population structure (Pritchard et al. 2000, Balding 2006, Zhao et al. 2007), control of this factor in several methodologies including the general linear model (GLM) (Pritchard et al. 2000), and the mixed linear model (Yu et al. 2005) is a prerequisite and is routinely implemented in GWAS.

The objectives of this study were to explore, quantify and provide a general overview of patterns of LD in a collection of mainly East African spring wheat designed for GWAS and to subsequently map and estimate the genetic effects of markers associated with APR to stem rust, leaf rust and stripe rust resistance in wheat.

MATERIALS AND METHODS

A comprehensive account of the plant material, genome-wide marker profiling, estimates of population structure and kinship are provided in Chapter 1 of this thesis. These resources are only briefly described in this Chapter.

Association mapping panel

A total of 290 accessions of spring wheat accessions, enriched mainly of lines released in East Africa (Appendix I) over a span of a century were chosen for evaluation of rust resistance. This panel excludes 7 lines that were inferred to be *durum* lines from population structure analyses.

Field plots and pathogen materials

Kenya (Njoro)

Field evaluations for stem rust resistance were conducted at Njoro in the 2011 main season (May to October) and in the offseason (November 2011 to April 2012). The main season is the wetter of the two seasons (Njau et al. 2010). Each line was sown on 75-cm wide raised beds in plots comprising two 70 cm rows, 20 cm apart with the 50 cm alleys. On each side of the plot in the middle of the alleys, a continuous row of a mixture of rust race *Ug99*- susceptible spreader wheat cultivars Chozi and Duma was sown. The experimental design was a randomized complete block (RCBD) with two replicates. A border of 8 rows of the spreader mixture was also planted to surround the field. To initiate an artificial stem rust epidemic, the spreader rows were inoculated using fresh urediniospores of stem rust race TTKSK (*Ug99*) collected from rust increase plots at Njoro. The urediniospores were first suspended in water and then injected into individual spreader plants prior to booting (growth stage Z35-Z37; Zadoks et al. 1974) at a distance

of approximately 1m. To further boost disease inoculum, spreader plants were also sprayed with a suspension of urediniospores in light mineral oil Soltrol 170 (Chevron Phillips Chemical Company, The Woodlands, TX) at least twice during stem elongation. No artificial inoculation was applied for stripe rust and evaluation for resistance to this disease was based on natural infection.

Minnesota (St. Paul, Crookston)

Rust evaluation plots were established at St. Paul in the spring wheat seasons of 2010, 2011 and 2012. The growing season at St. Paul is from April to August. Field plots were 2 m long single- rows with 20 cm row spacing and 2 m wide alleys planted in RCBD with two replicates. Spreader rows of a mixture of the highly rust susceptible wheat cultivars Thatcher, Morocco, and LMPG-6 were planted perpendicular to the test plots a week prior to planting the experimental plots. The spreader rows were inoculated with a mixture of *Puccinia triticina* races THBJ, MCDS and MBRJ. Both THBJ and MCDS are common in the current *P. triticina* population in the U.S. (Kolmer et al. 2007) while MBRJ was frequent in the 1990s (Long et al. 1998). To initiate stem rust epidemics, spreader rows were also inoculated with a mixture of common US *Puccinia graminis* f.sp *tritici* races QFCSC, QTHJC, MCCFC, RCRSC, RKQQC and TPMKC. Initially, inoculation was done at jointing (Z30 - Z39, Zadoks et al. 1974) and then repeated at the booting stage (Z35 - Z37, Zadoks et al. 1974). Race names follow North American rust nomenclature (Roelfs and Martens, 1988; Long and Kolmer, 1989).

Experimental plots at Crookston were similar to those at St. Paul. Evaluation for leaf rust resistance at this location was done in the 2012 spring wheat season with the

wheat test lines planted in RCBD with two replicates. Spreader rows were inoculated through injection using the same races as in the St. Paul environments.

Mexico (El batan, Toluca)

Field plots were established at CIMMYT's research fields at El batan and Toluca for leaf rust and stripe rust evaluations respectively in the 2011 crop season. The season in general runs mid-May to early- October, only slightly later at Toluca. At both locations, test plots were planted in double 1m long rows, with a 20 cm inter-row spacing, on top of 80 cm wide raised beds. In between beds, a 0.5 m wide alley was created. Experiments were planted as RCBD with 2 replicates. Disease spreader at both locations were seeded to the highly susceptible cultivar Morocco planted as hill plots on one side of the experimental plots adjacent to the alleys, and as long rows around the experimental block. Additionally, the spreader used to initiate stripe rust epidemics at Toluca also included eight lines in Avocet background that are highly susceptible to Mexican stripe rust races.

Leaf rust epidemics at El batan were initiated by spraying the spreader plants four weeks after sowing with urediniospores of the Mexican race MCJ/SP (Singh, 1991) suspended in soltrol oil, once a day for three consecutive days. Race MCJ/SP has avirulence/virulence formula *Lr2a, 2b, 2c, 3ka, 9, 16, 19, 21, 24, 25, 28, 29, 30, 32, 33, 34/1, 3, 3bg, 10, 11, 12, 13, 14a, 14b, 15, 17a, 18, 20, 22b, 23, 26, 27+31* (Singh, 1991). Disease spreader plots at Toluca were inoculated 4 weeks after planting with a mixture of three races having a combined virulence to *Yr3, Yr9, Yr17, Yr27* and *Yr31*.

Phenotyping

Disease reaction for all rusts was rated using the modified Cobb scale (Peterson et al. 1948) when the susceptible check lines had reached maximum severity. Invariably, the growth stage of most lines in the period of disease rating was beginning of anthesis to anthesis halfway (Z60 - Z65; Zadoks et al. 1974). Disease severity was scored in values ranging from 0 - 100 based on a visual assessment of the percentage of plant tissue (stem and leaf sheaths for stem rust and leaf for leaf rust and stripe rust) covered by disease. Severity scores for each reading were combined with an infection type, which describes the observed host response to infection which were scored as described in Knott, 1989 as: "R" (resistant) which depicts small uredinia surrounded by chlorosis or necrosis; "MR" (moderately resistant) *i.e.* medium sized uredinia surrounded by chlorosis or necrosis; "MS" (moderately susceptible) *i.e.* medium to large compatible uredinia without chlorosis and necrosis; "S" (susceptible) *i.e.* large compatible uredinia without chlorosis and necrosis. An infection type "M" was given where there was a clear overlap between MR and MS infection types.

A coefficient of infection (CI) was computed by combining disease severity and host response data into a single value. Thus, severity was multiplied with a constant for host response (Yu et al. 2011) where immune = 0.0, R = 0.2, MR = 0.4, M = 0.6, MS = 0.8 and S = 1.0.

Phenotypic correlation and best linear unbiased predictors (BLUPs)

Phenotypic correlations (r_p) between environments for the respective diseases were estimated based on Windows MS Excel® package. For association mapping analyses, BLUPs of coefficient of infection data were obtained for each genotype across

environment using function '*ranef*' in the lme4 package (R Development Core Team 2011).

SNP genotyping and data analysis

All wheat lines were genotyped at the USDA-ARS genotyping laboratory in Fargo, North Dakota using wheat 9,000 iselect SNP assay. After quality control, 6487 polymorphic SNPs were availed. Further processing to eliminate SNPs with a minor allele frequency (MAF) < 0.05% generated 6012 markers for this study.

Estimate of population structure

Inference of population structure was based on two methods: 'Discriminant Analysis of Principal Components' (DAPC) (Jombart 2010), and 'Structure' (Pritchard et al. 2000). The latter method provided the *Q* matrix corresponding to four inferred subpopulations assuming admixed model with independent allele frequencies (Figure 4, Chapter 1 this thesis), that was then incorporated as a covariate as described in the association mapping section described further below.

Linkage disequilibrium (LD) analyses

Genome-wide LD analysis was performed across the A, B, and D wheat genomes for the complete association mapping set. This study was further extended to six subpopulations identified from DAPC based structure analysis (a subpopulation of 7 *durum* wheat individuals excluded). LD was estimated as squared allele-frequency correlations (r^2) between pairs of SNP markers according to Weir (1996) using the PowerMarker v 3.25 software (Liu and Muse 2005). To depict extent of LD between pairs of loci, r^2 values were plotted against inter-marker genetic distance (cM). Locally

weighed polynomial regression (LOESS) curves were then fitted into the scatter plot using function '*smooth.spline'* of R (R Development Core Team, 2011).

Six hundred and thirty SNP markers (A genome = 273, B genome = 251, D genome = 106) at a genetic distance of 50 cM or more, were selected to estimate background LD. At the $\geq$ 50 cM inter-marker distances, loci were assumed to be unlinked and significant LD would likely reflect that due to other genetic and demographic factors rather than actual linkage. More specifically, the 95th percentile in the distributions of r^2 of the selected loci was estimated as the threshold r^2 (Breseghello and Sorrels 2006) to assume LD attributable to linkage. At its points of intersection with the LD decay curves, the threshold r^2 plotted as a horizontal line in the LD scatterplot provided estimates of extent of LD. To further explore patterns of the distribution of LD haplotype blocks within the wheat chromosomes, heat maps of estimated r^2 were created using package LD heat map of R (R Development Core Team, 2011).

Statistical significance of LD

For each estimated r^2 comparison-wise error rate *P*-values (here referred as *P*-values) were computed in PowerMarker (Liu and Muse, 2005) by implementing the Monte Carlo approximation to Fisher's exact *P*-value. A false discovery rate (FDR) (Benjamini and Hochberg 1995) established at 0.01 was used to determine the statistical significance of obtained *P*-values.

Association analysis

To find genomic regions associated with rust disease resistance, the compressed mixed linear model (MLM) in TASSEL v.3.0 was implemented (Yu et al. 2006, Bradbury et al. 2007, Zhang et al. 2010). A *P*-value was generated by fitting each SNP

marker into the MLM that has the form, $y = X\beta + Qv + u + e$, where y is the vector of the phenotypic values (BLUPs), X is the vector of SNP marker genotypes, β is the vector marker fixed effects to be estimated, Q is population structure matrix (derived from Structure analysis), v is a vector of fixed effects due to population structure, u is the vector of random effects and e is the vector of residuals. The variance of u is derived as, Var (u) = $2KV_{g,}$ where K represents the relative kinship matrix inferred from genotypes based on the proportion of shared alleles and V_g is the genetic variance. The variance of e is derived as Var (e) = V_R, where V_R is the residual variance.

In order to investigate the effectiveness of controlling for population structure in the association analyses, a general linear model (GLM) of the form, y = $X\beta$ which has no adjustment for population structure (*i.e*. naïve model) was also fit to all data sets. Subsequently for both MLM and GLM results, quantile - quantile (Q-Q) plots of estimated -log10 (P) were displayed using the observed P values and the expected P values and compared for inconsistency.

Threshold for declaring significant QTL

The critical P-values for assessing the significance of marker-trait associations in the MLM were calculated based on positive false discovery rate (pFDR; Q values), a multiple test correction method proposed by Storey 2002. These P-values were found to be highly stringent, probably owing to the use of all markers as independent tests in the correction. Considering the likelihood of over adjustment of the P-values owing to the mutual dependency of SNP in LD (Johnson et al. 2010), and the potential risk of type II error, a more liberal criteria was used to determine threshold P-values for marker-trait association. As suggested by Chan et al. 2010 and Pasam et al. 2012, the bottom 0.1

percentile of the distribution of *P*-values obtained can be considered as significant. Following this criteria, threshold *P*-values of ≤ 0.005 which roughly corresponds to the bottom 0.1 percentile in the present GWAS, was assumed to declare significant QTL for disease resistance.

In a bid to provide complimentary summary of declared putative QTL, manhattan plots were generated using a script written in R (R Development Core Team, 2011). Significant markers were also inspected for correspondence with genomic regions known to harbor QTL or rust resistance genes based on consensus maps in the CMap database (http://ccg.murdoch.edu.au/cmap/ccg-live/).

RESULTS

Phenotypic variation

Continuous variation was observed for all traits (Figures 1, 2, and 3) consistent with quantitative inheritance of APR to rust diseases. Estimated variance components for genotype were highly significant ($P \leq 0.001$) for the three diseases (Table 1). At the same significance threshold, genotype x environment interaction effect was also significant. Higher stem rust infection was observed in the offseason nursery in Kenya (*i.e.* SR_Kenya 2012) compared to the other four environments (Figure 1). Moreover, the difference between this season and the main season which displayed the lowest stem rust infection could probably be attributed to the wetter conditions in the offseason that favored heavier inoculum buildup. Stem rust disease infection in St Paul in 2012 was also relatively low. High temperatures especially towards the final crop growth stages could have contributed to this outcome mainly by impeding the rate of reinfection. Median leaf rust infection was highest in St Paul in 2011 and lowest at the El batan 2011 environment (Figure 2). Indeed, both the El batan nursery and St Paul 2010 were the only where maximum leaf rust coefficient of infection didn't exceed 85% but for a few outliers. In general, stripe rust infection was low at both testing environments despite heavy disease pressure observed in neighboring fields. This discrepancy probably suggests effects of a major gene (s) at relatively high frequency in the tested lines. Correlations of disease infection (Table 2) were in most instances moderate to high (range 0.5 - 0.8). The lowest correlation was that for stem rust infection between Njoro main season (*i.e*. SR_Njoro 2011) and St Paul 2010. Low correlations could be a reflection of the dynamics of site-specific environmental factors or differences relating to time to maturity of the different

genotypes at the diverse environments that might shorten or lengthen the duration of exposure to inoculum.

Population structure

Results of population structure are discussed in the upstream study of the molecular diversity and sub-structuring of the same panel here used for GWAS, as presented in Chapter 1 of this thesis. Briefly, 7 subpopulations were identified through discriminant analysis of principal components (DAPC) and were arbitrarily named East Africa1, East Africa2, East Africa3, North America, Mixed Identity1, Mixed Identity2 and Tetraploid. Hereafter, the same nomenclature is adopted for description purposes after discarding the tetraploid group of 7 durum wheat lines. Implementation of the model based Structure analysis pointed to at least 4 subpopulations highly concordant with the DAPC set of subpopulations.

Distribution, significance and patterns of pairwise LD

Histograms of LD (r^2) among pairs of SNP markers in the A-, B- and D- genomes of the complete set of 290 mapping individuals are shown in Figure 4. Notwithstanding the genome under study, these distributions in general depicted a common pattern inclusive of two clear maxima. Firstly, the largest proportion of marker pairs had r^2 values of 0.1 or less including a few with $r^2 = 0$, as would be expected for loci separated at long genetic distances and/or those in separate chromosomes where generations of recombination may have effectively attenuated ancestral LD. Secondly, marker pairs in perfect LD ($r^2 = 1$) - suggesting genetic linkage and predictably other causes including population structure, were observed to form the other extreme but smaller class in the

distributions across the genomes. Lastly, all other classes of pairwise LD were represented by fewer almost equal numbers of SNP pairs.

In addition, for the complete set of germplasm, the number of SNP pairs found to be in significant LD (FDR< 0.01) were respectively estimated at 61, 64, and 62% respectively for the A-, B- and D- genomes respectively. At the subpopulation level, the Mixed Identity1 had the least number of SNP markers in significant LD across all genomes while the East Africa1 had the highest in the A- genome and East Africa3 in the B- and D- genomes (Table 3).

A narrow range of the highest extent of significant intra-chromosomal LD after averaging across all marker pairs was observed in the combined population. LD extended at a genetic distance of 43 cM in chromosome 2A- the highest for the A- genome, 47 cM in chromosome 5B- the highest for the B- genome and 46 cM in chromosome 7D- the highest for the D genome. With respect to wheat subpopulations, chromosomes 2A and 6A had the highest intra-chromosomal LD among the A- genome chromosomes (Table 3, Appendix III) while this was observed to be frequently highest in chromosomes 2B and 5B in the B- genome set of chromosomes. Chromosomes 2D, 5D, 6D, and 7D were associated with higher long range LD compared to other genome D chromosomes. More specifically, LD extended as far as 60 cM on average in chromosome 6A in the Mixed Identity2 subpopulation, 35 cM in chromosome 2B both in the East Africa2 and North America subpopulations and 45 cM in chromosome 7D of East Africa1 subpopulation (Table 3).

A heat map aided exploration of patterns of LD within individual chromosomes revealed genomic regions falling across the whole spectrum of r^2 values with an

overrepresentation of regions with no or low LD. Haplotype blocks of high LD regions were discerned in the heat maps of all 21 chromosomes (Appendix V) as illustrated by chromosome 3B in Figure 5.

To further dissect patterns of LD for the complete mapping panel and its inherent subpopulations, LD decay curves fitted to scatterplots of r^2 were scrutinized. The 95th percentile of statistically significant r^2 was roughly 0.25 when averaged across subpopulations (Figure 6 and 7). In the A- and D- genomes, LD declined to 50% of its initial value (r^2 = 0.45 and r^2 = 0.46 respectively) at 17 cM in the entire population compared to B-genome whose initial value (r^2 = 0.44) declined to 14.6 cM (Figure 6). In the case of subpopulations, the level of initial LD varied considerably, more so in the A-genome (Figure 7). Values ranged from r^2 = 0.17 East Africa1) to r^2 = 0.58 (Mixed Identity1) in the A- genome; r^2 = 0.43 (East Africa1) to r^2 = 0.55 (Mixed Identity1) in the B-genome; and r^2 = 0.44 (Mixed Identity2) to r^2 = 0.58 (Mixed Identity1) in the D-genome. Initial LD declined fastest in the East Africa1 subpopulation reaching 50% at 13 cM (Table 4). This decay in LD, which perhaps reflects variation in subpopulation specific recombination processes was also notably rapid in the B- and D- genomes of the North America subpopulation where it declined to 50% at 12.5 cM and 13.5 cM respectively.

Association mapping

Population structure as a covariate

As reflected in the Q-Q plots (Figure 8), inclusion of population structure as a covariate in the MLM relative to the naïve model was highly effective in aligning each marker's observed *P*-value to the expected *P*-value under the assumption of uniform

distribution and no associations. These results were especially remarkable in the stripe rust resistance mapping implementation where *P*-values as low as 2.88 x 10^{-13} (*i.e.* $-\log_{10}(P) = 12.6$) were abridged to higher values *e.g.* 3.77 x 10^{-5} (*i.e.* $-\log_{10}(P) = 4.42$) in the current instance. Control of population structure achieved through MLM, thus proved most efficient in testing the null hypothesis of no marker-trait association vis-à-vis the alternative of significant marker-trait association present. In the Q-Q plots with MLM, SNP markers that deviate from the straight line (uniform distribution) are probably truly associated with the observed rust resistance phenotype.

Marker-trait associations

Stem rust resistance

A total of 16 SNP marker loci were found to be significantly ($P \leq 0.005$) associated with adult plant resistance (APR) to stem rust across the 5 testing environments. However, like was the case with leaf rust and stripe rust mapping, genomic regions declared significant did not survive the stringent thresholds invoked by multiple testing methods. In all cases very high pFDR (Q-value) were associated with the declared APR QTL which was attributed to probably an overly stringent control of population structure in the MLM and the small effects of identified markers. As such, SNP markers found significant were cautiously viewed as suggestive of QTL. With the exception of two significant markers whose map locations are unknown, all stem rust resistance associated SNPs were distributed across 7 chromosomes (Table 5a, Figure 9) and each explained between 2.9 to 4.7% (total = 56.4%) of the observed variation in APR resistance to the stem rust pathogen across environments. The SNP marker wsnp_CAP11_c3742_1796552 with a minor allele frequency of 0.07 is mapped at 12 cM

of chromosome 3B and probably represents the durably resistant APR gene *Sr2* found in the same chromosome arm in individuals descended from Yaroslav emmer wheat.

Leaf rust resistance

APR to the leaf rust pathogen across 5 testing environments was significantly ($P \leq 0.005$) associated with 28 SNP markers distributed among 10 chromosomes (Table 5b, Figure 9). These SNPs were found to explain between 2.7 and 5.6% (total= 72.8%) of the observed variation in all environments. Three of the significant markers have no known map location whereas on the basis of the wheat consensus map, four SNPs mapped closely to respective leaf rust resistance genes *Lr13* (chromosome 2B), *Lr19* (chromosome 6D), *Lr34* (chromosome 7B) and a QTL (chromosome 2D) (Figure 10). *Lr34* is among the few catalogued and widely characterized leaf rust APR genes.

Stripe rust resistance

Cumulatively across the two testing environments, 20 SNP markers declared to be significantly ($P \leq 0.005$) associated with stripe rust resistance explained 80.9% of the observed phenotypic variation. Individually, each of the SNPs explained between 3 to 6.4% of the total variation. These markers were distributed across 11 wheat chromosomes with the exception of two SNPs of unknown map location (Table 5c, Figure 11). A comparison with the wheat consensus map didn't yield any coincident QTL. However, the marker wsnp_Ku_c23305_33210628 at 63 cM of chromosome 2B matched close enough to the major gene of stripe rust resistance *Yr27.*

Signal for pleiotropy

Four genomic regions *i.e.* 91 - 94 cM, 155 - 162 cM, 24 - 26 cM, and 138 cM of chromosomes 2B, 3A, 4A, and 5A respectively were found to be associated with

resistance to both stem and stripe rusts, suggestive of pleotropic gene effects. Similarly, four regions 91 - 98 cM, 39 - 43 cM, 216 - 217 cM on chromosomes 3B, 5B, 6A and 7A overlapped in conferring resistance to leaf and stripe rusts. Following the same hypothetical model, only one region 110 - 111 cM on chromosome 2B was found to bear resistance for stem and leaf rusts while no region conferred resistance across the three rust types.

DISCUSSION

A key component of the genotype that a wheat breeder must be prepared to deliver is disease resistance to one of the three rusts of wheat (McIntosh, 1988). A hallmark of this concept is the success breeders have earned over the last century in developing rust resistant cultivars for different global environments, many having one or a combination of the numerous known *R*-genes. Recently, Sharma et al. 2012 have demonstrated the impact that breeding for rust resistance involving incorporation of high level resistances in high yielding lines have had in contributing to the overall genetic gain in CIMMYT elite spring wheat lines. In their work, these authors present a case for the role of additive minor genes that have been accumulated to confer near rust immunity in certain genotypes and highlighted the additional contribution of genes from the secondary genepool via CIMMYT's synthetic wheat program. Notwithstanding past success and considering mutable pathogens, particularly the rust fungi and the vulnerability they pose to world wheat productivity, identification of resistance through screening and tracking of resistance incorporated in wheat breeding populations remain of great priority globally. Essentially, the emergence and spread of new rust races as typified by the *Ug99* family of stem rust races (Singh et al. 2011) and the new races of stripe rust that may be adapted to unconventionally warmer temperatures (Milus et al. 2009) requires mobilization of scientific methods to break the barrier these diseases impose to "attainable yield" (Phillips 2010).

Association mapping, an application of linkage disequilibrium (LD) has been shown to be useful in dissecting resistance to wheat rust diseases, (Sajjad et al. 2012) particularly with regard to the quantitatively inherited APR. The effectiveness of whole

genome association studies for rust resistance as with other traits depends on the decay of LD initially present in a population at a rate determined by the genetic distance between loci and the number of generations since it arose (Mackay and Powell, 2006).

This study has highlighted patterns of intrachromosal LD in the bread wheat panel designed to map rust resistance genes/QTL. Using moderately sized genome-wide SNP data from the 9000 wheat SNP genotyping platform, extensive amounts of LD and its remarkable variation within chromosomes, among genomes and among subpopulations of the GWAS population has been elucidated. Overall, the extensive LD observed in the present mapping panel corroborate that self-pollinated plants have considerably high levels of LD (Gupta et al. 2005, Abdurakhmonov and Abdurkarimov 2008), that has been estimated to be about three orders of magnitude higher in wheat, a self-pollinating species, than in maize (*Zea mays* L.), an outcrossing species (Breseghello and Sorrels 2006). In barley (*Hordeum vulgare* L.) another selfing species, Zhang et al. 2009 reported that on average LD extends to 2.6 cM for a group of Canadian genotypes. In a population of 1816 elite US barley individuals, Hamblin et al. 2010 found average LD decay over a distance of 20 to 30 cM. In a collection of 102 tomato (*Solanum lycopersicum* L.) genotypes, Robbins et al. (2011) reported that LD decayed within over 6 -14 cM in contemporary processing varieties and 3 -16 cM within fresh market varieties.

Although the results of LD pattern assay will vary with the marker type utilized which may limit direct comparison of findings from different reports, the results of the extent of LD observed in this study is evidently largely congruent with similar works for hexaploid wheat. By assaying 170 lines derived from CIMMYT elite spring wheat yield trials (ESWYTs) using 813 DArT markers and 831 other markers including SSRs,

AFLPs and RFLPs, Crossa et al. (2007) observed an average length of 9.93 cM in LD blocks and that which extended approximately 65.0 and 87.4 cM in chromosomes 1B and 4B respectively. In their study of 189 Canadian elite bread lines based on 337 SSR markers, Somers et al. 2007 reported significant LD within a distance of 10 cM on most chromosomes except 1B, 3D, and 7D. These authors observed the highest extent of 41.2 cM on the short arm of chromosome 5A. Chao et al. (2010) utilized 394 SNP markers to survey LD patterns in a population of 478 spring and winter populations that span various US breeding programs and CIMMYT. In their report, these researchers detected LD of 20.8 - 27.1 cM in the D -genome, 20.4 - 21.5 cM in the B- genome, and 16.7 - 19.8 cM in the A- genome. Furthermore, Neumann et al. (2011) detected variable LD either along a linkage group or across the genome in a panel of 96 diverse winter wheats genotyped with 874 DArT markers. LD extending over 45 cM was observed on chromosome 7A.

The consensus of all the cited studies and now including the present research is that variation in LD patterns in hexaploid wheat is a function of population, genomic region and marker type but generally decay over fairly long genetic distances. Chao et al. (2010) suggested that observed divergence in the extent of LD between eight subpopulations of spring wheat studied might be due to unique breeding histories and selection pressures targeted to genes located in the different genomes during the process of cultivar development. A similar explanation can inform the differences observed among the six subpopulations in this study for the level of initial LD and the rate of its decay. With the exception of the Mixed Identity1 and Mixed Identity2 subpopulations, which could not be readily ascribed to specific descriptions, the other four subpopulations either originate from East Africa or North America (mainly Minnesota). In addition, the

East Africa1 subpopulation comprises of many modern Kenyan and Ethiopian cultivars and generally those released in the region after 1977. A considerable proportion of individuals in this subpopulation are derived from CIMMYT. The East Africa2 subpopulation in general are individuals released in the region between 1960 - 1977 while East Africa1 subpopulation is enriched of lines cultivated prior to 1960 including some landraces. This arbitrary temporal and spatial distinction of the six subpopulations might associate with specific breeding histories and selection pressures. East Africa1 subpopulation had the lowest initial LD across genomes and alongside North America subpopulation had the fastest decay. As pointed above, East Africa1 subpopulation includes most of the recent cultivars. It can be assumed that extensive crossing schemes used in developing breeding populations at CIMMYT as well as in the East African breeding programs might have allowed this subpopulation to capture more recombination events, subsequently removing not so tightly linked marker associations. On the other hand, high initial LD that decays slowly as observed in the Mixed Identity1 and Mixed Identity2 subpopulations might indicate fewer meiotic events to dissipate weak LD between loci. Additionally, unlike the other four subpopulations that had at least 60 individuals each, the Mixed Identity1 and Mixed Identity2 subpopulations had 8 and 12 genotypes respectively. Small population size is a factor that contributes to long-range LD considering that such subpopulations might have derived from a small number of founders, in which case the haplotypes present in the founders are more frequent than expected under equilibrium (Vinod, 2011).

Chromosomes 2A and 6A in A- genome, 2B in B- genome, and 7D in D- genome more frequently identified with the highest extent of LD across subpopulations in this

study. Notwithstanding contribution of possible demographic factors to this pattern, it is easy to implicate the impact of genes, more so those associated with adaptive traits that have been introduced to specific genomes during the process of cultivar development. It has been established that through a hitchhiking effect, where alleles get a lift in frequency from selection from neighboring alleles (Smith et al. 1974), LD may be generated among markers around the selected locus (Mackay and Powell, 2007). For instance, to adapt wheat cultivars to diverse ecologies the crop's breeding history has involved successive introduction of genes that control flowering including the *Vrn*- and *Ppd*- series. Such linkage blocks of favorable genes/QTL deliberately or inadvertently selected during the breeding of cultivars in the present mapping panel might have influenced the extensive LD observed in specific chromosomes. This outcome underscores the importance of controlling population structure in genome wide association mapping for rust resistance to ensure that only valid marker-trait associations and not those arising from population structure are reported.

Variation in the extent of LD along the chromosome affect the number of SNPs that can be chosen to capture a large fraction of the allelic variation of all SNP loci (*i.e.* tagSNPs, Tsui et al. 2005) in each genomic region to ensure that causal mutations are in LD with neighboring SNPs (Chao et al. 2010). For the present wheat mapping population, LD was on average 0.74 (median = 0.90) in the A- genome, 0.72 (median = 0.96) in the B- genome, and 0.74 (median = 0.98) in the D- genome for SNP markers separated by 1 cM. de Bakker et al. 2005 suggest LD values of 0.8 or higher as an acceptable threshold for tagSNP selection. At least considering the median LD values in the present study estimates, the levels are sufficiently high to allow selection of tagSNPs.

Assuming a 3,500 cM hexaploid wheat map and by placing markers evenly at distances of 1 cM (*i.e*. 3500 tagSNPs), causative mutations for trait under study would be found at 0.5 cM from one of the flanking markers and have an approximate LD of 0.74 in both A- and D- genomes and 0.72 in the B- genome. Equivalently, each marker would explain 74% of the QTL variance in the A- and D- genomes and 72% in the B- genome. To capture a higher proportion of the QTL variance, a larger mapping population would also be desirable. Besides, to achieve finer mapping resolution of say 0.2 cM at an LD of 0.8 or higher, Chao et al. (2010) estimated a total of 17,500 marker density for the same length of map in the example above. A further consideration for increasing probability of a marker that is in high LD with QTL arises from the idea that a marker and QTL must have similar minor allele frequency to be in LD (Newell et al. 2011). In this study approximately 6000 SNPs were available and are considered sufficient for coarse mapping. Projecting into the future, the need for higher marker coverage in wheat GWAS would greatly benefit from the increased marker density promised by the genotyping by sequencing technique (Poland et al. 2012).

By utilizing the MLM approach, this GWAS has identified 16, 28 and 20 SNP markers associated with APR to stem rust, leaf rust and stripe rust respectively at $P \leq 0.005$. A drawback in these results is that none of the detected loci remained significant after multiple testing corrections. Methods for multiple testing assume independence of tests where significance thresholds are determined by considering all markers as representing unique tests. The recognition that pairs of markers may not exemplify different tests due to LD is widely acknowledged (Johnson et al. 2010; Muller et al. 2011) and the question of what significance threshold is appropriate for GWAS is

somewhat unresolved (Dudbridge and Gusnanto, 2008). One of the classical research efforts in the application of GWAS to map rust resistance loci in wheat is that of Crossa et al. 2007 involving 170 wheat lines and 813 DArT markers as cited earlier. A total of 275 DArTs (63 for stem rust, 90 for leaf rust and 122 for yellow rust) were found associated with disease resistance and were spread across all chromosomes except 6D. Yu et al. 2011 conducted a GWAS to find loci significantly associated with stem rust race *Ug99* resistance in 276 current CIMMYT spring wheats profiled with over 2000 DArT markers. This research found 15 loci to be associated with APR to stem rust. In a different study involving dissecting APR to *Ug99* in 232 winter wheat lines using over 1200 DArT markers, Yu et al. 2012 detected 12 significant resistance associated loci. In all the rust resistance GWAS cited, groups of novel resistance loci or of loci confirming previously reported genomic regions in bi-parental populations were suggested. The annotation of the SNP markers declared significant in the present study was restricted to comparisons made to the wheat consensus maps in the CMap database (http://ccg.murdoch.edu.au/cmap/ccg-live/). In this way, a few of the mapped SNPs were identified to co-locate with known genes and/or QTL for resistance. It is likely that many of the resistance associated SNPs reported here may correspond to those reported in bi-parental populations. This probably may most effectively be confirmed through a meta-QTL analysis. Concurrently, validation of the significant loci through use of mapping populations free of population structure, *e.g.* a nested association mapping (NAM) panel, is crucial to identify consistently significant markers that can be used in marker-assisted rust disease breeding, more so in East Africa. Towards this end, a NAM population created from eight Kenyan wheat lines with demonstrated APR to stem rust has been

established and efforts to map for this based on screening done in Kenya, South Africa and Minnesota is in progress.

CONCLUSION

Among the different advances needed to enrich wheat productivity in the East Africa region is to mine favorable alleles within local breeding populations as well as other germplasm collections. However for this to be done efficiently, an increased understanding of the molecular basis of key traits and an expanding of phenotyping and genotyping of available plant material is necessary. Knowledge of the structure and patterns of LD in a crop, as this study has demonstrated, provides one angle from which to reflect on the molecular basis of phenotypic variation including for instance APR to rusts because many of the factors that impact on observed phenotypes including rates of recombination, selection, population sub-structuring and size, genetic isolation *etc.* have a direct bearing to LD. It is also acknowledged that information accumulated from LD studies is useful in designing experiments that can best help in exploiting useful variation. For instance, through GWAS in the present work, it has been reiterated that markers linked to APR to wheat rusts can be identified for potential use in breeding for durable resistance through marker assisted selection. However, APR is indicated to be under the control of numerous additive minor effect genes and the power to detect those in GWAS can be curtailed by overcompensation for population structure in the mixed model. This poses a challenge considering the high costs per data point associated with phenotyping for APR to rust in wheat. Rather than estimate individual marker effects followed by identification of significant markers for APR, it appears that looking into the future, the option provided by genomic selection where all markers with an effect in the phenotype are considered in the model, disregarding significance, might be most useful in breeding for durable resistance to wheat rusts based on APR genes.

Table 1 Analysis of variance of maximum disease severities for stem rust (SR), leaf rust (LR) and stripe rust (YR) reaction. Trait broad sense heritability ($h_b^{2)}$ estimates are provided in the last column.

Trait	Source	*df*	Mean square	*F* values
SR	Subpopulation	6	316.7	1695.0***
	Lines	271	43.2	231.3***
	Environments	4	4523.5	24208.3***
	Lines x Environments	979	17.7	94.6***
	Replicates/ Environments	5	1.3	6.8***
	Error	1256	0.2	
LR	Subpopulation	6	1118.1	1232.2***
	Lines	288	124.6	137.3***
	Environments	4	24340.2	26.8***
	Lines x Environments	1020	47.6	52.4***
	Replicates/ Environments	5	4.3	4.7***
	Error	1314	0.9	
YR	Subpopulation	6	973.6	599.6***
	Lines	243	404.5	249.1***
	Environments	1	19248.5	11854.7***
	Lines x Environments	228	52943.0	143.0***
	Replicates/ Environments	1	71.4	43.9***
	Error	478	1.6	

*** Significant at $P = 0.001$

Table 2 Correlation coefficients (r_p) for stem rust (SR), leaf rust (LR) and stripe rust (YR) maximum disease severity between different environments.

Trait	**Environment**	**r_p between environment**			
		SR_Njoro 2012	SR_St Paul 2010	SR_St Paul 2011	SR_St Paul 2012
SR	SR_Njoro 2011	0.64	0.27	0.36	0.37
	SR_Njoro 2012		0.32	0.43	0.41
	SR_St Paul 2010			0.70	0.69
	SR_St Paul 2011				0.81
		LR_St Paul 2011	LR_St Paul 2012	LR_El batan 2012	LR_Crookston2012
LR	LR_St Paul 2010	0.64	0.66	0.56	0.56
	LR_St Paul 2011		0.63	0.46	0.52
	LR_St Paul 2012			0.57	0.54
	LR_El batan 2012				
		YR_Njoro 2012			
YR	YR_ Toluca 2011	0.42			

Table 3 Summary of marker loci (%) in significant LD (FDR< 0.01) and the genetic distance of the highest extent of significant LD (mean) for the A-, B- and D- genomes of 6 subpopulations inferred from the GWAS wheat panel. Chromosomes associated with the long range LD are indicated in parenthesis.

Sub-population	SNP marker pair in significant LD (%)			Highest extent of significant LD (cM)		
	Genome			Genome (Chromosome)		
	A	*B*	*D*	*A*	*B*	*D*
East Africa1	50	49	43	52 (6A)	34 (2B)	45 (7D)
East Africa2	47	51	50	40 (2A)	35 (2B)	40 (7D)
East Africa3	43	52	51	42 (2A)	34 (2B)	37 (2D)
Mixed Identity1	22	11	14	27 (2A)	19 (5B)	20 (5D)
Mixed Identity2	24	26	29	60 (6A)	20 (2B)	36 (6D)
North America	44	45	46	52 (6A)	35 (2B)	44 (7D)

Table 4 Levels of initial LD (r^2) and genetic distance (cM) at which 50% of it is decayed among 6 subpopulations in the East Africa wheat GWAS panel.

Sub-population	Initial LD (r^2)			Genetic distance of 50 % LD decay (cM)		
	Genome			Genome		
	A	*B*	*D*	*A*	*B*	*D*
East Africa1	0.17	0.43	0.45	13.1	15.3	14.4
East Africa2	0.55	0.50	0.54	19.8	15.3	16.3
East Africa3	0.47	0.54	0.54	16.2	16.3	15.3
Mixed Identity1	0.58	0.55	0.58	25.0	19.2	17.3
Mixed Identity2	0.31	0.54	0.44	33.9	16.3	19.2
North America	0.35	0.49	0.50	27.2	12.5	13.5

Table 5a SNP markers associated with stem rust resistance (suggestive of QTL) through mixed linear model GWAS

Marker	Chr[a]	Position[b]	P-value	pFDR[c]	R^2 (%)[d]	MAF[e]
wsnp_RFL_Contig373_3974834	1A	64	0.002	0.36	3.5	0.33
wsnp_Ex_c14273_22230844	1B	18	0.001	0.36	3.6	0.25
wsnp_RFL_Contig2794_2564017	1B	34	0.002	0.36	3.3	0.34
wsnp_Ku_c11665_18999583	2B	94	0.002	0.36	3.2	0.13
wsnp_Ra_c14034_22046454	2B	110	0.002	0.36	3.2	0.47
wsnp_JD_c1472_2090800	2B	119	0.003	0.36	3.0	0.47
wsnp_JD_c14405_14144807	2B	160	0.000	0.36	4.7	0.37
wsnp_Ex_rep_c104125_88923836	3A	162	0.004	0.36	3.0	0.30
wsnp_CAP11_c3742_1796552	3B	12	0.001	0.36	4.2	0.07
wsnp_Ex_c4752_8482625	4A	26	0.001	0.36	3.9	0.06
wsnp_RFL_Contig3939_4369467	5A	96	0.001	0.36	3.7	0.06
wsnp_Ex_c55777_58153636	5A	138	0.004	0.36	2.9	0.45
wsnp_Ku_c40084_48381107	5B	188	0.001	0.36	4.1	0.32
wsnp_CAP12_c2231_1090724	5B	197	0.002	0.36	3.3	0.44
wsnp_Ex_c1512_2889138	-	-	0.001	0.36	3.8	0.07
wsnp_Ex_c10545_17226769	-	-	0.004	0.36	3.0	0.09

[a] Chr, Chromosome harboring QTL

[b] Position in centimorgan of peak $-\log_{10}$ (*P*-value)

[c] pFDR, positive false discovery rate or Q-value (Storey 2002)

[d] R^2 (Coefficient of determination), Percentage of phenotypic variance explained by QTL

[e] MAF, minor allele frequency

Table 5b SNP markers associated with leaf rust resistance (suggestive of QTL) through mixed linear model GWAS

Marker	Chr[a]	Position[b]	P-value	pFDR[c]	R^2 (%)[d]	MAF[e]
wsnp_BG314157A_Ta_2_1	1A	93	0.002	0.42	3.2	0.33
wsnp_Ra_c22110_31550370	1A	116	0.004	0.52	2.8	0.26
wsnp_Ku_c1176_2350548	2B	111	0.004	0.52	2.9	0.07
wsnp_Ex_c28627_37743031	2B	202	0.000	0.30	5.6	0.39
wsnp_RFL_Contig1892_1042675	2B	212	0.001	0.42	3.8	0.13
wsnp_Ra_c5532_9788185	3A	3	0.005	0.53	2.9	0.20
wsnp_Ra_c9738_16174002	3A	16	0.002	0.42	3.2	0.50
wsnp_Ku_c11052_18135847	3A	41	0.002	0.42	3.2	0.17
wsnp_RFL_Contig2011_1216801	3A	75	0.002	0.42	3.2	0.50
wsnp_Ex_c20250_29303152	3A	136	0.000	0.38	4.4	0.23
wsnp_RFL_Contig383_4163148	3B	91	0.002	0.42	3.3	0.07
wsnp_Ex_c6245_10887043	3B	105	0.005	0.53	2.7	0.23
wsnp_Ex_rep_c67054_65517671	4A	76	0.003	0.42	3.1	0.22
wsnp_Ex_c4148_7495656	4B	115	0.003	0.42	3.1	0.38
wsnp_Ex_c13777_21603644	5B	39	0.003	0.42	3.0	0.07
wsnp_JD_c10416_11077664	5B	64	0.001	0.42	3.9	0.21
wsnp_Ex_c54206_57165374	5B	102	0.005	0.53	2.7	0.09
wsnp_CAP12_c1663_836928	6A	189	0.000	0.30	5.2	0.39
wsnp_CAP12_rep_c4048_1842	6A	197	0.001	0.42	4.1	0.35
wsnp_Ex_c7546_12900094	6A	217	0.003	0.43	3.0	0.19
wsnp_Ra_c23253_32762188	7A	81	0.001	0.42	3.5	0.26
wsnp_JD_c12343_12604782	7A	96	0.003	0.42	3.1	0.24
wsnp_Ex_c19582_28564743	7A	101	0.003	0.42	3.0	0.18
wsnp_Ku_c5693_10079343	7A	169	0.003	0.42	3.1	0.25
wsnp_BF485305D_Ta_2_2	7D	36	0.002	0.42	3.2	0.06
wsnp_Ex_rep_c101523_8687949	-	-	0.002	0.42	3.4	0.09
wsnp_BQ168329A_Td_2_1	-	-	0.002	0.42	3.2	0.17
wsnp_Ra_c22017_31446553	-	-	0.003	0.42	3.2	0.08

[a] Chr, Chromosome harboring QTL

[b] Position in centimorgan of peak $-\log_{10}$ (*P*-value)

[c] pFDR, positive false discovery rate or Q-value (Storey 2002)

[d] R^2 (Coefficient of determination), Percentage of phenotypic variance explained by QTL

[e] MAF, minor allele frequency

Table 5c SNP markers associated with stripe rust resistance (suggestive of QTL) through mixed linear model GWAS

Marker	Chr[a]	Position[b]	P-Value	pFDR[c]	R^2 (%)[d]	MAF
wsnp_BE490041A_Ta_2_1	1A	43	0.003	0.20	4.9	0.33
wsnp_Ex_c3147_5816957	1B	89	0.001	0.20	4.5	0.21
wsnp_CAP11_rep_c4138_1957291	1B	124	0.002	0.29	3.6	0.10
wsnp_BE590634B_Ta_2_1	1D	49	0.001	0.25	4.0	0.46
wsnp_Ku_c23305_33210628	2B	63	0.001	0.20	4.8	0.41
wsnp_BE497494B_Ta_2_1	2B	76	0.004	0.36	3.5	0.20
wsnp_Ex_c7246_12443506	2B	91	0.003	0.32	3.5	0.22
wsnp_BE426418A_Ta_2_2	3A	155	0.004	0.32	3.5	0.12
wsnp_Ex_c33463_41948471	3B	98	0.003	0.36	3.2	0.10
wsnp_Ex_c8715_14590273	3B	126	0.004	0.32	3.4	0.10
wsnp_Ex_c22913_32130617	4A	24	0.001	0.25	4.0	0.08
wsnp_BF293620A_Ta_2_1	5A	123	0.002	0.28	3.7	0.28
wsnp_Ex_c8424_14192191	5A	138	0.005	0.36	3.1	0.33
wsnp_Ra_rep_c106477_90236168	5B	43	0.003	0.32	3.3	0.42
wsnp_Ex_c28973_38050174	6A	216	0.000	0.20	6.4	0.09
wsnp_BE590634B_Ta_2_5	7A	64	0.003	0.32	3.6	0.47
wsnp_Ku_c42539_50247426	7A	125	0.003	0.32	4.1	0.23
wsnp_Ex_c5839_10246915	7A	176	0.001	0.20	5.4	0.30
wsnp_Ra_c44957_51237572	-	-	0.001	0.20	5.4	0.28
wsnp_BE498419B_Ta_1_2	-	-	0.005	0.36	3.0	0.06

[a] Chr, Chromosome harboring QTL

[b] Position in centimorgan of peak $-\log_{10}(P\text{-value})$

[c] pFDR, positive false discovery rate or Q-value (Storey 2002)

[d] R^2 (Coefficient of determination), Percentage of phenotypic variance explained by QTL

[e] MAF, minor allele frequency

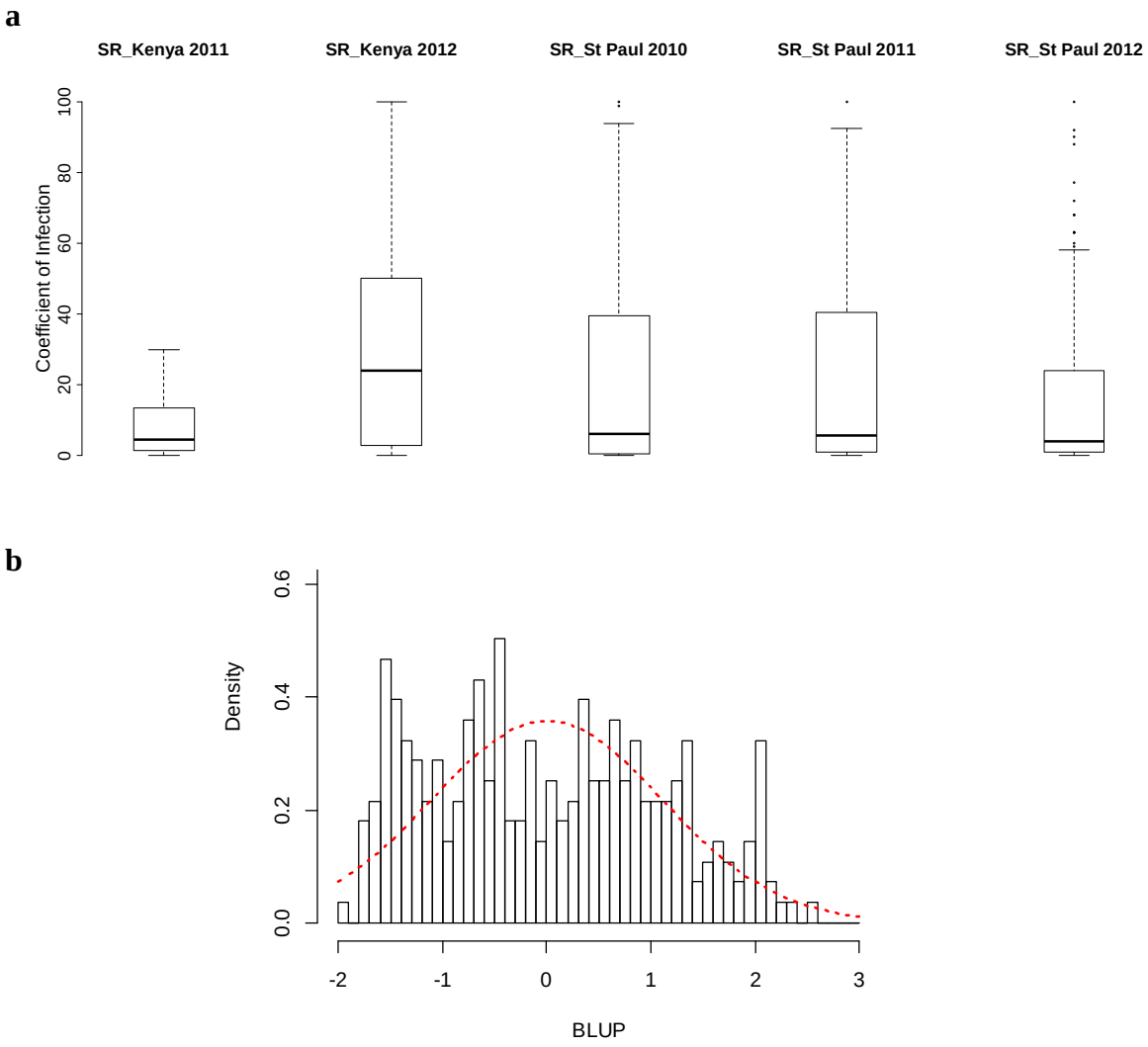


Figure 1 Phenotypic distribution of stem rust resistance. **(a)** Boxplots depicting differences in resistance summarized as coefficients of infection for the mapping population across five environments. Box edges represent the upper and lower quantile with median shown as the line in the middle of the box. Whiskers represent 1.5 times the quantile of the data with outliers shown as dots. (**b**) Histogram of BLUP values derived from stem rust coefficient of infection data across environments. The red dotted line within plot is the expected normal distribution.

