

Wie lange dauert eine Generation?

von Prof. Dr. med. Georg Hoffmann

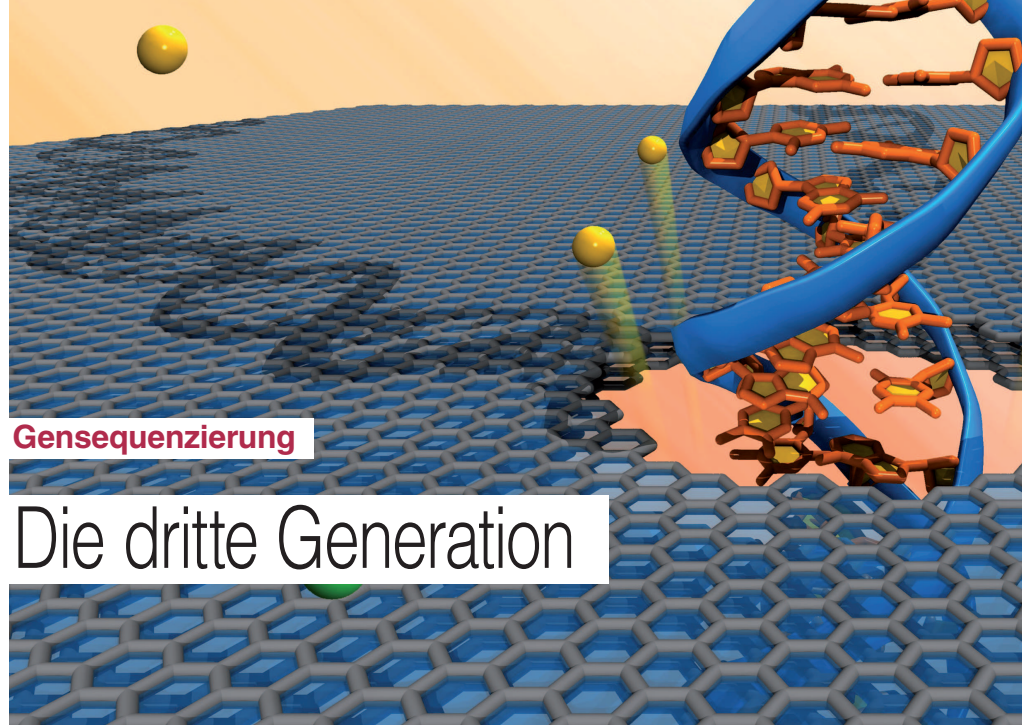
Wer die dritte Generation einer attraktiven Technologie auf den Markt bringen kann, hat gemeinhin besonders gute Erfolgchancen. Die erste Generation kämpft ja oft mit Geburtsfehlern, die zehnte oder zwanzigste lockt keinen Hund mehr hinter dem Ofen hervor. Aber „dritte Generation“: Das klingt bereits ausgereift und trotzdem noch innovativ.

So muss es nicht verwundern, dass viele Forscher und Unternehmen seit 2010 vom *Third Generation Sequencing*^[1] (TGS) sprechen, obwohl doch die zweite Generation gerade erst so richtig ins Rollen gekommen ist. In der Regel dauern nicht nur Menschen- sondern auch Gerätegenerationen etwa 25 Jahre (wobei kausale Zusammenhänge durchaus denkbar sind), und so lohnt ein kritischer Blick in die Geschichte der Gensequenzierung:

Die Geburt der ersten Generation datiert ins Jahr 1975, als es Frederick Sanger gelang, ein definiertes Oligonukleotid Base für Base auszulesen. Im Jahr 2000 entwickelte Jonathan Rothberg, der Gründer von *454 Life Sciences* ein Gerät der zweiten Generation zur gleichzeitigen Sequenzierung von Millionen Oligonukleotiden. Bis hierher stimmt der Zeitablauf.

