

Anforderungen an eine Bioinformatik- Infrastruktur in Deutschland zur Durchführung von bioökonomierelevanter Forschung

Empfehlungen des BioÖkonomieRats

Anforderungen an eine Bioinformatik- Infrastruktur in Deutschland zur Durchführung von bioökonomierelevanter Forschung

Empfehlungen des BioÖkonomieRats

Inhalt

Zusammenfassung	6
Einleitung	7
Bioökonomische Potentiale der modernen Biowissenschaften	7
Themenfelder der Bioinformatik	8
Empfehlungen	11
a) Infrastruktur	11
b) Optimierung der Nutzung von Rechnerkapazitäten	14
c) Die Entwicklung nachhaltiger Forschungs-, Handlungs- und Förderstrategien	15
Anhang	16
Aktuelle Bioinformatikstandorte – Beispiele potentieller Expertisezentren	16
Glossar	30

Zusammenfassung¹

Die Anwendung des Wissens und der Methoden aus den Biowissenschaften werden in den nächsten Jahren eine immer größere ökonomische Relevanz erlangen. Die Bioinformatik wird dabei eine wichtige Rolle spielen. Ein Ausbau der bioinformatischen Infrastruktur ist notwendig, damit die weitere Forschung wie auch die Nutzung wissenschaftlicher Erkenntnisse den bioökonomischen Anforderungen gerecht werden können. Zentrale Themenfelder im Bereich Bioinformatik sind der Aufbau von flexiblen, parallelisierten Einspeisungs-, Bereitstellungs- und Analysepipelines (Datenmanagement), die Weiterentwicklung statistischer Analysemethoden (Datenanalyse) sowie die Optimierung von Vorhersagemodellen (Datenaufbereitung). Zur optimalen Ausgestaltung dieser Themenfelder besteht folgender Handlungsbedarf:

- **Etablierung einer Bioinformatik-Infrastruktur aus einer Reihe von lokalen, gut ausgestatteten und spezialisierten Expertisezentren und einer übergreifenden, koordinierenden Instanz mit den folgenden Aufgaben:**
 - Vernetzung und Förderung der lokalen Expertisezentren zur Sicherung der Technologieentwicklung
 - Ausbau des Wissenstransfers zwischen biologischer Forschung und Bioinformatik
 - Etablierung von Standards in der Speicherung und Analyse von Daten
 - Freie Verfügbarkeit der benötigten Software-Werkzeuge und Standardisierung der Schnittstellen
- **Entwicklung nachhaltiger Forschungs-, Handlungs- und Förderstrategien zur**
 - Verbesserung der Bedingungen für eine gemeinsame Finanzierung von Verbundprojekten durch öffentliche Hand und Industrie
 - Förderung der Nachhaltigkeit der zur Verfügung stehenden Datenressourcen
- **Optimierung der Nutzung von Rechnerkapazitäten zur**
 - besseren Auslastung lokaler Ressourcen durch umfangreiche Ressourcenplanung
 - Verbesserung der Transferbedingungen von Daten im Cloud Computing (Kosten, Sicherheit, Arbeitsaufwand)
 - Bereitstellung von Supercomputern für spezifische Anwendungen

¹ Der BioÖkonomieRat dankt den Mitgliedern des Steering Committees Alfred Pühler, Frank Oliver Glöckner, Alexander Goesmann, Thomas Hartsch, Eric von Lieres, Klaus Mayer, Norbert Reinsch, Chris-Carolin Schön, Wolfgang Wiechert und Ralf Zimmer sowie allen Teilnehmern des Workshops „Bioinformatik“, die maßgeblich zu der Erarbeitung der vorliegenden Empfehlungen beigetragen haben.

Einleitung

Bioökonomische Potentiale der modernen Biowissenschaften

Biologie im Wandel

Mit dem Aufkommen eines großen Spektrums an neuen Methoden und Technologien hat sich die biologische Forschung in den vergangenen Jahren zu einer massiv datengenerierenden Wissenschaft gewandelt. Die gleichzeitige Entwicklung der Bioinformatik stellt die Voraussetzung für die Speicherung, den weltweiten Austausch und die Auswertung dieser Datenmengen dar. Die Kombination neuer Forschungstechnologien, wie z.B. Next Generation Sequencing, Hochdurchsatz-Präzisionsphänotypisierung oder sogenannten Omics-Technologien und bioinformatischer Werkzeuge zur Verknüpfung und Auswertung der gewonnenen Daten, ermöglicht ein tiefgreifendes Verständnis biologischer Zusammenhänge. Dieses reicht von detaillierten Kenntnissen der genetischen Ausstattung einzelner Spezies oder einzelner Organismen, über die Mechanismen der Ausprägung ihrer phänotypischen Merkmale, bis hin zu komplexen Wechselwirkungen, die innerhalb eines Ökosystems zum Tragen kommen.

Von der Grundlagenforschung zur angewandten Wissenschaft

Die wachsenden Erkenntnisse über die Mechanismen, die der Merkmalsausprägung in Organismen zugrunde liegen, generieren neue Möglichkeiten für die nachhaltige und wirtschaftliche Nutzung biologischer Ressourcen. Diese umfassen unter anderem die Entwicklung neuer biotechnologisch nutzbarer Verfahren, die gerichtete züchterische Verbesserung von Nutzpflanzen und Nutztieren sowie eine treffgenauere Ausrichtung von Pflanzenschutz und Tiermedizin. Auch ein tieferes Verständnis evolutionärer Zusammenhänge wird zur Entdeckung und Nutzung neuer biologischer Potentiale mit Hilfe der Biodiversitätsforschung beitragen. So lassen sich z.B. systembiologische Ansätze zur gezielten Ergänzung und Optimierung bisheriger Züchtungsverfahren entwickeln. Auch können ökonomische und nachhaltige Konzepte für die Gewinnung von Biokatalysatoren und bioaktiven Wirkstoffen aus verschiedenen Organismen durch die jüngst geschaffenen Möglichkeiten des direkten Zugriffs auf das genetische Material von nicht im Labormaßstab kultivierbaren Mikroorganismen entwickelt werden. Die heute und in Zukunft verfügbaren Daten erlauben zunehmend eine umfassende Modellierung sowohl der Vorgänge des Zentralstoffwechsels als auch einzelner ausgewählter Synthesewege. Diese Modelle können in der sog. „synthetischen Biologie“ zur Grundlage für die gezielte Neukonstruktion ganzer Stoffwechselwege in technologisch genutzten Organismen dienen.

Die derzeit nicht ausreichend an die Bedarfe der Forschung angepasste Infrastruktur im Bereich der Bioinformatik kristallisiert sich zunehmend als der limitierende Faktor für die zukünftige optimale Ausnutzung des gesamten bioökonomischen Potentials der modernen Biowissenschaften heraus. Dabei zeichnet sich in den verschiedenen Bereichen der biologischen Forschung, von der Grundlagenforschung bis zur angewandten Forschung, ein ähnlicher Handlungsbedarf im Bereich der Bioinformatik ab.

Einleitung

In vielen europäischen und außereuropäischen Staaten ist die zukunftsweisende Bedeutung der Bioinformatik bereits erkannt worden. Staaten wie z.B. die Niederlande (NBIC), die Schweiz (SIB) und Frankreich (ReNaBi) verfügen bereits über umfangreiche und gut organisierte Bioinformatik-Strukturprogramme, andere Länder wie z.B. Schweden (BILS) befinden sich hier gerade in der Aufbauphase. Auch international gibt es bereits Bemühungen zur Etablierung übergreifender Infrastrukturprogramme, um eine bessere Vernetzung und Austauschbarkeit der Daten zu erreichen (z.B. ELIXIR für den europäischen Raum). So verstärkt sich auch im Hinblick auf den internationalen Austausch, aber insbesondere auch auf die Konkurrenzfähigkeit der deutschen Forschung, die Dringlichkeit, den Aufbau einer deutschen Bioinformatik-Infrastruktur voranzutreiben.

Themenfelder der Bioinformatik:

Die Bandbreite der Bioinformatik reicht von der grundlegenden Problematik des Datenmanagements, hierzu zählen insbesondere die Datenhaltung und -strukturierung, bis hin zur Datenanalyse. Beispiele hierfür sind die statistische und quantitative Genetik, die Populationsgenetik sowie die übergreifende Metaanalyse, die biometrische Auswertung von Daten, die Integration verschiedener Datentypen und die Transformation von Erkenntnissen und Techniken aus der Grundlagenforschung hin zur angewandten Forschung und Entwicklung. Gerade die Modellierung und Simulation komplexer Systeme nehmen hier eine immer größere Rolle ein.

Datenmanagement

Im Bereich des Datenmanagements ist es essentiell, Systeme zu schaffen, welche die generierten Daten nicht nur speichern und strukturieren, sondern diese auch für die Auswertung und Interpretation zugänglich machen. Dabei gilt es, nicht nur die exponentiell wachsenden Datenmengen zu bewältigen, sondern auch der sehr großen Heterogenität der Primärdaten Rechnung zu tragen. Um diesen Aufgaben gerecht zu werden, ist der Aufbau von flexiblen, parallelisierten Einspeisungs-, Bereitstellungs- und Analysepipelines erforderlich. Intuitive Werkzeuge zur Visualisierung und Sichtung der Daten werden benötigt. Angesichts der enormen Datenmengen stellt sich vor allem die Herausforderung, effiziente Strategien zur Reduzierung der Datenkomplexität und -menge, z.B. von Primärdaten der Sequenzanalyse, durch geeignete Methoden der Datenreduktion und Datenkompression zu entwickeln.

Datenanalyse

Parallel zum Datenmanagement müssen neue effiziente, statistisch-mathematische Ansätze der Datenanalyse entwickelt werden. In der genetischen Analyse komplexer Merkmale, wie Ertrag und Ressourceneffizienz mittels Genomanalyse und Präzisionsphänotypisierung, entstehen hochdimensionale Datenmengen für deren optimale Nutzung die kontinuierliche Weiterentwicklung statistischer Analysemethoden notwendig ist. Die Integration und vergleichende Bewertung der Daten und Ergebnisse aus unterschiedlichen Forschungs- und Anwendungsbereichen (z.B. Molekularbiologie, Physiologie, Biodiversitätsforschung, Biotechnologie, Züchtung sowie Sensor-, Prozess- und Analysedaten aus der Pflanzen- und Tierforschung) ist die Basis für eine interdisziplinäre und translationale Forschung. Bereits heute gibt es eine Vielzahl von individuellen Datensätzen mit molekularer und phänotypischer Information sowie biologische Ressourcen und Analysewerkzeuge, deren optimale Nutzung erst durch das Zusammenführen der vorhandenen Informationen gewährleistet wird. Dies erfordert innovative bioinformatische Konzepte zur Etablierung von Wissensbanken (Knowledge Bases), welche die Verknüpfung individueller Datenbanken gewährleisten und somit auch die Integrationen heterogener Daten leisten. Sie sind für die Erstellung funktioneller Modelle und Simulationsansätze essentiell, die einen wesentlichen Baustein zukünftigen bioökonomischen Handelns darstellen.

Datenaufbereitung

Die rationale und datengetriebene Auswahl, Identifikation und Validierung geeigneter Modelle erfordert gleichermaßen die Verfügbarkeit lokaler und zentraler Rechenressourcen sowie die Entwicklung maßgeschneiderter und skalierbarer Software zur deren effizienter Nutzung. Beide Aspekte sind für die Kombination zweier besonders relevanter Optimierungsstrategien von herausragender Bedeutung: So sind zur optimalen Planung neuer Versuche oder der gezielten Leistungsoptimierung, z.B. von biotechnologischen Stoffumwandlungsprozessen, zum einen statistisch valide Abschätzungen fortgeplanter Datenunsicherheiten und zum anderen robuste Vorhersagen neuer Messungen notwendig.