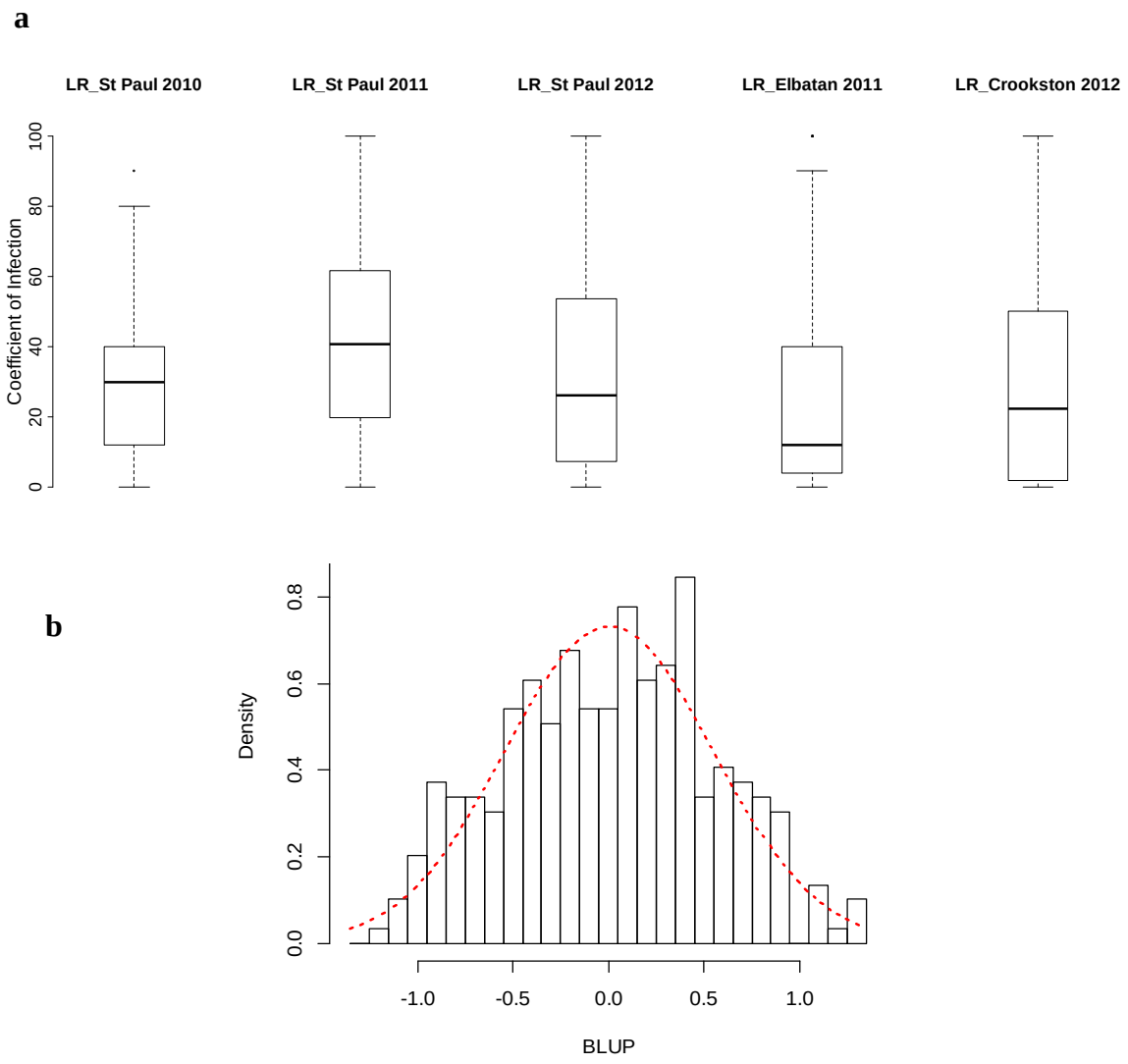
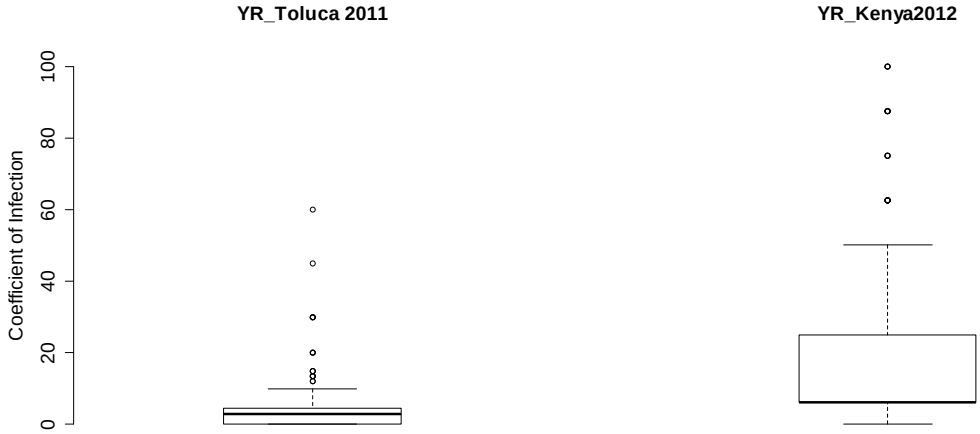


Figure 2 Phenotypic distribution of leaf rust resistance. **(a)** Boxplots depicting differences in resistance summarized as coefficients of infection for the mapping population across five environments. Box edges represent the upper and lower quantile with median shown as the line in the middle of the box. Whiskers represent 1.5 times the quantile of the data with outliers shown as dots. (**b**) Histogram of BLUP values derived from leaf rust data coefficient of infection across environments. The red dotted line within plot is the expected normal distribution.

a



b

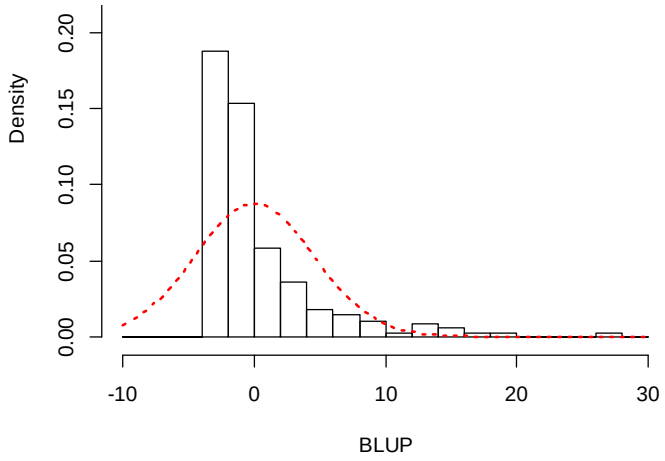


Figure 3 Phenotypic distribution of stripe rust resistance. **(a)** Boxplots depicting differences in resistance summarized as coefficients of infection for the mapping population across two environments. Box edges represent the upper and lower quantile with median shown as the line in the middle of the box. Whiskers represent 1.5 times the quantile of the data with outliers shown as dots. (**b**) Histogram of BLUP values derived

from stripe rust coefficient of infection data across environments. The red dotted line within plot is the expected normal distribution.

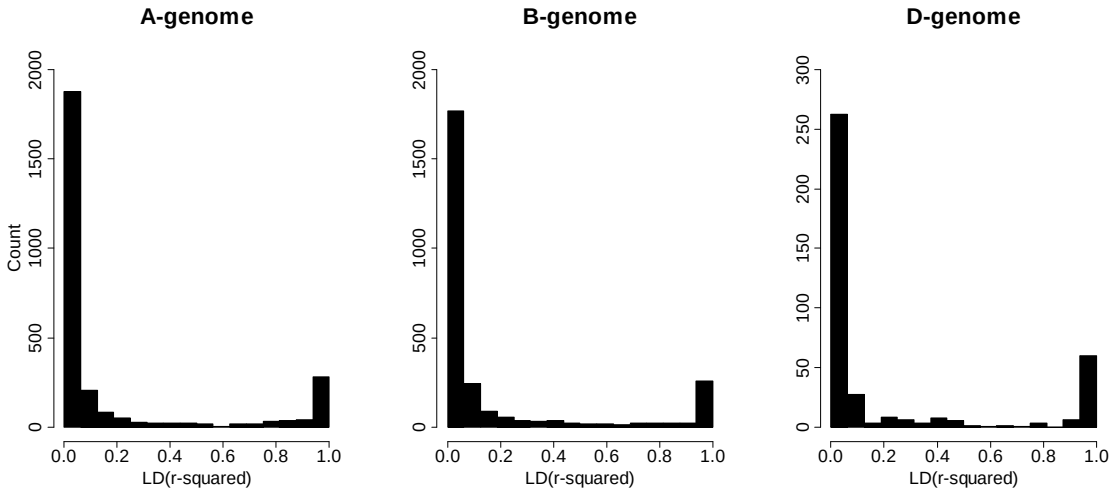


Figure 4 Distribution of r^2 values in the A-, B-, and D-genomes estimated for the East African wheat GWAS panel

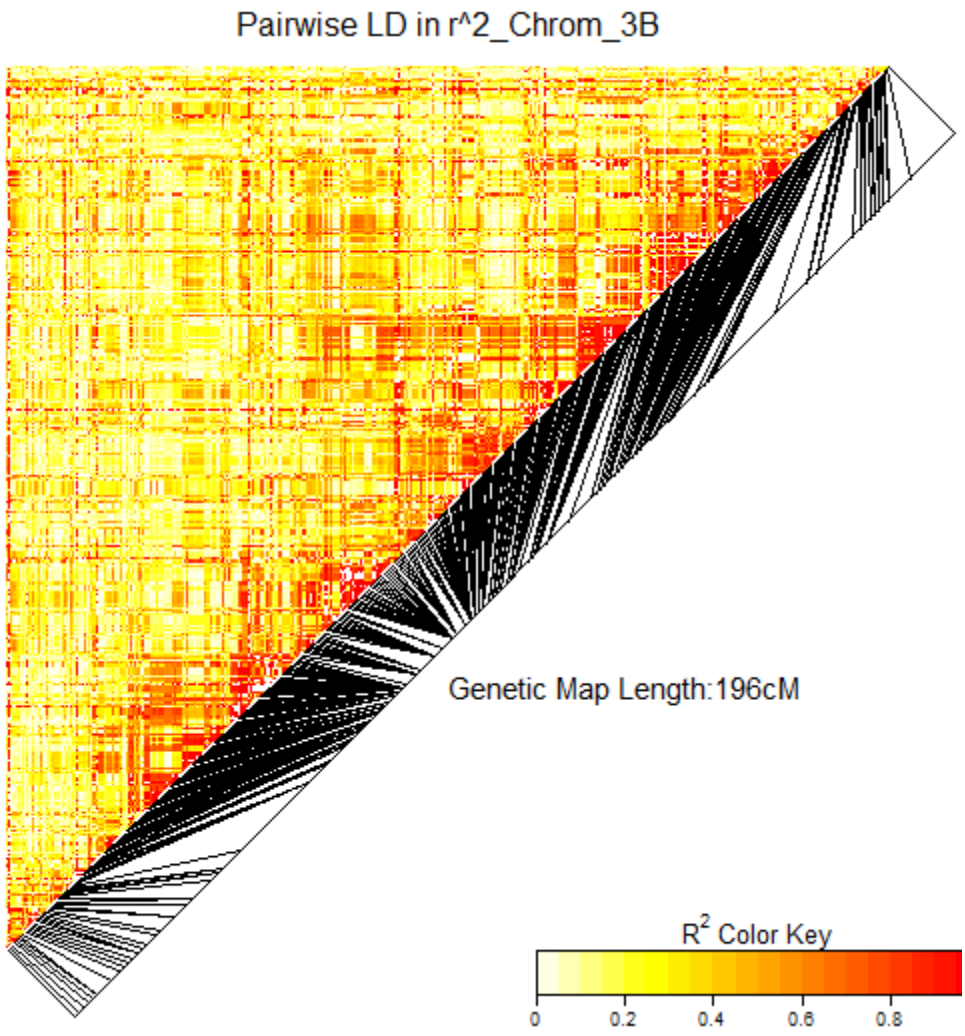


Figure 5 Pair-wise LD for wheat chromosome 3B exemplifying patterns of LD similarly observed across all chromosomes. Colored rectangles represent squared correlation (r^2) between SNP marker pairs. A color key is provided below the LD heat map showing r^2 values that corresponds to the color of the rectangle or region.

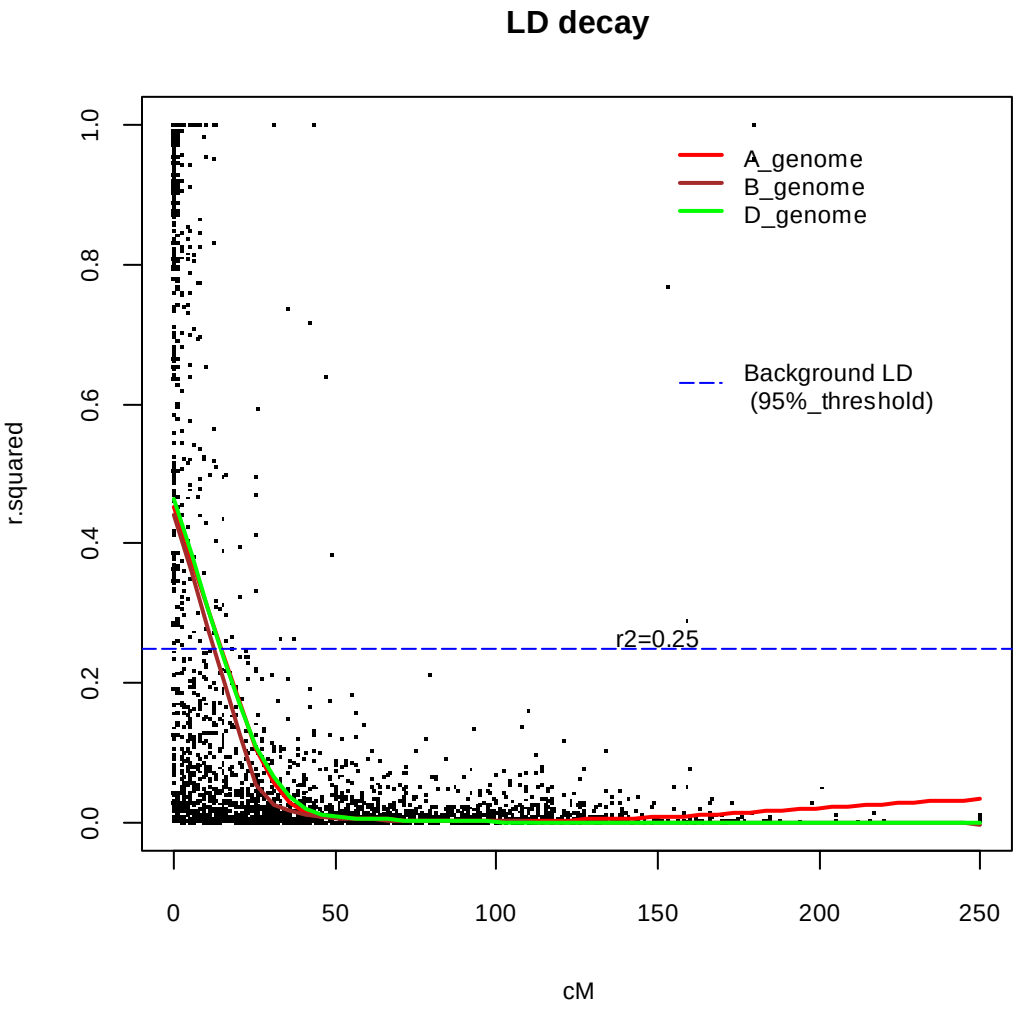


Figure 6 Scatterplots of the LD statistic r^2 as a function of genetic distance (cM) between pairs of SNP markers estimated for A-, B-, and D-genomes in the East African wheat GWAS panel. The locally weighted polynomial regression-based (LOESS) representing decay of r^2 along genetic distance is illustrated for each genome. LD critical threshold estimated from LD distribution of pairs of unlinked SNP markers is indicated by the dashed horizontal line.

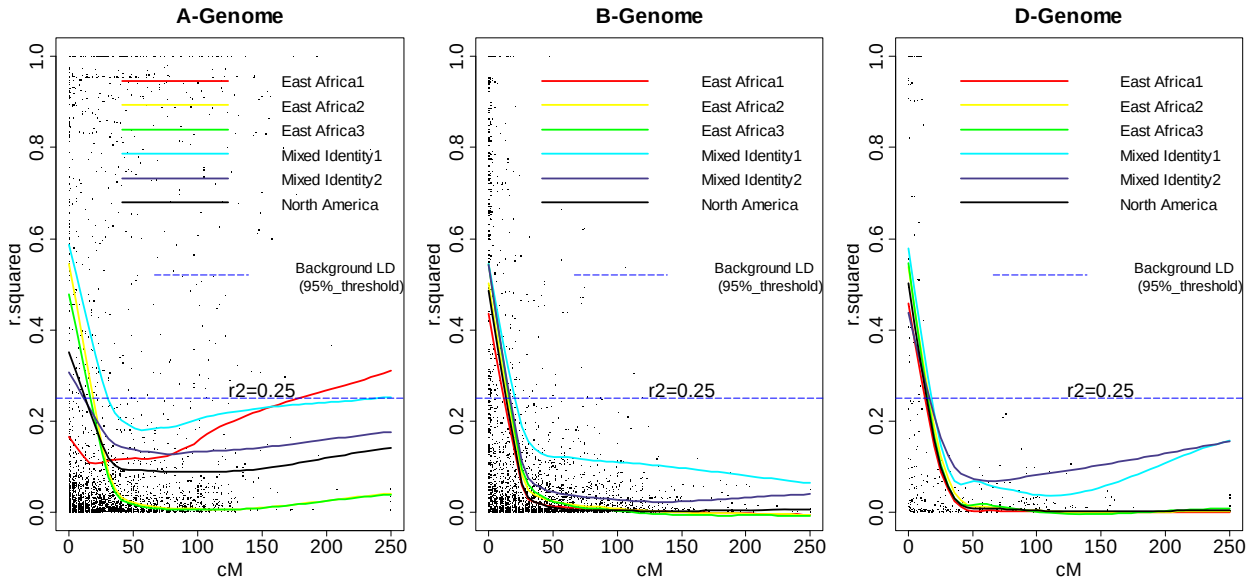


Figure 7 Scatterplots of the LD statistic r^2 as a function of genetic distance (cM) between pairs of SNP markers estimated for six subpopulations inferred in the East African wheat GWAS panel. For each subpopulation locally weighted polynomial regression-based (LOESS) representing decay of r^2 along genetic distance is illustrated in the A-, B- and D- genomes. LD critical threshold estimated from LD distribution of pairs of unlinked SNP markers is shown by the dashed horizontal line.

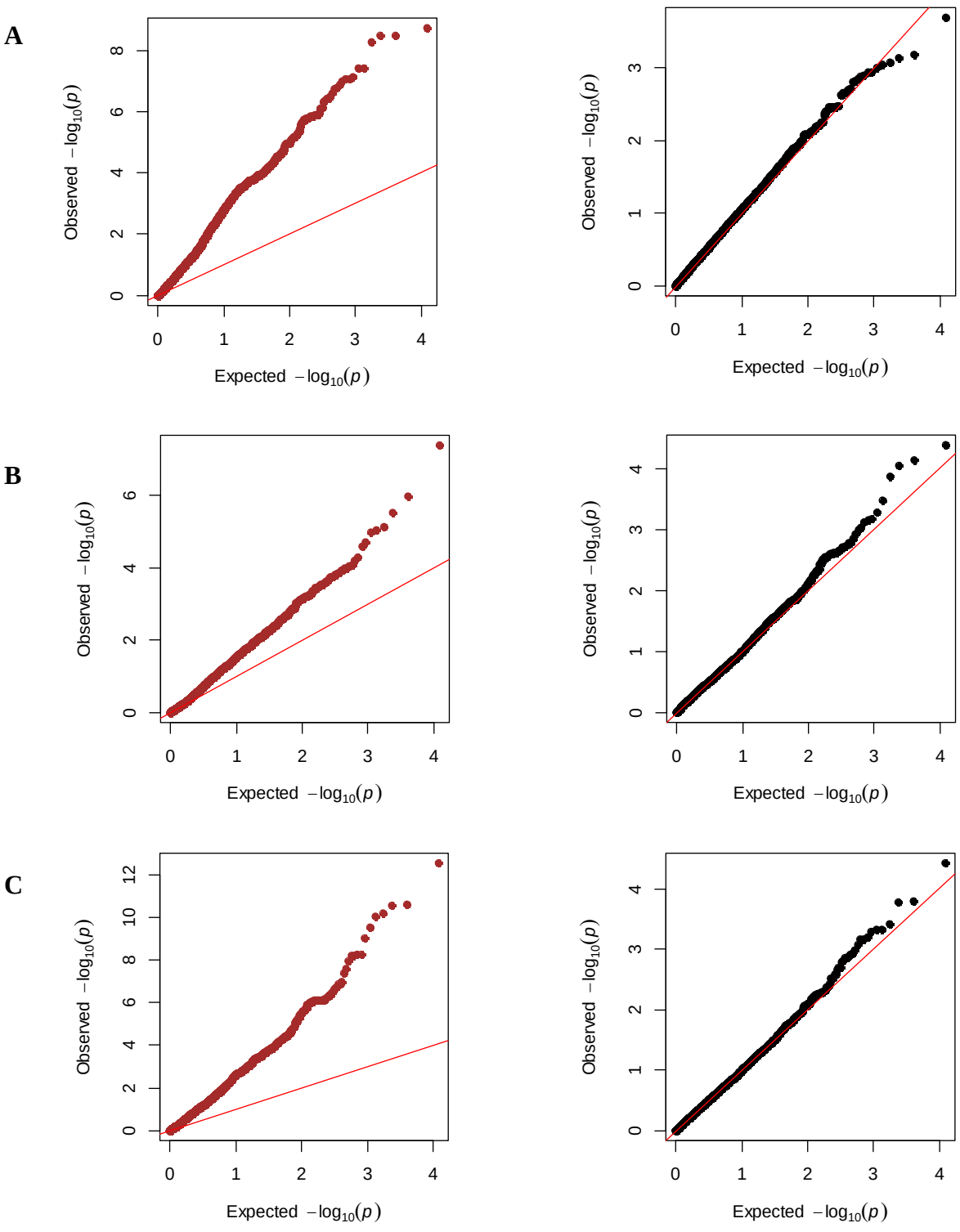


Figure 8 Quantile – Quantile plots for naïve (left panels) and mixed model (right panels) implemented in (**A**) stem rust-, (**B**) leaf rust-, and (**D**) stripe rust- GWAS.

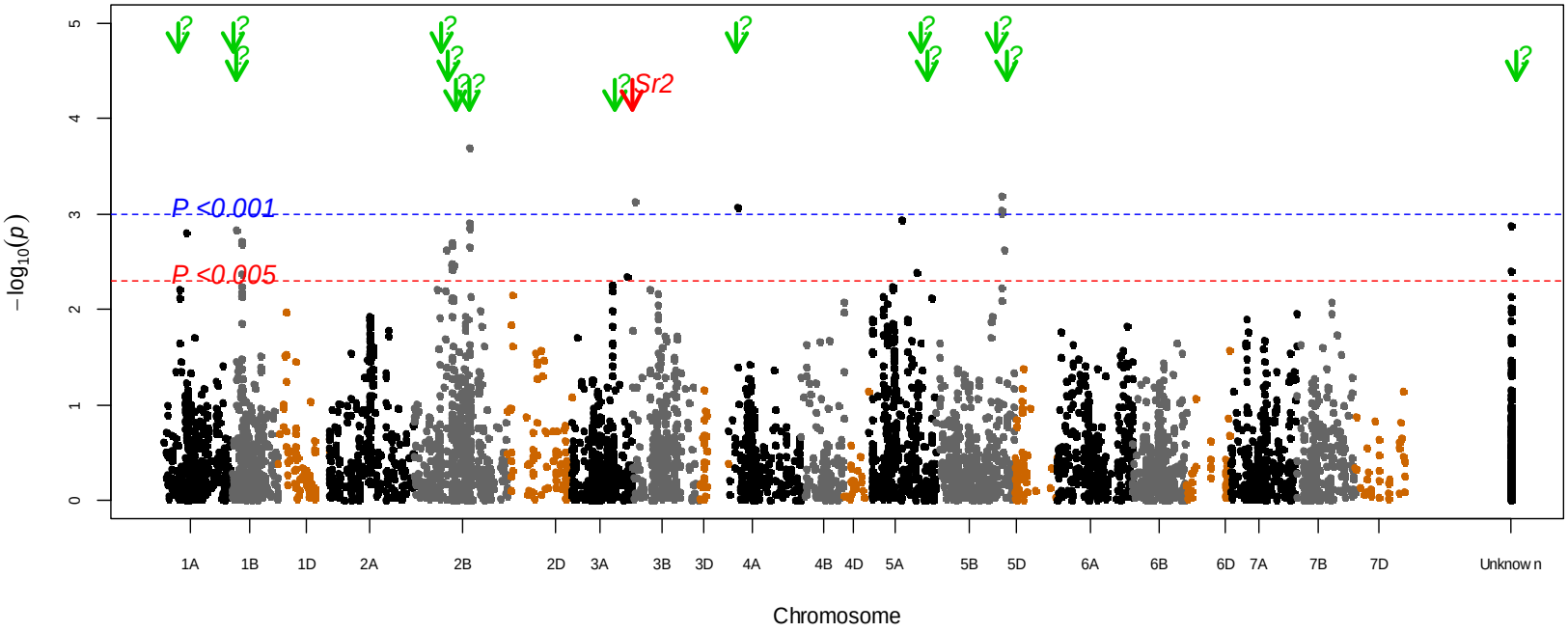


Figure 9 Manhattan plot indicating genomic regions putatively associated with stem rust resistance, marked by question marks and green arrows. The APR gene *Sr2* in chromosome 3B based on the wheat consensus map is shown in a red arrow. *x*- axis shows SNP along each chromosome; *y*- axis is the $-\log_{10}$ (*P*-value). Dashed horizontal lines designate thresholds for declaring significance.

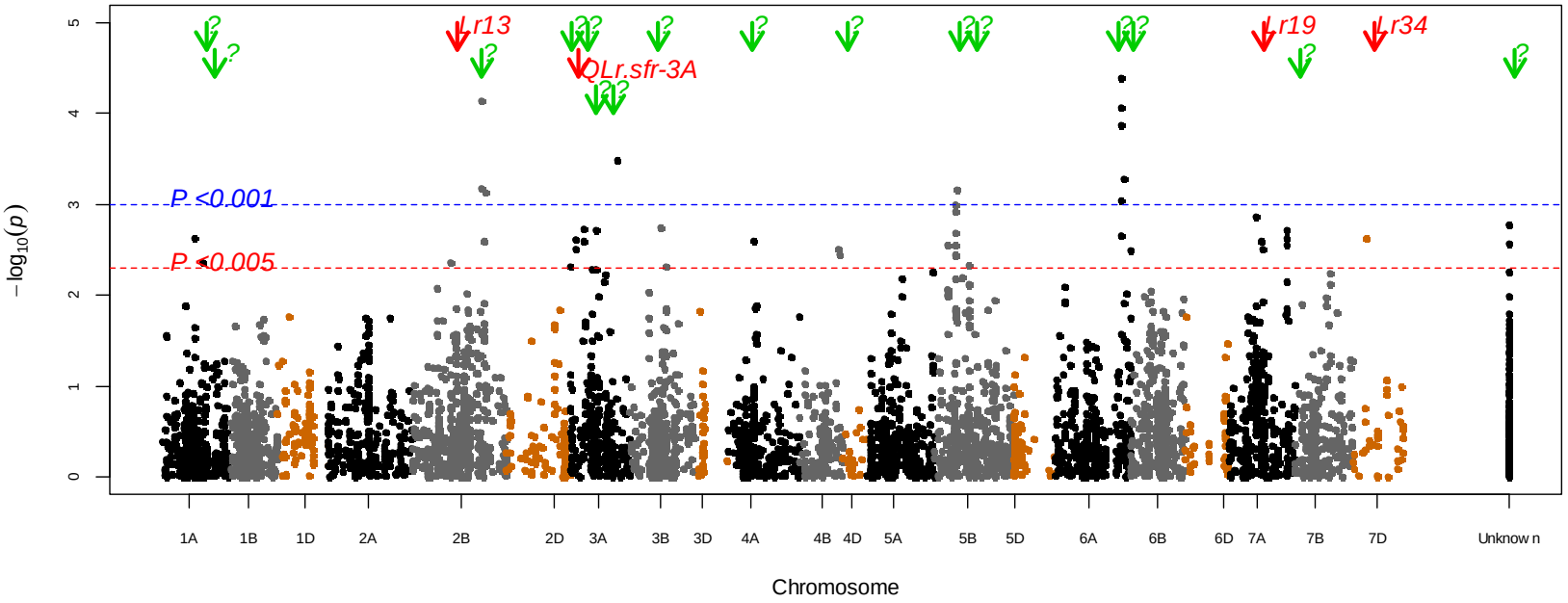


Figure 10 Manhattan plot indicating genomic regions putatively associated with leaf rust resistance, marked by question marks and green arrows. QTL and gene(s) from the wheat consensus map corresponding to mapped regions are shown in red arrows. *x*- axis shows SNP along each chromosome; *y*- axis is the $-\log_{10}$ (*P*-value). Dashed horizontal lines designate thresholds for declaring significance.

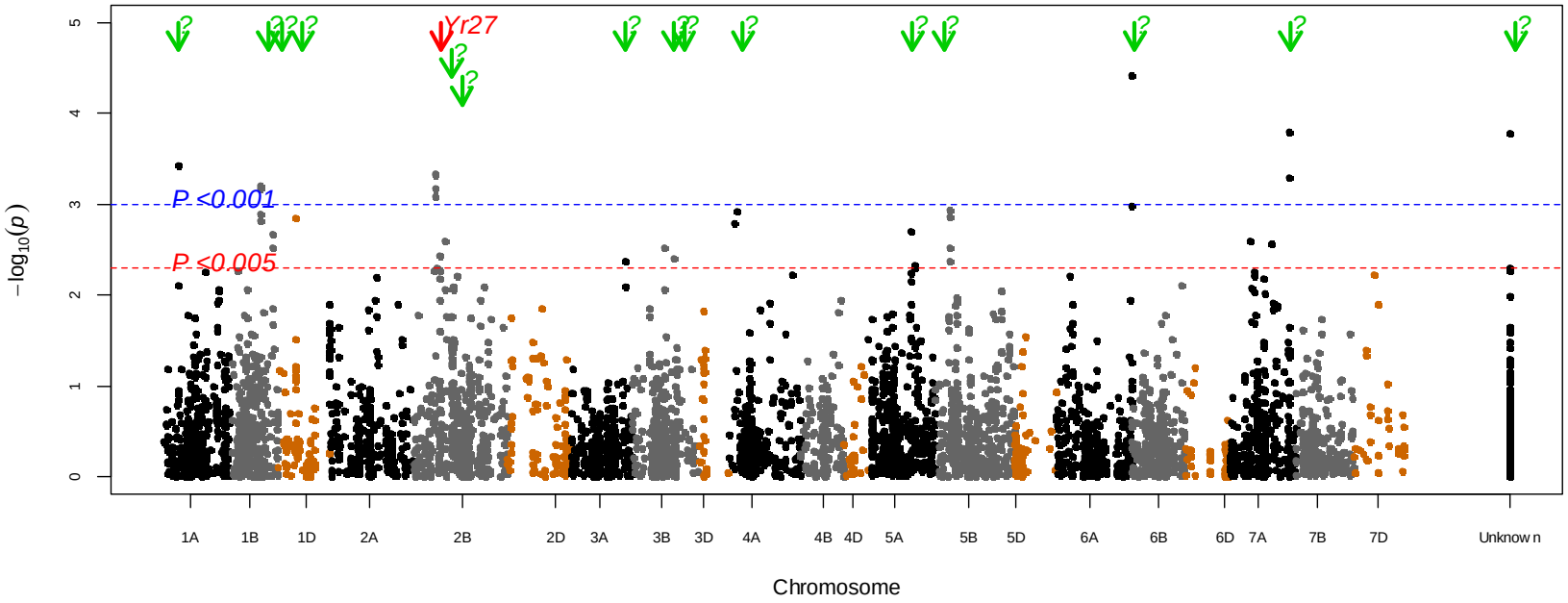


Figure 11 Manhattan plot indicating genomic regions putatively associated with stripe rust resistance, marked by question marks and green arrows. The major gene *Yr27* in chromosome 2B based the wheat consensus map is shown in a red arrow. *x*-axis shows SNP along each chromosome; *y*- axis is the $-\log_{10}$ (*P*-value). Dashed horizontal lines designate thresholds for declaring significance.

CHAPTER 3

Quantitative trait loci mapping for adult plant resistance to rust in bread wheat cultivar Kenya-Nyangumi

The challenge posed by race *Ug99* of wheat stem rust and lineal group of races has necessitated faster discovery of effective sources of resistance with the aim of utilizing characterized resistance in marker assisted selection and recently, genomewide selection for durable resistance. While the Kenyan bread wheat cultivar Kenya-Nyangumi is susceptible to *Ug99* and a few other related races, it has been observed to exhibit reasonably high adult plant resistance (APR) to these races under intense disease pressure in the Kenyan fields. To dissect this resistance and characterize the genomic regions conditioning it, the cross PBW343/Kenya-Nyangumi was made to generate 189 F_6 RILs. This population was evaluated for stem rust, stripe rust, and leaf rust in several environments in Kenya, Mexico and Minnesota and subsequently quantitative trait loci (QTL) controlling these diseases located based on composite interval mapping. Study of stripe rust and leaf rust related QTL was included after observing transgressive segregation for these diseases in the RIL population and cognizant that APR to wheat rusts often involve pleiotropic QTL. At a minimum LOD threshold of 2.5 chosen by permutation, eight, two and four minor effect QTL were detected for stem rust, stripe rust and leaf rust respectively. Observed lack of consistency of detected QTL across environments is attributed to QTL x environment interaction.

INTRODUCTION

The three rust diseases- stem rust (*Puccinia graminis* Pers. f. sp. *tritici*), leaf rust (*Puccinia triticina*) and stripe rust (*Puccinia striiformis* Westend. f. sp. *tritici* Erikss) are not only among the oldest plant diseases known to man but also the most important diseases of wheat (*Triticum aestivum* L.) globally (Roelfs 1992, Singh et al. 2008). The historical success of these biotrophic pathogens partly owes to their capacity to produce a large number of spores, which are readily disseminated over long distances, mainly by wind to infect wheat plants under favorable environmental conditions. Additionally, the genetic plasticity of these pathogens is well known. Often, no sooner are resistant wheat cultivar(s) deployed, virulent pathotypes increase in frequency and either render the resistant cultivar(s) vulnerable to infection or actually cause rust disease locally or on a continental scale (McIntosh 1988). A timely reminder of the potential of rust diseases is the recently emerged race *Ug99* of stem rust. First discovered in Uganda in 1999 (Pretorius et al. 2000), *Ug99*, also designated as TTKSK, based on North American rust nomenclature (Wanyera et al. 2006, Jin et al. 2008) has unique virulence to *Sr31*, a host resistance gene widely utilized in wheat worldwide and for which virulence had not been reported previously in the world (Singh et al. 2011). Apart from combining *Sr31* virulence to that of most genes of wheat origin (Singh et al. 2008), *Ug99* has since generated an illustrative evolutionary pathway leading to a "*Ug99* lineage" of races (Park et al. 2011, Singh et al. 2011) that are widespread in most of eastern and southern Africa (Hodson et al. 2011) and parts of the Middle East. Furthermore, it has been estimated that only about 5 - 10% of land in 22 African and Asian countries is planted with wheat

cultivars exhibiting inadequate resistance to *Ug99* and that approximately 85-95% of breeding materials from various countries are susceptible to this race (Singh et al. 2011).

Response to the threat posed by *Ug99* and other rust races include the conception of the Borlaug Global Rust Initiative (http://www.globalrust.org) affiliated with the 'Durable Rust Resistance in Wheat' (DRRW) project, whose research focuses on developing and deploying rust resistant wheat cultivars, more so for resource-poor farmers (wheatrust.cornell.edu) in the primary risk areas. Such global effort to curb wheat rusts recognizes that whereas these diseases can be economically controlled by use of foliar fungicides under high input agricultural production systems (Rosewarne et al. 2012), financial constraints and unavailability of chemicals amongst small scale wheat producers in developing nations may hinder chemical control. Concerns about environmental degradation through chemical pollution and lost ecosystem services suggest that development of durably resistant cultivars remains the most attractive option and is a priority objective in many wheat breeding programs.

Durable resistance is that which remains effective for a long period when applied on a large scale in a region that is undergoing regular epidemics of the pathogen (Johnson 1992). Breeding for resistance might involve a systematic incorporation of combinations of diverse resistance genes through limited or repeated backcrossing (Singh et al. 2011) and/or selection for resistance described as partial, slow rusting, general or adult plant resistance. This latter form of resistance is quantitatively inherited, and is often assumed to be more durable relative to that controlled by single major resistance genes since it is conditioned by more than one gene (McIntosh 1992, Kolmer, 1996).

In this study, the term adult plant resistance (hereafter APR) is adopted. APR is apparently race nonspecific and is characterized by lower frequencies of infections, lower uredinial size and less urediniospore production (Caldwell 1968, Chen and Line 1995, Liang et al. 2006). In the case of stem rust and leaf rust, absence of hypersensitivity is usually observed (Singh 2012). Accordingly, this resistance describes genes that confer moderate to inadequate resistance under high disease pressure and rather than completely halt fungal infections only slow the infection process in the adult plants (Kolmer 1996, Singh 2012) but not seedlings.

To date, only a few well characterized, non-hypersensitive often pleiotropic APR genes including *Sr2/Yr30*, *Lr34/Yr18*, *Lr46/Yr29*, *Lr67/Yr46*, and *Lr68* are catalogued and known to be effective in the field (Rosewarne et al. 2012; Singh 2012). A number of recent studies (*e.g*. Crossa et al. 2007, Bhavani et al. 2011, Yu et al. 2011, and Yu et al. 2012) have identified novel genomic regions associated with APR to wheat rusts using various mapping populations.

Several old tall Kenyan cultivars highly resistant to *Ug99* under field conditions have been detected (Singh et al. 2008), and some of them including "Kenya Plume" contain *Sr2* (Singh and McIntosh, 1986). The initial objective of the present study was to locate, map and estimate the genetic effects of APR genes that condition stem rust resistance in Kenya-Nyangumi (hereafter Nyangumi) wheat cultivar. After observing unambiguous transgressive segregation for leaf and yellow rust resistance, the mapping objective was elaborated to include characterizing quantitative trait loci (QTL) conditioning resistance for all the three rust pathogens.

MATERIALS AND METHODS

Plant materials

A segregating population of 189 F_6 recombinant inbred lines (RILs) was developed at CIMMYT from a cross between cultivars Nyangumi and PBW343. Nyangumi has the pedigree "Tezanos-Pintos-Precoz//Selkirk Enano*6/Lerma-Rojo-64/3/Africa-Mayo-48/4/Kenya-Swara/K-4500-6" and was released in 1979 and has remained a popular cultivar in the medium altitudes wheat growing zones of Kenya to date. Despite being susceptible at the seedling stage (Njau et al. 2009), this cultivar exhibits adult plant resistance to stem rust race TTKSK (*Ug99*) and related group of races. Moreover, it displays moderate resistance to leaf and yellow rusts in the field even under high inoculum pressure in Mexico. The second parent, PBW343 with the pedigree "Nord Desperez/Vogel9144//Kalyansona/BlueBird/3/Yaco/4/Veery#5" is a 'mega cultivar' that was cultivated on millions of hectares in India following its release in 1995. This cultivar has moderate to high susceptibility to stem rust in the field depending on the disease pressure.

