

# Assoziationskartierung in Gerste - Ein genomweiter und Kandidatengen-Ansatz

## Association mapping in barley - a genome wide and candidate gene approach

Inge E. Matthies<sup>1\*</sup>, Stephan Weise<sup>1</sup> und Marion S. Röder<sup>1</sup>

### Abstract

Diversity Arrays Technology (DArT) markers and Simple Sequence Repeats (SSRs) were applied to investigate genetic diversity (kinship), population structure, and extent of linkage disequilibrium (LD) on a genome wide level in European barley cultivars. While DArT technology detects the presence versus absence of individual fragments, SSRs comprise a co-dominant multiallelic marker system. DArT technology enables a whole genome profiling of crops such as barley without the need of sequence information. In total, 1915 DArT markers were polymorphic and 1088 of them were already mapped. Out of these, 253 were selected every 5 cM from the linkage map and used for LD and association studies for each chromosome. Population structure was also revealed in the spring and winter subpopulations. For the total set of 190 cultivars, five groups were found. The results for population structure employing SSR analysis were compared to the results obtained with DArTs. LD studies with both random marker sets showed clear differences about extent of LD depending on the population structure (spring or winter growth habit, row number) and according to the chromosomal location. The available data will be used to elucidate significant associations for important malting and kernel quality parameters with DArT and SSR data. Furthermore, in a structural genomic approach candidate genes with impact on malting quality were investigated and association results from two of them (Catalase1 and Phospholipase D-like protein) are presented exemplarily.

### Keywords

Association genetics, candidate genes, DArT, *Hordeum vulgare*, population structure, SNP

### Einleitung

Im Rahmen des GABI-GENOBAR-Projektes werden derzeit vier verschiedene Ansätze zur Detektion von signifikanten Beziehungen zu wichtigen Malz- und Kornqualitätsparametern verfolgt: (1) SNPs (single nucleotide polymorphisms) und INDELs (Insertionen-Deletionen) aus Kandidatengen mit bekannter Funktion im Kohlenstoff- und Stickstoffme-

tabolismus sowie zufällig im Genom verteilte (2) SSR- und (3) DArT-Marker sowie (4) Illumina-Marker.

Hier werden Beispiele für Assoziationsstudien basierend auf dem (a) genomweiten Ansatz mit DArTs und (b) Ergebnisse einiger der insgesamt 32 studierten Kandidatengene vorgestellt. Catalase 1 (*Cat1*) kommt in Peroxisomen vor und repräsentiert zusammen mit Superoxid-Dismutasen (SODs) und Hydroperoxidasen einen Teil des zellulären Abwehrmechanismus, in dem alle als Radikalfänger agieren. Das Phospholipase D-like Protein (*PLD*) spielt eine Rolle beim Abbau der Zellwände im Endosperm und hat demzufolge auch einen großen Einfluss auf die Malzqualität bei Gerste (*Hordeum vulgare* L.). PLD wird durch Abscisinsäure aktiviert und katalysiert die Hydrolyse von Phosphatidylcholin zu Phosphorsäure.

Das Ziel dieser Untersuchungen ist das Auffinden von hochsignifikanten Assoziationen zwischen den für Hochdurchsatz-Genotypisierung entwickelten molekularen Markern und den Malz- und Braueigenschaften der Gerste.

### Material und Methoden

Von allen Kandidatengen wurden genomische Fragmente mittels PCR generiert. Die dafür verwendeten spezifischen Primer wurden basierend auf öffentlich verfügbaren Sequenzen aus NCBI-Unigene (<http://www.ncbi.nlm.nih.gov/UniGene>) generiert. Nach Resequenzierung und Identifikation von UTR-Intron-Exon-Strukturen, Detektion vorhandener SNPs und INDELs wurden spezifische Pyrosequenzier-Marker-Assays entwickelt und die Genotypisierung von 190 Sorten durchgeführt. INDEL-Analysen erfolgten mittels Kapillarelektrophorese. Haplotypen wurden für jedes Gen in Abhängigkeit vom Polymorphiegrad definiert. Alle Methoden bzgl. DNA-Isolation, PCR-Amplifikation, Pyrosequenzierung und SNP- und INDEL-Detektion, Ermittlung der Populationsstruktur, des Kopplungsungleichgewichtes (LD, linkage disequilibrium) sowie zur Erforschung des Verwandtschaftsgrades wurden bereits bei MATTHIES et al. (2009a,b) beschrieben. Insgesamt wurden 1915 DArT-Marker bei Triticarte Pty (<http://www.triticarte.com.au>, WENZL et al. 2004) analysiert, 1088 davon sind bereits kartiert. Ein repräsentatives Set von 253 ausgewählten DArT-Markern (alle 5 cM über das gesamte Gerstengenom) sowie von 22 SSRs diente zur Ermittlung der Populationsstruktur und

