



Bundesministerium
für Bildung
und Forschung

**Schlussbericht zur BMBF Förderaktivität
Pflanzenbiotechnologie**

**Verbundvorhaben RYE-SELECT
„Genom-basierte Präzisionszüchtungsstrategie für Roggen“ – Teilprojekt D**

Förderkennzeichen: 0315946D

Zuwendungsempfänger: Technische Universität München, 80290 München

Ausführende Stelle: Technische Universität München – Wissenschaftszentrum
Weihenstephan – Lehrstuhl für Pflanzenzüchtung, Liesel-Beckmann-Str. 2, 85354 Freising

Projektleitung: Dr. Eva Bauer

Projektlaufzeit: 01.11.2011 bis 31.08.2015

Freising, Februar 2016

Technische Universität München, Lehrstuhl für Pflanzenzüchtung

Inhalt

I. KURZDARSTELLUNG	3
1. Aufgabenstellung.....	3
2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde.....	3
3. Planung und Ablauf des Vorhabens.....	3
4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde	5
5. Zusammenarbeit mit anderen Stellen	5
II. EINGEHENDE DARSTELLUNG	6
1. Verwendung der Zuwendung und erzielte Ergebnisse.....	6
2. Wichtige Positionen des zahlenmäßigen Nachweises	23
3. Notwendigkeit und Angemessenheit der geleisteten Arbeit	23
4. Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse	24
5. Bekannt gewordene Fortschritte auf dem Gebiet des Vorhabens bei anderen Stellen	24
6. Erfolgte und geplante Veröffentlichungen der Ergebnisse	24
III ERFOLGSKONTROLLBERICHT	26
1 Beitrag des Ergebnisses zu förderpolitischen Zielen.....	26
2 Wissenschaftlich-technische Ergebnisse des Vorhabens, erreichte Nebenergebnisse und gesammelte wesentliche Erfahrungen	26
3 Fortschreibung des Verwertungsplanes	26
4 Arbeiten, die zu keiner Lösung geführt haben	27
5 Präsentationsmöglichkeiten für mögliche Nutzer	27
6 Einhaltung der Ausgaben- und Zeitplanung.....	27
IV KURZFASSUNG.....	27

I. KURZDARSTELLUNG

1. Aufgabenstellung

Zur Ausschöpfung des Ertragspotentials von Roggen spielen genombasierte Züchtungsstrategien eine wichtige Rolle. Sie erlauben eine Steigerung der Selektionseffizienz und die Beschleunigung des Züchtungsprozesses. Neueste technologische Entwicklungen (Hochdurchsatz-Sequenzierung und -Genotypisierung) haben neue Möglichkeiten für die genomweite und merkmalspezifische Verbesserung von Elitematerial zur Entwicklung von hohertragreichen Hybriden ermöglicht.

RYE-SELECT hatte folgende konkrete Ziele:

- Verbesserung der molekularen Werkzeuge für Roggen durch Schaffung einer Genomsequenzressource und Entwicklung von hochdichten Genotypisierungsarrays
- Charakterisierung der Diversität in Roggen Genpools durch populationsgenetische Analysen als Voraussetzung für effizientes Genpool-Management
- Evaluierung des Potentials der Genomischen Leistungsvorhersage (*genomic prediction*, GP) zur Verbesserung von Ertrag und Ertragsstabilität in Roggen
- Identifizierung und Validierung von Genomregionen, die züchterisch relevante Merkmale wie Frosttoleranz, Ertrag und Qualitätsmerkmale in Hybridroggen beeinflussen

2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Roggen ist eine ökonomisch wichtige Nutzpflanze für Ernährung, Futtermittel und Bioenergie in Deutschland und Osteuropa. Das Verbundprojekt RYE-SELECT wurde als Zusammenarbeit von sechs Partnern aus vier akademischen Institutionen und einem Wirtschaftspartner initiiert, um neueste Erkenntnisse der Genomforschung zur Ertragssteigerung und Ertragsstabilität unter variierenden Umwelteinflüssen in der Roggenzüchtung zu nutzen. Neben der Technischen Universität München (TUM) waren KWS Lochow GMBH (KWL), das IPK Gatersleben (IPK), drei Arbeitsgruppen der Universität Hohenheim (HOH-BI, HOH-ZI, HOH-LSA) und das Julius Kühn-Institut (JKI) beteiligt. Alle Partner ergänzen sich mit umfangreichen Erfahrungen auf den Gebieten Roggenforschung und -züchtung, Populationsgenetik, Genomanalyse sowie Bioinformatik und haben bereits in einer Vielzahl vom BMBF- und BMELV/GFP- finanzierten Vorgängerprojekten zusammengearbeitet. RYE-SELECT konnte deshalb auf molekulare Ressourcen und Pflanzenmaterial aufbauen, die in vorangegangenen Projekten (GABI RYE-EXPRESS, GABI RYE-FROST) entstanden sind. Parallel zum Vorhaben RYE-SELECT wurde im AgroClustEr Synbreed die Genomische Leistungsvorhersage in Mais untersucht sowie Expertise bei der Entwicklung von hochdichten Genotypisierungsarrays erworben. Dadurch war es möglich Synergien zu nutzen und von dem Erfahrungsaustausch zu profitieren.

3. Planung und Ablauf des Vorhabens

RYE-SELECT war in drei Arbeitspakete (*work packages*, WP) organisiert (Abb. 1). In WP1 wurden durch Schaffung einer Genomsequenzressource und vergleichende Sequenzierung Polymorphismen detektiert, aus denen hochdichte Genotypisierungsarrays entwickelt werden konnten. In WP2 wurden Strategien zur Nutzung der Genomischen Leistungsvorhersage in Roggen entwickelt. Diese wurden durch populationsgenetische Analysen zur Charakterisierung züchterisch relevanter Genpools unterstützt. In WP3 wurden Genomregionen neu identifiziert bzw. validiert, die wichtige agronomische Merkmale wie Frosttoleranz, Qualität, Fertilität, und Ertragskomponenten bestimmen.

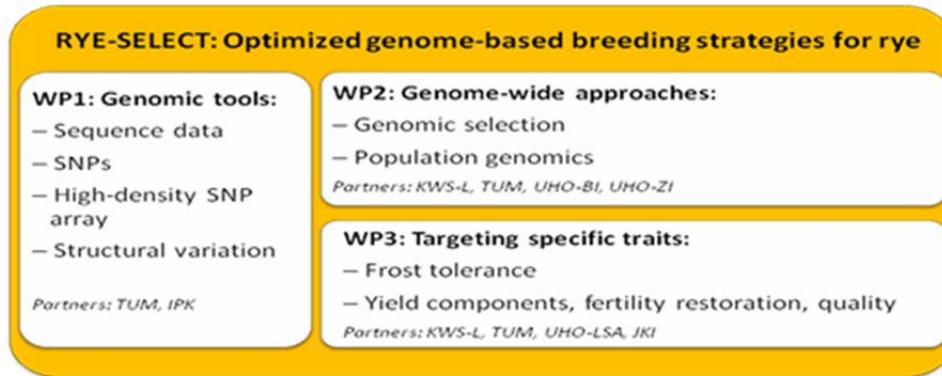


Abbildung 1: Projektstruktur von RYE-SELECT

Der Projektpartner TUM war an WP1, WP2, und WP3 beteiligt. Der zeitliche Ablauf und die vereinbarten Meilensteine sind in Abb. 2 dargestellt.

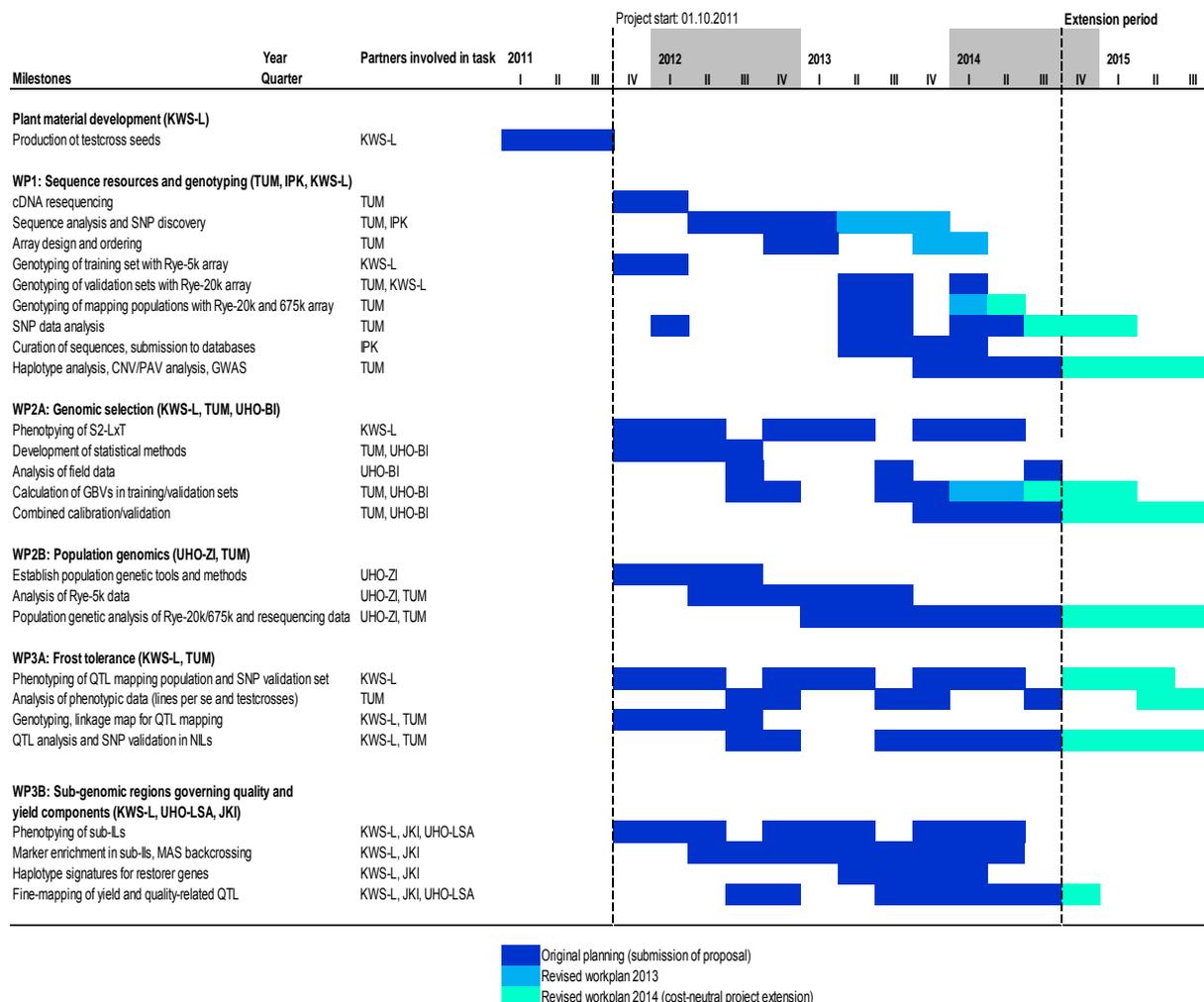


Abbildung 2: Zeitlicher Ablauf und Meilensteinplanung

Bedingt durch eine Verzögerung der Arbeitsabläufe in WP1 und WP2 und der Notwendigkeit zur Validierung von Ergebnissen in WP3 wurde eine zuwendungsneutrale Laufzeitverlängerung des Teilprojektes D um 11 Monate bis zum 31.08.2015 beantragt. Bis zum Ende des Verlängerungszeitraumes konnten alle Meilensteine erfolgreich bearbeitet werden.

4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Fortschritte in der Hochdurchsatz Sequenzierung und in Genotypisierungsmethoden eröffnen neue Möglichkeiten für eine effiziente Verbesserung wichtiger agronomischer Merkmale in Roggen. Für komplexe Merkmale, die durch viele Gene beeinflusst werden, wie z.B. dem Ertrag oder der Ertragsstabilität, müssen zunächst innovative, neue Konzepte entwickelt werden. Die vielversprechende Methode der Genomischen Leistungsvorhersage wurde ursprünglich für die Rinderzüchtung entwickelt und zeigte auch in ersten Studien Vorteile für Maiszuchtprogramme (Albrecht et al. 2011, Meuwissen et al. 2001). In RYE-SELECT sollte das Konzept der Genomischen Leistungsvorhersage für Roggen etabliert werden. Durch das Projekt GABI RYE-EXPRESS sind in Zusammenarbeit mit IPK eine Roggen EST-Kollektion und ein erster 5k SNP-Genotypisierungsarray entstanden (Haseneyer et al. 2011). Diese dienen als Ausgangspunkt für die Weiterentwicklung der genomischen Sequenzressourcen und von Genotypisierungsarrays für die genomische Leistungsvorhersage und die Analyse spezifischer Genomregionen in Roggen. In früheren GABI RYE Projekten wurden biologische Ressourcen wie Introgressionsbibliotheken geschaffen, sowie Donormaterial für die Analyse spezifischer agronomischer Merkmale entwickelt, die in RYE-SELECT weiter analysiert werden sollten. Hoch effiziente Phänotypisierungsplattformen für Feldversuche wurden durch den Wirtschaftspartner sowie HOH-LSA und JKI zur Verfügung gestellt.

5. Zusammenarbeit mit anderen Stellen

Es fand innerhalb und zwischen den einzelnen Arbeitspaketen eine enge Zusammenarbeit mit allen RYE-SELECT Projektpartnern statt. Bei der Analyse der Genomsequenzdaten gab es eine Kooperation mit der Arbeitsgruppe von Prof. Klaus F.X. Mayer vom Helmholtz Zentrum München, die eine Einbindung der neu geschaffenen Roggengenomressourcen in den größeren Kontext von Triticeae und anderen sequenzierten Gräsergenomen erlaubte. Zur Verankerung von Sequenzcontigs aus der Roggengenomressource anhand optischer Genomkarten wurde im Rahmen eines Pilotversuchs mit Prof. Jaroslav Dolezel (Institute of Experimental Botany, Olomouc, Tschechien) kooperiert. Im Bereich der Genomischen Leistungsvorhersage konnte RYE-SELECT von den Entwicklungen im AgroClustER Synbreed profitieren, in dem ähnliche Fragestellungen in Mais bearbeitet und Software- und Analysetools sowie ein hochdichter Genotypisierungsarray für Mais entwickelt wurden (Albrecht et al. 2011 und 2014, Unterseer et al. 2014, Wimmer et al. 2012).

