

# Technologierevolution in der Genomforschung

Von Thomas Hankeln, Hans Zischler und Erwin R. Schmidt

**Neue revolutionäre Sequenzierungstechnologien versprechen für Medizin und Biologie einzigartige Einsichten in das Erbmateriale von Lebewesen. In wenigen Jahren werden wir alle die DNA-Sequenz unseres eigenen individuellen Genoms bei moderaten Kosten entschlüsseln können. Die riesigen Datenmengen schaffen aber auch Probleme für die bioinformatische Verarbeitung und Interpretation. Das Nukleinsäureanalytik-Zentrum der Universität Mainz plant im Verbund mit dem Forschungsschwerpunkt „Rechnergestützte Forschungsmethoden in den Naturwissenschaften“ die zentrale Etablierung der neuen Verfahren.**

Manche Revolutionen dauern etwas länger. Immerhin 32 Jahre ist es her, dass Fred Sanger und Alan Coulson ein praktikables Verfahren zur Sequenzierung von DNA vorstellten. Bis heute hat die Sanger-Methodik, eine im Reagenzglas nachgeahmte DNA-Replikation, überlebt: Mit ihr wurden noch unter einem Milliarden-Dollar-Einsatz zwischen 1998 und 2003 die ersten Versionen des drei Milliarden Nukleotide umfassenden Humangenoms erstellt. Zwar wird die Sanger-Methode für kleine Projekte nicht aussterben, doch sehr viel schnellere und extrem kostengünstigere Sequenzierungsmethoden („Next-Generation-Sequencing“, NGS) revolutionieren derzeit die Genomforschung.

## NGS: Entdecke die Möglichkeiten!

Mit dem beabsichtigten Publicity-Effekt konnte per NGS das Genom des DNA-Helix-Entdeckers und Nobelpreisträgers James D. Watson für etwa 350.000 Dollar entschlüsselt werden. Die ebay-Auktion einer Humangenom-Sequenzierung im Mai 2009 zum Einstiegspreis von 68.000 US-Dollar war noch ein Werbegag der „Personal Genomics“-Firma KNOME, zeigt aber die Richtung an. Ziel der Humangenetik ist es, die vielen kleinen und großräumigen Unterschiede unseres Erbmaterials zu identifizieren. Es sind diese kleinen Unterschiede (mit einer Frequenz von etwa einem aus tausend Bausteinen), die uns als Lebewesen individuell machen. Seit Januar 2008 läuft das „1000 genomes project“. Hierbei werden durch Sequenzierung der Genome von Asiaten, Afrikanern und Europäern alle weltweit biomedizinisch relevanten Genunterschiede aufgespürt. Dies wird die Grundlage für eine individualisierte Medizin darstellen, bei der Medikamente auf den Genotyp abgestimmt verabreicht werden, um ihre Wirksamkeit zu verbessern

und Nebenwirkungen auszuschließen. Auch wird der Katalog der Genvarianten die Identifizierung solcher Gene beschleunigen, die für komplexe genetische Erkrankungen (Diabetes, Bluthochdruck, Krebs etc.) verantwortlich sind. So konnte kürzlich ein komplettes Genom von Blutkrebszellen sequenziert werden: Der Vergleich mit dem Genom aus gesunden Zellen des Patienten zeigte acht tumorspezifische Mutationen in Genen, die zuvor nie als krebisrelevant aufgefallen waren.

Auch die anderen Lebenswissenschaften sind elektrisiert ob der neuen Möglichkeiten: Molekulare Systematiker sequenzieren Genome, um aufgrund von Sequenzen Stammbäume von Tieren, Pflanzen und Bakterien zu erstellen. Dabei entdecken sie, dass klassische Gruppierungen aufgrund der molekularen Daten völlig neu sortiert und Lehrbücher umgeschrieben werden müssen. Dies hat erheblichen Einfluss darauf, wie wir die Evolution unseres eigenen Genoms betrachten. Schon werfen Evolutionsbiologen einen direkten Blick in die Vergangenheit unserer Spezies: Forscher haben per NGS eine erste Version des Genoms des Neandertalers rekonstruiert. Die Auswertung läuft noch: Wird es sich bestätigen, dass sich der Neandertaler vor etwa 600.000 Jahren von unserer eigenen Evolutionslinie abgetrennt hat und bis zu seinem Aussterben vor 25.000 Jahren als „Parallelgesellschaft“ ohne Genaustausch mit Homo sapiens existiert hat?

