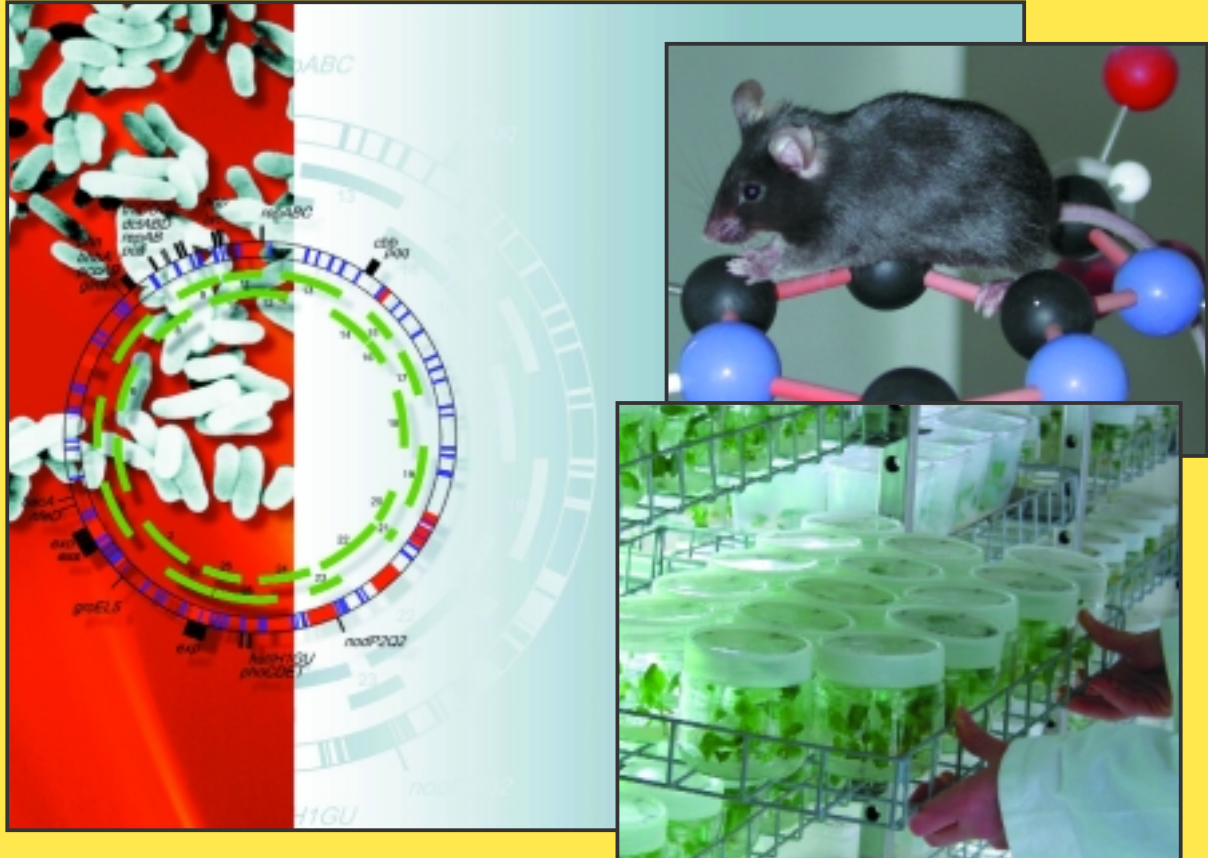


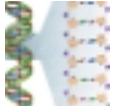
Forschung

an der Universität Bielefeld



Zukunftswissenschaft Genomforschung

- Vom Gen zum Genom zur Systembiologie
- Genomforschung und Bioinformatik
- Transkriptomforschung: Mikroarrays helfen bei der Aufklärung komplexer Genregulationsnetzwerke
- Proteomforschung: Vom Gen zur Funktion
- Stickstoff aus der Luft gegriffen
- Bakterien für die industrielle Produktion
- Das Genomprojekt *Clavibacter michiganensis*
- Optimierung von proteinreichen Samen aus Hülsenfrüchtlern
- Vom Aschenputtel zur Modellpflanze: Das nützliche Unkraut
- Wie menschlich ist die Maus?
- Finden, fast ohne zu suchen: Indexstrukturen in der Bioinformatik



S. 3

Vom Gen zum Genom zur Systembiologie

Entwicklung der bakteriellen Genomforschung
an der Universität Bielefeld

Alfred Pühler



S. 9

Genomforschung und Bioinformatik

Die Bielefelder Plattform zur Genomanalyse

Olaf Kaiser,
Folker Meyer



S. 14

Transkriptomforschung: Mikroarrays helfen bei der Aufklärung komplexer Genregulationsnetzwerke

Anke Becker



S. 18

Proteomforschung: Vom Gen zur Funktion

Mit der Sequenzierung ganzer Genome liegt das Buch auf dem Tisch –
nun müssen wir lernen, es zu lesen

Karsten Niehaus

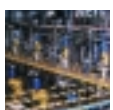


S. 23

Stickstoff aus der Luft gegriffen

Die Entschlüsselung des Genoms von *Sinorhizobium meliloti*
eröffnet neue Forschungsrichtungen

Stefan Weidner



S. 29

Bakterien für die industrielle Produktion

Genomforschung am Beispiel des Aminosäuren-produzierenden
Bakteriums *Corynebacterium glutamicum*

Jörn Kalinowski



S.33

Das Genomprojekt *Clavibacter michiganensis*

Neue Perspektiven zum Verständnis der Welkekrankheit der Tomate

Karl-Heinz Gartemann,
Rudolf Eichenlaub



S. 39

Optimierung von proteinreichen Samen aus Hülsenfrüchtlern

Genomforschung bei Pflanzen aus der Familie der Leguminosen

Helge Küster



S. 47

Vom Aschenputtel zur Modellpflanze: Das nützliche Unkraut

Methoden und Ziele der Genomforschung bei Pflanzen

Martin Sagasser,
Bernd Weisshaar



S.53

Wie menschlich ist die Maus?

Funktionelle Genomforschung am Modellorganismus der Biomedizin

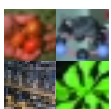
Thomas Schmitt-John,
Harald Jockusch



S. 63

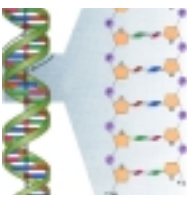
Finden, fast ohne zu suchen: Indexstrukturen in der Bioinformatik

Robert Giegerich,
Jens Stoye



S. 69

Abstracts



Vom Gen zum Genom zur Systembiologie

Entwicklung der bakteriellen Genomforschung an der Universität Bielefeld

Alfred Pühler

Fakultät für Biologie

■ Von den Anfängen der DNA-Sequenzierung zur Sequenzermittlung gesamter bakterieller Genome

Im Jahr 1953, also genau vor einem halben Jahrhundert, wurde das Zeitalter der Molekulargenetik eingeläutet. Watson und Crick publizierten damals die molekulare Struktur des Erbmoleküls Desoxyribonukleinsäure, die auch schlicht DNA genannt wird. Bei dieser Struktur handelt es sich um die berühmte DNA-Doppelhelix, so wie sie in Abbildung 1 dargestellt ist. Rein chemisch lässt sich die DNA als Leitermodell beschreiben. Die Holme bestehen aus Zucker-Phosphat-Ketten, während die Sprossen der Leiter von Basenpaaren gebildet werden. Wird dieses Leitermodell nun in sich verdreht, so entsteht die bekannte DNA-Doppelhelix.

Die Frage nach dem Informationsgehalt der DNA lässt sich anhand des Modells einfach erklären. Man muss lediglich wissen, dass in dem DNA-Molekül vier unterschiedliche Basen vorkommen, nämlich Adenin (A), Guanin (G), Cytosin (C) und Thymin (T). Der Informationsgehalt liegt nun in der Abfolge dieser vier Basen. Die genetische Information jedes beliebigen Organismus ist damit durch die Basensequenz seiner DNA festgelegt. Die Basensequenz wird durch Verwendung der Buchstaben A, G, C und T als lange Buchstabenkette dargestellt.

Nachdem das Konzept des Informationsgehaltes der DNA geklärt war, konzentrierten sich viele Gruppen auf die Entwicklung von Methoden zur Bestimmung der Basenabfolge definierter DNA-Fragmente. Im Jahr 1977 erfolgte der Durchbruch mit zwei verschiedenen Ansätzen. Maxam und Gilbert publizierten ein DNA-Sequenzierverfahren nach der chemischen Degradationsmethode, und Sanger beschrieb ein Verfahren nach der Ketten-

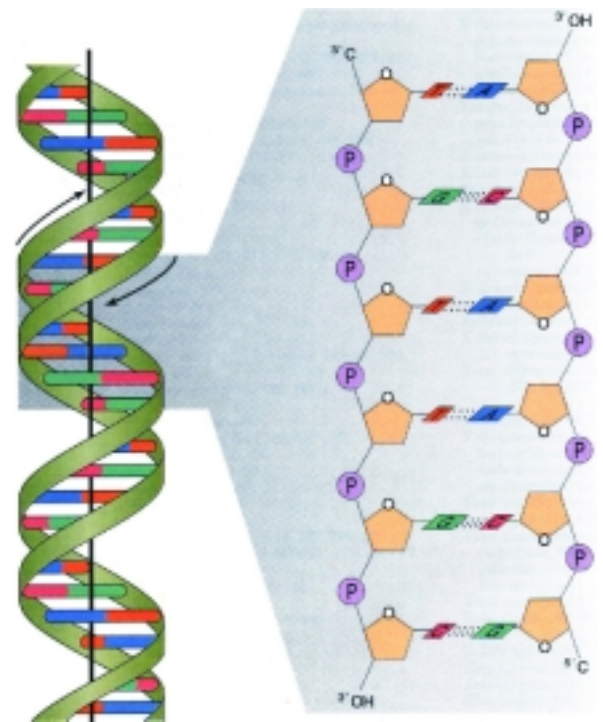


Abbildung 1: Die Struktur des Erbmoleküls
Dargestellt ist die von Watson und Crick im Jahr 1953 beschriebene DNA-Doppelhelix. Ein Ausschnitt dieser DNA-Doppelhelix ist als Leitermodell wiedergegeben. Die Zucker-Phosphat-Ketten bilden die Holme, die Basenpaare die Sprossen. Die zwei möglichen Basenpaarungen, nämlich A=T und G≡C, sind eingezeichnet.

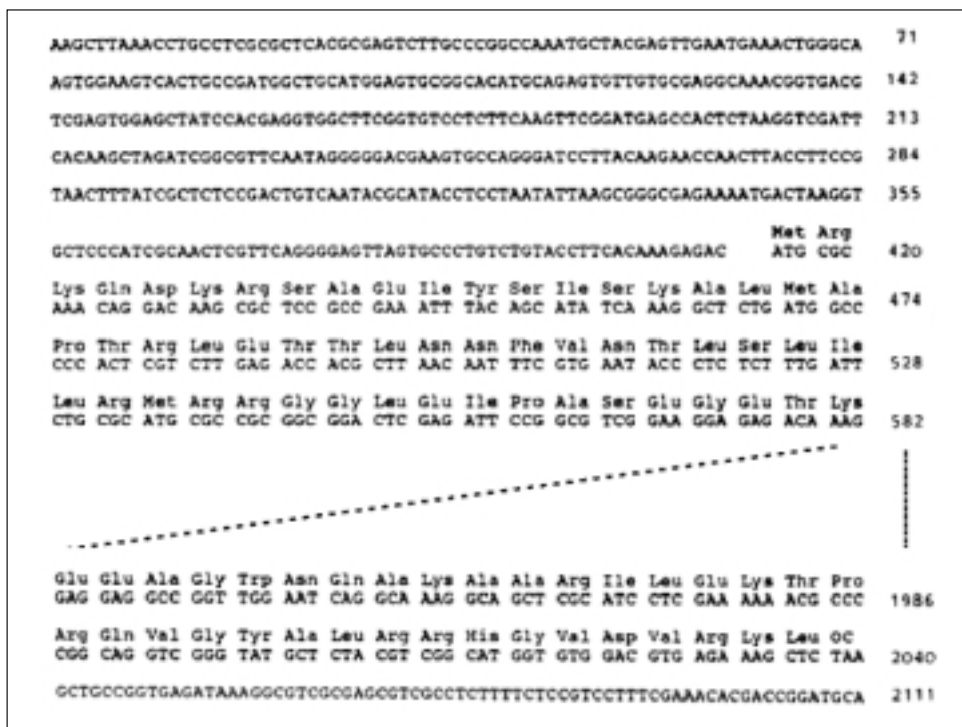


Abbildung 2: Identifizierung eines Regulorgens aus dem Bodenbakterium *Rhizobium meliloti*. Insgesamt wurden 2.111 Basen eines DNA-Strangs identifiziert. Auf diesem DNA-Strang liegt ein Gen, das für ein Regulatorprotein von 542 Aminosäuren kodiert. Dargestellt ist als Buchstabenkette der Beginn und das Ende des sequenzierten DNA-Strangs. Über den entsprechenden Triplets sind die kodierten Aminosäuren des identifizierten Genprodukts angegeben. Der große Mittelteil des DNA-Strangs ist aus Platzgründen nicht gezeigt.

abbruchmethode. Zunächst setzte sich das chemische Degradationsverfahren durch. In der Zwischenzeit wurde dieses Verfahren aber komplett vom Kettenabbruchverfahren verdrängt, da sich dieses automatisieren ließ.

An der Universität Bielefeld wurden die Methoden der DNA-Sequenzierung ab dem Jahr 1980 etabliert. Die erste bedeutende Publikation der Bielefelder Genetik erschien 1985 im EMBO-Journal. Sie handelt von der Sequenzermittlung eines interessanten Regulorgens des Bodenbakteriums *Rhizobium meliloti*. In Abbildung 2 sind die Sequenzierungsergebnisse wiedergegeben. Insgesamt wurde die Sequenz von 2111 Basen eines ausgewählten DNA-Fragments ermittelt. Eine Auswertung dieser Sequenz ergab, dass dieses DNA-Fragment ein Gen mit einer Länge von 1626 Basenpaaren trägt, das für ein Protein von 542 Aminosäuren kodiert. Diese Auswertung basiert auf der genetischen Kode-Tabelle, die jedem Basen-triplett eine Aminosäure zuordnet. Über einfache Methoden kann also aus einer vorliegenden Basensequenz die zugrundeliegende Genstruktur ermittelt werden.

Es dauerte nicht lange, bis man sich auch in Bielefeld an die Sequenzierung deutlich größerer DNA-Fragmente wagte. Im Jahr 1988 publizierten W. Arnold und Mitarbeiter im *Journal of Molecular*

Biology eine Arbeit mit dem Titel „Nucleotide sequence of a 24,206 base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*“. Auf diesem mehr als 24 Kilobasenpaare großen DNA-Fragment konnten dann immerhin schon 19 Gene identifiziert werden, die alle an der biologischen Stickstoff-Fixierung des Bodenbakteriums *Klebsiella pneumoniae* beteiligt sind.

Natürlich war es eines der Ziele der Bielefelder Sequenziergruppe, den Sprung von der Sequenzierung größerer DNA-Fragmente zur Sequenzierung ganzer bakterieller Genome zu schaffen. Im Jahre 2001 war es dann so weit. Das Genom des symbiontisch Stickstoff-fixierenden Bodenbakteriums *Sinorhizobium meliloti* wurde unter wesentlicher Mithilfe der Bielefelder Sequenziergruppe entschlüsselt. Das Genom, das aus einem Chromosom und zwei Megaplasmiden besteht, ist 6,7 Megabasenpaare groß. Insgesamt wurden auf diesem Genom 6204 Gene identifiziert. Über Einzelheiten der Sequenzierung sowie über das internationale Konsortium, das an dieser Arbeit beteiligt war, wird in diesem Heft separat berichtet.

Ein zweites Genomprojekt wurde von der Bielefelder Sequenziergruppe gerade abgeschlossen. Das Genom des industriell zur Aminosäureproduktion

verwendeten Stammes *Corynebacterium glutamicum* liegt sequenziert vor. Auf dem Genom, das eine Größe von 3,3 Megabasenpaare hat, wurden 3 002 Gene identifiziert. Die dazu verfasste Publikation ist 2003 im *Journal of Biotechnology* erschienen. Auch diesem Projekt, insbesondere den erzielten Ergebnissen, ist in diesem Heft ein separater Artikel gewidmet.

■ Bündelung der Bielefelder Genomforschungsaktivitäten in einem neu gegründeten Institut

Die Erfolge der Bielefelder Genetik auf dem Gebiet der Genomsequenzierung haben dazu geführt, dass zu Beginn dieses Jahres am Centrum für Biotechnologie ein Institut für Genomforschung gegründet wurde, das die verschiedenen Aspekte der Genomforschung zusammenfasst. In diesem Institut ist zunächst die Abteilung für Sequenzermittlung angesiedelt, die unter Einsatz von Sequenzierautomaten die Hochdurchsatzsequenzierung bakterieller Genome betreibt. Aber auch die zukunftsweisenden Methoden der Postgenomforschung, nämlich Transkriptom- und Proteomanalyse, sind Bestandteile des Instituts.

In der Abteilung für Transkriptomanalyse werden hochparallel alle aktiven Gene eines Organismus in Abhängigkeit von Umweltparametern ermittelt. Hierzu werden mittels Array-Technologie die Transkripte aktiver Gene des zu analysierenden Organismus bestimmt. Die Array-Technologie muss an dieser Stelle nicht näher erläutert werden, da ihr in diesem Heft ein Artikel gewidmet ist.

Eine weitere Abteilung des Instituts für Genomforschung beschäftigt sich mit der Analyse sämtlicher Proteine eines Organismus. Hierzu werden die isolierten Proteine zweidimensional aufgetrennt und einzelne Proteinspots über ein komplexes Verfahren unter Einbeziehung der Massenspektrometrie dem zugrunde liegenden Gen zugeordnet. Auch die Proteomanalyse wird in diesem Heft in einem separaten Artikel ausführlich beschrieben.

Einen Einblick in die apparative Ausstattung des Instituts für Genomforschung liefern zwei Abbildungen. Abbildung 3 zeigt einen Pipettierroboter, wie er zur Hochdurchsatzsequenzierung von Genomprojekten benötigt wird. In Abbildung 4 ist ein Massenspektrometer aus der Proteomabteilung abgebildet.

Besonders hervorgehoben werden muss noch die Abteilung für Bioinformatik des Centrum für Biotechnologie, die in der Zwischenzeit in „Bioinformatics Resource Facility“ umbenannt wurde. Diese Abteilung soll die große Datenflut bändigen, die sowohl aus Genomprojekten als auch aus Transkriptom- und Proteomanalysen erwächst. Die Bioinformatics



Abbildung 3: Verwendung eines Pipettierroboters für die Probenvorbereitung eines Sequenzierlaufs.



Abbildung 4: Einsatz eines Massenspektrometers in der Proteomforschung.

Resource Facility hat als Aufgabe, diese Datenflut zu erfassen, sie aber auch aufzuarbeiten und auszuwerten. Für die Annotation bakterieller Genome, worunter man die Generkennung und die Funktionsvorhersage versteht, wurde zum Beispiel das Programmpaket GenDB entwickelt, das bereits erfolgreich bei der Bearbeitung der *S. meliloti*- und *C. glutamicum*-Genomprojekte eingesetzt wurde. Insgesamt werden mit GenDB zur Zeit weltweit mehr als 20 bakterielle Genomprojekte gerechnet.

Tabelle 1: Aktuelle Forschungsprojekte am Institut für Genomforschung.

Förderorganisation	Projektbezeichnung
Deutsche Forschungsgemeinschaft (DFG)	Initiative „Bioinformatik und Genomforschung“
Deutsche Forschungsgemeinschaft (DFG)	Molekulare Grundlage der Mykorrhiza-Symbiosen (MolMyk)
Volkswagenstiftung (VW)	Genomics for bean productivity (Bean Genomics)
Bundesministerium für Bildung und Forschung (BMBF)	Genomforschung an Bakterien für den Umweltschutz, die Landwirtschaft und die Biotechnologie (GenoMik)
Europäische Union (EU)	Integrated structural, functional and comparative genomics of the model legume <i>Medicago truncatula</i>
Europäische Union (EU)	Construction of <i>Corynebacterium glutamicum</i> strains producing either L-valine or D-pantothenic acid – a rational approach using genome research

Das Institut für Genomforschung hat in der Zwischenzeit mit großem Erfolg eine Reihe von Forschungsprojekten eingeworben. Wie man Tabelle 1 entnehmen kann, treten als Forschungsförderer die Deutsche Forschungsgemeinschaft (DFG), die Volkswagenstiftung (VW), das Bundesministerium für Bildung und Forschung (BMBF) und die Europäische Union (EU) auf. Unter den Projekten befinden sich nicht nur bakterielle Genomprojekte. Das Mykorrhiza-Projekt der DFG und das Medicago-Projekt der EU widmen sich zum Beispiel dem Genom von *Medicago truncatula*, einer Pflanze, die wegen ihrer symbiontischen Stickstoff-Fixierung weltweit analysiert wird. Einzelheiten zum Mykorrhiza-Projekt werden in einem weiteren Artikel dieses Heftes detailliert geschildert. Das wohl umfangreichste Forschungsprojekt ist unter dem Namen „Genomforschung an Bakterien für den

Umweltschutz, die Landwirtschaft und die Biotechnologie“ aufgelistet.

■ Kompetenzzentrum der Universität Bielefeld steuert deutschlandweites Netzwerk für bakterielle Genomforschung

Das Bundesministerium für Bildung und Forschung fördert ein von der Universität Bielefeld koordiniertes Kompetenznetzwerk mit dem Thema „Genomforschung an Bakterien für den Umweltschutz, die Landwirtschaft und die Biotechnologie“. Das Kompetenznetzwerk bündelt deutschlandweit die Forschungsaktivitäten von Forschergruppen aus zwölf Universitäten, drei Großforschungseinrichtungen und zwei Firmen (Abbildung 5). Hauptaufgabe des Netzwerks ist es, die bakterielle Genomforschung auf den Gebieten Umweltschutz, Landwirtschaft und

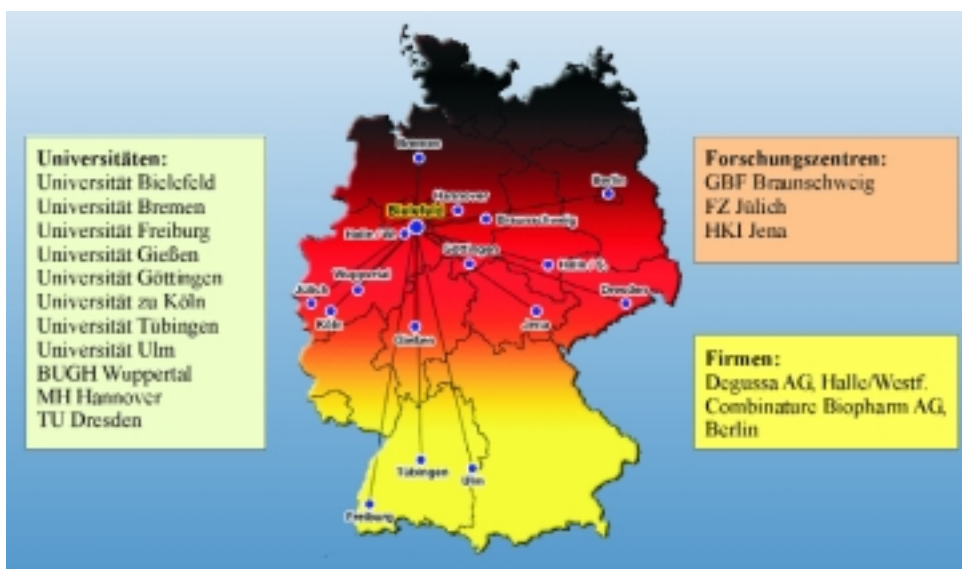


Abbildung 5:
 Die Struktur des Bielefelder Kompetenznetzwerks „Genomforschung an Bakterien für den Umweltschutz, die Landwirtschaft und die Biotechnologie“.

Biotechnologie wissenschaftlich voranzubringen und die Ergebnisse wirtschaftlich zu nutzen. Das an der Universität Bielefeld angesiedelte Kompetenzzentrum koordiniert das Kompetenznetzwerk und beteiligt sich an der Entwicklung von Technologien der Genom- und Postgenomforschung bei Bakterien.

Im Mittelpunkt der Arbeiten des Netzwerks steht die Sequenzermittlung von fünf bakteriellen Genomen, die zusammen einen Sequenzumfang von rund 30 Megabasenpaaren aufweisen. Man kann davon ausgehen, dass auf diesen fünf bakteriellen Genomen insgesamt fast 30 000 Gene identifiziert werden, also fast soviele Gene wie sie für das menschliche Genom vorhergesagt wurden. Die einzelnen Sequenzierprojekte sind in Tabelle 2 aufgelistet.

Bei *Azoarcus* sp. handelt es sich um ein Stickstoff-fixierendes Bakterium, das in der Lage ist, Kallagras und Reis mit gebundenem Stickstoff zu versorgen. Im Weiteren sind zwei pflanzenpathogene Bakterien aufgeführt. *Xanthomonas campestris* pv. *vesicatoria* befällt Paprikapflanzen, während *Clavibacter michiganensis* subsp. *michiganensis* vor allem Tomaten infiziert. Von der Sequenzierung der Genome dieser drei Bakterien erhofft man sich neue Erkenntnisse auf dem Gebiet der Bakterien-Pflanzen-Interaktion, die Einfluss auf landwirtschaftliche Praktiken nehmen.

Auf dem Umweltsektor ist die Sequenzierung eines ölabbauenden Bakteriums geplant. *Alcanivorax borkumensis* ist als obligater Kohlenwasserstoff-verwerter besonders für den Erdölabbau in marinen Habitaten geeignet. Auf dem Gebiet der Biotechnologie soll schließlich eines der größten bakteriellen Genome, nämlich das des Myxobakteriums

Sorangium cellulosum einer Sequenzanalyse unterzogen werden. Myxobakterien besitzen die Fähigkeit, biologisch aktive Wirkstoffe zu bilden, und sind daher von großer medizinischer Bedeutung.

Die in Tabelle 2 aufgelisteten Genomprojekte stehen praktisch alle vor dem Abschluss. In der Zwischenzeit wird über eine Fortsetzung des bakteriellen Genomforschungsnetzwerks nachgedacht. Es bietet sich natürlich an, dass nach Vorliegen der Genomdaten jetzt vor allem die Postgenomforschung in den Mittelpunkt rückt. Es ist vorgesehen, für die aufgelisteten Stämme sowohl in die Transkriptom- als auch in die Proteomforschung einzusteigen.

■ Bakterielle Genomforschung auf dem Weg zur Systembiologie

Die gesamte Biologie, insbesondere die Mikrobiologie, steht zur Zeit vor einem Paradigmenwechsel. In Zukunft wird man bei ausgewählten Organismen von der Analyse einzelner Gene zur Analyse der Gesamtheit aller Gene übergehen. Möglich wird dies durch die Entwicklung hochparalleler Methoden. Wie in diesem Artikel dargestellt, kann die Sequenzierung bakterieller Genome in der Zwischenzeit fast schon als Routinevorgang bezeichnet werden. Mittels Array-Technologie lässt sich über Transkriptanalysen die Gesamtheit aller Gene identifizieren, die unter bestimmten Umweltbedingungen aktiv sind. Auch die Gesamtheit aller Proteine einer Zelle lässt sich mit den Methoden der Proteomanalyse identifizieren. Zusätzlich werden zur Zeit Methoden entwickelt, die der Erfassung aller „Metabolite“ (kleine Stoffwechselprodukte) einer Zelle dienen.

Tabelle 2: Genomprojekte des Bielefelder Kompetenznetzwerks.

Bakterienspezies	Genomgröße	Projektleiter
<i>Azoarcus</i> sp. BH72	ca. 4,6 Mb	Prof. Dr. B. Reinhold-Hurek, Universität Bremen
<i>Clavibacter michiganensis</i> subsp. <i>michiganensis</i>	ca. 3,6 Mb	Prof. Dr. R. Eichenlaub, Universität Bielefeld
<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i>	ca. 5,4 Mb	Prof. Dr. U. Bonas, Universität Halle
<i>Alcanivorax borkumensis</i> SK2	ca. 3,2 Mb	Prof. Dr. K.N. Timmis, GBF Braunschweig
<i>Sorangium cellulosum</i> So ce 56	ca. 12,2 Mb	PD Dr. R. Müller, GBF Braunschweig

Die Gesamtsicht dieser hochparallelen Methoden wird in Abbildung 6 für eine Bakterienzelle wiedergegeben. Diese holistische Betrachtungsweise führt zwangsweise zur Systembiologie, die in Zukunft über „in silico-“ (d.h. computersimulierte) Vorhersagen biologischer Phänomene den Weg zu einer quantitativen prädiktiven Mikrobiologie ebnen soll. Dieser zukunftsorientierte Forschungsansatz wird augenblicklich in Bielefeld von Wissenschaftlern aus den Fachrichtungen Biologie, Chemie, Informatik, Mathematik und Physik intensiv diskutiert. Geplant ist eine Zusammenarbeit, die fakultätsübergreifend am Centrum für Biotechnologie angesiedelt sein soll. Als erster Schritt wurde im April dieses Jahres ein Internationaler Workshop abgehalten, der das vorhandene Wissen auf dem Gebiet der Systembiologie zusammenfasste. Man kann gespannt in die Zukunft blicken und hoffen, dass die Universität Bielefeld bei der Entwicklung der Systembiologie eine gewichtige Rolle spielen wird.

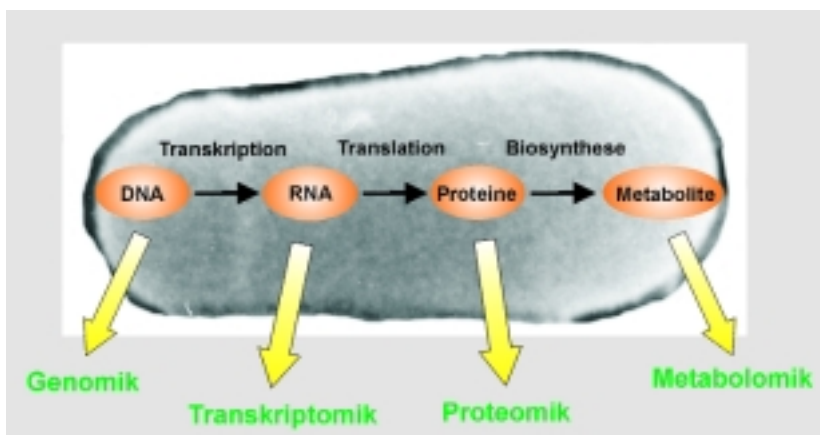
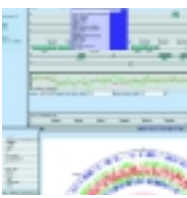


Abbildung 6: Systembiologie am Beispiel eines Bakteriums. Genomik, Transkriptomik, Proteomik und Metabolomik liefern die Datensätze zur Entwicklung der Systembiologie.



Alfred Pühler, Jahrgang 1940, absolvierte seine universitäre Ausbildung an der Friedrich-Alexander-Universität Erlangen-Nürnberg. Er studierte zunächst Physik, promovierte 1971 in Biologie und habilitierte sich 1976 in Mikrobiologie (Genetik). Im Jahre 1979 erhielt er den Ruf auf eine Professur am Lehrstuhl für Genetik an der Universität Bielefeld. Von 1992 bis 1994 bekleidete er das Amt des Prorektors für Forschung und wissenschaftlichen Nachwuchs an der Universität Bielefeld. Seit 1995 steht er dem Fachausschuss Biotechnologie der DECHEMA e.V. in Frankfurt vor. Der Zentralen Kommission für die Biologische Sicherheit am Robert Koch-Institut in Berlin gehört er seit 1995 an. Im Jahre 1999 wurde er in den Wissenschaftsrat berufen. Er ist Mitglied der Deutschen Akademie der Naturforscher Leopoldina und der Nordrhein-Westfälischen Akademie der Wissenschaften. In jüngster Zeit hat er an der Universität Bielefeld ein Zentrum für Genomforschung aufgebaut, das sich besonders der Genomanalyse von Mikroorganismen widmet.



Die Bielefelder Plattform zur Genomanalyse

Olaf Kaiser und
Folker Meyer

Fakultät für Biologie

Obwohl bereits 1995 das erste bakterielle Genom – des Krankheitserregers *Haemophilus influenzae* – sequenziert wurde und mittlerweile die Genome von annähernd 150 weiteren Bakterien vorliegen, ist Genomforschung an Bakterien auch heute noch ein aufwändiger und kostenintensiver Forschungsbereich. Um die bei der Sequenzierung eines Genoms anfallenden Datenmengen bewältigen und auswerten zu können, ist eine enge Zusammenarbeit von Biologen und Informatikern essentiell. Eine solche Zusammenarbeit wird an der Universität Bielefeld erfolgreich im Rahmen mehrerer Genomprojekte durchgeführt.

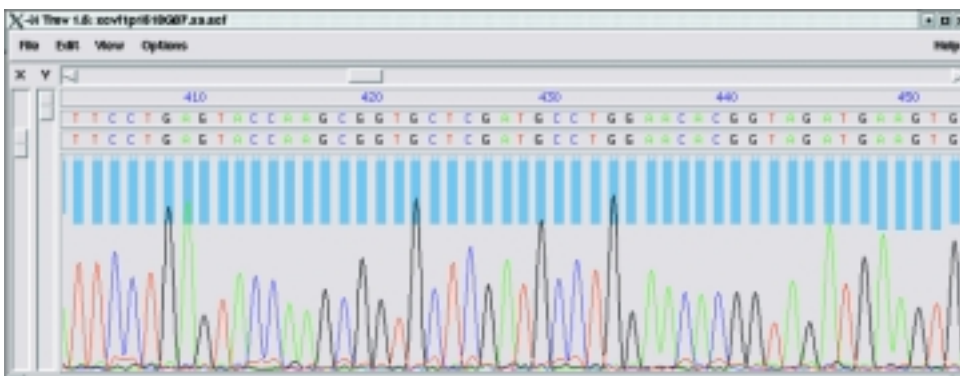


Abbildung 1
DNA-Sequenzinformation als Standardchromatogramm (SCF-File). Bei der Sequenzerstellung an modernen Sequenzierautomaten wird die Abfolge der vier Basen A (Adenin), C (Cytosin), G (Guanin) und T (Thymin), welche der genetische Code umfasst, ermittelt. Die Abfolge dieser vier Basen (Sequenz) kann in einem Chromatogramm dargestellt werden.

Die Genetik befasst sich bereits seit einigen Jahrzehnten mit der Analyse einzelner *Gene*. Ein *Gen* ist Bestandteil des sogenannten *Genoms*, das heißt der gesamten Erbsubstanz eines Organismus. Ein einzelnes Gen beschreibt in der Regel einen Teilbaustein der Zelle (zum Beispiel ein Enzym). Als Trägersubstanz der Erbinformation dient in allen Lebewesen die DNA (Desoxyribonukleinsäure). In der DNA ist die genetische Information als Abfolge von vier Basen festgelegt, die mit den Buchstaben A (Adenin), C (Cytosin), G (Guanin) und T (Thymin) kodiert werden. Für die Ermittlung dieser Buchstabenfolge (*der Sequenz*) sind die Techniken in den letzten Jahrzehnten immer weiter verfeinert worden. Durch die parallele Leistungssteigerung der verfügbaren Computersysteme und der Labortechnik ist nun die Analyse ganzer Genome möglich.