Bereits existierende Expertisezentren

In Teilen werden die hier angesprochenen Problemstellungen bereits bearbeitet und für einzelne Bereiche stehen ausgereifte Lösungen und Systeme schon zur Verfügung. Im Bereich der Mikrobiologie haben Forschungsinitiativen wie die GenoMik bzw. PathoGenoMik, oder das europäische Exzellenz-Netzwerk „Marine Genomics Europe“ in den vergangenen Jahren maßgeblich zum erfolgreichen Aufbau der mikrobiellen Genomforschung in Deutschland und Europa beigetragen. Auf dem Gebiet der Pflanzenforschung, wurde unter anderem durch die „GABI/Pflanzenbiotechnologie der Zukunft“-Initiative bereits eine Reihe exzellenter, international anerkannter und sichtbarer Expertisen in verschiedenen Bereichen der grünen Bioinformatik erarbeitet. Damit wurden hervorragende Voraussetzungen für die Etablierung einer mehrschichtigen grünen Bioinformatikplattform geschaffen.

Einleitung

Im Bereich der Tierforschung sind Projekte der FUGATO-Initiative zu nennen. In den Kompetenznetzen der Agrarforschung (Cropsense, Phänomics und Synbreed) ist die Bioinformatik bereits erfolgreich in agrar- und biowissenschaftliche Forschungsverbünde integriert. Zudem ist in Deutschland ein anerkanntes und vorbildhaftes universitäres System zur Ausbildung von Bioinformatik-Nachwuchswissenschaftlern implementiert.

Für den Aufbau der Expertisezentren bietet es sich daher an, zunächst bereits bestehende Bioinformatik-Standorte zu stärken und bei Bedarf zusätzlich die Entwicklung neuer Zentren zu fördern.

Im Bereich der Pflanzenforschung konnten mit dem Bioinformatikzentrum München, den Institutionen der sog. „ABCD / J“-Region (Aachen, Bonn, Köln, Düsseldorf, Jülich), dem Bioinformatik-Zentrum Gatersleben-Halle, der Region Tübingen/Hohenheim und dem Max-Planck-Institut für Molekulare Pflanzenphysiologie in Golm bereits fünf potentielle Expertisezentren beispielhaft aufgezeigt werden. Das Centrum für Biotechnologie der Universität Bielefeld (CeBiTec) bietet sich für den Bereich der Biotechnologie als Expertisezentrum an. Im Bereich der Modellbildung biochemischer Netzwerke und der Supercomputing-Anwendungen in der Systembiologie ist das Forschungszentrum Jülich ebenfalls international ausgewiesen. Die Region Bremen mit dem Max-Planck-Institut für Marine Mikrobiologie, der Jacobs University, dem Zentrum für Marine Umweltwissenschaften der Universität Bremen und dem Alfred-Wegener-Institut für Polar- und Meeresforschung in Bremerhaven stellt ein potentielles Expertisezentrum für den Bereich der Umweltmikrobiologie und Biodiversitätsforschung dar. Auf dem Gebiet der Tierforschung sind Rechenzentren wie Vit Verden und die Bayerische Landesanstalt für Landwirtschaft, Grub, das Leibniz-Institut für Nutztierbiologie in Dummerstorf sowie die Abteilung Tierzucht und Haustiergenetik der Universität Göttingen anzuführen.

Empfehlungen

a) Infrastruktur

Um eine möglichst effiziente, langzeitige und nachhaltige Nutzbarkeit der biologischen Daten für Forschung und kommerzielle Anwendung zu gewährleisten, sollte eine vernetzte und abgestimmte Bioinformatik-Infrastruktur aufgebaut werden, die auch den Aspekt der Translation von Forschungsergebnissen aus der Grundlagenforschung in die Anwendung berücksichtigt. Eine enge Zusammenarbeit von experimentellen und datenerzeugenden Strukturen auf der einen mit Entwicklung und Aufbau bioinformatischer Kompetenz auf der anderen Seite hat sich bereits als wichtige strukturelle Komponente erwiesen.

Der Schlüssel zur Entwicklung einer modernen, entwicklungsfähigen und effizienten Bioinformatik-Infrastruktur liegt in der Etablierung einer zweigleisigen Organisationsstruktur, die zum einen über eine Reihe von lokalen, gut ausgestatteten und spezialisierten Expertisezentren verfügt und zum anderen eine übergreifende Instanz zur Verfügung stellt, welche diese Zentren miteinander vernetzt und koordiniert. Dabei ist anzumerken, dass die geplante Bioinformatik-Infrastruktur nicht puristisch nur bioökonomie-relevanten Aufgaben vorbehalten sein sollte. Vielmehr ist zu prüfen, inwieweit auch eine Quervernetzung mit anderen Bereichen der Lebenswissenschaften als sinnvoll angesehen werden kann.

Lokale Expertisezentren

Die lokalen Expertisezentren gewährleisten durch die Bündelung von Know-how und technischer Ausstattung die Entwicklung von bioinformatischen Lösungsansätzen für spezifische Problemstellungen. Sie stellen die notwendigen Rechenkapazitäten nach Möglichkeit lokal zur Verfügung, für spezielle Anwendungen und Fragestellungen sollten darüber hinaus auch Super-Computing-Kapazitäten eingerichtet werden.

Als potentielle Expertisezentren kommen solche Standorte in Frage, die bereits auf ihrem Gebiet etabliert sind, in denen neben bioinformatischen Werkzeugen wie Software und Datenbanken auch ausreichend Computing-Ressourcen und Fachkräfte zur Verfügung stehen, die durch nationale und internationale Kooperationen gut in die Forschungslandschaft eingebunden sind und die nachhaltige Strukturen für die Ausbildung von wissenschaftlichem Nachwuchs sowie die Schulung von Nutzern aufweisen (siehe Anhang).

Durch die langfristige Förderung und Vernetzung lokaler Expertisezentren sollen die auf breiter Basis genutzten Bioinformatik-Technologien vorgehalten, weiterentwickelt und für die großflächige Nutzung in Forschungsprojekten zur Verfügung gestellt werden. Bei der apparativen Ausstattung ist ferner zu berücksichtigen, dass aufgrund der zum Teil sehr spezifischen Auswertungen oftmals individualisierte Speziallösungen erforderlich sind. Durch die zentralisierte Bereitstellung von bioinformatischem Wissen und bioinformatischen Dienstleistungen könnten insbesondere kleinere Forschergruppen und Neulinge auf dem Gebiet umgehend in die Lage versetzt werden, aus den Daten neues biologisches Wissen zu generieren ohne dabei selbst erst eine eigene Bioinformatik-Infrastruktur aufbauen zu müssen.

Empfehlungen

Zu den Aufgabenbereichen der Expertisezentren würde ferner die breit angelegte Schaffung von Kompetenzen in der bioinformatischen Auswertung von Genom- und Postgenomdaten zählen. Gemeinsame Trainingseinheiten haben sich im Rahmen verschiedener Netzwerke als hervorragendes Mittel für den Wissenstransfer zwischen den beteiligten Institutionen erwiesen. Darüber hinaus stärken sie die Vernetzung zwischen den Instituten. Dies führt zu einer nachhaltigen Stärkung der Genomforschung in Deutschland. Des Weiteren sollte die Ausbildung und Förderung des wissenschaftlichen Nachwuchses, beispielsweise in Graduiertenkollegs, ausgebaut werden. Das erleichtert in Folge den Brückenschlag zwischen Experten der Genomforschung und der Bioinformatik.

Die übergreifende, koordinierende Instanz

Die übergreifende Instanz wirkt als Koordinations-, Kontakt- und Informationsstelle zwischen den Expertisezentren, biologischen und bioinformatischen Forschungseinrichtungen sowie weiteren Anwendern und Interessensgruppen. Durch eine Förderung des Informationsaustauschs zwischen den verschiedenen Zentren sowie zwischen diesen und weiteren nationalen und internationalen Ansprechpartnern aus Forschung und Wirtschaft wird eine gemeinsame Basis geschaffen, von der aus der Vielfalt der bioinformatischen Themen und Aufgabenstellungen Rechnung getragen werden kann.

Durch die Vernetzung der lokalen Bioinformatik-Expertisezentren ist es möglich, die Technologieentwicklung in den verschiedenen Bereichen der angewandten Biologie wie auch der Grundlagenforschung gezielt zu fördern, Standards in der Speicherung und Analyse von Daten zu entwickeln und Konzepte zur Nachhaltigkeit der zur Verfügung stehenden Datenressourcen zu entwickeln.

Während es die Hauptaufgabe der Expertisezentren ist, spezifische Werkzeuge für die Auswertung der Forschungsdaten aus den jeweils verschiedenen Bereichen der Biowissenschaften bereitzustellen, befördert die übergreifende Instanz die Entwicklung und Nutzung gemeinsam benötigter Werkzeuge und Standards. Koordiniert durch die übergreifende Instanz nehmen so alle Expertisezentren an der Entwicklung gemeinsamer Grundlagen der Bioinformatik teil und dienen zudem als Ansprechpartner für spezifische Fragestellungen.

Der Aufbau von standardisierten Analyseschritten (Standard Operating Procedures), einheitlichen Schnittstellen und eine gewissenhafte Datendokumentation durch die Exzellenzzentren sind durch die übergreifende Instanz zu koordinieren. Dies wird vergleichende Analysen beschleunigen und die Qualität der Analysen nachhaltig sichern. Ähnliches gilt für die Bereitstellung spezialisierter Datenbanken und Tools für die Genom- und Biodiversitätsforschung. Der Zugang zu expertengeprüften Referenzdatensätzen (Biocuration) erweist sich zunehmend als Schlüsseltechnologie für die qualitativ hochwertige Auswertung von biologischen Daten sowie die Suche nach neuen Enzymen und Prozessen für biotechnologische Anwendungen.

Darüber hinaus wird die übergreifende Instanz als ein erster, vermittelnder Ansprechpartner in Fragen der Bioinformatik fungieren. Damit kann sie im Bereich der Wissenschaft dazu beitragen, den Wissenstransfer zwischen biologischer Forschung und Bioinformatik zu stärken. Auch wäre eine solche zentrale Anlaufstelle geeignet, den Austausch und die Zusammenarbeit von öffentlichen Forschungseinrichtungen und Wirtschaftsunternehmen zu befördern.

Für die Etablierung der übergreifenden Instanz empfiehlt es sich, in einem ersten Schritt eine „schlanke“ Koordinationsstruktur aufzubauen, welche zunächst die notwendige Vernetzung zwischen den bestehenden Zentren etabliert. Dies ließe sich durch das Einsetzen einer Steuergruppe erreichen, deren Mitglieder die Aktivitäten der verschiedenen Zentren koordinieren und die weitere Entwicklung lenken. Der Schwerpunkt der Aktivität dieser Gruppe sollte dabei zunächst auf dem Aufbau einer Netzwerkstruktur und der Koordination der Entwicklungs- und Standardisierungsprojekte sowie der hierzu notwendigen Ressourcen an den Expertisezentren sein. Ähnlich wie dies bereits in den Niederlanden (NBIC) und in Schweden (BILS) erfolgreich praktiziert wird, könnte eine enge Anbindung zwischen übergreifender Instanz und den Bioinformatikzentren dadurch erreicht werden, dass Mitarbeiter aus den Expertisezentren den Kern der Steuergruppe bilden. Langfristig sollte dann eine Überführung in eine breiter angelegte Institution mit festem Mitarbeiterstamm angestrebt werden, um den Support für Nutzer aus der Forschung, einen hohen Standard der Ausbildung in der Bioinformatik sowie die Kommunikation in die Öffentlichkeit auf hohem Niveau gewährleisten zu können.