Field plots and disease epidemics

Stem rust

The parents and RILs were evaluated for reaction to stem rust at Njoro, Kenya and St Paul, Minnesota. Field plots were sown in a randomized complete block design (RCBD) with two replicates at Njoro in the wetter main season (May to October) of 2010 and the off-season (November 2010 to April 2011). Each line was sown on 75 cm -wide raised beds in plots comprising double 70 cm- long rows, 20 cm apart with the provision of 50 cm alleys. On each side of the plot in the middle of the alleys, a continuous row of a mixture of susceptible spreader wheat cultivars Chozi and Duma, both carrying *Sr31*

(Njau et al. 2010) was sown. A border of 8 spreader rows was also planted to surround the field. To initiate the stem rust epidemic, spreader rows were inoculated using fresh urediniospores of stem rust race TTKSK (*Ug99*) collected from rust increase plots at Njoro. The urediniospores were first suspended in water and then injected into individual spreader plants prior to booting (growth stage Z35-Z37; Zadoks et al. 1974), at approximately 1m distances. To further boost disease inoculum, spreader plants were also sprayed with a suspension urediniospore in light mineral oil Soltrol 170 (Chevron Phillips Chemical Company, The Woodlands, TX) at least twice during stem elongation.

In Minnesota, rust evaluation plots were established at St. Paul in the spring wheat season of 2011 that span April to August. Field plots were 2 m long single- rows with 20 cm row spacing and 2m wide alleys planted in RCBD with two replicates. Spreader rows of a mixture of the highly rust susceptible wheat cultivars Thatcher, Morocco, and LMPG-6 were planted perpendicular to the test plots a week prior to planting the experimental plots. Artificial inoculation was performed initially at jointing (Z30-Z39; Zadoks et al. 1974) using a mixture of standard US races QFCSC, QTHJC, MCCFC, RCRSC, RKQQC and TPMKC. A repeat inoculation was done at booting stages (Z35-Z37; Zadoks et al. 1974).

Leaf rust

Test nurseries were planted in two locations- Ciudad Obregon in northwestern Mexico in 2010 and 2011 and St Paul, Minnesota in 2011 and 2012. In each environment (location/year combination), experimental design was a RCBD with two replications. At Ciudad Obregon, a test plot consisted of 1m- long double rows on a 75 cm wide raised beds with a 20 cm inter-row spacing. About two weeks post-sowing, spreader rows

(planted with cultivar Morocco) were inoculated by spraying a suspension of races MCJ/SP and MBJ/SP. These two Mexican leaf rust races share a similar avirulence/virulence formula: *Lr2a, 2b, 2c, 3ka, 9, 16, 19, 21, 24, 25, 28, 29, 30, 32, 33, 34/ 1, 3, 3bg, 10, 11, 12, 13,14a, 14b, 15, 17a, 18, 20, 22b, 23, 26, 27+31* (Singh, 1991). MBJ/SP is differentiated from MCJ/SP since it is only partially virulent on *Lr26* (Herrera et al. 2012). Seedlings of K. Nyangumi and PBW343 are susceptible to both races. At St Paul, Minnesota spreader rows planted to a blend of Thatcher, Morocco, and LMPG-6 were inoculated with a mixture of US leaf rust races THBJ, MCDS and MBRJ. The first two races are commonly found in the US while MBRJ was most prevalent in the 1990s.

Stripe rust

Field plots for stripe rust screening were established in RCBD, two replicates at CIMMYT's research fields at Toluca in the 2010 and 2011 crop season. Disease spreader comprising of 8 lines in Avocet background and Morocco were inoculated 4 weeks after planting with a mixture of 3 races having a combined virulence to *Yr3, Yr9, Yr17, Yr27* and *Yr31*. Both parents are susceptible to these races at the seedling stage.

Phenotyping

Disease severity for all three rust types was visually estimated based on the modified Cobb scale (Peterson et al. 1948) when the susceptible check lines in each trial had reached maximum severity. For QTL mapping the maximum disease severity (MDS) for each RIL and parents was recorded as the mean of two replicates in each environment.

Statistical analyses

Analysis of variance (ANOVA) was performed using function '*aov*' in R (R Development Core Team, 2011) where genotypes were modeled as fixed effects and environments and replicates as random effects. Broad sense heritabilities (h^2_b) of resistance to the three rust disease, expressed on a RIL-mean basis was estimated based on the formula $h^2_b = \sigma^2_g / (\sigma^2_g + \sigma^2_{ge} / e + \sigma^2_\varepsilon / re$ (Allard 1960). Respective variance components were estimated from ANOVA results as σ^2_g, genetic variance = (MS_f - MS_{fe})/re; σ^2_{ge}, genotype x environment interaction variance = ($MS_{fe} - MS_e$)/r and σ^2_ε, error variance = MS_e. MS_f is the mean square of genotype, MS_{fe} is the mean square of genotype x environment interaction, MS_e is the mean square of error, r is the number of replicates per environment and e is the number of environments.

Spearman rank correlation coefficients (r_s) of rust severity of the RILs for each disease across environments were estimated based on the mean disease severity data using R (R development team 2011).

The minimum number of genes/QTL controlling rust resistance in the PBW343/K. Nyangumi population was estimated using Wright's method (Wright 1984). To account for the level of inbreeding in the population this method was slightly modified by Singh et al. 1995 and has the form: $n = (GR)^2 / 4.27\ \sigma^2_g$, where n is the minimum number of effective number genes, GR is the genotypic range of the RILs estimated by multiplying the phenotypic range by heritability (h^2_b) and σ^2_g is the genetic variance.

Molecular marker assays

Genomic DNA was extracted from young leaves of the parents and RILs using the DNeasy Plant DNA extraction Kit (QIAGEN). For Diversity Arrays Technology (DArT®) marker assays, 500 - 1,000 ng samples of restriction grade DNA, suspended in TE with a final concentration of 50-100ng/Ml were shipped to Tricarte Pty. Ltd., Canberra, Australia (www.tricarte.com.au) for whole genome profiling (Wenzl et al. 2006, Neumann et al. 2011). DArT markers are named as described by Akbari et al. (2006) with a prefix 'wPt' or 'tPt' followed by a number corresponding to a particular clone in the genomic representation where 'w' and the first 't' in 'tPt' stands for wheat and triticale libraries respectively, 'P' for *Pst*I (primary restriction enzyme) and 't' for *Taq*I (secondary restriction enzyme). Across the complete mapping population, a total of 1355 DArT were typed with loci scored as present (1) or absent (0). After filtering monomorphic markers and those with <10% data points, 381 high quality loci were available for linkage mapping.

Linkage map construction and QTL analysis

Firstly, chromosomes were assigned to markers based on wheat Interpolated maps v. 4 from Tricarte Pty Ltd. (www.tricarte.com.au). Information from this host highlights that chromosomal assignments was established using several doubled haploid populations including 'Synthetic x Opata'. In the present work, the software JoinMap v. 4.0 (Van Ooijen 2006) was used to create linkage maps. Markers were assigned to linkage groups with a minimum threshold of logarithm of the odds (LOD) of 3.0 and genetic distance between markers was estimated based on the Kosambi mapping function (Kosambi, 1943). PLABQTL v. 1.2 (Utz and Melchinger, 1996) which implements the multiple

regression procedure of Haley and Knott (1992) was applied to localize QTL. For the different diseases, composite interval mapping (CIM) analysis incorporating the additive effects model was implemented through the "cov SEL" statement of PLABQTL using $\log_{10}$ transformed mean rust severity data. The function "cov SEL" automatically selects the most important markers to be used as cofactors through stepwise regression (Utz and Melchinger, 1996). Subsequently, for loci detected as significant at the same environment in the additive model, an additive x additive model was fit to the data to detect epistatic gene effects. The LOD threshold for declaring a significant QTL at a genome-wide type 1 error rate of $\alpha = 0.05$ in a single environment was set to 2.5 using 1,000 permutations (Churchill and Doerge 1994). The minimum distance between two putative QTL to be listed as separate was 10 cM. QTL effects were estimated as the proportion of phenotypic variance (R^2) explained by significant QTL. Lastly, to represent detected QTL locations, linkage maps and LOD contours were drawn using MapChart v. 2.1 (Voorrips 2002).

RESULTS

Disease expression and estimated number of resistance genes

Disease development for the different rust types was sufficient with the most susceptible RILs across environments recording maximum disease severities of 70 - 100%. As depicted by histograms of the maximum disease severity across environments for the different pathogens (Figure 1), disease severity distributions for the RILs were continuous which conforms to the polygenic inheritance of adult plant resistance to rust and its quantitative nature. However, most distributions deviated from normality and were generally skewed towards resistance which necessitated data transformation prior to QTL mapping. The resistant parent, Nyangumi was within the 0 - 20% phenotypic class with most of the pathogens across environments. Phenotypic ranges for PBW343 were largely dependent on the rust type tested. For this parent, stem rust maximum severities ranged from 30% observed in St Paul in 2011 to 70% in the Njoro offseason nursery (*i.e.* SR_Njoro 2010(O). In addition, for both leaf rust and stripe rust responses, the parental genotypes presented low and similar levels of maximum severities, excluding St Paul in 2012 where PBW343 had a maximum 60% leaf rust disease severity. In most environments, transgressive segregation was observed where a considerable number of RILs showed higher susceptibility than PBW343, an indication that this parent carries alleles that contribute resistance in the segregating population.

Notwithstanding the pathogen type assayed, phenotypic correlations (r_s) of mean disease severity were high ($P < 0.001$) when considered for the same location in different years but not across locations (Table 1). In this respect, repeatability for stem rust was highest in the Njoro location with r_s of 0.84 whereas leaf rust was most repeatable at

Ciudad Obregon with r_s of 0.68. Stripe rust was only screened at one location, Toluca and r_s was 0.61. Lack of genetic correlations for severities across locations reflects a large genotype x environment interaction effects in all rust trials. This outcome was reiterated in the ANOVA results (Table 2) where variance for genotype x environment component was highly significant ($P < 0.001$)

Trait heritability was estimated as 0.62 for leaf rust, 0.67 for stem rust and 0.83 for stripe rust resistance. The estimated number of effective genes segregating in the population based on Singh et al (1995), assuming no linkage, no epistasis, no dominance and loci of equal effects were approximately 7, 6 and 4 for stem rust, leaf rust and stripe rust respectively (Table 1).

Genetic linkage map

All the 381 DArT markers were mapped and placed in 37 linkage groups with some chromosomes represented by more than one linkage group. In total, the completed linkage map for QTL analysis covered a genetic distance of 1200 cM.

QTL for stem rust resistance

Nine significant QTL for resistance to stem rust with a LOD score ranging between 2.8 and 8.2 were detected on chromosomes 2D, 3B, 4A, 5B, 6A, 6B and 7B based on CIM (Table 3, Figure 2). These QTL are named following the standard nomenclature for designation of wheat QTL (McIntosh et al. 2003). Among the detected QTL, *QSr.umn-6B* at support interval 18 - 26 cM of chromosome 6B and *QSr.umn-7B.1* at support interval 10 - 14 cM of chromosome 7B were effective in both Njoro environments. Additionally, *QSr.umn-3B* within 28 - 48 cM interval was detected in the Njoro 2010 and St Paul 2011 environments. All other stem rust QTL were detected only

once at the chosen LOD = 2.5 threshold. Notably, *QSr.umn-2D.1* and *QSr.umn-2D.2* more than 14 cM apart, both detected in the Njoro 2011 environment had considerably large effects on stem rust resistance and together explained 24% of the total phenotypic variation (Table 3). Effects due to other detected stem rust QTL deemed minor QTL were consistently small and each of them individually explained between 2.2 and 8.7 % of the total variation.

QTL for leaf rust resistance

Four QTL, two each from Nyangumi and PBW343 were located on chromosome 2B, 3A, 6B and 7B (Table 3, Figure 2). These minor QTL had LOD scores between 2.7 and 4.0 and were not detected across environments. The QTL, *QLr.umn-7B* supported by the interval 8 - 12 cM which was contributed by PBW343 explained 7% of the total phenotypic variation in leaf rust resistance in the St Paul 2012 environment. This locus which perhaps has pleotropic effects is in the vicinity of the *QSr.umn-7B.1* detected in both Njoro environments for stem rust resistance.

QTL for stripe rust resistance

Two QTL both contributed by Nyangumi, were detected for stripe rust resistance in the two testing environments (Table 3, Figure 2). *QYr.umn-7D* at a support interval 12 - 22 cM of chromosome 7D was detected in Toluca 2010. This QTL explained only 3% of the phenotypic variation observed. The QTL *QYr.umn-4A*, within an interval of 4 - 8 cM of chromosome 4A, was detected at a LOD score of 6 and explained approximately 15% of the phenotypic variance observed in the mapping population at Toluca 2011. Besides, this QTL is within the same genomic location as *QSr.umn-4A* detected in St Paul stem rust screening in 2011, possibly pointing to pleiotropic effect of the same QTL.

Epistatic effects of rust resistance

Because only stem rust testing environments provided more than one QTL per environment, those QTL were tested for digenic epistasis in the multiple regression model of CIM. All pairwise QTL x QTL interactions were insignificant and only presented very small absolute effects ranging from -0.44 to 0.66, suggesting that additive effect of the minor stem rust QTL detected in the PBW343 x Nyangumi population played a more important role in enhancing resistance than did the epistatic effects.

DISCUSSION

Since the 'durable rust resistance project' (http://www.globalrust.org) coordinated effort to screen world wheat germplasm against TTKSK (*Ug99*) and associated stem rust races at Njoro, Kenya began, a number of old Kenyan wheat cultivars have been observed to express significant APR to these races even under intense disease pressure. In the present study we have presented the nature of this resistance in a RIL population developed from Nyangumi, a cultivar that beyond its observed resistance in the rust nurseries has maintained popularity with the small scale Kenyan farmers for nearly three decades. Moreover, the RIL segregating population initially designed to dissect stem rust resistance in Nyangumi was noted to display clear segregation, including transgressive segregation for leaf rust and yellow rust resistance in Ciudad Obregon and Toluca respectively, two of the traditional disease screening sites used by the CIMMYT wheat program. In addition, the detection of fully susceptible segregates in all environments for different rust diseases suggest the presence of different resistance genes in the two parental cultivars used to develop the mapping population in this study.

The chromosomal location of the *QSr.umn-3B* allele derived from Nyangumi is in the vicinity of *Sr2*, the only catalogued stem rust APR gene in wheat (Hare and McIntosh 1979). Further evidence to suggest involvement of APR resistance gene *Sr2* in K. Nyangumi is provided in Bhavani et al. 2011. In an elaborate APR mapping project involving six populations involving PBW343 as the susceptible parent, these authors located a QTL at exactly the same position as the QTL reported in the present study which they implicated to *Sr2*. This QTL has been mapped in the same chromosomal arm through genome-wide association studies (*e.g*. Crossa et al. 2007; Chapter 2, this thesis).

Bhavani et al. 2011 also located QTL for stem rust resistance at position 20 cM of chromosome 2D in a PBW343 x Kingbird population and position 21 cM of chromosome 6B in a PBW343 x Juchi population. These QTL might respectively correspond to *QSr.umn-2D.2* and *QLr.umn-6B* both from Nyangumi. Yu et al. 2012 reported a QTL within 27.3 and 31.9 cM in a winter wheat association mapping panel which probably co-locates with *QSr.umn-5B* presented here.

In their association mapping study, Crossa et al. 2007 identified a region of chromosome 2BS associated with adult plant leaf rust resistance. This QTL might match *QSr.umn-2B* in our study that locates within the same 95% confidence interval. The apparently pleiotropic QTL detected on chromosome 7B located at approximately 10 - 12 cM and derived from PBW343 is near a QTL for leaf rust resistance reported by Rosewarne et al. 2008 in the Avocet-S x Attila population. Attila, the source of this allele, is a sister line to PBW343 (Rosewarne et al. 2008), the susceptible parent in the present study.

Lu et al. 2009 reported a consistent and large effect QTL designated as *QYr.caas-7DS* at position 15.4 cM in two Chinese cultivars Libellula and Strampelli. The QTL *QYr.umn-7D* detected in K. Nyangumi is within the same support interval as *QYr.caas-7DS*. Similarly, in a doubled haploid mapping population derived from Kukri x Janz cross, Bariana et al. 2010 reported a QTL for yellow rust resistance in the genomic region of *QYr.caas-7DS*. Notably, the slow rusting pleiotropic gene *Lr34/Yr18* that confers high levels of resistance when combined with other minor genes (Singh and Rajaram 1993) is located on the same chromosome arm. *QYr.umn-7D* may thus correspond to *Lr34/Yr18*.

Several, perhaps new, unexploited QTL involving genomic regions that have not been implicated before for resistance to stem rust (*QSr.umn-2D.1*, *QSr.umn-4A*, *QSr.umn-5B*, *QSr.umn-6A*, and *QSr.umn-7B.1*), leaf rust (*QLr.umn-3A*, *QLr.umn-6B*), and yellow rust resistance (*QYr.umn.4A*) are reported in this study. Excluding the QTL on 2D, 5B and 6A all the other QTL are traced to Nyangumi. Future effort to validate these QTL in independent mapping panels will be necessary. For instance, in our separate work involving association mapping for rust resistance using a diverse collection of East African wheat (Chapter 2, this thesis), a QTL suggested to possibly have pleiotropic control of both stem rust and yellow rust is mapped close to the pleiotropic QTL on chromosome 4A in the present study.

Though significant epistatic interactions were not detected, this might not necessarily indicate absence of epistasis for resistance alleles *per se* because of the cancelling of allelic effects as suggested by the negative sign of the detected epistatic effects. In addition, the highly significant genotype x environment interaction observed in this study, given the diversity of the testing environments, might have purged likely epistatic interactions. In any case, lack of epistatic interaction would not be a sufficient condition to warrant dismissal of detected QTL for breeding for durable resistance given the moderate size of the RIL population used in in this study. The inconsistency and the small effects of the reported QTL indicate the well appreciated complexity of breeding for APR to rust in wheat. Granted, the probability of fixation of favorable alleles through the traditional QTL introgression and/or F_2 enrichment approaches (Bernado 2010) might be low. Recent methods involving genomic selection as illustrated by Rutkoski et al. 2011 might facilitate pyramiding of major and minor genes. From a different perspective,

accumulating all of the detected QTL in finished cultivars might not be the short term objective in a breeding program. Combination of at least two QTL in RILs in the present study led to a clear reduction in disease severity (Figure 3). Accumulation of between 4 and 5 genes is expected to retard disease progress to "near-immunity" (Singh et al. 2000) with only negligible disease levels at plant maturity even under high disease (Singh et al. 2008) conditions, typical in the East Africa highlands and other similar global rust disease hotspots.

Table 1 Phenotypic correlation coefficients (r_s) for stem rust (SR), leaf rust (LR) and stripe rust (YR) maximum disease severity between different environments, trait heritability (h^2) across environments and estimated effective number of genes (n)

Trait	**Environment**[a]	**r_s between environments**[b]					**(h^2)**	**No. of genes (n)**
		Njoro 2010 (O)	St Paul 2011	Obregon 2011	St Paul 2012	Toluca 2011		
SR	Njoro 2010 (M)	0.84***	0.08				0.67	6.78
	Njoro 2010 (O)		0.13					
LR	Obregon 2010		0.03	0.68***	0.10		0.62	6.41
	Obregon 2011		0.04		0.03			
	St Paul 2011				0.51***			
YR	Toluca 2010					0.61***	0.83	4.0

*** Significant at P< 0.001

[a] Njoro 2010 (M) and Njoro 2010 (O) refers to Njoro main season and offseason respectively

[b] Based on mean disease severity across two replications

Table 2 Analysis of variance of maximum disease severities for stem rust (SR), leaf rust (LR) and stripe rust (YR) reaction

Trait	Source	*df*	Mean square	*F* values
SR	Lines	188	1380	27.9***
	Environments	2	70461	1422.2***
	Lines x Environments	376	451	9.1***
	Replicate/ Environments	3	56	1.1
	Error	560	50	
LR	Lines	188	2061	129.9***
	Environments	3	30157	1899.9***
	Lines x Environments	558	935	58.9***
	Replicate/ Environments	4	61	3.9**
	Error	746	16	
YR	Lines	188	518	7.3***
	Environments	1	5056	70.8***
	Lines x Environments	188	91	1.3*
	Replicate/ Environments	2	6139	86***
	Error	374	71	

***, **, and * Significant at $P = 0.001$, 0.01, and 0.05 respectively

Table 3 Quantitative trait loci (QTL) detected by composite interval mapping for stem rust (SR) in 3 environments, leaf rust (LR) in 4 environments and stripe rust (YR) in 2 environments for 189 RILs derived from PBW343 x K. Nyangumi

Trait	Environment	Chr[a]	Pos (cM)[b]	QTL	Marker interval	LOD[c]	Add[d]	R^2 (%)[e]
SR	Njoro 2010 (M)	3B	30	*QSr.umn-3B*	wPt-733251_wPt-0021	2.8	13.9	8.7
		5B	10	*QSr.umn-5B*	wPt-7240_wPt-9724	3.5	-8.8	8.1
		6B	24	*QSr.umn-6B*	wPt-5971_wPt-1756	4.3	-26.6	4.6
		7B	12	*QSr.umn-7B.1*	wPt-0266_wPt-7887	3.6	19.5	3.6
	Njoro 2010 (O)	2D	12	*QSr.umn-2D.1*	wPt-730744_wPt-6850	6.0	-61.7	13.0
		2D	26	*QSr.umn-2D.2*	wPt-6574_wPt-2781	8.2	28.4	10.8
		6A	18	*QSr.umn-6A*	wPt-731250_wPt-731054	4.2	-1.9	9.8
		6B	22	*QSr.umn-6B*	wPt-6674_wPt-7207	4.6	15.1	3.2
		7B	12	*QSr.umn-7B.1*	wPt-0266_wPt-7887	3.7	13.1	4.1
		7B	42	*QSr.umn-7B.2*	wPt-9925_wPt-2273	5.0	9.3	2.2
	St Paul 2011	3B	30	*QSr.umn-3B*	wPt-733251_wPt-0021	4.0	50.9	4.7
		4A	4	*QSr.umn-4A*	wPt-0817_wPt-1091	3.0	18.8	3.4
LR	Obregon 2010	2B	50	*QLr.umn-2B*	wPt-1505_wPt-4701	4.0	-5.4	8.6
	Obregon 2011	6B	4	*QLr.umn-6B*	wPt-2297_wPt-667798	3.1	35.0	2.6
	St Paul 2011	3A	8	*QLr.umn-3A*	tPt-6376_wPt-8146	2.7	2.9	2.8
	St Paul 2012	7B	10	*QLr.umn-7B*	wPt-666424_wPt-734078	2.7	-9.3	7.0
YR	Toluca 2010	7D	14	*QYr.umn-7D*	wPt-663984_wPt-663848	2.8	34.0	3.0
	Toluca 2011	4A	6	*QYr.umn-4A*	wPt-6728_wPt-744614	6.0	9.5	14.5

[a] Chr, Chromosome harboring QTL

[b] Position in centimorgan of peak LOD

[c] LOD, peak logarithm of odds score

[d] Add, additive effect of resistance allele. Negative sign indicates resistance allele is from PBW343

[e] R^2 (Coefficient of determination), Percentage of phenotypic variance explained by detected QTL

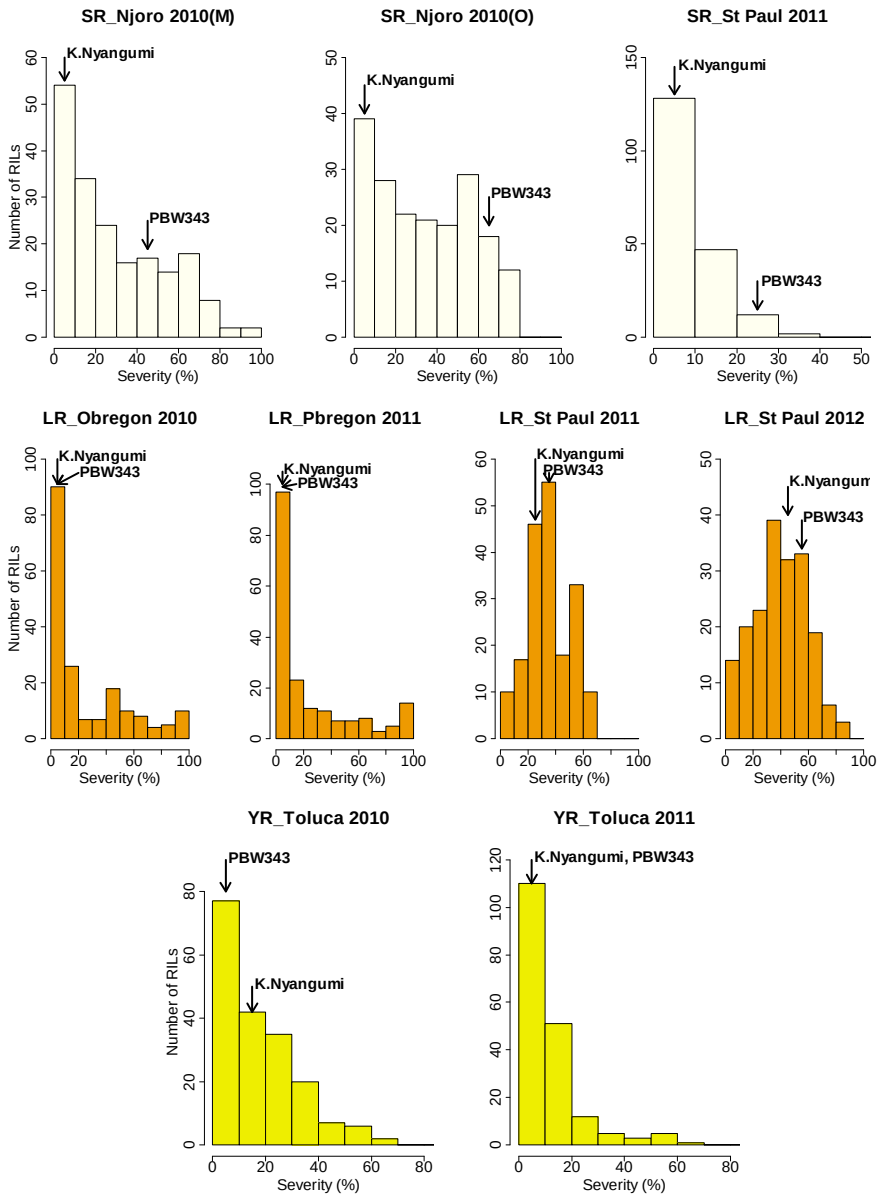


Figure 1 Distributions of disease severity for 189 PBW343 x K.Nyangumi RILs. Upper panel- last severity score for stem rust in Njoro, Kenya main season (SR_Njoro 2010(M), off-season (SR_Njoro 2010(O) and St Paul, Minnesota 2011; middle panel- last severity score for leaf rust in Obregon, Mexico in 2010 and 2011 and St Paul, Minnesota in 2011 and 2012; lower panel- last severity for yellow rust in Toluca, Mexico in 2010 and 2011.

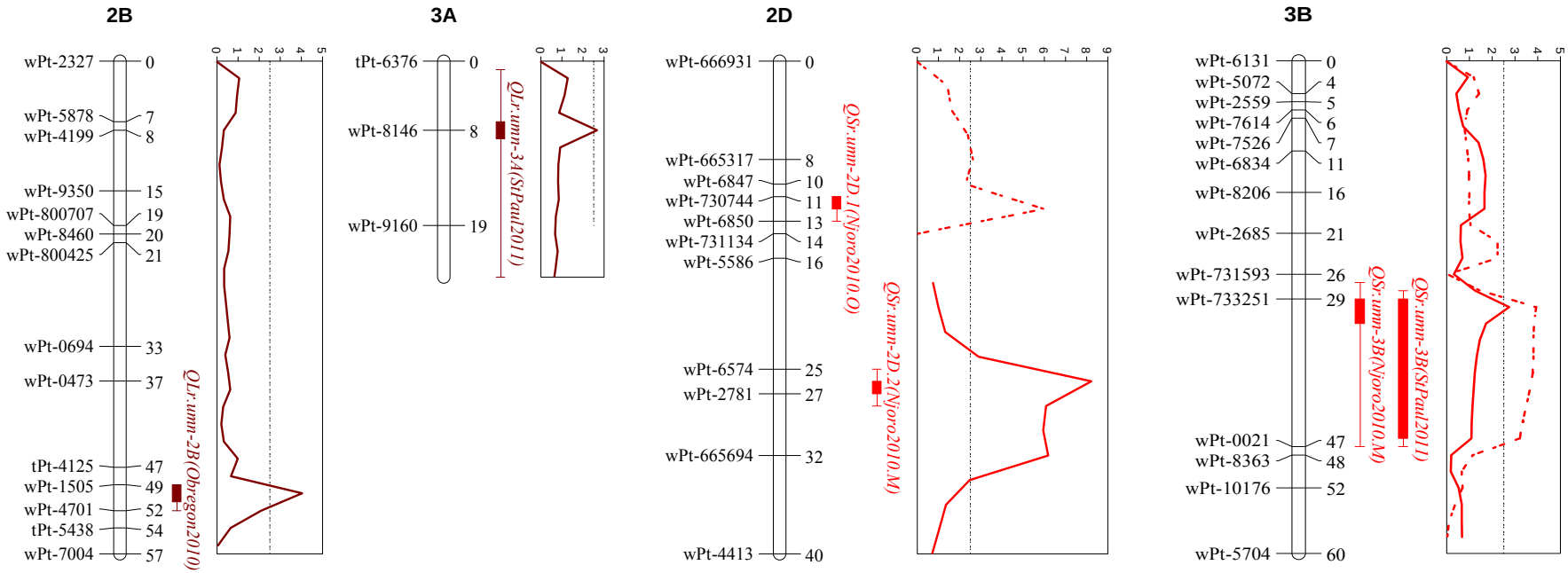


Figure 2 Linkage groups showing significant QTL for SR (red bars), LR (brown bars) and YR (blue bars), with corresponding LOD contours obtaind from CIM. The LOD significance threhold of 2.5 is indicated by a dashed line. In parenthesis after QTL name is the environment of detectio n. Genetic distances in centimorgans are indicated on the right of each linkage group.

Figure 2 *continued*

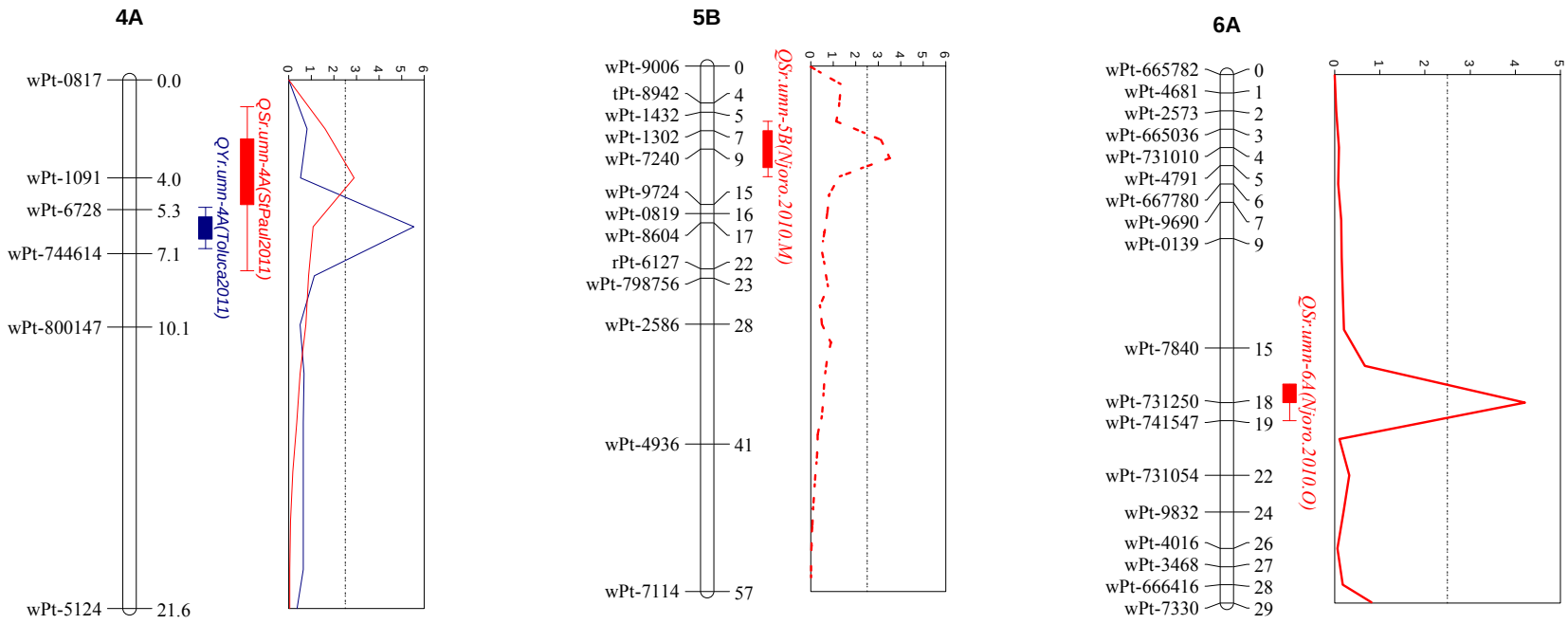
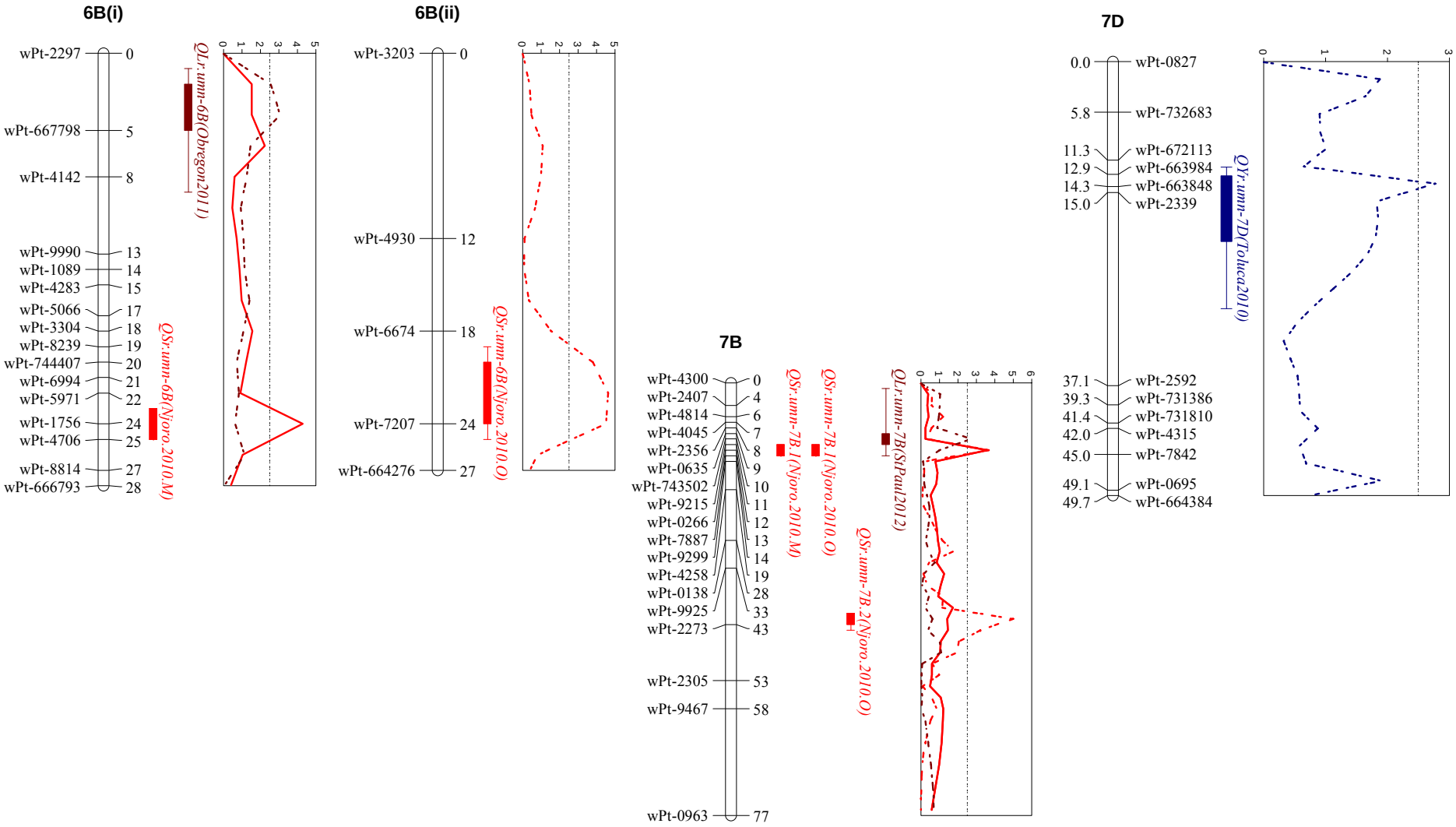


Figure 2 *continued*



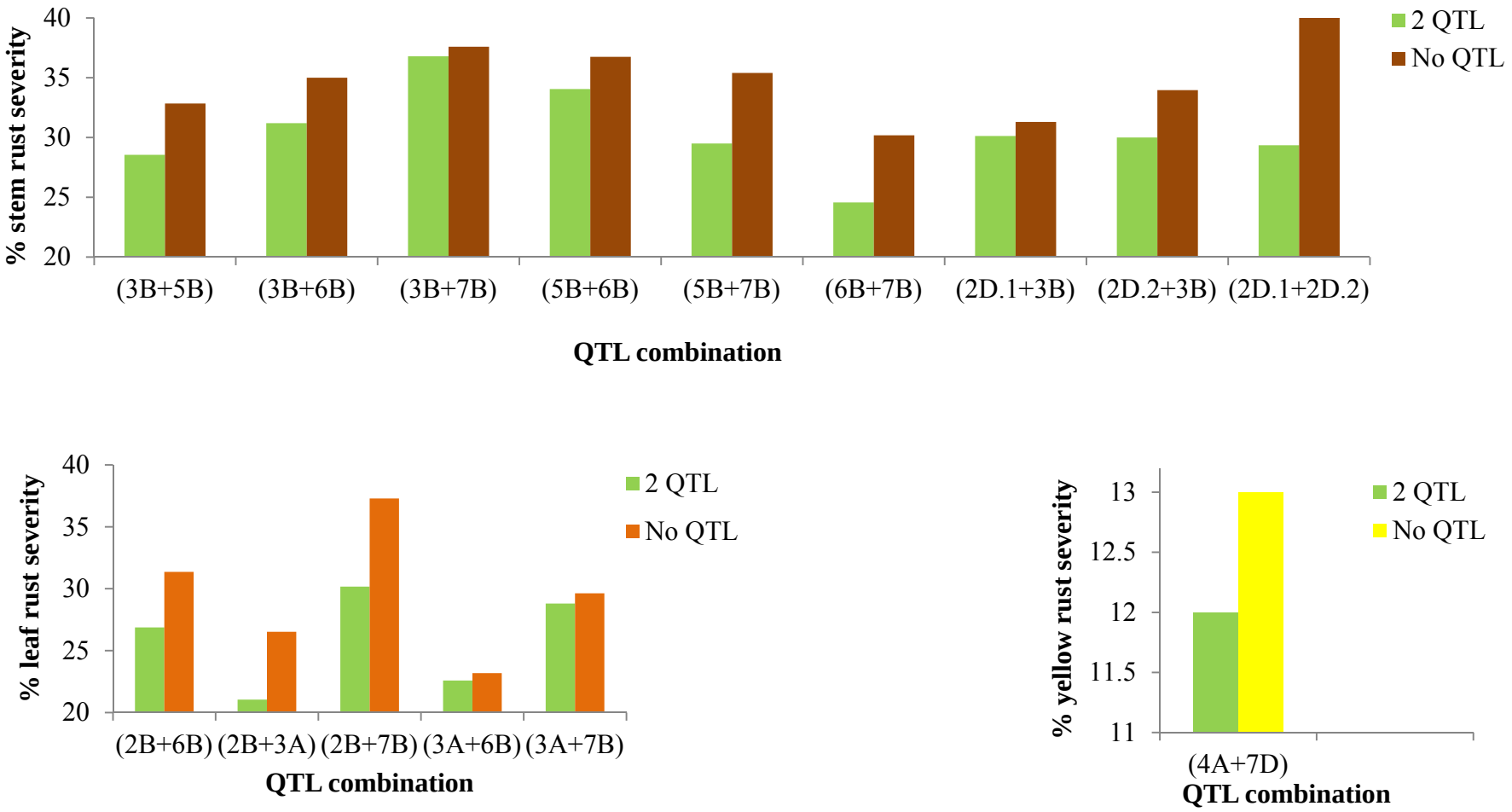


Figure 3 Effect of pyramiding pairs of QTL on rust severity illustrated for stem rust (top), leaf rust (bottom left) and yellow rust (bottom right). QTL are abbreviated to depict chromosomal location.

CHAPTER 4

Variation at the glutenin subunits loci, single kernel characterization and evaluation for grain protein in the East African bread wheat

In wheat breeding programs worldwide, selecting parental germplasm that meets certain technological criteria and use of those in developing breeding populations that segregate with the desired quality traits, is an overarching objective. Specifically, the success of a wheat cultivar in terms of adoption by farmers and popularity with processors and end users is ultimately a function of the various quality traits that define its uses as a food product. In order to identify superior breeding genotypes, it is necessary to isolate germplasm that carry preferred combinations of quality traits more so in adapted, high or moderately yielding lines within breeding programs. This study focused on 216 bread wheat lines associated with Ethiopia and Kenya breeding programs with an objective of identifying individuals that have superior alleles for bread making in combination with kernel hardness. Alleles at the high and low molecular weight glutenin subunits are reported for each line and suitability for bread making in most cases projected based on a quality score. Moreover, kernel hardness, a key determinant of a cultivar end use, is reported for all lines based on puroindoline composition. An important output from the present work is the identification of a few lines that are acknowledged as a suitable resource in breeding for bread making in the East African wheat programs of Ethiopia and Kenya.

INTRODUCTION

Wheat (*Triticum sp*.) is inextricably linked with the history of mankind. The utility of this species as a staple is illustrated not only by its unprecedented role in supporting the rise of major civilizations of Europe, West Asia and North Africa (Curtis 2002) but also its familiar position as a major source of calories and protein to billions of people today (Dubin and Brennan 2009). Moreover, with an average production of 620 million tons annually (Dubcovsky and Dvorak, 2007), wheat is of enormous economic importance worldwide (Slade et al. 2005).

The success of wheat derives from the unique technological properties of its flour (Morris 2002). A product of many constituents, wheat flour is used in making a whole range of foodstuffs such as leavened and unleavened breads, cookies, pastries, and various semolina products. It is enriched with gluten, a form of storage proteins commonly sub classified into glutenins and gliadins based on their solubility (Osborne, 1907). These components generally form about 80 to 85% of the total wheat protein (Veraverbeke and Delcour 2002) and are known to respectively influence the elasticity (strength) and extensibility (viscosity) of the dough properties essential for bread making (Payne et al. 1984). Glutenin can be separated into high-molecular-weight (80 - 120kDa) and low-molecular-weight (30 - 50kDa) glutenin subunits based on their electrophoretic ability in sodium dodecyl sulfate polyacrymide gel (Payne and Corfield 1979). In hexaploid wheats (*Triticum aestivum* L. genomes AABBDD), the high-molecular-weight glutenin subunits (HMW-GS) are encoded by pairs of closely-linked genes (*x*- and *y*-type) located on the long arms of the homoeologous chromosomes 1A, 1B, and 1D at the

Glu-1 locus (Shewry et al. 1989). These are designated- *Glu-A1, Glu-B1, Glu-D1* respectively. While six different HMW-GS would be expected in hexaploid bread wheat varieties, as a result of gene silencing (Fang et al. 2009), the genes encoding HMW-GS at the *Glu-A1* locus and the *y*-type HMW-GS at the *Glu-B1* locus are not always expressed, while that encoding y-type HMW-GS at the *Glu-A1* locus is generally not expressed (Veraverbeke and Delcour 2002). Consequently, the number of different HMW-GS commonly found in bread wheat is three to five (Fang et al. 2009). A large number of HMW-GS alleles exist at each locus (Payne and Lawrence 1983; McIntosh et al. 2003) which in turn affects the bread-making quality of different wheats (Liu et al. 2007). It is possible to make predictions of the bread making quality of wheat of a given allelic composition on the basis of quality scores proposed by Payne (1987) and Lukow et al. (1989). LMW-GS are encoded by the *Glu-A3*, *Glu-B3* and *Glu-D3* sets of loci on the short arms of homoeologous chromosomes 1A, 1B, and 1D (Payne 1987, Shewry et al. 1992). Allelic variation at the LMW-GS similar to HMW-GS has a contribution to properties of dough made with various wheat cultivars (D'Ovidio and Masci 2004). Besides, alleles at the glutenin subunits loci are useful markers to evaluate diversity in wheat genetic resources (Nevo and Payne 1987; Caballero et al. 2004) and these have in the past facilitated estimates for major wheat-producing countries (Xu et al. 2009).

