

24.01.2013 01:19

Wissenschaftler verfeinern Methoden zur Datenspeicherung in DNA-Form

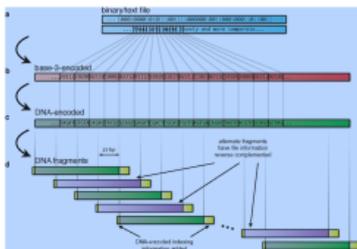
Wissenschaftler am European Bioinformatics Institute (**EMBL-EBI** [<http://embl.de/>]) beschreiben eine Methode, mittels der sich Daten digital in Form von DNA-Strängen zuverlässig speichern lassen. Die Idee, DNA-Moleküle als Speichermedium zu nutzen, ist nicht neu. Verlockend sind die DNA-Stränge wegen der hohen Datendichte. Die Autoren schätzen, dass sich "mindestens 100 Millionen Stunden hochauflösendem Videofilm in einer Menge DNA speichern lässt, die in eine Kaffeetasse passt". Die Methode, die in der Fachzeitschrift **Nature** [<http://www.nature.com/>] **vorgestellt** [<http://www.nature.com/nature/journal/vaop/ncurrent/full/nature11875.html>] wurde, erscheint nun deshalb attraktiv, weil sie die Fehlerrate des Verfahrens sehr stark senkt.

Die Menge an digitaler Information weltweit entspricht momentan ungefähr drei Zettabyte (zum Vergleich: 1 Zettabyte entspricht einer 1 mit 21 Nullen) und der Zuwachs an neuen digitalen Inhalten stellt die Archivare vor Herausforderungen. Festplatten sind teuer, während preiswerte Archivierungsmaterialien wie Magnetbänder innerhalb von Jahren zerfallen können. Dies ist vor allem für die Lebenswissenschaften ein Problem, denn die Speicherung riesiger Datenvolumen – einschließlich DNA-Sequenzen – macht einen wesentlichen Teil des wissenschaftlichen Datenbestandes aus.

"Wir wissen bereits, dass DNA ein äußerst robustes Speichermedium ist, denn wir können sie beispielsweise aus den Knochen des Wollmammuts extrahieren, die mehrere zehntausend Jahre alt sind und sie enthält immer noch sinnvolle Informationen," erläutert Nick Goldman vom EMBL-EBI. Fachleute dürften dabei vermutlich stutzen, denn die Kollegen lassen in der **Pressemittteilung** [http://embl.de/aboutus/communication_outreach/media_relations/2013/130123_Hinxton/index.html] unter den Tisch fallen, dass solche DNA nicht mehr am Stück ist und die teils sehr kleinen Fragmente aufwendig gereinigt und mühevoll in Reihe gebracht werden müssen, bevor man ihre Sequenz entziffert hat.

Kurze Fragmente sind aber zunächst kaum als Träger von großen Datenmengen vorstellbar. Nur wenn trocken und kühl gehalten, lässt sich DNA lange lagern. Dennoch zerbrechen lange Stränge auch dabei leicht zu Bruchstücken. Umweltfaktoren wie Feuchtigkeit, Wärme und Säure verstärken diesen Vorgang, sodass Jahrtausende alte DNA (ancient-DNA) meist nur noch aus Trümmern von unter 200 Bausteinen besteht. Diesen Effekt haben die EMBL-Forscher jedoch durchaus im Sinn gehabt.

Dabei ist das Lesen solcher kurzen oder auch längerer Stränge, die aus den vier Bausteinen A, T, G und C bestehen, relativ unkompliziert. Bisher war es das Schreiben, das dem Speichern von digitalen Informationen in Form von DNA-Strängen im Weg stand. Mit aktuellen Methoden lassen sich lediglich kurze Stränge beliebiger Sequenz künstlich herstellen und sowohl das Lesen als auch das Schreiben sind fehleranfällig, wenn zwei oder mehr gleicher DNA-Bausteine aufeinander folgen (z. B. TTTTT). Nick Goldman und Ewan Birney, stellvertretender Direktor am EMBL-EBI, haben diese Probleme im Prinzip mit zwei Verfahren gelöst.



