Final report for DFG-grant RO 1055/11-1

Title: Map-based cloning and functional validation of a QTL for grain size in wheat

PI:
Marion Röder, Group Gene- and Genome Mapping, IPK Gatersleben

Co-PI:
Jochen Kumlehn, Group Plant Reproductive Biology, IPK Gatersleben

Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3
06466 Stadt Seeland, OT Gatersleben

Duration of project:
1.02.2012- 31.1.2015

Internal Collaborator:

Dr. Nils Stein and Dr. Axel Himmelbach, Group Genome Diversity, IPK Gatersleben

External Collaborator:

Dr. Jaroslav Dolezel and Dr. Hana Simkova, Institute of Experimental Botany, Olomouc, Czech Republic

1. Goals of the project

The goals of the project were the further fine mapping and establishment of a physical map for QTL Q.*Tgw.ipk-7D* for grain size in wheat. The verification of the obtained candidate genes was established by sequencing in parental and selected near-isogenic lines (NILs) and the search for SNPs which alter amino acids in the coding sequence. A functional validation of selected candidate genes by knock-out using the RNA-guided nuclease (RGN) system (Shan et al., 2014) has been initiated.

2. Scientific results

2.1 Introduction

The previously described QTL for thousand-grain weight *QTgw.ipk-7D* associated with microsatellite marker *Xgwm1002-7D* was originally detected in a BC_2F_3 advanced backcross population of the winter wheat variety 'Prinz' and the synthetic wheat line W-7984 (lab designation: M6) (Huang *et al*. 2003). We developed near-isogenic lines (NILs) carrying introgressions of M6 in the genetic background of 'Prinz' with varying size on chromosome 7DS. The BC_4F_3 NILs had a 10% increase in thousand-grain weight compared to the control group and the recurrent parent 'Prinz' (Röder et al. 2008). The same QTL was detected in another population of winter wheat 'Flair' and synthetic wheat 'XX86' (Huang *et al*. 2004). By using homozygous recombinant lines developed from both populations, it was possible to fine-map *QTgw.ipk-7D* to an interval of approx. 1 cM flanked by markers *barc126*, *wmc405* and *Xgwm44* on wheat chromosome arm 7DS. From a chromosome arm 7DS-specific BAC library, BACs covering the region of *QTgw.ipk-7D* were isolated and their sequences were obtained by Next Generation Sequencing (NGS). Of the sequenced BACs, new microsatellite markers were developed and used for anchoring the BACs to the genetic map. Finally, the region of *QTgw.ipk-7D* was delimited to 14 BACs carrying ca. 16 predicted genes.

2.2 Phenotyping data

From the cross between 'Prinz' and 'M6', three different NIL-populations were derived (20 to 61 plant families per population). And from the cross between 'Flair' and 'XX86', one NIL-population with 84 plant families was derived. One NIL from each cross was chosen to be analyzed in more detail ('Prinz' 34d/1-116m; 'Flair' 130-1-1B-30G) in comparison with its recurrent parent.

Plants were grown in a greenhouse under field-like conditions and harvested in August 2014. Plant height (without roots and ears), grain mass and Harvest Index were given as mean of ten plants, whereas thousand grain weight (TGW), grain area and grain length were measured by MARVIN (GTA Sensorik GmbH, Neubrandenburg, Germany) and given as mean of six sample subsets. TGW, grain area and grain length were significantly different between NILs and their parents ($p<0.05$, T-test), plant height and Harvest Index (HI) were significant only for one NIL and its parent. (Fig. 1)

To determine fresh and dry weight of developing grains, 3x 10 grains per time point from 3 to 5 different ears were harvested. First differences between one NIL and its parent were observed at 18 daf (days after flowering), in the middle of the grain filling/storage phase. (Fig. 2)

