

Die Entwicklung der DNA-Sequenzierung – vom Genom eines Bakteriophagen zum Neandertaler**

Uschi Sundermann, Susanna Kushnir und Frank Schulz*

Genesequenzierung · Fluoreszenz ·
Genetischer Code · Nucleotide

Im Jahr 1977 publizierten Sanger et al. die erste jemals bestimmte genomische DNA-Sequenz: die etwa 5.000 Basenpaare eines Bakteriophagen-genoms.^[1] Im selben Jahr veröffentlichte Sanger die nach ihm benannte Methode zur Sequenzierung von DNA,^[2a] eine experimentelle Technik, die in den kommenden Jahrzehnten die moderne Biochemie revolutionieren sollte und für die Sanger zum zweiten Mal mit dem Nobelpreis geehrt wurde.^[2b] Mit dem Ziel der Entzifferung des menschlichen Genoms begann in den folgenden Jahren eine massive Weiterentwicklung der Sanger-Sequenzierung (Tabelle 1). Das 1990 initiierte Humangenomprojekt (HGP) führte zu einem fabrikähnlichen Ausbau der Sanger-Sequenzierkapazitäten in den beteiligten Instituten. Durch Optimierung und weitgehende Automatisierung der einzelnen Arbeitsschritte rückte die Aufklärung komplexer Genome in Reichweite.

In den frühen 1990er Jahren gelang dank dieser Entwicklung die Bestimmung kleinerer bakterieller Genome^[3] und 1996 bereits die des Genoms der Bäckerhefe, *Saccharomyces cerevisiae*.^[4] 2001, nach einem ganzen Jahrzehnt, wurde parallel vom Konsortium des Humangenomprojektes und Celera Genomics eine erste Sequenz des Humangenoms publiziert.^[5,6] Diese noch lückenhafte Sequenz wurde 2004 nahezu vervollständigt.^[7] Das Humangenomprojekt war ein Meilenstein, jedoch waren die dabei eingesetzten Methoden zeitaufwändig und teuer. Eine breite Anwendung, sei es in einer eventuellen personalisierten Medizin oder zur routinemäßigen Sequenzierung von Mikroorganismen, war zu diesem Zeitpunkt nach wie vor utopisch. Auch wenn die lau-

Tabelle 1: Vergleich von Kosten und Zeitaufwand verschiedener Sequenzierungstechniken. Nur die kommerziell verfügbaren Techniken der ersten und zweiten Generation sind berücksichtigt. Mbp: 1×10^6 Basenpaare; Gbp: 1×10^9 Basenpaare.^[10]

Jahr	\$/Mbp	Tage/Gbp	Anmerkung
1977	n.b.	n.b.	Dideoxy-Methode
1990	10×10^6	n.b.	Beginn des HGP
1995	1×10^6	n.b.	Einführung der Kapillarelektrophorese
1998	5×10^5	n.b.	
2002	9×10^4	260 ^[a]	Endphase des HGP
2005	60	3.1 ^[b]	Roche 454 GS FLX
2006	2	2.3 ^[b]	Illumina Solexa 1G
2007	2	1.6 ^[b]	AB SOLiD System

[a] Der Zeitbedarf wurde für die gesamte Sequenzierungskapazität des HGP errechnet. [b] Der Zeitbedarf wurde für einen einzelnen Sequenzierungsautomaten bestimmt.

fenden Kosten für die Sequenzierung einer einzelnen Nucleobase im Laufe des Humangenomprojektes dank der fortschreitenden Technik von ca. 10 US\$ pro Base auf 0.09 US\$ pro Base sanken (siehe Tabelle 1), lagen die Gesamtkosten für die Entzifferung des menschlichen Erbguts bei ca. drei Milliarden US\$.^[8–10]

In diesen Tagen jedoch wurde, nur wenige Jahre nach seiner Initiierung, das Neandertaler-Genomprojekt unter der Leitung von Svante Pääbo in Leipzig zum Erfolg geführt.^[11] Die etwa 3.2 Milliarden Basenpaare des komplexen Neandertaler-genoms wurden aus Jahrtausende alten kleinen DNA-Fragmenten bestimmt. Die Startbedingungen für dieses Projekt waren, bedingt durch den naturgemäß schlechten Zustand des alten genetischen Materials und seine vergleichsweise geringe Verfügbarkeit, erheblich ungünstiger als für das HGP. Ein signifikanter Entwicklungssprung in der Sequenzierung von DNA hat aber eine wesentlich schnellere und ökonomischere Entzifferung des Erbguts unseres prähistorischen Verwandten ermöglicht.

