Theor Appl Genet (2014) 127:2453–2463
DOI 10.1007/s00122-014-2389-5

ORIGINAL PAPER

Linkage disequilibrium and association analysis of stripe rust resistance in wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*) population in Israel

Hanan Sela · Smadar Ezrati · Pnina Ben-Yehuda · Jacob Manisterski · Eduard Akhunov · Jan Dvorak · Adina Breiman · Abraham Korol

Received: 5 August 2013 / Accepted: 28 August 2014 / Published online: 16 September 2014

Abstract

Key Message **Rapid LD decay in wild emmer population from Israel allows high-resolution association mapping. Known and putative new stripe rust resistance genes were found.**

Abstract Genome-wide association mapping (GWAM) is becoming an important tool for the discovery and mapping of loci underlying trait variation in crops, but in the wild relatives of crops the use of GWAM has been limited. Critical factors for the use of GWAM are the levels of linkage disequilibrium (LD) and genetic diversity in mapped populations, particularly in those of self-pollinating species. Here, we report LD estimation in a population of 128 accessions of self-pollinating wild emmer, *Triticum turgidum* ssp. *dicoccoides*, the progenitor of cultivated wheat, collected in Israel. LD decayed fast along wild emmer chromosomes and reached the background level within 1 cM. We employed GWAM for the discovery and mapping of genes for resistance to three isolates of *Puccinia striiformis,* the causative agent of wheat stripe rust. The wild emmer population was genotyped with the wheat iSelect assay including 8643 gene-associated SNP markers (wheat 9K Infinium) of which 2,278 were polymorphic. The significance of association between stripe rust resistance and each of the polymorphic SNP was tested using mixed linear model implemented in EMMA software. The model produced satisfactory results and uncovered four significant associations on chromosome arms 1BS, 1BL and 3AL. The locus on 1BS was located in a region known to contain stripe rust resistance genes. These results show that GWAM is an effective strategy for gene discovery and mapping in wild emmer that will accelerate the utilization of this genetic resource in wheat breeding.

Communicated by Heiko C. Becker.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-014-2389-5) contains supplementary material, which is available to authorized users.

H. Sela (✉) · S. Ezrati · P. Ben-Yehuda · J. Manisterski · A. Breiman
The Institute for Cereal Crops Improvement, Tel-Aviv University, P.O. Box 39040, 69978 Tel Aviv, Israel
e-mail: hans@tauex.tau.ac.il

E. Akhunov
Department of Plant Pathology, Kansas State University, Manhattan, KS 66506, USA

J. Dvorak
Department of Plant Sciences, University of California, Davis, CA 95616, USA

A. Korol
Department of Evolutionary and Environmental Biology and Institute of Evolution, University of Haifa, 31905 Haifa, Israel

Introduction

Wheat is globally the most widespread crop and shares with rice the second place in annual grain production (FAO 2010; http://www.faostat.fao.org). Preventing wheat yield losses due to pest and disease outbreaks is therefore of global importance. Among fungal pathogens that infect wheat, stripe (yellow) rust caused by *Puccinia striiformis* f. sp. *tritici* (PST) has recently become a major threat to wheat production because of the emergence of new aggressive races, some adapted to warmer climates (Wellings 2011).

More than 60 genes for stripe rust resistance have been identified in cultivated wheat and its wild relatives.

Unfortunately, most of these genes have succumbed to stripe rust in at least some parts of the world (Hovmoeller et al. 2008; Milus et al. 2009; Chen et al. 2010). Devastating stripe rust epidemics have recently been reported from the Middle East, North America and Europe (Chen et al. 2010; Nazari et al. 2011). It is therefore essential that new genes for stripe rust resistance are discovered and introgressed into elite wheat germplasm to protect it against this devastating disease.

Wild emmer (*Triticum turgidum* ssp. *dicoccoides*, genomes AABB), the progenitor of cultivated wheat, is a valuable source of new resistance against wheat diseases (Nevo et al. 2002), including stripe rust. Hybrids between wild emmer and cultivated tetraploid (genomes AABB) and hexaploid (genomes AABBDD) wheat are highly fertile making it possible to introgress resistance genes from wild emmer into cultivated wheat using standard breeding techniques (Gerechter-Amitai et al. 1984, 1989; Grama et al. 1984; Peng et al. 1999; Marais et al. 2005; Fu et al. 2009; Cheng et al. 2010). Wild emmer therefore holds a unique position among wheat relatives as a source of new genetic variation for the improvement of disease and pest resistance in wheat.

Wild emmer is discontinuously distributed across the Fertile Crescent and the center of its genetic diversity is in Israel (Feldman and Kislev 2007; Luo et al. 2007). Wild emmer is genetically subdivided into two populations: north-eastern population (Iran, Iraq, northern Syria and Turkey) and south-western population (Israel, Jordan, Lebanon, and Southern Syria) (Luo et al. 2007; Özkan et al. 2011; Zohary and Hopf 2000). Genetic evidence suggests that emmer was domesticated in south-eastern Turkey with secondary contribution from wild emmer in Israel (Ozkan et al. 2005; Luo et al. 2007).Wild emmer in Israel may therefore contain stripe rust resistance genes that are not present in cultivated wheat. We therefore focused our search for new sources of stripe rust resistance on wild emmer in Israel.

The discovery of new genes for disease resistance and their efficient deployment in wheat breeding require their mapping relative to molecular markers. Association mapping (AM) is an efficient way to simultaneously discover and map genes for disease resistance and other traits (Ersoz et al. 2007; Waugh et al. 2009; Wang et al. 2011; Sajjad et al. 2012). AM is based on linkage disequilibrium (LD) between a marker and a causal locus. Because LD extent in wild populations depends on a long history of recombination, it is possible to obtain finer mapping resolution with AM than with linkage studies of bi-parental mapping populations. However, AM is not free of limitations. Its effectiveness depends on the magnitude of LD in the investigated population and on the availability of a large number of mapped markers. Another limitation is that associations between a trait and markers in a population may not necessarily reflect physical proximity because LD is affected by population structure, demographic events and selection (Flint-Garcia et al. 2003; Rostoks et al. 2006; Zhang et al. 2009; Rafalski 2010; Wang et al. 2011). Spurious associations can be minimized by applying mixed linear models in the association analysis using parameters of kinship matrix as random effects (Kang et al. 2008; Zhang et al. 2009).

