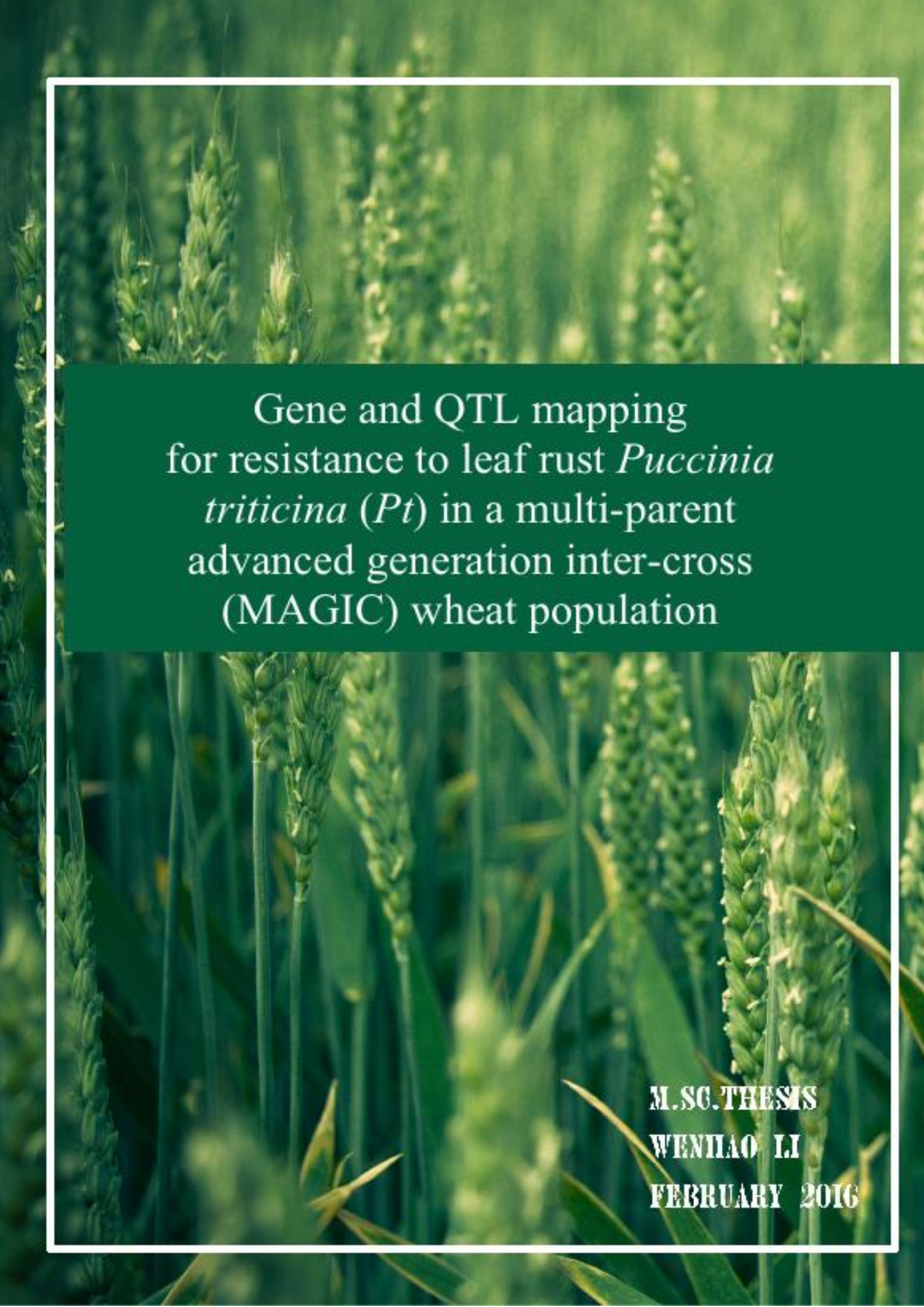


Gene and QTL mapping for resistance to leaf rust *Puccinia triticina* (*Pt*) in a multi-parent advanced generation inter-cross (MAGIC) wheat population

M.SC. THESIS
WENHAO LI
FEBRUARY 2016

WAGENINGEN UNIVERSITY

WAGENINGEN UR

Gene and QTL mapping for resistance to leaf rust *Puccinia triticina* (*Pt*) in a multi-parent advanced generation inter-cross (MAGIC) wheat population

M.Sc. Thesis, PBR-80436
Laboratory of Plant Breeding (PBR)
February 2016

Student:
Wenhao Li
Plant Biotechnology – Molecular Plant Breeding and Pathology

Supervisors:
Dr. Rients Niks (PBR)
Dr. Marco Bink (Biometris)
Tryntsje Boersma (PhD, Biometris, NIAB)

Acknowledgements

This project was conducted under the supervision of Dr.Rients Niks, Dr.Marco Bink and Tryntsje Boersma. As my supervisors, they were very generous to offer me help whenever I needed. Rients gave me thoughtful and thorough guidance in many aspects of my work like disease testing, results presentation and scientific writing. Marco provided me with lots of help in the part of data analysis. He generously offered me a working place in Biometris group where I finished the final part of my thesis. Tryntsje entrusted me with this thesis project and supported me a lot when I got lost. It was a great pleasure to work with them. This thesis project would not be completed without their supervision works. I would also like to take this opportunity to express my gratitude to my mother for supporting me studying in the Netherlands.

Contents

Figures

Tables

Summary

This MAGIC wheat population is a novel mapping population constructed in the National Institute of Agricultural Botany (NIAB) and some genetic researched have been done in this population, but not for leaf rust resistant. The objectives of this project are:

(1) Demonstrate the application of the MAGIC wheat population by mapping genes and QTLs for qualitative resistance and quantitative resistance to wheat leaf rust.

(2) Compare different approaches of association analysis in the MAGIC wheat population for validating or developing efficient analysis tool.

Phenotypic data for resistance (infection type and latency period) were collected as the example traits to validate the exploitation of MAGIC wheat population. Two ideal pathogen isolates were selected to screen the whole MAGIC population for collecting phenotypic data. The results of association analysis shows that two R genes for qualitative resistance were mapped and two putative QTLs for quantitative resistance were also detected. The discovered R genes and QTLs were traced back to the genome location of the eight founders to reveal the donors. At the end, literature study was done to compare previously reported genes with the discovered genes. All the results indicate that MAGIC wheat population can be a good resource for gene discovery. Different approaches have been tried in the association analysis. Statistical models designed for bi-parental population also worked well to detected R genes for qualitative resistance but can be improved. For quantitative resistance, the mapping approach using the previously estimated founder probabilities as input can detect the QTLs with small effects, while the models for bi-parental system are not very feasible.

1 Introduction and background

1.1 Mapping population

Plant breeding has been playing an important role in genetic improvement of the plant performance. Two main activities are involved in plant breeding: (1) exploring new variation in the germplasm and (2) assembling desirable phenotypes in the offspring (Koh *et al*, 2015). Various strategies and bio-technologies have been developed on these two aspects (Lusser *et al*, 2011). The idea of plant breeding is to expand and create valuable variation that satisfies breeder's demands, and select progeny or offspring with the positive traits. Traditional breeding strategies based on morphological traits are time-consuming and cumbersome. In order to speed up the whole breeding process, marker technologies have been developed. Instead of phenotypic selection, marker technologies provide a way to indirectly select superior plant. Particularly, molecular markers make it possible to map desirable alleles of a gene or quantitative trait loci (QTL). The selection on the basis of such markers is known as marker-assisted breeding (MAB), which can accelerate the breeding progress (Bernardo, 2008). However, without the appropriate strategies to identify the tightly trait-linked markers for MAB, it is difficult to fix the desirable genes in the progeny. The markers development requires an appropriate marker system, a suitable segregation population, an efficient data analysis tool, and the reliable phenotyping method.

Bi-parental population and natural germplasm are widely used for genetic study. Most of the segregation or mapping populations are derived from two parents. They are not difficult to construct and analysis tools are well established for them. Such populations, like recombinant inbred lines (RILs), near-isolated lines (NILs) and doubled haploids (DH), are widely used for QTL identification. However, there are some limitations of the bi-parental system. First of all, only two parents capture a small range of genetic diversity, and the segregation population is developed from the parents with the only one contrasting trait. Studying multiple traits is difficult in the same population. Secondly, the resolution of bi-parental population may not keep up with the high-resolution of genotyping. The limitation of recombination probability leads to the low mapping resolution. To circumvent the above limitations, linkage disequilibrium (LD) mapping or association mapping (AM) based on natural germplasm have been applied. Although AM captures greater genetic diversity than that of a bi-parental population, the power to detect the genomics region of interest cannot be guaranteed due to some limitations like population structure and the lack of a reliable high-density consensus Map. (Kim *et al*, 2007; Laurie *et al*, 2004; Yu *et al*, 2006)

Breeders and researchers came up with an intermediate method that can overcome the disadvantages of AM and bi-parental designs and combine their advantages. One of the approaches is to construct multi-parental advanced generation inter-cross (MAGIC) population. MAGIC is an extension of the advanced inter-cross method in which an inter-mated mapping population is created from multiple founder lines. The large number of parental founders increases the allelic and phenotypic diversity over traditional RILs (Kover *et al*, 2009; Mackay *et al*, 2014). The successive rounds of recombination cause LD to decay thereby increase the precision of QTL location (Mackay and Powell, 2007; Mackay *et al*, 2014). A MAGIC population of size 500 above is believed to detect single QTL explaining 5% of the phenotypic variation (Klasen *et al*, 2012; Valdar *et al*, 2006).

1.2 The construction of wheat MAGIC population

An eight-parent MAGIC wheat population has been constructed in NIAB through a controlled and balanced crossing scheme of eight parental wheat cultivars. In order to validate the utilization of this population for genetic map construction and QTL detection, the population has been genotyped by the Illumina Infinium iSlect 90K wheat assay and some traits have been evaluated and analyzed. The genetic diversity of this population has also been benchmarked to a bi-parental population and 64 wheat varieties. It is reported that this population is preferable for genetic study (Mackay *et al*, 2014; Verbyla *et al*, 2014). The construction of this MAGIC wheat population is comprised of founder selection, mixing and inbreeding.

The MAGIC wheat population is expected to cover wide genetic range by selecting founder lines that show wide phenotypic and geographic diversity. The eight modern cultivars (Table 1) were selected as founder lines to construct MAGIC wheat population (Mackay et al, 2014).

Table 1 Founder lines of the wheat MAGIC population.

Variety	Listing Year	Seed Yield(t/ha)a	NABIM Quality Group	Trait Attributes
Alchemy	2006	9.163	4	Yield, disease resistance, breeding use, soft
Brompton	2005	9.151	4	Hard feed, 1BL/1RS, OWBM-resistant
Claire	1999	8.654	3	Soft biscuit/distilling, slow apical development
Hereward	1991	7.683	1	High-quality benchmark 1 bread-making
Rialto	1994	8.377	2	Moderate bread-making, 1BL/1RS
Robigus	2003	9.053	3	Exotic introgression, disease resistance, breeding use, OWBM-resistant, Rht-B1
Soissons	1995	7.553	2	Bread-making quality, early flowering, Rht-B1
Xi19	2002	8.975	1	Bread-making quality, facultative type, breeding use

a.Yield adjusted for site and year effects as estimated by Mackay et al. (2011). (Mackay *et al*, 2011)
b.NABIM quality groups (http://www.nabim.com/):
1 (high bread-making); 2 (good bread-making); 3 (biscuit/cake); 4 (other).
cOWBM, orange wheat blossom midge.

