

Daten speichern mit DNA

Franziska Konitzer

Das Molekül DNA ist nicht nur als Träger unseres Erbguts praktisch: Wissenschaftler erkunden, ob es sich auch als Speichermedium für Daten eignen könnte. Im Prinzip funktioniert das bereits.

Im Dezember 2020, an einem Montag, war Google auf einmal weg. Ein Totalausfall: kein Gmail, kein YouTube, kein Analytics, kein Drive, kein Maps. Grund dafür war wohl ein Problem mit Googles Speicher und den erlaubten Speicherquoten. Das ist eine technisch stark vereinfachte Version der Ereignisse, aber die technischen Details sind auch nicht so wichtig. Wichtig ist, dass an diesem Beispiel klar wird, dass Daten nicht etwas sind, das diffus in irgendwelchen Clouds herumwabert. Daten benötigen Speicherplatz, früher auf Floppy Disks und CDs, heute auf Festplatten und Magnetbändern in immer größeren Datenzentren.

Es ist ein Problem, das Geodäten kennen: Laserscans zum Beispiel benötigen Speicherplatz, und das nicht zu knapp. Oder man nehme das Erdbeobachtungsprogramm »Sentinel«. Im Januar 2019 überschritten die Archivdaten erstmals ein Speichervolumen von zehn Petabytes, das entspricht 10.000 Terabytes. Bis zum Ende 2020 sollte unser »digitales Universum« Schätzungen zufolge 44 Zettabytes umfassen, das entspricht 44 Milliarden Terabytes. Wo und wie will man alle diese Daten aufheben?

Gibt es überhaupt genug Platz für den Speicherplatz?

Die Frage ist durchaus ernst gemeint. Schließlich wird die Menschheit auch in Zukunft nicht weniger Daten generieren, eher im Gegenteil. Ein ideales Speichermedium sollte daher nicht nur platzsparend sein, sondern auch energieeffizient. Und haltbar. Ein derzeit übliches Medium für die Langzeitspeicherung von digitalen Inhalten ist immer noch das Magnetband. Es hält rund dreißig Jahre. Handelsübliche Festplatten sind durchschnittlich nach zehn Jahren am Ende.

Auf der Suche nach Speichermedien der Zukunft interessieren sich manche Forscher für das Speichermedium der Vergangenheit, sozusagen für das allererste Speichermedium überhaupt. Die Rede ist vom Molekül unseres Erbguts, der Desoxyribonukleinsäure, kurz DNA. Die Abfolge aus den vier Basen Adenin, Thymin, Guanin und Cytosin liefert den Bauplan für verschiedene Aminosäuren. Aus den Aminosäuren wiederum entstehen Proteine. Auf der Erde verwenden alle Lebewesen die DNA als Träger der Erbinformation.

Die Vorteile von DNA als Speichermedium

»2015 ist es einem Forscherteam gelungen, das Genom des Wollmammuts aus einem viertausend Jahre alten Knochen,

der im arktischen Eis gefunden wurde, zu sequenzieren«, erzählt Christopher Takahashi von der University of Washington. Damit ist schon einmal klar, dass DNA sehr viel länger haltbar sein kann als Magnetbänder und Festplatten. »Außerdem braucht man keine Energie, um sie aufrechtzuerhalten«, fügt Takahashi hinzu. Das wirklich Verföhreische an der DNA ist aber der Platz, den sie braucht bzw. nicht braucht. Man stelle sich ein riesiges Datenzentrum von der Größe eines überproportionierten Baumarkts vor – und dann einen zuckerwürfelgroßen DNA-Datenblock, der die gleiche Menge an Daten enthielte.

Den DNA-Würfelzucker gibt es in dieser Form noch nicht. Aber prinzipiell Daten in DNA speichern? Das geht. 2012 veröffentlichte ein Forscherteam einen Artikel im renommierten Fachmagazin »Nature«, in dem sie der Fachwelt erstmals darlegten, dass die Datenspeicherung mit DNA tatsächlich funktioniert. Sie speicherten ein HTML-Dokument, elf Bilddateien und ein JavaScript-Programm.

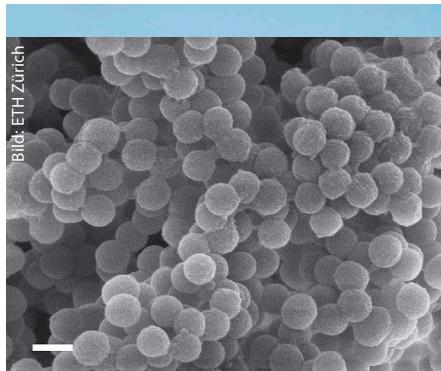
Vom Binärcode zu Nano-Glaskügelchen und wieder zurück

Prinzipiell kann man digitale Daten auf allem speichern, das einen Binärcode festhalten, also zwischen 0 und 1 unterscheiden kann. DNA-Moleküle mit ihren vier unterschiedlichen Basen sind somit für diese Aufgabe sogar überqualifiziert. »Man kann es beispielsweise so machen, dass man sagt: 00 ist Adenin, 01 ist Guanin und so weiter«, erzählt Robert Grass von der ETH Zürich. Er selbst arbeitet an der DNA-Datenspeicherung, seitdem er durch den oben erwähnten Fachartikel auf die Idee aufmerksam wurde.