■ Technik der DNA-Sequenzierung – Von der Base zum Genom

Um an das gesamte genetische Programm eines Organismus (seine Genomsequenz) zu gelangen, wird die Abfolge der vier Basen entschlüsselt, das heißt, das Genom des Organismus wird *sequenziert*. Die Bestimmung der Sequenz von DNA-Abschnitten ist ein beinahe vollständig automatisierter Vorgang, den neben Forschungseinrichtungen auch verschiedene Dienstleister anbieten. Die aktuelle Technik erlaubt jedoch nur die Bestimmung der Sequenz von kurzen DNA-Bruchstücken bis zu 1000 Basen. Daher wird ein Genom (bei Bakterien in der Regel zwischen 3 und 7 Mio. Basen) zunächst in viele kleine Einzelfragmente zerteilt. Aus dem Englischen abgeleitet wird dieser Vorgang als *Shotgun-Fragmentierung* bezeichnet. Das Genom wird in viele Teile „zerschossen“, die anschließend sequenziert werden (Abbildung 1 zeigt das Ergebnis einer solchen Sequenzierung). Diese Fragmente werden danach

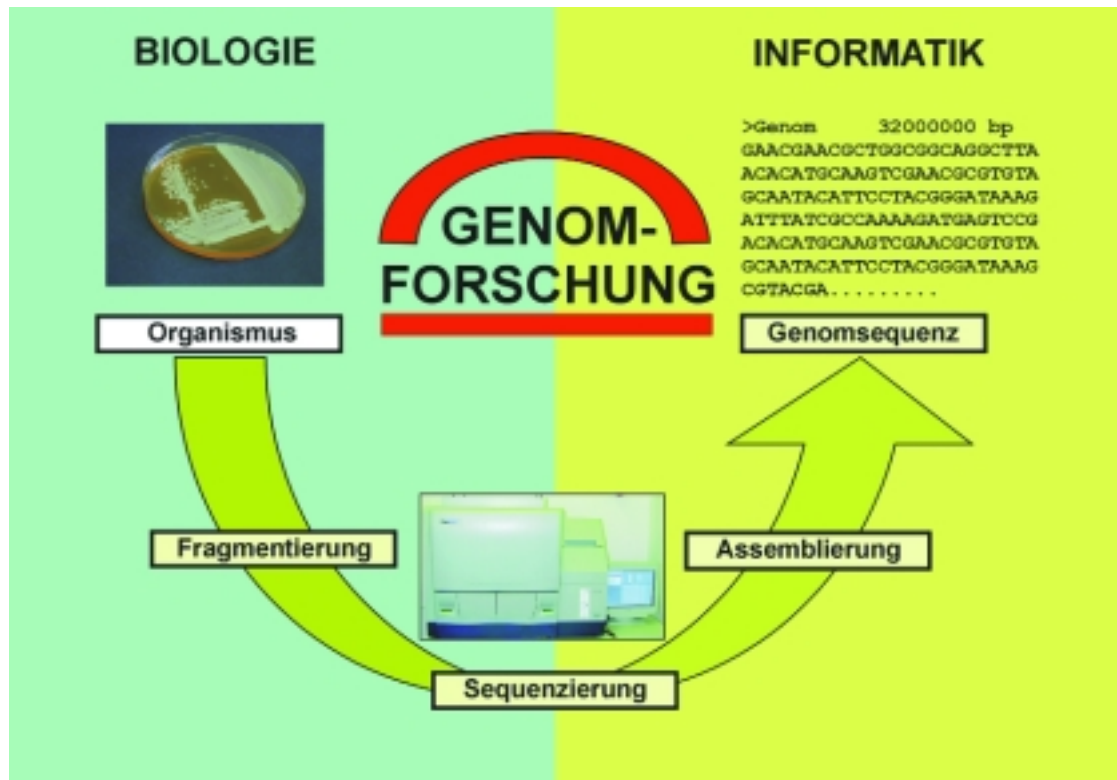


Abbildung 2

Schematische Darstellung der Vorgehensweise bei der Genomsequenzierung. Um aus der genomischen DNA die genomische Sequenz eines Bakteriums zu erhalten, muss diese sequenziert werden. Da mit der heutigen Technik keine genomische DNA im Hochdurchsatz sequenziert werden kann, sondern nur kleinere Bruchstücke, muss die genomische DNA in handhabbare Teile fragmentiert werden. Diese Teilstücke können sequenziert und anschließend wieder zu einem Ganzen, der genomischen Sequenz, assembliert werden.

im Computer virtuell wieder zusammengefügt. Der Vorgang des Zusammenfügens der Einzelfragmente auf Basis von Sequenzüberlappungen wird als *Genom-Assemblierung* bezeichnet. Der Ablauf einer solchen Genomsequenzierung ist ausgehend vom Organismus bis hin zur fertigen Genomsequenz in Abbildung 2 dargestellt.

■ Zusammenspiel von Biologie und Informatik

Die Anzahl der zu sequenzierenden Fragmente hängt bei der *Shotgun*-Sequenzierung nicht nur direkt von der Größe des Genoms ab, sondern auch von der Qualität der einzelnen Sequenzen (d.h. der Fehlerwahrscheinlichkeit pro Base in diesen Sequenzen). Für ein Bakterium mit 3 Mio. Basen werden in der Regel zwischen 20.000 und 30.000 Einzelfragmente sequenziert.

Die Einzelfragmente werden mit einer Assemblierungs-Software zu größeren Teilstücken (*Contigs*) zusammengesetzt. Die Entwicklung der Contig-Anzahl beschreibt genau den jeweiligen Stand des Genomprojekts. In Abhängigkeit von der Genomgröße kann mit der Anzahl der sequenzierten Fragmente die zu

erwartende Zahl der Contigs vorhergesagt werden. Um den Fortgang des Projekts zu überwachen, kann die reale Entwicklung der Contigs mit der theoretischen verglichen werden. In Abbildung 3 ist eine solche Entwicklung dargestellt. Erst nach der Assemblierung im Computer entsteht ein für die Forscher aus dem Bereich der Biologie interessanter Datensatz. Bis zu diesem Zeitpunkt bleibt die biologische Information in einer Unmenge an Sequenzdaten von Einzelfragmenten verborgen.

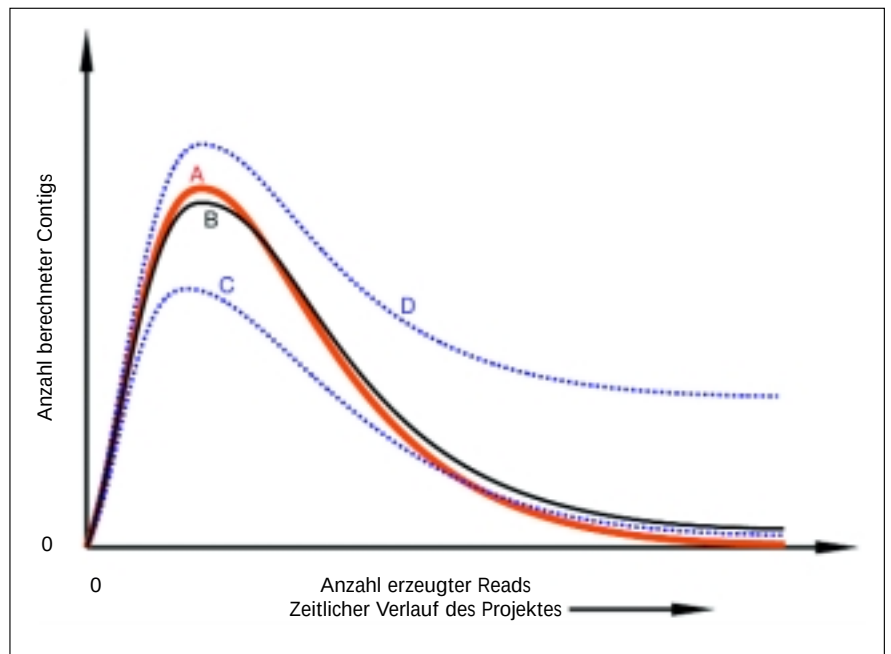
Daher wurde an der Universität Bielefeld ein Werkzeug (BioMake) erstellt, das den aktuellen Stand von Sequenzierung und Assemblierung visualisiert und so eine genaue Kontrolle über den Projektverlauf erlaubt.

■ Weitere Bearbeitung – Genvorhersage und Annotation

Mit der Ermittlung der Sequenz des bakteriellen Genoms ist die Arbeit jedoch nicht abgeschlossen. Es beginnt eine neue Phase der Analyse. Aus der Sequenz – das bedeutet aus der reinen Basenabfolge (eine Kette von „GATCAGTCATCGT...“ mit mehr-

Abbildung 3

Lander-Waterman-Plot-Visualisierung und Kontrolle des Fortschritts eines Genomprojekts. Der so genannte Lander-Waterman-Plot stellt die Anzahl der gelesenen Fragmente (reads) den zusammengefügteten Teilen des Puzzles (contigs) gegenüber. Die Kurve A zeigt unter Berücksichtigung der Faktoren Genomgröße und mittlere Leselänge der erzeugten Reads die theoretisch zu erwartende Entwicklung des Projekts. Zu Beginn des Projekts steigt die Contig-Anzahl stark an, da jeder neu erzeugte Read nur mit geringer Wahrscheinlichkeit mit einem anderen Read überlappt und statt dessen einen eigenen Contig darstellt. Bei Erzeugung weiterer Reads verbinden diese die vorhandenen Contigs in immer stärkeren Maße. Dies hat zur Folge, dass die Größe der einzelnen Contigs zunimmt und die Anzahl der Contigs abnimmt und zuletzt nur noch ein Contig, das heißt das Gesamtgenom, übrig bleibt. Die Kurven B,C,D beschreiben mögliche Verläufe realer Projekte. Kurve B zeigt ein optimal ablaufendes Projekt, während die Kurven C und D Projekte zeigen, bei denen entweder die Genomgröße falsch angegeben ist (Kurve C) oder Sequenzierprobleme auftreten (Kurve D).



ren Millionen Buchstaben) – kann auch der Experte fast nichts ablesen. Es gilt nun, die einzelnen Gene zu identifizieren und deren biologische Funktion zu beschreiben. Dieser Prozess erfordert eine enge Vernetzung der Genomforschung mit der Informatik, denn das Finden und Beschreiben von Genen erfolgt mittels Computerprogrammen.

Da zur Bearbeitung eines Genoms eine Vielzahl von Verfahren und Programmen notwendig ist (so

gibt es etwa verschiedene Arten von Genen), wird in der Regel die Beschreibung der Gene und deren Gen-Funktionen, die so genannte *Annotation*, von umfassenden Softwarepaketen – so genannten *Genomannotationssystemen* – unterstützt. Ein solches System (GenDB) wurde an der Universität Bielefeld entwickelt und wird bereits in einer Vielzahl von Projekten weltweit eingesetzt. Abbildung 4 vermittelt einen Eindruck von den Informationen, die

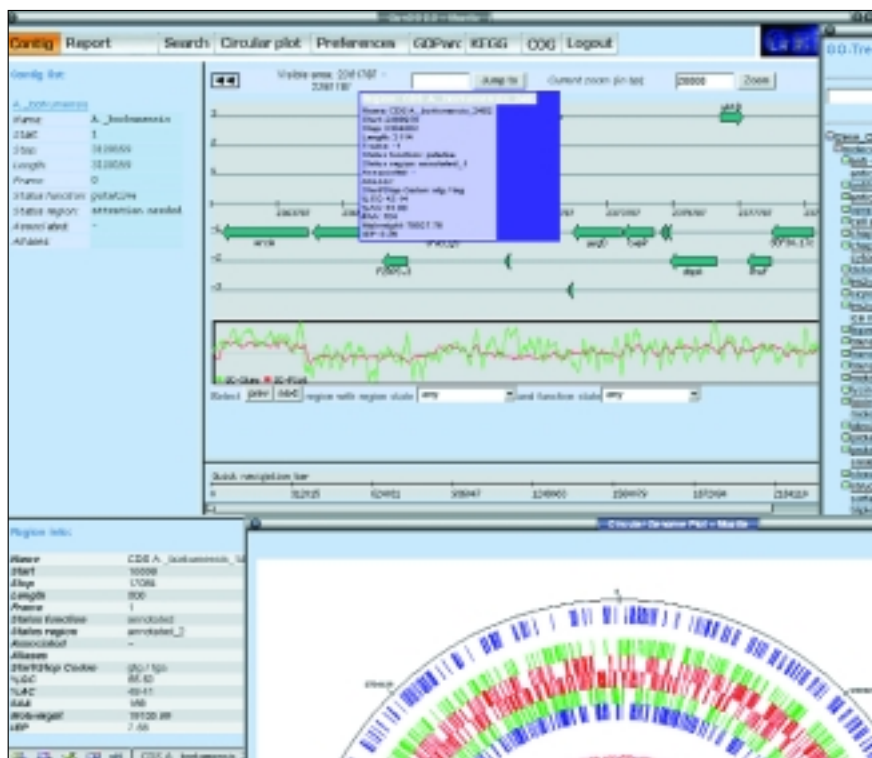


Abbildung 4
Das Annotations-Softwaresystem GenDB. Das an der Universität Bielefeld entwickelte und weltweit in unzähligen Projekten eingesetzte Annotations-Softwaresystem GenDB bietet den Forschern vielfältige Unterstützung bei der Genfunktionszuweisung.



Abbildung 5

Hardware-Infrastruktur. Für Genomforschung ist eine leistungsfähige Hardware-Infrastruktur unerlässlich. Die an der Universität Bielefeld für solche Projekte eingesetzte Hardware-Infrastruktur umfasst neben 50 CPUs in verschiedenen SUN-Applikationsservern weitere 90 CPUs in einem Compute-Cluster in SUN-GridEngine-Technologie. Ein Storage Area Network stellt sechs Terabyte an Festplattenspeicher online bereit mit einem Backup-System von 70 Terabyte Kapazität. Verbunden sind alle Komponenten über ein Gigabit-Netzwerk auf Lichtwellenleiter-Basis. Die Autoren danken Daniela Bartels und Peter Serocka für die Mithilfe bei der Erstellung der Abbildungen.

für ein Gen vom GenDB-System berechnet und dem Benutzer bei Bedarf dargestellt werden. Grundlage für diese Berechnungen ist eine leistungsfähige Computerplattform (siehe Abbildung 5), um die enormen Anforderungen an Rechenleistung und Speicherplatz zu erfüllen. Besonders aufwendig sind dabei die Datenbanksuchen. Jedes der Gene (ein Bakterium besitzt in der Regel mehrere tausend) wird mit allen bereits bekannten Genen in verschiedenen Datenbanken verglichen. Wird eine hinreichende Ähnlichkeit gefunden, kann die Beschreibung in der Datenbank genutzt werden, um die Funktion des jeweiligen Gens zu bestimmen. Durch den Einsatz von mehreren hundert parallel arbeitenden Prozessoren können solche Suchen für ganze Genome innerhalb weniger Tage bewältigt werden.

Die mit informatischen Methoden automatisch bestimmten möglichen Gen-Funktionen müssen jedoch von einem Experten aus der Biologie überprüft werden, da sowohl Datenbanken als auch Programme Fehler machen. Hinzu kommt die Tatsache, dass nicht alle Gene bereits in den Datenbanken vorhanden sind. In der Regel werden nur 40 bis 60 Prozent aller Gene eines Bakteriums durch informatische Methoden bestimmt. Um die Funktion der verbleibenden Gene aufzuklären, bedarf es neben der Weiterentwicklung der bioinformatischen Methoden auch weiterer Laborexperimente.

■ Aktuelle Projekte

Der dargestellte Weg von der Base bis zum annotierten Genom ist nicht rein theoretischer Natur, sondern wurde in den vergangenen zwei Jahren in mehreren Großprojekten an der Universität Bielefeld praktisch umgesetzt. Die Entschlüsselung der Erbinformation fünf bakterieller Genome aus den Bereichen Landwirtschaft, Umwelt und Biotechnologie wird an der Universität Bielefeld koordiniert. Aus dem Bereich Landwirtschaft werden die Genome der beiden pflanzenwuchsschädigenden Bakterien *Clavibacter michiganensis* subsp. *michiganensis* und *Xanthomonas campestris* pv. *vesicatoria* sowie das Genom des pflanzenwuchsfördernden Bakteriums *Azoarcus* sp. sequenziert. Aus dem Bereich Umwelt wird das Genom des Bakteriums *Alcanivorax borkumensis*, welches in der Lage ist, Erdöl abzubauen und zu verwerten, sequenziert, während im Bereich Biotechnologie das Genom des *Myxobakterium Sorangium cellulosum*, das mit ca. 12 600 000 Basen das größte bekannte bakterielle Genom besitzt, Untersuchungsobjekt ist. Die im Rahmen von Sequenzierprojekten benötigte Assemblierung des Genoms und die genaue Qualitätskontrolle der von Sequenzierdienstleistern gelieferten Sequenzdaten werden ebenfalls von Bielefelder Forschern durchgeführt.

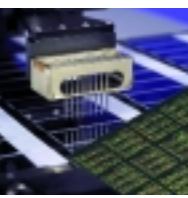
■ Ausblick

Die an der Universität Bielefeld praktizierte enge Zusammenarbeit von Biologen und (Bio-)Informatikern hat sich in den vergangenen Jahren als eine enorm leistungsfähige Plattform zur Durchführung und Weiterentwicklung von Forschungsprojekten im Bereich Genomforschung herauskristallisiert. Aktuelle vom Bundesministerium für Bildung und Forschung geförderte Forschungsprojekte und zukünftige Forschungsprojekte, die im Rahmen des 6. EU-Rahmenprogramms in Bielefeld bearbeitet werden, sind ein guter Indikator für den Erfolg der Plattform.

Dr. Olaf Kaiser studierte Biologie an der Universität Bielefeld und promovierte im Jahr 2001 am Lehrstuhl für Genetik bei Prof. Dr. Alfred Pühler mit der vergleichenden Analyse bakterieller Gemeinschaften der Rhizoplane (Wurzeloberfläche) von Wildtyp- und transgenen Rapspflanzen. Nach einem halbjährigen Industrieengagement im Geschäftsbereich DNA-Sequenzierung der IIT GmbH übernahm er im September 2001 den Bereich DNA-Sequenzkoordination im Technologieknoten des an der Universität Bielefeld angesiedelten Kompetenznetzwerks „Genomforschung an Bakterien für den Umweltschutz, die Landwirtschaft und die Biotechnologie“. Im Rahmen dieser Tätigkeit beschäftigt er sich zur Zeit mit der Planung und Durchführung von fünf bakteriellen Genomsequenzierungsprojekten und einigen kleineren Sequenzierungsprojekten.



Dr. Folker Meyer studierte Naturwissenschaftliche Informatik an der Universität Bielefeld und promovierte im Jahr 2002 bei Prof. Dr. Robert Giegerich und Prof. Dr. Alfred Pühler. Im Rahmen seiner Promotion beschäftigte er sich mit der Entwicklung eines Systems zur Analyse bakterieller Genomsequenzen. Aktuell steht die bioinformatische Betreuung der Genom- und Postgenomprojekte im Mittelpunkt seiner Arbeit. Neben der Promotion baute er die Service Bioinformatik-Gruppe im ehemaligen Zentrum für Genomforschung auf und leitet aktuell die im Aufbau befindliche Bioinformatics Resource Facility am Centrum für Biotechnologie der Universität Bielefeld.

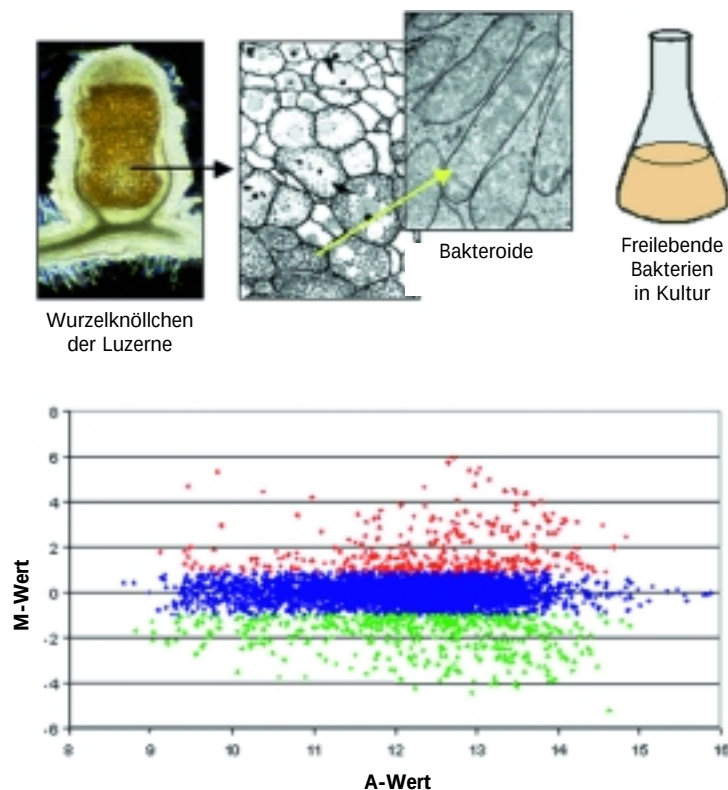


Transkriptomforschung: Mikroarrays helfen bei der Aufklärung komplexer Genregulationsnetzwerke

Anke Becker

Fakultät für Biologie

Im Blickpunkt der Genomforschung stand lange im Wesentlichen die Entzifferung der vollständigen Erbinformation von Organismen. In den letzten Jahren ist die Zahl der bekannten Genomsequenzen von Bakterien aber auch von höheren Organismen sprunghaft angestiegen. Die Bestimmung der Genomsequenz stellt die Grundlagen für weitere Analysen dar. Diese haben zum Ziel, die zelluläre Umsetzung der Genominformation zu beschreiben und zu verstehen. Die Übersetzung der Geninformation in Proteine ist ein essentieller Vorgang der Umsetzung von Erbinformation in zelluläre Eigenschaften. Die so genannten DNA-Chip- und Array-Techniken zur ganz-genomischen Analyse dieses Vorgangs haben sich explosionsartig entwickelt und sind heute aus der modernen Genomforschung nicht mehr wegzudenken. Ihre Anwendungsbereiche liegen sowohl in der Grundlagenforschung als auch in der medizinischen Diagnostik.



Aufbau und Ergebnis eines Mikroarray-Experiments zum Vergleich der Genaktivitäten in Bakteroiden und freilebenden Zellen des Bakteriums *Sinorhizobium meliloti*.

Im unteren Teil ist die Auswertung einer Hybridisierung dargestellt, in der ein Mikroarray mit ca. 6200 Genen gleichzeitig gegen mit roter Fluoreszenz markierte Transkripte aus Bakteroiden und mit grüner Fluoreszenz markierte Transkripte aus freilebenden Zellen hybridisiert wurde. Rote Punkte markieren solche Gene, die in Bakteroiden stärker aktiv sind, grüne Punkte zeigen solche Gene an, die in freilebenden Zellen stärker aktiv sind. Blaue Punkte deuten an, dass die entsprechenden Gene in beiden Bedingungen gleich stark aktiv sind. Die kombinierten Fluoreszenzintensitäten im roten und grünen Kanal wurden auf einer logarithmischen Skala gegen den Logarithmus des Verhältnisses der Intensitäten in beiden Kanälen aufgetragen.

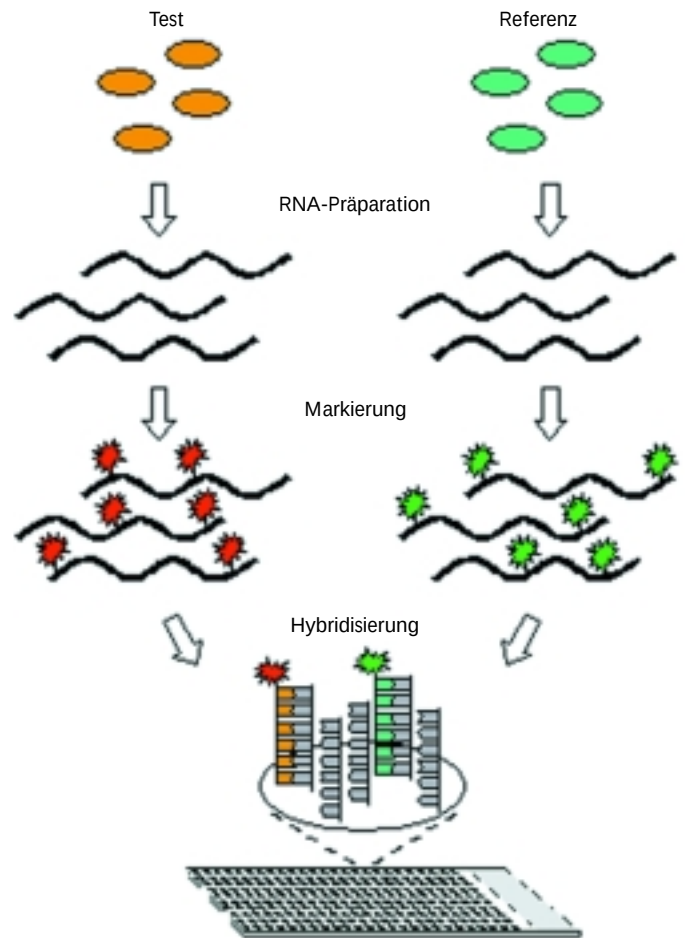
Die in der Genomsequenz enthaltene Information wird in zwei Schritten umgesetzt, die schließlich zur Produktion von Proteinen führen. Im ersten Schritt, der Transkription, wird von der Gensequenz eine Abschrift, die sogenannte Boten-RNA, erstellt. Diese dient als Vorlage für die Herstellung von Proteinen in einem als Translation bezeichneten Vorgang. Nicht alle Gene einer Zelle sind gleichzeitig aktiv und führen zur Herstellung von Proteinen. Die Regulation der Genaktivität bestimmt damit zum großen Teil die Eigenschaften einer Zelle.

Die Messung der Boten-RNA-Menge erlaubt Rückschlüsse auf die Aktivität eines Gens. Zur Messung der Boten-RNA-Menge macht man sich die Eigenschaft einzelsträngiger Nukleinsäuren wie der Boten-RNA zu Nutze, mit einzelsträngigen Nukleinsäuren komplementärer Sequenz Doppelstränge ausbilden zu können. Dieser als Hybridisierung bezeichnete Prozess ist sehr spezifisch für die jeweilige Nukleinsäuresequenz und eignet sich deshalb für genaue Analysen.

Die Methodik der Hybridisierung wird in der Molekularbiologie schon lange für den Nachweis und die Messung verschiedener Nukleinsäuren verwendet. Durch die Entwicklung der *Chip*- und *Array*-Technologie ist es gelungen, das Hybridisierungsverfahren für eine große Anzahl an Genen zu parallelisieren. Hierbei werden einzelsträngige DNA-Moleküle bekannter Sequenz als Sensoren für Boten-RNA-Moleküle in einem geordneten Raster auf einem Träger fixiert. Da mit solchen Verfahren die Gesamtheit der Gen-Abschriften, die auch als Transkripte bezeichnet werden, untersucht werden kann, spricht man von der Erforschung des Transkriptoms.

■ Methoden zur Herstellung von DNA-Array

DNA-Chips und *-Arrays* werden auf unterschiedliche Art und Weise hergestellt. Die Sensor-DNA-Moleküle werden zur Herstellung von *DNA-Chips* direkt auf der Trägeroberfläche synthetisiert. Hier kommen zum Beispiel photolithographische Verfahren zum Einsatz. *DNA-Arrays* werden dagegen durch Aufbringen der Sensor-DNA-Moleküle auf die Trägeroberfläche hergestellt. Von der Dichte der Rasterpunkte hängt es ab, ob von Makroarrays (niedrige Dichte) oder Mikroarrays (hohe Dichte) gesprochen wird. Am Institut für Genomforschung werden beide Array-Typen hergestellt, wobei der Trend sich aber in Richtung der hochdichten Mikroarrays entwickelt. Als Trägermaterialien dienen für Makroarrays häufig Nylonmembranen und für Mikroarrays Glas-Objektträger. Diese sind beschichtet, um die Fixierung der Sensor-DNA-Moleküle zu ermöglichen. Auf einem



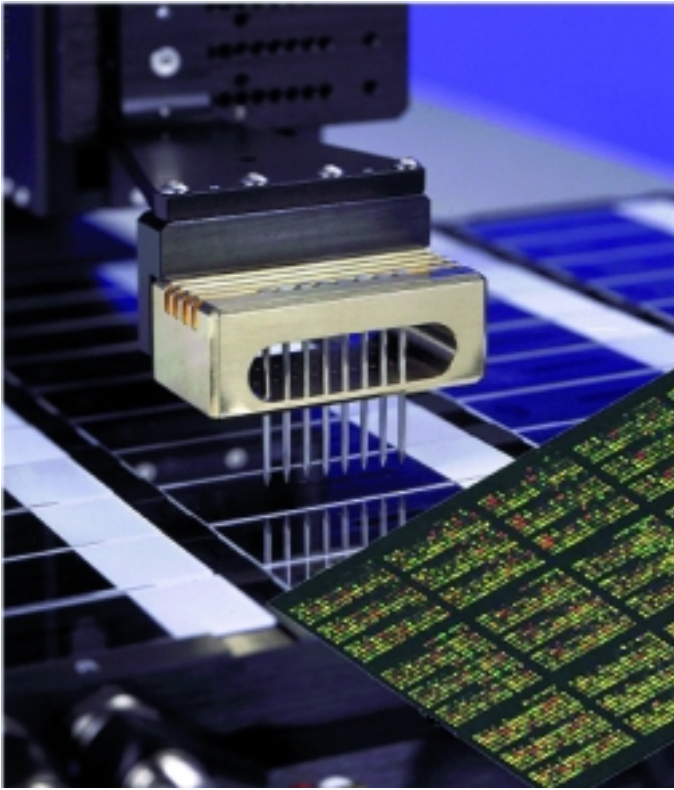
Prinzip der Array-Hybridisierung. Bei der Hybridisierung von Arrays werden die Genaktivitäten einer Test- und einer Referenz-Bedingung verglichen. Hierbei werden zunächst RNA-Moleküle aus beiden Bedingungen isoliert und anschließend mit verschiedenen Fluoreszenzfarbstoffen markiert. Durch Hybridisierung mit auf einem Träger in einem geordneten Raster fixierten Sensor-DNA-Molekülen wird die relative Menge der verschiedenen RNA-Moleküle in den zu vergleichenden Bedingungen ermittelt.

Glas-Objektträger können mit Hilfe moderner Robotik Sensor-DNA-Moleküle für mehrere tausend Gene aufgebracht werden.

■ Messung von Genaktivität mit Hilfe von Mikroarrays

Arrays werden vor allem eingesetzt, um bei höheren Organismen die Genaktivität in verschiedenen Geweben oder unter dem Einfluss verschiedener, zum Beispiel pharmazeutischer Substanzen zu untersuchen. Auch der Vergleich von Tumorgewebe mit normalem Gewebe ist ein weites Anwendungsfeld. Im Bereich der Mikrobiologie wird zum Beispiel die Genaktivität unter verschiedenen Kultivierungsbedingungen verglichen.

Um solche vergleichenden Analysen durchzuführen, werden zunächst die Abschriften der Gene unter den zu vergleichenden Bedingungen oder aus den



Ein mit Glasobjektträgern bestücktes Robotik-System wie es zur Herstellung von Mikroarrays verwendet wird. Im unteren rechten Teil ein Ausschnitt eines typischen Fluoreszenzbildes wie es bei der Auswertung von Mikroarray-Experimentdaten entsteht.

verschiedenen Geweben getrennt isoliert. Man bezeichnet diese zu vergleichenden Bedingungen auch als Test und Referenz. Anschließend erfolgt eine Kopplung von Fluoreszenzfarbstoffen an diese RNA-Transkripte, wobei rot fluoreszierende Farbstoffe an die Transkripte aus der Test-Bedingung und grün fluoreszierende Farbstoffe an die Transkripte aus der Referenz-Bedingung gebunden werden. Anschließend werden die rot und grün markierten Transkripte im gleichen Verhältnis gemischt und auf den Mikroarray gegeben. Erkennen die markierten RNA-Transkripte Sensor-Moleküle auf dem Glas-Objektträger, binden sie an diesen und können anschließend mit einem Fluoreszenz-Scanner identifiziert werden. Rot fluoreszierende Signale markieren dann solche Gene, die unter der Test-Bedingung stärker aktiv sind. Grün fluoreszierende Signale zeigen solche Gene an, die unter dieser Bedingung eine geringere Aktivität haben als unter der Referenz-Bedingung.

Die Stärke der Transkriptomforschung liegt darin, dass die Aktivität einiger tausend Gene in einem einzigen Experiment gemessen werden kann. Untersuchungen mit Hilfe der Transkriptomforschung führen zu einer enormen Datenfülle und erfordern sowohl einen hohen Bedarf an Speicherkapazität als auch an Software, die in der Lage ist, die Ergebnisse

effizient auszuwerten. Solche Systeme werden von der Bioinformatik-Gruppe des Centrums für Biotechnologie entwickelt und zur Verfügung gestellt.

■ Anwendungsbereiche

Bei Mikroarray- und Chip-Experimenten sind viele Varianten in der experimentellen Fragestellung denkbar. Die Einsatzmöglichkeiten sind vielseitig und umfassen Bereiche aus der Grundlagenforschung, der Agrarwissenschaft, der Lebensmittel- und Umwelttechnik sowie der Medizin. Die technischen Anforderungen in den verschiedenen Bereichen sind sehr ähnlich. Deshalb sind Zentren wie die Zentrale Einheit für Genomforschung an einer Reihe unterschiedlicher Projekte beteiligt und ermöglichen so Forschungsgruppen den Zugang zu modernen Techniken aus diesem Bereich der Genomforschung. Die Zentrale Einheit für Genomforschung ist hauptsächlich Kooperationspartner von Gruppen aus Deutschland und anderen europäischen Ländern, arbeitet aber auch mit Gruppen aus den USA und China zusammen.

■ Mikroarrays am Institut für Genomforschung

In der Zentralen Einheit für Genomforschung wurden bisher Mikroarrays für die Analyse der Genaktivität im symbiontischen Stickstoff-fixierenden Bodenbakterium *Sinorhizobium meliloti* und im Aminosäure-produzierenden Bakterium *Corynebacterium glutamicum* produziert. Diese werden im Rahmen von Projekten im Kompetenznetzwerk „Genomforschung an Bakterien für den Umweltschutz, die Landwirtschaft und die Biotechnologie“ verwendet. Für die Modellpflanze *Medicago truncatula* wurden ebenfalls Mikroarrays hergestellt. Diese werden im Genomforschungsprojekt MEDICAGO der Europäischen Union (<http://medicago.toulouse.inra.fr/EU>) in dem sich elf Partner aus Frankreich, Großbritannien, Deutschland, den Niederlanden und Ungarn zusammengeschlossen haben, genutzt. Darüber hinaus macht auch das Schwerpunktprogramm MolMyk der Deutschen Forschungsgemeinschaft (<http://molmyk.genetik.uni-bielefeld.de>), an dem 17 deutsche Arbeitsgruppen beteiligt sind, Gebrauch von diesen Mikroarrays.

■ Analyse von Regulationsnetzwerken im symbiontischen Bakterium *Sinorhizobium meliloti*

Ein Schwerpunkt der Arbeiten in der Zentralen Einheit für Genomforschung und des Clusters „Pflanzenwuchsfördernde Bakterien“ des in Bielefeld angesiedelten Kompetenznetzwerks liegt auf der

Erforschung des Transkriptoms des Bodenbakteriums *Sinorhizobium meliloti*. Dieses Bakterium geht mit Pflanzen wie der Luzerne eine Symbiose ein, in deren Verlauf die Bakterien in der Luft enthaltenen Stickstoff fixieren und der Pflanze zur Verfügung stellen. Dies ermöglicht den Pflanzen auch auf stickstoffarmen Böden zu wachsen, was von großer Bedeutung für die Landwirtschaft ist.

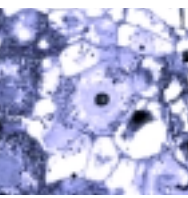
Die Bakterien besiedeln Knöllchenstrukturen an den Wurzeln der Wirtspflanzen und differenzieren in den Pflanzenzellen zu Bakteroiden, einem nicht mehr teilungsfähigen Stadium. Erst im Bakteroidstadium sind sie in der Lage, Stickstoff zu fixieren. Da dies ein sehr energieaufwendiger Prozess ist, müssen sie von der Pflanze mit Kohlenstoffverbindungen versorgt werden. Eine funktionierende Symbiose erfordert eine komplizierte Feinabstimmung zwischen der Pflanze und dem Bakterium. Durch die Erforschung der Genaktivität in Pflanze und Bakterium während der Symbiose werden Gene identifiziert, die für diese Abstimmung von Bedeutung sind. Hierbei ist die Mikroarray-Technik von großer Bedeutung, da sie eine globale Analyse der Genaktivitäten erlaubt. So wurde durch Mikroarray-Experimente gezeigt, dass sich die Genaktivität von freilebenden Bakterien und Bakteroiden drastisch unterscheidet.

*Hochschuldozentin Dr. Anke Becker studierte Biologie an der Universität Bielefeld und promovierte 1994 mit einem Thema aus dem Gebiet der bakteriellen Polysaccharidproduktion. Im Jahr 2000 erfolgte die Habilitation. Zur Zeit leitet sie die Transkriptomik-Abteilung der Zentralen Einheit für Genomforschung und die Transkriptomik Nachwuchsgruppe am Institut für Genomforschung. Ihre Arbeitsgruppe beschäftigt sich vor allem mit der funktionellen Analyse des Genoms des symbiontischen Modell-Bakteriums *Sinorhizobium meliloti*. Ein weiterer Schwerpunkt liegt auf der Entwicklung von Methoden in der Transkriptomforschung.*

■ Ausblick

Die Transkriptomik-Abteilung der Zentralen Einheit für Genomforschung an der Universität Bielefeld hat nach ihrer Gründung im Jahr 1999 ihren Platz in der Genomforschungslandschaft gefunden. Bei der Gründung haben die angewendeten Techniken noch wenig Beachtung gefunden. Dies hat sich in den letzten Jahren geändert. Durch die Vielseitigkeit der Anwendungsmöglichkeiten der Mikroarray-Technik und die steigende Zuverlässigkeit der Methoden werden die Möglichkeiten der Transkriptomforschung von vielen Arbeitsgruppen in den verschiedensten Forschungsgebieten genutzt. Die Mikroarray-Technik hat mittlerweile die Grenzen einer Technik, die zunächst nur in der Grundlagenforschung eingesetzt wurde, überschritten und wird bereits an vielen Stellen zur Diagnostik im Lebensmittelbereich, in der Umwelttechnik und der Medizin eingesetzt.