After the founders were determined, they were crossed in a way shown in Figure 1A. All founders were combined in a prescribed pattern, which is called a funnel. The individuals generated from eight-way crosses were processed to obtain homozygous stage by single seed descent (SSD). Theoretically, 315 unique funnels can be generated by considering all possible combinations through 2-way intercross to the final 8-way intercross (Figure 1B).

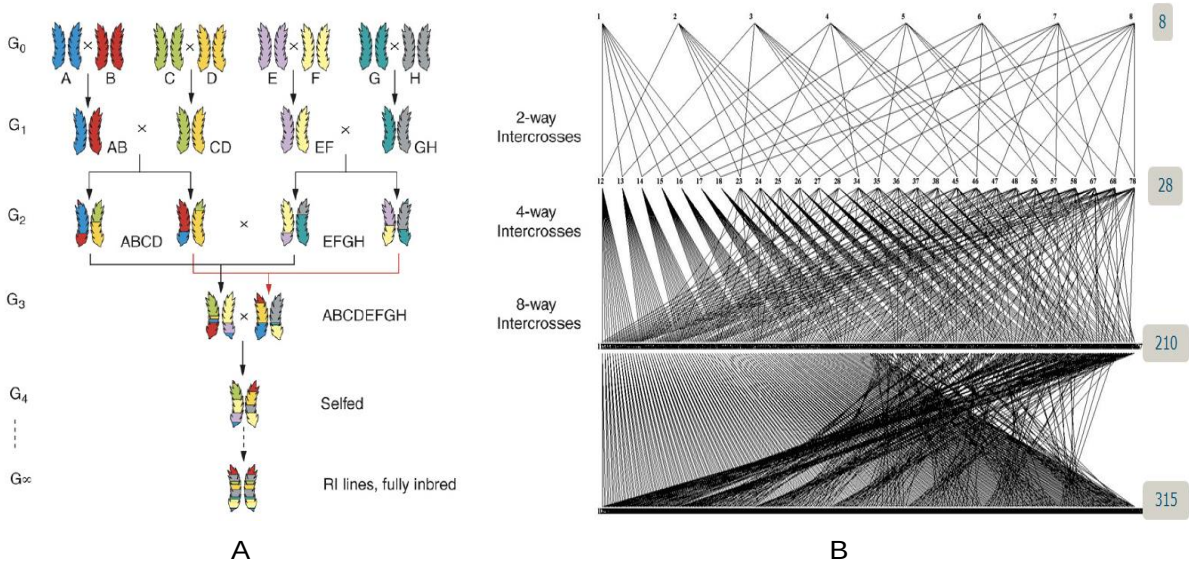


Figure 1 (A) The 8-way intercross of MAGIC as one funnel (Cavanagh *et al*, 2008); (B) the theoretical number of funnels generated from 8 founders (T.Boersma).

1.3 Wheat leaf rust resistance

In this project, gene and QTL mapping for leaf rust resistance was performed to validate the efficiency of this MAGIC wheat population for gene discovery. This trait is chosen for the following reasons: (1) leaf rust resistance is economically important for wheat yield and it has been widely studied; (2) leaf rust resistance can be assessed in both qualitative and quantitative aspects. With the help of appropriate methodological strategies, evaluating leaf rust resistance is expected to be relatively easy and precise compared to that of other quantitative traits such as plant height, yield and flowering time.

Wheat leaf rust, caused by *Puccinia triticina* (*Pt*), is one of the most widely distributed wheat diseases resulting in loss of grain production (Khan *et al*, 2013; Knott, 2012). Creating genetic resistance variety is one of the ways to control leaf rust. When it comes to genetic resistance, two classes of genes have been widely used in breeding program. The first class is called R gene or *Lr* gene in wheat. An effective R gene confers qualitative resistance that can extend form seedling stage to adult stage of the plant. The second class of genes, instead of giving complete resistance, can reduce the epidemic development of the pathogen and thus decrease infection severity in field trials compared to the severity on more susceptible wheat accessions. This is known as partial resistance (PR) or quantitative resistance. Qualitative resistance and quantitative resistance differ in genetic and biological aspects.

Qualitative resistance requires R (*Lr*) gene in wheat and the corresponding avirulence (*Avr*) gene in *Pt*. It is also known as gene-for-gene hypothesis (Flor, 1956). R gene can encode R-protein, which is able

to recognize the modification of virulence targets caused by Avr-encoded effector. The battle between R gene and *Avr* gene leads to hypersensitive reaction (HR) that can stop the further development of the pathogen (Van der Hoorn *et al*, 2002), leaving behind some necrosis or chlorotic flecks on leaves.

Quantitative resistance or PR, mainly based on multiple minor genes, is quantitatively inherited and not race-specific to *Pt*. The mechanisms behind quantitative resistance are diverse and multilayer. Most quantitative effect is due to variation in defence genes (Jones and Dangl, 2006; Vergne *et al*, 2010). The allelic variants show different levels of defence, thus, leading to different degree of PR between accessions. Some quantitative resistance is contributed by effector targets. The specific recognition between effector and their targets in plant can alter the expression of defence (Niks and Marcel, 2009). Therefore, the variation of effector targets among host accessions may result in different degree to which the pathogen can suppress PAMP-triggered immunity (PTI) and show difference in PR. R gene and PR gene can also work together. It is reported that effect of some *Lr* genes can be enhanced by the presence of some PR genes, and the stack of PR minor genes can result in "near immunity" (Singh *et al*, 2011).

2 Gene mapping for qualitative resistance

2.1 Materials and methods

2.1.1 Genotyping

DNA was extracted from F_5 generation during SSD procedure using the modified Tanksley extraction protocol (Fulton *et al*, 1995). The genomic DNA samples were then sent for genotyping using the wheat 90k iselect array (http://www.illumina.com/) from the Illumina platform. The arrays yielded 18606 raw SNPs data for 720 MLs, which are available on the website of NIAB (http://www.niab.com/MAGIC/). The quality control was performed to eliminate SNPs and samples of bad quality before processing. The filtering criteria are mainly based on genotyping quality scores (GC score), and SNPs with extremely low allele frequency (<1%) were also discarded. Other criteria were considered to refine SNPs or samples quality on the basis of the degree of missing value and deviation form Hardy–Weinberg equilibrium(HWE). For example, the large departure from HWE may indicate potential genotyping errors (Anderson *et al*, 2010; Turner *et al*, 2011). At the end, we got refined 17291 SNPs and 643 lines with high quality, which were used for phenotyping and association analysis.

2.1.2 Isolate selection

Infection type (IT) and latency period (LP) are two parameters for the assessment of resistance. IT scores indicate the intensity of the HR (McNeal et al, 1971). Based on the IT scores, MAGIC lines (MLs) can be determined to be resistant or susceptible as the phenotype of qualitative resistance. Among susceptible lines, different levels of resistance can be indicated by LP. As one of the prominent components of quantitative resistance, LP is defined as "the period elapsing from the moment of inoculation to the moment of becoming infectious" (Parlevliet, 1979; Shaner and Finney, 1980) that can be measured in monocyclic test. A precise inoculation method is also required. Settling tower, as a standardized method for the quantitative analysis of resistance, was used to guarantee homogeneous administration of inoculation (Eyal et al, 1968; Miclot et al, 2012).

The first step was to select the ideal isolate by testing different pathogen isolates on eight founders. The ideal isolate is expected to (1) have the wide virulence spectrum with at least one resistant founder and (2) if possible, give large variation of LP values among susceptible founder lines. The reason is that quantitative resistance can be masked by qualitative R-gene resistance, so an isolate that is avirulent to only one or two founders leaves enough susceptible MLs to map quantitative resistance. Virulence spectrum was determined by the IT scores that indicate the intensity of the HR on the scale of 0~9 (Table2) where rating 0~7 are considered as resistant type and 8~9 are considered as susceptible type (McNeal *et al*, 1971)

Table 2 Infection types in wheat infected by rust species.

Infection type	Description
0	No symptoms
1	Small necrotic or chlorotic flecks
2	Flecks somewhat larger
3	Minute uredosori surrounded by necrotic or chlorotic tissue
4	Between 3 and 4
5	Small uredosori surrounded by some necrotic or chlorotic tissue
6	Between 5 and 7
7	larger uredosori surrounded by some necrotic or chlorotic tissue
8	Uredosori surrounded by very faint chlorosis
9	Well-developed uredosori, no chlorosis or necrosis. There is often a pale halo around the uredosori
x	Various infection types on one and same leaf

LP of each pathogen isolate was determined in a monocyclic test by calculating at what time when 50% of eventual number of pustules became matured (LP50) during the disease development. The formula is described as:

$$LP50 = a + \left(\frac{b}{c}\right) * d$$

a= time from inoculation until last counting before 50% pustules were mature

b= time between the counting before and after 50% of pustules were mature

c= increase in number of pustules during period b

d= 50% counting minus number of pustules at the start of period b

The disease test started with evaluating eight founders with six isolates (Durum, Felix, Flamingo, INRA, Swiss and Ventus) provided by Dr. Rients Niks (Wageningen UR, plant breeding chair group). Disease test experiments were conducted in Unifarm. For each isolate, we prepared one tray, so we had six trays in total for all six isolates. The eight founders were sowed with 6 replicates in each tray. When the first leaves of seedlings were unfolded after around two weeks of growing in a pathogen-free compartment, those leaves were fixed by pins to stretch out and then the whole tray was put into settling tower. In settling tower, seedlings were inoculated by spraying rust urediniospores at a dosage from 2.5 mg to 7mg depending on the degree of freshness and storage time. After five minutes, the tray was placed in a dark dew chamber overnight (16h) at 18 degree centigrade with 100% humidity to promote the spore germination. The next morning, spore germination was checked under microscope. Next, the seedlings were moved into a greenhouse for infection development. About 5 days after the inoculation, the mature pustules were counted on the marked segment of each leaf. The number of matured pustules was checked once or twice a day until it stoped increasing. IT score was rated at the last round of pustule counting when a seedling showed unambiguous infectious pattern.

2.1.3 Screening MAGIC population with ideal isolate

After the ideal isolate had been selected, the MAGIC population was screened by the selected isolate to collect phenotypic data. In each tray, we prepared 24 MLs with 2 replications for each line. After two weeks, the first leave per ML was fixed and then inoculated using fresh rust spore at a dosage of 1mg per tray. Resistance evaluation was conducted during the development of disease similar to isolate selection on the eight founders. In this chapter, we focus on IT scores in the population as phenotypic data for qualitative resistance.

2.1.4 Genetic analysis

The consensus linkage map for gene mapping was constructed by Gardner (unpublished). The gene mapping approaches were based on single marker analysis (SMA) and simple interval mapping (SIM) that are described in Appendix A. The association analysis using SIM was performed with the help of some relevant packages like R/qtl (Broman *et al*, 2003). The scripts are provided in Appendix B. In order to cross-compare the results from two different mapping approaches, the generated test statistic p-values from SMA and LOD scores form SIM were extracted and normalized to the same scale for plotting in one figure. At the end, the identity-by-state (IBS) of founder alleles, which refer to the same alleles (Powell *et al*, 2010), was checked against the LOD profile to confirm the donors of detected genes.

2.2 Results

2.2.1 Isolate selection

All eight founders were inoculated by each rust isolate, and the IT and LP were evaluated. Table 3 presents the overview of virulence profile of the six isolates on eight founders. Isolate INRA and Ventus had the widest virulence spectrum. Robigus was evaluated as resistant to INRA and Ventus, for the necrosis and chlorotic flecks (IT=1) appeared on the inoculated leaves. For the other 7 founders, the rust uredosori were well developed on the inoculated leaves with a pale halo around it (IT=9), which was considered as susceptible. Another rust isolate, Flamingo, has the second widest virulence spectrum, which was virulent to five founders (IT score= 8 or 9). Swiss and Durum were virulent to one and three founders, respectively. All founders were resistant to Felix (IT score = 1 ~ 5).