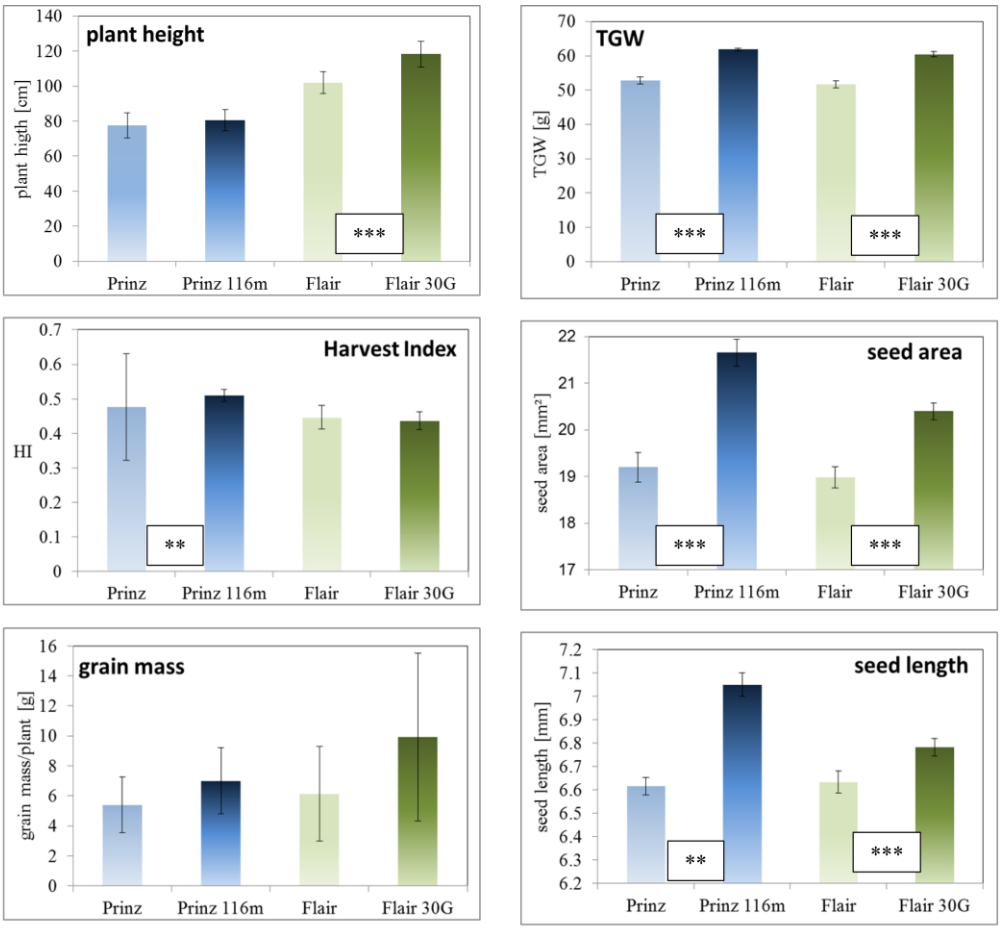


Fig. 1: Phenotypic analysis of NILs and recurrent parents. Asterisks mark significant differences between the NIL and its respective parent.

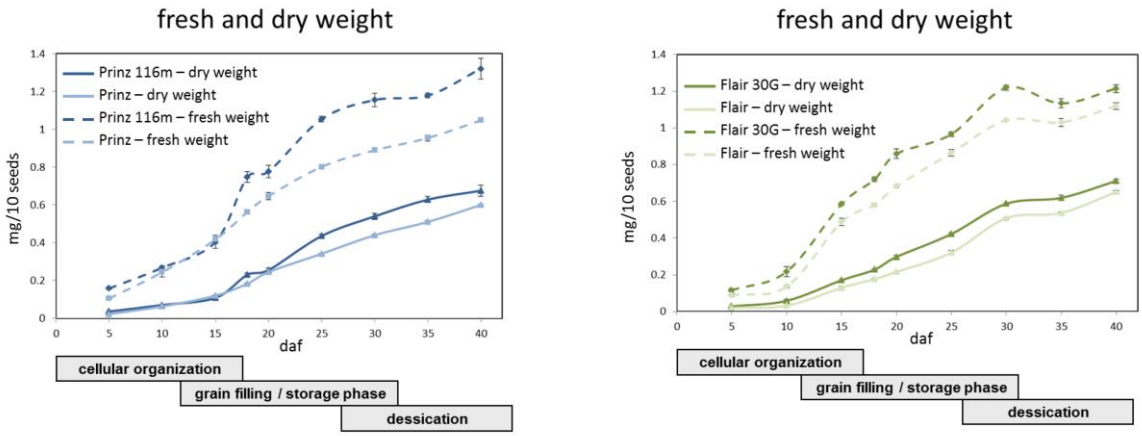


Fig. 2: Grain development of NILs and recurrent parents. Fresh and dry weight were determined based on 10 kernels per datapoint. Differences between NILs and recurrent parents appear from 18 daf (days after flowering).

In addition, the parents show an anthesis delay compared to the NILs. 'Prinz' is 4d (2014) or 6d (2010) later than 'Prinz' 34d/1-116m, whereas ears of 'Flair' 130-1-1B-30G emerged 1-2 days earlier than ears of 'Flair'. (Fig. 3)

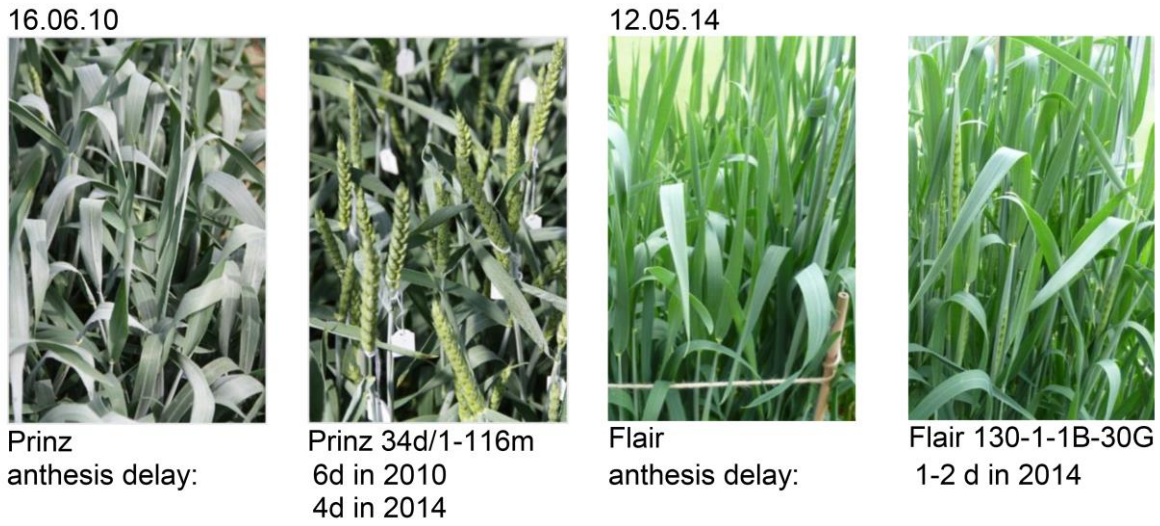


Fig. 3: Anthesis delay of NILs compared to recurrent parents.