Genome-wide association mapping (GWAM) in wild relatives of crops has been limited even though wild populations should be better suited for GWAM than crops. Wild progenitors of crops, such as wild emmer, usually harbor more genetic variation than their cultivated descendants (Salamini et al. 2002; Buckler and Thornsberry 2002; Haudry et al. 2007). Furthermore, LD decays more rapidly as a function of distance between loci in wild populations than in crops, especially in self-pollinating species, which increases mapping precision. For instance, in the self-pollinating *Glycine soja*, the wild progenitor of cultivated soybean *Glycine max*, LD r^2 values decayed to <0.1 within 100 kb, while in elite soybean cultivars the decay to $r^2 < 0.1$ was not observed even within 600 kb (Hyten et al. 2007). Similar situations may be present in the self-pollinating barley and wheat, members of the tribe Triticeae. In elite barley cultivars, LD extended over more than 200 kb but rapidly decayed over the first 1 kb in the wild forms (Caldwell et al. 2006). In cultivated durum wheat (*T. turgidum* ssp. *durum*), LD was reported to stretch for a long distance (Maccaferri et al. 2005), with high LD values found even between loci located 50 cM apart, and in bread wheat (*T. aestivum* ssp. *aestivum*) LD extent ranged from 1 to 30 cM in different studies (Chao et al. 2007, 2010 Somers et al. 2007; Crossa et al. 2007; Horvath et al. 2009). For comprehensive review, see Sajjad et al. (2012).

Here, we conducted a first genome-wide LD study in wild emmer using the recently developed wheat 9000-SNP iSelect assay (Cavanagh et al. 2013). We employed 2,278 SNP markers, polymorphic in a population of wild emmer in Israel, to investigate patterns of LD across the wild emmer genome and to evaluate the possibility of detecting SNP significantly associated with resistance to PST. Stripe rust resistance loci discovered here will be a valuable addition to genetic resources available to wheat breeders to improve this staple crop.

Methods

Plant material and stripe rust resistance assessment

A total of 128 accessions of wild emmer from 41 locations across Israel (see supplemental material for a full description of the accessions) were selected from the

collection of the Harold and Adele Lieberman Germplasm Bank, The Institute for Cereal Crops Improvement (ICCI), Tel-Aviv University. These accessions were evaluated for resistance to three isolates of PST: #5006, #5331, and #5341 (races 38E134, 6E0, and 134E150, respectively) (for race nomenclature, see Johnson et al. 1972). The stripe rust isolates were collected from bread wheat fields in Israel. Isolates #5006 and #5341 were chosen since they are virulent races representing the current stripe rust in Israel. They have three virulence differences between them. Isolate #5331 is less virulent than the other two on the cultivated differential sets; however, it is more virulent on wild emmer accessions (Cheng et al. 2010; supplementary on line data).

Three 1-week-old plants (replicates) per accession were inoculated with each of the three PST isolates. The test was conducted twice. The plants were inoculated by spraying PST spores (2 mg/ml) immersed in light mineral oil (Soltrol 170, Phillips Petroleum). The inoculated plants were left to dry at room temperature for 30 min and then kept in a dew chamber for 24 h at 100 % humidity: 12 h at 9 °C in the dark followed by 12 h day light at 15 °C. The plants were then grown at 70 % humidity under 12 h day/12 h night regime at 15 °C constant temperature. Disease response was scored for infection types (IT) 2 weeks after inoculation according to Qayoum and Line (1985). In case of discrepancies between replicates within a test or between the tests, the highest score was used.

Genotyping

All 128 accessions were genotyped with the Illumina 9K iSelect assay (Cavanagh et al. 2013). Ten accessions were genotyped twice for quality control. No differences were found between the duplicated results. SNP data were filtered to contain SNP with <10 % missing values and the minor allele frequency (MAF) >5 % (supplemental online material).

LD and association analysis

Information about the chromosomal location of the SNP was obtained from two maps: a map based on recombinant inbred lines (RIL) of a cross between *T. durum* cv Langdon (LDN) and *T. dicoccoides* accession PI 428082 (Jorgensen et al. 2012). The map includes 2325 SNP markers mapped at 897 loci. The parents of the RIL population differ by a reciprocal translocation between chromosomes 3B and 6B which affected the map of the two chromosomes. To overcome this limitation and increase the number of markers, an additional map was used in parallel based on the SynOp doubled haploid population (Saintenac et al. 2013). The SynOp map consists of 2258 SNP markers at 956 loci of the A and B genomes.

Unless stated otherwise, all of the statistical analyses and data manipulations were performed using R packages (R Core Team 2011). Principal coordinates analysis (PCO) was conducted using "ecodist" R package (Goslee and Urban 2007) to reveal the genetic relations between accessions and to unravel the population structure. Pairwise LD between SNPs was estimated by calculating r^2 and Chi-square p values as implemented in the "genetics" R package (Warnes et al. 2011).

We have also used the significance of association between SNPs in pairwise combinations where one of the SNP was recorded as a "phenotype" as a proxy of LD estimate. The significance of association was tested using the statistical models implemented in EMMA R package (Kang et al. 2008) and TASSEL (Bradbury et al. 2008). Linear mixed model association via likelihood ratio test (emma.ML.LRT; EMMA hereafter) implemented in EMMA takes into account the parameters of kinship matrix between genotypes as random effects and is expected to correct the effects of population structure that might cause spurious associations. The kinship matrix was calculated with EMMA and was estimated as a pairwise identical-by-state (IBS) allele-sharing matrix. The EMMA algorithm corrects p values for multiple testing.

The mixed linear model (MLM) implemented in TASSEL (Yu et al. 2006) was used with both the kinship matrix and the first three principal component analysis (PCA) vectors (both generated by TASSEL) as random and fixed effects, respectively, to correct for population structure. p values were adjusted for multiple tests by "p.adjust" R function using false discovery rate (FDR) method (Benjamini and Hochberg 1995; q values hereafter).