Table 3 Compatible (+) and incompatible (-) interactions of eight wheat founders and six isolates of wheat leaf rust (*Puccinia triticina*) with IT scores.

Founder	Isolate (IT score)					
	Durum	Felix	Flamingo	INRA	Swiss	Ventus
Alchemy	- (6)	- (3)	+ (8)	+ (9)	- (4)	+ (9)
Brompton	- (7)	- (1)	- (1)	+ (8)	- (1)	+ (9)
Claire	- (4)	- (3)	+ (8)	+ (9)	- (3)	+ (9)
Hereward	- (3)	- (5)	+ (9)	+ (9)	- (6)	+ (9)
Rialto	+ (8)	- (1)	- (1)	+ (8)	- (2)	+ (9)
Robigus	- (3)	- (3)	- (1)	- (1)	+ (9)	- (1)
Soissons	- (4)	- (4)	+ (9)	+ (9)	+ (9)	+ (9)
Xi19	- (4)	- (4)	+ (8)	+ (8)	+ (8)	+ (9)

Both INRA and Ventus can be considered as the ideal isolates to screen the whole population given the widest virulence spectrum. As for LP, the LP50 values within eight founders to INRA seem more stable than that to Ventus (Figure 2). Inoculated with INRA, Claire showed the shortest LP50 around 144.2h, while Brompton showed the relatively high LP50.

To sum up, INRA was chosen to screen the whole population due to the wide virulence spectrum as well as the relatively stable LP values within founders. The other isolate, Flamingo, was also selected to check whether there is any shared R gene with INRA, because Robigus was resistant to both INRA and Flamingo.

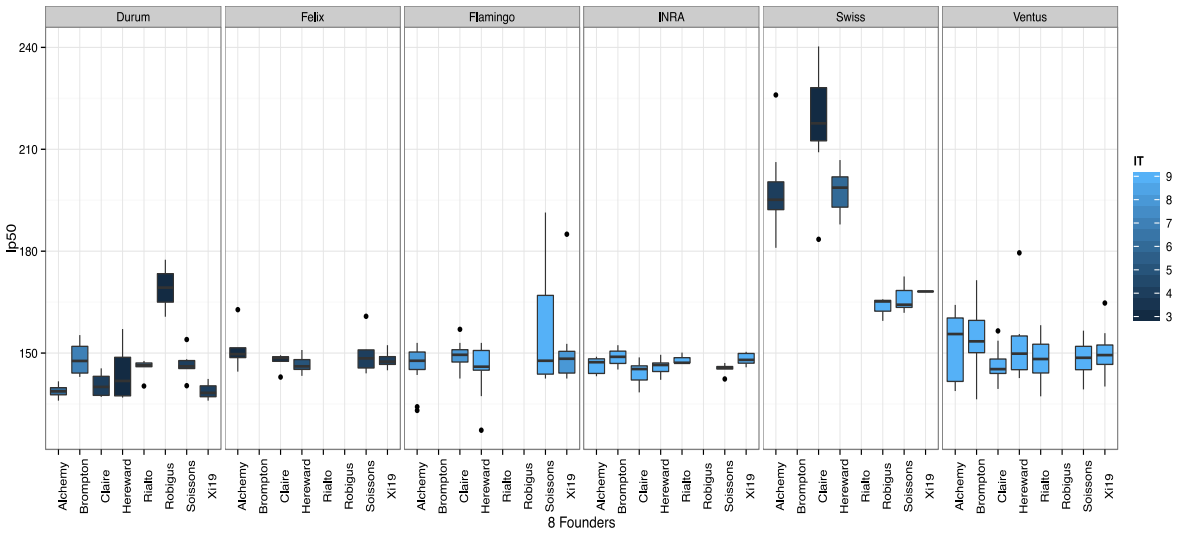


Figure 2 Latency period of six isolates on eight founders. The absolute LP value (LP50) against each founder lines to different isolate is presented in boxplot. The colour of box indicates the intensity of qualitative resistance (IT score). Relatively light blue indicates high IT score like 8 or 9, while the darker blue represent resistant lines with IT scores below 7.

2.2.1 Assessment of infection type

631 and 636 out 643 MLs of were successfully assessed to INRA and Flamingo, respectively. Missing values occurred due to some practical reasons, for example the seeds of some lines were not supplied and a few seeds were geminated poorly. After 2 weeks of pathogen development, IT scores were rated and the results are showed in table 4. Interestingly, for those resistant MLs to INRA except for one line, they were also resistant to Flamingo.

Table 4 The number of resistance and susceptible lines as response to two isolates of wheat leaf rust (*Puccinia triticina*).

Infection type	Isolate	
	INRA	Flamingo
Resistant lines (IT = 1)	48	116
Susceptible lines (IT = 8 or 9)	583	520

2.2.2 Association analysis results

The association between qualitative resistance and 17291 SNPs markers was analyzed using SMA and SIM mapping approaches (see Appendix A and Appendix B). The p-values (SMA) and LOD scores (SIM) were extracted and plotted together (Figure 3). The results form SMA and SIM both suggest that one resistance gene was detected to INRA and two detected to Flamingo.

An interval from 180.8cM to 181.8cM on chromosome 4A contains the resistance gene to INRA. This interval is defined by the large decline of adjusted p-values from the highest SNP at 180.9cM to its flanking SNPs. As response to Flamingo, the gene on chromosome 4A was mapped between 180.8cM and 182.9cM (peak SNP at 180.9cM), and the other gene was detected on chromosome 1B between 109.4cM and 119.0cM (peak SNP at 115.5cM). Some significantly associated SNPs identified to INRA and Flamingo on chromosome 4A are overlapped in the region around 180.9cM, which indicate that resistant MLs may share the same resistance gene. In this report, we denote the shared resistance gene on chromosome 4A as R1 and the other one detected on chromosome 1B as R2. Next, we move on to confirm the donors of R1 and R2.

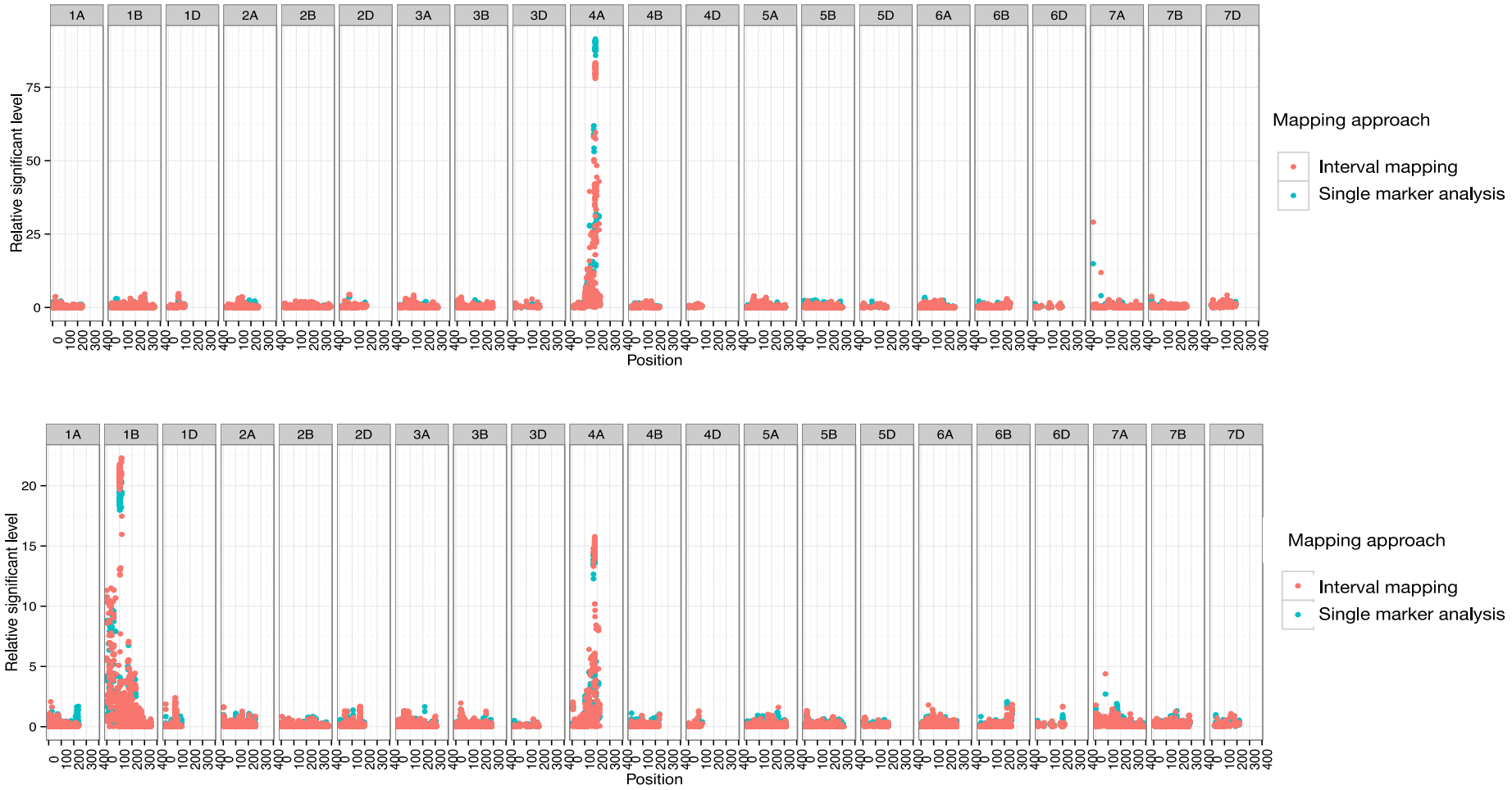


Figure 3 Gene mapping for qualitative resistance as response to isolate INRA (upper) and Flamingo (lower). Normalized –log10(p-values) and LOD scores was plotted on the y-axis against the position of SNPs on x-axis. Those red and blue dots represent the results from SIM and SMA, respective.

