

## **Forschungsvorhaben Teilprojekt: 0315957A**

**Thema:** Pflanzenbiotechnologie-Verbundvorhaben:

Entwicklung und Anwendung der neusten Generation genetischer Technologien bei Kulturpflanzen (NuGGET, Teilprojekt A)

## **Schlussbericht zu Nr. 3.2 BNBest-BMBF 98**

### **I.1. Aufgabenstellung**

Ziel des Pflanzenbiotechnologie-Verbundprojektes war es, durch die enge Zusammenarbeit von insgesamt 10 Partnern aus Akademie (6) und Wirtschaft (4), entscheidende Fortschritte in der Einrichtung, der Verbesserung und der Nutzung eines neuen Verfahrens für die schnelle Entdeckung von Kulturpflanzengenomen mit natürlichen oder induzierten Mutationen zu erzielen. Dazu sollte die „nächste Generation“ der Sequenzieretechnologien (NGS) eingesetzt werden. Mit der Verfügbarkeit dieser Verfahrensweise sollte der Entwicklung neuer Kulturpflanzen-Varietäten und Produkten sowie der Forschung mit und an Kulturpflanzen ein wichtiger Schub gegeben werden. Um dieses zu erreichen sollte die *SHOREmap* – Methodik, ein technologischer Durchbruch, der kurz zuvor in Modellorganismen entwickelt wurde, durch das NuGGET-Konsortium für die Anwendung bei den Kulturpflanzen Mais, Raps, Zuckerrübe und Gerste adaptiert werden. Dieses Verfahren erlaubt die simultane Entdeckung, Typisierung und Kartierung von Markern und die Untersuchung von Kandidatengenomen mittels NGS von DNA-Mischproben von Pflanzen mit mutantern oder Wildtyp-Phänotyp. Einerseits sollten die durch die hohe Komplexität der Kulturpflanzengenome gegebenen technischen und methodischen Herausforderungen durch Kulturart-spezifische Weiterentwicklungen überwunden werden. Andererseits sollten die eingerichteten und optimierten Verfahren zur schnellen molekularen Identifizierung von Kulturpflanzengenomen eingesetzt werden (z.B. Gene, die die pflanzliche Morphologie, die Produktion bestimmter Inhaltsstoffe, die phänologische Entwicklung, oder Komponenten komplexer Merkmale beeinflussen). Die im Verlaufe des Projektes entwickelten und optimierten Verfahren sollten unmittelbar den beteiligten Partnern aus der Industrie und der akademischen Forschung zugänglich gemacht werden und nach bzw. mittels Publikation der gesamten Wissenschaftsgemeinschaft, um die Grundlagen- und Anwendungs-bezogene Forschung zu unterstützen. Ebenso sollten die identifizierten Gene in weiterführende Analysen zu deren funktionellen Charakterisierung und in biotechnologische und züchterische Anwendungen eingespeist werden. Die Aufgaben in diesem Teilprojekt bezogen sich auf (i) die Gesamt-Koordination des Forschungsverbundes, der auf Grund seiner besonderen Größe und der spezifischen Aufgabenstellungen in fünf Teilgruppen untergliedert war, und (ii) auf die Optimierung und Anwendung der NGS-basierten Verfahren bei den Kulturpflanzen Mais und Gerste.

### **I.2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde**

Für die Durchführung des Teilprojekts 0315957A standen der ausführenden Stelle die notwendigen Labor- und Pflanzenkulturräume sowie die entsprechenden Groß- und Kleingeräte durch das Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) in Gatersleben zur Verfügung. Diese beinhalteten u.a. neuste Geräte und Einrichtungen für die NGS und die Pflanzenphänotypisierung.

An der durchführenden Stelle (IPK Gatersleben, Abt. Molekulare Genetik / AG Heterosis, Abt. Genbank / AG Genomdiversität, Abt. Cytogenetik und Genomanalyse / AG Bioinformatik und IT) bestanden umfangreiche Vorkenntnisse in den Bereichen molekulare Pflanzengenetik inkl. Erzeugung und Charakterisierung von Mutanten (mittels chemischer und biologischer Mutagene), der genetischen und physischen Kartierung und der Kartierungsbasierten Genisolierung, der tiefen Genotypisierung (u.a. mit Array-basierten Hochdurchsatz-Verfahren) und der Analyse komplexer und molekularer Merkmale. Ferner bestanden besonders für das Gerstengenom einzigartige Vorkenntnisse und Vorarbeiten und es lagen bereits intensive Erfahrungen mit der Erzeugung, der Handhabung und der Analyse von NGS-Daten vor.

### **I.3. Planung und Ablauf des Vorhabens**

Der Ablauf des Vorhabens entsprach weitgehend dem ursprünglichen und dem mit Zustimmung des Projektträgers geänderten (verlängerten) Arbeitsplan. Wie oben erwähnt (s.I.1), bestanden die Aufgaben in der Gesamt-Koordination des Forschungsverbundes und in der Optimierung und Anwendung der NGS-basierten Verfahren bei den Kulturpflanzen Mais und Gerste inklusive der experimentellen und bioinformatischen Aspekte. Dabei wurden insbesondere Beiträge zu den Mais-, Gerste- und Technologie-/Bioinformatik-Teilgruppen geleistet.

Im Folgenden sind die Arbeiten dargestellt, zu denen im Rahmen des Teilprojektes 0315957A beigetragen wurde.

1. Koordination des Gesamt-Verbundprojektes mit seinen 10 Partnern aus Wissenschaft und Wirtschaft. Dies beinhaltete die Ausarbeitung des Kooperationsvertrages, die Organisation der internen Projekttreffen, die Beobachtung des Arbeitsfortschrittes, sowie die Koordination der Abfassung von Projektberichten wie den Zwischenberichten der Projektpartner, des „Compact Intermediate Reports“ und des „Abschlussberichtes Pflanzenbiotechnologie für die Zukunft“.
2. Beiträge zu den Arbeiten der Mais-Teilgruppe, bestehend aus den Partnern P1 (IPK-TA), P2 (KWS) und P10 (IPK-US):
  - Erzeugung von Mischproben aus den von P2 bereit gestellten DNA-Proben, und NGS-DNA-Sequenzierung der DNA direkt und nach Anreicherung und Kartierung eines Blühzeitpunkt-QTL.
  - Identifizierung eines Blühzeitpunkts-Kandidatengens und Bestätigung seiner Identität .
  - Erzeugung einer neuen Mais-EMS-Mutantenpopulation der Inzuchtlinie PH207 und Phänotypisierung von M1-Individuen.
  - Charakterisierung der PH207 EMS-M2-Mutantenpopulation, Identifizierung neuer Mutanten und Erzeugung von Kartierungspopulationen für zwei der Mutanten .
  - Durchführung Kartierungen mit Hilfe der optimierten NGS-basierten Verfahren für die beiden neuen Mutanten und jeweils Identifizierung von Kandidatengenen, die der Verifikation und weiterführenden Untersuchungen zugeführt wurden.
  - Tiefgreifenden Sequenzanalyse der mittels NGS erhobenen DNA-Sequenzdaten wurde, deren Ergebnisse zur Identifizierung eines Blühzeitpunkt-Kandidatengens und zu jeweils einem Kandidatengener der beiden neu identifizierten Mutanten führten.
3. Beiträge zu den Arbeiten der Gersten-Teilgruppe, bestehend aus den Partnern P9 (IPK-NS) und P10 (IPK-US):
  - Erstellung rekombinanter F2-Pools und NGS-DNA-Sequenzierung unter Einsatz von „Exome-Capture“ für das *dsp1*-Gen und Identifizierung von Kandidatengenen

- Erstellung rekombinanter F2-Pools und NGS-DNA-Sequenzierung unter Einsatz von „Exome-Capture“ für das *lax-a*-Gen
  - Hoch auflösende Kartierung des *lax-a*-Gens in einer erweiterten F2-Population unter Nutzung tieferer rekombinanter F2-Pools und NGS-DNA-Sequenzierung unter Einsatz von „Exome-Capture“
  - Bestätigung der Kandidatengene für *dsp1* und *lax-a*.
4. Beiträge zu den Arbeiten der Technologie-/Bioinformatik-Teilgruppe, bestehend aus den Partnern P6 (MPI-DB), P8 (UBI) und P10 (IPK-US):
- Implementierung einer automatischen Pipeline zu Management von Sequenzrohdaten im IPK-Laborinformationsmanagementsystem (LIMS)
  - Adaptation von Pipelines für die parallelisierte Prozessierung von Hochdurchsatzsequenzdaten zur Detektion von Sequenzvariation
  - Entwicklung von R-Skripten zur Bestimmung und Visualisierung von Allelfrequenzen.

#### **I.4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde**

Zu Beginn des Projektes konnte auf umfangreiche publizierte Informationen (s.u., Literatur) und auf Kenntnisse zurückgegriffen werden, die bei der Analyse pflanzlicher Genome, insbesondere unter Einsatz der sich rasant entwickelnden NGS-Technologien und Typisierungsverfahren von DNA-Polymorphismen erzielt wurden. Mit den massiv erhöhten Durchsatzmöglichkeiten und den gleichzeitig dramatisch gesenkten Kosten, bekam die Nutzung von molekularen Markern eine stetig steigende Bedeutung in moderner Züchtungsforschung [1]. Die Marker-basierte Züchtung oder auch Präzisionszucht (MAB/MAS) ist ein effizientes Werkzeug, um komplexe Merkmale zu identifizieren, welche maßgeblich einen Anteil an Ertragsstabilität oder auch Stresstoleranz haben. Die Identifikation dieser Merkmale auf Basis des Phänotyps ist oft nur unter klar definierten Umweltbedingungen möglich. Komplexe Interaktionen zwischen der Umwelt und dem Phänotyp lassen sich durch phänotypische Selektion nur schwer voneinander trennen und verursachen dadurch einen großen Kosten- bzw. Arbeitsaufwand. Zu Beginn dieses Projektes beruhten die meisten Marker-basierten Zuchtprogramme auf gekoppelten Markern, welche durch Marker-Merkmal-Kopplung kartiert wurden. Da aber die Kopplung nicht vollständig und die Beziehung zwischen Marker Allelen und den für das Merkmal kausalen Allelen nicht gleichmäßig über den Genpool verteilt ist, ist der Nutzen von gekoppelten Markern limitiert. Im optimalen Fall würde ein Einzelnukleotid-Polymorphismus (SNP) in einem für das Merkmal kausalen Gen als Marker dienen. Die Identifikation von Sequenzvariationen innerhalb des kausalen Gens (und daraus resultierenden Markern) würde das Auftreten von falsch positiven Genotypen minimieren. Weiterführende Sequenzanalysen in größeren Genpools und die Entdeckung von neuen Mutationen (z.B. TILLING) erlauben die Identifikation neuer allelischer Variationen des kausalen Gens, welche direkt in Elite-Zuchtlinien eingekreuzt werden können. Des Weiteren eröffnet das Wissen über die Identität eines für das Merkmal kausalen Gens die Möglichkeit der gerichteten Veränderung des Genoms. Aufgrund dessen sind neue, hoch effiziente Methoden zur molekularen Identifikation von Genen (und ihrer allelischen Variation) von besonders hohem Wert für die moderne Pflanzenzucht. Der Gewinn an biologischem Wissen, wurde in den letzten zwei Jahrzehnten an der Modelnpflanze *Arabidopsis* demonstriert, ausgelöst durch die Ausnutzung von effizienteren molekularen Methoden zur Genidentifikation anhand von phänotypischen Veränderungen in Mutanten [2]. Vor Beginn des Projektes wurde durch einen Projektpartner (P6) eine neue Methode entwickelt, *SHOREmap* genannt, die die genomweite Genotypisierung mit der Kandidatengensequenzierung in einem einzigen Schritt kombiniert

