



Veröffentlichung der Ergebnisse von Forschungsvorhaben im BMBF-Programm

Pflanzenbiotechnologie-Verbundvorhaben

Auswahl und Identifikation molekularer Marker in spezifischen Regionen mit züchterisch wichtigen Eigenschaften und für die allgemeine genetische Kartierung in hexaploidem Weizen für die beschleunigte Pflanzenzüchtung

(SELECT) – Teilprojekt B

Förderkennzeichen:

0315949B

Zuwendungsempfänger:

Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK)

Corrensstr. 3

06466 Stadt Seeland OT Gatersleben

Projektkoordination:

Dr. Marion Röder

Laufzeit:

01.11.2011 bis 31.12.2015

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 0315949B gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei dem Autor

Inhaltsverzeichnis

I. Kurze Darstellung

1.	Aufgabenstellung	3
2.	Voraussetzungen, unter denen das Vorhaben durchgeführt wurde	3
3.	Planung und Ablauf des Vorhabens	4
4.	Wissenschaftlicher und technischer Stand, an den angeknüpft wurde	5
5.	Zusammenarbeit mit anderen Stellen	5

II. Eingehende Darstellung

1.	Der erzielten Ergebnisse	6
2.	Wichtige Positionen des zahlenmäßigen Nachweises	27
3.	Notwendigkeit und Angemessenheit der geleisteten Arbeit	27
4.	Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse	28
5.	Bekannt gewordene Fortschritte auf dem Gebiet des Vorhabens bei anderen Stellen	29
6.	Erfolgte und geplante Veröffentlichung der Ergebnisse	31
7.	Zitierte Literatur	32

I. Kurze Darstellung

1. Aufgabenstellung

Molekulare Marker sind heutzutage ein unverzichtbares Werkzeug in der Pflanzenzüchtung für alle wichtigen Kulturarten. Sie werden hauptsächlich für die Analyse der genetischen Diversität im gesamten Genom, für die Marker-gestützte Rückkreuzung, sowie für die Markierung individueller Gene oder QTL („quantitative trait loci“) eingesetzt. Um ein hohes Maß an Präzision in der Marker-gestützten Pflanzenzüchtung zu erreichen, werden immer größere Zahlen an molekularen Markern eingesetzt, welche entweder gleichmäßig im Genom verteilt sind oder spezifische chromosomale Regionen mit hoher Dichte absättigen. Die Markerentwicklung stellt vor allem in allopolyploiden Kulturpflanzen, wie beispielsweise in hexaploidem Kulturweizen mit seinen drei Genomen A, B und D, eine Herausforderung dar. Hier hat man einerseits SNPs („single nucleotide polymorphisms“) zwischen den verschiedenen Genomen, welche für die Genomkartierung nicht verwendbar sind, und andererseits SNPs innerhalb eines spezifischen Genomes, welche zwischen verschiedenen Weizensorten unterschiedlich sind und daher in der Genomkartierung einsetzbar sind.

Allopolyploide Species haben oft sehr große Genome. So beträgt die Genomgröße im hexaploiden Weizen >16Gbp und das Genom besteht zu über 90% aus hochrepetitiven Sequenzen. Aus diesem Grund war eine Referenzsequenz zu Projektbeginn noch nicht verfügbar. Daher war eine Strategie gesucht um das Gesamtgenom zu fraktionieren und die Einzelkopiefraktion, welche hauptsächlich die Gene enthält, zur Markerentwicklung zu nutzen.

Das Ziel des Projektes war es, Systeme zu entwickeln, um einerseits die Variabilität von Genen in spezifischen chromosomalen Regionen zu bestimmen und andererseits auch eine Abdeckung aller Gene des Weizengenoms zu erreichen. Dafür sollten sogenannte „Sequence Capture“ Techniken in Verbindung mit NGS („Next Generation Sequencing“) eingesetzt werden. Als Ziele wurden einige spezifische Regionen für Samengröße und Tausendkorngewicht im Weizengenom definiert. Daneben sollten einzelne Chromosomen mit weiteren Markern gesättigt werden. Die so entwickelten Marker sind für die Forschung im akademischen Bereich von Bedeutung, können aber auch kommerziell in der Pflanzenzüchtung eingesetzt werden.

2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Das IPK und TraitGenetics verfügen aus vorherigen Arbeiten über erste Informationen hinsichtlich spezifischer QTLs für Samengröße und Tausendkorngewicht. Diese Daten sollten im Rahmen des Projektes weiter genutzt werden

Als im Bereich der molekularen Markeranalyse tätige Firma verfügt TraitGenetics über ein umfangreiches Wissen im Bereich der molekularen Markeranalyse bei diploiden und polyploiden Nutzpflanzen. Dies betrifft insbesondere die Generierung und Auswertung von NGS-Daten, sowie die Hochdurchsatzgenotypisierung auf Arraybasis (hier hauptsächlich mit der Illumina-Technologie).

Das IPK verfügt über umfangreiches Wissen bezüglich der Datenauswertung von Assoziations- und QTL-Studien, welche durch zahlreiche Publikationen nachgewiesen sind. Ferner verfügt das IPK über Erfahrungen und die notwendigen technischen Ressourcen (Gewächshäuser, Vernalisationskammern, etc.) zur Generierung von genetischem Material für derartige Studien.

3. Planung und Ablauf des Vorhabens

Das Projekt bestand aus fünf Arbeitspaketen:

Im ersten Arbeitspaket sollte die Technik des Sequence Capture getestet und etabliert werden. Zu diesem Zweck wurden unterschiedliche Zielregionen im Weizengenom für ein Custom Sequence Capture System definiert. Dies umfassten eine Region mit einem publizierten QTL für Samengröße auf Chromosom 7D (Röder et al. 2008), mehrere chromosomale Regionen mit Marker-Merkmal-Assoziationen (MTAs) für Tausendkorngewicht (TKG) (Zanke et al. 2015), sowie die homeologen Gruppe 6 Chromosomen. Mit dem entwickelten Sequence Capture System sollten alle technischen und bioinformatischen Prozeduren etabliert werden.

Im zweiten Arbeitspaket sollte ein Whole Genome Exome Capture System entwickelt und genutzt werden. Mit diesem Exome Capture System sollte eine Reihe von Weizenlinien analysiert werden

Im dritten Arbeitspaket sollten zur Verifikation der genomischen Regionen für Samengröße und Tausendkorngewicht Kreuzungskombinationen von Weizensorten sowohl als F2-Kartierungspopulationen, wie auch spezifische Rückkreuzungslinien für einzelne QTL entwickelt werden. Ausgewählte SNPs aus dem ersten Arbeitspaket für gezielte genomische Regionen mit QTL für Samengröße sollten zur Verifikation des Ansatzes auch in KASP-Marker umgewandelt und auf unterschiedlichem Pflanzenmaterial einschließlich einer öffentlich verfügbaren Kartierungspopulation (ITMI-DH International Triticeae Mapping Initiative – doubled haploids, Sorrells et al. 2011) kartiert werden.

Im vierten Arbeitspaket sollten die aus dem Whole Genome Exome Capture erhaltenen SNPs zur Kartierung von Populationen aus Arbeitspaket 3 und einem repräsentativen Set von europäischen Weizensorten eingesetzt werden. Dazu sollte ein SNP-Genotypisierungsarray entwickelt und eingesetzt werden um speziell für ausgewählte genomischen Regionen mit QTL für Samengröße und

Tausendkorngewicht, aber auch für alle 21 Weizen-Chromosomen, eine sehr hohe Markerdichte zu erhalten.

Das fünfte Arbeitspaket umfasste die umfassende begleitende bioinformatische, genetische und populationsgenetische Analyse. In diesem Projektteil sollten alle Daten gesammelt werden und weiter ausgewertet werden.

4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Vor Beginn des Projektes (zur Zeit der Antragstellung in 2010) war noch keine Referenzsequenz für das Weizengenom verfügbar und auch die SNP-Entwicklung steckte noch in den Kinderschuhen. SNP-Marker in ca. 2500 Genen waren in dem Projekt WheatHapMap (<http://wheat.pw.usda.gov/SNP>) erhältlich. Jedoch nur ein Teil dieser Marker war für die Genotypisierung aktueller Sorten praktisch einsetzbar. Erste Publikationen mit kleinen Datensets zeigten, dass der ILLUMINA Golden Gate Assay auch für allopoloide Kulturarten wie Weizen oder Raps einsetzbar ist (Akhunov et al. 2009, Durstewitz et al. 2010).

Sequence Capture-Technologien wurden vor Projektbeginn insbesondere in der Humangenetik eingesetzt (Teer und Mullikin 2010), und ersten Studien für diploide Kulturarten wie Mais (Fu et al. 2010) waren veröffentlicht.

Umfangreiche Studien zur Identifikation von QTL und Marker/Trait-Assoziationen für Samengröße und Tausendkorngewicht waren ebenfalls noch nicht vorhanden und auch keine Absättigung einzelner Regionen mit Markern für solche QTL.

5. Zusammenarbeit mit anderen Stellen

Ursprünglich war es geplant, das Projekt mit drei Partnern durchzuführen. Durch das Ausscheiden eines Partners im ersten Jahr des Projektes verblieben das IPK und TraitGenetics als Kooperationspartner und die Arbeitsprojekte des Projektes mussten entsprechend angepasst werden (siehe wissenschaftlicher Bericht). Während des gesamten Projektes erfolgten regelmäßig Projekttreffen und gemeinsame Besprechungen zum Fortgang des Projektes zwischen dem IPK und TraitGenetics.

II. Eingehende Darstellung

1. Der erzielten Ergebnisse

Arbeitspaket 1: Entwicklung eines ersten Capture Systems

Identifikation der Zielregionen und 'Bait selection' (IPK, TraitGenetics)

Für den Vortest des Sequence Capture wurden verschiedene Zielregionen im Weizengenom definiert, für welche die Methode des „Sequence Capture“ (Abbildung 1) getestet werden sollte.