Kernel hardness (endosperm texture) of wheat has been suggested as perhaps the largest single factor affecting wheat end product quality (Morris 2002, Giroux et al. 2003). This trait is considered imperative by growers, millers and bakers for their intended end use (Morris, 2002). Moreover, the criteria of classifying wheat as soft, medium soft, hard, medium hard and extra hard based on grain hardness (Kent and Evers,

1994) is universally accepted for differentiating the world trade of wheat grain (Pasha et al. 2010). Hard and soft wheats are macroscopically similar. To classify wheat cultivars into the different textural classes, various techniques according to cohesion, crushing and grinding have been implemented. Single Kernel Characterization System (SKCS), for instance crushes individual kernels and uses algorithms based on the force-defoliation profile data (Gaines 1996), which is then processed to provide the weight, size, moisture and hardness of the kernel. In addition, grain texture has in the past been evaluated based on near infra-red spectroscopy (NIR) techniques (Brites et al. 2008). A key determinant of kernel softness is friabilin [also known as grain softness protein (Jolly et al. 1993)]. Friabilin is an endosperm specific starch associated fraction of cereal grain protein (Pasha et al. 2010). This protein comprises a family of 13 closely related proteins found on water washed wheat starch (Morris et al. 1994) with the main component being the membrane-bound proteins puroindoline a and b (Rahman et al. 1994). Functional or wild type of both puroindolines gives soft-textured grains (Pasha et al. 2010). Hard wheat indicates the lack of one or more of the puroindoline proteins. It is widely accepted that puroindolines have important roles in bread making. They are known to provide finer crumb structure of bread and influence dough tenacity and extensibility.

Several objectives were envisaged in the present study: (i) to characterize the HMW and LMW glutenin subunit composition in the historical and recent east African bread wheat lines (ii) to explore the genetic variation at the HMW-GS and LMS-GS glutenin loci (iii) to investigate relative puroindoline content and total crude protein among samples (iv) to profile wheat samples for kernel hardness, and kernel size by implementing SKCS.

MATERIALS AND METHODS

Plant materials

Seeds of 216 bread wheat lines were assembled for this study. This set includes over 80% of cultivars grown in Kenya and Ethiopia in the past century as well as a few advanced breeder lines and landraces historically associated with the region. To enhance the study, pedigrees of these materials were thoroughly scrutinized to identify commonly occurring parental lines that are the foundation of the east African wheat. On this basis, a total of 46 frequently utilized parents were availed. Hence three distinct sources of accessions; Kenya, Ethiopia and founder, representing discrete subpopulations were identified for this study (Tables 2a, 2b, and 2c). To obtain suitable samples for glutenin sub-unit assays, grain of each of the lines was multiplied in cd Obregon, Mexico during the 2011 growing season. Samples for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) were obtained by thoroughly cleaning grain pooled from two replicates. Similarly, samples for puroindoline protein studies and kernel characterization were prepared from grain increased in the greenhouse at St Paul, Minnesota in the winter of 2011.

Sample extractions SDS- PAGE electrophoresis

Glutenin protein extracts were separated at the CIMMYT wheat quality laboratory, El batan Mexico using the method proposed by Singh et al. 1991 with minor modifications. Briefly, 8 μl of sample extract was obtained by adding 20 mg of crushed whole meal endosperm sample to 0.75 ml of 50 % propan-1-ol which was then mixed thoroughly for 30 minutes under a thermomixer comfort® at 65^0C. This was necessary to

remove gliadin, albumin and globulin proteins. Following centrifugation for 2 min at 10,000 rpm, a pellet was recovered. Further steps involved addition of 0.1ml of 1.4% vinylpridine solution and full range glutenin extraction at 90 ^{0}C using 2% SDS containing 1.5% dithiothreitol and centrifugation at 1400 rpm (5 min) and 10,000 rpm for a further 5 min. The specifications for SDS-PAGE were: separation gel (15.0% concentration), cross linker (1.3%) at pH of 8.5. Protein fractionation was achieved by electrophoresis in a vertical gel in a discontinuous Tris-HCL-SDS buffer system. The gels were stained with bromophenol blue and visually analyzed for allele identification after running them at a current of 12.5 mA.

Extraction of puroindoline protein was performed at the wheat quality and carbohydrate laboratory at North Dakota State University, Fargo USA. A total of 216 lines were available for this assay. Four standards namely, Alpowa-1-soft white spring wheat (wild type), Alpowa-2-Null (deletion line, no puroindolines), Alpowa-3-Null a (Pina-null, Pina-D1b/Pinb-D1a), Alpowa-4-Null b (Pinb-null, Pina/Pinb-D1g), which were provided by Dr Craig Morris of Washington State University, USA were included in the analyses. Also evaluated was the soft wheat cultivar 'Becker', and a lane for a broad range MW standard protein from BioRad (*www.bio-rad.com*). For all samples, extraction followed the procedure proposed by Morris et al. 1994: 50mg whole wheat flour was added to 1ml 50% 2-propanol and 50% 0.1N NaCl and subsequently vortexed at 50^0 C for approximately 2hr. After centrifuging, 300 ml of supernatant was recovered per sample and transferred to a tube containing 300 ml non-reducing SDS-PAGE sample buffer. Approximately 200 mg of solid sucrose was then added. Tubes were subsequently vortexed at 50^0 C for 0.5hr, heated at 95^0 C for 5 min and then centrifuged. 50ml aliquots

were loaded on SDS-PAGE gradient (15-30% acrylamide) gels prepared in-house (14cm x 16cm x 1.5mm) and run overnight on a Hoeffer SE600 system. To facilitate identification of puroindoline bands, the proteins were fixed in trichloroacetic acid solution and then stained with CBB-R in a solution of methanol, acetic acid and water followed by destaining. Gel imaging and analysis was achieved using Biorad GelDoc XR system with BioRad Quantity One imaging software. This software draws small rectangles around the protein bands of interest on the gel photographs and then measures the relative intensity of the gray/ black within the rectangles. This data is then assumed to correspond to the protein concentration. As such, puroindoline bands were identified located just below the 14.4 KDa mark of the MW standard protein. Estimated protein concentrations for the respective samples were then divided by that of Alpowa-1-soft (calibrated as 1) to obtain a relative puroindoline content value.

HMW-GS and LMW-GS allele naming and Quality scores

The nomenclature proposed by Payne and Lawrence (1983) for HMW-GS and Singh et al. 1991 for LMW-GS was adopted in describing alleles for these proteins. This was also supplemented with the guidelines in the catalogue of gene symbols for wheat (McIntosh et al. 2003). Comparisons were made with observed profiles in the standard genotypes- Pavon, Opata and Pictic. A quality score depicting the bread making quality of each wheat sample as outlined by Payne et al. 1987 was obtained to correspond with the observed combination of HMW-GS alleles (Table 1).

Statistical analysis of genetic diversity

Allelic frequency was calculated by summing all lines with the allele within a subpopulation (Kenya, Ethiopia, and founder) and simply dividing this total by the

number of lines for the subpopulation. To further summarize variation at the glutenin loci, several statistics were obtained: (1) allele per locus *A* was a count of the alleles present at each of the *Glu-A1, Glu-B1, Glu-D1, Glu-A3, Glu-B3,* and *Glu-D3* loci. This absolute number is sensitive to sample size (Weising et al. 1995). To correct for that, effective number of alleles, A_e was calculated by taking into account the allele frequency as: $A_e = 1/\sum p_i^2$, where p_i is the frequency of each allele observed across all investigated loci. (2) Expected heterozygosity was estimated as: $H_e = 1-\sum p_i^2$. (3) Shannon's information index was obtained as: $S = -\sum p_i \log_2 (p_i)$ respectively.

Total crude Protein

The crude protein content of wheat samples were measured according to AACC approved method 46-30.1 (AACC International, 11th edition) using the nitrogen combustion method on a LECO FP428 (Leco, St Joseph MI). Samples were weighed (~ 0.15-0.20g) into foil cups and the percent nitrogen (N) was measured after burning. Total crude protein was calculated as N x 5.25 on a 12% moisture basis.

Kernel characterization

A single kernel characterization system (SKCS) (SKCS 4100, Perten Instruments, Springfield, IL) was used to determine the kernel quality characteristics of the wheat samples based on AACC approved Method 55-31.01 (AACC International, 11th edition). Kernel hardness (g), kernel weight (mg) and kernel diameter (mm) were estimated from an average of 300 kernels per sample. In this analysis, the Ethiopian landrace durum line MG07766 was included as a check for extremely hard grain. The SKCS isolates wheat kernels, weighs them and then crushes them in a progressively narrower gap followed by a toothed rotor and a crescent (Perten Instruments 1995). The crushing force and

electrical conductivity between the rotor and electrically isolated crescent are measured. Subsequently, all data are processed by integrated computer software for 300 kernels to provide means and standard deviations for weight, size, moisture, hardness index and hardness index distributions of all kernels (Gaines et al. 1996).

RESULTS

An annotated example of observed electrophoretic profiles for HMW-GS in a few of the wheat lines studied is shown in Figure 1. The allelic composition at the complex HMW gene loci (*Glu-A1, Glu-B1, Glu-D1*) in 116 Kenya, 52 Ethiopia and 33 founder lines is summarized in Tables 2a, 2b, and 2c respectively. All the 201 lines could be differentiated into 42 groups based on HMW-GS allelic composition. Twelve of these groups were represented by single individuals, for example combination 1, 13+16, 2+12 was recorded only in Kenyan cultivar Mbuni while 1, 13+16, 5+10 was only found in Duma. The largest group comprising of 24 lines across all subpopulations had the combination 2*, 7+8, 2+12. In more than 60% of the groups (allelic combinations), a quality score (Payne, 1987) could be assigned ranging from 4 to 10 (Tables 2a, 2b, and 2c). Higher scores are usually associated with superior bread making qualities with a score of 10 considered optimal. In all those lines that could be decisively assigned a quality score, at least one of the constituent allele was undefined in Payne's scale.

A total of 19 HMW *Glu-1* alleles were detected; 3 at *Glu-A1,* 11 at *Glu-B1,* and 5 at *Glu-D1* (Table 3). Similarly, at the LMW *Glu-3* loci, 22 alleles were identified; 7 at *Glu-A3,* 11 at *Glu-B3,* and 3 at *Glu-D3* (Table 4). Also provided in Tables 3 and 4 are frequencies of the respective detected glutenin alleles across the entire population and in the Kenyan, Ethiopian, and founder populations. At the *Glu-A1* locus, allele *Glu-A1b* designating subunit 2* was the most frequent- 58%, while allele *Glu-B1b* (subunit 7+8) and *GluD1d* (subunit 5+10) were more frequent at 34% for *Glu-B1* and 55% for *Glu-D1* loci respectively (Table 3). Alleles occurring at a frequency of 5% or less, for example *Glu-B1cb* (subunit 7^{OE}+8) at *Glu-B1* were considered rare. In this regard, a total of 9 rare alleles were observed at the HMW *Glu-1* loci including 3 undesignated alleles (subunits

13+8, 7+17, 1.5+10.5) which are probably novel and whose effects on bread making quality needs to be determined. The most frequent alleles at the LMW *Glu-3* loci were *Glu-A3c* -40%, *GluB3b*-29% and *GluD3*b-52%. At these loci, 8 rare alleles with frequencies of 5% or less were observed (Table 4).

The allele *Glu-D1d* (subunit 5+10) deemed to have excellent effects on bread making quality based on Payne scale had the highest frequency (60%) in the Kenyan subpopulation. Although this allele was at a considerably high frequency in the Ethiopian (50%) and founder (46%) subpopulations, alleles *Glu-A1b* (subunit 2^*) and *Glu-D1a* (subunit 2+12) which are predicted to be slightly inferior to allele *Glu-D1d*, were at higher frequencies in both of these populations respectively. Moreover, compared to the Kenyan and Ethiopian subpopulations, the founder population had a higher frequency of the inferior alleles including *Glu-A1c* (subunit Null) and *Glu-B1d* (subunit 6+8). Three lines; Shina, Sofumar, and Kubsa, all from Ethiopia, were found to carry the allele *Glu-B1a* (subunit 7) that is also considered deleterious to bread making quality.

A summary of the genetic diversity statistics obtained for each of the three subpopulations and across the full set of lines is given in Table 5. At the HMW loci, *Glu-B1* on the long arm of chromosome 1B showed the highest genetic diversity (H_e = 0.8) across the whole population compared to *Glu-A1* (H_e = 0.6) and *Glu-D1* (H_e = 0.5) on the same arm. The genetic diversity at the *Glu-B1* locus was also consistently higher relative to the other two HMW loci in each of the subpopulations ranging from 0.7 in Ethiopian and founder subpopulations to 0.8 in the Kenyan subpopulation. This result was also in agreement with the number of effective alleles (A_e), estimated to be a maximum of 5 in the Kenyan subpopulation, and to Shannon information index (S) that ranged from 2 in

the founder lines to 2.5 in the Kenya subpopulation. Among the LMW loci, *Glu-B3* on the short arm of chromosome 1B had the highest diversity in all subpopulations. For instance, this locus was observed to be characterized by a maximum of 7 effective alleles in the Ethiopian subpopulation. Overall, chromosome 1D (loci *Glu-D1* and *Glu-D3*) displayed the lowest diversity for the glutenin subunit alleles in the present subpopulations of wheat germplasm compared to chromosomes 1A (loci *Glu-A1* and *Glu-A3*) and 1B (loci *Glu-B*1 and *Glu-B3*).

A summary of puroindoline protein assay and other grain characteristics for 106 Kenya, 47 Ethiopia, and 32 founder lines is provided as Appendix IV. Puroindoline content, relative to the wild type cultivar Alpowa-1-soft (calibrated as content =1), ranged from 0.02 in the founder line BW21 to 1.98 in the line Kenya-paa. The mean relative puroindoline content was higher in the Kenyan subpopulation (0.75, median= 0.75) than in the Ethiopian (0.52, median=0.45) and founder subpopulation (0.62, median=0.67). As expected, relative puroindoline content was negatively correlated to actual physical texture of the grain samples as estimated through the SKCS process (Figure 2). Through SKCS profiling, 43%, 59%, and 48% of the lines were classified as hard in the Kenyan, Ethiopian, and founder subpopulations respectively (Table 6). The frequency of soft and very soft wheat classes was lowest in the Ethiopian subpopulation and highest in the Kenyan lines. Only the check durum wheat line MG0766 had a hardness index exceeding 90, on a scale of 100 and was classified as very hard.

On a 12% moisture basis, % crude protein ranged from as low as 10.0 in the founder line Yaktana 54A to 18.6 in the Ethiopia line Beltista. Both of these lines are soft wheats. Among the hard wheat class, Kenya-heroe with a crude protein content of

11.59% was the lowest compared to 17.85% in the Ethiopia landrace Sindi5. Generally, the mean crude protein in all subpopulations was above 14.5% (Appendix IV). Kernel weight and kernel diameter, both considered to directly impact on total yield were highly correlated ($r = 0.87$) and were on average higher in the Ethiopian subpopulation than in the other two subpopulations. The top 8 lines across the whole population had kernel weights on average higher than 51g. Six of these lines also had the highest kernel diameter. In addition, all these wheat lines were identified to be in the soft wheat class.

In view of results across assays, 12 lines which are all hard wheats that combine the best alleles at the HMW-GS loci, have crude protein content above 14% and have above average kernel weight and kernel diameter were identified. These lines (bolded in Tables 2a, 2b, and 2c) which are considered desirable parental genotype in breeding for bread making quality include Duma, Njoro BWII, CianoF67, Kenya Swara, K6290Bulk, ET-12-D4, Mbega, Kenya Nyangumi, Kenya 184-P, Pavon76, Kenya Popo, and Sindi 6.

DISCUSSION

Wheat in the form of bread provides more nutrients to the world population than any other single food source (Pena 2000). Globally, wheat is consumed in over 175 countries with an average consumption of 67kg/ capita in 2003, which represents more than one third of the minimum food requirements of most adults (Dubin and Brennan 2009). Bread wheat has been cultivated in the East Africa countries- Kenya and Ethiopia for more than a century with virtually all production used to meet for local demand. In recent years, the demand for bread and related food products in the region has increased sharply owing to changed diets that favor those over traditional food sources, and generally due to increased human populations. It follows that developing cultivars adapted to the East Africa region has been of paramount interest and is of priority in both Kenya and Ethiopia. It is widely accepted that breeding for high yields and yield stability as well as maintaining resistance to pests and pathogens are essential pillars of breeding programs for any crop (Tester and Langridge 2010). In the case of bread wheat, breeding for bread making quality is a top precedence because unlike maize and rice, the other major cereal crops, nearly all the wheat consumed by mankind is processed rather than eaten as whole grain. Improving quality is heavily dependent on understanding the complexities of endosperm storage proteins. Moreover, wheat grain of certain textural classes, depending on production region and the intended end use, attracts different premiums from food processors and is an important consideration in designing wheat breeding programs. An accepted consensus is that the first and essential step in plant breeding involves selection of parents to make crosses (Allard 1960, Fehr 1987, Bernado 2010). Selection of parents might best be achieved by collecting genetic information on

potential parents prior to making crosses. This not only saves resources that would otherwise have been wasted from making inferior crosses for the target trait but it is also understood that the cross with the highest progeny mean and largest genetic variance, likely if desirable parents are selected beforehand, has the potential to produce the best lines (Bernado 2003).

Extensive allelic variation in the glutenin subunit loci was found in the east Africa and founder bread wheat germplasm in the present study with a total of 19 HMW-GS alleles and 22 LMW-GS alleles respectively. While in some of the lines information on glutenin composition had previously been established (e.g. Tarekegne et al. 2005, Dessalegn et al. 2011), the present results help validate those findings. Most importantly, information about many of the historical Kenya and Ethiopia germplasm hitherto uncharacterized is an important step towards identifying superior parents. Overall, there was no clear historical trends in the glutenin composition across all lines sampled. This result indicates that lines that have superior combination of glutenin alleles are generally randomly distributed across release decades. In their work involving 42 Ethiopian grown bread wheat lines, a large proportion of which was also included in the present study, (Tarekegne et al. 2005) found that with five alleles, *GluB1* locus was the most variable amongst the three HMW-GS loci. This is consistent with results presented here which identified *GluB1* to have the highest diversity with an estimated four effective alleles. An additional overlap in the two studies is that within Ethiopian bread wheat allele *GluA1b* (subunit 2^{*}) was most frequent (> 70%) at the *GluA1* locus. In a comprehensive study involving 718 wheat cultivars and advanced lines from 20 countries, Jin et al. (2011) observed frequencies of 39.7, 83.3, 20.4, and 72.2% for allele *GluA1b* (subunit 2^{*}) in

Australia, Canadian, Chinese, and US wheats respectively. This allele was also common in the Kenya germplasm with a frequency of 54.3%. In Tarekegne et al. (2005), allele *GluB1c* (subunit 7+9) was identified as most prevalent (57.1%) among the Ethiopian bread wheat. Here, allele *GluB1b* (subunit 7+8) was found to be the most frequent at the *GluB1* locus for this group of genotypes. The discrepancy could have arisen from the larger sample size utilized in the present work.

In contrast to the east African wheat population, the allele *GluA1c* (subunit Null) has been reported to be the most prevalent in Chinese wheat. Nakamura (2000) reported a frequency of 80.4 and 74.1% for *GluA1c* in Chinese and Japanese wheat respectively. Fang et al. 2009 observed frequencies > 96% in several groups of Chinese endemic wheats. In old and modern Iranian bread wheat, Shahnejat-Bushehri et al. (2006) reported a frequency of 46% for *GluA1c*. In this study this allele, considered to be deleterious to bread making quality, was only at an elevated frequency within the founder germplasm. Although in proportionately lower frequencies in the east Africa wheat, similar to the Chinese and Japanese wheat, alleles *GluB1b* (subunit 7+8) and *GluA1c* (subunit 7+9) are the most common at the *GluB1* locus. However, at this locus allele *GluB1i* (subunit 17+18) which is rare in the 405 Chinese and Japanese varieties reported in Nakamura's study is quite enriched in the east African material, especially those from Ethiopia. Another clear difference in the germplasm from the different regions is that east African wheat seems to be richer with allele *GluD1d* (subunit 5+10) at the *GluD1* locus than either Chinese or Japanese wheat. In fact this allele has been found to be rare and presented variable frequencies in the European genepool- being extremely rare in the Swedish wheat and fairly common in Norwegian wheats (Tohver 2007).

The LMW-GS allele *GluA3b* is present in 3 and 7 Kenyan and Ethiopia lines respectively. This allele, alongside *GluB3b*, is known to have positive effects in bread making (Jin et al. 2011). Most reassuring, the lines identified to carry the *GluA3b* allele have suitable combinations of HMW-GS alleles, with Payne quality scores of 8 to 10. This indicates a potential advantage in using these genotypes as parents in bread quality improvement populations, particularly in east Africa. Indeed, the allele *GluB3b* is very frequent in Kenyan and founder wheat and provided lines that carry this allele are also positive for superior combinations of alleles at other loci, they are also good parental candidates in developing breeding populations. At the *GluA3* locus, alleles found most common in this study across Kenyan, Ethiopian, and founder subpopulations *i.e. GluA3c* and *GluA3e* were observed to fall between 29.2 – 100% and 0 – 33.3% respectively in Jin et al. 2011 research.

Apart from the insights derived from the glutenin subunit fingerprints, the incorporation of SKCS and puroindoline protein assays to empirically distinguish east African wheat kernel textural classes provides a precedence in selecting parents that combine desired trait combinations which as mentioned before are likely to provide superior progeny. Incorporating the extra information on grain texture, this study identified 12 genotypes that potentially could be used to develop breeding populations segregating for superior bread making characteristics. Considering the prerequisite that parental genotypes must also present high mean and large variance for yield, and the need that they must avail desired yield stability and resistance to pests and diseases to be of optimal breeding benefit, the list of lines highlighted could be reduced further to include these needs. Of the 12 lines, Duma and Njoro BWII are essentially rated amongst the best

performers for yield in the lower altitude drier areas and the wetter wheat growing zones of Kenya respectively. Given their susceptibility to prevalent stem rust disease races (Njau et al. 2009), these genotypes can be crossed with germplasm identified to have good yield potential, superior grain quality and rust resistance, but which may not necessary be adapted to the Kenya environment. Additionally, the lines Kenya Swara, Kenya Nyangumi, Kenya popo have been detected to have adult plant resistance to stem rust (Njau et al. 2009), including *Sr2* (Singh et al. 2008) which is considered more durable under the intense disease pressures predominant in Kenya. Using these germplasm to breed for rust resistance, good bread making performance and perhaps better yield stability appears be a worthy strategy.

Table 1 HMW *Glu-1* allelic variation and associated Payne score for bread-making quality

Glu-1A	*Glu-1A*	*Glu-1D*	Payne score[a]
		5 + 10	4
1; 2*	7 + 8; 17 + 18; 13 + 16		3
	7 + 9	2 + 12; 3 + 12	2
Null	7; 6 + 8	4+12	1

[a] Higher Payne score corresponds to better bread-making quality.

Table 2a Allelic profiles at the HMW loci (*Glu-A1, Glu-B1 and Glu-D1*) and predicted quality score for 116 Kenya bread wheat lines

Glu-A1	*Glu-B1*	*Glu-D1*	Quality score[a]	Line[b]
1	7+8	5+10	10	**1010 F3 SEL. 7**, 1016.P.2, Kenya Cheetah, Kenya Civet, **Mbega**
1	13+16	5+10	10	**Duma**
2*	7+8	5+10	10	10101AM2 (L), 1010 F3 SEL. 13 C, Kenya 6297-2
				Goblet, **Kenya Ngiri**, **Kenya Nyangumi**, **Kenya Nyati**
				H 441, **Kenya Swara**, Morris, R 64
2*	17+18	5+10	10	Kenya Kudu, Kenya Leopard, **Kenya -184-P**, Kenya -5
				Kenya Popo, **Njoro Bw II**, Reliance 261M
1	7+9	5+10	9	Kenya Mbweha, Kenya Heroe, Kenya Sungura, PW Thatcher
				1012 B.1. (L), 1200.M, 688 F4 SEL 3
2*	7+9	5+10	9	291 J.1.I.1 , Kenya Farmer, Kenya Mamba, Kwale, Pasa
				Reliance , Token-Ken, Trophy, Beacon-Ken, Bonza 63, Brewster
N	7+8	5+10	8	1076.D.7, Kenya-318-AJ-4 A-1,
N	17+18	5+10	8	Chozi, Kenya -358-AC, Mentor
1	6+8	5+10	8	Regent
1	7+8	2+12	8	RFN
1	13+16	2+12	8	Mbuni
2*	13+16	2+12	8	Kenya Yombi
2*	6+8	5+10	8	Kenya Kongoni, 869 B.3.B.5, Kenya page
2*	7+8	2+12	8	Romany, Fronthatch, Kenya 6820, Kenya Chiriku, Gem, Kenya Nyoka
				Kenya-58, Menco, Pewter, Primex
2*	7+17	5+10	8	Kenya B-256-G, Kenya Governor
2*	17+18	2+12	8	338 AA 1 A 2

[a] Quality score indicated as '-' suggest scoring not possible based on Payne et al. (1987).

[b] Lines in bold combine desirable HMW glutenin alleles and other grain quality traits and are deemed suitable breeding parents.

Table 2a *continued*

Glu-A1	*Glu-B1*	*Glu-D1*	Quality score	Line
2*	7+9	2+12	7	Kentana Yaqui, Ngamia, Salmayo
N	7+9	5+10	7	Kenya Ibis, Kenya Jay, Tama
1	7+9	2+12	7	Pitcher, Santa Elena
2*	7+9	2+12	7	Bailey, Bonny, Bounty, Fanfare, Inia66
1	7+9	3+12	7	Kenya 8
N	7+8	2+12	6	1010 F3 SEL. 4, 1061.K.1, 1061.K.4, Gabrino, Kenya Plume, Lenana,
N	17+18	2+12	6	Kenya 155
1	6+8	2+12	6	321 BT 11 B1
1	6+8	3+12	6	Kenya 131
N	7+9	2+12	5	Catcher, Fury
N	13+8	2+12	-	1016 P.1
N	13+19	1.5+12	-	Kenya -1
N	13+19	2+12	-	Kenya-294-B-2 A-3, Kenya- 117A, P. Walker Munro, Rhodesian sabanero
1	13+19	5+10	-	Kenya 501, FL I Kenya 9, Kenya 117C,
1	7+17	2+12	-	Kenya 7
1	7^{oe}+8	5+10	-	Kenya Paka, Kenya Tembo, Kenya Kanga, Kenya Nungu
2*	13+19	2+12	-	Equator, Sabanero
2*	13+19	5+10	-	Kenya-318.O.3B.2, Africa Mayo, Kenya Grange
2*	7^{oe}+8	5+10	-	Kenya Fahari, Kenya Zabadi
2*	13+19	3+12	-	B F2 36 C1 L, Kenya Standard

[a] Quality score indicated as '-' suggest scoring not possible based on Payne et al. (1987).

Table 2b Allelic profiles at the HMW loci (*Glu-A1, Glu-B1 and Glu-D1*) and predicted quality score for 52 Ethiopia bread wheat lines

Glu-A1	*Glu-B1*	*Glu-D1*	Quality score	Line
1	7+8	5+10	10	BTC Barce, MG 07793, MG 07795, **Sindi 6**
2*	7+8	5+10	10	**K 6290 Bulk**, Dodota, **ET-12-D4**, Galema
2*	17+18	5+10	10	Madda walabu, **Pavon 76**, Katar, KKBB
1	7+9	5+10	9	KBG-01, Abola, Dashen
2*	7+9	5+10	9	Wetera, Hawi, Megal, HAR-407, Simba, Tura, Tusie
1	7+8	2+12	8	Meraro, K-6106-8
2*	7	5+10	8	Shina, Sofumar, Kubsa
2*	7+8	2+12	8	Gara, ET-13 A2, Aggia Sindi, CI 14393, Mitike, K6295-4A, Sindi 1, Sindi 2, Sindi 5, Sindi 7, Sindi 8
2*	17+18	2+12	8	Wabe, Batu, Laketch
N	7+9	5+10	7	Digelu
1	7+9	3+12	7	HRS 55
2*	7+9	2+12	7	Jiru, Dereseglen,
N	7+8	2+12	6	Sindi 4
N	17+18	2+12	6	Beltista
1	6+8	2+12	6	Enkoy
2*	6+8	2+12	6	Bobicho
N	7+8	1.5+10.5	-	Dure
2*	13+19	2+12	-	MG 07768, Bale

[a] Quality score indicated as '-' suggest scoring not possible based on Payne et al. (1987).

[b] Lines in bold combine desirable HMW glutenin alleles and other grain quality traits and are deemed suitable breeding parents.

Table 2c Allelic profiles at the HMW loci (*Glu-A1, Glu-B1 and Glu-D1*) and predicted quality score for 33 founder bread wheat lines

Glu-A1	*Glu-B1*	*Glu-D1*	Quality score	Line
1	7+8	5+10	10	Sonora63
1	17+18	5+10	10	Ciano F67
2*	7+8	5+10	10	Bonito, Kentana 48, **Reward**,
2*	17+18	5+10	10	Bluebird
1	7^{oe}+8	5+10	10?	Tobari 66
1	7+9	5+10	9	Bonanza
2*	7+9	5+10	9	Bonza, Gradenero, Newthatch, Bobwhite
N	7+8	5+10	8	Red Egyptian, McMurachy
1	6+8	5+10	8	Yaktana 54A
2*	7+8	2+12	8	Bage, II-50-17, Lerma rojo
2*	17+18	2+12	8	Timstein, Zaragoza 75, Aguilera, Gabo, Kalyanosona
N	7+8	2+12	6	Wis. 245, Frocor 2328, Mentana, Penjamo 62
N	17+18	2+12	6	Olescens dwarf
1	6+8	2+12	6	Supremo
N	6+8	2+12	4	Exchange
N	13	2+12	-	184
N	13+19	2+12	-	Egyptian Na 95
1	13+19	2+12	-	Bobin

[a] Quality score indicated as '-' suggest scoring not possible based on Payne et al. (1987).

[b] Line in bold combine desirable HMW glutenin alleles and other grain quality traits and is deemed a suitable breeding parent.

Table 3 Allelic frequency at *Glu-A1*, *Glu-B1*, and *Glu-D1* loci in 168 east Africa (Kenya, Ethiopia) and founder wheat lines

HMW locus	Subunit	Allele[a]	Source						Total	
			Kenya		Ethiopia		Founder			
			No.	Freq. (%)	No.	Freq. (%)	No.	Freq. (%)	No.	Freq. (%)
Glu-A1										
	1	*a*	30	25.8	11	21.1	7	21.2	48	23.9
	2*	*b*	63	54.3	37	71.1	16	48.5	116	57.7
	Null	*c*	23	19.8	4	7.7	10	30.3	37	18.4
Glu-B1										
	7	*a*	0	0	3	5.7	0	0	3	1.5
	7+8	*b*	35	30.2	20	38.4	13	39.4	68	33.8
	7^{OE}+8	*cb*	6	5.2	0	0	1	3.0	7	3.5
	7+9	*c*	34	29.3	14	26.9	5	15.2	53	26.4
	6+8	*d*	6	5.2	2	3.8	3	9.1	11	5.5
	13+16	*f*	3	2.6	0	0	0	0	3	1.5
	13+19	*g*	16	13.8	2	3.8	2	6.1	20	10.0
	17+18	*i*	12	10.3	11	21.2	8	24.2	31	15.4
	13	*as*	0	0	0	0	1	3.0	1	0.01
	13+8	*?*	1	0.01	0	0	0	0	1	0.01
	7+17	*?*	3	2.6	0	0	0	0	3	1.5
Glu-D1										
	2+12	*a*	41	35.3	24	46.2	18	54.5	83	41.3
	3+12	*b*	4	3.4	1	1.9	0	0	5	2.5
	5+10	*d*	70	60.3	26	50.0	15	45.5	111	55.2
	1.5+12	*aj*	1	0.01	0	0	0	0	1	0.01
	1.5+10.5	*?*	0	0	1	1.9	0	0	1	0.01

[a]Allele naming follows McIntosh et al. 2003

Table 4 Allelic frequency at *Glu-A3*, *Glu-B3*, and *Glu-D3* loci in 168 east Africa (Kenya, Ethiopia) and founder wheat lines

LMW Locus	Allele[a]	Source						Total	
		Kenya		Ethiopia		Founder			
		No.	Freq. (%)	No.	Freq. (%)	No.	Freq. (%)	No.	Freq. (%)
Glu-A3									
	a	6	5.2	2	3.8	0	0	8	4.0
	b	3	2.6	8	15.4	3	9.1	14	7.0
	c	44	37.9	24	46.2	12	36.4	80	39.8
	d	7	6.0	6	11.5	4	12.1	17	8.5
	e	53	45.7	10	19.2	10	30.3	73	36.3
	f	3	2.6	2	3.8	2	6.1	7	3.5
	g	0	0	0	0	2	6.1	2	0.01
Glu-B3									
	a?	0	0	0	0	1	3.0	1	0.0
	b	41	35.3	7	13.5	9	27.3	58	28.9
	c	5	4.3	4	7.7	1	3.0	10	5.0
	c'	1	0.9	0	0	0	0	1	0.0
	d	3	2.6	3	5.8	1	3.0	7	3.5
	e	4	3.4	1	1.9	4	12.1	9	4.5
	g	2	1.7	8	15.4	1	3.0	11	5.5
	g?	9	7.8	3	5.8	3	9.1	15	7.5
	h	31	26.7	14	26.9	7	21.2	52	25.9
	i	13	11.2	4	7.7	5	15.2	21	10.5
	j	7	6.0	8	15.4	1	3.0	16	7.9
Glu-D3									
	a	33	28.4	6	11.5	8	24.2	47	23.4
	b	56	48.3	32	61.5	17	51.5	105	52.2
	c	27	23.3	14	26.9	8	24.2	49	24.4

[a]Allele naming follows McIntosh et al. 2003

Table 5 Diversity at HMW and LMW loci in the Kenya, Ethiopia and founder subpopulations of bread wheat

	Source															
	Kenya				Ethiopia				Founder				Across			
Locus	*A*	A_e	H_e	*S*	*A*	A_e	H_e	*S*	*A*	A_e	H_e	*S*	*A*	A_e	H_e	*S*
Glu-A1	3	2.5	0.6	1.4	3	1.9	0.5	1.2	3	2.7	0.6	1.5	3	2.3	0.6	1.4
Glu-B1	9	4.8	0.8	2.5	7	3.9	0.7	2.2	6	3.9	0.7	2.0	11	4.7	0.8	2.5
Glu-D1	4	1.7	0.4	1.5	4	2.2	0.5	1.1	2	2.0	0.5	1.0	5	2.1	0.5	1.1
Glu-A3	6	2.8	0.6	1.8	6	3.5	0.7	2.2	6	3.8	0.7	2.2	7	3.3	0.7	2.1
Glu-B3	10	4.5	0.8	2.6	9	6.6	0.9	2.9	9	5.9	0.8	2.8	11	5.6	0.8	2.8
Glu-D3	3	2.7	0.6	1.5	3	2.2	0.5	2.6	3	2.6	0.6	1.5	3	2.9	0.7	1.5

A, Mean number of alleles per locus

A_e, Effective number of alleles per locus

H_e, Genetic diversity index

S, Shannon's index of diversity

Table 6 Frequency distribution among 5 hardness classes in 122 Kenyan, 60 Ethiopian, and 46 founder bread wheat lines

		Source							
		Kenya		Ethiopia		Founder		Total	
Class	Index[a]	No.[b]	Freq. (%)[c]	No.	Freq. (%)	No.	Freq. (%)	No.	Freq. (%)
Very hard	> 90	0	0	1	0	0	0	0	0
Hard	75-89	53	43	35	59	22	48	109	48
Med. Hard	55-74	0	0	0	0	0	0	0	0
Soft	35-54	29	24	6	10	7	15	42	19
Very soft	< 34	40	33	18	31	17	37	75	33

[a] Hardness index, resistance to deformation derived from SKCS

[b] No., Number of lines

[c] Freq. (%), Frequency

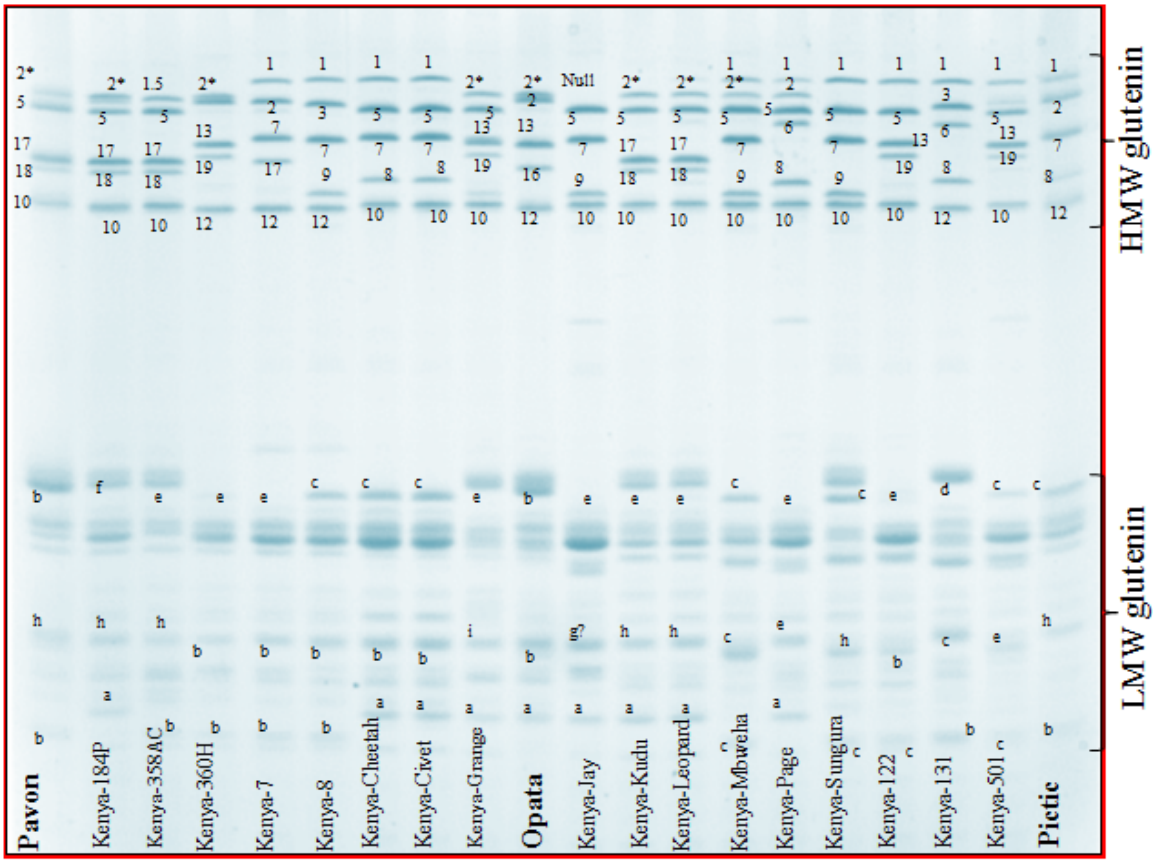


Figure 1 SDS-PAGE profiles of HMW-GS and LMW-GS alleles representing a section of bread wheat lines assayed. Standards (Pavon, Opata, and Pictic) are indicated in bold.

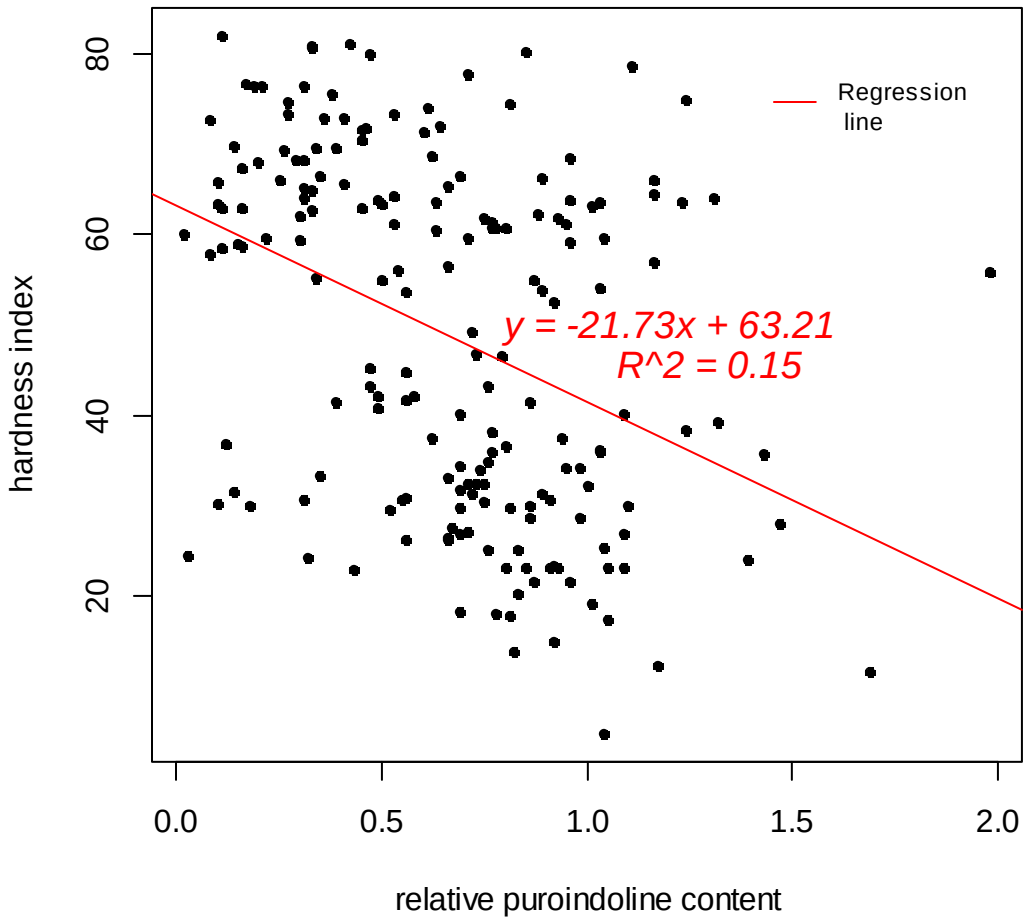


Figure 2 Relationship between relative puroindoline content (obtained with respect to Alpowa-1-soft) and hardness index among 185 East Africa (Kenya, Ethiopia) and founder bread wheat lines.

REFERENCES

AACC International. 2010. Approved Methods of Analysis, 11th Edition Methods 46-30.1. Available online only, AACC International: St. Paul.

AACC International. 2010. Approved Methods of Analysis, 11th Edition Methods 55-31.01. Available online only, AACC International: St. Paul.

Abdallah, J. M., B. Goffinet, C. Cierco-Ayrolles, and M. Perez-Enciso. 2003. Linkage disequilibrium fine mapping of quantitative trait loci: A simulation study. Genetics, Selection, Evolution: GSE 35 (5): 513-32.

Abdurakhmonov, I. Y., and A. Abdukarimov. 2008. Application of association mapping to understanding the genetic diversity of plant germplasm resources. International Journal of Plant Genomics 2008: 1-18.

Afzal, S. N., M.I. Haque, M.S. Ahmedani, S. Bashir, and A.R. Rattu. 2007. Assessment of yield losses caused by *Puccinia striiformis* triggering stripe rust in the most common wheat varieties. Pak. J. Bot 39 (6): 2127-34.

Akbari, M., P. Wenzl, V. Caig, J. Carling, L. Xia, S. Yang, G. Uszynski, V. Mohler, A. Lehmensiek, and H. Kuchel. 2006. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theoretical and Applied Genetics 113 (8): 1409-1420.