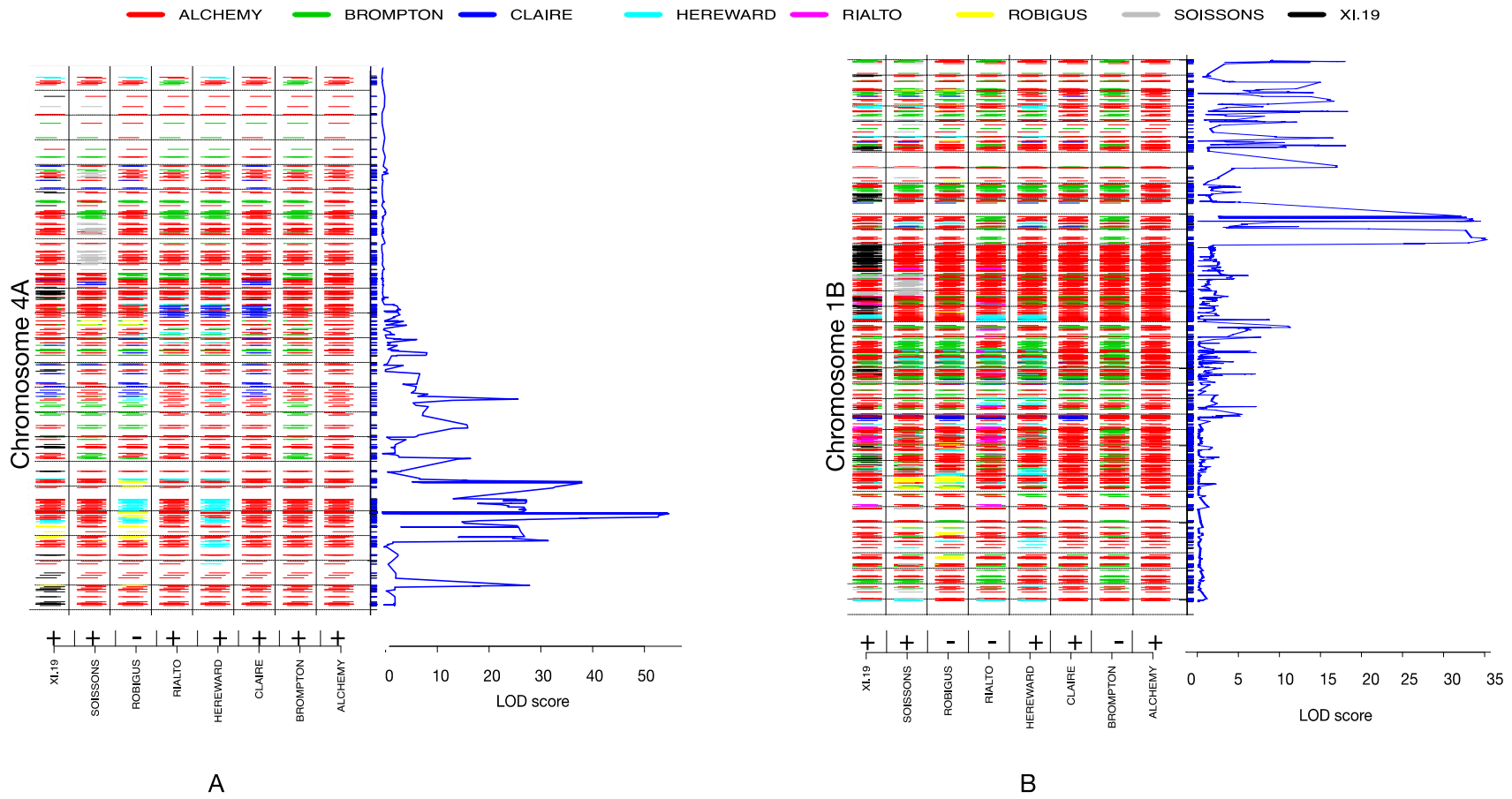


Figure 4 Confirmation of donors of R1 and R2 from IBS founder alleles. (A) IBS founder alleles on chromosome 4A against LOD score to INRA. (B) IBS founder alleles on chromosome 1B against LOD score to Flamingo. The linkage maps are placed with short arms on top. At any given (SNP) position on the linkage maps, the founders with the same colour share the same IBS allele. The founder Alchemy in red was used as reference and the same alleles on other founders are in red too. Likewise, any same IBS allele shared with Brompton but different from Alchemy is in green. The lines are jittered for better visualization of the tightly positioned SNPs.

2.2.3 The donors of R genes

Checking the IBS of founder alleles at the position of peak SNP provided a method to trace R genes back to donors. The IBS of the founder genotypes on chromosome 4A and 1B, where R1 and R2 were detected, was plotted against the LOD profile under INRA and Flamingo using SIM (Figure 6).

On chromosome 4A (Figure 4A), Robigus has a unique IBS founder allele (in yellow) at the place of the peak SNP (180.9cM), while other seven founders share the same IBS founder allele with Alchemy. It is confirmed that Robigus, as the only resistant founder to INRA, delivered the R1 gene to the MAGIC wheat population. On chromosome 1B (Figure 4B), the founders with unique allele (in green) at the position of peak SNP (116.4cM) are the founders (Brompton and Rialto) that were resistant to Flamingo, so Brompton and Rialto are believed to be the donors of R2.

In summary, two R genes were detected and their donors were revealed. R1 gene, located on chromosome 4AL and inherited from Robigus, conferred resistance to INRA and Flamingo; R2 gene, located on chromosome 1BS and inherited from Brompton and Rialto, gave resistance to Flamingo.

2.3 Discussion and conclusion

2.3.1 Gene-for-gene model in qualitative resistance

In this chapter, the results are consistent with our assumption that multiple disease tests can be performed on the same wheat MAGIC population and multiple race-specific resistance genes can be identified. By executing two rounds of disease tests with different isolates INRA and Flamingo, we detected two corresponding resistance genes R1 and R2 and successfully traced them back to their donors. All findings in this chapter are also in agreement with the theories that qualitative resistance is monogenetic inheritance and can be explained by gene-for-gene model.

The differential interaction can be observed in the virulence profile of six isolates on eight founders. For instance, isolate INRA was virulent to Rialto but avirulent to Robigus, while Swiss was inversely avirulent to Rialto but virulent to Robigus. This race-specific interaction is believed to be the results of interaction of the R gene in the plant and corresponding *Avr* gene in the pathogen, also known as gene-for-gene hypothesis (Flor, 1956).

In this study, the resistance gene R1 was detected on chromosome 4AL to both INRA and Flamingo. The donor of R1 is the shared resistant founder Robigus. Based on gene-for-gene model, these findings can be explained by assuming that pathogen isolate INRA and Flamingo carry the corresponding *Avr*1 gene which made them avirulent to founder Robigus as well as the 48 MLs that have inherited R1 from Robigus through eight-way cross. Flamingo was avirulent to Brompton and Rialto, which indicate that a different avirulent gene *Avr*2 was recognized by corresponding resistance genes R2 from Brompton and Rialto (Table 5). The possible loss-of-function mutation from *Avr2* to *avr2* in INRA impaired gene-for-gene interaction between INRA with Brompton and Rialto, and hence, these two founders were unlucky infected by INRA.

Table 5 Gene-for-gene model for qualitative resistance between 3 founders and 2 isolates.

Founder (R gene)	Isolate (*Avr* gene)	
	INRA (*Avr*1, *avr*2)	Flamingo (*Avr*1, *Avr*2)
Brompton (R2)	+	-
Rialto (R2)	+	-
Robigus (R1)	-	-

R1- *Avr*1 interaction gives the resistance of Robigus (R1) to pathogen isolate INRA (*Avr*1) and Flamingo (*Avr*1); R2-*Avr*2 interaction gives the resistance of Brompton (R2) and Rialto (R2) to Flamingo (*Avr*2); no interaction between R2 and *avr*2 explains the susceptibility of Brompton (R2) and Rialto (R2) to INRA (*avr*2).

Literature study was done to compare reported *L*r genes with the detected R1 and R2 genes in this study (Table 6). Among previously reported 46 race-specific (qualitative) leaf rust resistance genes (Li *et al*, 2014), only *Lr*28 (McIntosh *et al*, 1982) is reported to locate on chromosome 4AL. *Lr*28, derived from *Aegilops speltoides*, is an effective gene for resistance through the whole lifespan of the plant in many region of the word (Naik *et al*, 1998). *Lr*28 was identified with flanked RAPD marker by a distance of 2.4± 0.016 cM (Cherukuri *et al*, 2005) and linked to SSR marker *Xwmc497* at a distance of 2.9cM (Pallavi *et al*, 2015). In this report we mapped R1 to around 1cM. On chromosome 1BS, no *Lr* genes have been reported (McIntosh *et al*). The most likely *Lr* gene that denoted as R2 in this study is *Lr*71. *Lr*71 is flanked by marker *Xgwm18* and *Xbarc*187 crossing the centromere and possibly locates on the short arm of chromosome 1B, given the unsolved centromere region (Singh *et al*, 2013).

Table 6 Comparison with previously reported race-specific genes.

Denotation	Gene symbol	Synonyms	Chromosome	Markers
R1	*Lr*28		4AL	*Xwmc*497 (~2.9 cM)
R2	*Lr*71	LrAK12c	1B centromere region not resolved	*Xgwm*18 and *Xbarc*187

2.3.2 Mapping approaches for qualitative resistance in wheat MAGIC population

In this chapter, the mapping approaches we used for association study were SMA and SIM with refined genotypic data as input. SMA and SIM have been widely applied to bi-parental system, and they also worked well in MAGIC for mapping R genes. The benefit of using biallelic SNPs in regression model is that the process is computationally simple and straightforward and thereby provides a chance to quickly screen the whole genome. Such approaches based on regression method was firstly applied to a four-way cross system (Xu, 1996). In Mackay's (2014) research, the marker scores were also used as input for mapping awning gene using mixed model accounting for stratum generated from different funnels, and the *B1* awning locus was successfully mapped to a 7.5 cM interval on chromosome 5AL. Qualitative trait like awning and race-specific resistance are conferred by a major locus that strongly influence the trait, so that the major locus can be mapped by simply using marker scores as input, while the estimation of founder probabilities seems not very necessaryin this case. Huang (2015) also used simulation method to prove that the result of interval mapping using biallelic SNPs is consistent with that based on founder probabilities when the trait is contributed by major locus.

However, in this study, there is a limitation of gene mapping for qualitative resistance against Flamingo using SMA and SIM in linear regression model. For resistance to INRA, only one resistance

gene is involved, similar to the scenario in Mackay's study that awning is conferred by only one *B1* gene. Since all the resistance MLs can be explained by R1, the statistic model accounting for the allele's effect of only R1 worked well, so that R1 can be mapped in a relative precise region. On the contrary, resistance to Flamingo is not as straightforward as that to INRA. The resistant MLs under Flamingo can be divided into three groups which were inherited with R1, R2 and R1+R2, respectively. While we performed association mapping to Flamingo, all the resistance MLs were pooled, so the genetic effects of R1 and R2 were not statistically distinguished. Therefore, the detection power of R gene to Flaming was decreased. That is why, as response to Flamingo, R1 was mapped at the interval 2 times larger than that to INRA; the interval region of R2 against Flamingo is 10 times larger than that of R1 detected against INRA. Therefore, the statistic model should be improved in this scenario. R1 was the shared resistance gene to isolate INRA and Flamingo, which is responsible for 47 resistance MLs to Flamingo. When we want to map R2, we can treat R1 as cofactor to account for some part of the resistant MLs and thus the effect of the R2 can be separated. In this model, R1 serves as the proxy to increase the gene mapping precision of R2.

3 QTL mapping for quantitative resistance

3.2 Materials and methods

3.2.1 Genotyping and phenotyping

The phenotyping method for collecting LP data is descried in the previous chapter, and we also use the same genotypic data for association analysis. LP was evaluated among the 583 susceptible lines which are R1-free to INRA. In order to account for the possible nonuniformity and fluctuation of environmental conditions or manual evaluation error in different batches, the average LP50 values of Brompton per batch was used as reference to calculate the relative latency period (RLP) for MLs in the same batch. In order to validate the reliability of phenotyping, 12 MLs with long LP and 12 MLs with relative short LP among 583 MLs were re-evaluate in one tray with 2 replicates for each line.

$$\mathrm{RLP} = \mathrm{LP50} * \frac{100}{\mathrm{LP50\ of\ the\ refrence\ line}}$$

3.2.2 Genetic analysis

The same linkage map and refined genetic data as described in chapter 2 were used to map QTLs for quantitative resistance. In MAGIC population, the inputs of association analysis for quantitative resistance were marker scores and founder probabilities. The founder probabilities of each MLs were estimated by a hidden Markov model (HMM) based-framework called reconstructing ancestry blocks bit by bit (RABBIT) (Zheng *et al*, 2015). The results of SMA and composite interval mapping (CIM) were benchmarked to the results using founder probabilities as input. The involved models are described in Appendix A.

The association analysis of phenotypic data (RLP) and genotypic data (SNPs) or founder probabilities was performed using R program based on some packages such as HAPPY (Mott *et al*, 2000), R/qtl (Broman *et al*, 2003) and ASReml (Gilmour *et al*, 2009) etc. The threshold of significant level in CIM and SMA models was determined by permutation test, during which the phenotype data was randomly sampled 1000 times while genotype data remained intact. A set of maximum LOD scores or –log10 (P-value) were collected and the 95th percentile was set as threshold. Some part of the script is provided in Appendix B.

