

Our site uses cookies to deliver relevant messages during your visit, such as presenting our latest offers and allowing easier access for subscribers. To receive the best experience, please allow cookies. [View our cookies policy](#), or [manage your cookies](#).

Allow

The Economist

Subscribe



Virtuous spirals

DNA may soon be used to store computer data

But it is not just a question of base pairs becoming bits



123RF

[Print version](#) | [Science and technology](#) >

Oct 20th 2018



DNA IS THE oldest information-storage system known. It predates every other, from pencil and paper to computer hard drives, by billions of years. But attempts to employ it to store data

generated by people, as opposed to data needed to bring those people (and every other living thing) into being in the first place, have failed.

The reason is not so much technological difficulty as cost. Encoding a single gigabyte in DNA would run up a bill of several million dollars. Doing so on a hard drive costs less than a cent. Catalog, a biotechnology firm in Boston, hopes to bring the cost of DNA data-storage below \$10 per gigabyte. That is still on the pricey side. But for really large storage requirements a second ratio also comes into play: gigabytes stored per cubic metre.

Hard drives take up space. Their storage ratio is about 30m gigabytes per cubic metre. Catalog's method can store 600bn gigabytes in the same volume. For organisations such as film studios and particle-physics laboratories, which need to archive humongous amounts of information indefinitely, the ratio of the two ratios, as it were, may soon favour DNA.

The obvious temptation when designing a DNA-based storage system is to see the ones and zeros of binary data and the chemical base pairs (AT and GC) of deoxyribose nucleic acid as equivalent, and simply to translate the one into the other, with each file to be stored corresponding to a single, large DNA molecule. Unfortunately, this yields molecules that are hard for sequencing machines to read when the time comes to look at what data the DNA is encoding. In particular, there are places in computer data that consist of long strings of either ones or zeros. DNA sequencers have difficulty when faced with similarly monotonous strings of base pairs.

Catalog has taken a different tack. The firm's system is based on 100 different DNA molecules, each ten base pairs long. The order of these bases does not, however, encode the binary data directly. Instead, the company pastes these short DNA molecules together into longer ones. Crucially, the enzyme system it uses to do this is able to assemble short molecules into long ones in whatever order is desired. The order of the short molecular units within a longer molecule encodes,

according to a rule book devised by the company, the data to be stored. Starting with 100 types of short molecule means trillions of combinations are possible within a longer one. That enables the long molecules to contain huge amounts of information.

The cost savings of Catalog's method come from the limited number of molecules it starts with. Making new DNA molecules one base pair at a time is expensive, but making copies of existing ones is cheap, as is joining such molecules together. The Catalog approach also means it is harder for data to be misread. Even if a sequencing machine gets a base or two wrong, it is usually possible to guess the identity of the ten-base-pair unit in question, thus preserving the data.

Catalog's combinatorial approach does mean that more DNA is needed per byte stored than other DNA-based methods require. This increases both the time and the cost of reading it to recover the stored data in electronic form for processing. Overall, though, the method promises to have significant advantages over its predecessors.

The next task is to translate that promise into reality. To this end, Catalog is working with Cambridge Consultants, a British technology-development firm, to make a prototype capable of writing about 125 gigabytes of data to DNA every day. If this machine works as hoped (it is supposed to be ready next year), the company intends to produce a more powerful device, able to write 1,000 times faster, within three years. The second age of DNA information storage may then, at last, begin.

[Print version](#) | [Science and technology](#) >

Oct 20th 2018



 [About The Economist](#)



