

# Genomforschung bei Getreide

N. STEIN

## Einleitung

Die kleinkörnigen Getreide (Weizen, Gerste und Roggen) sind, gemäß Gesamtanbaufläche und Gesamtproduktion, die wichtigsten Kulturarten in Europa (FAOSTAT 2005). Sie liefern die wesentliche Grundlage für die menschliche Ernährung sowie als Tierfutter und als nachwachsende Rohstoffe. Um diese essentielle Ressource langfristig auch unter veränderten Anbaubedingungen, sich ändernden Klimabedingungen sowie neuen Anforderungen an Getreide-Produkte zu sichern und zu verbessern, bedarf es eines grundlegenden Verständnisses der biologischen Prozesse, die an der Ausbildung agronomischer Eigenschaften beteiligt sind. Einen wesentlichen Beitrag hierzu leistet die weitere Aufklärung der genetischen bzw. genomischen Grundlagen der Merkmalsausprägung. Umfangreiche Ressourcen wurden für die Analyse von Modellpflanzen entwickelt, die in der Erstellung der vollständigen Sequenz des Arabidopsis und des Reis-Genoms mündeten [1, 2]. Diese essentiellen genomischen Ressourcen ermöglichen völlig neue Ansätze in der Aufklärung pflanzenbiologischer Prozesse, und resultierende Erkenntnisse können große Relevanz für eine Übertragung auf unsere Kulturpflanzen besitzen. Jedoch nicht alle Kulturarten-spezifischen Fragestellungen lassen sich durch das Studium der Modellpflanzen beantworten. Z.B. kann Reis aufgrund einer nur ca. 50 %igen Kolinearität zu den Triticeae-Genomen (Weizen, Gerste, Roggen) [3, 4] nur in der Hälfte der Fälle als Modell zur Genisolierung in diesen Kulturarten dienen. Es besteht daher ein grosser Bedarf an der Verbesserung der genomischen Infrastruktur für die Genomanalyse in den Getreiden.

Die Voraussetzungen für die Getreidegenomanalyse haben sich bereits innerhalb der vergangenen 5 Jahre gravierend verbessert und es besteht Grund zu der

Annahme, dass sich diese Situation in den kommenden Jahren weiter verbessern wird. Im Rahmen dieses Vortrags soll ein Überblick über den derzeitigen Stand der Getreidegenomforschung, sowie über absehbare zukünftige Entwicklungen, vermittelt werden.

## Getreidegenomanalyse - die Modelle

### Reis

Wichtigste Modellpflanze für die Getreidegenomanalyse ist der Reis (*Oryza sativa*). Anfang der neunziger Jahre wurde der Grundstein für die eingehende Analyse der Gräsergenome gelegt. Die vergleichende Kartierung von RFLP Markern offenbarte eine über weite Bereiche lineare Konservierung der Anordnung von Markern/Genen (Kolinearität) zwischen Reis und anderen Gräsergenomen. Hieraus wurde das zirkuläre Modell der Gräsergenomkolinearität abgeleitet [5] woraus das Dogma entstand, dass das Reisgenom als Modell für die Genisolierung in den weitaus komplexeren, 10-40x fach umfangreicheren Genomen der Getreide-Kulturarten dienen könnte [6]. Mittlerweile ist das Reisgenom (aus verschiedenen Unterarten) sequenziert (Tabelle 1) [2] und diese Information wird als integraler Bestandteil aller Strategien zur Genomanalyse in übrigen Gräsern genutzt.

