Half a Million DVDs of Data Stored in Gram of DNA

By ScienceNow
01.24.13
10:30 AM

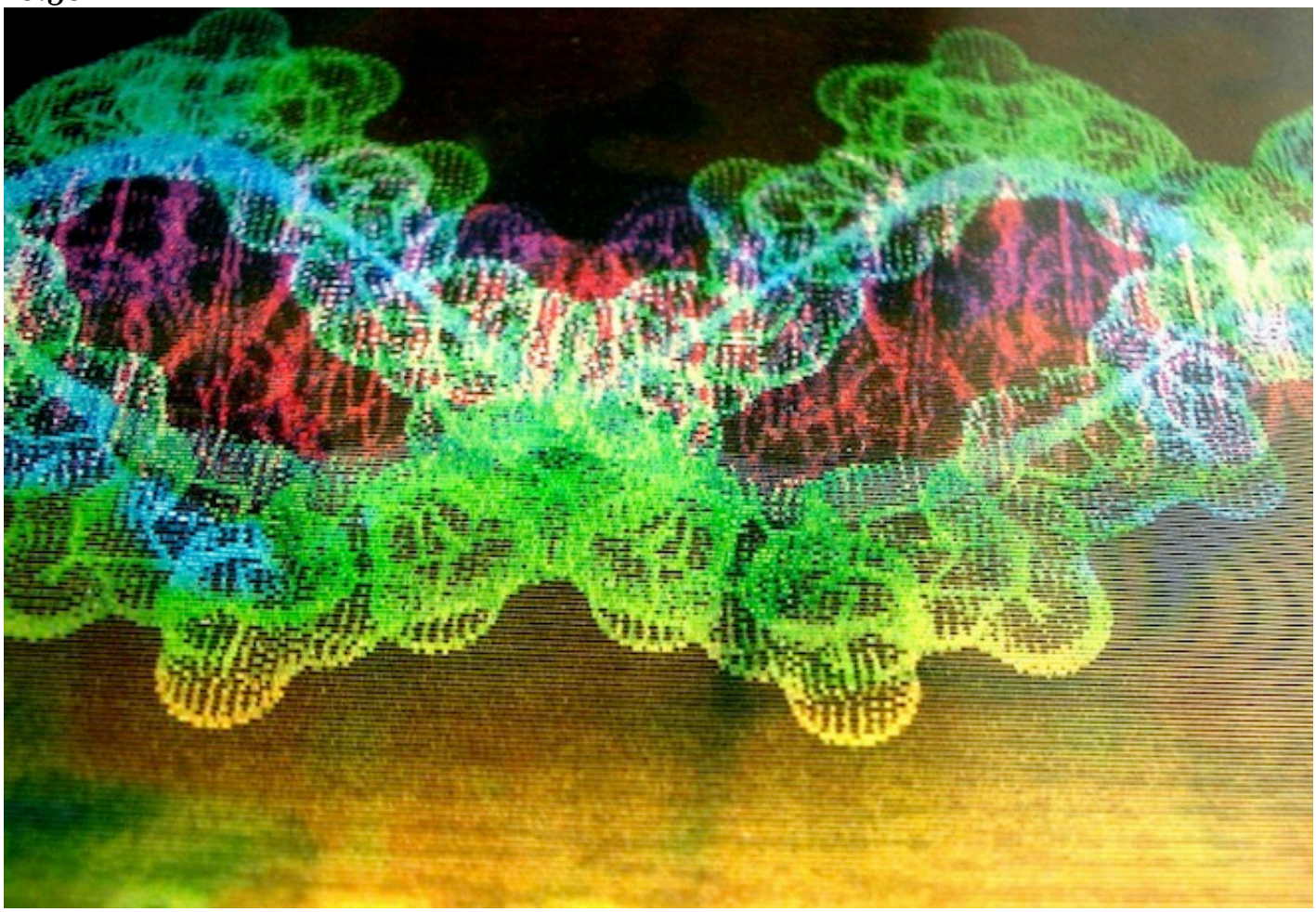


Image: Steve Jurvetson/Flickr

By Robert F. Service, *Science*NOW

Paleontologists routinely resurrect and sequence DNA from woolly mammoths and other long-extinct species. Future paleontologists, or librarians, may do much the same to pull up Shakespeare's sonnets, listen to Martin Luther King Jr.'s "I have a dream" speech, or view photos. Researchers in the United Kingdom report today that they've encoded these works and others in DNA and later sequenced the genetic material to reconstruct the written, audio, and visual information.

The new work isn't the first example of large-scale storage of digital information in DNA. Last year, researchers led by bioengineers Sriram Kosuri and George Church of Harvard Medical School reported that they stored a copy of one of Church's books in DNA, among other things, at

a density of about 700 terabits per gram, more than six orders of magnitude more dense than conventional data storage on a computer hard disk. Now, researchers led by molecular biologists Nick Goldman and Ewan Birney of the European Bioinformatics Institute (EBI) in Hinxton, U.K., report online today in *Nature* that they've improved the DNA encoding scheme to raise that storage density to a staggering 2.2 petabytes per gram, three times the previous effort.

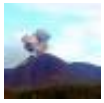
To do so, the team first translated written words or other data into a standard binary code of 0s and 1s, and then converted this to a trinary code of 0s, 1s, and 2s—a step needed to help prevent the introduction of errors. The researchers then rewrote that data as strings of DNA's chemical bases: As, Gs, Cs, and Ts. At the storage density achieved, a single gram of DNA would hold 2.2 million gigabits of information, or about what you can store in 468,000 DVDs. What's more, the researchers also added an error correction scheme, encoding the information multiple times, among other tricks, to ensure that it could be read back with 100% accuracy.