<sup>1</sup> Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstraße 3, D-06466 GATERSLEBEN

\* Ansprechpartner: Inge E. MATTHIES, [matthies@ipk-gatersleben.de](mailto:matthies@ipk-gatersleben.de)

**Tabelle 1: Anzahl der DArT- und SSR-Marker für die Ermittlung der Populationsstruktur, des Kopplungsungleichgewichtes (LDs) und des Verwandtschaftsgrades sowie für die genomweiten Assoziationsstudien bei 190 Gerstensorten**

**Table 1: Number of DArT and SSR markers used for linkage disequilibrium (LD)-studies, investigation of population structure, kinship and association studies in a set of 190 barley cultivars**

Chromosom	Anzahl DArT-Marker		Anzahl SSR-Marker
1H	141	35	4
2H	197	44	5
3H	190	44	3
4H	67	21	2
5H	155	41	2
6H	138	31	3
7H	200	37	3
Gesamt	(kartiert)	(Subset, alle 5cM)	
	1088	253	22

des Verwandtschaftsgrades (*Tabelle 1*). Zufällig verteilte Marker, wie SSRs und DArTs, wurden für die Bestimmung (1) des LD, (2) der Populationsstruktur, (3) des Verwandtschaftsgrades verwendet.

Phänotypische Daten wurden aus den statistischen Jahrbüchern der deutschen Braugerstengemeinschaft und Landesortenversuchen der letzten 20 Jahre zusammengetragen und in der Datenbank MetaBrew verwaltet (WEISE et al. 2009). Insgesamt 29 Korn- und Malzqualitätsparameter wurden für die Assoziationsstudien mit genomweiten DArT-Markern und spezifischen SNPs aus den Kandidatengen bei An-

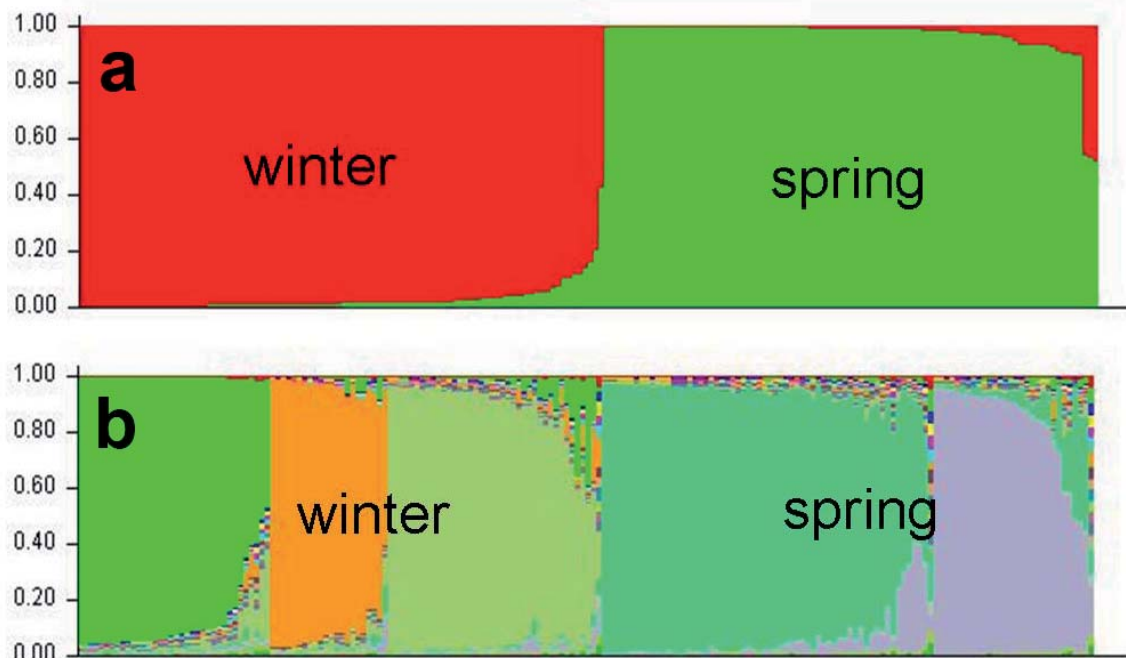
nahme des Generellen und Gemischten Linearen Modells (GLM und MLM) nach SEARLE (1987) berücksichtigt und mit TASSEL nach BRADBURY et al. (2007) verrechnet. Dazu wurde eine Pipeline zur Automatisierung der Berechnungen entwickelt (KUENNE et al. 2007).