Auch Ökologen haben Spannendes vor: NGS ermöglicht es, bislang eher exotische und auf Genomebene kaum bekannte Spezies aus wichtigen Ökosystemen zu analysieren. So wurde soeben ein Genkatalog von Riff-bildenden Korallen erstellt, um daraus Gene zu extrahieren, die für eine Klima-Anpassung von Korallen wichtig sind. Auch in der Mikrobiologie ist NGS ein Durchbruch: Die relativ kleinen Genome von Bakterien (nur zirka ein Tausendstel des Humangenoms) werden zu Hunderten sequenziert und verglichen. Dabei zeigt sich zum Beispiel, dass Bakterien durch den „horizontalen Gentransfer“ natürlicherweise DNA aus der Umwelt aufnehmen, in ihrem Genom etablieren und daraus neue Eigenschaften entwickeln können. Weitere Anwendungen findet NGS beispielsweise in der qualitativen und quantitativen Analyse von Genprodukten wie RNA-Molekülen (Transkriptom). Durch „deep sequencing“ ist es möglich, nur in sehr geringer Kopienzahl (statistisch weniger als ein Molekül pro Zelle) vorliegende Transkripte nachzuweisen. Mit der NGS-Transkriptomanalyse können

gezielt diejenigen Bereiche des Genoms erfasst werden, die funktionell wichtig sind und Aussagen über die Regulation der Genaktivität in verschiedenen Entwicklungsstadien und Geweben eines Organismus liefern.

**Bye bye Sanger: die neue Generation von Sequenzierungsverfahren**

Das Wesen von NGS besteht in dem Verzicht auf langsame und kostspielige Abläufe. Abgeschafft wurden die klassische Klonierung von DNA, durch die man homogenes Erbmateriale für die Sequenzierung erhielt, und der unpraktische Sequenz-Leseschritt durch Gelelektrophorese. Die Durchführung der Sequenzierungsreaktion im Mikro- oder zukünftig gar Nanomaßstab senkt den Reagenzienverbrauch drastisch. Drei NGS-Verfahren konkurrieren derzeit auf dem Markt, basieren aber auf unterschiedlichen Prinzipien und werden zum Teil unterschiedliche Anwendungsgebiete in der Genomforschung haben.

Das meist verbreitete Verfahren von 454 Life Sciences/Roche ist eine Weiterentwicklung der bereits Anfang der 1990er Jahre erdachten Pyrosequenzierung (Abb. 1A). Zunächst wird die zu sequenzierende DNA (etwa ein komplettes Genom oder Kopien der Gentranskripte = cDNA) physikalisch zerlegt. Die Bruchstücke werden einzeln an 20µm-Mikrokügelchen (Beads) gekoppelt und daran heftend durch eine Polymerasekettenreaktion (PCR) klonal vermehrt; das heißt, eine Kugel trägt viele identische Kopien eines bestimmten DNA-Moleküls. Die Beads werden nun einzeln in Löcher einer sogenannten Picotiter-Platte gefüllt. Bei einer Million Mikro-Reaktionslöchern (Wells) pro Platte können ebenso viele verschiedene Moleküle gleichzeitig sequenziert werden. Bei der eigentlichen Sequenzreaktion wird wie bei Sanger ein Einzelstrang des zu sequenzierenden DNA-Moleküls als Vorlage genommen und mithilfe eines Primers und einer Polymerase zum Doppelstrang ergänzt („sequencing-by-synthesis“). Immer wenn ein Nukleotid richtig, das heißt komplementär zum Vorlagenstrang, eingebaut wird, kann das dabei freigesetzte Pyrophosphat (PPi) mithilfe eines in den Wells befindlichen Enzymsystems zunächst in den Energielieferanten Adenosintriphosphat (ATP) und dann in einen zu messenden Lichtblitz umgewandelt