Inzwischen ist es sogar möglich, einzelne Nukleinsäurestränge in rasender Geschwindigkeit ohne vorherige Amplifikation zu sequenzieren – zweifellos ein enormer Fortschritt. Ob die hier vorgestellten Systeme aber wirklich schon zu einer neuen Generation gehören oder nur Vorläufer und Zwischenstadien auf einem langen Weg darstellen, wird sich erst in einigen Jahren erweisen. Man darf gespannt sein, ob das bis 2025 dauert.

[1] Schadt E. et al. A window into third-generation sequencing. *Hum Mol Genet* 2010; doi: 10.1093/hmg/ddq416



Gensequenzierung

Die dritte Generation

Das *Next Generation Sequencing* (NGS) hat die Nukleinsäuresequenzierung dramatisch beschleunigt und verbilligt. Aber es gibt keine Atempause: Mit dem *Single Molecule Sequencing* (SMS) kündigt sich bereits der nächste Technologiesprung an.

Vor ziemlich genau drei Jahren, im Mai 2011, kam es in Norddeutschland zum weltweit größten Ausbruch einer *E. coli*-Infektion¹ mit blutigen Durchfällen und schweren Nierenkomplikationen. Rund 4.000 Menschen erkrankten, fast 1.000 entwickelten ein hämolytisch-urämisches Syndrom (HUS), 53 starben.

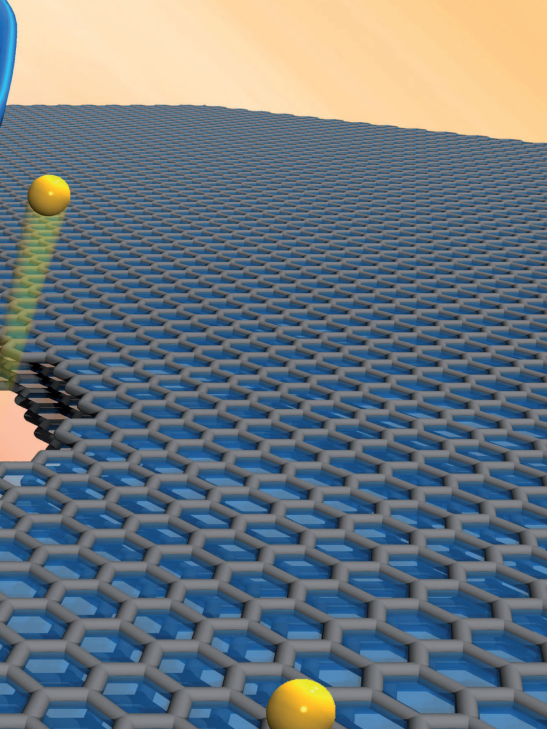
Nur wenige Wochen später, als die Epidemiologen noch fieberhaft nach der Infektionsquelle fahndeten, konnten Forscher eines internationalen Teams im *New England Journal of Medicine* bereits die biochemischen Ursachen aufzeigen, die zu dieser unerwarteten Aggressivität geführt hatten. Mithilfe eines damals brandneuen Verfahrens zur Gensequenzierung hatten sie in kürzester Zeit die kompletten Genome von zwölf verschiedenen *Coli*-Stämmen analysiert und herausgefunden, dass ein normalerweise harmloser Darmbewohner des Serotyps O104:H4 durch Gentransfer gefährliche Eigenschaften wie etwa Shiga-Toxin-2-Produktion und Antibiotikaresistenz erworben hatte.

¹EHEC = enterohämorrhagische *Escherichia coli*

Möglich wurde die schnelle Aufklärung unter anderem durch die SMRT-Sequenzierung (*Single Molecule Real-Time*), die gerade erst für ausgewählte Anwender auf den Markt gekommen war. Mit dieser Technik konnten die Untersucher anhand von kurzen Lichtblitzen das Enzym DNA-Polymerase in Echtzeit bei der Ablesung eines einzelnen Nukleinsäurestrangs beobachten – und das nicht nur einfach, sondern zigtausendfach in einem Gerät mit 75.000 parallel angeordneten Messzellen, von denen (aus technischen Gründen) etwa jede fünfte ein Enzymmolekül enthielt.