2.3 Comparison of the physical map with the genetic map

The physical map of the QTL region *QTgw.ipk-7D* was established based on results of the lab of Jaroslav Dolezel , Olomouc, Czech Republic. It contains one large contig with 38 sequenced BACs and several gaps. It starts with two overlapping BACs containing the flanking markers barc126 or wmc405. This is followed by a gap of unknown size and the contig with 38 BACs. According to a comparison of our sequence data with the optical map of chromosome 7DS, the gap between our BAC-contig and the BACs 99N06 and 76M16 which contain the marker Xgwm44 is about 1 Mbp (Helena Staňková and Hana Šimková, not published). In Fig. 4, only BACs which contain analyzed markers are shown.

In collaboration with Nils Stein (IPK Gatersleben) 16 BACs of the respective genomic region were also retrieved from barley and the sequences were obtained.

Fig. 4 gives an overview of the QTL *QTgw.ipk-7D* and compares this region in wheat with barley, rice *Sorghum bicolor* and *Brachypodium distachyon*. For the three sequenced species, we show the region of 16 candidate genes. In barley we show the minimum tiling path (MTP) of a BAC contig in this region, whereas in wheat all sequenced BACs between the new flanking marker 100H_c9 and the BAC containing gene 33 are shown.

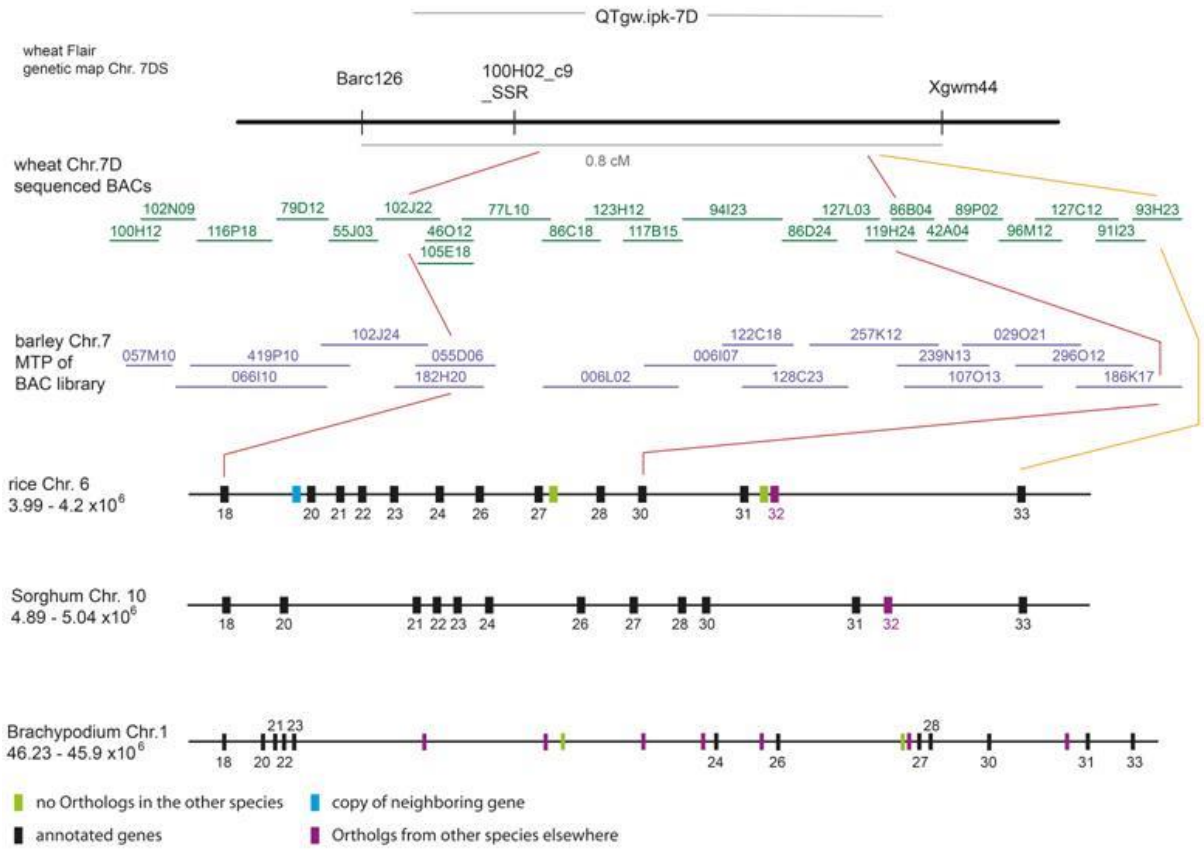


Fig. 4: Comparison of region *QTgw.ipk-7D* in wheat with barley, rice, sorghum and *Brachypodium.*

2.4 Synteny of 16 candidate genes

To identify candidate genes, the next generation sequencing (NGS) sequences derived from the selected BACs were analyzed with two gene prediction tools: For the TriAnnot platform, only assembled sequence contigs larger than 10 kb were used (http://urgi.versailles.inra.fr/Species/Wheat/Triannot-Pipeline), whereas for riceGAAS (http://RiceGAAS.dna.affrc.go.jp), all assembled sequence contigs larger than 1 kb were used. All predicted genes were named a candidate gene, provided they have a known function and mRNA or cDNA homologs in the databases. A short summary of all predicted 16 candidate genes is given in Table 1.

The 16 candidate genes show a high synteny in the five analyzed species (Annex, Fig. S1). They have the same order like in rice, with three exceptions:

- In wheat, three genes are duplicated compared to rice. The *OXIDOREDUCTASE* and the *RIBOSOMAL PROTEIN L39* are located in between the two variants of the *MYB*-factor and of the *ETHYLENE RESPONSE FACTOR 1*, respectively, whereas the two variants of the *hAT* family protein are probably next to each other.
- In barley there are three copies of the *MYB*-factor next to each other. But the two copies with closer homology to *MYB*-factor A or *MYB*-factor B from wheat are probably not expressed, because they have premature stop-codons. And the predicted *MYB*-factor shows less than 60 % homology to *MYB*-factor A or *MYB*-factor B in wheat.