Proteomforschung: Vom Gen zur Funktion

Mit der Sequenzierung ganzer Genome liegt das Buch auf dem Tisch – nun müssen wir lernen, es zu lesen.

Karsten Niehaus

Fakultät für Biologie

Die Erbinformation besteht, wie zum 50-jährigen Jubiläum der Entschlüsselung der DNA-Struktur durch alle Medien geht, aus einem Code von nur vier „Buchstaben“. Für einige Organismen wurde die gesamte Erbinformation, bestehend aus einigen Millionen Buchstaben, entschlüsselt. Die Gesamtheit dieser Information wird als Genom bezeichnet. Wie aber wird dieser Code umgesetzt in die Vielfalt der uns bekannten Natur? Wie wird mit Hilfe dieser vier Buchstaben die Form einer Schnecke, die Größe eines Baumes oder unsere Augenfarbe festgelegt? Diese Fragen sind nicht gelöst, aber einen ersten Hinweis auf ihre Beantwortung bekommen wir, wenn wir dem Informationsfluss in der Zelle folgen.



Wie wird aus einem Samenkorn eine Pflanze, die wächst, blüht und erneut Samen produziert? Wie richtet die Pflanze ihre Blätter zum Licht aus und bildet längere Wurzeln auf trockenerem Boden? Wie erkennt eine Pflanze den Befall durch krankheitserregende Bakterien und unterscheidet diese von nützlichen Bakterien? Die Entschlüsselung des Erbguts verspricht einen Zugang zur Beantwortung dieser Fragen. Wie die einzelnen Worte einer Sprache, sind die Gene, die das Erbgut bilden, die „Informationseinheiten“ der Zelle. Und so wie sich ein Satz nur durch das Zusammenspiel seiner Wörter erklärt, erklären sich die komplexen Leistungen eines Lebewesens durch das Zusammenspiel seiner Gene. Der genetische Code wird jedoch zunächst in Eiweiße (Proteine) übersetzt, ein Vorgang, den man mit dem Dechiffrieren einer Nachricht vergleichen könnte.

■ Ein Gen – Ein Protein – Eine Funktion?

Von definierten Genen werden in der Zelle Abschriften, so genannte Boten-(messenger-) RNAs, erstellt. Dieser Vorgang wird als Transkription bezeichnet. Die Gesamtheit aller zu einem definierten Zeitpunkt vorhandenen mRNAs bildet das so genannte Transkriptom. Während das Genom uns sagt, welche Gene in einem Organismus vorhanden sind, zeigt das Transkriptom, welche Gene gerade aktiv sind. Die mRNAs können in Proteine übersetzt werden, ein Vorgang der als Translation bezeichnet wird, da hier der Vier-Buchstaben-Code der Nukleinsäuren in die Aminosäuren der Proteine übersetzt wird. Proteine geben der Zelle ihre Struktur und bewirken alle Stoffwechselvorgänge. Atmung, Photosynthese, Bewegung; jeder dieser Vorgänge wird direkt durch Proteine (Enzyme) verursacht. Die Gesamtheit der Proteine bildet das Proteom. Stoffwechsel ist an kleine organische Verbindungen, die Metabolite, gebunden. Entsprechend den vorgehenden Definitionen wird die Gesamtheit aller Metabolite als Metabolom bezeichnet (Abbildung 1).

Das Proteom ist für das Verständnis der Lebensvorgänge von besonderem Interesse. Biologische Funktionen, wie Zellteilung, Differenzierung oder spezifische Reaktionen auf die Umwelt, erfordern komplexe „Anpassungsleistungen“ der einzelnen Zelle. Solche Anpassungsleistungen sind häufig mit einer Änderung in der Proteinausstattung und/oder Proteinmodifikation sowie Proteinlokalisierung in einer Zelle verbunden. So haben sekretierte Proteine in aller Regel andere Aufgaben und Eigenschaften als solche, die sich im Zellkern finden, und diese unterscheiden sich von denen, die zum Beispiel in die dem Stoffwechsel der Zelle dienenden Mitochondrien transportiert werden. Darüber hinaus kommt es im Rahmen einer Aktivierung spezifischer Proteinfunktionen zu einer chemischen Veränderung des betreffenden Proteins. Beispielsweise kann ein durch Phosphorylierung (Übertragung so genannter Phosphatgruppen) modifiziertes Protein aus dem Cytoplasma (der flüssigen Grundsubstanz innerhalb der Zellmembran) in den Zellkern wandern und dort die Aktivität verschiedener Gene stimulieren. Ziel der Proteomforschung ist es, die Gesamtheit aller Proteine und ihrer Modifikationen zu erfassen. Da die Lokalisierung in der Zelle mit der Funktion korreliert werden kann, sind auch Daten zur Proteinlokalisierung von steigendem Interesse.

Bakterien enthalten einige tausend Proteine, je nach Art und der Wachstumsphase, in der sie sich befinden. Dies entspricht in etwa der Zahl der Gene, die für Proteine kodieren (ca. 3000 bis 12000). In Eukaryonten (Pflanzen, Tiere) spielt neben der höheren Zahl der Gene (ca. 12000 bis 40000), ein differenzielles Bearbeiten der mRNA und eine hohe Variabilität in der Modifikation von Proteinen eine große Rolle. Ein um wenige Aminosäuren verkürztes, mit einer Phosphatgruppe oder einem Zuckerrest (Oligosaccharid) modifiziertes Protein kann so aktiviert werden oder eine neue Funktion erhalten. Modifizierte Proteine müssen daher als „eigenständige“ Proteine betrachtet werden. In Bakterien werden so aus 3000 proteinkodierenden Genen unter Umständen 4000 verschiedene Proteine. Beim Menschen mit seinen ca. 35 000 Genen wird heute von bis zu einer Million Proteinen ausgegangen.

■ Technologie und Technik: Proteomanalyse

Ziel der Proteomanalyse ist es, möglichst viele Proteine einer Zelle oder eines Organismus zu identifizieren und ihre quantitativen Veränderungen sowie Modifikationen zu erfassen. Dieser neue und ganzheitliche Anspruch zielt auf ein Verständnis der Zusammenhänge ab. Der oft als reduktionistisch

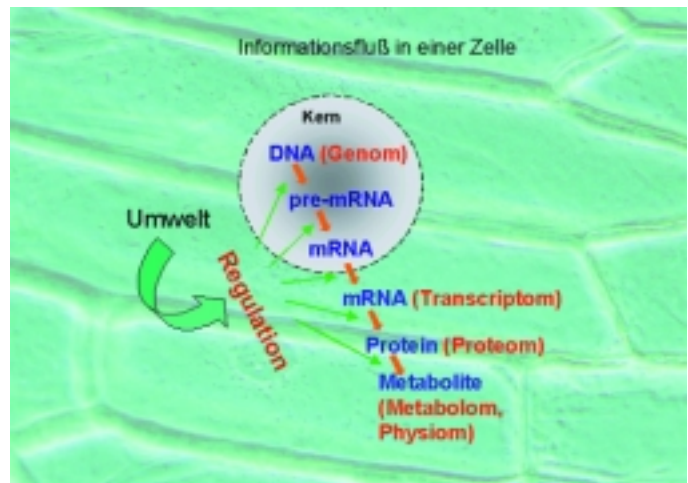


Abbildung 1: Wie wird die Erbinformation realisiert? Von definierten Genen auf der DNA werden Abschriften (mRNA) erstellt, die in Proteine übersetzt werden können. Wenn es sich bei den Proteinen um Enzyme handelt, können diese den Umsatz von Metaboliten katalysieren. Die jeweilige Gesamtheit der DNA, mRNA, Proteine und Metabolite wird als Genom, Transkriptom, Proteom und Metabolom bezeichnet. Es handelt sich jedoch nicht um einen linearen Informationsfluss von der DNA zu den folgenden Ebenen, sondern um ein vielfach rückgekoppeltes Netzwerk, in dem keine der Ebenen unabhängig von den anderen agiert.

angesehene Ansatz der Molekularbiologie findet hier zu einer übergeordneten Betrachtungsweise der biologischen Vorgänge zurück.

Technisch betrachtet kann der Ablauf einer Proteomanalyse in drei Schritte unterteilt werden: i.) Die Proteinisolierung, ii.) die Trennung der einzelnen Proteine und iii.) ihre Identifikation. Um das Proteom von Bakterien und Pflanzen zu untersuchen, wurde in Bielefeld eine Arbeitsgruppe für Proteomforschung aufgebaut. Dies geschah zunächst am Lehrstuhl für Genetik, dann im Rahmen der DFG-Initiative „Bioinformatik und Genomforschung“. Alle erforder-

lichen Techniken konnten erfolgreich implementiert werden. Während die Proteinisolierung relativ einfach ist, stellt die Trennung eines Proteingemisches hohe Anforderungen an die verwendete Technik. Für die meisten Analysen wird die zweidimensionale Gelelektrophorese (2D-Gel) verwendet, die Proteine nach ihrer Ladung und der Masse trennt. Durch Anfärbung werden die getrennten Proteine sichtbar gemacht (Abbildung 2). Häufig werden diese Experimente differentiell angelegt. So wird zum Beispiel das Proteom einer Bakterienkultur unter optimalen Bedingungen wachsend mit dem der gleichen Kultur unter Stressbedingungen verglichen. Proteine, die nur unter Stress gebildet werden, könnten mit einer Anpassung an diese Bedingung korreliert sein. Findet man solche Proteine in 2D-Gelen, ist der nächste Schritt ihre Identifikation. Dazu werden die Proteine mit proteinspaltenden Enzymen (in der Regel Trypsin) in definierte Bruchstücke (Peptide) zerlegt. Die Größe der Peptide – oder genauer gesagt deren Masse – dient dabei als „Fingerabdruck“ des betreffenden Proteins. Die Masse der Peptide wird mit Hilfe der Massenspektrometrie bestimmt. Bereits vor drei Jahren konnte in Bielefeld ein Massenspektrometer vom Typ Bruker Biflex III beschafft werden, das kürzlich durch ein Bruker ultraflex ergänzt wurde. Die mit den Massenspektrometern erstellten Peptidmassenlisten werden mit den aus den Genomdaten vorher-

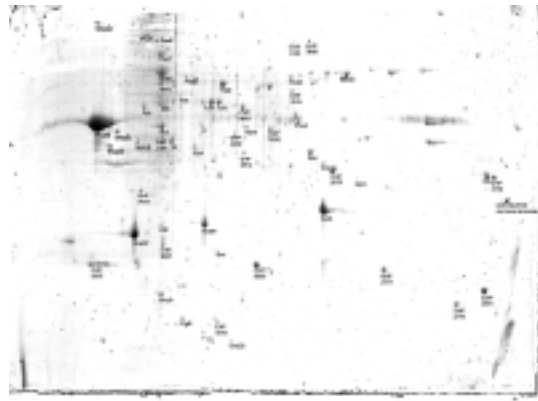


Abbildung 2: Auftrennung und Nachweis von Proteinen durch zweidimensionale Gelelektrophorese (2D-Gel). Die Proteine werden nach ihrer Ladung (horizontal) und ihrer Masse (vertikal) getrennt. Jeder „Fleck“ entspricht dabei einem Protein. Bei dem hier gezeigten Beispiel handelt es sich um Proteine, die von dem phytopathogenen Bakterium *Xanthomonas* in die Umwelt abgegeben werden.

gesagten Peptidmassen aller Gene des Organismus verglichen. Stimmen die gemessenen und vorhergesagten Peptidmassen überein, sind das Protein und sein korrespondierendes Gen identifiziert (Abbildung 3). Zur Identifikation von Proteinmodifikationen und für die Bestimmung kurzer Abschnitte der Aminosäuresequenz von Proteinen ist ein weiteres Massenspektrometer (Thermo Finnigan LCQ deca) vorhanden.

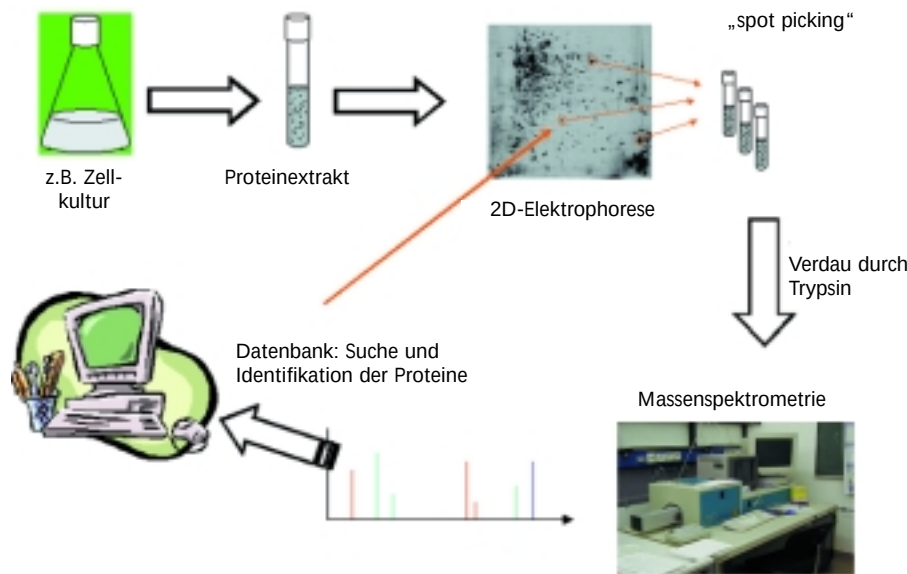


Abbildung 3: Ablauf eines Proteom-Experiments. Der erste Schritt der Analyse besteht aus der Extraktion der Proteine. Da Bakterien einige 1000, höhere Organismen wie Pflanzen und Tiere viele 10 000 verschiedene Proteine besitzen, müssen diese vor einer Identifizierung zunächst getrennt werden. Dazu dient die zweidimensionale Gelelektrophorese (2D-Gel). Die getrennten Proteine werden durch Färbung sichtbar gemacht. In dieser „Landkarte der Proteine“ kann nach Auffälligkeiten gesucht werden, zum Bei-

spiel nach solchen Proteinen, die nur bei Stress auftreten. Um herauszufinden, um welche Proteine es sich handelt, werden diese aus dem Gel ausgeschnitten und in Bruchstücke zerlegt. Die Größe der Bruchstücke wird mit einem Massenspektrometer bestimmt. Das Muster dieser Bruchstücke ist charakteristisch für das betreffende Protein. Mit Hilfe von Computern kann so das Protein identifiziert werden.

■ Wie kommunizieren Bakterien mit Pflanzen?

Wissenschaftliches Ziel der Arbeiten innerhalb der Arbeitsgruppe ist die Aufklärung der molekularen Mechanismen für die Erkennung von pathogenen wie symbiontischen Bakterien durch Pflanzen. Als wesentliche Elemente dieser Erkennung sollen Signalmoleküle und Komponenten der pflanzlichen Signalübermittlung identifiziert werden. In Analogie zur Entwicklung neuer Medikamente könnten Komponenten der pflanzlichen Signalübermittlung auch Angriffspunkte neuartiger Pflanzenbehandlungsmittel sein. Die Entscheidung der Pflanze für Kooperation oder Abwehr wird in der extrazellulären Matrix von Bakterium und Pflanze initiiert und später auch dort manifestiert. Die molekularen Mechanismen der Pathogen-Symbiont-Erkennung sind an außerzelluläre Proteine und Membranproteine gebunden.

Das pflanzenpathogene Bakterium *Xanthomonas campestris* pv. *campestris* (Xcc) ist der Erreger der Schwarzaderfäule an der Wirtspflanze Weißkohl (*Brassica oleracea*) und löst am Nichtwirt Paprika (*Capsicum* spp.) Abwehrreaktionen aus (Abbildung 4). Langfristiges Ziel dieses Vorhabens ist es, Virulenzfaktoren des Bakteriums zu identifizieren. Auf Grundlage der genomischen Sequenz von Xcc werden die ausgeschiedenen Proteine über den Proteom-Ansatz bestimmt. Bisher konnten schon 70 sekretierte Proteine identifiziert werden (Abbildung 2). Unter diesen Proteinen fanden sich drei, die die Abwehr der Pflanze induzieren. Von besonderem Interesse sind differentiell – also nur in Gegenwart der Wirtspflanze – sekretierte Proteine. Gene, die für solche Proteine kodieren, sollen durch genetische Veränderung ausgeschaltet werden, um Virulenzfaktoren zu identifizieren, die für die Interaktion mit der Wirtspflanze essentiell sind.

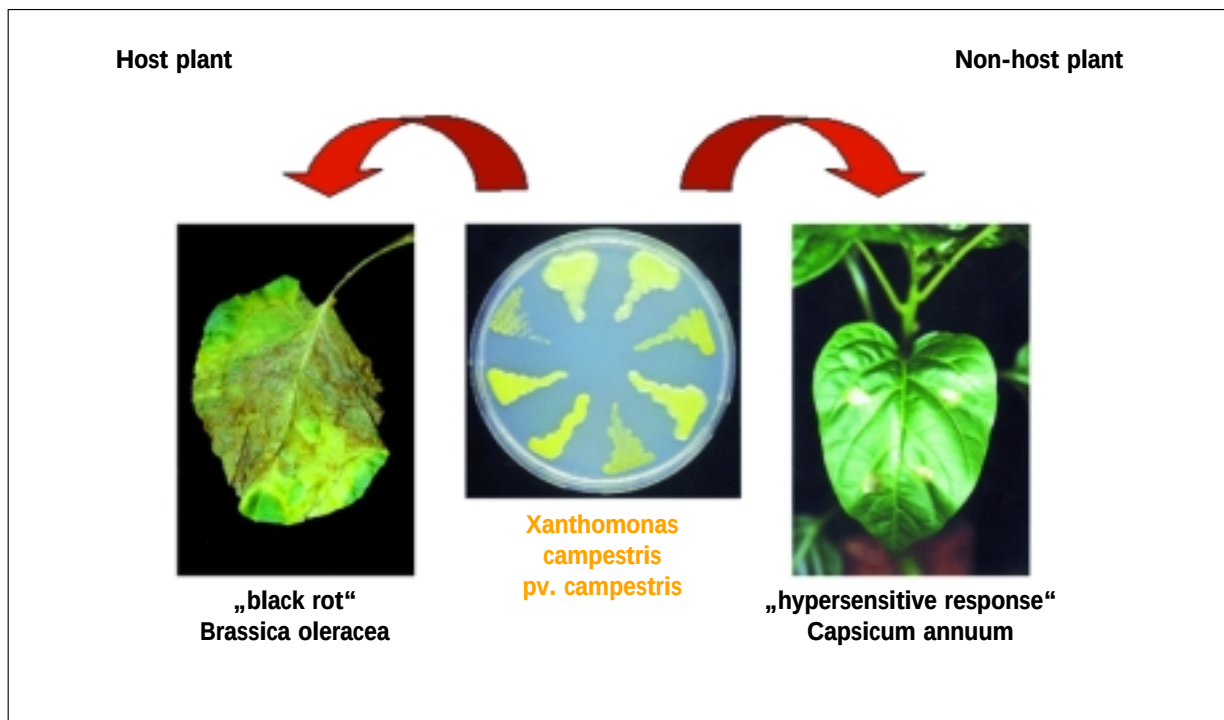


Abbildung 4: Interaktion des phytopathogenen Bakteriums *Xanthomonas campestris* pv. *campestris* mit Wirts- und Nichtwirtspflanzen. *X. campestris* pv. *campestris* kann auf Petrischalen mit einem festen Medium angezogen werden (Bildmitte). Die Bakterien erzeugen hier eine große Menge eines gelben Schleims, der in der Lebensmittelindustrie als Verdickungsmittel in Salatsoßen oder anderen Zubereitungen verwendet wird. Bei einer Infektion von Wirtspflanzen (links) kommt es zur Ausbreitung der Bakterien und zum Absterben der Pflanze. Die Krankheit wird als Schwarzaderfäule bezeichnet. Nicht-Wirtspflanzen wie die Paprika (rechts) reagieren auf eine Infektion mit Pflanzenabwehr. Es kommt nicht zur Erkrankung. Wissenschaftlich wie wirtschaftlich interessant ist hier die Frage nach den Mechanismen der Erkennung von Pflanze und Bakterium.

Das zweite bearbeitete Modell ist die Symbiose zwischen dem Bodenbakterium *Sinorhizobium meliloti* und der Wirtspflanze *Medicago truncatula* (Abbildung 5). Hier handelt es sich um ein für Bakterien und Pflanze nützliches Zusammenleben. Die Bakterien erhalten von der Pflanze Kohlenhydrate und geben gebundenen Stickstoff an die Pflanze ab. Zu Beginn der Symbiose infizieren die Bakterien die Wurzel und induzieren die Bildung eines neuen Organs, des so genannten Wurzelknöllchens. Hier infiziert *S. meliloti* die Pflanzenzellen, ohne die Abwehr der Pflanze zu stimulieren. Auch in diesem Forschungsvorhaben werden die vom Bakterium sekretierten Proteine untersucht. Erste Resultate zeigen, dass die symbiontischen Bakterien im Unterschied zum Pathogen weniger „aggressive“ Enzyme abgeben. Es finden sich jedoch auch viele Proteine mit bislang unbekannter Funktion. Vergleichend wird dazu auch das Proteom der pflanzlichen Zellwand untersucht. Hier zeigt sich, dass die Zellwand ein sehr dynamischer Raum ist, der für die Interaktion mit symbiontischen wie pathogenen Bakterien von besonderer Bedeutung ist.

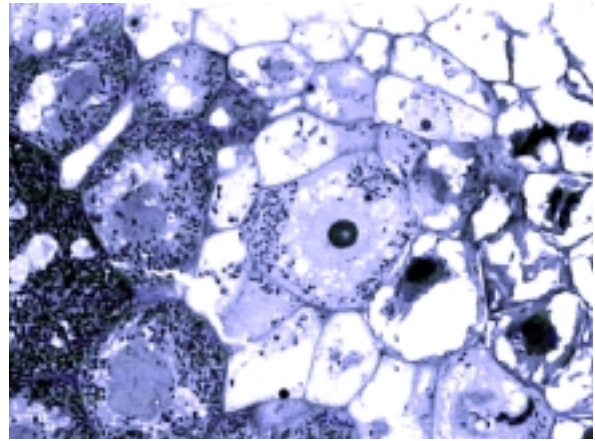


Abbildung 5: Die *Sinorhizobium-meliloti-Medicago-truncatula*-Symbiose. Mikroskopischer Schnitt durch infiziertes Pflanzengewebe. In den pflanzlichen Zellen finden sich hunderte von eingedrungenen Bakterien (dunkle Stäbchen). Offensichtlich kommt es jedoch nicht zur Schädigung der Pflanze. Im Unterschied zur Interaktion mit einem Krankheitserreger ist die Symbiose zwischen Rhizobien und ihren Wirtspflanzen für beide Partner nützlich. Die Pflanze liefert den Bakterien energiereiche Nährstoffe, während die Bakterien die Pflanze mit fixiertem Stickstoff versorgen.

Das Verhalten einer Zelle lässt sich also nur als regulatorisches Netzwerk aus allen Komponenten der verschiedenen Ebenen der Gene, mRNAs, Proteine, Metabolite und der Zell-Struktur begreifen. Dies ist der Ansatz der Systembiologie. Die großen Fortschritte in der Bioinformatik erlauben es, die Resultate der einzelnen Disziplinen zu verknüpfen. Dadurch entsteht gerade ein neues Bild der Zelle.



Hochschuldozent Dr. Karsten Niehaus studierte Biologie in Köln und Bielefeld und habilitierte sich 1999 in Zellbiologie und Genetik. Seine Forschungsschwerpunkte sind die molekularen Grundlagen der Bakterien-Pflanzen Erkennung. Seit 2001 werden diese Arbeiten durch den Aufbau der „DFG-Nachwuchsgruppe für Proteomforschung“ erweitert.



Stickstoff aus der Luft gegriffen

Die Entschlüsselung des Genoms von *Sinorhizobium meliloti* eröffnet neue Forschungsrichtungen

Stefan Weidner

Fakultät für Biologie

Stickstoff ist in der Luft mit einem Anteil von 79 Prozent reichlich vorhanden, jedoch ist er in dieser molekularen Form für Tiere und Pflanzen nicht verwertbar. Pflanzen aus der Familie der Schmetterlingsblütler (zum Beispiel Bohnen, Erdnüsse, Klee und Luzerne) besitzen häufig symbiontische Bakterien, die Stickstoff aus der Luft binden und den Pflanzen in Form von Ammonium verfügbar machen. Eines dieser symbiontischen Bakterien ist *Sinorhizobium meliloti*. Seine vollständige DNA-Sequenz wurde im Jahr 2001 entschlüsselt. Dies eröffnete die Möglichkeit, gezielte gentechnische Veränderungen vorzunehmen und dessen Auswirkungen auf die Gesamtheit aller Gene und Proteine von *S. meliloti* untersuchen zu können.



Abbildung 1: Stickstoff-Fixierung von *Sinorhizobium meliloti* in Symbiose mit Luzerne. Zieht man Luzerne in einem stickstoffarmen Boden an (rechte Pflanze), so ist ihr Wuchs sehr kümmerlich. Versetzt man den Boden jedoch mit *S. meliloti*-Bakterien, wächst die Pflanze deutlich besser. *S. meliloti* und Luzerne gehen eine Symbiose ein, in deren Verlauf es zur Bildung von Wurzelknöllchen kommt. In den Knöllchen leben die *S. meliloti*-Bakterien und fixieren Stickstoff aus der Luft, den sie den Pflanzen in Form vom Ammonium verfügbar machen.

■ Stickstoff-Fixierung in der Symbiose

Fast jeder kennt den Stickstoff-Kreislauf zwischen Atmosphäre und Boden aus dem Biologie-Unterricht. Der größte Teil des in den Boden eingebrachten Stickstoffs wird biologisch im Boden fixiert (ca. 140 Mio t pro Jahr), entweder durch freilebende oder mit Pflanzen in Symbiose lebende Mikroorganismen. Erstaunlich ist, dass dies bei normaler Umgebungstemperatur und Normaldruck geschieht, denn die industrielle Ammoniak-Synthese nach dem Haber-Bosch-Verfahren erfolgt katalytisch bei 500 °C und 200 bar. Besonders symbiontische Bakterien (Rhizobien) tragen mit ca. 300 kg Stickstoff pro Hektar und Jahr enorm zur Stickstoff-Fixierung bei. Rhizobien können mit Pflanzen aus der Familie der Schmetterlingsblütler (Leguminosen) eine Symbiose eingehen, in deren Verlauf es zur Ausbildung von Wurzelknöllchen kommt, in denen die Bakterien den molekularen Stickstoff aus der Luft fixieren und den Pflanzen als Ammoniak verfügbar machen (Abbildung 1). Der so fixierte Stickstoff wird von der Pflanze für die Bildung von Nukleinsäuren, Proteinen und Vitaminen benötigt. Die Bakterien werden im Gegenzug in den Wurzelknöllchen von der Pflanze mit Kohlenhydraten versorgt. Leguminosen gedeihen aufgrund dieser symbiontischen Stickstoff-Fixierung häufig auch auf kargen Böden. Die Bedeutung der Leguminosen für die Ernährung der Weltbevölkerung ist enorm, denn zu dieser Pflanzen-Familie gehören Körner-Leguminosen wie Sojabohne, Erbse, Bohne, Linse oder Erdnuss und Futter-Leguminosen wie Klee und Luzerne. Es ist daher nicht verwunderlich, dass diese Bakterien-Pflanzen-Interaktion seit vielen Jahren erforscht wird, wobei die Symbiose zwischen *S. meliloti* und Luzerne (*Medicago truncatula*) mittlerweile als Modellsystem gilt. Da die Entschlüsselung des Erbguts (Genom) von Bakterien gut etabliert ist, war die vollständige Sequenzierung des Genoms von *S. meliloti* ein konsequenter nächster Schritt (Abbildung 2).

■ Das *S. meliloti*-Genomprojekt

Das Genom von Bakterien besteht meist nur aus einem ringförmigen Chromosom und evtl. aus kleineren, zusätzlichen, ringförmigen DNA-Molekülen, die als Plasmide bezeichnet werden und die für das Bakterium nicht lebensnotwendig sind. Das Genom von *S. meliloti* besteht dagegen aus drei großen DNA-Molekülen: dem eigentlichen ringförmigen Chromosom und zwei als Megaplasmid pSymA und Megaplasmid pSymB bezeichneten

DNA-Molekülen (Abbildung 2). Das Chromosom hat eine Größe von 3,65 Megabasen (Mb) und liegt damit in der gleichen Größenordnung, wie man es von anderen Bakterien kennt. Das Megaplasmid pSymA hat eine Größe von 1,35 Mb und das Megaplasmid pSymB eine Größe von 1,68 Mb. Entsprechend dieser Dreiteilung des Genoms wurde auch das Genom-Sequenzierungsprojekt durchgeführt. Das Chromosom wurde im Rahmen eines EU-Projektes unter Bielefelder Beteiligung sequenziert, die DNA-Sequenz von Megaplasmid pSymA an der Stanford Universität (USA) erstellt. Die Etablierung der Sequenz von Megaplasmid pSymB erfolgte im Rahmen eines Projektes des Bundesministeriums für Bildung und Forschung an der Universität Bielefeld, wobei ein Teil der Sequenz auch an der McMaster Universität (Kanada) erstellt wurde.

Für die Sequenzierung ganzer Genome können verschiedene Strategien angewandt werden. Da DNA-Sequenzierung, also das Auslesen der Abfolge der vier DNA-Bausteine A (Adenin), C (Cytosin), G (Guanin) und T (Thymin), immer nur in relativ kurzen Anschnitten erfolgen kann, ist in jedem Fall eine Unterteilung des zu sequenzierenden DNA-Moleküls notwendig. Im Falle von *S. meliloti* kam eine zweistufige Strategie zur Anwendung, die als *Ordered-Shotgun*-Sequenzierung bezeichnet wird. Zunächst wurde das Genom in große Stücke, in so genannte BAC-Klone unterteilt. Die Methode bringt es mit sich, dass man nach dieser Unterteilung nicht weiß, in welcher Reihenfolge diese BAC-Klone hintereinander gehören. Daher ist zunächst die richtige Anordnung zu ermitteln. Die BAC-Klone müssen lückenlos in einem BAC-Contig (*contiguous*: angrenzend, einander berührend) angeordnet werden. Eine solche BAC-Contig-Karte des Megaplasmid pSymB ist in Abbildung 3 gezeigt. Das vollständige Megaplasmid pSymB konnte durch 26 BAC-Klone lückenlos dargestellt werden. Anschließend wird die DNA-Sequenz der einzelnen BAC-Klone erstellt. Da auch die BAC-Klone für die eigentliche Sequenzierung noch zu groß sind, werden sie zunächst wiederum in kleine Stücke, so genannte *Shotgun*-Klone, unterteilt, die direkt für die DNA-Sequenzierung eingesetzt werden. Folglich wird erst die Sequenz der einzelnen BAC-Klone ermittelt, um dann die Sequenzen der BAC-Klone zur Gesamt-Sequenz zusammen zu setzen. Zwar erscheint diese Strategie aufwendig, sie hat jedoch den Vorteil, dass immer große zusammenhängende Bereiche fertig sequenziert werden. Bei anderen Strategien entstehen größere Bereiche erst zum Ende der Sequenzierung.

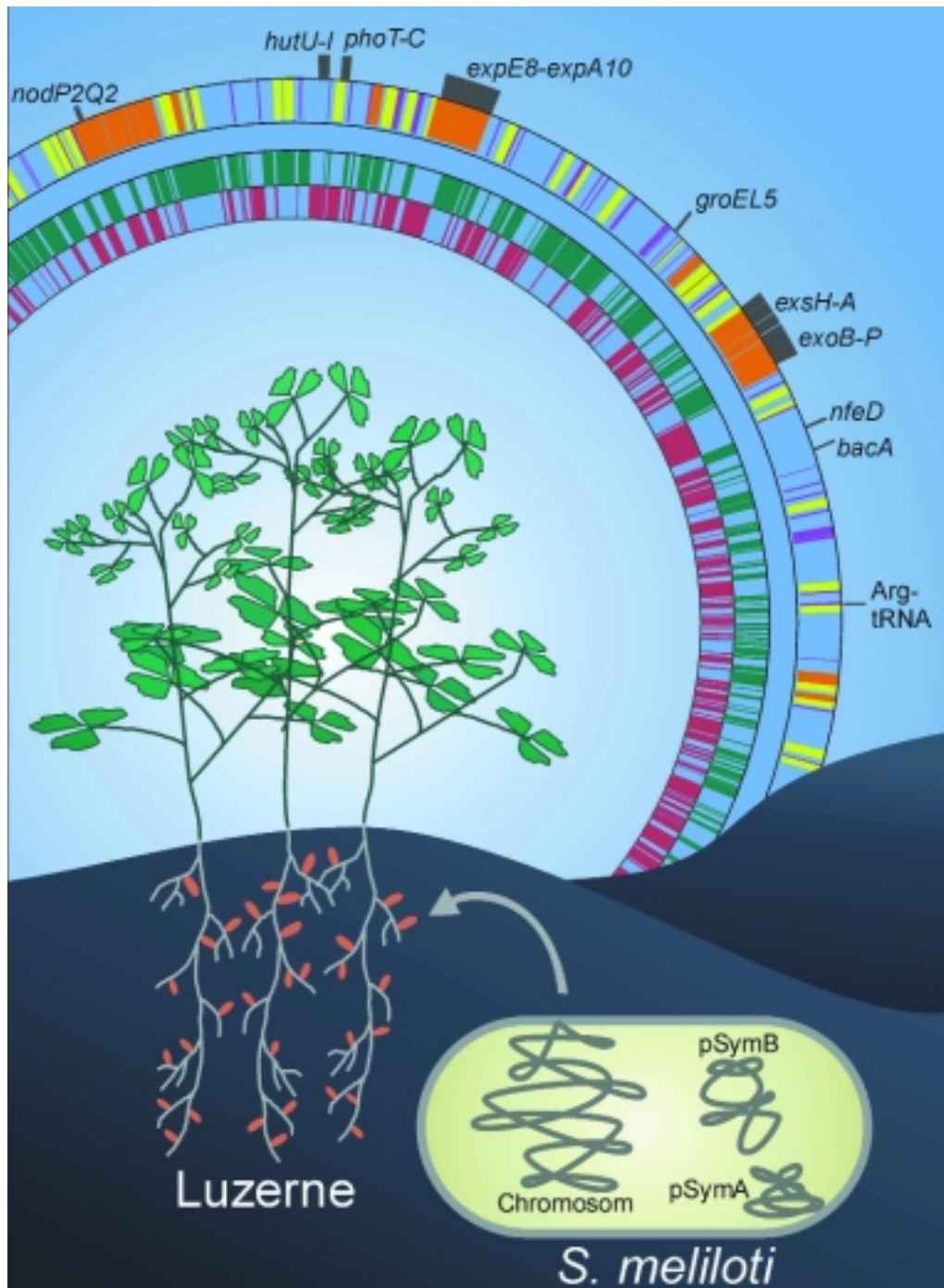


Abbildung 2: Die Symbiose zwischen *S. meliloti* und Luzerne stellt ein Modellsystem dar, das seit vielen Jahren erforscht wird. Das Erbgut von *S. meliloti* besteht aus drei Teilen, dem Chromosom, dem Megaplasmid pSymA und dem Megaplasmid pSymB. Die DNA-Sequenzen aller drei Moleküle wurden in einem internationalen Konsortium ermittelt. Der Kreisausschnitt im Hintergrund zeigt einen Ausschnitt aus der genetischen Karte von pSymB.

■ Unterschiedlicher Ursprung der drei DNA-Moleküle von *S. meliloti*

Ist die DNA-Sequenz eines Genoms ermittelt, müssen die darin verschlüsselten Gene identifiziert und ihnen eine mögliche Funktion zugewiesen werden. Dieser als Annotation bezeichnete Prozess erfordert einen intensiven Einsatz von Software, die von Seiten der Bioinformatik erstellt wird. Im mehrstufigen

Annotationsprozess erfolgt zunächst eine Vorhersage der Position und Länge möglicher Gene im DNA-Molekül. Anschließend werden für alle potentiellen Gene beziehungsweise die darin verschlüsselten Proteine Funktionsvorschläge auf Basis von Computeranalysen und Vergleichen mit Datenbanken erstellt. Letztlich ist jedoch der Wissenschaftler gefragt, die Funktionsvorschläge zu kontrollieren und die endgültige Annotation vorzunehmen.

