# Whealbi

## Wheat and barley Legacy for Breeding Improvement

Grant agreement number: FP7-613556

**Collaborative Project**
**SEVENTH FRAMEWORK PROGRAMME**

# Deliverable 5.1

## *Meta-dataset including functional validation of candidate genes for race-nonspecific pathogen resistance in barley*

**Due date:** M42

**Actual submission date:** M60

**Project start date:** January 1st, 2014 **Duration:** 60 months

**Workpackage concerned:** 5

**Concerned workpackage leader:** UZH

**Dissemination level:** PU

# Table of contents

# Glossary and Definitions

| | |
|---|---|
| ANOVA | analysis of variance |
| *Bgh* | *Blumeria graminis* f. sp. *hordei* |
| BLUEs | best linear unbiased estimations |
| bp | base pair |
| CGs | candidate genes |
| ExCom | Executive Committee |
| GWAS | genome-wide association study |
| ILA | infected leaf area |
| kb | kilo base |
| OX | overexpression |
| MAF | minor allele frequency |
| MLM | mixed linear model |
| *mlo* | *mildew locus* O |
| RNAi | RNA interference |
| SI | susceptibility index |
| siRNA | small interfering RNA |
| SNP | single nucleotide polymorphism |
| TIGS | transient-induced gene silencing |

# Summary

**Objectives** The aim of this task was to identify and functionally validate new candidate genes (CGs) or new alleles of CGs for race-nonspecific pathogen resistance in barley.

**Rationale:** The results of the phenotyping of 267 barley genotypes in response to two poly-virulent *Blumeria graminis* f. sp. *hordei* (*Bgh*) isolates were reported in D3.4 and formed the basis for the work described here. After a meta-analysis for known resistance genes and in combination with the provided SNP data of 403 Whealbi genotypes, 201 accessions were selected for the calculation of a genome-wide association study (GWAS) using a mixed linear model (MLM). After correcting for multiple testing, the significant SNPs were used to identify CGs based on the current reference genome of barley (Mascher *et al.* 2016). To further validate the CGs, a 'megablast' analysis against related species (like *Aegilops tauschii* and *Brachypodium distachyon*) was performed and the allele status of the confirmed CGs were determined on the basis of the significant SNPs. In consequence of performed amplification tests, the four most promising CGs were selected for functional validation. The generated RNAi (RNA interference) constructs were used for transient assays based on particle bombardment to achieve a transient gene silencing in single cells from a resistant and a susceptible genotype. Additionally, the defined full-length alleles from three of the CGs were cloned in an overexpression (OX) vector and they were also tested by particle bombardment.
The deliverable had to be postponed to M60 because the necessary phenotyping was delayed in the beginning because of lacking seeds and the work required more time than assumed in the beginning due to technical problems.

**Teams involved:** IPK.

# Identification of candidate genes for race-nonspecific resistance in barley against *Blumeria graminis* f. sp. *hordei.*

## 1.1. Material and Methods

### 1.1.1. Genome-wide association study (GWAS)

In D3.4, the results of the phenotyping of 267 barley accessions against the two poly-virulent *Bgh* isolates D35/3 (JKI-242) and RiIII (JKI-75) were reported. Additionally, a meta-analysis was carried out to assess the presence of known resistance genes. The 13 identified *mlo* (*mildew locus* O) carriers and the ten race-specific resistant lines were excluded from the GWAS to achieve a bias for race-nonspecific associations. A final set of 201 genotypes was selected for the association study, because from 47 of our phenotyped barley accessions no exome capture data were provided. From the selected genotypes, the best linear unbiased estimations (BLUEs) for the two isolates were calculated as outlined by Henderson (1975) and Piepho *et al.,* 2003. Moreover, an *in silico* mixed inoculation was performed, which was named *Max*, because always the higher BLUE value from the two isolates was selected in order to increase the bias for race-nonspecific associations.
The genotypic data were again filtered for a minor allele frequency (MAF) of 5 %, which lead to a final set of 424 567 SNPs (single nucleotide polymorphisms). The GWAS itself was carried out as MLM based on the approach suggested by Yu *et al.,* 2006 using the R package ASReml-R (Butler *et al.,* 2009). To correct for the underlying population structure, the Rogers' distances (Reif *et al.,* 2005) were used to estimate the marker-derived kinship matrix. For the determination of the significant threshold, a multiple testing correction called simple$M$ (Gao *et al.,* 2008) was applied and the following criteria were used: (1) the percentage cutoff C was set to 99.8%, (2) the correlation matrix was calculated chromosome wide, (3) the effective number of independent tests $M_{eff\_G}$ was 1414, and (4) $\alpha_e$ was set to 0.05.