indem sie eine große Anzahl von phänotypisch selektierten Rekombinanten sequenziert. Für einen Machbarkeitsnachweis wurde von P6 das Prinzip der Analyse von zusammengefassten Segregaten (BSA) auf eine F<sub>2</sub> Kartierungspopulation, bestehend aus allelischen Gruppen von EMS induzierten Arabidopsis Mutanten, angewandt [3,4]. Eine im Referenzhintergrund induzierte Mutante wurde dazu in einen polymorphen Stamm gekreuzt, welcher sich von der Referenz in ungefähr 0.5% der Nukleotid Positionen unterschied. Mehrere hundert Rekombinanten in der F<sub>2</sub> Generation wurden aufgrund des gleichen mutierten Phänotyps ausgewählt, eine DNS Probe wurde aus vereinigttem Material erstellt und diese wurde mit einer ca. zwanzigfachen Tiefe mit Hilfe der Illumina Technologie sequenziert. Die Sequenzdaten wurden dann mit Hilfe der SHORE Pipeline [5] auf das Referenzgenom abgebildet, welche vorher von P6 entwickelt wurde. Eine neue Softwareerweiterung, *SHOREmap*, erlaubte die gleichzeitige genetische Abbildung und Identifikation von kausalen Mutationen [6]. Die so gefundenen Mutationen innerhalb einer identifizierten Region wurden als Eingabe für *SHOREmap* ANNOTATE benutzt, und nicht synonyme Mutationen in dem Gen AT4G35090 mit der kürzesten Distanz zum Peak, wurden bestätigt. Dabei wurden unterschiedliche Dichten der genetischen Marker und unterschiedliche Sequenzabdeckungen untersucht. Unter Bedingungen mit hoher Markerdichte war eine sechsfache Genomabdeckung ausreichend. Sogar eine geringe Markerdichte (30 kb/Marker) war für eine Mutationsidentifikation ausreichend, solange eine zwanzigfache Genomabdeckung gewährleistet war. BSA sichert die Methode gegen falsche Bonitur von Rekombinanten ab, was eine erhebliche Fehlerquelle für konventionelle Kartierungsversuche sein kann. Es wurde gezeigt, dass der gleiche Datensatz, mit geringen Präzisionsverlusten, auch für die *de novo* Markervorhersage (via *SHOREmap* DENOVO) verwendet werden kann, wenn eine dichte Karte von SNP Markern noch nicht verfügbar ist. Eine hohe Genomabdeckung (min. elffach) war hier ausschlaggebend für eine *de novo* Markervorhersage. In der Zwischenzeit hat P6 *SHOREmap* erfolgreich für die Kartierung von über einem Duzend weiterer EMS Mutationen benutzt (damals unveröffentlicht). Wie von P6 vor Beginn des Projektes gezeigt [7], konnte *SHOREmap* Mutationen auch in Arten ohne Referenzgenom identifizieren, wenn ein weiterer Schritt eingeführt wurde, welcher für den genetischen Hintergrund korrigierte. Eine wichtige Voraussetzung für *SHOREmap* sind Art spezifische fortschrittliche Ressourcen der Genomsequenzanalyse, welche für die Zielpflanzen im NuGGET Projekt vorhanden waren (mit substanziellem Beitrag der teilnehmenden Projektpartner). Des Weiteren waren Mutanten Populationen und/oder Mutationen oder natürliche Variationen, welche wichtige Merkmale beeinflussen für die Partner des Projektes sofort verfügbar:

**Mais** gehört zu den wichtigsten Nutzpflanzen weltweit, und dominiert in Bezug auf Produktionsmengen von Nahrungsmitteln und Futtermitteln [8]. An Mais wird intensive Hybridzüchtung betrieben, mit dem Ziel den Korn und Silage Ertrag und die Ertragsstabilität zu erhöhen. Außerdem arbeiten neue Zuchtprogramme seit mehreren Jahren auf eine höhere Ausbeute von Mais als Energiepflanze hin. Ein wichtiger Faktor zur Anpassung an einheimische Bedingungen und zur Erhöhung der Biomasse ist der Blüh- und Reifezeitpunkt [9,10]. Mehrere Studien konnten QTL für den Blühzeitpunkt identifizieren [11,12]. Wichtige QTL für den Blühzeitpunkt wurden auf Chromosom 8 [13] und 10 [14] gefunden. P2 konnte einen wichtigen Blühzeitpunkt QTL auf Chromosom 10, mit additiven Effekten von vier bis fünf Tagen Differenz bis zum Blühzeitpunkt identifizieren, und produzierte damit Introgressionslinien, mit denen eine segregierende Kartierungspopulation erstellt werden konnte. Unter Berücksichtigung der hohen Relevanz der Anpassung des Blühzeitpunktes für die Züchtungsforschung von P2 und aufgrund der Verfügbarkeit von gut beschriebenem genetischem Material (Vererbung des QTL folgt Mendelschen Regeln), wurde dieses Merkmal für eine Machbarkeitsstudie ausgewählt. In diesem Projekt wurde eine neue EMS induzierte Mutanten Population erzeugt, und nach Mutationen durchsucht, welche Einfluss auf den Blühzeitpunkt und/oder das vegetative Wachstum hatten. An einer auf diese Weise

neu identifizierten Mutante wurden die optimierten Technologien angewandt, und die kausale Mutation schnell und sicher identifiziert. Die erwarteten anfallenden Kosten für so ein Experiment erlaubten es nicht weitere Mutationen mit dem Budget zu identifizieren. Gesetzt den Fall, dass aufgrund des technischen Fortschritts die Kosten bis dahin dramatisch reduziert worden wären, wäre in Betracht gezogen worden, das verbleibende Budget für die Identifikation weitere Mutationen einzusetzen. Seit Jahrzehnten ist Mais ein Objekt der molekularen/genetischen Forschung und zahlreiche Gene wurden durch positionelle Klonierung identifiziert; zum Beispiel, *CRP1* [15], *GL1* [16], *Vgt1* [17], *ZEP*, *NCED*, und *ABA* [18]. Weiter wurden viele QTL mit Einfluss auf Stickstoffverwertung, Kornertrag, Wurzelarchitektur, Pflanzenhöhe, Kornauströcknung und Korngewicht, Monolignol (BM Gene) und ABA Biosynthese gefunden [18, 19, 20, 21]. Sowohl die Verfügbarkeit des veröffentlichten 2,3 Gb Genoms der Linie B73 [22], als auch die laufenden Arbeiten an Genomen der Linien Mo17 [23], Palomero Toluqueño [24] und anderen [24a], lieferten eine gute ausgezeichnete Ausgangsbasis für die Anwendung von *SHOREmap* in Mais. Jedoch mussten die Methoden, mit Hinblick auf eine extreme Vielfalt im DNS Gehalt [25,26], angepasst werden. Aufgrund seines gut charakterisierten Genoms, und in Anbetracht seiner hohen ökonomischen Relevanz, wurde Mais ausgewählt, um als Präzedenz für Nutzpflanzen mit großen und komplexen Genomen zu dienen.

**Gerste**, die Nummer Drei der in Deutschland angebaute Getreidesorten, hat ein großes diploides Genom (5,1 Gb) mit einer hohen Komplexität (>80 % repetitiver DNS). Gerste dient als akzeptiertes diploides Modell für Brotweizen, der weltweit wichtigsten Art der Nutzpflanzen. Die hochkonservierte Syntenie zwischen den Genomen von Weizen und Gerste wird offensichtlich in einer Vielzahl von wichtigen konservierten Merkmalen (z.B. Antwort auf die Photoperiode, Antwort auf die Vernalisation, einige Schädlingsresistenz Loci). Unter der Schirmherrschaft des Internationalen Gerste Sequenzierung Konsortiums, IBSC (<http://barleygenome.org>), wurde an der Sequenzierung des Gerste Genoms gearbeitet [27]. Folgende Ressourcen wurden zu großen Teilen von dem GABI-FUTURE-BARLEX Konsortium erstellt: (i) eine BAC Contig Karte mit dreizehnfacher Sequenzabdeckung, (ii) „paired-end“ Sequenzierung von ~300.000 BAC Klonen, (iii) 3.500 sequenzierte BACs, die Gene enthalten, (iv) „high-density“ Gen basierte Markerkarten [28, 29, 30], (v) Zuordnung von ungefähr 20.000 Genen zu BACs, (vi) ein lineares Modell der Reihenfolge von über 20.000 Genen entlang aller Chromosomen in Gerste (~3.000 pro Gersten Chromosom, „genome zipper model“ [31, 32] GABI-FUTURE BARLEX, Manuskript ist eingereicht). Weiterhin wurde die Whole-Genome-Shotgun-Methode gefolgt von einer *de novo* Assemblierung an zwei fast isogenetischen Gerste Akzessionen getestet: Obwohl nur eine achtfache Sequenziertiefe erreicht wurde, jeweils auf der Illumina GAIIX Plattform, konnten 90 % aller bekannten Gene in Gerste contigs für wenigstens 60 % ihrer Länge zugeordnet werden und mehr als 90 % von einem Set aus 4.600 Genbasierten SNP Loci [30] konnten in den Sequenz contigs identifiziert werden (GABI-FUTURE-BARLEX und ERA-PG-BARCODE Konsortien, unveröffentlichte Information). Aufgrund dieser Ergebnisse wurde für das haploide Genom der Kulturvarietät Morex Whole-Genome-Shotgun-Sequenzierung mit einer fünfzigfachen Sequenziertiefe initiiert (GABI-FUTURE-BARLEX Konsortium, unveröffentlichte Daten). Dieser Datensatz lieferte im Grunde die assemblierte genomische Sequenzinformation aller Gene in Gerste. 800 Mb der assemblierten WGS contigs aus Morex (15 % des Genoms) konnten Chromosomenarmen zugeordnet werden. Die Integration mit der physikalischen Karte war in Arbeit. Dadurch wurden Informationen über physikalisch gekoppelte Gensequenzen bereitgestellt, welche die Basis für eine erfolgreiche Anwendung der *SHOREmap* Strategie in Gerste bildeten. Das Ziel dieses Sub-Projektes war der Beweis, dass der vorgeschlagene *SHOREmap* Ansatz ein erster Schritt in Richtung der Isolation von agronomisch wichtigen Merkmalen in Zentromernähe in Triticeae (Gerste, Weizen, Roggen) Chromosomen war. Diese Regionen beherbergen mehrere hundert Gene, die sich nicht durch

Rekombinationsgenetik auflösen ließen, welche obligatorisch für Positionsklonierung war. Neben dem Beweis der Durchführbarkeit von der Isolation von agronomisch wichtigen Genen in Gerste, bauten die Ergebnisse eine Brücke, um den Ansatz auf den noch wichtigeren Brotweizen (*Triticum aestivum*) zu übertragen sobald die genomischen Ressourcen in Weizen den gleichen Entwicklungsstand erreichen. Die finanzielle Begrenzung dieses Sub-Projektes erlaubte nur die Identifikation der zwei vorgeschlagenen Loci. Eine direkte Übertragung der Technologie zu Weizen war im Rahmen dieses Projektes nicht möglich. Im Falle einer erfolgreichen Isolation der zwei Gene in Gerste wird die Übertragbarkeit des gewonnenen biologischen Wissens genutzt werden können, um die allelische Variabilität in den orthologen Gene in Weizen einschätzen zu können. Da das Projekt auf der Benutzung neuer Technologien beruhte, konnte die Geschwindigkeit des technischen Fortschritts nicht abgeschätzt werden, und bei fallenden Kosten wurden mit den freigewordenen Mitteln weitere Sequenzierarbeiten durchgeführt. Gesetzt den Fall, wurden die zusätzlichen Mittel für die Isolation weiterer Ertragsbasierter Loci eingesetzt.