werden (Abb.1A, B). Nacheinander werden die vier DNA-Bausteine Adenin (A), Guanin (G), Cytosin (C) und Thymin (T) hinzugefügt (Sequenzierzyklus 1). Zwischen den einzelnen Nukleotidzugaben erfolgen Waschschriffe, die das System „zurücksetzen“. Wenn der Matrizenstrang zum Beispiel ein Adenin enthält, leuchtet das Well ausschließlich bei der Zugabe von Thymin auf. Danach erfolgt der nächste Sequenzierzyklus, wieder mit den vier Zugabeschritten. Wenn hier zum Beispiel der Lichtblitz bei G-Zugabe entsteht und dreimal so stark ist wie der Blitz in Zyklus 1, so lautet die Sequenz bis hierhin „TGGG“. Die Darstellungsform solcher Sequenzdaten („Flowgram“) ist in Abbildung 1C gezeigt. Durchschnittlich etwa 400 Nukleotide (geplant sind bis 1.000) können so derzeit pro Well gelesen werden. Diese Leselänge kommt dem Sangerverfahren nahe und ermöglicht eine passable Aufstellung (Assemblierung) von Teilsequenzen eines Genoms. Die 454-Technologie gilt daher als Methode der Wahl für die *de novo*-Sequenzierung unbekannter Genome (Mensch, Bakterienstämme etc.) und die Sequenzierung von Transkriptomen lassen sich mit der vielseitigen 454-Technik durchführen. Die Datenmengen sind mit bis zu 500 Mega-Basenpaaren (MBp) Sequenzinformation pro zehn Stunden Gerätelaufl bereits erheblich (Sanger-Kapillarsequencer: ein MBp pro Tag), liegen jedoch unterhalb der beiden anderen Systeme, die daher für rein quantitative Applikationen (Auszählen von DNA-Schnipseln, wie etwa zur Messung der Genexpression) besser geeignet sind. Schwächen hat die 454-Technologie prinzipbedingt beim Lesen langer Homopolymer-Abschnitte, die jedoch in den wichtigen kodierenden Genbereichen nicht so häufig sind: Es ist schwierig, die Lichtintensität nach Einbau von zum Beispiel 20 C-Nukleotiden gegenüber nur 19 Bausteinen zu diskriminieren. Dennoch liegt insbesondere bei ausreichend hoher Redundanz (das heißt mehrfachem Sequenzieren derselben DNA-Region) die Lesegenauigkeit bei etwa 99 Prozent.

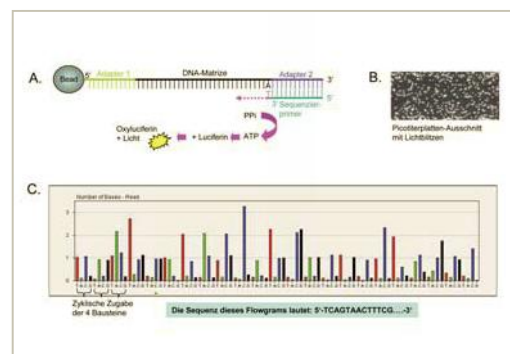
Die zwei weiteren Verfahren der Firmen Illumina („Genome Analyzer II“) und ABI („SOLiD“) sind technisch ebenfalls ausgesprochen elegant. Beide produzieren erheblich größere Datenmengen, aber mit 35 bis 100 Basenpaaren (Bp) viel kürzere Leseweiten pro Teil-Sequenzierung (Read). Für eine detaillierte Erklärung dieser Verfahren und der bereits angekündigten Technologien der dritten Generation verweisen wir aus redaktionstechnischen Gründen auf eine erweiterte Version dieses Manuskripts (als PDF-File herunterzuladen auf <http://molgen.biologie.uni-mainz.de>).

**Bioinformatische Herausforderungen**

So effizient die neuen Hochdurchsatz-Sequenzierungsverfahren auch sind, so groß ist die Herausforderung für die Bioinformatik, diese Datenmengen zu speichern, sie auszuwerten und sinnvolles Wissen

Abb. 1: Die 454-Pyrosequenzierung.

- A. Prinzip der Sequenzreaktion durch DNA-Synthese.
- B. Detektion der Lichtsignale auf Picotiterplatten-Ausschnitt.
- C. Darstellung des Sequenzierungsergebnisses als „Flowgram“.