[<http://www.heise.de/newsticker/meldung/Wissenschaftler-verfeinern-Methoden-zur-Datenspeicherung-in-DNA-Form-1790445.html?view=zoom;zoom=1>] Byte-Strings, beispielsweise von einer Datei, werden mittels Huffman-base-3-encoding in Zeichen-Strings ohne Zeichenwiederholungen umgewandelt; die Länge der Nutzdaten wird auf 20 Zeichen begrenzt. 

Im Kern wandeln sie Byte-Strings, beispielsweise von einer Datei, mittels eines Huffman-Algorithmus in Base-3-Code und anschließend in DNA-Code um, sodass Zeichenwiederholungen ausgeschlossen sind. Zusätzlich begrenzen sie die Länge der Nutzdaten. Damit nun eine beliebig lange Datei als DNA-Strang enkodiert werden kann, setzen sie überlappende Stränge ein, die mittels zusätzlicher Zeichen indiziert sind, um die Position der Fragmente innerhalb der gesamten Abbildung beim Lesevorgang wieder ermitteln zu können. Die Kodierung lässt nur im Prinzip keine Wiederholungen zu – bei der Synthese kann in sehr seltenen Fällen dennoch etwas schief gehen. Weil sich die

[http://www.heise.de/newsticker/meldung/Wissenschaften/Verfahren-um-Fragmente-überlappen-(das-Informationen-Methoden-zur-Datenspeicherung-in-DNA-Form-1790445.html?view=zoom;zoom=1]

Fragmente überlappen (das Informationsabbild ist redundant), kann man die Fehler aber leicht erkennen, und so die fehlerhaften Zeichen beim

Lesevorgang korrigieren.

Erste Tests hat das Verfahren bereits bestanden. Birney und Goldman haben einige Dateien mit ihrem Verfahren in die Zeichenfolge der DNA übersetzt und das Resultat vom kalifornischen Unternehmen Agilent synthetisieren lassen. "Das Ergebnis sieht aus wie eine winzige Menge Staub," erklärte Emily Leproust von Agilent. Die DNA-Fragmente haben dann EMBL-Mitarbeiter sequenziert und anschließend mit umgekehrtem Verfahren fehlerfrei dekodiert.

"Wir haben einen fehlertoleranten Code entwickelt, der sich einer molekularen Form bedient, von der wir wissen, dass sie unter günstigen Bedingungen 10.000 Jahre oder länger halten kann", so Nick Goldman. "Solange es jemanden gibt, der den Code kennt, wird man ihn lesen können – falls man im Besitz einer Maschine ist, die DNA lesen kann."

Das Verfahren steht damit natürlich erst am Anfang, denn wie üblich gibt es noch eine Reihe praktischer Probleme zu lösen, wie die Autoren selbst einräumen. Aufgrund ihrer Dichte und Langlebigkeit erscheint die DNA-Synthese ohnehin attraktiv. Das neue Verfahren bringt nun eine Verwertung in Reichweite. Die nächsten Schritte werden sein, die Kodierung zu perfektionieren und die praktischen Probleme zu lösen, um das DNA-Speichermodell wirtschaftlich rentabel zu machen. ([dz \[mailto:dz@ct.de\]](mailto:dz@ct.de))

Permalink: <http://heise.de/-1790445> [http://heise.de/-1790445]



Auch auf heise online:

- Voreingestelltes WPA2-Passwort bei Belkin-Routern leicht zu berechnen**
- Bundesregierung streitet über Vorratsdatenspeicherung vs. Quick Freeze**
- Audio-Captchas: Die meisten lassen sich knacken**
- Roboter aus DNA-Strängen**
- Google lässt Map/Reduce patentieren**
- Google will Speicherung von Suchanfragedaten nicht weiter begrenzen**

Mehr zum Thema **Genetik** [http://www.heise.de/thema/Genetik] **DNA** [http://www.heise.de/thema/DNA]