

# Genomanalyse an Pflanzen - Überblick über die Thematik und Auswirkungen für die Pflanzenzüchtung

W. FRIEDT, R. SNOWDON und C. OBERMEIER

## Allgemeiner Stand der Genomforschung

In den letzten Jahren sind enorme Fortschritte in der Genomanalyse an Pflanzen gemacht worden, deren Auswirkungen auf die Pflanzenzüchtung sich u. a. in der kartengestützten Klonierung (Positionsklonierung) und Identifizierung relevanter Gene, basierend auf der vergleichenden Genomanalyse von Kulturpflanzen und Modellpflanzen zeigen. Mit der stetig und rasch zunehmenden Generierung von Nukleinsäuresequenz-Daten in öffentlichen Datenbanken für eine Vielzahl von Kulturpflanzenarten und der Verfügbarkeit neuer Hochdurchsatz-Analysetechniken darf für die nahe Zukunft eine deutliche Effizienzsteigerung bei der Identifizierung von Genen mit Bedeutung für die praktische Pflanzenzüchtung erwartet werden.

Mit Hilfe der Genomanalyse lässt sich genetische Variation bei Nutzpflanzen und ihren Verwandten direkt auf der Ebene der Erbsubstanz beschreiben (strukturelle Genomanalyse). Ferner können Genotyp- und Umwelteffekte sowie deren Wechselwirkung im Hinblick auf die Genexpression und Merkmalsausprägungen (Phänotyp) aufgedeckt werden (funktionelle Genomanalyse). Die Erforschung vollständiger Genome von Organismen und funktioneller Wechselwirkungen ihrer Gene wird auch als Genomik bezeichnet.

Die bisherige Entwicklung der Genomanalyse an Pflanzen kann in drei zeitliche Phasen gegliedert werden (BOREVITZ und ECKER 2004). Die erste Phase war noch durch eine Ein-Gen-ein-Phänotyp Sichtweise geprägt; es kamen z.B. Markeranalysen anhand von Restriction Fragment Length Polymorphisms (RFLPs) (vergl. SNOWDON und FRIEDT 2004), Einzelgen-Sequenzanalysen und Hybridisierungsanalysen un-

ter Verwendung von Arrays mit geringer Dichte (dot blots, Northern blots) zur Anwendung. In der zweiten Phase überwogen dagegen Techniken wie Single Nucleotide Polymorphisms (SNPs), Gesamt-Genomische Sequenzierungsanalysen und Hybridisierungsanalysen unter Verwendung von Arrays mit mittlerer Dichte zur Identifizierung multipler Gene, die an der Ausprägung spezifischer Phänotypen beteiligt sind. In der gegenwärtigen dritten Phase der Genomanalyse an Pflanzen ist die vergleichende gesamt-genomische Sequenzierung verwandter Kulturarten in Kombination mit Hybridisierungsanalysen unter Verwendung von Arrays mit extrem hoher Dichte für Einzelindividuen-Genotypisierungen zu beobachten.

Whole-genome arrays (WGAs) erlauben es Transkription, alternatives Splicing, Protein-DNA-Bindeverhalten und den epigenetischen Zustand des gesamten Genoms zeitlich parallel zu verfolgen um die Evolution, Dynamik und Vernetzung von Genen zu verstehen, die komplexen Phänotypen (quantitativen Merkmalen) zugrunde liegen.

