

Hans-Peter Piepho

Inhalte und Perspektiven der Bioinformatik für die Agrarwissenschaften

Bioinformatik ist eine äußerst breite und vielseitige, sich rasant entwickelnde interdisziplinäre Wissenschaft im Bereich der Life Science Forschung. Sie hat zum Ziel, Kenntnisse und Methoden der Biologie, Chemie, Mathematik, Informatik und Biostatistik zu nutzen, um die Entschlüsselung der Erbinformationen bei Mensch, Tier und Pflanze voranzutreiben und diese Information praktisch nutzbar zu machen. An mehreren Universitäten werden oder wurden Studiengänge und Forschungsschwerpunkte aufgelegt. Die DFG hat kürzlich fünf Universitäten für ihre Förderinitiative ausgewählt (Tübingen, Bielefeld, München, Leipzig, Saarbrücken). Im Rahmen der Agrarwissenschaften befindet sich die Bioinformatik an der Schnittstelle verschiedener Fachdisziplinen mit Bezug zur Genetik, wie z.B. Pflanzen- und Tierzüchtung, Phytomedizin, Pflanzen- und Tierernährung, Biometrie und Agrarinformatik.

Die neuesten Entwicklungen von Hochdurchsatz- (High-Throughput) Techniken zur Generierung genombezogener biologischer Daten lassen sich unter zwei Begriffe fassen: Genomsequenzierung und DNA Chip Technologie (Mikroarrays, Expressionsanalyse). Diese Techniken gewinnen in vielen agrarwissenschaftlichen Forschungsprojekten zunehmend an Bedeutung. In den letzten Jahren wurden die DNA Sequenzen verschiedener Organismen weitgehend entschlüsselt (Arabidopsis, Reis). In Zukunft wird es um die Aufklärung der Funktion und Wirkungsweise von Genen gehen (Postgenomics). Hierbei wird die multiparallele Analyse von Expressionsprofilen auf der Ebene von Transkriptom und Proteom mit Hilfe von Mikroarrays (DNA Chips) eine grosse Rolle spielen, welche besondere Anforderungen an die biostatistische Auswertung stellt (Normalisierung, Datentransformation, Multiplizität beim Test auf differentielle Expression, Verfahren zur Gruppierung von Genen, z.B. Support Vector Machines, Clusteralysen, Self-organized Maps oder Mixture Models). Ein weiterer wichtiger Bereich bioinformatischer Anwendungen im Rahmen der Züchtung ist das Auffinden von Kandidatengenomen durch Homologiesuche in sequenzierten und annotierten Genomen (Arabidopsis, Reis). Hierbei wird auf öffentliche Online-Sequenzdatenbanken zurückgegriffen, wobei statistische Suchalgorithmen wie BLAST und Hidden-Markov-Modelle verwendet werden.

In Zukunft werden die in den 90er Jahren dominierenden genetischen DNA Marker (RFLPs, AFLPs, Mikrosatelliten) zunehmend durch Expressionsdaten (cDNA, Proteine) und Sequenzdaten ergänzt werden. Es werden bioinformatische Tools benötigt, welche es erlauben, genetische Markerdaten, Expressionsdaten und phänotypische Leistungsdaten miteinander zu korrelieren und so genetische Netzwerke zu identifizieren sowie Genorte für quantitative Merkmale (Quantitative Trait Loci, QTL) aufzuspüren und für die Züchtung nutzbar zu machen. Entsprechende zukünftige Strategien werden mit dem Begriff "genetische Genetik" belegt und stellen eine immense Herausforderung für das junge Gebiet der Bioinformatik dar. Die schiere Datenfülle aus High-Throughput Anwendungen übersteigt bisher übliche Datenmengen bei weitem, was hohe Anforderungen an die Entwicklung informatischer Werkzeuge stellt. Demzufolge werden in vielen Züchtungsunternehmen bioinformatische Einheiten neu gegründet, die die wesentliche Aufgabe haben, die Erfassung und Speicherung großer Datenmengen sowie den

Datenfluss zwischen verschiedenen Gruppen innerhalb des Unternehmens zu gewährleisten und bioinformatische Tools für die Datenauswertung zur Verfügung zu stellen. Die Leistungsfähigkeit dieser Einheiten wird in zunehmendem Maß über Erfolg und Misserfolg von Züchtungsprojekten entscheiden.