### 1.1.2. Identification of CGs

All SNPs with a $-\log_{10}$(p-value) ≥ 4.45 were considered as significant. Because of the focus on race-nonspecific resistance, all significant SNPs form the trait *Max* were used for the determination of the CGs. The significant SNPs that were only significant in the D35/3 and/or RiIII trait were treated as potential race-specific resistances. At least 2 kb of the flanking genomic regions (current reference genome, Mascher *et al.,* 2016) of the significant SNPs from the trait *Max* were extracted with the help of the IPK Galaxy Server and the sequences were used for a BLAST analysis with the BARLEX Server (Colmsee *et al.,* 2015). Based on these results, the corresponding Gene-IDs and gene positions were used to assign the CGs. To confirm the annotated gene models, the 33 identified CGs were verified via 'megablast' analysis with the NCBI (National Center for Biotechnology Information, 1988) Server. The candidate gene models were compared with similar sequences of (predicted) genes in related species like *Aegilops tauschii* subsp*. tauschii* or *Brachypodium distachyon* to confirm especially the start and stop codons of the candidates. Moreover, only these candidates were selected for further analyses where potential cDNA clones and expressed sequence tags from barley could be identified with the help of the BARLEX Server.

The twelve confirmed CGs were analysed concerning their expression in the reference genotype Morex and via PCR based amplification of the determined (full-length) alleles. Moreover, they were analysed in regard of their allele status based on the significant SNPs (trait *Max*). The results were used to select the four most promising CGs for the functional validation tests. For the defined alleles, the average infected leaf area (ILA) was calculated from the BLUEs and to detect significant differences between the alleles, we performed an unpaired, two-tailed t-test using GraphPad Prism 7.01 for Windows.

### 1.1.3. Functional validation tests

To carry out the functional validation, a method called biolistic gene transfer or particle bombardment was used. In these experiments, the CGs were either transiently silenced or overexpressed in single epidermis cells.

#### 1.1.3.1. Transient-induced gene silencing (TIGS)

This approach is based on RNAi via expression of hairpin constructs which produce small interfering RNAs (siRNAs). The SiFi tool (Douchkov *et al.,* 2014) was used to predict the region with the most efficient siRNAs, but without potential off-targets. The cloning of the four selected CGs and the bombardment were performed similarly as described by Douchkov *et al.,* 2005. The bombarded leaves of the resistant (WB-052) or the susceptible genotype (Morex, WB-101) were inoculated 48 h after bombardment with *Bgh* isolate CH4.8 (JKI-247) with a spore density of 450-530 spores/mm$^2$ and the relative susceptibility index was calculated according to the following formula:

$$susceptibility\ index\ (SI) = \frac{\sum GUS\text{-}stained\ cells\ with\ haustorium}{\sum total\ number\ of\ GUS\text{-}stained\ cells}$$

$$relative\ SI\ [\%] = \left(\frac{SI\ of\ test\ gene}{SI\ of\ empty\ vector\ control}\right) * 100$$

To detect possible significant differences, the mean values of the relative SIs per construct from five independent experiments were used for an ordinary one-way ANOVA (analysis of variance) in comparison with the empty vector control, followed by Dunett's multiple comparisons test using GraphPad Prism 7.01 for Windows.

#### 1.1.3.2. Overexpression (OX) approach

For this approach, the defined full-length alleles form three CGs were cloned into pIPKTA9 (Zimmermann *et al.,* 2006), which led to an (over-) expression of the CG alleles under control of the cauliflower mosaic virus 35S promoter. The cloning of the fourth CG failed. The bombardment was performed similarly as described by Douchkov *et al.,* 2005, but without the additional randomized pinning of the leaves after bombardment. The bombarded leaves of WB-052 or Morex were inoculated after 4 h with *Bgh* isolate CH4.8 with a spore density of 150-250 spores/mm$^2$ and the relative SI was calculated. The determination of significant differences was done as described for the TIGS approach, but only from four independent experiments.