Akhunov, E., C. Nicolet, and J. Dvorak. 2009. Single nucleotide polymorphism genotyping in polyploid wheat with the illumina GoldenGate assay. Theoretical and Applied Genetics 119 (3): 507-517.

Akhunov, E. D., A. R. Akhunova, O. D. Anderson, J. A. Anderson, N. Blake, M. T. Clegg, D. Coleman-Derr, E. J. Conley, C. C. Crossman, and K. R. Deal. 2010. Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genomics 11 (1): 702-723.

Allard, R.W. 1960. Principles of plant breeding. New York, john wiley & sons. USA.

Allen, A. M., G. L. A. Barker, S. T. Berry, J. A. Coghill, R. G. William, S. Kirby, P. Robinson, R. C. Brenchley, R. D'Amore, and N. McKenzie. 2011. Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat. Plant Biotechnology Journal 9: 1086-1099.

Anderson, J.A., G.A. Churchill, J.E. Autrique, S.D. Tanksley, and M.E. Sorrells. 1993. Optimizing parental selection for genetic linkage maps. Genome 36 (1): 181-186.

Antao, T., A. Lopes, R. J. Lopes, A. Beja-Pereira, and G. Luikart. 2008. LOSITAN: A workbench to detect molecular adaptation based on an fst-outlier method. BMC Bioinformatics 9 (1): 323-327.

Balding, D. J. 2006. A tutorial on statistical methods for population association studies. Nature Reviews Genetics 7 (10): 781-791.

Bariana, H.S., U.K. Bansal, A. Schmidt, A. Lehmensiek, J. Kaur, H. Miah, N. Howes, and CL McIntyre. 2010. Molecular mapping of adult plant stripe rust resistance in wheat and identification of pyramided QTL genotypes. Euphytica 176 (2): 251-260.

Barrett, B.A., K.K. Kidwell, and P.N. Fox. 1998. Comparison of AFLP and pedigree-based genetic diversity assessment methods using wheat cultivars from the Pacific Northwest. Crop Science 38 (5): 1271-1278.

Bazin, E., K. J. Dawson, and M. A. Beaumont. 2010. Likelihood-free inference of population structure and local adaptation in a bayesian hierarchical model. Genetics 185 (2): 587-602.

Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London. Series B: Biological Sciences 263 (1377): 1619-1626.

Belzile, F. B., D. J. Somers, T. B. Banks, P. 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 (6): 557-567.

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. Series B (Methodological): 289-300.

Bernardo, R. 2003. Parental selection, number of breeding populations, and size of each population in inbred development. Theoretical and Applied Genetics 107 (7): 1252-1256.

Bhavani, S., RP Singh, O. Argillier, J. Huerta-Espino, S. Singh, P. Njau, S. Brun, S. Lacam, and N. Desmouceaux. 2011. Mapping durable adult plant stem rust resistance to the race *Ug99* group in six CIMMYT wheats. Paper presented at 2011 BGRI Technical Workshop, June, 13-16, University of Minnesota and USDA Cereal Disease Lab, St Paul, Minnesota.

Borner, A., V. Korzun, and A. J. Worland. 1998. Comparative genetic mapping of loci affecting plant height and development in cereals. Euphytica 100: 245-248.

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32 (3): 314-331.

Bouchet, S., D. Pot, M. Deu, J. F. Rami, C. Billot, X. Perrier, R. Rivallan, L. Gardes, L. Xia, and P. Wenzl. 2012. Genetic structure, linkage disequilibrium and signature of selection in sorghum: Lessons from physically anchored DArT markers. PloS One 7 (3): e33470.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23 (19): 2633-2635.

Breseghello, F., and M. E. Sorrells. 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172 (2): 1165-1177.

Brites, Carla Moita, Santos, Carla Alexandra Lourenço dos, Ana Sofia Bagulho, and Maria Luísa Beirão-da-Costa. 2008. Effect of wheat puroindoline alleles on functional properties of starch. European Food Research and Technology 226 (5): 1205-12.

Brown, T. A., M. K. Jones, W. Powell, and R. G. Allaby. 2009. The complex origins of domesticated crops in the fertile crescent. Trends in Ecology & Evolution 24 (2): 103-109.

Brown, W. L. 1983. Genetic diversity and genetic vulnerability-an appraisal. Economic Botany 37 (1): 4-12.

Bryan, G.J., P. Stephenson, A. Collins, J. Kirby, JB Smith, and MD Gale. 1999. Low levels of DNA sequence variation among adapted genotypes of hexaploid wheat. Theoretical and Applied Genetics 99 (1): 192-198.

Buckler, E. S., and J. M. Thornsberry. 2002. Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5: 107-111.

Caballero, L., LM Martin, and JB Alvarez. 2004. Variation and genetic diversity for gliadins in Spanish spelt wheat accessions. Genetic Resources and Crop Evolution 51 (7): 679-86.

Caldwell, R. M. 1968. Breeding for general and/or specific plant disease resistance. Paper presented at Proc. of the Int. Wheat Genetics Symp., Third. Butterworth & Co., Sydney, Australia.

Cavanagh, C. *et al.* Tracing the selection history of wheat by high-density SNP genotyping of world populations (*In press*).

Chan, E. K. F., H. C. Rowe, and D. J. Kliebenstein. 2010. Understanding the evolution of defense metabolites in arabidopsis thaliana using genome-wide association mapping. Genetics 185 (3): 991-1007.

Chao, S., W. Zhang, E. Akhunov, J. Sherman, Y. Ma, M. C. Luo, and J. Dubcovsky. 2009. Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L.) cultivars. Molecular Breeding 23 (1): 23-33.

Chao, Shiaoman, Jorge Dubcovsky, Jan Dvorak, Ming-Cheng Luo, Stephen Baenziger, Rustam Matnyazov, Dale Clark, et al. 2010. Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). BMC Genomics 11 (1): 727-743.

Chen, X.M. 2005. Epidemiology and control of stripe rust (*Puccinia striiformis* f. sp. *tritici*) on wheat. Canadian Journal of Plant Pathology 27 (3): 314-337.

Chen, X.M., and R. F. Line. 1995. Gene action in wheat cultivars for durable, high-temperature, adult-plant resistance and interaction with race-specific, seedling resistance to *Puccinia striiformis*. Phytopathology 85 (5): 567-572.

Chevenet, F., C. Brun, A. L. Bañuls, B. Jacq, and R. Christen. 2006. TreeDyn: Towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics 7 (1): 439.

Churchill, G. A., and R. W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138 (3): 963-971.

Committee on Genetic Vulnerability of Major Crops. 1972. Genetic Vulnerability of Major Crops. Natl. Acad. Sci., Washington, DC.

Condón, F., D. C. Rasmusson, E. Schiefelbein, G. Velasquez, and K. P. Smith. 2009. Effect of advanced cycle breeding on genetic gain and phenotypic diversity in barley breeding germplasm. Crop Science 49 (5): 1751-1761.

Crisci, J. L., and J. D. Jensen. 2012. Evolution of the human genome: Adaptive changes. *eLS*.

Crossa, J., J. Burgueño, S. Dreisigacker, M. Vargas, S. A. Herrera-Foessel, M. Lillemo, R. P. Singh, R. Trethowan, M. Warburton, and J. Franco. 2007. Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 177 (3): 1889-913.

Curtis, B.C., S. Rajaram, and H. Gómez Macpherson. 2002. Bread what Improvement and production. FAO, Rome.

de Bakker, P. I. W., R. Yelensky, I. Pe'er, S. B. Gabriel, M. J. Daly, and D. Altshuler. 2005. Efficiency and power in genetic association studies. Nature Genetics 37 (11): 1217-1223.

Dessalegn, T., C.S. Van Deventer, M.T. Labuschagne, and H. Martens. 2011. Allelic variation of HMW glutenin subunits of Ethiopian bread wheat cultivars and their quality. African Crop Science Journal 19 (2): 55-63.

Dixon, G.E. 1960. A review of wheat breeding in Kenya. Euphytica 9 (2): 209-221.

Donini, P., J.R. Law, RMD Koebner, JC Reeves, and RJ Cooke. 2000. Temporal trends in the diversity of UK wheat. TAG Theoretical and Applied Genetics 100 (6): 912-917.

D'Ovidio, R., and Stefania M. 2004. The low-molecular-weight glutenin subunits of wheat gluten. Journal of Cereal Science 39 (3): 321-39.

Dubcovsky, J., and J. Dvorak. 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. Science 316 (5833): 1862-1866.

Dubin, H.J., and J. P. Brennan. 2009. Combating stem and leaf rust of wheat: Historical perspective, impacts, and lessons learned Intl Food Policy Res Inst.

Dudbridge, F., and A. Gusnanto. 2008. Estimation of significance thresholds for genomewide association scans. Genetic Epidemiology 32 (3): 227-234.

Dvorak, J., P. E. McGuire, and B. Cassidy. 1988. Apparent sources of the A genomes of wheats inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. Genome 30 (5): 680-689.

Estoup, A., P. Jarne, and J. M. Cornuet. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Molecular Ecology 11 (9): 1591-604.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology 14 (8): 2611-2620.

Evans, L.E., J.W. Martens, G.J. Green, and E.A. Hurd. 1969. Sources of resistance to wheat stem rust in east Africa. Canadian Journal of Plant Science 49 (6): 649-654.

Excoffier, L., T. Hofer, and M. Foll. 2009. Detecting loci under selection in a hierarchically structured population. Heredity 103 (4): 285-298.

Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under linux and windows. Molecular Ecology Resources 10 (3): 564-567.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure: Extensions to linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

Fang, J., Y. Liu, J. Luo, Y. Wang, P.R. Shewry, and G. He. 2009. Allelic variation and genetic diversity of high molecular weight glutenin subunit in Chinese endemic wheats (*Triticum aestivum* L.). Euphytica 166 (2): 177-82.

FAOSTAT, 2009. Agriculture organization of the United Nations. Statistical Database.

Fehr, W. R., 1987. Principles of cultivar development. Volume 2. Macmillan publishing company.

Flint-Garcia, S. A., J. M. Thornsberry, and B. IV. 2003. Structure of linkage disequilibrium in plants. Annual Review of Plant Biology 54 (1): 357-374.

Foll, M. 2012. BayeScan v2. 1 user manual. Ecology 20: 1450-1462.

Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A bayesian perspective. Genetics 180 (2): 977-993.

Fu, Y. B., G. W. Peterson, J. K. Yu, L. Gao, J. Jia, and K. W. Richards. 2006. Impact of plant breeding on genetic diversity of the Canadian hard red spring wheat germplasm as revealed by EST-derived SSR markers. Theoretical and Applied Genetics 112 (7): 1239-1247.

Fu, Y. B., and D. J. Somers. 2009. Genome-wide reduction of genetic diversity in wheat breeding. Crop Science 49 (1): 161-168.

Gaines, CS, PF Finney, LM Fleege, and LC Andrews. 1996. Predicting a hardness measurement using the single-kernel characterization system. Cereal Chemistry. 73 (2): 278.

Germán, S., A. Barcellos, M. Chaves, M. Kohli, P. Campos, and L. de Viedma. 2007. The situation of common wheat rusts in the southern cone of America and perspectives for control. Crop and Pasture Science 58 (6): 620-630.

Gianibelli, M.C., O.R. Larroque, F. MacRitchie, and C.W. Wrigley. 2001. Biochemical, genetic, and molecular characterization of wheat glutenin and its component subunits. Cereal Chemistry 78 (6): 635-646.

Ginkel, M., and S. Rajaram. 1993. Breeding for durable resistance to diseases in wheat: An international perspective. Durability of Disease Resistance: 259-272.

Giroux, M.J., Sripo, T., Gerhardt, S. and Sherwood, J.E. 2003. Puroindolines: their role in grain hardness and plant defense. In Biotechnology and Genetic Engineering Reviews (Stephen, E.H., *ed*.), Vol. 20, pp. 277–290. Hampshire: Intercept.

Glazmann, J. C., Kilian, B., Upadhyaya, H. D., & Varshney, R. K. (2010). Assessing genetic diversity for crop improvement. Current Opinion in Plant Biology, 13: 1-7.

Gupta, P. K., S. Rustgi, and P. L. Kulwal. 2005. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Molecular Biology 57 (4): 461-485.

Gupta R.B., Shepherd K.W.: Two-step one-dimensional SDS-PAGE analysis of LMW subunits of glutelin.Variation and genetic control of the subunits in hexaploid wheats. Theoretical and Applied Genetics 1990, 80(1)**:**65-74.

Haley, C. S., and S. A. Knott. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69 (4): 315.

Hamblin, M. T., A. M. Casa, H. Sun, S. C. Murray, A. H. Paterson, C. F. Aquadro, and S. Kresovich. 2006. Challenges of detecting directional selection after a bottleneck: Lessons from sorghum bicolor. Genetics 173 (2): 953-964.

Hamblin, M. T., T. J. Close, P. R. Bhat, S. Chao, J. G. Kling, K. J. Abraham, T. Blake, W. S. Brooks, B. Cooper, and C. A. Griffey. 2010. Population structure and linkage disequilibrium in US barley germplasm: Implications for association mapping. Crop Science 50 (2): 556-566.

Hare, R.A., and R.A. McIntosh. 1979. Genetic and cytogenetic studies of durable adult-plant resistances in hope and related cultivars to wheat rusts. Zeitschrift Fur Pflanzenzuchtung 83 : 350-367.

Haudry, A., A. Cenci, C. Ravel, T. Bataillon, D. Brunel, C. Poncet, I. Hochu, S. Poirier, S. Santoni, and S. Glemin. 2007. Grinding up wheat: A massive loss of nucleotide diversity since domestication. Molecular Biology and Evolution 24 (7): 1506-1517.

Heisey, P. W., M. A. Lantican, and H. J. Dubin. 2002. Impacts of international wheat breeding research in developing countries, 1966-97. CIMMYT.

Herrera-Foessel, S. A., R. P. Singh, J. Huerta-Espino, G. M. Rosewarne, S. K. Periyannan, L. Viccars, V. Calvo-Salazar, C. Lan, and E. S. Lagudah. 2012. Lr68: A new gene conferring slow rusting resistance to leaf rust in wheat. Theoretical and Applied Genetics: 1-12.

Hodson, DP. 2011. Shifting boundaries: Challenges for rust monitoring. Euphytica 179 (1): 93-104.

Hoisington, D., M. Khairallah, T. Reeves, J. M. Ribaut, B. Skovmand, S. Taba, and M. Warburton. 1999. Plant genetic resources: What can they contribute toward increased crop productivity? Proceedings of the National Academy of Sciences 96 (11): 5937-5943.

Horton, M. W., A. M. Hancock, Y. S. Huang, C. Toomajian, S. Atwell, A. Auton, N. W. Muliyati, A. Platt, F. G. Sperone, and B. J. Vilhjálmsson. 2012. Genome-wide patterns of genetic variation in worldwide arabidopsis thaliana accessions from the RegMap panel. Nature Genetics 44 (2): 212-216.

Huang, B. E., A. W. George, K. L. Forrest, A. Kilian, M. J. Hayden, M. K. Morell, and C. R. Cavanagh. 2012. A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotechnology Journal: 826-839.

Huang, X. Q., M. Wolf, M. W. Ganal, S. Oxford, R. Koebner, and M. S. Röder. 2007. Did modern plant breeding lead to genetic erosion in European winter wheat varieties? Crop Science 47 (1): 343-349.

Hufford, M. B., P. Lubinksy, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. R. Ibarra. 2012. The genomic signature of crop-wild introgression in maize. Preprint.

Jackson E.A., Morel M.H., Sontag-Strohm T, Branlard G, Metakovsky E.V., Redaelli R: Proposal for combining the classification systems of alleles of *Gli-1* and *Glu-3* loci in bread wheat (*Triticum aestivum* L.). *J Genet Breed* 1996, 50**:**321-336.

Jannink, J. L., and B. Walsh. 2002. Association mapping in plant populations. Quantitative Genetics, Genomics and Plant Breeding, CAB International: 59-68.

Jeffreys, H. 1961. Theory of probability. Oxford University Press, New York.

Jennings, H. S. 1917. The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage. Genetics 2 (2): 97.

Jin, H., J. Yan, R.J. Peña, X.C. Xia, A. Morgounov, L.M. Han, Y. Zhang, and Z.H. He. 2011. Molecular detection of high-and low-molecular-weight glutenin subunit genes in common wheat cultivars from 20 countries using allele-specific markers. Crop and Pasture Science 62 (9): 746-54.

Jin, Y., L.J. Szabo, Z.A. Pretorius, R.P. Singh, R. Ward, and T. Fetch Jr. 2008. Detection of virulence to resistance gene *Sr24* within race TTKS of *Puccinia graminis* f. sp. *tritici*. Plant Disease 92 (6): 923-926.

Johal, G. S., P. Balint-Kurti, and C. F. Weil. 2008. Mining and harnessing natural variation: A little MAGIC. Crop Science 48 (6): 2066-2073.

Johnson, R. 1992. Past, present and future opportunities in breeding for disease resistance, with examples from wheat. Euphytica 63 (1): 3-22.

Johnson, R. 1978. Practical breeding for durable resistance to rust diseases in self-pollinating cereals. Euphytica 27 (2): 529-540.

Johnson, R. C., G. W. Nelson, J. L. Troyer, J. A. Lautenberger, B. D. Kessing, C. A. Winkler, and S. J. O'Brien. 2010. Accounting for multiple comparisons in a genome-wide association study. BMC Genomics 11 (1): 724.

Jolly, C.J., S. Rahman, AA Kortt, and T.J.V. Higgins. 1993. Characterization of the wheat mr 15000 "grain-softness protein" and analysis of the relationship between its accumulation in the whole seed and grain softness. Theoretical and Applied Genetics 86 (5): 589-97.

Jombart, T. 2012. A tutorial for discriminant analysis of principal components (DAPC) using adegenet 1.3-4. cran.r-project.org/web/packages/adegenet/.../adegenet-dapc.pdf.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genetics 11 (1): 94-108.

Kalinowski, S. T. 2004. Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. Conservation Genetics 5 (4): 539-543.

Kalinowski, S. T. 2005. hp-rare 1.0: A computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes 5 (1): 187-189.

Kent, N. L., & Evers, A. D. 1994. Kent's technology of cereals (4th ed.). An introduction for students of food science and agriculture. UK: Pergamon Publication.

Kihara, H. 1982. Wheat studies: Retrospect and prospects. Elsevier Scientific Publishing Company.

Kilian, B., and H. Ozkan. 2012. A case study for barley. Agro biodiversity Conservation: Securing the Diversity of Crop Wild Relatives and Landraces: 186.CABI Publishing.

Kilpatrick, R.A. 1975. New wheat cultivars and longevity of rust resistance, 1971-75. Agricultural Research Service, US Department of Agriculture.

Kinyua, M. G., B. Otukho, and O. S. Abdalla. 2000. Developing wheat varieties for the drought-prone areas of Kenya: 1996–1999. The 11th Regional wheat workshop for Eastern, Central and Southern Africa. Addis Ababa, Ethiopia, CIMMYT pp 45-52.

Knott, D. R. 1989. The wheat rusts-breeding for resistance. Springer-Verlag. London UK.

Kolmer, J. A. 1996. Genetics of resistance to wheat leaf rust 1. Annual Review of Phytopathology 34 (1): 435-455.

Kolmer, J. A, Y. Jin, and DL Long. 2007. Wheat leaf and stem rust in the United States. Crop and Pasture Science 58 (6): 631-8.

Kosambi, D. D. 1943. The estimation of map distances from recombination values. Annals of Human Genetics 12 (1): 172-5.

Kraakman, A. T. W., R. E. Niks, P. M. Van den Berg, P. Stam, and F. A. Van Eeuwijk. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168 (1): 435-46.

Kraakman, A. T. W., F. Martinez, B. Mussiraliev, F.A. Van Eeuwijk, and R. E. Niks. 2006. Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. Molecular Breeding 17 (1): 41-58.

Lathbury, R.J. 1947. The development of wheat varieties in Kenya. Emp. J. Exp. Agric 15: 177-88.

Law, C.N., E. Suarez, T.E. Miller, and A. J. Worland. 1998. The influence of the group 1 chromosomes of wheat on ear-emergence times and their involvement with vernalization and day length. Heredity 80 (1): 83-91.

Leonard, K. J., and L. J. Szabo. 2005. Stem rust of small grains and grasses caused by puccinia graminis. Molecular Plant Pathology 6 (2): 99-111.

Lewontin, R.C., and J. Krakauer. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74 (1): 175-195.

Liang, S. S., K. Suenaga, Z.H. He, Z. L. Wang, H. Y. Liu, D. S. Wang, R. P. Singh, P. Sourdille, and X. C. Xia. 2006. Quantitative trait loci mapping for adult-plant resistance to powdery mildew in bread wheat. Phytopathology 96 (7): 784-789.

Line, R. F., and X. Chen. 1995. Successes in breeding for and managing durable resistance to wheat rusts. Plant Disease 79 (12): 1254-1255.

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

Liu, Y., Z.Y. Xiong, Yong-Gang He, P. R. Shewry, and Guang-yuan He. 2007. Genetic diversity of HMW glutenin subunit in Chinese common wheat (Triticum aestivum L.) landraces from hubei province. Genetic Resources and Crop Evolution 54 (4): 865-74.

Long D.L., Kolmer J. A. 1989. A North American system of nomenclature for *Puccinia recondita* f. sp. *tritici.* Phytopathology 79:525–29

Long, D. L., K. J. Leonard, and J. J. Roberts. 1998. Virulence and diversity of wheat leaf rust in the United States in 1993 to 1995. Plant Disease 82 (12): 1391-400.

Lookhart, G. L., K. Hagman, and D. D. Kasarda. 1993. High-Molecular-Weight glutenin subunits of the most commonly grown wheat cultivars in the US in 1984. Plant Breeding 110 (1): 48-62.

Lu, H., and R. Bernardo. 2001. Molecular marker diversity among current and historical maize inbreds. Theoretical and Applied Genetics 103 (4): 613-617.

Lu, Y., C. Lan, S. Liang, X. Zhou, D. Liu, G. Zhou, Q. Lu, J. Jing, M. Wang, and X. Xia. 2009. QTL mapping for adult-plant resistance to stripe rust in Italian common wheat cultivars libellula and strampelli. Theoretical and Applied Genetics 119 (8): 1349-1359.

Lukow O. M., Payne P. I., Tkachuk R. 1989. The HMW glutenin subunit composition of Canadian wheat cultivars and their association with bread-making quality. J Sci Food Agric 46:451 - 460.

Luciani, T.J., 2009. RSvgDevice: An R SVG Graphics Device.http://www.darkridge.com/wjake/RSvg/ (accessed 11.07.12).

Mackay, I., and W. Powell. 2007. Methods for linkage disequilibrium mapping in crops. Trends in Plant Science 12 (2): 57-63.

Mago, R., G. Brown-Guedira, S. Dreisigacker, J. Breen, Y. Jin, R. Singh, R. Appels, ES Lagudah, J. Ellis, and W. Spielmeyer. 2011. An accurate DNA marker assay for stem rust resistance gene Sr2 in wheat. TAG Theoretical and Applied Genetics 122 (4): 735-44.

Malinga, J. N., M. G. Kinyua, A. W. Kamau, J. K. Wanjama, J. O. Awalla, and R. S. Pathak. 2007. Biotypic and genetic variation within tropical populations of russian wheat aphid, *diuraphis noxia* (kurdjumov) in Kenya. J. Entomol 4: 350-61.

Mariac, C., Lea Jehin, A. A. Saidou, A. C. Thuillet, M. Couderc, P. Sire, H. Judge, H. Adam, G. Bezançon, and J. L. Pham. 2011. Genetic basis of pearl millet adaptation along an environmental gradient investigated by a combination of genome scan and association mapping. Molecular Ecology 20 (1): 80-91.

Mariam, H. G. 1991. Wheat production and research in Ethiopia: A historical perspective. Addis Ababa: IAR/CIMMYT.

Martin, D. J., and B. G. Stewart. 1990. Dough stickiness in rye-derived wheat cultivars. Euphytica 51 (1): 77-86.

Makanda, D. W., and J. F. Oehmke, 1995. Kenya's Wheat Agriculture: Past, Present and Future. Staff Paper # 95-54, Department of Agricultural Economics, Michigan State University.

McCouch, S. 2004. Diversifying selection in plant breeding. PLoS Biology 2 (10): e347.

McDonald, B. A., and C. Linde. 2002. Pathogen population genetics, evolutionary potential, and durable resistance. Annual Review of Phytopathology 40 (1): 349-379.

McFadden, E. S., and E. R. Sears. 1946. The origin of *Triticum spelta* and its free-threshing hexaploid relatives. Journal of Heredity 37 (3): 81-9.

McIntosh, R. A. 1988. The role of specific genes in breeding for durable stem rust resistance in wheat and triticale. Breeding Strategies for Resistance to the Rusts of Wheat. CIMMYT, El Batan, Mexico: 1-9.

McIntosh, R. A. 1992. Close genetic linkage of genes conferring adult-plant resistance to leaf rust and stripe rust in wheat. Plant Pathology 41 (5): 523-527.

McIntosh, R. A., C. R. Wellings, and R. F. Park. 1995. Wheat rusts: An atlas of resistance genes. CSIRO Publishing.

McIntosh, R. A., Y. Yamazaki, KM Devos, J. Dubcovsky, WJ Rogers, and R. Appels. 2003. Catalogue of gene symbols for wheat.

Miedaner, T., and V. Korzun. 2012. Marker-assisted selection for disease resistance in wheat and barley breeding. Phytopathology 102 (6): 560-566.

Milus, E. A., K. Kristensen, and M. S. Hovmøller. 2009. Evidence for increased aggressiveness in a recent widespread strain of *Puccinia striiformis* f. sp. *tritici* causing stripe rust of wheat. Phytopathology 99 (1): 89-94.

Mohammadi, S. A., and B. M. Prasanna. 2003. Analysis of genetic diversity in crop plants-salient statistical tools and considerations. Crop Science 43 (4): 1235-1248.

Morgounov, A. I., V. A. Vlasenko, A. McNab, and H. J. Braun. 1996. Wheat breeding: Objectives, methodology, and progress. CIMMYT.

Morris, C.F, G.A Greenblatt, A.D. Bettge, and H.I. Malkawi. 1994. Isolation and characterization of multiple forms of friabilin. Journal of Cereal Science 20 (2): 167-74.

Morris, C. F. 2002. Puroindolines: The molecular genetic basis of wheat grain hardness. Plant Molecular Biology 48 (5): 633-47.

Müller, B.U., B. Stich, and H.P. Piepho. 2010. A general method for controlling the genome-wide type I error rate in linkage and association mapping experiments in plants. Heredity 106 (5): 825-831.

Muñoz-Amatriaín, M., L. Cistué, Y. Xiong, H. Bilgic, AD Budde, MR Schmitt, KP Smith, PM Hayes, and GJ Muehlbauer. 2010. Structural and functional characterization of a winter malting barley. Theoretical and Applied Genetics 120 (5): 971-84.

Nakamura, H. 2000. Allelic variation at high-molecular-weight glutenin subunit loci, glu-A1, glu-B1 and glu-D1, in Japanese and Chinese hexaploid wheats. Euphytica 112 (2): 187-93.

Narum, S. R., and J. O. N. E. Hess. 2011. Comparison of FST outlier tests for SNP loci under selection. Molecular Ecology Resources 11: 184-94.

Nazari, K., A. Yahyaoui, O. Abdalla, M. Nachit, F. Ogbonnaya, R. Brettell, S. Rajaram, and S. Aleppo. 2010. Wheat Rust Diseases in Central and West Asia and North Africa (CWANA) and Breeding for the Multiple Disease Resistance. ICARDA/CIMMYT

Nei, M. 1972. Genetic distance between populations. American Naturalist: 283-292.

Neumann, K., B. Kobiljski, S. Denčić, RK Varshney, and A. Börner. 2011. Genome-wide association mapping: A case study in bread wheat (*Triticum aestivum* L.). Molecular Breeding 27 (1): 37-58.

Nevo, E. 2005. Population genetic structure of wild barley and wheat in the near east fertile crescent: Regional and local adaptive patterns. Cereal Genomics: 135-163.

Nevo, E., and P.I. Payne. 1987. Wheat storage proteins: Diversity of HMW glutenin subunits in wild emmer from Israel. Theoretical and Applied Genetics 74 (6): 827-36.

Newell, M. A., D. Cook, N. A. Tinker, and J. L. Jannink. 2011. Population structure and linkage disequilibrium in oat (*Avena sativa* L.): Implications for genome-wide association studies. Theoretical and Applied Genetics 122 (3): 623-32.

Nielsen, R. 2005. Molecular signatures of natural selection. Annual Review of Genetics 39: 197-218.

Njau, P. N., R. Wanyera, G. K. Macharia, J. Macharia, R. Singh, and B. Keller. 2009. Resistance in Kenyan bread wheat to recent eastern African isolate of stem rust, *Puccinia graminis* f. sp. *tritici*, *Ug99*. Journal of Plant Breeding and Crop Science 1 (2): 22-27.

Njau, P. N., Y. Jin, J. Huerta-Espino, B. Keller, and R. P. Singh. 2010. Identification and evaluation of sources of resistance to stem rust race *Ug99* in wheat. Plant Disease 94 (4): 413-9.

Oraguzie, N. C., P. L. Wilcox, E. H. A. Rikkerink, and H. N. Silva. 2007. Linkage disequilibrium. Association Mapping in Plants: 11-39.

Ornella, L., S. Singh, P. Perez, J. Burgueño, R. Singh, E. Tapia, S. Bhavani, S. Dreisigacker, H. J. Braun, and K. Mathews. 2012. Genomic prediction of genetic values for resistance to wheat rusts. The Plant Genome 5 (3): 136-148.

Osborne, T. B. 1907. The proteins of the wheat kernel. Carnegie institution of Washington.

Park, R., T. Fetch, D. Hodson, Y. Jin, K. Nazari, M. Prashar, and Z. Pretorius. 2011. International surveillance of wheat rust pathogens: Progress and challenges. Euphytica 179 (1): 109-117.

Parlevliet, J. E. 1975. Partial resistance of barley to leafrust, *Puccinia hordei*. Effect of cultivar and development stage on latent period. Euphytica 24 (1): 21-27.

Parlevliet, J. E. 2002. Durability of resistance against fungal, bacterial and viral pathogens; present situation. Euphytica 124 (2): 147-156.

Pasam, R. K., R. Sharma, M. Malosetti, F. A. van Eeuwijk, G. Haseneyer, B. Kilian, and A. Graner. 2012. Genome-wide association studies for agronomical traits in a worldwide spring barley collection. BMC Plant Biology 12 (1): 16.

Pasha, I., F.M. Anjum, and C.F. Morris. 2010. Grain hardness: A major determinant of wheat quality. Food Science and Technology International 16 (6): 511-22.

Payne, P.I., and Kathryn C. 1979. Subunit composition of wheat glutenin proteins, isolated by gel filtration in a dissociating medium. Planta 145 (1): 83-8.

Payne, P.I., and G.J. Lawrence. 1983. Catalogue of alleles for the complex gene loci. *glu-A1, glu-B1 and glu-D1* which code for high molecular weight subunits of glutenin in hexaploid wheat. Journal of Cereal Res Common: 29-35.

Payne, P.I., E.A. Jackson, L.M. Holt, and C.N. Law. 1984. Genetic linkage between endosperm storage protein genes on each of the short arms of chromosomes 1A and 1B in wheat. Theoretical and Applied Genetics 67 (2): 235-43.

Payne, P.I., Genetics of wheat storage protein and the effect of allelic variation on pan bread quality**.** Annu Rev Plant Physiol 1987, 38**:**141-153.

Payne, P.I., Mark A. Nightingale, Anatole F. Krattiger, and Linda M. Holt. 1987. The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties. Journal of the Science of Food and Agriculture 40 (1): 51-65.

Peña, R.J., B.C. Curtis, and S. Rajaram. 2002. Wheat for bread and other foods. Bread Wheat Improvement and Production. Food and Agriculture Organization of the United Nations. Rome: 483-542.

Peng, J. H., D. Sun, and E. Nevo. 2011. Domestication evolution, genetics and genomics in wheat. Molecular Breeding 28 (3): 281-301.

Peterson, R. F., A. B. Campbell, and A. E. Hannah. 1948. A diagrammatic scale for estimating rust intensity on leaves and stems of cereals. Canadian Journal of Research 26 (5): 496-500.

Petit, R. J., A. El Mousadik, and O. Pons. 2008. Identifying populations for conservation on the basis of genetic markers. Conservation Biology 12 (4): 844-855.

Phillips, R. L. 2010. Mobilizing science to break yield barriers. Crop Science 50 (Supplement_1): S, 99-S-108.

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.

Pretorius, Z. A., R. P. Singh, W. W. Wagoire, and T. S. Payne. 2000. Detection of virulence to wheat stem rust resistance gene *Sr31* in *puccinia graminis*. f. sp. *tritici* in Uganda. Plant Disease 84 (2): 203-.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155 (2): 945-959.

Pritchard, J. K., M. Stephens, N. A. Rosenberg, and P. Donnelly. 2000. Association mapping in structured populations. American Journal of Human Genetics 67 (1): 170.

R Development Core Team. 2011 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.http://www.R-Project.org

Rafalski, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology 5 (2): 94-100.

Rahman, S., Christopher J. Jolly, John H. Skerritt, and Andrea Wallosheck. 1994. Cloning of a wheat 15-kDa grain softness protein (GSP). European Journal of Biochemistry 223 (3): 917-25.

Rajaram S., Mann C.E., Ortiz-Ferrara G., Mujeeb-Kazi A 1983. Adaption, stability and high yield potential of certain 1B/1R CIMMYT wheats. Proc 6th Int Wheat Genet Syrup Kyoto, Japan, 613-621.

Reif, J.C., A.E. Melchinger, X.C. Xia, M. L. Warburton, D. A. Hoisington, S. K. Vasal, G. Srinivasan, M. Bohn, and M. Frisch. 2003. Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. Crop Science 43 (4): 1275-82.

Ren, Y., Z. He, J. Li, M. Lillemo, L. Wu, B. Bai, Q. Lu, H. Zhu, G. Zhou, and J. Du. 2012. QTL mapping of adult-plant resistance to stripe rust in a population derived from common wheat cultivars naxos and shanghai 3/Catbird. Theoretical and Applied Genetics: 125 (6)1211-1221.

Reynolds, M. P., A. J. D. Pask, and D. Mullan. 2012. Physiological breeding I: Interdisciplinary approaches to improve crop adaptation. CIMMYT.

Robbins, M. D., S. C. Sim, W. Yang, A. Van Deynze, E. van der Knaap, T. Joobeur, and D. M. Francis. 2011. Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato. Journal of Experimental Botany 62 (6): 1831-45.

Roelfs, A.P., and J.W. Martens. 1988. An international system of nomenclature for *Puccinia graminis* f. sp. *tritici*. Phytopathology 78:526–533.

Roelfs, A. P., R. P. Singh, and E. E. Saari. 1992. Rust diseases of wheat: Concepts and methods of disease management. CIMMYT.

Rosenberg, N. A., T. Burke, K. Elo, M. W. Feldman, P. J. Freidlin, M. A. M. Groenen, J. Hillel, A. Mäki-Tanila, M. Tixier-Boichard, and A. Vignal. 2001. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. Genetics 159 (2): 699-713.

Rosewarne, G. M., R. P. Singh, J. Huerta-Espino, and G. J. Rebetzke. 2008. Quantitative trait loci for slow-rusting resistance in wheat to leaf rust and stripe rust identified

with multi-environment analysis. Theoretical and Applied Genetics 116 (7): 1027-1034.

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 an avocet × pastor wheat population. Theoretical and Applied Genetics 124 (7): 1283-1294.

Ross-Ibarra, J., P. L. Morrell, and Brandon S. G. 2007. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proceedings of the National Academy of Sciences 104 (Suppl 1): 8641-8648.

Roussel, V., J. Koenig, M. Beckert, and F. Balfourier. 2004. Molecular diversity in French bread wheat accessions related to temporal trends and breeding programs. Theoretical and Applied Genetics 108 (5): 920-930.

Rutkoski, J. E., E. L. Heffner, and M. E. Sorrells. 2011. Genomic selection for durable stem rust resistance in wheat. Euphytica 179 (1): 161-73.

Saintenac, C., D. Jiang, and E. D. Akhunov. 2011. Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. Genome Biol 12: R88.

Sajjad, M., S. H. Khan, and A. M. Kazi. 2012. The low down on association mapping in hexaploid wheat (*Triticum aestivum* L.). Journal of Crop Science and Biotechnology 15 (3): 147-58.

Samborski, D. J. 1985. Wheat leaf rust. The Cereal Rusts 2: 39-59.

Sanderson, M. J., and L. Hufford. 1996. Homoplasy: The recurrence of similarity in evolution. Academic Press.

Semagn, K., Å. Bjørnstad, and M. N. Ndjiondjop. 2010. An overview of molecular marker methods for plants. African Journal of Biotechnology 5 (25).

Semagn, K., C. Magorokosho, B. S. Vivek, D. Makumbi, Y. Beyene, S. Mugo, BM Prasanna, and M. L. Warburton. 2012. Molecular characterization of diverse CIMMYT maize inbred lines from eastern and southern Africa using single nucleotide polymorphic markers. BMC Genomics 13 (1): 113-123.

Shahnejat-Bushehri, Ali-Akbar, Masoud Gomarian, and Bahman Yazdi-Samadi. 2006. The high molecular weight glutenin subunit composition in old and modern bread wheats cultivated in Iran. Crop and Pasture Science 57 (10): 1109-14.

Sharma, R.C., J. Crossa, G. Velu, J. Huerta-Espino, M. Vargas, TS Payne, and RP Singh. 2012. Genetic gains for grain yield in CIMMYT spring bread wheat across international environments. Crop Science 52 (4): 1522-1533.

Singh, N. K., K. W. Shepherd, and G. B. Cornish. 1991. A simplified SDS-PAGE procedure for separating LMW subunits of glutenin. Journal of Cereal Science 14 (3): 203.

Slade, A. J., Susan I. F., Dayna L., Michael N. S., and Daniel F. 2004. A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nature Biotechnology 23 (1): 75-81.

Shen, R., J. B. Fan, D. Campbell, W. Chang, J. Chen, D. Doucet, J. Yeakley, M. Bibikova, E. Wickham Garcia, and C. McBride. 2005. High-throughput SNP genotyping on universal bead arrays. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 573 (1): 70-82.

Shewry, P. R. 2009. Wheat. Journal of Experimental Botany 60 (6): 1537-1553.

Shewry, P.R., N. G. Halford, and A.S. Tatham. 1989. The high molecular weight subunits of wheat, barley and rye: Genetics, molecular biology, chemistry and role in wheat gluten structure and functionality. Oxford Surveys of Plant Molecular and Cell Biology 6: 163-219.

Shewry, P. R., N. G. Halford, and A. S. Tatham. 1992. High molecular weight subunits of wheat glutenin. Journal of Cereal Science 15 (2): 105-20.

Sim, S. C., A. Van Deynze, K. Stoffel, D. S. Douches, D. Zarka, M. W. Ganal, R. T. Chetelat, S. F. Hutton, J. W. Scott, and R. G. Gardner. 2012. High-density SNP genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding. Plos One 7 (9): e45520.

Singh, R. P., and R. A. McIntosh. 1986. Genetics of resistance to *Puccinia graminis tritici* and *Puccinia recondita tritici* in Kenya plume wheat. Euphytica 35 (1): 245-256.

Singh, R. P. 1991. Pathogenicity variations *of Puccinia recondita* f. sp. *tritici* and *Puccinia graminis* f. sp. *tritici* in wheat-growing areas of Mexico during 1988 and 1989. Plant Disease 75 (8): 790-794.

Singh, R. P., and S. Rajaram. 1993. Genetics of adult plant resistance to stripe rust in ten spring bread wheats. Euphytica 72 (1): 1-7.

Singh, R. P., H. Ma, and S. Rajaram. 1995. Genetic analysis of resistance to scab in spring wheat cultivar frontana. Plant Disease 79 (3): 238-240.

Singh, R. P., J. Huerta-Espino, S. Rajaram, B. Barna, and Z. Kiraly. 2000. Achieving near-immunity to leaf and stripe rusts in wheat by combining slow rusting resistance genes. Paper presented at Proceedings of the 10th Cereal Rusts and Powdery Mildews Conference, Budapest, Hungary, 28 August-1 September 2000.

Singh, R. P., H. M. William, J. Huerta-Espino, and G. Rosewarne. 2004. Wheat rust in Asia: Meeting the challenges with old and new technologies. Paper presented at New directions for a diverse planet: Proceedings of the 4th International Crop Science Congress, Brisbane, Australia.

Singh, R. P., D. P. Hodson, J. Huerta-Espino, Y. Jin, P. Njau, R. Wanyera, S. A. Herrera-Foessel, and R. W. Ward. 2008. Will stem rust destroy the world's wheat crop? Advances in Agronomy 98: 271-309.

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: 465-81.

Singh, R. P. Pros and cons of utilizing major, race-specific resistance genes versus partial resistance in breeding rust resistant wheat. Paper presented at 2012 BGRI Technical Workshop, September1-4. Beijing China.

Smith, J. M., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genet Res 23 (1): 23-35.

Somers, D. J., R. Kirkpatrick, M. Moniwa, and A. Walsh. 2003. Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome 46 (3): 431-437.

Soto-Cerda, B. J., and S. Cloutier. 2012. Association mapping in plant genomes. Genetic Diversity in Plants. InTech, Rijeka: 29-54.

Stich, B., J. Möhring, H. P. Piepho, M. Heckenberger, E. S. Buckler, and A. E. Melchinger. 2008. Comparison of mixed-model approaches for association mapping. Genetics 178 (3): 1745-1754.

Storey, J. D. 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 (3): 479-498.

Stucki, D., B. Malla, S. Hostettler, T. Huna, J. Feldmann, D. Yeboah-Manu, S. Borrell, L. Fenner, I. Comas, and M. Coscollà. 2012. Two new rapid SNP-typing methods for classifying mycobacterium tuberculosis complex into the main phylogenetic lineages. PloS One 7 (7): e41253.

Szalma, S. J, E. S. Buckler, M. E. Snook, and M. D. McMullen. 2005. Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. Theoretical and Applied Genetics 110 (7): 1324-1333.

Tanksley, S. D., and S. R. McCouch. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277 (5329): 1063-1066.

Tanno, K., and G. Willcox. 2006. How fast was wild wheat domesticated? Science 311 (5769): 1886.

Tarekegne, A., and M. T. Labuschagne. 2005. Relationship between high molecular weight glutenin subunit composition and gluten quality in Ethiopian-grown bread and durum wheat cultivars and lines. Journal of Agronomy and Crop Science 191 (4): 300-7.

Tester, M., and P. Langridge. 2010. Breeding technologies to increase crop production in a changing world. Science 327 (5967): 818-22.

Thornsberry, J. M., M. M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, and E. S. Buckler. 2001. Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics 28 (3): 286-289.

Thuillet, A. C., T. Bataillon, S. Poirier, S. Santoni, and J. L. David. 2005. Estimation of long-term effective population sizes through the history of durum wheat using microsatellite data. Genetics 169 (3): 1589-99.

Tsui, K. W., and M. M. Y. Waye. 2005. A haplotype map of the human genome. Nature 437 (7063): 1299-320.

Utz, H. F., and AE Melchinger. 1996. PLABQTL: A program for composite interval mapping of QTL. J Quant Trait Loci 2 (1): 1-5.

Van De Wouw, M., T. Van Hintum, C. Kik, R. Van Treuren, and B. Visser. 2010. Genetic diversity trends in twentieth century crop cultivars: a meta-analysis. Theoretical and Applied Genetics 120 (6): 1241-52.

Van Inghelandt, D., A. E. Melchinger, C. Lebreton, and B. Stich. 2010. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theoretical and Applied Genetics 120 (7): 1289-1299.

Van Ooijen, J. W. 2006. Joinmap 4. Software for the Calculation of Genetic Linkage Maps in Experimental Populations. Kyazma B. V., Wageningen, Netherlands

Varshney, R. K., A. Graner, and M. E. Sorrells. 2005. Genic microsatellite markers in plants: Features and applications. Trends in Biotechnology 23 (1): 48-55.

Veraverbeke, W. S., and J. A. Delcour. 2002. Wheat protein composition and properties of wheat glutenin in relation to breadmaking functionality. Critical Reviews in Food Science and Nutrition 42 (3): 179-208.

Vigouroux, Y., M. McMullen, C. T. Hittinger, K. Houchins, L. Schulz, S. Kresovich, Y. Matsuoka, and J. Doebley. 2002. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proceedings of the National Academy of Sciences 99 (15): 9650-9655.