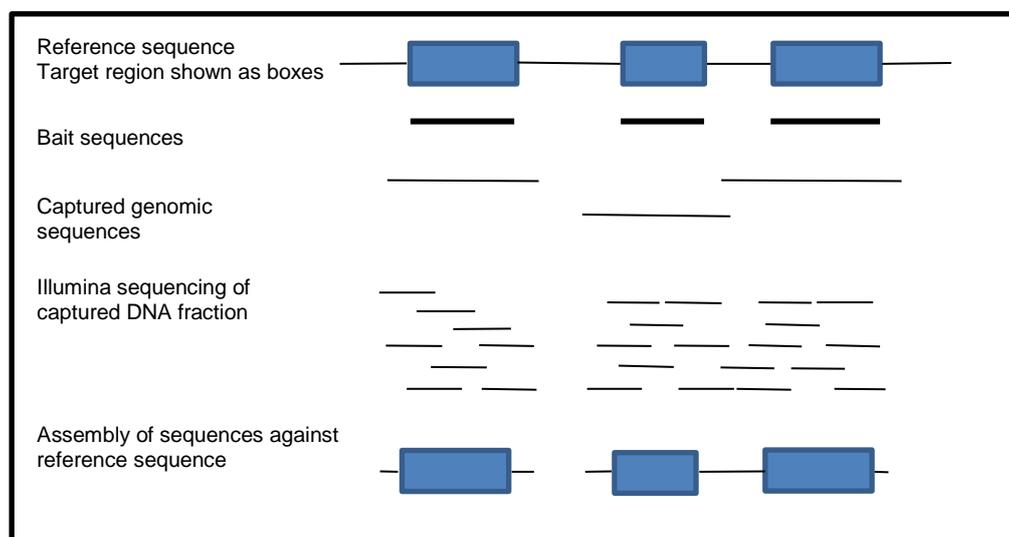


Abbildung 1: Prinzip und Ablauf des Sequence Capture für spezifische Bereiche eines Genoms

- **Genomische Region für Tausendkorngewicht QTL *QTgw.ipk-7D* (Bereich 1)**

Die QTL-Region *QTgw.ipk-7D* wurde auf Chromosomenarm 7DS identifiziert, feinkartiert (Röder et al. 2008) und ist inzwischen von einem sequenzierten Weizen BAC-Contig abgedeckt. Für diesen QTL liegen aus den Kreuzungen Prinz × M6 und Flair × XX86 Kartierungspopulationen und spezifische Rekombinante vor. Für diesen Bereich wurde die gesamte Sequenz des BAC-Contigs assembliert und nachfolgend durch das Screening der Graingenes Wheat Repeats Datenbank (Triticeae Repeat Data Base, <http://wheat.pw.usda.gov/ITMI/Repeats/>) und der TIGR Triticum Repeats Datenbank in Segmente aus repetitiven und potenziellen Single Copy Sequenzen sortiert (Abbildung 2).

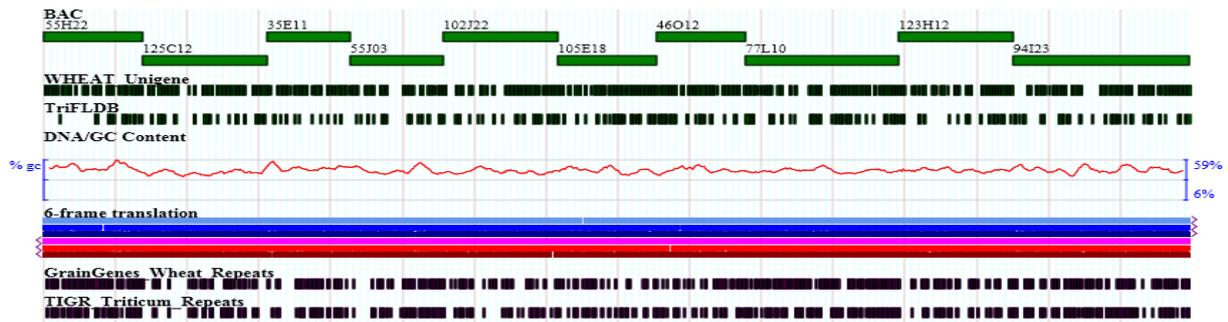


Abbildung 2: Analyse des BAC-Contigs auf Chromosomenarm 7DS hinsichtlich des Bestandes an repetitiven Sequenzen. Der gesamte Bereich des Contigs umfasst etwa 1,9 Mbp. Die markierten Bereiche in der Analyse (GrainGenes Wheat Repeats und TIGR Triticum Repeats) zeigen Regionen mit repetitiven Sequenzen, welche für die Bait-Entwicklung ausgeschlossen wurden.

- **Mit Tausendkorngewicht (TGW) assoziierte genomische Regionen (Bereich II und III)**

In dem Projekt GABI-WHEAT wurden Daten zu Marker-Merkmalbeziehungen (MTAs) zwischen phänotypischen Daten und Mikrosatellitenmarkern in einem Assoziationskartierungsansatz etabliert. Dort und in weiteren in dem SELECT Projekt durchgeführten Untersuchungen wurden zahlreiche Assoziationen für TKG identifiziert (siehe auch unter Arbeitspaket 3). Wir haben eine Reihe von genomischen Regionen mit stabilen und hoch signifikanten MTAs für TKG selektiert, welche einen steigernden Effekt für das TKG haben (Tabelle 1).

Tabelle 1: Ausgewählte TKG-assozierte Markerbereiche

Nr.	Marker	Gruppe	Position cM	Nr.	Marker	Gruppe	Position cM
1	WMS1521	CHR1B	59	12	WMS1454	CHR 5D	0
2	WMS0539	CHR 2D	90,934	13	WMS0427	CHR 6A	0
3	WMS0320	CHR 2D	145,336	14	WMS0219	CHR 6B	45,641
4	WMS0369	CHR 3A	184,507		WMS0219	CHR 6B	53,555
5	BARC0012	CHR 3A	219,387	15	CFD0075	CHR 6D	128,154
6	WMS0376c	CHR 3B	98,788	16	GMD0132	CHR 6D	128,958
7	WMC0722	CHR 4A	0	17	WMS0469	CHR 6D	177,35
8	WMC0283	CHR 4A	72,143	18	WMS0808	CHR 7B	59,810
9	WMS0495	CHR 4B	42,728				
10	BARC0330	CHR 5A	27,165				
11	WMC0415	CHR 5B	132,158				

Die assoziierten Mikrosatellitenmarker wurden auf der im Arbeitspaket 3 beschriebenen ITMI-DH Population genotypisiert. Basierend auf der Position der Mikrosatellitenmarker in einem SNP-Framework konnten über Synteniedaten mit Reis und *Brachypodium* Gene identifiziert werden, welche in den entsprechenden Regionen lagen (+/- 10 cM links und rechts der entsprechenden Position). Zum Bait-Design wurden zwei Ansätze gewählt. Einmal wurden nur die kodierenden Bereiche

der Gene (Exons) verwendet und das zweite Mal für dieselben Gene auch Introns und nicht-translatierte flankierende Regionen (ca. 500 bp). Ziel sollte es sein zu testen, ob es möglich ist, mehr SNPs zu identifizieren, wenn Sequenzen außerhalb der Exons als Baits verwendet werden, da die Exons unter kontinuierlicher Selektion stehen und damit weniger Variation zeigen, als nicht-kodierende Sequenzen.

• **Zusätzliche Marker auf der Gruppe 6 Chromosomen (Bereich IV)**

Chromosom 6A trägt wichtige Gene oder QTL und ist das Chromosom, welches von Deutschland offiziell innerhalb des internationalen Weizengenom Sequenzierungskonsortium (IWGSC; <http://www.wheatgenome.org/>) bearbeitet wird (BMBF-Projekt TRITEX, Leiter Dr. Thorsten Schnurbusch). Daher war eine Markerabsättigung für die homeologe Chromosomengruppe 6 von Interesse und Gene dieser Chromosomengruppe wurden zur Auffüllung des nutzbaren Baitvolumens verwendet.

Etablierung der ‚Sequence Capture‘ Methode für die ausgewählten Regionen (TraitGenetics, IPK)

Insgesamt wurden 24 Weizenlinien (Auswahl der Linien und DNA-Präparation erfolgten durch das IPK, Tabelle 2) mit der ‚Sequence Capture‘ Technologie der Firma Agilent (SureSelect^{XT} Target Enrichment System for Illumina Paired-End Sequencing Library) für die ausgewählten Regionen durchgeführt. Die dafür notwendigen experimentellen Vorgehensweisen konnten erfolgreich etabliert werden. Die gewonnene DNA wurde auf einem Illumina MiSeq Sequenziergerät sequenziert.

Tabelle 2: Liste der Weizenlinien für ‚Sequence Capture‘ und ihre Zuordnung zu den spezifischen Regionen

W7984 (M6 =Synthetic	Region I (Chr. 7D)
Prinz	
Prinz34a/3-1 No. 51	
Prinz34a/3-1 No. 73	
XX86	
Flair	
Flair-130-1-1B No. 82C	
Flair-130-1-1B No. 30B	
Bobwhite	
Mulan	
Ephoros	
Hereford	
Expert	
Sogood	
Intérét	
Mikon	
Inspiration	

Opata	Region IV (Chr. 6A)
W7984 (M6	
Bussard	
Recital	
Apache	
Renan	
Sperber	
Victo	
CHS	

Auswertung der Sequenzdaten aus dem Sequence Capture (TraitGenetics)

Bei der bioinformatischen Auswertung erfolgte zuerst eine Analyse hinsichtlich der Anreicherung für die selektierten Bereiche. Es wurde festgestellt, dass bis zu 90% der Reads der Referenzsequenz zugeordnet werden konnten. Durchschnittlich waren es etwa 80% der Reads. Nach der bioinformatischen Analyse haben sich folgende Ergebnisse für die vier Bereiche (Tabelle 3) ergeben.

Tabelle 3: Ausbeute an Roh-SNPs für die einzelnen Regionen. Dabei handelt es sich um die gesamte Anzahl der identifizierten SNPs (umfasst intergenomische und intragenomische SNPs).