## 1.2. Results

### 1.2.1. Analysis of potential race-nonspecific CGs

On the basis of the reported phenotyping results including the previously described meta-analysis for the presence of known resistance genes (D3.4) and the provided exome capture data, an association study with the BLUEs form 201 genotypes was performed. The Figure 1 depicts the resulting Manhattan plots for the three traits: the *in silico* mixed inoculation *Max* and the single isolates (RiIII and D35/3).

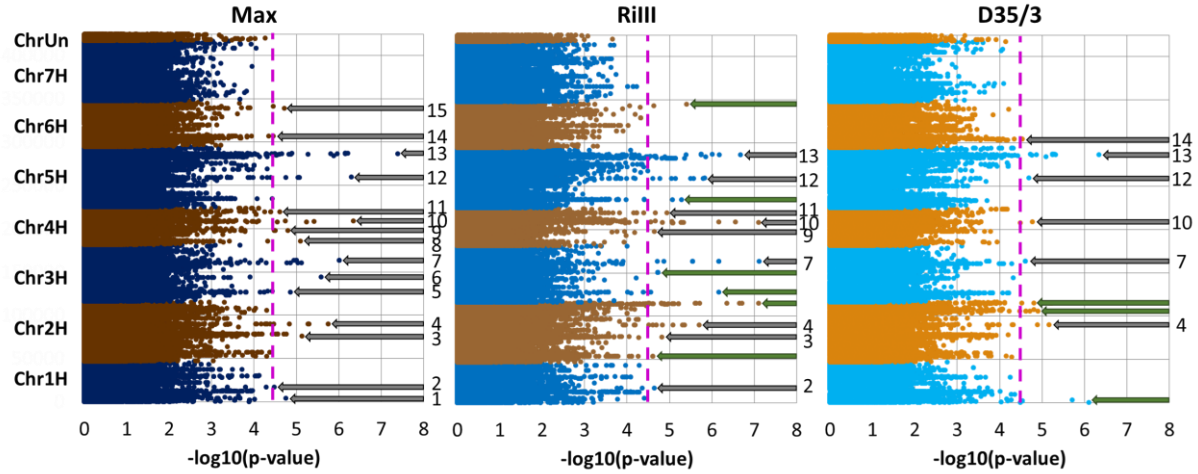


**Figure 1: Manhattan plots of the –log$_{10}$-transformed p-values for the three traits (*Max*, RiIII and D35/3).**
The significance threshold of -log$_{10}$(p-value) = 4.45 is displayed as pink dotted line and the 15 potential race-nonspecific loci were numbered and labelled with grey errors. The assumed race-specific loci were labelled with green errors.

Fifteen potential race-nonspecific loci were identified in the trait *Max* on the chromosomes 1H to 6H. An analysis of the significant SNPs revealed 33 annotated genes based on the current reference genome and the corresponding gene models were further confirmed via a 'megablast' analysis with related species. The Table 1 gives an overview of the twelve confirmed CGs including the total number of SNPs, the number of significant SNPs and the number of determined alleles. All confirmed CGs display only two alleles even if more than one SNP was significant and one of the two alleles was always more frequent in the analysed genotype panel.

**Table 1: Overview of the twelve CGs with confirmed annotation.**
The column 'Candidate' represents the CGs, which annotation were confirmed after the 'megablast' analysis and the column 'Loci number' refers to the numbering of the loci in Figure 1. The column 'Total number of SNPs' designates the number of all SNPs that where located within the annotated gene model and the column 'Number of significant SNPs' represents the number of significant SNPs for the Max trait. These SNPs were used for the determination of the alleles and their total number per CGs is given in the column 'Number of alleles'.

| Candidate | Loci number | Total number of SNPs | Number of significant SNP | Number of alleles |
|---|---|---|---|---|
| WB-CG-6 | 5 | 37 | 1 | 2 |
| WB-CG-7 | 6 | 8 | 1 | 2 |
| WB-CG-8 | 7 | 26 | 3 | 2 |

**Table 2: (*continued*)**

| Candidate | Loci number | Total number of SNPs | Number of significant SNP | Number of alleles |
|---|---|---|---|---|
| WB-CG-11 | 8 | 48 | 1 | 2 |
| WB-CG-13 | 10 | 17 | 1 | 2 |
| WB-CG-14 | 10 | 24 | 1 | 2 |
| WB-CG-17* | 12 | 16 | 1 | 2 |
| WB-CG-19* | 12 | 18 | 1 | 2 |
| WB-CG-20 | 13 | 13 | 1 | 2 |
| WB-CG-23* | 13 | 18 | 1 | 2 |
| WB-CG-28* | 13 | 27 | 2 | 2 |
| WB-CG-31 | 13 | 54 | 5 | 2 |

The asterisk (*) indicates the selected CG's for further analyses.