- In wheat and barley there are two genes inserted compared to rice*, XBAT32* and *BTB/POZ* containing protein are located between *ETHYLENE RESPONSE FACTOR 1* and *PGPS/D10*.

In addition, gene *PGPS/D10* (differentially expressed cDNA-clone for pollen germination in Petunia) in wheat is 156 bp longer in the 5' region compared with the homologues from the other species. In barley the same sequence (94% identity) is present, but has a 1 bp deletion at position 26 (in reference to the wheat starting ATG) and therefore a premature stop-codon.

Table 1: Comparison of 16 candidate genes in wheat and barley

wheat				**barley**		
gene function	BAC_contig	Exons	Unigene	BAC_contig	Exons	Unigene
put. Proteinkinase	7D046O12_c2	9	TC414900, CJ693419, TC444442	055D06_c3	>9	TC242954, TC244989
put. Uncharacterized protein LOC100382581;	7D046O12_c2 , 7D105E18_c1	2	CA620389, AL829386	006I07_c2, 006L02_c5	2	TC257260, BY857232
put. Oxidoreductase	7D046O12_c2	8	TC457299, CJ670376	006I07_c2, 006L02_c5	6	TC268595, TC261327
put. Uncharacterized protein LOC100382581 Myb, DNA-binding	7D077L10_c11	2	TC337080, CA742239	006I07_c2, 006L02__c5	2	TC255742, DN157756
putative ethylene response factor1, AP2 domain	7D123H12_c7	2	CA621787, BE516792	128C23_c3	3	TC266556, BI958427
ribosomal protein L39	7D123H12_c2	22	AY846827, CJ563260,	006I07_c8, 06L02_c3	1	TC241452, TC222704
putative ethylene response factor1	7D123H12_c2	2	CA622305, CA621787, BE516792	128C23_c3	3	TC266556, BI958427
putative XBAT32	7D094I23_c2	11	TC455094, TC417319, BJ244067, CA639287	107O13_c1	8	TC203895, TC199576, CB883363, TC247207
put. BTB/POZ domain containing prot.; protein binding	7D094I23_c17	1	CA666507, TC423824, TC443226, BQ246251	257K12_c1	1	TC200526, TC204314, TC265124, TC279604
put. PGPS/D10	119H24_117	2	BE428919	296O12_c1	2	BM370638, TC251667

put. ethylene-responsive element-binding protein	119H24_116	1	TC290865, TC290182, DR734500, BJ306475	186K17_c1	1	TC270288, BI957055
putative beta-1,3-glucan synthase	119H24_116	8	CJ658004	186K17_c2	5	TC240327, TC265275
putative uncharacterized protein LOC100193128	89P02_c19 , 42A04_c2	7	AL819528	morex_contig_1561853 morex_contig_156087	9	AK369705
putative hAT family dimerisation domain containing protein	91I23_c21, 127C12_c20	4	CJ716875, CA677290, TC304725	morex_contig_45950/morex_contig_1560870	6/8	MLOC_61880.1 /MLOC_11341.2
Two component response regulator ARR18	93H23_c3, 127C12_c25	7	CJ716875, CA677290, TC304725	morex_contig_45950/morex_contig_1560870	6/8	MLOC_61880.1 /MLOC_11341.2

2.5 Sequencing of 16 candidate genes

In order to detect SNPs at chromosome 7D, sequencing of candidate genes in the parental lines and the NILs was performed (Fig. 5). One of the candidate genes could be responsible for observed differences in grain size if there are SNPs between 'Prinz' and 'M6' and these SNPs have the same variant in 'Prinz' and NIL_{Prinz} as well as in 'M6' and NIL_{M6} (b in Fig. 5). If candidate genes have no differences between NIL_{Prinz} and NIL_{M6} or if 'M6' has the same SNP like NIL_{Prinz} (or vice versa), then it could not explain differences in grain size (a and c in Fig. 5)

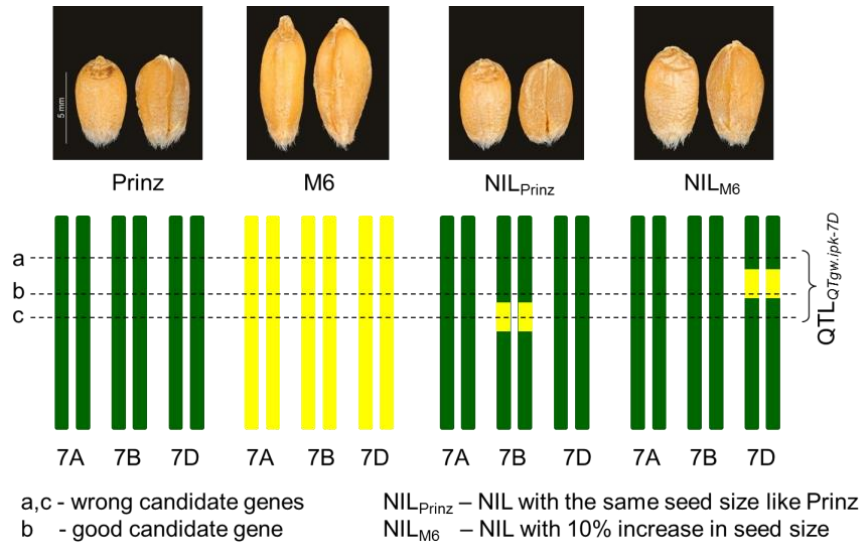


Fig. 5: Grain phenotype and strategy of genome identification

Until now, 13 candidate genes were analyzed by cloning gene-specific PCR products into a cloning vector and sequencing inserts from about ten single clones per line. 7D-specific SNPs and other 7D-specific sequence variations are summarized in Table 2. Only twelve SNPs induce an amino acid substitution. All other detected 7D-specific sequence variations are silent mutations or are located in non-coding regions.