#### Literatur:

- [1] McCouch S (2004) Diversifying Selection in Plant Breeding. PLoS Biol 2(10): e347. doi:10.1371/journal.pbio.0020347
- [2] The Plant Journal (2010) 61, Special Issue `Arabidopsis: A rich harvest 10 years after completion of the genome sequence
- [3] Giovannoni, J. J., Wing, R. A., Ganal, M. W. & Tanksley, S. D. (1991). Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucleic Acids Res 19, 6553-6558
- [4] Michelmore, R. W., Paran, I. & Kesseli, R. V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl. Acad. Sci. USA 88, 9828-9832
- [5] Ossowski, S. et al. (2008). Sequencing of natural strains of *Arabidopsis thaliana* with short reads. Genome Res 18, 2024-2033
- [6] Schneeberger, K. et al. (2009). *SHOREmap*: simultaneous mapping and mutation identification by deep sequencing. Nat. Methods 6, 550-551
- [7] Laitinen, R. A., Schneeberger, K., Jelly, N. S., Ossowski, S. & Weigel, D. (2010) Identification of a spontaneous frame shift mutation in a nonreference *Arabidopsis* accession using whole genome sequencing. Plant Physiol. 153, 652-654.
- [8] <http://www.agrostats.com/world-statistic/world-production-statistics.html>
- [9] Richards RA. 2006. Physiological traits used in the breeding of new cultivars for water-scarce environments. Agricultural Water Management 80, 197–211.
- [10] Craufurd PQ, Wheeler TR (2009) Climate change and the flowering time of annual crops Journal of Experimental Botany, 9: 2529–2539
- [11] Chardon F, Virlon B, Moreau L, Falque M, Joets J, Decousset L, Murigneux A, Charcosset A (2004) Genetic Architecture of Flowering Time in Maize As Inferred From Quantitative Trait Loci Meta-analysis and Synteny Conservation With the Rice Genome. Genetics 168:2169-2185
- [12] Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C et al (2009) The Genetic Architecture of Maize Flowering Time. Science 325:714-718
- [13] Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashv S, Bruggemann E, Li B, Hainey CF, Radovic S, Zaina G, Rafalski J-A, Tingey SV, Miao G-H, Phillips RL, Tuberosa R (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. PNAS 107(30):11376-11381

- [14] Tian F, Stevens NM, and Buckler ES (2009) Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. *PNAS*. 106: 9979-9986
- [15] Fisk DG, Walker MB, Barkan A (1999) Molecular cloning of the maize gene *crp1* reveals similarity between regulators of mitochondrial and chloroplast gene expression *The Embo Journal* 18:2621-2630
- [16] Sturaro M, Hartings H, Schmelzer E, Velasco R, Salamini F, Mario Motto M (2005) Cloning and Characterization of *GLOSSY1*, a Maize Gene Involved in Cuticle Membrane and Wax Production. *Plant Physiology* 2005138: 478–489.
- [17] Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashv S, Bruggemann E, Li B, Hainey CF, Radovic S, Zaina G, Rafalski J-A, Tingey SV, Miao G-H, Phillips RL, Tuberosa R (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *PNAS* 107(30):11376-11381
- [18] Capelle V, Remoué C, Moreau L, Reyss A, Mahé A, Massonneau A, Falque M, Charcosset A, Thévenot C, Rogowsky P, Coursol S, Prioul J-L (2010) QTLs and candidate genes for desiccation and abscisic acid content in maize kernels *BMC Plant Biology* 10:2
- [19] Hirel B, Bertin P, Quilleré I, et al. (2001) Towards a better understanding of the genetic and physiological basis for nitrogen use efficiency in maize. *Plant Physiology* 125:1258-1270
- [20] Tuberosa R, Salvi S, Sanguineti MC, Landi P, Maccaferri M, Landi P (2003). Searching for quantitative trait loci controlling root traits in maize: a critical appraisal. *Plant and Soil*, 255 (1), 35-54.
- [21] Yi W, Ji Y, Zhengfeng Z, Yonglian Z (2006) The comparative analysis based on maize integrated QTL map and meta-analysis of plant height QTLs *Chinese Science Bulletin* 51: 18 2219-2230
- [22] Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S et al (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics *Science*:326: 1112 -1115
- [23] Rokhsar D, Chapman J, Mitros T, Goodstein D. (2009). Update on the Mo17 genome sequencing project. *Maize Genetics Conference Abstracts*, 51:T13.
- [24] Vielle-Calzada JP, Martínez de la Vega O, Hernández-Guzmán G, Ibarra-Laclette E, Alvarez-Mejía C, Vega-Arreguín JC, Jiménez-Moraila B, Fernández-Cortés A, Corona-Armenta G, Herrera-Estrella L & Herrera-Estrella A. (2009). The Palomero genome suggests metal effects on domestication. *Science* 326: 1078
- [24a] Lai, J., Li, R., Xu, X., Jin, W., Xu, M., Zhao, H., Xiang, Z., Song, W., Ying, K., Zhang, M., et al. (2010). Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat. Genet.* 42: 1027-1030.
- [25] Buckler E, Gaut B, McMullen M (2006) Molecular and functional diversity of maize. *Current Opinion in Plant Biology* 9: 172–176.
- [26] Messing J, Dooner H (2006) Organization and variability of the maize genome. *Current Opinion in Plant Biology* 9: 157–163.
- [27] Schulte D, Close TJ, Graner A, Langridge P, Matsumoto T, Muehlbauer G, Sato K, Schulman AH, Waugh R, Wise RP, Stein N (2009) The International Barley Sequencing Consortium--At the Threshold of Efficient Access to the Barley Genome. *Plant Physiol* 149:142-147
- [28] Stein N, Prasad M, Scholz U, Thiel T, Zhang H, Wolf M, Kota R, Varshney RK, Perovic D, Graner A (2007) A 1000 loci transcript map of the barley genome – new anchoring points for integrative grass genomics. *Theor Appl Genet* 114:823-839
- [29] Sato K, Nankaku N, Takeda K (2009) A high-density transcript linkage map of barley derived from a single population. *Heredity* 103:110-117
- [30] Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao S, Varshney R, Szucs P, Sato K, Hayes PM, Matthews DE, Kleinhofs A, Muehlbauer GJ, DeYoung J, Marshall DF,

Madishetty K, Fenton RD, Condamine P, Graner A, Waugh R (2009) Development and Implementation of High-Throughput SNP Genotyping in Barley. *BMC Genomics* 10:582

[31] Mayer, K.F.X., Taudien, S., Martis M., Simkova, H., Suchankova, P., Gundlach, H., Wicker, T., Petzold, A., Felder, M., Steuernagel, B., Scholz, U., Graner, A., Platzer, M., Dolozel, J., and Stein, N. (2009) Gene content and virtual gene order of barley chromosome 1H. *Plant Physiology* 151(2): 496-505. DOI:10.1104/pp.109.142612

[32] Mayer, K.F.X., Martis, M., Hedley, P., Šimková, H., Liu, H., Morris, J.A., Steuernagel, B., Taudien, S., Roessner, S., Gundlach, H., Kubaláková, M., Suchánková, P., Murat, F., Felder, M., Nussbaumer, T., Graner, A., Salse, J., Endo, T., Sakai, H., Tanaka, T., Itoh, T., Sato, K., Platzer, M., Matsumoto, T., Scholz, U., Doležel, J., Waugh, R., and Stein, N. (2011). Unlocking the barley genome by chromosomal and comparative genomics. *The Plant Cell* 23(4): 1249-63. doi: 10.1105/tpc.110.082537.

### **I.5. Zusammenarbeit mit anderen Stellen**

Das Projektvorhaben wurde in enger Zusammenarbeit mit den anderen beteiligten Partnern im Verbundprojekt NuGGET durchgeführt. In regelmäßigen Intervallen fanden Projekttreffen statt, mindestens einmal pro Jahr am Standort eines beteiligten Partners sowie jeweils am Rande des jährlichen stattfindenden PLANT2030-Statusseminars. Bei diesen Treffen wurde der Projektfortschritt besprochen, weiterführende Arbeiten geplant und Interaktionen mit anderen Partnern innerhalb und außerhalb des Pflanzenbiotechnologie-Programms diskutiert. Auf diese Weise konnten alle erforderlichen Informationen zwischen den Partnern ausgetauscht werden. Spezifische Kooperationen und zusätzliche Treffen erfolgten in Bezug auf den Austausch von Zwischenergebnissen, von vorläufigen Daten sowie der Unterzeichnung der Kooperationsverträge beteiligten Subgruppen.

### **II.1. Verwendung der Zuwendung und des erzielten Ergebnisses im Einzelnen**

Das NuGGET-Verbundprojekt, an dem insgesamt 10 Partner beteiligt waren, wurde in seiner Gesamtheit durch den Zuwendungsempfänger (P1, IPK-TA) koordiniert (s. Koordinations-AP). Die weiteren Arbeiten des Zuwendungsempfängers und die erzielten Ergebnisse beziehen sich auf seine geleisteten Beiträge zu den Arbeiten der Mais-, Gerste- und Technologie-/Bioinformatik-Teilgruppen (s. Mais-APs, Gerste-APs und Technologie-/Bioinformatik-APs).

#### **Zu Koordinations-AP:**

Die Projektkoordination des aus 10 Partnern aus Wissenschaft und Industrie bestehenden Konsortiums wurde während der gesamten Projektlaufzeit durchgeführt. Neben vielen intensiven Telefonkonferenzen wurden während dieser Zeit an verschiedenen Standorten wie in Tübingen (Januar 2012), in Gatersleben (Dezember 2012) und in Kiel (Oktober 2013) jeweils Treffen des Gesamtverbundes organisiert. Zudem fanden vielfältige (auch informelle) Treffen der jeweiligen Teilgruppenpartner statt, die der Abstimmung der Arbeiten in den Teilgruppen, dem Datenaustausch und der Koordination der Datenanalysen dienten. Nach dem fristgerechten Auslaufen einiger der vorher insgesamt aus 10 Partnern bestehenden Teilprojekte zum 31.07. 2014 wurde die Koordination der noch verbliebenen bzw. zuwendungsneutral verlängerten Teilprojekte bis zum endgültigen Auslaufen des Verbundprojekts (am 31.07.2015) fortgeführt. Nach den nun abschließend ausgeführten Analysen aller erhobenen Daten ist ein weiteres Treffen aller Partner zum Erfahrungs- und Ergebnisaustausch geplant.

Der Zuwendungsempfänger (IPK Gatersleben) war zudem federführend bei den Verhandlungen zum Kooperationsvertrag des Gesamtverbundes und hat die Ausarbeitung der



ggf. erforderlichen ergänzenden Verträge der jeweils beteiligten Subgruppen begleitet. Er war auch für die Koordination der rechtzeitigen Abfassung der Projektberichte verantwortlich, sowie die Koordination der jährlichen Zwischenberichte der beteiligten Projektpartner und deren Abgabe an den PTJ wie auch für das Verfassen des „Compact Intermediate Report“ sowie des „Abschlussberichtes Pflanzenbiotechnologie für die Zukunft“.

Weiterhin wurden die Auftritte und Präsentationen der einzelnen NuGGET Subgruppen bei den alljährlich stattfindenden PLANT2030 Status Seminaren koordiniert.

(Zu M-AP1 (durchgeführt durch P2, KWS): Es wurden F2:F3 Familien hergestellt und phänotypisch charakterisiert, welche aus der Kreuzung einer Introgressionslinie mit spätem Blühzeitpunkt (Elter 1) und einer Inzuchtlinie mit frühem Blühzeitpunkt (Elter 2) entstanden. Aus Individuen der F2:F3 Familien isolierte DNA-Proben wurden P1 (IPK) bereitgestellt (siehe Schlussbericht von P2, KWS).

