

Blick zurück in die Zukunft der Bioinformatik

VON PEER BORK, HEIDELBERG

■ Vor 20 Jahren wurden Bioinformatiker belächelt oder als Datenpiraten verunglimpft. Heute kommt die Biologie nicht mehr ohne sie aus.

Nach dem Aufwachen das Lesen der frühmorgendlichen 50 E-mails *en bloc* – im

Folgenden die relevanten Nachrichten im Schnelldurchlauf: Ein Manuskript, das nun schon fast ein Jahr im Begutachtungsprozess ist, wurde abgelehnt, weil Statistik alleine nicht gut genug ist und mehr Experimente ran müssen; irgend jemand hat im institutsweiten Computercluster Chaos angerichtet und ein Petabyte Speicher ist voll – womit die Daten aus dem Sequenzierer nicht mehr automatisch in unsere Rechner aufgenommen werden können.

Mein (alter) USB-Stick speichert 1 Gigabyte und da gehen alle meine Präsentationen des Jahres drauf, warum braucht man also eine Million-mal mehr Speicher dieser Tage, um arbeiten zu können? Als Bioinformatiker hatte man es schon mal einfacher...

Auf dem Weg zur Arbeit dann das Erinnern an die „Gute alte Zeit“: Ich bin 1987 als Diplomand in die Bioinformatik eingestiegen; der Begriff trat gerade seinen Siegeszug an und vereinte alle, die mit dem Computer die Biologie verändern wollten. Das waren meist Leute, die Proteinstrukturen analysierten oder in öffentlichen Sequenzdatenbanken auf Entdeckungsreise gingen; teilweise aber auch Theoretiker und Modellierer, die den Begriff erst mit der Zeit langsam annahmen.

Mit den exponentiell wachsenden Daten in öffentlichen Datenbanken (eine riesige Errungenschaft in der Biologie und zusammen mit dem Internet Innovationsmotor) war es damals schon wie kurz nach dem Urknall: Die Bioinformatik expandierte rasant, wurde zu etwas Be-

deutendem ... und zerfällt vielleicht bald in viele Teildisziplinen. Schon heute hat man massenhaft Galaxien mit eigenen Gesetzen und nur wenige schauen über die eigene hinaus. Zu Sequenz- und Strukturdaten gesellten sich in den 90er Jahren viele andere quantifizierbare Datentypen: Die „-omics“ waren geboren.

Microarrays klingen zwar heute schon wie eine veraltete Technologie, aber damit ging die Transkriptomik los – zusammen mit der Genomsequenzierung waren sie der Vorbote von Automatisierung in der Biologie und einhergehender Computerisie-

„Die Bioinformatik ist teilweise nur noch ein Sammelbegriff unter dem sich verschiedenste Teildisziplinen gruppieren...“

rung sowie systembiologischen Ansätzen. Inzwischen ist nicht nur die Molekularbiologie infiltriert. Mit der Proteomik geht es in Richtung Biochemie und auch die Zellbiologie kommt ohne computergesteuerte Bildanalyse nicht mehr aus und entwickelt eine regelrechte Filmindustrie.

Auch Chemiker und Mediziner, die Daten meistens sorgsam unter Verschluss hielten, erkennen mittlerweile die Vorteile des freien Datenaustausches. Pharmafirmen lassen Daten klinischer Studien extern von Bioinformatikern analysieren, die auch federführend bei der Auswertung elektronischer Patientenkarteen sind (in Ländern, in denen dies technisch und gesetzlich möglich ist).

Die Bioinformatik ist teilweise nur noch ein Sammelbegriff unter dem sich verschiedenste Teildisziplinen gruppieren und nicht alle, die mit dem Computer in der Biologie oder angrenzenden Gebieten arbeiten, würden sich mit dem Begriff identifizieren. Und hier reden wir nicht nur von den Computerbiologen, die biologische Fragestellungen bearbeiten und Bioinformatik eher als Methoden-, Ressourcen- und Infrastrukturentwicklung sehen. Auch jemand, der 90 Prozent mit

Laborjournal-Titelseiten 1994-2014



Laborjournal 10/2012 thematisierte die Problematik von Mensch-Tier-Mischwesen. Moderne biologische Methoden verwischen die Grenzen zwischen Mensch und Tier.

Computeranalyse beschäftigt ist und für die Sequenzierung eher Serviceeinrichtungen heranzieht, sieht sich eher als Genomiker zu dessen Profil eben auch Bioinformatik gehört.