Mit allen genomweiten Markern wurden insgesamt 190 europäische Gerstensorten untersucht, welche zu gleichen Anteilen aus Sommer- und Winterformen bestehen. Beim Kandidatengenansatz wurden Assoziationsstudien mit einer Auswahl von 140 Sorten (davon 60 Sommergersten) aus ursprünglich 342 genotypisierten Sorten durchgeführt.

## Ergebnisse und Diskussion

### *Populationstruktur und Kopplungsungleichgewicht (Linkage disequilibrium)*

Die verwendete Gerstenpopulation lässt sich klar in Winter- und Sommerformen unterteilen. Insgesamt konnten fünf Untergruppen für alle 190 Gerstensorten gefunden werden. Dabei entfallen drei Subgruppen auf die 94 zweizeiligen Sommergersten und zwei weitere Subgruppen auf die 96 Wintergersten (*Abbildung 1*). Letztere lassen sich aber nicht klar nach ihrer Zeiligkeit differenzieren. Ein hohes Kopplungsungleichgewicht ist eine wichtige Voraussetzung für das Auffinden von signifikanten Assoziationen (THORNSBERRY et al. 2001, FLINT-GARCIA et al. 2003). Dieses kann beispielsweise in einigen Regionen für das Chromosom 4H gezeigt werden (*Abbildung 2*).



**Abbildung 1: Populationsstruktur in 190 europäischen Gerstensorten ermittelt mit 22 SSR-Markern. (a) Sommer- und Winterformen (b) fünf Subpopulationen, davon entfallen drei auf Wintergerste.**

**Figure 1: Population structure of the total set of 190 European cultivars studied with 22 SSRs illustrated by bar plots. (a) Q2-SSR is differentiating winter from spring barley and (b) Q5-SSR shows the subclustering in these main groups.**