Einige allgemeine Eigenschaften der drei DNA-Moleküle von *S. meliloti*

Eigenschaft	Chomosom	pSymA	pSymB	Genom
Länge (bp)	3 654 135	1 354 226	1 683 333	6 691 694
G+C-Gehalt	62,7%	60,4%	62,4%	62,1%
kodierte Proteine	3341	1293	1570	6204
Waisen-Gene	5%	11,5%	12,3%	8,2%

Die Annotation des Genoms von *S. meliloti* brachte einige Überraschungen mit sich. Einige Daten der Annotation sind in der Tabelle zusammengefasst. Insgesamt liegen im Genom die Informationen für 6204 Proteine vor. Dies ist deutlich mehr als beispielsweise beim „Haustierchen“ der Bakteriengenetiker, dem Darmbakterium *Escherichia coli*, das für 4289 Proteine kodiert. Obwohl mittlerweile das Genom zahlreicher Bakterien vollständig sequenziert und annotiert wurde, finden sich in neu sequenzierten Genomen immer wieder Gene, die keinerlei Ähnlichkeit zu allen bisher bekannten Genen aufweisen.

Im Genom von *S. meliloti* handelt es sich bei insgesamt über 8 Prozent aller Gene um solche „Waisen-Gene“. Weitere Analysen der drei DNA-Moleküle zeigten, dass sie sich in vielen Eigenschaften voneinander unterscheiden, die teilweise Rückschlüsse auf die Evolution von *S. meliloti* zulassen. Ein charakteristisches Merkmal von DNA ist ihr G+C-Gehalt. Er gibt an, zu wie viel Prozent die DNA aus den Bausteinen G und C zusammengesetzt ist. Wie der Tabelle zu entnehmen ist, unterscheidet sich das Megaplasmid pSymA von Chromosom und Megaplasmid pSymB. Auch weitere Charakteristika,

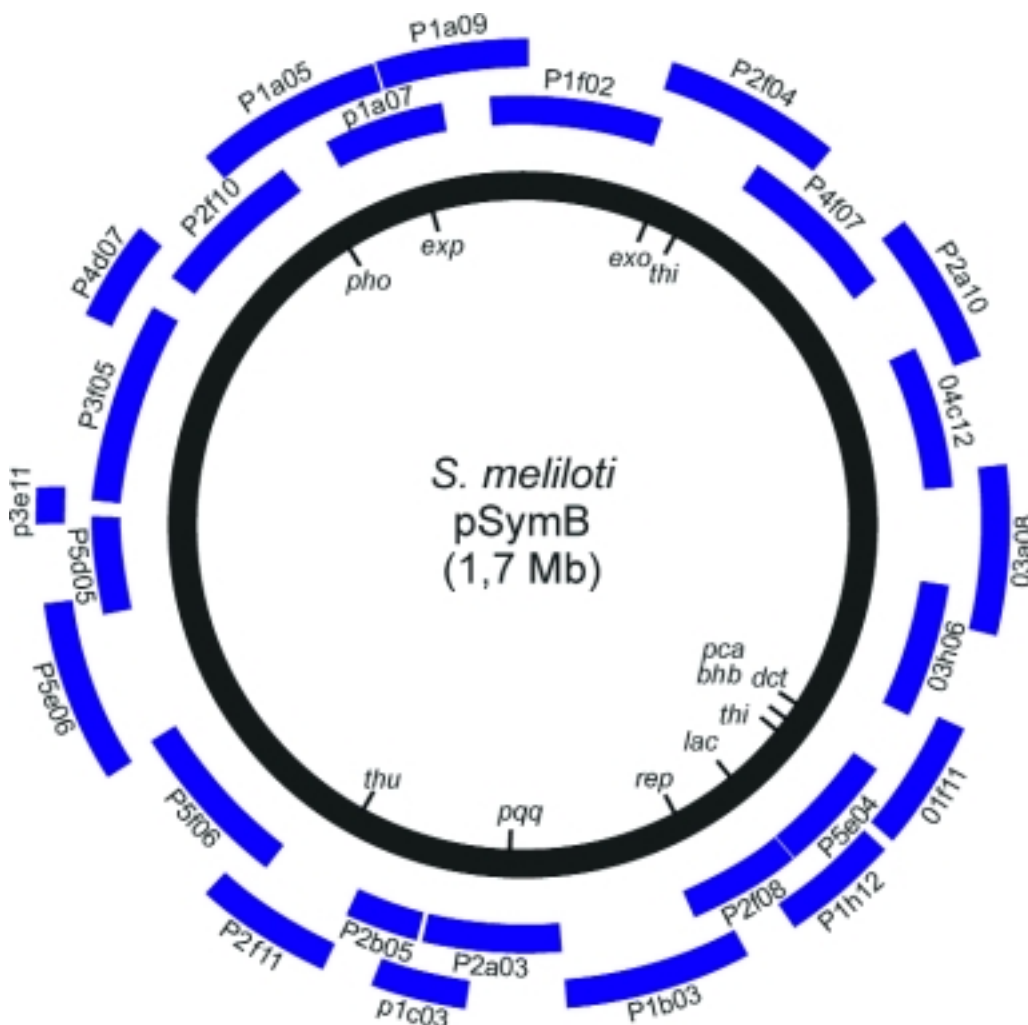


Abbildung 3: BAC-Klon-Contig des Megaplasmid pSymB von *S. meliloti*. Der innere Kreis stellt das Megaplasmid pSymB mit einigen Genen und Genregionen dar. Außen herum sind alle 26 BAC-Klone gezeigt, die einander überlappen und pSymB lückenlos abdecken. Die BAC-Klone wurden einzeln sequenziert und ihre Sequenzen zur Gesamt-DNA-Sequenz von Megaplasmid pSymB zusammengesetzt.

die hier nicht näher aufgeführt sind, deuten darauf hin, dass sich Chromosom und pSymB ähnlicher sind und das Megaplasmid pSymA einige abweichende Merkmale aufweist. Die Annotation der Gene hat ergeben, dass das Chromosom dem eines typischen Bakteriums entspricht, wie es auch aus anderen Genomprojekten bekannt ist. Der „Erwerb“ des Megaplasמידs pSymB hat *S. meliloti* die Fähigkeit verliehen, verschiedenste Nährstoffe, wie sie im Boden vorkommen, aufnehmen und verwerten zu können. Weiterhin trägt pSymB die Information für die Bildung verschiedener Zuckermoleküle (Polysaccharide), die auf der Oberfläche der bakteriellen Zelle angelagert werden und eine entscheidende Rolle bei der wechselseitigen Erkennung von Bakterium und Pflanze und bei der Besiedelung der Wurzelknöllchen spielen. Das andere Megaplasmid, pSymA, ist dagegen notwendig, damit es überhaupt zur Ausbildung der Symbiose kommen kann. Nahezu alle Gene, die für die symbiontische Stickstoff-Fixierung notwendig sind, liegen auf diesem DNA-Molekül. Auch die Fähigkeit, trotz geringerer Verfügbarkeit von Sauerstoff, im Knöllchen leben und atmen zu können, verdankt *S. meliloti* Genen von pSymA. Es war lange eine ungeklärte Frage, ob *S. meliloti* für das Überleben tatsächlich auf beide Megaplasמידe angewiesen ist. Experimentell ist es tatsächlich gelungen, *S. meliloti* von pSymA zu befreien und unter Laborbedingungen lebensfähige *S. meliloti*-Kulturen zu erhalten. Mit Megaplasmid pSymB ist dies bisher nicht gelungen. Dies ist jedoch nicht verwunderlich, da die Annotation gezeigt hat, dass einige Gene auf pSymB lokalisiert sind, die für *S. meliloti* essentiell sind. Auf pSymA konnten keine lebensnotwendigen Gene ausgemacht werden. Vergleicht man diese und weitere Eigenschaften der drei DNA-Moleküle, lässt sich der Schluss ziehen, dass zunächst ein Bakterium existierte, das nur ein Chromosom trug. Im Laufe der Evolution scheint dieses Bakterium dann das Megaplasmid pSymB erworben zu haben. Das Megaplasmid pSymA scheint sich – evolutiv betrachtet – erst vor „kürzerer“ Zeit zu den beiden anderen Molekülen gesellt zu haben.

■ Was kommt nach der Sequenzierung?

Die Verfügbarkeit der vollständigen Sequenz eines Genoms eröffnet neue Ansätze zur weiteren Forschung an *S. meliloti*. Die Aufklärung der Funktion und Bedeutung der einzelnen Gene für bestimmte Stoffwechselvorgänge, bei der symbiontischen Stickstoff-Fixierung und bei anderen Fragestellungen, kann nun gerichtet und zielstrebig vorgenommen werden. Dank der verfügbaren Sequenz ist man in der Lage, gezielt einzelne Gene, die laut ihrer Annotation in den untersuchten Prozess involviert sein könnten, durch Mutation auszuschalten und anschließend die Auswirkungen der Mutation analysieren zu können. Eine völlig neue Dimension bekommt die Analyse von *S. meliloti* durch die Techniken der Transkriptom- und der Proteomanalyse (siehe die Beiträge von Anke Becker und Karsten Niehaus in diesem Heft).



Stefan Weidner studierte Biologie an der Universität Bielefeld mit dem Schwerpunkt Mikrobiologie, Molekularbiologie sowie Genetik. Nach dem Diplom 1993 promovierte er 1996 ebenfalls in Bielefeld. Er koordinierte Verbundprojekte des Bundesministeriums für Bildung und Forschung und ist Leiter der Arbeitsgruppe „Rhizobien-Genomforschung“ am Lehrstuhl für Genetik der Universität Bielefeld. Seit 2003 ist er Geschäftsführer des Bielefelder Centrums für Biotechnologie (CeBiTec).

Bakterien für die industrielle Produktion

Genomforschung am Beispiel des Aminosäuren-produzierenden Bakteriums *Corynebacterium glutamicum*

Jörn Kalinowski

Fakultät für Biologie

Aminosäuren sind die Bausteine aller Proteine. Sie werden entweder vom Organismus selbst hergestellt oder müssen als essentielle Aminosäuren mit der Nahrung aufgenommen werden. Für den Menschen sind etwa zehn der insgesamt 20 Aminosäuren essentiell. Bei Mensch und Tier kommt der Ausgewogenheit der Mengen der aufgenommenen Aminosäuren eine besondere Bedeutung zu. Die jeweils in einer zu geringen Menge aufgenommene Aminosäure begrenzt die Verwertung der übrigen, im Überschuss aufgenommenen Aminosäuren. Dies führt bei der Tierernährung zu geringeren Mastserfolgen und einem Stickstoff-Überschuss in der Gülle. Da die meisten Futtermittel eher unbalancierte Aminosäure-Zusammensetzungen haben, wird das Futter in der modernen Tierzucht durch gezielte Zugabe einzelner, limitierender Aminosäuren optimiert. Neben der Tierernährung finden Aminosäuren Anwendungen als Grundstoffe für chemische Produkte, pharmazeutische Wirkstoffe und als Bestandteile von Infusionslösungen.

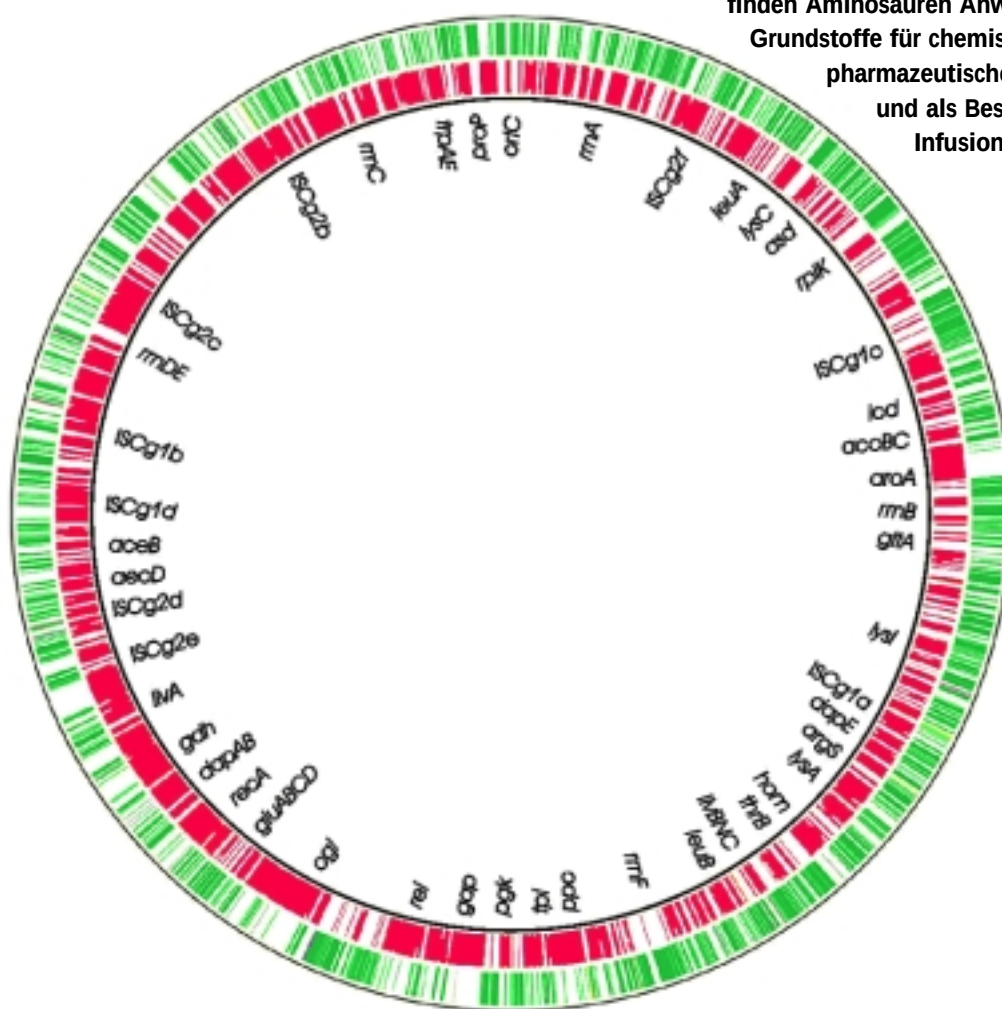


Abbildung 3: Das zirkuläre Genom von C. glutamicum, Stamm ATCC 13032, umfasst etwa 3,28 Millionen Bausteine (Mb). Die mehr als 3000 Gene, die auf beiden Strängen des Genoms liegen, sind als rote und grüne Striche eingezeichnet. Eine Reihe von wichtigen Genen ist auf der Innenseite bezeichnet.



Abbildung 1: Eine industrielle Produktionsanlage für Aminosäuren. Hier produzieren Mikroorganismen in großen Rührkesseln (Fermentern) Aminosäuren im Tonnen-Maßstab. Gezeigt ist eine Reihe von Vor- (obere Reihe) und Hauptkultur-Fermentern (untere Reihe) und das komplexe Netzwerk von Zuleitungsrohren, über die die Mikroorganismen mit Nährstoffen versorgt werden.

■ Die biotechnologische Herstellung von Aminosäuren

Aminosäuren werden industriell entweder durch Aufarbeitung von Eiweiß-haltigen Abfällen (z.B. Tierkadaver, Federn, Haare), durch chemische Synthese oder mit Hilfe biotechnologischer Verfahren hergestellt. Dabei hat die biotechnologische Produktion unter modernen Produktionsgesichtspunkten (Risikopotential von Tierprodukten, Umweltschutz, Nachhaltigkeit) die weitaus besten Entwicklungsaussichten. In diesen biotechnologischen Verfahren wird die Fähigkeit bestimmter Mikroorganismen genutzt, aus einfachen Rohstoffen wie Zuckern und Salzen hochwertige Aminosäuren in großen Mengen und ausschließlich in ihrer biologisch verwertbaren Form zu produzieren. Für diese Produktionsprozesse ist neben dem „Haustier“ der Biotechnologie, *E. coli*, vor allem auch das Bodenbakterium *Corynebacterium glutamicum* geeignet, das bereits als Wildstamm die Aminosäure Glutaminsäure ausscheidet. Durch Mutagenese- und Selektionsverfahren konnten auch *C. glutamicum*-Stämme für die Produktion einer Reihe anderer, wirtschaftlich bedeutsamer Aminosäuren entwickelt werden.

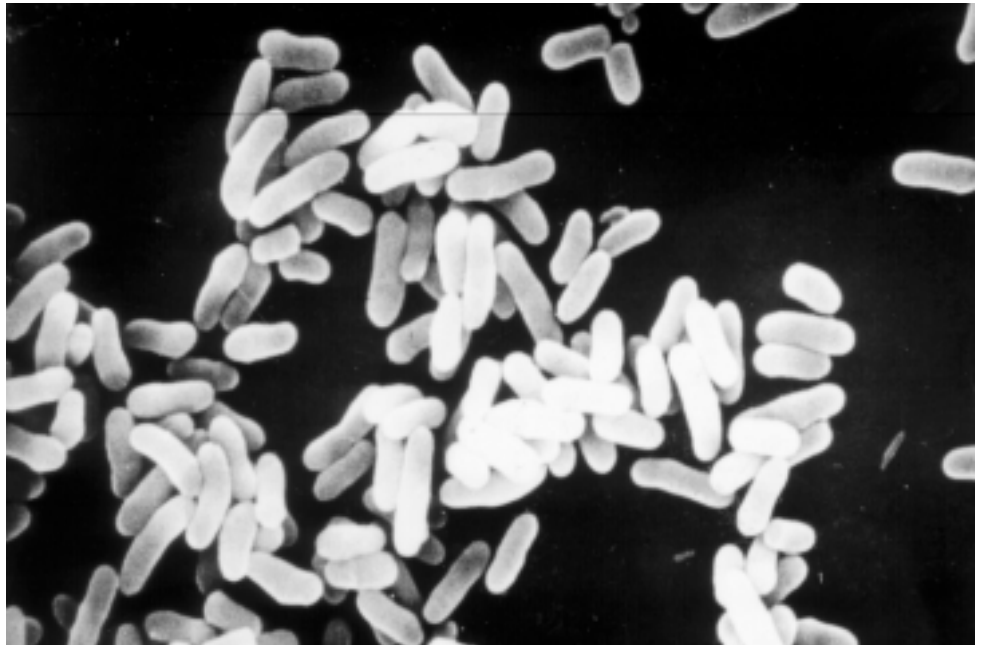
Diese Produktionsprozesse, Fermentationen genannt, laufen in großen Kesseln (Fermentern) bis zum Maßstab von mehreren hundert Kubikmetern ab (Abbildung 1). Auf diese Weise wird der Weltmarkt von zur Zeit mehr als zwei Millionen Tonnen Aminosäuren jährlich zum größten Teil von fermentativ gewonnenen Aminosäuren abgedeckt.

■ Der Produktionsorganismus *Corynebacterium glutamicum*

Der wirtschaftlich bedeutendste Mikroorganismus für die Aminosäure-Produktion ist das harmlose Bodenbakterium *Corynebacterium glutamicum* (Abbildung 2), das in den Fünfzigerjahren des vergangenen Jahrhunderts zuerst in Japan entdeckt und bald danach bereits industriell eingesetzt wurde. *C. glutamicum* zeichnet sich neben seiner Eigenschaft, Aminosäuren in großen Mengen in das die Zellen umgebende Nährmedium auszuscheiden, vor allem durch sein schnelles Wachstum und seine geringen Ansprüche an die Qualität der dazu benötigten Nährstoffe aus. Daher können zur Produktion von hochwertigen Aminosäuren zum Beispiel Abfallstoffe aus der Zuckerindustrie wie Melasse oder Maisquellwasser eingesetzt werden. Da der Produktionsorganismus die Klassifikation GRAS („generally regarded as safe“) besitzt, kann er nach Hitzebehandlung zur Abtötung der Zellen im Produkt verbleiben und seine Biomasse mit verfüttert werden. Dies führt zu ausgesprochen kostengünstigen und abfallarmen Produktionsprozessen.

Um den Produktionsstamm in Bezug auf seine Aminosäure-Produktivität zu optimieren, wird normalerweise die Methode des Mutanten-„Screenings“ angewendet, bei der Bakterien zunächst mit Chemikalien oder UV-Bestrahlung „mutagenisiert“ (zum Mutieren gebracht) und danach die Mutanten auf verbesserte Produktionseigenschaften geprüft werden. In der Vergangenheit konnten so durch

Abbildung 2: Elektronenmikroskopische Aufnahme von Bakterien der Art *Corynebacterium glutamicum* (500-fache Vergrößerung).



multiple Mutagenese-Runden sehr effiziente *C.-glutamicum*-Stämme erzeugt werden, die im Verlauf einer typischen Fermentation (48 bis 96 Stunden) mehr als 100 Gramm einer reinen Aminosäure je Liter Nährmedium produzieren konnten. Die verwendeten Mutagenese-Prozeduren haben allerdings den Nachteil, dass sich hier im Laufe der Stammentwicklung neben den produktionsfördernden Mutationen auch solche anhäufen, die das Wachstum des Produktionsstamms negativ beeinflussen. Um hier Abhilfe zu schaffen, bietet es sich an, mit Hilfe der Gentechnik nur die produktionsfördernden Mutationen auf ein Wildtyp-Bakterium zu übertragen, um so eine optimale Verbindung von guten Wachstums- und Produktionseigenschaften zu erzeugen.

Geeignete gentechnische Methoden für die Übertragung genetischer Informationen in *C. glutamicum* wurden seit 1985 am Lehrstuhl für Genetik der Universität Bielefeld entscheidend mitentwickelt. Erstmals wurden hier Verfahren für den gezielten Genaustausch oder die gezielte Entfernung beliebiger Geninformation aus dem Erbgut von *C. glutamicum* entwickelt. Mit Hilfe der entwickelten Methoden wurde eine Reihe von Genen analysiert, die zum Beispiel für die Produktion der Aminosäure Lysin von entscheidender Bedeutung sind (Pühler und Kalinowski, 1995).

■ Das *C.- glutamicum*-Genomprojekt

Der Ansatz, für die Aminosäureproduktion relevante Gene einzeln zu suchen und zu analysieren, erfährt in Untersuchungen der gesamten Geninformation (Genom) des Produktionsorganismus eine entschei-

dende Weiterentwicklung. Solche Analysen gesamter Genome mit mehreren Millionen Bausteinen sind, gestützt auf technische Fortentwicklungen bei den Analysegeräten und der Computersoftware, seit etwa zehn Jahren möglich und heutzutage auf dem Weg, eine Standard-Technologie zu werden.

Die Entschlüsselung des *C.-glutamicum*-Genoms war aufgrund der besonderen wirtschaftlichen Bedeutung dabei natürlich ein vorrangiges Ziel. Erwartet wurde, mittels Computer-Unterstützung neue Gene zu identifizieren, die für die Aminosäureproduktion wichtig sind, und diese zu patentieren. Es kam dabei zu einem „Wettrennen“ mehrerer Konkurrenzprojekte, die von japanischen, amerikanischen und deutschen Industrieunternehmen finanziert wurden. Eines der Projekte wurde in Bielefeld koordiniert und von der Degussa AG finanziert.

Das *C.-glutamicum*-Genom umfasst etwa 3,28 Millionen Bausteine (Abbildung 3, siehe Seite 29). Es konnten mit Hilfe computergestützter Auswertungsverfahren über 3000 Gene identifiziert werden. Zur weiteren Auswertung und zur Funktionsvorhersage dieser Gene wurden in Bielefeld eigene Computerprogramme entwickelt, die sowohl hochwertige Analyseprogramme bereitstellen als auch aktuellen Vorort-Zugriff auf die weltweit verteilten Gendatenbanken ermöglichen. Die Integration aller Daten aus diesen Analysen liefert eine Basis von überragender Bedeutung sowohl für die molekularbiologische Forschung an *C. glutamicum* als auch für die Rekonstruktion des Stoffwechsels zur Identifizierung produktionsrelevanter Gene (Kalinowski et al., 2003).

Mit Hilfe dieser Analysen ließen sich für eine Reihe von Genen Funktionsvorhersagen machen, die

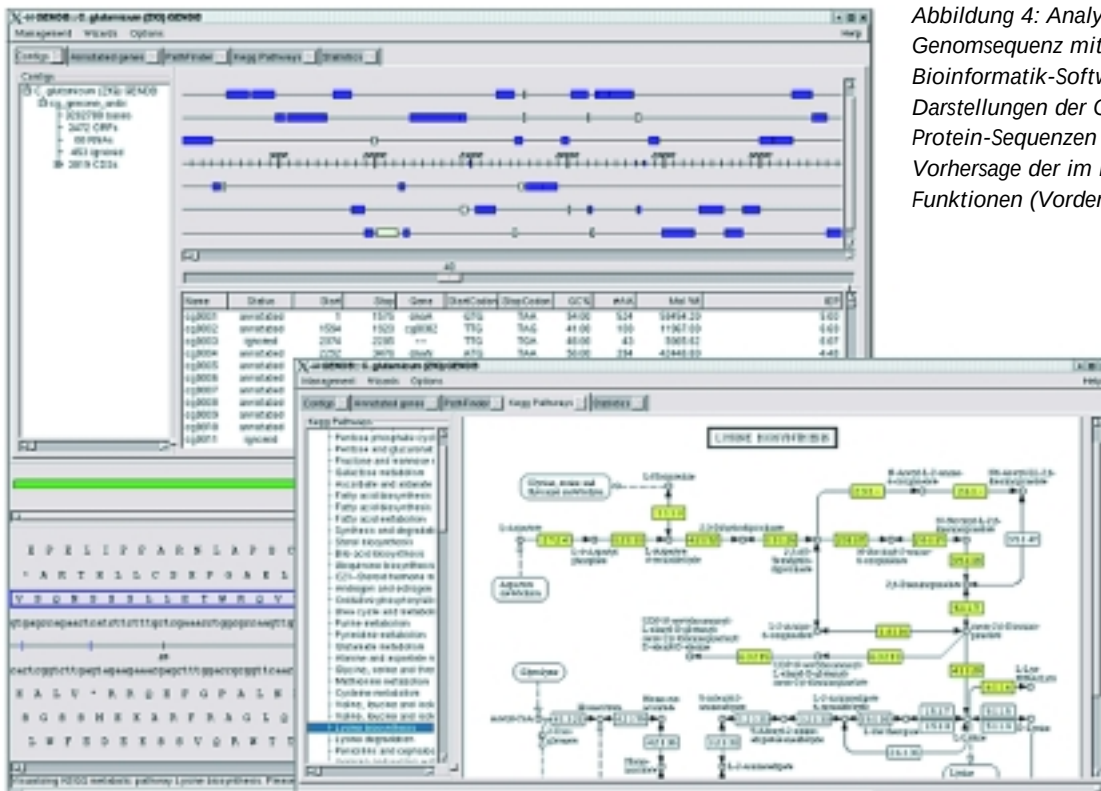


Abbildung 4: Analyse der *C.-glutamicum*-Genomsequenz mit der in Bielefeld entwickelten Bioinformatik-Software GenDB. Gezeigt sind die Darstellungen der Gene und der DNA- bzw. Protein-Sequenzen (Hintergrund) sowie die Vorhersage der im Lysin-Stoffwechsel gefundenen Funktionen (Vordergrund).

eine Beteiligung an der Synthese oder am Transport von Aminosäuren nahelegen (Abbildung 4). Solche „Kandidatengene“ wurden dann für die detaillierte Funktionsuntersuchung ausgewählt und mit den bereits erwähnten gentechnischen Methoden *C.-glutamicum*-Stämme hergestellt, in denen das Gen entweder entfernt oder die Genfunktion verstärkt wurde. Durch den Vergleich dieser Mutanten mit dem Ausgangsstamm in Bezug auf die Aminosäureproduktion konnten aus der Schar der Kandidaten diejenigen Gene ausgewählt werden, die tatsächlich von besonderer Bedeutung für die Aminosäureproduktion sind. Es wird erwartet, dass die optimale Kombination dieser Gene die fermentative Aminosäureproduktion in den nächsten Jahren in die Nähe des theoretischen Maximums bringt.

■ Literatur

J. Kalinowski et al. (2003), The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *Journal of Biotechnology* 104: 5–25.

A. Pühler und J. Kalinowski (1995), Molekulargenetik am Beispiel Aminosäure-produzierender Corynebakterien. *Biologie in unserer Zeit* 25: 221–228.



Dr. Jörn Kalinowski studierte Biologie an der Universität Bielefeld und promovierte hier 1990 mit einer Arbeit zur Genetik des Bakteriums *Corynebacterium glutamicum*. Anschließend war er Mitarbeiter am Lehrstuhl für Genetik der Universität Bielefeld. Seit dem Jahr 2000 leitet er das Zentrum für Genomforschung der Universität Bielefeld. Seine Forschungsschwerpunkte sind Genetik und Genomforschung bei Corynebakterien, Proteom- und Transkriptomforschung bei Mikroorganismen, Bioinformatik mikrobieller Genome und Biotechnologie der Aminosäureproduktion.



Das Genomprojekt *Clavibacter michiganensis*

Neue Perspektiven zum Verständnis der Welkekrankheit der Tomate

Karl-Heinz Gartemann
und Rudolf Eichenlaub

Fakultät für Biologie

Wie bei Mensch und Tier treten auch bei Pflanzen Krankheiten auf, die von Bakterien, Viren oder Pilzen hervorgerufen werden. Im Falle der Tomate ist dies beispielsweise die bakterielle Welkekrankheit, die durch das Bakterium *Clavibacter michiganensis* subsp. *michiganensis* ausgelöst wird. Diese Krankheit kann bis zum Absterben der Tomatenpflanze führen. Die Aufklärung der Mechanismen auf molekularer Ebene, die für die Ausprägung der Welkekrankheit verantwortlich sind, bildet eine der Grundlagen zu einer erfolgreichen Bekämpfung der Krankheit oder aber der Vermeidung einer Infektion.



Abbildung 1: In der Mitte ist eine normale Tomatenpflanze im Alter von etwa acht Wochen zu sehen. Links und rechts Pflanzen, die vier Wochen vorher mit *Clavibacter michiganensis* subsp. *michiganensis* infiziert wurden und die Symptome der Welkekrankheit zeigen, die bei der rechten Pflanze schon zum Absterben geführt hat.

Die Tomate gehört wie Kartoffel, Aubergine und Paprika zu den landwirtschaftlich bedeutenden und in großem Maßstab angebauten Nachtschattengewächsen. In kühlen und feuchten Anbaugeländen herrschen bei Nachtschattengewächsen durch Pilze hervorgerufene Krankheiten vor. Ein Beispiel ist die Krautfäule der Kartoffel, verursacht durch den Pilz

Phytophthora, die 1842–45 in Irland zur „großen Hungersnot“ führte. In wärmeren, trockeneren Anbaugeländen dagegen sind auch bakterielle Erkrankungen bedeutsam, zum Beispiel die Ringfäule der Kartoffel, die durch das Bakterium *Clavibacter michiganensis* subsp. *sepedonicus* verursacht wird. Ein nah verwandtes Bakterium *Clavibacter michi-*

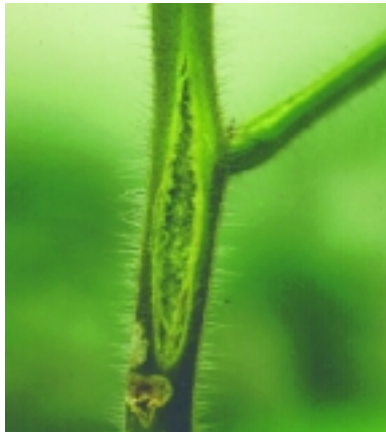


Abbildung 2: Typische Symptome der durch *Clavibacter* ausgelösten Welkekrankheit der Tomate. Oben ein Beispiel der unifacialen Fiederblattwelke, in der Mitte eine Sprossläsion und im Bild unten infizierte Tomatenfrüchte, die ‚bird's eyes‘ aufweisen.

ganensis subsp. *michiganensis* (Cmm), mit dem wir an unserem Lehrstuhl seit mehreren Jahren arbeiten, infiziert Tomaten und verursacht die bakterielle Welke der Tomate. Beide Pflanzenkrankheiten führen weltweit zu erheblichen Ertragsverlusten in der Landwirtschaft, so dass für beide Erreger in der EU und vielen anderen Staaten Quarantänebestimmungen gelten, die eine Verbreitung der Erkrankungen verhindern sollen, denn eine Bekämpfung der einmal ausgebrochenen Krankheit ist bisher nicht möglich, und es sind auch keine resistenten Tomatensorten verfügbar.

■ Das Krankheitsbild der Tomatenwelke

Abhängig vom Zeitpunkt und dem Ausmaß der Infektion treten verschiedene typische Symptome an der Tomate auf (Abbildungen 1 und 2). *Cmm* besiedelt nach der Infektion einen Teil des pflanzlichen Gefäßsystems, das Xylem, das für den Transport von Wasser und Nährsalzen verantwortlich ist (Abbildung 3). Die Bakterien vermehren sich im Xylem und breiten sich über die gesamte Pflanze aus. Allein die Vermehrung der Bakterien im Xylem führt bereits zu einer geringen Wachstumseinbuße der Pflanze, die wahrscheinlich durch eine Unterversorgung mit Wasser in den Blättern und im Sproß hervorgerufen wird.

Die eigentliche Krankheit dagegen äußert sich im Verwelken der Pflanze, das typischerweise zuerst einseitig an einem Teil der Fiederblätter eines Blattes beginnt, während die gegenüberliegenden Fiederblätter noch unbeeinflusst sind. In späteren Stadien verläßt *Cmm* das Xylem und dringt in die umliegenden Gewebe ein. Der Stengel der Pflanze kann aufplatzen, es entstehen so sogenannte Sprossläsionen, und die Pflanze stirbt ab. Allerdings können manche infizierte Pflanzen noch zur Blüte kommen und reife Tomaten ausbilden. Diese Früchte weisen charakteristische Flecken („bird's eyes“, Abbildung 1) auf, da sie ebenfalls von *Cmm* infiziert sind. Das Bakterium dringt sogar in die Tomatensamen ein, die somit eine Hauptquelle für erneute Infektionen sind.

Eine weitere Infektionsquelle sind abgestorbene Tomatenpflanzen im Boden, in und an denen *Cmm* für mehrere Jahre überleben kann. Neu angepflanzte Tomaten werden dann über kleine Wunden im Wurzel- oder Sprossbereich infiziert (Abbildung 4, siehe Seite 36). Eine dritte Infektionsquelle ist das Beschneiden der Tomatenpflanzen, bei dem *Clavibacter* durch kontaminierte Werkzeuge übertragen wird.

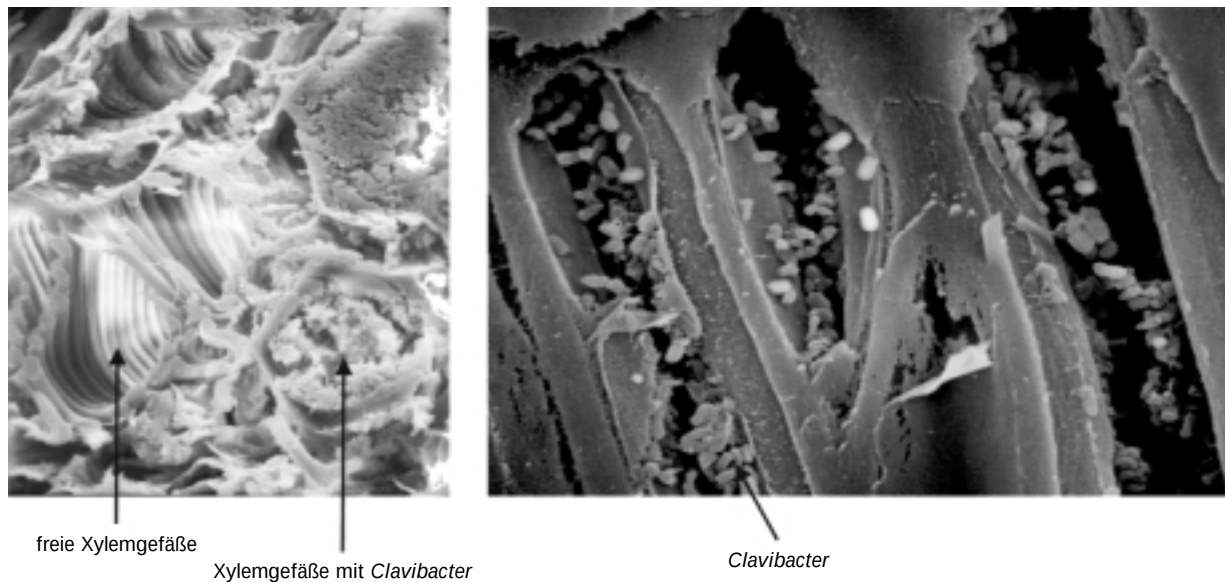


Abbildung 3: Besiedlung des Xylems durch *Clavibacter michiganensis*. Das Xylem sind Gefäßbahnen, die die ganze Pflanze durchziehen und hauptsächlich für den Wassertransport verantwortlich sind. *Clavibacter* besiedelt das Xylem und nutzt die Gefäße zur Ausbreitung über die Pflanze. Die beiden elektronenmikroskopischen Aufnahmen zeigen Xylemgefäße im Querschnitt (rechts) und im Längsschnitt (links).