EMMA generated pairwise p values between SNP were regressed on the genetic distance between SNPs. Non-linear regression analysis was performed to fit the data using $-\log_{10} p = \exp(a + b \times d)$ as a proxy and employing "nls" function in R, where p is the p value of LD, d is the genetic distance between SNP sites and a and b are the regression estimates.

MLM as implemented in TASSEL and EMMA was used to calculate the association between SNP and IT.

The individuals in the association mapping panel were classified based on IT to resistant (IT = 1–3 score values) and susceptible (IT = 5–9 score values). Since the tests with the three isolates have shown almost identical results, the classification summarized the highest IT of all three isolates. The proportion of resistant plants was low (15.5 %), Therefore, the association analysis was conducted on a subset of 64 accessions where the proportion of resistant plants was 24 %. The subset was selected by plotting a neighbor-joining tree of the

Table 1 Responses of wild emmer accessions to isolates #5006, #5331, and #5341 of *Puccinia striiformis*

	Number (percent) of accessions for each infection type[a]					
Isolate	1	3	5	7	9	Total
#5006	16 (15.5)				87 (84.5)	103
#5331	8 (7.8)	6 (5.8)	4 (3.9)		84 (81.6)	103
#5341	11 (10.7)	8 (7.8)	6 (5.8)	2 (1.9)	76 (73.8)	103

[a] Infection types: resistant = 1–3, susceptible = 7–9, intermediate = 5, according to Qayoum and Line (1985)

accessions using the nj function in "ape" R package and selecting the cluster (clade) that contained all but one of the resistant accessions. The *p* values obtained from the association analysis were adjusted for multiple tests by two methods: (1) calculating *q* values. (2) Inferring the adjusted *p* value relative to empirical distribution of *p* values. The empirical distribution of *p* values was calculated by EMMA using the 2,050 polymorphic SNP in the northern *horanum* subset as phenotypes and as genotypes. Adjusted *p* value was estimated using the formula adp = *w*/2,100,225 where adp is the adjusted *p* value, w is the number of *p* values in the empirical distribution lower than the original *p* value and 2,100,225 is the number of *p* values in the empirical distribution (empirical *p* values hereafter).

Results

Stripe rust resistance

The 103 accessions of wild emmer *horanum* race from 35 locations representing the range of the race distribution in Israel were tested for resistance to PST isolates #5006, #5331, and #5341. Except for 3 accessions, the classification of the accessions to resistant (IT $\leq$ 3) and susceptible (IT $\geq$ 5) classes was the same for the three isolates. The reactions to isolate #5006 were more extreme both on the resistant and susceptible side, while the reactions to isolates #5331 and #5341 showed the tendency to be intermediate (Table 1). Overall, 15.5 % of the accessions were classified as resistant to PST (IT $\leq$ 3) and the rest were classified as susceptible (IT $\geq$ 5). Most of the resistant accessions were from northern Israel (Fig. 1).

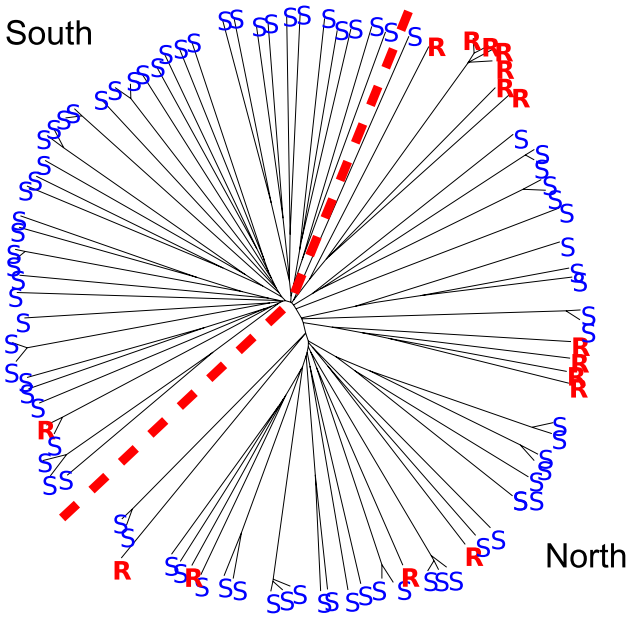

Fig. 1 Neighbor-joining tree of 103 genotypes of the *horanum race*. *R* resistant genotypes, *S* susceptible genotypes. *Dashed line* separates between the north cluster that was used in the GWAM and the south cluster

LD analysis

All 128 wild emmer accessions were genotyped with the wheat 9K Infinium assay. Four accessions which had an excess of heterozygous SNP loci were removed, leaving 124 accessions for LD analysis. To investigate LD patterns and to test the ability of the association mapping methods to correct for population structure, a matrix of LD *p* values obtained with each of the three methods: standard LD, EMMA, and TASSEL-MLM was constructed for all pairs of 950 mapped SNP markers (Fig. 2). Each method produced low *p* value LD both between SNP on a single chromosome (intra-chromosomal LD) and between SNP on different chromosomes (inter-chromosomal LD) (Fig. 2). The highest number of inter-chromosomal LD was produced by the Standard LD method (Fig. 2, top), while EMMA and TASSEL-MLM showed much lower numbers (Fig. 2, middle).