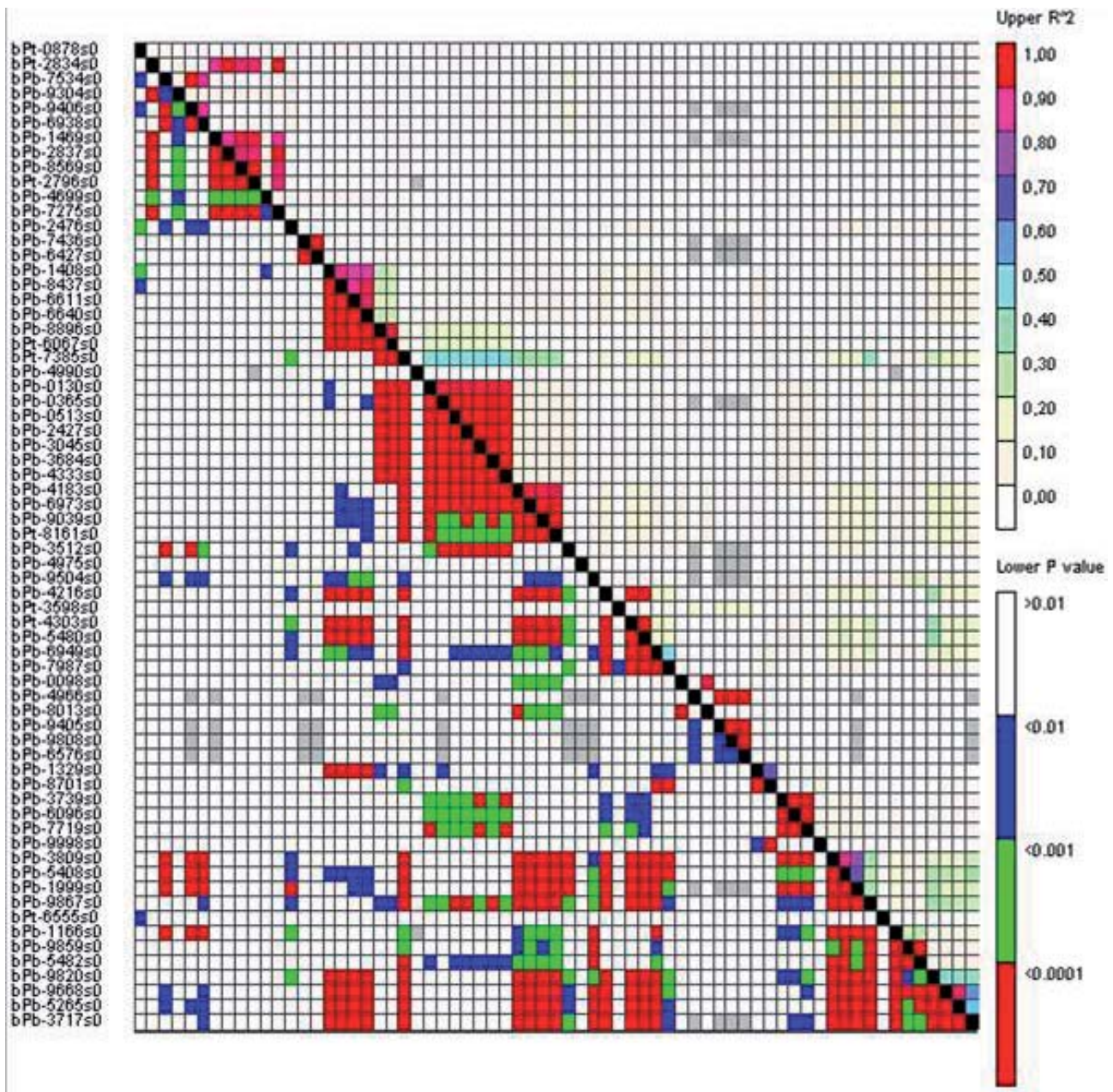


Abbildung 2: Kopplungsungleichgewicht ermittelt mit 67 DArT-Markern für alle Gerstensorten auf 4H.

Figure 2: Linkage disequilibrium on 4H determined with 67 DArT markers and all 190 barley cultivars

### Genomweite Assoziationen mit DArT-Markern

Mit allen 1915 DArT-Markern, von denen 1088 bereits kartiert sind, wurden bei Assoziationsstudien mit 29 Malz- und Kornqualitätsparametern zahlreiche signifikante Beziehungen aufgedeckt. Ausgewählte Ergebnisse für zwei Merkmale werden für das Chromosom 4H in *Abbildung 3* gezeigt. Dabei ergibt sich ein hochsignifikanter Bereich ( $P < 0,001$ ) für den Malzparameter Viskosität zwischen 61,6 und 65,3 cM sowie bei 67,9 cM sowohl für die gesamte Population unter Berücksichtigung der Populationsstruktur als auch für die Wintersubpopulation (*Abbildung 3a*). Für Tausendkornmasse (TKM) zeigen nur die Wintersorten eine signifikante Korrelation zu den DArT-Markern lokalisiert auf 4H bei 60,5 und 72,3 cM (*Abbildung 3b*).

### Assoziationsstudien in Kandidatengen

Für beide Kandidatengene *Cat1* und *PLD* wurden genomische Fragmente von 919 bp bzw. 598 bp amplifiziert und sequenziert. Jeweils zwei Hochdurchsatzmarker konnten mittels Pyrosequenzier-Assays für jedes Gen entwickelt werden. Bei der Genotypisierung von 342 Sorten konnten drei Haplotypen für *Cat1* bzw. zwei für *PLD* basierend auf den SNP-Mustern gefunden werden (*Abbildung 4*).