Großes Zukunftspotenzial

„Diese dritte Generation der Sequenziersysteme wird die Mikrobiologie revolutionieren“ prophezeite Prof. Eric Schadt, Bioinformatiker an der renommierten *Mount Sinai School of Medicine* in New York, nach dem EHEC-Erfolg. Mag gut sein, dass Schadt einen etwas zu euphorischen Blick hatte, da er 2011 noch als wissenschaftlicher Direktor bei der Herstellerfirma Pacific Biosciences arbeitete. Neutrale Experten sind sich einig, dass



Vermessung eines DNA-Einzelstrangs auf atomarer Ebene beschrieb; eine Weiterentwicklung dieses Prinzips, die *Transmission Electron Microscopy*, wurde 2010 vorgestellt und letztes Jahr an Aeon Biowares in Delaware verkauft; wann die Technik marktreif wird, ist allerdings offen.

Nanotechnologie

Das Titelbild zeigt das Prinzip einer weiteren Sequenzieretechnik, die man der dritten Generation zuordnet. Hier wird ein Nukleinsäure-Strang durch eine nur wenige Nanometer große Öffnung einer ultradünnen Membran gezwängt – so, als würde man einen Wollfaden durch ein winziges Nadelöhr ziehen. Bei den im Bild dargestellten Sechsecken handelt es sich um Nanotubes aus Kohlenstoff, es können aber auch Lipid-Bilayer wie in einer Zellmembran oder komplexe Proteinkonstrukte mit enzymatischer Aktivität sein.

Beim Durchtritt durch die „Nanoporen“ ändern sich mechanische, chemische, elektrische oder optische Parameter in Abhängigkeit vom jeweiligen Nukleotid. Die gemessenen Signale spiegeln somit die Basenabfolge wider. Ähnlich wie bei der oben beschriebenen Elektronenmikroskopie hat

auch dieses Verfahren den Charme, dass es native Nukleinsäuren analysieren kann und so potenzielle Fehlerquellen – zum Beispiel Strukturartefakte durch Fluoreszenzmarkierung – vermeidet.

Am weitesten fortgeschritten scheint die Entwicklung einer „disruptiven“ Technik bei Oxford Nanopore Technologies (UK) zu sein: Hier erfolgt beim Durchtritt durch die Nanopore eine enzymatische Spaltung des Nukleinsäurestrangs; bei der Translokation des freigesetzten Nukleotids durch die Nanopore entsteht im elektrischen Feld eine basenspezifische Ladungsänderung, die aufgezeichnet werden kann. Die Firma hat ein nur streichholzsachtelgroßes Gerät zur Sequenzierung von Bakterien- und Virengenomen angekündigt. Angeblich liegt der künftige Preis pro Genom unter einem Euro.

Auch IBM ist in diesem Feld tätig, scheint aber von einem marktfähigen System noch weiter entfernt zu sein. Beeindruckend sind jedenfalls die theoretischen Lesegeschwindigkeiten der eingesetzten Transistortechnik: Bis zu 500 Millionen Nukleotide pro Sekunde sollen erzielbar sein. Limitierend ist auch hier das zeitliche Auflösungsvermögen des Detektors.

die SMRT-Technik wegen zu vieler Lesefehler noch nicht für die medizinische Diagnostik geeignet ist. Aber wie hoch ihr Zukunftspotenzial eingeschätzt wird, erkennt man allein schon an den 75 Millionen Dollar, die Roche Ende 2013 für eine Partnerschaft mit Pacific Biosciences ausgab; der Pharmakonzern will seine derzeitige Sequenzier-Plattform der zweiten Generation mithilfe von „PacBio“ modernisieren.