To assess the structure in the investigated population and its effect on inter-chromosomal LD, a principal coordinates (PCO) graph based on the genetic distances between wild emmer accessions was constructed (Fig. 3). The first and second vectors accounted for 32 and 4 % of diversity, respectively. The PCO analysis revealed a clear population

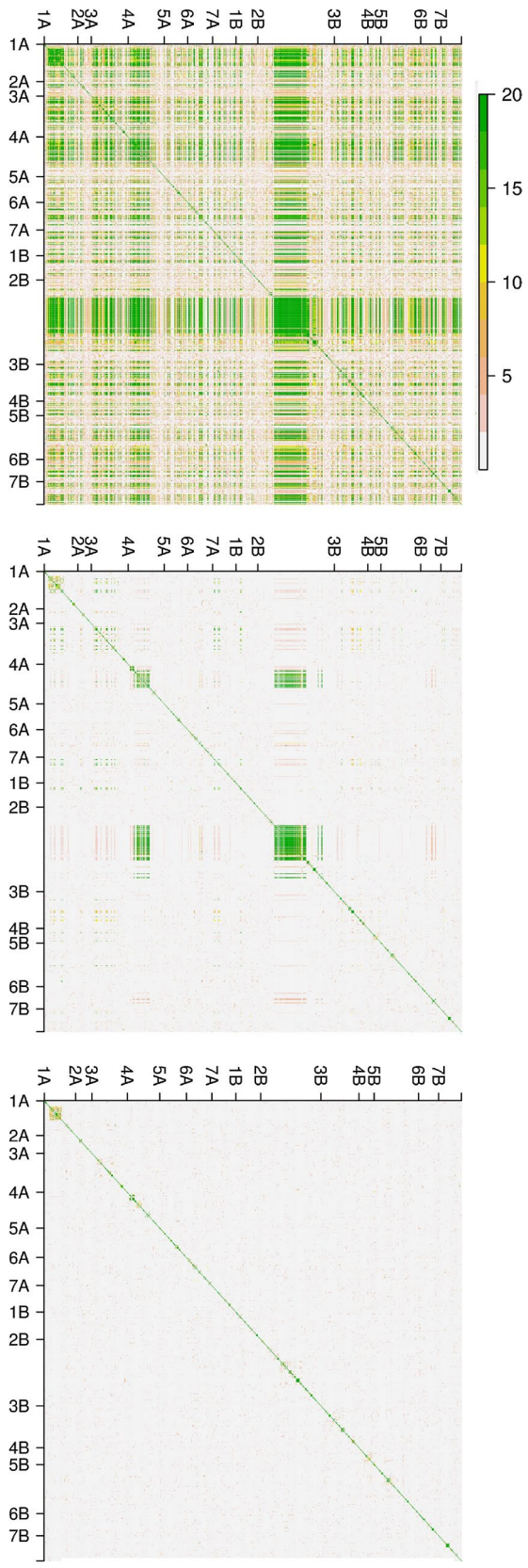

Fig. 2 Plots of $-\log_{10} p$ values of LD between SNP across the wild emmer genome. Chromosomal locations (*1A* through *7B*) of SNP are at the *top* and *left* of the plots. The *top* plot, (*judaicum* + *hornum* set) shows standard LD Chi-square $-\log_{10} p$ values. The *middle* plot (*judaicum* + *hornum* set) shows $-\log_{10} p$ values of associations computed with EMMA software with correction for population structure. The *bottom* plot, (*horanum* subset) shows $-\log_{10} p$ values of associations computed with EMMA software with correction for population structure. *Colors* from *gray* to *green* show increasing $-\log_{10} p$ values (see *color scale* at the *right* of plots) *Green dots* located away from the diagonal represent spurious associations. The scale of the chromosomes reflects the number of polymorphic SNP per chromosome

structure resulting in two clusters: a major cluster including 103 accessions (80 %) and a minor cluster including the remaining 21 (20 %) accessions. The minor cluster consisted of race *judaicum* as we could judge based on the collection sites of the accessions around the Lake of Galilee and their large grain sizes. The major cluster consisted of the *horanum* race (Poyarkova 1988; Poyarkova et al. 1991).

To check whether population structure was the cause of the spurious associations, the *judaicum* cluster was removed from the analysis. The removal of these accessions resulted in the removal of 180 (19 %) of the 950 mapped SNP markers that were not polymorphic in the remaining accessions. The removed SNP markers were not evenly distributed across the wild emmer genome: 36 % of the SNP from chromosome 2B were removed, while no SNP was removed from chromosome 1A (Table 2). Compared to the entire dataset, the LD matrix of the remaining subset of race *horanum* accessions showed greatly reduced inter-chromosomal spurious associations. The significant associations were mostly located along the diagonal of the LD plot (Fig. 2, bottom) and were between adjacent SNP markers. The LD p values decayed rapidly within 1 cM or less, and only small LD blocks were observed (Fig. 4). The largest block consisted of 15 SNP from the same locus on the SynOp map (Fig. 2, bottom).

The best fit for the correlation between the LD p values produced by EMMA and genetic distances is given by the function $-\log_{10} p = \exp(a + b \times d)$, where p is the p value of LD, d is the genetic distance in cM between the SNP, and a and b are the regression estimates ($r^2 = 0.27$).

Genome-wide association mapping

Because only 15.5 % of the accessions were resistant, and all but one were from the northern region, a subset of 64 accessions from the northern region was selected. The selection was done by clustering the accessions on neighbor-joining tree and selecting the clade that included all but one of the resistant accessions (Fig. 1). The subset had a

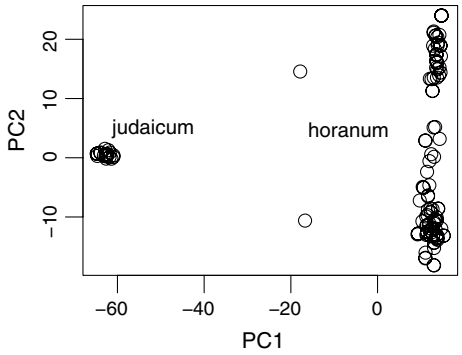

Fig. 3 Principal coordinates analysis (PCO) based on the genetic distances between genotypes

more balanced ratio of resistant and susceptible accessions (1/3). Furthermore, the selection has removed a major clade with almost no resistant genotypes that would have caused a bias in the AM because of population structure effect. Out of 8643 SNP in the assay, 2050 were polymorphic in the northern *horanum* subset (MAF > 5 %). Association analysis was conducted between the polymorphic SNP and IT of the 64 *horanum* accessions. Of these SNP, 1,168 were mapped on the RIL map (Jorgensen et al. 2012) or on the SynOP map (Saintenac et al. 2013). Significant marker trait associations (MTA) were observed (Table 3; Fig. 5). EMMA and TASSEL-MLM agreed on the five most significant MTA: SNP #2578 and #3169, on the same locus on chromosome arm 1BS, two SNP #1982 and #7817 on the