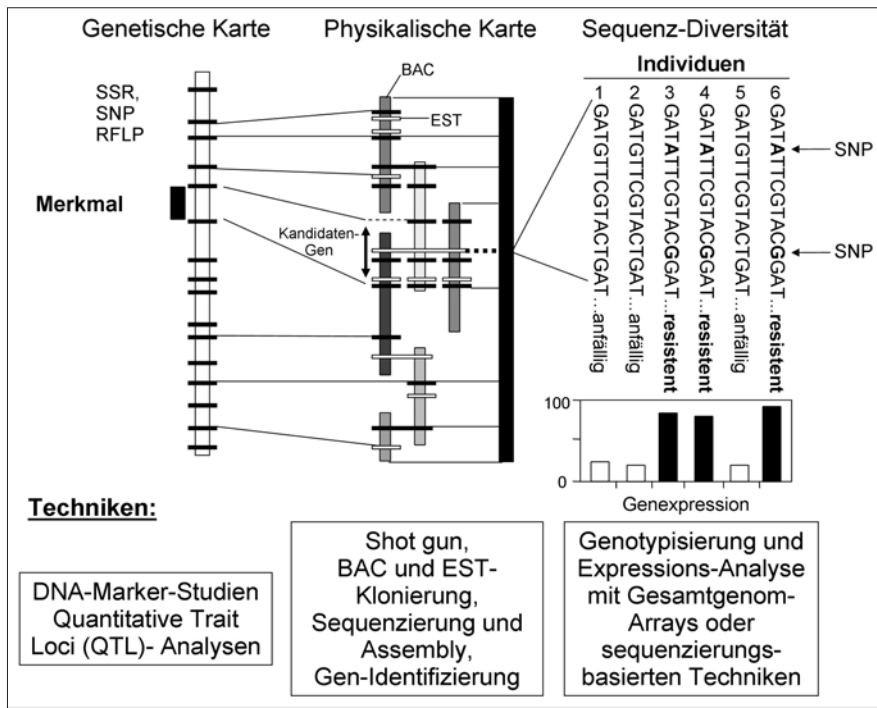
Wichtige Meilensteine in der Genomanalyse an Pflanzen stellen die Veröffentlichungen der vollständigen genomischen Nukleinsäuresequenzen der Modellpflanzen *Arabidopsis thaliana* (Ackererschmalwand) und *Oryza sativa* (Reis) in den Jahren 2000 und 2005 dar. Darüber hinaus sind in den letzten Jahren BAC- (Bacterial Artificial Chromosome) bzw. YAC- (Yeast Artificial Chromosome) Klonkollektionen und physische Karten für eine Reihe wichtiger Kulturpflanzenarten etabliert worden. Ferner wird eine rasch wachsende Zahl von Expressed Sequence Tag (EST)-Einzelsequenzen generiert (bis dato 10,2 Millionen pflanzlicher ESTs), die eine große Spannbreite von Genen umfassen, in den verschie-

densten Gewebetypen exprimiert werden und an verschiedensten Entwicklungsvorgängen, Stoffwechselwegen und der Synthese von Sekundärmetaboliten beteiligt sind. Mit Hilfe der stetig ansteigenden genomischen Datenmengen für Kulturpflanzenarten ist es in zunehmendem Maße und mit steigender Auflösung möglich, DNA-Marker aus genetischen und physischen Karten zu integrieren, funktionelle Kandidatengene zu identifizieren und eine funktionelle Kartierung von Genen für komplexe Merkmale zu erreichen.

Die in der funktionellen Genomanalyse von Pflanzen angewendeten Methoden beinhalten zwei grundlegend verschiedene Ansätze: In der klassischen Genetik (forward genetics) ist der Ausgangspunkt eine erkennbare Mutation (Mutante = Phänotyp) und das Ziel ist die genetische Aufklärung der Mutation und die molekulare Charakterisierung der (des) beteiligten Gene(s); die Analyse geht also vom Phänotyp aus und ist gerichtet auf die Aufklärung des zugrunde liegenden Genotyps. Im Gegensatz hierzu ist bei der Reversen Genetik (reverse genetics) der Ausgangspunkt ein Gen (Kandidatengen) in Form der betreffenden cDNA oder des kodierten Proteins, und das Ziel ist die Analyse der Funktion des Gens bzw. Allels, das durch gezielte Mutagenese in Zellen oder Organismen generiert worden ist. Die Analyse geht hier mithin vom Genotyp aus und ist gerichtet auf die Analyse der Funktion des veränderten Gens im Hinblick auf die Entstehung eines Phänotyps. In der vorwärtsgerichteten Genomanalyse an Pflanzen werden Methoden wie z. B. die QTL-Kartierung, die Erstellung physischer Karten, die vergleichende Genotypisierung und die Expressionsanalyse mittels Whole-genome arrays und Positionsklonierung eingesetzt (Abbildung 1).