■ Die Virulenzfaktoren des Erregers der Tomatenwelke

Clavibacter michiganensis subsp. *michiganensis* besitzt zwei Plasmide, pCM1 und pCM2, ringförmige DNA-Moleküle, die zusätzlich zum bakteriellen Chromosom vorhanden sind und keine für das Überleben des Bakteriums essentiellen Gene tragen. Sie bieten aber oft unter bestimmten Umweltbedingungen einen Überlebensvorteil, wie zum Beispiel Plasmide, die Resistenzen gegen Antibiotika tragen. Diese Plasmide können unter natürlichen Bedingungen in einem Teil der Bakterienpopulation verloren gehen. Sie können aber auch künstlich im Labor entfernt werden, ohne dass das Bakterium im Wachstum beeinträchtigt wird. Ein Verlust beider Plasmide führt zu dem *Cmm*-Stamm CMM100, der keine Krankheitssymptome mehr hervorruft, jedoch die Tomate effektiv infiziert und besiedelt. Bakterienstämme, die nur jeweils eines der Plasmide enthalten, zeigen Symptome, jedoch einen abgeschwächten Krankheitsverlauf. Dementsprechend sollten auf beiden Plasmiden Gene liegen, deren Genprodukte direkt an der Ausprägung der Welke beteiligt sind.

Um herauszufinden, welche Gene dies sind, bringt man mit gentechnischen Methoden kürzere DNA-Fragmente der beiden Plasmide in den Plasmid-freien *Cmm*-Stamm CMM100, der keine Welke mehr

hervorruft. Anschließend infiziert man im Labor Tomaten mit diesen Bakterien und untersucht, ob wieder Welke auftritt. Dies ist der Fall, wenn ein Gen, dessen Produkt für die Welkeausprägung notwendig ist, auf dem DNA-Fragment vorhanden ist.

Mittels dieser Technik gelang es, zwei Gene zu identifizieren, die direkt für die Ausbildung der Welke verantwortlich sind. Auf dem Plasmid pCM1 ist dies das Gen für eine Cellulase, ein Protein, das an der Zersetzung von Cellulose, die u.a. in pflanzlichen Zellwänden vorkommt, beteiligt ist. Auf dem zweiten Plasmid pCM2 wurde ein Gen *pat-1* entdeckt, dessen genaue Funktion bisher noch unbekannt ist, dessen Genprodukt aber verwandt mit bestimmten Proteasen ist, die auch bei Mensch und Tier vorkommen.

Werden diese beiden Gene aus dem Krankheitserreger entfernt, führt dies zu einem Verlust der Virulenz (der Ansteckungsfähigkeit). Das Wiedereinbringen der beiden Gene in den Stamm CMM100 restauriert die Fähigkeit zur Ausbildung der Welkekrankheit. Die Existenz nicht virulenter Stämme wie CMM100, die die Tomate kolonisieren, bedeutet, dass die Gene, die essentiell für die erfolgreiche Infektion und Ausbreitung in der Pflanze sind, nicht auf den Plasmiden, sondern auf dem Chromosom liegen.

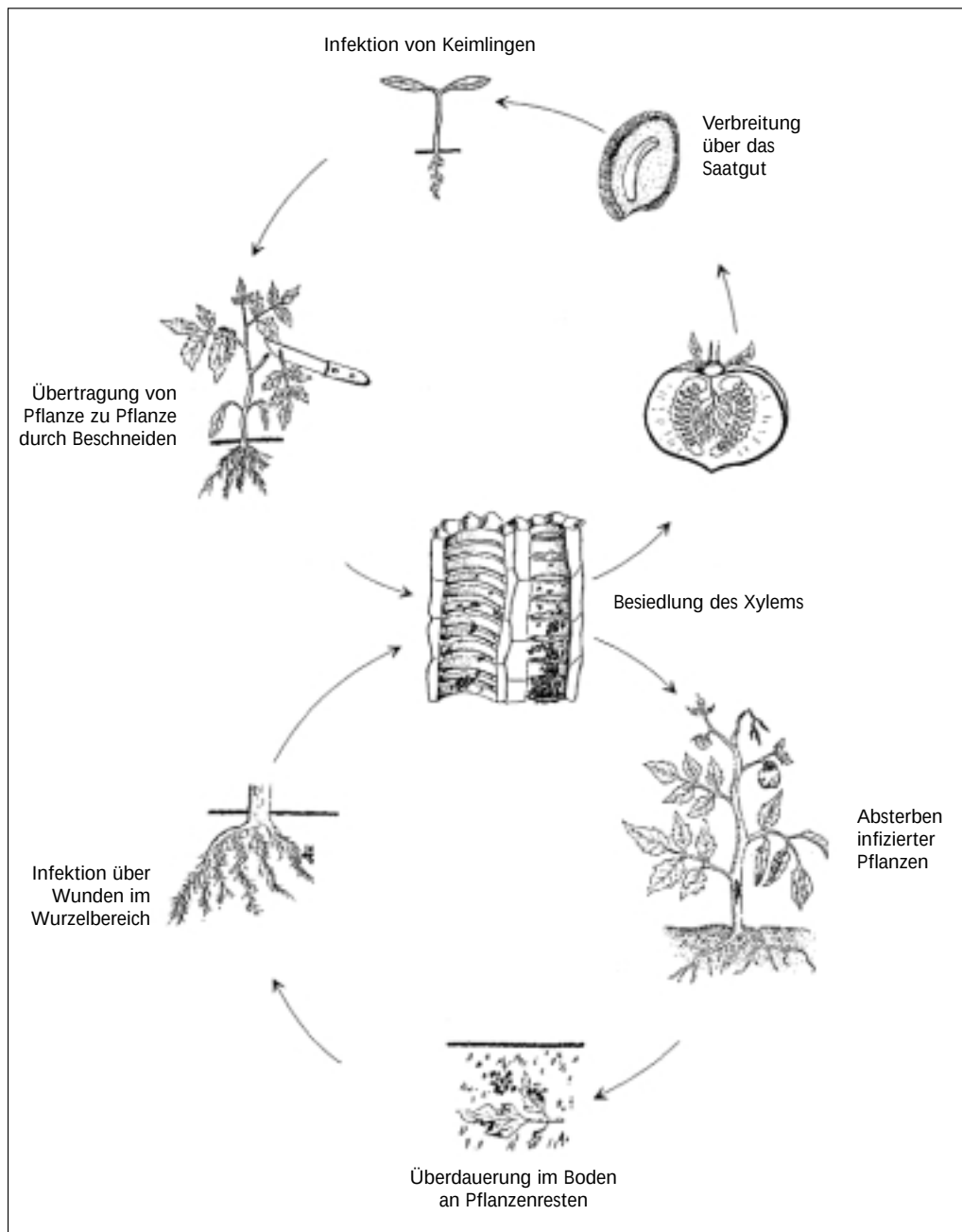


Abbildung 4: Der Infektionszyklus von *Clavibacter michiganensis subsp. michiganensis*.

■ Die Besiedelung der Pflanze als Voraussetzung zur Krankheitsauslösung

Das Vorhandensein der beiden Virulenzgene alleine reicht jedoch noch nicht aus, um die Welkekrankheit auszulösen. Die Untersuchung verschiedener *Cmm*-Feldisolate aus Israel und den Niederlanden zeigte, dass einige nicht virulent sind, das heißt keine Welke auslösen, obwohl sie die Gene für eines oder beide

Virulenzgene *celA* und *pat-1* besitzen. Denselben Phänotyp zeigen im Labor erzeugte Mutanten, denen eine größere chromosomale Region von etwa 100 000 Basenpaaren, die etwa 100 Gene enthalten kann, fehlt. Alle diese Stämme sind nicht in der Lage, die Pflanze erfolgreich zu besiedeln und sich in der Pflanze zu vermehren, und dementsprechend können sie die Welkekrankheit nicht auslösen.

Zwei unterschiedliche Prozesse könnten bei derartigen Stämmen gestört sein. Zum einen kann bereits das Eindringen der Bakterien in die Pflanze, zum anderen das Überleben und die Vermehrung im Xylem beeinträchtigt sein. Beide Prozesse werden von einer Vielzahl von Genen unterschiedlichster Funktion (zum Beispiel die Erkennung der Tomate, die Verwertung von Nahrungsquellen in der Pflanze oder die Unterdrückung einer pflanzlichen Abwehrreaktion) beeinflusst, die häufig additiv oder synergistisch wirken. Die Identifizierung der Komponenten derartiger multifaktorieller Systeme ist schwierig und stellt eine zentrale Herausforderung in der Erforschung der Pathogenitätsdeterminanten dar.

■ Die Sequenzierung des Chromosoms von *Clavibacter michiganensis* subsp. *michiganensis*

Zur Zeit wird im Rahmen eines vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Projektes die vollständige Nukleotidsequenz von *Cmm* bestimmt. Das Chromosom von *Cmm* enthält ungefähr 3,4 Millionen Basenpaare. Für kleine überlappende Stücke von etwa 500 Basenpaaren Länge wird die exakte Abfolge dieser Basenpaare (Sequenz) bestimmt. Aus sehr vielen dieser Einzelsequenzen kann mittels Computerprogrammen die Gesamtsequenz des Chromosoms rekonstruiert werden. Ausgehend von der Gesamtsequenz können die einzelnen Gene identifiziert und in Proteine übersetzt werden. *Cmm* sollte etwa 3000 Gene/Proteine enthalten. Im nächsten Schritt wird für jedes einzelne Protein ein Datenbankenvergleich durchgeführt, mittels dessen die Verwandtschaft zu Proteinen aus anderen Organismen festgestellt werden kann. Dies ermöglicht oft eine erste Aussage über die mögliche Funktion des Genes/Proteins, falls nah verwandte Proteine bereits bekannter Funktion beschrieben sind.

■ Das Genomprojekt als Hilfsmittel zur Aufspürung von Pathogenitätsgenen

Nach Fertigstellung der Basensequenz sollen Gene identifiziert werden, die am Infektions- und Besiedlungsprozess oder an der Ausprägung der Welke beteiligt sein könnten. Zum einen ist – ausgehend von der Sequenzinformation – teilweise direkt eine Abschätzung der Funktion des kodierten Proteins möglich. Zum anderen kann dies über einen Vergleich mit bereits bekannten Genen/Proteinen anderer, besser untersuchter pflanzenpathogener Krankheitserreger erreicht werden. Ob so identifizierte

Gene wirklich eine Rolle im Krankheitsablauf spielen, muss dann über eine gezielte Inaktivierung der Gene und sich anschließende Virulenztests an Tomaten überprüft werden.

In der Zukunft kann die Identifizierung von Genen/Proteinen, die für die Kolonisation und Welkeausprägung notwendig sind, ein Ausgangspunkt für die Entwicklung von Behandlungsmethoden oder zur Züchtung resistenter Tomatensorten sein.

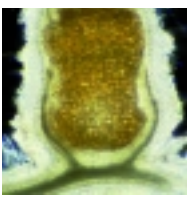
Zudem werden die Daten aus dem Genomprojekt eine Verbesserung der Diagnostik für *Cmm* erlauben. Sehr genaue und schnelle diagnostische Verfahren basieren heute zumeist auf dem Nachweis spezifischer DNA-Sequenzen (z.B. bei Vaterschaftstests und in der Kriminalistik). Die Kenntnis der Genomsequenz sollte die Auswahl von Genen, die sowohl hochspezifisch für *Cmm* (im Gegensatz zu anderen nah verwandten Bakterien) als auch für krankheits-erregende *Cmm*-Stämme (im Gegensatz zu nicht virulenten Stämmen) sind, zur Entwicklung verbesserter Nachweisverfahren ermöglichen. Damit könnte die Saatgutzertifizierung, die zur Vermeidung von *Cmm*-Infektionen nötig ist, vereinfacht und zuverlässiger werden.



Prof. Dr. Rudolf Eichenlaub studierte Biologie an den Universitäten Frankfurt und Bochum. 1973 promovierte er an der Ruhr-Universität Bochum und arbeitete dort bis 1983 als wissenschaftlicher Assistent am Lehrstuhl Biologie der Mikroorganismen, unterbrochen durch einen zweijährigen Forschungsaufenthalt an der University of California, San Diego. 1981 habilitierte er sich mit Arbeiten zur DNA-Replikation des Plasmides F. 1983 folgte er einem Ruf auf eine Professur für Genetik an der Universität Hamburg. Seit 1985 ist er Lehrstuhlinhaber für Gentechnologie/Mikrobiologie an der Fakultät für Biologie der Universität Bielefeld. Seine Forschungsgebiete umfassen den Abbau von Schadstoffen durch Bodenbakterien und die Biologie von Bakterien, die Pflanzenkrankheiten hervorrufen.



Dr. Karl-Heinz Gartemann studierte Biologie an der Universität Bielefeld und schloss sein Studium 1992 mit dem Diplom ab. 1999 promovierte er dort mit Arbeiten zum bakteriellen Schadstoffabbau. Seitdem arbeitet er als wissenschaftlicher Assistent an der Untersuchung pflanzenpathogener Bakterien.



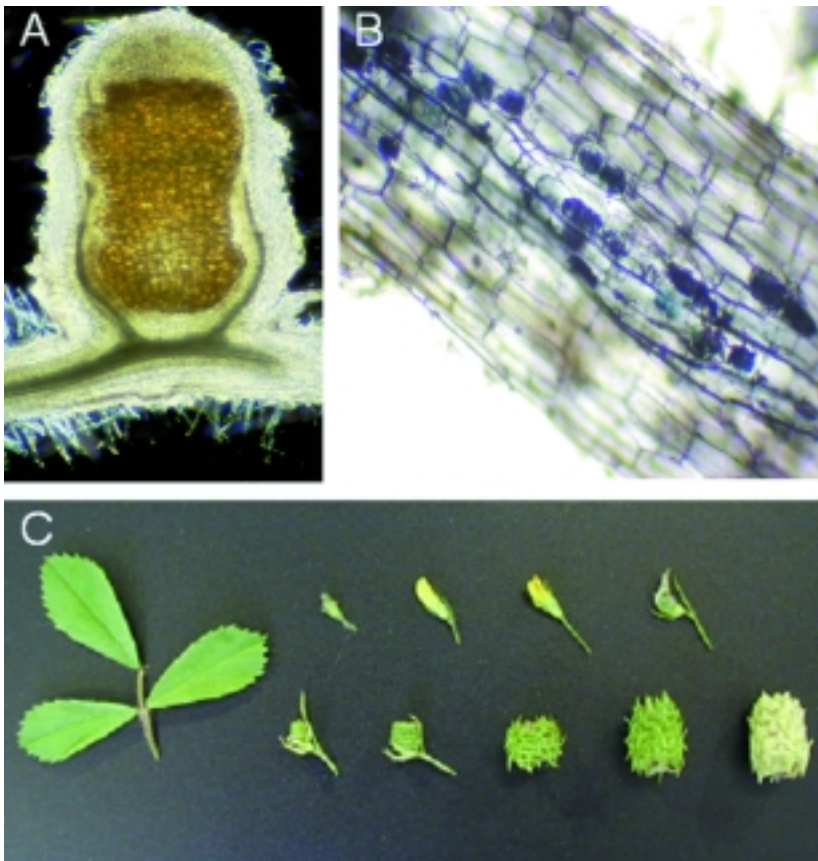
Optimierung von proteinreichen Samen aus Hülsenfrüchtlern

Genomforschung bei Pflanzen
aus der Familie der Leguminosen

Helge Küster

Fakultät für Biologie

Die Familie der Leguminosen ist eine der artenreichsten Pflanzenfamilien. Bekannte europäische Vertreter sind die Erbse und die Luzerne. Zwei Eigenschaften machen Leguminosen zu interessanten Objekten für Pflanzengenetiker. Einerseits ist dies die Fähigkeit, im Wurzelbereich Symbiosen sowohl mit Bakterien als auch mit Pilzen einzugehen, durch die eine optimale Versorgung mit Stickstoff, Phosphaten und Mineralien erreicht wird. Andererseits weisen Blätter und Samen dieser Pflanzen einen hohen Anteil an Proteinen auf. Da dies von landwirtschaftlicher Bedeutung ist, rücken Leguminosen zunehmend ins Zentrum von Untersuchungen, die sich moderner Verfahren der Genomforschung bedienen. Langfristiges Ziel ist es, die Züchtung von Linien zu ermöglichen, die als Quelle hochwertiger pflanzlicher Proteine im Rahmen einer nachhaltigen Landwirtschaft dienen können.



Charakteristische Organe der Modell-Leguminose *Medicago truncatula*.

A. Schnitt durch ein Wurzelknöllchen, das in einer Symbiose mit dem Bodenbakterium *Sinorhizobium meliloti* entstanden ist. Im zentralen Bereich des Knöllchens sind die symbiontischen Bakterien in der Lage, Luftstickstoff zu binden.

B. Mykorrhizierte Wurzel, in der pilzliche Fäden und bäumchenförmige Arbuskeln, die einige Wurzelzellen komplett ausfüllen, durch Anfärbung lila hervorgehoben sind. Durch die Arbuskeln importieren die Wurzelzellen Phosphate und Mineralien, die von den pilzlichen Fäden im Boden aufgenommen und in die Wurzel transportiert werden.

C. Entwicklungsreihe von Blüten und Hülsen. Links zum Vergleich ein Blatt von *Medicago truncatula*. Charakteristisches Merkmal von *Medicago truncatula* sind die tönnchenförmigen Hülsen, die mit spitzen Ausläufern versehen sind.

Medicago-truncatula-Genomprojekte unter Beteiligung der Universität Bielefeld

Europäische Union

Projekt im Fünften Rahmenprogramm:
„MEDICAGO: Integrated structural, functional and comparative genomics of the model legume *Medicago truncatula*“
<http://medicago.toulouse.inra.fr/EU>



Deutschland

Deutsche Forschungsgemeinschaft Schwerpunktprogramm:
„Molekulare Grundlagen der Mykorrhiza-Symbiosen“
<http://MoIMyk.Genetik.Uni-Bielefeld.DE>



Zusammenstellung von Medicago-truncatula-Genomprojekten, an denen der Lehrstuhl für Genetik beteiligt ist.

Mit über 650 Gattungen und ca. 18 000 Arten stellen Leguminosen (Schmetterlingsblütler) die drittgrößte Familie höherer Pflanzen. Die Vielfalt dieser Familie zeigt sich darin, dass ihre Mitglieder sowohl Kräuter als auch Bäume umfassen. Weltweit stellen Leguminosen gleich nach den Getreiden die Familie mit der größten Bedeutung in der Landwirtschaft dar. Dies begründet sich darin, dass Leguminosen aufgrund der Ausbildung proteinreicher Blätter und Samen sowohl ausgezeichnetes Viehfutter als auch wertvolle Bestandteile der menschlichen Ernährung darstellen.

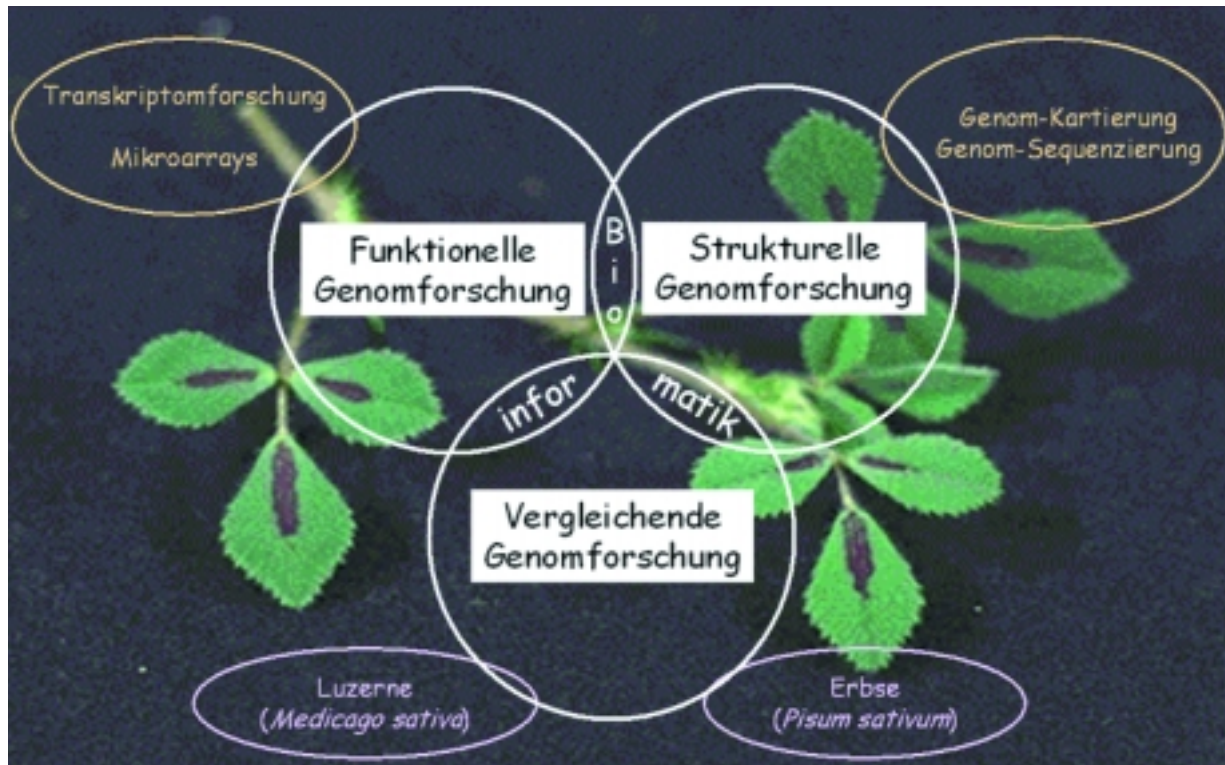
■ **Überführung von Luftstickstoff in organische Verbindungen durch Symbiose mit Bodenbakterien**

Eine herausragende Eigenschaft der Leguminosen ist sicherlich ihre Fähigkeit, eine Symbiose mit Bodenbakterien aus der Familie der Rhizobien einzugehen. Im Zuge dieser Symbiose werden von den bakteriellen Mikrosymbionten so genannte Wurzelknöllchen an den Pflanzenwurzeln erzeugt und anschließend besiedelt (siehe auch: *Forschung an der Universität Bielefeld* 19/1999, S. 38–45). Die Bakterien werden hierbei in die pflanzlichen Zellen aufgenommen und sind jetzt in der Lage, molekularen Stickstoff aus der Luft in Ammonium zu überführen und der Pflanze zur Verfügung zu stellen. Zur Durchführung dieser sehr energieaufwendigen biologischen Stickstoff-Fixierung werden die Bakterien mit Kohlenstoff-Ver-

bindungen versorgt, die letztlich aus der pflanzlichen Photosynthese stammen. Aus ökologischer Sicht ist die biologische Stickstoff-Fixierung besonders vorteilhaft, da sie ein Wachstum der Leguminosen ohne zusätzliche Stickstoffdüngung erlaubt. Die Produktion dieser „chemischen“ Düngemittel trägt durch Verbrennung von fossilen Brennstoffen zum Treibhauseffekt bei, und eine unsachgemäße Anwendung von Nitratdüngern kann zur Auswaschung der Nitrate ins Grundwasser und zu einer nachfolgenden Eutrophierung (Überdüngung) von Gewässern führen.

■ **Aufnahme von Phosphaten und Mineralien aus dem Boden durch Mykorrhizasymbiose mit Bodenpilzen**

Im Gegensatz zum spezifischen Vorkommen der Wurzelknöllchensymbiose bei Leguminosen sind etwa 80 Prozent aller Landpflanzen in der Lage, Mykorrhizasymbiosen mit Glomus-Pilzen auszubilden. Bei einer Form der Mykorrhiza, der Endomykorrhiza, kommt es – ganz ähnlich zur Wurzelknöllchensymbiose – zu einer Besiedlung von Wurzelzellen durch einen pilzlichen Zellfaden. Charakteristisches Symbioseorgan ist das Arbuskel, eine stark verzweigte Endstruktur des Zellfadens, die große Teile der infizierten Wurzelzellen ausfüllt. Durch diese Endosymbiose optimieren Pflanzen ihre Versorgung mit Phosphaten und Mineralien aus dem



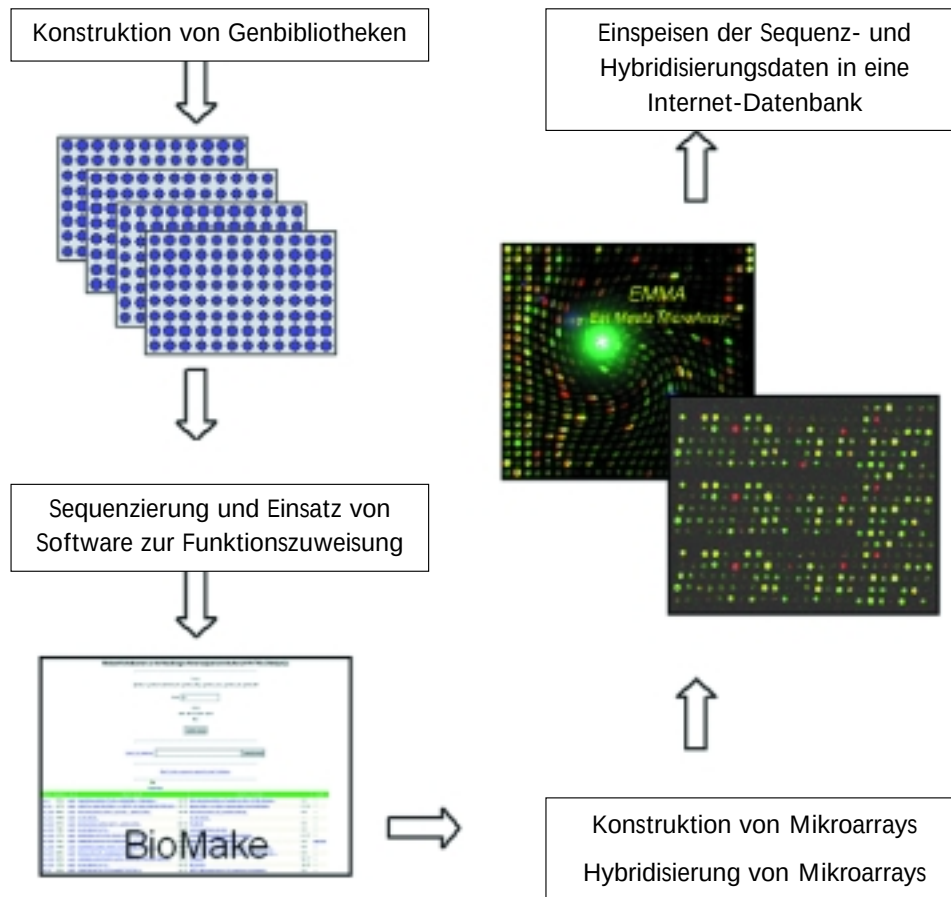
Organigramm des EU-Projekts „MEDICAGO“. Das Projekt integriert drei Teilbereiche der Genomforschung: funktionelle Genomforschung, strukturelle Genomforschung und vergleichende Genomforschung. Die zentrale Position der Bioinformatik zur Verknüpfung der erhaltenen Daten ist angedeutet. Wichtige Techniken, die in der funktionellen und strukturellen Genomforschung angewandt werden, sind angegeben.

Boden. Auch in dieser Symbiose versorgt die Pflanze den pilzlichen Mikrosymbionten mit energiereichen Kohlenstoffverbindungen.

■ Große Mengen an Proteinen in Blättern und Samen der Leguminosen

Neben der einzigartigen Fähigkeit zur Ausbildung von zwei verschiedenen Wurzelsymbiosen zeichnen sich Leguminosen durch die Bildung ungewöhnlich proteinreicher Blätter oder Samen aus. Anhand dieser Merkmale kann zwischen Futter- und Körnerleguminosen unterschieden werden. Eine bekannte Vertreterin der Futterleguminosen ist die Luzerne, als wichtige Körnerleguminose soll hier die Erbse genannt werden. Letztlich ist die Bildung proteinreicher Organe eine Konsequenz aus der Fähigkeit zur biologischen Stickstoff-Fixierung, die es den Leguminosen erlaubt, ungewöhnlich große Mengen Stickstoff-

haltiger Aminosäuren zu produzieren und in Form von Proteinen zu speichern. Diese Eigenschaft macht Pflanzen der Gattungen *Pisum* (Erbsen), *Vicia* (Ackerbohnen), *Phaseolus* (Bohnen), *Lens* (Linsen) und *Cicer* (Kichererbsen) zu wichtigen Bestandteilen auch der menschlichen Ernährung. Insbesondere in Afrika, Asien und Südamerika liefern diese Leguminosen einen essentiellen Beitrag zur menschlichen Ernährung, während der Verzehr von Leguminosen in Europa rückläufig ist. Im Zuge der BSE-Krise werden in letzter Zeit allerdings zunehmend heimische Körnerleguminosen als Alternative für die Produktion hochwertiger Nahrungsproteine gesehen. Ein unter dem Gesichtspunkt des Aufbaus einer nachhaltigen Landwirtschaft wünschenswerter Nebeneffekt wäre, dass die ernährungsphysiologisch hochwertigen proteinreichen Samen unter Ausnutzung der biologischen Stickstoff-Fixierung und somit ohne den Einsatz von Nitratdüngern produziert werden könnten.



Organisation des DFG-Schwerpunktprogramms „MoIMyk: Molekulare Grundlagen der Mykorrhiza-Symbiosen“. Das Projekt basiert auf der Sequenzierung von Transkriptsequenzen aus Mykorrhizawurzeln und einer anschließenden Funktionszuweisung mit Hilfe der im Projekt entwickelten Software BioMake. Die für die Sequenzierung nötigen Genbibliotheken werden von den Projektpartnern konstruiert und in der Zentralen Einheit für Genomforschung an der Universität Bielefeld sequenziert. Ausgehend von den Sequenzermittlungen werden Mikroarrays konstruiert und in Zusammenarbeit mit den Projektpartnern hybridisiert. Die erhaltenen Daten werden in einer Projektdatenbank über das Internet für die Projektpartner zur Verfügung gestellt.

■ Warum ist es sinnvoll, Leguminosen mit Methoden der Genomforschung zu analysieren?

In den USA werden Leguminosen auf ca. 25 Prozent der landwirtschaftlich nutzbaren Fläche angebaut. Im Gegensatz dazu beschränkt sich der Anbau von Leguminosen in Europa auf nur ca. 5 Prozent der Ackerfläche. Dies liegt vor allem an der Anfälligkeit der existierenden Linien gegen diverse Umweltstressfaktoren und Pathogene sowie an der Tatsache, dass die mit Leguminosen zu erzielenden Erträge stärkeren Schwankungen unterworfen sind als die Erträge aus dem Anbau von Getreiden. Um diese Nachteile in Zukunft ausgleichen zu können und den Anbau von Leguminosen somit auch wirtschaftlich attraktiv zu machen, ist es erforderlich, genetische

Komponenten zu identifizieren, die sowohl Resistenzen gegen Pathogene und Umweltfaktoren vermitteln als auch dazu beitragen können, Erträge langfristig zu stabilisieren. Es liegt auf der Hand, dass komplexe Eigenschaften wie Resistenzen, Ertragsstabilität und -qualität nicht von einzelnen Genen, sondern von Netzwerken miteinander interagierender Gene und Genprodukte kontrolliert werden. Die Aufklärung dieser Interaktionen zur Gewinnung von Informationen für die Pflanzenzüchtung erfordert es, Forschungsstrategien einzusetzen, die in der Lage sind, möglichst alle, mindestens aber sehr viele Gene eines Organismus gleichzeitig zu analysieren. Ein solcher Ansatz ist ohne die Anwendung moderner Methoden der Genomforschung undenkbar.

■ *Medicago-truncatula* als geeignete Modellpflanze für die Genomforschung an Leguminosen

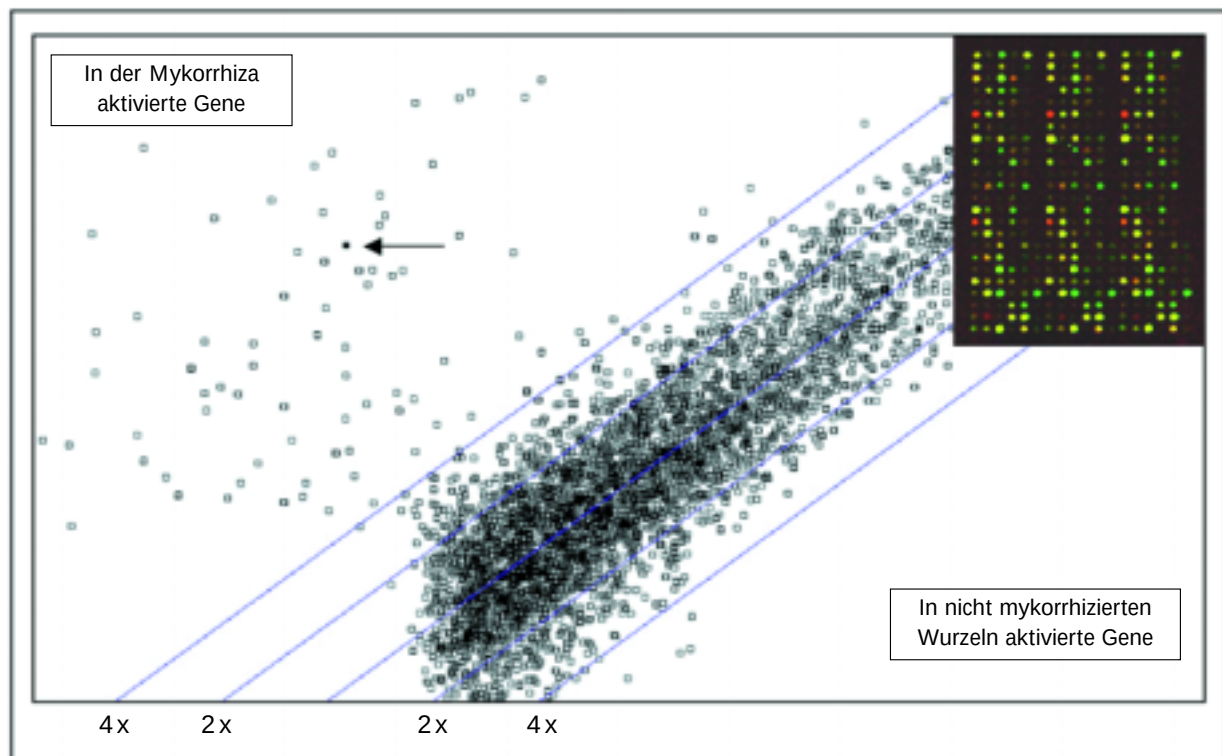
Bisher wurde in der Pflanzengenomforschung vor allem die Modellpflanze *Arabidopsis thaliana* aus der Familie der Kreuzblütler bearbeitet, und diese Pflanze war es auch, deren Genom vor zwei Jahren zuerst entschlüsselt werden konnte. Allerdings ist *Arabidopsis thaliana* als Kreuzblütler nicht in der Lage, Wurzelknöllchensymbiosen einzugehen. Außerdem gehört *Arabidopsis* zu den wenigen Pflanzen, die keine Mykorrhizasymbiose ausbilden können. Darüber hinaus sind *Arabidopsis*-Samen

zwar ein gutes Modell für ölreiche Samen, zum Beispiel des Rapses, aber kein geeignetes Modell für Pflanzen mit proteinreichen Samen. Dies gilt auch für den Reis, dessen Genom ebenfalls bereits entschlüsselt werden konnte.

Die Bündelung der Forschungsaktivitäten auf ein Modell für Kreuzblütler (*Arabidopsis*) und ein Modell für Getreide (Reis) war so erfolgreich, dass auch innerhalb der Leguminosen nach einer geeigneten Modellpflanze gesucht wurde. Sie sollte alle Voraussetzungen aufweisen, um eine Genomsequenzierung durchführen sowie moderne Verfahren der Genomforschung einsetzen zu können. In den letzten Jahren wurden mit dem Hornklee *Lotus japonicus* und mit *Medicago truncatula*, einer im Mittelmeerraum beheimateten Verwandten der Luzerne, solche Modell-Leguminosen identifiziert. Insbesondere *M. truncatula* ist nah verwandt mit landwirtschaftlich bedeutenden Futter- und Körnerleguminosen, die aufgrund ihrer ungewöhnlich großen Genome als Modell-Leguminosen ausscheiden. Durch den hohen Verwandtschaftsgrad kann allerdings davon ausgegangen werden, dass Ergebnisse im Modellsystem *M. truncatula* auf Futter- und Körnerleguminosen übertragen werden können. Mittlerweile haben sich in den USA und Europa eine Reihe von Genomforschungsprojekten formiert, die verschiedene Aspekte der Biologie der Leguminosen anhand des Modellsystems *M. truncatula* analysieren. Im Zentrum dieser Aktivitäten steht ein Genomsequenzierprojekt der „*Legume Genome Initiative*“, das in den nächsten Jahren mit der Sequenz der Modell-Leguminose *M. truncatula* die dritte pflanzliche Genomsequenz aufklären soll.