Dieser Entwicklungssprung läutet nun eine neue Ära in der biochemischen Forschung ein. Die ersten Schritte auf dem Weg zu einer wirklich breiten Anwendbarkeit der Genomsequenzierung gelangen unabhängig voneinander durch „sequencing by synthesis“, entwickelt von 454 Life Sciences unter der Leitung von Jonathan Rothberg^[12] und „multiplex polony sequencing“, entwickelt von Shendure et al.^[13] Beide Gruppen nutzten Fluoreszenzmikroskope als Detektoren in der Sequenzierung, welche die simultane Sequenzierung

[*] U. Sundermann, Dr. S. Kushnir, Prof. Dr. F. Schulz
Technische Universität Dortmund, Fakultät für Chemie
Otto-Hahn-Straße 6, 44221 Dortmund (Deutschland)
Fax: (+49) 231-1332498
E-Mail: frank3.schulz@tu-dortmund.de
Homepage: <http://www.chemie.tu-dortmund.de/schulz>

U. Sundermann, Prof. Dr. F. Schulz
Max-Planck-Institut für Molekulare Physiologie
Abteilung für Chemische Biologie
Otto-Hahn-Straße 11, 44227 Dortmund (Deutschland)

[**] Die Autoren danken dem Beilstein-Institut zur Förderung der chemischen Wissenschaften für die großzügige finanzielle Unterstützung sowie dem Fonds der Chemischen Industrie für ein Liebig-Stipendium an F.S. und ein Doktorandenstipendium an U.S. U.S. ist Mitglied der International Max-Planck-Research School of Chemical Biology.

mehrere Hunderttausend DNA-Fragmente in winzigen Mengen ermögliche – eine enorme Steigerung gegenüber den Sequenzierungen im 96er-Format der Didesoxy-Methode. Die massive Parallelisierung trug maßgeblich dazu bei, dass bereits die erste Version des Sequenzierungsautomaten von 454 Life Sciences Genomsequenzierungen zu einem Sechstel der Kosten der Sanger-Methode versprach. Zu Beginn der Entwicklung hatte die Sequenzierung durch Synthese zunächst Startschwierigkeiten, da sie durch eine kleinere Leselänge der einzelnen Sequenzierungsläufe und geringere Genauigkeit der Reaktionen eingeschränkt war; in beiden Parametern war die Didesoxy-Methode auf ihrem hohen Entwicklungsstand der neuen Technik überlegen. Die „sequencing by synthesis“-Technik stand jedoch damals, anders als die Sanger-Methode, erst am Anfang ihrer Entwicklung, und die Länge der Fragmente, die ohne Einsatz der Bioinformatik in einem Stück sequenziert werden konnten (die Leselänge) stieg nach nur kurzer Entwicklungszeit von 100 über 250 auf heute 400 bis 500 Basen.^[14] Kurz nach Einführung der 454-Sequenzierungsplattform kamen zwei weitere konkurrierende Systeme auf den Markt: die Illumina-Solexa-Plattform und das SOLiD Sequencing von Applied Biosciences. Diese beiden Techniken senken die Kosten pro Basenpaar und erhöhen den Probendurchsatz gegenüber der 454-Technik noch einmal signifikant. Der Preis hierfür ist allerdings eine geringere Leselänge, was die De-novo-Sequenzierung von Genomen besonders in repetitiven Sequenzabschnitten erschweren kann.

Gemeinsam bilden sie die so genannte 2. Generation der Sequenzierungstechnik mit teilweise komplementären Stärken und Schwächen (die gegenwärtige Verbreitung der Systeme in der Literatur ist in Abbildung 1 dargestellt). Allen Techniken der zweiten Generation ist ein einleitender Fragmentierungsschritt der zu sequenzierenden genomischen DNA gemeinsam, ein Schritt, der in Analogie zur Sanger-Sequenzierung abläuft. Im Anschluss jedoch werden die DNA-Fragmente nicht kloniert und in vivo amplifiziert, wie

bei der Sanger-Technik, sondern durch Polymerasekettenreaktion (PCR) für die Sequenzierung verfügbar gemacht (Abbildung 2). Die nachfolgende Sequenzierung verläuft