3.3 Results

3.3.1 Assessment of latency period

RLP of the 583 susceptible MLs under INRA was evaluated and the frequency distribution is plotted (Figure 5). The overall mean of RLP is 98.36 with variance of 19.36. We used LP values of the seven susceptible founders to INRA from isolate selection (section 2.2.1) to calculate their RLP with Brompton as reference line. In the MAGIC population, around 44.5% of the MLs had RLP values lower than Claire, and around 29.2% of the MLs had RLP values higher than Brompton.

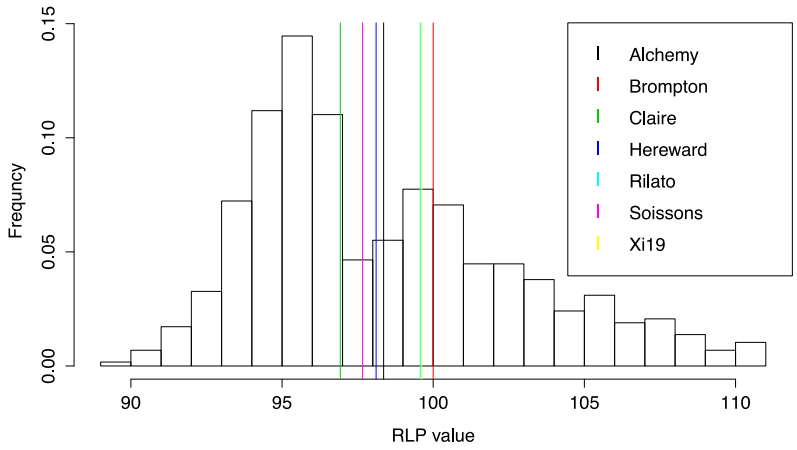


Figure 5 Frequency distribution of relative latency period (RLP) in the MAGIC wheat population.

From each tray, the MLs with the highest LP value and the lowest LP value were selected to repeat the LP evaluation under INRA. There is a significant deference (P-value < 0.01) between long-LP MLs and short-LP MLs, but some MLs were cross-confounding and not quite distinct from each other (Figure 6).

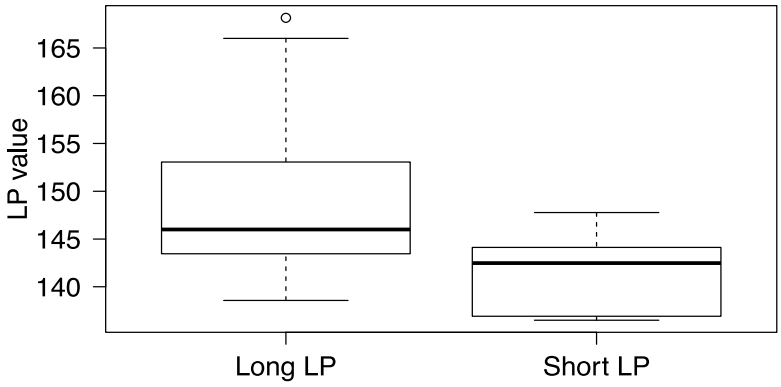


Figure 6 Re-evaluated LP values form long-LP and short-LP MLs

3.3.2 QTL mapping

Based on the founder probabilities, QTLs for quantitative resistance were identified on chromosome 1BS and chromosome 6BL (Figure 7). Brompton and Rialto contributed the long LP gene on chromosome 1BS and Robigus contributed to long LP on chromosome 6BL. Their QTL effects are around 0.1σ to the whole population.

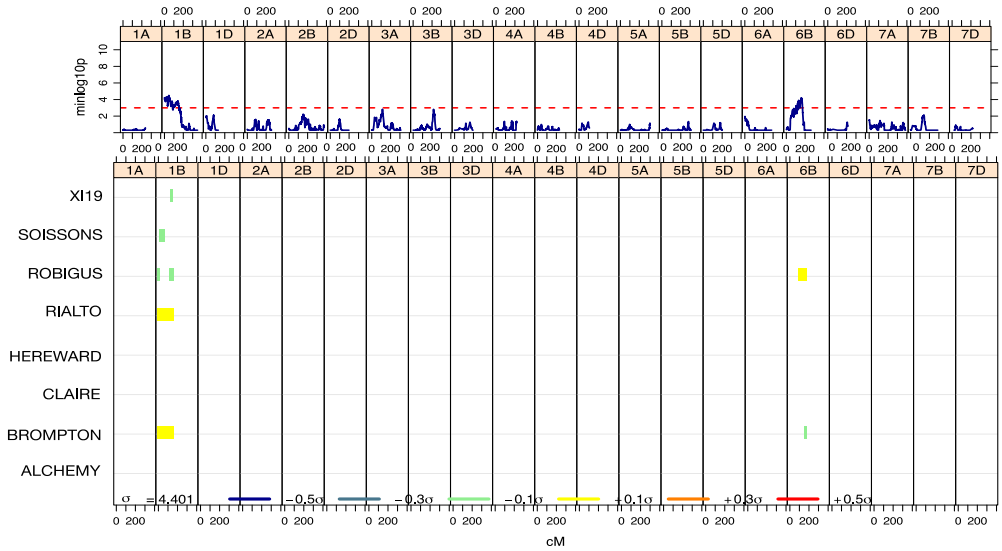


Figure 7 QTLs profile based on founder probabilities (upper) and the parental haplotype effect (lower) of LP. Significant QTLs are determined by exceeding the threshold of 3 and the haplotype effect sizes are indicated in different colors.

To compare the QTL detection power of different mapping methods, the QTL profiles form CIM and SMA were extracted from Appendix B and benchmarked to the mapping approach based on founder probabilities (Figure 8). The threshold was set at 3.9 by permutation test. There are some differences of QTLs detection among different mapping approaches. QTLs detected on chromosome 1B and 6B are also detected by CIM and SMA. However, the QTLs detected on chromosome 3B and 4A were unexpected, given the fact that they were not detected by the approach using founder probabilities as input.

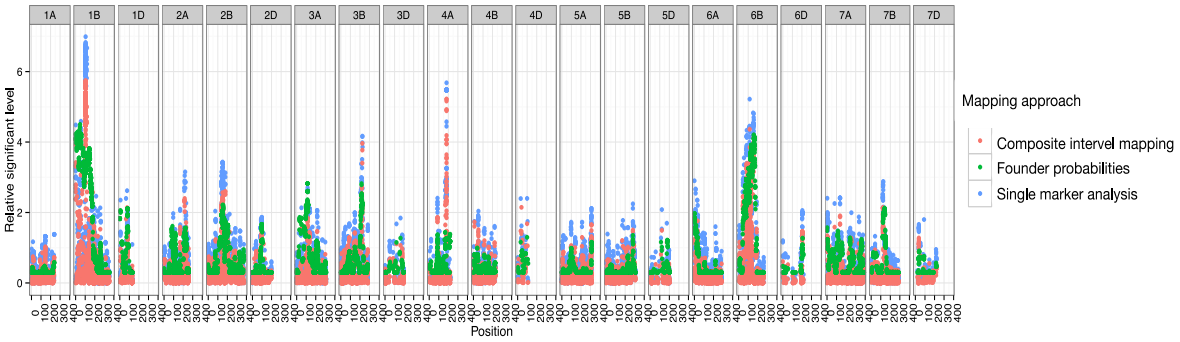


Figure 8 QTLs profile based on composite interval mapping (red) and single marker analysis (blue) against the model with founder profanities as input (green).

3.4 Discussion

3.4.1 Polygenetic inheritance in quantitative resistance

In this study, the results of QTLs mapping indicates that two QTLs coffering the quantitative resistance on chromosome 1BS and 6BL. That is why those susceptible MLs under INRA showed quantitative differences in degree of infection with meticulous observation and showed the transgressive inheritance compared to eight founders. It might be explained that abundant QTLs existing in eight founders contribute to different levels of resistance to INRA. Each QTL may have a small effect to LP level, and the by-chance accumulation of them through eight-way inter-cross leaded to the transgressive segregation.

The wheat chromosome 1BS is functionally active in multiple disease resistance. Several QTLs for quantitative resistance and some *Lr* genes have been mapped on this region (Li *et al*, 2014; McIntosh *et al*). Chromosome 1BS is like the battlefield for the plant-pathosystems gene competence. In this study, we detected a qualitative resistance gene R2 on chromosome 1BS as response to Flamingo. The other isolate, INRA, break down R2 by altering Avr2, and thus become virulent to the donors of R2 (Brompton and Rialto). However, the suppression may not be complete, given the fact that a QTL for quantitative resistance was detected on chromosome 1BS and the long LP was donated by Brompton and Rialto, but the QTL need to be mapped more precisely to confirm that in future work. The other QTL detected on chromosome 6BL was conferred by Robigus. The quantitative resistance also occurred on Robigus when inoculated with INRA, but the effect was not observed because it had been masked by the completely expressed resistance conferred by R1. The quantitative resistance might be observed on Robigus only when a new virulent isolate appears to break down R1, which is refer to residual resistance.

In previous study, five QTLs for quantitative resistance to leaf rust have been reported on chromosome 1BS (Messmer *et al*, 2000; Rosewarne *et al*, 2012; Schnurbusch *et al*, 2004; Singh *et al*, 2009; William *et al*, 1997) and three QTLs on 6BL (Chu *et al*, 2009; Rosewarne *et al*, 2012; William *et al*, 2006) (Table 7).

Table 7 Comparison with previously reported QTLs partial resistance.

QTL	Chr.	Donor	Marker Interval	Reference
QLr.sfrs-1BS	1BS	Forno	Xpsr949–Xgwm18	Messmer et al., 2000
QLr.sfr-1BS	1BS	Forno	Xgwm604–OA93	Schnurbusch et al., 2004
QLr.cimmyt-1BS.1	1BS	Parula	Xcmtr03–500	William et al., 1997
QLr.cimmyt-1BS.2	1BS	Pastor	wPT5580–wPT3179	Rosewarne et al., 2012
QLr.pbi-1BS	1BS	Beaver	1BL/1RS	Singh et al., 2009
QLr.fcu-6BL	6BL	TA4152–60	Xbarc5–Xgwm469.2	Chu et al., 2009
QLr.cimmyt-6BL.1	6BL	Pastor	wPT6329–wPT5176	Rosewarne et al., 2012
QLr.cimmyt-6BL.2	6BL	Pavon 76	XpAGGmCGA1	William et al., 2006

Because these QTLs have been detected from different marker panels and mapping populations, their positions may differ due to different population sample size, recombination rate and accuracy of genotyping or phenotyping methods. Therefore, it is a challenge to precisely cross-compare the positions of putative QTLs. With increasing QTLs detected for more traits in this wheat MAGIC population, it is valuable to integrate results from different studies by QTL meta-analysis in future study (Lillemo and Lu, 2015; Maccaferri *et al*, 2015).

3.4.2 Mapping approaches for quantitative resistance in wheat MAGIC population

Implementing mixed model and inputting founder probabilities are the effective approaches for QTLs mapping in this wheat MAGIC population, and this approach was also tried to map QTLs for height and yield (see Appendix C), while the general QTLs mapping approaches in bi-parental population are not effective in MAGIC. By the method of SMA and CIM, ghost QTLs were detected on chromosome 3B and 4A. The allele frequency for the peak SNP (Tdurum_contig39549_245) on chromosome 4A is as low as 8% which is believed to generate the false positive. Since only Robigus conferred the long LP on chromosome 6B, the effect of genetic factor was not strong enough to be detected by approaches based on bi-parental system. In MAGIC population, the genome segment of progenies is the mosaic of eight founders. In order to increase the power and precision of QTL detection, it is wise to use founder probabilities to account for each founder's effect instead of simply using bi-allelic information without knowing the haplotype effect.