Table 2: Summary of identified 7D-specific sequence variations between parents

Gene or protein	No. of Sequence variations	located in: Exons	Introns	UTRs
PROTEIN KINASE	19	11 (3*)	6	2
MYB factor A	6	2	2	2
OXIDOREDUCTASE	18	6 (3*)	10	2
MYB factor B	18	7 (2*)	9	2
ETHYLENE RESPONSE FACTOR 1 A	0			
ETHYLENE RESPONSE FACTOR 1 B	0			
put. *XBAT32*	7^{+}	1*	6	0
put. BTB/POZ domain containing protein	1^{+}	0	0	1
PGPS/D10	5^{+}	3 (1*)	2	0
ethylene-responsive element-binding protein	0			
Metallophosphoesterase domain cont. protein	1^{+}	0	1	0
hAT family dimerisation domain containing protein	1^{+}	1*	0	0
Two component response regulator	3^{+}	2 (1*)	0	1

* SNP inducing an amino acid change $^{+}$ only detected in 'Flair' x 'XX86'

2.6 Mapping and new markers

BAC-derived new SSR-markers as well as candidate gene-specific new CAPS and InDel-markers were developed. All types were analyzed in the four crossing populations. Together with the phenotype data, this was used for QTL calculations with program Qgene version 4.3.10 (http://coding.plantpath.ksu.edu/qgene/) to evaluate LOD scores in the single marker regression mode and with GenStat v14 as Single Trait Linkage Analysis.

When the project started, there were already 9 segregating markers analyzed, located between the flanking markers barc126 and Xgwm44. In the new available BAC sequences, 111 potential SSR loci were identified (with repetitive motifs > 7 repetitions). From these potential SSR loci, 30 have a triplet repeat and 81 a doublet repeat. From 52 analyzed potential SSR loci, 8 turned out to be

polymorphic between parents and segregating in the populations. These 8 SSR-loci are located on 5 different BACs.

From 16 candidate genes, 13 were sequenced to identify sequence variations which might be responsible for the QTL. From these identified sequence variations, we developed 7 CAPS-markers and 7 InDel-markers located on 5 different BACs (without repetition in other BACs).

2.7 QTL calculation

All mapped markers were subject to QTL calculation by single marker regression. In the population 'Prinz' 34a/4, LODs are above 3 only for the traits TGW2009, grain area2009, grain length2009 and grain length2010. The highest LODs are calculated for the markers 96M_c23 to Xgwm44 in all traits in the year 2009. For trait length2010, highest LODs are located at the borders of the analyzed region (marker Xswm10 to barc154 and from 96M_c23 to Xgwm974). Population 34d/1 gives a different picture: LODs are similar in the region wmc436 to Xgwm974 or from gpw229 to Xgwm974 for one trait. Between the traits, LODs ranged from 3 to 13.8, with three exceptions where the LODs are smaller than 3. So with these calculations no assumption for the localization of the QTL within this region could be done. In the 'Flair'-population 130-1-1B LODs are higher than 3 between the markers 100H_c9 and Xgwm676. Between W12C2 and Xgwm676 LODs are smaller than in the region between 100H_c9 and W12C2. So our assumption would be that the QTL is located between the markers 100H_c9 and W12C2 (Supplemental Figure S2)

Because QTL-calculation did not delimit the region, we looked at all lines with recombination events between the markers barc126 and Xgwm974 in one population and whether these lines have small or large grains. In the population 'Flair' 130-1-1B, all lines which have the 'XX86' genotype between the markers 100H12_c9 and 86C18_c4 have large grains (one exception line 78E) and all lines which have 'Flair' genotype in this region have small grains. The lines 99B and 99D delimits the right border and line 122D delimits the left border with marker 100H12_c9 being outside of the region (Supplemental Figure S3). This leads us to the conclusion that the smallest region for the QTL determining seed size is located between the markers 100H12_c9 and 86C18_c4. This region includes four candidate genes: one *ZINC FINGER PROTEIN 103*, a *PROTEIN KINASE*, an *OXIDOREDUCTASE* and a *MYB* factor.

2.8 Expression analysis of six candidate genes

Expression analysis of NILs and the respective recurrent parents yielded various patterns after different time points in grain development.

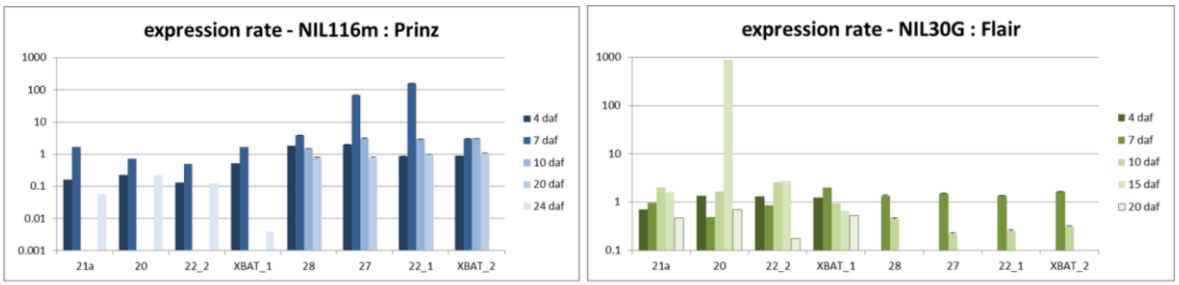


Fig. 6: Relative expression of 6 candidate genes in NIL116m and in NIL30G. Expression was normalized to *UBIQUITIN* and shown in relation to the recurrent parent. Expression rates larger than 1 denotes that the expression in the NIL is higher than in the parent and if the expression rate is below 1, the expression in the NIL is less than in the parent.