Table 2 Total numbers of SNP per chromosome found in the set of all 124 analyzed wild emmer accessions and the *horanum* subset after the removal of the *judaicum* accessions and mean map distance between SNP based on SynOP genetic map

Chr.	Chr. length cM	Whole set		*Horanum* subset		
		No. of SNP	Average distance between SNP cM	No. of SNP	Average distance between SNP cM	Percent of removed SNP[a]
1A	157	77	2	58	2.4	25
1B	145	50	2.9	46	2.9	8
2A	163	30	5.4	30	5.4	0
2B	217	174	1.2	111	2	36
3A	197	84	2.3	65	3.1	23
3B	163	76	2.1	70	2.7	8
4A	183	82	2.2	60	2.8	27
4B	109	30	3.6	28	3.6	7
5A	188	53	3.5	49	3.5	8
5B	264	90	2.9	82	3.1	9
6A	148	57	2.6	47	3	18
6B	169	46	3.7	37	4.4	20
7A	197	53	3.7	44	4.5	17
7B	166	48	3.5	43	3.8	10

[a] The percent of SNP in the chromosome that were polymorphic in the whole set but not in the *horanum* subset

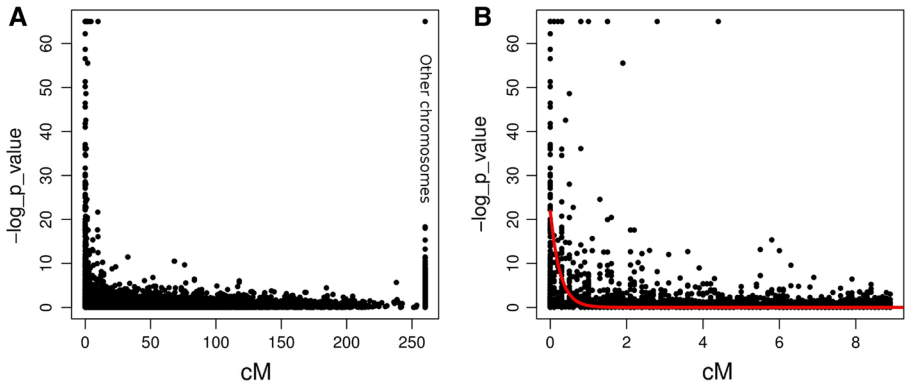

Fig. 4 Plot of $-\log_{10} p$ values of LD as a function of the genetic distance between SNP. *p* values were calculated by EMMA. **a** Plot of all values. **b** Zoomed into 10 cM scale, *red line* represents the regression line ($r^2 = 0.27$). Regression formula is $-\log_{10} p = \exp(a + b \times d)$, where *p* is the *p* value of LD and *d* is the genetic distance between the SNP and *a* and *b* are the regression coefficients

Table 3 SNP associated with stripe rust resistance

	SNP index[a]	2041	2578	3169	1982/7817[e]
	Chromosome arm	1BL	1BS	1BS	3AL
	Chromosomal location cM[b]	84	57	57	88
	Odds ratio[c]	0	131	∞	24
	MAF %	14	28	40	16
EMMA	*p* value	1.1e−7	1.2e−12	1.3e−12	2.2e−8
	q value	4.7e−5	1.3e−9	1.3e−9	1.1e−5
	Empirical *p* value	0.0015	0.001	0.001	0.0014
TASSEL	*p* value	2.7e−5	5.8e−7	6.1e−7	9.0e−6
	q value	0.01	0.0006	0.0006	0.004
	r^{2d}	0.34	0.52	0.51	0.39

[a] See supplemental online material for SNP names

[b] Location of the SNP in the 90k consensus map (http://129.130.90.211/snp/)

[c] Odds ratio was calculated from 2 × 2 contingency table of the two SNP alleles and resistance/susceptibility scores

[d] r^2 for the marker is calculated based on a formula for r^2 for a generalized least squares GLS model

[e] The two markers are in complete LD

same locus on 3AL and SNP #2041 on 1BL. TASSEL *p* values were higher than EMMA *p* values. Since EMMA produced lower *q* values, a second method of adjustment was implemented. An empirical *p* values were calculated as the fraction of all *p* values of the empirical distribution lower than each of the original *p* values. In this way, *p* values were adjusted relative to empirical *p* values calculated from the same population. The empirical *p* values for the first five MTA of EMMA were all significant ($p < 0.002$). The five best MTA from EMMA and TASSEL-MLM are summarized in Table 3. The first two MTA on 1BS have the same location on the genetic maps. They are located ~1.5 kb apart on the same contig in the survey sequence data base (http://wheat-urgi.versailles.inra.fr/Seq-Repository). However, they are not in complete LD ($r^2 = 0.75$, $p = 2.8e{-}13$). The locus on 1BL (#2014) is in medium LD with the loci on 1BS ($r^2 = 0.5$, $p = 0.0009$). While the *p* value of its MTA is lower than that of the MTA on 1BS, the odds ratio of the 1BL MTA (#2041) is 0, meaning that no susceptible genotype has allele A. The odds ratio of SNP #3169 is infinity, meaning that all genotypes with allele A are susceptible.

Discussion

Stripe rust resistance and its association mapping

A total of 15.5 % of 103 accessions from 35 locations in Israel were resistant to PST infection, indicating that wild