#### **Zu M-AP2:**

DNA-Proben der jeweils 200 frühesten und 200 spätesten blühenden Individuen der F2:F3 Familien (Details M-WP1) wurde vereint und zusammen mit den zwei Eltern Linien mit Illumina NGS Technologie sequenziert. Alle vier Proben (2 Pools, 2 Eltern) wurden auf der Illumina HiSeq 2000 Plattform sequenziert. Der Ertrag der Sequenzierung waren 600.000 bis 900.000 kurze Lesepaare mit 2 x 100 Nukleotiden. Die Daten wurden in M-WP7 weiter analysiert (siehe unten). Zusätzlich wurden in ihrer Komplexität reduzierte DNS Proben, sowohl von den Pools als auch den Eltern, mit einer höheren Sequenziertiefe auf der Illumina HiSeq 2000 mit einem Ertrag von 160 bis 250 Millionen Lesepaaren mit 2 x 100 Nukleotiden sequenziert. Die Daten wurden in M-AP7 weiter analysiert (siehe unten). Mit beiden Ansätzen war es möglich den Blühzeitpunkt QTL auf der Zentralen Region auf Chromosom 10 zu identifizieren (s. M-AP7 für Details). Bei den sequenzierten DNA Pools wurde nur eine sehr geringe Rekombinationsrate in der Nähe des Blühzeitpunkt QTL gefunden. Als Konsequenz wurde die Kartierungsstrategie angepasst, sodass, anstelle von Pools, DNA Proben von individuell markierten und in ihrer Komplexität reduzierten Rekombinanten, welche vorher von P2 (KWS) mit flankierenden Markern identifiziert wurden, sequenziert wurden. Dabei wurden DNA Proben von 10 Rekombinanten und den davon abstammenden und homozygoten Pflanzen mit sequenziert. Die Pflanzen wurden von P2 phänotypisiert und bereitgestellt. DNA wurde individuell von 33 der F3:F4 und 22 der F2:F3 nach Verdau mit ApeKI mit einer Tiefe von 1,5 bis 1,9 Millionen Lesepaaren sequenziert. Die Daten wurden an P10 für Sequenzanalysen weitergereicht (s. M-AP7 für Details). Dabei zeigte sich, dass die erreichte Markerdichte in der QTL Region für eine Kartierung aller Rekombinationspunkte relativ zum Blühzeitpunkt QTL ausreichte. Dieses Ergebnis machte die tiefere Sequenzierung der Zielregion überflüssig, da kein weiterer Auflösungsgewinn zu erwarten war. Bei näherer Betrachtung des 95,5 kb großen Abschnitts des QTLs zeigte sich, dass ein einzelnes annotiertes Gen in dieser Region auf dem B73 Referenzgenom ist. Es wurde dadurch zum eindeutigen Kandidatengen für den Blühzeitpunkt QTL.

#### **Zu M-AP3:**

Das einzige annotiertes Gen im B73 Referenzgenom innerhalb der QTL Region (95,5 kb) Gen ist im Projektverlauf an anderer Stelle als wichtigster genetischer Faktor für die Anpassung an die Tageslänge identifiziert worden. Damit war die Verifizierung des mit Hilfe der NGS-Methodik identifizierten Gen gegeben und es waren für das Erreichen der Projektziele keine weiteren Analysen an TILLING Mutanten erforderlich. Eine abschließenden Analyse der Sequenzvariation der Allele der spät blühenden Introgressionslinie (Elter 1) und der früh blühenden Inzuchtlinie (Elter 2) ist in Abstimmung mit P2 (KWS) für weiterführende Arbeiten geplant.

#### **Zu M-AP4:**

Nachdem in den ersten Phasen des Projektes die EMS-Behandlung von Maispollen verbessert wurde und die Verfahrensweisen zu dessen Handhabung für die Bestäubung von Kolben optimiert wurden, konnte mit der Erstellung einer EMS mutagenisierten Mais Population begonnen werden: Zunächst wurden insgesamt 350 Pflanzen im Gewächshaus und 963 Pflanzen auf dem Feld mit EMS behandelten Pollen befruchtet. Zusammen belief sich der Ertrag auf 10.582 individuelle  $M_1$  Maiskörner. Die Keimfähigkeit und Überlebenswahrscheinlichkeit der  $M_1$  Körner war mit 33% geringer im Vergleich zu Körnern des Wild Typs PH207. Die Anwesenheit von EMS induzierten Mutationen im Genom der  $M_1$  Körner wurde exemplarisch an 20 Individuen mittels eines „genotyping-by-sequencing (GBS)“-ähnlichen Verfahrens auf der Illumina HiSeq Plattform getestet. Um das Auftreten und die Frequenz der EMS induzierten Mutationen in der Population unabhängig zu evaluieren, wurden die gleichen 20 Individuen auch mittels RNAseq analysiert. Nachdem die im Dunkeln gekeimten Sämlinge in das Licht überführt worden sind, wurde RNA isoliert und in cDNA überschrieben. Die so gewonnene cDNA wurden auf der Illumina HiSeq Plattform sequenziert. Die Daten wurden an P10 für weiterführende Sequenzanalysen übergeben (s. M-WP7 für Details). Im Frühling 2013 wurden die Phänotypisierung der 10.582  $M_1$  Körner mit Hilfe der Phänotypisierungsplattform des IPK (LemnaTec Scanalyzer System für große Pflanzen) und auf dem Feld initiiert. Anfang April 2013 wurden 1.600, vierzehn Tage alte  $M_1$  Sämlinge (2.000 im Gewächshaus vorgekeimt) auf die Phänotypisierungsanlage überführt und das Wachstum und die morphologischen Charakteristika der Pflanzen untersucht. Für die ersten vier Wochen wurden die Pflanzen unter reduzierten Temperaturen angezogen (15°C Tag/ 10°C Nacht), um für den Mai typische Bedingungen zu simulieren. In den darauffolgenden zwei Wochen wurden die Temperaturen auf 20°C am Tag und 15°C in der Nacht angehoben. Nach den sechs Anzuchtwochen auf der LemnaTec wurden die Pflanzen für vier Tage im Freien adaptiert, und dann Mitte Juni ins Feld ausgepflanzt. Wenn nötig wurden die Pflanzen gewässert, gedüngt und von Unkraut befreit. Mindestens jeden zweiten Tag wurden die Pflanzen visuell auf dem Feld bonitiert. Zweiunddreißig, bei der Bonitur auffällige Pflanzen, wurden zusätzlich bis zur Reife ( $M_2$ ) im Gewächshaus angezogen und geerntet. Die verbliebenen 8.500  $M_1$  Körner wurden Ende April 2013 für 12 Tage im Gewächshaus vorgekeimt und nach einer dreitägigen Akklimatisation im Freien konnten 7.000 Sämlinge auf das Feld ausgepflanzt werden. Darauf folgte ein starker Wetterumschwung mit sehr kalten Temperaturen und starken Regenfällen, was zum Absterben der Population führte. Nur 86 Pflanzen (1.22%) überlebten diese Wetterbedingungen mit dramatischen Entwicklungsverzögerungen. Die verbleibenden Pflanzen (1.600 von der Phänotypisierungsanlage und 86 vom Feld) wurden zwischen Juni und Ende August selbst befruchtet bzw. gekreuzt, falls sie nicht genug Pollen produzierten. Die befruchteten Kolben wurden für weitere vier Wochen eingetütet, um Fremdbefruchtung zu verhindern. Zwischen September und Oktober 2013 wurden die Kolben geerntet und für weitere vier Wochen unter trocknenden Bedingungen gelagert. Insgesamt wurden 1.771 Kolben befruchtet. Davon produzierten 167 Kolben (9.42%) keine Körner. Die verbleibenden 1.604 Kolben ( $M_2$ ) wurden von 1.153 selbst befruchteten und 451 gekreuzten Pflanzen produziert. Aufgrund der massiven witterungsbedingten Verluste auf dem Feld, wurde sofort mit der Erzeugung einer neuen  $M_1$  Population begonnen. Zwischen September 2013 und April 2014 wurden 309 PH207 Pflanzen im Gewächshaus mit EMS behandelten Pollen befruchtet. Von diesen Pflanzen konnten nochmal 9.893 individuell mutagenisierte  $M_1$  Körner, im Durchschnitt 32 Körner pro Kolben, geerntet werden. Im April 2014 wurden davon 6.595  $M_1$  Körner für zehn Tage vorgekeimt, von denen 6.000 Sämlinge Anfang Mai 2014 ins Feld ausgepflanzt wurden. Die verbleibenden 3.298  $M_1$  Körner wurden für zukünftige Anforderungen eingelagert. Wenn nötig wurden die Pflanzen gewässert, gedüngt und von

Unkraut befreit. Der größte Teil der  $M_1$  Pflanzen auf dem Feld konnte selbst bestäubt werden. Nur 75  $M_1$  Pflanzen wurde mit Fremdpollen von Nachbarpflanzen gekreuzt, da sie selbst keinen fruchtbaren Pollen produzierten. Alle befruchteten Kolben wurden vor Fremdbefruchtung geschützt. Im Oktober 2014 konnten von 5.500 Pflanzen Kolben mit  $M_2$  Körnern geerntet werden. Die verbleibenden 500 Pflanzen gingen durch Insekten oder Vogelfraß verloren. Die  $M_2$  Körner wurden geerntet, nachdem die geernteten Kolben weitere vier Wochen getrocknet wurden. Der Ertrag belief sich auf durchschnittlich 250 Körner pro Kolben. Zusammen mit den  $M_2$  Körnern aus 2013 wurden mehr als 7.000 Kolben mit durchschnittlich 160 Körnern produziert.