Mit dem Aufschwung und der steigenden Notwendigkeit der Datenaufarbeitung änderte sich auch die Einstellung der Lebenswissenschaftler zu den Bioinformatikern. Ende der 80er Jahre noch belächelt, überwog in den 90er Jahren Skepsis, Neid (Versinnbildlichung von Bioinformatikern als Datenpiraten, die von der Arbeit anderer leben und publizieren) und Furcht vor dem Neuen. Die Versuche, vielerorts notgedrungen Bioinformatik als Service zu etablieren, gingen meistens nicht auf und so wurde das notwendige Übel akzeptiert. Ende der 90er Jahre gab es, stimuliert

durch viele Förderprogramme, einen Boom von Professuren und in diesem Millennium wird Bioinformatik als ein essentieller Bestandteil biologischer Forschung angesehen und hält in fast jedem Labor Einzug.

Wenn man den Bioinformatiker so definiert, dass er mehr als 50 Prozent seiner Zeit am Computer verbringt, ohne Manuskriptarbeit oder Kommunikation (E-mail, Skype etc.), sind bei uns am EMBL laut internen Umfragen inzwischen mehr als 40% aller Wissenschaftler einzurechnen – Tendenz steigend. Damit ist die Bioinformatik als Kategorie in der Zitations-Statistik für das *Laborjournal* nicht mehr geeignet: Zu viele würden sich beschweren, nicht berücksichtigt worden zu sein. Der Aufschwung des eigenen Wissenschaftsgebietes ist für einen ehemals „Belächelten“ zwar eine Genugtuung, macht das wissenschaftliche Leben aber auch viel komplexer und komplizierter.

In den „Goldenen“-90ern konnte man als Bioinformatiker (zumindest in der Sequenzanalytik) in wenigen Tagen durch eine Datenbankrecherche ein biologisch durchaus relevantes Ergebnis erzielen (zum Beispiel eine Domäne entdecken und funktionell charakterisieren) und ein leicht zu strukturierendes Manuskript entwerfen; mein Rekord war vom Start der Suche bis zum Manuskripteinreichen drei Tage (in der Zeit, als vor 20 Jahren das erste Mal das *Laborjournal* am EMBL auftauchte).

Diese Effizienz hatte auch ihre Kehrseiten, denn ich wurde auch schon von Mitgliedern einer Berufungskommission als Scharlatan bezeichnet, weil ich im Rah-

men der Habilitation einen dreiwöchigen praktischen Kurs mit vier Studenten in eine Publikation münden ließ: Dies wäre keine richtige Wissenschaft, wo man sich normalerweise Jahre abmühen muss, bis man zu richtigen Ergebnissen kommt und vermittelt Studenten ein falsches Bild. Im allgemeinen war man aber damals ein gesuchter Spezialist und konnte so ziemlich alle Anfragen bedienen (wenn sie denn interessant waren) und wegen der noch nicht ausufernden Datenbankgrößen waren

Rechenleistung und Speicherplatz noch nicht limitierend.

Während man früher in einer „datenkargen“ Biologie hauptsächlich mit der Auswertung eines experimentellen Ergebnisses beschäftigt war, werden heute „Big Data“ mehr und mehr auch zur Gene-

rierung von Hypothesen herangezogen, die mit Experimenten validiert werden müssen. Die Bioinformatik ist also im Erkenntnisgewinnungsprozess deutlicher involviert. Die Iteration mit den Experimentatoren, wie in der Systembiologie üblich, ist zwar spannend, aber Ergebnisse ziehen sich halt hin. Die Erarbeitung des Studiendesigns mit verschiedensten experimentellen Komponenten impliziert Wichtigkeit, bringt aber auch mehr Verantwortung. Ein Computereperiment in den Sand zu setzen oder Programmierfehlern aufzusitzen ist zwar unschön, hat aber weder zeitlich noch finanziell die Auswirkungen eines fehlgeschlagenen komplexen Experimentes.

Die meisten unserer jungen Bioinformatiker am Institut sind in experimentellen Gruppen und müssen sich selbst um ihren Bedarf kümmern, da die entsprechenden Gruppenleiter andere Stärken als Computerwissen haben. Heute funktionieren zwar viele Analysen Web-basiert auf dem PC. Oft werden externe Datenbanken und Programme aber lokal gespeichert (häufig mehrmals) und die zentralen Rechner mit nicht immer sinnvollen und effektiven „Jobs“ gefüllt.