Bereich	# SNPs
I	14.466
II	19.479
III	22.756
IV	49.464
	106.165

Die SNP-Zahlen entsprachen den Erwartungen. Beispielsweise konnten etwas mehr SNPs in der Region III (Gene mit Exons und Introns plus flankierende Sequenzen) als in Region II (nur Exons) identifiziert werden.

Im Nachfolgenden wurden bioinformatische Auswertungswerkzeuge entwickelt, um die SNPs zwischen den drei Weizengenomen von solchen, welche zwischen den einzelnen Linien auftreten, zu unterscheiden. Bedingt durch die hohe Coverage der einzelnen Bereiche (>100 x) konnten diese zwei Klassen von SNPs klar unterschieden werden. Nach der Eliminierung der intergenomischen SNPs (SNPs zwischen den drei Genomen – diese stellen den Großteil der SNPs dar) sind insgesamt 21870 potenziell nutzbare Marker verblieben (Tabelle 4).

Tabelle 4: Anzahl der pro Region erhaltenen nutzbaren SNPs

Bereich	Anzahl SNPs
I	3.409
II	3.980
III	4.694
IV	9.787
	21.870

Zu direkter Validierung dieser SNPs wurde ein repräsentativer Teil anschließend als KASP-Marker auf einem Weizenpanel getestet. Auf die Ergebnisse der Untersuchungen wird in Arbeitspaket 3 eingegangen.

Insgesamt konnten aus dem Arbeitspaket 1 folgende Schlüsse gezogen werden:

- Die gesamte Prozedur des Sequence Capture ausgehend vom Bait-Design über die eigentlichen experimentellen Arbeiten bis zur NGS-Analyse konnte erfolgreich etabliert werden.
- Die notwendigen bioinformatischen Analysewerkzeuge konnten ebenfalls entwickelt werden und bei ausreichender Sequenzabdeckung (>100 x) können SNPs zwischen den Genomen (Intergenomisch) des Weizens von SNPs zwischen einzelnen Linien (intra-genomisch) klar unterschieden werden.
- Es konnte eine relativ hohe Anzahl von nutzbaren SNPs für die einzelnen genomischen Bereiche identifiziert werden.

Arbeitspaket 2: Entwicklung und Nutzung eines Capture Systems für das Weizengenom

Etablierung eines Sequence Capture Systems für die meisten Weizengene (TraitGenetics)

Im Rahmen dieses Projektteiles wurde erwogen, ein eigenes Sequence Capture System für Weizen zu etablieren. Ursprünglich war dabei daran gedacht möglichst alle Weizengene zusammen mit den Introns und 5' und 3'-flankierenden Sequenzen als Baits zu benutzen. Die Experimente aus dem Arbeitspaket 1 (siehe auch Tabelle 4) haben aber gezeigt, dass dies keine große Verbesserung gegenüber einem reinen Exome Capture System darstellt (nur 15% mehr SNPs). Dieser Gewinn von 15% Markern im Vergleich zwischen Exome und ganzen Genen als Baits erscheint als gering wenn man berücksichtigt, dass Introns und flankierende Sequenzen normalerweise unter geringerem Selektionsdruck stehen. Die Hauptursache für den geringen Gewinn liegen nach gründlicher Analyse jedoch darin, dass mit einem Exome Capture auch flankierende Sequenzen (bis zu jeweils mehreren hundert bp links und rechts eines Exons) mit sequenziert werden und somit nicht nur die meisten Exons, sondern auch ein Teil der flankierenden Regionen der Gene abgedeckt werden.

Gleichzeitig wurde zu diesem Zeitpunkt ein öffentliches Exome Capture System der Firma Nimblegen/Roche für Weizen (WEC) verfügbar, welches von einer Forschergruppe (Jordan et al., 2015) etabliert wurde. Mit diesem System werden 107 Mb an genischen single copy Sequenzen (278181 einzelne Targets) abgedeckt. Dies entspricht 321 Mb an Sequenz über alle drei Weizengenome, da diese simultan detektiert werden. Aus diesem Grund und weil durch den Ausfall des dritten Projektpartners an einem Bait-Design nur mit reduziertem Aufwand gearbeitet

werden konnte, wurde dieses bereits verfügbare System für die weiteren genomweiten Analysen verwendet.

Nutzung des Nimblegen Sequence Capture Systems für 25 Weizenlinien (TraitGenetics)

Insgesamt wurde mit dem WEC-System das Exome Capture für 25 von dem IPK ausgewählte und zur Verfügung gestellte Linien durchgeführt. Die Etablierung des technischen Ablaufes der Prozedur erfolgte zuerst mit einem Exome Capture System von Gerste, da hier die Situation einfacher ist, weil Gerste diploid ist und so die Daten schneller analysiert werden konnten. Ferner konnte die Sequenzierung für Gerste auf dem MiSeq durchgeführt werden und war damit schneller zu realisieren und auf Erfolg zu überprüfen. Nachdem die Analyse bei Gerste erfolgreich war, wurden die 25 Weizenlinien bearbeitet. Basierend auf den Ergebnissen aus Projektteil 1, wurde die angereicherte Fraktion dieser Linien mit einer Sequenzabdeckung von durchschnittlich 100x auf einem Illumina HiSeq-Sequenzierungsgerät (2 x 100 bp paired end) analysiert.

Bioinformatische Auswertung der erhaltenen Daten (TraitGenetics)

Die nachfolgende bioinformatische Analyse der Daten für die 25 Linien erfolgte mit den in Projektteil 1 entwickelten bioinformatischen Methoden. Der erste Schritt war dabei die Bestimmung der Menge und des Prozentsatzes der erhaltenen Reads auf die Referenzsequenzen kartieren.

Zusammenfassend wurden pro Linie zwischen 253 und 455 Millionen Reads erhalten. Davon konnten pro Linie zwischen 55% und 79% der Reads auf der Referenzsequenz kartiert werden. Mit zwischen 19 und 33 Gbp an kartierten Reads konnte eine Abdeckung von 60-103x der Referenzsequenz (gesamt 321 Mb über alle drei Genome) erreicht werden. In Tabelle 5 sind diese Daten für alle untersuchten Linien dargestellt.

Die Anzahl der identifizierten SNPs und Insertionen/Deletionen im Vergleich zur Referenzsequenz ist für jedes/n Chromosom(enarm) in der folgenden Tabelle 6 dargestellt. Dabei handelt es sich um Rohpolymorphismen, auf die noch keinerlei Selektionskriterien angewandt wurden.

Tabelle 5: Darstellung der 25 Linien für welche das Sequence Capture durchgeführt wurde (mapped = Anzahl der Reads, welche auf die Referenzsequenz passen; total = Anzahl der gesamten Reads für die entsprechende Linie; %mapped = Prozentsatz der passenden Reads; map-length = Gesamtumfang der passenden Reads in bp).

plant	mapped	total	%mapped	map_length
Apache	216.455.080	332.425.316	65,1%	21.861.963.080
Bobwhite	189.952.727	253.904.810	74,8%	19.185.225.427
Bussard	205.875.216	370.914.790	55,5%	20.793.396.816
CHS	316.450.985	399.357.362	79,2%	31.961.549.485
Ephoros	220.295.925	362.002.184	60,9%	22.249.888.425
Expert	213.394.448	350.572.802	60,9%	21.552.839.248
Flair	274.741.867	369.463.924	74,4%	27.748.928.567
Flair30B	232.917.200	351.462.682	66,3%	23.524.637.200
Flair82C	262.675.049	430.025.080	61,1%	26.530.179.949
Hereford	231.649.896	349.694.544	66,2%	23.396.639.496
Inspiration	311.665.834	392.579.592	79,4%	31.478.249.234
Interet	325.286.608	409.267.112	79,5%	32.853.947.408
M6	319.028.373	402.597.454	79,2%	32.221.865.673
Mikon	282.130.060	355.772.166	79,3%	28.495.136.060
Mulan	279.282.146	455.913.096	61,3%	28.207.496.746
Opata	245.199.135	370.998.444	66,1%	24.765.112.635
Prinz	300.163.812	378.294.734	79,3%	30.316.545.012
Prinz34a	189.024.497	334.215.028	56,6%	19.091.474.197
Prinz51	275.760.328	350.677.374	78,6%	27.851.793.128
Recital	285.668.381	431.067.862	66,3%	28.852.506.481
Renan	246.203.379	372.036.618	66,2%	24.866.541.279
Sogood	326.942.123	412.224.434	79,3%	33.021.154.423
Sperber	232.216.074	354.234.406	65,6%	23.453.823.474
Victo	237.120.291	361.374.652	65,6%	23.949.149.391
XX86	231.433.623	316.326.416	73,2%	23.374.795.923

Tabelle 6: Anzahl der in den 25 Linien insgesamt beobachteten Polymorphismen pro Chromosom bzw. Chromosomenarm.

Chromosom	Deletion	Insertion	SNPs	Chromosom	Deletion	Insertion	SNPs
1AL	22.400	137.879	2.272.349	5AL	25.873	222.322	4.039.514
1AS	11.998	104.568	1.911.768	5AS	12.541	143.756	2.847.208
1BL	25.636	143.265	2.519.319	5BL	36.890	238.081	4.200.090
1BS	18.813	130.418	2.374.467	5BS	11.643	101.919	2.128.419
1DL	17.700	180.677	3.067.163	5DL	18.270	153.916	2.241.419
1DS	10.134	93.977	1.768.768	5DS	8.864	110.170	2.181.566
2AL	27.162	189.883	3.116.374	6AL	17.032	122.006	1.867.716
2AS	24.576	159.653	2.906.113	6AS	19.466	154.514	2.546.057
2BL	37.537	229.017	4.063.211	6BL	25.601	140.166	2.412.070
2BS	26.831	166.849	3.130.903	6BS	21.742	129.102	2.276.314
2DL	21.578	150.655	2.008.404	6DL	14.097	135.687	2.336.650
2DS	12.355	97.931	1.399.182	6DS	9.906	112.966	2.104.566
3AL	19.188	122.031	2.171.762	7AL	21.902	136.855	2.227.436
3AS	12.735	95.147	1.765.115	7AS	19.854	110.164	1.672.415
3B	65.254	426.360	7.278.087	7BL	25.664	139.533	2.017.177
3DL	14.726	93.298	1.176.680	7BS	16.852	114.380	1.910.371
3DS	13.315	87.964	1.323.855	7DL	15.860	146.868	2.341.974
4AL	28.639	181.891	3.012.128	7DS	17.738	136.563	1.998.022
4AS	14.161	125.093	2.286.746				
4BL	14.698	101.309	1.688.945				
4BS	15.591	145.007	2.856.059				
4DL	12.644	128.889	1.973.471				
4DS	7.129	93.139	1.789.857				

Wie zu erwarten war, kann man erkennen, dass die Anzahl der SNPs deutlich höher ist, als die Anzahl der Insertionen/Deletionen. Die Anzahl der Polymorphismen pro

Linie schwankt zwischen 900000 und mehr als 1,3 Millionen SNPs. Die meisten Unterschiede zur Referenz wurden, wie erwartet, in Linie M6 (Synthetic) beobachtet.