© T. Hankeln

über das Genom daraus zu generieren. In der Zeitschrift „Nature Biotechnology“ wurde dies kürzlich mit dem Versuch verglichen, seinen Wasserdurst aus einem Feuerwehrschauch heraus zu stillen. Aufgrund der seriellen Fluoreszenzmessung in jedem Sequenzierungsschritt produzieren alle NGS-Techniken Bild-Rohdaten in bisher nicht gekanntem Umfang. Das SOLiD-System erfordert 15 Terabyte (TB) an Speicher für die reine Arbeitsumgebung sowie 30-40 TB für mittelfristige Datenarchivierung. Dabei ist eine Langzeitlagerung der wertvollen Rohdaten ratsam. Charakteristisch für die 454-Pyrosequenzierung sind zum Beispiel falsche Abschätzungen der Nukleotidzahl in Homopolymer-Nukleotidabschnitten, artifizielle Baseninsertionen und Fehler durch unterschiedliche Matrizen-Moleküle an einem Bead. Die sogenannten Base-Calling-Algorithmen werden jedoch stetig verbessert und eine Re-Analyse alter Läufe ist daher sinnvoll. Parallel zum Problem der Archivierung stoßen die althergebrachten Laborprotokollbücher an ihre Grenzen und müssen durch professionelle Labor-Managementsysteme ersetzt werden.

Entscheidend ist jedoch, dass leider eine Gemeinsamkeit von Sangertechnik und NGS weiter besteht: Alle derzeitigen Verfahren sind methodisch bedingt auf eine Leselänge beschränkt, die zwischen 35 Nukleotiden bei SOLiD und etwa 1.000 Nukleotiden bei Sanger liegt. Daher müssen die DNA-Moleküle komplexer Genome zunächst physikalisch zerstückelt und dann in Form von Millionen relativ kleinen Stücken sequenziert werden („Shotgun“-Verfahren, Abb. 2A). Erst der Computer kann aus diesen Schnipseln durch Erkennen von Überlappungen der Nukleotidabfolge zwischen den Teilabschnitten wieder die gewünschte Sequenz in originaler Moleküllänge rekonstruieren. Gemeinhin wird dies mit einem Puzzle aus Millionen von Teilen verglichen, ohne dass man allerdings das fertige Bild vor Augen hat (Abb. 2B). Diese *de novo*-Assemblierung zusammenhängender DNA aus Sequenzschnipseln zu sogenannten Contigs (contiguous sequence) ist als NP-schweres informatisches Problem ohne exakte Lösung bekannt. Mathematisch haben wir es hier mit einem „shortest common superstring“-Problem bzw. Hamilton-Graph zu tun (Abb. 3A). Die Teilsequenzen sind dabei als Knoten dargestellt, eine Sequenzüberlappung zwischen ihnen als Kante. Ziel der Assemblierung ist es, den kürzesten Weg zu rekonstruieren, bei dem alle Knoten nur einmal angesteuert werden. Abbildung 3A zeigt, dass schon bei Auftreten geringfügiger Sequenz-Doppelungen zwischen Reads mehrere Rekonstruktionspfade möglich sind. Die von NGS erzeugten mittellangen bis kurzen Reads verschärfen dieses Problem: Je kürzer und zahlreicher die Schnipsel sind, desto problematischer gestaltet sich der Assemblierungsvorgang. So wurde gezeigt, dass 750 Bp lange Sanger-Reads das Neisseria-Bakteriengenom immerhin in 59 Contigs von meist mehr als 1.000 Nukleotiden Länge assemblieren können, während

70 Bp-short Reads mehr als 1.800 Contigs produzieren. Hier tut also Algorithmenentwicklung Not, die auf die Verarbeitung von Millionen kurzer Reads speziell abgestimmt ist.

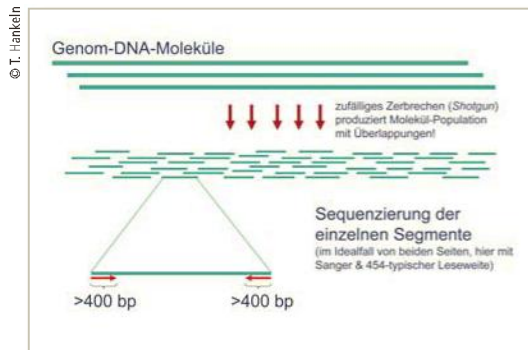


Abb. 2A: Genomsequenzierung im Shotgun-Verfahren.