Das Arbeitsprofil eines Bioinformatikers ist breiter gefächert und geht mit ständiger Weiterbildung einher. Noch vor wenigen Jahren konnte man sich auf eine Omics-Methode und ein entsprechendes Teilgebiet spezialisieren, und nur wenige

konnten „skripten“ und analysieren. Heute ist beides gefordert und man muss methodisch und in der Biologie fit sein – auch gut informiert – um die verschiedensten existierenden Programme wohlausgewählt zu eigenen „Pipelines“ zusammen zu stöpseln. Diese Vielfalt bedeutet aber auch fast tägliche Weiterbildung und nicht immer hilft da das Internet...

Mit dem ersten Kaffee im Büro (den gab's auch schon vor 20 Jahren) ist dies auch das Thema des ersten Vormittagsmeetings: Wie kann man 200 Bioinformatiker in allen Ecken des Hauses überhaupt erfassen, zusammenführen und ihnen Informationen zukommen lassen, sowie Redundanz vermeiden?

Wir haben bei uns das Projekt Bio-IT ins Leben gerufen, das dem Rechnung tragen soll, aber auch einer komplexen Logistik bedarf: Doktoranden und Postdocs werden bei uns schon im „Laufzettel“ am ersten Arbeitstag auf die Bioinformatik aufmerksam gemacht und erhalten eine entsprechende persönliche Einleitung. Diverse interne Kurse, mehrere virtuelle Computerbiologiezentren mit Servicesprechstunden in bestimmten Fokusgebieten, Erfahrungsaustauschveranstaltungen (teilweise mit Pizza-Anreiz) und eine aufwendige interne Webseite, die von CPU-Beschaffenheiten, zentralen Programmen und Datenbanken, bis hin zu Tipps alles anbieten muss, gehören zum BIO-IT Paket.

Nicht zu vergessen eine Arbeitsgruppe mit Mitgliedern aus allen wissenschaftlichen Disziplinen, damit auch keiner zu kurz kommt. All dies mit der Vorgabe, nicht zu viel „Top-down“ zu organisieren und den Bedarf richtig einzuschätzen. Als ich anfang, gab es 20 Hanseln im Institut, die den Computer als Hauptarbeitsmittel in der Biologie sahen, und wir haben nach

dem Mittag Tischtennis gespielt. Auch das war „Community building“ und klar „Bottom-up“; die Redundanz war minimal wie die Daten, mit denen wir

umgingen, aber ab einer bestimmten Institutsgröße geht das nicht mehr.

Die Mitteilung vom Handy (automatisch durch das institutsweite Kalendersystem initiiert) erinnert, dass als nächstes ein Telefoninterview mit einem Journalisten verabredet ist (Öffentlichkeitsarbeit wird ja immer wichtiger und die entsprechende Abteilung hat schon mehr als zehn Leute – vor 20 Jahren gab es da noch niemanden). Eigentlich ist es ja schön, wenn sich Leute für unsere Arbeit interessieren und der Be-

„Mit dem Aufschwung und der steigenden Notwendigkeit der Datenaufarbeitung änderte sich auch die Einstellung der Lebenswissenschaftler zu den Bioinformatikern.“

„Damit ist die Bioinformatik als Kategorie in der Zitations-Statistik für das Laborjournal nicht mehr geeignet.“

griff Bioinformatiker auch im Fernsehen zu bewundern ist. Aber da fast jede Woche irgendein Termin dieser Art ansteht, summiert sich das zu einer Menge Zeit, die von der eigentlichen Arbeit, der Bioinformatik abgeht...

Nach dutzenden Unterschriften, kleineren Entscheidungen und einigen kurzen Eins-zu-Eins Gesprächen dann am späten Vormittag der Höhepunkt des Tages: Eine richtige wissenschaftliche Besprechung, ein echtes „Brainstorming“ über ein Projekt, das Kreativität verlangt. Wegen der gestiegenen Komplexität der Daten ist es einfacher, in kleinen Teams zu arbeiten, in denen die Mitglieder unterschiedliche Ausbildungsprofile haben. Da ist das Zusammenstecken der Köpfe meist fruchtbar. Allerdings geht es auch hier meistens um das clevere „Putzen“ von Rohdaten, die vergleichbar gemacht werden müssen, um die Optimierung von Programmen, die in vertretbarer Zeit eine Analyse ermöglichen

sowie um die unabhängige Bestätigung von Ergebnissen. Da keiner die Wahrheit kennt, sind Programmfehler schwer zu finden und gute Tests essentiell.