Bei der Vielzahl der gefundenen SNPs stellt sich auch die Frage nach der Häufigkeitsverteilung – d.h. wie oft (in wie vielen Linien) ein Polymorphismus gefunden wird. Bei einmalig (in einer Linie) beobachteten Polymorphismen ist eine gewisse Vorsicht angebracht, da die Sequence Capture Prozedur doch recht komplex ist und besonders durch PCR-Amplifikation auch Fehler in die Sequenzen eingebaut werden können. Bei einer Anzahl von 24 bzw. 25 Abweichungen im Vergleich zur Referenz ist davon auszugehen, dass die Referenz wiederum an der entsprechenden Position einen Fehler enthält. Die Häufigkeitsverteilung der Polymorphismen ist in der folgenden Abbildung 3 angegeben. Insgesamt sind mehr als 5 Mio. nutzbare SNPs vorhanden von denen mehr als eine Million SNPs in mehr als einer Linie und weniger als 23 Linien vorkamen.

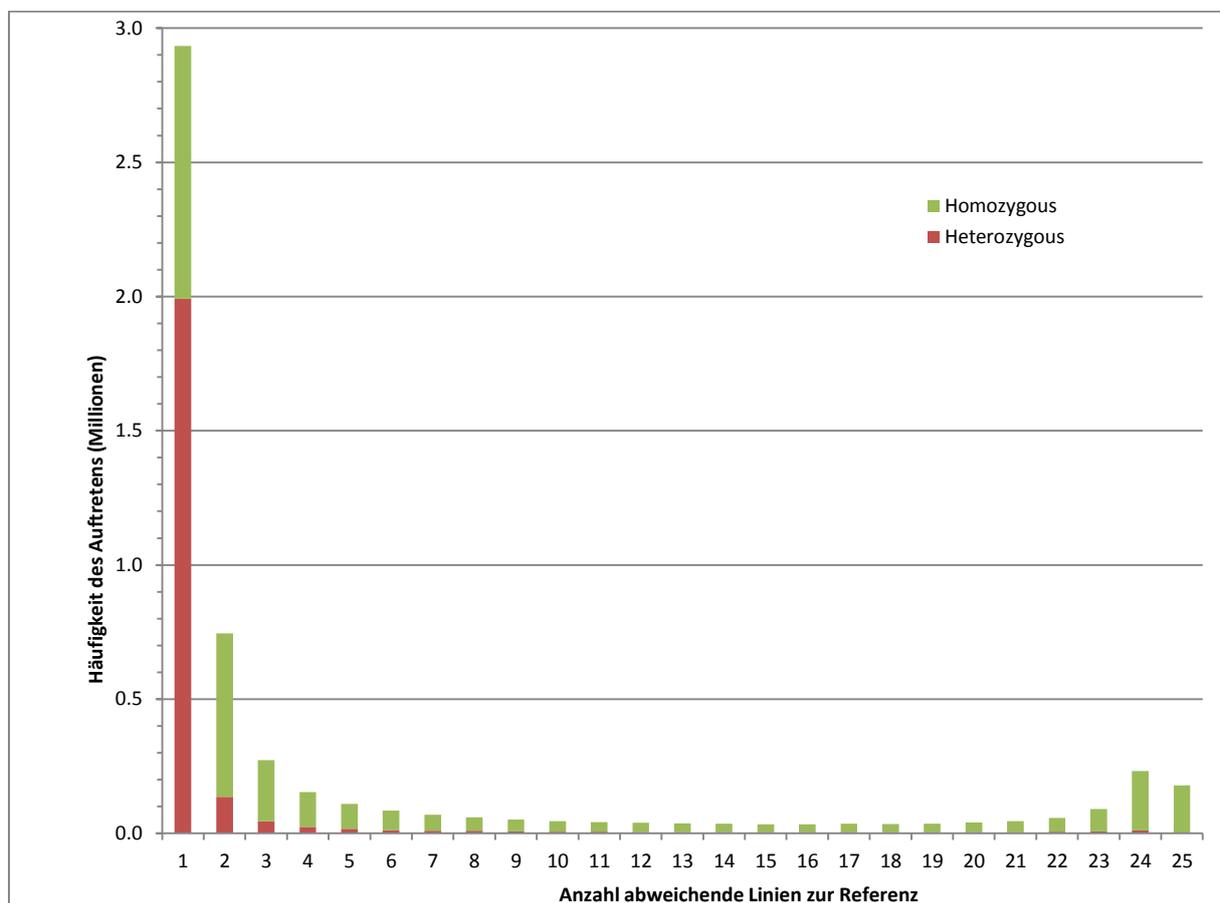


Abbildung 3: Häufigkeitsverteilung der beobachteten SNPs bezogen auf die gesamte Anzahl der untersuchten Linien.

Arbeitspaket 3: Genetisches Material und Validierung der identifizierten Marker für Samengröße und Tausendkorngewicht durch gezielte Markeranalyse

Materialentwicklung (IPK)

Die ITMI-DH (International Triticeae Mapping Initiative - Doppelt Haploide) Kartierungspopulation stellt eine relativ neue international genutzte Kartierungspopulation für Genotypdaten dar, welche mit neuen Markertechniken generiert werden (Sorrells et al. 2011; Poland et al. 2012). Wir haben Saatgut für 174 Linien aus dieser Population von Prof. Mark Sorrells, Cornell University, USA erhalten und genomische DNA isoliert. Diese Kartierungspopulation wurde als allgemeine Referenzpopulation genutzt.

Für die Feinkartierung eines Teils der Regionen mit QTLs für Samengröße und Tausendkorngewicht wurden für vier Kreuzungskombinationen Rückkreuzungslinien entwickelt, welche jeweils für 1-3 TKG QTLs segregieren. Mit Hilfe eines marker-gestützter Rückkreuzungsprogrammes wurden Introgressionen mit definierten genomischen Regionen für TKG QTL selektiert. In der Generation BC₂-F₂ wurde für homozygote Linien selektiert (Tabelle 7).

Tabelle 7: Selektion von homozygoten BC₂F₂-Linien für einzelne QTL in den vier Populationen

Population	Merkmal	Anzahl der selektierten Loci und Chromosomen	Anzahl der getesteten BC₂F₂-Linien	Anzahl der selektierten homozygoten Introgressionslinien
Ephoros x Mulan	TKG	3; Chr. 2D, 4A, 6D	408	180
Mikon x Inspiration	TKG	1; Chr. 3A	624	43
Interet x Sogood	TKG	1; Chr. 2D	144	41
Hereford x Expert	TKG	1; Chr. 1B	132	40

Aus den vier genannten Kreuzungskombinationen wurden parallel vier F₂-Kartierungspopulationen mit jeweils mehr als 200 Individuen erstellt, welche für die Kartierung der SNPs aus dem 135K Affymetrix Array (siehe Arbeitspaket 4) eingesetzt wurden.

Weitere Charakterisierung der QTL für Tausendkorngewicht (IPK)

Aus den Ergebnissen des ursprünglichen GABI-WHEAT-Projektes lagen umfangreiche Daten zu Marker/Merkmalassoziationen für Tausendkorngewicht vor. Beispielsweise wurden SNP-Daten und phänotypischen Daten in Kooperation mit

dem parallel laufenden VALID-Projekt (FKZ 0315947) weiter ausgewertet um präzisere Ergebnisse und Bestätigungen für die in diesem Projekt untersuchten QTL zu bekommen. Die Ergebnisse dieser Analysen sind in der Publikation von der Projektmitarbeiterin Frau Dr. C. Zanke (Zanke et al. 2015) detailliert dargestellt und werden deshalb hier nicht noch einmal im Detail wiederholt. Insgesamt konnten so die einzelnen QTL genauer beschrieben werden.

Entwicklung von SNP-Markern für die Verifikation und Kartierung der gefundenen Polymorphismen in Arbeitspaket 1 (IPK, TraitGenetics)

Basierend auf strikten technischen Kriterien (Coverage zwischen 20 und 200; nur homozygote SNP-calls) konnten 4052 SNPs vorhergesagt werden. Aus diesem Set wurden 191 KASP-Marker entwickelt, um die vorhergesagten SNPs in einer experimentellen Markeranalyse zu validieren (Tabelle 8). Von den 191 Markern waren 183 funktional und davon waren dann 119 KASP-Marker auf den entsprechenden Kreuzungen polymorph und konnten auf der 7D-NIL Population (Region 1), den vier BC₂-F₂ Populationen (Regionen 2 und 3) und der ITMI-DH Population (Region 4) kartiert werden.

Tabelle 8: Zusammenfassung der vorhergesagten SNPs und der getesteten KASP-Marker

Bereich	Vorhergesagte SNPs	Entwickelte KASP Marker	Polymorphe KASP Marker
I	491	84	57
II	457	22	12
III	657	36	27
IV	2447	49	23
Summe	4052	191	119

Feinkartierung und weitere Analyse der Marker (IPK, TraitGenetics)

Im nächsten Schritt wurden 39 polymorphe KASP-Marker für ausgewählte Regionen der Bereiche 2 und 3 auf den BC₂-F₂ Linien der vier Rückkreuzungspopulationen (Arbeitspaket 3.1) getestet und validiert.

