

Veröffentlichung der Ergebnisse von Forschungsvorhaben im BMBF-Programm

B I O L O G I E

"Das diesem Bericht zugrundeliegende BMBF-Forschungsvorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 0315063C gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor".

Forschungsvorhaben: GABI-FUTURE-Verbundvorhaben: 'Erschließung des genetischen Potentials von Roggen mittels Etablierung einer Ressource zur funktionellen Genomanalyse des exprimierten ('EXPRESSed') Anteil des Roggengenoms (GABI-RYE EXPRESS)'(Teilprojekt C)

Förderkennzeichen: 0315063C

Zuwendungsempfänger: Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Postfach 11 29, 85758 Oberschleißheim

Ausführende Stelle: Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH) - Institut für Bioinformatik (IBI), Ingolstädter Landstr. 1, 85764 Oberschleißheim

Projektleitung: Herr Dr. Mayer

Laufzeit: 01.01.2008 bis 31.12.2010

Project (+ Acronym)	Title	<i>GABI-RYE EXPRESS: Unlocking the genetic potential of rye by establishing a functional genomics resource for the EXPRESSED portion of the rye genome, Teilvorhaben C</i>
Identific. Number (FKZ)		0315063C
Grant (+ Investigator)	holder Principle	K. Mayer
Project Partners (if appl.)		WZW, Technische Universität München IPK Gatersleben
Project Koordinator		Dr. Eva Bauer
Subcontractor(s)		n.a.

I. Kurzbeschreibung der Projektfragestellungen

Im Rahmen des GABI RYE Express Projektes wurde eine grosse Anzahl von Roggen ESTs mittels NGS Technologie aus verschiedenen Kultivaren sequenziert. Diese Ressourcen dienen dazu kultivarspezifische *assemblies* durchzuführen, genomische Variationen -SNPs, InDels- zu detektieren und eine *transcriptomics* Resource für Roggen zu entwickeln. Diese Ressource sollte zu u.a. zum Design eines oligobasierenden Roggengenotypisierungs Arrays verwendet werden (siehe Berichte der Teilvorhaben A & B). Spezifische Aufgabe des Teilprojektes C war es diese genomweiten Transkriptomdaten in den Kontext mit anderen genomischen Referenzen aus den Gräsern und Triticeae zu stellen. Dazu sollten mehrere Arbeitskomponenten und –pakete beigesteuert werden:

- (a) die Entwicklung geeigneter genomischer Referenzen und dem Aufbau entsprechender Datenbankreferenzierungssysteme sowie
- (b) die Entwicklung geeigneter und nutzerfreundlicher Visualisierungssysteme zur Darstellung dieser Daten und schliesslich
- (c) die Darstellung und Konzeption einer integrativen Darstellung von genomischen Roggengenen im Kontext mit Referenzgenomen und unter Einbeziehung integrativer Ansätze.

Zu Beginn des Vorhabens war öffentlich nur ein kleiner Datensatz öffentlich verfügbarer Roggensequenzen vorhanden (rund 10000 EST Sequenzdaten). Untersuchungen im Rahmen des Projektes zeigten jedoch dass diese Resource hochgradig kontaminiert war und zu rund 2/3 tatsächlich aus Weizensequenzen bestand. Folglich war der tatsächlich verfügbare Datensatz nochmals deutlich geringer und die im Rahmen des Projektes generierten, strukturierten und analysierten Daten umso wertvoller.

II ZIELE

- geplante Ziele
- Erreichte und nicht erreichte Ziele
- Gründe für Nichterreichung von Zielen

Wesentliche **Ziele** des Teilvorhabens waren

- (1) die Entwicklung und Datenbankeinbindung von wichtigen Referenzgenomen im Kontext des Projektes; d.h. Fokussierung auf Gräsergenome,
- (2) die Entwicklung und Implementation eines vergleichenden Visualisierungstools und
- (3) die Einbindung und Darstellung von im Vorhaben generierten Roggen-spezifischen Daten. Projekt

Wichtige **Ergebnisse** konnten im geförderten Projektzeitraum erzielt werden:

- (1) Die Analyse des *Sorghum* Genoms wurde abgeschlossen und im Rahmen internationaler Einbindung zur Publikation gebracht (Paterson et al., 2009, Nature 457, 551-556). Eine *Sorghum* Datenbank wurde implementiert die die vollständigen Daten des *Sorghum* Genoms enthält (<http://mips.gsf.de/proj/plant/jsf/sorghum/index.jsp>).
- (2) Die Analyse des *Brachypodium* Genoms als wichtiges, nahe verwandtes Genom von Roggen (und *Triticeae* im Allgemeinen) wurde durchgeführt und abgeschlossen (The Brachypodium Genome Initiative, 2010, Nature, 463, 763-768). Ein vollständiges und assembliertes Genom wurde im Berichtszeitraum verfügbar gemacht. Umfassendes Training von Analysepipelines und Integration einer Vielzahl von homologen und heterologen Evidenzen wurde durchgeführt und eine genomweite Analyse nach genischen und repetitiven Elementen durchgeführt. Evaluierungen zeigen eine für *de novo* Analysen außergewöhnlich hohe Qualität und Spezifität, die an das jahrelang intensiv kuratierte *Arabidopsis* Genom heranreicht. Damit steht erstmals ein evolutionär und funktionell ein qualitativ äußerst hochwertiges und nah verwandtes Referenzgenom für Roggen (und andere *Triticeae*) zur Verfügung. Eine entsprechende *Brachypodium* Datenbank wurde implementiert, die die vollständigen Daten des Genoms enthält (<http://mips.helmholtz-muenchen.de/plant/brachypodium/index.jsp>).
- (3) Entwicklung und Implementation eines vergleichenden Visualisierungstools wurde vorangetrieben und abgeschlossen. Die umfangreichen und komplexen Genom- und genomassoziierten Daten können aus der Datenbank und via web Zugang durch vielschichtige Suchkriterien abgerufen werden und werden dann in tabellarischer Form dargestellt und kommuniziert. Für solche Datentypen ist jedoch eine nur auf textuelle Suche beschränkte Suche nicht zeitgemäß und wenig komfortabel. Visuell unterstützte Suchen wurden deshalb implementiert. Im Rahmen des Projektes wurde eine elektronische Genomvisualisierungsplattform, CrowsNest, entwickelt und in die Webdarstellung eingebunden und implementiert, die den Erfordernissen nach zum Einen komfortabler und intuitiver Suche, zum anderen der Darstellung großer und komplexer Daten gerecht wird. CrowsNest erlaubt es zum Einen einzelnen Genome skalierbar, vom Gesamtgenom bis zum Einzelgen zu visualisieren und bildet zum Anderen einen komfortable vergleichende Genomics Plattform die es erlaubt zwei verwandte Genome im Vergleich darzustellen. Auch hier ist eine Genomebene bis hin zur vergleichenden Einzelgen-Darstellung unterstützt. Diese Plattform bildet ein robustes Fundament für die Darstellung kompletter Genome, jedoch auch von einzelnen Regionen bis hin zur vergleichenden Darstellung von Einzelgenen. Weiter Informationen und das CrowsNest Interface können unter <http://mips.helmholtz-muenchen.de/plant/crowsNest/index.jsp> erreicht werden.
- (4) Die Genome von Reis, *Brachypodium* und *Sorghum* wurden in die Plattform

eingebunden, die Darstellung von Roggengenen wird in absehbarer Zeit folgen. Dieser Arbeitsschritt konnte im Berichtszeitraum nicht abgeschlossen werden. Dafür sind im wesentlichen wissenschaftliche Gründe ursächlich die zum Zeitpunkt der Antragsstellung noch nicht absehbar waren. Derzeitige Assoziationsregeln und chromosomale Anordnungen beruhen auf Homologiebeziehungen zwischen in RYE EXPRESS erzeugten Roggen EST Assemblierungen und den jeweiligen Sequenzorthologen in den Referenzorganismen. Diese Zuordnung kann und wurde durchgeführt. Mit kürzlich verfügbar gewordenen Daten und analytischen Möglichkeiten kann diese auf bloßer Sequenzähnlichkeit beruhende Ansatz bedeutend verfeinert werden. Zum einen ist eine hochauflösende genetische Markerkarte für Roggen entwickelt worden und zum anderen wurde eine integrativer Analyseansatz (GenomeZipping) entwickelt der es erlaubt mit Hilfe von hochauflösenden Markerdaten und Vergleich gegen Referenzgenome geordnete Pseudochromosome zu errechnen. Dieser Ansatz wurde bereits durchgeführt und wird in naher Zukunft publiziert werden. Die Ergebnisse übertreffen zur Antragstellung antizipierte Ziele bei weitem da nicht nur sequenzbasierte Orthologiebeziehungen sondern deren genaue chromosomale Positionierung erreicht wurde was hinsichtlich der chromosomalen und genomischen Struktur und der Synteniestructur ein wegweisender Schritt ist. Diese Ergebnisse übertreffen die ursprünglichen Vorhabensziele erheblich.

III Vorraussichtlicher Nutzen & Verwertbarkeit der Ergebnisse

Beitrag der Ergebnisse zu den förderungspolitischen Zielen

Die in GABI RYE Express erreichten Ziele stehen im Einklang mit mehreren forschungspolitischen Zielen des GABI-3 Programmes. Die Einzelkomponenten des Projektes stellen Schlüsseltechnologie für die Analyse des transkribierten Teils des Roggengenoms deren speziellen funktionellen Eigenschaften bereit und stellen sie in den Kontext mit anderen genomischen Referenzierungssystemen. Innovative technologische Komponenten zur Analyse, Strukturierung und Darstellung molekularer Roggen Information wurden entwickelt, implementiert und angewandt.