#### *Catalase 1 (Cat1)*

Bei *Cat1* wurden neun SNPs in vier Introns und einem Exon detektiert. Nur die synonymen SNP7 im Exon und SNP8 im benachbarten Intron konnten in Pyrosequenziermarker umgewandelt werden. Diese beiden SNPs zeigten in der

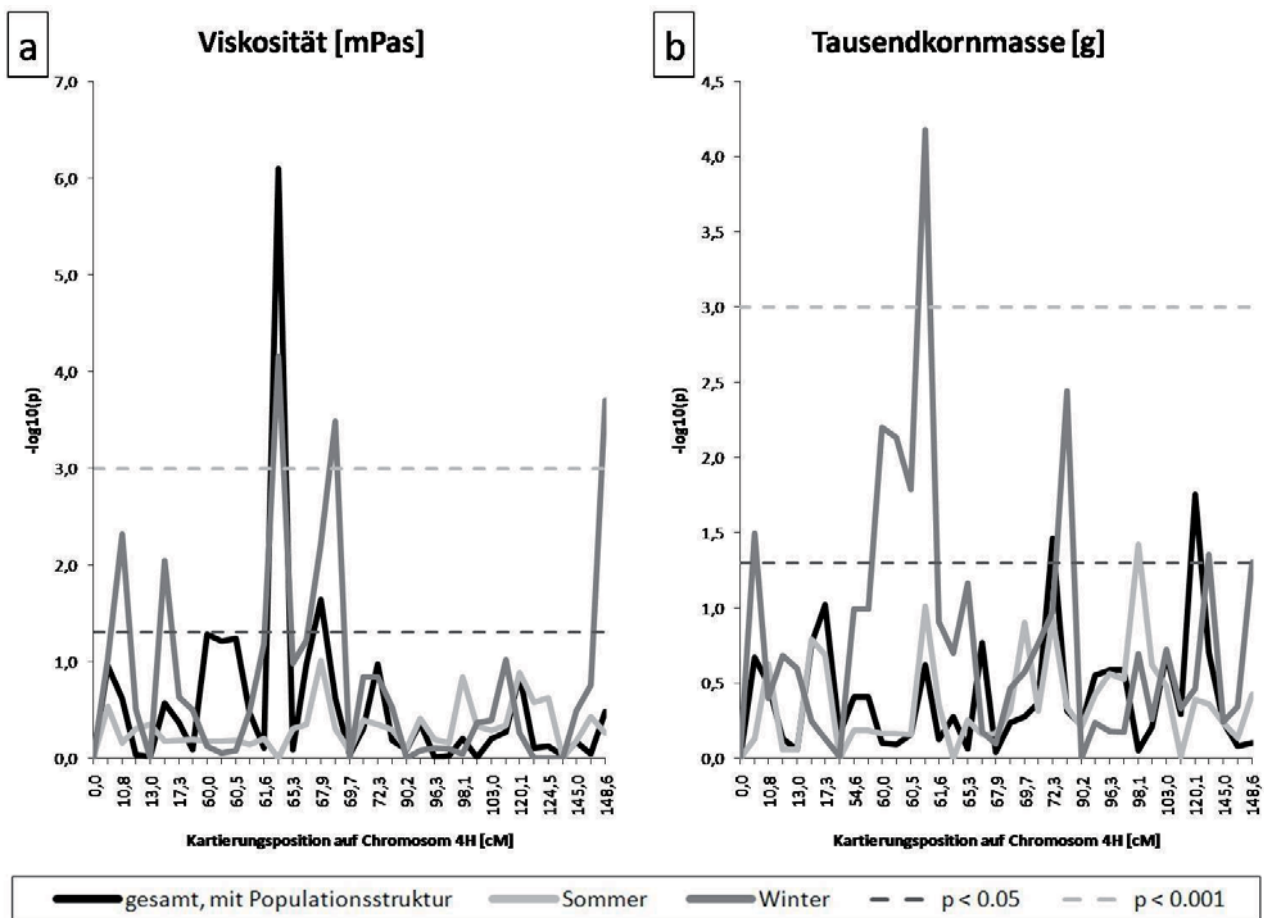


Abbildung 3: Assoziationsstudien mit DaRT-Markern auf Chromosom 4H der Gerste für die Merkmale Viskosität (a) und Tausendkornmasse (b). Insgesamt wurden 190 Gerstensorten unter Berücksichtigung der Populationsstruktur sowie die Sommer- und Winterformen getrennt verrechnet. Signifikante p-Werte sind als  $-\log_{10}$  dargestellt.