II. EINGEHENDE DARSTELLUNG

1. Verwendung der Zuwendung und erzielte Ergebnisse

WP1: Genomische Ressourcen und Werkzeuge

1) *Whole Genome Shotgun* Sequenzierung

Die rasch voranschreitenden Entwicklungen der *Next Generation Sequencing* Techniken wurden genutzt, um eine Roggen Genomreferenzsequenz durch *Whole Genome Shotgun* (WGS) Sequenzierung zu generieren. Für die Erstellung der Referenzsequenz wurde die stark ingezüchtete und damit weitgehend homozygote Winterroggenlinie Lo7 ausgewählt. Außerdem wurden zehn diverse Inzuchtlinien resequenziert, von denen je die Hälfte aus dem Saat- bzw. Polleneltherpool stammen und die damit die beiden für die Hybridzüchtung wichtigen Materialgruppen repräsentieren. Als Vertreter einer Roggenwildart wurde eine Einzelpflanze aus einer *Secale vavilovii* Akzession sequenziert. TUM hat die DNA-Proben dem Dienstleister Eurofins Genomics GmbH (Ebersberg) zur Sequenzierung verschiedener Banken mit unterschiedlichen Fragmentgrößen auf einer Illumina HiSeq 2000 Plattform übergeben und die erzeugten Sequenzdaten im *fastq* Format anschließend an den Projektpartner IPK für die bioinformatischen Analysen weitergeleitet. Zusammen mit IPK wurde eine Strategie zur Assemblierung des Referenzgenoms und zur Analyse der Resequenzierungsdaten für die Charakterisierung von Polymorphismen entwickelt. Die Ergebnisse der bioinformatischen Analysen und die Charakterisierung der Roggengenomressource sind im Bericht des IPK eingehend dargestellt. Von IPK wurden den Projektpartnern sowohl ein umfassendes Assembly der Referenzlinie Lo7 als auch Einzelassemblies der resequenzierten Linien zur Verfügung gestellt, die für sequenzbasierte Analysen wie z.B. die Entwicklung genspezifischer Marker oder Sequenzvergleiche durch BLAST-Analysen in anderen Arbeitspaketen (WP2B, WP3) erfolgreich genutzt wurden.

Aufgrund der Größe des Roggengenoms und des WGS Sequenzierungsansatzes hat die Qualität der Referenzsequenz zwar noch nicht den gleichen Stand erreicht wie dies derzeit in international deutlich intensiver bearbeiteten Kulturarten wie Mais, Weizen oder Gerste der Fall ist. Dennoch ist eine hervorragende Ausgangsbasis geschaffen, die unter Verwendung weiterer Technologien wie z.B. der Sequenzierplattform PacBio im Rahmen zukünftiger Projekte kontinuierlich verbessert werden kann. Längere Sequenzinformationen würden zu längeren Contigs und damit zu einer besseren Abdeckung des Genoms beitragen, was auch die Annotation codierender Sequenzen qualitativ weiter verbessern würde. Ein weiterer vielversprechender Ansatz zur Verbesserung genomischer Sequenzressourcen, für strukturelle Genomanalysen und zur Verankerung von WGS Sequenzcontigs entlang einer physikalischen Karte sind sogenannte „optische Genomkarten“ wie sie von der Fa. BioNano unter Nutzung der *Irys* Plattform angeboten werden. Voraussetzung für die Verankerung von WGS Contigs entlang einer derartigen physikalischen Karte sind Contigs/Scaffolds mit einer Größe von mindestens 20 kb. Im gegenwärtigen Lo7 Assembly erfüllt zwar nur ein geringer Anteil von 2 % (22.281 Contigs/Scaffolds) diese Voraussetzung, dennoch wurde ein Pilotversuch begonnen mit der Arbeitsgruppe von Prof. J. Dolezel (Olomouc, Tschechien) zur Erstellung einer physikalischen Genomkarte für die sequenzierte Linie Lo7, um über die Verankerung der WGS Contigs Einsichten in die physikalische Genomstruktur bei Roggen zu gewinnen. Erste Testläufe mit der aus Wurzelspitzen nach Flow-Sorting isolierten hochmolekularen DNA und Nutzung der beiden Nicking-Endonukleasen *NbBvC1* und *NtBspQ1* verliefen vielversprechend und wiesen eine Dichte der Markierungssignale von 9,0 bzw. 11,4 Signalen pro 100 kb auf. Für weitere Analysen wird das Enzym *NtBspQ1* verwendet, mit dem ein N50-Wert von 240 kb für die kartierten Moleküle erreicht wurde. Aufgrund zwischenzeitlicher technischer Probleme mit der *Irys* Plattform sind diese Analysen derzeit noch nicht abgeschlossen. Es wird erwartet, dass eine ca. 200-fache Genomabdeckung erreicht werden kann, wodurch eine physikalische Genomkarte in hoher Qualität und Genomabdeckung möglich sein sollte, die eine wichtige Basis für die weitere Verbesserung der im Projekt erstellten Roggengenomsequenzressource darstellt.

2) Entwicklung eines hochdichten Genotypisierungsarrays

Aufgrund der Genomgröße von Roggen (ca. 8 Gb) und des für Fremdbefruchter typischen raschen Abfalls des Kopplungsphasenungleichgewichts (*linkage disequilibrium* – LD) (Li et al. 2011a) bestand ein dringender Bedarf für die Etablierung einer deutlich größeren Anzahl genomweiter Marker als bisher verfügbar. Die Resequenzierung von zehn Inzuchtlinien und von *S. vavilovii* verfolgte daher zwei Ziele. Zum einen sollte die Diversität in Roggen anhand von Sequenzvarianten (*single nucleotide variants* – SNVs, Insertionen/Deletionen – Indels) umfassend genomweit charakterisiert werden. Des Weiteren bildeten die ca. 8,3 Mio. detektierten Sequenzvarianten die Basis für die Entwicklung eines hochdichten Genotypisierungsarrays für Roggen unter Verwendung der Affymetrix Axiom® Technologie. Zusammen mit IPK wurden Filterungskriterien festgelegt, anhand derer die Anzahl der Varianten auf rund 600.000 (600k) eingeschränkt wurde. Dabei sollte eine möglichst gleichmäßige Genomabdeckung, die Abdeckung codierender Regionen und eine gute Repräsentierung der Diversität der beiden Genpools sowie der Wildart gewährleistet werden. Um eine optimale Übertragbarkeit der Ergebnisse mit früheren Studien zu gewährleisten wurden 10.250 technisch funktionale Marker des im Projekt RYE-EXPRESS entwickelten 5k Arrays (Haseneyer et al. 2011) sowie eines von Projektpartner KWL entwickelten 16k Arrays ausgewählt, der auch in WP2A zum Einsatz kam. Anhand dieses Arrays konnten die *in silico* vorhergesagten Varianten experimentell überprüft, sowie Daten für die Erstellung einer hochdichten genetischen Kopplungskarte zur Verankerung von Sequenzcontigs und für populationsgenetische Analysen generiert werden.

Insgesamt 576 DNA-Proben für die Analyse der 600k Arrays wurden von Partner KWL und von TUM bereitgestellt und auf der Affymetrix Gene Titan® Plattform am Lehrstuhl für Tierzucht der TUM prozessiert. Unter den 576 Proben befanden sich eine Kartierungspopulation von 133 rekombinante Inzuchtlinien (RILs) mit den zugehörigen Eltern, 287 S2-Linien für weitere Analysen in WP2A zur Abschätzung der genomischen Vorhersagegenauigkeit mit hochdichten SNP-Daten, ein Diversitätspanel mit 82 Elitelinien und 43 Proben von genetischen Ressourcen für populationsgenetische Analysen in WP2B, sowie Weizen-Roggen-Additionslinien, Triplets (A, B, A x B), technische Replikate und Wasserproben für Qualitätskontrollen. Von insgesamt 571 DNA-Proben waren bei zwei Ausfällen 569 auswertbar (99,6%). Die *Call Rates* der Roggenproben lagen dabei in der Regel über 90%. Lediglich Weizen-Roggen-Additionslinien und die Weizensorte Chinese Spring lagen mit 74-78% erwartungsgemäß deutlich niedriger. Die Auswertungsroutine der Affymetrix Genotypisierungsarrays teilt die SNPs in verschiedene Klassen ein, je nach den Mustern, die sich bei der Clusterung der Hybridisierungssignale ergeben. Am besten geeignet sind sogenannte „*Poly High Resolution*“ (PHR) SNPs, die eine klare Zuordnung in zwei homozygote (AA, BB) und ein heterozygotes (AB) Cluster erlauben. Die Klasse der sogenannten „*Off-Target Variants*“ (OTV) enthält SNPs, bei denen ein viertes Cluster mit geringerer Signalintensität detektiert wird. Hierbei handelt es sich um zusätzliche Sequenzvarianten, die neben den eigentlichen Ziel-SNPs (*target SNPs*) nachgewiesen werden. In der Klasse „*No Minor Homozygotes*“ (NMH) fehlt eines der homozygoten Cluster, während die Klasse „*Mono High Resolution*“ (MHR) monomorphe SNPs enthält. Schließlich gibt es noch die Klassen „*Other*“ und „*Call Rate Below Threshold*“ (CBT), die verschiedene Qualitätskriterien nicht erfüllen und daher als nicht auswertbar bzw. nicht funktionierend betrachtet werden. Insgesamt verteilen sich die SNPs anteilig wie folgt auf die sechs Klassen: PHR 39,3%, OTV 7,6 %, NMH 5,8%, MHR 7,5%, *Other* 39,6%, CBT 0,2%. Für weitere Auswertungen sind die Klassen PHR, OTV und gegebenenfalls NMH von Interesse, die zusammen 52,7% der gesamten SNPs entsprechen. Dieser Anteil entspricht in etwa den Erwartungen für SNPs, die zuvor nicht experimentell validiert und rein auf Basis von *in silico* Analysen ausgewählt wurden (unterseer et al. 2014). Von den 10.250 SNPs des Illumina Arrays sind 8.922 (87,0%) in den Klassen PHR, OTV, NMH oder MHR enthalten und damit experimentell auch auf einer zweiten Genotypisierungsplattform validiert. Auch dieser Wert entspricht den Erfahrungen aus anderen Species zur Übertragbarkeit zwischen Genotypisierungsplattformen. Insbesondere die Gruppe der ca. 236k PHR SNPs ist eine äußerst wertvolle und

vielseitige Ressource für die spätere Nutzung auf Arrays mit geringerer Markerzahl für spezifische Fragestellungen oder zur Konvertierung in KASP-Markerassays zur Untersuchung spezifischer Genomregionen, etwa für die Feinkartierung von Genen. Die Toolbox für die genom-basierte Züchtung im Roggen konnte mit der hohen Anzahl experimentell validierter Marker deutlich erweitert werden.

3) Analyse von Sequenz- und Genotypisierungsdaten

Für die Erstellung einer hochdichten genetischen Karte wurden in der RIL-Population Lo7 x Lo225 ca. 128k polymorphe SNPs identifiziert. Um eine möglichst zuverlässige genetische Karte zu erstellen erfolgte eine Filterung der SNPs auf Marker mit maximal 5 % Fehlwerten und höchstens 5 % heterozygoter Genotypen. Von den ursprünglich 131 RILs wurden zwei Linien aufgrund zu hoher Heterozygotie ausgeschlossen, sodass die Kartierung auf 129 RILs basierte. Die Kartierung erfolgte mit der Software ASMap (Taylor und Butler 2014) unter Nutzung des MSTmap Algorithmus (Wu et al. 2008). Die Zuordnung der Kopplungsgruppen zu den sieben Chromosomen 1R bis 7R erfolgte anhand bereits kartierter Marker (Martis et al. 2013). Nach stringentem Ausschluss problematischer oder ungekoppelter Marker enthält die finale Karte 87.820 Marker, die 44.371 Lo7 WGS Contigs und 3.022 Contigs aus früheren Projekten (GABI RYE-EXPRESS bzw. Sequenzressourcen des Partners KWL) repräsentieren und damit eine Verankerung dieser Contigs entlang des Roggen-genoms erlauben (Tabelle 1). Diese bisher für Roggen einmalige Referenzkarte mit sehr hoher Markerdichte zeigte eine gute Übereinstimmung mit früheren Roggenkarten. Die genetische Karte wurde von HOH-ZI als Basis für genomweite Screens nach Selektionssignalen im Rahmen der populationsgenetischen Analysen in WP2B genutzt (siehe Bericht HOH-ZI).

Tabelle 1: Kennzahlen der hochdichten genetischen Karte aus der RIL-Population Lo7 x Lo225 und der Verankerung von Genen anderer Gräserarten im Rye Genome Zipper v2

Chromosom	1R	2R	3R	4R	5R	6R	7R	Total
Genetische Karte:								
Kartenlänge [cM]	154,3	180,9	166,7	206,8	214,6	181,7	139,9	1244,9
Anzahl kartierter Marker	12.340	13.050	10.055	10.517	13.382	15.525	12.951	87.820
Mittlerer Abstand zwischen Markern [cM]	0,013	0,014	0,017	0,020	0,016	0,012	0,011	0,014
Anzahl Loci in der Karte	290	280	290	324	362	302	230	2078
Mittlerer Abstand zwischen Loci [cM]	0,532	0,646	0,575	0,638	0,593	0,602	0,608	0,599
Maximaler Abstand zwischen Loci [cM]	2,03	3,76	2,87	3,89	3,92	5,28	2,43	5,28
Gesamtzahl repräsentierter Sequenzcontigs	6.570	6.786	5.156	5.900	6.740	8.722	7.519	47.393
Anzahl repräsentierter Lo7 WGS Contigs	6.125	6.255	4.773	5.531	6.262	8.322	7.103	44.371
Rye Genome Zipper v2:								
Anzahl flicDNAs aus Gerste	1.624	1.770	1.656	1.298	1.805	1.452	1.616	11.221
Anzahl Gene Brachypodium	1.438	1.842	1.710	1.101	1.822	1.205	1.330	10.448
Anzahl Gene Reis	1.284	1.629	1.583	1.023	1.386	1.092	1.174	9.171
Anzahl Gene Sorghum	1.353	1.621	1.608	977	1.495	1.028	1.160	9.242
Anzahl nicht-redundanter Sequenzcontigs	6.220	6.860	5.753	6.035	6.981	7.168	6.941	45.958

Die Karte war Grundlage für die Erstellung einer neuen und deutlich umfassenderen Version v2 des sogenannten *Rye Genome Zipper* (Martis et al. 2013), einer Integration von genetischen Kartierungsinformationen in Roggen mit Sequenzressourcen aus Reis, Brachypodium, Sorghum und Gerste anhand von Sequenzhomologien und Syntäniebeziehungen (Martis et al. 2013) über eine Zusammenarbeit mit der Arbeitsgruppe von Klaus F.X. Mayer vom Helmholtz Zentrum München (Tabelle 1). Diese weitere wertvolle Ressource für die Roggengenomforschung wird der wissenschaftlichen Gemeinschaft im Rahmen einer Publikation zugänglich gemacht (in Vorbereitung) und erlaubt über die direkte Verankerung mit genomischen Ressourcen aus verwandten Gräserarten einen gezielten Zugriff auf die entsprechenden Sequenzressourcen.

WP2: Genomweite Ansätze

2A Genomische Selektion

Das Projekt hatte zum Ziel, die genomische Leistungsvorhersage (*genomic prediction*, GP) in Roggen Elitematerial zu etablieren und verschiedene Einflussfaktoren auf die Genauigkeit dieser Vorhersage zu evaluieren, um die Vorhersagemodelle und damit die Selektion von Individuen aus Züchtungspopulationen optimieren zu können. Mit Hilfe der Leistungsvorhersage kann der Zuchtwert von Individuen vorhergesagt werden, für die bisher keine phänotypischen Daten aus Feldversuchen vorliegen. Verglichen mit stammbaumbasierten Methoden lässt die GP, die auf der markerbasierten tatsächlichen Verwandtschaft der Individuen beruht, eine höhere Genauigkeit erwarten.