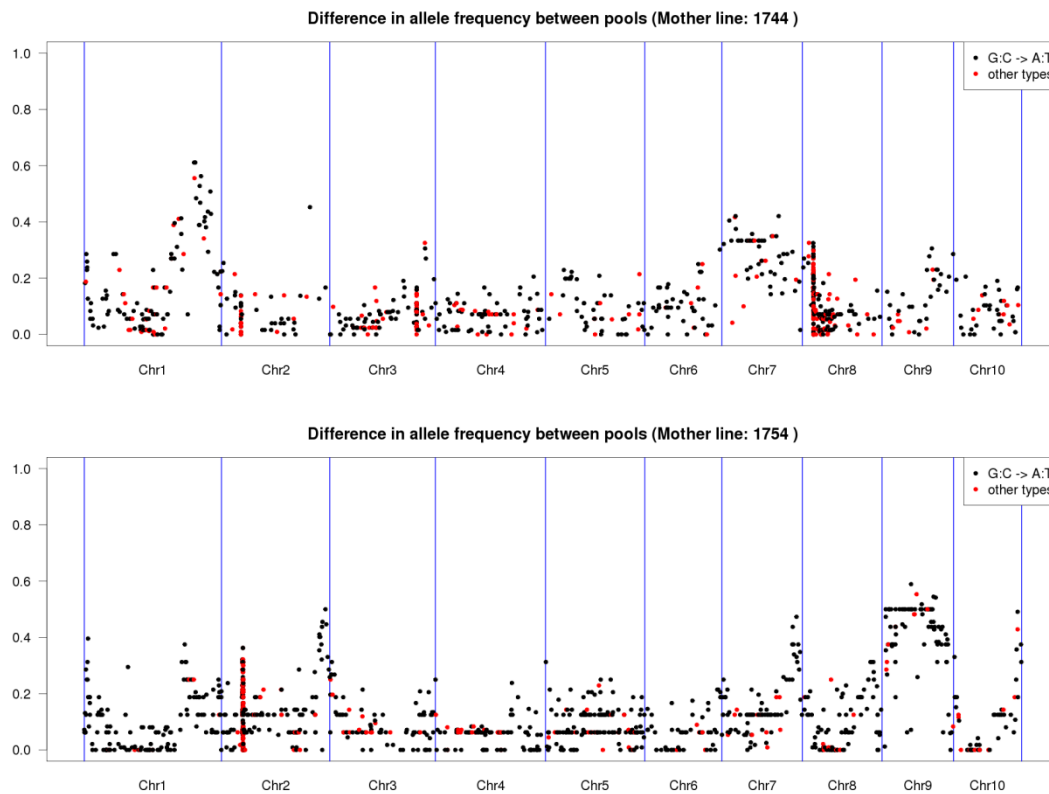
#### **Zu M-AP5:**

Um die Qualität der PH207-EMS-Population in Hinsicht der Mutationsfrequenz und Detektierbarkeit der induzierten Mutationen zu evaluieren, wurde eine zufällig ausgewählte Teilmenge von 256 Individuen der  $M_2$  Population für ein Mutanten Screening auf der IPK-Phänotypisierungsanlage angezogen. Dazu wurden 8 Körner per  $M_2$  Individuum ausgepflanzt und phänotypisch mit der automatischen Hochdurchsatz Phänotypisierungsplattform (LemnaTec) und der Hochdurchsatz Bildanalyse Software IAP in Kombination mit visueller Bonitur untersucht. Pflanzen von zweiundzwanzig der individuellen Familien (22; 8,6%) zeigten visuell eindeutig identifizierbare Unterschiede im vegetativen Wachstum wie Zwergwuchs (3 Linien), Albinismus (8 Linien), blassgrünen Habitus (4 Linien), oder auch chlorotische Flecken (5 Linien). Weitere dreiundzwanzig Linien (9%) litten unter einem starken Verlust der Keimfähigkeit (weniger als 38% Keimfähigkeit). Segregationsanalysen der  $M_2$  Linien bestätigten die Vererbung eines semi-dominanten blassgrünen Phänotyps, während alle anderen Linien Anzeichen einer typisch rezessiven Vererbung, den Mendelschen Regeln folgend, aufwiesen. Detailliertere Analysen wurden für die Linie mit dem semi-dominanten blassgrünen Phänotyp und für eine rezessive Linie mit Zwergwuchs durchgeführt. Vergleichende Bildanalyse (IAP) anhand von den  $M_3$  Nachkommen der zwergwüchsigen und blassgrünen Linien konnte eindeutig zwischen Schwesterpflanzen mit und ohne Phänotyp unterscheiden, indem Parameter wie durchschnittlicher Farbwert, Pflanzenkompaktheit, die Intensität der gelben Fluoreszenz und, als Beispiel für die Struktur des Habitus, die Fläche des kleinstmöglichen die Pflanze umschließenden Rechteckes ausgewertet wurden. Um den gemessenen Anstieg der Fluoreszenz in der Linie mit Zwergwuchs zu erklären, wurde die mit Hilfe funktioneller Chlorophyllfluoreszenz  $\Phi_{(PSII)}$  in  $M_3$  Pflanzen bestimmt, ein sensibler Indikator für photosynthetische Leistungsfähigkeit. Im Vergleich zum Wildtyp war die Chlorophyll Fluoreszenz der blassgrünen Mutanten stark reduziert und in der Zwergmutante signifikant erhöht.

#### **Zu M-AP6:**

Individuelle DNA-Proben von phänotypisch eindeutig segregierenden  $M_2$  Pflanzen mit zwergwüchsigen (8x) und blassgrünen (8x) Phänotyp zusammen mit individuellen DNA Proben der korrespondierenden Schwesterlinien (2x8) ohne mutierten Phänotyp (zusammen 32 Pflanzen) wurden per „Whole-Genome-Shotgun“ Methode auf vier Flowcells bis zu einer vierzehn bis vierundzwanzigfachen Sequenzierentiefe auf der Illumina HiSeq 2000 Plattform sequenziert (Tabelle 1). Pro individuell sequenzierter Pflanze lag der Ertrag zwischen 321 und 556 Millionen Lesepaaren. Wir nutzten die Rechenoperationen für das Mapping-by-Sequencing nach der bereits veröffentlichten Methode von P10 (Mascher *et al.*, 2014). Die Lesepaare wurden auf das B73 Referenzgenom mit BWA abgebildet und die Variationen (SNP) mit SAMtools evaluiert. In weiterführenden Analysen in der Programmiersprache R konnten ~2.000 hoch qualitative EMS induzierte Mutationen detektieren. Die Allel-Frequenzen für die Mutanten Population mit sowie für die Population ohne Phänotyp sind in Abbildung 1 dargestellt. Durch Phänotypisierung der  $M_3$  Populationen einer jeden individuell

sequenzierten M<sub>2</sub> Pflanze, konnte in den Sequenzdaten nach SNPs gefiltert werden, die genau dem bestimmten allelischen Zygotiemuster in jeder einzelnen M<sub>2</sub> folgen. Die kausalen Mutationen konnten so auf den langen Arm von Chromosom 1 (Zwergwuchs) und Chromosom 10 (blassgrün) kartiert werden. Die Intervalle hatten eine Länge zwischen 0.9 (Chr.10, 144-144,9 Mb; blassgrün) und ~5 Mb (Chr.1, 240-245 Mb; Zwerg). In beiden Fällen liegt in den Intervallen jeweils ein Kandidatengen, für dessen Mutation ein vergleichbarer Phänotyp beschrieben wurde (Tabelle 2).



**Abbildung 1:** Unterschiede in der Allele Frequenz zwischen Mutanten Pools mit und ohne Phänotyp in den segregierenden Populationen Zwerg (1744, oben) und blassgrün (1754, unten). Die schwarzen Punkte zeigen segregierende EMS typische Punktmutationen in den beiden Populationen (ca. 2000 SNPs).

Tabelle 1: Daten Alignment von individuell sequenzierten gDNA Lesepaaren für jeder der Mutanten mit (*zwerg*, *blassgrün*) und ohne (*ZWERG*, *BLASSGRÜN*) mutierten Phänotyp

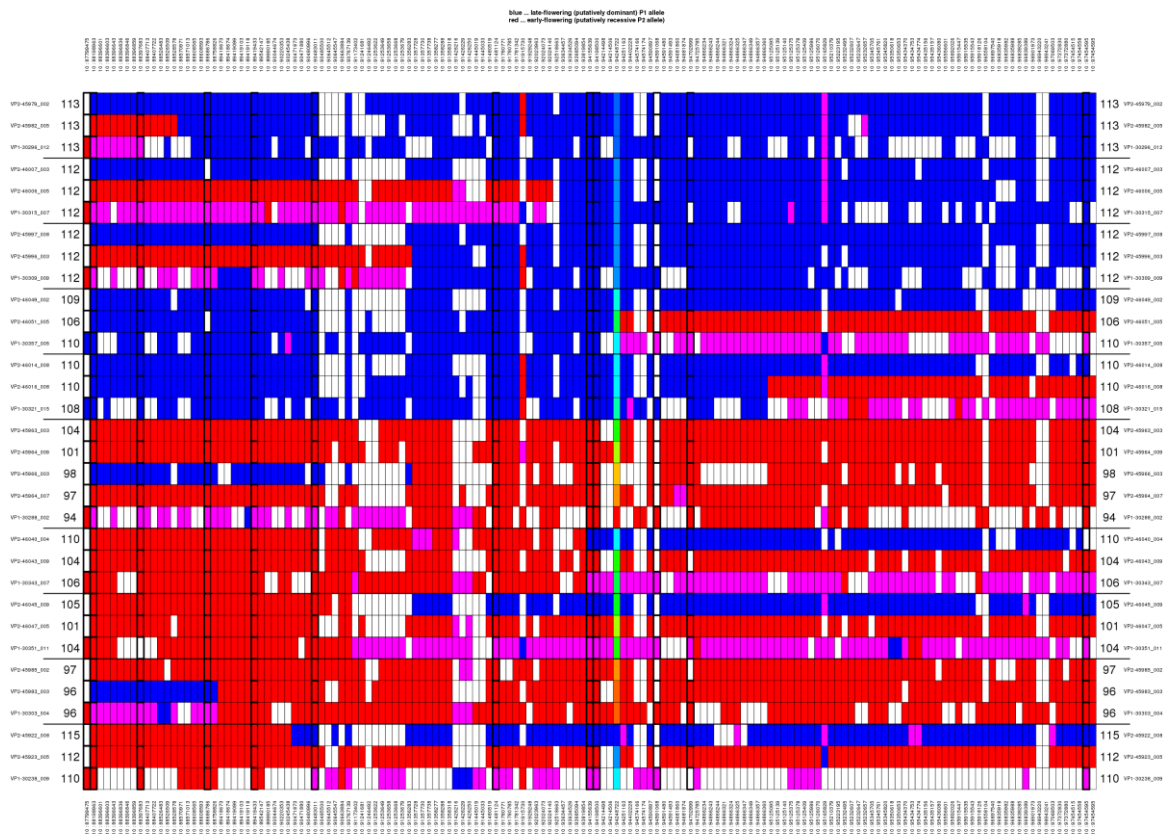
Mutanten n = 32	DNA- Stück/ Pflanze (gesamt)	DNA- Stück/ Pflanze (kartiert)	%	Lesepaare/ Pflanze (zugeordnet)	%	DNA- Stück/ Pflanze (einzigartig)	%	Geschätzte Sequenzier- tiefe
<i>zwerg</i> n=7	~ 432,9 M	~ 429,9 M	99,3	~ 428,6 M	99	~ 215,3 M	49,7	19 x
<i>ZWERG</i> n=9	~ 437,6 M	~ 434,5 M	99,3	~ 433,2 M	98,9	~ 218,5 M	49,9	19 x
<i>blassgrün</i> n=8	~ 457,1 M	~ 453,9 M	99,3	~ 452,7 M	99	~ 229,7 M	50,3	20 x
<i>BLASSGRÜN</i> n=8	~ 448,3 M	~ 445,3 M	99,3	~ 444,1 M	99,1	~ 222,6 M	49,7	19 x

Tabelle 2: Segregationsdaten der M<sub>3</sub> wurden für die Bestimmung des Genotyps der sequenzierten M<sub>2</sub> genutzt. SNPs, welche dem erwarteten Zygotiemuster in der Population entsprachen, wurden dem Langenarm von Chr. 1 (*zwerg*) und dem Langenarm von Chr. 10 (*blassgrün*) zugeordnet.