Vinod, K. K. Association mapping in crop plants. Centre for Advanced Faculty Training in Genetics and Plant breeding. Jan 20-Feb 9, 2011 Tamil Nadu Agric. University, Coimbatore, India.

Voorrips, R. E. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. Journal of Heredity 93 (1): 77-78.

Wang, T., G. Chen, Q. Zan, C. Wang, and Y. Su. 2012. AFLP genome scan to detect genetic structure and candidate loci under selection for local adaptation of the invasive weed mikania micrantha. PloS One 7 (7): e41310.

Wanyera, R., M. G. Kinyua, Y. Jin, and RP Singh. 2006. The spread of stem rust caused by *Puccinia graminis* f. sp. *tritici*, with virulence on Sr31 in wheat in eastern Africa. Plant Disease 90 (1): 113-114.

Wanyera, R., J. K. Macharia, S. M. Kilonzo, and J. W. Kamundia. 2009. Foliar fungicides to control wheat stem rust, race TTKS (*Ug99*), in Kenya. Plant Disease 93 (9): 929-932.

Weir, B., and C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370.

Weir, B. S. 1996. Genetic data analysis II. Methods for Discrete Population Genetic Data. University of Washington, USA.

Weising, K., H. Nybom, K. Wolff, and Günter K. 2005. DNA fingerprinting in plants: Principles, methods, and applications. CRC.

Wenzl, P., H. Li, J. Carling, M. Zhou, H. Raman, E. Paul, P. Hearnden, C. Maier, L. Xia, and V. Caig. 2006. A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics 7 (1): 206-227.

White, J., J. R. Law, I. MacKay, K. J. Chalmers, J. S. C. Smith, A. Kilian, and W. Powell. 2008. The genetic diversity of UK, US and Australian cultivars of *Triticum aestivum* measured by DArT markers and considered by genome. Theoretical and Applied Genetics 116 (3): 439-453.

William, M., P. Langridge, R. Trethowan, S. Dreisigacker, and J. Crouch. 2008. Genomics of wheat, the basis of our daily bread. Genomics of Tropical Crop Plants: 515-548.

Wright, S. 1984. Evolution and the genetics of populations, volume 3: Experimental results and evolutionary deductions. Vol. 3. University of Chicago press. USA.

Wright, Stephen I., and Brandon S. Gaut. 2005. Molecular population genetics and the search for adaptive evolution in plants. Molecular Biology and Evolution 22 (3): 506-519.

Xu, Li-Li, Wei Li, Yu-Ming Wei, and You-Liang Zheng. 2009. Genetic diversity of HMW glutenin subunits in diploid, tetraploid and hexaploid Triticum species. Genetic Resources and Crop Evolution 56 (3): 377-91.

Yan, L., A. Loukoianov, G. Tranquilli, M. Helguera, T. Fahima, and J. Dubcovsky. 2003. Positional cloning of wheat vernalization gene VRN1. Proc Nat'l Acad Sci USA 100: 6263-6268.

Yu, J., G. Pressoir, W. H. Briggs, I. V. Bi, M. Yamasaki, J. F. Doebley, M. D. McMullen, B. S. Gaut, D. M. Nielsen, and J. B. Holland. 2005. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38 (2): 203-208.

Yu, L. X., A. Lorenz, J. Rutkoski, R. P. Singh, S. Bhavani, J. Huerta-Espino, and M. E. Sorrells. 2011. Association mapping and gene–gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. Theoretical and Applied Genetics 123 (8): 1257-1268.

Yu, L. X., A. Morgounov, R. Wanyera, M. Keser, S. K. Singh, and M. Sorrells. 2012. Identification of *Ug99* stem rust resistance loci in winter wheat germplasm using genome-wide association analysis. Theoretical and Applied Genetics: 1-10.

Zadok, J. C., T. T. Chang, and C. F. Konzak. 1974. A decimal code for the growth stages of cereals. Weed Research 14 (6): 415-421.

Zhang, L. Y., S. Marchand, N. A. Tinker, and F. Belzile. 2009. Population structure and linkage disequilibrium in barley assessed by DArT markers. TAG Theoretical and Applied Genetics 119 (1): 43-52.

Zhang, Z., E. Ersoz, C. Q. Lai, R. J. Todhunter, H. K. Tiwari, M. A. Gore, P. J. Bradbury, J. Yu, D. K. Arnett, and J. M. Ordovas. 2010. Mixed linear model approach adapted for genome-wide association studies. Nature Genetics 42 (4): 355-360.

Zhao, K., M. J. Aranzana, S. Kim, C. Lister, C. Shindo, C. Tang, C. Toomajian, H. Zheng, C. Dean, and P. Marjoram. 2007. An arabidopsis example of association mapping in structured samples. PLoS Genetics 3 (1): e4.

Zhao, K., C. W. Tung, G. C. Eizenga, M. H. Wright, M. L. Ali, A. H. Price, G. J. Norton, M. R. Islam, A. Reynolds, and J. Mezey. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nature Communications 2: 467-476.

Zhao, X. L., T. C. Zheng, X. C. Xia, Z. H. He, D. Q. Liu, W. X. Yang, G. H. Yin, and Z. F. Li. 2008. Molecular mapping of leaf rust resistance gene LrZH84 in Chinese wheat line Zhou 8425B. Theoretical and Applied Genetics 117 (7): 1069-1075.

Zhu, C., M. Gore, E. S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. The Plant Genome 1 (1): 5-20.

Zhu, Y. L., Q. J. Song, D. L. Hyten, C. P. Van Tassell, L. K. Matukumalli, D. R. Grimm, S. M. Hyatt, E. W. Fickus, N. D. Young, and P. B. Cregan. 2003. Single-nucleotide polymorphisms in soybean. Genetics 163 (3): 1123-1134.

Appendix I A list of 297 wheat lines (excluding 7 durum lines) evaluated for diversity and population structure and GWAS related studies. Year of release for some entries is an approximation. Cluster assignment through STRUCTURE or DAPC analyses are provided in the last column

Entry	Variety/Line	Status	Where released	Year of release	Pedigree	Cluster Assignment	
						STRUCTURE	DAPC
1	184	-	-	-	-	East Africa3	East Africa3
2	2375	-	-	-	-	East Africa3	North America
3	3438	-	-	-	-	East Africa2	East Africa2
4	8952	-	-	-	-	East Africa3	Mixed Identity2
5	11383	-	-	-	-	East Africa1	Mixed Identity1
6	1010AM2(L)	Cultivar	Kenya	1969	II-50-17/KENYA-184-P	East Africa2	East Africa2
7	1010F3SEL.13C	Cultivar	Kenya	1969	II-50-17/KENYA-184-P	East Africa3	East Africa2
8	1010F3SEL.4	Cultivar	Kenya	1969	II-50-17/KENYA-184-P	East Africa3	East Africa2
9	1010F3SEL.7	Cultivar	Kenya	1969	II-50-17/KENYA-184-P	East Africa3	East Africa2
10	1012B.1.(L)	Cultivar	Kenya	1969	MENTANA/KENYA//BAGE/3/KENYA-184-P	North America	Mixed Identity2
11	1016.P.2	Cultivar	Kenya	1969	360H /3/FRONTANA // KENYA 58 / NEWTHATCH	East Africa2	East Africa2
12	1016P.1	Advanced line	Kenya	1969	KENYA-360-H/II-50-17	East Africa2	East Africa2
13	1061.K.1	Advanced line	Kenya	-	MIDA // McMURACHY / EXCHANGE /3/ RIO NEGRO	East Africa2	East Africa2
14	1061.K.4	Advanced line	Kenya	-	MIDA // McMURACHY / EXCHANGE /3/ RIO NEGRO	East Africa2	East Africa2
15	1076.D.7	Cultivar	Kenya	1969	LEE/FRONTANA//KENYA-184-P	East Africa3	East Africa2
16	1200.M	Advanced line	Kenya	-	CI 12632 /3* La PREVISION	East Africa3	East Africa3
17	2002A	-	-	-	-	East Africa3	East Africa3
18	2414-11	-	-	-	-	East Africa3	Mixed Identity2
19	291J.1.I.1	Advanced line	Kenya	-	AUSTRALIA 26 / KENYA 58	East Africa3	East Africa3
20	321BT11B1	Cultivar	Kenya	1960	AUSTRALIAN-45-C-5/KENYA-117-A	East Africa3	East Africa3
21	338AA1A2	Cultivar	Kenya	1951	AUSTRALIAN-27/KENYA-192-Q	East Africa3	East Africa3
22	688F4SEL3	Advanced line	Kenya	1969	KENYA-294M //2* MARQUIS/ AGROPYRON ELONGATUM	North America	North America
23	690F4SEL.D.1	Advanced line	Kenya	1969	KENYA-360-H//2*MARQUIS/AGROPYRON ELONGATUM	East Africa3	East Africa3
24	8-11-1-2-12	-	-	-	-	East Africa1	East Africa1
25	869B.3.B.5	-	-	-	-	North America	Mixed Identity2
26	A99ar	Cultivar	Minnesota	1982	GLENLEA / ZARAGOZA	East Africa3	North America
27	Abola	Cultivar	Ethiopia	1997	BOBWHITE/BUCKBUCK	East Africa1	East Africa1
28	AfricaMayo	Cultivar	Kenya	1960	AFRICA/MAYO-48	East Africa3	Mixed Identity1
29	AggiaSindi	-	-	-	-	East Africa3	East Africa3
30	Aguilera	Founder	Mexico	-	LV-MEX	East Africa3	East Africa3
31	Anahuac	Cultivar	Minnesota	1978	II-12300//LERMA-ROJO-64/SIETE-CERROS-66/3/NORTENO-67	East Africa1	East Africa1
31	Anahuac	Cultivar	Minnesota	1978	II-12300//LERMA-ROJO-64/SIETE-CERROS-66/3/NORTENO-67	East Africa1	East Africa1
32	Angus	Cultivar	Minnesota	1978	THATCHER/2*SUPREZA/3/FRONTANA//KENYA-58/NEWTHATCH/7/PEMBINA//FRONTANA/5*THATCHER/6/MIDA//KENYA117A/2*THATCHER/3/FRONTANA/4*THATCHER/4/MN-III-58-4/5/KENYA-58/NEWTHATCH//3*LEE	North America	North America
33	Arabe	-	-	-	-	East Africa3	East Africa3
34	B192	-	-	-	-	East Africa3	East Africa3
35	Bacup	Cultivar	Minnesota	1996	NUY-BAY/PIONEER-2375//MARSHALL,USA	East Africa3	East Africa3

Appendix I *continued*

36	**Bage**	**Cultivar**	**Brazil**	**1938**	**1068.36/LA-ESTANZUELA-2787-C**	East Africa3	North America
37	**Bailey**	**Cultivar**	**Kenya**	**1966**	**/4*THATCHER/3/THATCHER//KENYA-58/NEWTHATCH/4/THATCHER/5/FRONTANA/4*THATCHER**	East Africa3	Mixed Identity2
38	**Bale**	**-**	**-**	**-**	**-**	North America	North America
39	**Batu**	**Cultivar**	**Ethiopia**	**1984**	**GALLO/CUCKOO//KAVKAZ/SUPER X**	East Africa3	East Africa3
40	**Beacon-Ken**	**Cultivar**	**Kenya**	**1968**	**FRONTANA / KENYA 58 // NEWTHATCH /3/3* BONZA**	East Africa1	East Africa1
41	**Beltista**	**-**	**-**	**-**	**-**	East Africa2	East Africa2
42	**BF236C1L**	**Advanced line**	**Kenya**	**-**	**-**	East Africa3	East Africa3
43	**Bluebird**	**Cultivar**	**Mexico**	**1969**	**CIANO-67(SIB)//SONORA-64/KLEIN-RENDIDOR/3/II-8156**	East Africa1	Mixed Identity1
44	**Bobicho**	**Cultivar**	**Ethiopia**	**2002**	**PEREGRINE/PF70354/KALYANSONA/BLUEBIRD/ALONDRA/3/MARINGA**	East Africa3	East Africa3
45	**Bobin**	**Cultivar**	**Australia**	**1925**	**THEW/STEINWEDEL**	East Africa3	East Africa3
46	**Bobwhite**	**Cultivar**	**Mexico**	**1977**	**AVRORA//KALYANSONA/BLUEBIRD/3/(SIB)WOODPECKER**	East Africa1	East Africa1
47	**Bonanza**	**Cultivar**	**Kansas**	**1969**	**PITIC-62/(SIB)CHRIS//SONORA-64**	East Africa1	East Africa1
48	**Bonito**	**Cultivar**	**Kenya**	**1973**	**BONZA 'S'/4/ FRONTANA // THATCHER/ TRIGO GLUTINOSA/3/ MENTANA/5/2* LERMA// SELKIRK/ LERMA /3/ WISCONSIN 245/4/ MENKEMEN 626**	East Africa2	East Africa2
49	**Bonny**	**Cultivar**	**Kenya**	**1966**	**YAQUI-53/2*BONZA**	North America	East Africa2
50	**Bonza**	**Cultivar**	**Colombia**	**1970**	**YAQUI-50/KENTANA-48**	East Africa2	East Africa2
51	**Bonza63**	**Founder**	**Colombia**	**1963**	**RIO-NEGRO/2*BONZA-55**	East Africa2	East Africa2
52	**Borah**	**Cultivar**	**Minnesota**	**1974**	**NO-58/THATCHER//THATCHER/KENYA-FARMER/3/MN-III-58-1//FRONTANA/3*THATCHER**	East Africa3	North America
53	**Bounty**	**Cultivar**	**Kenya**	**1966**	**TIMSTEIN/2*KENYA//BONZA**	East Africa3	East Africa2
54	**Breadwheat23**	**-**	**-**	**-**	**-**	East Africa1	East Africa1
55	**Brewster**	**Cultivar**	**Kenya**	**1966**	**FRONTANA/4*THATCHER/3/THATCHER//KENYA-58/NEWTHATCH/4/THATCHER/5/FRONTANA/4*THATCHER**	North America	North America
56	**BTCBarce**	**-**	**-**	**-**	**-**	East Africa3	East Africa3
57	**BW21**	**-**	**Founder**	**-**	**-**	East Africa1	East Africa1
58	**Cajeme71**	**Cultivar**	**Kenya**	**1971**	**CIANO SIB /3/ SONORA 64 / KLEIN RENDIDOR // SIETE CERROS 66**	East Africa1	East Africa1
59	**Catcher**	**Cultivar**	**Kenya**	**1963**	**THATCHER/SANTA-CATALINA//FROCOR**	East Africa2	East Africa2
60	**Ceres**	**Cultivar**	**Minnesota**	**1924**	**MARQUIS/KOTA**	North America	North America
61	**Chinesespring**	**-**	**-**	**-**	**-**	East Africa3	East Africa3
62	**Chozi**	**Cultivar**	**Kenya**	**1998**	**F-12-7/COCORAQUE 75//GENARO 81**	East Africa1	East Africa1
63	**Chris**	**Cultivar**	**Minnesota**	**1965**	**FRONTANA/3*THATCHER/3/KENYA-58/NEWTHATCH//2*THATCHER**	North America	North America
64	**CI14393**	**Cultivar**	**Ethiopia**	**1975**	**FROCOR*2/4/COMETA/3/ NEWTHATCH// MENTANA/ MENKEMEN**	East Africa2	East Africa2
65	**CI2401**	**Cultivar**	**Mexico**	**1967**	**PITIC-62/(SIB)CHRIS//SONORA-64**	East Africa3	East Africa3
66	**CianoF67**	**Cultivar**	**Minnesota**	**1963**	**KLEIN-TITAN/3*THATCHER/3/II-44-29/2*THATCHER**	East Africa1	East Africa1
67	**Cocoraque75**	**Founder**	**Mexico**	**1975**	**II-12300//LERMA-ROJO-64/II-8156/3/NORTENO-67**	East Africa1	East Africa1
68	**Crim**	**Cultivar**	**Minnesota**	**1963**	**KLEIN-TITAN/3*THATCHER/3/II-44-29/2*THATCHER**	East Africa3	North America
69	**Dashen**	**Cultivar**	**Ethiopia**	**1984**	**KAVKAZ/BUHO//KALYANSONA/BLUEBIRD**	East Africa1	East Africa1
70	**Dereseglen**	**Cultivar**	**Ethiopia**	**1974**	**CI8154//2*FEDERATION**	East Africa2	East Africa2
71	**Digelu**	**Cultivar**	**Ethiopia**	**-**	**SHANGHAI #7/KAUZ**	East Africa3	Mixed Identity2
72	**Diredawa**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa2	East Africa2
73	**Dodota**	**Cultivar**	**Ethiopia**	**2001**	**BLUEJAY/COCORAQUE F 75//PARULA/BOBWHITE**	East Africa1	East Africa1

Appendix I *continued*

74	**Duma**	**Cultivar**	**Kenya**	**1993**	**AURORA/UP301//GALLO/SUPER X/3/PEWEE/4/MAIPO/MAYA 74//PEWEE**	East Africa1	East Africa1
75	**Dure**	**Cultivar**	**Ethiopia**	**2001**	**-**	East Africa3	East Africa2
76	**EgyptianNa95**	**Cultivar**	**Kenya**	**-**	**KENYA-U / KENYA 9MIA-3**	East Africa3	East Africa3
77	**Enkoy**	**Cultivar**	**Ethiopia**	**1974**	**HEBRAND SEL/WISCONSIN 245/SUPRESA/3/2*FROCOR//FRONTANA/YAQUI/4/AGUILERA**	East Africa2	East Africa2
78	**EnkoyU**	**-**	**-**	**-**	**-**	East Africa2	East Africa2
79	**Equator**	**Cultivar**	**Kenya**	**1920**	**AUSTRALIAN-VARIETY**	East Africa3	East Africa3
80	**Equator1**	**Cultivar**	**Kenya**	**1960**	**(S)EQUATOR**	East Africa3	East Africa3
81	**Era**	**Cultivar**	**Minnesota**	**1970**	**II-55-10/4/PEMBINA/II-52-329/3/II-53-388/III-58-4//II-53-546**	North America	North America
82	**ET-12-D4**	**Cultivar**	**Ethiopia**	**1981**	**MAMBA/UQ105**	East Africa1	East Africa1
83	**ET-13A2**	**Cultivar**	**Ethiopia**	**1981**	**ENKOY/UQ105**	East Africa2	East Africa2
84	**Exchange**	**Cultivar**	**Indiana**	**1963**	**WARDEN,GBR/HYBRID-ENGLISH**	East Africa3	East Africa3
85	**Fanfare**	**Cultivar**	**Kenya**	**1964**	**FROCOR//FRONTANA/YAQUI**	East Africa2	East Africa2
86	**Federation**	**Cultivar**	**Australia**	**1901**	**YANDILLA/PURPLE-STRAW[113];PURPLE-STRAW/YANDILLA**	East Africa3	East Africa3
87	**Fletcher**	**Cultivar**	**Minnesota**	**1970**	**II-55-10/4/PEMBINA/II-52-329/3/II-53-388/III-58-4//II-53-546**	East Africa3	East Africa3
88	**FLIKenya9**	**Advanced line**	**Kenya**	**-**	**-**	North America	North America
89	**Florence**	**Cultivar**	**Canada**	**1892**	**ALPHA,1888/HARD-RED-CALCUTTA**	North America	North America
90	**Frocor2328**	**Advanced line**	**Colombia**	**1951**	**FRONTANA // C.O. / C.R.**	East Africa2	East Africa2
91	**Frontana**	**Cultivar**	**Brazil**	**1930**	**FRONTEIRA/MENTANA**	East Africa2	East Africa2
92	**Fronthatch**	**Cultivar**	**Kenya**	**1963**	**FRONTANA / KENYA58 // NEWTHATCH**	East Africa2	East Africa2
93	**Fury**	**Cultivar**	**Kenya**	**1964**	**FROCOR/MENTANA/KENYA-2/MCMURACHY/YAQUI-50**	East Africa2	East Africa2
94	**Gabo**	**Cultivar**	**Australia**	**1955**	**TIMSTEIN/KENYA-58//GABO**	East Africa3	East Africa3
95	**Gabrino**	**Cultivar**	**Kenya**	**1963**	**KENTANA/RIO-NEGRO//GABO-54**	East Africa3	East Africa2
96	**Galema**	**Cultivar**	**Ethiopia**	**1995**	**4777*2//FKN/GABO-AUS/3/PAVON F 76**	East Africa1	East Africa1
97	**Gamtoos**	**Cultivar**	**South Africa**	**-**	**KAVKAZ/(SIB)BUHO//KALYANSONA/BLUEBIRD**	East Africa1	East Africa1
98	**Gandum-i-Fasai**	**Landrace**	**Iran**	**-**	**LV-IRN**	East Africa3	East Africa3
99	**Gara**	**Cultivar**	**Ethiopia**	**1984**	**AVRORA//KALYANSONA/BLUEBIRD/3/(SIB)WOODPECKER**	East Africa1	East Africa1
100	**Gem**	**Cultivar**	**Kenya**	**1964**	**BT908 / FRONTANA // CAJEME 54**	East Africa3	East Africa2
101	**Giza155**	**Cultivar**	**Egypt**	**1968**	**GIZA-144/3/MIDA,USA/CADET,USA//2*HINDI-62**	East Africa3	East Africa3
102	**Goblet**	**Cultivar**	**Kenya**	**1967**	**GABO-54/LERMA-52//GABO/3/KENYA/GENERAL-URQUIZA**	East Africa3	Mixed Identity2
103	**Gradenero**	**Cultivar**	**Mexico**	**-**	**-**	East Africa3	East Africa3
104	**H441**	**Advanced line**	**Kenya**	**-**	**REWARD/ CI 12632**	North America	North America
105	**Halt**	**Cultivar**	**Colorado**	**-**	**SUMNER/CO-820026//PI-372129/3/TAM-107**	East Africa3	East Africa3
106	**HAR1685**	**Cultivar**	**Ethiopia**	**1994**	**NORD DEPREZ/VG9144//KALYANSONA/BLUEBIRD/3/YACO/4/VEERY #5**	East Africa1	East Africa1
107	**HAR-407**	**Cultivar**	**Ethiopia**	**1987**	**KAVKAZ/BUHO//KALYANSONA/BLUEBIRD**	East Africa1	East Africa1
108	**Hatcher**	**Cultivar**	**Colorado**	**2004**	**YUMA/T-57//TAM-200/3/4*YUMA/4/KS-91-H-184/VISTA**	East Africa3	East Africa3
109	**Hawi**	**Cultivar**	**Ethiopia**	**2000**	**CHILERO/PARULA**	East Africa1	East Africa3
110	**Hard-Federation**	**Founder**	**Australia**	**-**	**S)FEDERATION**	East Africa3	East Africa1
111	**HJ98**	**Cultivar**	**Minnesota**	**1998**	**W-8814/NORAK**	East Africa1	North America
112	**Hope**	**Cultivar**	**South Dakota**	**1927**	**YAROSLAV-EMMER/MARQUIS**	North America	North America

Appendix I *continued*

113	**HRS55**	**Founder**	**-**	**-**	**-**	East Africa3	East Africa3
114	**II-50-17**	**Cultivar**	**Minnesota**	**1967**	**FRONTANA//KENYA-58/NEWTHATCH**	East Africa2	East Africa2
115	**Impala**	**Cultivar**	**South Africa**	**1954**	**KOALISIE/HOPE**	East Africa3	East Africa3
116	**Inia66**	**Cultivar**	**Ethiopia**	**1971**	**LERMA ROJO 64/SONORA 64**	East Africa2	East Africa2
117	**Jiru**	**Cultivar**	**Ethiopia**	**2006**	**-**	East Africa1	East Africa1
118	**Justin**	**Cultivar**	**Minnesota**	**1962**	**CONLEY/ND-40-2**	North America	North America
119	**K-360-H**	**Cultivar**	**Kenya**	**-**	**KENYA 294M7C6C / KENYA 184P2A1E**	East Africa3	East Africa2
120	**K-6106-8**	**Cultivar**	**Ethiopia**	**1977**	**CI-8154/2*FROCOR/3/2*GABO-54/36896//II-53-526**	East Africa2	East Africa3
121	**K6290Bulk**	**Cultivar**	**Ethiopia**	**1977**	**AFRICA MAYO/*ROMANY**	East Africa2	East Africa2
122	**K6295-4A**	**Cultivar**	**Ethiopia**	**1980**	**ROMANY/GABO-GAMENYA**	East Africa2	East Africa2
123	**Kalyanosona**	**Cultivar**	**India**	**1967**	**FRONTANA // KENYA 58/ NEWTHATCH/3/NORIN 10 /BREVOR/4/ GABO 55**	East Africa1	East Africa1
124	**Katar**	**Cultivar**	**Ethiopia**	**1999**	**COOK/VEE''S''//DOVE''S''/SERI/3/BJY''S''**	East Africa1	East Africa1
125	**KBG-01**	**Cultivar**	**Ethiopia**	**-**	**(300/SM+501M)/HAR 1709**	East Africa1	East Africa1
126	**Kentana48**	**Cultivar**	**Mexico**	**1948**	**KENYA-C-9906/MENTANA**	North America	North America
127	**KentanaYaqui**	**Cultivar**	**Kenya**	**1960**	**KENTANA-48/YAQUI-48**	East Africa2	East Africa2
128	**Kenya**	**Advanced line**	**Kenya**	**1929**	**K-25550;AFRC-6510**	East Africa3	East Africa3
129	**Kenya 291 J.1.I.1**	**Advanced line**	**Kenya**	**-**	**AUSTRALIAN-26-A/KENYA-58**	East Africa3	East Africa3
130	**Kenya-1**	**Cultivar**	**Kenya**	**1961**	**S)LV-KEN**	East Africa3	East Africa3
131	**Kenya-117A**	**Advanced line**	**Kenya**	**-**	**MARQUIS / AGUILERA 8**	East Africa3	East Africa3
132	**Kenya117C**	**Advanced line**	**Kenya**	**-**	**MARQUIS / AGUILERA 8**	North America	North America
133	**Kenya-122**	**Advanced line**	**Kenya**	**-**	**MARQUIS/AGUILERA 8**	North America	North America
134	**Kenya-131**	**Cultivar**	**Kenya**	**-**	**FLORENCE / AGUILERA 8**	East Africa3	East Africa3
135	**Kenya155**	**Advanced line**	**Kenya**	**-**	**-**	East Africa3	East Africa3
136	**Kenya-184-P**	**Cultivar**	**Kenya**	**1951**	**RELIANCE/KENYA-73-D**	North America	East Africa3
137	**Kenya-294-B-2A-3**	**Advanced line**	**Kenya**	**-**	**AUSTRALIAN-26-A/KENYA-117-A**	East Africa3	Mixed Identity1
138	**Kenya-318.O.3B.2**	**Advanced line**	**Kenya**	**-**	**KENYA-112/CERES**	North America	North America
139	**Kenya-318-AJ-4A-1**	**Advanced line**	**Kenya**	**-**	**KENYA-112/CERES**	North America	Mixed Identity2
140	**Kenya-358-AC**	**Advanced line**	**Kenya**	**-**	**KENYA-184-P/KENYA-294-B**	East Africa3	East Africa3
141	**Kenya-362-B-1A**	**Advanced line**	**Kenya**	**1956**	**EQUATOR / KENYA 294.M . 7.C.6 .C.**	East Africa3	East Africa3
142	**Kenya-5**	**Advanced line**	**Kenya**	**1960**	**S)LV-KEN**	East Africa3	East Africa3
143	**Kenya501**	**Advanced line**	**Kenya**	**-**	**-**	East Africa3	East Africa3
144	**Kenya-58**	**Advanced line**	**Kenya**	**-**	**RED EGYPTIAN / KENYA BF3B10V1**	East Africa3	East Africa3
145	**Kenya6297-2**	**-**	**-**	**-**	**-**	East Africa2	East Africa2
146	**Kenya-6297-2**	**Advanced line**	**Kenya**	**-**	**ROMANY // WISCONSIN 245 / SIPRESA 51 /3/2* FROCOR / FRONTANA / YAQUI /4/ ANHINGA**	East Africa2	East Africa2
147	**Kenya6820**	**Cultivar**	**Kenya**	**-**	**KENYA 4500-35 / KENYA SWARA**	East Africa2	East Africa2
148	**Kenya7**	**Advanced line**	**Kenya**	**-**	**-**	East Africa3	East Africa3
149	**Kenya8**	**Advanced line**	**Kenya**	**-**	**-**	East Africa3	East Africa3
150	**KenyaB-256-G**	**Cultivar**	**Kenya**	**-**	**KENYA-U/KENYA-9-M-1-A-3**	East Africa3	East Africa3
151	**KenyaCheetah**	**Cultivar**	**Kenya**	**-**	**WARIGO/STERLING**	East Africa3	East Africa3

Appendix I *continued*

152	**KenyaChiriku**	**Cultivar**	**Kenya**	**1989**	**KTB/(SIB)CARPINTERO**	East Africa2	East Africa2
153	**KenyaCivet**	**Cultivar**	**Kenya**	**1966**	**CI 12632 /3* KENYA 354**	East Africa3	East Africa3
154	**KenyaFahari**	**Cultivar**	**Kenya**	**1977**	**TOBARI-66/3/SRPC-527-67//CI-8154/2*FROCOR**	East Africa1	East Africa1
155	**KenyaFarmer.**	**Cultivar**	**Kenya**	**1954**	**AUSTRALIAN-27/KENYA-192**	East Africa3	East Africa3
156	**KenyaFL.1.158**	**Advanced line**	**Kenya**	**-**	**-**	East Africa3	East Africa3
157	**KenyaGovernor**	**Cultivar**	**Kenya**	**1925**	**-**	East Africa3	East Africa3
158	**KenyaGrange**	**Cultivar**	**Kenya**	**1966**	**KENYA-360-F/GRANADERO-KLEIN**	East Africa3	East Africa3
159	**KenyaHeroe**	**Cultivar**	**Kenya**	**1998**	**MBUNI/SRPC-64//YRPC-1**	East Africa3	East Africa1
160	**KenyaHunter**	**Cultivar**	**Kenya**	**1964**	**EQUATOR II / KENYA 310.0 . 33.2 // HOPE / TIMSTEIN /3/ REGENT**	East Africa3	East Africa3
161	**Kenya-Ibis**	**Cultivar**	**Kenya**	**2008**	**KWALE/DUMA**	East Africa1	East Africa1
162	**KenyaJay**	**Cultivar**	**Kenya**	**1966**	**EQUATOR/KENYA-318**	East Africa3	East Africa3
163	**Kenya-Kanga**	**Cultivar**	**Kenya**	**1977**	**MENCO/4/WISCONSIN-245/SUPREMO 51/3/2*FEDERATION /FRONTANA /YAMESEK**	East Africa2	East Africa2
164	**KenyaKifaru**	**Advanced line**	**Kenya**	**1977**	**TOBARI-66*3/3/WISCONSIN-245//CI-8154/2*FROCOR**	East Africa1	East Africa1
165	**KenyaKongoni**	**Cultivar**	**Kenya**	**1981**	**CI-8154/2*FROCOR//3*ROMANY/4/WISCONSIN-245/II-50-17/CI-8154//2* FROCOR/3/TOBARI-66**	East Africa2	East Africa2
166	**KenyaKudu**	**Cultivar**	**Kenya**	**1966**	**KENYA-131/KENYA-184-P**	East Africa3	East Africa3
167	**KenyaLeopard**	**Cultivar**	**Kenya**	**1966**	**LAGAEDINHI /3* KENYA 381P // CI 12632 /3* KENYA 354P**	East Africa3	East Africa3
168	**KenyaMamba**	**Advanced line**	**Kenya**	**1962**	**AFRICA-MAYO-48/4/WISCONSIN-245/SUPREMO-51/3/2* FEDERATION/ FRONTANA/YAMESEK**	East Africa3	East Africa2
169	**Kenya-Mbweha**	**Cultivar**	**Kenya**	**1974**	**CI-8154/2*FROCOR/3/2*GABO-54/36896//II-53-526**	East Africa3	East Africa2
170	**KenyaNgiri**	**Cultivar**	**Kenya**	**1979**	**CI-8154/2*FROCOR//WHEAT-RYE-TRANSLOCATION/SANTA-CATALINA/3/MANITOU/4/2*TOBARI-66**	East Africa1	East Africa1
171	**KenyaNungu**	**Cultivar**	**Kenya**	**1975**	**WISCONSIN-245/II-50-17//CI-8154/2*FEDERATION/3/2*TOBARI-66**	East Africa1	East Africa1
172	**KenyaNyangumi**	**Cultivar**	**Kenya**	**1979**	**TEZANOS-PINTOS-PRECOZ//SELKIRK-ENANO*6/LERMA-ROJO-64/3/AFRICA-MAYO-48/4/KENYA-SWARA/K-4500-6**	East Africa1	East Africa1
173	**Kenya-Nyati**	**Cultivar**	**Kenya**	**1973**	**AFRICA-MAYO/2*ROMANY**	East Africa3	East Africa2
174	**KenyaNyoka**	**Cultivar**	**Kenya**	**1975**	**CI-8154/2*FEDERATION//3*ROMANY**	East Africa2	East Africa2
175	**Kenya-Paa**	**Cultivar**	**Kenya**	**1981**	**KAVKAZ/3/CIANO-67/CHRIS//OLESENS-DWARF**	East Africa3	North America
176	**Kenyapage**	**Cultivar**	**Kenya**	**1963**	**MENTANA/KENYA-58//BAGE/3/KENYA-184-P**	East Africa3	East Africa3
177	**KenyaPaka**	**Cultivar**	**Kenya**	**1975**	**WISCONSIN-245/II-50-17//CI-8154/2*TOBARI-66**	East Africa1	East Africa1
178	**KenyaPloughman**	**Cultivar**	**Kenya**	**1950**	**CERES/KENYA-112-E-8-L-5**	North America	North America
179	**KenyaPlume**	**Cultivar**	**Kenya**	**1965**	**MIDA/MCMURACHY//EXCHANGE/3/KENYA-184-P**	East Africa2	East Africa2
180	**KenyaPopo**	**Cultivar**	**Kenya**	**1982**	**KLEIN-ATLAS/TOBARI-66//CENTRIFEN/3/BLUEBIRD/4/KENYA-FAHARI**	East Africa1	East Africa1
181	**KenyaStandard**	**Cultivar**	**Kenya**	**1930**	**-**	East Africa3	East Africa3
182	**KenyaSungura**	**Cultivar**	**Kenya**	**1969**	**ID 1877/MORRIS**	North America	North America
183	**KenyaSwara**	**Cultivar**	**Kenya**	**1972**	**CI-8154/2*FROCOR/3/TIMSTEIN/2*KENYA//Y-59.2.B**	East Africa1	East Africa1
184	**KenyaTembo**	**Cultivar**	**Kenya**	**1975**	**WISCONSIN-245/II-50-17//CI-8154/2*TOBARI-66**	East Africa1	East Africa1
185	**Kenya-Tumbili**	**Cultivar**	**Kenya**	**1984**	**KTB/GIZA-155//NADADORES-63/T-238-1-5-8-17-10/3/KLEIN-ATLAS /TOBARI-66//CENTRIFEN/BLUEBIRD**	East Africa1	East Africa1
186	**KenyaYombi**	**Cultivar**	**Kenya**	**1998**	**MBUNI/SRPC-64//YRPC-5**	East Africa3	East Africa1
187	**KenyaZabadi**	**Cultivar**	**Kenya**	**1979**	**CORRECAMINOS/INIA-67//K-4500-2/3/KENYA-SWARA//TOBARI-66/CIANO-67**	East Africa1	East Africa1
188	**Kitt**	**Cultivar**	**Minnesota**	**1975**	**II-55-14/II-60-15**	North America	North America

Appendix I *continued*

189	**KKBB**	**Cultivar**	**Ethiopia**	**1982**	**KAVKAZ/KALYANSONA/BLUEBIRD**	East Africa1	East Africa1
190	**Kruger**	**Cultivar**	**Kenya**	**1987**	**KAVKAZ/3/SONORA 64/CIANO F 67//INIA F 66/4/MAYA 74//BLUEBIRD/INIA F 66**	East Africa3	Mixed Identity1
191	**Kubsa**	**Cultivar**	**Ethiopia**	**1994**	**NORD-DESPREZ/VG-9144//KALYANSONA/BLUEBIRD/3/YACO/4/VEERY**	East Africa1	East Africa1
192	**Kwale**	**Cultivar**	**Kenya**	**1987**	**KAVKAZ/TANORI-71/3/MAYA-74(SIB)//BLUEBIRD/INIA-66**	East Africa1	East Africa1
193	**KZK-1177**	**Founder**	**-**	**-**	**-**	East Africa3	East Africa3
194	**Lahota**	**-**	**-**	**-**	**-**	East Africa3	Tetraploid
195	**Laketch**	**Cultivar**	**Ethiopia**	**1970**	**PENJAMO-62/GABO 55**	East Africa1	East Africa1
196	**Lee**	**Cultivar**	**Minnesota**	**1950**	**HOPE/TIMSTEIN**	East Africa3	Mixed Identity2
197	**Lenana**	**Cultivar**	**Kenya**	**1963**	**YAQUI- 48 / KENTANA- 48**	East Africa3	East Africa2
198	**Lermarojo**	**Cultivar**	**Mexico**	**1955**	**LERMA-50/YAQUI-48//MARIA-ESCOBAR*2/SUPREMO-211**	East Africa2	East Africa2
199	**Maddawalabu**	**Cultivar**	**Ethiopia**	**2000**	**TANORI F 71/3/Fn/Th/Nar 59 *2/4/Bol'S'**	East Africa1	East Africa1
200	**Mahon**	**Landrace**	**Ethiopia**	**-**	**BLES-DE-MAHON**	East Africa3	East Africa3
201	**Marquillo**	**Cultivar**	**Minnesota**	**1926**	**MARQUIS/(TR.DR)IUMILLO**	North America	North America
202	**Marquis**	**Cultivar**	**Minnesota**	**1907**	**HARD-RED-CALCUTTA**	North America	North America
203	**Marshall**	**Cultivar**	**Minnesota**	**1982**	**ERA/WALDRON**	North America	North America
204	**Mbega**	**Cultivar**	**Kenya**	**1993**	**BONANZA/YECORA-70/3/F-35-75//KALYANSONA/BLUEBIRD**	East Africa1	East Africa1
205	**Mbuni**	**Cultivar**	**Kenya**	**1987**	**ZARAGOZA-75/3/LD-357-E/THATCHER//GALLO**	East Africa1	East Africa1
206	**McMurachy**	**Cultivar**	**Canada**	**1958**	**RC-1373/RED-EGYPTIAN**	East Africa3	East Africa3
207	**Mcvey**	**Cultivar**	**Minnesota**	**1999**	**NING-8331/MN-87029//MN-89068**	East Africa3	North America
208	**Megal**	**Cultivar**	**Ethiopia**	**1997**	**F371/TRM//BUC''S''/3/LIRA''S''**	East Africa1	East Africa1
209	**Menco**	**Cultivar**	**Kenya**	**1963**	**MENTANA / KENYA // FRONTANA / CINCO**	East Africa2	East Africa2
210	**Mentana**	**Cultivar**	**Mexico**	**1913**	**RIETI/WILHELMINA//AKAKOMUGI**	East Africa2	East Africa2
211	**Mentor**	**Cultivar**	**Kenya**	**1967**	**MENGAVI/3/SPICA/KODA//GABO**	East Africa3	Mixed Identity1
212	**Meraro**	**Cultivar**	**Ethiopia**	**2005**	**-**	East Africa3	East Africa1
213	**MG07759**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	Mixed Identity1
214	**MG07761**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	Tetraploid
215	**MG07762**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	Tetraploid
216	**MG07766**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	Tetraploid
217	**MG07768**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	East Africa3
218	**MG07780**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	Tetraploid
219	**MG07782**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	Tetraploid
220	**MG07793**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	East Africa3
221	**MG07795**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	East Africa3
222	**Mida**	**Cultivar**	**Minnesota**	**1944**	**MERCURY/RL-625**	North America	North America
223	**Minnpro**	**Cultivar**	**Minnesota**	**1990**	**MN-72299/MN-74115**	East Africa3	North America

Appendix I *continued*

224	**Mitike**	**Cultivar**	**Ethiopia**	**1993**	**BOBWHITE/REICHENBACHII**	East Africa2	East Africa2
225	**MN72131**	**-**	**Minnesota**	**-**	**-**	East Africa3	North America
226	**Moran**	**Cultivar**	**Minnesota**	**1967**	**KENYA-58/THATCHER//THATCHER/KENYA-FARMER THATCHER//KENYA-117**	North America	North America
227	**Morris**	**Cultivar**	**Kenya**	**-**	**A/MIDA/3/FRONTANA/4*THATCHER/4/THATCHER/5/FRONTANA /4*THATCHER**	North America	North America
228	**Napayo**	**Cultivar**	**Minnesota**	**1972**	**MANITOU*2/RL-4124.1**	North America	North America
229	**Newthatch**	**Cultivar**	**Minnesota**	**1944**	**HOPE/THATCHER//2*THATCHER**	North America	North America
230	**Ngamia**	**Cultivar**	**Kenya**	**-**	**BUCKY/MAYA-74/4/BLUEBIRD//HD-832/OLESENS DWARF/3/CIANO 67 /PENJAMO 62**	East Africa3	East Africa1
231	**NjoroBwII**	**Cultivar**	**Kenya**	**2007**	**IAS-58/4/KALYANSONA/BLUEBIRD//CAJEME-F-71/3/ALONDRA/5/BOBWHITE**	East Africa1	East Africa1
232	**Norm**	**Cultivar**	**Minnesota**	**1992**	**MN-73167/MN-81070**	North America	North America
233	**NP761**	**Founder**	**India**	**1941**	**PUSA-52/PUSA-165**	East Africa3	East Africa3
234	**Olescensdwarf**	**Cultivar**	**Zimbabwe**	**1970**	**NORIN-10/MARA,ITA//ANGOLAN-X-2-50**	East Africa3	East Africa3
235	**P.WalkerMunro**	**Advanced line**	**Kenya**	**-**	**-**	East Africa2	East Africa2
236	**Pasa**	**Cultivar**	**Kenya**	**1989**	**BUCK BUCK/CHAT**	East Africa1	East Africa1
237	**Pavon76**	**Cultivar**	**Ethiopia**	**1982**	**VICAM 571//CIANO F67/SIETE CERROS T**	East Africa1	East Africa1
238	**PBW343**	**Cultivar**	**India**	**1995**	**ND/VG-1944//KALYANSONA//BLUEBIRD/3/YACO(SIB)/4/VEERY-5**	East Africa1	East Africa1
239	**Penjamo62**	**Cultivar**	**Mexico**	**-**	**FKN/NORIN 10 BREVOR**	East Africa3	Mixed Identity2
240	**Pewter**	**Cultivar**	**Kenya**	**1964**	**PW-327,USA/5*THATCHER**	North America	North America
241	**Pitcher**	**Founder**	**-**	**-**	**-**	East Africa2	East Africa2
242	**Polk**	**Cultivar**	**Minnesota**	**1968**	**THATCHER / SUPREZA /3/ KENYA 58 / NEWTHATCH // FRONTANA**	North America	North America
243	**Primex**	**Cultivar**	**Kenya**	**1969**	**90875,MEX**	North America	East Africa2
244	**PWThatcher**	**Cultivar**	**Kenya**	**-**	**THATCHER/AGENT**	North America	North America
245	**Quamy**	**-**	**-**	**-**	**-**	East Africa3	Tetraploid
246	**R64**	**Cultivar**	**Kenya**	**1953**	**DIENTE DE CAMELLO/CERES**	North America	North America
247	**RedEgyptian**	**Cultivar**	**Mexico**	**-**	**RED EGYPTIAN**	East Africa3	East Africa3
248	**Redman**	**Cultivar**	**Minnesota**	**1946**	**REGENT,CAN/CANUS**	North America	North America
249	**Regent**	**Cultivar**	**Kenya**	**1939**	**H44/REWARD**	North America	North America
250	**Reliance**	**Cultivar**	**Kenya**	**1933**	**KANRED/MARQUIS**	North America	North America
251	**Reliance261M**	**Advanced line**	**Kenya**	**-**	**RELIANCE / KENYA 68**	East Africa3	East Africa3
252	**Reward**	**Cultivar**	**Canada**	**1928**	**MARQUIS/PRELUDE**	North America	North America
253	**RFN**	**Cultivar**	**Kenya**	**1949**	**(S)SABANERO**	East Africa3	East Africa3
254	**Rhodesiansabanero**	**Advanced line**	**Kenya**	**-**	**MARQUIS / AGUILERA 8**	East Africa3	East Africa3
255	**Ripper**	**Founder**	**-**	**-**	**-**	East Africa3	East Africa3
256	**RL1377**	**Advanced line**	**Kenya**	**-**	**MARQUIS / AGUILERA 8**	East Africa1	East Africa1
257	**Romany**	**Cultivar**	**Ethiopia**	**1970**	**COLOTANA 261-51 / YAKTANA 54A**	East Africa2	East Africa2
258	**Sabanero**	**Advanced line**	**Kenya**	**1934**	**-**	East Africa3	East Africa3
259	**Salmayo**	**Cultivar**	**Kenya**	**1963**	**SALLES/MCMURACHY//MAYO-48**	East Africa2	East Africa2
260	**SantaElena**	**Cultivar**	**Kenya**	**1969**	**SANTA-CATALINA-6/THATCHER//FROCOR**	East Africa2	East Africa2
261	**Selkirk**	**Cultivar**	**Minnesota**	**1953**	**MCMURACHY/EXCHANGE//3*REDMAN**	North America	North America