Mapping for quantitative resistance in wheat magic population successfully reveals the genetic architecture for long LP values, but the resolution is not as high as we expected, given the high-resolution mapping population, high-density linkage map and advance mapping algorithm. We assessed latency period in monocyclic test as the component of quantitative resistance. It is convenient and not very time-consuming. However, the latency period might be temperature-sensitive and easily influenced by other conditions and thereby LP was subject to experimental error. Consequently, the large variance of LP is due to non-genetic factor, while the effect of genetic factor was relatively neglected and thus the true QTLs could not identified in a promisingly precise position. In future study, other important components for quantitative resistance, such as infection frequency and spore production, along with latency period, can be re-assessed in the population to cross-confirm or refine the position of detected QTLs.

4 General Discussion and Future Prospects

4.1 General discussion

The wheat MAGIC population is a promising resource for gene isolation and QTL mapping. Previous study by Mackay (2014) has shown that this population has captured 74% of genetic diversity of 64 wheat varieties and genomes of MLs are highly recombined. With the help of high-throughput marker system (wheat 90k SNP array) and advanced statistical tools accounting for haplotype effect, this population is believed to be a potential platform for genetic dissection of complex trait. It was the first time to implement MAGIC population in studying wheat leaf rust resistance in this project. We have executed several rounds of disease test with multiple *Pt* isolates on this population along with the founders. As results, two race-specific resistance genes and two QTLs for quantitative resistance have been detected and founder contribution has been revealed.

MAGIC population has its unique strength in genetic study over bi-parent system and natural germplasm. Firstly, it can be served as a permanent resource for gene or QTL mapping for multiple traits, while most of bi-parental population are constructed for the only one contrasting trait between them and thus are not suitable to study other traits. Compared to a set natural germplasm, MAGIC is more related to breeding germplasm because the eight founders are modern cultivars. Therefore the wheat population can also be utilized in breeding program. However, the issue of MAGIC might be the difficulty of testing epistasis. Due to the diversity of founder alleles and large number of markers, testing for pairwise interactions requires exhaustive analysis approach and much larger sample size of the population.

Various mapping approaches used in this study have shown their strengths and limitations. The statistical model using marker scores is straightforward and computationally easy, which is perforable to fast scan the genome for locating the trait-associated markers. Since the linkage map is high-throughput, approaches based on interval mapping like SIM and CIM that can estimate the genotype of missing SNPs, are not very necessary. Instead, SMA is more efficient in the association study for qualitative resistance. However, the limitations of using marker scores are as follows: first, the founder contribution is unknown. We have to check the IBS founder allele to identify the donors of putative genes or QTLs. Second, it tends to generate false positive and biased statistic test if the allele

frequency is too low. At the last, only the major gene with large effect can be mapped, while minor QTLs cannot be detected effectively. Given the above limitations, we recommend to the use founder probability in QTL mapping for some quantitative traits with low heritability, which can detect the minor QTLs and reveal haplotype effects. If there is any limitation with this approach in this study, it might be that we did not account for the genetic background of funnels and environmental effect.

4.2 Future prospective

Wheat MAGIC population has shown its magic in gene discovery and genetic study. We can image more and more agronomic traits will be assessed and genetic research will continue in this population. Future studies may focus on more complex traits. For instance, resistance to abiotic stress like salt tolerance and drought tolerance can be evaluated on various conditions, providing insight to the gene-by-environment interaction and mechanism of abiotic stress tolerance. With increasing association analysis carrying on, it is possible to identify more pleiotropic genes, which makes the population valuable in breeding program.

MAGIC population can also serve as breeding population. In future, Muti-parent advanced generation recurrent selection (MAGReS) can be applied in breeding program (Huang *et al*, 2015). The identified favourable alleles from association study are in turn applied to select superior MLs. The MAGIC population can be used as training population for constructing genomics selection (GS) model. Those positive alleles can be pyramid with the help of GS.

Appendix A. Statistic models in association analysis

- Single marker analysis

The simplest mapping method is based on single marker analysis (SMA), also known as point analysis (Soller *et al*, 1976). Each individual is assigned to groups based on the single marker genotype. By testing the significance of difference between phenotypic means in the marker genotype, the trait-associated markers can be detected. The genetic model of SMA can be described as follows:

$$y_i = \mu + x_i\alpha + e_i$$

Where y_i is the adjusted phenotypic value of *i*-th individual; μ is the grand phenotypic mean for each individual; α represents the genotype effect given the single marker alleles x_i of *i*-th individual, and e_i is the random error term assumed to be $e_i \sim N(0, {\sigma_e}^2)$. For each SNP marker, the null hypothesis (H_0: $\alpha = 0$) assuming there is no genotype effect will be tested. The p-values form the multiple testing of each SNP is converted to -log10 scale as the indication of significant level. The most likely trait-associated SNPs are indicated by the peak p-value on the profile plot. The association analysis of SMA is independently perform based on the single markers, thus, linkage map is not necessarily required.

- Simple Interval mapping

The simple interval mapping (SIM) was first discussed in 1989 (Lander and Botstein, 1989), which become a standard QTL mapping procedure and then it has been further extended to many other mapping procedures. The genetic model of SIM is as flowering:

$$y_i = \mu + x_{j|i}\alpha + e_i$$

The terms of y_i, μ, α and e_i in SIM are the same as in SMA. The biggest difference is that the single marker alleles x_i is replaced by $x_{j|i}$, estimated by the conditional genotypes. The maximum likelihood estimation (MLE) can be used to obtain the unknown parameters of u, α and ${\sigma_e}^2$ given the observed phenotypic value y_i and estimated genotypic data c_i. The MLE for those parameters is also estimated assumed there is no QTL anywhere in the marker interval. The likelihood of a putative QTL presenting in the marker interval is indicted by the LOD scores, which are generated from above two MLEs.

- Composite interval mapping

The composite interval mapping (CIM) uses the identified QTLs in SIM as cofactors to search for other marker intervals on the same chromosomes. The peak markers is first selected and then revaluated together with the second peak marker. If both of them are significant, they are treated as cofactors to search for the third highest marker until the whole genome is scanned. The cofactors serve as proxies accounting for genetic background, thus this approach is believed to decrease the chance of detecting ghost QTLs and increases the precision of QTLs position (Jansen 1994; Zeng 1994). We supposed there are m+1 ordered marker, the statistical model of CIM is expressed as:

$$y_i = \mu + \alpha^q x_i^q + \sum_{c=1}^{m+1} \alpha^c x_i^c + e_i$$

where y_i, μ and e_i, as described in previous models, represent the adjusted phenotype of i-th individual, ground mean of all individuals and random error term; α^q denotes the effect of QTL q, and α^c denotes the effect of cofactor c; x_i^q and x_i^c are genetic predictors estimated form conditional genotypes at QTL q and cofactor c.

- Use founder probabilities as input for QTL mapping

The inputting of the genome for QTL mapping can base on marker score and founder probabilities. Because eight founders are involved in the mapping population design, the genomic segment of each MLs is the made up of the eight founders. SMA, SIM and CIM with markers scores as input for QTLs mapping, which provides rare information about the haplotype mosaics construction of each MLs. Therefore, the donor of a certain trait might be difficult to identify even though the QTLs are detected. The founder probability p_{if}^l indicate the likelihood that the *i*-th ML is derived founder *f* at the locus *L*, which can be estimated by hidden Markov model (HMM) (Zheng *et al*, 2015). The common model used in MAGIC population is fixed-effects linear model based on the calculated p_{if}^l (Kover *et al*, 2009). The phenotype y_i in ML *i* can be modelled given a QTL segregate at locus *L* in which the phenotypic effect due to founder haplotype isα_f.

$$y_i = \sum_f p_{if}^l \alpha_f + e_i$$

Appendix B. The R script for association mapping using SMA, SIM and CIM

- **SMA**

Required packages

```
library(dplyr)
library(ggplot2)
```

The linkage map

```
snp_map<-read.csv("arranged_geno.csv")[,1:3]
load("geno_data.RData")
```

The genotypic data

```
snp_data<-read.csv("arranged_geno.csv")[,-c(2,3)]
```

The phenotypic data

```
phenoall<-read.csv("phenoall.csv")
```

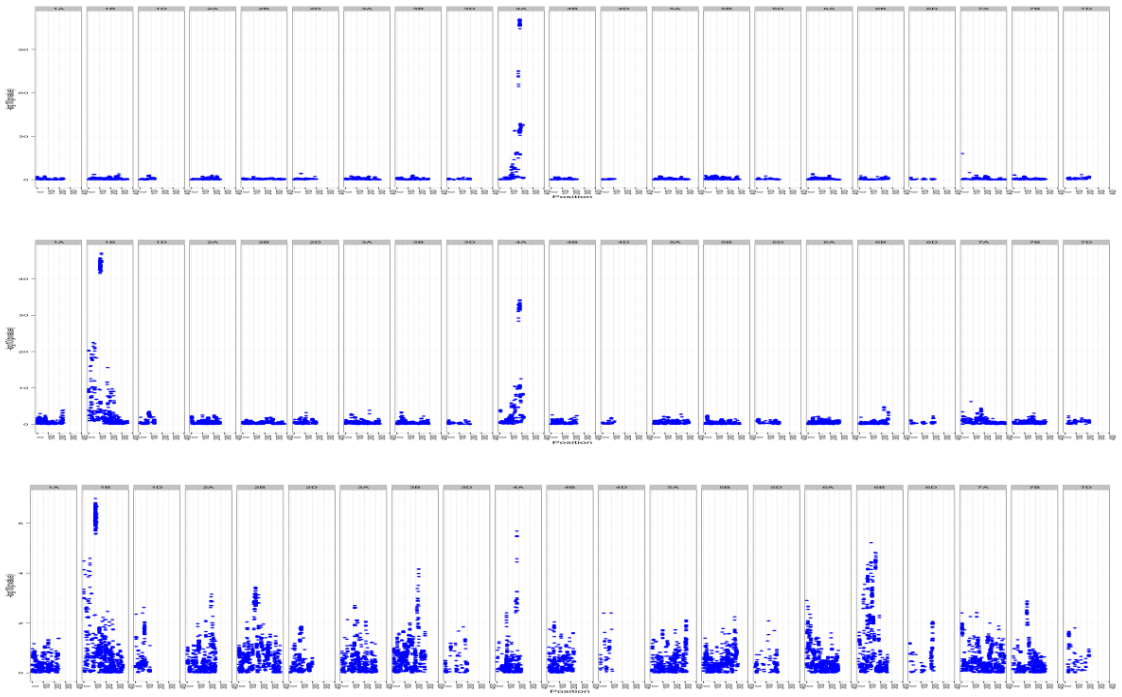
The function of SMA

```
SMA.mapping<-function(trait) {
    index<-grep(trait, colnames(phenoall))
    if (class(phenoall[,index]) == "factor") {
        sigCal<-function(snp){
            chi<-chisq.test(snp,phenoall[,index])
            p_value<-chi$p.value
            return(p_value)}
    } else if (class(phenoall[,index]) == "numeric" || class(phenoall[,inde
x])=="integer") {
        sigCal<-function(snp){
            mod<-summary(lm(phenoall[,index]~snp))
            p_value<-mod$coefficients[2,4]
            return(p_value)}
    }
    all_pvalue<-apply(genoall[,-1],2 , sigCal)
    all_pvalue<-data.frame(SNPs=colnames(genoall[,-1]), pvalue=unlist(all_p
value))
    merged<-merge(all_pvalue,snp_map, by="SNPs")
    merged<-arrange(merged, Chrom., Position)
    merged$indice<-1:length(snp_data$SNPs)
    merged$pvalue<- -log10(merged$pvalue)
    filename<-paste(trait,"_SMA_mod.csv",sep = "")
    write.csv(merged,filename)
    figure<-ggplot(merged,aes(Position, pvalue))+geom_point(col="blue")+the
me_bw()+theme(axis.text.x=element_text(angle = 90))+facet_grid(.~Chrom.)+yl
ab("-log10(pvalue)")
```

```
    return(figure)
}
```