Wegen der Vielzahl von Aktivitäten an bundesdeutschen Forschungseinrichtungen im Bereich der Life Sciences kommt es beim Aufbau bioinformatischer Einheiten im Agrarbereich darauf an, die komparativen Vorteile strategisch zu nutzen und nicht Dinge zu duplizieren, die anderswo bereits mit großer kritischer Masse durchgeführt werden. Der Vorteil agrarwissenschaftlicher Forschungseinrichtungen liegt in der starken Kompetenz auf dem Gebiet der Tier- und Pflanzenzüchtung sowie in der vorwiegend direkt anwendungsbezogenen Forschung. Im agrarwissenschaftlichen Bereich liegt das Schwergewicht der Forschung auf ökonomisch relevanten phänotypischen Leistungsmerkmalen wie Ertrag, Resistenz und Nährstoffeffizienz bei Pflanzen oder täglichen Zunahmen und Milchleistung bei Tieren. All diese Merkmale sind im hohen Maße komplex, also durch zahlreiche Gene sowie durch vieldimensionale metabolische Pathways und Synthesewege determiniert. Der komparative Vorteil agrarwissenschaftlicher Forschungseinrichtungen gegenüber eher grundlagenorientierten Einrichtungen liegt bei der Expertise hinsichtlich der Analyse, Interpretation und praktischen Nutzbarmachung von Leistungsdaten. Ein wesentliches Ziel ist in diesem Zusammenhang die Weiterentwicklung bioinformatischer Werkzeuge für Assoziationsanalysen zwischen Leistungsdaten auf Ebene des Pflanzenbestandes oder des Einzeltiers einerseits (Makroebene) und Daten auf zellulärer und subzellulärer Ebene, also auf der Ebene von Genom, Transkriptom, Proteom und metabolischer Information andererseits (Mikroebene). Durch solche Werkzeuge kann die in Zukunft exponentiell wachsende Fülle an Information auf der Mikroebene ökonomisch für die Tier- und Pflanzenzüchtung nutzbar gemacht werden.

Im Zuge der "Molekularisierung" der Züchtungsforschung gibt es seit Jahren einen Trend weg von Feld und Stall hin zu Reagenzglas, Gel und Chip, der sich durch die neue Generation von Techniken der Expressionsanalyse noch verstärkt. Bei der Euphorie hinsichtlich des Informationsgehaltes molekularer Daten darf jedoch die Bedeutung der Gewinnung möglichst akurater phänotypischer Leistungsdaten nicht in Vergessenheit geraten. Diese Binsenweisheit manifestiert sich u.a. dadurch, dass in der Liste der "-ome" (Metabolom, Pysionom, etc.) neuerdings auch der Begriff "Phänom" genannt wird. Bei vielen Agrarwissenschaftlern wird dieser Begriff Schmunzeln hervorrufen, deutet er doch an, dass der gute alte Versuch (Feldversuch, Gewächshausversuch, Tierversuch) zur Gewinnung phänotypischer Daten noch nicht ausgedient hat und integraler Bestandteil einer erfolgreichen bioinformatischen Gesamtstrategie in den Agrarwissenschaften sein muss.

Prof. Dr. Hans-Peter Piepho vertritt das Fachgebiet Bioinformatik an der Universität Hohenheim (Institut für Pflanzenbau und Grünland, Fruwirthstrasse 23, 70599 Hohenheim, Tel. 0711-459 2386, e-mail: piepho@uni-hohenheim.de).