### Brachypodium

Die Tatsache, dass das Reisgenom in Teilbereichen nur begrenzte Kolinearität zu den Triticeae Genomen besitzt, hat die Aufmerksamkeit seit kurzem auf eine weitere Modellpflanze gelenkt - *Brachypodium distachyon* und *sylvaticum*. Eine engere stammesgeschichtliche Verwandtschaft zu den Triticeae (Divergenz der Spezies vor ca. 40 Mio Jahren gegenüber ca. 70 Mio Jahren zwischen Reis und den Triticeae), eine kleines Genom, kurze Generationszeit sind typische Kriterien, die *Brachypodium* zu einem inte-

ressanten ergänzenden Modell in der Getreidegenomanalyse machen [7]. Ein internationales Konsortium wurde gebildet mit dem erklärten Ziel die wesentlichen Voraussetzungen zu schaffen, um *Brachypodium* zu einem wirksamen Modell in der Getreidegenomanalyse zu entwickeln (<http://www.brachypodium.org/>). Erst kürzlich wurde der Beschluss gefasst, das Genom von *Brachypodium distachyon* vollständig zu sequenzieren (Shotgun-Sequenzierung, <http://www.jgi.doe.gov/sequencing/why/CSP2007/brachypodium.html>).

## Genomanalyse bei Getreiden

### Mais und Sorghum

Mais ist weltweit unter den drei wichtigsten Kulturarten und ist das Getreide mit der höchsten Wertschöpfung für Züchtungsindustrie und Landwirte. Dies erklärt zum Teil, warum die Ressourcen zur Genomanalyse in Mais gegenüber denen für kleinkörnige Getreide wesentlich weiter entwickelt sind (Tabelle 1).

In einem gemeinsamen Programm des NSF/USDA/DOE der USA wurde die Sequenzierung des Maisgenoms beschlossen ([http://www.nsf.gov/news/news\\_summ.jsp?cntn\\_id=104608](http://www.nsf.gov/news/news_summ.jsp?cntn_id=104608)). Aufgrund des 4-fach kleineren Genoms und der hervorragenden Kolinearität zum Maisgenom kommt der *Sorghum*-Hirse eine wichtige Modellfunktion in der Maisgenomforschung zu. Die Shotgun-Sequenzierung des Sorghum-Genoms ist ebenfalls bereits in der Bearbeitung (<http://www.jgi.doe.gov/sequencing/why/CSP2006/sorghum.html>). Diese Information wird eine sehr zentrale und wichtige Rolle bei der Zusammensetzung und Annotation des Maisgenoms zugemessen. Neben den Genomsequenzen aus Reis und *Brachypodium* werden die Mais- und Sorghum-Genomsequenzen einen elementaren und unmittelbaren Einfluss auf die Genomanalyse in den Triticeae-Getreidearten haben.

**Autor:** Dr. Nils STEIN, Leibniz Institut für Pflanzengenetik und Kulturpflanzenforschung Gatersleben (IPK), Corrensstraße 3, D-06466 GATERSLEBEN



Tabelle 1: Stand der Genom-Sequenzierung in verschiedenen Getreidearten

	ESTs <sup>1)</sup>	GSS <sup>2)</sup>	genomische Sequenzen(>70000 bp)	
<i>O. sativa</i>	1,211,064	252,775	> 400 Mb	full genome <sup>3)</sup>
<i>Z. mays</i>	1,160,490	2,014,160	> 400 Mb	3756 BACs <sup>4)</sup>
<i>T. aestivum</i>	855,066	7,999	9.5 Mb	71 BACs
<i>H. vulgare ssp. vulgare + spontaneum</i>	461,482	2,033	2.4 Mb	16 BACs
<i>S. bicolor</i>	204,208	573,724	4.9 Mb	36 BACs + WGS <sup>5)</sup>
<i>S. propinquum</i>	20,881	23,674	-	-
<i>B. distachyon</i>	20,449	-	0.5 Mb	3 BACs + WGS <sup>5)</sup>
<i>T. monococcum</i>	10,139	-	2.5 Mb	15 BACs
<i>S. cereale</i>	9,195	-	-	-
<i>T. turgidum ssp. durum</i>	8,924	-	2.9 Mb	18 BACs
<i>A. sativa</i>	7,632	-	-	-
<i>A. tauschii</i>	-	5,055	0.62 Mb	5 BACs

1) Expressed Sequence Tags, dbEST release 081806 und 110306 (<http://www.ncbi.nlm.nih.gov/dbEST>) 2) Genomic Survey Sequences, dbGSS release 081806 (<http://www.ncbi.nlm.nih.gov/dbGSS>) 3) [1] 4) status November 15<sup>th</sup>, 2006 5) Shotgun-Sequenzierung (WGS) durch JGI/DOE (<http://www.jgi.doe.gov/sequencing>)