■ Das Genomforschungsprojekt MEDICAGO der Europäischen Union

Der Lehrstuhl für Genetik an der Universität Bielefeld ist unter Nutzung der Infrastruktur der Zentralen Einheit für Genomforschung an verschiedenen nationalen und internationalen Projekten beteiligt, die sich mit der Leguminosen-Genomforschung am Modell *M. truncatula* beschäftigen. Zunächst ist hier das Projekt „MEDICAGO: Integrated structural, functional and comparative genomics of the model legume *Medicago truncatula*“ zu nennen, das vom Centre National de la Recherche Scientifique in Toulouse, Frankreich, koordiniert wird (<http://medicago.toulouse.inra.fr/EU>). In diesem Projekt haben sich elf Partner aus Frankreich, Großbritannien, Deutschland, den Niederlanden und Ungarn zusammengeschlossen, um – basierend auf



Ergebnis einer typischen Mikroarray-Hybridisierung. Dargestellt ist die Auswertung einer Hybridisierung, in der ein Mikroarray mit ca. 6000 Genen gleichzeitig gegen mit roter Fluoreszenz markierte Transkripte aus Mykorrhizawurzeln und mit grüner Fluoreszenz markierte Transkripte aus nicht mykorrhizierten Wurzeln hybridisiert wurde. Ein Auszug des Fluoreszenzbildes dieser Hybridisierung ist oben rechts zu sehen. Rote Signale markieren solche Gene, die in der Mykorrhiza aktiviert sind, grüne Signale zeigen solche Gene an, die in nicht mykorrhizierten Wurzeln aktiviert sind bzw. in der Mykorrhiza reprimiert (unterdrückt) sind. Durch Überlagerung entstehende gelbe Signale deuten an, dass die entsprechenden Gene in beiden Bedingungen gleich stark aktiv sind. Die relativen

Fluoreszenzintensitäten wurden anschließend auf einer logarithmischen Skala gegeneinander aufgetragen. In der Mykorrhiza aktivere Gene ergeben Datenpunkte im linken oberen Bereich, in der Mykorrhiza reprimierte Gene werden durch Datenpunkte im Bereich unten rechts angezeigt. Blaue Linien deuten die unten angegebenen Aktivierungsfaktoren an, die zentrale blaue Linie markiert Gene, die in beiden Bedingungen gleich stark aktiv sind. Durch einen Pfeil hervorgehoben ist ein ca. 40-fach aktiviertes Gen, das einen sogenannten Phosphat-Transporter spezifiziert. Dieses Protein sorgt dafür, dass das in den pilzlichen Zellfäden angelieferte Phosphat von den Pflanzen in die Wurzelzellen aufgenommen werden kann.

der Erstellung einer Karte des *M. truncatula*-Genoms (strukturelle Genomforschung) – die Voraussetzungen für die Sequenzierung des Genoms dieser Pflanze zu schaffen. Neben den Kartierungsarbeiten liegt ein besonderer Schwerpunkt auf der Transkriptomforschung (funktionelle Genomforschung), mit deren Hilfe von den verschiedenen Projektteilnehmern sowohl die Knöllchen- und Mykorrhizasymbiose als auch die Entwicklung von Blüten und Hülsen untersucht wird. Die Untersuchung der Entwicklung von Blüten und Hülsen wird in Bielefeld durchgeführt und bedient sich der Methodik der Mikroarray-Hybridisierung, die im folgenden Abschnitt näher erläutert wird. Mit Hilfe der vergleichenden Genomforschung hat sich das Konsortium zum Ziel gesetzt, die am

Modellsystem *M. truncatula* gewonnenen Ergebnisse auf landwirtschaftlich relevante Körnerleguminosen zu übertragen. Hierfür wird an der Erstellung vergleichender Genomkarten der verschiedenen Leguminosen gearbeitet.

■ Mykorrhizasymbiose von *Medicago-truncatula* im Schwerpunktprogramm MolMyk der Deutschen Forschungsgemeinschaft

Seit dem Jahr 2000 wird das Schwerpunktprogramm „MolMyk: Molekulare Grundlagen der Mykorrhiza Symbiosen“ (<http://molmyk.genetik.uni-bielefeld.de>) der Deutschen Forschungsgemeinschaft (DFG) vom Lehrstuhl für Genetik koordiniert. Im Rahmen dieses

Projekts versuchen 17 deutsche Arbeitsgruppen, solche Gene zu identifizieren, die für die Mykorrhizasymbiose wichtig sind. Dies geschieht auf zwei Ebenen. Zunächst haben wir gezielt nur solche Gene sequenziert, die in mykorrhizierten Wurzeln aktiv sind. Im Gegensatz zur Genomsequenzierung, die letztlich in der Lage ist, alle Gene eines Organismus erfassen, bedienen wir uns eines Tricks, um schnell an für die Mykorrhizasymbiose relevante Gene zu kommen. Wir haben hierfür nur die Abschriften der Gene (so genannte RNA-Transkripte) aus mykorrhizierten Wurzeln isoliert und anschließend deren Sequenz bestimmt. So konnten wir innerhalb weniger Monate fast 3500 Gene identifizieren, die in mykorrhizierten Wurzeln von *M. truncatula* aktiv sind. Schätzungsweise sind dies gut 10 Prozent der Gene von *M. truncatula*.

Der zweite Ansatz bediente sich der Transkriptomforschung unter Anwendung von so genannten „DNA-Mikroarrays“. Hierbei handelt es sich um Glas-Objektträger, auf die Proben für einige tausend Gene aufgebracht werden können. Die Mikroarrays werden anschließend einer so genannten Hybridisierung unterzogen. Für dieses Verfahren wurden getrennt sowohl die Abschriften solcher Gene isoliert, die in Mykorrhizawurzeln aktiv sind, als auch die Abschriften solcher Gene, die in nicht mykorrhizierten Wurzeln vorkommen. Anschließend erfolgt eine Kopplung von Fluoreszenzfarbstoffen an die RNA-Transkripte, wobei rot fluoreszierende Farbstoffe an die Transkripte aus Mykorrhiza- und grün fluoreszierende Farbstoffe an die Transkripte aus Nicht-Mykorrhiza-Wurzeln gebunden werden. Danach wurden die rot und grün markierten Transkripte im gleichen Verhältnis gemischt und auf den Mikroarray gegeben. Erkennen die markierten RNA-Transkripte Gene auf dem Glas-Objektträger, binden sie an diesen und können anschließend mit einem Fluoreszenz-Scanner identifiziert werden. Rot fluoreszierende Signale markieren dann solche Gene, die in Mykorrhizawurzeln im Vergleich zu nicht mykorrhizierten Wurzeln stärker aktiv sind. Grün fluoreszierende Signale zeigen solche Gene an, die in mykorrhizierten Wurzeln eine geringere Aktivität haben als in nicht mykorrhizierten Wurzeln. Solche Aktivitätsanalysen auf Ebene der Transkripte geben erste Hinweise auf die Relevanz von Genen für die Mykorrhizasymbiose.

Die Stärke der Transkriptomforschung liegt darin, dass die Aktivität einiger tausend Gene in einem einzigen Experiment nachgewiesen werden kann. Natürlich sind bei solchen Analysen viele Varianten in der experimentellen Fragestellung denkbar, so die Untersuchung der Bedeutung von Pflanzenhormonen

oder Umweltfaktoren für die Genexpression in Mykorrhizawurzeln. Ein weiterer Ansatz besteht darin aufzuklären, inwieweit es Übereinstimmungen zwischen der Genaktivierung in der Wurzelknöllchen- und der Mykorrhizasymbiose gibt. Durch die Identifizierung solcher „Symbiosin“-Gene erhoffen wir uns Hinweise auf gemeinsame Prozesse in diesen bedeutenden Endosymbiosen.

Untersuchungen mit Hilfe der Transkriptomforschung führen zu einer enormen Datenfülle und erfordern sowohl einen hohen Bedarf an Speicherkapazität als auch an Software, die in der Lage ist, die Ergebnisse effizient auszuwerten. Besonders wichtig für das MolMyk-, aber auch das MEDICAGO-Projekt ist daher die Zusammenarbeit mit der Bioinformatik-Gruppe der Zentralen Einheit für Genomforschung, die dafür verantwortlich ist, die Ergebnisse unserer Mikroarray-Hybridisierungen in zentralen Projekt-Datenbanken vorzuhalten.

■ Ausblick

Die gesteigerte Bedeutung der Genomforschung an Leguminosen zur Verbesserung von Sorten im Bereich Futtermittel und Nahrungsmittelproduktion

hat unter „New strategies to improve grain legumes for food and feed“ zur Aufnahme in das Sechste Rahmenprogramm der EU-Forschungsförderung geführt. Mittlerweile hat sich unter Beteiligung der Universität Bielefeld ein Konsortium gebildet, das einen Antrag auf Förderung eines „Integrated Project“ in diesem Bereich gestellt hat. Parallel zu konkreten Bemühungen um Forschungsförderung zeichnen sich in den verschiedenen internationalen Genomforschungsprojekten an Leguminosen zunehmend Tendenzen ab, Aktivitäten zu vernetzen, Redundanzen in der Forschung zu vermeiden und eine internationale „Legume Genomics Community“ aufzubauen. Es ist zu erwarten, dass basierend auf der Infrastruktur der Zentralen Einheit für Genomforschung auch die Universität Bielefeld hier ihren Platz finden wird.



*Dr. Helge Küster hat an der Universität Bielefeld und an der University of Edinburgh Biologie studiert und 1995 mit einem Thema aus dem Gebiet der Wurzelknöllchenbildung bei der Ackerbohne *Vicia faba* promoviert. Zur Zeit leitet er die Arbeitsgruppe „Legume Genomics“ am Lehrstuhl für Genetik und der International NRW Graduate School in Bioinformatics and Genome Research. Seine Gruppe beschäftigt sich vor allem mit der Transkriptomforschung bei der Modell-Leguminose *Medicago truncatula* unter Einsatz von Sequenzierungen und Mikroarray-Hybridisierungen. Basierend auf den Daten aus der Transkriptomforschung interessiert sich die Gruppe für die Aufklärung der Funktion und der Expressionsregulation solcher Gene, die entweder in verschiedenen Wurzelsymbiosen oder in der Entwicklung von Blüten und Hülsen aktiv sind. In diesem Zusammenhang liegt ein weiterer Arbeitsschwerpunkt der Gruppe auf der Konstruktion und Analyse transgener Leguminosen.*



Vom Aschenputtel zur Modellpflanze: Das nützliche Unkraut

Methoden und Ziele der Genomforschung bei Pflanzen

Martin Sagasser und
Bernd Weisshaar

Fakultät für Biologie

Pflanzen und pflanzliche Produkte sind die Basis unserer Ernährung. Außerdem dienen sie als Ausgangsmaterial für Medikamente, als Industrierohstoffe und als Bau- und Brennstoff. Die Vielfalt der Stoffwechselleistungen von Pflanzen und ihre große Variabilität und Komplexität werden vom Menschen seit Jahrtausenden genutzt. Darin zeigt sich die immense Bedeutung der Pflanzen für den Menschen. Andererseits lassen genau diese Eigenschaften Untersuchungen an Pflanzen oft besonders schwierig werden. So sind auch heute noch viele pflanzliche Entwicklungs- und Regulationsprozesse nicht oder nur ansatzweise verstanden. Ein Weg, die offenen Fragen zu beantworten, beginnt auf der Ebene der Gene. Genetische Information zu lesen und zu verstehen bildet die Grundlage dafür, Pflanzen insgesamt zu verstehen. Mit diesem Wissen können Nutzpflanzen noch gezielter gezüchtet werden, um auf begrenzter Fläche und mit geringerem Einsatz von Schutzmitteln und Dünger höhere Erträge zu erzielen.



Ackerschmalwand (Arabidopsis thaliana), Rosettenblätter junger Pflanzen. Während die Blätter der linken Pflanze deutlich behaart sind (Wildtyp), sind die der rechten glatt. Die rechte Pflanze trägt eine Mutation in einem Transkriptionsfaktor-Gen. Die sichtbare Veränderung lässt darauf schließen, dass dieses Gen bei der Regulation der Haarbildung eine Rolle spielt.



Rapsblüten. Vier radialsymmetrisch angeordnete Blütenblätter geben der Familie der Kreuzblütler ihren Namen. Auch die Ackerschmalwand gehört zu dieser Familie. Die enge Verwandtschaft erleichtert die Übertragung und Anwendung der am Modell gewonnenen Erkenntnisse.

Traditionell haben Pflanzenforscher die Fragen, die sie interessierten, an ganz unterschiedlichen Arten untersucht. Mais, Sojabohne, Kartoffel oder Weizen haben zwar große Bedeutung für die menschliche Ernährung, als Forschungsobjekte haben sie aber ihre Tücken: Sie benötigen viel Platz und Zeit für ihre Entwicklung und haben häufig extrem große Genome. Die Gesamtheit der genetischen Information des Weizens beispielsweise ist fünfmal so groß wie die des Menschen. Weizen enthält zudem nicht nur zwei, sondern gleich mehrere Chromosomensätze (Polyploidie), was die Untersuchung der Funktionen einzelner Gene und ihrer Varianten sehr erschwert. Ähnliches gilt bei vielen Kulturpflanzen.

Anfang der 80er Jahre setzte sich in der Forschungsgemeinschaft mehr und mehr die Erkenntnis durch, dass das Ziel, die Physiologie und Entwicklung von Pflanzen vollständig zu verstehen, nur erreicht werden kann, wenn sich die diesbezüglichen Anstrengungen auf eine einfache Modellpflanze konzentrieren. Da alle Blütenpflanzen eng verwandt sind, konnte man davon ausgehen, dass die komplette Sequenzierung aller Gene einer repräsentativen Pflanzenart und die Untersuchung der von ihnen kodierten Proteine es ermöglichen würde, grundlegende Erkenntnisse über alle Pflanzen zu gewinnen. Die Wahl der Forscher für ein Modell für Blütenpflanzen fiel damals auf die Ackerschmalwand, *Arabidopsis thaliana*.

■ Das Aschenputtel unter den Kohlgewächsen

Ackerschmalwand ist ein unscheinbares, auch in unseren Feldern und Gärten verbreitetes „Unkraut“. Als Kreuzblütler ist sie nahe verwandt mit Raps, Kohl und Senf. Sie durchläuft ihren Lebenszyklus von der Keimung bis zur Produktion neuer Samen jedoch innerhalb von nur sechs bis acht Wochen. Damit stehen zum Beispiel Ergebnisse von Kreuzungsexperimenten sehr schnell zur Verfügung. Ackerschmalwand ist so klein, dass auf einem Quadratmeter ca. 300 Pflanzen angezogen werden können. Ihr Genom ist mit 125 Mio. Basenpaaren pro einfachem Chromosomensatz eines der kleinsten unter den Pflanzen und 24-mal kleiner als das des Menschen. Es ist „diploid“ (mit einem doppelten, also vollständigen Chromosomensatz ausgestattet) und enthält im Vergleich zu anderen Pflanzen nur sehr wenig sich wiederholende und wenig informative, so genannte repetitive DNA. Diese günstigen Eigenschaften haben Ackerschmalwand so beliebt gemacht, dass weltweit viele tausend Wissenschaftler mit ihr arbeiten.

■ Mehr Fragen als Antworten

Die Entschlüsselung der Genomsequenz der Ackerschmalwand wurde im Jahr 2000 abgeschlossen. Zum ersten Mal stand nun die vollständige DNA-Sequenzinformation einer Blütenpflanze zur Verfügung.

und damit Informationen über alle Gene, die für das Leben einer Pflanze nötig sind. Aber wie so oft in der Wissenschaft wurden auch hier mit einer Erkenntnis mehr neue Fragen aufgeworfen als beantwortet. Eine Vielzahl der 29 000 Gene, deren Existenz aus der Sequenz abgeleitet wurde, war bislang unbekannt und in ihrer Bedeutung für die Pflanze unverstanden. Zum Teil waren zwar Aussagen über die Funktionen einzelner Gene möglich, nicht aber über die Zusammenhänge, in denen sie wirken. Ziel der Forschung ist nun, diese neu gefundenen Gene den biologischen Prozessen in Pflanzen zuzuordnen und dadurch sowohl die Genfunktionen als auch die Entwicklungs- und Regulationsprozesse von (Blüten-)Pflanzen besser zu verstehen.

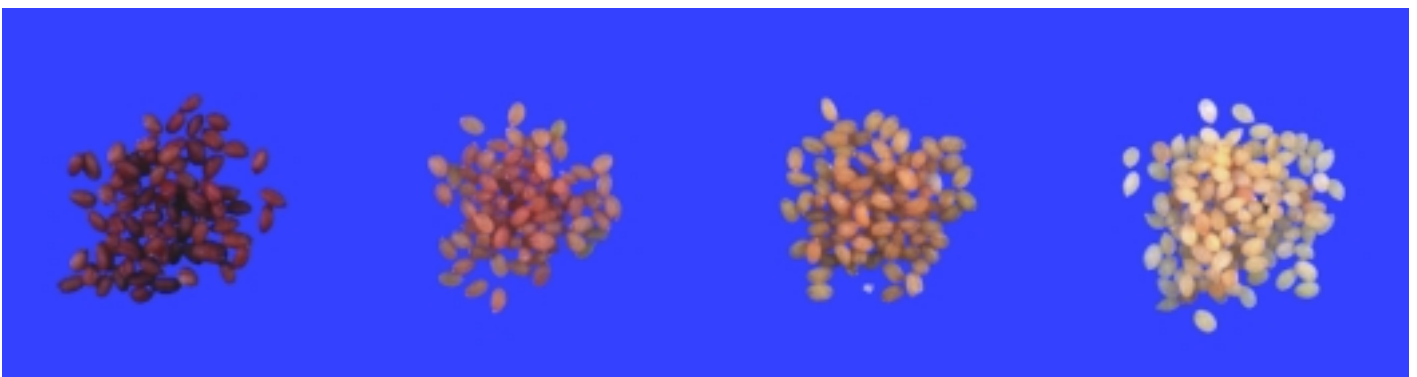
■ Molekulare Schalter gesucht

Am Lehrstuhl für Genomforschung der Universität Bielefeld, der im Februar 2003 seine Arbeit aufnahm, werden vor allem Fragestellungen aus dem Bereich der Kontrolle der Genaktivität bearbeitet. Vielfach ist nämlich nicht nur das Vorhandensein oder die Abwesenheit eines Gens von Bedeutung für den Organismus, sondern insbesondere die korrekte zeitliche und räumliche Regulation der Genaktivität im Verlauf der Entwicklung einer Pflanze oder als Antwort auf äußere Reize. Deutlich wird dies zum Beispiel bei Abwehrgenen, die beim Angriff eines Pathogens, wie beispielsweise eines Pilzes oder Bakteriums, schnell genug und am richtigen Ort angeschaltet werden müssen.

Die Genexpression und ihre Stärke wird von Transkriptionsfaktoren vermittelt. Dies sind Proteine, die an bestimmte DNA-Sequenzelemente in den regulatorischen Regionen (Promotoren) ihrer Zielgene binden. Transkriptionsfaktoren sind die zentralen Schalter der Entwicklung aller Lebewesen. Aber gerade pflanzliche Genome enthalten ungewöhnlich viele Gene, die für Transkriptionsfaktoren kodieren; Ackerschmalwand beispielsweise etwa 1800. Manche dieser Gene gehören Genfamilien an, die nur bei Pflanzen gefunden werden. Eventuell brauchen Pflanzen diese Gene und die damit möglichen Regulationsmechanismen, um auf Umweltreize zu reagieren und sich anzupassen, wohingegen Tiere vielen negativen Umwelteinflüssen einfach ausweichen können.

■ Weglaufen geht nicht

Einige Familien von Transkriptionsfaktor-Genen haben weit über 100 kaum voneinander unterscheidbare Mitglieder. Eine der Fragen, denen am Lehrstuhl nachgegangen wird, ist daher, welche Prozesse genau durch die einzelnen Faktoren reguliert werden. Ein möglicher Weg zur Identifizierung der Funktion eines Gens ist, es zu zerstören und die Auswirkungen des Verlustes auf Aussehen und Entwicklung der Pflanze zu untersuchen. Anders als in Hefe oder bei Mäusen ist es bei Pflanzen nicht möglich, einzelne Gene gezielt auszuschalten. Darum ist man hier darauf angewiesen, sehr große Zahlen von Pflanzen mit zufälligen Mutationen im Genom



Unterschiedlich gefärbte Samen von drei Arabidopsis-Mutanten im Vergleich zum Wildtyp (dunkelbraun). Auch Raps hat gewöhnlich dunkelbraune Samen. Die braunen Pigmente der Samenschale sind jedoch schlecht verdaulich, so dass der eiweißreiche Rückstand, der bei der Ölgewinnung entsteht, nur eingeschränkt als Tierfutter genutzt werden kann. Mit Hilfe der Mutanten kann erforscht werden, wie in Arabidopsis die Bildung der braunen Farbstoffe gesteuert wird. Die gewonnenen Erkenntnisse können dann zur Züchtung gelbsamiger Rapsorten genutzt werden.



Gewebekultur. Mit dieser Methode kann eine große Anzahl Pflanzen innerhalb kurzer Zeit auf engstem Raum unter kontrollierten und sterilen Bedingungen angezogen werden. An solchem Material kann beispielsweise gezielt die Abwehrreaktion gegen ein bestimmtes Pathogen untersucht werden.

zu erzeugen und dann diejenigen zu finden, in denen das zu untersuchende Gen getroffen wurde. Entsprechende Populationen umfassen bis zu 100 000 Pflanzen, so dass der Einsatz von Computern und automatisierten Verfahren nötig ist, um die Pflanzen mit der gewünschten Mutation zu identifizieren.

Auch bei der Untersuchung der Mutanten werden heute vermehrt automatisierte Verfahren eingesetzt. Mit ihrer Hilfe lässt sich beispielsweise die Gesamtheit aller zu einem bestimmten Zeitpunkt unter bestimmten Bedingungen in einem Gewebe aktiven Gene (Transkriptom), die Gesamtheit der vorhandenen Proteine (Proteom) und der Stoffwechselprodukte (Metabolom) untersuchen und mit dem Wildtyp (also der Pflanze, in der das zu untersuchende Gen intakt ist) vergleichen.

Anschauliche Beispiele für Untersuchungen, bei denen Mutanten eine wichtige Rolle spielen, sind die Bildung von pflanzlichen Haaren oder Farbstoffen wie den Flavonoiden. Bestimmte Pigmente dieser Klasse treten standardmäßig im Verlauf ihrer Entwick-

lung auf. In Blättern dagegen werden Flavonoide auch als Reaktion auf Stress-Situationen wie UV-Bestrahlung, Hitze, Kälte, Nährstoffmangel oder Trockenheit gebildet. Mit Hilfe der beschriebenen experimentellen Ansätze konnte gezeigt werden, dass jede der beiden Reaktionen durch das Zusammenwirken unterschiedlicher Mitglieder der gleichen Transkriptionsfaktorfamilien reguliert wird.

■ Vom Modell in die Praxis

Die Experimente an Ackerschmalwand liefern Erkenntnisse über die Regulation von Genen, die Funktionsweise der zugehörigen Proteine, ihre Interaktionen mit anderen Proteinen, ihre Evolution und vieles mehr. Da es sich bei diesen Arbeiten häufig um Grundlagenforschung handelt, die mit öffentlichen Mitteln finanziert wird, ist ein Großteil der Ergebnisse in internationalen Datenbanken frei zugänglich und kann zur Erkundung anderer Pflanzenarten genutzt

werden. Zum Beispiel können Gene, die es Acker-schmalwand ermöglichen, sich gegen Krankheits-erreger zu verteidigen, in anderen Arten, wie etwa dem verwandten Raps, aber auch der Zuckerrübe oder dem Weizen, gezielt gesucht und daraufhin überprüft werden, ob sie dort die gleiche Funktion haben. So müssen die Baupläne dieser Pflanzen nicht vollständig entschlüsselt werden, um Informationen über Gene zu gewinnen, die diejenigen Merkmale bestimmen, die für Pflanzenzüchter und Landwirtschaft von Bedeutung sind.

■ Karten weisen den Weg

Neben Informationen über einzelne Gene und die Zusammenhänge, in denen sie wirken, stellt die funktionelle Genomanalyse der Pflanzenzüchtung aber auch molekulare Marker zur Verfügung. Diese ermöglichen es, sich in einem Genom zu orientieren. Mit ihrer Hilfe können regelrechte Karten erstellt werden, die sicht- und messbare Eigenschaften von Pflanzen einzelnen Genen oder Chromosomenregionen zuordnen. Mit Hilfe solcher molekularer Marker lässt sich feststellen, ob im Genom einer Pflanze bestimmte Eigenschaften, zum Beispiel Krankheitsresistenzen, vorhanden sind. Züchter, die allein mit klassischen Methoden arbeiten, müssen bei der Entwicklung neuer Sorten nach jedem Kreuzungsversuch warten, bis die Pflanzen ausgewachsen sind, um sie auf die gewünschten Eigenschaften hin zu untersuchen. Bei Verwendung molekularer Marker dagegen lassen sich Kreuzungen sehr viel gezielter planen und die erwünschten Merkmale in der Nachkommenschaft eindeutig und frühzeitig feststellen. Dies ist besonders wichtig, wenn es um sehr komplexe Eigenschaften geht, die durch mehr als ein Gen bestimmt werden. Mit dem Einsatz der markergestützten Züchtung kann die Entwicklung neuer Sorten, die derzeit durchschnittlich 15 bis 20 Jahre dauert, erheblich verkürzt werden. So hilft heute ein „Unkraut“ mit, in Zukunft schneller und gezielter zu Sorten mit verbesserten Eigenschaften (Widerstandsfähigkeit gegenüber biotischen oder abiotischen Faktoren, Inhaltsstoffe, Ertrag etc.) zu gelangen.



Martin Sagasser studierte Biologie an der Technischen Universität Braunschweig. Am Max-Planck-Institut für Züchtungsforschung in Köln promovierte er über die Regulation der Bildung von Pflanzenfarbstoffen. Er arbeitete in einem Wissenschaftsverlag und ist seit April 2003 Mitarbeiter am Lehrstuhl für Genomforschung. Seine Forschung zielt vor allem auf die Aufklärung der Funktion einiger neuartiger regulatorischer Proteine in Pflanzen.



Bernd Weisshaar studierte Biologie an der Universität Köln und promovierte über Genregulation bei DNA-Tumorviren. Nach zwei Jahren als Postdoc wurde er 1991 Arbeitsgruppenleiter am Kölner Max-Planck-Institut für Züchtungsforschung. Neben seiner wissenschaftlichen Tätigkeit auf dem Gebiet der Genregulation baute er dort die DNA-Analytik auf. Im Februar 2003 übernahm er den Lehrstuhl für Genomforschung an der Universität Bielefeld. Seine Forschungsschwerpunkte liegen im Bereich der Genomik pflanzlicher Transkriptionsfaktoren und der Regulation der Flavonoidbiosynthese.



Wie menschlich ist die Maus?

Funktionelle Genomforschung am Modellorganismus der Biomedizin

Thomas Schmitt-John
und Harald Jockusch

Fakultät für Biologie

Die Genome von Maus und Mensch haben eine ähnliche Größe: In ca. drei Milliarden „Buchstaben“ (3 Gigabasen) umfassenden DNA-Molekülen ist die Information für nur etwa 30.000 bekannte Gene und vermutlich noch viele unbekannte Funktionen niedergelegt. Die Entsprechung der Gene bei Maus und Mensch einschließlich ihrer Anordnung auf Chromosomenabschnitten ist erstaunlich hoch. Die Funktionelle Genomforschung an der Maus möchte erklären, wie mit einer relativ bescheidenen Anzahl von Genen komplexeste Leistungen gesteuert werden können. Spontan entstandene oder gentechnisch erzeugte Mutationen bei der Maus können wertvolle Aufschlüsse über menschliche Erbkrankheiten und familiäre Dispositionen für Krankheiten geben. Durch den Vergleich der Buchstabenabfolgen zwischen Mensch und Maus findet man DNA-Abschnitte auch außerhalb von proteincodierenden Bereichen, die trotz der 75 Millionen Jahre getrennter Stammesentwicklung der beiden Säuger in ihrer Sequenz konserviert geblieben sind. Man schreibt diesen erst kürzlich entdeckten Basen-Sequenzen regulatorische Funktionen zu. Auf dem Gebiet der Funktionellen Genomforschung mit biomedizinischem Bezug gibt es an der Universität Bielefeld eine enge Zusammenarbeit von Molekulargenetikern am Lehrstuhl für Entwicklungsbiologie und Molekulare Pathologie mit Bioinformatikern in der Technischen Fakultät.

Eine genetisch getestete Maus, gesunde Trägerin eines Gendefekts für eine Muskeldystrophie, beschnuppert ein „Kohlenstoffatom“ (schwarze Kugel) in der Base Adenin im Molekülmodell einer DNA-Doppelhelix. Der weiße Bindungsarm links ist eine „Wasserstoffbrücke“, die eine wichtige Rolle für die Basenpaarung und damit für die identische Verdopplung des Erbmateri als spielt.



■ Was ist und wozu studiert man Funktionelle Genomik?

„Ist die DNA ein Code?“ – mit dieser Frage begann eine nachdenkliche Betrachtung über „Codes, Metaphern und Genome“ in einer in biowissenschaftlichen Laboratorien verbreiteten Zeitschrift. Erst einmal die Antwort: „Nein!“ – „DNA“, deutsch auch „DNS“, „Desoxyribonukleinsäure“, ist ein Makromolekül und damit eine wägbare chemische Substanz – ein „Code“ ist eine Übersetzungsvorschrift, also etwas Immaterielles, Abstraktes. Die in der Frage angesprochene Gleichheit kann also schon aus Gründen der Logik gar nicht existieren und damit auch nicht Gegenstand einer vernünftigen Frage sein. Die Frage heißt also eigentlich: Welche Eigenschaften lebender Organismen sind in welcher Weise in deren DNA codiert? Wir kennen den klassischen „Genetischen Code“, die Übersetzungsvorschrift der Nukleinsäuresprache (Basenabfolge) in die Proteinsprache (Aminosäureabfolge) seit nunmehr fast 40 Jahren. Nur zehn Jahre hat es von der Entdeckung der Watson-Crick-Doppelhelix, die in diesem Frühjahr ihr 50-jähriges Jubiläum gefeiert hat, bis zum „Knacken“ des Genetischen Codes gedauert!

Man könnte fragen: Was gibt es eigentlich heute noch Umwälzendes an den Genomen von Mensch und Maus zu erforschen, nachdem die ca. 3 Milliarden lange Buchstabenfolge der beiden Säugerarten mittlerweile von öffentlichen Datenbanken (www.ensembl.org) abgerufen werden kann? Wie so oft in den Naturwissenschaften hat auch auf diesem Gebiet fortschreitende Erkenntnis und Ansammlung von Daten neue Komplikationen gebracht. Man weiß jetzt, dass in der DNA-Doppelhelix weit mehr als die Bauvorschrift für Proteine steckt, obwohl bereits die Zahl möglicher Proteine mit ca. 100 000 weit höher ist als die Zahl dafür codierender „klassischer“ Gene. Man wusste schon seit langem, dass von etlichen Genen auch Nukleinsäuren abgelesen werden, zum Beispiel solchen, die als „Helfer“ bei der Eiweißsynthese mitwirken und somit nicht identisch sind mit der zu übersetzenden Boten-RNA. In den letzten Jahren hat man herausgefunden, dass kleine Ribonukleinsäuren – bisher übersehen oder als Abbauprodukte missdeutet – an vielen regulatorischen Prozessen in der Zelle mitwirken (vgl. *Laborjournal* Heft 7-8, 2003). Schließlich können Basenabfolgen in der DNA auch die Struktur der DNA selbst und damit ihr Verhalten, zum Beispiel bei der Zellteilung oder der Ablesbarkeit benachbarter Gene, beeinflussen.

Die Frage nach der funktionellen Bedeutung oder dem „Informationsgehalt“ von Basenabfolgen be-

trifft den Softwareaspekt des Genoms. Sie ist Gegenstand des derzeit wohl wichtigsten Teilgebiets der heutigen Genetik, das als „Funktionelle Genomforschung“ oder „Funktionelle Genomik“ bezeichnet wird. Bei den folgenden Betrachtungen befassen wir uns nur mit den Genen, die auf den Chromosomen im Zellkern lokalisiert sind, nicht mit dem vergleichsweise winzigen Genom der Mitochondrien (Zellorganellen, die der Energiegewinnung dienen), das beim Menschen nur 16 000 Basenpaare umfasst.

Auf drei Wegen versucht man heute, der Buchstabentitanei großer Genome einen Sinn abzugewinnen.

(1) Man wendet Suchprogramme an, die aus Erfahrungen an bekannten Genen funktionelle Einheiten, zum Beispiel Start und Ende von Genen, Befehle, welche Sequenzbereiche in Aminosäureabfolgen von Proteinen übersetzt werden sollen, regulatorische Sequenzen usw. erkennen können. Diese Aufgabe der Bioinformatik ist nicht trivial, da die genetische Schrift keine „Satzzeichen“ und keine „Kapitelüberschriften“ kennt, sondern „ohne Punkt und Pause“ über Hunderte von Millionen von Basenpaaren jedes einzelnen Chromosoms (und damit durchgehenden DNA-Fadens) läuft und überdies von langen Abschnitten offenbar bedeutungsloser, sich wiederholender („repetitiver“) Sequenzen durchsetzt ist.

(2) Man verändert experimentell begrenzte Bereiche im Genom, entweder durch eine Schrotschussmethode (die Veränderungen sind dabei zufällig im Genom verteilt) oder durch gezielte Ausschaltung einzelner Gene. Experimentelle Mutantenerzeugung ist natürlich nur auf Versuchstiere anwendbar. Sie wird derzeit weltweit intensiv an der Labormaus durchgeführt. Eine für die funktionelle Genomik relevante Auskunft erhält man, wenn die Mutation genau im Genom lokalisiert wurde und Veränderungen in physiologischen Leistungen oder im Erscheinungsbild („Phänotyp“) festgestellt wurden.

(3) Man beruft sich auf ein Prinzip der molekularen Evolutionsforschung, dass Bausteinabfolgen, die über Jahrmillionen der Zerrüttung durch sich anhäufende zufällige Kopierfehler standgehalten haben, also „konserviert“ wurden, eine für seinen Träger wesentliche Funktion haben müssen. Um solche Sequenzen zu finden, vergleicht man die Basenabfolgen von Genomen verschiedener Spezies, zum Beispiel einigermaßen nahe verwandter Organismen wie Mensch und Maus, die sich in der Stammesgeschichte der Tiere vor ca. 75 Millionen Jahren getrennt haben. Für die Maus wären das etwa 300 Millionen Generationen,

also reichlich Gelegenheiten, zufällige Veränderungen im Genom anzuhäufen. Ob solche Mutationen, einfache Basenaustausche („Punktmutationen“) bis zu Verlusten, Verschiebungen oder Verdopplungen von Sequenzbereichen, Folgen für den Organismus haben, hängt davon ab, ob ein funktionell wichtiger Bereich des Genoms betroffen ist. Umgekehrt schließt man aus dem weitgehenden Erhalt einer Basenabfolge in der Evolution, dass diese eine wichtige Funktion hat. Individuen mit veränderter Basenabfolge wären im Allgemeinen in dieser Funktion beeinträchtigt und würden im „Kampf ums Dasein“ gegenüber ihren arteigenen Konkurrenten unterliegen. Daher werden sich solche Mutationen, auch wenn sie immer wieder auftreten, nicht durchsetzen können.

Die unter (1) genannten Methoden der Bioinformatik werden in einem weiteren Artikel in dieser Ausgabe des Bielefelder Forschungsmagazins von Jens Stoye und Robert Giegerich beschrieben. In diesem Artikel werden die Ansätze (2) „Genausschaltung“ und (3) „Vergleichende Genomik“ mit Beispielen aus der Bielefelder Forschung illustriert.

■ Wie menschlich ist die Maus?

In Abbildung 1 werden einige biologische Eigenschaften von Mensch und Maus verglichen. Beide sind Säugetiere, aber es gibt doch erhebliche Unterschiede: die Lebenserwartung unterscheidet sich etwa um den Faktor 40, das Gewicht um den Faktor 3000! Da verwundert es nicht, wenn man als Modell für ernährungsphysiologische Fragen, Herz-Kreislauf-Krankheiten oder Operationsmethoden lieber auf das Schwein als Modell für den Menschen zurückgreift. Genetik und vor allem Funktionelle Genomik erfordert aber große Individuenzahlen und eine schnelle Generationenfolge. Allein schon aus Gründen von Platz und Kosten kommen in diesem Bereich nur Kleinsäuger in Betracht. Die Abbildung 1 zeigt, dass die Maus in der Organisation ihres Genoms, also einer prinzipielleren Eigenschaft als Lebenserwartung und Körpergröße, dem Menschen doch sehr ähnlich ist.

