



## Ein Virus wird entdeckt

### Forschung und Diagnostik mit dem Next Generation Sequencing

**Milchbauern in Nordrhein-Westfalen verzeichneten 2011 einen deutlichen Rückgang der Milchleistung von Kühen, Fieber und andere Krankheitssymptome. Aber wo lag die Ursache? Wissenschaftlerinnen und Wissenschaftlern am Friedrich-Loeffler-Institut (FLI) gelang es, einem bisher unbekanntem Krankheitserreger auf die Spur zu kommen. Mit Hilfe einer modernen molekularbiologischen Methode, der Metagenomanalyse, fanden sie an Proben von erkrankten Tieren Hinweise auf Viren, die bis dahin in Europa nicht aufgetreten waren. Dieses Ergebnis brachte die Wissenschaftler auf die Fährte des unbekanntem Virus und ermöglichte dessen Isolierung und Nachweis in Proben weiterer erkrankter Tiere.**

Die Genome aller Organismen und Viren bestehen aus Nucleinsäuren (Desoxyribonucleinsäuren, englisch abgekürzt DNA, bzw. Ribonucleinsäuren, englisch abgekürzt RNA) und speichern deren Baupläne. Als solche sind sie wichtige Quellen für Forschung und Diagnostik zur Identifizierung und Charakterisierung eines Organismus. Daher wurde schon bald nach der Identifikation der DNA als dem Informationsspeichermedium der Zellen an der Entschlüsselung dieser Information gearbeitet.

Die Genome bestehen aus vielen hintereinander gehängten Nucleotiden, den Bausteinen, die die Information mithilfe von vier Basen verschlüsseln. Die Basen sind die Buchstaben des DNA-Alphabets. Jeweils drei Basen bilden ein Wort, das als Triplet oder Codon bezeichnet wird. Die im Genom gespeicherten, aus vielen

Tripletts bestehenden, Sätze (= Gene) beinhalten die Information zur Bildung der für den Organismus erforderlichen Eiweiße, die die funktionellen Bestandteile der Zelle sind. Somit gibt die Abfolge der vier Buchstaben des DNA-Alphabets Auskunft über Eigenschaften des Organismus. Aus diesem Grund entwickelten Wissenschaftler Methoden zur Bestimmung der Basenabfolge oder Sequenzierung der DNA. Von den verschiedenen Methoden zur Sequenzierung setzte sich zunächst die Kettenabbruch-Methode nach Sanger durch. Gleichzeitig wurde an alternativen Verfahren zur DNA-Sequenzierung gearbeitet. In den frühen 2000er Jahren wurden die ersten dieser Verfahren zur Marktreife gebracht und fortan in Abgrenzung zum first generation sequencing, der für hohen Durchsatz optimierten Sequenzierung nach Sanger, unter der Bezeichnung next generation sequencing (NGS) zusammengefasst.

#### Entschlüsselung des Genoms mit neuen Methoden

Hinter NGS verbergen sich verschiedene Methoden, die allesamt zur Bestimmung der DNA-Sequenz nicht die klassische Kettenabbruchmethode nach Sanger nutzen. Inzwischen werden innerhalb des NGS second und third generation sequencing unterschieden. Wesentlicher Unterschied ist, dass beim second generation sequencing die Vervielfältigung der DNA notwendig ist, um bei der Sequenzierung messbare Signalstärken zu erhalten. Beim third generation sequencing wird die DNA direkt sequenziert. Dies hat den Vorteil, dass die ermittelten Sequenzen weniger Artefakte enthalten, die bei der Vervielfältigung der DNA auftreten.

Ungeachtet der Unterschiede zwischen den Verfahren, gibt es gemeinsame Merkmale. Neben einem deutlich höheren Probendurchsatz gehört zu diesen Gemeinsamkeiten die Möglichkeit, die zu sequenzierende DNA mit molekularbiologischen Verfahren direkt in eine sogenannte DNA-Bibliothek umzuwandeln. Als DNA-Bibliothek wird eine Mischung zahlloser DNA-Moleküle, die die DNA der Ausgangsprobe möglichst genau repräsentieren und für die Sequenzierung geeignet sind, bezeichnet. Da es sich bei den eingesetzten Methoden um ungerichtete Verfahren handelt, die alle DNA-Moleküle mit annähernd identischer Effizienz umsetzen, werden alle enthaltenen DNA-Moleküle mit gleicher Wahrscheinlichkeit in funktionelle Bestandteile der DNA-Bibliothek überführt und sequenziert. Jetzt ist es möglich, die Basenabfolge eines jeden DNA-Moleküls ohne jegliche Vorkenntnisse zu bestimmen. Die großen Datensätze er-



Abbildung 1: Besonders betroffene Arten: Rinder und Schafe

lauben es entweder, Genome mit großer Zuverlässigkeit der untersuchten DNA-Abschnitte zu sequenzieren oder einen breiten Überblick über die Zusammensetzung der DNA einer Probe zu erhalten. Durch die genannten Merkmale ergeben sich diverse Einsatzgebiete für die zur Verfügung stehenden NGS-Verfahren. Insbesondere die großen Datensätze, die generiert werden, sind dabei von herausragender Bedeutung, weil sie einen enormen Informationsgehalt bergen.