## Weizen, Gerste, Roggen

Die Genomgröße sowie der sehr hohe Anteil an repetitiver DNA galt lange Zeit als unüberwindliche Hürde bei der Isolierung von Genen aus den Getreidearten Weizen, Gerste und Roggen. Nach der Schaffung genomischer DNA Bibliotheken bestehend aus Klonen mit > 100 kb DNA Insertionen/Klon (z.B. BAC Bibliotheken, Übersicht in [8]) wurden jedoch innerhalb der vergangenen 10 Jahre unter erheblichem Aufwand die ersten Gene aus Weizen und Gerste isoliert (Tabelle 2). Durch umfangreiche Sequenzierung exprimierter Gene (sogenannte EST Sequenzierung, EST = expressed sequence tag) kennt man heute einen Großteil der in Gerste und Weizen enthaltenen Gene. Diese Information wurde sehr erfolgreich zur Entwicklung Gen-basierter Marker genutzt, so dass heute für Weizen und Gerste sehr dichte Gen-Karten existieren, anhand derer man den Grad der Kolinearität zwischen Gräsergenomen neu bewerten kann [4, 9]. Die Verfügbarkeit Genom-umfassender Microarrays sowie die Entwicklung neuer Hochdurchsatz-Markertechnologien [10-12] wird es zumindest in Gerste in naher Zukunft erlauben, eine mehrere tausend Gene umfassende, genetische SNP-Marker Karte zu erstellen. Trotz der bereits enorm verbesserten genomischen Ressourcen, stellt die kartengestützte Isolierung eines Gens aus Gerste oder Weizen einen kaum vertretbaren Aufwand an Zeit und Arbeitsleistung dar. Dies wird sich erst durch eine Verfügbarkeit physischer Karten sowie einer zumindest teilweisen Sequenzierung einzelner Triticeae-Genome substantiell ändern. Phy-

Tabelle 2: Positionell klonierte Gene aus Triticeae Spezies

Gen	Funktion	Referenz
<b>Weizen (2n), <i>Triticum monococcum</i></b>		
<i>VRN1</i>	Homology to MADS-box proteins	[13]
<i>VRN2</i>	ZCCT-like transcription factor	[14]
<b>Weizen (6n), <i>Triticum aestivum</i></b>		
<i>LR10</i>	CC-NBS-LRR	[15]
<i>LR21</i>	NBS-LRR	[16]
<i>PM3</i>	NBS-LRR	[17]
<i>Q</i>	<i>APETALA (AP2)</i> -like transcription factor	[18]
<i>PH1</i>	CDC2 ?	[19]
<b>Gerste, <i>Hordeum vulgare</i></b>		
<i>MLO</i>	Seven-transmembrane protein, homology to G-protein coupled receptors	[20]
<i>RAR1</i>	Zinc-binding protein (CHORD domain)	[21]
<i>ROR2</i>	Syntaxin	[22]
<i>MLA</i>	CC-NBS-LRR	[23]
<i>RPG1</i>	Receptor kinase	[24]
<i>RYM4</i>	Translation initiation factor eIF4E	[25]
<i>VRS1</i>	Homoeobox transcription factor	[26]
<i>PPD-1</i>	Pseudo-Response-Regulator (PRR)	[27]

Tabelle aktualisiert nach [8]

sische Karten befinden sich für das Weizen D-Genom (*A. tauschii*), für einzelne Weizenchromosomen sowie für das Gerste-Genom im Aufbau. Internationale Konsortien haben sich zusammengefunden um Ihre Aktivitäten zum Erreichen der physischen Karten sowie der Genomsequenz für Triticeae-Genome effizient zu bündeln und zu koordinieren (International Wheat Genome Sequencing Consortium IWGSC, <http://www.wheatgenome.org/>; International Barley Genome Sequencing Consortium IBSC, <http://barleygenome.org/>; European Triticeae Genomics Initiative ETGI, <http://www.etgi.org/>).