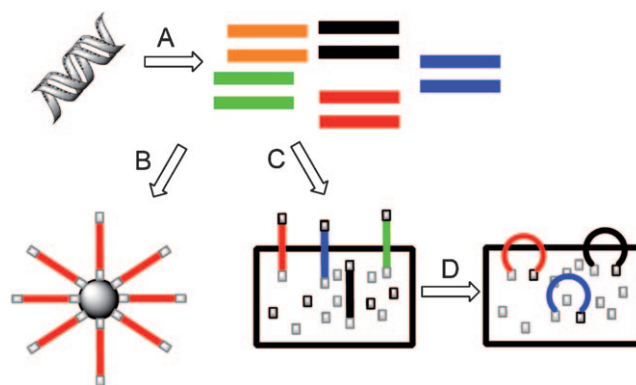


Abbildung 2. Übersicht über die einzelnen Schritte in einer Sequenzierungstechnik der 2. Generation. In einem ersten Schritt wird die genomische DNA durch Scherkräfte in kleine Abschnitte unterteilt (A). Anschließend werden diese in vitro mit kleinen DNA-Fragmenten verknüpft und, im Fall der 454-Technik, der Polony-Sequenzierung (Polony = Polymerase Colony) und des SOLiD-Systems, auf Festphasenkügelchen immobilisiert (B). Durch eine Emulsions-PCR wird die DNA amplifiziert, ein Schritt, der den in der Sanger-Methode intrinsischen „cloning bias“ vermeidet. Die Festphasenkügelchen werden vereinzelt und die Sequenzierungsreaktionen ohne chromatographische Schritte durch Fluoreszenzmikroskopie detektiert. In der Solexa-Technik werden die mit Oligonucleotiden markierten DNA-Fragmente durch Hybridisierung auf einer festen Oberfläche immobilisiert, eine Amplifikation erfolgt durch Bridge-PCR (D). Dadurch bilden sich Cluster identischer DNA-Sequenzen auf der Oberfläche, die anschließend in Sequenzierungsreaktionen ebenfalls fluoreszenzmikroskopisch detektiert werden.

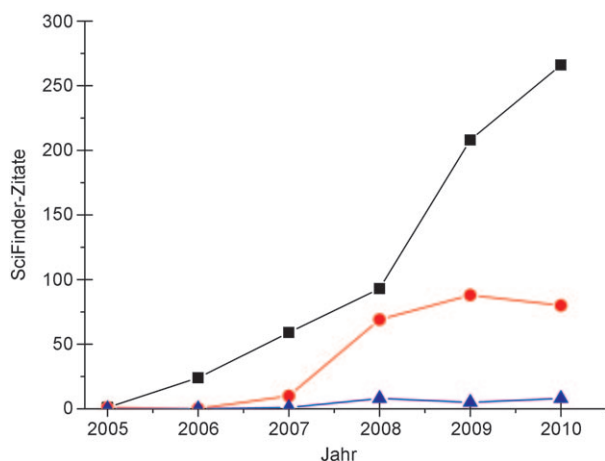


Abbildung 1. Ergebnis einer Literatursuche in SciFinder nach den kommerzialisierten Sequenzierungstechniken der 2. Generation: Zahl der gelisteten Zitate pro Jahr (Stand: Juni 2010, Werte für 2010 extrapoliert). ■: Roche 454 GS FLX; ●: Illumina Solexa Sequencer; ▲: Applied Biosystems SOLiD System.

nach unterschiedlichen Prinzipien, in allen Fällen werden aber hochauflösende Fluoreszenzmikroskope mit CCD-Kameras als Detektoren eingesetzt. Dies ermöglicht eine massive Parallelisierung und damit einen sehr großen Probendurchsatz bei vergleichsweise geringen Kosten pro Basenpaar (siehe Tabelle 1). Das Neandertaler-Genomprojekt setzte auf eine Kombination aus dem 454-GS-FLX- und Illumina-Solexa-GAII-System.^[11] Aus an vier verschiedenen Standorten geborgenen Knochen wurden jeweils wenige hundert Milligramm Knochensubstanz zur Gewinnung der DNA eingesetzt.

Das sehr hohe Kontaminationsrisiko durch humane oder sonstige Fremd-DNA erforderte den Einsatz speziell entwickelter Techniken. Teil der Maßnahmen war die Verwendung projektspezifischer Sequenzen in den zur Immobilisierung der genomischen DNA eingesetzten kleinen Sequenzabschnitten. Zudem zeigten erste Untersuchungen, dass die gewonnenen Proben bis zu 99% mikrobiellen Ursprungs waren, eine Konsequenz der langen Lagerzeit. Die Gruppe um Pääbo setzte daher Restriktionsenzyme ein, die bevorzugt GC-reiche bakterielle DNA-Sequenzen spalten. Hierdurch gelang eine 4- bis 6-fache Anreicherung der Neandertaler-DNA. Durch diese Maßnahme wurde zwar zwangsläufig die Abdeckung des Genoms reduziert, jedoch wurde dieser Verlust in Kauf genommen, um mit einer drastisch verbes-