Mutanten	Segregierende SNPs	EMS-typische SNPs [%]	SNPs passend zu Zygotiemuster	Chr.	Intervall auf Chr. 1. SNP – letzter SNP	Kandidatengene in der Region
<i>zwerg</i>	1.357	56,4	3	1L	240 – 245 Mbp	<i>anther ear1</i>
<i>blassgrün</i>	3.536	50,3	4	10L	144 – 144,9 Mbp	<i>white2</i>

### Zu M-WP7:

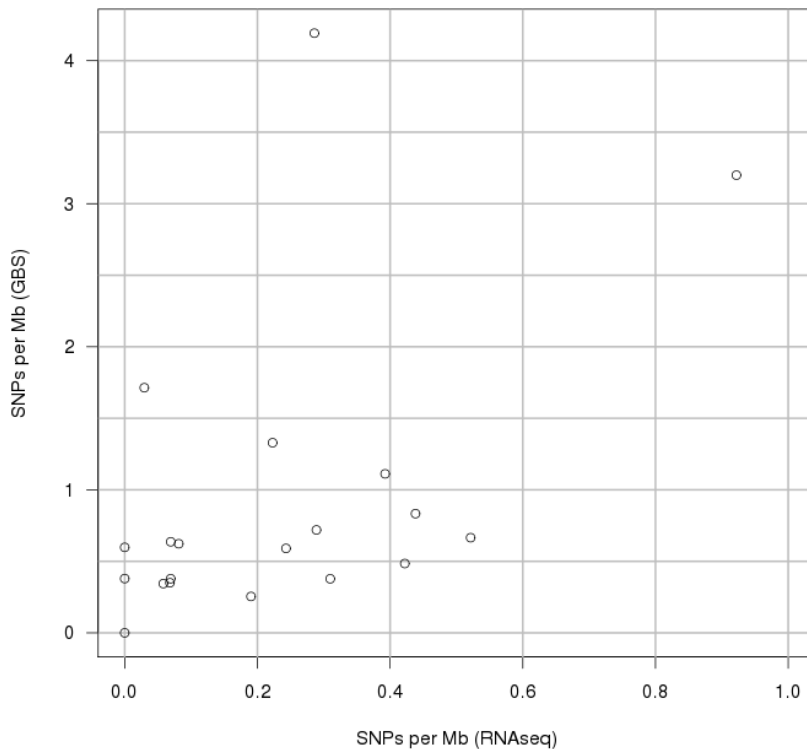
Durch die Bestimmung der Koordinaten eines 6 Mb umspannenden Intervalls auf dem langen Arm von Chromosom 10, welches den Blühzeitpunkt QTL enthält, konnten die Anzahl von Proteinkodierenden Genen und die Menge der repetitiven DNS bestimmt werden, um einen Sequenz „Capture-Array“ zu designen. Phänotypische gleiche Pools aus einer F<sub>2</sub>:F<sub>3</sub> Population, segregierend nach einem frühen Blühzeitpunkt, wurden für GBS und WGS sequenziert. Ein bioinformatisches Analysewerkzeug für Lesepaarkartierung und der Detektion von Sequenzunterschieden zum B73 Referenzgenom wurde entwickelt. In beiden Experimenten, WGS und GBS, wurde ein Peak für die Allel Frequenz auf Chromosom 10 in der gleichen Region des Kartierungsintervalls von P2 (KWS) gefunden. Um die Mutationsfrequenz der neu entwickelten Mutanten abschätzen zu können wurden 20 M<sub>1</sub> Pflanzen und der Wild Typ per GBS sequenziert. Die Mutationsfrequenz wurde auf ~1,4 SNPS pro Metabasen geschätzt. Eine feine Kartierung des Blühzeitpunkt QTL auf Chromosom 10 wurde mit einer begrenzten Anzahl von rekombinanten DNA Proben von der KWS aus dem Juni 2012 durchgeführt. Um die Auflösung der Kartierung weiter zu erhöhen wurden Sequenzdaten von ausgewählten Rekombinanten und deren Nachkommen analysiert. Die DNA Proben wurden einzeln markiert, um sie in den Sequenzbibliotheken wiederzufinden und sie einzelnen genotypischen Gruppen anhand ihrer Sequenzvariation zuordnen zu können. Die Sequenzdaten wurden mit unserem etablierten bioinformatischen Algorithmus analysiert. Die Lesepaare wurden auf das B73 Referenzgenom mit Hilfe von BWA abgebildet. Der Ertrag der Sequenzierung lag zwischen 1,5 und 9 Millionen Lesepaaren pro sequenziertem Individuum. Ungefähr 70% der Lesepaare konnte unverwechselbar auf dem B73 Referenzgenom abgebildet werden. Unterschiede zwischen den Genotypen und Genomabschnitten wurde mit dem Programm SAMtools untersucht. Hierfür war eine minimal fünffache Sequenziertiefe erforderlich. Zwischen den zwei Elter Pflanzen wurden ~20.000 segregierende SNPs mit weniger als 50% fehlenden Datenpunkten gefunden. Für die einzelnen Individuen wurden zwischen 4.000 und 18.000 genotypische Unterschiede (SNPs) gefunden. Sogar mit einem stringenteren Umgang mit fehlenden Datenpunkten (weniger als 20% fehlende Datenpunkte) wurden immer noch ~4.000 SNPs identifiziert. Innerhalb der Zielregion für den Blühzeitpunkt QTL (Chr.10) waren ~50 SNPs zu finden. Die Genotypen wurden in R mit einer genomweiten Auflösung und in der Zielregion visualisiert (Abbildung 1). Die Ergebnisse aus der Sequenzgestützten Genotypisierung waren in voller Übereinstimmung mit den Ergebnissen aus der Einzelmarkeranalyse von P2 (KWS). Die Position des Blühzeitpunkt QTL wurde durch einen genomweiten QTL Scan mit Rqtl bestätigt. Mit der bildlichen Darstellung der Genotypen war uns eine Identifikation der nahen Rekombinanten möglich.



**Abbildung 2:** Die bildliche Darstellung der Genotypen von ausgewählten Rekombinanten in der Nähe des Blühzeitpunkt QTL auf Chromosom 10.

Um die abgeschätzte Mutationsrate der EMS Population bestätigen zu können, wurde ein RNAseq Datensatz der gleichen 20 Individuen analysiert, die auch mittels GBS untersucht wurden. Dazu passten wir unsere Programme an Transkriptionsdaten an, indem wir die Software Tophat, für gespleißte Sequenzalignments, benutzten. SNPs wurden mit SAMtools identifiziert. Dafür war eine zwanzigfache Sequenziertiefe unabdingbar, um verlässlich Individuum basierte Heterozygotie identifizieren zu können (max. 33% fehlenden Datenpunkte erlaubt). Die kumulative Größe der Zielsequenz in RNAseq (25 Mb) war vergleichbar mit der der Sequenzgestützten Genotypisierung (21 Mb). Die geschätzte Mutationsrate korrelierte zwischen GBS und RNAseq Daten (Pearson's  $r = 0.5$ , Abbildung 2). Die durchschnittliche SNP Frequenz was jedoch geringer in den RNAseq Daten als in den GBS Daten (0,5 SNPs pro Mb vs. 1,4 SNPs pro Mb). Wir erklären dies mit einer potentiell geringeren SNP Frequenz, welche wahrscheinlich dem Selektionsdruck geschuldet ist, der nicht synonyme Mutationen in den transkribierten Exons aussortiert. Des Weiteren werden Transkripte mit „nonsense“ Punktmutationen wahrscheinlich gar nicht erst exprimiert, was eine Detektion durch RNS Sequenzierung verhindert.

### EMS SNP frequency in RNAseq vs. GBS data



**Abbildung 3:** Die geschätzte SNP Frequenz der RNAseq Daten (x-Achse) verglichen mit der geschätzten SNP Frequenz der GBS Daten (y-Achse).

#### Zu B-AP1:

Gemeinsam mit Kooperationspartnern am James Hutton Institute und der University of Minnesota/St. Paul sowie Roche/NimbleGen wurde ein neuartiger „Flüssigphasen-Exome-Capture-Assay“ mit 90 Mbp Zielsequenz basierend auf 4,2 Mio. Motiven entwickelt. Zu seinem Test in der NGS-basierten Kartierung wurden rekombinante Pools für das „awn roughness“ Merkmal aus der Morex x Barke RIL-Population (je 40 Rekombinante mit den alternativen Merkmalsausprägungen) gebildet und die angereicherte DNA einer tiefen NGS unterzogen. Die erhaltenen Daten erlaubten eine Kartierung des „awn roughness“ Locus innerhalb eines einzelnen BAC-Klons des für die Region verfügbaren „minimal tiling path“ BAC-Contigs. Eine Kartierung des *dsp.ar*-Lokus erfolgte mit Hilfe der Exome-Capture-NGS tiefer Pools von Rekombinanten, die aus 3000 F2-Individuen selektiert wurden. Dabei wurde zunächst nur eine geringe Zahl an Kandidatengen gefunden. Die Verfahrensweise wurde daher um einen RNA-Sequenzierungs-Ansatz erweitert und nach Bekanntwerden der Allelität von *dsp* und *erectoides-a* (*ert-a*) Mutationen wurde in Kooperation mit C. Dockter (Carlsberg Lab., Copenhagen) drei Gruppen von insgesamt 41 unabhängig induzierten *ert-a*-Mutanten in die Untersuchungen einbezogen. Die für jede der drei Gruppen aus gepoolten DNAs erzeugten Exome-Capture-Banken wurden jeweils einer tiefen NGS-DNA-Sequenzierung unterzogen. Allerdings war auch mit Hilfe dieser Daten keine eindeutige Identifizierung eines Kandidatengens möglich, da keine potenziell ursächlichen Mutationen erkennbar waren und auf Grund der geringen Rekombinationsrate in der Region eine Priorisierung auf Basis der genetischen Position nicht möglich war.

Der entwickelte NGS-basierte Kartierungsansatz wurde jedoch erfolgreich für eine weitere positionale Gen-Identifizierung genutzt: Für eine Röntgen-induzierte *many-noded dwarf* (*mnd*) – Mutante wurde eine F2-Kartierungspopulation erstellt und DNA-Pools von 20

Individuen mit mutantem sowie 40 Individuen mit Wildtyp Phänotyp einer Exome-Capture-NGS unterzogen. Mit Hilfe der entwickelten bioinformatischen Analyse-Pipeline unter der Nutzung der assemblierten WGS-Sequenz des Gerste-Kultivars Barke wurden die SNP-Allelfrequenzen visualisiert und es konnte eine klare Positionierung des Locus innerhalb des Intervalls von Pos. 80 cM bis Pos. 110 cM des Chromosoms 5H erzielt werden. Mittels Untersuchung der erzielten Sequenzieretiefe in dieser Region wurde für ein einzelnes Gen die Anwesenheit im Wildtyp-Pool und Abwesenheit im mutanten Pool festgestellt, dessen Deletion in der Mutante durch PCR bestätigt wurde.

#### **Zu B-AP2:**

Auf der Basis von zunächst 1000 F2 Individuen ausgewählte Rekombinante der Kartierungspopulation für die *lax-a* Mutation wurden genutzt, um zwei phänotypische Pools von je 20 Genotypen zu erstellen und einer NGS RNA-Sequenzierung zu unterziehen. Mit Hilfe der für diese und die in B-AP1 erzeugten Daten erstellten bioinformatischen Analyse-Pipeline und der Nutzung der physisch verankerten Assemblies einer WGS-Sequenz des Gerste-Kultivars Morex konnte für das *lax-a* Gen eine klare Zuordnung zu einem genomischen Intervall erreicht werden.

#### **Zu B-AP3:**

Die für den *lax-a* Locus verfügbare Kartierungspopulation wurde auf 2000 F2-Individuen erhöht und auf Basis flankierender Marker-Daten wurden tiefe Pools von Rekombinanten erstellt, um die genetische Auflösung in der Region dieses Locus zu steigern. Auf der Basis des Typs der Rekombinanten erstellte Pools wurden nach individueller Markierung („Barcoding“) der Exome-Capture-NGS unterzogen und es wurden die Allelfrequenzen der DNA-Polymorphismen in der Umgebung des *lax-a*-Locus ermittelt, die exakt den zuvor bestimmten Rekombinationsereignissen entsprachen. An Hand der Kriterien (i) einer Übereinstimmung von SNP-Frequenzen mit den *lax-a* Phänotypen in den rekombinanten Pools und (ii) fehlender Repräsentierung von Sequenzen in den Exome-Capture-NGS-Daten auf Grund potenzieller Deletion wurden 9 Kandidaten-Contigs des WGS-Genom-Referenzassembly von Morex und ein deletiertes Gen identifiziert. Unter den dort lokalisierten Genen wies nur das deletierte Gen einen Expressionsunterschied in den RNA-Seq.-Daten auf und stelle daher das wahrscheinlichste Kandidatengen dar. Seine Deletion in der *lax-a* Mutante BW457 wurde mittels PCR verifiziert und die Kartierung von WGS-Sequenzdaten dieser Mutante wies eine ca. 450 kb große Deletion nach, in der nur das eine Kandidatengen lokalisiert war.