Für die Einrichtung einer koordinierenden Struktur ist es auf der organisatorischen Ebene zunächst notwendig, dass sich ein kleiner Kreis von Personen bereit erklärt, die Ausarbeitung des Konzepts der übergreifenden Instanz zu übernehmen und die hierfür notwendige Vernetzung mit den als Expertisezentren in Frage kommenden Forschungseinrichtungen, mit Wirtschaftsunternehmen und mit internationalen Institutionen zu schaffen. Zudem ist ein Finanzierungsmodell zu finden, das es erlaubt, eine solche auf eine langfristige Wirkung angelegte Institution nachhaltig zu fördern. Eine anfängliche Finanzierung durch das BMBF wird dabei angestrebt.

b) Optimierung der Nutzung von Rechnerkapazitäten

Die in den letzten Jahren rasant beschleunigte Datengenerierung und die daraus resultierenden Anforderungen im Bereich der Datenauswertung, insbesondere für Sequenzanalysen und für statistische Vergleiche von Genotypen und Phänotypen, sind mit Einzelplatzrechnern nicht zu bewältigen. Analoge Entwicklungen sind bei den Simulationsmethoden für die modellbasierte Datenauswertung und Versuchsplanung in der statistischen Genetik und der Systembiologie zu beobachten. Darüber hinaus fällt der typische Rechenbedarf einzelner Arbeitsgruppen nicht kontinuierlich an, sondern ist durch temporäre Spitzenlasten gekennzeichnet.

Nutzung lokaler Ressourcen

Lokale Server oder Rechnercluster sind zur Deckung des Grundbedarfs sinnvoll und weit verbreitet. Die Verwendung lokaler Ressourcen ist vor allem dann von Vorteil, wenn interaktiv mit den Daten gearbeitet werden soll. Die Auslastung lokaler Rechnercluster kann grundsätzlich durch die alternierende oder simultane Ausführung rechenintensiver Aufgaben aus anderen Bereichen verbessert werden. Dies erfordert allerdings eine umfangreiche Ressourcenplanung aller beteiligten Arbeitsgruppen.

Nutzung externer Ressourcen – Cloud Computing

Um unregelmäßig auftretende Lasten bewältigen zu können, ist es oft vorteilhaft, die Kapazitäten zentraler Rechenzentren zu nutzen. Externe Rechenkapazitäten können heute in großem Umfang gemietet werden (Cloud Computing). Dabei entstehen jedoch im Vergleich mit lokalen Servern und Clustern höhere Kosten und die Sicherheit vertraulicher Daten kann in der Regel nicht oder nur sehr aufwändig gewährleistet werden. Die mangelnde Datensicherheit wird insbesondere bei Zusammenarbeiten mit Privatunternehmen als äußerst problematisch angesehen. Desweiteren entstehen durch die Notwendigkeit, große Datenmengen wiederholt in die Cloud transferieren zu müssen, erhebliche Zusatzaufwände. Der beständigen Weiterentwicklung der Cloud-Technologien in Hinblick auf Sicherheit und Leistungsfähigkeit wird aus Sicht der Bioinformatik eine hohe Priorität eingeräumt.

Nutzung externer Ressourcen – Supercomputer

Im wissenschaftlichen Umfeld stehen Speicher und Rechenkapazitäten für aufwändige Berechnungen durch Rechenzeitkontingente auf Supercomputern in großem Umfang zur Verfügung. Supercomputer sind allerdings für ungeplante Spitzenlasten nur bedingt geeignet, da Kontingente innerhalb bestimmter Zeitfenster abgerufen werden müssen. Supercomputerzentren können daher als sehr gute Ergänzung zur notwendigen Pflege und dem Ausbau bioinformatischer Hardwarekapazitäten gesehen werden, nicht jedoch als alleinige technische Lösung.

Um sowohl lokale Rechner und Rechnercluster als auch zentrale Großrechner nutzen zu können, ist die freie Verfügbarkeit der benötigten Software-Werkzeuge und eine Standardisierung der Schnittstellen Voraussetzung.

c) Die Entwicklung nachhaltiger Forschungs-, Handlungs- und Förderstrategien

Verknüpfung von akademischer Forschung und privaten Unternehmen

Im Hinblick auf die Rolle, welche der Bioinformatik zukünftig in der Bioökonomie zukommen wird, ist eine vermehrte gemeinsame Finanzierung von Verbundprojekten durch die öffentliche Hand und die Industrie anzustreben. Doch bestehen zurzeit noch zahlreiche Hürden, die ein Zusammenwirken von akademischer Forschung und privaten Unternehmen erschweren. Diese reichen von der Problematik der Publikation der Forschungsdaten aus Public Private Partnerships bis hin zu patentrechtlichen Fragestellungen.

Nachhaltigkeit der Datenressourcen

Von fundamentaler Bedeutung für alle nachgeschalteten analytischen und wissensgenerierenden Prozesse sind die Konzepte zur Nachhaltigkeit der zur Verfügung stehenden Datenressourcen. Hierzu gehören neben den experimentell erzeugten Daten auch deren Metadaten, wie die Beschreibung des Experiments, Herkunft und Art des verwendeten biologischen Materials oder der verwendeten Analysemethoden. Um eine solche Beschreibung zu vereinheitlichen und für Analysemethoden leichter zugänglich zu machen, bieten sich ‚kontrollierte Vokabulare‘ oder noch zu erstellende Ontologien an. So ist neben der Sicherung des Zugangs der Forscher zu modernsten Geräten und Technologien, zur Generierung von Daten, wie sie z.B. zur Hochdurchsatzsequenzierung, der Transkriptomik, Proteomik und Metabolomik notwendig sind, auch zu gewährleisten, dass die gewonnenen Daten für ein breites Spektrum von Anwendungen und über lange Zeiträume hinweg zur Verfügung stehen und leicht wiederauffindbar sind.

Anhang

Aktuelle Bioinformatikstandorte – Beispiele potentieller Expertisezentren

a) Bereich Pflanze

Bioinformatikzentrum München	
Beteiligte Institutionen	Institut für Bioinformatik und Systembiologie am Helmholtz Zentrum München (HMGU) Technische Universität München (TUM)
Spezialisierung	U. a. umfassende Expertise in der Analyse, Darstellung und Bereitstellung von Genomen klassischer pflanzlicher Modellorganismen (Arabidopsis, Medicago, Brachypodium) sowie von Kulturpflanzen (Reis, Tomate, Gerste, Weizen, Roggen, Mais, Sonnenblume). Analyse von Next-Generation-Sequencing Daten und ihre Verknüpfung mit genomischen, evolutionären und funktionell biologischen Fragestellungen bei Modellorganismen und Nutzpflanzen und -tieren. Entwicklung statistischer Methoden zur Analyse quantitativer Merkmale, zur funktionalen Analyse der Biodiversität und zur Aufklärung wichtiger ertragsbildender Mechanismen. Weitreichende Expertise in den Bereichen Statistische Genetik, Pflanzen- und Tierzüchtung und Molekularbiologie, laufende Berufungsverfahren zur Populationsgenetik und Biostatistik.
Vernetzung	Koordination bzw. Beteiligung an zwei langjährig angelegten Forschungsverbänden: dem vom BMBF geförderten AgroClustEr Synbreed und dem DFG Sonderforschungsbereich „Molecular mechanisms regulating yield and yield stability in plants“ Die beteiligten Gruppen sind eng mit nationalen und internationalen Konsortien und Initiativen in Europa und den USA wie z. B. dem Internationalen Weizen und Gersten Genom Sequenzierungskonsortien (IWGSC & IBSC) und der europäischen Pflanzengenominfrastrukturplattform transPlant vernetzt. Umfangreiche Beteiligung an Projekten der GABI / Pflanzenbiotechnologie der Zukunft-Initiative. Zusammenarbeit mit Wirtschaftsunternehmen in Public-Private-Partnerships.
Ausbildung	Die beiden Institutionen stellen gemeinsam die Ausbildung von talentiertem Nachwuchs in den Studiengängen Bioinformatik, Agrarwissenschaften, Biologie und Molekulare Biotechnologie der TUM sicher. Für die Entwicklung von Schnittstellenkompetenz der Studierenden und Doktoranden und um die interdisziplinäre Vernetzung verschiedener Fachdisziplinen zu sichern, engagieren sich beide Einrichtungen gemeinsam in Seminarreihen und Summerschools.
Personal	Ca. 45 Personen in den Bereichen Bioinformatik, Quantitative Genetik, Tier- und Pflanzenzüchtung, Ausweitung durch laufende Berufungsverfahren Populationsgenetik (W2) und Biostatistik (W3)
Software	R-Paket synbreed zur genomischen Vorhersage komplexer Phänotypen
Datenbanken	Keine Angabe
Computerinfrastruktur	Mächtige Compute Cluster am HMGU, Leibniz-Rechenzentrum und Munich Center of Advanced Computing

Pflanzenbioinformatik Standort Gatersleben-Halle

Beteiligte Institutionen	<p>Martin-Luther-Universität Halle-Wittenberg (MLU) mit dem Institut für Informatik</p> <p>Leibniz-Institut für Pflanzenbiochemie in Halle (IPB)</p> <p>Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung in Gatersleben (IPK)</p>
Spezialisierung	<p>Die Analyse von Next-Generation- Sequencing Daten (Assemblierung und Diversitätsstudien, RNA-Seq und ChIP-Seq); die Entwicklung von Datenbanken, Datenintegration sowie Information Retrieval; die Analyse biologischer Netzwerke; die Bildanalyse (Mikroskopiebildern, DNA-Signalen, Hochdurchsatzphänotypisierung von Pflanzen); die angewandte Informatik im Bereich Metabolomics und Massenspektrometrie; die Systembiologie (Modellierung des Stoffwechsel und Flussanalysen); die Visualisierung und visuelle Datenanalyse biologischer Daten; und die molekulare Phylogenie.</p>
Vernetzung	<p>Vielfältige nationale und internationale Kooperationen mit Ländern wie Griechenland, Grossbritannien, Türkei, Schweden, Schweiz, Spanien, Finnland, Frankreich, Österreich, Niederlande, Japan, Australien, USA, Israel, Kanada, Russland, Iran. Zudem gibt es Firmenkooperationen (BASF Plant Sciences, Bayer Crop Sciences, Boehringer, KWS, und vielen anderen Pflanzenzüchtungsunternehmen).</p>
Ausbildung	<p>Neben dem seit 1999 existierenden, nun auslaufenden Diplomstudiengang Bioinformatik (MLU Halle) gibt es einen Bachelor- und Masterstudiengang Bioinformatik (MLU Halle), Promotionen in Bioinformatik (MLU Halle) sowie umfangreiche Lehrtätigkeiten für Bioinformatik-Module (Uni Kiel), den Bachelor- und Masterstudiengang Biotechnologie (Hochschule Anhalt, Standort Köthen) und den Bachelorstudiengang Informatik (Hochschule Harz, Standort Wernigerode).</p>
Personal	<p>MLU: 3 Professuren mit weiteren 4 Haushalts-(HH) und 3 Drittmittel(DM)-finanzierten Wissenschaftlerstellen</p> <p>IPB: eine Gruppe mit 4 Wissenschaftlern (Gruppenleiter, 2 HH- und 1 DM-Stelle)</p> <p>IPK: 6 Arbeitsgruppen (davon 2 rein DM-finanziert) auf dem Gebiet der Pflanzenbioinformatik mit insgesamt aktuell 29 Wissenschaftlern (7 HH und 22 DM)</p>
Software	<p>Es wurden u. a. folgende Tools entwickelt:</p> <p>Alida (Dokumentation von Datenanalysen); MiToBo (Microscope Image Analysis Toolbox); Vanted (Analyse von Omics-Daten im Netzwerkkontext); SBGN-ED (Systems Biology Graphical Notation Editor); LAILAPS (Suchmaschine zum Information Retrieval zur nutzerspezifischen Relevanzbewertung); Datenbanken und Datenintegration von Labordatenmanagement bis zur domänenübergreifenden Datenanalyse; Bioconductor Pakete (xcms, CAMERA, Rdisop, mzR); Tools zur Metabolitenidentifikation; IAP (Bildanalyse von Hochdurchsatzphänotypisierungsdaten); HIVE (Integrative Analyse multimodaler Daten); FBASimVis (Flux Balance Analyse); KGML-ED (KEGG pathway Editor); CentiBin/CentiLin (Zentralitätsanalysen in Netzwerken); Jstacs (Bibliothek für statistische Analysen und Sequenzklassifikation); MotifAdjuster/MiMB/Dispom (Transkriptionsfaktor-Bindungsstellen Annotation und Vorhersage); und PHHMM (Analyse von Array-CGH Daten).</p>
Datenbanken	<p>Beteiligung an der Entwicklung von öffentlichen Datenbanken im Bereich Pflanzenbioinformatik (in Auswahl):</p> <p>GBIS (Bundeszentrales Genbankinformationssystem); MetaCrop (Informationssystem für Metabolismus in Pflanzen) und MassBank (Massenspektrometrie Referenzdaten). Weiterhin wird aktiv in Standardisierungsgremien mitgearbeitet für Standards in der Systembiologie (SBML, SBGN); der Zitierbarkeit von wissenschaftlichen (Primär-)Daten (DataCite - DOI); der Proteomics Initiative (mzML, TraML) und der Repräsentation biologischer und experimenteller Meta- und Primärdaten (ISA-TAB).</p>