Figure 3: Association studies with DaRT markers on barley chromosome 4H. Results for the traits viscosity (a) and thousand grain weight (b) are shown for the total set of 190 cultivars with correction for population stratification and the subsets of spring and winter barley which are plotted together. Significances of the calculated p-values are shown as  $-\log_{10}$  (p-value)

Subpopulation von 60 Sommergersten signifikante Assoziationen mit den Parametern Viskosität, Endvergärungsgrad und Hektolitergewicht, was sich auch in den Ergebnissen für die Haplotypen Cat1\_H1 und Cat1\_H2 widerspiegelt (Tabelle 2). Die Varianz für Malzextrakt und Viskosität wird von Cat\_H1 zu 32% bzw. zu 9% definiert. Während Cat1\_H2 24% der phänotypischen Varianz für den Endver-

gärungsgrad erklärt, wird dieser nur zu 14% von Cat1\_H1 in den Sommersorten bestimmt.

*Phospholipase-D-like protein (PLD)*

Vom kodierenden Gen für das Phospholipase-D-like Protein (PLD) konnten jeweils ca. 300 bp vom letzten Exon und der sich anschließenden 3'UTR bei den Referenzgenotypen

Tabelle 2: Signifikante Assoziationen für SNPs und Haplotypen des *Cat1*-Kandidatengenes mit einigen Korn- und Malzqualitätsparametern bei Annahme des Generellen Linearen Modells (GLM) für alle 140 Sorten (T) bei Berücksichtigung der Populationsstruktur bzw. für jede Teilpopulation bestehend aus 60 Sommer- (S) und 80 Wintergersten (W)

Table 2: Association results for SNPs and haplotypes for the *Cat1* gene identified by the General Linear Model (GLM) for certain malting and grain quality parameters calculated either with all cultivars (T) with regard of population structure or each subset of 60 spring (S) and 80 winter barleys (W)

Parameter	SNP7	SNP8	Cat1_H1	Cat1_H2	Cat1_H3
Viskosität (8,6% mPas)	S*	S**	S*	S***	n.s.
Vergärbarer Extrakt (%)	T* S***	n.s.	S***	T*	S***
Endvergärungsgrad (%)	S**	S***	S***	S***	T*
Malzqualitätsindex (MQI)	T* S***	n.s.	T* S***	T* S*	S**
Hektolitergewicht (kg hl <sup>-1</sup> )	T* S***	S*	S***	T* S*	S**