The twelve confirmed genes were analysed in regard of their expression in the reference genotype Morex and we further tried to amplify the different alleles in various genotypes. On the basis of these results, two CGs form loci 12 and two CGs form loci 13 were selected for further.

Moreover, the effect of the different alleles on the ILA (Figure 2) was evaluated. As described above, always a major allele (A2) and a minor allele (A1) were identified and the results indicate that the less frequent allele (A1) is significantly associated with resistance in all four CGs.

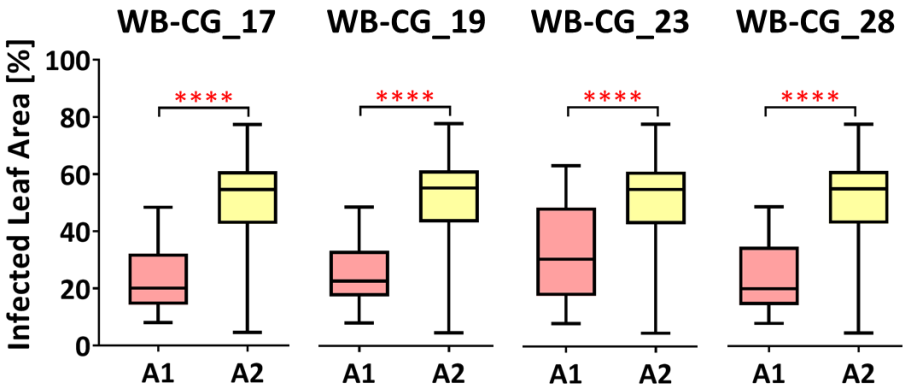


**Figure 2: Box plots of the average infected leaf area (ILA) for the defined alleles**
The significant SNPs form the four selected CGs (WB-CG_17, WB-CG_19, WB-CG_23, WB-CG_28) were used to define the allele status and the average ILA (in %) from the BLUEs were calculated for the different alleles (A1 and A2). The significance was determined with an unpaired, two-tailed t-test with GraphPad Prism 7.01. **** p-value <0,0001

### 1.2.2. Functional validation of CGs

The first validation approach that was carried out, was a transient silencing based on RNAi. The generated constructs should produce siRNAs and thus silence the CGs in the bombarded epidermis cells. The Figure 3 gives an overview of the obtained results.