Computerinfrastruktur	High Performance Cluster (90 Knoten/200 GigaByte Hauptspeicher); SMP Rechner (8x4-core Opteron/ 256 GigaByte Hauptspeicher); Hierarchical Storage Management (HSM) system (~65 TeraByte / 9 TeraByte online Zugriff); High Performance Cluster (1840 Kerne / ~2.17 TB RAM); 3D Visualisierungsstation; und IPB Compute Cloud (650 CPU Kerne, zentrales SAN Speichernetzwerk)
-----------------------	--

Bioinformatik-Aktivitäten in Tübingen und Hohenheim

Beteiligte Institutionen	<p>Interfakultäres Zentrum für Bioinformatik Tübingen (ZBIT): Eberhard-Karls-Universität Tübingen, Universitätsklinik Tübingen, MPI für Entwicklungsbiologie, MPI für Intelligente Systeme,</p> <p>Friedrich-Miescher-Laboratorium Hohenheim: Universität Hohenheim (Fakultät Agrarwissenschaften), Landesanstalt für Pflanzenzüchtung</p>
Spezialisierung	<p>ZBIT: Verschiedene Gebiete der Bioinformatik</p> <p>Hohenheim: Statistische Genomik</p>
Vernetzung	<p>Hohenheim: Verschiedene internationale Kooperationen im Rahmen von GABI / PLANT2030 und SYNBREED. Es besteht eine wissenschaftliche Kollaboration zwischen dem Institut für Pflanzenzüchtung (Schmid) und dem MPI Tübingen (Weigel).</p>
Ausbildung	<p>ZBIT: Die Universität Tübingen hat 1998 den ersten Studiengang für Bioinformatik in Deutschland eingerichtet. Heute findet die Ausbildung über BSc-/MSc-/PhD-Programmen in Bioinformatik statt. Derzeit gibt es ca. 220 Studierende, 50 Doktoranden, die an mehreren Graduiertenschulen beteiligt sind.</p> <p>Hohenheim: In Hohenheim findet die Ausbildung zur Bioinformatik und statistischen Genomik in den BSc und MSc-Programmen 'Agrarwissenschaft' mit den Schwerpunkten Pflanzen- und Tierzüchtung, im MSc-Programm 'Crop Sciences' mit der Vertiefung 'Plant Breeding' sowie in den PhD-Programmen statt. Derzeit gibt es ca. 100 Studierende und 20 Doktoranden in diesem Bereich.</p>
Personal	<p>ZBIT: 14 Arbeitsgruppen auf verschiedenen Gebieten der Bioinformatik</p>
Software	<p>ZBIT: In Tübingen werden zahlreiche Softwarepakete entwickelt, die für die grüne Bioinformatik verwendet werden, u. a.: Metagenomik (MEGAN Paket), Phylogenien (SplitsTree), Galaxy server (gene prediction, cis-elements etc.), Short read assembly (LOCAS), Molecular modeling (BALL), Proteomics (OpenMS), Systembiologie (BN++ [BioMiner]), Integrierte next generation sequencing Analyse (SHORE Paket), NGS aligner (QPALMA, PALMapper, GenomeMapper) und NGS Transkriptom Analyse (rQUANT).</p> <p>Hohenheim: In Hohenheim wurde eine Reihe von Softwarepaketen für die statische Genomik entwickelt. Beispiele sind: Genetische Kartierung (PLABQTL), Simulation von Geno- und Phänotypen (phenosim, R-hypred) und die Auswertung von Züchtungsprogrammen (PLABSTAT, R-selectiongain, R-mvngGrAd).</p>
Datenbanken	Keine Angabe
Computerinfrastruktur	<p>ZBIT: In Tübingen sind jeweils ein Cluster an der Universität und auf dem Max-Planck-Campus, die von Personal mit Dauerstellen betreut werden.</p> <p>Hohenheim: In Hohenheim betreiben die Fachgebiete des Instituts für Pflanzenzüchtung einen Cluster mit mehr als 100 Knoten, der ab 2012 von einem Mitarbeiter mit einer Dauerstelle betreut werden wird. Zusätzliche externe Rechenkapazitäten werden über Kooperationen genutzt.</p>

Max-Planck-Institut für Molekulare Pflanzenphysiologie, Golm	
Beteiligte Institutionen	<p>Max-Planck-Institut für Molekulare Pflanzenphysiologie (MPIMP, Golm, Zentrale Infrastrukturgruppe Bioinformatik),</p> <p>Universität Potsdam (Golm, Bioinformatikgruppe, Lehrstuhl Prof. Selbig)</p>
Spezialisierung	<p>Omics-Datenmanagement und -analyse: Entwicklung von Datenbanken zur Verwaltung von Omics-Daten, speziell von Metabolomics-Daten und Next Generation Sequenzierdaten, Markeridentifizierung, z. B. im Kontext von Züchtung, Entwicklung und Anwendung von statistischen Methoden der Omics-Datenanalyse;</p> <p>Systembiologie: Analyse von Omics-Daten vor dem Hintergrund von Signal- und Stoffwechselwegen, Netzwerkrekonstruktion aus Omics-Daten;</p> <p>Genomweite Assoziationsstudien: Entwicklung von Werkzeugen zur Detektierung von Genotyp-Phänotyp-Assoziationen;</p> <p>RNA: Untersuchungen der Sequenz-Struktur-Funktions-Beziehungen von RNA Molekülen, insbesondere nichtkodierende RNA (miRNA), Entwicklung und Bereitstellung von Methoden zur funktionellen Klassifizierung von RNA</p>
Vernetzung	<p>Mit experimentell arbeitenden Arbeitsgruppen des MPIMP und der Universität Potsdam;</p> <p>Austausch mit Bioinformatikgruppen der Region (Leibniz Institut für Pflanzenbiochemie, Halle, Humboldt-Universität Berlin, MPIMG Berlin-Dahlem);</p> <p>Zahlreiche nationale und internationale Kontakte über Projekte: MPIMP: u. a. Uni Erlangen, IMB-Aachen, Universität Wien, Aberdeen University;</p> <p>Uni Potsdam: U. a. IPK Gatersleben, Ludwig-Maximilians-Universität München, National Institute of Biology Ljubljana, 8 Europäische Partner eines EU-MC-ITN</p>
Ausbildung	<p>Universität Potsdam: Master Bioinformatik;</p> <p>Doktorandenausbildung am MPIMP und an der Universität Potsdam</p>
Personal	<p>MPIMP: Ca. 300 Mitarbeiter, davon: MPIMP-Bioinformatik: 9 Mitarbeiter (1 Gruppenleiter, 4 Postdocs, 2 Programmierer, 2 Doktoranden), weiterhin Bioinformatik-orientierte Mitarbeiter in zahlreichen Arbeitsgruppen;</p> <p>Universität Potsdam: AG-Bioinformatik (Prof. Selbig): 10 + Mitarbeiter (1 Lehrstuhlinhaber, 6 Postdocs, 2 Doktoranden, 1 Systemadministrator, studentische Mitarbeiter); Bioinformatik-orientierte Mitarbeiter der AG-Mathematische Modellierung und Systembiologie (Prof. Huisinga); mehrere Bioinformatik-orientierte Mitarbeiter in anderen Arbeitsgruppen</p>
Software	<p>Vielzahl von selbstentwickelten stand-alone und webbasierten Softwarewerkzeugen zu Omics-Datenanalyse (MetaGeneAlyse und pcaMethods zur statistischen Datenauswertung);</p> <p>Spezialisierte Software zur Auswertung von Proteomics (IOMACS)- und Metabolomicsexperimenten (TagFinder, GoBioSpace U. a.);</p> <p>Kommerzielle Software (CLC, Statistica, Matlab, Mathematica), Public Domain Software (R, MeV etc.);</p> <p>Betriebssysteme: Linux, Windows; Programmiersprachen / -umgebungen: Python, Perl, Java, C, .Net, C#, R, MATLAB; Webprogrammierung, Datenbanken: SQL, MySQL, Postgres</p>

Anhang

Datenbanken	Golm Metabolome Database (GMD, GC/MS Daten), ChlamyCyc (Chlamydomonas Stoffwechselwege, Gene und Proteine), RLoop (RNA-Loop-Strukturen), NGS Small-Reads-DB, TROST (Kartoffel Trockenstressdaten), ChlExDa (Chlamydomonas Experimentelle Daten), AraNet (Expressionskorrelationsnetzwerke in Modellpflanzen); GABI-PD, GABI-Primärdatenbank
Computerinfrastruktur	MPIMP: 12 Server, 40 TB Plattenplatz, Anzahl der Kerne: 88 Uni-Potsdam: 10 Host-Rechner mit 96 Rechenkernen, 10 Arbeitsplatzrechner, 48 TB zentraler Festplattenspeicher

b) Bereich Tier

Vit Verden	
Beteiligte Institutionen	VIT ist Dienstleistungsrechenzentrum für die organisierte Tierzucht (z. B.: Landeskontrollverbände, Rinder-, Pferde-, Schaf- und Schweine-Zuchtorganisationen, Genotypisierungslabore)
Spezialisierung	EDV-Anwendungen und genetisch statistische Auswertungen im Bereich Tierproduktion und Tierzucht
Vernetzung	Dt. Universitätsinstitute und Forschungsanstalten im Bereich Tierzucht Andere nationale und internationale Rechenzentren und Forschungseinrichtungen im Bereich Tierzucht
Ausbildung	Promovierte Diplomagraringenieure (Spezialisierung Tierzucht), Programmierer, DB-Experten, Bioinformatiker
Personal	Ca. 120 Mitarbeiter (Fachabteilungen, Programmierer, Systembetreuung)
Software	Eigene Softwareentwicklung (JAVA, Fortran, SAS)
Datenbanken	Oracle
Computerinfrastruktur	Verteilte Serversysteme, Linux-Cluster

Georg-August-Universität Göttingen, Abteilung Tierzucht und Haustiergenetik

Beteiligte Institutionen	Georg-August-Universität Göttingen, Gesellschaft für wissenschaftliche Datenverarbeitung mbh Göttingen (GWDG)
Spezialisierung	Wissenschaftliches Rechnen mit vielfältigen Datenstrukturen, Verarbeiten von tierzuchtspezifischen Daten, Aufbereitung von Daten aus der Hochdurchsatz-Genotypisierung, Hochdurchsatz-Phänotypisierung und Next-Generation Sequencing, Genomische Modelle, Assoziationskartierung, Populationsgenetik
Vernetzung	Synbreed, Fugato-Projekte, DFG-GRK 1664 'Skalenprobleme in der Statistik', Zentrum für Statistik der Universität Göttingen