20 = *PROTEIN KINASE*; 21a = *MYB* factor A; 22_1 = *OXIDOREDUCTASE*, with 2 different primer pairs amplified; 27 = *PGPS/D10*; 28 = ethylene-responsive element binding-protein; XBAT_1 = *XBAT32*, with 2 different primer pairs amplified.

2.9 Outlook: Silencing of two candidate genes in planta

The smallest region including the QTL contains four genes: a *ZINC FINGER PROTEIN*, a *PROTEIN KINASE*, a *MYB* factor and an *OXIDOREDUCTASE*. To analyze whether one of these genes is the responsible gene for higher TGW, we plan to silence them via RNA-guided nuclease technology in the wheat lines 'Certo' and 'Bobwhite' which contain the same alleles of the respective candidate genes like 'Flair' and 'Prinz'.

For two genes, we chose three different target sites and start cloning. Prior to generating stable transgenic plants, we plan to test the chosen target sequences in a transient assay. To this end, we needed to clone our target sequences in two different plasmids.

The plasmid p6i-2x35S-TE9 (DNA-Cloning-Service, Hamburg, Germany) will be used for generation of stable transgenic plants via *Agrobacterium*-mediated transfer of the transgene into pre-cultured immature embryos of wheat. Since binary vectors are more difficult to handle during cloning procedures assembly of *Cas9* endunuclease and target specific protospacer with guide-RNA scaffold each with own promotor and terminator is generated in a high copy vector called pGH193-insert.

The second intermediate plasmid pGH189 is used for the transient assay and contains a GFP coding sequence driven by a constitutive promoter. After insertion of our target sequence the *gfp* gene become nonfunctional and was designated pGH189-insert. Co-bombardment into barley leaves leads to reconstitution of *gfp* in 1/3rd of the cases rendering cells fluorescent which could be counted under a fluorescent microscope.

Table 3 gives an overview of the chosen target sites in the *PROTEIN KINASE* and in the *MYB* factor B and for the status of cloning. Testing of the plasmids in the transient assay has been initiated in the group of Jochen Kumlehn.

Table 3: Targets for gene silencing using the CRISP/Cas system.

target sequence	position	description	name	pGH189 transient	pGH193 shuttle vector	P6i-2x35S-TE
Protein kinase						
TACTCTGGGTACGAACTCCA	+1122	Exon 1	A	+	+	+
TGCGAACAACGTCTATCCTG	+2274	Exon 1	B	+	+	+
TCCGAGCCCTAGGACGGCGG	-28	5' UTR	C	+	+	+
***MYB* factor B**						
CGCGCCAAATACCGCATACT	+619	Exon 1	D	+	+	+
CGCGGTCGGCCCGTCTCGCC	+528	Exon 1	E	+	+	+
GGCGGCAAGAAAAGATGCCG	-14	5' UTR/Exon 1	F	+	+	+

2.10 Industry collaborations resulting from this research

This research has resulted in an interaction with the company Bayer CropScience, who is testing several of the developed NILs for their performance in the field considering TGW and grain yield.

3. Summary

The genomic region of thousand grain weight QTL Q.*Tgw.ipk-7D* in wheat was cloned in a BAC contig of 38 sequenced BACs in wheat and 16 sequenced BACs of the syntenic region in barley. Synteny was established to the annotated genome sequences of rice, sorghum and *Brachypodium*. A total of 16 candidate genes were detected. Further fine mapping and marker development narrowed down the region to the four most likely candidates, one *ZINC FINGER PROTEIN 103*, a *PROTEIN KINASE*, an *OXIDOREDUCTASE* and a *MYB* factor. These were characterized by expression analysis in developing grains. Functional verification of the target genes by knockout via the RNA-guided nuclease technology has been initiated.

The developed near isogenic lines carrying Q.*Tgw.ipk-7D* showed increased thousand grain weight, grain area and grain length, when tested under field like conditions in a green house. Currently a number of the lines are evaluated for their agronomic performance in the field in collaboration with an industrial partner.

Overall, the obtained results can be considered as a major step towards the identification of genes underlying a TGW-QTL in wheat.

Cited Literature:

Huang XQ, Cöster H, Ganal MW, Röder MS (2003) Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). Theor Appl Genet **106**: 1379-1389

Huang XQ, Kempf H, Ganal MW, Röder MS (2004) Advanced backcross QTL analysis in progenies derived from a cross between a German elite winter wheat variety and a synthetic wheat (*Triticum aestivum* L.). Theor Appl Genet **109**: 933-943

Röder MS, Huang XQ, Börner A (2008) Fine mapping of the region on wheat chromosome 7D controlloing grain weight. Funct Integr Genomics **8**: 79-86

Shan Q, Wang Y, Li J, Gao C (2014) Genome editing in rice and wheat using the CRISPR/Cas system. Nature Protocols **9**: 2395-2410

Presented Lectures:

JAENECKE, C.: Map-based cloning and functional validation of a QTL for grain size in wheat. Plant Genetics and Breeding Technologies, Vienna / Austria, 18.-20.02.13

Presented Posters:

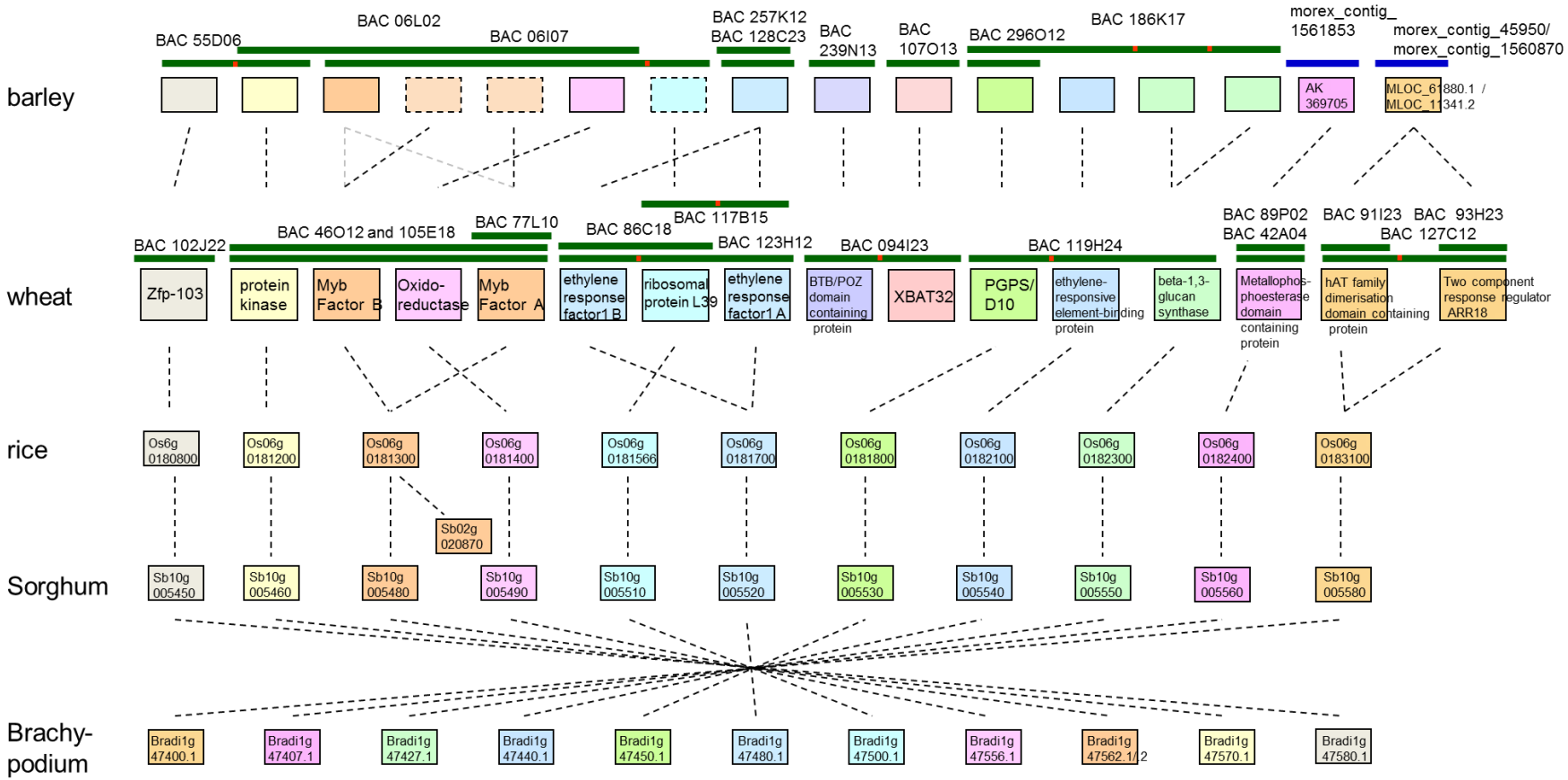
JAENECKE, C., N. WEICHERT, H. ŠIMKOVÁ, J. DOLEŽEL, D. WOLFF, U. SCHOLZ, A. HIMMELBACH, N. STEIN, M.S. RÖDER: Grain size QTL region *QTgw.ipk-7D* in wheat: sequence analysis and synteny to related grass species. GPZ: 10th conference – Plant Stress and Genomics, Halle/Saale, 18.-20.09. 2012

JAENECKE, C., C. ZANKE, N. WEICHERT, H. ŠIMKOVÁ, J. DOLEŽEL, D. WOLFF, U. SCHOLZ, A. HIMMELBACH, N. STEIN, M.S. RÖDER: Grain size QTL region *QTgw.ipk-7D* in wheat: sequence analysis and synteny to related grass species. Workshop Molekulare Präzisionszüchtung, Gatersleben, 12.-13.09.2013