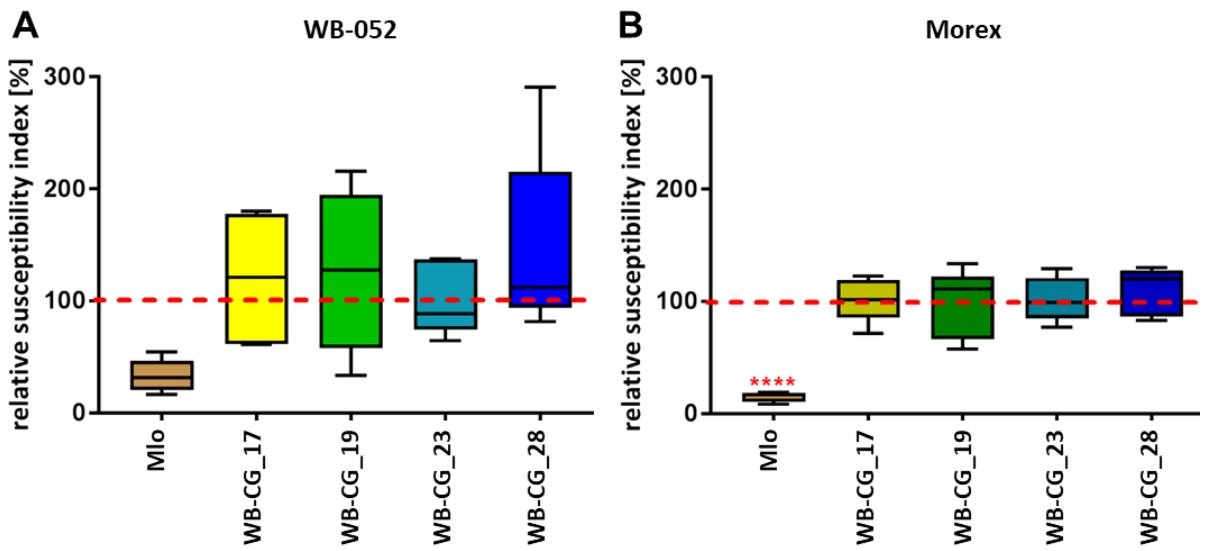


**Figure 3: Box plots of the relative SI values for the four CGs in the TIGS approach**
The relative SIs were calculated according to the above given formulas from five independent experiments for the resistant genotype WB-052 (**A**) and the susceptible genotype Morex (WB-101) (**B**). A construct which silence the *Mlo* gene (Mlo) was used as positive control in all experiments and the significance was tested via ordinary one-way ANOVA in comparison with the empty vector control (which was set to 100 % and is designated as red dotted line), followed by Dunett's multiple comparisons test using GraphPad Prism 7.01, **** p-value <0,0001

In case of the TIGS approach for the resistant genotype (Figure 3A), neither the positive control (*Mlo*) nor the tested CGs showed any significant effect. In contrast to this, the positive control was significantly reduced in the susceptible genotype (Figure 3A), nevertheless none of the CGs constructs showed an effect.
Additionally, an OX approach was performed with the full-length alleles from three CGs. The results are depicted in Figure 4. In case of the susceptible genotype (Figure 4B), none of the constructs showed an effect, but in case of the resistant genotype (Figure 4A), the overexpression of allele A1 from WB-CG_17 led to a significantly increased relative SI and thus to an elevated susceptibility.

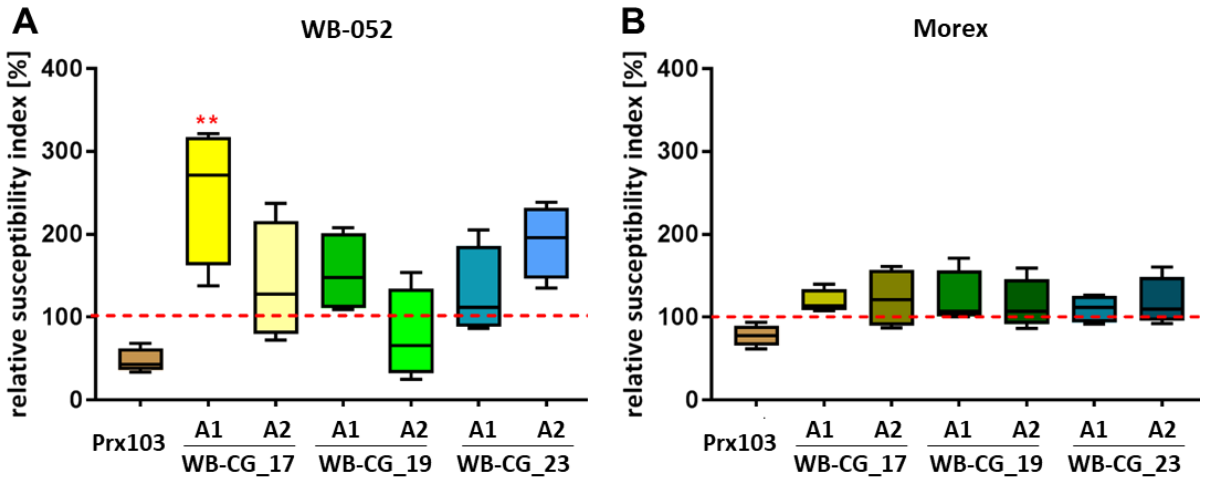


**Figure 4: Box plots of the relative SI values for the three CGs in the OX approach**
The relative SIs were calculated according to the above given formulas from four independent experiments for the resistant genotype WB-052 (**A**) and the susceptible genotype Morex (WB-101) (**B**). The minor allele (A1) and the major allele (A2) were overexpressed for the indicated CGs. A construct which overexpresses a wheat peroxidase (Prx103) was used as positive control in all experiments and the significance was tested via ordinary one-way ANOVA in comparison with the empty vector control (which was set to 100 % and is designated as red dotted line), followed by Dunett's multiple comparisons test using GraphPad Prism 7.01, ** p-value <0,005