Qualitative resistance under INRA, Flamingo and quantitative resistance

```
par(mfrow=c(3,1))
SMA.mapping("IT_INRA")
SMA.mapping("IT_Flamingo")
SMA.mapping("RLP")
```



- **Accounting for funnels using mixed linear model**

Required packages and function

```
library(lme4)
library(lmerTest)
SMA.mapping.funnel<-function(trait) {
    index<-grep(trait, colnames(phenoall))
    sigCal<-function(snp){
        mod<-anova(lmer(phenoall[,index]~(1|funnel)+(1|funnel:plant)+ snp,
data=phenoall))
        p_value<-mod[6]
        return(p_value)}
    all_pvalue<-apply(genoall[,-1],2 , sigCal)
    all_pvalue<-data.frame(SNPs=colnames(genoall[,-1]), pvalue=unlist(all_p
value))
    merged<-merge(all_pvalue,snp_map, by="SNPs")
    merged<-arrange(merged, Chrom., Position)
    merged$indice<-1:length(snp_data$SNPs)
    merged$pvalue<- -log10(merged$pvalue)
```

```
    filename<-paste(trait,"_funnel_mod.csv",sep = "")
    write.csv(merged,filename)
    figure<-ggplot(merged,aes(Position, pvalue))+geom_point(col="blue")+the
me_bw()+theme(axis.text.x=element_text(angle = 90))+facet_grid(.~Chrom.)+yl
ab("-log10(pvalue)")
    return(figure)
}
```

- **SIM**

Required packages

```
library(qtl)
```

The linkage map, genotypic data and phenotypic data, the format was prepared according to the r/qtl package's requirements and saved as Rdata.

```
load("IM.RData")
```

The function of SIM, the threshold (h=3.9) was determined by permutation

```
SIM.mapping<-function(trait){
    index<-grep(trait, colnames(magic$pheno))
    if (class(magic$pheno[,index])=="numeric") {
        magic<-calc.genoprob(magic)
        magic<-jittermap(magic)
        magic<-sim.geno(magic)
        out<-scanone(magic,pheno.col = index,method = "imp")
        write.csv(out, "IM_mod.csv")
        figure<-plot(out, ylab = "LOD score",col="blue")+abline(h=3.9,col="
red")
        return (figure)

    } else {
        if(index==3) {
            magic$pheno$IT_INRA<-as.character(magic$pheno$IT_INRA)
            magic$pheno$IT_INRA[!is.na(magic$pheno$IT_INRA)& magic$pheno$IT
_INRA=="R"]<-0
            magic$pheno$IT_INRA[!is.na(magic$pheno$IT_INRA)& magic$pheno$IT
_INRA=="S"]<-1
            magic$pheno$IT_INRA<-as.numeric(magic$pheno$IT_INRA)
            out<-scanone(magic, pheno.col = 3, model = "binary")
            filename<-paste(trait,"_SIM_mod.csv",sep = "")
            write.csv(out,filename)
            figure<-plot(out,ylab = "LOD score",col="blue")
            return (figure)
        }
        else if (index==4) {
            magic$pheno$IT_Flamingo<-as.character(magic$pheno$IT_Flamingo)
            magic$pheno$IT_Flamingo[!is.na(magic$pheno$IT_Flamingo)& magic
$pheno$IT_Flamingo=="R"]<-0
            magic$pheno$IT_Flamingo[!is.na(magic$pheno$IT_Flamingo)& magic
$pheno$IT_Flamingo=="S"]<-1
            magic$pheno$IT_Flamingo<-as.numeric(magic$pheno$IT_Flamingo)
            out<-scanone(magic, pheno.col = 4, model = "binary")
```

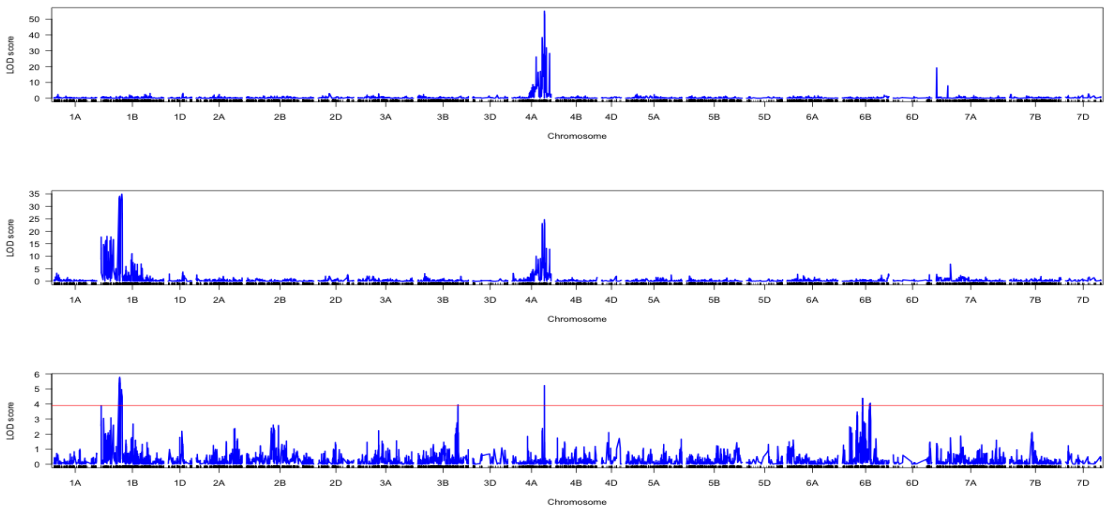
```
            filename<-paste(trait,"_SIM_mod.csv",sep = "")
            write.csv(out,filename)
            figure<-plot(out,ylab = "LOD score",col="blue")
            return (figure)
        }
        else{
            magic$pheno$Awns<-as.character(magic$pheno$Awns)
            magic$pheno$Awns[!is.na(magic$pheno$Awns)& magic$pheno$Awns=="N
"]<-0
            magic$pheno$Awns[!is.na(magic$pheno$Awns)& magic$pheno$Awns=="Y
"]<-1
            magic$pheno$Awns<-as.numeric(magic$pheno$Awns)
            out<-scanone(magic, pheno.col = 13, model = "binary")
            filename<-paste(trait,"_SIM_mod.csv",sep = "")
            filename<-paste(trait,"_SIM_mod.csv",sep = "")
            write.csv(out,filename)
            figure<-plot(out,ylab = "LOD score",col="blue")
            return (figure)
        }

    }

}
```

Qualitative resistance under INRA, Flamingo and qauntitative resistance

```
par(mfrow=c(3,1))
SIM.mapping("IT_INRA")
SIM.mapping("IT_Flamingo")
SIM.mapping("RLP")
```



- **CIM**

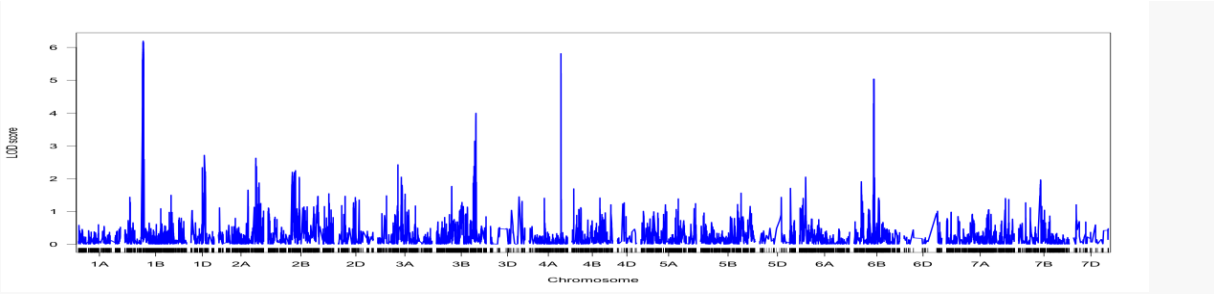
The function of CIM

```
CIM.mapping<-function(trait){
    index<-grep(trait, colnames(magic$pheno))
    magic<-jittermap(magic)
```

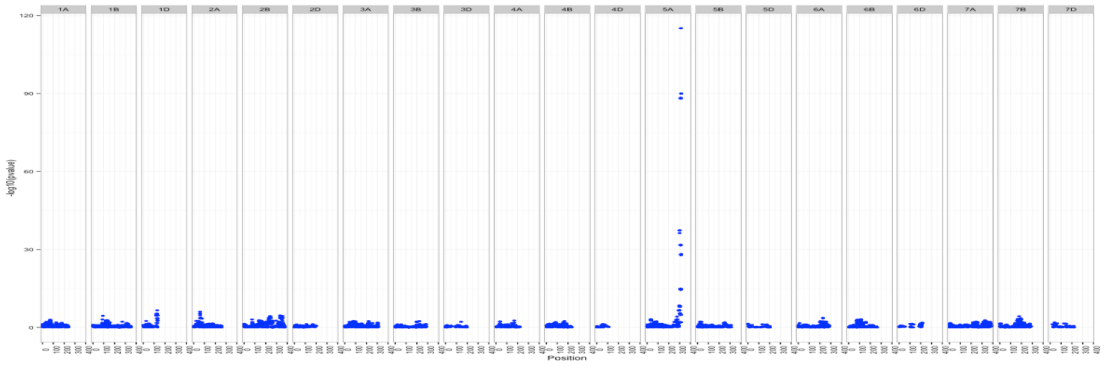
```
    magic<-sim.geno(magic)
    out<-cim(magic,pheno.col = index,method = "imp")
    filename<-paste(trait,"_CIM_mod.csv",sep = "")
    write.csv(out, filename)
    figure<-plot(out, ylab = "LOD score", col="blue")
    return (figure)
}
```