Appendix I *continued*

262	**Seri**	**Cultivar**	**Minnesota**	**1982**	**KAVKAZ/(SIB)BUHO//KALYANSONA/BLUEBIRD**	East Africa1	East Africa1
263	**Shield**	**Cultivar**	**Minnesota**	**1986**	**COTEAU(CI-17749,s)/(CI-17801,w)DAWN**	East Africa3	North America
264	**Shina**	**Cultivar**	**Ethiopia**	**1998**	**GOLDEN-VALLEY(GOV)/AZTECA-67//MUSALA/3/R-37/GHL-121//KALYANSONA/BLUEBIRD/4/ANI**	East Africa1	East Africa1
265	**Simba**	**Cultivar**	**Ethiopia**	**2000**	**PARULA/VEERY #6//MYNA/VULTURE**	East Africa1	East Africa1
266	**Sindi1**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	East Africa3
267	**Sindi2**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	East Africa3
268	**Sindi4**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	East Africa3
269	**Sindi5**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	East Africa3
270	**Sindi6**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa1	East Africa1
271	**Sindi7**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	East Africa3
272	**Sindi8**	**Landrace**	**Ethiopia**	**-**	**-**	East Africa3	East Africa3
273	**Sirbo**	**Cultivar**	**Ethiopia**	**2001**	**VS73.600/MRL/3/BOBWHITE//YECORA F 70/TRIFON**	East Africa1	East Africa1
274	**Sofumar**	**Cultivar**	**Ethiopia**	**2000**	**4777(2)//FKN/GB/3/PAVON F 76**	East Africa1	East Africa1
275	**Sonora**	**-**	**Mexico**	**-**	**LV-Sonora**	East Africa2	East Africa2
276	**Sonora63**	**Cultivar**	**Ethiopia**	**1975**	**YAKTANA-54//NORIN-10/BREVOR/3/2*YAQUI-54**	East Africa1	Mixed Identity2
277	**Supremo**	**Cultivar**	**Mexico**	**1948**	**SURPRESA//HOPE/MEDITERRANEAN**	East Africa3	East Africa3
278	**Tama**	**Cultivar**	**Kenya**	**1963**	**YAKTANA-54/LERMA-52**	East Africa3	East Africa2
279	**Thatcher**	**Cultivar**	**Minnesota**	**1934**	**MARQUIS/(TR.DR)IUMILLO//MARQUIS/KANRED**	North America	North America
280	**Timstein**	**Cultivar**	**Minnesota**	**1939**	**STEINWEDEL/GAZA**	East Africa3	East Africa3
281	**Tobari66**	**Cultivar**	**Mexico**	**1966**	**TEZANOS-PINTOS-PRECOZ/SONORA-64-A**	East Africa1	Mixed Identity1
282	**Token-Ken**	**Cultivar**	**Kenya**	**1966**	**TIMSTEIN/2*KENYA//YAQUI-50**	North America	North America
283	**Tom**	**Cultivar**	**Minnesota**	**-**	**-**	East Africa3	North America
284	**Trophy**	**Cultivar**	**Kenya**	**1968**	**TIMSTEIN/2*KENYA-RF-324//2*YAQUI-50**	North America	East Africa2
285	**Tura**	**Cultivar**	**Ethiopia**	**1999**	**ARO-YR-SEL-60-1989**	East Africa1	East Africa1
286	**Tusie**	**Cultivar**	**Ethiopia**	**1997**	**COOK/VEERY//DOVE/SERI M82**	East Africa1	East Africa1
287	**Vance**	**Cultivar**	**Minnesota**	**1989**	**ND-560/MN-7595**	East Africa3	North America
288	**Verde**	**Cultivar**	**Minnesota**	**1995**	**MN-7663/SBY-354-A**	East Africa3	North America
289	**Wabe**	**Cultivar**	**Ethiopia**	**1994**	**MIRLO/BUCKBUCK**	East Africa1	East Africa1
290	**Waldron**	**Cultivar**	**Minnesota**	**1968**	**JUSTIN/ND-81**	East Africa3	North America
291	**Wared**	**Cultivar**	**Minnesota**	**1974**	**II-55-10/4/PEMBINA/II-52-329/3/II-53-388/III-58-4//II-53-546**	North America	North America
292	**Wetera**	**Cultivar**	**Ethiopia**	**2000**	**MONCHO''S''-BUCKBUCK''S''**	East Africa1	East Africa1
293	**Wheaton**	**Cultivar**	**Minnesota**	**1983**	**CRIM/2*ERA//BUITRE/GALLO**	North America	North America
294	**Wis.245**	**Cultivar**	**Wisconsin**	**1954**	**PD-2666-A-2-2-2-15-6-3*3/(TR.TI)D-357-1**	North America	North America
295	**Yaktana54A**	**Cultivar**	**Mexico**	**1954**	**YAQUI-48/KENTANA-48//FRONTANA**	East Africa2	East Africa2
296	**Yaqui50**	**Cultivar**	**Minnesota**	**1950**	**NEWTHATCH/MARROQUI-588**	North America	East Africa2
297	**Zaragoza75**	**Cultivar**	**Mexico**	**1975**	**MENGAVI/II-8156**	East Africa1	East Africa1

Appendix II SNP marker loci used in estimation of population structure in 297 wheat lines

SNP No.	Marker	Chrom.	Pos. (cM)	MAF	SNP	SNP No.	Marker	Chrom.	Pos. (cM)	MAF	SNP
1	wsnp_Ex_c10657_17376086	1A	8	0.05	[T/C]	38	wsnp_BE403956B_Ta_2_1	1B	39	0.03	[A/G]
2	wsnp_Ex_c5469_9655281	1A	15	0.26	[A/C]	39	wsnp_BE489692B_Ta_2_2	1B	46	0.38	[T/C]
3	wsnp_Ex_c64327_63176640	1A	27	0.4	[A/G]	40	wsnp_Ex_c5388_9526777	1B	51	0.43	[A/G]
4	wsnp_Ku_c42878_50516167	1A	33	0.41	[A/G]	41	wsnp_Ex_c26620_35859364	1B	56	0.17	[T/C]
5	wsnp_Ra_c26191_35761997	1A	37	0.43	[T/C]	42	wsnp_BE495786B_Ta_2_1	1B	61	0.35	[A/C]
6	wsnp_Ku_c34659_43981982	1A	43	0.1	[A/G]	43	wsnp_CAP11_c543_375403	1B	66	0.24	[A/G]
7	wsnp_Ex_c12117_19381492	1A	47	0.14	[T/G]	44	wsnp_BE446240B_Ta_2_1	1B	71	0.13	[A/G]
8	wsnp_Ex_c3253_5994376	1A	58	0.37	[A/G]	45	wsnp_Ra_rep_c103071_87770006	1B	76	0.35	[T/C]
9	wsnp_Ex_rep_c105244_89727546	1A	63	0.41	[A/G]	46	wsnp_Ex_c26860_36084209	1B	81	0.13	[A/G]
10	wsnp_Ex_c43454_49770425	1A	68	0.46	[T/C]	47	wsnp_BG606986B_Ta_2_1	1B	87	0.45	[T/C]
11	wsnp_BE605063A_Ta_2_2	1A	73	0.02	[T/C]	48	wsnp_Ex_c194_381656	1B	92	0.49	[T/C]
12	wsnp_JD_c722_1080592	1A	78	0.06	[T/C]	49	wsnp_BF200640B_Ta_2_1	1B	97	0.15	[A/G]
13	wsnp_JD_c9291_10157217	1A	83	0.03	[T/C]	50	wsnp_Ex_rep_c67098_65573819	1B	102	0.38	[A/G]
14	wsnp_Ex_c12165_19452361	1A	88	0.11	[T/C]	51	wsnp_JD_c14411_14148961	1B	108	0.39	[A/G]
15	wsnp_BG314157A_Ta_2_1	1A	93	0.34	[A/G]	52	wsnp_CAP7_c4778_2155754	1B	117	0.41	[T/C]
16	wsnp_Ex_c20489_29564938	1A	98	0.29	[T/C]	53	wsnp_CAP12_c1337_682282	1B	122	0.26	[T/C]
17	wsnp_BM137879A_Ta_2_1	1A	103	0.43	[T/C]	54	wsnp_RFL_Contig4298_4993464	1B	127	0.49	[A/G]
18	wsnp_Ku_c17322_26392311	1A	108	0.47	[T/C]	55	wsnp_Ku_c2797_5284087	1B	132	0.19	[T/C]
19	wsnp_Ku_c33917_43336035	1A	113	0.29	[A/G]	56	wsnp_BE446672B_Ta_2_1	1B	137	0.13	[T/C]
20	wsnp_Ex_c33452_41938013	1A	118	0.33	[T/C]	57	wsnp_Ex_c1597_3045682	1B	142	0.12	[A/G]
21	wsnp_Ex_c1359_2604298	1A	123	0.45	[T/G]	58	wsnp_Ex_c278_538285	1D	0	0.3	[T/C]
22	wsnp_Ex_c6488_11266589	1A	128	0.27	[T/C]	59	wsnp_Ku_c3682_6786230	1D	16	0.35	[T/C]
23	wsnp_BF474340A_Ta_2_1	1A	135	0.23	[T/C]	60	wsnp_Ex_c1358_2600929	1D	21	0.33	[T/C]
24	wsnp_Ex_c19353_28290393	1A	143	0.26	[A/C]	61	wsnp_CAP11_c2307_1200406	1D	26	0.43	[T/C]
25	wsnp_CAP11_c146_160903	1A	151	0.43	[A/G]	62	wsnp_RFL_Contig4025_4499792	1D	33	0.18	[A/C]
26	wsnp_Ex_c10631_17340809	1A	156	0.09	[A/G]	63	wsnp_Ex_c3372_6195001	1D	46	0.11	[A/C]
27	wsnp_Ra_c2227_4304970	1A	163	0.41	[T/C]	64	wsnp_CAP11_c1043_618449	1D	51	0.4	[A/G]
28	wsnp_Ex_c9343_15514687	1A	168	0.43	[T/G]	65	wsnp_BE424100D_Ta_1_1	1D	57	0.19	[T/C]
29	wsnp_JD_c628_947299	1A	174	0.31	[T/C]	66	wsnp_BE517805D_Ta_1_1	1D	62	0.43	[T/C]
30	wsnp_Ex_c102566_87702963	1A	180	0.03	[A/G]	67	wsnp_Ku_c53270_57959459	1D	68	0.3	[A/G]
31	wsnp_Ex_c52086_55808126	1A	187	0.26	[A/G]	68	wsnp_RFL_Contig2036_1264133	1D	75	0.04	[T/G]
32	wsnp_Ku_c13229_21142792	1B	9	0.08	[T/C]	69	wsnp_Ex_c26626_35864468	1D	81	0.28	[A/G]
33	wsnp_Ex_c5780_10153638	1B	14	0.29	[T/C]	70	wsnp_Ex_c25974_35235456	1D	86	0.45	[A/G]
34	wsnp_BE405834B_Ta_2_2	1B	19	0.22	[T/C]	71	wsnp_Ku_c7266_12551048	1D	91	0.44	[T/G]
35	wsnp_CAP8_c4576_2228073	1B	24	0.39	[T/C]	72	wsnp_JD_c6544_7697578	1D	96	0.38	[T/C]
36	wsnp_Ex_c11976_19193550	1B	29	0.36	[A/G]	73	wsnp_Ex_rep_c111610_93458148	1D	104	0.44	[T/C]
37	wsnp_Ex_c1429_2745237	1B	34	0.14	[T/C]	74	wsnp_Ex_c35886_43949442	1D	146	0.26	[T/C]

Appendix II *continued*

SNP No.	Marker	Chrom.	Pos. (cM)	MAF	SNP	SNP No.	Marker	Chrom.	Pos. (cM)	MAF	SNP
75	wsnp_Ex_c19516_28481857	2A	0	0.24	[A/G]	113	wsnp_Ex_c2388_4476302	2B	20	0.47	[T/G]
76	wsnp_Ex_c11827_18986376	2A	6	0.21	[T/G]	114	wsnp_Ra_rep_c106119_89961852	2B	31	0.08	[A/C]
77	wsnp_Ex_c26406_35653828	2A	12	0.23	[A/G]	115	wsnp_Ex_c1602_3055066	2B	35	0.23	[T/G]
78	wsnp_Ex_c61879_61748626	2A	19	0.31	[A/G]	116	wsnp_CAP11_rep_c8489_3664985	2B	40	0.07	[A/G]
79	wsnp_CAP12_c510_276361	2A	27	0.12	[A/G]	117	wsnp_Ex_c14711_22788263	2B	44	0.49	[A/G]
80	wsnp_Ex_c21709_30869795	2A	33	0.06	[T/C]	118	wsnp_JD_c5064_6183978	2B	49	0.27	[A/G]
81	wsnp_Ex_c19556_28530231	2A	43	0.47	[T/C]	119	wsnp_BF146221B_Ta_2_2	2B	54	0.01	[A/G]
82	wsnp_Ex_c5412_9564346	2A	52	0.28	[A/G]	120	wsnp_Ex_c7285_12506938	2B	62	0.45	[A/G]
83	wsnp_JD_rep_c48914_33168544	2A	62	0.15	[T/G]	121	wsnp_RFL_Contig2744_2471775	2B	69	0.13	[A/G]
84	wsnp_Ex_c15822_24204224	2A	77	0.43	[A/C]	122	wsnp_BE497494B_Ta_2_1	2B	77	0.21	[A/C]
85	wsnp_Ex_rep_c101526_86881496	2A	83	0.49	[T/C]	123	wsnp_Ex_c6537_11338763	2B	81	0.1	[T/C]
86	wsnp_BG314532A_Ta_2_1	2A	88	0.39	[A/G]	124	wsnp_Ex_c27867_37030229	2B	88	0.39	[T/C]
87	wsnp_Ex_c32079_40793255	2A	96	0.14	[T/G]	125	wsnp_Ku_c4507_8157580	2B	93	0.33	[A/G]
88	wsnp_Ex_rep_c66800_65171198	2A	104	0.43	[T/G]	126	wsnp_CAP12_rep_c4678_2134259	2B	107	0.16	[A/G]
89	wsnp_Ku_rep_c110147_94431847	2A	109	0.24	[A/G]	127	wsnp_BE404601B_Ta_2_1	2B	113	0.23	[T/G]
90	wsnp_Ex_rep_c66377_64570189	2A	114	0.28	[T/C]	128	wsnp_Ex_c5239_9272511	2B	118	0.41	[T/C]
91	wsnp_Ex_c36242_44232473	2A	123	0.2	[T/C]	129	wsnp_BF291736B_Ta_1_1	2B	126	0.35	[A/G]
92	wsnp_bg606625A_Ta_2_1	2A	129	0.15	[T/G]	130	wsnp_Ku_c2561_4878036	2B	131	0.45	[A/G]
93	wsnp_Ex_rep_c102538_87682273	2A	135	0.08	[T/C]	131	wsnp_BF202975B_Ta_2_1	2B	135	0.34	[A/C]
94	wsnp_CAP11_c24_62115	2A	141	0.4	[T/C]	132	wsnp_Ku_rep_c71198_70910111	2B	141	0.33	[T/G]
95	wsnp_Ex_c28204_37349164	2A	149	0.47	[T/C]	133	wsnp_Ex_c15985_24399118	2B	145	0.34	[A/G]
96	wsnp_Ex_rep_c105158_89662129	2A	163	0.14	[A/G]	134	wsnp_Ex_rep_c66545_64828225	2B	149	0.34	[A/C]
97	wsnp_Ex_c7829_13320738	2A	169	0.38	[A/G]	135	wsnp_Ex_c29445_38480890	2B	157	0.06	[A/G]
98	wsnp_Ra_c32271_41304469	2A	181	0.28	[A/C]	136	wsnp_Ex_c2097_3932976	2B	161	0.24	[T/C]
99	wsnp_Ra_c3378_6318431	2A	187	0.33	[T/C]	137	wsnp_Ex_rep_c67543_66165372	2B	165	0.44	[T/G]
100	wsnp_Ex_c15133_23323937	2A	193	0.07	[A/G]	138	wsnp_Ex_c54998_57670603	2B	169	0.13	[T/C]
101	wsnp_CAP11_c1711_934478	2A	202	0.23	[A/G]	139	wsnp_Ex_c10796_17575074	2B	185	0.47	[T/C]
102	wsnp_Ex_c24964_34219020	2A	207	0.25	[A/C]	140	wsnp_Ex_c18503_27349536	2B	189	0.18	[T/C]
103	wsnp_Ex_c19003_27913936	2A	212	0.12	[T/C]	141	wsnp_Ex_c12922_20472434	2B	193	0.28	[T/C]
104	wsnp_JD_c13946_13810300	2A	216	0.17	[A/G]	142	wsnp_Ex_c2203_4130096	2B	197	0.31	[A/G]
105	wsnp_Ex_c14639_22695091	2A	223	0.17	[T/G]	143	wsnp_Ex_c28627_37743031	2B	203	0.39	[T/C]
106	wsnp_Ex_rep_c108004_91402649	2A	230	0.24	[A/G]	144	wsnp_JD_c6010_7167084	2B	207	0.3	[A/G]
107	wsnp_Ex_c63909_62932893	2A	235	0.13	[A/G]	145	wsnp_Ex_c45468_51254978	2B	212	0.1	[T/C]
108	wsnp_CAP11_c1737_946813	2A	241	0.16	[T/C]	146	wsnp_Ex_c13865_21720466	2B	216	0.22	[A/G]
109	wsnp_Ex_c11560_18632777	2B	0	0.44	[T/C]	147	wsnp_Ex_rep_c101581_86924089	2B	226	0.31	[T/G]
110	wsnp_Ex_c851_1654297	2B	5	0.12	[T/C]	148	wsnp_BG605258B_Ta_2_7	2B	231	0.48	[T/C]
111	wsnp_Ex_c259_497455	2B	10	0.13	[A/G]	149	wsnp_BG605258B_Ta_2_1	2B	241	0.05	[A/G]
112	wsnp_Ex_c13686_21480826	2B	15	0.22	[A/G]	150	wsnp_Ex_rep_c72569_70908990	2B	249	0.23	[A/G]

Appendix II *continued*

SNP No.	Marker	Chrom.	Pos. (cM)	MAF	SNP	SNP No.	Marker	Chrom.	Pos. (cM)	MAF	SNP
151	wsnp_Ra_c17622_26522072	2B	254	0.22	[A/G]	187	wsnp_RFL_Contig4752_5707729	3A	46	0.03	[A/G]
152	wsnp_CAP11_c3226_1588070	2B	260	0.34	[T/C]	188	wsnp_Ex_c22435_31629303	3A	52	0.31	[A/G]
153	wsnp_Ex_c5193_9204522	2B	265	0.16	[A/G]	189	wsnp_Ex_c25653_34914467	3A	56	0.19	[A/G]
154	wsnp_Ex_c31064_39902843	2B	271	0.21	[T/C]	190	wsnp_Ex_c6217_10848574	3A	60	0.44	[A/G]
155	wsnp_Ex_c14779_22892053	2D	0	0.34	[T/C]	191	wsnp_Ex_c14420_22402673	3A	64	0.18	[T/C]
156	wsnp_JD_rep_c63957_40798121	2D	5	0.33	[T/C]	192	wsnp_Ex_rep_c101340_86719239	3A	69	0.46	[A/G]
157	wsnp_Ex_c12250_19568265	2D	10	0.05	[A/C]	193	wsnp_Ex_c28310_37444843	3A	73	0.16	[A/G]
158	wsnp_Ex_c1668_3168723	2D	15	0.03	[T/C]	194	wsnp_Ex_c12850_20377830	3A	77	0.15	[T/C]
159	wsnp_BG275030D_Ta_2_2	2D	40	0.06	[T/C]	195	wsnp_Ex_c32003_40728918	3A	81	0.38	[A/G]
160	wsnp_CAP12_c812_428290	2D	55	0.36	[A/G]	196	wsnp_Ku_c9316_15647115	3A	85	0.11	[A/G]
161	wsnp_CAP12_c1503_764765	2D	59	0.28	[T/C]	197	wsnp_Ku_c4886_8753646	3A	98	0.25	[A/G]
162	wsnp_Ex_c5412_9565527	2D	63	0.21	[T/C]	198	wsnp_Ex_c1335_2556442	3A	102	0.22	[A/G]
163	wsnp_Ra_c67199_65253620	2D	71	0.18	[T/C]	199	wsnp_JD_c2743_3678590	3A	106	0.27	[T/C]
164	wsnp_Ku_c8712_14751858	2D	78	0.17	[T/C]	200	wsnp_Ex_c4094_7399975	3A	111	0.09	[A/G]
165	wsnp_BE488779D_Ta_1_2	2D	82	0.06	[A/G]	201	wsnp_Ex_c8884_14841846	3A	117	0.14	[A/G]
166	wsnp_RFL_Contig3561_3746066	2D	97	0.37	[A/G]	202	wsnp_Ex_rep_c67786_66472676	3A	122	0.14	[T/C]
167	wsnp_Ex_c2251_4218338	2D	103	0.12	[A/G]	203	wsnp_CAP11_rep_c4226_1995152	3A	127	0.34	[T/C]
168	wsnp_Ra_rep_c116793_96612614	2D	108	0.14	[A/G]	204	wsnp_BG262734A_Ta_2_3	3A	132	0.44	[A/G]
169	wsnp_Ex_rep_c69782_68740893	2D	114	0.25	[A/G]	205	wsnp_Ex_c20250_29303152	3A	136	0.23	[T/G]
170	wsnp_CAP7_c2782_1329707	2D	131	0.14	[T/G]	206	wsnp_CAP8_c6939_3242530	3A	147	0.46	[T/C]
171	wsnp_Ex_c18250_27065775	2D	140	0.16	[T/G]	207	wsnp_BE604885A_Ta_2_1	3A	152	0.37	[A/G]
172	wsnp_Ra_rep_c71290_69343893	2D	145	0.33	[A/G]	208	wsnp_BE426418A_Ta_2_2	3A	156	0.12	[T/C]
173	wsnp_Ex_rep_c66522_64795143	2D	151	0.1	[A/G]	209	wsnp_CAP11_rep_c8581_3702222	3A	161	0.1	[A/G]
174	wsnp_Ku_c2713_5145312	2D	156	0.22	[T/C]	210	wsnp_Ex_c25082_34346512	3A	165	0.22	[T/C]
175	wsnp_Ra_rep_c108284_91604017	2D	163	0.3	[A/G]	211	wsnp_BM137927A_Ta_2_1	3A	173	0.2	[T/G]
176	wsnp_BE445431D_Ta_2_1	2D	168	0.44	[A/G]	212	wsnp_Ra_c5532_9788185	3B	1	0.2	[T/C]
177	wsnp_RFL_Contig463_5510887	2D	173	0.03	[T/C]	213	wsnp_Ex_c2723_5047696	3B	5	0.25	[A/G]
178	wsnp_RFL_Contig2659_2346243	2D	184	0.14	[A/G]	214	wsnp_Ex_c8386_14128029	3B	9	0.31	[A/C]
179	wsnp_Ex_c25132_34396655	3A	2	0.46	[T/C]	215	wsnp_Ku_c55746_59522526	3B	13	0.3	[T/C]
180	wsnp_Ex_c2573_4788116	3A	6	0.34	[T/C]	216	wsnp_Ex_rep_c67033_65490126	3B	19	0.26	[A/G]
181	wsnp_JD_c2722_3653988	3A	14	0.43	[T/C]	217	wsnp_Ex_c19993_29024127	3B	23	0.38	[T/G]
182	wsnp_Ex_c8409_14170476	3A	19	0.32	[A/G]	218	wsnp_Ku_c663_1368085	3B	27	0.37	[T/C]
183	wsnp_BE495175A_Ta_1_1	3A	24	0.1	[T/C]	219	wsnp_Ra_c488_1027573	3B	33	0.33	[T/C]
184	wsnp_Ex_c15674_24004513	3A	29	0.34	[A/G]	220	wsnp_RFL_Contig2177_1500201	3B	38	0.44	[T/C]
185	wsnp_Ex_c33765_42199371	3A	35	0.37	[A/G]	221	wsnp_Ku_c9009_15185650	3B	53	0.29	[A/G]
186	wsnp_Ex_c742_1458743	3A	42	0.18	[A/C]	222	wsnp_Ex_c33140_41677458	3B	58	0.44	[T/C]

Appendix II *continued*

SNP No.	Marker	Chrom.	Pos. (cM)	MAF	SNP	SNP No.	Marker	Chrom.	Pos. (cM)	MAF	SNP
223	wsnp_JD_c828_1226159	3B	62	0.13	[A/G]	256	wsnp_Ex_c20386_29451037	4A	33	0.27	[T/C]
224	wsnp_Ex_c19372_28312715	3B	66	0.21	[T/C]	257	wsnp_Ex_c55245_57821568	4A	38	0.45	[A/G]
225	wsnp_Ra_c32055_41111615	3B	71	0.23	[A/G]	258	wsnp_Ku_rep_c102527_89481571	4A	43	0.47	[A/C]
226	wsnp_Ex_c22630_31827919	3B	75	0.27	[T/G]	259	wsnp_Ku_rep_c101597_88729428	4A	48	0.23	[A/G]
227	wsnp_Ex_c18624_27492167	3B	81	0.19	[A/G]	260	wsnp_Ku_c32619_42216179	4A	53	0.44	[T/C]
228	wsnp_BE494474B_Ta_2_1	3B	86	0.07	[A/G]	261	wsnp_BE445427A_Ta_2_1	4A	58	0.41	[T/C]
229	wsnp_BE445348B_Ta_2_1	3B	91	0.07	[T/G]	262	wsnp_Ex_c7011_12080274	4A	63	0.09	[T/G]
230	wsnp_JD_c5643_6802088	3B	96	0.19	[T/C]	263	wsnp_BE398523A_Ta_2_1	4A	68	0.25	[A/C]
231	wsnp_BE499016B_Ta_2_1	3B	102	0.15	[T/C]	264	wsnp_Ex_rep_c104859_89444355	4A	73	0.18	[A/G]
232	wsnp_Ex_c3040_5616383	3B	111	0.03	[T/C]	265	wsnp_BE404977A_Ta_1_1	4A	78	0.18	[T/C]
233	wsnp_Ra_c10710_17570054	3B	116	0.28	[T/C]	266	wsnp_JD_c8309_9321723	4A	83	0.19	[A/G]
234	wsnp_RFL_Contig204_1272643	3B	121	0.36	[T/G]	267	wsnp_Ex_c17338_26018247	4A	89	0.35	[T/C]
235	wsnp_Ex_c13505_21253168	3B	126	0.03	[A/G]	268	wsnp_Ex_c13031_20625900	4A	98	0.42	[A/G]
236	wsnp_Ex_c3907_7088011	3B	132	0.17	[A/G]	269	wsnp_Ex_rep_c67958_66693031	4A	102	0.16	[A/G]
237	wsnp_Ex_rep_c101457_86818055	3B	139	0.3	[T/G]	270	wsnp_Ex_c5470_9657856	4A	109	0.3	[T/C]
238	wsnp_Ex_c2500_4671165	3B	157	0.49	[A/C]	271	wsnp_Ex_c612_1213451	4A	116	0.22	[T/C]
239	wsnp_CAP11_c59_99317	3B	164	0.1	[A/G]	272	wsnp_Ra_rep_c87547_79842909	4A	125	0.08	[A/G]
240	wsnp_Ex_c2536_4728768	3B	171	0.13	[A/G]	273	wsnp_Ex_c41313_48161689	4A	132	0.14	[A/G]
241	wsnp_Ex_c56591_58653386	3B	176	0.18	[T/C]	274	wsnp_Ex_c2352_4405961	4A	137	0.14	[T/C]
242	wsnp_Ex_c16569_25082760	3B	181	0.35	[T/C]	275	wsnp_CAP11_c8366_3622210	4A	145	0.2	[A/C]
243	wsnp_Ex_c45877_51547406	3B	195	0.25	[A/G]	276	wsnp_Ku_c8391_14261321	4A	151	0.15	[T/C]
244	wsnp_Ex_c10630_17338753	3D	1	0.47	[A/G]	277	wsnp_Ex_c1246_2393978	4A	165	0.2	[T/G]
245	wsnp_Ex_c13629_21411429	3D	5	0.15	[T/C]	278	wsnp_Ra_c4400_7986499	4A	174	0.07	[T/G]
246	wsnp_Ex_c14027_21925629	3D	12	0.11	[T/C]	279	wsnp_Ku_c7102_12271493	4A	181	0.08	[T/C]
247	wsnp_Ex_c28705_37809171	3D	12	0.11	[A/G]	280	wsnp_Ex_c1246_2393249	4A	189	0.4	[T/C]
248	wsnp_Ra_c17636_26538543	3D	16	0.15	[T/C]	281	wsnp_Ex_c13953_21831752	4A	198	0.1	[T/C]
249	wsnp_Ex_c5061_8986366	3D	21	0.25	[T/C]	282	wsnp_Ex_c10955_17794520	4A	206	0.36	[T/C]
250	wsnp_Ra_rep_c69820_67401482	3D	85	0.12	[A/G]	283	wsnp_Ku_c12399_20037334	4B	0	0.41	[T/C]
251	wsnp_Ex_c13354_21047873	4A	0	0.03	[A/G]	284	wsnp_Ex_c29500_38529808	4B	8	0.01	[T/C]
252	wsnp_Ex_c9928_16346945	4A	12	0.46	[A/C]	285	wsnp_Ex_c10347_16946522	4B	15	0.44	[T/C]
253	wsnp_Ra_c14920_23225219	4A	17	0.24	[A/G]	286	wsnp_Ex_c6739_11646407	4B	21	0.24	[A/G]
254	wsnp_Ex_c28429_37553452	4A	22	0.26	[T/C]	287	wsnp_Ex_c7362_12622736	4B	26	0.24	[A/C]
255	wsnp_Ex_c4752_8482625	4A	27	0.38	[A/G]	288	wsnp_BE442666B_Ta_2_1	4B	37	0.13	[A/G]

Appendix II *continued*

SNP No.	Marker	Chrom.	Pos. (cM)	MAF	SNP	SNP No.	Marker	Chrom.	Pos. (cM)	MAF	SNP
289	wsnp_Ex_c8913_14881924	4B	42	0.3	[A/C]	330	wsnp_Ex_c19519_28487099	5A	92	0.15	[T/C]
290	wsnp_Ex_c18433_27269748	4B	57	0.18	[A/G]	331	wsnp_Ra_c17216_26044790	5A	98	0.23	[A/G]
291	wsnp_Ex_c3119_5763762	4B	62	0.22	[A/G]	332	wsnp_BE405849A_Ta_1_1	5A	103	0.12	[T/C]
292	wsnp_Ra_c18498_27571740	4B	67	0.32	[T/C]	333	wsnp_Ex_c17523_26244256	5A	119	0.41	[T/C]
293	wsnp_Ku_rep_c69863_69343983	4B	72	0.1	[T/C]	334	wsnp_AJ612027A_Ta_2_1	5A	124	0.48	[A/G]
294	wsnp_Ex_c26807_36031771	4B	77	0.11	[T/C]	335	wsnp_Ex_c621_1231444	5A	136	0.28	[T/C]
295	wsnp_Ra_c1146_2307483	4B	83	0.13	[T/G]	336	wsnp_BF484028B_Td_2_1	5A	141	0.01	[A/G]
296	wsnp_Ex_c13764_21585737	4B	89	0.03	[A/G]	337	wsnp_BG607308A_Ta_2_1	5A	146	0.05	[A/G]
297	wsnp_Ex_c5187_9195120	4B	96	0.37	[T/C]	338	wsnp_Ex_c94356_82888150	5A	151	0.34	[T/C]
298	wsnp_Ex_c38704_46166494	4B	103	0.07	[A/G]	339	wsnp_Ex_c5461_9636197	5A	156	0.14	[T/G]
299	wsnp_CAP7_c599_312057	4B	107	0.28	[T/C]	340	wsnp_Ex_c18107_26909127	5A	163	0.15	[T/C]
300	wsnp_BG604404B_Ta_2_1	4B	112	0.33	[A/G]	341	wsnp_Ex_c1481_2831499	5A	169	0.34	[A/G]
301	wsnp_Ex_c4148_7494801	4B	117	0.32	[T/C]	342	wsnp_Ex_c31154_39982416	5A	177	0.46	[T/C]
302	wsnp_Ex_c29200_38260949	4B	123	0.12	[T/G]	343	wsnp_Ex_c12684_20157261	5A	182	0.36	[T/C]
303	wsnp_Ex_rep_c107564_91144523	4D	1	0.15	[A/G]	344	wsnp_Ex_c32414_41076471	5A	187	0.37	[T/C]
304	wsnp_CAP11_c356_280910	4D	8	0.12	[A/G]	345	wsnp_Ex_c3620_6612294	5A	193	0.33	[A/G]
305	wsnp_RFL_Contig2828_2623246	4D	13	0.49	[A/C]	346	wsnp_CAP11_c3658_1764767	5B	0	0.03	[T/C]
306	wsnp_Ku_c9140_15390166	4D	16	0.1	[T/G]	347	wsnp_Ex_c607_1204733	5B	11	0.26	[A/C]
307	wsnp_BF202706D_Ta_1_1	4D	19	0.2	[A/G]	348	wsnp_Ex_c2459_4591695	5B	15	0.26	[A/G]
308	wsnp_Ex_rep_c67296_65839761	4D	23	0.24	[A/G]	349	wsnp_Ex_c12927_20480163	5B	20	0.23	[T/C]
309	wsnp_Ex_c12450_19850925	4D	28	0.1	[T/C]	350	wsnp_Ku_c568_1187615	5B	32	0.49	[T/G]
310	wsnp_Ex_c15977_24392258	4D	33	0.13	[T/C]	351	wsnp_Ex_c214_421541	5B	38	0.24	[T/C]
311	wsnp_Ex_rep_c79748_75305162	4D	40	0.06	[T/G]	352	wsnp_Ex_c58012_59490259	5B	44	0.29	[A/G]
312	wsnp_BF473052D_Ta_2_1	4D	48	0.23	[A/G]	353	wsnp_Ex_c5915_10378599	5B	49	0.23	[T/C]
313	wsnp_BE591739D_Ta_2_3	4D	53	0.49	[T/C]	354	wsnp_BQ166999B_Ta_2_1	5B	54	0.46	[T/G]
314	wsnp_Ku_rep_c70792_70439377	4D	61	0.06	[T/C]	355	wsnp_CAP8_c1594_914839	5B	64	0.34	[A/G]
315	wsnp_Ex_c34252_42593715	4D	70	0.06	[A/G]	356	wsnp_Ex_c15304_23532301	5B	69	0.16	[A/G]
316	wsnp_RFL_Contig4136_4696148	5A	0	0.13	[T/C]	357	wsnp_Ku_c6464_11320381	5B	69	0.35	[T/C]
317	wsnp_Ex_c25707_34968426	5A	11	0.1	[T/C]	358	wsnp_BE497820B_Ta_2_2	5B	80	0.17	[T/C]
318	wsnp_Ku_c9559_16000086	5A	18	0.36	[T/C]	359	wsnp_CAP7_c3386_1586636	5B	85	0.04	[T/C]
319	wsnp_Ex_c11992_19213872	5A	22	0.38	[T/C]	360	wsnp_Ex_c33455_41940691	5B	92	0.35	[A/G]
320	wsnp_BE499835A_Ta_2_1	5A	31	0.2	[A/G]	361	wsnp_Ex_c10902_17715268	5B	97	0.15	[T/C]
321	wsnp_Ex_c356_698872	5A	39	0.21	[A/G]	362	wsnp_CAP11_c948_571287	5B	102	0.19	[A/G]
322	wsnp_Ex_c10231_16783750	5A	44	0.39	[T/C]	363	wsnp_Ex_c18941_27841286	5B	110	0.23	[A/G]
323	wsnp_Ku_c933_1913402	5A	52	0.26	[A/G]	364	wsnp_CAP12_c303_168418	5B	116	0.34	[A/G]
324	wsnp_Ex_c11874_19050989	5A	57	0.09	[A/G]	365	wsnp_JD_rep_c49950_34066505	5B	122	0.21	[T/C]
325	wsnp_Ex_rep_c101757_87065169	5A	65	0.27	[T/C]	366	wsnp_JD_rep_c63083_40243538	5B	127	0.01	[A/G]
326	wsnp_Ex_c16127_24561453	5A	70	0.2	[A/G]	367	wsnp_Ku_c3294_6125586	5B	135	0.4	[T/C]
327	wsnp_Ex_c4921_8763500	5A	75	0.26	[A/G]	368	wsnp_Ex_c210_411604	5B	141	0.26	[A/C]
328	wsnp_Ku_c35386_44598937	5A	80	0.16	[A/G]	369	wsnp_Ex_c56866_58816359	5B	146	0.22	[A/C]
329	wsnp_Ku_rep_c103857_90489584	5A	85	0.2	[A/G]	370	wsnp_Ex_c13485_21225504	5B	152	0.17	[A/G]

Appendix II *continued*

SNP No.	Marker	Chrom.	Pos. (cM)	MAF	SNP	SNP No.	Marker	Chrom.	Pos. (cM)	MAF	SNP
371	wsnp_Ra_c5212_9291784	5B	157	0.28	[A/G]	413	wsnp_Ex_rep_c102807_87894515	6A	116	0.35	[A/G]
372	wsnp_Ex_c3042_5618219	5B	162	0.06	[T/C]	414	wsnp_Ku_rep_c102901_89769309	6A	122	0.17	[T/C]
373	wsnp_Ex_c35398_43558614	5B	168	0.02	[T/C]	415	wsnp_CAP12_c928_486444	6A	127	0.33	[A/C]
374	wsnp_BG607308B_Ta_2_1	5B	173	0.46	[T/G]	416	wsnp_Ex_rep_c105594_89968727	6A	132	0.3	[T/C]
375	wsnp_Ex_c53170_56501500	5B	179	0.28	[T/C]	417	wsnp_BE495143A_Ta_2_1	6A	139	0.31	[A/G]
376	wsnp_Ex_c8019_13598142	5B	184	0.43	[T/C]	418	wsnp_Ex_rep_c68010_66754171	6A	144	0.02	[T/C]
377	wsnp_Ex_c29130_38196906	5B	194	0.24	[A/G]	419	wsnp_Ex_c11439_18459047	6A	154	0.06	[T/C]
378	wsnp_Ex_c7196_12357989	5B	202	0.32	[T/G]	420	wsnp_Ra_c2270_4383252	6A	171	0.33	[A/C]
379	wsnp_Ku_c4427_8029592	5B	207	0.49	[A/C]	421	wsnp_Ex_rep_c101766_87073440	6A	177	0.47	[T/C]
380	wsnp_Ex_c3175_5864291	5B	213	0.43	[A/G]	422	wsnp_Ku_c29287_39194579	6A	183	0.19	[A/G]
381	wsnp_Ra_c5637_9955966	5B	218	0.13	[A/G]	423	wsnp_CAP12_c1663_836928	6A	189	0.4	[T/C]
382	wsnp_Ex_c24031_33277293	5B	224	0.13	[A/G]	424	wsnp_Ku_c28854_38769308	6A	194	0.01	[A/G]
383	wsnp_Ex_c11055_17927668	5D1	0	0.11	[A/G]	425	wsnp_BF291974A_Ta_2_1	6A	198	0.11	[A/G]
384	wsnp_CAP11_c209_198432	5D1	5	0.29	[A/G]	426	wsnp_Ex_c26771_35998816	6A	205	0.47	[A/G]
385	wsnp_Ex_c33327_41834973	5D2	13	0.12	[T/C]	427	wsnp_Ex_c5446_9616983	6A	210	0.4	[A/G]
386	wsnp_Ex_c14182_22122844	5D2	18	0.21	[T/C]	428	wsnp_Ku_rep_c71567_71302229	6A	216	0.16	[T/C]
387	wsnp_CAP8_c2589_1356390	5D2	23	0.11	[T/G]	429	wsnp_Ex_c1143_2195598	6B	0	0.21	[A/C]
388	wsnp_Ex_c31741_40494572	5D2	32	0.04	[T/C]	430	wsnp_RFL_Contig2223_1603535	6B	5	0.16	[T/C]
389	wsnp_JD_rep_c62958_40146122	5D2	38	0.12	[T/C]	431	wsnp_JD_c15167_14703349	6B	13	0.41	[T/C]
390	wsnp_Ku_c7276_12565357	5D2	44	0.49	[A/G]	432	wsnp_Ex_rep_c115803_95396724	6B	18	0.26	[A/G]
391	wsnp_RFL_Contig1127_166848	5D2	58	0.06	[A/G]	433	wsnp_BE404947B_Ta_2_12	6B	23	0.27	[T/C]
392	wsnp_RFL_Contig1749_876559	5D2	84	0.49	[T/C]	434	wsnp_Ex_c19873_28889591	6B	28	0.05	[T/G]
393	wsnp_RFL_Contig2996_2877869	5D2	91	0.02	[A/C]	435	wsnp_BQ167224B_Ta_2_1	6B	33	0.09	[A/C]
394	wsnp_Ex_c1278_2449191	5D2	106	0.08	[A/G]	436	wsnp_RFL_Contig2738_2459768	6B	38	0.19	[T/C]
395	wsnp_Ex_c508_1008029	5D2	118	0.17	[T/C]	437	wsnp_Ex_c7907_13427724	6B	44	0.1	[T/C]
396	wsnp_Ex_c22984_32207214	5D2	123	0.23	[A/G]	438	wsnp_BE488206B_Ta_2_1	6B	49	0.1	[A/G]
397	wsnp_Ku_c26310_36272732	6A	1	0.07	[T/C]	439	wsnp_Ex_rep_c66552_64838102	6B	54	0.22	[T/C]
398	wsnp_Ex_c21633_30782312	6A	7	0.09	[T/C]	440	wsnp_Ex_c17349_26035281	6B	59	0.2	[T/C]
399	wsnp_CAP11_c2142_1128735	6A	15	0.43	[A/G]	441	wsnp_Ex_c13816_21656469	6B	64	0.19	[T/C]
400	wsnp_CAP8_c6680_3136899	6A	22	0.22	[T/C]	442	wsnp_Ex_c1725_3267186	6B	73	0.21	[T/C]
401	wsnp_Ku_rep_c71785_71517370	6A	32	0.08	[A/G]	443	wsnp_Ex_c3858_7011837	6B	78	0.38	[T/C]
402	wsnp_Ex_c9763_16125630	6A	37	0.17	[A/C]	444	wsnp_Ex_c1145_2198295	6B	83	0.4	[A/G]
403	wsnp_Ex_c13223_20866191	6A	42	0.29	[T/C]	445	wsnp_Ex_c6143_10747643	6B	88	0.46	[A/C]
404	wsnp_BQ159493A_Ta_2_2	6A	53	0.16	[A/G]	446	wsnp_Ex_c3854_7003399	6B	94	0.31	[T/C]
405	wsnp_JD_c2180_3000498	6A	61	0.32	[A/G]	447	wsnp_Ex_c11348_18326158	6B	99	0.26	[T/C]
406	wsnp_BE490604A_Ta_2_1	6A	76	0.49	[T/C]	448	wsnp_RFL_Contig4180_4784206	6B	104	0.39	[A/G]
407	wsnp_BF483993A_Ta_2_1	6A	81	0.15	[A/G]	449	wsnp_CV776265A_Ta_2_2	6B	109	0.22	[A/G]
408	wsnp_CAP7_c1839_907899	6A	87	0.16	[T/C]	450	wsnp_CAP12_c475_258416	6B	117	0.09	[A/G]
410	wsnp_BF202329A_Ta_2_2	6A	98	0.09	[T/G]	451	wsnp_Ex_c1383_2651887	6B	122	0.19	[A/G]
411	wsnp_CAP12_c1226_628910	6A	104	0.28	[A/G]	452	wsnp_Ex_c7713_13153321	6B	128	0.15	[A/C]
412	wsnp_Ex_c34545_42832894	6A	111	0.44	[T/G]	453	wsnp_Ex_c11379_18370982	6B	143	0.11	[A/G]