Von den 77 funktionalen KASP-Markern aus Bereich I waren 61 KASP-Marker polymorph. Das bedeutet, dass 80% der getesteten KASP-Marker für diesen Bereich den Erwartungen entsprachen. Dies ist sehr zufriedenstellend und die Quote liegt

damit im oberen Bereich der Erfolgsquote für Weizen-SNPs generell. Zur weiteren Bestätigung wurden sechs KASP-Marker innerhalb der QTL-Region *QTgw.ipk-7D* (Bereich 1) auf Chromosomenarm 7D, welche mit 12 assemblierten und annotierten BACs abgedeckt war, fein kartiert (Abbildung 4).

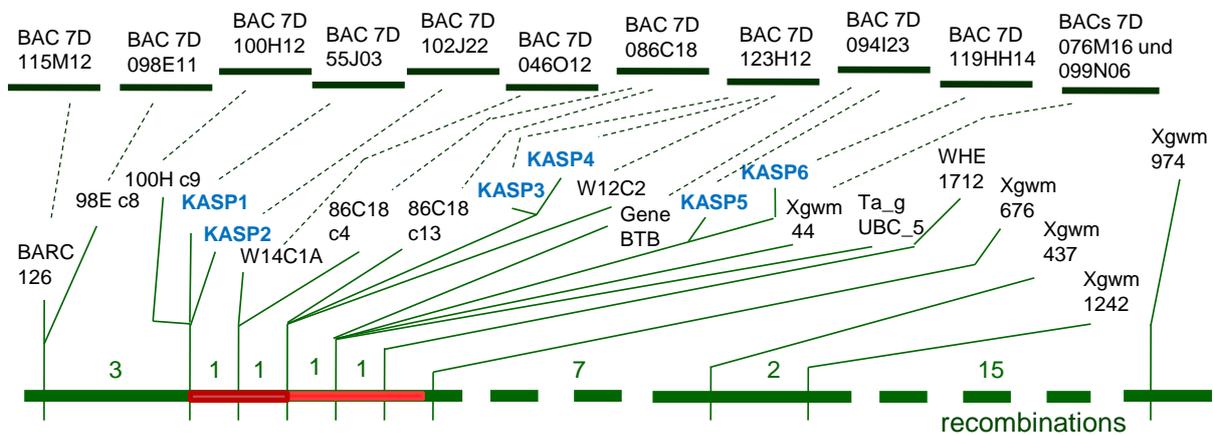


Abbildung 4: Genetische und physikalische Karte der QTL-Region *QTgw.ipk-7D*. Die neu entwickelten KASP-Marker sind blau dargestellt.

Zusammenfassend lässt sich damit feststellen, dass die im Sequence Capture identifizierten Marker zum überwiegenden Teil auch tatsächlich in den selektierten Genombereichen kartieren und damit der gesamte Ansatz zur Markeranreicherung funktioniert hat. Damit war die Entwicklung eines großen und teuren Genotypisierungsarrays (Arbeitspaket 4) gerechtfertigt.

Arbeitspaket 4: Validierung einer großen Zahl identifizierter SNPs in Weizensorten und segregierendem Material mit Hilfe eines Genotypisierungsarrays

Design eines Affymetrix-Arrays zur genetischen Kartierung vieler identifizierter SNP-Marker (TraitGenetics)

Im ursprünglichen Projektplan war beabsichtigt, einen Genotypisierungsarray mit etwa 10000 Markern auf der Illumina-Plattform zur Analyse von Weizensorten und Kartierungspopulationen zu verwenden. Während der Projektbearbeitung wurde aber von der Firma Affymetrix ein neues Genotypisierungssystem auf Arraybasis etabliert und kommerziell angeboten. Dieses System erlaubt es mehr als 10x so viele SNP-Markern auf einem experimentellen Array mit nur geringfügig höheren Kosten als bei der Illumina-Plattform zu untersuchen. Da wir aus dem Sequence Capture mehrere Millionen SNPs identifiziert hatten, wurde zusammen mit dem Projektpartner

entschieden, statt einem 10K Illumina-Array einen 135K Affymetrix-Array zu entwickeln.

Aus dem Sequence Capture Experiment, welches in Arbeitspaket 2 beschrieben wurde, wurden alle Roh-SNPs (etwa 5 Mio.) als Pool für das Arraydesign ausgewählt. In dem ersten Selektionsschritt durften weiter verwendete SNPs nicht intergenomisch sein und die untersuchte Position mussten durch mind. 6 Pflanzen (=1/4 des Panel) abgedeckt sein.

Diese Kandidaten-SNPs des SNP-Pools mussten in einem zweiten Schritt folgende technische Kriterien erfüllen:

- 2 oder 3 verschiedene Genotypen wurden detektiert (AB, AH, ABH)
- Die Allelfrequenz des seltenen Allels musste $\geq 25\%$ sein
- In der flankierenden Sequenz des jeweiligen SNPs sollten innerhalb von 50 bp upstream und 10 bp downstream (bzw. umgekehrt) keine weiteren Polymorphismen beobachtet werden
- Die jeweilige Coverage sollte unter 200x liegen um duplizierte Genombereiche zu eliminieren

In einem dritten Schritt wurden folgende genetische Selektionskriterien angewendet:

- Kopiezahl der entsprechenden flankierenden Sequenz in der gesamten bekannten Weizengenomsequenz unter 5 Kopien (via BLAST-Analyse)
- Der Haplotyp sollte weitgehend unique sein (flankierende SNPs nicht im Kopplungsgleichgewicht in den untersuchten Linien)
- Für Chromosom 7D wurden alle verfügbare SNPs doppelt (vorwärts und rückwärts) berücksichtigt
- Für die Tausendkorngewicht-Target-Regionen wurden doppelt so viele SNPs ausgewählt wie für die Nicht-Target-Regionen

Im Rahmen einer weiteren Überprüfung dieser letztendlich ausgewählten SNPs hinsichtlich ihrer Qualität wurde festgestellt, dass ca. 4.000 dieser SNPs auch auf dem bekannten 90k illumina Wheat Chip (Wang et al., 2014) und ca. 1.500 auf dem 35K Affymetrix Wheat Chip (Auswahl von Markern aus dem Winfield et al. 2015 Array) vertreten waren. Damit konnte davon ausgegangen werden, dass unsere identifizierten SNPs echt sind und auf dem Array nutzbar sein sollten. Tabelle 9 zeigt die Zahl der SNPs für die ausgewählten Regionen der Kartierungspopulationen auf dem 135K Array.

Tabelle 9: Verteilung der kartierbaren SNPs auf den Target-Regionen aller Kartierungspopulationen

Chrom	Population					Total
	Hereford	Mikon	Mulan	Opata	Sogood	
	Expert	Inspiration	Ephoros	M6	Interet	
1BL	2.348					2.348
1BS	875					875
2DL			1.298		1.475	2.773
2DS			1.402		1.201	2.603
3AL		3.464				3.464
3AS		1.694				1.694
4AL			2.654			2.654
4AS			580			580
5BL		7.188				7.188
5BS		1.131				1.131
6AL				3.885		3.885
6AS				3.099		3.099
6BL				4.126		4.126
6BS				4.075		4.075
6DL			1.460	3.352		4.812
6DS			775	2.947		3.722
total	3.223	13.477	8.169	21.484	2.676	49.029

Für die technische Prüfung durch den Arrayhersteller wurden insgesamt ca. 630.000 Kandidaten-SNPs an Affymetrix verschickt. Für diese SNPs wurden von Affymetrix Empfehlungen zur Nutzung abgegeben. Diese Empfehlungen wurden für die finale Auswahl der SNPs berücksichtigt.

Die Selektion der weiteren genomweit verteilten SNPs erfolgte anhand der o.g. Auswahlkriterien mit perl-scripten. Tabelle 10 zeigt die endgültige Verteilung der selektierten SNPs auf dem 135K Array.

Tabelle 10: Chromosomale Verteilung der meisten Marker auf dem 135K Array basierend auf der zu diesem Zeitpunkt verfügbaren Weizengenomsequenz:

Chromosome	# Markers	Chromosome	# Markers
1AL	4.780	4DL	1.656
1AS	1.695	4DS	878
1BL	5.164	5AL	6.000
1BS	2.088	5AS	1.698
1DL	2.946	5BL	6.662
1DS	1.098	5BS	1.457
2AL	4.496	5DL	3.950
2AS	3.669	5DS	1.154
2BL	4.762	6AL	2.858
2BS	3.675	6AS	2.599
2DL	3.304	6BL	3.832
2DS	2.505	6BS	3.519
3AL	4.835	6DL	2.438
3AS	2.177	6DS	1.788
3B	9.450	7AL	3.252
3DL	3.680	7AS	3.659
3DS	2.075	7BL	3.555
4AL	4.196	7BS	2.328
4AS	1.696	7DL	2.203
4BL	2.904	7DS	2.863
4BS	2.834	total	132.378

SNP Analyse von Weizenlinien mit dem 135K Affymetrix Array (TraitGenetics, IPK)

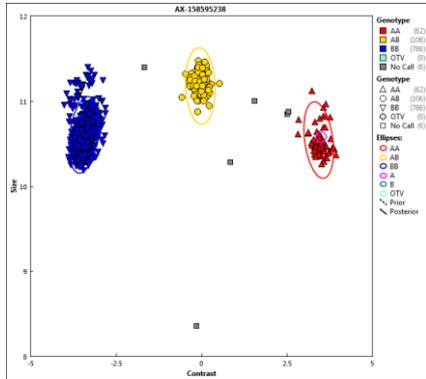
Mit dem Array wurden 960 Proben (10 Arrays) erfolgreich analysiert. Die untersuchten Proben bestanden zum einen aus 192 diversen Winterweizenlinien, die bereits vorher mit einer großen Anzahl von SNP Markern erfolgreich getestet worden waren (Illumina Infinium 90K Array, Affymetrix Wheat Breeders 35K Array). Dieses Sortenset ist ausgezeichnet geeignet, um eine aussagekräftige Qualitätskontrolle der neuen SNP Marker im Vergleich zu bereits auf Arrays analysierten SNP-Markern zu gewährleisten. Desweiteren wurden die in dem Projekt erzeugten vier spaltenden Populationen mit je 192 Proben getestet (Ephoros/Mulan, Mikon/Inspiration, Hereford/Expert und Intérét/Sogood). Die Daten für die einzelnen Marker wurden mit der Affymetrix Axiom Workbench (Axiom Analysis Suite AXAS) zunächst in sechs Klassen (Conversion Types) eingestuft.