#### Wie funktioniert NGS?

Das Verfahren des Genome Sequencers wird beispielhaft in Abbildung 2 beschrieben. Die DNA-Moleküle werden auf eine für die Sequenzierung passende Länge fragmentiert (A). Anschließend wird die DNA in einem molekularbiologischen Verfahren enzymatisch mit speziellen Adaptern versehen und so in eine DNA-Bibliothek umgewandelt (B, C). Diese Adapter dienen der sortenreinen Vervielfältigung der Moleküle der DNA-Bibliothek, um bei der Sequenzierung messbare Signalstärken zu erhalten. Die Amplifizierung erfolgt in einer sogenannten Emulsions-Polymerase Kettenreaktion (emPCR®). In der emPCR® ist jedes Wassertropfenchen der Wasser-in-Öl-Emulsion ein individueller Mikroreaktor, in dem sich jeweils ein DNA-Bead (ein kleines Kügelchen, das die DNA binden kann) und ein Molekül der DNA-Bibliothek befinden (D). Dadurch wird die Sortenreinheit der Vervielfältigung gewährleistet. Nach der PCR (E) sind alle synthetisierten Kopien des Ausgangs-DNA-Moleküls fest an die Oberfläche der DNA-Beads gebunden. Diese Sammlung von Millionen Kopien eines einzelnen DNA-Moleküls wird als DNA-Klon bezeichnet. Im Anschluss an die emPCR® werden die Beads mit den an ihrer Oberfläche gebundenen Kopien eines Moleküls der ursprünglichen DNA-Bibliothek gereinigt und für die Sequenzierung vorbereitet (F). Dazu werden die Beads in eine sogenannte Pico Titer Platte (PTP) geladen (G). Diese PTP hat mehrere