Ausbildung	Promovierte Agrarwissenschaftler, Diplom-Agrarwissenschaftler, Diplom-Mathematiker, Diplom-Physiker
Personal	Abteilungsleiter, 2 wiss. Mitarbeiter, 2 Postdocs, 10 Doktoranden, 2 Programmierer
Software	Mathematische Software (Mathematica, Maple) Gängige statistische Software (Statistica, SPSS, SAS, R, etc.) Tierzuchtspezifische Software (z. B. ASReml, VCE, PEST, ZPLAN+, ...) Software für Sequenzdaten; Verarbeitung von Rohdaten (z. B. PHRAP, PHRED) Sequenzvergleiche und Multialignment (z. B. BLAST, bwa, Phylip, SAMtools) Sequenzdatenbanken (z. B. EMBL) Programmiersprachen R, C++, Python, Fortran
Datenbanken	MySQL und Oracle-Datenbanken bei der GWDG
Computerinfrastruktur	GWGD: Cluster mit mehreren Parallelrechnern (Intel Xeon -und AMD Opteron-Systeme) mit Batch-Systemen, insgesamt über 5000 Cores und über 18000 GB RAM GWGD Systeme zur Datensicherung und Serververwaltung Verschiedene eigene Server, z. B. AMD Opteron mit 24 Kernen, 2,1 GHz Taktung, 128 GB RAM

Leibniz-Institut für Nutztierbiologie

Beteiligte Institutionen	Leibniz-Institut für Nutztierbiologie, Wilhelm-Stahl-Allee 2, 18196 Dummerstorf
Spezialisierung	Integrative Bioinformatik beim Nutztier (Rind, Schwein) speziell für Leistungsmerkmale, funktionale Merkmale und Verhaltensmerkmale Statistische Genomik, genetische Statistik, Schätzung genetischer Parameter, Zuchtwertschätzung, Zuchtplanung, Populationsgenetik in der Züchtung Ontologie für Verhaltensmerkmale
Vernetzung	Kompetenznetzwerk Phänomics Universität Rostock, Lehrstuhl für Systembiologie und Bioinformatik Institut für Neuro- und Bioinformatik, Universität zu Lübeck, Institut für Tierzucht und Tierhaltung, Christian-Albrechts-Universität Kiel Institute for Bioengineering and Food Science, Biostatistics Group, University of Life Science Aas, Norway Vit Verden, Zuchtverbände, Leistungsprüfungsorganisationen und andere Praxispartner
Ausbildung	R-Kurse für Internationale Leibniz-Graduiertenschule DiVa und FBN-Doktoranden Biomarkerlab für Studenten der Molekularen Biotechnologie und der Informatik der Uni Rostock Gene-Set-Enrichment als Teil der Vorlesung Molekulare Bioinformatik II in den Masterstudiengängen Molecular Life Sciences und Informatik der Universität zu Lübeck Modul Lineare und gemischte lineare Modelle im Masterstudiengang Tierwissenschaften der Universität Rostock

Anhang

Personal	Arbeitsgruppe Biomathematik und Bioinformatik mit vier Wissenschaftlern Arbeitsgruppe Tierzucht und Haustiergenetik mit vier Wissenschaftlern Nachwuchsgruppe Integrative Bioinformatik beim Rind (2 DM-Stellen für fünf Jahre) Postdoktorandenstelle zur Ontologieentwicklung (1 DM-Stelle für fünf Jahre)
Software	Simulation von nutztiertypischen Genotypverteilungen und Phänotypen; Algorithmenentwicklung für integrative Bioinformatik beim Nutztier; Ontologieentwicklung mit Schwerpunkt Tierverhalten
Datenbanken	Projektdatenbank für Kompetenznetzwerk Phänomics; Projektdatenbank für integrative Bioinformatik beim Rind
Computerinfrastruktur	Das FBN besitzt zurzeit fünf Compute-Server mit insgesamt 124 Knoten. Außerdem besteht die Möglichkeit, externe Rechnerkapazitäten (2 Cluster mit 30 und 10 Knoten) der Universität Rostock zu nutzen (Kooperationsvertrag).

Bayerische Landesanstalt für Landwirtschaft, Grub

Beteiligte Institutionen	Bayerische Landesanstalt für Landwirtschaft, Grub
Spezialisierung	Genomische Zuchtwertschätzung von Rindern und Schweinen Genomweite Assoziationsstudien bei Rindern und Schweinen
Vernetzung	TU München Christian-Albrechts Universität zu Kiel Uni Hohenheim ZuchtData GmbH, Wien Landesamt für GeoInformation und Landentwicklung, Kornwestheim Landeskuratorium der Erzeugerringe für tierische Veredelung, Bayern
Ausbildung	Keine Angabe
Personal	6 Wissenschaftler, 2 Doktoranden, 2 Programmierer
Software	Verwendete Software: R, SAS; Beagle; findhap V2; ASReml, DMU, MiX99,... Selbst entwickelte Software: unter Fortran und Perl
Datenbanken	Genotypen: Alle Genotypen der Rinderrassen, Fleckvieh und Braunvieh in Deutschland, ca. 15.000 Genotypen Fleckvieh, ca. 5.000 Genotypen Braunvieh, hauptsächlich Illumina 54K Bead-Chip, teilweise Illumina 777K Bead-Chip (ca. 1.500), Genotypen von 2.000 Schweinen der Rassen Deutsche Landrasse und Deutsches Edelschwein, Illumina 60K Bead-Chip Phänotypen: Leistungsdaten der o.g. Rassen in 44 Merkmalen seit 1990, Abstammungsdaten der o.g. Rassen seit 1950, Zuchtwerte der o.g. Rassen in 44 Merkmalen Datenbanksysteme: OraclemySQL
Computerinfrastruktur	Windows Arbeitsplatzrechner (Standard 4GB RAM), 5 Linux-Workstations unter Debian: Dual-Xeon 4- bzw. 6-Core Ausstattung, RAM-Ausstattung 16, 19 bzw. 64 GB je Maschine, Anbindung an 4 TB Festplattenserver, 1 Linux Workstation mit Oracle-DB, 1 IBM 550Q mit 2 Power5 Quadcore Proz. und 64 GB RAM unter AIX5L

c) Bereich Mikrobiologie und Biotechnologie

Technologieplattform Bioinformatik am Centrum für Biotechnologie (CeBiTec) der Universität Bielefeld	
Beteiligte Institutionen	Die „ Bioinformatics Resource Facility “ (BRF) – eine aus der Bioinformatik-Initiative der DFG (2001) resultierende wissenschaftlich ausgerichtete Dienstleistungs- und Entwicklungseinrichtung – administriert die von den Einheiten des CeBiTec genutzten Rechner, und sie unterstützt die diversen Großprojekte der Genomforschung insbesondere durch die Integration und Neuentwicklung von Datenbankanwendungen zur effizienten Speicherung der anfallenden Primärdaten sowie durch die Implementierung von Software zur Analyse großer genombasierter Datenmengen.
Spezialisierung	Das Spektrum der zumeist biotechnologisch ausgerichteten Forschungsprojekte reicht von der Analyse mikrobieller Genome und Metagenome über die Bearbeitung von Pilzen (insbesondere Hefen), Algen und Pflanzen (Ackerschmalwand, Zuckerrübe, Weinrebe, Raps) bis hin zu tierischen Zellkulturen (z. B. Chinese Hamster Ovary Cells).
Vernetzung	Neben dem Einsatz der entwickelten Systeme in verschiedenen Genom- und Postgenomprojekten am CeBiTec greifen auch zahlreiche externe Partner in nationalen und internationalen Kooperationen auf die Bielefelder Infrastruktur zu. Insgesamt sind aktuell ca. 500 interne und mehr als 2.700 externe Benutzer registriert, davon stammen ca. 55% aus Deutschland. Ausgewählte Projekte: GenoMik, GenoMik-Plus, GenoMik-Transfer, PathoGenoMik, Marine Genomics Europe, Grain Legumes, GABI-Kat, NuGGET, AnnoBeet, SysMap, GK Bioinformatik, CLIB Graduate Cluster,...
Ausbildung	Im Bereich Lehre ist die bioinformatische Ausbildung an der Universität Bielefeld in den Bachelor- bzw. Masterstudiengängen „Bioinformatik und Genomforschung“ und „Naturwissenschaftliche Informatik“ verankert, sowie im Masterstudiengang „Genombasierte Systembiologie“. Die bioinformatische Ausbildung wird getragen durch 4 Lehrstühle (E. Baake – Biomathematik & Theoretische Bioinformatik; J. Stoye – Genominformatik; R. Giegerich – Praktische Informatik; R. Hofestädt – Bioinformatik & Medizinische Informatik) und weitere Arbeitsgruppen (T. Nattkemper – Biodata Mining & Neuroinformatik; NWG A. Sczyrba – Computational Metagenomics & Single Cell Genomics; NWG A. Goesmann – Computational Genomics).
Personal	Die BRF ist neben der Leiterstelle mit 6 dauerhaften Mitarbeiterstellen für die Systemadministration ausgestattet, die allesamt von der Universität finanziert werden. Ferner werden kontinuierlich 3 studentische Hilfskräfte für Routinetätigkeiten eingesetzt, deren Finanzierung Haushaltsmitteln des CeBiTec entnommen wird. Zur gebündelten Vertretung der Interessen der verschiedenen Arbeitsgruppen am CeBiTec wurde ein BRF-Koordinationsausschuss eingerichtet, der in enger Zusammenarbeit mit den Systemadministratoren die Weiterentwicklung der CeBiTec-Rechnerinfrastruktur plant und gestaltet.

Software	<p>Neben der Administration und Weiterentwicklung der technischen Infrastruktur arbeitet die BRF darüber hinaus auf dem Gebiet der angewandten Bioinformatik, indem sie aktiv Softwarelösungen für Hochdurchsatz-Analysen auf dem Gebiet der Genom- und Postgenomforschung entwickelt. Schwerpunkte bilden dabei die DNA-Sequenzanalyse und Genomannotation (SAMS, GenDB, EDGAR SARUMAN, Conveyor) inklusive der Rekonstruktion von Stoffwechselwegen (CARMEN), die Hochdurchsatz-Analyse im Bereich der Transkriptomik (EMMA), Proteomik (QuPE) und Metabolomik (MeltDB), sowie das allgemeine Datenmanagement und die Visualisierung (ProMeTra). Die genannten Systeme werden durch eine Integrationsschicht namens „BRIDGE“ miteinander verbunden. Weitere Software-Tools des Instituts für Bioinformatik am CeBiTec: CARMA, CPA, r2cat, Gecko, GISMO, REGANOR, QAlign, PASSTA, BACCardl, Genlight, RNACast, RNASHapes, Genalyzer, PathFinder, RNAhybrid, VANESA, PathAligner, TACOA, BIIGLE, etc. Neben den zuvor genannten webbasierten Programmpaketen und der virtuellen Arbeitsumgebung stellt der Bielefelder Bioinformatik-Server (BiBiServ) weitere Bioinformatik-Anwendungen für anonyme Benutzer zur Verfügung. Mit seiner Web-Services-basierten Programmierschnittstelle bietet er eine zusätzliche Möglichkeit, externen Benutzern neu entwickelte Werkzeuge auf einer etablierten Plattform zur Verfügung zu stellen.</p>
Datenbanken	<p>DAWIS-M.D., CardioVINEdb, BioDWH, RAMEDIS, STCDB, Coryne-RegNet, Myco-RegNet, BIOIMAX, etc.</p>
Computerinfrastruktur	<p>Der Hardware-Park des CeBiTec umfasst heute eine Rechenleistung von ca. 25 TeraFLOPS (796 CPUs bzw. 4.024 CPU Kerne), eine Online-Speicherkapazität von 433 Terabyte und eine Brutto-Backup-Kapazität von ca. 1,4 Petabyte. Zur langfristigen Archivierung der Rohdaten und zur täglichen Sicherung der Daten werden in der BRF drei Bandsysteme mit einer möglichen Endausbaukapazität von 16 Petabyte eingesetzt. Zur Prozessierung der Primärdaten und zur weiteren Hochdurchsatzanalyse steht ein Compute-Cluster zur Verfügung. Dieser wird unter anderem für vielfältige DNA-Sequenzanalysen wie beispielsweise die Annotation von Genomen oder Metagenomanalysen eingesetzt.</p> <p>Für spezielle Bioinformatik-Anwendungen mit z. B. sehr hohem Speicherbedarf stehen diverse Applikationsserver bereit, die mit bis zu 96 CPU-Cores und maximal 1024 GB Arbeitsspeicher ausgestattet sind. Außerdem sind hier Spezialrechner zu nennen, die besonders für die spezifischen Anforderungen der Bioinformatik ausgelegt sind. So werden 4 Server mit insgesamt 12 TimeLogic DeCypher-FPGA-Karten zur Beschleunigung von BLAST-Analysen eingesetzt. Für GPU-basierte Ansätze, wie beispielsweise das Read Mapping mit der in Bielefeld entwickelten Software SARUMAN wurden drei IBM iDataPlex-Server mit jeweils zwei NVIDIA Tesla M2070-GPU-Karten beschafft. Für die eigenständige Entwicklung von FPGA-basierten Algorithmen steht außerdem ein Convey HC-1ex-System mit vollem Arbeitsspeicherausbau zur Verfügung. Aufsummiert über die vergangenen 10 Jahre entspricht diese Ausstattung einem Investitionsumfang von mehr als 6 Mio. Euro aus DFG-, BMBF- und EU-Projekten sowie aus Sonderzuwendungen und Eigenmitteln der Universität.</p> <p>Ein weiterer wichtiger Bestandteil der Bielefelder Bioinformatik-Infrastruktur ist die virtuelle Arbeitsumgebung basierend auf Sun Ray-Thin-Clients. Durch den Einsatz der energiesparenden und kostengünstigen Terminals werden heute mit minimalem Administrationsaufwand mehr als 350 Bioinformatik-Arbeitsplätze bereitgestellt. Dabei kann die BRF auf einen beinahe unterbrechungsfreien und äußerst stabilen zwölfjährigen Betrieb verweisen. Insbesondere wurde die Tauglichkeit eines derartigen virtuellen Arbeitsplatzes für den weltweiten Einsatz bereits durch einen erfolgreichen Betrieb der Terminals im WAN u. a. in Europa, Südamerika und Asien belegt.</p>