Luca D'Urbino

Aussie rules

What the world can learn from Australia

It is perhaps the most successful rich economy

LEADERS >



The man who spoke out

Obituary: Jamal Khashoggi

OBITUARY >



Poverty in California

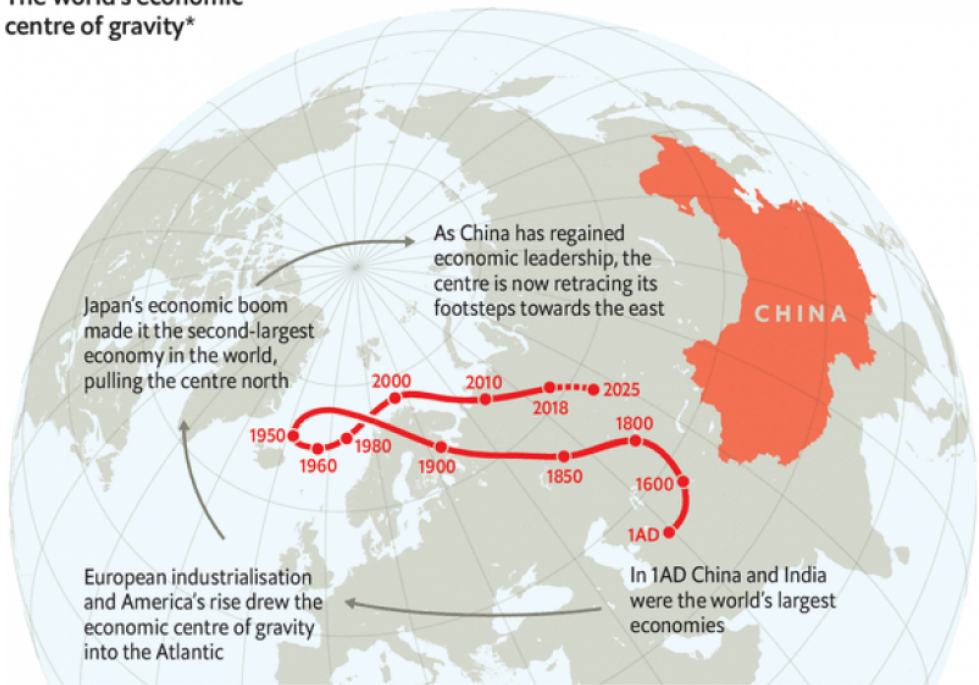
Why one of America's richest states is also its poorest

UNITED STATES >

A new hegemon

The Chinese century is well under way

The world's economic centre of gravity*



*The economic centre of the globe is calculated using an average of countries' locations weighted by their GDP

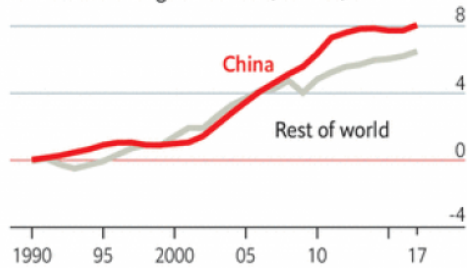
GDP per person

% change
1990-2017
1,000

China's GDP per person grew by 903% from 1990 to 2017

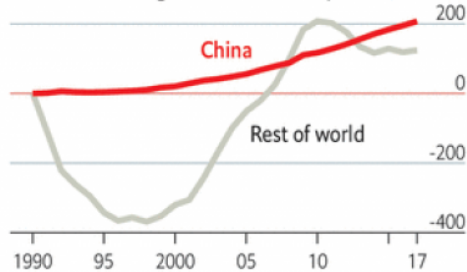
CO₂ emissions

Cumulative change since 1990, tonnes, bn

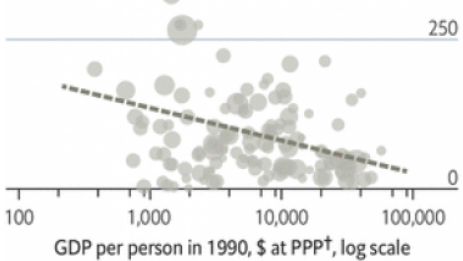


Military spending

Cumulative change since 1990, 2016 prices, \$bn

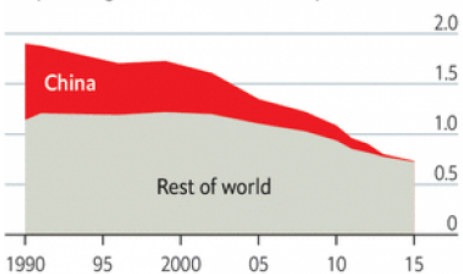


Circle size = population, 2017



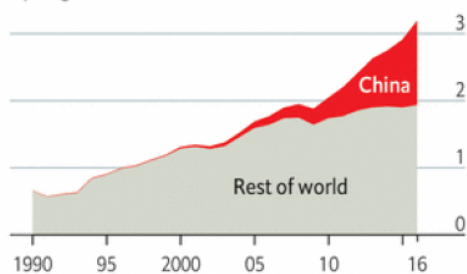
Poverty

People living on less than \$1.90 a day†, bn



Patent applications

By origin, m



Sources: Economist Intelligence Unit; Global Carbon Project; Maddison Project Database; SIPRI; World Bank; World Intellectual Property Organisation; The Economist

†Prices at 2011 purchasing-power parity

[Subscribe](#)

[Contact us](#)

[Help](#)

[Open Future](#)

Keep updated



[About The Economist](#)

[Advertise](#)

[Reprints](#)

[Careers](#)

[Media Centre](#)

[Terms of Use](#)

[Privacy](#)

[Cookie policy](#)

[Manage cookies](#)

[Accessibility](#)

[Modern Slavery Statement](#)

Copyright © The Economist Newspaper Limited 2018. All rights reserved.