**Autoren:** Prof. Dr. Dr. h.c Wolfgang FRIEDT, Dr. Rod SNOWDON und Dr. Christian OBERMEIER, Lehrstuhl für Pflanzenzüchtung, Interdisziplinäres Forschungszentrum für Umweltsicherung (IFZ), Justus-Liebig-Universität Giessen, Heinrich-Buff-Ring 26-32, D-35392 GIESSEN, wolfgang.friedt@agrار.uni-giessen.de





**Abbildung 1: Schematische Übersicht zur Genomanalyse.** Genetische Karte mit verschiedenen DNA-Marker-Typen (SSR: Simple Sequence Repeats, SNP: Single Nucleotide Polymorphism, RFLP: Restriction Length Fragment Polymorphism). Durch Lokalisierung einiger dieser Marker und Expressed sequence tags (ESTs) auf einer physikalischen Karte, die aus den Nucleinsäuresequenzen überlappender Bacterial Artificial Chromosome (BAC)-Klone erstellt wurden, lassen sich genetische und physikalische Karten integrieren, und Regionen mit Kandidatengenen für interessierende Merkmale können eingegrenzt werden. Die DNA-Sequenz des gesamten Genoms stellt die endgültige physikalische Karte dar. Zusätzlich zur Konsensus-Sequenz ist die Sequenz-Diversität in Kandidatengenen-Regionen von Interesse, die auf eine mögliche Korrelation mit Genexpressions-Karten und phänotypische Daten getestet werden kann (basierend auf RAFALSKI 2002, verändert).

**Tabelle 1: Genomische Ressourcen für Brassica-Arten und die Modellpflanze *Arabidopsis thaliana*;** aktuelle Anzahl von Nucleinsäuresequenzen verschiedener Kategorien in öffentlich zugänglichen Datenbanken, zusammengestellt aus: National Center for Biotechnology Information (NCBI), European Molecular Biology Laboratory (EMBL), The Arabidopsis Information Resource (TAIR), BrassicaDB und Brassica Genome Gateway 2006 (Berücksichtigt wurden Daten für *B. napus*, *B. rapa*, *B. oleracea* und *B. nigra*, Stand Oktober 2006).

Kategorie	Brassicaceae	<i>A. thaliana</i>
BACs (Bacterial Artificial Chromosomes)	533	1.730
ESTs (Expressed Sequence Tags)	129.830	622.972
GSSs (Genome Survey Sequences)	808.039	440.439
Maps (Marker based genetic linkage)	102	462
RFLPs (Restriction Fragment Length Polymorphisms)	1.539	644
SSRs (Simple Sequence Repeats)	404	322
Sequences (genomic+EST+GSS nucleotide)	946.729	1,229.207

Für die Reverse Genetik stehen z.B. bei *Arabidopsis* umfangreiche Kollektionen von Transposon-Tagging Mutanten und TILLING-(Targeting Induced Local Lesions IN Genomes) Populationen zur Verfügung, die mittels vergleichender Genomikansätze die Analyse der Funk-

tion von Kandidatengenen an Kulturpflanzen ermöglichen. Eine weitere, zunehmend häufiger genutzte Methode der Reversen Genetik bei Modell- und Kulturpflanzen-Arten ist die Methode des Gene silencing (RNA interference) ausgewählter Kandidatengene mit dem Ziel

der Aufklärung ihrer Funktion. Die genetische Kartierung von Merkmalen, gefolgt von funktionellen biochemischen und zellbiologischen Analysen bleibt essentiell für ein umfassendes Verständnis der zugrunde liegenden gen-regulatorischen Netzwerke und die Nutzung ausgewählter Gene für die praktische Pflanzenzüchtung. In der näheren Zukunft ist die Integration von unterschiedlichen Arten genomischer Daten und die Entwicklung leistungsfähiger bioinformatischer Methoden ein wichtiges Ziel, um eine effiziente Nutzung der verfügbaren und stetig steigenden Datenmengen für die Pflanzenzüchtung zu erreichen.