Herstellung der Teil-Sequenzierungen (Reads)

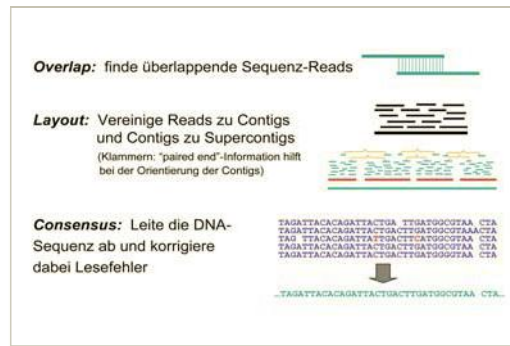


Abb. 2B. Assemblierung der Gesamtsequenz in drei Schritten („overlap-layout-consensus“-Ansatz).

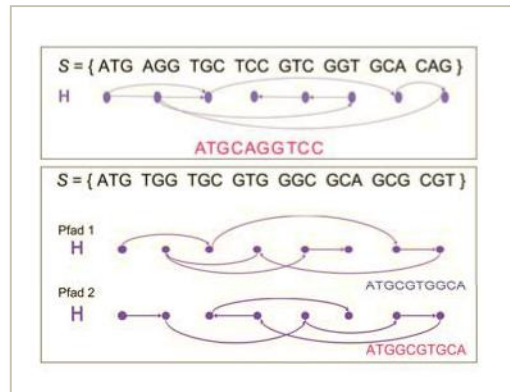
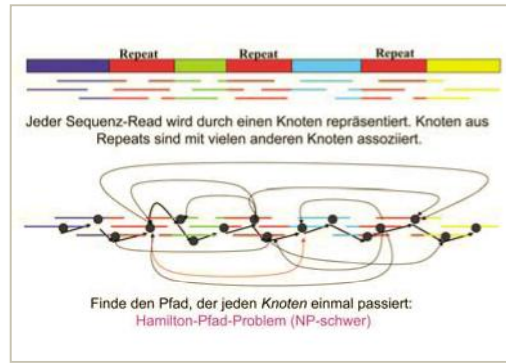


Abb. 3A: Rekonstruktion der Gesamtsequenz durch Graphentheoretische Verfahren.

Hamilton-Graphen (H). Die Reads (S) bestehen in diesem einfachen Beispiel aus Trinukleotiden. Das Beispiel unten ist nicht eindeutig lösbar, da zwei Pfad-Möglichkeiten für die Assemblierung bestehen.

Das größte Problem bei der Assemblierung ist, dass komplexe Genome von eukaryotischen, vielzelligen Organismen oft viele Millionen von ganz ähnlichen, sich nahezu perfekt wiederholenden „repetitiven“ Sequenzen besitzen, deren mögliche Funktionen im Genom noch unklar sind. Etwas voreilig werden sie oft als Genom-Müll bezeichnet. Das Humangenom besteht zu fast 50 Prozent aus repetitiver DNA, darunter allein mehr als eine Million Repeats vom Typ „Alu“ mit 300 Nukleotiden Länge. Ist die Leselänge der Sequenzierungsmethode „n“ kleiner als die Repeatlänge „n“, so kann kein Standard-Assemblierungsalgorithmus Ordnung schaffen, da er nicht erkennen kann, an welche Stelle des Genoms genau eine bestimmte repetitive Kopie gehört (Abb. 3B). Im Hamilton-Graph verweist jeder Repeat-Knoten

Abb. 3B. Hamilton-Pfade sind problematisch, wenn eine Sequenz mit Repeats durchsetzt ist (verändert nach [www.bioalgorithms.info](http://www.bioalgorithms.info)).



auf viele andere ebensolche Knoten. Jedes Repeat ist dann ein Bruchpunkt für die Assemblierung. Das Resultat sind extrem stark fragmentierte Assemblies: So zeigen Simulationen, dass das Genom des Fadenwurms *Caenorhabditis elegans* (110 Millionen Nukleotide lang) mit 50 Bp-Reads nur zur Hälfte und nur zu Contigs von maximal 10.000 Nukleotiden zusammengesetzt werden kann, ein natürlich unbefriedigendes Ergebnis. Sanger-Reads mit 1.000 Bp Leseweite haben hingegen im Humangenom die meisten Repeat-Hürden erfolgreich genommen und Contigs in nahezu vollständiger Chromosomenlänge ermöglicht.