■ Gene und Krankheiten

Die klassische Genetik hat ihren Ausgang von der Beobachtung von Mutationen genommen. Diese Veränderungen in der Erbsubstanz Desoxyribonukleinsäure (abgekürzt DNS oder DNA) führen in vielen Fällen zu sichtbaren Unterschieden zwischen

Mensch



Homo sapiens sapiens
Johann Wolfgang von Goethe

Lebenslaufdaten

Pränatalentwicklung: 9 Monate
Erwachsen nach 18 Jahren
Mittleres Körpergewicht
50–75 kg
Generationszeit 25–30 Jahre
Lebenserwartung etwa 80 Jahre

Genetik

Chromosomen 1–22, X und Y
24.261 Gene
3.232.141.503 Basenpaare

Maus



Mus musculus laboratorius
Stamm NMRI

Lebenslaufdaten

Pränatalentwicklung: 21 Tage
Sexuelle Reife nach 4–6 Wochen
Mittleres Körpergewicht
25–30 g
Generationszeit 8 Wochen
Lebenserwartung etwa 2 Jahre

Genetik

Chromosomen 1–19, X und Y
24.948 Gene
2.497.038.443 Basenpaare

Abbildung 1: Mensch und Maus in Steckbriefen
Man sieht, dass sich Mensch und Maus in den genetischen Daten ähnlicher sind, als in den Lebenslaufdaten. Die angegebenen Zahlen zu den Genomen spiegeln den gegenwärtigen Forschungsstand (Juli 2003) wider, so dass sich die Anzahl der Basenpaare und der Gene sicherlich im Verlauf der nächsten Monate verändern wird. Der abgebildete *Homo sapiens* hatte das Alter von 81 Jahren (Gemälde von Johann Joseph Schmeller, 1830).

„Mutante“ und „Wildtyp“ (wie man bei Versuchorganismen sagt). Durch Mutationen, die beim Menschen zu Veränderungen von wichtigen Enzymen und damit zu Erbkrankheiten führen, ist man schon um die vorletzte Jahrhundertwende auf die enge Beziehung zwischen Proteinen und Genen aufmerksam geworden, zu einer Zeit also, als die chemische Natur der Erbsubstanz noch vollkommen unbekannt war. Diese Beziehung, deren Herzstück die Übersetzungsvorschrift zwischen der Basenabfolge der Nukleinsäuren und der Aminosäureabfolge der Proteine, der „Genetische Code“, ist, wurde zur Grundlage der modernen Molekulargenetik.

In Abbildung 2 ist eine „Krankheits-Genkarte“ der Maus gezeigt, wobei als Beispiele vor allem neurodegenerative und Muskelkrankheiten gewählt wurden, da diese in der Medizin wichtigen Erbdefekte seit vielen Jahren in der Arbeitsgruppe „Entwicklungsbiologie und Molekulare Pathologie“ im Rahmen der Sonderforschungsbereiche 223 (1985–1996)

■ Mutationen nach Maß

Schon zur Zeit der klassischen Genetik im vorigen Jahrhundert hat man erkannt, dass durch bestimmte energiereiche Strahlen und reaktionsfreudige Chemikalien Schäden in der DNA und damit Mutationen erzeugt werden können. Die dabei angewendeten Methoden beruhten aber auf dem Schrotschuss-Prinzip: Die Gene wurden rein zufällig vom Mutagen getroffen, und man konnte nie vorhersagen, welche Mutanten entstehen würden. Diese Situation hat sich in den letzten Jahren grundlegend gewandelt. Neben den vorhandenen, ohne Zutun des Menschen entstandenen „Spontanmutanten“ der Maus und solchen, die mit „Schrotschussmethoden“ erzeugt wurden, stehen heute Ergebnisse raffinierter Gen-ausschaltungsmethoden zur Verfügung (Abbildung 3). Auf dem Umweg über embryonale Stammzellen können nicht nur bekannte Gene gezielt ausgeschaltet werden – man erhält dann so genannte knock-out- oder KO-Mäuse – die funktionelle Ausschaltung kann auch auf bestimmte Gewebe beschränkt oder zeitlich reguliert werden. Auch die Kombination beider Methoden „gewebsspezifisch“ und „auf Kommando“ ist möglich. Es können auch normale Gene der Maus durch „kranke Gene“ des Menschen ersetzt werden („Knock-in“). Was ist der Sinn dieser gentechnischen Akrobatik? Die Aktivität vieler Gene wird im Laufe des Lebens eines Organismus mehrfach in verschiedenen Kontexten abgerufen. Ist zum Beispiel eine Genfunktion vor der Implantation in die Gebärmutter, später aber nochmals zur Differenzierung eines Organs und weiterhin zur Regeneration bei Verwundung dieses Organs von Bedeutung, so wird ein normales Gen-KO zum Absterben der frühen Embryonen führen. Es werden also nie Tiere mit einem „Phänotyp“ zu sehen sein, und die späteren Funktionen des Gens bleiben dem Forscher verborgen. Über sie erfährt man erst dann etwas, wenn die Funktion des Gens zu einem späteren Zeitpunkt im Leben der Maus ausgeschaltet wird („Konditionaler KO“). Der Sinn eines Knock-in eines menschlichen Krankheitsgens ist die Erzeugung eines möglichst realitätsnahen Tiermodells für eine Erbkrankheit (Genokopie und Phänokopie). Versuche dieser Art laufen zur Zeit an der Universität Bielefeld, um die Funktion von bestimmten Genen (z.B. Neuregulin-1) im Nervensystem zu untersuchen.

■ Vergleichende Anatomie von Genomen

Dem Wissenschaftszweig „Vergleichende Anatomie“ haftet ein gewisser Museumsgeruch an. Die

vergleichende Zoologie auf der Ebene der Gene ist aber ein derzeit blühendes Forschungsgebiet. Vergleicht man die Anordnung sich entsprechender („homologer“) Gene auf den Chromosomen von Maus und Mensch, so findet man nur bei den Geschlechts-Chromosomen X und Y eine Ähnlichkeit im Genbestand. So liegt zum Beispiel das Gen für das Protein Dystrophin, das mit geschlechtsspezifischen Funktionen nichts zu tun hat und dessen Ausfall zur Degeneration der Muskeln führt, bei Mensch und Maus auf dem X-Chromosom. Das Gen, das die Entwicklung zum männlichen Geschlecht anstößt, liegt bei beiden Arten auf dem Y-Chromosom. Bei den übrigen Chromosomen, den „Autosomen“, scheint auf den ersten Blick keine einfache Beziehung für die chromosomalen Orte bei beiden Arten zu bestehen. Ein zweiter Blick zeigt jedoch, dass es auf den Chromosomen der Maus ganze Blöcke von Genen gibt, die in Gengehalt und -anordnung bei beiden Arten übereinstimmen. Eine solche Situation bei einem Abschnitt auf dem Arm p des Chromosoms 2 des Menschen und Chromosom 11 der Maus ist in Abbildung 4 gezeigt. Man erkennt eine fast perfekte Übereinstimmung in Gengehalt, Genreihenfolge und Leserichtung bei der Synthese der Boten-RNA an der DNA, die sich über 75 Millionen Jahre erhalten hat. Projiziert man solche Genombereiche von der Chromosomenkarte der Maus auf die des Menschen (oder umgekehrt), so muss man dazu eins der Genome in etwa 200 verschieden große Stücke schneiden. Diese Stücke wurden während der vergangenen 75 Millionen Jahre in jeweils einem der beiden Genome gegeneinander verschoben. Hat dieser Befund eine Nutzenanwendung? Ja, er hat: Man kann auf dieser Basis mit Kenntnis der Lokalisation eines Gens bei einer Art die Lokalisation des entsprechenden Gens bei einer anderen, verwandten Art voraussagen. Auf diese Weise wurde zum Beispiel die chromosomale Lage des menschlichen Myotonie-Gens aufgrund der bekannten Lage des Myotonie-Gens der Maus „erraten“ (vgl. *Forschung an der Universität Bielefeld*, Nr. 6, 1992).

Bei dieser Betrachtung haben wir über bekannte, wohl definierte Gene gesprochen, die zumeist durch eine protein-codierende Sequenz definiert sind. Das heißt, dass an diesen Genen eine Boten-RNA mit einer Basensequenz gebildet wird, die ihrerseits die Information zur Verkettung von Aminosäuren zu einem Protein trägt. Der Nutzen des Genomvergleichs geht aber weiter. Man kann mit ihm uralte Botschaften im genetischen Text aufspüren, die nichts mit der Struktur von Proteinen zu tun haben.

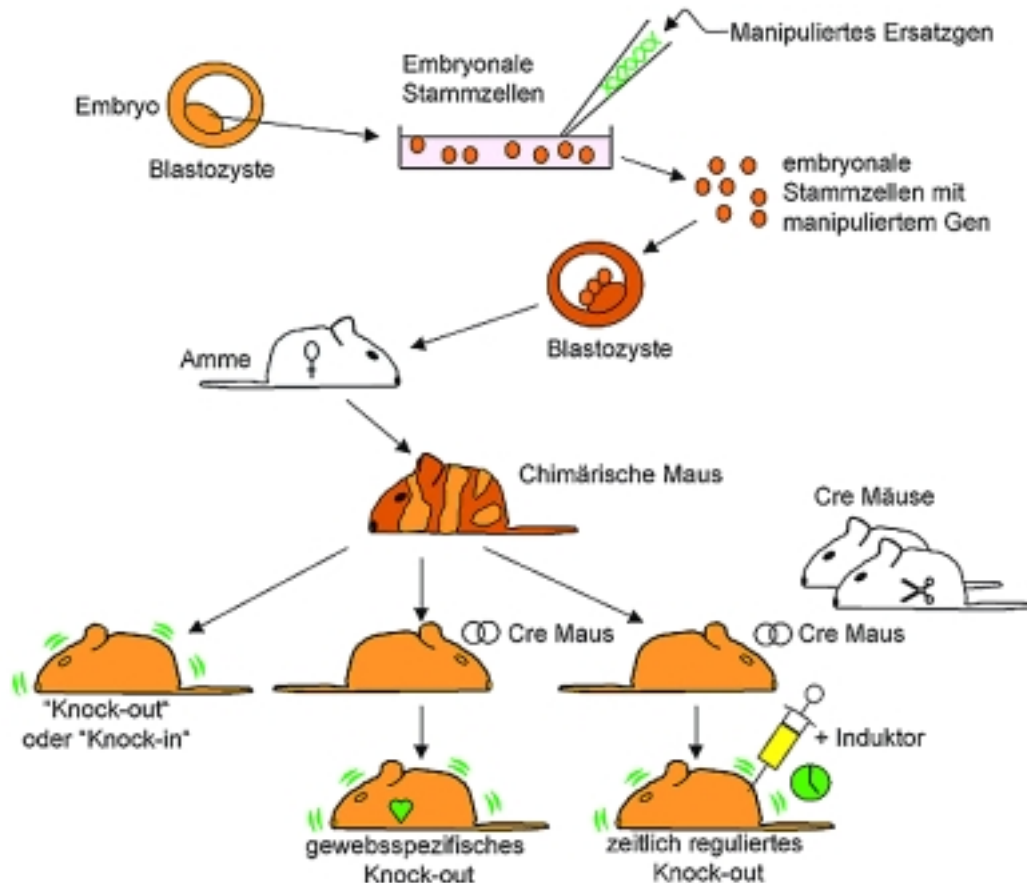


Abbildung 3: Mutationen nach Maß: Manipulation von Genen zur Analyse ihrer Funktion. Die modernen Methoden der Genausschaltung („Knock-out“) oder des Gensatzes („Knock-in“) verlaufen über den Umweg über (ursprünglich aus Blastozysten isolierte) embryonale Stammzellen, die unter Zellkulturbedingungen mit der DNA eines manipulierten „Ersatzgens“ behandelt werden, das mit einer gewissen Wahrscheinlichkeit das authentische Wildtyp-Gen von seinem Platz verdrängt. Die in der gewünschten Weise genetisch veränderten Stammzellen können isoliert, weiter kultiviert und schließlich wieder in eine Blastozyste eingebracht werden, wo sie an der Entwicklung einer chimärischen Maus teilnehmen. Im günstigen Fall beteiligen sich die genetisch veränderten Stammzellen auch an der Bildung der Keimzellen, so dass schließlich reinerbige mutierte „knock-out-Mäuse“ erzeugt werden können, falls der komplette Ausfall des betroffenen

Gens mit der Entwicklung und Lebensfähigkeit der Maus vereinbar ist. Dies ist zwar erstaunlich oft, aber nicht immer der Fall. Daher erhält man in einigen Fällen wertvollere Informationen, wenn der Funktionsausfall des Gens auf bestimmte Gewebe oder einen bestimmten Zeitraum beschränkt wird. Dies erreicht man, indem man das authentische Gen durch ein ebenfalls funktionsfähiges Gen ersetzt, das vorgefertigte Schnittstellen für das „Scherenzym“ Cre-Recombinase aus einem Bakterienvirus enthält. Das Gen für dieses Enzym kreuzt man mithilfe eines entsprechenden transgenen Mausstamms in den Stamm mit dem ausschneidbaren Ersatzgen ein. Das Cre-Recombinase-Gen kann nach Bedarf mit Regulatorsequenzen versehen sein, die die Synthese des Scherenenzym auf bestimmte Gewebe begrenzen, durch von außen zugegebene Substanzen regulierbar machen oder beides.

■ Urschreie im Genom

Beim direkten Vergleich homologer genomischer Sequenzen von Mensch und Maus fällt auf, dass es neben den protein-codierenden Genen eine größere Zahl kürzerer Sequenzen gibt, die zwischen den Arten ähnlich geblieben sind. Solche Homologien werden als „Konservierte nicht-codierende Sequenzen“ (CNS) bezeichnet (Abbildung 4). Man findet zum Beispiel in der in Abbildung 4 gezeigten Region von 1,3 Millionen Basenpaaren mehr als 200 CNS-

Elemente, die mehr als 75 % Identität zu ihrer homologen Sequenz in der anderen Spezies aufweisen und länger sind als 100 Basenpaare. Es ist schwer vorstellbar, dass diese Sequenzen rein zufällig die 75 Millionen Jahre getrennte Evolution von Mensch und Maus mit so wenigen Veränderungen überstanden haben, während benachbarte Sequenzen drastisch verändert wurden. Man muss daher annehmen, dass diese CNS-Elemente einem evolutionären Selektionsdruck unterliegen und somit irgend eine biologische Funktion erfüllen. Die Funktion der CNS könnte in

der Regulation der Synthese oder dem „Zuschneiden“ der Boten-RNA liegen. Sie könnten auch eine Rolle bei der Strukturierung der DNA mit Proteinen zum Chromatin der Chromosomen spielen. In Bielefeld wird im Rahmen unserer Kooperation mit der Gruppe von Robert Giegerich ein Werkzeug entwickelt, mit dem automatisch große genomische Sequenzen verglichen, CNS-Elemente von Genen und repetitiven Sequenzen getrennt und damit identifiziert und schließlich in Datenbanken abgelegt werden. Die weitere funktionelle Analyse wird aber wieder in Zellkultur und durch Mutagenese der Maus durchgeführt.

normale Funktion des betroffenen Gens allenfalls geringfügig beeinträchtigen und im gesunden Menschen keine Krankheiten auslösen. Solche Polymorphismen werden mit statistischen Methoden im Genom aufgespürt. Ihre genomweite Verbreitung spiegelt die Vernetzung von Genwirkungen für physiologische Funktionen und Krankheitsgeschehen wider. Man erhofft sich von der Identifizierung der polymorphen Gene neben diagnostischen und prognostischen Erkenntnissen die Möglichkeit gezielter Entwicklung von spezifischen Pharmaka. Die Kartierung von QTL im Genom wird daher in großen

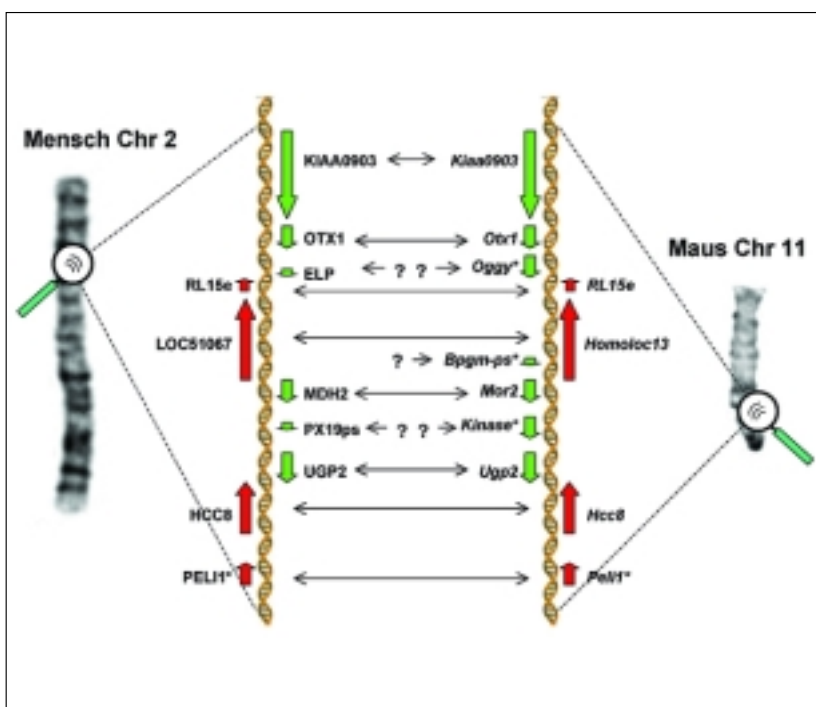


Abbildung 4: Spuren der Stammesgeschichte im Genom: Entsprechung von Gen-Positionen auf Chromosomenabschnitten. Auf Chromosom 11 der Maus (rechts) gibt es einen Abschnitt von ca. 1,3 Millionen Basenpaaren, dem ein etwa ebenso langer Abschnitt auf dem menschlichen Chromosom 2p (links) entspricht. Die genauere Analyse dieses Falls von „konservierter Syntanie“ in Bielefeld hat gezeigt, dass relative Anordnung und Leserichtung trotz 75 Millionen Jahren getrennter Stammesentwicklung im Wesentlichen erhalten geblieben sind. Die Gene sind als breite rote oder grüne Pfeile dargestellt, deren Spitze die „Leserichtung“ bei der Abschrift der DNA-Basenabfolge in die Boten-RNA (messenger RNA, mRNA) darstellt. Die schmalen Doppelpfeile verbinden entsprechende (homologe) Gene in beiden Spezies. Bei Einzelpfeilen mit Fragezeichen ist das entsprechende Gen entweder nicht vorhanden oder auf einem anderen Chromosom lokalisiert. *, Gene, die aufgrund unserer Untersuchungen in Bielefeld entdeckt wurden. Die graphische Darstellung der goldenen Doppelhelix (nicht maßstabsgetreu!) haben wir vom Festkolloquium zum 50. Jahrestag des Watson-Crick-Modells in Cambridge, England, April 2003, übernommen.

■ Netzwerk der Genwirkungen: Risikofaktoren und modifizierende Gene

Bei den meisten „Volkskrankheiten“, wie Herz-Kreislauf-Erkrankungen, Diabetes, Krebs und Alzheimerscher Krankheit, gibt es familiäre Häufungen, die auf die Beteiligung genetischer Faktoren hinweisen. Solche Risikofaktoren beeinflussen die Wahrscheinlichkeit des Eintretens einer Krankheit und/oder die Schwere des Verlaufs. Da sie eine quantitative Wirkung – verstärkend oder abschwächend – haben, werden die entsprechenden Gene im Angelsächsischen als „Quantitative Trait Loci“ (QTL) bezeichnet. Oft handelt es sich bei den „Übeltätern“ um minimale Sequenzunterschiede (Polymorphismen), die die

epidemiologischen Studien angegangen. Allerdings lässt sich der Effekt eines ererbten Polymorphismus beim Menschen nur sehr schwer nachweisen, da sich die exogenen Faktoren wie Ernährung und andere Lebensgewohnheiten, die ebenfalls eine positive oder negative Wirkung auf die untersuchten Krankheiten haben können, den genetischen Effekt überlagern. Bei der Maus lassen sich die exogenen Faktoren standardisieren, unter anderem durch eine moderne Tierhaltung mit perfekten Hygienebedingungen, wie sie die neue Versuchstieranlage der Universität Bielefeld bieten wird. Dank der Verfügbarkeit von genetisch homogenen Inzuchtstämmen, mit denen die meisten der großen Volkskrankheiten untersucht werden können, und der Möglichkeit, einheitliche

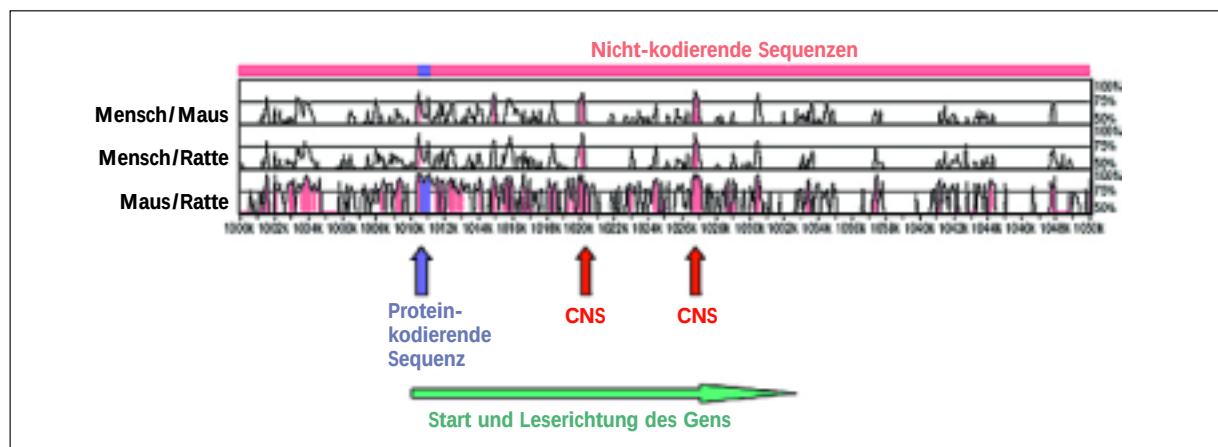


Abbildung 5: Die kleine RNA, das unbekannte Wesen
Die Übersetzung der Nucleinsäureschrift in die Proteinschrift folgt bestimmten Regeln, aufgrund derer man Protein-codierende Abschnitte auf der DNA mit den Methoden der Bioinformatik mehr oder weniger verlässlich erkennen kann. Solche Regeln gibt es offenbar nicht, wenn an der DNA ein wichtiges RNA-Stück überschrieben wird, das nicht in ein Protein übersetzt wird. Daher verwendet man in diesem Fall die vergleichende Genomik als Hilfsmittel für die Funktionelle Genomik. Ein DNA-Abschnitt, der nicht für Proteine codiert, aber dennoch zwischen evolutionär entfernten Arten wie Maus und Mensch oder Ratte und Mensch konserviert geblieben ist, wird

„conserved noncoding sequence“ oder „CNS“ genannt (die Molekulargenetiker hat es nicht gestört, dass diese Abkürzung bereits für ihr eigenes Gehirn vergeben war). In der Abbildung sind zwei CNS in einem kleinen Ausschnitt (50 000 Basenpaare) des Neuregulin-Gens zu sehen. Sie liegen zwischen Protein-codierenden Abschnitten, von denen einer, am linken Ende des Gens, gezeigt ist. Sind die verglichenen Arten evolutionär zu nahe verwandt wie Ratte und Maus (stammesgeschichtliche Trennung vor 10–20 Millionen Jahren), so „sieht man vor lauter Bäumen den Wald nicht“, das heißt, es gibt zu viele Sequenzähnlichkeiten, um die Besonderheit der CNS zu erkennen.

Stämme mit den gewünschten Genkombinationen gezielt züchten und analysieren zu können, ist es bei der Maus möglich, QTLs sehr viel effizienter aufzuspüren als beim Menschen.

In Bielefeld wird seit vielen Jahren die „wobbler-Maus“ (Gensymbol „wr“, s. Abbildung 2) als Modell für Spinale Muskelatrophie (SMA) untersucht. Spinale Muskelatrophien sind unheilbare Krankheiten, bei denen eine Degeneration von motorischen Nervenzellen in Hirnstamm und Rückenmark zum fortschreitenden Abbau („Atrophie“) der Skelettmuskulatur führt. Schon bei den Kreuzungen unseres Mausstamms mit der wobbler-Mutation mit anderen Mausstämmen (wie sie zur Kartierung des Gens auf Chromosom 11 durchgeführt werden mussten) war aufgefallen, dass es zu starken Unterschieden im Schweregrad der Muskelatrophie kam. Dieser Beobachtung, die auf die Existenz von modifizierenden Genen oder QTLs schließen ließ, sind wir in den letzten Jahren nachgegangen: Durch gezielte Verpaarung wurden drei QTLs kartiert. Einer dieser Loci konnte innerhalb des Chromosoms 14 der Maus

eingengt werden. In diesem Genombereich waren neben anderen Genen ein Gen für ein nerven-spezifisches Zytoskelettprotein (Neurofilament-Untereinheit) lokalisiert, und in diesem Gen konnte tatsächlich in dem zur Verpaarung benutzten „Partner-Mausstamm“ eine ungewöhnliche Aminosäure-abfolge entdeckt werden, die möglicherweise für eine Verstärkung des neurologischen Phänotyps der wobbler-Maus verantwortlich ist. Man findet in den Nervenzellen der „modifizierten wobbler-Mäuse“ auffällige Verklumpungen des Neurofilament-Proteins, wie sie auch bei einigen neurodegenerativen Erkrankungen des Menschen beobachtet werden.

■ Schlussfolgerungen und Ausblick

Aus dem Vorhergehenden sollte klar geworden sein, dass eine intensive Forschung an den Genomen von Maus und Mensch auch oder gerade nach der Veröffentlichung der Sequenzen aus Sicht der Grundlagenforschung und Biomedizin sinnvoll und notwendig ist. Detaillierte Analysen bestimmter Bereiche der Genome von Maus und Mensch führen auch

heute noch zur Entdeckung neuer Gene oder regulatorischer Sequenzen, und hier können auch kleinere Gruppen an Universitäten, nicht nur „Sequenzierfabriken“ mit zahlreichen Wissenschaftlern und Technikern, wichtige Beiträge leisten. Dies gilt umso mehr für die Funktionsanalyse von Genen, vor allem durch Genmanipulationen an der Maus, da hierzu die Kombi-

nation molekulargenetischer mit entwicklungsbiologischer und physiologischer Expertise gefragt ist. Ein noch völlig offenes Gebiet ist das Zusammenspiel der Gene bei komplexen Leistungen des Gehirns, dem man sich derzeit mit einer Kombination von Molekulargenetik (Herstellung gewebspezifischer und zeitlich regulierbarer Genausschaltungen), Neuroanatomie und Verhaltensforschung zu nähern versucht.



Privatdozent Dr. Thomas Schmitt-John (links) studierte Biologie in Frankfurt/Main und promovierte am Institut für Organische Chemie der Universität Frankfurt. Nach einer kurzen Phase in einem Auftragsforschungs-Unternehmen wechselte er für drei Jahre als Postdoc an das GSF-Institut für Säugetiergenetik (München), das damals von Prof. Dr. Rudi Balling geleitet wurde. Seit 1995 arbeitet er in Bielefeld am Lehrstuhl Entwicklungsbiologie und Molekulare Pathologie, unter anderem als Teilprojektleiter im Sonderforschungsbereich 549 (seit 1998). Im Jahr 1999 gründete er mit zwei weiteren Partnern das Biotech-Unternehmen Praenadia GmbH in Münster, dessen Geschick er seither mitsteuert. Im Jahr 2001 habilitierte er sich für das Fach „Genetik“.

Professor Dr. Harald Jockusch (rechts) studierte Biologie und Chemie in Frankfurt/Main, Tübingen und München und promovierte 1966 mit einer am Max-Planck-Institut für Biologie in Tübingen durchgeführten Untersuchung auf dem Gebiet der Molekularen Pflanzenvirologie. 1968–1970 Postdoc in Madison, Wisconsin, USA. 1971 Habilitation für das Fach Biologie in Tübingen. 1975–1977 Dozent am Biozentrum der Universität Basel. 1977–1981 Professor für Neurobiologie in Heidelberg. 1981

bis heute Professor für Entwicklungsbiologie und Molekulare Pathologie an der Universität Bielefeld, dort 1988-1991 Prorektor für Forschung und wissenschaftlichen Nachwuchs. Mitbegründer und erster Koordinator des Schwerpunkts „Muskeelforschung“ der Deutschen Forschungsgemeinschaft. Mitglied des wissenschaftlichen Beirats der Deutschen Gesellschaft für Muskelkranke und des Nutzerbeirats der Meeresbiologischen Anstalten des Alfred-Wegener-Instituts Bremerhaven. Mitbegründer und 1993 bis 1996 Sprecher des Sonderforschungsbereichs 223 „Pathomechanismen zellulärer Wechselwirkungen“, Mitbegründer und 1998 bis heute Sprecher des SFB 549 „Prozessierung und Signalwirkung extrazellulärer Makromoleküle“. Projektleiter im Graduiertenkolleg „Strukturbildungsprozesse“ der Fakultät für Mathematik.

Das Foto zeigt die Autoren vor dem neuen zentralen Tierhaus der Universität Bielefeld. Das „Maushaus“ ist ein fast fensterloses Gebäude, das an den Flügel W (Biologie) des Universitätshauptgebäudes angebaut und in diesem Herbst seiner Bestimmung übergeben wird. Modernste Lüftungs- und Hygienetechnik sorgt für höchste Haltungsstandards für die Zucht von Mäusen für die biomedizinische Forschung in Bielefeld.

Westfälisch-Lippische Universitätsgesellschaft

– Verein der Freunde und Förderer e.V. –

Werden Sie Mitglied!

Eine Universität braucht Freunde. Sie braucht Kontakt zur Öffentlichkeit. Sie braucht das Engagement aller Bürger und aller privaten oder öffentlichen Körperschaften und Firmen der Region. Sie braucht auch die Verbundenheit ihrer Studenten über die Studienzeit hinaus. Sie braucht die Unterstützung aller Kreise zur Erfüllung ihrer Aufgaben. Dies gilt erst recht für eine junge und expandierende Universität. Wir laden Sie zum Beitritt ein!

Was wir tun:

Förderung von Forschung und Lehre der Universität: Unterstützung von Fakultäten, Einrichtungen und Forschungsvorhaben.

Pflege der Beziehungen zwischen Universität und Bevölkerung: Durchführung und Unterstützung von Vorträgen, Konzerten, Kunstausstellungen, Sportveranstaltungen usw.

Hilfe bei Problemen der Studentenschaft: Förderung von Studentenwohnheimen, Vergabe von Auslandsstipendien, praxisorientierte Betreuung (Gesprächskreise, Firmenbesichtigungen) und andere Aktivitäten.

Förderung wissenschaftlicher Arbeiten: Jährliche Verleihung von Preisen für hervorragende Habilitations- und Dissertationsarbeiten. Unterstützung von förderungswürdigen wissenschaftlichen Arbeiten im Allgemeinen.

Vertiefung der Beziehungen zwischen Universitätsgesellschaft und Lehrkörper: Veranstaltungsreihe „Fakultäten stellen sich vor“.

Wir brauchen den gut ausgebildeten Nachwuchs. Deshalb tun wir etwas dafür.

Die Westfälisch-Lippische Universitätsgesellschaft – Verein der Freunde und Förderer e.V. – wurde 1966 durch Persönlichkeiten aus Politik, Kultur, Wirtschaft und Wissenschaft gegründet. Zu ihren Mitgliedern gehören heute Bürger aus allen Schichten der Bevölkerung, Personen aus Handel und Industrie sowie öffentliche Körperschaften und Firmen.

Vorsitzender des Kuratoriums: Dr. Werner Efinger, Bielefeld, **stellvertretender Vorsitzender:** Prof. Dr. Dieter Timmermann, Bielefeld, **weitere Mitglieder:** Detlef Adler, Bielefeld, Dr. Dietmar Baumeister, Bielefeld, Otto Clüsener, Bielefeld, Peter Ebertz, Bielefeld, Margrit Harting-Kohlhase, Espelkamp, Dirk Ulrich Hindrichs, Bielefeld, Wolfgang Kaeller, Bielefeld, Wilhelm Krömer, Minden, Helmut Kruse, Detmold, Karen Leffers, Werther-Isingdorf, Friedrich Lohmann, Bielefeld, Rudolf Miele, Gütersloh, Dr. Peter von Möller, Bielefeld, Günther Rempel, Bielefeld, Dr. Christian Schäferbarthold, Minden, Joachim Schultz-Tornau, Bielefeld, Herbert Sommer, Bielefeld, Dr. Peter G. Ulrich, Bielefeld, Reinhold Trinius MdL, Porta Westfalica, Dr. Dr. Jürgen Weitkamp, Lübbecke, Vertreter der Universität: Hans-Jürgen Simm, Prof. Dr. Michael Röckner, Dr. Lothar van Laak, Stefan Bielefeld

Vorsitzender des Vorstandes: Ortwin Goldbeck, Bielefeld, **stellvertretender Vorsitzender:** Oberbürgermeister Eberhard David, Bielefeld, **Geschäftsführer und Schatzmeister:** Prof. Dr. h.c. Helmut Steiner, Bielefeld, **weitere Mitglieder:** Bürgermeister Thomas Gabriel, Herford, Dr. Siegfried Luther, Gütersloh, Wolf-Dieter Meier Scheuven, Bielefeld, Walter Maaß, Bielefeld, Hans-Georg Vogt, Bielefeld, Dr. Rainer Wend, Bielefeld, **Ehrenmitglieder:** Rudolf August Oetker, Ernst Graumann †, Prof. Dr. Karl Peter Grottemeyer, Dr. Kurt Schober †, Gerd Seidensticker, Walter Stich.

Die Mitglieder werden durch die Herausgabe von Mitteilungen und Mitgliederversammlungen auf dem Laufenden gehalten. Außerdem erhalten sie kostenlos die Bielefelder Universitätszeitung.

Alle Mitglieder der Universitätsgesellschaft erhalten auch das Forschungsmagazin der Universität Bielefeld kostenlos!

Die Universitätsgesellschaft ist steuerlich als gemeinnützige Institution anerkannt.

Mitgliedschaft

Wenn Sie Mitglied werden wollen, wenden Sie sich bitte an:
Westfälisch-Lippische Universitätsgesellschaft,
Prof. Dr. h.c. Helmut Steiner, Wilhelmstr. 3a, 33602 Bielefeld,
Tel. (05 21) 12 43 47, Fax. (05 21) 5 21 33 33
email Anschrift: HelmutSteiner@t-online.de

Ich/Wir wünsche(n)

☐ eine Firmenmitgliedschaft
zu einem Jahresbeitrag von 260,- EUR
Die zuständige Kontaktperson in der Firma ist

☐ eine Einzelmitgliedschaft
zu einem Jahresbeitrag von 30,- EUR

Bitte senden Sie die Beitrittsbestätigung an:

Name

Firma

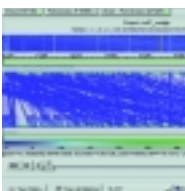
Straße

PLZ/Ort

Datum

Unterschrift

Die Überweisung erfolgt auf das Konto-Nr. 0 669 499 (BLZ 480 700 20) der Deutschen Bank AG in Bielefeld.



Finden, fast ohne zu suchen: Indexstrukturen in der Bioinformatik

Robert Giegerich
und Jens Stoye

Technische Fakultät

Die Bioinformatik wird häufig als eine Hilfswissenschaft zur Unterstützung der (Molekular-)Biologie und Genomforschung angesehen. Dennoch haben an verschiedenen Stellen Fragestellungen aus der Biologie auch Teilgebiete der Informatik so stark beeinflusst, dass sich daraus neue Forschungsrichtungen innerhalb der Informatik ergeben haben. Eine solche ist die strukturelle Mustererkennung in großen Textbeständen, bei der es darauf ankommt, bestimmte, durch ihre Struktur gekennzeichnete Textmuster zu finden, ohne die Texte immer wieder aufs Neue komplett durchsuchen zu müssen.