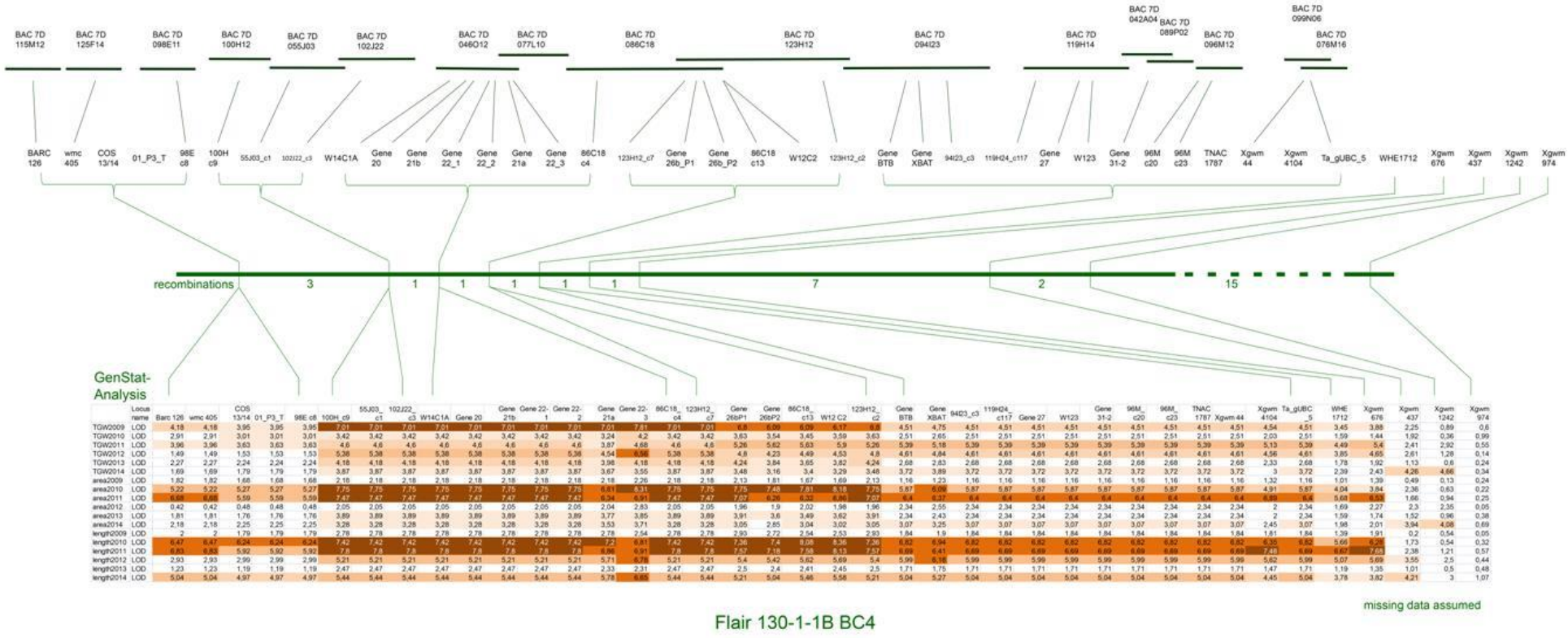
JAENECKE, C., N. WEICHERT, H. ŠIMKOVÁ, H. STAŇKOVÁ, J. DOLEŽEL, D. WOLFF, U. SCHOLZ, A. HIMMELBACH, N. STEIN, M.S. RÖDER: Grain size QTL region *QTgw.ipk-7D* in wheat: sequence analysis and synteny to related grass species. EUCARPIA conference, Wernigerode, 29.06.-04.07. 2014

JAENECKE, C., N. WEICHERT, H. ŠIMKOVÁ, H. STAŇKOVÁ, J.DOLEŽEL, F. ANDERSCH, U. SCHOLZ, S. BEIER, A. HIMMELBACH, N. STEIN, M.S. RÖDER: Analysis of candidate genes for grain size QTL *QTgw.ipk-7D* in wheat. Institute's day IPK, Gatersleben, 8.-9.10.2014

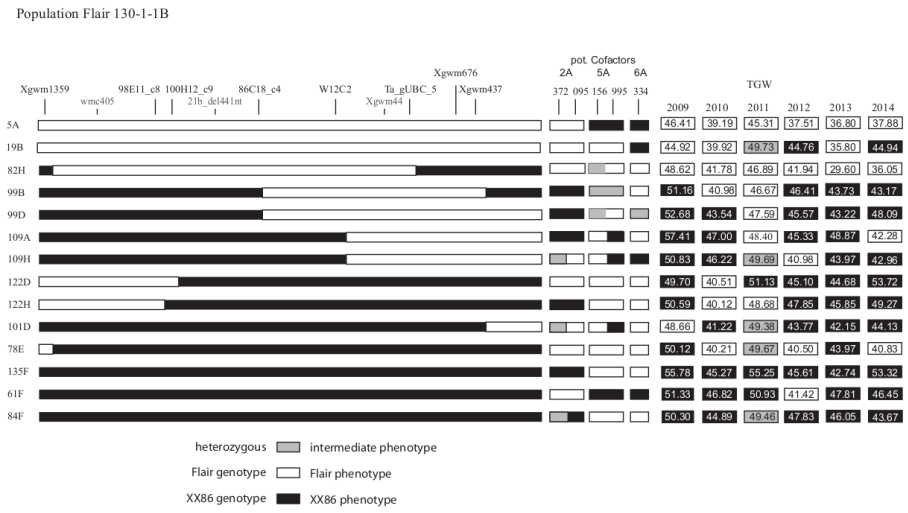
Annex:



Supplemental Figure S1: Synteny relationships of 16 candidate genes in wheat, barley, rice, *Sorghum bicolor* and *Brachypodium distachyon.* Colored rectangles represent genes (one color one gene) where a transcript exists (rice, *Sorghum bicolor* and *Brachypodium distachyon*) or is assumed (barley and wheat). Rectangles with dashed lines represent genes, where the transcript probably does not exist, because of frameshift mutations and early stop codons. Green bars represent BACs covering the genomic region of *QTgw.ipk-7D* in wheat or barley and red dots mark the border between sequence contigs within a BAC. Blue bars represent sequencing contigs.



Supplemental Figure S2: Combined physical and genetic map and results of QTL-analysis of population 'Flair' 130-1-1B BC4. In the upper part of the picture are shown all BACs with analyzed markers in their putative order. In the next line are all analyzed markers in their putative order. In the next line is the genetic map of the population 'Flair' 130-1-1B shown with number of recombinations between two groups of markers. The bottom shows a table with LOD scores of 18 analyzed traits calculated with GenStat v14 and Single Trait Linkage Analysis. LODs above 3 are colored.



Supplemental Figure S3: Recombination events and trait TGW in population 'Flair' 130-1-1B BC4. The lines 99B and 99D delimits the right border and line 122D delimits the left border with marker 100H12_c9 being outside of the region. This leads us to the conclusion that the smallest region for the QTL determining seed size is located between the markers 100H12_c9 and 86C18_c4.