d) Bereich Pflanze und Mikrobiologie

Bioinformatik in der ABCD / J Region	
Beteiligte Institutionen	RWTH Aachen Universität Bonn Universität Düsseldorf Forschungszentrum Jülich MPI für Pflanzenzüchtungsforschung Köln
Spezialisierung	<p>RWTH Aachen: Die RWTH Aachen hat sich auf dem Gebiet der weißen Systembiologie (Prof. Blank) verstärkt und eine Bioinformatikgruppe ist von der Medizin ausgeschrieben.</p> <p>Universität Bonn: Die Universität Bonn besitzt mit dem Institut für Nutzpflanzenwissenschaften und Ressourcenschutz mit Prof. Léon Expertise auf dem Gebiet der Zuchtwertschätzung. Prof. Ewert leitet die Modellierungsarbeiten an der Landwirtschaftlichen Fakultät der Universität Bonn (v. a. Ertragsmodellierung). Eine weitere Professur für statistische Genetik ist in Planung.</p> <p>Forschungszentrum Jülich: Europäisches Pflanzen Phänotypisierungs Netzwerk (EPPN). Weiterhin findet sich am IBG1 (Prof. Wiechert) große Expertise in der weißen Biotechnologie und dort auch in der Modellierung von Netzwerken und biotechnologischen Prozessen.</p> <p>MPI für Pflanzenzüchtungsforschung Köln: Expertise auf dem Gebiet der Pflanzenzüchtung und -genetik</p> <p>Universität Düsseldorf: Metagenomics, metabolische Netzwerke</p>
Vernetzung	Kooperationen bestehen u. a. mit INRA Frankreich, International Arabidopsis Informatics Consortium, International Plant Phenotyping Network, International Tomato Annotation Group, International Medicago Genome Annotation Group, IPK, iPlant, MIPS Munich, MPI Golm & Tübingen, Universität Bielefeld, University Toronto: BAR Viewer, Perth Plant Energy Biology Center SUBA, PlantsDB, Tomato / potato trait, phenotype and mapping database, Wageningen University, Fraunhofer Institute (Aachen Fraunhofer FIT: Life Science Informatics, Bonn: Fraunhofer SCAI-Bioinformatics)
Ausbildung	<p>RWTH Aachen: Bisher Bioinformatikmodule im Studiengang Biotechnologie, Bioinformatik soll nun in die Biologie eingebracht werden.</p> <p>Universität Bonn: MSc Life Science Informatics, B-IT Center, Bonn; Bioinformatics im MSc Crop Science.</p> <p>Universität Düsseldorf: Für Biologen besteht ein umfangreiches Bioinformatik Curriculum (http://www.molevol.de/~bioinf/).</p>

Anhang

<p>Personal</p>	<p>RWTH Aachen: Ab Mitte 2012 wird eine Hausstelle als Rat / Oberrat besetzt.</p> <p>Universität Bonn: Nachwuchsgruppenleiterin als Akad. Rätin, Systemadministrator unbefristet</p> <p>Universität Düsseldorf: Zwei Lehrstühle Bioinformatik innerhalb des Studiengangs Informatik</p> <p>Forschungszentrum Jülich: Für die geplante Nachfolge der Gabi Primärdatenbank: eine Anschlussfinanzierung von 2 FTE durch das Forschungszentrum im Rahmen der Berufungszusage an Prof Usadel. Die AG von Prof. Usadel am FZJ wird ebenfalls zusätzliches grundfinanziertes Personal umfassen.</p> <p>Im Rahmen des DPPN (Deutsches Pflanzenphänotypisierungsnetzwerk): Rückgriff auf IT Struktur am IBG2 (Prof Schurr), Kapazitäten des Jülich Plant Phenotyping Center (JPPC) sollen weiter ausgebaut werden.</p> <p>Am IBG1: eine Modellierungsabteilung (Biochemische Netzwerke und biotechnologische Prozesse)</p> <p>Am IBG2: eine Modellierungsgruppe (Struktur-Funktions-Modelle Wurzel und Spross).</p> <p>Mit dem BioSC assoziiert: Das MPI für Pflanzenzüchtungsforschung Köln mit drei Gruppen mit direktem Bezug zur Grünen Bioinformatik (Dr. Schneeberger (NGS Mapping), Dr. Jimenez-Gomez (adaptive genomics and genetics) und Dr. Stich (Quantitative crop genetics) sowie mit Prof. Koornneef).</p>
<p>Software</p>	<p>RWTH Aachen: Mercator (MapMan Annotation), Robin (Microarrayanalyse), R-Robin (RNA Seq Analysen), MapMan, PageMan (Visualisierung von Omics-Daten), Corto</p> <p>Universität Bonn: Funktionsannotation: PhyloFun, AHRD, R-Pakete für ChIP-chip / ChIP-Seq (ChipR), Aggregatoren/Workflowtools für web services</p> <p>Universität Düsseldorf: PhyloPythiaS</p> <p>Forschungszentrum Jülich: OMIX (Netzwerk-Editor), 13CFLUX (Stoffflussanalyse), CADET (Chromatographie)</p>
<p>Datenbanken</p>	<p>RWTH Aachen: MapMan (Funktionelle Klassen), CSB.DB (Korrelationsdatenbanken)</p> <p>Uni Köln: Datenbank Aramemnon (Gruppe von Prof. Flügge)</p> <p>Universität Bonn: AFAWE, Funktionsvorhersagen http://afawe.mpipz.mpg.de</p> <p>Forschungszentrum Jülich: Gabi Primärdatenbank Nachfolge; phänotypische Datenbanken, Kopplung Phänotyp-Genotyp-Datenbanken</p>
<p>Computerinfrastruktur</p>	<p>Forschungszentrum Jülich: Supercomputing Center, sowie Cluster am IBG1 und IBG2</p>

e) Bereich Umweltmikrobiologie und Biodiversitätsforschung

Bremer Infrastruktur für Umweltmikrobiologie und Biodiversitätsforschung	
Beteiligte Institutionen	<p>Max-Planck-Institut für Marine Mikrobiologie Bremen</p> <p>Jacobs Universität Bremen</p> <p>Zentrum für Marine Umweltwissenschaften, Universität Bremen (MARUM)</p> <p>Alfred-Wegener-Institut für Polar- und Meeresforschung Bremerhaven (AWI)</p>
Spezialisierung	<p>MPI-Bremen / Jacobs University: Der bioinformatische Schwerpunkt des MPI-Bremen/Jacobs University ist die mikrobielle Diversitäts- und Genomforschung. Neben der Sequenzanalyse und Klassifikation (Binning) stehen dabei vor allem der Aufbau und Betrieb von Referenzdatenbanken (SILVA-Projekt) sowie die Datenintegration von Diversitäts- und Funktionsdaten mit Umweltparametern im Vordergrund (Megx-Projekt). Desweiteren beteiligt sich das MPI/Jacobs aktiv an der Entwicklung von Metadatenstandards, Austauschformaten und Ontologien, um den Datenaustausch und Interoperabilität von Daten und Datenbanken zu verbessern.</p> <p>AWI / MARUM: Langzeitarchivierung und Publikation von biologischen Umweltdaten. Zudem hat PANGAEA sich in den letzten Jahren aktiv am Aufbau von Geodateninfrastrukturen und der Entwicklung einschlägiger Standards beteiligt. Daten werden typischerweise über zentrale Portalendienste zur Verfügung gestellt, wobei PANGAEA® einerseits in der Rolle als „Daten- und Metadatenverteiler“ und andererseits als zentraler Netzwerkbauer, Portalbetreiber und Broker zwischen den unterschiedlichen e-Infrastrukturen auftritt. Über diverse Metadatenstandards und Protokolle (OGC-CS, OAI-PMH, DiGIR, ABCD) werden diverse Portale und Suchmaschinen mit den Inhalten aus PANGAEA® versorgt. PANGAEA war maßgeblich an der Entwicklung der Zitierfähigkeit von Daten und am Aufbau von DataCite beteiligt. Seit 2009 werden Dienste bereitgestellt, welche die dynamische Cross-Referenzierung von Daten und Artikeln ermöglichen, u. a. aus Science Direct (z. B.: http://dx.doi.org/10.1016/j.biocon.2010.04.009).</p> <p>AWI: Aktuelle Anwendungen der Bioinformatik am AWI sind z. B. die Modellierung ökologischer Nischen für Diatomeen (Projekt der Hustedt-Diatomeen Sammlung), die Erfassung der zeitlichen Variation mikro-planktonischer Algengemeinschaften unter globalen Veränderungen, die Aufschlüsselung toxischer Pathways in Dinoflagellaten (Shellfish Poisoning, z. B. <i>Azadinium spinosum</i>) sowie die molekulare Charakterisierung von Lebensgemeinschaften in Meereis (Projekt MacSealce) und biotechnologische Anwendungen kalteangepasster in-situ ölabbauender Meeresbakterien. Desweiteren sind zu nennen: Teilnahme an Genomsequenzierungsprojekten zu Schlüsselorganismen, transkriptomische Untersuchungen der Adaptation / Acclimation in höheren Organismen sowie auf Ökosystem-Ebene.</p>
Vernetzung	<p>MPI-Bremen / Jacobs University: MPI/Jacobs ist mit seiner bioinformatischen Expertise an einer Reihe von aktuellen Forschungsprojekten beteiligt: BMBF-Projekt MIMAS (Microbial Interactions in MARine Systems), SAW-Leibniz-Projekt ATKIM (Abbaubarkeit von arktischem, terrigenem Kohlenstoff im Meer), EU-Projekte MAMBA (Marine Metagenomics for new Biotechnological Applications), EuroMarine (Integration of European Marine Research Networks of Excellence), BioVeL (Biodiversity Virtual e-Laboratory) und koordiniert das EU-Projekt Micro B3 (Biodiversity, Bioinformatics, Biotechnology). Darüber hinaus ist MPI/Jacobs an verschiedenen nationalen und europäischen Infrastrukturprojekten beteiligt: DFG-Projekt CIBAS (Center for integrative Biodiversity Analysis and Synthesis), EU-Projekte EuroFleets (Towards an alliance of European research fleets), EMBRC (European Marine Biological Resource Centers) und MIRRI (Microbial Resource Research Infrastructure). Desweiteren wurden enge Kontakte zum European Bioinformatics Institute (EBI) und speziell zum ELIXIR Projekt aufgebaut (European Life Sciences Infrastructure for Biological Information).</p>