Im WP2A wurden über die ursprünglich geplanten Datenmengen hinaus umfangreiche phänotypische und genotypische Daten sowie Stammbauminformationen über bis zu zehn Generationen aus sechs Selektionszyklen der Jahre 2009 bis 2014 vom Wirtschaftspartner KWL bereitgestellt. In jedem Jahr gab es Datensätze aus Leistungsprüfungen zur Erfassung der allgemeinen Kombinationsfähigkeit (*general combining ability*, GCA) im deutschen und polnischen Zuchtprogramm des Pollenelterpools (Tabelle 2). Bei den Selektionskandidaten handelte es sich um S₂ Inzuchtlinien, die als Testkreuzungen jeweils mit zwei Hybridtestern pro Jahr an 7-8 Standorten in Deutschland bzw. Polen geprüft wurden. Es wurden insgesamt 12 agronomische Merkmale erfasst, im Folgenden soll der Focus auf den wichtigsten Merkmalen Kornertrag (*grain dry matter yield*, GDY), Wuchshöhe (*plant height*, PHT) und Tausendkorngewicht (*thousand kernel weight*, TKW) liegen.

Tabelle 2: Übersicht der Daten für die genomische Leistungsvorhersage. DE: Deutschland, PL: Polen

Jahre	Stammbauminformation		Genotypisierungsdaten	
	DE	PL	DE	PL
2009	454	455	413	406
2010	454	374	416	368
2011	456	785	299	260
2012	455	806	288	264
2013	620	628	298	360
2014	449	808	449	808

In Zusammenarbeit mit Partner HOH-BI wurde zunächst eine Pipeline für die zweistufige Analyse der phänotypischen Daten über Orte und Jahre etabliert, um adjustierte Mittelwerte für die Vorhersagemodelle zu erhalten. Die Ergebnisse dieser Arbeiten wurden bereits publiziert (Bernal-Vasquez et al. 2014) und sind im Bericht von Partner HOH-BI genauer beschrieben.

Die auf gemischten Modellen (*best linear unbiased prediction*, BLUP) basierenden stammbaum- und genom-basierten Zuchtwerte wurden sowohl innerhalb als auch zwischen Selektionszyklen geschätzt und validiert. Zur Berechnung der markerbasierten Verwandtschaftsmatrix und für die genomische Leistungsvorhersage standen jeweils 10.416 SNP Marker eines 16k Array zur Verfügung, von denen 5.607 genetisch kartiert waren. Die Vorhersagegenauigkeiten (*prediction accuracies*) wurden jeweils anhand geeigneter Kreuzvalidierungsszenarien (*cross validation*, CV) geschätzt, bei denen jeweils 80 % der Individuen aus einem oder mehreren Zyklen ein Kalibrationsset (*calibration set*, CS) und 20% der Individuen das Validationsset (*validation set*, VS) bildeten (Albrecht et al. 2011). Die etablierten Auswertungsroutinen für die phänotypischen Daten und für die Leistungsvorhersage wurden im Lauf des Projekts auch beim Wirtschaftspartner KWL etabliert und werden bereits in Zuchtprogrammen genutzt.

In der züchterischen Verbesserung von Elitezuchtmaterial mit Hilfe genombasierter Selektion ergeben sich im Wesentlichen zwei Szenarien: die Vorhersage i) innerhalb von Zuchtzyklen und ii) über Zuchtzyklen hinweg. In beiden Fällen spielt die Verwandtschaft des jeweils verwendeten genetischen Materials eine wesentliche Rolle. Bei der Erfassung der GCA in frühen Zyklen hat sich in der rein phänotypischen Selektion ein optimiertes Schema etabliert, bei dem anhand zweier Tester an vielen Orten die Leistung geprüft wird (Wilde 1996). In einem zweiten Ansatz wurde daher untersucht, inwieweit die Vorhersage innerhalb von Testern und über Tester hinweg übereinstimmt, um die Allokation von Ressourcen (Tester, Orte, Linien) in Roggenzuchtprogrammen im Rahmen der genomischen Leistungsvorhersage möglicherweise neu zu optimieren.

1) Vorhersage innerhalb von Zuchtzyklen und über Zuchtzyklen hinweg

Die folgenden Ergebnisse stellen zunächst den systematischen Vergleich verschiedener Analysemodelle und Vorhersageszenarien für Datensätze der Jahre 2009-2012 (Zyklus 1 bis Zyklus 4) exemplarisch für das deutsche Zuchtprogramm dar. Die Ergebnisse für die Daten aus dem polnischen Zuchtprogramm sind grundsätzlich vergleichbar. Anhand der Daten aus dem deutschen Zuchtprogramm wurden die Vorhersagemodelle innerhalb von Zuchtzyklen und über Zuchtzyklen hinweg verglichen, um die Einflussfaktoren auf die Genauigkeit der Leistungsvorhersage zu evaluieren. Um die Ergebnisse vergleichbar zu machen und die Stichprobengröße als verzerrenden Faktor auszuschließen, wurde aus jedem Jahr eine zufällige Stichprobe von genotypisierten und phänotypisierten 260 S₂-Linien herangezogen (Summe: 1.040 Individuen). Die Testkreuzungsmittelwerte unterschieden sich signifikant ($p < 0,01$) zwischen Zyklen (Tabelle 3). Die Varianzkomponenten für Genotyp und Genotyp x Ort waren in allen Fällen hochsignifikant, wobei die Genotyp x Ort Interaktion in allen Fällen kleiner als die genotypische Varianzkomponente war. Die Heritabilitäten für die Merkmale GDY, PHT und TKW für die Jahre 2009 bis 2014 lagen im mittleren bis hohen Bereich (0,76-0,94).

Die berechneten stammbaum- bzw. markerbasierten Verwandtschaftsmatrizen für die Vorhersagemodelle zeigten keine ausgeprägte Populationssubstruktur innerhalb bzw. über Zyklen hinweg, mit Ausnahme einer größeren Familie in Zyklus 1 (Abb. 3A). Im Venn-Diagramm ist ersichtlich, dass die insgesamt 203 Vorfahrenlinien, für die Stammbaumdaten vorlagen eine Konnektivität über Zyklen hinweg ermöglichen (Abb. 3B). Für Zyklus 4 ist eine etwas geringere Überlappung mit den vorherigen Zyklen anhand gemeinsamer Vorfahren ersichtlich, was sich in einigen der im Folgenden dargestellten Ergebnisse widerspiegelt.

Bei der Vorhersage innerhalb von Zyklen (Kreuzvalidierungsszenario CV1) zeigte der Vergleich der beiden Methoden zur Leistungsvorhersage in allen Fällen die Überlegenheit der genomischen Vorhersage (GBLUP) gegenüber der stammbaumbasierten Methode (*pedigree-based*, PBLUP) (Abb. 4). Die Vorhersagegenauigkeit anhand von GBLUP lag je nach Merkmal in einem mittleren bis hohen Bereich von 0,44 bis 0,75 und zeigt dabei das hohe Potential der genomischen Leistungsvorhersage in Elitezuchtprogrammen. Im Mittel

über die vier Zyklen waren die GBLUP Vorhersagegenauigkeiten am höchsten für GDY (0,68) und etwas geringer für PHT und TKW (je 0,63).

Tabelle 3: Testkreuzungsmittelwerte mit Standardfehlern (s.e.), Heritabilität (h^2) und Genotyp x Ort Varianzkomponenten für die drei Merkmale GDY, PHT und TKW in vier Zyklen (Cycle 1 – 2009 bis Cycle 4 - 2012).

Zyklus	GDY [dt ha ⁻¹]			PHT [cm]			TKW [g]		
	Mittel ± s.e	h^2	$\sigma^2_{g \times l}$ ^a	Mittel ± s.e	h^2	$\sigma^2_{g \times l}$	Mittel ± s.e	h^2	$\sigma^2_{g \times l}$
Cycle 1	90,8 ± 0,19	0,86	0,52	130,4 ± 0,26	0,91	0,33	36,4 ± 0,11	0,90	0,19
Cycle 2	78,3 ± 0,23	0,86	0,24	126,7 ± 0,29	0,94	0,17	33,1 ± 0,11	0,80	0,64
Cycle 3	81,6 ± 0,20	0,77	0,87	109,5 ± 0,28	0,89	0,61	37,0 ± 0,11	0,76	0,60
Cycle 4	91,5 ± 0,23	0,83	0,72	124,3 ± 0,21	0,94	0,19	34,6 ± 0,10	0,87	0,43

^a Skaliert anhand der jeweiligen genotypischen Varianzkomponenten.

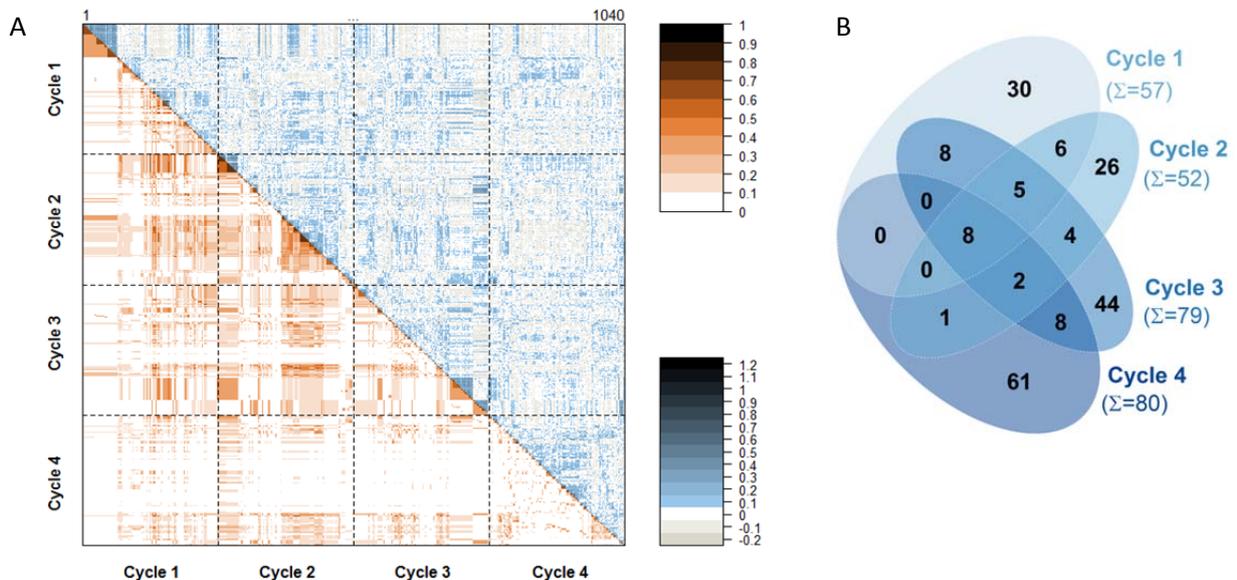


Abbildung 3: A) Verwandtschaftsverhältnisse innerhalb und zwischen Zuchtzyklen der Jahre 2009 (Cycle 1) bis 2012 (Cycle 4). Unterhalb der Diagonale ist die stammbaum- und über der Diagonale die marker-basierte Verwandtschaftsmatrix dargestellt. Farbtafeln zeigen die Werteskala für die jeweiligen Verwandtschaftsmatrizen. B) Gemeinsame Vorfahren innerhalb und zwischen Zuchtzyklen 1 bis 4. Die Summe der pro Zyklus auftretenden Vorfahren (inklusive doppelt verwendeter Linien) ist jeweils in Klammern angegeben.

Für die Vorhersage über Zyklen hinweg wurden verschiedene Szenarien untersucht. Die Vorhersage über Zyklen hat zum Ziel, die Informationen bereits geprüfter Individuen zu akkumulieren und damit den genotypischen Wert zukünftiger Selektionskandidaten vorherzusagen, was zu einer Verkürzung der Züchtungszyklen führen und damit den Selektionsgewinn erhöhen könnte. Zunächst wurden für alle paarweisen Kombinationen jeder Zyklus genutzt um einen der anderen Zyklen vorherzusagen (Szenario CV2). Im Mittel über alle sechs möglichen Vorhersagen zukünftiger Zyklen (*forward prediction*) lag die Vorhersagegenauigkeit für GDY mit GBLUP bei 0,50 im Vergleich zu 0,36 für PBLUP (Abb. 5A und 5D). Für die Merkmale PHT und TKW war die Vorhersage über Zyklen mit PBLUP nicht möglich (Abb. 5E und 5F).

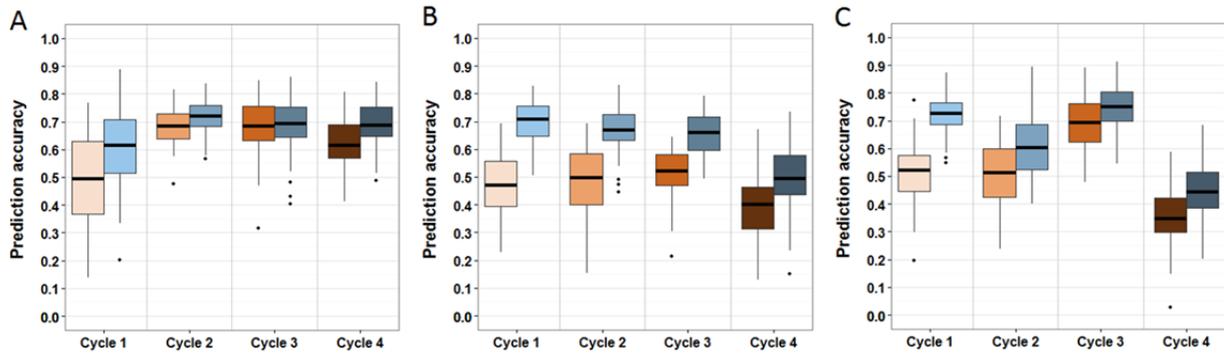


Abbildung 4: Vorhersagegenauigkeit mit Szenario CV1 innerhalb von Zuchtzyklen (*Cycle 1* bis *Cycle 4*) anhand von PBLUP (orange/braun) bzw. GBLUP (blau) für die Merkmale A) Kornertag (GDY), B) Wuchshöhe (PHT) und C) Tausendkorngewicht (TKW). Boxplots zeigen jeweils den Mittelwert (Horizontale), die obere und die untere Quartile aus 10 x 5-facher Kreuzvalidierung. Kalibrationsset: 208 Individuen, Validationsset: 52 Individuen.