serten Qualität der Sequenzierung das Genom mit 1-facher Abdeckung zu entziffern. Ein weiterer wesentlicher Aspekt der experimentellen Arbeit war der Grad an Kontamination mit menschlicher DNA. Diese hatte im Vorfeld der Veröffentlichung zu Kritik an dem Projekt geführt.^[15] Pääbo et al. nutzten zwei verschiedene statistische Analysen ihrer Daten und schlossen aus beiden übereinstimmend, dass der Anteil menschlichen Genmaterials in den Proben bei unter 1 % liegt. Zur Analyse der Sequenz wurde das Erbgut von fünf modernen Menschen (eines San aus Südafrika, eines Yoruba aus Westafrika, eines Papua Neuguineers, eines Han-Chinesen und eines französischen Europäers) entschlüsselt und mit den Genomen des Neandertalers und des Schimpansen sowie dem humanen Referenzgenom verglichen. Interessanterweise wurde dabei eine signifikante Verwandtschaft des Neandertalergenoms mit den europäischen und ostasiatischen, jedoch nicht mit den afrikanischen Genomen festgestellt. Der Fluss an genetischer Information lief nahezu vollständig in Richtung des modernen Menschen und macht einige Prozent unseres heutigen Genoms aus. Der Neandertaler steht Pääbo et al. zufolge dem heutigen Menschen genetisch sehr nahe, es handelte sich um einen engen Verwandten des anatomisch modernen Menschen, und eine partielle genetische Vermischung der beiden Spezies fand sehr wahrscheinlich nach dem Verlassen des afrikanischen Kontinents statt. Die Entzifferung des Neandertaler-Genoms demonstriert die enorme Kapazität der zweiten Generation an Sequenzierungstechniken, die innerhalb kurzer Zeit zu einer weiten Verbreitung der Systeme geführt hat. Die Anwendungen der 2. Generation sind breit gestreut und reichen von der Analyse von Metagenomen und Transkriptomen über das akkurate Sequenzieren („deep sequencing“) zur Identifikation von Einzelnucleotidpolymorphismen bis hin zu De-novo-Sequenzierungen wie im Fall des Neandertaler-Genoms. Das häufig diskutierte 1000\$-Genom liegt mit diesen Systemen allerdings noch in weiter Ferne.^[9] Daher wird bereits seit einigen Jahren an einer dritten Generation gearbeitet, der „next next generation“.^[10] Von diesen neuen Sequenzierungstechniken hat noch keine die Marktreife, geschweige denn eine Anwendung jenseits einzelner Modellversuche, erreicht. Wenn aber die Entwicklung in dem derzeit atemberaubenden Tempo weiter voranschreitet, handelt es sich wohl hierbei nur noch um eine Frage einiger Jahre, bis die Sequenzierung von Genomen zu einer Routinetechnik in biochemischen Labors geworden sein wird. Die Auswirkungen dieser Entwicklung auf die Natur- und Lebenswissenschaften ebenso wie auf die menschliche Gesellschaft sind derzeit noch nicht klar abzusehen – doch dass sie gewaltig sein werden, dürfte sicher sein.

Eingegangen am 25. Juni 2010

Online veröffentlicht am 22. September 2010

- [1] F. Sanger, G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchison, P. M. Slocumbe, M. Smith, *Nature* **1977**, 265, 687–695.
- [2] a) F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. USA* **1977**, 74, 5463–5467; b) F. Sanger, *Angew. Chem.* **1981**, 93, 937–944.

- [3] F. R. Blattner, G. Plunkett, III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, Y. Shao, *Science* **1997**, 277, 1453–1462.
- [4] R. A. Clayton, O. White, K. A. Ketchum, J. C. Venter, *Nature* **1997**, 387, 459–462.
- [5] International Human Genome Consortium, *Nature* **2001**, 409, 860–921.
- [6] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. D. Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R.-R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Y. Wang, A. Wang, X. Wang, J. Wang, M.-H. Wei, R. Wides, C. Xiao, C. Yan et al., *Science* **2001**, 291, 1304–1351.
- [7] International Human Genome Consortium, *Nature* **2004**, 431, 931–945.
- [8] F. S. Collins, M. Morgan, A. Patrinos, *Science* **2003**, 300, 286–290.
- [9] R. F. Service, *Science* **2006**, 311, 1544–1546.
- [10] P. K. Gupta, *Trends Biotechnol.* **2008**, 26, 602–611.
- [11] R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H.-Y. Fritz, N. F. Hansen, E. Y. Durand, A.-S. Malaspina, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prufer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Hober, B. Hoffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de La Rilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, S. Pääbo, *Science* **2010**, 328, 710–722.
- [12] M. Margulies, M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J.-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, J. M. Rothberg, *Nature* **2005**, 437, 376–380.
- [13] J. Shendure, G. J. Porreca, N. B. Reppas, X. Lin, J. P. McCutcheon, A. M. Rosenbaum, M. D. Wang, K. Zhang, R. D. Mitra, G. M. Church, *Science* **2005**, 309, 1728–1732.
- [14] J. M. Rothberg, J. H. Leamon, *Nat. Biotechnol.* **2008**, 26, 1117–1124.
- [15] J. D. Wall, S. K. Kim, *PLoS Genet.* **2007**, 3, e175.