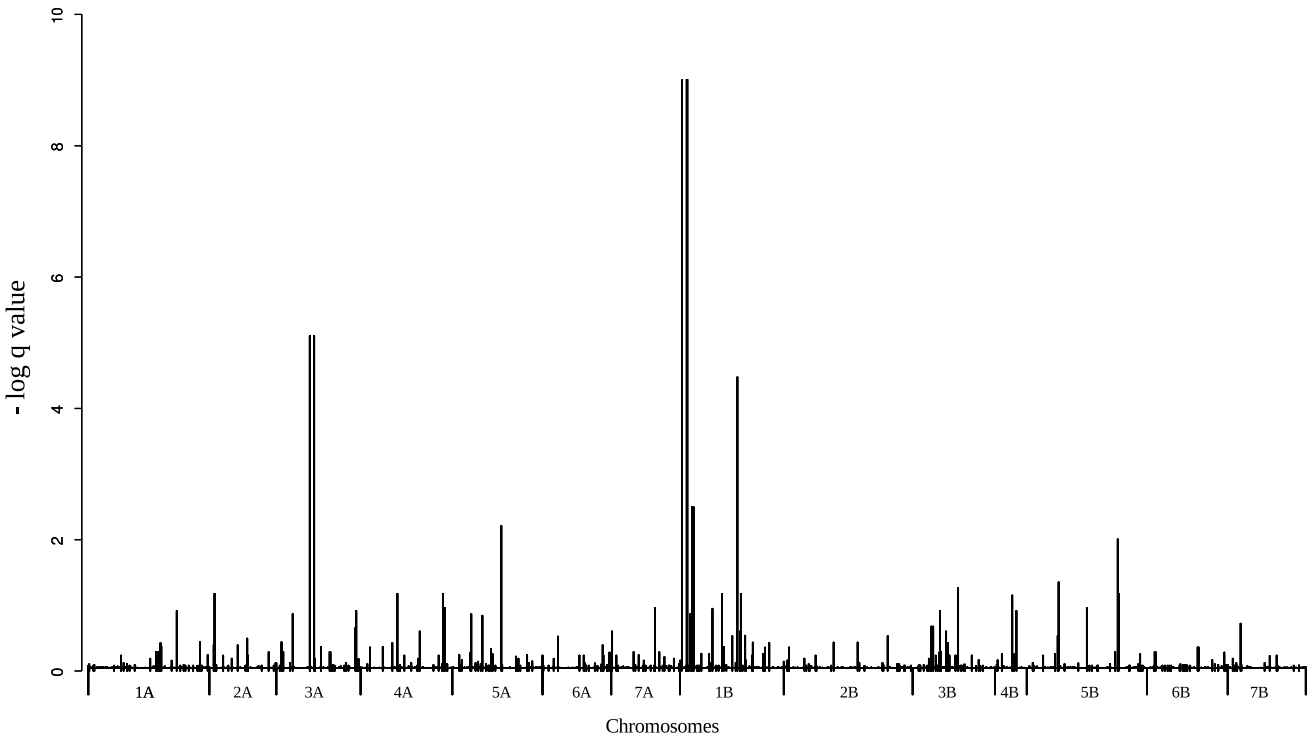

Fig. 5 Genome wide association analysis of response to stripe rust isolates. *p* values were calculated using EMMA. *Y* axis, the $-\log_{10} q$ values of the association. *X* axis, SNP ordered along the chromosomes

emmer in Israel is a valuable source of PST resistance. More resistant accessions were found in the northern part of Israel than in the southern part. This distribution of resistance coincides with more favorable environmental conditions for stripe rust development in northern Israel than in the southern Israel (Cheng et al. 2010).

Most of the responses to PST were extreme, resulting in either high or low IT, which implies the involvement of major *R* genes. Therefore, the association analysis was focused on the extreme responses aiming at mapping major *R* genes.

SNP markers on chromosome arms 1BS, 1BL, and 3AL were significantly associated with stripe rust resistance in the wild emmer race *horanum* in Israel. Resistance locus detected on 1BS may coincide with the location of *Yr15* or *YrH52* that are derived from wild emmer, while the loci on 1BL and 3AL, do not coincide with any known wheat *Yr* gene, suggesting that the association analysis may have detected novel *Yr* resistance gene (Peng et al. 2000; Mcintosh et al. 2013). However, more genotypes from the northern populations are needed to validate these two novel loci. It is possible that the locus on 1BL has a high MTA because of its proximity to the 1BS locus and not because it is a resistance gene locus.

The distribution of PST resistance was uneven; resistance was concentrated in few populations, particularly those from the Mt. Hermon region and from the Safed region. The accessions from these regions that were included in the GWAM resulted in rediscovery of known PST resistance genes. In search for new resistance genes, more thorough screening of wild emmer is needed. To find new resistance loci, we recommend to screen and genotype more accessions out of the two above-mentioned regions.

GWAM employing TASSEL-MLM and from EMMA pointed to the same set of markers but the TASSEL *p* values were higher than those of EMMA. The major difference between the two algorithms is that TASSEL employs the kinship matrix and the first 3 PC to correct for population structure, while EMMA employs only the kinship matrix. Since the level of *p* values obtained from the two algorithms differed, setting an a priori threshold of *p* values at 0.05 or any other value may lead to low detection power or high FDR. It is therefore better to report *p* values relative to the distribution of *p* values either with the FDR adjustment or by comparing them to an empirical distribution of *p* values calculated from the same population.

LD estimation

The two compared methods of compensation for population structure in association analysis, TASSEL and EMMA, have removed most of the spurious associations. Nevertheless, they could not correct all of them in the highly structured wild emmer population unless the *judaicum* race cluster was excluded. Wang et al. (2011) reported that EMMA analysis using only kinship matrix was the best method in association analysis of barley cultivars in terms of low number of false positives and high detection power. However, in our study, both EMMA and TASSEL generated spurious associations between chromosomes which were eliminated only by analyzing the *horanum race* separately from the *judaicum* race. We, therefore, recommend constructing a LD matrix using the SNP as "traits" in association analysis to check whether the employed algorithm performs well in a specific population, and only then to conduct association analysis of the target trait(s). This should be done even when using algorithms assumed capable to correct for population structure. When used properly, the algorithms should not generate many significant interchromosomal associations.

Even though wild emmer is a highly self-pollinating species (Golenberg 1988), we observed a rapid LD decay, reaching a background level within 1 cM. The rapid LD decay observed in wild emmer is similar to that reported for natural populations of other self-pollinating plants such as *Arabidopsis* and wild barley (Morrell et al. 2005; Nordborg et al. 2002; Caldwell et al. 2006; Kim et al. 2007). Compared to cultivated wheat, the LD range in wild emmer is shorter (Maccaferri et al. 2005; Crossa et al. 2007; Cavanagh et al. 2013). This difference is most likely caused by recent population bottlenecks and artificial selection accompanying wheat domestication and improvement in recent history of cultivated wheat (Cavanagh et al. 2013). While short LD range is useful for precise mapping of alleles underlying phenotypic trait variation in association mapping studies, it requires much greater SNP density for covering the entire genome. Kim et al. (2007) estimated that ~140,000 SNP were needed to cover the *A. thaliana* genome. The genomic resources being developed for wheat (Brenchley et al. 2012; Feuillet et al. 2012) will accelerate the discovery of new wheat SNP and the development of denser SNP arrays for more detailed LD analyses in wheat. Additionally, SNP discovery in wild emmer may have to be undertaken to get a comprehensive coverage of the wild emmer genome.