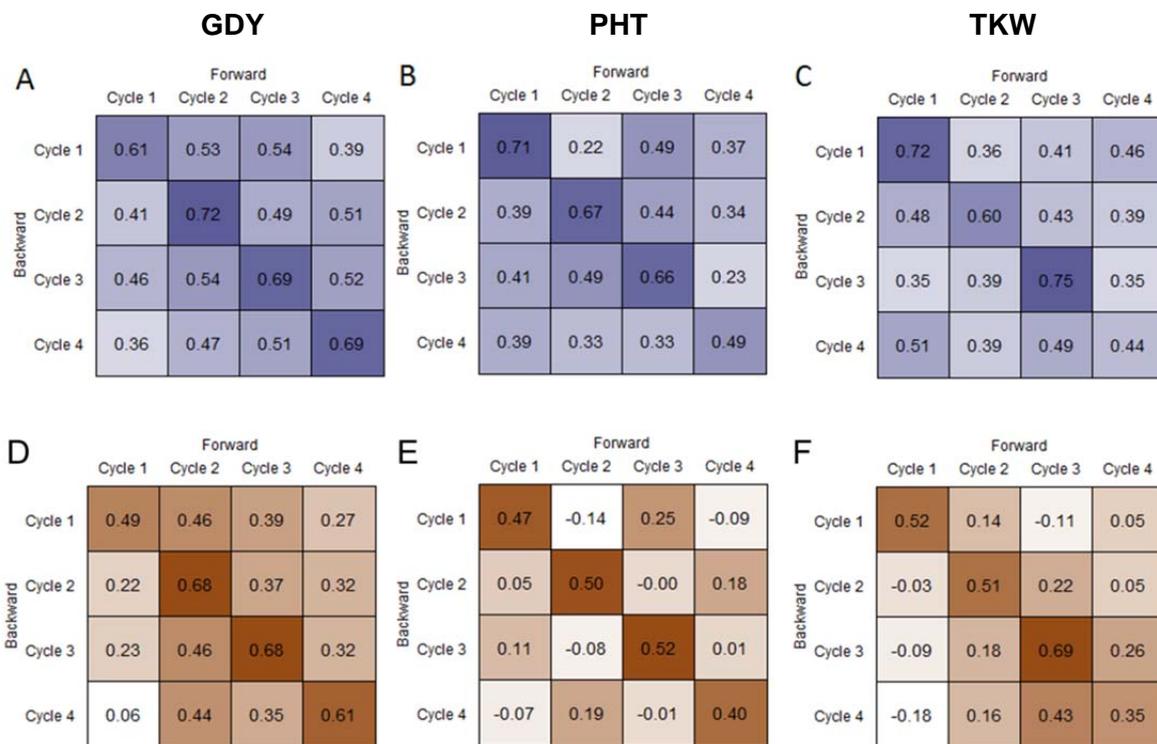


Abbildung 5: Vorhersagegenauigkeit innerhalb von Zyklen (CV1, Diagonale) und in paarweisen Vorhersagen zwischen Zyklen (CV2) für zukünftige oder zurückliegende Zyklen (oberhalb bzw. unterhalb der Diagonale). Die obere Hälfte zeigt die Werte für GBLUP (blau) an, die untere Hälfte für PBLUP (braun). A) und D): GDY, B) und E): PHT, C) und F): TKW. Die Werte beruhen auf GBLUP bzw. PBLUP mit 10 x 5-facher Kreuzvalidierung bei konstanter Größe der Kalibrations- bzw. Validationssets (N=208 bzw. N=52).

Der Einfluss der Verwandtschaftsverhältnisse zwischen Kalibrations- und Validationsset in den 12 möglichen paarweisen CV2 Szenarios auf die Vorhersagegenauigkeit wurde anhand der Korrelation zwischen den maximalen Verwandtschaftskoeffizienten U_{max} aus der markerbasierten Verwandtschaftsmatrix mit den jeweiligen GBLUP Vorhersagewerten für GDY untersucht (Abb. 6). Die Korrelation für GDY war signifikant positiv ($r = 0,67$; $p < 0,02$), während die Korrelationen für PHT und TKW nicht signifikant waren.

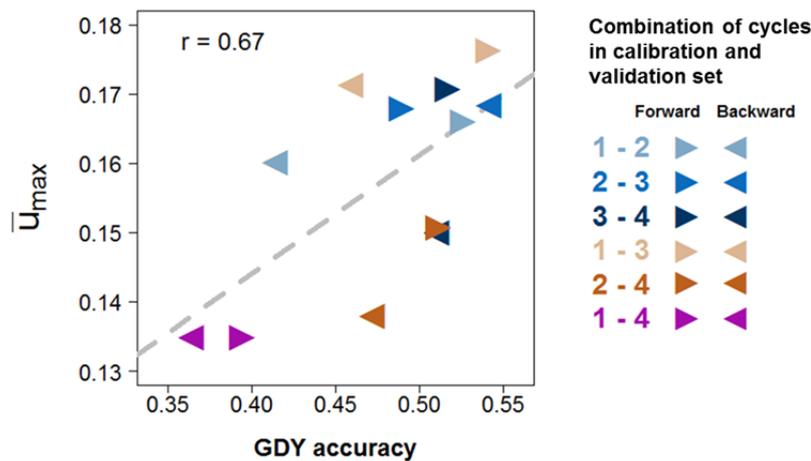


Abbildung 6: GBLUP Vorhersagegenauigkeit über Zyklen im CV2 Szenario für GDY (*GDY accuracy*) und mittlere maximale Verwandtschaft \bar{U}_{max} zwischen Kalibrations- und Validationsset. Farbige Dreiecke zeigen die verschiedenen Zyklen im Kalibrations-/Validationsset für die Vorhersage zukünftiger bzw. vergangener Zyklen. Die Werte beruhen auf GBLUP mit 10 x 5-facher Kreuzvalidierung bei konstanter Größe von Kalibrations- bzw. Validationsset (N=208 bzw. N=52).

Liegen Daten aus mehreren Zyklen vor, dann können Daten aggregiert und der Vorteil der größeren Stichprobenumfänge im Kalibrationsset genutzt werden um die Vorhersagemodelle zu verbessern. Dabei ist denkbar, dass die in einem Zyklus geprüften Individuen mit in das kombinierte Kalibrationsset eingehen, um weitere, noch ungeprüfte Individuen dieses Zyklus vorherzusagen. Diese Vorgehensweise könnte im praktischen Zuchtprogramm dazu beitragen, die Prüfumfänge im Feld zu reduzieren und trotzdem Informationen zum genotypischen Werte ungeprüfter Individuen eines gegebenen Zyklus zu erhalten. Im Szenario CV3 wurden Kalibrationssets von 208, 416, 624 und 832 Individuen verwendet um die maximale Information zu nutzen. Es wurden jeweils Individuen aus vier Zyklen ins Kalibrationsset genommen um weitere Individuen aus einem der Zyklen vorherzusagen (Abb. 7).

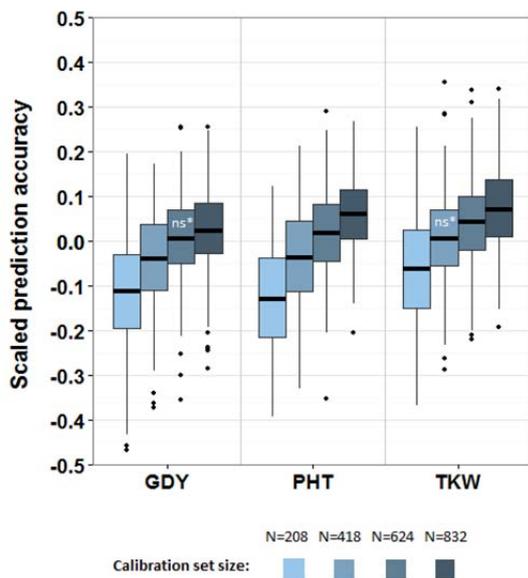


Abbildung 7: Effekt der Stichprobengröße im Kalibrationsset auf die GBLUP Vorhersagegenauigkeit für GDY, PHT und TKW. Die dargestellten Werte sind jeweils auf die merkmalspezifische Vorhersagegenauigkeit innerhalb von Zyklen (CV1) skaliert. Boxplots zeigen den Mittelwert (Horizontale), die obere und die untere Quartile aus 10 x 5-facher Kreuzvalidierung bei unterschiedlichen Größen des Kalibrationssets und einer konstanten Größe des Validationssets (N=52). Mit ns* gekennzeichnete Datensätze unterscheiden sich nicht signifikant ($p > 0.05$) von der mittleren Vorhersagegenauigkeit innerhalb von Zyklen (CV1).

Bei größeren Individuenzahlen im Kalibrationssets zeigte sich ein klar positiver Trend der mittleren GBLUP Vorhersagegenauigkeit für alle drei Merkmale. Bei der maximalen Stichprobengröße von 832 war die Vorhersage signifikant höher als die mittleren Vorhersagegenauigkeiten in CV1 oder den paarweisen CV2 Szenarios. Die Vorhersage mit PBLUP konnte kaum von der Aggregation der Daten profitieren (Daten nicht gezeigt).

Zusammenfassend zeigen die Ergebnisse, dass die Vorhersagegenauigkeit von GBLUP für alle drei Merkmale bei Aggregation der Daten aus mehreren Selektionszyklen signifikant ansteigt. Dieser Informationsgewinn kann daher auch in praktischen Zuchtprogrammen effektiv genutzt werden. Bei der Vorhersage über Jahre stieg die Vorhersagegenauigkeit vor allem bei größeren Kalibrationssets an, während der vermutete Einfluss von genetischen Faktoren wie spezifischer Testereffekte oder Umweltfaktoren gering war. Neben den Stichprobengrößen wurde als weiterer wichtiger Einflussfaktor auf die Vorhersagegenauigkeit die maximale Verwandtschaft zwischen Kalibrations- und Validationsset identifiziert. Optimale Kalibrationssets sollten daher auch über Zyklen hinweg die Verwandtschaftsverhältnisse adäquat repräsentieren. Weitere Interpretationen der Daten werden derzeit in einer umfassenden Publikation zusammengefasst, die in Kürze zur Begutachtung eingereicht werden soll.

2) Vorhersage innerhalb von Testern und über Tester hinweg

In den Daten des polnischen Zuchtprogramms aus den Jahren 2009 bis 2012 wurden jeweils 260 bis 406 S_2 Linien mit zwei verschiedenen Testern an mindestens drei gemeinsamen Orten phänotypisiert. Anhand dieser Daten wurde explizit die Vorhersage innerhalb und über Tester genauer analysiert, um Aufschlüsse über eine optimale Ressourcenallokation bezüglich der Anzahl zu prüfender Linien und zu verwendender Tester zu erhalten. Dazu wurden Tester-spezifische adjustierte Mittel der phänotypischen Werte für Kornertrag (GDY), Wuchshöhe (PHT) und Tausendkorngewicht (TKW) berechnet und die phänotypischen sowie genotypischen Korrelationen zwischen den Testkreuzungsleistungen mit zwei unterschiedlichen Testern bestimmt. Tabelle 4 zeigt für die verschiedenen Zyklen und Merkmale die Heritabilitäts-Schätzungen und die phänotypischen und genotypischen Korrelationen zwischen den Testern. Es zeigten sich im Allgemeinen hohe genotypische Korrelationen zwischen den zwei Testern für die verschiedenen Jahre, für Kornertrag lagen diese in allen vier Zyklen bei 0,8 oder höher. Lediglich für Zyklus 2 und die Merkmale PHT und TKW wurde nur eine mittlere genotypische Korrelation geschätzt (0,44 und 0,63).

Tabelle 4 Anzahl der analysierten Linien (N) im polnischen Datensatz in den vier Zyklen (Zyklus 1 – 2009 bis Zyklus 4 – 2012) welche mit zwei Testern phänotypisiert wurden; Heritabilität (h^2), phänotypische Korrelation (ρ_P) und genotypische Korrelation (ρ_G) zwischen Testern für die Merkmale GDY, PHT und TKW in den vier Zyklen.

		Zyklus 1	Zyklus 2	Zyklus 3	Zyklus 4
N		406	368	260	249
GDY	h^2	0,66	0,58	0,58	0,69
	ρ_P	0,63	0,59	0,57	0,70
	ρ_G	0,81	0,86	0,79	0,88
PHT	h^2	0,71	0,77	0,74	0,75
	ρ_P	0,77	0,37	0,76	0,62
	ρ_G	0,99	0,44	0,98	0,74
TKW	h^2	0,73	0,46	0,62	0,41
	ρ_P	0,69	0,35	0,63	0,39
	ρ_G	0,86	0,63	0,81	0,64

Des Weiteren wurde die genom-basierte Vorhersagefähigkeit innerhalb und über Tester verglichen. Dazu wurde die Vorhersagegenauigkeit von GBLUP innerhalb eines Zyklus und Testers mittels 10x5-facher Kreuzvalidierung ermittelt unter Verwendung der phänotypischen Daten welche mit einem Tester erzeugt wurden. Zur Bestimmung der Vorhersagegenauigkeit über Tester wurden die geschätzten genotypischen Werte welche mit Tester 1 (Tester 2) erzeugt wurden mit den phänotypischen Werten von Tester 2 (Tester 1) verglichen. Die

Ergebnisse aus der Vorhersage sind für die Merkmale Kornertrag und Pflanzenhöhe in Abbildung 8 dargestellt.

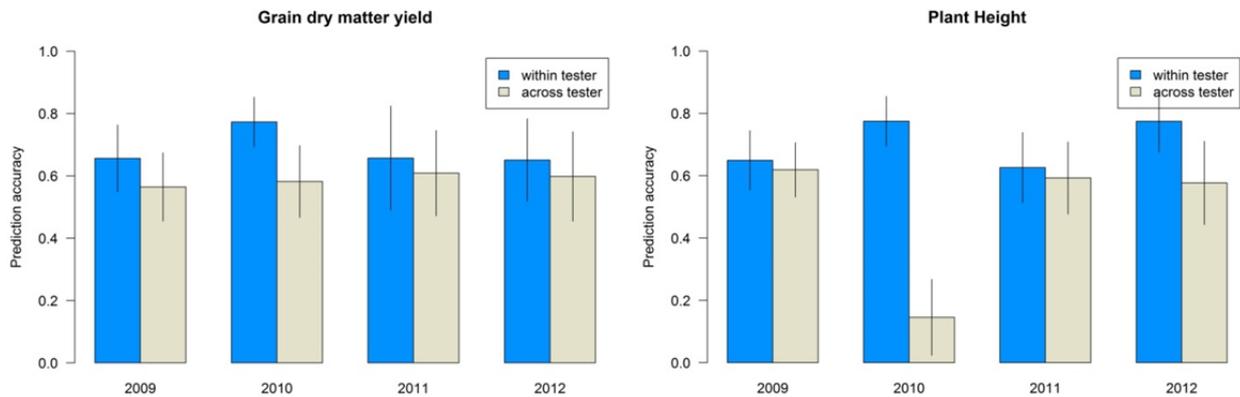


Abbildung 8: Mittlere genom-basierte Vorhersagegenauigkeiten aus 10x5-facher Kreuzvalidierung innerhalb und über Tester in den Zyklen 1-4 (2009-2012) des polnischen Zuchtprogramms für die Merkmale Kornertrag (links) und Pflanzenhöhe (rechts).

Innerhalb der Tester zeigten sich hohe Vorhersagegenauigkeiten zwischen 0,63 und 0,77 über die vier Zyklen und beide Merkmale. In den meisten Fällen ging die Genauigkeit nur gering zurück wenn über die Tester vorhergesagt wurde. Dies war durch die hohen genotypischen Korrelationsschätzungen zwischen den Testern zu erwarten. Eine Ausnahme wurde für PHT in Zyklus 2 beobachtet. Im Gegensatz zur Vorhersage innerhalb der Tester war hier die Vorhersagegenauigkeit über Tester sehr niedrig. Innerhalb dieses Zyklus zeigte sich jedoch für dieses Merkmal auch nur eine mittlere genotypische Korrelation von 0,44.