Anschließend kann diese Basenabfolge synthetisiert werden. Dabei entsteht nicht gezielt ein einzelner, langer DNA-Strang, sondern sehr viele kurze Stränge, die bis zu 200 Basen lang sind, also theoretisch 400 Bits speichern können. »Wir erhalten viele solche identischen Stränge, zwischen hundert und tausend«, sagt Grass. »Eigentlich synthetisieren wir also etwas zu viel Information.«

Und wohin damit? Das ist das Spezialgebiet von Robert Grass. »Bei der Stabilität von DNA gibt es zwei Seiten«, erzählt er. Einerseits ist sie sehr stabil, wie am Beispiel des Genoms des Wollmammuts ersichtlich ist. »Andererseits ist DNA im Labor nicht so stabil«, sagt Grass. Trocken ist besser als nass. Kalt ist besser als warm. In Laboren wird die DNA üblicherweise in Gefriertruhen aufbewahrt und bei Bedarf gefriergetrocknet verschickt. Für die Datenspeicherung wäre das kein energieeffizientes Verfahren.

»Wir brauchen also etwas, das die DNA vor dem Zerfall schützt«, so Grass. Er hat dafür spezielle Nano-Glaspartikel entwickelt. Die DNA wird dabei in rund hundert Nanometer Durchmesser große Glaskügelchen eingeschlossen. Das Forscherteam, zu dem Christopher Takahashi gehört, experimentiert in Zusammenarbeit mit dem US-Unternehmen



Ein Rasterelektronenmikroskop hat diese Aufnahme gemacht: Es zeigt mit DNA gefüllte Nano-Glaskügelchen. Leider sieht man die DNA selbst auch bei sehr hohen Auflösungen nicht.



Sag mal »Hallo«: Dieses vollautomatische System zur DNA-Datenspeicherung wurde von Forschern der University of Washington zusammen mit dem Unternehmen Microsoft entwickelt. Damit ist es gelungen, das Wort »Hallo« vollautomatisch in DNA zu speichern und wieder abzurufen.

Bild: Microsoft/The University of Washington

Microsoft mit DNA, die auf dünnen Glasplatten getrocknet wird.

Die Daten aus der DNA hinausbekommen kann man dann bei Bedarf mit üblichen DNA-Sequenzierungsverfahren, wie sie in Laboren weltweit eingesetzt werden – wenn auch meist für medizinische Zwecke.

Aber ...

Diese Kurzvariante unterschlägt einige wichtige Details. Einerseits wäre da der Aspekt, dass kein Code der Welt perfekt geschrieben und ausgelesen werden kann. »Man braucht eine Fehlerkorrektur«, sagt Reinhard Heckel von der TU München, der zusammen mit Robert Grass an der DNA-Datenspeicherung arbeitet. Er hat dafür entsprechende Algorithmen entwickelt. Diese können sich zwar an bereits existierende Fehlerkorrektur-Algorithmen aus der klassischen Informationstechnik anlehnen, müssen aber auch auf die Besonderheiten der DNA angepasst werden, beispielsweise die Tatsache, dass man nur sehr kurze DNA-Sequenzen schreiben und lesen kann, und damit die Daten auf relativ kurzen Strängen vorliegen.

Auch müssen sich Forscher vorher Gedanken machen, wie die Daten ausgelesen werden sollen. So will es einem Start-Up vor einigen Jahren gelungen sein, die gesamte englischsprachige Wikipedia – 16 Gigabyte – auf DNA zu speichern. Was, wenn man aber nun nur einen Eintrag »lesen« möchte, ohne die gesamte DNA zu sequenzieren? »Prinzipiell ist auch gezieltes Auslesen möglich«, sagt Robert Grass. Dafür müssen die Stränge und Sequenzen aber vorher mit entsprechenden Indizes versehen werden.

Derzeit würde es aber auch noch eine ganze Weile dauern, bis man die Wikipedia wieder ausgelesen hätte. »Mit unseren derzeitigen Methoden würden wir einen Tag brauchen, um ein Gigabyte an Daten in DNA zu speichern und daraus verschieden Dateien wieder auszulesen«, erzählt

Christopher Takahashi. »Wenn man mehr Sequenzierungsgeräte hat, könnte man auch das gesamte Gigabyte an einem Tag auslesen.«

Der Datenspeicher der Zukunft?

Derzeit ist die DNA-Datenspeicherung also noch ziemlich langsam – und ziemlich teuer. Trotzdem ist das Interesse da. Neben wissenschaftlichen Forschergruppen wie denen von Reinhard Heckel und Robert Grass arbeiten mehrere Start-Ups an entsprechenden Verfahren. Unter den großen Softwareunternehmen ist es vor allem Microsoft, das Interesse an der DNA-Datenspeicherung hat und das deshalb mit den Forschern der University of Washington zusammenarbeitet. Dieser Kollaboration ist es bereits vor einiger Zeit gelungen, ein vollautomatisches System zu demonstrieren, also eines, das keine Labortechniker mehr für die Zwischenschritte benötigt.

Entwicklungsbedarf sehen Reinhard Heckel und Robert Grass vor allem bei der Synthese der DNA. Während das Auslesen von DNA im Zuge der Entschlüsselung des menschlichen Genoms mit viel Zeit und Geld bedacht wurde – das Human Genome Project wurde im Jahr 2003 abgeschlossen – ist bei der Synthese bislang kein ähnlicher Schritt erfolgt. »Es gibt viele Anwendungen, bei denen man DNA lesen will, aber nicht so viele Anwendungen für das Schreiben«, erklärt Reinhard Heckel.

Christopher Takahashi hält die Datenarchivierung für ein vielversprechendes Einsatzgebiet, denn hier kann die DNA all ihre Vorteile voll ausspielen. Er sagt: »Wir müssen nun einmal auf chemische Prozesse statt auf Elektronen warten. Deshalb halte ich es für unwahrscheinlich, dass die DNA jemals für relative schnelle Speicheranwendungen in PCs und in Servern genutzt wird, für die wir heutzutage Speichersticks und Festplatten nutzen.«

Kontakt: f.konitzer@gmail.com