Wie SMRT funktioniert, ist im Kasten unten beschrieben. Es gibt aber noch viele weitere SMS-Techniken. Der älteste Ansatz stammt bereits aus dem Jahr 1959, als Robert Feynman vom *California Institute of Technology* die elektronenmikroskopische

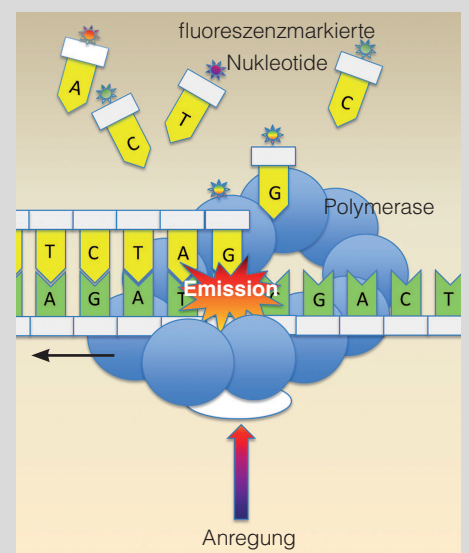
SMRT – eine „smarte“ Technologie

SMRT steht für *Single Molecule Real-Time (Sequencing)*. Zu Recht spricht man diese elegante Technik wie *smart* aus.

Ein einzelnes Polymerasemolekül wird am Boden der Messkammer über einem winzigen Loch fixiert. Dort synthetisiert das Enzym einen Nukleinsäure-Doppelstrang. Die Nukleotide (A, C, G und T) sind mit unterschiedlichen Fluoreszenzfarbstoffen markiert, die von unten her angeregt werden. Das Loch ist allerdings so klein (< 1 nm), dass das Licht nur das gerade eingebaute Nukleotid erreicht – alle übrigen liegen im Dunkel.

Und nun kommt der Clou: Im Moment des Einbaus wird der Fluoreszenzfarbstoff abgesprengt und taucht wieder ins Dunkel der Messkammer ein, so dass das Emissionslicht nur einen kurzen Blitz auslösen kann. Der Detektor empfängt zu jedem Nukleotid die passende Wellenlänge für Bruchteile einer Millisekunde.

Theoretisch baut die Polymerase etwa tausend Nukleotide pro Sekunde (!) ein. Da aber kein Detektor der Welt so schnell messen kann, muss die Geschwindigkeit durch gentechnische Manipulationen auf etwa drei Nukleotide pro Sekunde gedrosselt werden. In puncto Geschwindigkeit ist also noch viel Luft nach oben.



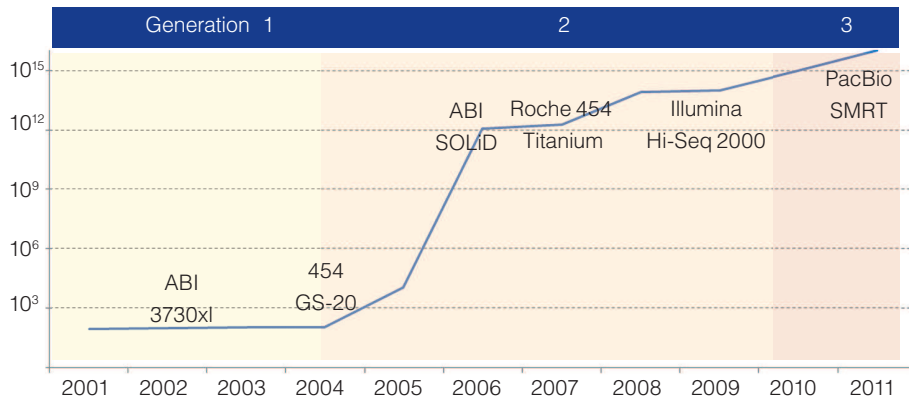
In der Praxis bieten die Systeme der dritten Generation bislang allerdings – was Geschwindigkeit und Preis betrifft – noch nicht die erhofften Vorteile; denn was möglicherweise dank eleganter Messtechnik gewonnen wird, geht durch höhere Fehlerraten wieder verloren.

Zugang zur funktionellen Genomik, also zur Auswirkung des genetischen Bauplans auf die Transkription und Translation.