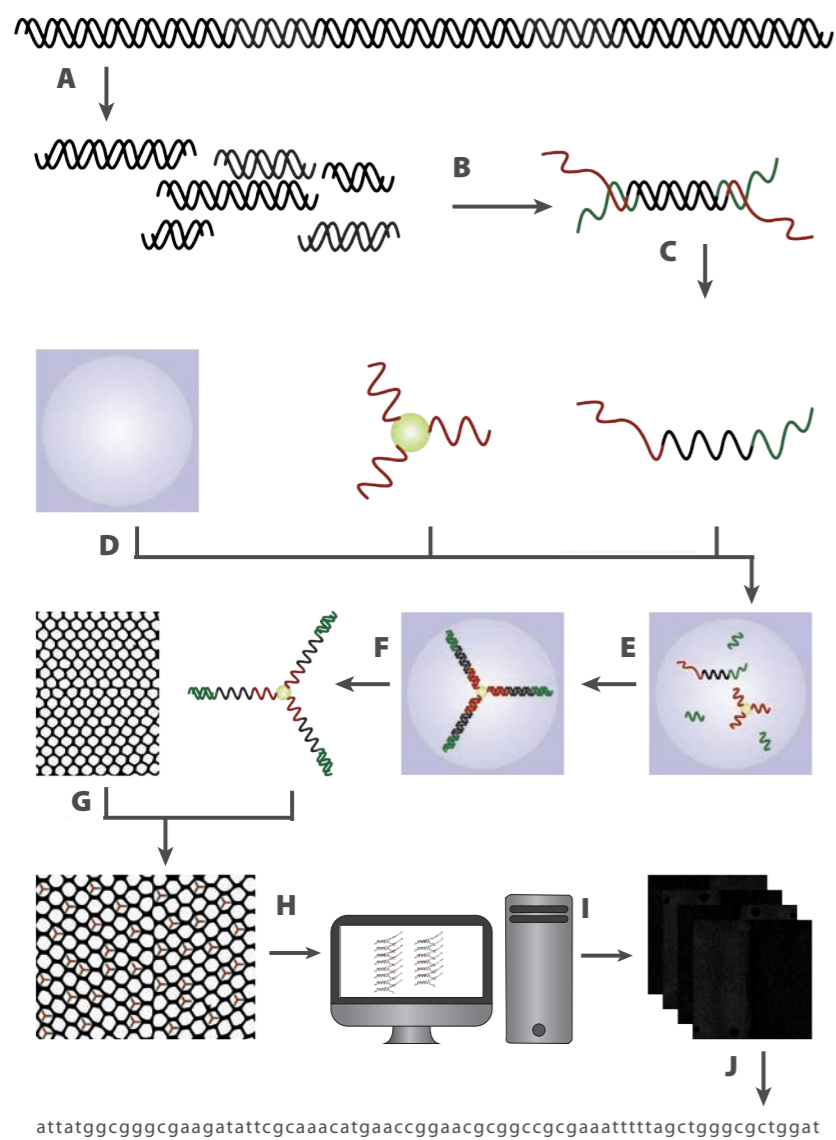


Abbildung 2: Prinzip des Next Generation Sequencing am Beispiel des Genome Sequencer FLX

Millionen kleine Vertiefungen die Wells genannt werden. In den Wells sind die DNA-Beads für die Dauer der Sequenzierung fixiert. Die beladene PTP wird ins Gerät überführt (H), das die Sequenzierung durchführt. Infolge eines Nukleotideinbaus wird ein Lichtsignal generiert, das von einer hochempfindlichen CCD-Kamera für alle Wells der PTP simultan detektiert und als Bilddatei gespeichert wird (I). Die Lichtmenge ist direkt proportional zur Anzahl der eingebauten Nukleotide. Daher kann aus dem durch die Kamera detektierten Signal die Anzahl der in einem Schritt eingebauten Nukleotide berechnet werden. Zusammen mit der Kenntnis des Nukleotids, das zur Lichtemission führte, kann die Gerätesoftware die Basensequenz jedes einzelnen, in einem Well der PTP befindlichen, DNA-Klons berechnen (J).

### Der Krankheit auf der Spur

Ein wichtiges Einsatzgebiet des NGS ist die sogenannte Metagenomanalyse. Diese dient dazu, die Zusammensetzung einer Population im Hinblick auf die enthaltenen Organismen und Viren zu untersuchen. Das primäre Ziel ist in diesem Fall aber nicht die detaillierte Beschreibung einzelner Organismen, sondern die Identifizierung möglichst vieler der enthaltenen Organismen, um einen umfassenden Überblick über die Diversität der Population zu erhalten. Die NGS ergibt aufgrund ihrer Vorteile ein nahezu verzerrungsfreies Bild der untersuchten Population. Abbildung 3 zeigt schematisch den Ablauf der Untersuchung. Zunächst wird eine geeignete Probe gewonnen und in eine sogenannte DNA-Bibliothek überführt. Diese DNA-Bibliothek repräsentiert alle Bestandteile der gesamten Population gleichermaßen. Bei der Sequenzierung der DNA-Bibliothek wird eine große Anzahl (bis zu mehrere Millionen für eine Probe) kurzer DNA-Sequenzabschnitte gewonnen. Die Identifikation der in der Probe vorkommenden Organismen geschieht auf der Basis der ermittelten DNA-Sequenzen durch Vergleich mit bekannten Sequenzen bekannter Organismen. Dies ist möglich, weil die Genomsequenzen für bestimmte Organismen spezifisch sind. Daher geht man bei Ähnlichkeit zweier Sequenzen von einem ähnlichen Ursprung aus.

Die Metagenomanalyse wurde im Jahr 2011 eingesetzt, um die Ursache eines vielfach in Nordrhein-Westfalen beobachteten Rückganges der Milchleistung bei Kühen aufzuklären. Die Symptomatik der Erkrankung war unspezifisch, so dass gezielte Untersuchungen auf bestimmte, zu den Symptomen passende, Krankheiten ohne Ergebnis blieben. Daher wurden Proben von betroffenen Tieren in einer Metagenomanalyse untersucht. Bei der Analyse der mittels NGS erhaltenen Sequenzen gelang auch die Zuordnung von Sequenzen zu Viren, die bis dahin in Europa nicht aufgetreten waren. Diese Viren