Die folgende Abbildung 5 zeigt Beispiele für die Klassifizierung der Marker mit hohen und niedrigen FLD Score Werten. Der FLD-Score ist hoch, wenn alle allele calls eng geclustert sind und die drei Clusterzentren möglichst weit voneinander entfernt

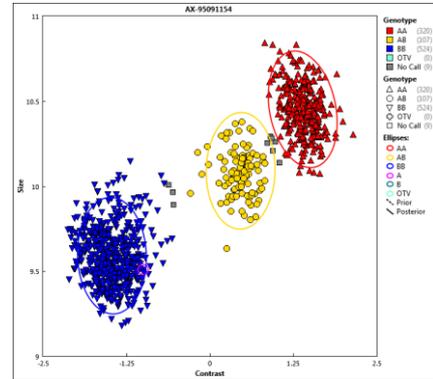
liegen. Bei Markern mit hohem FLD-Score können die Allele zuverlässiger klassifiziert werden.

Conversionstyp Poly High Resolution (PHR)

Hoher FLD Score

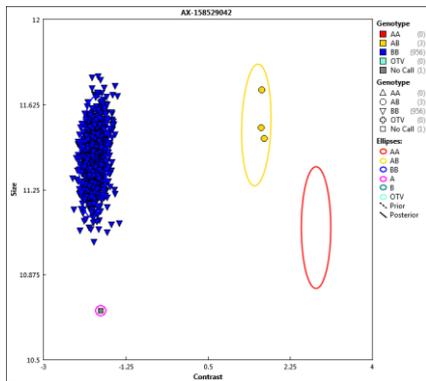


niedriger FLD Score

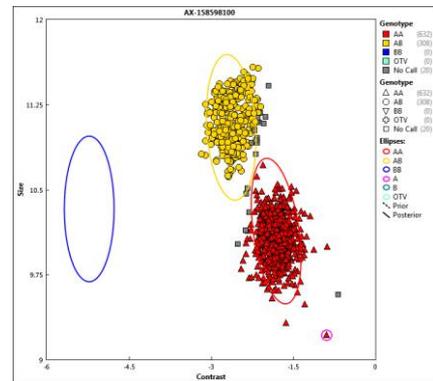


Conversionstyp No Minor Hom (NMH)

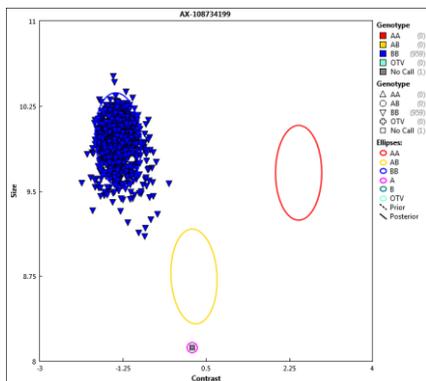
Hoher FLD Score



niedriger FLD Score

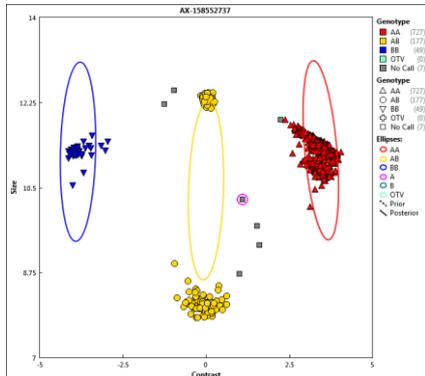


Conversionstyp Mono High Resolution (MHR)



Conversionstyp Off Target Variants (OTV)

Hoher FLD Score



niedriger FLD Score

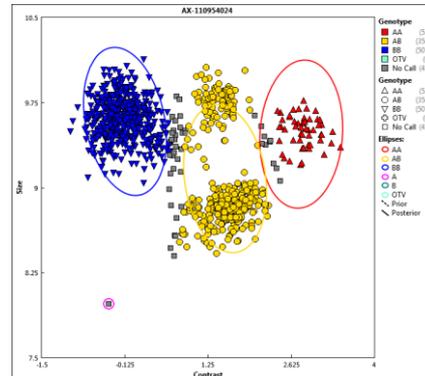


Abbildung 5: Beispiele für die Klassifizierung der SNP Marker in die wichtigsten Conversionstypen, jeweils mit hohen und niedrigen FLD Score Werten

Insgesamt ergab sich die in Tabelle 11 dargestellte Einstufung in die einzelnen Conversionstypen. Die Nutzbarkeit der Marker bezüglich der Einordnung in die Conversion Types wurde dann wie folgt zu bewertet:

Tabelle 11: Einstufung der SNP Marker in die Conversionstypen

PolyHighResolution:	36,483 (27%)	nutzbar
NoMinorHom:	41,283 (30%)	nutzbar
MonoHighResolution:	23,081 (17%)	nicht nutzbar
CallRateBelowThreshold:	8,240 (6%)	nicht nutzbar
OffTargetVariants (OTV)	2,437 (2%)	teilweise nutzbar
Other:	25,256 (18%)	teilweise nutzbar

Gesamt 136,780 (100%)

Die weitere Klassifizierung der Marker kann über die Call Rate (sollte im Schnitt über 97,5% liegen), über die Anzahl an Clustern (Nclus 2 oder 3) oder die Allel-Frequenz (Minor Allele Frequency, MAF) vorgenommen werden. Diese Klassifikation ist allerdings abhängig vom untersuchten Material.

Zusammenfassend ist festzustellen, dass 56,7% der Marker (77765) von sehr guter Qualität sind (Conversionstypen PolyHighResolution und NoMinorHom). Die Marker der Kategorie NoMinorHom sind dabei nur im Weizensortenpanel polymorph und nicht in den Kartierungspopulationen.

Zusätzliche nutzbare Marker können noch aus den anderen Conversionstypen (siehe Anmerkungen in Tabelle 11) gewonnen werden. Wenn man diese Markerzahlen

addiert, kommt man zu etwa 100000 nutzbaren Markern. Dies entspricht etwa 75% der ursprünglich auf dem Array eingesetzten Marker. Damit ist die Erfolgsquote deutlich höher als bei dem Infinium 90K Array von Wang et al. (2014). Das zeigt, dass unsere bioinformatische und genetische Markerauswahl sehr zuverlässig war und mit diesen Daten eine detaillierte genetische Analyse möglich ist.

Arbeitspaket 5: Umfassende bioinformatische und genetische Analyse der erhaltenen Daten (TraitGenetics, IPK)

Die Analyse aller erhaltenen Daten in diesem Arbeitspaket umfasste einerseits die fortlaufende Auswertung der Daten der einzelnen experimentellen Arbeitspakete in Vorbereitung auf die nächsten Arbeitsschritte und andererseits die analytische Auswertung der erhaltenen Daten. Die Auswertung der experimentellen Daten in den laufenden Arbeitspaketen ist weitgehend in den Berichten zu den einzelnen Arbeitsschritten 1-4 dargestellt und es wird hier nicht weiter darauf eingegangen.

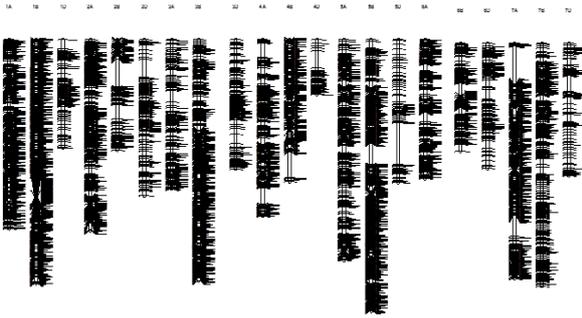
Vielmehr werden hier exemplarisch die Ergebnisse der analytischen Auswertung der genetischen Kartierung der Marker auf dem 135K Affymetrix Array dargestellt, da eine Absättigung der genetischen Karte, insbesondere einzelner Regionen mit interessanten QTL für Samengröße und Tausendkorngewicht, ein Hauptziel des Projektes war. Hierfür wurden zunächst alle Marker aller Conversionstypen in Betracht gezogen, die zwischen den Eltern der Kartierungspopulationen polymorph waren und die Polymorphismen auch in den F2-Nachkommen der Populationen zeigten. Anschließend wurde für jede Population eine vorläufige Karte erstellt. Die weitere Qualitätskontrolle wurde dann anhand der originalen Genotypisierungsdaten vorgenommen. Alle Marker, die die genetische Karte bzw. die Kopplungsgruppen unnötig verlängert haben – also jeweils einen großen Abstand zu ihren flankierenden Markern zeigten – wurden dabei eliminiert, da es sich bei solchen Markern meist um qualitativ minderwertige Marker handelt.

Mit den verbliebenen zuverlässigen SNP-Markern wurde anschließend die genetische Karte für jede Population erstellt. Tabelle 12 zeigt die in jeder Population kartierten Marker und Abbildung 6 die jeweilige genetische Karte.