Beyond demonstrating DNA's superlative information storage abilities, Goldman, Birney, and their colleagues also asked when such a technology might be worth implementing. Institutions such as the Large Hadron Collider, a particle accelerator in Geneva, Switzerland, produce on the order of 15 petabytes of data each year. So the need for vast archival storage is growing rapidly. Now, such institutions commonly archive data by storing it on magnetic tape. Keeping that data safe over many decades requires rewriting it at regular intervals, adding to the cost of preservation. DNA, on the other hand, can be stable for thousands of years if kept in a cool, dry place. Goldman also notes that the costs of synthesizing DNA, which corresponds to writing the code, as well as sequencing, or reading out the code, are dropping fast. According to the EBI researchers, at current rates, DNA data storage is now cost-effective for only data that need to be archived for 600 years or more. But if the costs of DNA synthesis—currently the most expensive part of the enterprise—drop 100-fold, that break-even number would drop to about 50 years.

Harvard's Kosuri calls the latest study "good work." But he says that cost won't be the hitch. For starters, he notes, once you write a batch of data in DNA, you can't change it or rewrite over it, as is often done with other data storage technologies. And you can't access any particular piece of information, but rather must sequence large swaths of DNA to find what you've archived.

So even though DNA's data storage densities are off the charts, it may still be worth putting those family photos on a DVD for now.

This story provided by Science*NOW, the daily online news service of the journal* Science.



doran Volcano Update: El Reventador Producing Lava Flows and Ash Fall, urahua Settles Down

How Do You Get Candy to the Space Station?

Back to top

Tweet 541

Reddit Digg Stumble Upon Email

19 comments · 529 reactions

★ 0

Leave a message...

Discussion | Community | Share

Tim Harrington · 2 days ago

I'd like to see someone encode a bunch of pirated movies into DNA form, splice the sequences into genetically engineered rats, and release them in a pest-ridden metropolitan area. Presto: the MPAA picks up the tab for all future extermination costs.

26 · Reply · Share ›

CrackerJacker ➔ Tim Harrington · 2 days ago

If they also engineer them to glow in the dark, could they project the movies as well?

6 · Reply · Share ›

ThatsHowISeeIt ➔ Tim Harrington · 2 days ago

Except I don't think rats hunt rats :)

7 | 1 · Reply · Share ›

RobertPiven ➔ Tim Harrington · 2 days ago

like Eric explained I am inspired that any one able to make $6603 in four weeks on the network.

0 | 1 · Reply · Share ›

NuprinBoy · 2 days ago

No question the density and stability make DNA perfect for archival storage. But are there chemical reaction speed limitations for how fast DNA data could be written or read? I wonder how long it took to encode the book into DNA and how long it took to read it back.

3 · Reply · Share ›

NewGawker · 2 days ago

How long before we can copy all the information in our brain and put it into a robot

1 · Reply · Share ›

FelixKrull · 2 days ago

Half a Million DVDs...?

I hate imperial units.

1 · Reply · Share ›

lakshwadeep · 2 days ago

Theoretically, you could read parts of the DNA by translating sections into RNA, using different promoters and things like sigma factors or DNA methylation/demethylation.

1 · Reply · Share ›

P.F. Bruns · 2 days ago

I had that done. Now I have decades of "Lucy" reruns stuck in my head.

Hey, wait! If we do this with an Australopithecus, we could have an "I Love 'Lucy'" marathon!

1 · Reply · Share ›

scott taggart · 2 days ago

a research grant looking for a problem. this has no practical use or value (note the sited 50-100 year possible cost justification). and, relating it to actual cellular dna is just dumb. I have no idea how much dna is in a cell but it's probably on the order of a few micrograms?

0 · 4 · Reply · Share ›

JeramieH3 → scott taggart · 2 days ago

> a research grant looking for a problem

You're right, 640KB should be enough for everybody.

6 · Reply · Share ›

Rohit Singh → scott taggart · 2 days ago

The point is not what use it has. Science is not done for use, but for the sake of doing it.

And who knows? Maybe someday there will be a use for this that we can't even imagine right now.

1 · Reply · Share ›

ERSorensen → scott taggart · a day ago

Yeah, this is just like those idiots who decided we should go to the moon. Non-targeted research has never led to anything useful. Except cell phones... and the internet...

0 · Reply · Share ›

m12345 · a day ago

I can see there is lots wrong with this story and the underlying research.

You cannot and will never be using DNA to store information, you might as well say I can store my information as farts in a jar.

0 · 1 · Reply · Share ›

ERSorensen ↗ m12345 • a day ago

Please, don't talk shit unless you're willing to back it up with some facsimile of a fact or argument. You're just making people scroll farther.

2 ^ v · Reply · Share ›

Ryan Walsh • a day ago

So when do we start running a "search" for previously stored data that may be out there...maybe efforts like SETI are just misplaced, we may be looking in the wrong places.

0 ^ v · Reply · Share ›

david gibbs • a day ago

All this code just happened by chance? Or was it designed by a Creator? You have an AMAZING brain. YOU DECIDE.

0 ^ v · Reply · Share ›

simmortal • a day ago

This is fascinating research. Like all advances in science, they may not have immediate practical uses and in fact be worse than existing techniques but, if it is a disruptive technology then it will improve over time and become the "new norm" for how we encode data for long term storage.

Remember Asimov's foundation and the encyclopedia galactica? Perhaps this is not so Scifi after all..

0 ^ v · Reply · Share ›

Kermit Frazier • 2 days ago

Inject "War and Peace" into an ovum and what will hatch???

0 ^ v · Reply · Share ›

Comment feed Subscribe via email