Die bemerkenswerteste Eigenschaft der dritten Generation ist aber wohl die Signalverarbeitung in Echtzeit, weil sie kinetische Messungen erlaubt. So werden

Abschließend sei die HLA-Typisierung als bedeutsame Applikation von SMS erwähnt. Die HLA-Gene haben die höchste bei Eukaryonten beschriebene Komplexität und mit Abstand die meisten Polymorphismen. Trotz ihrer beträchtlichen klinischen Bedeutung sind mit Systemen der ersten und zweiten Generation bisher weniger als 10 Prozent auf dem Allelniveau sequenziert. Nur mit viel Aufwand konnte bestimmt werden, ob funktionell zusammengehörige Polymorphismen gemeinsam oder getrennt auf die väterlichen und mütterlichen Allele verteilt sind.

Systeme der dritten Generation arbeiten viel schneller als die erste und bieten deutlich höhere Leselängen von mehr als 1000 bp als die zweite. Deshalb kann man nun in kurzer Zeit viel mehr Polymorphismen erfassen und auch dann allelspezifisch zuordnen, wenn sie mehrere 100 Basenpaare voneinander entfernt liegen. 🌸



Der Durchsatz der Sequenziergeräte (kbp/Lauf) nahm von der ersten zur zweiten Generation sprunghaft zu, während sich bei der dritten Generation bislang nur ein moderater Zuwachs abzeichnet.

Da SMS ja definitionsgemäß mit Einzelmolekülen arbeitet, schlägt sich jeder zufällige Fehler 1:1 im Leseergebnis nieder. Kann beispielsweise ein Nukleotid beim Nanopore-Verfahren die Öffnung nicht in der vorgesehenen Zeit passieren, so wird sie mehrfach gezählt und täuscht eine Insertion vor. Man muss folglich ein Vielfaches an Wiederholungsläufen durchführen und entsprechend mehr Daten auswerten, um vergleichbare Zuverlässigkeit zu erzielen.

beispielsweise methylierte Basen beim SMRT-Verfahren langsamer prozessiert, sodass man ohne zusätzlichen Aufwand epigenetische Veränderungen des Genoms erfassen kann. Im Fall der eingangs erwähnten EHEC-Epidemie konnte auf diese Weise gezeigt werden, dass die Inaktivierung von Genen für die veränderte Pathogenität des Erregers eine Rolle spielt.



Prof. Dr. med. Rudolf Gruber
Mitglied der Redaktion

Es gibt aber dennoch eine Reihe von Vorzügen, die die Sequenzierung von Einzelmolekülen auf mittlere Sicht sehr attraktiv machen. Am augenfälligsten ist natürlich der Wegfall der vorgeschalteten DNA- oder RNA-Amplifikation mittels PCR bzw. Reverse-Transkriptase-PCR. Das bedeutet Zeitgewinn bei der Präanalytik und Elimination eines fehleranfälligen Schritts. Außerdem verschafft die direkte Sequenzierung verschiedener RNA-Spezies (mRNA, rRNA, miRNA etc.) der Forschung einen verbesserten

	Erste Generation	Zweite Generation	Dritte Generation
Typische Vertreter	ABI 3730/3730xl	Roche 454 Illumina Ion Torrent	Pacific Bioscience Oxford Nanopore IBM
Lesegenauigkeit	hoch	hoch	mittel
Leselänge	mittel	kurz bis mittel	lang
Durchsatz	niedrig	hoch	mittel
Kosten pro Base	hoch	niedrig	mittel
Kosten pro Lauf	niedrig	hoch	niedrig
Zeitbedarf	Stunden	Tage	Stunden

Sequenziergeräte der ersten Generation basieren auf der Kettenabbruchmethode von Sanger, bei der pro Zyklus nur eine Base bestimmt werden kann. Bei der zweiten Generation erfolgen Millionen von Lesezyklen parallel, bei der dritten werden Tausende von Nukleinsäure-Einzelmolekülen ohne vorherige Amplifikation gleichzeitig sequenziert.