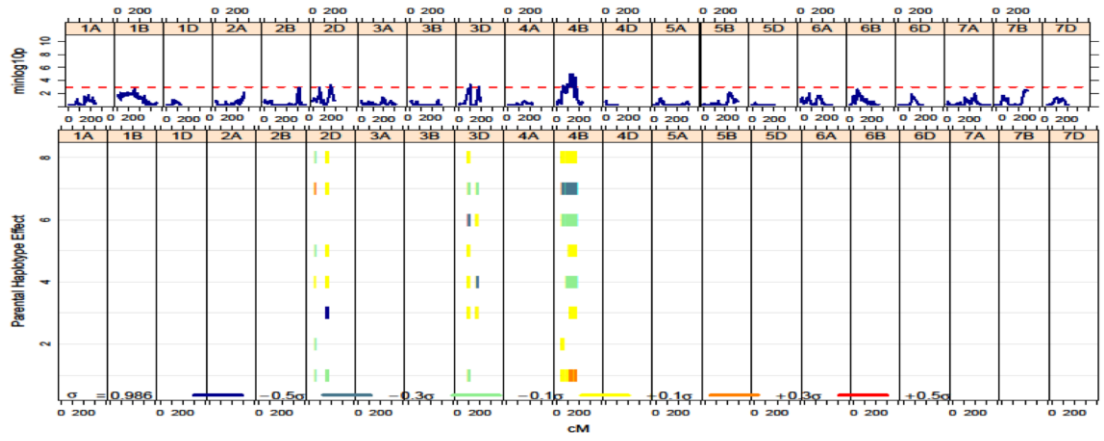
Qunatitative resistance RLP

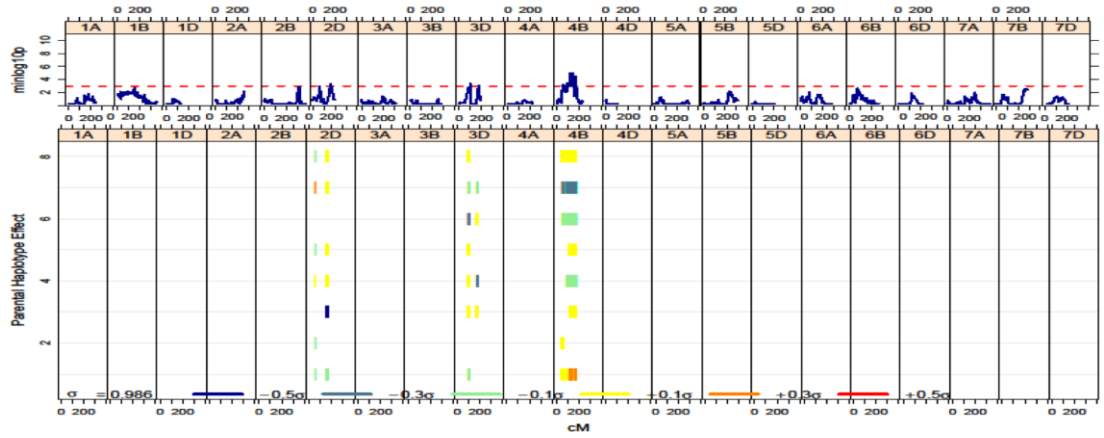
```
CIM.mapping("RLP")
```



Appendix C. Association mapping for awning, yield and height







References

Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT (2010). Data quality control in genetic case-control association studies. *Nature protocols* **5**(9)**:** 1564-1573.

Bernardo R (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Science* **48**(5)**:** 1649-1664.

Broman KW, Wu H, Sen Ś, Churchill GA (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**(7)**:** 889-890.

Cavanagh C, Morell M, Mackay I, Powell W (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current opinion in plant biology* **11**(2)**:** 215-221.

Cherukuri DP, Gupta SK, Charpe A, Koul S, Prabhu KV, Singh RB *et al* (2005). Molecular mapping of Aegilops speltoides derived leaf rust resistance gene Lr28 in wheat. *Euphytica* **143**(1-2)**:** 19-26.

Chu C-G, Friesen T, Xu S, Faris J, Kolmer J (2009). Identification of novel QTLs for seedling and adult plant leaf rust resistance in a wheat doubled haploid population. *Theoretical and applied genetics* **119**(2)**:** 263-269.

Eyal Z, Clifford BC, Caldwell RM. (1968). *Vol. 58*. AMER PHYTOPATHOLOGICAL SOC 3340 PILOT KNOB ROAD, ST PAUL, MN 55121, pp 530-&.

Flor H (1956). The complementary genic systems in flax and flax rust. *Adv Genet* **8:** 29-54.

Fulton TM, Chunwongse J, Tanksley SD (1995). Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Molecular Biology Reporter* **13**(3)**:** 207-209.

Gilmour AR, Gogel B, Cullis B, Thompson R, Butler D (2009). ASReml user guide release 3.0. *VSN International Ltd, Hemel Hempstead, UK*.

Huang BE, Verbyla KL, Verbyla AP, Raghavan C, Singh VK, Gaur P *et al* (2015). MAGIC populations in crops: current status and future prospects. *Theoretical and Applied Genetics* **128**(6)**:** 999-1017.

Jones JD, Dangl JL (2006). The plant immune system. *Nature* **444**(7117)**:** 323-329.

Khan MH, Bukhari A, Dar ZA, Rizvi SM (2013). Status and strategies in breeding for rust resistance in wheat. *Agricultural Sciences* **4**(06)**:** 292.

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S *et al* (2007). Recombination and linkage disequilibrium in Arabidopsis thaliana. *Nature genetics* **39**(9)**:** 1151-1155.

Klasen J, Piepho H, Stich B (2012). QTL detection power of multi-parental RIL populations in Arabidopsis thaliana. *Heredity* **108**(6)**:** 626-632.

Knott DR (2012). *The wheat rusts—breeding for resistance*, Vol 12. Springer Science & Business Media.

Koh H-J, Kwon S-Y, Thomson M (2015). Current Technologies in Plant Molecular Breeding.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD *et al* (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in Arabidopsis thaliana. *PLoS Genet* **5**(7)**:** e1000551.

Lander ES, Botstein D (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**(1)**:** 185-199.

Laurie CC, Chasalow SD, LeDeaux JR, McCarroll R, Bush D, Hauge B *et al* (2004). The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics* **168**(4)**:** 2141-2155.

Li Z, Lan C, He Z, Singh RP, Rosewarne GM, Chen X *et al* (2014). Overview and application of QTL for adult plant resistance to leaf rust and powdery mildew in wheat. *Crop Science* **54**(5)**:** 1907-1925.

Lillemo M, Lu Q. (2015). *Abstracts*.

Lusser M, Parisi C, Plan D, Rodríguez-Cerezo E. (2011). ISBN 978-92-79-19715-4, ISSN 1018-5593, doi: 10. 2791/ 54761, Publications Office of the European Union, Luxembourg.

Maccaferri M, Zhang J, Bulli P, Abate Z, Chao S, Cantu D *et al* (2015). A genome-wide association study of resistance to stripe rust (Puccinia striiformis f. sp. tritici) in a worldwide collection of hexaploid spring wheat (Triticum aestivum L.). *G3: Genes| Genomes| Genetics* **5**(3)**:** 449-465.

Mackay I, Horwell A, Garner J, White J, McKee J, Philpott H (2011). Reanalyses of the historical series of UK variety trials to quantify the contributions of genetic and environmental factors to trends and variability in yield over time. *Theoretical and Applied Genetics* **122**(1)**:** 225-238.

Mackay I, Powell W (2007). Methods for linkage disequilibrium mapping in crops. *Trends in plant science* **12**(2)**:** 57-63.

Mackay IJ, Bansept-Basler P, Barber T, Bentley AR, Cockram J, Gosman N *et al* (2014). An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. *G3 (Bethesda, Md)* **4**(9)**:** 1603-1610.

McIntosh R, Dubcovsky J, Rogers W, Morris C, Appels R, Xia X *et al* CATALOGUE OF GENE SYMBOLS FOR WHEAT: 2013-2014.

McIntosh R, Miller T, Chapman V (1982). Cytogenetical studies in wheat. XII. Lr28 for resistance to Puccinia recondita and Sr34 for resistance to Puccinia graminis tritici. *Zeitschrift fur Pflanzenzuchtung= Journal of plant breeding*.

McNeal F, Konzak CF, Smith E, Tate W, Russell T (1971). *A uniform system for recording and processing cereal research data*, Vol 34. USDA-ARS.

Messmer M, Seyfarth R, Keller M, Schachermayr G, Winzeler M, Zanetti S *et al* (2000). Genetic analysis of durable leaf rust resistance in winter wheat. *Theoretical and Applied Genetics* **100**(3-4)**:** 419-431.

Miclot A-S, Wiedemann-Merdinoglu S, Duchêne E, Merdinoglu D, Mestre P (2012). A standardised method for the quantitative analysis of resistance to grapevine powdery mildew. *European journal of*

plant pathology **133**(2)**:** 483-495.

Mott R, Talbot CJ, Turri MG, Collins AC, Flint J (2000). A method for fine mapping quantitative trait loci in outbred animal stocks. *Proceedings of the National Academy of Sciences* **97**(23)**:** 12649-12654.

Naik S, Gill K, Rao VP, Gupta V, Tamhankar S, Pujar S *et al* (1998). Identification of a STS marker linked to the Aegilops speltoides-derived leaf rust resistance gene Lr28 in wheat. *Theoretical and applied genetics* **97**(4)**:** 535-540.

Niks RE, Marcel TC (2009). Nonhost and basal resistance: how to explain specificity? *New Phytologist* **182**(4)**:** 817-828.

Pallavi J, Singh A, Rao U, Prabhu K (2015). Identification, Validation of a SSR Marker and Marker Assisted Selection for the Goat Grass Derived Seedling Resistance Gene Lr28 in Wheat. *Journal of Plant Pathology & Microbiology* **2015**.

Parlevliet JE (1979). Components of resistance that reduce the rate of epidemic development. *Annual Review of Phytopathology* **17**(1)**:** 203-222.

Powell JE, Visscher PM, Goddard ME (2010). Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics* **11**(11)**:** 800-805.

Rosewarne G, Singh R, Huerta-Espino J, Herrera-Foessel S, Forrest K, Hayden M *et al* (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.

Schnurbusch T, Paillard S, Schori A, Messmer M, Schachermayr G, Winzeler M *et al* (2004). Dissection of quantitative and durable leaf rust resistance in Swiss winter wheat reveals a major resistance QTL in the Lr34 chromosomal region. *Theoretical and Applied Genetics* **108**(3)**:** 477-484.

Shaner G, Finney RE (1980). New sources of slow leaf rusting resistance in wheat. *Phytopathology* **70**(12)**:** 1183-1186.

Singh D, Mohler V, Park R (2013). Discovery, characterisation and mapping of wheat leaf rust resistance gene Lr71. *Euphytica* **190**(1)**:** 131-136.

Singh D, Simmonds J, Park R, Bariana H, Snape J (2009). Inheritance and QTL mapping of leaf rust resistance in the European winter wheat cultivar 'Beaver'. *Euphytica* **169**(2)**:** 253-261.

Singh R, Huerta-Espino J, Bhavani S, Herrera-Foessel S, Singh D, Singh P *et al* (2011). Race non-specific resistance to rust diseases in CIMMYT spring wheats. *Euphytica* **179**(1)**:** 175-186.

Soller M, Brody T, Genizi A (1976). On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *TAG Theoretical and Applied Genetics* **47**(1)**:** 35-39.

Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, Crenshaw AT *et al* (2011). Quality control procedures for genome-wide association studies. *Current protocols in human genetics***:** 1.19. 11-11.19. 18.

Valdar W, Flint J, Mott R (2006). Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics* **172**(3)**:** 1783-1797.

Van der Hoorn RA, De Wit PJ, Joosten MH (2002). Balancing selection favors guarding resistance proteins. *Trends in plant science* **7**(2)**:** 67-71.

Verbyla AP, George AW, Cavanagh CR, Verbyla KL (2014). Whole-genome QTL analysis for MAGIC. *Theoretical and Applied Genetics* **127**(8)**:** 1753-1770.

Vergne E, Grand X, Ballini E, Chalvon V, Saindrenan P, Tharreau D *et al* (2010). Preformed expression of defense is a hallmark of partial resistance to rice blast fungal pathogen Magnaporthe oryzae. *BMC plant biology* **10**(1)**:** 206.

William H, Hoisington D, Singh R, Gonzalez-de-Leon D (1997). Detection of quantitative trait loci associated with leaf rust resistance in bread wheat. *Genome* **40**(2)**:** 253-260.

William H, Singh R, Huerta-Espino J, Palacios G, Suenaga K (2006). Characterization of genetic loci conferring adult plant resistance to leaf rust and stripe rust in spring wheat. *Genome* **49**(8)**:** 977-990.

Xu S (1996). Mapping quantitative trait loci using four-way crosses. *Genetical Research* **68**(02)**:** 175-181.

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF *et al* (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics* **38**(2)**:** 203-208.

Zheng C, Boer MP, van Eeuwijk FA (2015). Reconstruction of genome ancestry blocks in multiparental populations. *Genetics* **200**(4)**:** 1073-1087.