Vernetzung	<p>AWI / MARUM: PANGAEA ist akkreditiertes Welt Datenzentrum innerhalb des ICSU World Data System (WDS) sowie des WMO Information System (WIS) und war in den letzten 15 Jahren in mehr als 140 nationalen, europäischen und internationalen Projekten aktiv eingebunden (aktuell IODP (NSF), EUR-OCEANS und EUROMARIN, ESONET (NoE), EMSO (CP), EPOCA (CP), CoralFish (CP), EUROBASIN (IP), HYPOX (CP), EMODNET Bio und Tara-Oceans, sowie im nationalen Rahmen BIOACID, INTERDY-NAMIK und SOPRAN (dt. SOLAS) – vollständige Liste www.pangaea.de/projects). Darüber hinaus unterhält PANGAEA eine breite Kooperation mit Wissenschaftsverlagen (Elsevier, Springer, Wiley, AGU u. a.).</p> <p>AWI: Das AWI hat in den letzten Jahren u. a. in Forschungsverbänden (z. B. Marine Genomics Europe Network of Excellence) und Genom-Sequenzier-Konsortien (Micromonas, Th. pseudonana, E. siliculosus, F. cylindrus, E. huxleyi, Ch. Crispus, Glaciecola), in Transkriptom-Sequenzierprojekten (Krill, P. brachycara, Hyas, S. latissima, Meereis-Metatranscriptom), aber auch in der Programmforschung (z. B. Küstenforschung, Harmful Algal Blooms, Ökologische Chemie) und Nachwuchsgruppen (PLANKTOSENS) ein weitreichendes Forschungs- und Anwendungsprofil in den modernen Omics-Methoden (heute vorwiegend Next Generation Sequencing und Microarrays) entwickelt und ausgebaut.</p>
Ausbildung	<p>Bachelor Studiengang Applied Computational Mathematics mit Spezialisierung Bioinformatic an der Jacobs University. Vorlesung und Praktikum Bioinformatik an der Universität Bremen und im Master Studiengang Marine Microbiology der International Max Planck Research School. Regelmäßige Bioinformatik Workshops und on-site Training von Nutzern, online Tutorials. Mitarbeit von Studenten an Projekten im Rahmen von internships, guided research Modulen und als studentische Hilfskräfte.</p>
Personal	<p>MPI / Jacobs: 9 Postdocs, 9 Doktoranden, 2 Masterstudenten, 2 Techniker, 1 Team Assistant, 1 Gruppenleiter</p> <p>AWI / MARUM: 5 Postdocs, 2 Techniker, 3 Data manager, 1 Gruppenleiter</p> <p>AWI Rechenzentrum / Bio / Bioinformatik: 4 Postdocs</p>
Software	<p>Diversität und Phylogenie (ARB/SILVA), Klassifikation, Binning (TETRA, TaxSOM, TaxoMeter), Standardisierung (MetaBar, CDinFusion), Annotation (JCoast), Datenintegration (Megx.net)</p> <p>AWI / MARUM: Für den Bau von Portalen und die Anbindung von Datenanbietern hat die PANGAEA® Gruppe in den letzten Jahren eine Open Source Software entwickelt (Schindler & Diepenbroek 2008), die aufgrund ihrer modularen Gestaltung beliebige Metadatenstandards (ISO19xxx, DIF, Dublin Core, Darwin Core etc.) unterstützt. Die Software wird für diverse Projekte eingesetzt (u. a. IODP, CARBOCHANGE, EPOCA, ESONET/EMSO, HYPOX, C3-GRID). PANGAEA setzt zudem die Datenwarehouse Software von Sybase (IQ) ein, die primär als Vorstufe zur Kompilation von Datenprodukten verwendet wird.</p> <p>AWI: Vergleichende Genomik (Phylogena), Mikroalgen-Communities (Pyloassigner), vergleichende Metagenomik (MGMC), Mikro-satelliten-Marker-Design (STAMP)</p>
Datenbanken	<p>MPI-Bremen / Jacobs University: SILVA: Die europäische Datenbank für ribosomale RNA Sequenzen (www.arb-silva.de)</p> <p>Die mikrobielle Biodiversitätsforschung beruht in erster Linie auf der Analyse von Markergenen. Die ribosomale RNA hat sich dabei als „Gold“-Standard etabliert und die Anzahl an öffentlich verfügbaren rDNA Sequenzen steigt seit Jahren mit einer Verdoppelungsrate von 12-18 Monaten exponentiell an (derzeit rund 2.7 Millionen Sequenzen, Stand Januar 2012). Für die Analyse dieser Datenflut sind spezialisierte Referenzdatenbanken und Softwaretools von entscheidender Bedeutung. Um diesem Rechnung zu tragen, wurde das ARB und SILVA Datenbankprojekt vor über 20 Jahren gegründet. ARB und SILVA sind international anerkannte Werkzeuge für die Bearbeitung, Kuratierung und Analyse von rDNA Sequenzen in der Biodiversitätsforschung, für die industrielle Qualitätskontrolle und medizinische Diagnostik.</p>

	<p>Megx.net (www.megx.net): Megx.net wurde 2005 als die erste integrierte Datenbank im Bereich der Umweltmikrobiologie entwickelt und erlaubt den konzertierten Zugriff auf mikrobielle Genominformationen und Biodiversität im Umweltkontext. Globale Umweltparameter werden dabei „on the fly“ aus ozeanographischen Datenquellen generiert. Die enge Vernetzung von Megx.net mit öffentlichen Sequenz- und Umweltdatenrepositorien wie EMBL-EBI/ENA und PANGAEA, in Kombination mit intuitiver Visualisierung der Ergebnisse, erlaubt den Nutzern einen dynamischen Blick auf Biodiversität und Funktion im Umweltkontext.</p> <p>AWI / MARUM: PANGAEA® – Publisher for Earth & Environmental Science (ICSU World Data Center) (www.pangaea.de)</p> <p>Die breitspektral über die gesamten Geo-, Bio- und Umweltwissenschaften verteilten Datenbestände des WDC-MARE dienen der Erforschung globaler Umweltveränderungen. Der Schwerpunkt liegt auf georeferenzierbaren Daten aus den Bereichen der Ozeanographie, Meeresgeologie, Paläoozeanographie und Meeresbiologie. Operationelle Plattform ist das Informationssystem PANGAEA. Das System beinhaltet gegenwärtig ca. 450.000 Datensätze mit mehr als 6,5 Milliarden Datenpunkten zu ca. 40.000 verschiedenen Parametern aus allen Weltmeeren und Kontinenten und finanziert sich mehrheitlich über Projektdatenmanagement und den Aufbau von Geodaten-Infrastrukturen.</p> <p>AWI: PLANKTONNET Biodiversitätsplattform, Hustedt Diatom Research Centre (Sammlungsdaten)</p>
Standardisierung und Ontologien	<p>MPI-Bremen / Jacobs University: Genomic Standards Consortium (www.gensc.org)</p> <p>Gegründet 2005 in Oxford hat sich das mit internationalen Wissenschaftlern besetzte Genomic Standards Consortium (GSC) zur Aufgabe gemacht, Richtlinien für eine möglichst kompakte, aber dennoch repräsentative Menge an wünschenswerten Zusatzdaten für Sequenzinformationen zu entwerfen. Daraus entstand zunächst der MIGS (Minimum Information about a Genome Sequence) – und MIMS (Minimum Information about a Metagenome Sequence) -Standard für Genom- und Metagenominformationen. Nach weiteren Entwicklungsjahren konnte das Konsortium kürzlich den MIMARKS (Minimum Information about a MArker gene Sequence) -Standard und die MixS-Spezifikationen (Minimum Information about any (x) Sequence) veröffentlichen. Ebenfalls unter dem Dach des GSC ist die Entwicklung von Ontologien z. B. für die Habitatklassifikation durch die Environment Ontology angesiedelt. MPI-Bremen / Jacobs Universität Bremen sind im GSC federführend und verwalten die zentrale Datenbank für die bereits etablierten Standards und Spezifikationen und eine GSC Referenzimplementierung in XML.</p>
Computerinfrastruktur	<p>MPI-Bremen: 500 Cores als Cluster mit 60Tb Storage (ausfallsicher, permanent überwacht), Webserver, Archivspeicher für Sequenzanalyse, Phylogenie, Annotation, Datenbanken und Services.</p> <p>AWI: 12 Knoten Vektorrechner NEC SX8R, 3,3 TFlop/s, 56 TByte GFS Dateisystem, Ozean/Meereis/Paleoklima-Modelle. 24 Core Dual-Opteron-Cluster, Genom-Annotation, Phylogenie und Transcriptomics. 1 SMP Knoten 16 Core Opteron, 32GB RAM, Assembly/Mapping 454-ILLUMINA Genomics/Transcriptomics, phylogenetic placement of 454 sequencing data, Large-scale niche-modeling, Metagenomic Monte-Carlo Markov-Chain Bayesian statistics. 1 SGI UV100-20 Blades, 160 cores Intel E7-883, 2,56 TB RAM, 96TB Dateisystem InfiniteStorage, Ozeanmodellierung, Datenassiliation, Transcriptomics-Annotation/-Mapping, Genome-Assembly, high-throughput phylogenetic placement of 454 sequencing data, Meta-genomic annotations. 2 PByte Archivspeicher SL8500 (LTO/3), Netapp skalierbare Speicher-Systeme.</p> <p>Nutzung der Services: International über Webpages und Webservices. Sun (Oracle) Secure Global Desktop (Webbasierend) und Sun-Ray Thin-Clients für verteiltes arbeiten an virtuellen Arbeitsplätzen. Kooperation mit Firmen über die in Bremen ansässige Ribocon GmbH (2005 aus dem MPI ausgegründet). Am AWI kommen zusätzlich Galaxy Workflows zum Einsatz.</p>

Glossar

BILS: Bioinformatics Infrastructure for Life Sciences. Dezentrale nationale Forschungsinfrastruktur für Bioinformatik in Schweden, die vom Schwedischen Forschungsrat unterstützt wird.

Biocuration: Beinhaltet die Übersetzung und Integration von biologischen Daten in eine Datenbank, die eine Verknüpfung der Daten mit wissenschaftlicher Literatur und weiteren Datensätzen ermöglicht.

Biodiversität: Konzept zur Beschreibung der Vielfalt des Lebens auf den drei Ebenen Ökosysteme, Arten und Gene. Als vierte Ebene gilt die Vielfalt der Wechselbeziehungen innerhalb und zwischen den anderen drei Ebenen, welche als funktionale Biodiversität bezeichnet wird.

Biokatalysatoren: Biokatalysatoren sind polymere Biomoleküle, die biochemische Reaktionen in Organismen beschleunigen, indem sie die Aktivierungsenergie der Reaktionen herab- oder (seltener) heraufsetzen.

Cloud-Computing: Cloud-Computing umschreibt den Ansatz, abstrahierte IT-Infrastrukturen (z.B. Rechenkapazität, Datenspeicher, Netzwerkkapazitäten oder auch fertige Software) dynamisch an den Bedarf angepasst über ein Netzwerk zur Verfügung zu stellen.

Cropsense: Netzwerk für komplexe Sensorik zur Nutzpflanzenforschung, Züchtung und Bestandessteuerung.

Datenintegration: Zusammenführung von Daten aus einer Vielzahl verschiedener Quellen.

ELIXIR: Die European Life-science Infrastructure for Biological Information ist eine paneuropäische Initiative zum Aufbau einer nachhaltigen europäischen Bioinformatik-Infrastruktur.

FUGATO: Vom Bundesministerium für Bildung und Forschung (BMBF) gefördertes Forschungsprogramm zur Funktionellen GenomAnalyse im Tierischen Organismus.

GABI/Pflanzenbiotechnologie der Zukunft: Vom Bundesministerium für Bildung und Forschung (BMBF) sowie durch privatwirtschaftliche Unternehmen gefördertes Forschungsprogramm auf dem Gebiet der zukunftsorientierten Pflanzenbiotechnologie.