#### **Zu B-AP4:**

Zur Verifikation des in B-AP3 identifizierten *LAX-A*-Kandidatengens wurden TILLING-Screens in der GABI-TILL Gersten-Population des IPK durchgeführt und 17 nicht-synonyme Mutationen detektiert. Darunter befand sich eine Mutante mit einem Leucin zu Phenylalanin Aminosäureaustausch, der mit einem *lax-a*-typischen Phänotyp einherging. Eine weitere große Serie von 28 unabhängigen *lax-a* mutanten Allelen wurde im Rahmen einer Kooperation mit S. Taketa (Okayama Univ., Japan) herangezogen, um die Identität des *LAX-A*-Gens eindeutig zu bestätigen.

Auch für das Kandidatengen, das für den *mnd*-Locus identifiziert worden war, wurden TILLING-Screens durchgeführt und es wurden 18 nicht-synonyme Punktmutationen sowie eine vorzeitige Stop-Mutation identifiziert. Mit Hilfe einer größeren Zahl von Gersten-Akzessionen, die von der 'Nordic Genebank' bereitgestellt wurden und als *mnd*-Mutanten beschrieben waren, wurde die Identität des identifizierten Gens auf Grund eines breiten Spektrums von beobachteten Mutationen bestätigt.



### Zu TB-AP2:

Die Verwaltung von Sequenzrohdaten, die am IPK vom der Hochdurchsequenziergerät „Illumina HiSeq2000“ produziert werden, wurde im IPK-internen Laborinformationsmanagementsystem (LIMS) implementiert, die eine nahezu automatisierte Abspeicherung unterstützt. Die Rohdaten im FASTQ-Format werden automatisch in ein hierarchisches Speichersystem (HSM) transferiert, wo mehrere redundante Kopien auf Festplatten und Magnetbändern, die Datenintegrität gewährleisten. Die Speicherung der Daten und die Erfassung der Metadaten erfolgt in einer Form, so dass eine Submission zu öffentlichen Sequenzdatenbanken wie beispielsweise EMBL/ENA vereinfacht wird. Für die Gerstendaten ist diese Submission bereits erfolgt. Die Accession-Nummern sind in der erfolgten Publikation aufgeführt.

In der Anfangsphase des Projektes wurde eine Anfrage an alle Projektpartner gestellt, um den Bedarf für ein projektpartnerübergreifendes Forschungsdatenmanagement (Archivierung, Zugriff und Integration von Datensätzen diverse Domänen) zu eruieren. Da diese Anfrage negative beschieden wurde, wurde dieser Ansatz nicht weiterverfolgt. Stattdessen, wurde eine projektinterne Webseite aufgesetzt, auf der bestehende, für das Projekt relevante Datenbanken und von den Projektpartnern entwickelte Softwaretools aufgelistet wurden.

### Zu TB-AP4:

Am IPK bestehende Analysepipelines zur Sequenzvariationsdetektion mittels Hochdurchsatzsequenzierung wurden angepasst, um den Shoremap-Ansatz (Mapping-by-sequencing) zu implementieren. Die Vorprozessierung und Alignierung von Sequenzreads gegen ein Referenzgenom sowie die Identifikation von Einzelnukleotidevarianten benutzen die weitverbreitete Kommandozeilentools cutadapt, BWA und SAMtools (Abbildung 4). Die Berechnung werden parallelisiert unter Ausnutzung von Multiprozessor-Rechenservers am IPK durchgeführt. Zur Allelfrequenzabestimmung und Visualisierung wurden eigene Skripte in der R-Programmierungsumgebung entwickelt. Die mit dieser Software durchgeführten Analysen sind in den Mais- und Gerstenarbeitspaketen dargelegt. Die Pipelines und Ergebnisse wurden im Rahmen der Projektpublikationen beschrieben und öffentlich zugänglich gemacht.

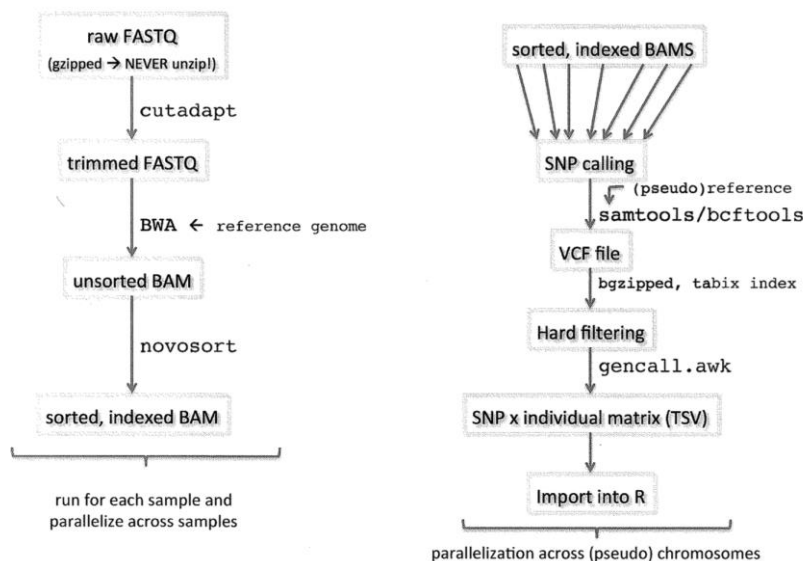


Abbildung 4: Analyse-Workflow zur Verarbeitung von Genotyping-by-Sequencing-Daten.

## II.2. Die wichtigsten Positionen des zahlenmäßigen Nachweises

- Innerhalb der ursprünglich vorgesehenen Projektlaufzeit (2011-2014) wurden 3 verschiedene Projekttreffen an den Standorten Tübingen, Gatersleben und Kiel organisiert.
- von P2 (KWS) wurden 2.146 DNA Proben von allen F2:F3 Familien sowie von weiteren 33 F3:F4 Familien an P1 (IPK-TA) zu Analyse geliefert.
- 4 gepoolte DNA Proben (2x Pool aus 200 F<sub>2</sub>, beide Eltern) wurden mit einem Ertrag von 600k bis 900k gepaarten Lesepaaren mit der Illumina NGS Technologie sequenziert.
- 4 in ihrer Komplexität reduzierte gepoolte DNA Proben (2x Pool aus 200 F<sub>2</sub>, beide Eltern) wurden mit einem Ertrag von 160 bis 250 Millionen gepaarten Lesepaaren mit der Illumina NGS Technologie sequenziert.
- DNA wurde individuell von 33 der F3:F4 und 22 der F2:F3 nach Verdau mit ApeKI mit einer Tiefe von 1,5 bis 1,9 Millionen Lesepaaren sequenziert.
- 70% der Lesepaare konnte unverwechselbar auf dem B73 Referenzgenom abgebildet werden.
- 20.000 segregierende SNPs wurden zwischen den Elter Pflanzen gefunden.
- Für die einzelnen Individuen wurden zwischen 4.000 und 18.000 genotypische SNPs gefunden.
- Mit stringenterem Filter wurden im Schnitt immer noch 4.000 SNPs identifiziert.
- Innerhalb des Blühzeitpunkt QTL (Chr.10) waren 50 SNPs zu finden
- 1 Kandidatengen wurde für den Blühzeitpunkt QTL auf Chromosom 10 identifiziert.
- von 1.313 in den Jahren 2012 und 2013 mit EMS-behandelten Pollen bestäubten Pflanzen wurden 10.582 individuelle M<sub>1</sub> Maiskörner gewonnen.
- 20 M1-Individuen wurden auf DNA und RNA Ebene auf der Illumina HiSeq 2000 Plattform sequenziert, um die Frequenz der EMS induzierten Mutationen zu dokumentieren.
- Die Mutationsfrequenz liegt durch GBS bestimmt bei 1,4 SNP pro Megabasen.
- Die Mutationsfrequenz liegt durch RNAseq bestimmt bei 0.5 SNP pro Megabasen.
- 1.600 M<sub>1</sub> Maispflanzen wurden 6 Wochen auf der IPK-Phänotypisierungsanlage (LemnaTec) phänotypisch analysiert und dann ins Freiland ausgepflanzt.
- Im Jahr 2013 wurden insgesamt 7.600 M1-Pflanzen ins Feld ausgepflanzt; wg. widriger Witterungsverhältnisse konnten jedoch nur 1.771 Kolben (M<sub>2</sub>) geerntet werden.
- Von 309 im Gewächshaus in den Jahren 2013 und 2014 erneut mutagenisierten PH207 Pflanzen konnten weitere 9.893 M<sub>1</sub> Körner geerntet werden.
- 6.000 M<sub>1</sub> Pflanzen wurden im Jahr 2014 ins Feld ausgepflanzt und produzierten 5.500 Kolben mit durchschnittlich 250 M<sub>2</sub> Körnern
- insgesamt wurden mehr als 7.000 Kolben mit durchschnittlich 160 M<sub>2</sub> Körnern produziert.
- 256 zufällig ausgewählte M<sub>2</sub> Familien wurden mit je 8 Individuen auf der IPK-Phänotypisierungsanlage (LemnaTec) phänotypisch evaluiert und zudem visuell bonitiert.
- Das vegetative Wachstum von 8,6% dieser Linien war beeinträchtigt.
- Die meisten Mutationen wurden rezessiv, eine auch semi-dominant vererbt.
- Automatisierte Bildanalyse konnte, an 2 Beispielen gezeigt, die Mutanten eindeutig identifizieren.
- Die Chlorophyll Fluoreszenz der Zwergmutante und der blassgrüne Mutante war signifikant unterschiedlich zum Wild Typ.

- 32 M<sub>2</sub> Mutanten wurden individuell sequenziert und Sequenziertiefen von 14 - 24 x erreicht.
- Für jedes individuell sequenzierte Individuum konnten zwischen 321 und 556 Millionen Lesepaare erreicht werden.
- Im Schnitt wurden für jede der Mutanten 2.000 EMS induzierte Punktmutationen gefunden.
- Mit Hilfe von M<sub>3</sub> Phänotypen, konnten die Zwergmutation (Chr.1, 240 – 245 Mb) und die blassgrüne Mutation (Chr.10, 144 – 144,9 Mb) auf dem Genom kartographiert werden und ein Kandidatengen isoliert werden.
- Für Gerste wurde ein neuartiger „Flüssigphasen-Exome-Capture-Assay“ mit 90 Mbp Zielsequenz basierend auf 4,2 Mio. Motiven entwickelt.
- Die Kartierungspopulation für den *lax-a*-Lokus wurde von 1000 auf 2000 Individuen erhöht.
- Der entwickelte Exome-Capture-NGS-Ansatz wurde für Pools von 2x 40 rekombinanten Individuen bzgl. des „awn roughness“ Merkmals angewendet. Ebenso für rekombinate Pools des *dsp.ar*-Lokus, die aus 3000 F2 Individuen selektiert worden waren, für drei Guppen von insgesamt 41 *ert-a*-Mutanten, für rekombinate Pools des *lax-a*-Lokus, die aus 2000 F2 Individuen selektiert worden waren, und für phänotypische Pools von 20 und 40 *mnd* bzw. Wildtyp F2-Individuen.
- Für rekombinante Pools des *dsp.ar*-Lokus sowie des *lax-a* Lokus wurden NGS RNA-Sequenzierungen durchgeführt.
- Für das *LAX-A*-Kandidatengen wurden mittels TILLING-Screens 17 nicht-synonyme Mutationen identifiziert.
- Für das *LAX-A*-Kandidatengen wurden mittels TILLING-Screens 17 nicht-synonyme Mutationen identifiziert.
- Für das *MND*-Kandidatengen wurden mittels TILLING-Screens 18 nicht-synonyme Mutationen und eine vorzeitige Stop-Mutation identifiziert.
- Für die Gerste wurden zwei Gene, *LAX-A* und *MND*, mit Hilfe des entwickelten bzw. optimierten NGS-basierten Verfahrens molekular identifiziert.
- Für die Verwaltung der am IPK produzierten NGS-Sequenzrohdaten (vom „Illumina HiSeq2000“-Gerät) wurde im IPK-internen Laborinformationsmanagementsystem (LIMS) ein Verfahren implementiert, das eine nahezu automatisierte Abspeicherung unterstützt.
- Es wurde eine projektinterne Webseite aufgesetzt, auf der die für das Projekt relevante Datenbanken und entwickelte Softwaretools aufgelistet wurden.
- Am IPK bestehende Analysepipelines zur Sequenzvariationsdetektion mittels Hochdurchsatzsequenzierung wurden angepasst, um den Shoremap-Ansatz (Mapping-by-sequencing) zu implementieren.
- In der R-Programmierungsumgebung wurden Skripte zur Allelfrequenzabestimmung und Visualisierung entwickelt, die für die Analyse der Mais- und Gerstendaten eingesetzt wurden.