## 1.3. Conclusion

The performed association studies led to the identification of 15 potential race-nonspecific resistance loci on the chromosomes 1H to 6H. Within these 15 loci, 33 CGs were annotated on the basis of the current reference genome and from twelve of them the annotated gene models could be confirmed via 'megablast' analysis. The significant SNPs were used to evaluate the allele status of the twelve CGs. In all CGs, a less frequent (minor) allele and a more frequent (major) allele was identified even if more than one SNP was significant.
Based on amplification tests, the four most promising CGs from chromosome 5H were selected for further analyses. To determine the effect of the different alleles, the average ILA for the two alleles per CGs was calculated and the results indicate that the minor allele is always associated with resistance. The selected CGs were used in TIGS and OX assays to functionally validate them. In the silencing approach none of the CGs showed an effect but the OX of the minor allele (A1) of WB-CG_17 led to a significant increased susceptibility.

## 1.4. References

**Butler D. G., Cullis B. R., Gilmour A. R. and Gogel B. J.** (2009) *ASREML-R Reference Manual. Release 3.0.* Technical Report, Queensland Department of Primary Industries, Australia.

**Colmsee C., Beier S., Himmelbach A., Schmutzer T., Stein N., Scholz U., Mascher M.** (2015) *BARLEX – the barley draft genome explorer.* Molecular Plant **8:** 964-966
URL: http://apex.ipk-gatersleben.de/apex/f?p=284:10:::::: (10.12.2018)

**Douchkov D., Nowara D., Zierold U., Schweizer P.** (2005) *A high throughput gene-silencing system for the functional assessment of defense-related genes in barley epidermal cells.* Molecular Plant-Microbe Interaction Journal **18:** 755-761

**Douchkov D., Lück S., Johrde A., Nowara D., Himmelbach A., Rajaraman J., Stein N., Sharma R., Kilian B., Schweizer P.** (2014) *Discovery of genes affecting resistance of barley to adapted and non-adapted powdery mildew fungi.* Genome Biology **15:**518

**Gao X., Starmer J., Martin E. R.** (2008) *A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms.* Genetic Epidemiology **32:** 361-369

**Henderson C. R.** (1975) *Best linear unbiased estimation and prediction under a selection model.* Biometrics **31:** 423-447

**Mascher M., Gundlach H., Himmelbach A., Beier S., Twardziok S. O., Wicker T., Radchuk V., Dockter C., Hedley P. E., Russell J., Bayer M., Ramsay L., Liu H., Haberer G., Zhang X.-Q., Zhang Q., Barrero R. A., Li L., Taudien S., Groth G., Felder M., Hastie A., Šimkova H., Staňkova H., Vrana J., Chan S., Munoz-Amatriain M., Ounit R., Wanamaker S., Bolser D., Colmsee C., Schmutzer T., Aliyeva-Schnorr L., Grasso S., Tanskanen J., Chailyan A., Sampath D., Heavens D., Clissold L., Cao S., Chapman B., Dai F., Han Y., Li H., Li X., Lin C., McCooke J. K., Tan C., Wang P., Wang S., Yin S., Zhou G., Poland J. A., Bellgard M., Borisjuk L., Houben A., Doležel J., Ayling S., Lonardi S., Kersey P., Langridge P., Muehlbauer G. J., Clark M. D., Caccamo M., Schulman A. H., Mayer K. F. X., Platzer M., Close T. J., Scholz U., Hansson M., Zhang G., Braumann I., Spannagl M., Li C., Waugh R., Stein N.** (2017) *A chromosome conformation capture ordered sequence of the barley genome.* Nature **544:** 427-448

**National Center for Biotechnology Information (NCBI)** (1988) Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information.
URL: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch (10.12.2018)

**Piepho H.P., Büchse A., Emrich K.** (2003) *A hitchhiker's guide to mixed models for randomized experiments*. Journal of Agronomy and Crop Science **189:**310–322

**Reif J. C., Melchinger A. E., Frisch M.** (2005) *Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management.* Crop Science **45:** 1-7

**Yu J., Pressoir g., Briggs W. H., Vroh Bi I., Yamasaki M., Doebley J. F., McMullen M. D., Gaunt B. S., Nielsen D. M., Holland J. B., Kresovich S., Buckler E. S.** (2006) *A unified mixed-model method for association mapping that accounts for multiple levels of relatedness.* Nature Genetics **38:** 203-207

**Zimmermann G., Bäumlein H., Mock H.-P., Himmelbach A., Schweizer P.** (2006) *The multigen family encoding Germin-Like proteins of barley. Regulation and function in basal host resistance*. Plant Physiology **142:** 181-192