2B Populationsgenomische Analysen

1) Diversität in Elitepools und genetischen Ressourcen

Ein repräsentatives Sortiment an Inzuchtlinien aus den beiden für die Hybridzüchtung genutzten genetischen Pools (Pollen- und Saatelterpool) sowie ein diverses Sortiment an genetischen Ressourcen wurde anhand der Genotypisierungsdaten des Arrays untersucht, um die Differenzierung der Pools zu charakterisieren. Es standen 47 Pollenelterlinien, 38 Saatelterlinien und 46 Proben aus genetischen Ressourcen zur Verfügung. Für eine Hauptkoordinatenanalyse basierend auf Rogers' Distanzen (*principal coordinate analysis*, PCoA) wurden 179,147 Marker der Klasse PHR mit maximal 5 % fehlenden Werten und einer Minorallelfrequenz > 0,01 verwendet. Die PCoA zeigte eine klare Trennung der beiden Elitepools, jedoch keine ausgeprägte Populationsstruktur innerhalb der Elitepools (Abb. 9).

Die genetischen Ressourcen waren erwartungsgemäß wenig differenziert und lagen zwischen den beiden divergenten Elitepools. Ein Teil der Akzessionen genetischer Ressourcen überlappte mit dem Saatelterpool, nur eine der Akzessionen lag im Bereich des Pollenelterpools. Derartige Information zu genetischen Ressourcen kann dazu beitragen, die genetische Diversität im Elitematerial zu verbreitern, ohne dabei die klare Trennung der Elitepools für die Hybridzüchtung zu gefährden. Der neue hochdichte Genotypisierungsarray kann dazu genutzt werden über genomweite Haplotypvergleiche die Zuordnung genetischer Ressourcen zu einem der beiden Elitepools effizienter zu gestalten.

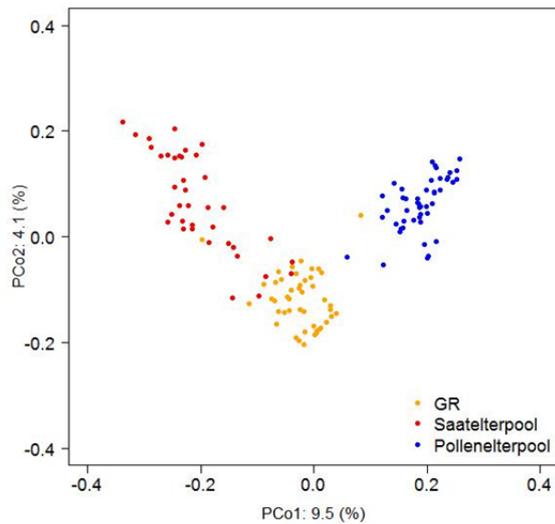


Abbildung 9: Hauptkoordinatenanalyse von Elitelinien des Pollen- und Saatelterpools (blau bzw. rot) sowie genetischer Ressourcen (GR, orange) basierend auf 179,147 Markern. PCo1 und PCo2 sind die beiden ersten Hauptkoordinaten. In Klammern ist der Anteil der durch diese Hauptkoordinaten erklärten Varianz angegeben.

2) Genomweite Assoziationsstudie

Basierend auf den Daten des in WP2A dargestellten Datensatzes von vier Selektionszyklen (2009-2012) wurde eine genomweite Assoziationsstudie (GWAS) durchgeführt. Hierbei wurden die ersten sechs Hauptkomponenten der marker-basierten Verwandtschaftsmatrix genutzt, um für mögliche Familienstruktur im Material zu korrigieren. Abbildung 10 zeigt das Ergebnis für Genomregionen mit Assoziationen zu den Merkmalen GDY, PHT und TKW. Die Analyse zeigt, dass Daten, die in Zuchtprogrammen für die genomische Selektion erhoben werden gleichzeitig verwendet werden können, um Genomregionen zu identifizieren, die agronomisch wichtige Merkmale beeinflussen.

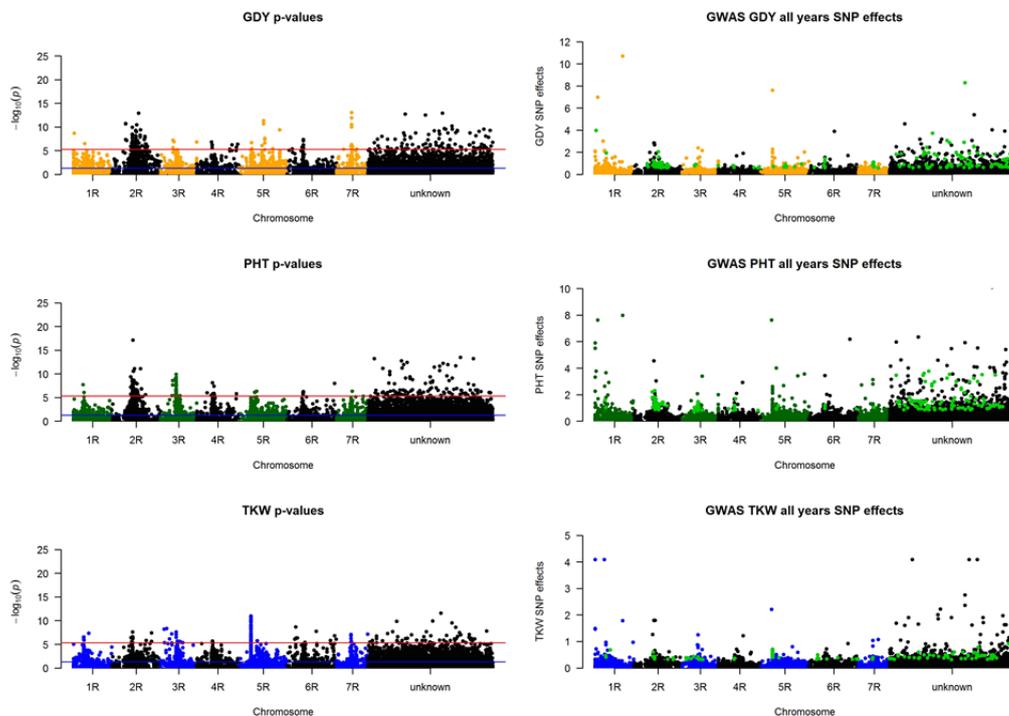


Abbildung 10: Genomweite Assoziationsstudie für vier Zyklen des deutschen Zuchtprogrammes für die Merkmale Kornertrag (GDY, orange), Wuchshöhe (PHT, dunkelgrün) und Tausendkorngewicht (TKW, blau). Links sind die $-\log_{10}(p)$ Werte dargestellt, rechts die geschätzten SNP-Effekte. Der Grenzwert für p -Wert = 0,05 ist links als blaue horizontale Linie, der Bonferroni-korrigierte p -Wert als rote Linie dargestellt. In hellgrün sind die SNPs hervorgehoben, die signifikante p -Werte (Bonferroni) aufwiesen.

Werden nur die signifikanten SNPs ($p < 0,05$, Bonferroni-korrigiert) betrachtet, dann zeigt sich eine Überlappung insbesondere für die Merkmale GDY und PHT (Abb. 11). Dies entspricht vermutlich der bekannten Korrelation der beiden Merkmale. Die vorliegenden Daten können dazu dienen zusammen mit den nunmehr vorliegenden Genomsequenzressourcen Kandidatengene zu identifizieren, die die einzelnen Merkmale beeinflussen.

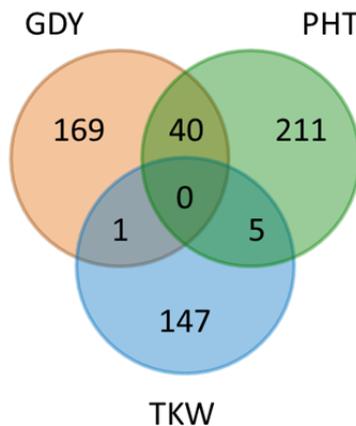


Abbildung 11: Venn-Diagramm für die Anzahl der signifikanten SNPs ($p < 0,05$, Bonferroni-korrigiert), die in der GWAS für die Merkmale Kornertrag (GDY), Wuchshöhe (PHT) und Tausendkorngewicht (TKW) für einzelne Merkmale bzw. mehrere Merkmale identifiziert wurden.

WP3: Analyse spezifischer agronomischer Merkmale

3A Frosttoleranz

Zur Identifizierung von Genomregionen, die einen Einfluss auf die Frosttoleranz von Winterroggen haben, wurde eine QTL-Analyse an einer für Frosttoleranz spaltenden Winterroggenpopulation durchgeführt. Die Kartierungspopulation wurde aus einer Kreuzung aus der Europäischen Elitelinie Lo157 und einem Gameten der hochgradig frosttoleranten kanadischen Population ‚Puma‘ erstellt.

1) Genotypisierung und genetische Kartierung

Von der Kartierungspopulation wurden 263 F_3 und 249 F_4 Linien mit 384 genomweiten SNPs des Rye-5k Array und 192 F_5 Linien mit genomweiten SNPs aus dem Rye-16k Array genotypisiert. Nach Aufbereitung der Datensätze standen in der F_4 Generation 211 Linien mit 180 polymorphen SNPs und in der F_5 Generation 192 Linien mit 2.950 polymorphen SNPs zur Verfügung. Zusätzlich wurden unter Nutzung der in WP1 generierten Sequenzressource und aus der Sequenzierung von BAC-Klonen einer bei INRA-CNRGV verfügbaren Roggen BAC-Bank Primer für verschiedene Kandidatengene (*Mybs3*, *Cbf11*, *Cbf9*, *Cbf12*, *Cbf18*, *Dhn3*, *Vrn1* und ein *Cbf*-Gen mit dem provisorischen Namen *Cbfo*) etabliert, die mögliche Komponenten der Signaltransduktionskette für Froststress darstellen. Die Primer wurden für die Genotypisierung der F_4 und F_5 Generation mit CAPS (*cleaved amplified polymorphic site*) und KASP (kompetitive Allel-spezifische PCR) Marker-Assays verwendet.

Als Basis für die QTL Kartierung wurde eine genetische Karte aus 211 F_4 Linien erstellt. Darin wurden 158 SNPs aus dem Rye-5k Array kartiert und zusätzlich fünf Kandidatengene (*Mybs3*, *VRN1*, *Cbf9*, *Cbf12* und *Cbf18*) integriert. Mit 8 (7R) bis 30 (1R) Markern pro Chromosom weist die genetische Karte der F_4 Generation eine Gesamtlänge von 1.171 cM und eine durchschnittliche Markerintervalllänge von 7,9 cM auf. Da zwischenzeitlich der Rye-16k Array verfügbar wurde, bestand die Möglichkeit, auch die F_5 Generation mit einem größeren Markerdatensatz zu genotypisieren, der sich somit auch für genomweite Leistungsvorhersagen eignet. Anhand dieser Daten wurde eine genetische Karte aus 192 F_5 Linien erstellt, die 2.346 SNPs aus dem Rye-16k Array und sieben Kandidatengene (*Mybs3*,

Dhn3, *Cbf9*, *Cbf11*, *Cbf12*, *Cbf18* und *Cbfo*) enthält. Die Anzahl der Marker pro Chromosom lag hier zwischen 150 (7R) und 486 (5R) bei einer Gesamtlänge von 1.357 cM und einer durchschnittlichen Markerintervalllänge von 1,2 cM. Auf Grund der bei der Verwendung nicht adaptierter genetischer Ressourcen in einigen Genomregionen auftretenden gestörten Spaltung, konnten jedoch nicht alle Marker und Kandidatengene aus der genetischen Karte der F₄ Generation auch in die Karte der F₅ Generation integriert werden. Dennoch stimmen beide Karten gut miteinander und mit der hochdichten Lo7xLo225 Karte (WP1) überein.

2) Phänotypisierung und Analyse phänotypischer Daten

F₄ und F₅ Linien sowie F₃ und F₄ Testkreuzungen der Kartierungspopulation wurden in den Jahren 2012, 2013 und 2014 in einer kontrollierten Umwelt, dem Frosttest, und in Feldversuchen an zwei russischen (Lipezk und Kasan) und drei kanadischen Standorten (Minto, Portage la Prairie, Saskatoon) für Frosttoleranz phänotypisiert. Die Phänotypisierung wurde vom Projektpartner KWL durchgeführt. Für den Frosttest unter kontrollierten Bedingungen wurden die Pflanzen sieben Wochen bei 2 bis 3 °C vernalisiert und anschließend in die Gefrierkammer gestellt. Im Verlauf einer Woche wurde die Temperatur auf -20 bis -23°C abgesenkt und anschließend wieder auf 5°C angehoben. Nach einer zweiwöchigen Regenerationsphase bei 8 bis 10°C wurde das Merkmal Regenerationsfähigkeit (*recovery after freezing*, REC) in Boniturnoten von 1 (anfällig) bis 9 (tolerant) erhoben. Im Feld wurden in Alpha-Gitteranlagen mit zwei bis drei Wiederholungen pro Ort die Merkmale Entwicklungsstand nach Winter (*development after winter*, DAW) in Boniturnoten von 1 (anfällig) bis 9 (tolerant) und die Überlebensrate (*survival after winter*, SAW) in Prozent erfasst.

Das im Frosttest erfasste Merkmal REC und die beiden in den Feldversuchen erhobenen Merkmale DAW und SAW wurden für Linien und Testkreuzungen getrennt analysiert. Stark abweichende Beobachtungen wurden durch Ausreißertests identifiziert und entfernt. In einem gemischten Model wurden unter Berücksichtigung von Wiederholungs- und Blockeffekten für jedes Jahr im Frosttest und für jede Jahr-Ort-Kombination im Feld („Umwelt“) adjustierte Mittelwerte für jeden Genotyp berechnet und Varianzkomponenten sowie die Wiederholbarkeit geschätzt. Datensätze aus Jahren bzw. Umwelten mit signifikanter genotypischer Varianz und einer Wiederholbarkeit von über 10 % wurden in eine kombinierte Analyse über Jahre im Frosttest bzw. über Orte und Jahre im Feld einbezogen. Das gemischte Model der kombinierten Analyse berücksichtigte neben Block- und Wiederholungseffekten noch Umwelt- und Genotyp-Umwelt-Interaktions-Effekte. Aus der kombinierten Analyse wurden adjustierte Mittelwerte für jeden Genotyp sowie Varianzkomponenten und Heritabilitäten für jedes Merkmal geschätzt. Alle phänotypischen Auswertungen wurden mit der Software R (Development Core Team, 2013) durchgeführt. Für die gemischten Modelle wurde das Paket ASReml-R (Butler et al., 2009) verwendet. Mittelwerte, Varianzkomponenten und Heritabilitäten aus der kombinierten Analyse sind in Tabelle 5 dargestellt.