### **II.3. Notwendigkeit und Angemessenheit der geleisteten Arbeit**

In diesem Projekt wurden neue Verfahrensweisen für die Detektion agronomisch wichtiger Gene und deren allelische Variation entwickelt, optimiert und genutzt. Damit wurden die Möglichkeiten der zukünftigen Entwicklung neuen genetischen Materials mit verbesserter Ertragsbildung und Ertragsstabilität wesentlich erhöht, indem die Detektion genetischer Variation wesentlich beschleunigt und erleichtert wird, die in direkter Koppelung zu Merkmalen mit kommerziellem Interesse stehen. Die erhaltenen Ergebnisse und die geleistete Arbeit entsprechen somit in vollem Umfang den Zielen des Förderprogramms.

Mit den innerhalb der (verlängerten) Projektlaufzeit erzielten Ergebnisse wurden alle

entscheidenden Projektziele erreicht: Es wurden umfangreiche methodisch-technische Weiterentwicklungen und Optimierungen der NGS-basierten Identifizierung und Kartierung von Mutationen erzielt, die es erlauben diese zuvor nur an Modellorganismen angewendete Verfahrensweise auch an den Kulturpflanzen Mais und Gerste mit sehr komplexen Genomen effizient einzusetzen. Die Effektivität der Methodik und der auch konzeptionell weiterentwickelten Verfahrensweisen wurde bei beiden Kulturarten an mehreren Beispielen sowohl für natürlich vorkommende als auch für induzierte Mutationen/Sequenzvariationen belegt und es wurden jeweils Gene identifiziert, die agronomisch bedeutende Nutzpflanzeigenschaften (Blühzeitpunkt, Ährenarchitektur) beeinflussen. Die auf diese Weise identifizierten Gene und deren allelischen Varianten stehen für weiterführende Untersuchungen und mögliche wirtschaftliche Anwendungen zur Verfügung. Mit der Erstellung einer sehr großen neuen und sehr gut charakterisierten Mais-Mutantenpopulation im genetischen Hintergrund einer Maislinie, die wesentlich zur Begründung des europäischen Zuchtmaterials beigetragen hat (PH207), ist zudem eine höchst wertvolle experimentelle Ressource geschaffen worden. In Kombination mit der verbesserten und beschleunigten Verfahrensweise für die Identifizierung von durch Mutationen betroffenen Gene, die phänotypische Veränderungen verursachen, ist damit eine neuartige Grundlage für die Grundlagen- und die Anwendungs-bezogene Forschung an Kulturpflanzen in Deutschland geschaffen worden, von der sowohl die akademische wie auch die industrielle Forschung sehr profitieren wird (insbesondere im Hinblick auf die durch andere Fördermaßnahmen weiterentwickelten und implementierten Pflanzenphänotypisierungsverfahren und Plattformen).

#### **II.4. Voraussichtlicher Nutzen (Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplanes)**

In diesem Teil-Projekt wurden sehr wichtige Erkenntnisse und Erfahrungen über die Nutzung von NGS-Verfahren für die Identifizierung und Kartierung von Mutationen gewonnen, die zunächst nur für Modellorganismen mit der kleinen und kompakt strukturierten Genomen zur Verfügung standen. Mit den in diesem Projekt erzielten Weiterentwicklungen der Verfahren und ihrer an einschlägigen Anwendungsbeispielen belegten universellen Anwendbarkeit können sie unmittelbar für die Grundlagenforschung und die angewandte Forschung an Kulturpflanzen eingesetzt werden. Die unmittelbare Einbindung der Industriepartner in das Verbundprojekt stellt eine direkten Technologie- und Kenntnistransfer in die Anwendung hinsichtlich der Stärkung und Beschleunigung von Züchtungsverfahren sicher. Besondere Bedeutung erhalten die erzielten methodischen Fortschritte, die implementierten Datenanalyseverfahren und die geschaffenen neuen Ressourcen (Mutantenpopulation) vor dem Hintergrund der parallel (durch andere Fördermaßnahmen des BMBF) entwickelten und eingerichteten Verfahren und Installationen zur Hochdurchsatzphänotypisierung von Pflanzen. Die sich aus der Kombination der beiden Technologieplattformen ergebende Synergie wird sowohl akademischen wie auch industriellen Pflanzenforschern ermöglichen in wesentlich beschleunigter Weise neues geistiges Eigentum zu entwickeln; neue Kenntnisse über Kulturpflanzengene und deren Variation, die wirtschaftlich wichtige Merkmale von Pflanzen bestimmen. Die erhaltenen Erkenntnisse können zudem leicht auf weitere Systeme übertragen werden, in denen die Identifizierung merkmalsbestimmender genetischer Faktoren von Bedeutung sind.

#### **II.5. Während der Durchführung des Vorhabens dem ZE bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen**

Vergleichbare Ergebnisse wurden im Verlaufe des Projektes von anderen Stellen nicht bekannt.

Die im Projektverlauf an anderer Stelle erfolgte molekulare Identifizierung eines

Blühzeitpunkt-bestimmendes Gen in Mais belegte die korrekte Kartierung und Erkennung des für den Blühzeitpunktunterschied von P1 und P2 verantwortlichen Gens.

## II.6. Erfolgte oder geplante Veröffentlichungen des Ergebnisses

Jost M, Taketa S, Mascher M, Himmelbach A, Yuo T, Shahinnia F, Druka A, Schmutzer T, Steuernagel B, Platzer M, Taudien S, Morgante M, Waugh R, Scholz U and Stein N (2016) A homolog of *Blade-On-Petiole 1* and 2 (*BOP1/2*) controls internode length and homeotic changes of the barley inflorescence. *Plant Physiol* accepted with revision.

Jost M, Mascher M, Kuon J-E, Himmelbach A, Aßfalg A, Beier S, Scholz U, Graner A and Stein N (2016) Mapping-by-sequencing of the gene *HvMnd5H*. manuscript in preparation.

Mascher M, Jost M, Kuon J-E, Himmelbach A, Aßfalg A, Beier S, Scholz U, Graner A and Stein N (2014) Mapping-by-sequencing accelerates forward genetics in barley. *Genome Biology* **15**:R78.

Pankin A, Campoli C, Dong X, Kilian B, Sharma R, Himmelbach A, Saini R, Davis SJ, Stein N, Schneeberger K and von Korff M (2014) Mapping-by-Sequencing Identifies HvPHYTOCHROME C as a Candidate Gene for the early maturity 5 Locus Modulating the Circadian Clock and Photoperiodic Flowering in Barley. *Genetics*:10.1534/genetics.1114.165613.

Mit dem in Kürze erwarteten Abschluss der Verifikation der in der PH207-Mutantenpopulation neu identifizierten und kartierten Mais-Genloci ist die Veröffentlichung des für Mais entwickelten und angewendeten Verfahrens sowie der Mutantenpopulation geplant.

#### IV. Kurzfassung (Berichtsblatt; Document Control Sheet)

1. ISBN oder ISSN	2. Berichtsart (Schlussbericht oder Veröffentlichung) Schlussbericht
3. Titel Forschungsvorhaben Teilprojekt: 0315957A Thema: Pflanzenbiotechnologie-Verbundvorhaben: Entwicklung und Anwendung der neusten Generation genetischer Technologien bei Kulturpflanzen (NuGGET, Teilprojekt A)	
4. Autor(en) [Name(n), Vorname(n)] Altmann, Thomas Heuermann, Marc Scholz, Uwe Mascher, Martin Stein, Uwe	5. Abschlussdatum des Vorhabens 31.07.2015
	6. Veröffentlichungsdatum
	7. Form der Publikation
8. Durchführende Institution(en) (Name, Adresse) Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) Gatersleben Corrensstraße 3 06466 Stadt Seeland, OT Gatersleben	9. Ber. Nr. Durchführende Institution
	10. Förderkennzeichen 0315957A
	11. Seitenzahl 23
12. Fördernde Institution (Name, Adresse)  Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. Literaturangaben 37
	14. Tabellen 2
	15. Abbildungen 4
16. Zusätzliche Angaben	
17. Vorgelegt bei (Titel, Ort, Datum)	
18. Kurzfassung Im Pflanzenbiotechnologie-Verbundvorhaben „Entwicklung und Anwendung der neusten Generation genetischer Technologien bei Kulturpflanzen (NuGGET, Teilprojekt A)“ wurden neue Verfahrensweisen für die Detektion agronomisch wichtiger Gene und deren allelische Variation entwickelt, optimiert und genutzt: Es wurden umfangreiche methodisch-technische Weiterentwicklungen und Optimierungen der NGS-basierten Identifizierung und Kartierung von Mutationen erzielt, die es erlauben diese zuvor nur an Modellorganismen angewendete Verfahrensweise auch an den Kulturpflanzen Mais und Gerste mit sehr komplexen Genomen effizient einzusetzen. Die Effektivität der Methodik und der auch konzeptionell weiterentwickelten Verfahrensweisen wurde bei beiden Kulturarten an mehreren Beispielen sowohl für natürlich vorkommende als auch für induzierte Mutationen/Sequenzvariationen belegt und es wurden jeweils Gene identifiziert, die agronomisch bedeutende Nutzpflanzeigenschaften (bei Mais: Blühzeitpunkt, (vegetatives) Wachstum, Pigmentierung; bei Gerste: Ährenarchitektur, Habitus) beeinflussen. Die in diesem Projekt identifizierten Gene und deren allelischen Varianten stehen für weiterführende Untersuchungen und mögliche wirtschaftliche Anwendungen zur Verfügung. Mit der Erstellung einer sehr großen neuen und sehr gut charakterisierten Mais-Mutantenpopulation im genetischen Hintergrund einer Maislinie, die wesentlich zur Begründung des europäischen Zuchtmaterials beigetragen hat (PH207), ist zudem eine höchst wertvolle experimentelle Ressource geschaffen worden. In Kombination mit der verbesserten und beschleunigten Verfahrensweise für die Identifizierung von durch Mutationen betroffenen Gene, die phänotypische Veränderungen verursachen, ist damit eine neuartige Grundlage für die Grundlagen- und die Anwendungs-bezogene Forschung an Kulturpflanzen in Deutschland geschaffen worden, von der sowohl die akademische wie auch die industrielle Forschung sehr profitieren wird (insbesondere im Hinblick auf die durch andere Fördermaßnahmen weiterentwickelten und implementierten Pflanzenphänotypisierungsverfahren und Plattformen. Mit den innerhalb der (verlängerten) Projektlaufzeit erzielten Ergebnisse wurden alle entscheidenden Projektziele erreicht.	

19. Schlagwörter Mapping by Sequencing, Nest Generation Sequencing, Gerste, Mais	
20. Verlag	21. Preis