### Stand der Genomforschung bei Ölpflanzen - Beispiel Brassica-Arten

Als Ergebnis einer natürlichen Kreuzung zwischen Rübren (*Brassica rapa*) und Kohl (*B. oleracea*) gehört Raps (*B. napus*) gemeinsam mit weiteren Wild- und Nutzpflanzen der Gattung Brassica zur Familie der Kreuzblütler (Brassicaceae). Für die Genomforschung bei diesen Spezies ist es von großem Nutzen, dass auch die Modellpflanze *Arabidopsis thaliana* mit ihrem vergleichsweise kleinen Genom (ca. 130 Mbp) zu dieser Familie gehört und daher für Ansätze der vergleichenden Genomik hier ideal eingesetzt werden kann. Durch wiederholte Polyploidisierung im Lauf der Evolution sind die Brassica-Genome deutlich größer (*B. napus* ca. 1200 Mbp). So war es sehr nützlich, dass bereits im Jahre 2000 das *Arabidopsis*-Genom vollständig sequenziert wurde. Darüber hinaus werden in rascher Folge weitere genomische Ressourcen verfügbar gemacht, die zum großen Teil in öffentlichen Datenbanken frei zugänglich sind (Tabelle 1).

Zudem werden in jüngster Zeit aber auch solche Ressourcen für die Brassica-Arten selbst erarbeitet, wobei derzeit *B. rapa* (Chinakohl, Rübren) eine Vorreiterrolle zukommt (LIM et al. 2006). Aus der rasch wachsenden Datenmenge ergeben sich zunehmend bessere Möglichkeiten für die vergleichende Genomforschung bei Raps unter Nutzung der Erkenntnisfülle aus der *Arabidopsis*-Forschung. Insbesondere hinsichtlich ertragssichernder Merkmale (Wuchstyp, Entwicklungsverlauf, Krankheits- und

**Tabelle 2: Ausgewählte Beispiele nutzbarer molekularer Marker für die Rapszüchtung**

Merkmalskomplex	Spezifische Eigenschaft(en)
Pflanzentyp	Wuchshöhe (Zwergwuchs) Blühzeit (Beginn, Dauer) Blütenblattlosigkeit Samenfarbe (schwarz bzw. gelb)
Cytoplasmatische Männliche Sterilität (CMS)	Ogura CMS Polima CMS Fertilitätsrestoration
Toleranz gegen abiotischen Stress	Kälteresistenz, Winterhärte
Krankheitsresistenz	Phoma-Resistenz Resistenz gegen Weißrost Virus-Resistenz (TMV, TuYV)
Samenölqualität	Erucasäure-, Öl-, Linol-, Linolensäure-Gehalt Glucosinolat-Gehalt und Muster Faser-, Lignin- und Tannin-Gehalt

Stress-Resistenz) und Qualitätseigenschaften (Öl- und Schrotqualität) kann hieraus besonderer Nutzen gezogen werden.

### Anwendung von Ergebnissen der Genomforschung in der Rapszüchtung

Bereits seit einiger Zeit werden molekulare Markertechniken für die Erstellung genetischer Karten und die Identifikation molekularer Marker für wesentliche Nutz- und Verwertungseigenschaften in der Rapszüchtung eingesetzt (SNOWDON and FRIEDT 2004). Allerdings ist die Zahl molekularer Marker, die auch in unterschiedlichem Zuchtmaterial eingesetzt werden können, relativ überschaubar (ausgewählte Beispiele: *Tabelle 2*). Hier erhofft sich die Züchtungspraxis zweifellos noch größere Fortschritte.