Genom: Die Gesamtheit der genetischen Information eines Organismus.

Genomik: Forschungsgebiet, das sich mit Organismen auf der Ebene ihrer genomischen Daten auseinandersetzt.

GenoMik: Vom Bundesministerium für Bildung und Forschung (BMBF) im Jahr 2001 gestartete Forschungs- und Förderinitiative „Genomforschung an Mikroorganismen – GenoMik“ zur Schaffung der strukturellen und inhaltlichen Voraussetzungen für die Nutzung des Potentials von Mikroorganismen durch globale genombasierte Forschungsansätze.

Hochdurchsatz-Präzisionsphänotypisierung: Automatisierte Methode, bei der im Hochdurchsatz eine große Anzahl von Phänotypisierungen durchgeführt wird.

Knowledge Bases: Spezielle Datenbanken für das Wissensmanagement.

Metabolom: Die Gesamtheit der Metabolite eines Organismus.

Metabolomik: Forschungsgebiet, das sich mit Organismen auf der Ebene der Metabolite auseinandersetzt.

Metadaten: Daten, die Informationen über andere Daten enthalten.

Modellbasierte Datenauswertung: Statistische Datenauswertung unter Benutzung von Modellen, die dem jeweiligen Problem angepasst sind und versuchen, mögliche Mechanismen der zugrundeliegenden Prozesse zu erfassen.

NBIC: Netherlands Bioinformatics Centre. Niederländisches Bioinformatik-Netzwerk mit Expertise in den Bereichen Forschung, Lehre und Support.

Next Generation Sequencing: Neue Methoden der DNA-Sequenzierung, die einen erhöhten Durchsatz erlauben.

Omics-Technologien: Zusammenfassende Bezeichnung der Technologien, mit denen jeweils die Gesamtheit einer organismischen Systemebene analysiert wird, z. B. alle Gene (Genomics), alle Transkripte (Transcriptomics), alle Proteine (Proteomics) oder alle Metabolite (Metabolomics).

Ontologie: Sprachlich gefasste und formal geordnete Darstellungen einer Menge von Begrifflichkeiten und der zwischen ihnen bestehenden Beziehungen in einem bestimmten Gegenstandsbereich. Sie werden dazu genutzt, „Wissen“ in digitalisierter und formaler Form zwischen Anwendungsprogrammen und Diensten auszutauschen.

PathoGenoMik: Förderrichtlinie des Bundesministeriums für Bildung und Forschung zur Förderung transnationaler Forschungsprojekte innerhalb des ERA-NET PathoGenoMics „Transnationale Pathogenomik: Prävention, Diagnose, Behandlung und Monitoring von humanen Infektionskrankheiten“ im Rahmenprogramm „Biotechnologie – Chancen nutzen und gestalten“.

Phänomics: Vom Bundesministerium für Bildung und Forschung (BMBF) gefördertes Kompetenznetz zur Agrar- und Ernährungsforschung. Es vertritt einen systembiologischen Ansatz zur Genotyp-Phänotyp-Abbildung bei den Nutztieren Rind und Schwein.

Phänotypisierung: Quantitative Analyse von Schlüsselfunktionen und -strukturen von Organismen und biologischen Systemen und der zugrunde liegenden physiologischen, molekularen und genetischen Mechanismen.

Postgenomdaten: Biologische Daten, die das zelluläre Geschehen in seiner Gesamtheit analysieren und damit über die rein genetische Ebene der Datenerfassung hinausgehen.

Primärdaten: Sequenzdaten von DNA-, RNA- und Proteinmolekülen.

Proteom: Die Gesamtheit aller Proteine, die zu einem bestimmten Zeitpunkt in einem Organismus vorkommen.

Proteomik: Forschungsgebiet, das sich mit Organismen auf der Ebene ihrer Proteine auseinandersetzt.

Rechnercluster: Eine Anzahl von vernetzten Computern. Ziel des „Clustering“ besteht meistens in der Erhöhung der Rechenkapazität oder der Verfügbarkeit gegenüber einem einzelnen Computer.

ReNaBi: Réseau National des plates-formes Bioinformatiques. Französische Bioinformatik-Netzwerkstruktur.

SIB: Swiss Institute of Bioinformatics. Föderation von Bioinformatik-Forschungsgruppen führender Schweizer Universitäten und des Swiss Federal Institutes of Technology.

Standard Operating Procedures: Arbeitsanweisung, die das Vorgehen innerhalb eines Prozesses beschreibt.

Supercomputer: Die jeweils schnellsten Rechner ihrer Zeit. Typisches Merkmal eines modernen Supercomputers ist seine große Anzahl an Prozessoren, die auf gemeinsame Peripheriegeräte und einen teilweise gemeinsamen Hauptspeicher zugreifen können. Supercomputer werden häufig für Computersimulationen im Bereich des Hochleistungsrechnens eingesetzt.

Synbreed: Durch das Bundesministerium für Bildung und Forschung gefördertes Kompetenznetz zur Gründung eines interdisziplinären Zentrums für die genombasierte Züchtungsforschung bei Nutzpflanzen und Nutztieren. Verbund von Wissenschaftlern aus Pflanzen- und Tierzüchtung, Molekularbiologie, Bioinformatik und Humanmedizin unter Beteiligung universitärer, außeruniversitärer und industrieller Kooperationspartner.

Synthetische Biologie: Fachgebiet im Grenzbereich von Molekularbiologie, Organischer Chemie, Ingenieurwissenschaften, Nanobiotechnologie und Informationstechnik mit dem Ziel, biologische Systeme und Mikroorganismen mit Hilfe standardisierter Bausteine zu konstruieren.

Systembiologie: Biowissenschaft, die sich zum Ziel setzt, die komplexen und dynamisch ablaufenden biologischen Prozesse von Zellen oder Organismen in ihrer Gesamtheit zu verstehen.

Transkriptom: Die Gesamtheit der Transkripte, die zu einem bestimmten Zeitpunkt in einem Organismus vorkommen.

Transkriptomik: Forschungsgebiet, das sich mit Organismen auf der Ebene ihrer Transkripte auseinandersetzt.

Mitglieder des Steering Committees „Workshop Bioinformatik“

Prof. Dr. Frank Oliver Glöckner

MPI für Marine Mikrobiologie / Jacobs Universität Bremen

Dr. Alexander Goesmann

Centrum für Biotechnologie / Universität Bielefeld

Dr. Thomas Hartsch

GeneData AG

Dr. Eric von Lieres

Forschungszentrum Jülich

Dr. Klaus Mayer

Munich Information Centre for Protein Sequences (MIPS) /
Helmholtz Zentrum München

Prof. Dr. Alfred Pühler (Vorsitzender)

Centrum für Biotechnologie, Universität Bielefeld

Prof. Dr. Norbert Reinsch

FBN Dummerstorf

Prof. Dr. Chris-Carolin Schön

Technische Universität München

Prof. Dr. Wolfgang Wiechert

Forschungszentrum Jülich

Prof. Dr. Ralf Zimmer

Ludwig-Maximilians-Universität München

Mitglieder des Forschungs- und Technologierats Bioökonomie

Prof. Dr. Dr. h.c. Reinhard F. Hüttl

(Vorsitzender)

Wissenschaftlicher Vorstand des Helmholtz-Zentrums Potsdam Deutsches Geoforschungszentrum, Präsident acatech, Professor für Bodenschutz und Rekultivierung an der Brandenburgischen Technischen Universität Cottbus

Dr. Dr. h.c. mult. Andreas J. Büchting

(stellv. Vorsitzender) Vorsitzender des Aufsichtsrats KWS SAAT AG

Prof. Dr. Bernd Müller-Röber

(stellv. Vorsitzender)

Professor für Molekularbiologie, Max-Planck-Institut für Molekulare Pflanzenphysiologie und Universität Potsdam

Prof. Dr. Dr. h.c. Joachim von Braun

(stellv. Vorsitzender)

Direktor am Zentrum für Entwicklungsforschung (ZEF), Universität Bonn

Prof. Dr. Achim Bachem

Vorstandsvorsitzender des Forschungszentrums Jülich GmbH

Dr. Helmut Born

Generalsekretär des Deutschen Bauernverbands e.V.

Prof. Dr. Hannelore Daniel

Technische Universität München, Lehrstuhl für Ernährungsphysiologie

Prof. Dr. Utz-Hellmuth Felcht

Managing Director, One Equity Partners Europe, München, Mitglied des acatech Präsidiums

Prof. Dr. Thomas Hirth

Leiter des Fraunhofer-Instituts für Grenzflächen- und Bioverfahrenstechnik und des Instituts für Grenzflächenverfahrenstechnik, Universität Stuttgart

Prof. Dr. Folkhard Isermeyer

Präsident des Johann Heinrich von Thünen-Instituts Braunschweig, Bundesforschungsinstitut für Ländliche Räume, Wald und Fischerei

Dr. Stefan Marcinowski

Mitglied des Vorstands BASF SE
Vorsitzender des Vorstands der Deutschen Industrievereinigung Biotechnologie (DIB)

Prof. Dr. Dr. h.c. Thomas C. Mettenleiter

Präsident des Friedrich-Loeffler-Instituts Insel Riems, Bundesforschungsinstitut für Tiergesundheit

Dr. Dr. h.c. Christian Patermann

Berater für die wissensbasierte Bioökonomie des Landes Nordrhein-Westfalen

Prof. Dr. Alfred Pühler

Centrum für Biotechnologie, Universität Bielefeld

Prof. Dr. Manfred Schwerin

Professor für Tierzucht, Universität Rostock und Vorstand des Leibniz-Instituts für Nutztierbiologie, Dummerstorf

Prof. Dr. Wiltrud Treffenfeldt

Direktorin für Forschung und Entwicklung Europa, Mittlerer Osten und Afrika, Dow Europe, Horgen, Schweiz

Prof. Dr. Fritz Vahrenholt

Vorsitzender der Geschäftsführung RWE Innogy GmbH

Dr. Holger Zinke

Vorstandsvorsitzender BRAIN AG

Prof. Dr. Alexander Zehnder

(ständiger Gast)

Direktor des Water Research Institute der Universität Alberta, Edmonton, Kanada

Der BioÖkonomieRat dankt dem Bundesministerium für Bildung und Forschung als Mittelgeber sowie acatech – Deutsche Akademie der Technikwissenschaften für die administrative Begleitung.

Besonderer Dank gilt den externen Gutachtern für ihre wertvollen Hinweise zu dem vorliegenden Papier. Die Inhalte der Empfehlungen liegen in der alleinigen Verantwortung des BioÖkonomieRats.

Die Arbeit des BioÖkonomieRats wurde durch eine Geschäftsstelle unterstützt:

Dr. Claus Gerhard Bannick (Leiter Geschäftsstelle)

Dr. Andrea George (wiss. Mitarbeiterin)

Dr. Katja Leicht (wiss. Mitarbeiterin)

Petra Ortiz Arrebato (Assistenz)

Ulrike von Schlippenbach (wiss. Mitarbeiterin)

Dr. Elke Witt (wiss. Mitarbeiterin)

Dr. Eva Wendt (wiss. Mitarbeiterin)

Julian Braun, Martin Schmidt (studentische Mitarbeiter)

IMPRESSUM

Herausgeber

Forschungs- und Technologierat Bioökonomie (BÖR)

© BÖR, Berlin (2012)

Konzept und Gestaltung

Oswald + Martin Werbeagentur, Berlin

Druck

Brandenburgische Universitätsdruckerei

ISSN 1869-1404, ISBN 978-3-942044-66-0, (Druckausgabe), ISBN 978-3-942044-67-7 (Online-Version)

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Nationalbibliografie.

Detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.



Herausgeber
Forschungs- und Technologierat Bioökonomie (BÖR)
© BÖR, Berlin (2012)

Kontakt
Geschäftsstelle des BioÖkonomieRats
Charlottenstraße 35–36
10117 Berlin
Tel.: 030 767718911
Fax: 030 767718912
E-Mail: info@biooekonomierat.de
Internet: www.biooekonomierat.de