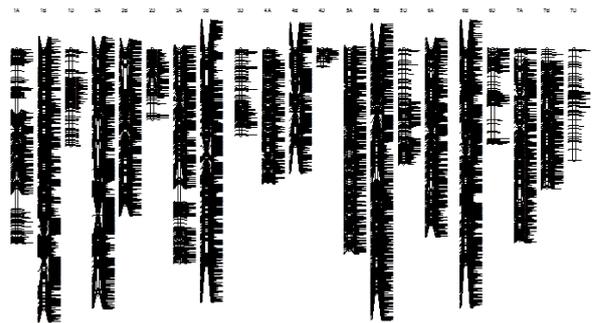
Tabelle 12: Anzahl der in den einzelnen Populationen kartierten Marker

Population	Anzahl kartierte Marker
Ephoros x Mulan	13383
Mikon x Inspiration	21663
Hereford x Expert	12956
Interet x Sogood	15730

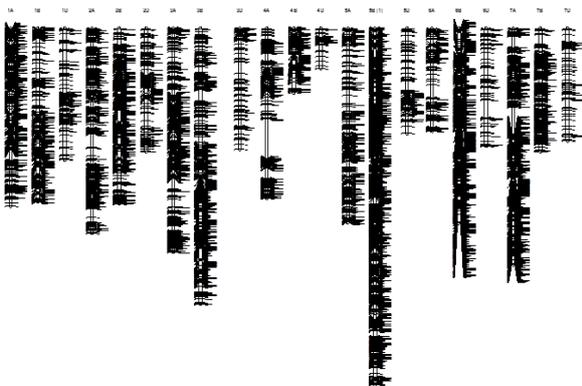
Ephoros x Mulan



Mikon x Inspiration



Hereford x Expert



Interet x Sogood

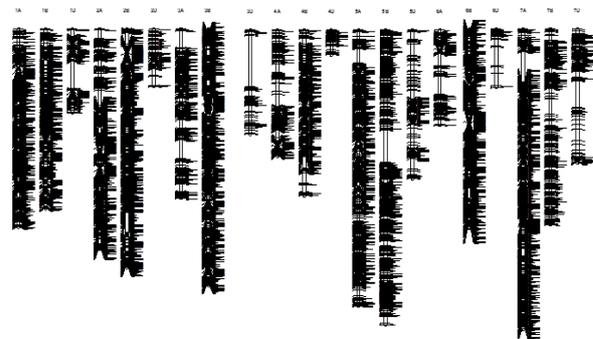


Abbildung 6: Genetische Karten der vier Populationen zu Markerabsättigung spezifischer Regionen mit interessanten QTL für Tausendkorngewicht.

Zusammen wurden in den vier analysierten Kartierungspopulationen 41.105 oder 30% der Marker auf dem Array genetisch kartiert. Die Abbildung 7 zeigt eine sogenannte BIN Map – eine zusammengefasst Karte aller vier Populationen, wobei die Länge der Chromosomen jeweils auf 100 gesetzt und die einzelnen Positionen der Marker entsprechend berechnet wurden. In der Abbildung 7 ist jeder Kartenort nur durch je einen Marker repräsentiert da alle Marker den Rahmen der Darstellung sprengen würden. Die Verteilung aller Marker über die einzelnen Chromosomen ist wie erwartet aber nicht gleichmäßig, da ja teilweise Marker für bestimmte QTL-Regionen angereichert wurden. Die angereicherten Chromosomen sind sehr gut mit Marker abgesättigt, wohingegen andere Chromosomen nur spärlicher besetzt sind.

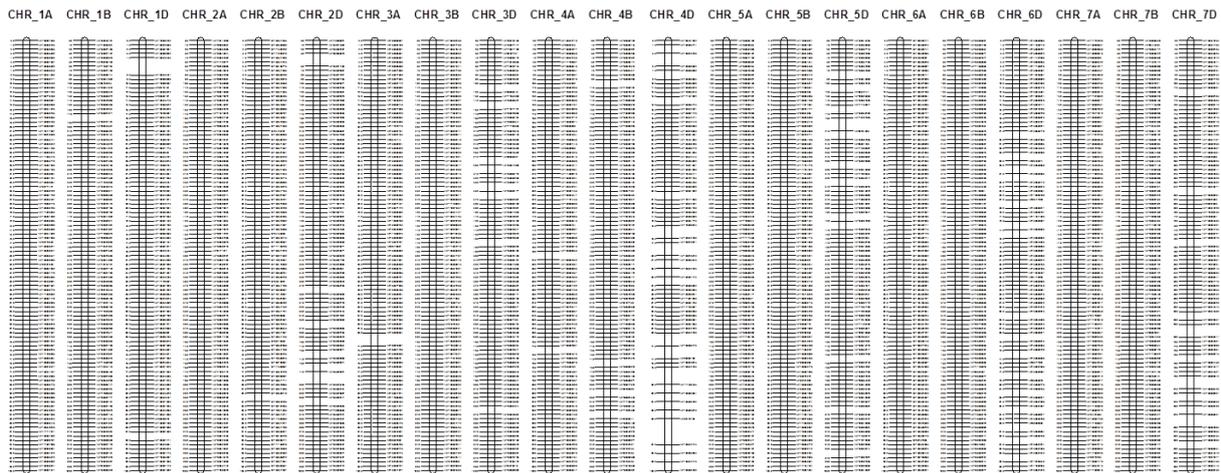


Abbildung 7: Genetische BIN Map der SNP Marker über alle vier untersuchten F2-Populationen. Die Marker-Orte werden jeweils nur von einem SNP Marker repräsentiert.

Im Zentromerbereich sind die Zahlen für die Marker pro Ort deutlich höher als an den Randbereichen (Telomeren) der Chromosomen. Generell sind die Chromosomen des D-Genoms deutlich weniger mit Markern besetzt als die für das A- und B-Genom. Wenn man die Daten der mit Markern angereicherten Chromosomen mit anderen genetischen Karten auf SNP-Basis vergleicht und die einzelnen Chromosomen in den einzelnen Karten genauer analysiert, dann ist auch eine klare Markeranreicherung auf diesen Chromosomen erkennbar. Die präzisen Analysen für die einzelnen Populationen und Chromosomen sind aber sehr umfangreich und noch in Arbeit.

Für eine detaillierte Betrachtung der Nutzbarkeit von SNP Marker Arrays in Sortenmaterial ist eine sogenannte Haplotypenstrukturanalyse erforderlich. Dazu werden die Markerallele farblich unterschiedlich visualisiert (Abbildung 8). Marker, die jeweils exakt gleiche Genotypisierungsmuster in allen untersuchten Sorten aufweisen, also zu 100% in LD sind, werden in einen Haplotypenblock eingeordnet. Bedingt durch die Markerauswahl in der genetischen Selektion gab es sehr wenige große Haplotypenblöcke (Tabelle 13).

Tabelle 13: Anzahl und Größe der Haplotypenblöcke (HX) in den untersuchten Weizensorten

marker / HX	HX	marker	average
> 100	9	1590	177
50 - 100	14	893	64
10 - 49	431	7186	17
5 - 9	868	5498	6
1 - 4	10624	15610	1.5
sum	11946	30777	

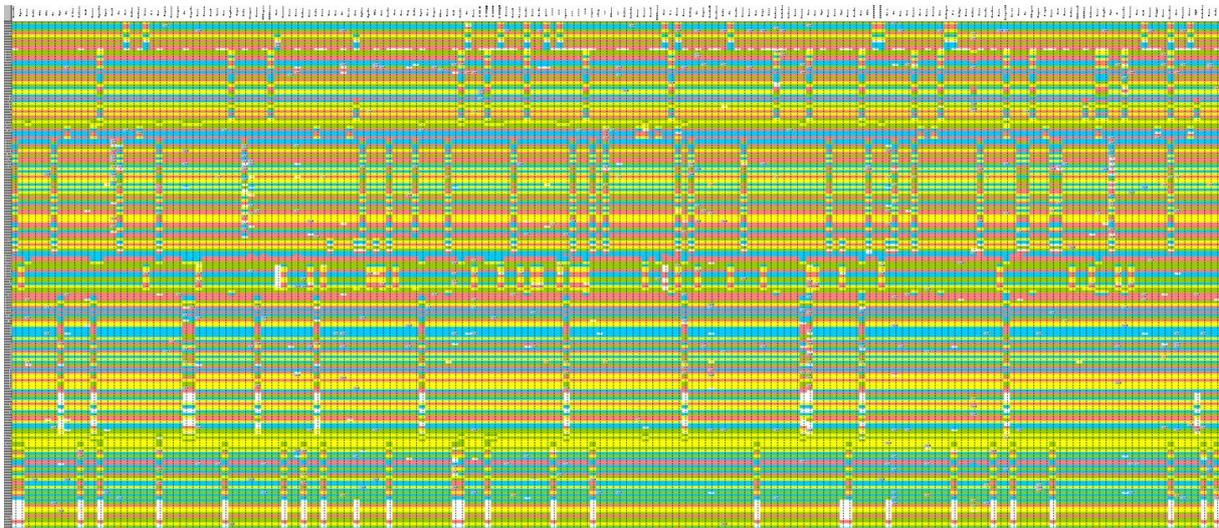


Abbildung 8: LD-Block Struktur der SNP Marker (von oben nach unten) des 135K Arrays in beispielhaft ausgewählten Europäischen Winterweizensorten (von links nach rechts). Die Marker sind nach LD-Blöcken sortiert.

Es ist festzuhalten, dass die Analyse des 135K Affymetrix Arrays auf den Weizensorten qualitativ sehr hochwertige Resultate geliefert hat, die im Vergleich zu bereits existierenden Arrays eine Reihe von neuen informativen SNP Markern geliefert haben und Lücken in der genetischen Karte des Weizengenoms schließen konnten.

Gesamtbetrachtung über das SELECT-Projekt

Insgesamt konnte das SELECT-Projekt für die meisten Fragestellungen weitgehend erfolgreich abgeschlossen werden. Die gängigen Sequence Capture Verfahren (Agilent und Nimblegen) und bioinformatische Verfahren für eine qualitativ hochwertige Datenauswertung konnten etabliert werden. Abweichend zur ursprünglichen Planung wurde im Rahmen des Projektes ein öffentlich verfügbares

Exome Capture System für das Weizengenom verwendet und es wurden keine eigenen Baits entwickelt. Für eine Reihe von QTL für Samengröße und Tausendkorngewicht konnten Linien (NILs) zur Feinkartierung der entsprechenden Regionen entwickelt werden. Durch die Entwicklung eines 135K Affymetrix Arrays konnte gegenüber der ursprünglichen Planung eine mehr als 10x höhere Anzahl an SNPs untersucht werden. Die Daten zeigen, dass es möglich ist, gezielt für spezifische Regionen des Weizengenoms zusätzliche Marker zu identifizieren und diese auf segregierenden Populationen zu kartieren.