Für alle drei Merkmale wurde für Linien *per se* signifikante genotypische Varianz beobachtet. Die Heritabilität der Linien *per se* war für das in der kontrollierten Umwelt erfasste Merkmal REC mit 0,79 wesentlich höher als für die Merkmale DAW (0,38) und SAW (0,48) im Feld. Die Heritabilitäten der Testkreuzungen für die Merkmale im Feld waren wie zu erwarten niedriger als die der Linien. Für alle drei Merkmale hatten die Testkreuzungen höhere Mittelwerte als die Linien *per se*. Die genetische Korrelation zwischen Linien und Testkreuzungen war im Frosttest (REC) sehr hoch mit 0,98 und hochsignifikant (Abb. 12), für die Merkmale im Feld (DAW, SAW) aber wesentlich geringer und nicht signifikant. Zwischen den beiden Phänotypisierungsplattformen (Feld, kontrollierte Umwelt) wurden für beide Merkmalspaare (REC und DAW, REC und SAW) für Linien *per se* genetische Korrelationen von 0,72 berechnet. Das spricht dafür, dass ein Test unter kontrollierten Bedingungen ausreichende Aussagekraft für die Frosttoleranz im Feld besitzt, um für ein Vorscreening von Linien *per se* genutzt zu werden.

Tabelle 5: Populationsmittelwert, Genotyp- ($\hat{\sigma}_g^2$), Genotyp-Umwelt-Interaktions- ($\hat{\sigma}_{ge}^2$) und Fehler-Varianzkomponenten ($\hat{\sigma}^2$) sowie Heritabilität (\pm Standardfehler) für die drei Merkmale REC (*recovery after freezing*), DAW (*development after winter*) und SAW (*survival after winter*) aus der phänotypischen Analyse über Umwelten für Linien *per se* und Testkreuzungen.

Merkmal	Linien <i>per se</i>			Testkreuzungen		
	REC (Boniturnote)	DAW (Boniturnote)	SAW (%)	REC (Boniturnote)	DAW (Boniturnote)	SAW (%)
Individuen	197	187	183	204	192	195
Umwelten	4*	6	6	4	5	6
Mittelwert	5,60 \pm 0,11	5,81 \pm 0,05	57,43 \pm 0,99	6,13 \pm 0,07	6,87 \pm 0,03	69,92 \pm 0,48
$\hat{\sigma}_g^2$	2,00 \pm 0,24	0,20 \pm 0,05	76,77 \pm 16,78	0,89 \pm 0,10	0,03 \pm 0,01	7,26 \pm 4,42
$\hat{\sigma}_{ge}^2$	0,47 \pm 0,07	0,56 \pm 0,06	165,19 \pm 18,62	0,00 \pm 0,02	0,09 \pm 0,02	55,36 \pm 8,79
$\hat{\sigma}^2$	0,61 \pm 0,16	0,77 \pm 0,03	245,02 \pm 9,37	0,34 \pm 0,01	0,45 \pm 0,02	183,69 \pm 6,78
Heritabilität	0,79 \pm 0,02	0,38 \pm 0,07	0,48 \pm 0,06	0,86 \pm 0,02	0,26 \pm 0,08	0,16 \pm 0,08

*Vier Umwelten unter kontrollierten Bedingungen: 2012, 2013 und 2014 bei -21°C, 2013 zusätzlich bei -23°C

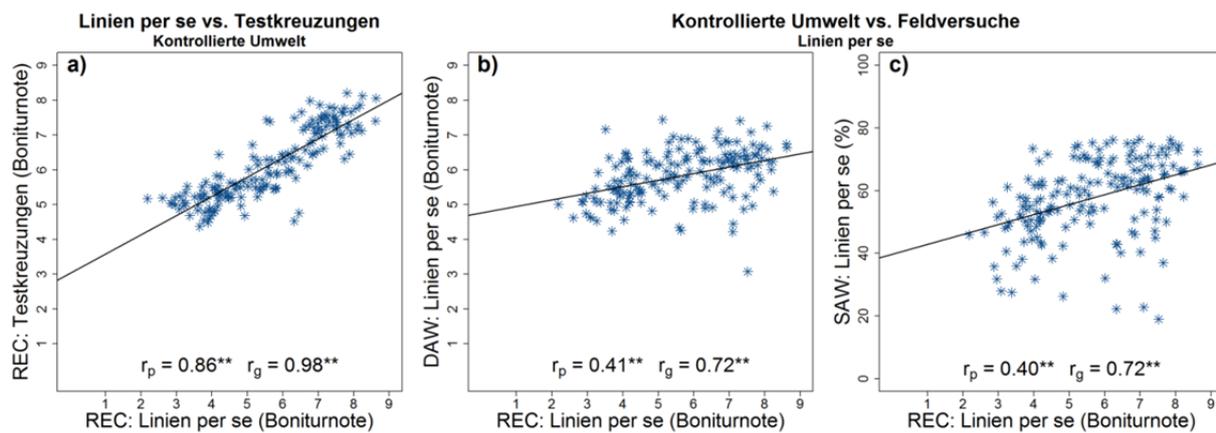


Abbildung 12: Phänotypische (r_p) und genetische (r_g) Korrelation **a)** zwischen Linien *per se* und Testkreuzungen in der kontrollierten Umwelt für das Merkmal REC (*recovery after freezing*), bzw. in Linien *per se* zwischen REC und **b)** dem im Feld erfassten Merkmal DAW (*development after winter*) sowie **c)** dem im Feld erfassten Merkmal SAW (*survival after winter*).

3) QTL Analyse für Frosttoleranz

Die QTL Analyse wurde auf Basis der genetischen Karte der F₄ Generation durchgeführt. Es wurden phänotypische Daten der F₄ und F₅ Linien *per se* und der F₃ und F₄ Testkreuzungen aller Merkmale aus Frosttest und Feldversuchen verwendet. Die Analyse wurde anhand adjustierter Mittelwerte sowohl für einzelne Umwelten/Jahre als auch einer kombinierten Analyse über Umwelten/Jahre durchgeführt. Die Ergebnisse der QTL Kartierung über Umwelten sind in Abbildung 10 dargestellt. Es wurden für die drei Merkmale insgesamt 13 QTL identifiziert. Drei QTL auf den Chromosomen 4R, 5R und 7R wurden konsistent in verschiedenen Jahren und Orten und in der kombinierten Analyse über Umwelten detektiert (Abbildung 13). Der Haupt-QTL auf Chromosom 5R wurde für alle drei Merkmale sowohl in den Datensätzen der Linien *per se* als auch der Testkreuzungen kartiert. Durch diesen QTL wurde in einzelnen Datensätzen bis zu 67,3 % der phänotypischen Varianz erklärt. Der zur QTL Position am nächsten gelegene flankierende Marker war das Kandidatengen *Cbf9*. In 1 cM Abstand davon konnte das Kandidatengen *Cbf12* kartiert werden. Aus Weizen und Gerste ist die Kolokalisation eines Clusters von *Cbf*-Transkriptionsfaktor-Genen mit dem

Frosttoleranzlocus *Fr-2* auf dem langen Arm von Chromosom 5 bereits bekannt (Knox et al., 2008; Pasquariello et al., 2014) und konnte hiermit auch in Winterroggen bestätigt werden. Das Frosttoleranzallel des großen QTL auf 5R wurde vom frosttoleranten kanadischen Elter Puma beigetragen und stellt damit einen großen Gewinn für die Züchtung europäischer Elitelinien mit verbesserter Frosttoleranz dar. Da Puma als nicht adaptierte genetische Ressource anfällig für Inzuchtdepression ist, muss bei der Nutzung des QTL auf 5R in der Marker-basierten Selektion sichergestellt werden, dass keine enggekoppelten Gene mit nachteiligem Effekt auf weitere gewünschte Merkmale übertragen werden. Das Frosttoleranzallel an zwei anderen in der kombinierten Analyse identifizierten QTL auf den Chromosomen 4R und 7R wurde von der Elitelinie Lo157 beigetragen. Die Nutzung dieser beiden QTL für die Züchtung europäischer Elitelinien unter Verwendung flankierender Marker in der Marker-basierten Selektion sollte damit problemlos gelingen.

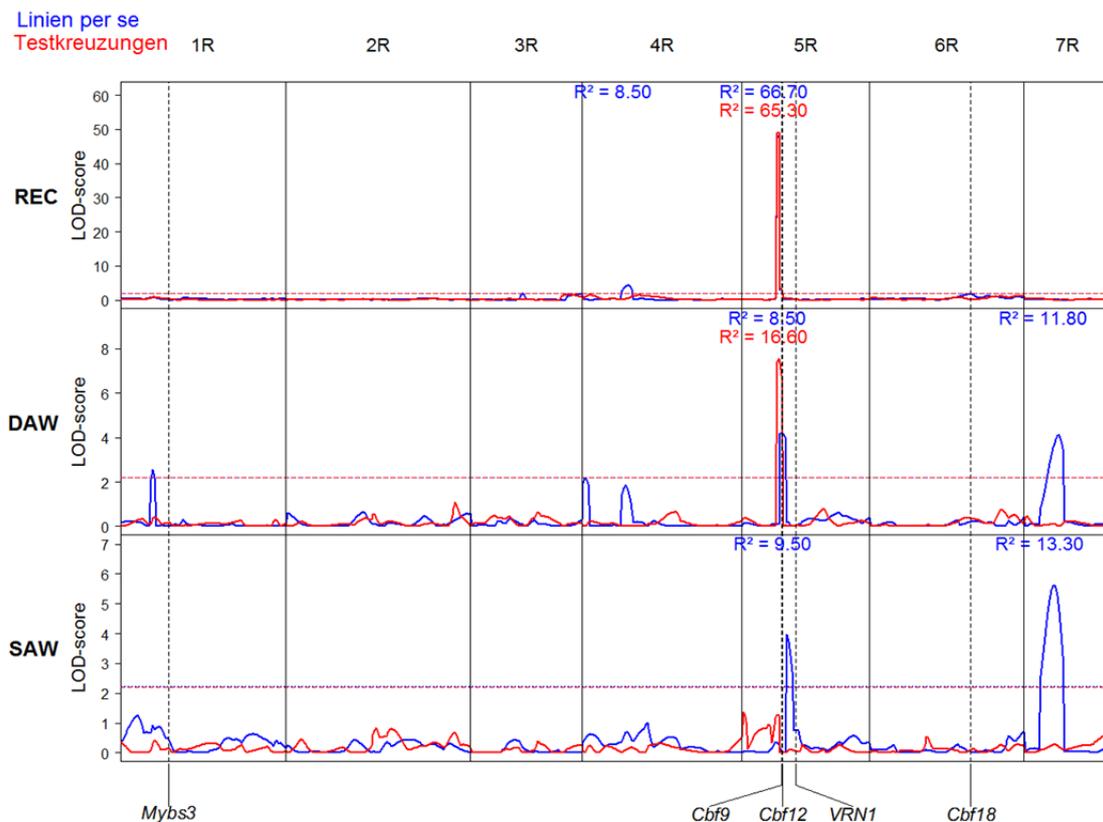


Abbildung 13: LOD-Profil der QTL Analyse über Umwelten in der F_4 Generation für die Merkmale REC (*recovery after freezing*), DAW (*development after winter*) und SAW (*survival after winter*). Der jeweilige Signifikanzgrenzwert ist als gestrichelte horizontale Linie eingezeichnet. Zu jedem QTL der den Grenzwert überschreitet, ist der Anteil der erklärten phänotypischen Varianz (R^2 %) angegeben. Die Position der in die genetische Karte integrierten Kandidatengene (*Mybs3*, *Vrn1*, *Cbf9*, *Cbf12* und *Cbf18*) ist mit vertikalen gestrichelten Linien gekennzeichnet.

4) Genomische Leistungsvorhersage für Frosttoleranz

Um Perspektiven für die Marker-basierte Selektion für das komplexe Merkmal Frosttoleranz zu untersuchen, wurde die Vorhersagegenauigkeit der QTL Analyse mit genomischen Vorhersagemodellen verglichen. Dieser Vergleich wurde unter Verwendung der dichten genetischen Karte der F_5 Generation angestellt, in der sieben Kandidatengene für Frosttoleranz integriert sind. Für die genomische Vorhersage wurde ein Modell mit Variablenselektion, LASSO (*Least absolute shrinkage and selection operator*) sowie verschiedene GBLUP Modelle verwendet, in denen ein oder mehrere mit QTL gekoppelte

Marker als fixe Effekte oder gewichtet (Zhao et al., 2014) berücksichtigt wurden. Die Vorhersagefähigkeit jedes Modells wurde in einer fünffachen Kreuzvalidierung mit zwanzig Wiederholungen ermittelt. Mit genomischer Vorhersage (LASSO, GBLUP) wurden für alle drei Merkmale sowohl für Linien *per se* als auch für Testkreuzungen immer bessere Vorhersagefähigkeiten erreicht als in der QTL Analyse (Abbildung 14). Die beste Vorhersagefähigkeiten lieferten entweder das LASSO-Modell oder ein GBLUP-Modell, in dem der flankierende Marker des Haupt-QTL auf Chromosom 5R entweder alleine oder in Kombination mit dem flankierenden Marker der QTL auf Chromosom 4R oder 7R als fixer Effekt oder gewichtet in das GBLUP Modell integriert wurde. Im Frosttest wurde die beste Vorhersagefähigkeit mit 0,88 (Linien *per se*) erreicht. Für die beiden Merkmale aus den Feldversuchen lagen die besten Vorhersagefähigkeiten zwischen 0,25 (SAW mit Linien *per se*) und 0,51 (DAW mit Testkreuzungen). Damit ist sowohl in der kontrollierten Umwelt als auch für die Merkmale im Feld eine Selektion mittels genomischer Vorhersagemodelle als Alternative zur klassischen Marker-basierten Selektion zu betrachten, da anzunehmen ist, dass weitere Genomregionen mit geringen Effekten in Summe zur Frosttoleranz beitragen, die in der QTL-Analyse nicht erfasst wurden.

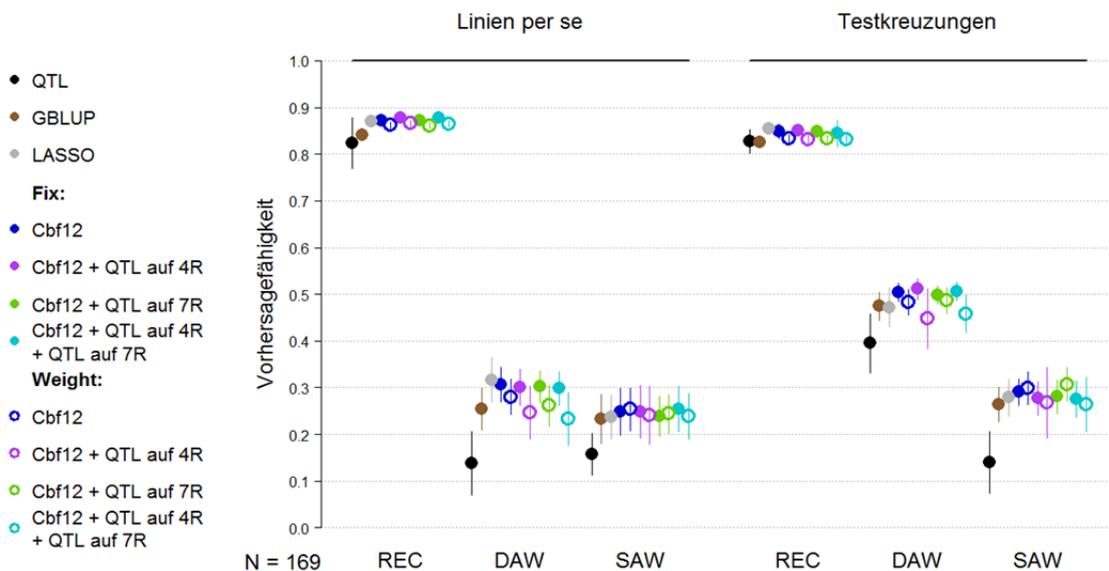


Abbildung 14: Mittlere Vorhersagefähigkeit für Linien *per se* und Testkreuzungen und die Merkmale REC (*recovery after freezing*), DAW (*development after winter*) und SAW (*survival after winter*) mit dem QTL Modell und verschiedenen genomweiten Vorhersagemodellen. Es ist die Vorhersagefähigkeit für das QTL Modell und die genomischen Vorhersage-Modelle LASSO und GBLUP ohne bzw. mit fixen oder gewichteten Effekten nach Kreuzvalidierung dargestellt.