n.s., nicht signifikant; \*, p<0,05; \*\*, p<0,01; \*\*\*, p<0,001

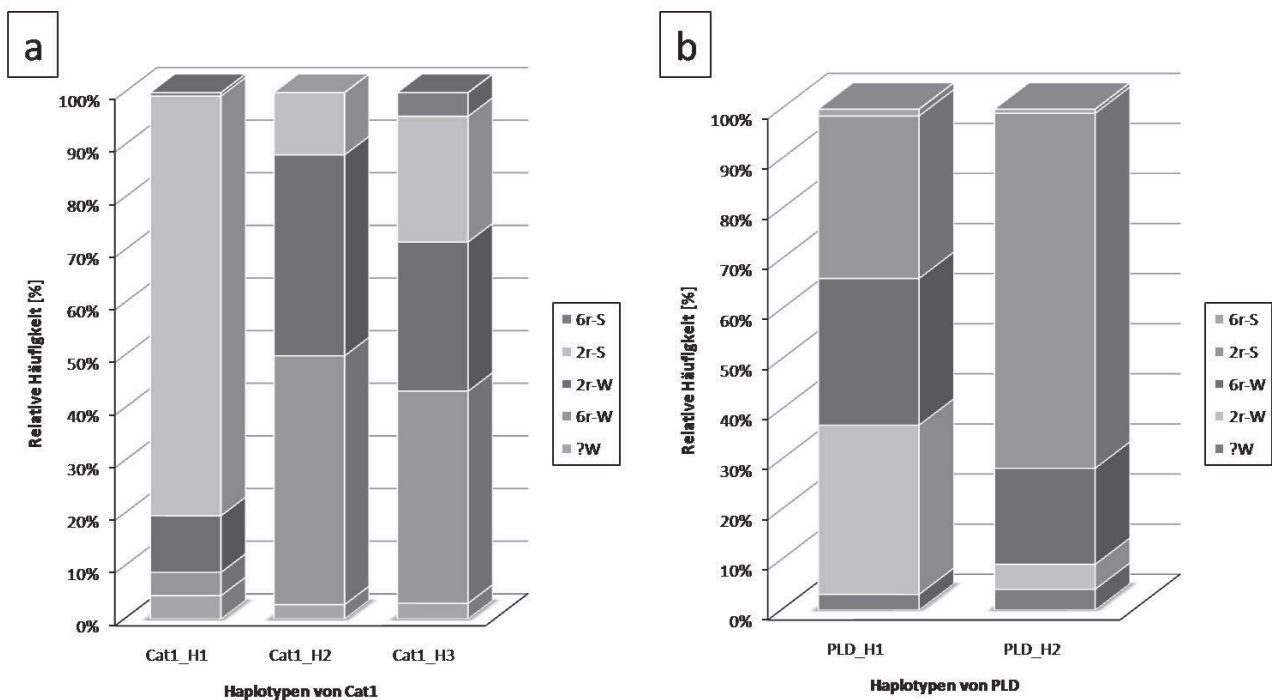


Abbildung 4: Relative Haplotypenverteilung der Kandidatengene *Cat1* (a) und *PLD* (b) in 342 Gerstensorten (S, Sommergersten; W, Wintergersten; 2r, 2-zeilig; 6r, 6-zeilig; ?, Wintergersten mit unbekannter Zeiligkeit)

Figure 4: Haplotype distribution of the genes *Cat1* (a) and *PLD* (b) in a set of 342 barley cultivars related to the population substructure according to growth habit and row number (S, spring; W, winter; 2r, two-row; 6r, six-row; ?, winter barleys with unknown row-number)

sequenziert werden. In letzterer wurden vier SNPs detektiert und zwei davon konnten in Pyrosequenziermarker umgewandelt werden. Assoziationsstudien unter Annahme des GLM zeigten signifikante Beziehungen entweder mit allen 140 Sorten bzw. in den Subpopulationen von 60 Sommer- oder 80 Wintersorten. Der Haplotyp PLD\_H2 hat einen signifikanten Einfluss auf Malzextrakt, Farbe des Malzes sowie dem MQI und erklärt entsprechend 27%, 23% bzw. 26% der phänotypischen Varianz dieser Merkmale innerhalb der Sommersorten. Hoch signifikante Assoziationen ( $p < 0,001$ ) konnten jeweils für SNP1 und SNP3 mit den Merkmalen Endvergärungsgrad, Malzextrakt, Friabilimeter und MQI detektiert werden. Beide Marker erklären bis zu 30% der phänotypischen Varianz für diese Merkmale (Tabelle 3).

Beide Kandidatengene, *Cat1* (Steptoe  $\times$  Morex) und *PLD* (OWB), konnten am distalen Ende des langen Arms vom Gersten-Chromosom 7H kartiert werden.

## Schlussfolgerungen und Ausblick

Mit Hilfe der entwickelten funktionalen Marker aus dem genomweiten und Kandidatengenansatz lässt sich die Züchtung von Braugersten mit verbesserten Korn- und Malzqualitätseigenschaften durch markergestützte Selektion erheblich beschleunigen. Durch Korrelation der erhaltenen molekulargenetischen Daten mit den sortenspezifischen Malzeigenschaften können kausale Zusammenhänge zwischen Genotyp und Phänotyp gefunden werden. Dabei kann