Nach dem Ende des Projektes erfolgen noch weitere Auswertungen der erhaltenen großen Markerdatensets und eine detaillierte Charakterisierung des entwickelten genetischen Materials auf Tausendkorngewicht. Wenn die NILs soweit bestätigt sind, wird in der Zukunft versucht, die entsprechenden QTL weiter einzugrenzen und langfristig molekular zu isolieren.

2. Wichtige Positionen des zahlenmäßigen Nachweises

Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK)

Position	€
Personal	202.768,30
Verbrauchsmaterial	150.488,55
Reisekosten	2.647,37
Overhead	20.276,83
Unteraufträge	13.761,16
Sonstige Vorhabenkosten	0,00
Gesamt	389.942,21

3. Notwendigkeit und Angemessenheit der geleisteten Arbeit

Am IPK wurden folgende Arbeiten durchgeführt:

- Sequence Capture:
Das IPK unterstützte die TraitGenetics GmbH bei den Arbeiten für ‚Sequence Capture‘, sowohl für das Agilent System als auch für das Nimblegen System (‚Exome Capture‘). IPK stellte die DNA zur Verfügung, war an Qualitätsanalysen und Konzentrationsbestimmungen beteiligt und unterstützte die praktischen Arbeiten zum ‚Sequence Capture‘.
- Markergestützte Selektion für spezifische QTLs:
IPK entwickelte das Pflanzenmaterial für die näher untersuchten Tausendkorngewicht-QTL Regionen. Dazu wurden für vier Kreuzungskombinationen BC2-S2 Rückkreuzungslinien entwickelt und mit Hilfe von molekularen Markern homozygote Introgressionslinien selektiert. Dabei erfolgte für die markergestützte Rückkreuzungsanalyse einzelner QTLs für Korngröße/Tausendkorngewicht die Foreground-Analyse am IPK und die Background (genomweite) Analyse mit einem Array bei TraitGenetics. Die erhaltenen Linien werden derzeit im Feld des IPKs vermehrt und auf den Phänotyp TGW bonitiert.
- Entwicklung von Kartierungspopulationen:
Daneben war das IPK für die Entwicklung von Kartierungspopulationen verantwortlich. Dies umfasste die Anzucht und Vermehrung der verfügbaren

internationalen ITMI-DH Kartierungspopulation, sowie von vier selbst entwickelten Kartierungspopulationen, welche in Zusammenarbeit mit TG zur Kartierung des 135K Affymetrix-Chips verwendet wurden.

- Datenauswertung:

Die weitere Analyse der generierten Genotypisierungsdaten erfolgte durch das IPK und TraitGenetics zusammen.

4. Voraussichtlicher Nutzen und Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans

Durch das SELECT-Projekt konnte das IPK für alle Methoden und Analyseschritte zur Identifikation von Polymorphismen in angereicherten Fraktionen des Weizengenoms wichtige Erfahrungen sammeln, so dass diese Methode auch zukünftig in anderen Forschungsprojekten Anwendung finden kann.

Die aus dem Projekt erhaltenen genetischen Daten (vier kartierte Populationen und die Analyse eines Sortenpanels) liefern uns zusätzliche Informationen im Bereich des generellen Polymorphismusniveaus im Zuchtmaterial und bei dem Vergleich genetischer und physikalischer Informationen beim Weizen.

Letztendlich können mit den Kartierungsinformationen aus angereicherten Genombereichen mit interessanten QTLs für Korngröße/Tausendkorngewicht in der Zukunft die entsprechenden QTL-Regionen über die ebenfalls in dem Projekt erzeugten NILs näher eingegrenzt, genauer kartiert und langfristig die darunter liegenden Gene identifiziert werden.

5. Bekannt gewordene Fortschritte auf dem Gebiet des Vorhabens bei anderen Stellen

Während des Vorhabens sind folgende Fortschritte in den dargestellten Bereichen bei Weizen bekannt geworden, welche alle Arbeitspakete betreffen:

- Sequence Capture

Saintenac et al. (2011) Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biology* 12:R88

Winfield et al. (2012) Targeted re-sequencing of the allohexaploid wheat exome. *Plant Biotechnology Journal* 10:733-742

Allen et al. (2013) Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal* 11:279-295

Jordan et al. (2015) A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homeologous genomes. *Genome Biology* 16:R48

Die dargestellten Publikationen beschreiben die exemplarische Entwicklung und Verwendung von Sequence Capture bei tetraploidem und hexaploidem Weizen. In der Publikation von Jordan et al., 2015 wird das im Projekt verwendete Nimblegen Exome Capture System beschrieben. Die beschriebenen Verwendungen unterscheiden sich von der hier beschriebenen Nutzung.

- Affymetrix-Genotypisierung bei Weizen

Winfield et al. (2015) High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol J.* doi: 10.1111/pbi.12485

Die dargestellte Publikation beschreibt die erste Verwendung der Affymetrix-Technologie zur Genotypisierung bei Weizen. Der beschriebene Array ist weitgehend nur für interspezifisches Material verwendbar und erhält nur relativ wenige SNPs, welche zwischen Sorten polymorph sind.

- QTLs für Korngröße/Tausendkorngewicht

Wang et al. (2012) Identifying loci influencing 1,000-kernel weight in wheat by microsatellite screening for evidence of selection during breeding. *PLOS ONE* 7: e29432

Tyagi et al. (2015) Interval mapping and meta-QTL analysis of grain traits in common wheat (*Triticum aestivum* L.). *Euphytica* 201:367-380

Griffiths et al. (2015) Genetic dissection of grain size and grain number trade-offs in CIMMYT wheat germplasm. *PLOS ONE* 10:e0118847

Mohler V et al. (2016) Considering causal genes in the genetic dissection of kernel traits in common wheat. *Journal of Applied Genetics* (published online: 23 April 2016; DOI 10.1007/s13353-016-0349-2)

Huang Y, Kong Z, Wu X, Cheng R, Yu D, Ma Z (2015) Characterization of three wheat grain weight QTLs that differentially affect kernel dimensions. *Theoretical Applied Genetics* 128:2437-2445

Li Q, Zhang Y, Liu T, Wang F, Liu K, Chen J, Tian J (2015) Genetic analysis of kernel weight and kernel size in wheat (*Triticum aestivum* L.) using unconditional and conditional QTL mapping. *Molecular Breeding* 35:194

Wu et al. (2015) High-density genetic linkage map construction and QTL mapping of grain shape and size in the wheat population Yande1817 x Beinong6. *PLOS ONE* 10:e0118144

Su Z, Jin S, Lu Y, Zhang G, Chao S, Bai G (2016) Single nucleotide polymorphism tightly linked to a major QTL on chromosome 7A for both kernel length and kernel weight in wheat. *Molecular Breeding* 36:15

Die beschriebenen Publikationen beschreiben verschiedene Untersuchungen zu QTLs für Korngröße/Tausendkorngewicht (genomweite Untersuchungen und Charakterisierung einzelner QTLs). Sie sind praktisch alle parallel oder nach der Publikation aus dem SELECT-Projekt (Zanke et al. 2015) erschienen.

6. Erfolgen und geplante Veröffentlichungen des Ergebnisses

Publikationen in der Laufzeit des SELECT-Projektes (Stand Mai 2016)

Zanke, C., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., Argillier, O., Stiewe, G., Hinze, M., Neumann, F., Eichhorn, A., Polley, A., Jaenecke, C., Ganal, M.W., Röder, M.S.: Analysis of main effect QTL for thousand grain weight in European winter wheat (*Triticum aestivum* L.) by genome-wide association mapping. *Frontiers in Plant Science* 6:644 (2015)

Weitere Publikationen sind geplant. Sie werden den Affymetrix-Array und die genetischen Kartierungsdaten aus der Arrayanalyse in Bezug auf Korngröße/Tausendkorngewicht enthalten.

Vorträge bei internationalen Tagungen (nur Vortragende/r dargestellt)

In diesen Vorträgen wurden auch Teile des SELECT-Projektes dargestellt.

Ganal M.: Association Mapping in European Winter Wheat. Wheat Conference, Sydney, Australia, 20-25. 9. 2015

Röder M.: Association Mapping in European Winter Wheat. Plant Genetics and Breeding Technologies II, Vienna, Austria, Feb. 1-2, 2016

7. Zitierte Literatur:

Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theor Appl Genet* 119:507-517

Durstewitz G, Polley A, Plieske J, Luerksen H, Graner EM, Wieseke R, Ganal MW (2010) SNP discovery by amplicon sequencing and multiplex SNP genotyping in the allopolyploid species *Brassica napus*. *Genome* 53:948-956

Fu Y, Springer NM, Gerhardt DJ, Ying K, Yeh CT, Wu W, Swanson-Wagner R, D'Ascenzo M, Millard T, Freeberg L, Aoyama N, Kitzman J, Burgess D, Richmond T, Albert TJ, Barbazuk WB, Jeddloh JA, Schnable PS (2010). Repeat subtraction-mediated sequence capture from a complex genome. *Plant J* 62: 898-909

Jordan et al. (2015) A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homeologous genomes. *Genome Biology* 16:R48

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLOS ONE* 7: e32253

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

Teer JK, Mullikin JC (2010) Exome Sequencing: The Sweet Spot before Whole Genomes. *Hum Mol Genet* 19: R145-R151

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G,

Wieseke R, Plieske J; International Wheat Genome Sequencing Consortium, Lillemo M, Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo MC, Dvorak J, Morell M, Dubcovsky J, Ganai M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E. (2014) Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol J.* 12:787-96