Appendix II *continued*

SNP No.	Marker	Chrom.	Pos. (cM)	MAF	SNP	SNP No.	Marker	Chrom.	Pos. (cM)	MAF	SNP
454	wsnp_Ex_rep_c69373_68312188	6B	148	0.46	[A/G]	497	wsnp_BF483039A_Ta_2_1	7A	186	0.45	[T/C]
455	wsnp_CAP8_rep_c4776_2319078	6B	153	0.16	[A/C]	498	wsnp_JD_c1219_1766330	7A	194	0.31	[T/C]
456	wsnp_Ex_c6731_11634168	6B	153	0.27	[T/G]	499	wsnp_CAP11_c589_398835	7B	0	0.36	[A/C]
457	wsnp_Ex_c1988_3742022	6D	1	0.13	[A/C]	500	wsnp_Ra_c56439_58965239	7B	12	0.42	[A/G]
458	wsnp_Ex_c4894_8723248	6D	6	0.26	[A/G]	501	wsnp_Ex_c24376_33618864	7B	18	0.24	[A/G]
459	wsnp_Ra_c25255_34824465	6D	14	0.3	[A/G]	502	wsnp_Ex_c23755_32994701	7B	25	0.22	[A/G]
460	wsnp_BQ161779D_Ta_2_1	6D	28	0.16	[T/C]	503	wsnp_CAP11_c802_502739	7B	30	0.13	[T/C]
461	wsnp_Ex_c30754_39633791	6D	74	0.22	[T/C]	504	wsnp_RFL_Contig3316_3390729	7B	38	0.45	[T/C]
462	wsnp_Ku_c46270_53051831	6D	84	0.23	[A/G]	505	wsnp_Ex_c20139_29186540	7B	43	0.01	[A/G]
463	wsnp_BM137835D_Ta_2_1	6D	114	0.29	[T/C]	506	wsnp_Ku_c7052_12196253	7B	48	0.32	[T/C]
464	wsnp_Ex_c4518_8119503	6D	119	0.38	[A/G]	507	wsnp_BQ169669B_Ta_2_2	7B	54	0.09	[A/G]
465	wsnp_Ex_c14984_23137713	6D	125	0.11	[T/C]	508	wsnp_Ex_c2847_5257068	7B	59	0.11	[T/C]
466	wsnp_Ex_c43009_49439922	7A	4	0.17	[A/G]	509	wsnp_Ku_c10572_17445600	7B	64	0.25	[T/C]
467	wsnp_Ex_c14654_22713620	7A	9	0.09	[T/C]	510	wsnp_Ex_c1446_2775275	7B	71	0.22	[A/C]
468	wsnp_Ex_c24167_33416760	7A	14	0.3	[A/G]	511	wsnp_Ex_c22955_32173776	7B	77	0.43	[T/G]
469	wsnp_BG313770A_Ta_2_1	7A	20	0.4	[T/C]	512	wsnp_BF474379B_Ta_2_1	7B	82	0.04	[T/C]
470	wsnp_JD_c513_777807	7A	32	0.34	[T/G]	513	wsnp_BF482403B_Ta_1_1	7B	88	0.45	[T/C]
471	wsnp_Ex_c13248_20898211	7A	37	0.22	[A/G]	514	wsnp_JD_c2701_3626787	7B	96	0.18	[T/C]
472	wsnp_BF473884A_Ta_1_1	7A	42	0.22	[A/G]	515	wsnp_Ex_c8963_14948293	7B	101	0.43	[T/C]
473	wsnp_Ku_c57198_60433631	7A	46	0.35	[T/C]	516	wsnp_BE605194B_Ta_2_7	7B	106	0.39	[A/G]
474	wsnp_Ex_c20062_29096408	7A	52	0.18	[T/C]	517	wsnp_Ex_c56425_58548095	7B	116	0.23	[A/G]
475	wsnp_Ex_c40247_47349166	7A	57	0.24	[T/C]	518	wsnp_Ex_c17176_25816766	7B	123	0.37	[T/C]
476	wsnp_Ex_c1334_2553027	7A	64	0.12	[A/G]	519	wsnp_Ra_c39394_47110214	7B	129	0.3	[A/C]
477	wsnp_Ex_c10094_16590615	7A	69	0.18	[A/G]	520	wsnp_CAP7_c4513_2055201	7B	135	0.27	[A/C]
478	wsnp_bf474966A_Ta_2_1	7A	75	0.24	[A/G]	521	wsnp_Ku_c707_1465779	7B	140	0.25	[A/G]
479	wsnp_Ex_c26560_35803210	7A	80	0.13	[T/C]	522	wsnp_Ex_rep_c69316_68246280	7B	150	0.07	[T/C]
480	wsnp_Ra_c5251_9356517	7A	85	0.33	[A/G]	523	wsnp_Ex_c3930_7127883	7B	157	0.47	[A/G]
481	wsnp_JD_c15333_14824351	7A	93	0.45	[A/G]	524	wsnp_Ex_c12556_19992074	7B	164	0.26	[A/G]
482	wsnp_Ra_c18364_27416387	7A	97	0.4	[T/C]	525	wsnp_RFL_Contig3269_3313084	7B	170	0.19	[T/C]
483	wsnp_Ex_c1146_2200823	7A	102	0.48	[T/C]	526	wsnp_BE443044D_Ta_1_1	7D	0	0.04	[A/G]
484	wsnp_Ex_c15341_23591761	7A	107	0.11	[T/C]	527	wsnp_JG_c10610_3538832	7D	8	0.49	[A/G]
485	wsnp_BQ160404A_Ta_1_1	7A	117	0.15	[A/G]	528	wsnp_CAP8_rep_c9647_4198594	7D	17	0.35	[T/C]
486	wsnp_Ex_c658_1293865	7A	125	0.41	[A/G]	529	wsnp_Ra_c6864_11931483	7D	28	0.24	[T/C]
487	wsnp_Ex_c53843_56941644	7A	130	0.12	[A/G]	530	wsnp_Ku_c17320_26385567	7D	34	0.37	[A/G]
488	wsnp_Ku_c19943_29512612	7A	134	0.14	[A/G]	531	wsnp_Ex_c10430_17064001	7D	47	0.47	[A/C]
489	wsnp_Ex_c10617_17320668	7A	141	0.38	[T/C]	532	wsnp_Ex_c2123_3988735	7D	57	0.1	[A/G]
490	wsnp_Ex_c53442_56678505	7A	146	0.14	[T/C]	533	wsnp_be493897D_Ta_1_1	7D	68	0.04	[A/G]
491	wsnp_Ex_c9971_16412345	7A	155	0.29	[T/C]	534	wsnp_Ex_c20320_29383285	7D	95	0.2	[T/C]
492	wsnp_Ku_c26118_36079171	7A	160	0.05	[A/G]	535	wsnp_Ku_c27221_37168065	7D	102	0.09	[A/C]
493	wsnp_RFL_Contig2805_2579582	7A	165	0.32	[T/C]	536	wsnp_BE490643D_Ta_2_1	7D	126	0.16	[T/C]
494	wsnp_Ku_c5693_10079343	7A	170	0.25	[T/C]	537	wsnp_Ex_c17914_26681837	7D	135	0.24	[T/C]
495	wsnp_Ex_c53387_56639804	7A	175	0.46	[T/C]	538	wsnp_Ex_c5884_10325223	7D	140	0.06	[A/G]
496	wsnp_CAP11_c639_424134	7A	181	0.36	[A/G]						

Appendix III Mean highest extent of significant LD (FDR ≤ 0.01) in cM across chromosomes in six subpopulations inferred from the East Africa GWAS panel. Values in parenthesis indicate median.

Chrom	East Africa1	East Africa2	East Africa3	Mixed Identity1	Mixed Identity2	North America	Combined
1A	34 (21)	23 (8)	25 (8)	25 (6)	10 (1)	23 (6)	27 (12)
2A	49 (36)	40 (27)	42 (25)	27 (0)	14 (0)	36 (19)	43 (19)
3A	35 (25)	22 (4)	26 (15)	10 (0)	27 (3)	23 (5)	30 (18)
4A	32 (15)	18 (8)	18 (9)	22 (11)	40 (22)	20 (8)	27 (13)
5A	38 (28)	25 (1)	22 (4)	21 (0)	42 (27)	40 (27)	30 (12)
6A	52 (38)	29 (10)	25 (5)	14 (0)	60 (57)	52 (37)	31 (11)
7A	40 (29)	27 (6)	19 (5)	26 (3)	38 (24)	39 (26)	28 (12)
1B	21 (9)	23 (14)	22 (16)	5 (0)	12 (1)	17 (7)	24 (18)
2B	34 (20)	36 (20)	34 (15)	10 (0)	20 (7)	35 (14)	46 (30)
3B	20 (8)	22 (8)	18 (6)	1 (0)	6 (0)	19 (5)	32 (22)
4B	17 (7)	18 (7)	18 (7)	0 (1)	11 (2)	21 (4)	25 (16)
5B	28 (9)	24 (7)	22 (6)	19 (1)	8 (0)	30 (7)	47 (34)
6B	21 (8)	19 (11)	17 (11)	6 (1)	7 (2)	15 (7)	28 (17)
7B	16 (3)	22 (8)	23 (8)	6 (0)	8 (2)	15 (3)	38 (24)
1D	15 (6)	22 (7)	16 (8)	6 (2)	24 (1)	12 (5)	22 (10)
2D	26 (5)	35 (8)	37 (13)	4 (0)	10 (2)	33 (8)	42 (29)
3D	5 (5)	11 (6)	12 (5)	2 (2)	5 (6)	5 (5)	5 (5)
4D	16 (12)	22 (19)	15 (13)	31 (-)	10 (6)	19 (13)	19 (16)
5D	9 (0)	10)1)	8 (1)	6 (2)	8 (2)	3 (1)	13 (3)
6D	27 (1)	15 (1)	22 (1)	3 (0)	36 (4)	22 (1)	20 (2)
7D	45 (33)	40 (39)	27 (9)	0 (0)	35 (14)	44 (41)	46 (41)

Appendix IV A summary of various grain characteristics among 123 Kenyan, 63 Ethiopian, and 46 founder wheat lines

Variety/ Line	Source	Relative Puroindoline[a]	Crude protein (%)[b]	Kernel weight	Kernel diameter
1010AM2(L)	Kenya	0.91	15.74	52.54	3.15
1010F3SEL.13C	Kenya	0.92	16.68	37.09	2.80
1010F3SEL.4	Kenya	0.54	16.17	48.28	3.17
1010F3SEL.7	Kenya	-	-	37.70	2.76
1012B.1.(L)	Kenya	-	-	38.69	2.89
1016.P.2	Kenya	0.56	13.19	44.30	3.03
1016P.1	Kenya	0.67	16.38	39.73	2.67
1061.K.1	Kenya	1.24	15.84	31.83	2.24
1061.K.4	Kenya	0.86	14.07	43.76	2.73
1076.D.7	Kenya	0.69	16.18	46.25	2.98
1200.M	Kenya	-	-	33.37	2.51
291J.1.I.1	Kenya	-	16.42	43.42	3.03
321BT11B1	Kenya	0.12	15.89	45.03	3.02
338AA1A2	Kenya	0.19	14.76	35.93	2.59
688F4SEL3	Kenya	0.49	14.75	37.34	2.85
AfricaMayo	Kenya	0.88	14.00	42.33	2.90
Bailey	Kenya	0.60	14.15	41.18	3.02
Beacon-Ken	Kenya	1.05	15.31	33.63	2.28
BF236C1L	Kenya	1.10	15.99	45.06	2.86
Bonny	Kenya	1.01	14.64	39.84	2.89
Bonza63	Kenya	0.93	14.15	34.31	2.37
Bounty	Kenya	1.17	13.64	38.56	2.52
Brewster	Kenya	0.85	-	37.37	2.85
Catcher	Kenya	0.94	14.09	37.87	2.62
Chozi	Kenya	0.33	15.77	45.90	3.10
Duma	Kenya	0.16	14.43	47.78	3.19
Equator	Kenya	0.76	15.78	30.54	2.20
Equator1	Kenya	-	-	35.10	2.53
Fanfare	Kenya	-	-	38.49	2.52
FLIKenya9	Kenya	0.50	14.95	42.86	2.99
Fronthatch	Kenya	0.95	-	42.10	2.78
Fury	Kenya	0.30	16.62	46.57	3.09
Gabrino	Kenya	0.82	13.41	48.28	2.92
Gem	Kenya	1.04	14.44	45.51	2.95
Goblet	Kenya	0.71	14.11	45.60	2.92
H441	Kenya	0.89	-	37.36	2.74
Impala	Kenya	1.23	17.39	39.56	2.80
Inia66	Kenya	-	-	41.85	2.87
K-360-H	Kenya	-	-	37.63	2.64
KentanaYaqui	Kenya	0.69	-	47.68	2.96
Kenya-1	Kenya	0.56	15.12	43.40	2.92

[a] Puroindoline content is relative to Alpowa-1-soft

[b] Units are recorded in % protein converted from nitrogen x 5.25 and expressed on 12% moisture basis

Appendix IV *continued*

Variety/ Line	Source	Relative Puroindoline	Crude protein (%)	Kernel weight	Kernel diameter
Kenya117C	Kenya	1.43	12.81	40.90	2.84
Kenya-122	Kenya	0.39	13.00	41.39	2.84
Kenya131	Kenya	-	-	27.75	2.23
Kenya155	Kenya	0.92	17.48	44.49	2.76
Kenya-184-P	Kenya	0.11	16.05	40.92	3.08
Kenya-294-B-2A-3	Kenya	0.49	18.50	42.29	2.88
Kenya-318.O.3B.2	Kenya	0.72	13.19	36.13	2.69
Kenya-318-AJ-4A-1	Kenya	0.81	16.58	38.66	2.78
Kenya-358-AC	Kenya	0.35	16.66	43.64	3.10
Kenya-5	Kenya	0.92	14.12	43.98	2.99
KenyaCivet	Kenya	0.66	-	38.40	2.76
KenyaFahari	Kenya	1.16	15.89	47.19	3.02
KenyaFarmer.	Kenya	0.91	14.37	44.12	3.02
Kenya501	Kenya	0.75	14.76	44.44	2.91
Kenya-58	Kenya	0.86	13.56	34.79	2.50
Kenya6297-2	Kenya	-	-	39.54	2.76
Kenya6820	Kenya	1.32	16.13	42.47	2.96
Kenya7	Kenya	0.15	17.72	39.73	2.71
Kenya8	Kenya	0.03	16.38	36.67	2.76
KenyaB-256-G	Kenya	0.66	14.78	34.85	2.45
KenyaCheetah	Kenya	0.74	14.01	35.81	2.59
KenyaChiriku	Kenya	-	-	47.26	2.87
KenyaFL.1.158	Kenya	0.69	11.93	36.82	2.69
KenyaGovernor	Kenya	1.03	14.15	43.09	2.98
KenyaGrange	Kenya	0.69	16.96	45.76	3.00
KenyaHeroe	Kenya	0.08	11.59	49.80	3.11
KenyaHunter	Kenya	0.47	14.34	37.04	2.80
Kenya-Ibis	Kenya	1.16	13.57	36.83	2.66
KenyaJay	Kenya	0.42	13.96	36.35	2.80
Kenya-Kanga	Kenya	1.00	14.88	44.03	2.93
KenyaKifaru	Kenya	0.78	15.31	34.34	2.43
KenyaKongoni	Kenya	1.69	14.91	48.90	2.93
KenyaKudu	Kenya	0.71	17.30	37.72	2.58
KenyaLeopard	Kenya	0.77	17.63	42.07	2.82
KenyaMamba	Kenya	-	-	46.67	3.23
KenyaNgiri	Kenya	0.69	15.00	38.27	2.65
KenyaNungu	Kenya	0.89	14.98	42.52	2.75
KenyaNungu	Kenya	1.03	15.15	44.27	2.83
KenyaNyangumi	Kenya	0.41	16.54	41.37	2.89
Kenya-Nyati	Kenya	0.77	13.59	43.90	2.86
KenyaNyoka	Kenya	0.87	13.00	36.00	2.52
Kenya-Paa	Kenya	1.98	16.47	38.70	2.83
Kenyapage	Kenya	0.47	-	33.68	2.42
KenyaPaka	Kenya	1.04	14.35	42.29	2.73
KenyaPloughman	Kenya	0.73	11.69	37.06	2.68
KenyaPlume	Kenya	0.76	15.95	52.77	3.26
KenyaPopo	Kenya	1.31	14.60	40.89	2.77
KenyaStandard	Kenya	0.89	15.41	50.23	3.12

Appendix IV *continued*

Variety/ Line	Source	Relative Puroindoline	Crude protein (%)	Kernel weight	Kernel diameter
Kenya117C	Kenya	1.43	12.81	40.90	2.84
Kenya-122	Kenya	0.39	13.00	41.39	2.84
Kenya131	Kenya	-	-	27.75	2.23
Kenya155	Kenya	0.92	17.48	44.49	2.76
Kenya-184-P	Kenya	0.11	16.05	40.92	3.08
Kenya-294-B-2A-3	Kenya	0.49	18.50	42.29	2.88
Kenya-318.O.3B.2	Kenya	0.72	13.19	36.13	2.69
Kenya-318-AJ-4A-1	Kenya	0.81	16.58	38.66	2.78
Kenya-358-AC	Kenya	0.35	16.66	43.64	3.10
Kenya-5	Kenya	0.92	14.12	43.98	2.99
KenyaCivet	Kenya	0.66	-	38.40	2.76
KenyaFahari	Kenya	1.16	15.89	47.19	3.02
KenyaFarmer.	Kenya	0.91	14.37	44.12	3.02
Kenya501	Kenya	0.75	14.76	44.44	2.91
Kenya-58	Kenya	0.86	13.56	34.79	2.50
Kenya6297-2	Kenya	-	-	39.54	2.76
Kenya6820	Kenya	1.32	16.13	42.47	2.96
Kenya7	Kenya	0.15	17.72	39.73	2.71
Kenya8	Kenya	0.03	16.38	36.67	2.76
KenyaB-256-G	Kenya	0.66	14.78	34.85	2.45
KenyaCheetah	Kenya	0.74	14.01	35.81	2.59
KenyaChiriku	Kenya	-	-	47.26	2.87
KenyaFL.1.158	Kenya	0.69	11.93	36.82	2.69
KenyaGovernor	Kenya	1.03	14.15	43.09	2.98
KenyaGrange	Kenya	0.69	16.96	45.76	3.00
KenyaHeroe	Kenya	0.08	11.59	49.80	3.11
KenyaHunter	Kenya	0.47	14.34	37.04	2.80
Kenya-Ibis	Kenya	1.16	13.57	36.83	2.66
KenyaJay	Kenya	0.42	13.96	36.35	2.80
Kenya-Kanga	Kenya	1.00	14.88	44.03	2.93
KenyaKifaru	Kenya	0.78	15.31	34.34	2.43
KenyaKongoni	Kenya	1.69	14.91	48.90	2.93
KenyaKudu	Kenya	0.71	17.30	37.72	2.58
KenyaLeopard	Kenya	0.77	17.63	42.07	2.82
KenyaMamba	Kenya	-	-	46.67	3.23
KenyaNgiri	Kenya	0.69	15.00	38.27	2.65
KenyaNungu	Kenya	0.89	14.98	42.52	2.75
KenyaNungu	Kenya	1.03	15.15	44.27	2.83
KenyaNyangumi	Kenya	0.41	16.54	41.37	2.89
Kenya-Nyati	Kenya	0.77	13.59	43.90	2.86
KenyaNyoka	Kenya	0.87	13.00	36.00	2.52
Kenya-Paa	Kenya	1.98	16.47	38.70	2.83
Kenyapage	Kenya	0.47	-	33.68	2.42
KenyaPaka	Kenya	1.04	14.35	42.29	2.73
KenyaPloughman	Kenya	0.73	11.69	37.06	2.68
KenyaPlume	Kenya	0.76	15.95	52.77	3.26
KenyaPopo	Kenya	1.31	14.60	40.89	2.77
KenyaStandard	Kenya	0.89	15.41	50.23	3.12

Appendix IV *continued*

Variety/ Line	Source	Relative Puroindoline	Crude protein (%)	Kernel weight	Kernel diameter
KenyaSungura	Kenya	0.64	-	35.12	2.71
KenyaSwara	Kenya	0.61	15.87	42.68	2.88
KenyaTembo	Kenya	0.47	13.76	46.25	2.88
Kenya-Tumbili	Kenya	0.93	12.97	40.71	2.77
KenyaYombi	Kenya	0.14	13.26	46.43	3.02
KenyaZabadi	Kenya	0.27	13.69	43.05	2.88
K-Mbweha	Kenya	-	-	41.28	2.72
Kwale	Kenya	0.66	15.28	46.26	2.85
Lenana	Kenya	0.73	14.50	44.89	2.97
Mbega	Kenya	0.25	14.88	41.59	2.72
Mbuni	Kenya	-	14.20	50.47	3.09
Menco	Kenya	0.85	14.03	48.39	2.95
Mentor	Kenya	0.31	15.10	41.21	3.05
Morris	Kenya	1.09	17.10	32.36	2.50
Ngamia	Kenya	-	-	42.26	3.09
NjoroBwII	Kenya	0.53	17.54	44.85	2.85
P.WalkerMunro	Kenya	0.78	15.99	43.26	2.77
Pasa	Kenya	1.11	15.75	38.50	2.75
Pewter	Kenya	1.03	15.12	39.45	2.94
Pitcher	Kenya	0.32	14.71	42.80	2.83
Primex	Kenya	1.39	15.59	50.51	3.11
PWThatcher	Kenya	0.80	13.29	40.32	2.92
R64	Kenya	0.35	15.67	43.50	3.02
Regent	Kenya	0.63	-	36.05	2.69
Reliance	Kenya	0.50	14.22	42.11	2.90
Reliance261M	Kenya	0.14	13.40	46.31	2.99
Rhodesiansabanero	Kenya	1.47	14.10	42.85	2.87
Romany	Kenya	0.66	14.38	52.09	3.23
Sabanero	Kenya	0.43	14.56	39.83	2.75
Salmayo	Kenya	0.96	13.30	40.30	2.82
SantaElena	Kenya	1.05	16.97	49.67	3.17
Tama	Kenya	-	14.75	46.97	2.98
Token-Ken	Kenya	0.62	14.62	40.74	2.79
Trophy	Kenya	-	14.13	43.38	3.07
Abola	Ethiopia	0.96	14.81	46.03	2.86
AggiaSindi	Ethiopia	0.11	14.96	44.74	2.99
Arabe	Ethiopia	-	-	39.77	2.77
Batu	Ethiopia	0.10	14.15	44.64	2.90
Beltista	Ethiopia	0.81	18.57	48.92	3.00
Bobicho	Ethiopia	0.52	14.52	44.87	2.97
BTCBarce	Ethiopia	-	-	44.66	2.99
CI14393	Ethiopia	0.95	13.23	40.03	2.80
Dashen	Ethiopia	0.08	15.49	44.28	3.07
Dereseglen	Ethiopia	0.49	13.64	38.94	2.58
Digelu	Ethiopia	-	-	44.03	2.86
Diredawa	Ethiopia	0.72	13.99	40.25	2.68
Dodota	Ethiopia	0.16	12.96	43.73	2.91
Dure	Ethiopia	-	-	42.03	2.79
EnkoyE	Ethiopia	0.18	15.38	35.17	2.53
EnkoyU	Ethiopia	0.98	17.55	37.53	2.67

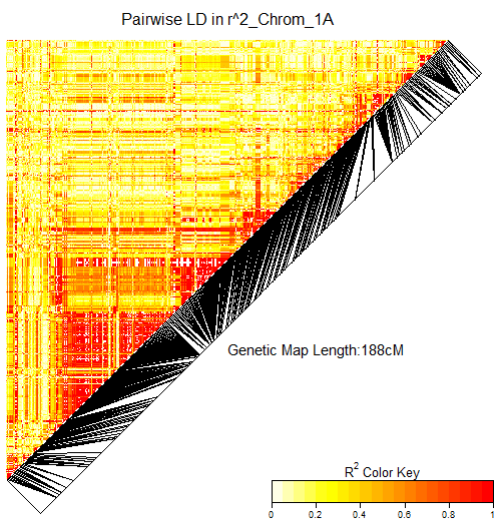
Appendix IV *continued*

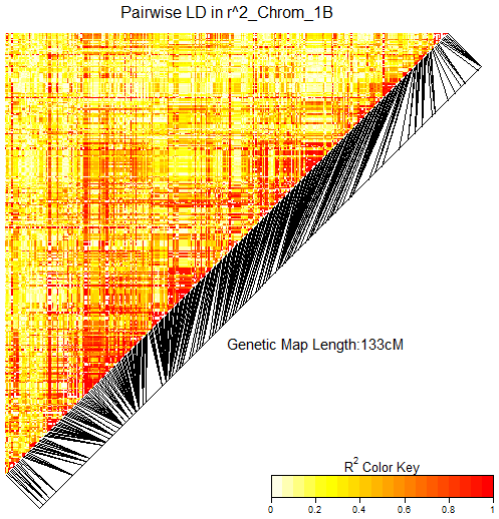
Variety/ Line	Source	Relative Puroindoline	Crude protein (%)	Kernel weight	Kernel diameter
ET-12-D4	Ethiopia	0.46	-	41.89	2.89
ET-13A2	Ethiopia	1.03	14.63	43.48	2.95
Galema	Ethiopia	0.41	12.61	41.26	2.90
Gara	Ethiopia	0.63	12.68	44.33	2.97
HAR1685	Ethiopia	0.33	15.30	46.94	3.09
HAR-407	Ethiopia	-	12.95	39.13	2.74
Hawi	Ethiopia	-	-	56.34	3.33
HRS55	Ethiopia	0.83	14.06	37.76	2.76
Jiru	Ethiopia	0.45	13.11	42.89	2.91
K-6106-8	Ethiopia	-	-	41.36	2.91
K6290Bulk	Ethiopia	1.16	14.51	42.63	2.83
K6295-4A	Ethiopia	0.58	14.44	41.53	2.94
Katar	Ethiopia	0.31	12.59	43.93	3.05
KBG-01	Ethiopia	1.24	16.30	37.78	2.66
KKBB	Ethiopia	0.31	12.07	40.35	2.64
Kubsa	Ethiopia	0.71	13.89	45.53	3.05
Laketch	Ethiopia	0.71	14.22	38.92	2.69
Maddawalabu	Ethiopia	0.56	13.59	54.39	3.28
Megal	Ethiopia	0.75	14.79	44.39	3.11
Meraro	Ethiopia	-	-	38.62	2.81
MG07761	Ethiopia	-	-	37.56	2.68
MG07766	Ethiopia	-	-	34.63	2.56
MG07768	Ethiopia	0.10	16.61	42.52	2.71
MG07780	Ethiopia	-	-	46.75	3.08
MG07782	Ethiopia	-	-	45.28	2.91
MG07793	Ethiopia	-	-	45.43	3.07
MG07795	Ethiopia	0.56	14.30	48.74	3.13
Mitike	Ethiopia	0.38	12.97	40.31	2.79
Pavon76	Ethiopia	0.45	14.39	40.92	2.81
Shina	Ethiopia	0.11	13.90	46.34	3.03
Simba	Ethiopia	0.20	14.04	48.21	3.15
Sindi 3	Ethiopia	0.76	18.44	46.20	2.90
Sindi1	Ethiopia	0.75	14.45	43.54	2.90
Sindi2	Ethiopia	0.22	17.60	47.14	2.95
Sindi4	Ethiopia	0.31	13.94	47.85	3.21
Sindi5	Ethiopia	0.27	17.85	42.68	2.93
Sindi6	Ethiopia	0.33	-	40.07	2.81
Sindi7	Ethiopia	1.01	16.27	45.31	2.83
Sindi8	Ethiopia	0.86	16.98	42.68	2.75
Sirbo	Ethiopia	0.17	14.42	35.71	2.77
Tura	Ethiopia	0.21	12.75	35.14	2.57
Tusie	Ethiopia	0.39	15.14	42.24	2.98
Wabe	Ethiopia	0.36	15.54	44.61	2.96

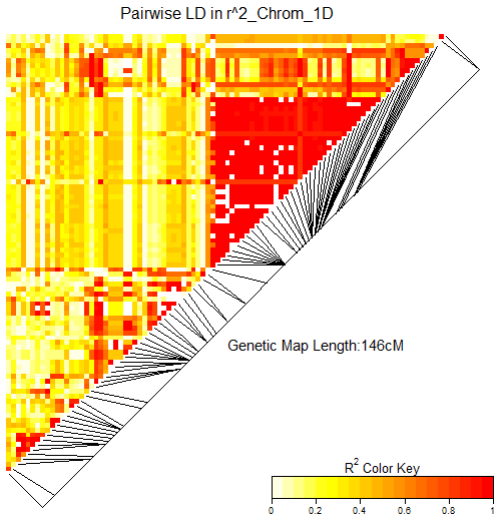
Appendix IV *continued*

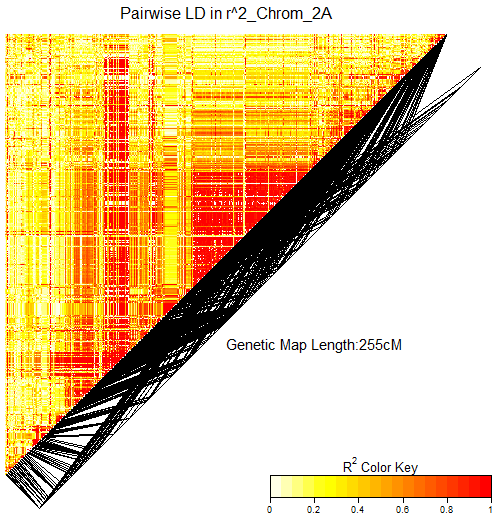
Variety/ Line	Source	Relative Puroindoline	Crude protein (%)	Kernel weight	Kernel diameter
Wetera	Ethiopia	0.34	15.72	46.63	3.05
184	Founder	0.79	13.42	39.91	2.69
2375	Founder	0.29	14.94	44.43	2.98
Aguilera	Founder	0.16	16.15	50.84	3.20
Bage	Founder	0.80	13.70	50.93	3.07
Bluebird	Founder	-	-	48.15	3.05
Bobin	Founder	0.56	11.13	55.88	3.46
Bobwhite	Founder	-	-	34.53	2.62
Bonanza	Founder	0.53	15.51	40.80	2.80
Bonito	Founder	0.69	14.89	44.43	2.88
Bonza	Founder	1.04	11.70	40.52	2.62
BW21	Founder	0.02	16.74	45.80	2.97
Cajeme71	Founder	-	-	43.00	2.94
Ceres	Founder	0.96	15.15	31.53	2.40
CianoF67	Founder	0.30	15.70	42.83	2.88
Cocoraque75	Founder	0.26	16.17	38.75	2.59
Exchange	Founder	0.77	14.56	40.99	2.72
Federation	Founder	-	-	41.61	2.85
Frocor2328	Founder	0.80	17.39	36.64	2.51
Frontana	Founder	-	-	35.77	2.44
Gabo	Founder	-	-	41.70	2.94
Giza155	Founder	0.62	14.52	47.56	3.05
Gradenero	Founder	1.09	14.93	42.85	2.90
II-50-17	Founder	0.81	16.49	44.25	2.94
Kalyanosona	Founder	0.53	14.05	43.64	2.91
Kentana48	Founder	-	-	46.52	2.88
Lermarojo	Founder	0.66	14.59	46.31	2.78
McMurachy	Founder	0.33	14.43	29.33	2.35
Mentana	Founder	-	-	40.72	2.54
Mida	Founder	0.96	16.52	46.20	3.06
Newthatch	Founder	-	-	32.98	2.62
Olescensdwarf	Founder	0.10	13.37	46.89	3.05
Penjamo62	Founder	-	-	46.38	2.96
RedEgyptian	Founder	-	13.89	38.14	2.84
Reward	Founder	0.77	16.67	38.66	2.86
Selkirk	Founder	0.55	15.22	49.77	3.05
Seri	Founder	0.45	15.81	40.57	2.97
Sonora	Founder	0.98	14.75	42.23	2.77
Sonora63	Founder	-	-	51.81	3.24
Supremo	Founder	0.69	17.75	48.70	2.97
Timstein	Founder	0.31	15.79	41.60	2.96
Tobari66	Founder	0.87	14.41	42.23	2.80
Wis.245	Founder	-	-	40.43	2.91
Yaktana54A	Founder	0.83	9.99	40.98	2.70
Yaqui50	Founder	0.34	13.32	40.66	2.81
Zaragoza75	Founder	1.09	13.53	40.27	2.90

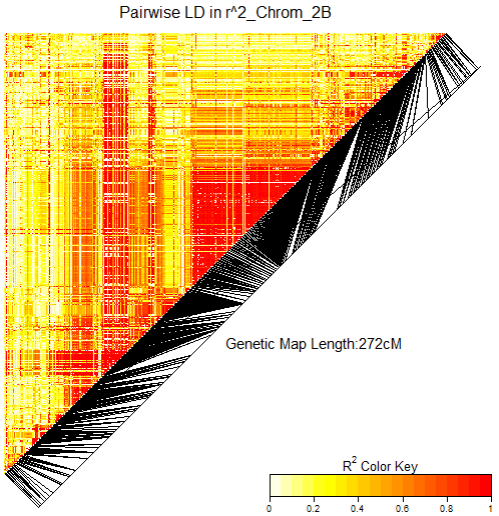
Appendix V Heatmaps of 21 wheat chromosomes indicating varition in intrachromosomal linkage disequlibrium

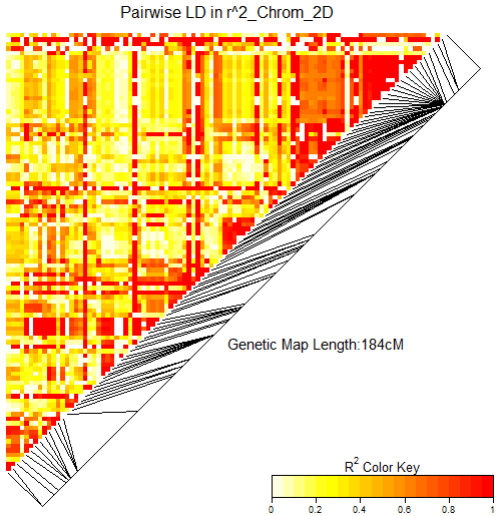




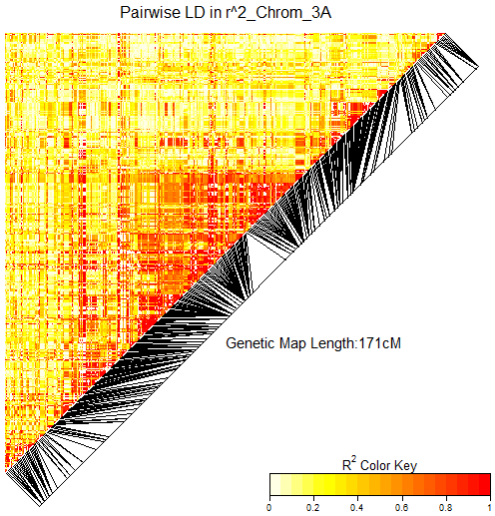


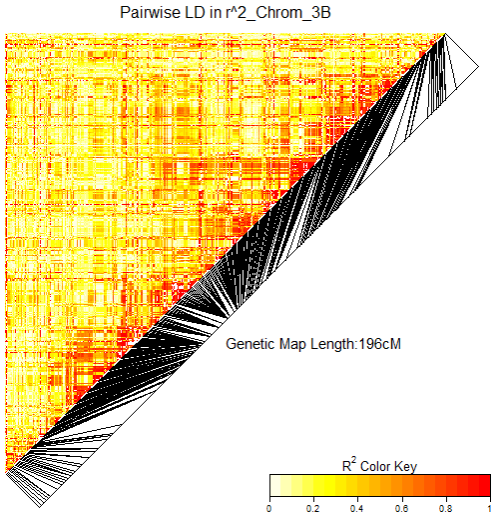


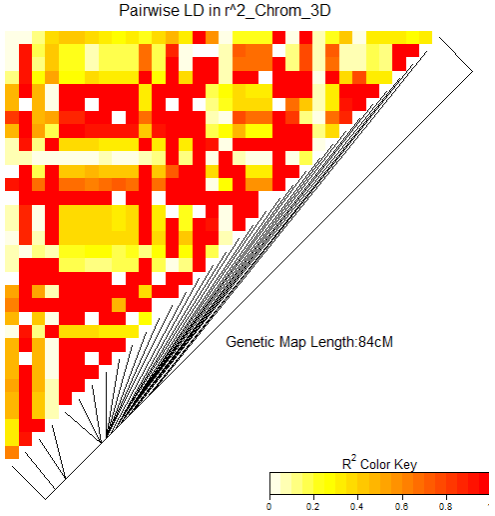


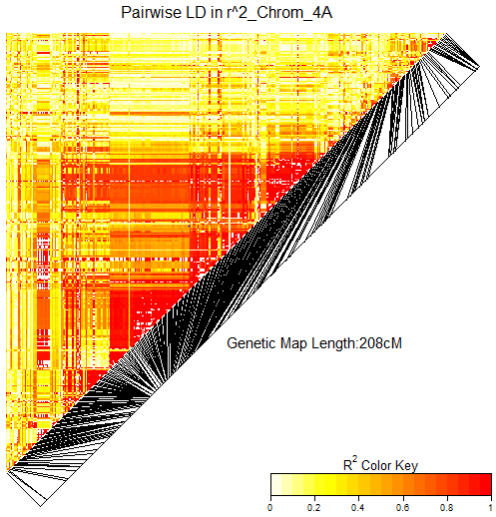


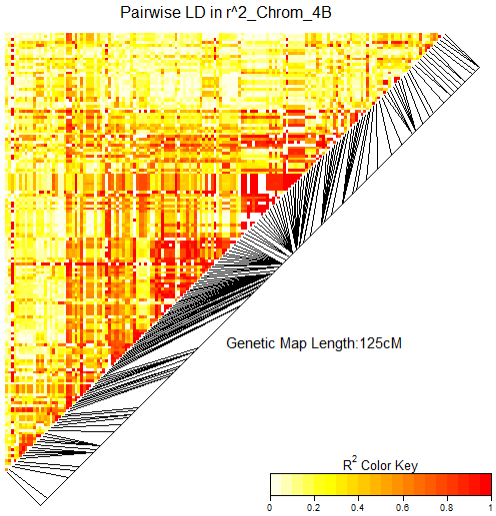
Appendix V *continued*

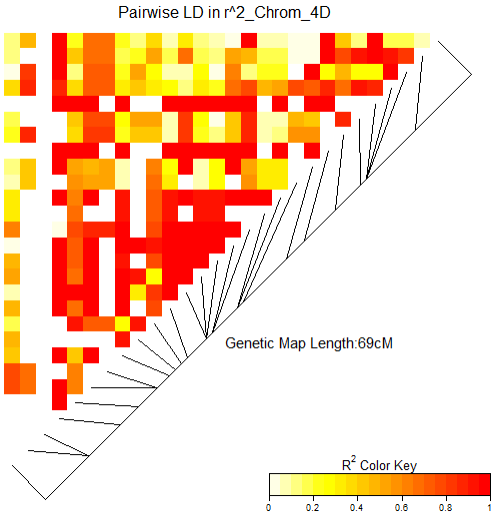




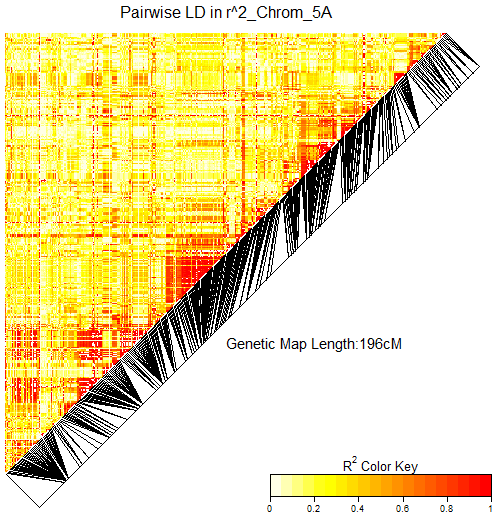


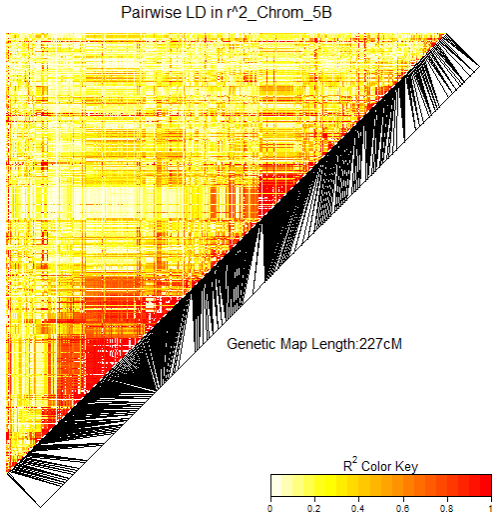


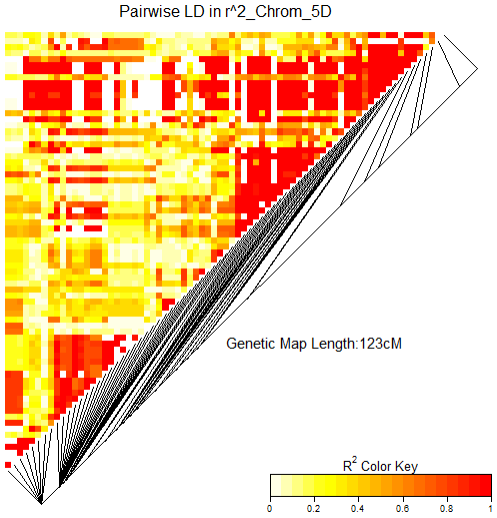


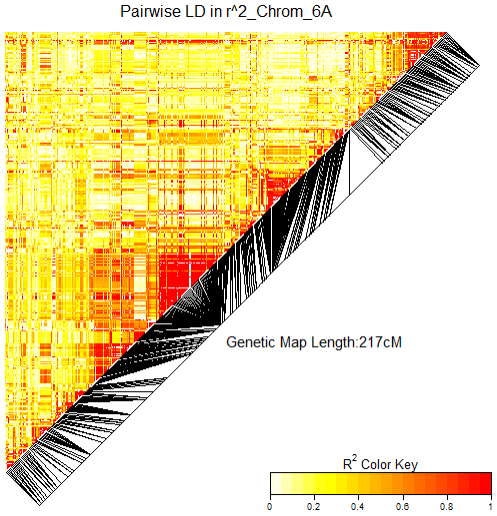


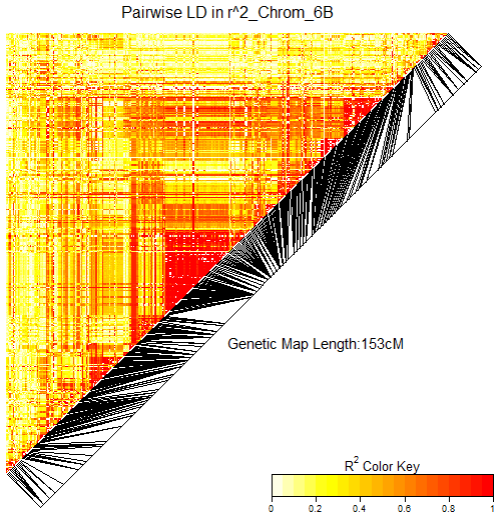
Appendix V *continued*

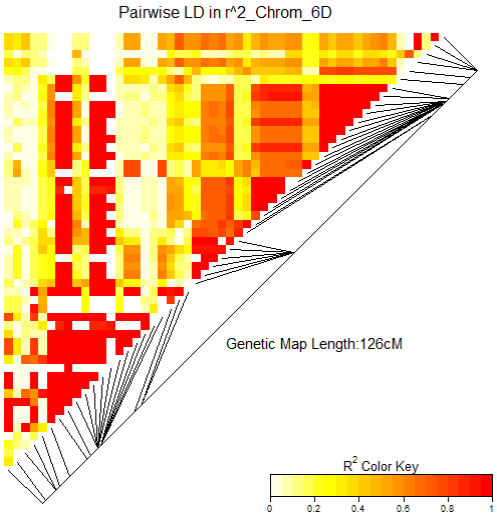












Appendix V *continued*