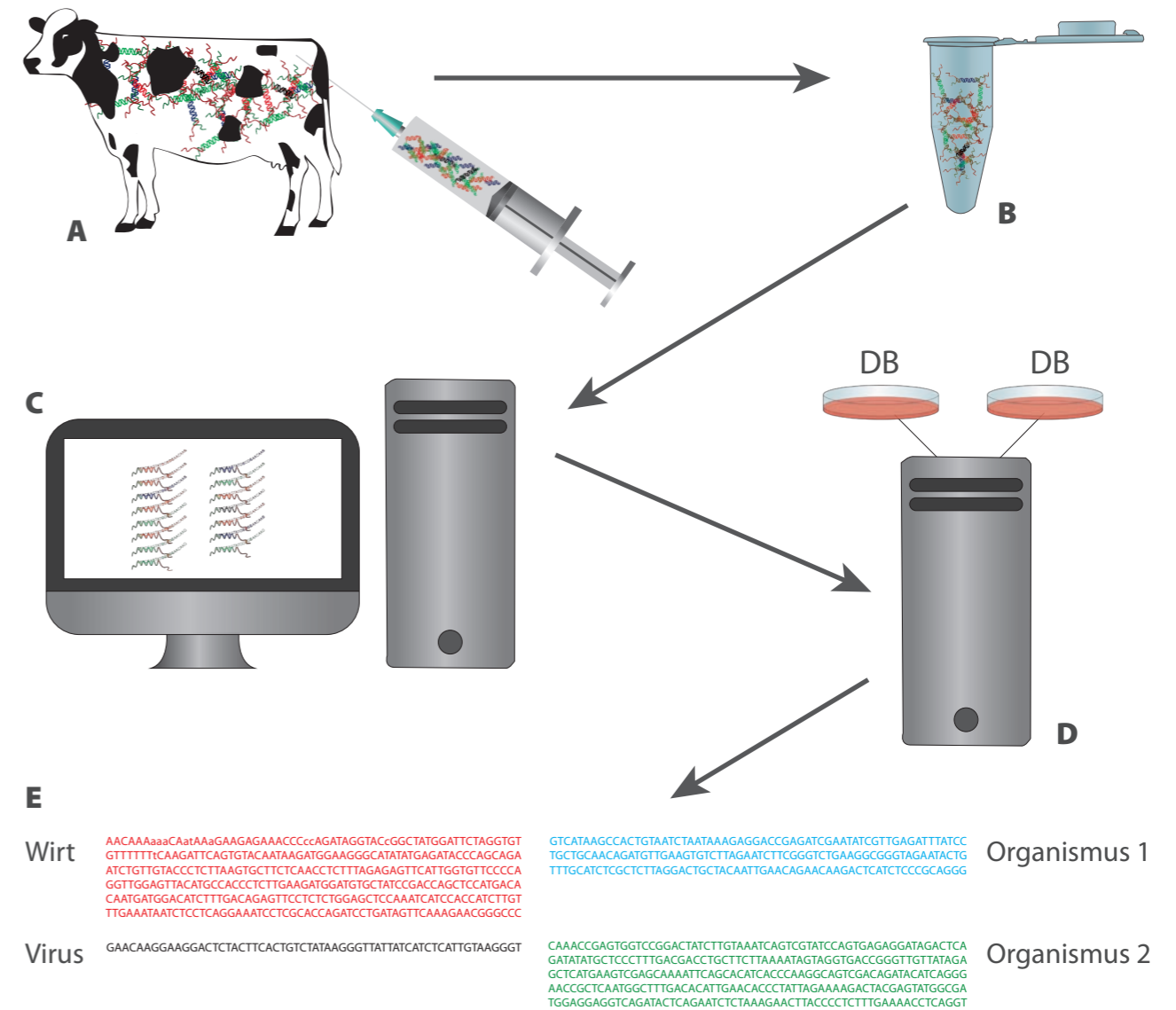


Abbildung 3: Übersicht über die wichtigsten Schritte einer diagnostischen Metagenomanalyse. Aus der Probe eines erkrankten Tieres werden alle Nukleinsäuren isoliert und daraus eine DNA-Bibliothek hergestellt (A, B). Anschließend wird die individuelle Basensequenz der einzelnen Moleküle der DNA-Bibliothek ermittelt (C). Der erzeugte Datensatz enthält bis zu einigen Millionen individueller DNA-Sequenzen. Diese werden dann mit Hilfe geeigneter bioinformatischer Prozeduren den nächstverwandten Taxa zugeordnet (D-E).

gehörten zu den Arten Akabane- und Shamonda-Virus, die zur Familie der Bunyaviren gehören. Viren der gefundenen Arten traten bis dahin nur in Asien, Australien und Afrika auf und verursachen eine wie in Nordrhein-Westfalen beobachtete Symptomatik. Aufgrund dieser eindeutigen Hinweise war es in der Folge sowohl möglich, das Virus in einer passenden Zellkultur zu kultivieren als auch entsprechende RNA-Moleküle in Proben weiterer infizierter Tiere nachzuweisen. Das neu entdeckte Virus wurde wie bei der Benennung neuer Viren üblich nach seinem Fundort Schmallenberg Virus genannt.



**Dr. Dirk Höper**  
Friedrich-Loeffler-Institut, Institut für Virusdiagnostik,  
Greifswald – Insel Riems

E-Mail: dirk.hoeper@fli.bund.de