Als Lösungsansatz für NGS-Verfahren bietet es sich an, sogenannte „paired end“-Information zu benutzen. Dabei werden die DNA-Moleküle eines Genoms vor der Sequenzierung so getrimmt, dass sich deren zu sequenzierende Enden in einem vom Experimentator bestimmten festen Nukleotidabstand befinden. Diese Abstandsinformation wird bei der Assemblierung genutzt, um repetitive Bereiche zu überbrücken. Doch auch neue algorithmische Möglichkeiten tun sich auf. Pevzner und Kollegen haben beispielsweise für die Assemblierung anstatt des NP-schweren Hamilton-Pfads die Variante eines Euler-Pfad-Ansatzes („de Bruijn-Graph“) vorgeschlagen, der effiziente Lösungen in linearer Zeit verspricht. Hierbei werden anstatt der Knoten (Sequenzen) alle Kanten (Nukleotidüberlappungen) nur einmal begangen.

Ungeachtet der aktiven bioinformatischen Forschung zeigen unsere Vergleiche implementierter Assemblierungswerkzeuge deutliche Unterschiede in ihrer Performance: So variierte kürzlich die Anzahl von erstellten Contigs eines 45.000 Reads umfassenden cDNA-Sequenzierungsprojekts zwischen 400 und 7.000. Natürlich ist so ein Ergebnis durch Assemblierungsparameter bedingt, wie etwa die Länge und Match-Qualität der erforderlichen Überlappung zweier Sequenzen. Oftmals verhalten sich gerade die von NGS-Unternehmen mitgelieferten „Komplettlösungen“ wie eine „Schwarze Box“ mit intern definierten Einstellungen. Auch geeignete Benchmarking-Datensätze sind rar, so dass es schwer fällt, das derzeit bestfunktionierende Tool zu identifizieren. Darüber hinaus stoßen kleine Arbeitsgruppen, selbst solche mit kleinen Cluster-Architekturen, bei umfangreiche-

ren Assemblierungsprojekten rechnertechnisch an Grenzen. Hier wird es Aufgabe sein, die universitären Rechenzentren und NGS-Lieferanten zur Anpassung der Firmensoftware auf die hauseigenen Systeme zu bewegen.

Die Erstellung von Contigs für Genomsequenzen oder mRNA-Transkripte stellt zudem nur den Anfang der bioinformatischen Analyse dar. Erst danach erfolgt die Suche nach dem Informationsgehalt der Sequenzen. Diese *downstream*-Analyse erfordert erneut die intensive Zusammenarbeit von Molekularbiologen und Bioinformatikern. Hier wird ein entscheidendes Erfolgskriterium der neuen Sequenzierungstechniken liegen.

### Genomforschung und NGS in Mainz

Die Genomforschung hat Tradition in Mainz. Schon Ende der 1970er Jahre wurden in der AG Schmidt kleine Genomabschnitte sequenziert. Nach Einführung der fluoreszenzbasierten Sanger-Sequenzierung 1990 konnten wir uns an internationalen Genomprojekten beteiligen, so zum Beispiel 1996 an der Entschlüsselung des Hefe-Genoms. Eine deutlich anspruchsvollere Größenordnung war das Humangenomprojekt. Wir haben den damals innovativen Ansatz der „komparativen Genomik“ verfolgt, indem wir parallel einen humangenetisch interessanten Chromosomenabschnitt des Menschen (eine Million Nukleotide auf Chromosom 11p15.3) und die entsprechende Region des Mausgenoms (Chromosom 7) sequenziert haben. Durch Vergleich der homologen Sequenzabschnitte konnten wir in den zunächst anonymen Sequenzen die konservierten Genstrukturen exakt definieren. Mit Einführung der NGS-Technologie wird die Diversität und Zahl der Genomik-Projekte nun erheblich angewachsen:

■ In der AG Zischler (Anthropologie) wird eine völlig neue Genklasse, piRNA genannt, untersucht. piRNA-Genen kommt eine wichtig Funktion in der Abwehr „springender“ Nukleinsäuren (Transposons) zu, die über evolutionäre Zeiträume in der Lage sind, Genome regelrecht zu parasitieren. Durch Katalogisierung der vermutlich mehr als 50.000 piRNA-Genloci im Menschen und in nicht-humanen Primaten soll erforscht werden, wie piRNA-Pools die Besiedlung von Genomen mit springender DNA kontrollieren. Nur NGS-Methoden können mit angemessenem Aufwand die erforderlichen Daten produzieren.

■ Die prähistorische Populationsgenetik steht im Mittelpunkt des Interesses der AG Burger (Anthropologie, Palaeogenetik). Im Spurenlabor untersuchen die Anthropologen alte DNA aus archäologischen Skeletten des Menschen und seiner Haustiere. Ziel ist die Rekonstruktion der Besiedlungsgeschichte Europas und Zentralasiens. Durch NGS kann die bisherige Datenmenge etwa zwanzigfach erhöht werden, wo-

mit eine detaillierte Rekonstruktion prähistorischer demographischer Dynamik möglich wird.

■ Die AGs Lieb (Zoologie) und Hankeln (Molekulargenetik) arbeiten im Rahmen des DFG-Schwerpunktprogramms 1174 „Deep Metazoan Phylogeny“ mit großen Sequenzdatensätzen (sogenannte Phylogenomik) an der Aufklärung der stammesgeschichtlichen Stellung und des Genrepertoires exotischer Tiergruppen, für die es bislang keine Gen(om)information gibt. Hierzu zählen extrem seltene Molluskenarten aus der Tiefsee sowie Rädertierchen mit ungewöhnlichen Fähigkeiten zur Trockenheitstoleranz und Strahlungsresistenz.

■ In Kooperation mit der Uni Frankfurt bearbeitet die AG Hankeln auch ökologische und evolutionsbiologische Fragestellungen bei Zuckmücken, deren Larven in Gewässer-Ökosystemen einen Großteil der Biomasse ausmachen. Obgleich Chironomiden etablierte Modellorganismen der Ökotoxikologie darstellen, ist ihr Genom quasi unerforscht. Durch NGS-Transkriptomanalyse sollen Gene identifiziert werden, die eine klimatische Anpassung von Chironomiden steuern und eine Grundlage für Artbildungsprozesse darstellen.

■ Als Mitglied im EU-Konsortium „EUROGrow“ identifiziert die AG Schmidt (Molekulargenetik) durch NGS-Transkriptomanalyse Gene, die beim Aufbau von Knochen und Knorpel eine Rolle spielen. Ziel ist die Aufklärung der Genexpressionsmuster bei der Differenzierung mesenchymaler Stammzellen. Die *in vitro*-Differenzierung dieser Zellen zu Chondroblasten und Chondrozyten ist in der regenerativen Medizin von großem Interesse.

Das neue Zentrum für Genomsequenzierung soll an die vorhandenen Nukleinsäureanalytik-Einrichtungen angegliedert werden. Wir planen zunächst die Etablierung der 454-Technologie, die insbesondere eine *de novo*-Sequenzierung erlaubt. In einer zweiten Ausbaustufe soll für komplementäre Anwendungen eine Short-Read-Plattform etabliert werden. Eine adäquate Verarbeitung und Auswertung der Daten erfordert entsprechend leistungsfähige Rechneranlagen und bioinformatische Kompetenz. Ein Kernstück des neuen NGS-Zentrums ist daher der Forschungsschwerpunkt „Rechnergestützte Forschungsmethoden in den Naturwissenschaften“, der einen klaren Standortvorteil für die Universität darstellen wird.