Ein wichtiger Punkt ist jedes Mal auch die Interpretation der Ergebnisse. Bei vergleichender Analyse freut man sich über Korrelationen, oft sind diese aber nur indirekt. So ist es schon vorgekommen, dass die beobachtete Veränderung der Darmflora in Patienten nichts mit der eigentlichen Krankheit zu tun hat, auf die man es abgesehen hat, sondern mit einem Arzneimittel, das viele Patienten nehmen müssen, oder mit geänderten Ess- und Lebensgewohnheiten nach einer Diagnose. Zumindest macht es aber Spaß,

an wissenschaftlichen Fragestellungen zu tüfteln, dafür wurde man ja lange ausgebildet – keiner hat mich je gelehrt, wie man bei steigendem Aufkommen organisatorischer Anforderungen genügend Zeit für die Wissenschaft behält...

Nach dem Mittag lassen die Organisationstätigkeiten im nächsten E-mail-Block auch sofort grüßen: „Time sheet“-Erinnerung der Personalabteilung, Klärungsbedarf bei einem Computerkauf durch die Finanzabteilung, eine Anfrageliste der lokalen Ethikkommission hinsichtlich eines Experimentes, eine Nachfrage vom „Grant Office“ wegen zugewiesener Mittel, und unsere Rechtsabteilung hat einen 12-Seiter verfasst, um die Zusammenarbeit mit einer Universität im Ausland eindeutig zu regeln.

Wobei hier unsere Technologietransfer GmbH nicht ganz einverstanden ist, da einige Klauseln nach Service aussehen und somit über sie abgehandelt werden müssen – laut interner Regel Nummer 1342 [Nummer von der Redaktion geändert].

Klar, ich verdiene mehr als meine Postdocs und somit verdiene ich auch diese Art von E-mails, aber auch die Postdocs kriegen heute einiges von der gestiegenen Professionalität ab: sie müssen viele interne Berichte verfassen, zu angeordneten Treffen im Rahmen der Sachbeihilfen fahren und dürfen stundenweise elektronisch ihre Arbeitszeit

dokumentieren (hier lässt die EU-Administration grüßen), das zählt sicher nicht zu Bioinformatik und passt auch nicht in das Klischee der wissenschaftlichen Freiheiten. Vor 20 Jahren musste ich nach Abschluss der Arbeit für eine Sach-

beihilfe auch meinen Bericht einschicken und war mir nie sicher, ob den jemals jemand lesen wird, aber das war nach zwei oder drei Jahren ungestörter Forschung...

Bei aller Professionalität ist am frühen Nachmittag auch mal wieder meine wissenschaftliche Erfahrung gefragt: Eine Publikation wird vorbereitet und da müssen Prioritäten gesetzt und Abbildungen verbessert werden. Da kann ich jetzt auch mal meckern. Zum Beispiel darüber, dass die meisten Biologen mit Boxplots immer noch auf dem Kriegsfuß stehen und, dass man für den Leser vereinfachen muss. Hier hat sich, so zumindest mein Eindruck, noch nicht viel geändert, abgesehen davon, dass eine Software das Erstellen der Abbildungen vereinfacht – wenn man sie bedienen kann. Ich brauchte damals als Postdoc auch Rat für die Präsentation und habe ihn auch bekommen...

Jetzt heißt es, unbemerkt nach Hause zu verschwinden, da ich noch zwei Stunden in Ruhe ein Manuskript über Genomannotation editieren und an einem Grant zur Diagnose von Magenkrebs basteln muss. Der E-mail-Block am späten Nachmittag spricht dann aber wieder eine andere Sprache: Dreimal „Dear Sir“, das heißt Postdoc-Anfragen aus Indien, zweimal die Bitte um ein Gutachten für eine Zeitschrift und einmal für eine Sachbeihilfe. Des Weiteren mehrere Einladungen zu kommerziell organisierten Konferenzen und Angebote von allen möglichen Firmen, beim Projekt- oder Grantmanagement oder bei der Bioinformatik selber behilflich zu sein.

Es wird also nicht nur in den Instituten selbst mehr Wissenschaft organisiert (bei einer gleichbleibenden Anzahl von Wissenschaftlern). Um die Wissenschaft herum floriert auch eine immer größere Wirtschaft. Vor 20 Jahren waren da nur

„Da kann ich jetzt auch mal meckern. Zum Beispiel darüber, dass die meisten Biologen mit Boxplots immer noch auf dem Kriegsfuß stehen und dass man für den Leser vereinfachen muss.“

Laborjournal-Titelseiten 1994-2014



In Laborjournal 3/2013 lautete das Titelthema: Kontaminierte Zelllinien. Zelllinien, die durch andere kreuzkontaminiert sind, bremsen den wissenschaftlichen Fortschritt.

die Wissenschaftsverlage, deren teure Fachzeitschriften unter anderem durch die gegenseitige Begutachtung ohne Bezahlung sehr profitabel waren. Seiteneinsteiger wie das kostenfreie *Laborjournal* waren damals hochwillkommen...