5) Validierung von SNP-Effekten in Kandidatengenen

Der Projektpartner KWL stellte zusätzlich phänotypische Daten aus zwei Jahren von Feldstandorten und aus dem Frosttest in der kontrollierten Umwelt für die Validierung von SNP Effekten in drei Kandidatengenen zur Verfügung, die in einer kandidatengenbasierten Assoziationsstudie signifikant mit Frosttoleranz assoziiert waren (GABI RYE-FROST Projekt, Li et al. 2011b). Die Daten wurden an Paaren nah-isogener Linien (NILs) erhoben, die aus S₂ Linien mittel- und osteuropäischer Winterroggenpopulationen entwickelt wurden. Die NIL-Paare spalten an den zu validierenden SNPs auf, weisen aber einen homogenen genetischen Hintergrund auf, vergleichbar mit einer *Bulked Segregant Analysis*. Für jeweils einen SNP in den Kandidatengenen *Cbf12* und *Cbf14* konnte ein signifikanter Effekt auf die an den Feldstandorten erfassten Merkmale DAW und SAW nachgewiesen werden. Für einen SNP im Kandidatengen *Cbf15* konnte ein signifikanter Effekt auf alle drei Merkmale (REC, DAW, SAW) in beiden Phänotypisierungsplattformen bestätigt werden (Abbildung 15).

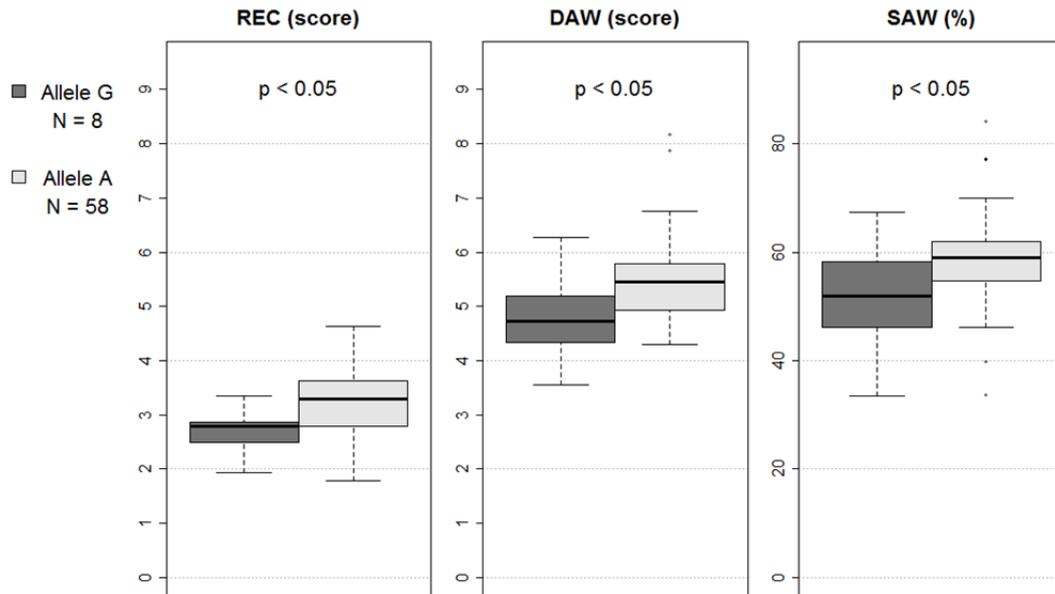


Abbildung 15: Effekt des zu validierenden SNPs in *Cbf15* auf die Merkmale REC (*recovery after freezing*), DAW (*development after winter*) und SAW (*survival after winter*) in 66 nah-isogenen Linien, die entweder das Frosttoleranz-erhöhende Allel „A“ oder das alternative Allel „G“ aufweisen.

Damit konnte der Effekt von SNPs in den drei Kandidatengen aus der früheren Assoziationsstudie auf Frosttoleranz validiert werden. Der Effekt von *Cbf14* ließ sich in der Validierung nicht über alle Umwelten und Merkmale hinweg bestätigen. Sehr schiefe Aufspaltung und damit sehr unausgeglichene Stichprobengrößen können dafür verantwortlich sein. Allerdings ist *Cbf14* vermutlich eng mit *Cbf15* gekoppelt und würde bei einer Marker-basierten Selektion auf *Cbf15* teilweise mitberücksichtigt. Eine Einbeziehung von *Cbf15* und *Cbf12* in die Marker-basierte Selektion auf Frosttoleranz ist aufgrund dieser Validierungsergebnisse sehr erfolgversprechend.

Zitierte Literatur:

- Albrecht T, Wimmer V, Auinger HJ, Erbe M, Knaak C, Ouzunova M, Simianer H, Schön C-C (2011) Genome-based prediction of testcross values in maize. *Theor Appl Genet* 123:339-350
- Albrecht T, Auinger HJ, Wimmer V, Ogutu JO, Knaak C, Ouzunova M, Piepho HP, Schön C-C (2014) Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. *Theor Appl Genet* 127:1375-1386
- Butler D, Cullis B, Gilmour A, Gogel B (2009) ASReml-R reference manual. Queensland Department of Primary Industries and Fisheries, Toowoomba.
- Development Core Team R (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
- Haseneyer G, Schmutzer T, Seidel M, Zhou R, Mascher M, Schön C-C, Taudien S, Scholz U, Stein N, Mayer K, Bauer E (2011) From RNA-seq to large-scale genotyping - genomics resources for rye (*Secale cereale* L.). *BMC Plant Biol* 11:131
- Knox AK, Li CX, Vagujfalvi A, Galilba G, Stockinger EJ, Dubcovsky J (2008) Identification of candidate CBF genes for the frost tolerance locus Fr-A(n)2 in *Triticum monococcum*. *Plant Mol Biol* 67:257-270
- Li Y, Böck A, Haseneyer G, Korzun V, Wilde P, Schön C-C, Ankerst D, Bauer E (2011b) Association analysis of frost tolerance in rye using candidate genes and phenotypic data from controlled, semi-controlled, and field phenotyping platforms. *BMC Plant Biol* 11:146

- Li Y, Haseneyer G, Schön C-C, Ankerst D, Korzun V, Wilde P, Bauer E (2011a) High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. *BMC Plant Biol* 11:6
- Martis MM, Zhou R, Haseneyer G, Schmutzer T, Vrana J, Kubalaková M, König S, Kugler KG, Scholz U, Hackauf B, Korzun V, Schön CC, Dolezel J, Bauer E, Mayer KF, Stein N (2013) Reticulate evolution of the rye genome. *Plant Cell* 25:3685-3698
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829
- Pasquariello M, Barabaschi D, Himmelbach A, Steuernagel B, Ariyadasa R, Stein N, Gandolfi F, Tenedini E, Bernardis I, Tagliafico E, Pecchioni N, Francia E (2014) The barley frost resistance-H2 locus. *Funct Integr Genomics* 14:85-100
- Taylor J, Butler D (2014) ASMap: An (A)ccurate and (S)peedy linkage map construction package for inbred populations that uses the extremely efficient MSTmap algorithm. (R package version 0.3-3)
- Unterseer S, Bauer E, Haberer G, Seidel M, Knaak C, Ouzunova M, Meitinger T, Strom TM, Fries R, Pausch H, Bertani C, Davassi A, Mayer KFX, Schön C-C (2014) A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics* 15:823
- Wilde P (1996) Multi-stage selection for combining ability among pollen parent lines in hybrid rye breeding. In: Geiger HH (ed) *International Symposium on Rye Breeding & Genetics*, Göttingen, Germany, pp 15-25
- Wimmer V, Albrecht T, Auinger H-J, Schön C-C (2012) synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics* 28:2086-2087
- Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genetics* 4:e1000212
- Zhao Y, Mette MF, Gowda M, Longin CFH, Reif JC (2014) Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity* 112:638-645

2. Wichtige Positionen des zahlenmäßigen Nachweises

Siehe Anlage „Verwendungsnachweis Mittel“.

3. Notwendigkeit und Angemessenheit der geleisteten Arbeit

Alle Arbeiten waren zur Erreichung der gesetzten Projektziele notwendig und wurden in angemessenem Umfang durchgeführt. Während der Projektlaufzeit wurde durch eine Anpassung in der Vorhabenbeschreibung bzw. im Arbeitsplan den aktuellen technischen Entwicklungen im Bereich der Hochdurchsatz-Sequenzierung bzw. -Genotypisierung Rechnung getragen. Die im Projekt geschaffenen phänotypischen und genomischen Daten stehen den beteiligten Projektpartnern zur Verfügung und stellen für die Roggenzüchtungsforschung eine wichtige Grundlage für zukünftige Forschungsarbeiten und Kooperationen dar. RYE-SELECT hat zur Etablierung der Hochdurchsatz-Sequenzierung in Roggen, der Schaffung einer umfangreichen Genomsequenzressource, der Entwicklung eines hochdichten SNP-Array und umfangreichen Genotypisierungsdaten in Roggen geführt. Sowohl die Grundlagenforschung als auch die angewandte Züchtung wird von der substanziellen Erweiterung der zur Verfügung stehenden genomischen Ressourcen für Roggen profitieren. Die Auswertung der Daten hat bereits zu einer gemeinschaftlichen Publikation geführt, weitere Publikationen sind in Vorbereitung.

4. Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse

RYE-SELECT hat einen wichtigen Beitrag zur Etablierung der genomweiten Leistungsvorhersage in Elitematerial sowie zur Analyse spezifischer genomischer Regionen geleistet. Mit SNPs, die anhand des hochdichten 600K SNP-Array experimentell validiert wurden ist kostengünstige und sehr umfangreiche Genotypisierung großer Probenzahlen im Hochdurchsatz möglich. Der populationsgenomische Ansatz führte zu gut annotierten genetischen Polymorphismen für Diversitätsstudien von Züchtungspools. Der entwickelte „molekulare Werkzeugkasten“ mit einer ersten Version einer Referenzsequenz stellt die Basis für zukünftige genombasierte Forschung im Roggen dar und erlaubt vergleichende Genomstudien im Kontext der Triticeae Weizen und Gerste sowie mit Reis, Brachypodium und Sorghum. Dies wird in Zukunft auch verstärkt funktionale Analysen in Roggen ermöglichen.

Das beteiligte Züchtungsunternehmen hat das Konzept der genomischen Zuchtwertschätzung anhand von SNP-Markern in der angewandten Züchtung implementiert. Ergebnisse des Projekts sind Voraussetzung für eine Optimierung der Ressourcenallokation in der Hybridzüchtung unter Einsatz der genomischen Selektion und tragen damit zu einer gesteigerten Selektionseffizienz bei. Züchtungsunternehmen werden die Saatgutmärkte Zentral- und Osteuropas durch Hohertragsorten vergrößern, von denen Landwirte und letztlich auch Verbraucher profitieren werden.

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

Im Berichtszeitraum fanden sich keine für das Projekt relevanten Ergebnisse von anderen Arbeitsgruppen.

6. Erfolgte und geplante Veröffentlichungen der Ergebnisse

1) Publikationen in Zeitschriften mit „peer-review“ Verfahren

Bernal-Vasquez A-M, Möhring J, Schmidt M, Schönleben M, Schön C-C, Piepho H-P (2014) The importance of phenotypic data analysis for genomic prediction - a case study comparing different spatial models in rye. *BMC Genomics* 15:646

2) Beiträge auf Konferenzen

a) Vorträge:

Bauer E, Barilar I, Beier S, Bernal A-M, Erath W, Gundlach H, Hackauf B, Haseneyer G, Korzun V, Martis M, Mayer KFX, Miedaner T, Möhring J, Piepho H-P, Schmid K, Schmidt M, Schmutzer T, Schön C-C, Schönleben M, Scholz U, Spannagl M, Wilde P (2014) New horizons for genome-based breeding in rye. XXII. Plant and Animal Genome Conference, San Diego, CA, USA. January 10-15, 2014, W435

Bauer E, Barilar I, Gundlach H, Hackauf B, Korzun V, Martis M, Mayer KFX, Schmid K, Schmutzer T, Schön C-C, Scholz U, Trost E (2015) Rye - don't be afraid of an 8 Gb genome jigsaw. EUCARPIA - International Conference on Rye Breeding and Genetics 24-26 June 2015, Wroclaw, Poland, pp 32-33 (invited talk)

Erath W, Bauer E, Fowler B, Gordillo A, Haseneyer G, Korzun V, Ponomareva M, Schmidt M, Schmiedchen B, Wilde P, Schön C-C (2014) Identification of genomic regions involved in frost tolerance in winter rye. EUCARPIA Cereals Section – ITMI Joint Conference, Wernigerode, 29.06.2014 - 04.07.2014

Erath W, Bauer E, Gordillo A, Korzun V, Schmidt M, Schmiedchen B, Wilde P, Schön C-C (2014) Identification of genomic regions involved in frost tolerance in winter rye. GPZ 2014: Genetic Variation in Plant Breeding, Kiel, 23.09.2014 - 25.09.2014

- Korzun V, Gordillo A, Schmidt M, Schmiedchen B, Wilde P, Bauer E, Erath W, Schön C-C, Schönleben M, Schmutzer T, Scholz U, Bernal A-M, Möhring J, Piepho H-P, Miedaner T, Barilar I, Schmid K, Hackauf B (2015) RYE-SELECT: Genomic tools and population genetics for rye. Statusseminar Potsdam
- Korzun V, Wilde P, Bauer E, Erath W, Haseneyer G, Schön C-C, Schönleben M, Beier S, Schmutzer T, Scholz U, Bernal A-M, Möhring J, Piepho H-P, Miedaner T, Schmid K, Hackauf B (2013) RYE-SELECT: Genome-based breeding strategies for rye. Statusseminar Potsdam
- Schönleben M, Bauer E, Bernal-Vasquez A-M, Piepho H-P, Schmidt M, Korzun V, Wilde P, Gordillo A, Schön C-C (2014) RYE-SELECT - genomic prediction in rye. Statusseminar in Potsdam
- Schönleben M, Bauer E, Bernal-Vasquez A-M, Piepho H-P, Schmidt M, Korzun V, Wilde P, Gordillo A, Schön C-C (2014) „Genomic prediction in hybrid rye breeding“, Proceedings of the joint EUCARPIA Cereal Section & ITMI Conference Wernigerode, Germany
- Schönleben M, Bauer E, Schmidt M, Korzun V, Wilde P, Gordillo A, Schön C-C (2015) Genomic prediction across breeding cycles in rye. EUCARPIA - International Conference on Rye Breeding and Genetics 24-26 June 2015, Wroclaw, Poland, p 31

b) Poster:

- Wilde P, Korzun V, Bauer E, Schön C-C, Scholz U, Piepho H-P, Schmid K, Miedaner T, Hackauf B (2012) RYE-SELECT: Genome-based precision breeding strategies for rye. Statusseminar Potsdam
- Schönleben M, Bauer E, Bernal-Vasquez A-M, Piepho H-P, Schmidt M, Korzun V, Wilde P, Gordillo A, Schön C-C (2013) RYE-SELECT: Genomic selection in hybrid rye breeding . Statusseminar Potsdam
- Erath W, Bauer E, Fowler B, Gordillo A, Korzun V, Ponomareva M, Schmidt M, Schmiedchen B, Wilde P, Schön C-C (2015) RYE-SELECT: Identification of genomic regions involved in frost tolerance in winter rye. Statusseminar Potsdam
- Schönleben M, Bauer E, Schmidt M, Korzun V, Wilde P, Gordillo A, Schön C-C (2015) RYE-SELECT: Genomic prediction across breeding cycles in rye. Statusseminar Potsdam
- Hackauf B, Bauer E, Miedaner T (2015) RYE-SELECT: Approaching the genetic diversity of restorer genes for hybrid rye breeding. Statusseminar Potsdam

3) Geplante Veröffentlichungen in Zeitschriften mit „peer-review“ Verfahren (eingereicht bzw. in Vorbereitung, Titel und Autoren vorläufig)

- Auinger H-J, Bauer E, Gordillo A, Korzun V, Lehermeier C, Piepho H-P, Schmidt M, Schön C-C, Schönleben M, Wilde P. Accuracy of genomic prediction across testers in rye.
- Auinger H-J, Bauer E, Geiger HH, Gordillo A, Korzun V, Lehermeier C, Piepho H-P, Schmidt M, Schön C-C, Schönleben M, Wilde P. Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye.
- Bauer E, Schmutzer T, Barilar I, Trost E, Martis M, Gundlach H, Hackauf B, Schmidt M, Mascher M, Korzun V, Mayer KFX, Schmid K, Schön C-C, Scholz U. Structure and diversity of the rye genome.
- Erath W, Bauer E, Fowler B, Gordillo A, Korzun V, Ponomareva M, Schmidt M, Schmiedchen B, Wilde P, Schön C-C. Identification of genomic regions involved in frost tolerance in winter rye.
- Hackauf B, Bauer E, Korzun V, Miedaner T. Fine mapping of the restorer gene *Rfp1c* in rye (*Secale cereale* L.).
- Rajaraman J, Douchkov D, Lück S, Hensel G, Rutten T, Höfle C, Hückelhoven R, Klinkenberg J, Trujillo M, Bauer E, Schmutzer T, Kumlehn J, Schweizer P. A shift from disease susceptibility to resistance in barley by partial duplication of a U-box E3 ligase gene. (eingereicht)

III ERFOLGSKONTROLLBERICHT

1 Beitrag des Ergebnisses zu förderpolitischen Zielen

Das Verbundvorhaben RYE-SELECT wurde im Rahmenprogramm „Biotechnologie – Chancen nutzen und gestalten“ der Förderinitiative zur Pflanzenbiotechnologie der Zukunft durch das Bundesministerium für Bildung und Forschung (BMBF) unterstützt. Mit dem Projekt wurde ein interdisziplinärer Forschungsverbund von sechs akademischen und einem Wirtschaftspartner etabliert. Mit RYE-SELECT ist es hervorragend gelungen moderne Erkenntnisse der Pflanzengenomforschung in die züchterische Praxis zu überführen. Mit Hilfe von innovativen, genombasierten Züchtungsstrategien trägt das Projekt zu einer Steigerung des Ertragsniveaus und der Ertragsstabilität in der ökonomisch wichtigen Nutzpflanze Roggen und damit zu einer nachhaltigen Landnutzung bei. Die geschaffenen Ressourcen und Methoden stehen den Partnern auch für künftige gemeinschaftliche Forschungsprojekte zur Verfügung und werden im Rahmen von Publikationen öffentlich verfügbar gemacht. Die Wettbewerbsfähigkeit der deutschen Saatzuchtindustrie in Regionen mit ungünstigen klimatischen Bedingungen (Osteuropa, Kanada), in denen keine Alternative zum Roggenanbau besteht, wurde durch das Projekt RYE-SELECT gestärkt. Somit leistet das Projekt einen sehr erfolgreichen Beitrag zu den förderpolitischen Zielen der BMBF Förderinitiative.

2 Wissenschaftlich-technische Ergebnisse des Vorhabens, erreichte Nebenergebnisse und gesammelte wesentliche Erfahrungen

In RYE-SELECT wurden in großem Umfang Genomsequenzressourcen von elf Roggenlinien und einer Roggenwildart geschaffen. Die Assemblierung der Sequenzen und Detektion von über 8,3 Mio. Sequenzvarianten (SNPs, Indels) ermöglichte die Entwicklung eines 600k Genotypisierungsarrays. Durch den Array konnten 52,7% der Sequenzvarianten experimentell validiert werden und sind für zukünftige Analysen nutzbar. Die Erstellung einer hochdichten genetischen Karte mit mehr als 88k Markern erlaubte die Verankerung von 44.371 Sequenzcontigs entlang des Roggengenoms. Anhand der SNP-Daten konnte durch populationsgenetische Analysen die Diversität und Struktur von Elitematerial und genetischen Ressourcen charakterisiert sowie Selektionssignale in Roggen detektiert werden, die die starke Differenzierung der Elitegenpools belegen.

Das Potential der genomischen Leistungsvorhersage anhand genomweiter Marker wurde in Roggen Zuchtmaterial ermittelt. Dabei konnten sowohl innerhalb von Selektionszyklen als auch über Zyklen hinweg mittlere bis hohe Vorhersagegenauigkeiten für agronomisch relevante Merkmale wie Ertrag, Wuchshöhe und Tausendkorngewicht erzielt werden. Die etablierten Analysepipelines und Vorhersagemodelle wurden bereits beim Wirtschaftspartner in Zuchtprogramme integriert um die Selektion verbesserten Zuchtmaterials zu beschleunigen.

Eine verbesserte Frosttoleranz des mitteleuropäischen Roggenmaterials ist erforderlich, um für Hybridroggen neue Anbauregionen mit stark kontinentalem Klima zu erschließen. Durch QTL-Kartierung wurden Genomregionen in Roggen identifiziert, die neue Allele zur Verbesserung der Frosttoleranz tragen. Mit Hilfe Marker-gestützter Selektion oder genomischer Leistungsvorhersage können Linien mit hoher Frosttoleranz identifiziert und für züchterische Zwecke genutzt werden.

3 Fortschreibung des Verwertungsplanes

Die Arbeiten im Projekt bilden die Grundlage von zwei Dissertationsschriften (Wiltrud Erath, Manfred Schönleben). Diese werden in den nächsten Monaten eingereicht werden und öffentlich zugänglich sein.

Der entwickelte „molekulare Werkzeugkasten“ mit der Genomsequenzressource und dem 600k SNP Array stellt eine wertvolle Basis für zukünftige genombasierte Forschung im Roggen und verwandter Spezies dar. Aufbauend auf den in RYE-SELECT erstellten Ressourcen, Daten und Analysemethoden können zukunftsweisende Arbeiten zur Etablierung genomweiter, Marker-basierter Selektionsstrategien in der Kulturpflanze Roggen durchgeführt werden. Der entwickelte Genotypisierungsarray ist weiterhin verfügbar und über das Projektende hinaus nutzbar. Die zur Detektion verwendeten Methoden wurden in eine Analysepipeline für Hochdurchsatz-Genotypisierungsdaten integriert, die ebenfalls dauerhaft zur Verfügung steht. Es ist geplant, die erfolgreiche, interdisziplinäre Zusammenarbeit mit den Projektpartnern weiter zu vertiefen. Mehrere Publikationen der Projektergebnisse sind in Vorbereitung. Über diese Veröffentlichungen werden die Ressourcen und Ergebnisse der wissenschaftlichen Gemeinschaft zugänglich gemacht.

4 Arbeiten, die zu keiner Lösung geführt haben

Keine.

5 Präsentationsmöglichkeiten für mögliche Nutzer

Die Ergebnisse des Projektes wurden auf mehreren nationalen und internationalen Konferenzen präsentiert. Wesentliche Ergebnisse des Projektes wurden bereits in einer internationalen Zeitschrift publiziert, weitere Publikationen sind in Vorbereitung. Die ausführliche Liste der bereits erfolgten und noch geplanten Ergebnispräsentationen bzw. Veröffentlichungen findet sich im Teil II dieses Berichts.

6 Einhaltung der Ausgaben- und Zeitplanung

Durch Verzögerungen bei der Bereitstellung der genotypischen Daten (600k SNP Array) und durch die Erhebung zusätzlicher phänotypischer Daten zur Absicherung der Ergebnisse zur Frosttoleranz ergaben sich Verspätungen im ursprünglichen Projektzeitplan. Mit der Genehmigung einer zuwendungsneutralen Verlängerung des Bewilligungszeitraums um 11 Monate konnten alle vorgesehenen Arbeitspakete erfolgreich untersucht werden. Alle Untersuchungen konnten im Rahmen des ursprünglich aufgestellten Finanzplans durchgeführt werden.

IV KURZFASSUNG

Siehe separate Dokumente „Berichtsblatt“ (auf Deutsch und Englisch).

Berichtsblatt

1. ISBN oder ISSN	2. Berichtsart (Schlussbericht oder Veröffentlichung) Schlussbericht
3. Titel Verbundprojekt RYE-SELECT „Genom-basierte Präzisionszüchtungsstrategien für Roggen“ – Teilprojekt D	
4. Autor(en) [Name(n), Vorname(n)] Bauer, Eva	5. Abschlussdatum des Vorhabens August 2015
	6. Veröffentlichungsdatum Februar 2016
	7. Form der Publikation Bericht
27	9. Ber. Nr. Durchführende Institution
	10. Förderkennzeichen 0315946D
	11. Seitenzahl 25
12. Fördernde Institution (Name, Adresse) Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. Literaturangaben 17
	14. Tabellen 5
	15. Abbildungen 15
16. Zusätzliche Angaben	
17. Vorgelegt bei (Titel, Ort, Datum)	
18. Kurzfassung In RYE-SELECT wurden in großem Umfang Genomsequenzressourcen von elf Roggenlinien und einer Roggenwildart geschaffen. Die Assemblierung der Sequenzen und Detektion von über 8,3 Mio. Sequenzvarianten (SNPs, Indels) ermöglichte die Entwicklung eines 600k Genotypisierungsarrays. Durch den Array konnten 52,7% der Sequenzvarianten experimentell validiert werden und sind für zukünftige Analysen nutzbar. Die Erstellung einer hochdichten genetischen Karte mit mehr als 88k Markern erlaubte die Verankerung von 44.371 Sequenzcontigs entlang des Roggengenoms. Anhand der SNP-Daten konnte durch populationsgenetische Analysen die Diversität und Struktur von Elitematerial und genetischen Ressourcen charakterisiert sowie Selektionssignale in Roggen detektiert werden, die die starke Differenzierung der Elitegenpools belegen. Das Potential der genomischen Leistungsvorhersage anhand genomweiter Marker wurde in Roggen Zuchtmaterial ermittelt. Dabei konnten sowohl innerhalb von Selektionszyklen als auch über Zyklen hinweg mittlere bis hohe Vorhersagegenauigkeiten für agronomisch relevante Merkmale wie Ertrag, Wuchshöhe und Tausendkorngewicht erzielt werden. Die etablierten Analysepipelines und Vorhersagemodelle wurden bereits beim Wirtschaftspartner in Zuchtprogramme integriert um die Selektion verbesserten Zuchtmaterials zu beschleunigen. Eine verbesserte Frosttoleranz des mitteleuropäischen Roggenmaterials ist erforderlich, um für Hybridroggen neue Anbauregionen mit stark kontinentalem Klima zu erschließen. Durch QTL Kartierung wurden Genomregionen in Roggen identifiziert, die neue Allele zur Verbesserung der Frosttoleranz tragen. Mit Hilfe Marker-gestützter Selektion oder genomischer Leistungsvorhersage können Linien mit hoher Frosttoleranz identifiziert und für züchterische Zwecke genutzt werden.	
19. Schlagwörter Roggen, Genomsequenzierung, SNP-Array, Diversität, Genomische Leistungsvorhersage, QTL-Kartierung, Frosttoleranz	
20. Verlag	21. Preis

Document Control Sheet

1. ISBN or ISSN	2. type of document (e.g. report, publication) Final report
3. title RYE-SELECT „Genome-based precision breeding strategies for rye“ – Part D	
4. author(s) (family name, first name(s)) Bauer, Eva	5. end of project August 2015
	6. publication date February 2016
	7. form of publication Report
8. performing organization(s) (name, address) Technische Universität München TUM School of Life Sciences Weihenstephan Department of Plant Sciences Plant Breeding Liesel-Beckmann-Str. 2 85354 Freising / Germany	9. originator's report no.
	10. reference no. 0315946D
	11. no. of pages 27
12. sponsoring agency (name, address) Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. no. of references 17
	14. no. of tables 5
	15. no. of figures 15
16. supplementary notes	
17. presented at (title, place, date)	
18. abstract In RYE-SELECT comprehensive genome sequence resources from eleven rye inbred lines and one wild species were established. Assembly of the sequences and detection of more than 8.3 million sequence variants (SNPs, Indels) enabled the development of a 600k SNP genotyping array. Experimental validation of 52.7% of the variants was possible and made these variants available for future analyses. The construction of a ultra-dense genetic map with more than 88k markers allowed anchoring of 44,371 sequence contigs along the rye genome. In population genetic analyses based on the SNP data, the diversity and structure of rye elite material and genetic resources was characterized. The diversity analyses and genomewide screens for signatures of selection revealed the strong differentiation of the elite genepools. The potential of genomic selection in rye was evaluated in elite breeding material. Genomic prediction reached intermediate to high accuracies both within and across breeding cycles for the agronomically important traits grain dry matter yield, plant height and thousand kernel weight. The established analysis pipelines and prediction models were integrated in the breeding programs of the industrial partner to speed up breeding processes. Improved frost tolerance in central European rye breeding material is necessary to extend hybrid rye growing areas to more continental climates. Through QTL mapping genomic regions were identified which carry novel favorable alleles for the improvement of frost tolerance. Aided by marker-assisted selection or genomic prediction, lines with superior frost tolerance can be identified.	
19. keywords Rye, genome sequencing, SNP array, diversity, genomic prediction, QTL mapping, frost tolerance	
20. publisher	21. price