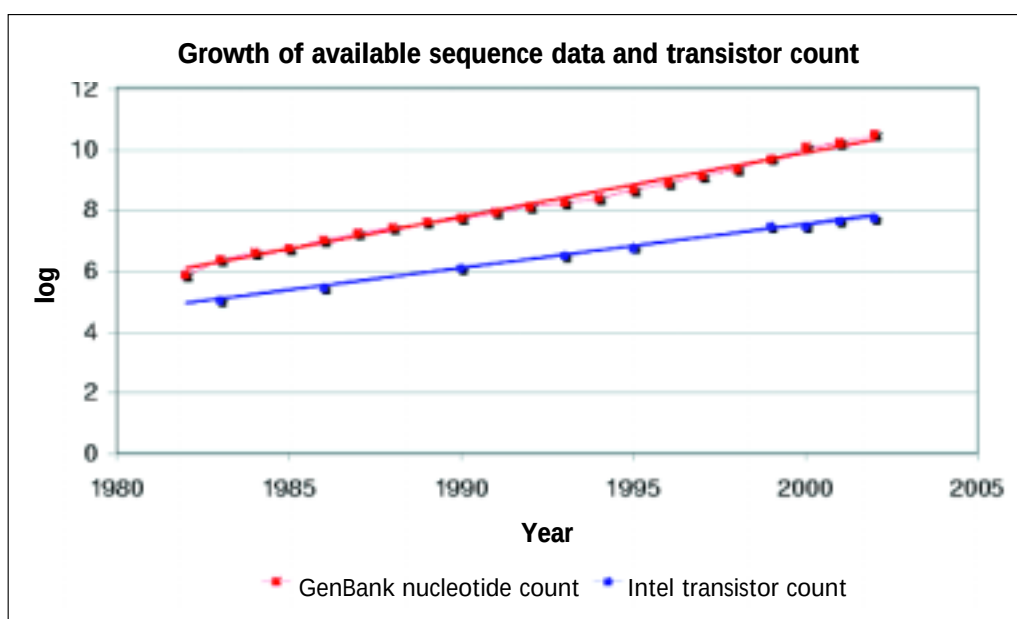


Abbildung 1
Der Anstieg der Integrationsdichte moderner Computerchips wird von dem Anstieg der verfügbaren Genomsequenzen noch deutlich übertroffen.

In den vergangenen Jahren ist die Menge in öffentlichen Datenbanken gespeicherter biologischer Sequenzdaten, meist in Form von DNS- oder Proteinsequenzen, extrem angestiegen, nicht zuletzt forciert durch die immensen Anstrengungen im Rahmen des Humangenomprojekts. Ein Dilemma, das mit diesem Datenwachstum einhergeht, wird sichtbar, wenn man den Anstieg der Menge der biologischen Sequenzdaten mit dem allseits bekannten ebenfalls enormen Anstieg der Integrationsdichte und Geschwindigkeit der jeweils aktuellen Computerchips vergleicht: Während die Chip-Integrationsdichte sich etwa alle 18 Monate verdoppelt (was nach dem Mitgründer der Firma Intel gerne als

„Moore's law“ bezeichnet wird), verdoppelt sich die Menge der in den biologischen Sequenzdatenbanken gespeicherten DNA- und Proteinsequenzen seit ca. 1980 sogar schon etwa alle 14 Monate. Hieraus ergibt sich, dass eine Gesamtanalyse aller verfügbaren Genomdaten, selbst bei Verwendung der jeweils aktuellsten Computergeneration, mit zunehmender Zeit immer länger dauern wird (siehe Abbildung 1). Das exponentielle Wachstum der Größe der Lücke zwischen den beiden Kurven macht klar, dass die einzige Lösung dieses Dilemmas die Entwicklung neuer Analysemethoden ist, deren Geschwindigkeit im Idealfall nur noch sehr wenig von der Größe der zu untersuchenden Daten abhängt.

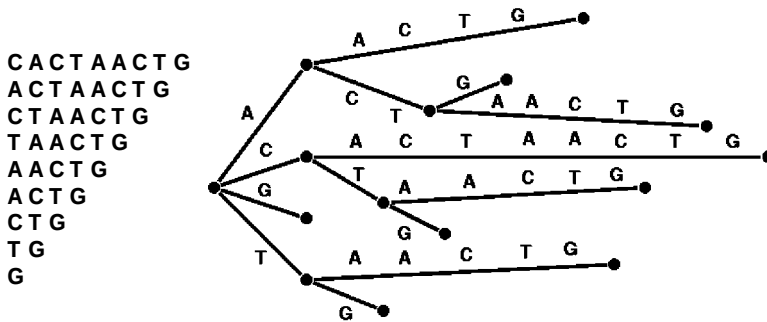


Abbildung 2
Suffixbaum des Textes CACTAACTG. Links sind alle Suffixe des Textes aufgelistet. Im rechts dargestellten Baum kann man diese wiederfinden, wenn man bei der ganz links liegenden „Wurzel“ des Baumes beginnend die Zeichen entlang den Kanten des Baumes liest.

■ Aber kann es so etwas geben?

In der Tat wird in der Informatik schon seit längerer Zeit an Methoden gearbeitet, die es ermöglichen, die Dauer einer Datenanalyse unabhängig von der Größe der zu analysierenden Datenmengen zu machen. In der Regel wird dabei nach dem folgenden Schema vorgegangen: Sobald die Daten bekannt sind, werden einmalig Zeit und Speicherplatz investiert, um die Daten so vorzuverarbeiten, dass sie hinterher schneller analysiert werden können, als dies ohne eine Vorverarbeitung möglich wäre. Man spricht davon, dass ein *Index* der Daten erstellt wird.

Im Folgenden wollen wir einen Index vorstellen, der sich in der Bioinformatik als wahres „Orakel“ herausgestellt hat: den Suffixbaum. An der Technischen Fakultät der Universität Bielefeld beschäftigen sich zwei Arbeitsgruppen (AG Praktische Informatik und AG Genominformatik) mit dieser Datenstruktur.

■ Der Suffixbaum und die Repeat-Suche

Für sich gesehen ist ein Suffixbaum eine ganz einfache Datenstruktur (siehe Abbildung 2), aber die dahinter verborgenen Potenziale werden schon bei einer so einfachen Fragestellung wie der Suche nach exakten Teilworten einer langen Sequenz sichtbar: Wenn man zum Beispiel testen will, ob die Mustersequenz *TAATA* Teilwort des Textes *CACTAACTG* ist, so muss man nur von links aus in den Baum in Abbildung 2 „hineinlaufen“ und dabei den Zeichen der Abfragesequenz folgen. Man wird feststellen, dass *T*, *A* und *A* passen, dass dann aber keine Fortsetzung *TA* mehr möglich ist, *TAATA* also nicht in *CACTAACTG* vorkommt. Das beachtliche an diesem Vorgehen ist, dass die Zeit zur Beantwortung der Anfrage nur von der Länge der Mustersequenz abhängt und nicht von der Länge des Textes.

Bevor wir auf eine weitere wichtige Anwendung des Suffixbaums eingehen wollen, die Repeat-Suche, soll kurz erwähnt werden, dass es neben dem Suffixbaum noch andere Indexstrukturen gibt, die sich auch zur Textanalyse einsetzen lassen. Die meisten solcher Strukturen sind allerdings mehr oder weniger mit dem Suffixbaum verwandt und unterscheiden sich hauptsächlich im Speicherbedarf oder leicht in der Funktionalität. Durch seine einfache Struktur ist der Suffixbaum aber am besten geeignet, die Ideen der Indizierung von Texten darzustellen, deshalb wollen wir uns hier darauf beschränken.

Besonders gut eignen sich Suffixbäume zum Entdecken wiederholt auftretender Sequenzabschnitte, Repeats genannt. Bei der Betrachtung von Abbildung 2 lässt sich erkennen, dass beispielsweise das längste Repeat in *CACTAACTG*, *ACT*, der am weitesten rechts liegenden Verzweigung im Suffixbaum entspricht. Zum Ermitteln aller Repeats in einem Text haben wir das Werkzeug REPuter entwickelt, das zunächst den Suffixbaum erstellt und daraufhin die – dann sehr einfache – Ermittlung der Repeats durchführt. Dabei ist das Werkzeug nicht auf die in dem obigen Beispiel dargestellten exakten Repeats beschränkt, sondern es werden auch Repeats mit einigen Fehlern zugelassen.

Die Fragestellung der Repeat-Suche resultiert aus vielen Anwendungen in der Genomforschung, von denen wir nun fünf beschreiben. Wie man sehen wird, spielt auch die Visualisierung von Ergebnissen eine wichtige Rolle.

■ Prüfung von assemblierten Sequenzen

Zur Sequenzierung werden DNA-Moleküle in kleine Abschnitte zerstückelt. Anschließend müssen sie als Sequenzen wieder zusammengefügt (assembliert) werden. Damit dies möglich ist, sequenziert man in mehrfacher Überdeckung. Dabei kommt es immer

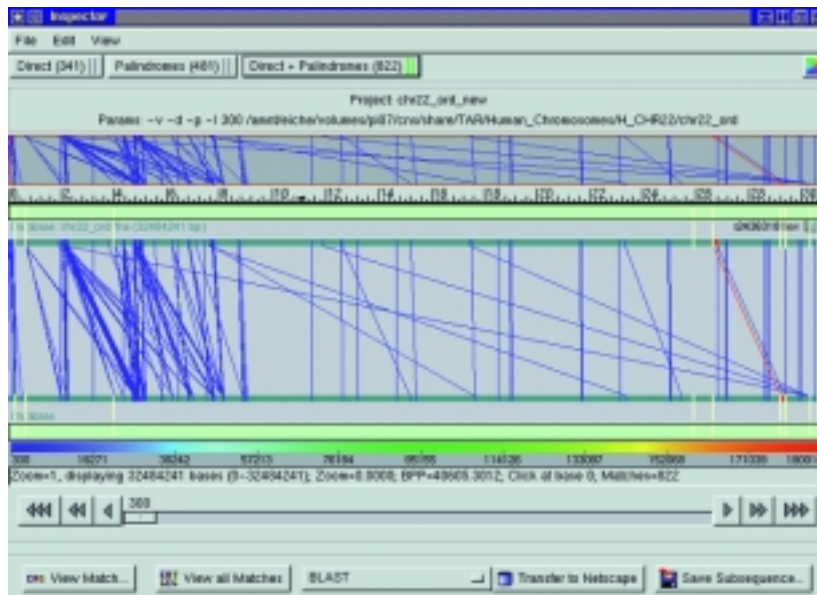


Abbildung 3
Prüfung von assemblierten Sequenzen. Wir sehen Repeats ab einer Länge 200 im menschlichen Chromosom 22, dargestellt in 11 Fragmenten (Contigs). Die Repeats werden als Linien eingetragen; ihre Länge ist durch einen Farb-Code von schwarz bis lila dargestellt. Hier hat man das ungewöhnliche Bild, dass alle Repeats kurz erscheinen (und daher schwarz) gegenüber dem längsten (lila), das sage und schreibe 190014 Nukleotide umfasst. Ein Repeat dieser außerordentlichen Länge kann eigentlich nur ein Artefakt sein; tatsächlich wurde es in einer späteren Version des Genoms entfernt.

wieder vor, dass lange Sequenzabschnitte mehrfach an unterschiedlichen Teilen aufgebaut und am Ende auch mehrfach (statt einfach) eingesetzt werden. Überraschenderweise übersehen die Assemblier-Programme gerade sehr lange Duplikate. Dies liegt daran, dass der Vergleich langer Sequenzen rechnerisch sehr aufwendig ist – es sei denn, man verwendet unsere Indexstrukturen. Abbildung 3 zeigt eine solche Situation.

■ Krankheiten beruhen auf genomischen Deletionen

Treten genomische Abschnitte mehrfach auf, kann es bei der Replikation zu Löschungen ganzer Abschnitte (Deletionen) führen. Liegen Gene auf dem gelöschten Abschnitt, kann der so begründete genetische

Defekt zu einem Krankheitsbild führen. Je nach Lage und Anzahl der Repeats kann es zu Deletionen unterschiedlicher Länge und unterschiedlich schwerem Krankheitsbild kommen. Eine solche Situation „sieht“ man in Abbildung 4.

■ Eindeutige Sequenzabschnitte

Oft sucht man nach charakteristischen Sequenzabschnitten – solchen, die im gesamten Genom nur einmal auftreten und sogar dann noch eindeutig sind, wenn man eine geringe Zahl von Abweichungen erlaubt. Will man zum Beispiel DNA-Microarrays herstellen, ist die Eindeutigkeit der Proben entscheidend für die Interpretierbarkeit der Ergebnisse. Ein Durchprobieren einzelner Abschnitte ist rechnerisch nicht praktikabel.

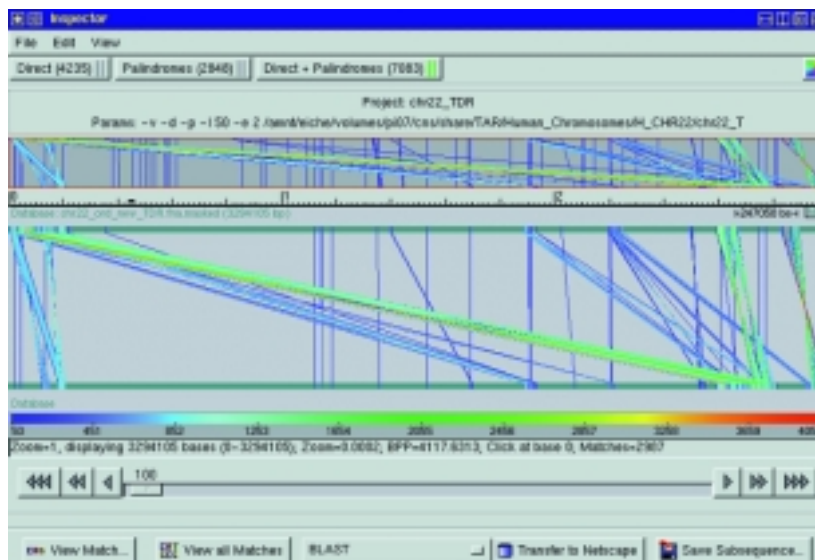


Abbildung 4
Duplikation von Genomabschnitten. Die Abbildung zeigt einen Ausschnitt des menschlichen Chromosoms 22 mit einer ungewöhnlichen Repeat-Region. Ein typisches netzartiges Muster entsteht durch mehrfach (teilweise invertiert) auftretende Abschnitte. Die Deletionen, die zwischen ihnen entstehen können, entsprechen unterschiedlicher Ausprägung des unter dem Namen DiGeorge/Velo-cardio-facial-syndrome bekannten Krankheitsbildes.

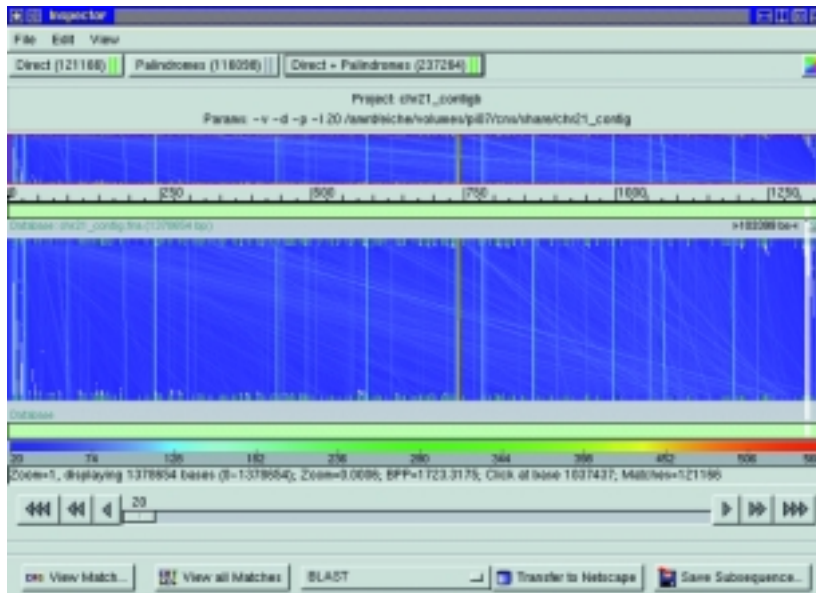


Abbildung 5
Ein Abschnitt des menschlichen Chromosoms 21. Auf den ersten Blick scheint alles voller Repeats.

Wir gehen einen scheinbaren Umweg, der aber viel effizienter ist. Mit REPuter bestimmen wir alle Repeats ab einer Mindestlänge k , mit Fehlertoleranz f . Da k relativ klein ist, treten solche Repeats in Massen auf, und fast das ganze Bild wird schwarz – vgl. Abbildung 5. Aber nur fast! In der Nahaufnahme von Abbildung 6 werden Repeat-freie Regionen sichtbar. Die hier liegenden Abschnitte der Länge k haben die gewünschte Eindeutigkeit, selbst bei Duldung von f Fehlern.

■ Vergleich von Genstrukturen

Programme zur Vorhersage „eukaryotischer“ Gene (von Pflanze, Tier oder Mensch im Gegensatz zu bakteriellen Genen) finden nur selten die perfekte „Intron-Exon-Struktur“. Diese bezieht sich auf die Tatsache, dass für ein Protein kodierende Gene in höheren Lebewesen häufig nicht in einem Stück auf dem DNA-Strang abgelegt sind, sondern in Abschnitten, den so genannten „Exons“, die von nicht für das Protein kodierenden DNA-Abschnitten, den „Introns“ unterbrochen werden. Auf die Intron-Exon-Struktur bezogen hilft ein Vergleich der unabhängig vorhergesagten Genstrukturen verwandter Organismen. In Abbildung 7 wird das pellino-Gen von Mensch und Maus verglichen.

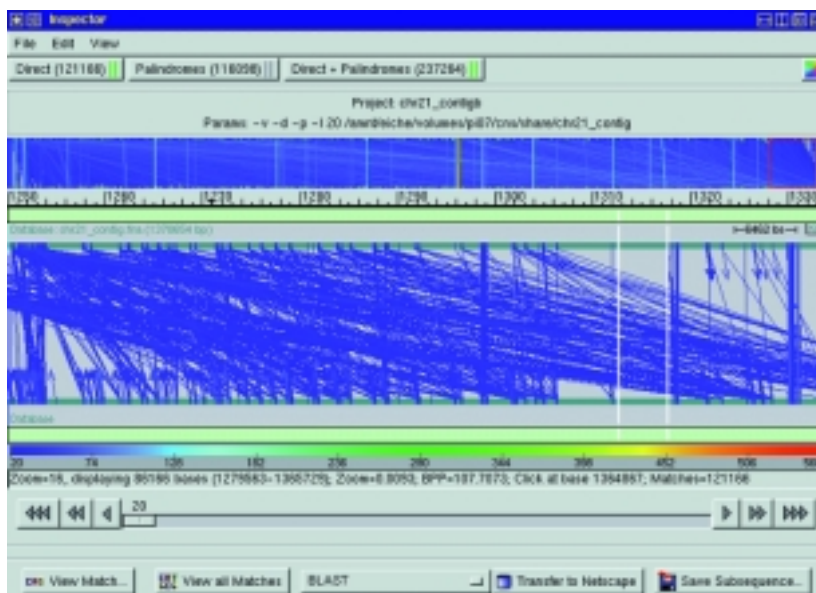


Abbildung 6
Nahaufnahme von Abbildung 5, auf der (zwischen den weißen Linien) ein Bereich sichtbar wird, der keine Repeats enthält.

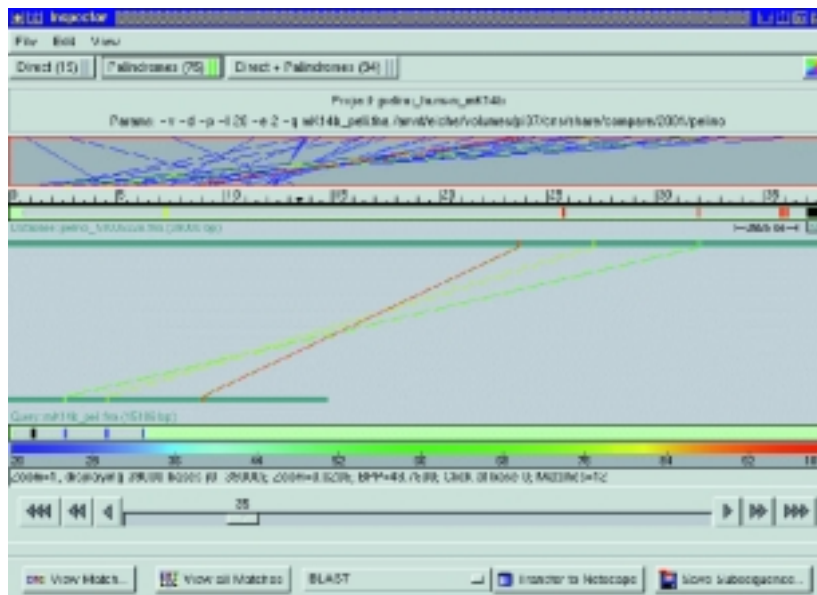


Abbildung 7
Vergleich der Struktur des pellino-Gens von Mensch (links der weißen Linie) und Maus (rechts). Vorhergesagte Genabschnitte sind grün markiert. Die lila Linie zeigt einen weiteren stark zwischen Mensch und Maus konservierten Bereich – möglicherweise ein Exon, das übersehen wurde.

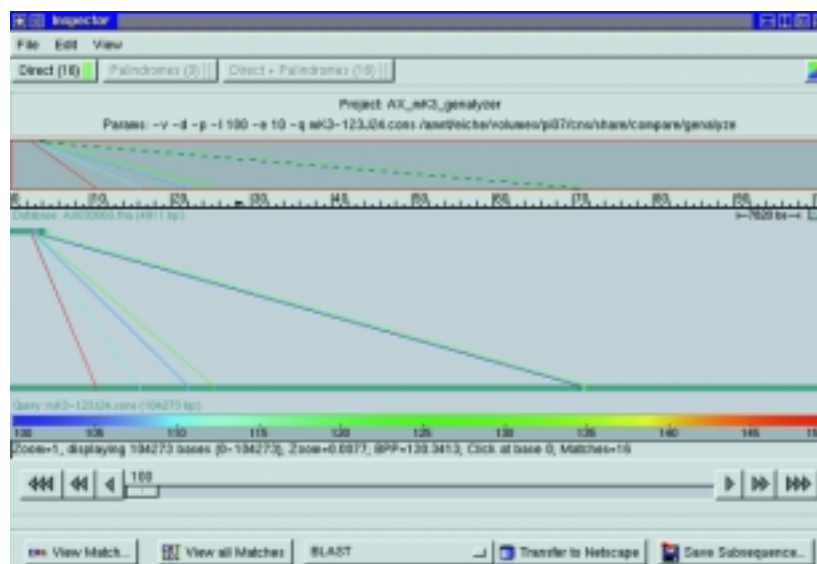


Abbildung 8
cDNA/EST mapping. Hier wird die cDNA des menschlichen Gens Kiaa0903 mit einem Abschnitt des Maus-Chromosoms 11 verglichen. Man sieht auf einen Blick, dass (und wie) das Kia0903-Gen in der Maus durch lange Introns weit separiert ist.

Die Autoren danken
Dr. Jomuna Choudhuri für
das Generieren der Grafiken.

■ Abgleich genomischer Sequenzen mit cDNA und EST

ESTs sind Sequenzen exprimierter und gespleißter DNA (das heißt, die oben angesprochenen nicht kodierenden Introns sind hier herausgeschnitten). Vergleicht man sie mit genomischer DNA, erhält man Hinweise auf die Lage möglicher Gene. EST-Daten werden in großen Mengen bestimmt, so dass auch hier der Vergleich eine Frage der Effizienz ist. Abbildung 8 zeigt eine Anwendung von REPuter auf dieses Problem.

■ Schlussbemerkung

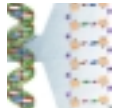
Wir haben fünf Beispiele für Anwendungen der automatischen Repeat-Analyse beschrieben. Während sicherlich einige dieser Fragestellungen auch mit ad-hoc-Methoden, die den Text ein- oder mehrfach durchsuchen, beantwortet werden könnten, ist es doch sehr schnell ersichtlich, dass REPuter mit Hilfe des Suffixbaums in der Lage ist, die verschiedensten Fragestellungen schnell und exakt zu beantworten.



Prof. Dr. Robert Giegerich studierte Mathematik und Informatik an der TU München und an der Stanford University, USA. Er promovierte 1981 an der TU München mit einer Arbeit auf dem Gebiet der Übersetzer-erzeugenden Systeme. Seit 1989 ist er Professor für Praktische Informatik an der Universität Bielefeld, wo er seit 1992 das Forschungs- und Lehrgebiet Bioinformatik aufgebaut hat. Er ist derzeit Sprecher des Graduiertenkollegs Bioinformatik und verantwortlich für diverse Bielefelder Aktivitäten im Rahmen der DFG-Initiative Bioinformatik sowie der „NRW International Graduate School in Bioinformatics and Genome Research“.



Prof. Dr. Jens Stoye studierte Naturwissenschaftliche Informatik an der Universität Bielefeld. Nach der Promotion 1997 erhielt er ein Stipendium des DAAD für einen einjährigen Aufenthalt als Postdoc an der University of California in Davis. Nach einem weiteren Postdoktorat am Deutschen Krebsforschungszentrum in Heidelberg wurde er Gruppenleiter für Algorithmische Bioinformatik am Max-Planck-Institut für Molekulare Genetik in Berlin. Seit März 2002 ist er Professor für Genominformatik an der Technischen Fakultät der Universität Bielefeld.



Alfred Pühler

From Genes to Genomes to Systems Biology

Development of the Bacterial Genome Research at Bielefeld University

The development of sequence analysis at Bielefeld University is presented by a retrospect on three publications. The first of these concerns the sequence of an interesting regulatory gene of *Rhizobium meliloti*, the second the sequence of the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*, and the third the sequence of the complete genome of *Sinorhizobium meliloti*. In the second section of the article, the formation of a Genome Research Institute as part of the Centre for Biotechnology at Bielefeld University is highlighted. In this institute, developing techniques in the field of genomics, transcriptomics, proteomics and bioinformatics are pursued. The institute has already attracted large scale research projects from different sponsor organisations. The most important research project concerns the Germany-wide competence network on Genome Research of Bacteria Relevant for Agriculture, Environment and Biotechnology. This competence network is financially supported by the Federal Ministry of Education and Research is coordinated by the Bielefeld Genome Research Institute. Finally, the article outlines the future perspective. The next step in development will deal with Systems Biology, since the newly developed techniques of genome and postgenome research permit a holistic study of bacterial cells. In particular, bioinformatics represents the dominating discipline which is necessary to handle the huge amount of data produced and to develop in silico models of living cells.



Olaf Kaiser · Folker Meyer

Genomeresearch and Bioinformatics

The Bielefeld Platform of Analyzing Genomes

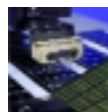
The sequencing and subsequent sequence analysis of *Haemophilus influenzae* in 1995 was the starting point of a new era. Since then, more than one hundred complete genomes have been published. Sequencing and analyzing a complete genome, however, is still a complex, expensive project. At Bielefeld University, a close collaboration of Biologists and Computer Scientists has been established and successfully applied to the analysis of five bacterial genomes. Two successful software platforms have been established in the framework of these genome projects, and more are currently under development.



Robert Giegerich · Jens Stoye

Index Structures in Bioinformatics

Bioinformatics is often described as connecting “wet” biology to „theoretical“ computer science, providing (important) infra-structural support and transfer of knowledge. Some questions originating from computational biology, however, have initiated intense research in new or re-activated areas of computer science itself, including the analysis of large textual data sets and the search for structural patterns in long strings. The problem in bioinformatics is that the amount of available biological data grows much faster than microprocessors accelerate, so that it is mandatory to develop analysis methods that run faster than proportional to the data size. The basic idea to make this possible is that as soon as the data are available, computer time and memory are invested once in order to pre-process the data and build an “index” which later permits to speed up queries. A very powerful and flexible index for questions arising in bioinformatics is the so-called suffix tree. It is particularly well suited for finding various kinds of textual repeats. The demand for repeat analysis results from many applications in genome research, a few of which are presented in this contribution.



Anke Becker

Transcriptomics: Microarray Technology Helps Elucidating Complex Gene Regulation Networks

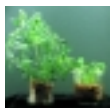
While deciphering of genome sequences is moving into ever higher gear, more than experimental procedures for the investigation of biological functions of individual DNA sequence elements and encoded gene products become important for further analyses. Experimental approaches to elucidate the function of DNA sequences from a more global perspective have become increasingly crucial. The analysis of global gene expression patterns (transcriptomics) is a key area, since how a cell or an organism develops and differentiates and how it adapts to variable conditions is determined to a large part by the profile of gene expression. DNA array technology, where thousands of different DNA sequences are arrayed in a defined matrix on different supports in growing densities, is rapidly becoming the method of choice for gene expression profiling. The applications of microarrays extend beyond the boundaries of basic research into environmental monitoring, pharmacology and medical diagnostics. In recent years, DNA-microarray technology has overcome the initial state and emerged as an important tool to meet many of these challenges. As a partner collaborating in a number of German and European research networks, the Transcriptomics Group at the Institute for Genome Research of Bielefeld University is dedicated to provide technical support, training and advice to researchers in this field of genome research.



Karsten Niehaus

Proteomics: From Gene to Function

The sequencing of whole genome of several model organisms is finished or will be finished in the near future. The obvious vision is now to go "From Genes to Functions". A central focus of the postgenomic research will be to understand how cellular phenomena arise from the connectivity of genes, RNAs, proteins and metabolites. Realising the potential of post-genome science will be a highly interdisciplinary process requiring close collaboration between a wide range of expertise, including experts from physics, chemistry, biology and bioinformatics. Understanding complex biological systems requires the integration of experimental and computational research leading to a systems biology approach. For the years to come, proteomics will be a central part of this approach. Proteins are encoded by genes; but there are neither genes, gene transcription, nor enzymes and metabolites without proteins. This does not mean that proteins are the "most important" players in this game. But from the regulatory point of view, proteins are the receptors, mediators and executives that control virtually all processes in a living cell. Establishing "maps" of the type and quantity of all proteins under given environmental conditions is a first step on the way from genes to functions.



Stefan Weidner

Nitrogen Grasped from the Air

The largest amount of nitrogen in soil is fixed by free-living or by symbiotic micro-organisms. Bacteria of the group Rhizobia are capable of establishing symbiosis with legume plants like peas, beans, soybeans or peanuts, and contribute the major part of fixed nitrogen. Therefore, this symbiosis is of enormous agronomic importance. The entire tripartite genome of the model organism *Sinorhizobium meliloti*, the micro-symbiont of alfalfa, has been sequenced in an international consortium under participation of Bielefeld University. A detailed analysis of the three genomic replicons revealed interesting insights about the evolution of this bacterium and a better understanding of the intimate bacterial-plant interaction. The availability of the entire genomic sequence initiates new research strategies and new research areas like transcriptome and proteome analysis.



Jörn Kalinowski

Genomeresearch Using the Example of the Aminoacid-Producing Bacteria *Corynebacterium glutamicum*

Amino acids are building blocks for the proteins of all living organisms. Some of these, the essential amino acids, have to be taken up with food or feed. Pure amino acids are commonly added to improve the nutritional quality of feed. Amino acids are mostly produced by fermentation processes involving bacteria. The most important of these is the harmless, soil-inhabiting microbe *Corynebacterium glutamicum*. *C. glutamicum* strains developed by classical mutagenesis and selection processes today produce over 100 grams of a pure amino acid per litre of culture medium. The *C. glutamicum* entire genome sequence comprising 3.28 million base pairs was established in a project led by Bielefeld University and the company Degussa AG. It is expected to use the genomic information in addition to genetic engineering techniques to optimize amino acid production processes in the near future.



Karl-Heinz Gartemann · Rudolf Eichenlaub

Bacterial Wilt and Canker of Tomato

Clavibacter michiganensis michiganensis is the causal agent of bacterial wilt and canker of tomato. Two genes essential for virulence were identified which are encoded on plasmids. The genes required for effective colonisation of the plant, a prerequisite for the development of disease symptoms, are located on the bacterial chromosome. Currently, a genome project to determine the complete nucleotide sequence of *Cmm* is conducted. This will permit a survey of the chromosome for genes involved in virulence and colonisation for the first time. The results may open up new possibilities of disease control and may improve diagnostics of *Cmm* infections.



Helge Küster

Genomeresearch on Plants of the Family of Legumes

Legumes have the unique ability to enter the nitrogen-fixing root nodule symbiosis and the mycorrhiza symbiosis that contributes to the acquisition of phosphates and minerals by the plant. In addition to their capacity to enter root endosymbioses, legumes are characterised by leaves and seeds that are unusually rich in proteins. This property renders legumes an essential component of animal and human diet. Since other model plants like *Arabidopsis* or rice cannot enter root endosymbioses and do not form protein-rich seeds, the Mediterranean legume *Medicago truncatula*, a close relative of the key European legume crops, was chosen as a model for genomics studies. The Department of Genetics is involved in different projects dedicated to legume genomics using the model plant *M. truncatula*. These projects are the European Union 5th Framework project "MEDICAGO: Integrated structural, functional and comparative genomics of the model legume *Medicago truncatula*" and the Deutsche Forschungsgemeinschaft Focus programme "MolMyk: Molecular Basics of Mycorrhizal Symbioses" that is coordinated in Bielefeld.



Thomas Schmitt-John · Harald Jockusch

How Human is the Mouse?

Functional Genomics on the Model Organism of Biomedicine.

Man and mouse differ dramatically in their biology and physiology, but their genomes are surprisingly similar. This is true for the total number of (classical) genes, their internal structure and their grouping on chromosome segments. The challenge of functional genomics is to explain how the complexity of a human being is achieved with only about 30,000 genes. As research disciplines, human and mouse genomics support each other: Only in the mouse can genes be manipulated to study their function. Gene knock-outs can now be realised in a tissue specific and time-controlled manner. Functionally important DNA segments that do not code for proteins can be identified as sequences conserved between man and mouse. Biomedical aspects of comparative and functional genomics are being studied at Bielefeld University, in collaboration between the Department of Developmental Biology and Molecular Pathology, and the Research Group in Practical Informatics.



Martin Sagasser · Bernd Weisshaar

Methods and Aims of Genomeresearch on Plants

Plants are the basis of our diet, they provide medicinal drugs as well as raw materials for industry and construction. Despite the importance of plants for our life, many aspects of their life cycle, physiology, development and regulation are still poorly understood. To fully understand plants, we need to read and interpret their genomic information. This allows us to learn more about the molecular basis of traits like yield or stress tolerance, which are determined by far more than a single genetic factor. This knowledge can contribute to more directed and efficient plant breeding approaches.

■ Abbildungen

Abbildungen S. 15, 18 und 23: Karsten Niehaus ■ Abbildung S. 25: D. Jording ■ Abbildung S. 30: Degussa AG Düsseldorf ■ Abbildung S. 33: A. Bermpohl ■ Abbildung S. 34: H. Jahr ■ Abbildung S. 35: H. Jahr, Rudolf Eichenlaub ■ Abbildung S. 36: nach A. Bermpohl ■ Abbildungen S. 39: Dieter Kapp (A), Martin Vieweg (B), Christian Firnhaber (C) ■ Abbildung S. 47: Marc Jacoby ■ Abbildung S. 48: Norddeutsche Pflanzenzucht-Lembke ■ Abbildung S. 49: Maret Kalda/Martin Sagasser ■ Abbildung S. 51: Norddeutsche Pflanzenzucht-Lembke ■ Abbildung S. 61: P. Heimann ■ Weitere Abbildungen wurden von den Autoren zur Verfügung gestellt ■ Autorenporträts, soweit nicht von den Autoren zur Verfügung gestellt: Norma Langohr

■ Impressum

Herausgeber: Universität Bielefeld, Informations- und Pressestelle ■ Redaktion: Dr. Gerhard Trott (verantwortlich), Dr. Hans-Martin Kruckis ■ Übersetzungen: Günter Seib ■ Satz und Gestaltung: Thomas P. Kiper, Hunteweg 28, 33689 Bielefeld ■ VMK Druckerei GmbH, Faberstr. 17, 67590 Monsheim ■ Anzeigenverwaltung: VMK Verlag für Marketing und Kommunikation GmbH & Co. KG, Faberstr. 17, 67590 Monsheim, Tel. 06243/909-0, Fax 06243/909-400, E-mail: info@vmk-verlag.de, Internet: <http://www.vmk-verlag.de> ■ Erscheinungsweise: in der Regel einmal jährlich ■ Auflage: 4500 ■ Anschrift von Redaktion und Vertrieb: Informations- und Pressestelle der Universität Bielefeld, Postfach 10 01 31, 33501 Bielefeld, Tel. 0521/106-4146, Fax 0521/106-2964 ■ Internet: www.uni-bielefeld.de/Presse/fomag/

■ ISSN 0937-2873

Das Magazin »Forschung an der Universität Bielefeld« kann für 5 Euro pro Jahr abonniert werden ■ Die Mitglieder der Westfälisch-Lippischen Universitätsgesellschaft erhalten das Magazin kostenlos ■ Der ungekürzte Nachdruck von Beiträgen ist unter Nennung des Autors und der Quelle frei ■ Wir bitten um Belegexemplare.