■ **Summary**

Life science research is currently being revolutionized by novel ultrahigh-throughput methods for deciphering genomic information of organisms. Within a few years, these new DNA sequencing technologies (termed “Next-Generation Sequencing”, NGS) will most probably allow us to know our own personal genome at reasonable costs, with

enormous biomedical impact. The huge amounts of sequence data produced by NGS, however, create big challenges for bioinformatics methods to keep pace. The University of Mainz “Competence Center for Nucleic Acid Analysis” and the research focus “Computational Sciences Mainz” are joining forces to centrally establish NGS technology.

■ **Kontakt**

Univ.-Prof. Dr. Thomas Hankeln  
 Institut für Molekulargenetik, gentechnische Sicherheitsforschung und Beratung  
 Johannes Gutenberg-Universität Mainz  
 Johann-Joachim-Becher-Weg 30a  
 D-55128 Mainz  
 Tel. +49 (0) 6131-39 23 277  
 Fax +49 (0) 6131-39 24 585  
 Email: hankeln@uni-mainz.de

**Univ.-Prof. Dr. Thomas Hankeln**



Privat

Thomas Hankeln, Jahrgang 1959, hat Biologie und Geographie an der Ruhr-Universität Bochum studiert und dort 1990 promoviert. Nach seinem Wechsel nach Mainz hat er Themen der Genomforschung bearbeitet und sich 1998 mit Arbeiten zur Molekularen Evolution von Genfamilien im Fach Genetik habilitiert. Seit 2001 ist Thomas Hankeln C3-Professor am Institut für Molekulargenetik der Johannes Gutenberg-Universität. Gegenwärtige Forschungsgebiete umfassen die Identifizierung von Genen durch vergleichende Genomanalyse, die Funktionsaufklärung solcher Gene in Tiermodellen sowie die Verwendung von Genomdaten zur phylogenetischen Systematik. Thomas Hankeln ist Mitgründer des seit 1998 bestehenden Biotechnologieunternehmens GENterprise GmbH, das im Bereich der Genomforschung Service-Dienstleistungen anbietet sowie Forschungs- und Entwicklungsprojekte durchführt.

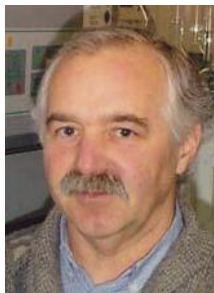
**Univ.-Prof. Dr. Hans Zischler**



Privat

Hans Zischler, Jahrgang 1957, studierte Biologie an den Universitäten Hohenheim und Tübingen. Er hat seine Doktorarbeit am Max-Planck-Institut (MPI) für Psychiatrie in München angefertigt und wurde 1991 promoviert. Nach Postdoc-Zeiten am MPI für Psychiatrie und an der Ludwig-Maximilians-Universität München übernahm er 1997 die Leitung der Arbeitsgruppe Primatengenetik am Deutschen Primatenzentrum in Göttingen. 2002 wurde er auf den Lehrstuhl für Anthropologie der Universität Mainz berufen. Seine Forschungsschwerpunkte liegen in der Analyse von evolutionären Mustern und Prozessen innerhalb der Divergenz nicht-humaner Primaten und des Menschen.

**Univ.-Prof. Dr. Erwin R. Schmidt**



Privat

Erwin R. Schmidt, Jahrgang 1949, studierte von 1967 bis 1973 Biologie und Chemie an der Justus Liebig-Universität in Gießen und hat 1975 dort in Genetik promoviert. Von 1975 bis 1978 war er DFG Postdoc an der Ruhr-Universität Bochum im Fachbereich Biologie, ab 1978 wissenschaftlicher Assistent im Institut für Genetik im FB Medizin. Er habilitierte 1985 in Genetik. 1989 folgte er einem Ruf auf eine Professur für Molekulargenetik nach Mainz. 1992 erhielt er den Ruf auf eine Professur für Genetik in Stuttgart-Hohenheim, kehrte aber nach einem Jahr Stuttgart als Professor für Molekulargenetik zurück nach Mainz. 1994 wurde er Leiter des unter seiner Mitwirkung neu gegründeten Instituts für Molekulargenetik, gentechnologische Sicherheitsforschung & Beratung. 1998 gründete er zusammen mit Kollegen die Biotechnologiefirma GENterprise. Seit April 2008 dient er dem Fachbereich Biologie als Dekan.