Tabelle 3: Signifikante Assoziationen für SNPs und Haplotypen für das Phospholipase D-like Protein (PLD) kodierende Gen mit einigen Korn- und Malzqualitätsparametern bei Annahme des Generellen Linearen Modells (GLM) für alle 140 Sorten (T) bei Berücksichtigung der Populationsstruktur bzw. für jede Teilpopulation bestehend aus 60 Sommer- (S) und 80 Wintergersten (W)

Table 3: Association results and significant SNPs and haplotypes of the Phospholipase D-like Protein (PLD) gene for some chosen malting quality parameters calculated either with all cultivars (T) with regard of population structure or each subset of 60 spring (S) and 80 winter barleys (W)

Parameter	SNP1	SNP3	PLD_H1	PLD_H2
Farbe (EBC)	T** S***	T** S***	T* S*	T** S***
Extrakt (%)	S*** W*	S*** W*	W*	S***
Endvergärungsgrad (%)	T** S***	T** S***	T* S**	n.s.
Friabilimeter (%)	S*** W**	S*** W**	W**	S*
Verzuckerungszahl VZ45 (%)	S**	S**	n.s.	n.s.
Malzqualitätsindex (MQI)	S*** W**	S*** W**	S* W**	S*** W**

n.s., nicht signifikant; \*,  $p < 0,05$ ; \*\*,  $p < 0,01$ ; \*\*\*,  $p < 0,001$

gezeigt werden, wie hoch der genetische Beitrag zu einem Malzmerkmal sein kann, was nicht nur für Züchter, sondern auch für die Malz- und Brauindustrie von Interesse ist. Durch Entwicklung neuer Gerstensorten mit verbesserten physiologischen Leistungen wird eine weitere Optimierung des Zucht- und Mälzungsprozesses ermöglicht.

## Danksagung

Wir danken Ellen Weiß und Manja Franke für die technische Durchführung der SNP- und SSR-Studien, der SU Biotec und der KWS-Lochow AG für die Mitarbeit bei der Genotypisierung einiger Kandidatengene, sowie Triticarte für die DArT-Analysen. Dem BMBF wird für die Finanzierung der beiden GABI-Projekte MALT (0313125A) und GENOBAR (0315066C) herzlich gedankt.

## Literatur

- BRADBURY PJ, ZHANG Z, KROON DE, CASSTEVENS RM, RAMDOSS Y, BUCKLER ES, 2007: TASSEL Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633-2635.
- FLINT-GARCIA SA, THORNSBERRY JM, BUCKLER ES, 2003: Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* 54:357-374.
- KUENNE C, GROSSE I, MATTHIES IE, SCHOLZ U, SRETENOVIC-RAJICIC T, STEIN N, STEPHANIK A, STEUERNAGEL B, WEISE S, 2007: Using data warehouse technologies in crop plant bioinformatics. *J Integr Bioinformatics* 4: 88.
- MATTHIES IE, WEISE S, RÖDER MS, 2009a: Association of haplotype diversity in the  $\alpha$ -amylase gene *amyl1* with malting parameters in barley. *Mol Breed* 23: 139-152.
- MATTHIES IE, WEISE S, FÖRSTER J, RÖDER MS, 2009b: Association mapping and marker development of the candidate genes (1 $\rightarrow$ 3), (1 $\rightarrow$ 4)- $\beta$ -D-glucan-4-glucanohydrolase and (1 $\rightarrow$ 4)- $\beta$ -xylan-endohydrolase 1 for malting quality in barley. *Euphytica* 170: 109-122.
- SEARLE, 1987: *Linear models for unbalanced data*. Wiley, New York.
- THORNSBERRY JM, GOODMAN MM, DOEBLY J, KRESOVICH S, NIELSEN D, BUCKLER ES, 2001: Dwarf8 polymorphisms associate with variation in flowering time. *Nat Genet* 28:4673-4680.
- WEISE S, SCHOLZ U, RÖDER MS, MATTHIES IE, 2009: MetaBrew: A comprehensive database of malting quality traits in brewing barley. *Barley Genet Newsl* 39:1-4.
- WENZL P, CARLING J, KUDRNA D, JACCOUD D, HUTTNER E, KLEINHOF A, KILIAN A, 2004: Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proc Natl Acad Sci* 101: 9915-9920.