Die erreichten Ergebnisse stellen eine nachhaltige Integrative Roggeninformationsresource für *Trititiceae* fokussierte GABI Projekte als auch für die internationale *community* dar. Sie bildet eine datenreiche und strukturierte Basis für die funktionelle Charakterisierung des Roggengenoms sowie eine notwendige Datenbasis zur Wissens- und Mehrwertgenerierung via vergleichender funktioneller Analyse.

Das Projekt hat kontinuierlich relevante pflanzengenomische Daten die öffentlich verfügbar wurden integriert und an der Generierung und Weiterentwicklung dieser Ressourcen partizipiert.

Wissenschaftlich technische Ergebnisse des Vorhabens

Technische Entwicklung

- Etablierung und Weiterentwicklung roggenpezifischer Ressourcen und bioinformatischer Getreidegenomressourcen
- Entwicklung von Bioinformatik und Datenressourcen für die strukturierte Assoziation genomischer und funktioneller Daten zur Handhabung zukünftiger Roggen- und Trititiceaegenom assoziierter Projekte
- Entwicklung eines skalierbaren und flexiblen Visualisierungstools zur grafischen Darstellung von Syntenedaten zwischen Gräsergenomen
- Entwicklung eines flexiblen und skalierbaren Datenbankframeworks zur Speicherung und Darstellung vergleichender Gräsergenomdaten.

Pflanzengenom-Forschung

- Umfassende Syntenie (Meta und Feinstruktur Ebene) Analysen von Grassgenomen
- Entwicklung eines umfassenden Transkriptomdatensatzes aus Roggen

- Polymorphismusdaten aus Roggen
- Detailliertes genetisches mapping framework für Roggen
- Detailliertes Genomstrukturvergleiche und –beziehungen zwischen Triticeae genomen
- Publikationen
- wissenschaftliche Kooperationen
- Schaffung strukturierter genomischer Ressourcen für die vergleichende und integrative Gräsergenomforschung

Bioinformatische Forschung und Entwicklung

- Weiterentwicklung bioinformatischer Analytik für die pflanzliche Genomforschung
- Bereitstellung strukturierter, analysierter und funktionell integrierter Daten im pflanzengenomischen Kontext
- Vergleichende und kombinatorische Analyse von *in silico* analytischer und experimenteller Daten, Applikation adäquater Analysemethoden und Entwicklung von Analysemethoden zur Hochdurchsatzanalytik von angereicherten Sequenzbibliotheken und genomischen Datensätzen

Forschung zum Züchtungswesen

- Bereitstellung neuartiger molekularer und funktioneller Ressourcen für die Nutzung in der Pflanzenzucht
- Nutzung der detektierten, strukturierten und analysierten Loci für die Identifikation von funktions- und züchtungsrelevanten molekularen Komponenten
- Bereitstellung strukturierter analytischer Ergebnisse zur Nutzung in Kulturpflanzen

Fortschreibung des Verwertungsplanes

O.g. Ergebnisse führten bereits zu Veröffentlichung in wissenschaftlichen Zeitschriften. Weitere Veröffentlichungen die auf den erhobenen Daten und Analysen aufbauen sind in der Vorbereitung. Alle Komponenten der Arbeiten bilden eine fundamentale Basis für weitere Anschlussprojekte im nationalen und internationalen Rahmen. Einige Projekte sind bereits gefördert. Die im Projekt durchgeführten Arbeiten bildeten eine fundamentale Basis für Weiterentwicklung pflanzengenomischer Projekte, insbesondere von komplexen Nutzpflanzengenomen, sowohl auf nationaler als auch auf internationaler Ebene. Sie sind deshalb von unmittelbarem Interesse für kollaborative Projekt in der Züchtungsindustrie. Alle Daten wurden in geeigneten Datenbanken strukturiert und abgelegt und verfügbar gemacht.

Arbeiten die zu keiner Lösung geführt haben

n.a.

Präsentationsmöglichkeiten

Teile der Arbeiten sind bereits in internationalen Kongressen und in wissenschaftlichen Zeitschriften veröffentlicht worden. Weitere Publikationen sind in der Vorbereitung.

Verbesserung der Forschungsstruktur

- Etablierung von diversen genomischen Datenbanken; Weiterentwicklung von Pflanzengenomdatenbanken
- Verbesserung des Zugangs zu strukturierten und hochintegrierten Pflanzengenomdaten
- Weiterentwicklung der pflanzengenomorientierten und züchterischen Forschungs- und Entwicklungsstruktur
- kontinuierliche elektronische Publikation der Daten; Web Darstellung

Bereitstellung qualitativ hochwertiger und umfangreich analysierter/funktionell integrierter Genomdatensätze als Grundlage funktioneller Genomanalytik und molekularen Züchtungsentwicklung