The *judaicum–horanum* race separation

The wild emmer population in Israel showed a strong subdivision along the *judaicum* and *horanum* morphological classification. The origin of the *judaicum* race and the reason for its genetic differentiation from the *horanum* race remain enigmatic. The gene flow between the two races is very limited, even though they grow side by side in some locations (Poyarkova et al. 1991). It was suggested that race *judaicum* originated from hybridization of wild emmer

with durum followed by selection for wild emmer traits (Blumler 1997). Genetic diversity of the *judaicum* population is reduced compared to that of the *horanum* race (Luo et al. 2007) and genetic differentiation of the two races is concentrated in specific regions of the genome, particularly on chromosomes 2B and 4A. Both chromosomes play important roles in differentiation of wild and domesticated wheat (Dvorak et al. 2006, 2012). Chromosome 5A harbors the *Q* gene essential for free-threshing habit in durum and bread wheat but was not enriched for inter-chromosomal LD. If race *judaicum* originated from hybridization of wild emmer with cultivated wheat, our data suggest that the cultivated parent was wheat without the *Q* gene. This is consistent with the failure to find genetic proximity between race *judaicum* and *T. durum* (Luo et al. 2007).

Concluding remarks

Association mapping in wild populations is expected to have superior resolution compared to association mapping in cultivars because of greater genetic diversity usually harbored in wild populations and shorter LD. Our success with GWAM of stripe rust resistance genes is consistent with these assumptions. We showed that spurious associations resulting from population structure can be detected and resolved. The association mapping of stripe rust resistance in wild emmer resulted in the discovery of three putative resistance loci. With the increasing threat of stripe rust to global wheat crop (http://www.globalrust.org/traction), these loci should be validated in biparental populations and tests for allelism with known PST loci in vicinity, and if different from them, incorporated into wheat breeding programs.

Authors contributions AB, AK, EA and JD conceived and managed the project. SE, PB, and JM performed the resistance tests and the plant propagation. EA conducted the genotyping. HS performed the statistical analysis and wrote the manuscript. AK advised about the statistical analysis. AK, EA, JD and HS reviewed and edited the manuscript.

Acknowledgments This work was supported by grant No IS-4137-08 from the US-Israel BARD foundation, grants 2009-65300-05638 and 2011-68002-30029 (Triticeae-CAP) from the USDA National Institute of Food and Agriculture.

Conflict of interest The authors declare no conflict of interest.

Ethical standards The authors declare that ethical standards are met, and all the experiments comply with the current laws of the country in which they were performed.

References

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 57:289–300

Blumler MA (1997). Introgression of durum into wild emmer and the agricultural origin question. In: Damania AB, Valkoun J, Willcox G, Qualset CO (eds) The origins of agriculture and crop domestication. ICARDA, IPGRI, FAO and UC/GRCP, pp 252–268

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2008) TASSEL: software for association mapping of complex trait in diverse samples. Bioinformatics 23:2633–2635

Brenchley R, Spannagl M, Pfeifer M, Barker GLA, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705–710

Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5:107–111

Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. Genetics 172:557–567

Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S et al (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc Nat Acad Sci 110:8057–8062

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

Chao S, Dubcovsky J, Dvorak J, Luo MC, Baenziger SP, Matnyazov R, Clark DR, Talbert LE, Anderson JA, Dreisigacker S (2010) Population-and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). BMC Genomics 11:727

Chen X, Penman L, Wan A, Cheng P (2010) Virulence races of *Puccinia striiformis* f. sp. *tritici* in 2006 and 2007 and development of wheat stripe rust and distributions, dynamics, and evolutionary relationships of races from 2000 to 2007 in the United States. Can J Plant Pathol 32:315–333

Cheng J, Yan J, Sela H, Manisterski J, Lewinsohn D, Nevo E, Fahima T (2010) Pathogen race determines the type of resistance response in the stripe rust-*Triticum dicoccoides* pathosystem. Physiol Plant 139:269–279

Crossa J, Burgueño J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M, Singh RP, Trethowan R, Warburton M, Franco J (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 177:1889–1913

Dvorak J, Akhunov ED, Akhunov AR, Deal KR, Luo MC (2006) Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. Mol Biol Evol 23:1386–1396

Dvorak J, Deal KR, Luo MC, You FM, von Borstel K, Dehghani H (2012) The origin of spelt and free-threshing hexaploid wheat. J Hered 103(3):426–441

Ersoz ES, Yu J, Buckler ES (2007) Applications of linkage disequilibrium and association mapping in crop plants. In: Genomics-assisted crop improvement. Springer, Dordrecht, The Netherlands, pp 97–120

Feldman M, Kislev ME (2007) Domestication of emmer wheat and evolution of free-threshing tetraploid wheat. Isr J Plant Sci 55:207–221

Feuillet C, Stein N, Rossini L, Praud S, Mayer K, Schulman A, Eversole K, Appels R (2012) Integrating cereal genomics to support innovation in the triticeae. Funct Integr Genomics 12:573–583

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

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

Gerechter-Amitai ZK, Sharp EL, Reinhold M (1984) Temperature-sensitive genes for resistance to *Puccinia Striiformis* in *Triticum dicoccoides*. Euphytica 33:665–672

Gerechter-Amitai Z, Silfhout CH, Grama A, Kleitman F (1989) *Yr*15—a new gene for resistance to *Puccinia Striiformis* in *Triticum dicoccoides* sel. Euphytica 43:187–190

Golenberg E (1988) Outcrossing rates and their relationship to phenology in *Triticum dicoccoides*. Theor Appl Genet 75:937–944