Neben der Menge und Qualität des Rapsöls gewinnt die Beschaffenheit des Rapsextraktionsschrotes (RES) eine zunehmende Bedeutung für die Wertschöpfung der Rapssaat. Das RES hat aufgrund eines relativ hohen Proteingehalts verbunden mit einer günstigen Aminosäurezusammensetzung prinzipiell einen hohen Futterwert. Aber unerwünschte

Begleitstoffe (sekundäre Inhaltsstoffe) stören die Futter-Aufnahme und Verwertung: Einerseits wirken sich Glucosinolate bei der Verfütterung an monogastrische Nutztiere (Hühner, Schweine) ungünstig aus. Andererseits bedingt der hohe Rohfasergehalt (Lignin u.a.) von RES aus schwarzer Rapssaat nicht allein eine geringe Energiedichte, sondern weitere Rohfaserkomponenten (z.B. Tannine) beeinträchtigen die Proteinverwertung durch das Tier, aber auch die angestrebte Gewinnung von Protein-Isolaten und Konzentraten zum Zwecke der Humanernährung.

Sowohl beim Glucosinolat- als auch beim Rohfasergehalt handelt es sich um quantitative Merkmale. Aus klassisch-genetischen und QTL-Analysen ist seit längerem bekannt, dass beide Merkmalskomplexe durch eine Reihe von Genloci kontrolliert werden (vgl. BADANI et al. 2006). Da die betreffenden Gene in *Arabidopsis* bereits kloniert wurden und auch vergleichende *Brassica/Arabidopsis*-Genomkarten schon vorliegen (vgl. PARKIN et al. 2005), können nunmehr anhand vergleichender Genomik die betreffenden Gene in *Brassica* identifiziert und funktionell charakterisiert werden. Zunächst geben QTL Hinweise auf die

Lokalisation entsprechender Gene im Rapsgenome. Sodann können anhand vergleichender Karten und geeigneter Marker (RFLPs, SSRs) die zugehörigen Gene in *Arabidopsis* auffindig gemacht werden. Entsprechende *Arabidopsis*-Sequenzen dienen dann für die Identifikation analoger Rapssequenzen und die Erstellung von Contigs auf dem Weg zur Klonierung des jeweiligen vollständigen Gens aus Raps.

### Danksagung

Wir bedanken uns bei der DFG und dem BMBF für die großzügige Förderung unserer Forschungsarbeiten. Besonderer Dank gilt allen Mitarbeiterinnen und Mitarbeitern am IPZ Giessen für engagierte und tatkräftige Mitwirkung in den verschiedenen Genomforschungsprojekten.

### Literatur

- BADANI, A.G., R.J. SNOWDON, B. WITTKOP, F.D. LIPSA, R.BAETZEL, R. HORN, A.G. DE HARO, R. FONT, W. LUEHS and W. FRIEDT, 2006: Co-localisation of a partially dominant gene for yellow seed colour with a major QTL influencing acid detergent fibre (ADF) content in different crosses of oilseed rape (*Brassica napus*). *Genome* (im Druck).
- BOREVITZ, J.O. and J.R. ECKER, 2004: Plant genomics: The third wave. *Annu. Rev. Genomics Hum. Genet.* 5:443-77.
- LIM, Y.P., P. PLAHA, S.R. CHOI, T. UHM, C.P. HONG, J.W. BANG and Y.K. HUR, 2006: Toward unraveling the structure of *Brassica rapa* genome. *Physiologia Plantarum* 126:585-591.
- PARKIN, I.A., S.M. GULDEN, A.G. SHARPE, L. LUKENS, M. TRICK, T.C. OSBORN and D.J. LYDIATE, 2005: Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 171:765-781.
- RAFALSKI, J.A., 2002: Plant genomics: Present state and a perspective on future developments. *Briefings in Functional Genomics and Proteomics* 1: 80-94.
- SNOWDON, R.J. and W. FRIEDT, 2004: Molecular markers in *Brassica*/oilseed breeding: Current status and future possibilities. *Plant Breeding* 123:1-8.