Abends geht es dann zum Flughafen, an der Reisetätigkeit hat sich, abgesehen davon, dass man seinen Vortrag jetzt im Flieger zusammenstellt, nicht so viel geändert. Dass es mehr wird, liegt am Jobprofil, als Postdoc kam ich auf zehn Reisen pro Jahr, aber auch vor 20 Jahren hatte ich als Gruppenleiter schon ca. 20, nur gingen die fast alle zu wissenschaftlichen Konferenzen. Heute ist der CO₂-Ausstoß mit dem Flugzeug nicht viel höher, aber oft ist eine Reise gleich mit mehreren Veranstaltungen

verknüpft. Konferenzen werden dabei zunehmend von Begutachtungen, Projekt- oder Managementtreffen verdrängt.

Da stellt sich unwillkürlich die Frage, wohin die Reise geht...mit der Bioinformatik. Wird die Biologie so quantitativ wie die Physik? Wird es das Jobprofil Bioinformatiker noch geben, da jeder Biologe Bioinformatik als Handwerk parat haben

„Eigentlich ist Bioinformatik heute viel spannender, vielfältiger – auch erfolgsträchtiger – als vor 20 Jahren und kann Disziplinübergreifend eingesetzt werden.“

muss, so wie heute schon Textverarbeitungs- oder Präsentationsprogramme? Oder wird der Bioinformatiker gefragter sein denn je, weil die Datengenerierung billiger wird, die Analyse aber immer anspruchsvoller? Beides spricht dafür,

dass der Erwerb von Bioinformatik-Kenntnissen eine gute Investition ist. Aber wie wird Bioinformatik im Alltag aussehen? Werden alle experimentelle Daten in der

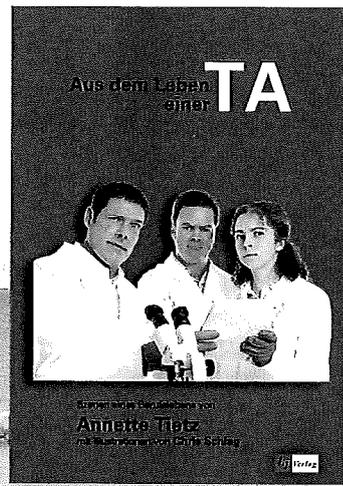
„Cloud“ verwaltet, vernetzt und bioinformatisch faktisch anonym analysiert? Wird ein besseres Datenverständnis zusammen mit „Big Data“ zu einer Kultur führen, in der auch Fehlschläge publiziert werden, um bessere Statistiken zu bekommen (da krankt es noch gewaltig in der Biologie und in anderen Naturwissenschaften)?

Eigentlich ist Bioinformatik heute viel spannender, vielfältiger – auch erfolgsträchtiger – als vor 20 Jahren, und kann Disziplinübergreifend eingesetzt werden. Im Vergleich mit der „Guten alten Zeit“ erreicht ein Bioinformatiker heute pro Zeiteinheit viel mehr und wichtigere Ergebnisse, trotz aller kreativitäts- und effizienzhemmender „Professionalität“ im Arbeitsalltag.

Wäre toll, wenn ich heute nochmal als Doktorand einsteigen könnte...

Peer Bork ist Leiter der „Structural and Computational Biology Unit“ am Europäischen Laboratorium für Molekularbiologie (EMBL) in Heidelberg.

Suchen Sie eine richtig gute Urlaubslektüre?



Nur bei uns!

Ein humoriger Blick auf die wirklichen Probleme dieser Welt: defekte Kaffeemaschinen, unverständliche Vorträge, miesgelaunte Chefs, oder noch schlimmer: gutgelaunte Chefs. Die führen garantiert etwas im Schilde.“

Annette Tietz: „Aus dem Leben einer TA“. 210 Seiten, Softcover, erschienen 2012. Preis: 12,80 € (inkl. MwSt. und Versand)

Bestellmöglichkeiten:

- <http://www.laborjournal.de/rubric/shop/shop.lasso>
- per Email an versand@laborjournal.de (bitte mit vollständiger Lieferadresse)

Für alle, die im Urlaub nicht ganz auf das Labor verzichten wollen