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. J Stat Softw 22:1–19

Grama A, Gerechter-Amitai ZK, Silfhout CH (1984) Additive gene action for resistance to *Puccinia Striiformis* f. sp. *tritici* in *Triticum dicoccoides*. Euphytica 33:281–287

Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glemin S (2007) Grinding up wheat: a massive loss of nucleotide diversity since domestication. Mol Biol Evol 24:1506–1517

Horvath A, Didier A, Koenig J, Exbrayat F, Charmet G, Balfourier F (2009) Analysis of diversity and linkage disequilibrium along chromosome 3B of bread wheat (*Triticum aestivum* L.). Theor Appl Genet 119:1523–1537

Hovmoeller MS, Yahyaoui AH, Milus EA, Justesen AF (2008) Rapid global spread of two aggressive strains of a wheat rust fungus. Mol Ecol 17:3818–3826

Hyten DL, Choi I, Song Q, Shoemaker RC, Nelson RL, Costa JM, Specht JE, Cregan PB (2007) Highly variable patterns of linkage disequilibrium in multiple soybean populations. Genetics 175:1937–1944

Johnson R, Stubbs RW, Fuchs E, Chamberlain NH (1972) Nomenclature for physiologic races of *Puccinia Striiformis* infecting wheat. Trans Br Mycol Soc 58:475–480

Jorgensen C, Distelfeld A, Luo M, Korol A, Dvorak J (2012) Genetic dissection of the domestication syndrome in tetraploid wheat. Plant and animal genome XX, San Diego, CA, USA

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 39:1151–1155

Luo MC, Yang ZL, You FM, Kawahara T, Waines JG, Dvorak J (2007) The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theor Appl Genet 114:947–959

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15:271–290

Marais GF, Pretorius ZA, Wellings CR, McCallum B, Marais AS (2005) Leaf rust and stripe rust resistance genes transferred to common wheat from *Triticum dicoccoides*. Euphytica 143:115–123

McIntosh RA, Yamazaki Y, Dubcovsky J, Rogers J, Morris C, Appels R and Xia XC (2013) Catalogue of gene symbols for wheat. In: 12th International Wheat Genetics Symposium, Yokohama

Milus EA, Kristensen K, Hovmøller MS (2009) Evidence for increased aggressiveness in a recent widespread strain of *Puccinia Striiformis* f. sp. *tritici* causing stripe rust of wheat. Phytopathology 99:89–94

Morrell PL, Toleno DM, Lundy KE, Clegg MT (2005) Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. Proc Nat Acad Sci 102:2442–2447

Nazari K, Hodson D, and Hovmoller K (2011) Yellow rust in CWANA 2010–2011, BGRI technical workshop, Minnisota

Nevo E, Korol AB, Beiles A, Fahima T (2002) Evolution of wild emmer and wheat improvement: population genetics, genetic resources, and genome organization of wheat's progenitor, *Triticum dicoccoides*. Springer, Berlin

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J et al (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 30:190–193

Ozkan H, Brandolini A, Pozzi C, Effgen S, Wunder J, Salamini F (2005) A reconsideration of the domestication geography of tetraploid wheats. Theor Appl Genet 110:1052–1060

Özkan H, Willcox G, Graner A, Salamini F, Kilian B (2011) Geographic distribution and domestication of wild emmer wheat (*Triticum dicoccoides*) Genet Resour Crop Evol 58:11–53

Peng JH, Fahima T, Roder MS, Li YC, Dahan A, Grama A, Ronin YI, Korol AB, Nevo E (1999) Microsatellite tagging of the stripe-rust resistance gene *YrH52* derived from wild emmer wheat, *Triticum dicoccoides*, and suggestive negative crossover interference on chromosome 1B. Theor Appl Genet 98:862–872

Peng J, Korol AB, Fahima T, Röder MS, Ronin YI, Li YC, Nevo E (2000) Molecular genetic maps in wild emmer wheat, triticum dicoccoides: genome-wide coverage, massive negative interference, and putative quasi-linkage. Genome Res 10:1509–1531

Poyarkova H (1988) Morphology, geography and infraspecific taxonomics of *Triticum dicoccoides* Körn. A retrospective of 80 years of research. Euphytica 38:11–23

Poyarkova H, Gerechter-Amitai Z, Genizi A (1991) Two variants of wild emmer (*Triticum dicoccoides*) native to Israel: morphology and distribution. Can J B 69:2772–2789

Qayoum A, Line RF (1985) High-temperature, adult-plant resistance to stripe rust of wheat. Phytopathology 75:1121–1125

Rafalski JA (2010) Association genetics in crop improvement. Curr Opin Plant Biol 13:174–180

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci 103:18656–18661

Saintenac C, Jiang D, Wang S., Akhunov E (2013) Sequence-based mapping of the polyploid wheat genome. G3: Genes| Genomes| Genetics

Sajjad M, Khan SH, Kazi AM (2012) The low down on association mapping in hexaploid wheat (*Triticum aestivum* L.). J Crop Sci Biotechnol 15:147–158

Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R, Martin W (2002) Genetics and geography of wild cereal domestication in the near east. Nat Rev Genet 3:429–441

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

R Core Team (2011) R: a language and environment for statistical computing. http://www.R-project.org/

Wang M, Jiang N, Jia T, Leach L, Cockram J, Waugh R, Ramsay L, Thomas B, Luo Z (2011) Genome-wide association mapping of agronomic and morphologic trait in highly structured populations of barley cultivars. Theor Appl Genet 124:233–246

Warnes G, Gorjanc G, Leisch F, Man M (2011) Genetics: Population Genetics (http://CRAN.R-project.org/package=genetics)

Waugh R, Jannink JL, Muehlbauer GJ, Ramsay L (2009) The emergence of whole genome association scans in barley. Curr Opin Plant Biol 12:218–222

Wellings CR (2011) Global status of stripe rust: a review of historical and current threats. Euphytica 179:1–13

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhang Z, Buckler ES, Casstevens TM, Bradbury PJ (2009) Software engineering the mixed model for genome-wide association studies on large samples. Brief Bioinform 10:664–675

Zohary D, Hopf M (2000) Domestication of plants in the old world: the origin and spread of cultivated plants in west Asia, Europe, and the Nile Valley. Oxford University Press, USA