## Ausblick

Der rasante Fortschritt auf dem Gebiet der Entwicklung neuer Sequenzieretechnologien [28] wird auch die Genomana-

lyse bei Getreiden entscheidend beeinflussen. Bereits heute sind Techniken verfügbar, die innerhalb weniger Stunden, Sequenzinformation im Umfang von 20-100 Millionen Nukleotiden liefern. Obwohl weitere technische Verbesserungen erforderlich sind, haben diese Techniken absehbar auch eine große Bedeutung in Fragestellungen zur Sequenzierung umfangreicher Genome mit hohem Anteil repetitiver DNA [29]. Innerhalb der nächsten 5-10 Jahre werden große Bereiche der Genome unserer wichtigsten Kulturarten teilweise oder ganz sequenziert werden/worden sein. Es ist daher an der Zeit, die unter heutigen und zukünftigen Gesichtspunkten, wichtigsten, agronomischen Merkmale zu definieren und das Pflanzenmaterial für eine präzise, hochauflösende, genetische Lokalisierung dieser meist quantitativen Merkmale zu erstellen, sowie neue Tech-

niken zur hoch-präzisen, effizienten Erfassung phänotypischer Merkmale zu entwickeln.

## Literatur

1. The Arabidopsis Genome Initiative: *Nature* 2000, **408**(6814):796-815.
2. International Rice Genome Sequencing Project: *Nature* 2005, **436**:793-800.
3. GAUT, B.S., 2002: *New Phytol*, **154**(1):15-28.
4. STEIN, N. et al., 2006: *TAG*: under revision.
5. MOORE, G. et al. 1995: *Current Biol*, **5**(7):737-739.
6. BENNETZEN, J.L. et al., 1997: *Genome Res*, **7**:301-306.
7. DRAPER, J. et al., 2001: *Plant Physiol*, **127** (4):1539-1555.
8. STEIN, N. et al., 2004: In: *Cereal Genomics*. Edited by Gupta P, Varshney R. Dordrecht, The Netherlands: Kluwer Academic Publishers, 331-360.
9. SORRELLS, M.E. et al., 2003: *Genome Res*, **13**(8):1818-1827.
10. CUI, X. et al., 2005: *Bioinformatics*, **21**(20): 3852-3858.
11. ROSTOKS, N. et al., 2005: *Genome Biol*, **6**:R54.
12. ROSTOKS, N. et al., 2006: *PNAS*:0606133103.
13. YAN, L. et al., 2003: *PNAS*, **100**(10):6263-6268.
14. YAN, L. et al., 2004: *Science*, **303**:1640-1644.
15. FEUILLET, C. et al., 2003: *PNAS*, **100**(25): 15253-15258.
16. HUANG, L. et al., 2003: *Genetics*, **164**(2): 655-664.
17. YAHIAOUI, N. et al., 2003: *Plant J.*, **37**: 528-538.
18. FARIS, J.D. et al., 2003: *Genetics*, **164**(1): 311-321.
19. GRIFFITHS, S. et al., 2006: *Nature*, **439** (7077):749.
20. BUSCHGES, R. et al. 1997: *Cell* 1997, **88**(5): 695-705.
21. SHIRASU, K. et al., 1999: *Cell* 1999, **99**(4): 355-366.
22. COLLINS, N.C. et al., 2003: *Nature*, **425** (6961):973-977.
23. WEI, F.S. et al., 1999: *Genetics*, **154**(2):1929-1948.
24. BRUEGGEMAN, R. et al., 2002: *PNAS*, **99** (14):9328-9333.
25. STEIN, N. et al., 2005: *Plant J.*, **42**(6):912-922.
26. KOMATSUDA, T.: personnel communication.
27. TURNER, A. et al., 2005: *Science*, **310**(5750): 1031-1034.
28. SERVICE, R., 2006: *Science*, **311**:1544-1546.
29. WICKER, T. et al., 2006: *BMC Genomics*, **7**:275.