# Speed reading from storage

## DNA can hold information securely but we need ways to quickly find the data

S ANANTHANARAYANAN

Deoxyribonucleic acid, the molecular code found in the nuclei of living cells, is the most compact and hardy data storage system we have encountered. The nucleus of microscopic cells contains a giant molecule, billions of units long and compactly folded, with the code for all the proteins that the cells of an organism would produce, and hence all the characteristics of the organism itself. The code then survives, essentially unaltered through the many times that cells divide, and the DNA molecule has been found reasonably intact in frozen animal remains that are even thousands of years old.

It has been estimated that a gram of DNA could store 215 petabytes, or 215 million gigabytes of data. In comparison, the best computer data recording systems store data in terabytes, or thousands of gigabytes and may weigh half a kilogram. And then, computer data storage devices degrade with age and what is worse, the technology that is used to create or read data from them keeps getting outdated. An efficient method to write digital data into a DNA-like molecular record and then retrieve the data would be a vast improvement in computing and data management.

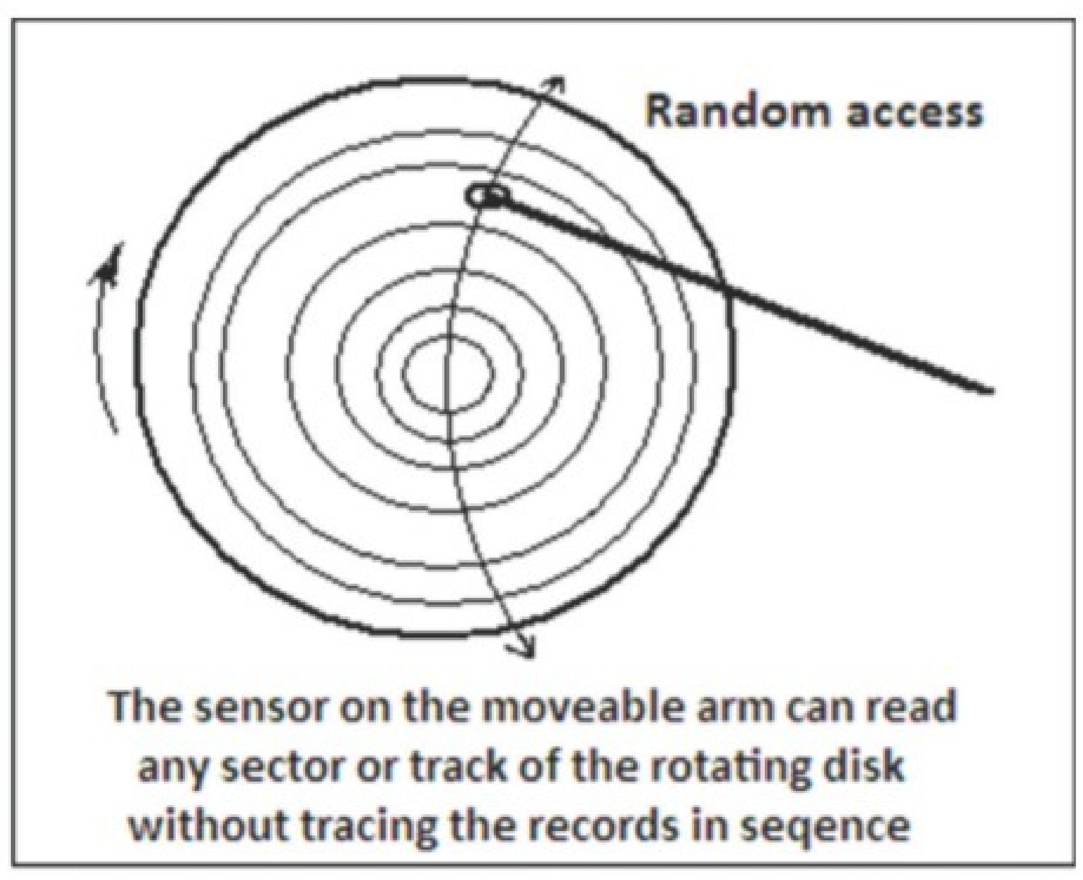
While methods to write digital data into the DNA molecule have been developed, reading a reasonable quantity of data, once it is stored inside a DNA molecule is still a complicated process. This is the feature that is standing in the way of DNA storage growing into a practical strategy for preserving and handing large data. Lee Organick, Siena Dumas Ang, Yuan-Jyue Chen, Randolph Lopez, Sergey Yekhanin, Konstantin Makarychev, Miklos Z Racz, Govinda Kamath, Parikshit Gopalan, Bichlien Nguyen, Christopher N Takahashi, Sharon Newman, Hsing-Yeh Parker, Cyrus Rashtchian, Kendall Stewart, Gagan Gupta, Robert Carlson, John Mulligan, Douglas Carmean, Georg Seelig, Luis Ceze and Karin Strauss, from the University of Washington and Microsoft Research, Washington, describe, in their paper in the journal, Nature Biotechnology, an orders-of-magnitude improvement in the recording and speed of retrieval of sizeable data from a DNA record.

The DNA molecule consists of a pair of chains, or backbones, of units, called nucleotides or bases with "side chains" of four kinds of molecular groups along its length. Information is coded by the sequence of the four kinds of side chain groups — they are called, C, G, A and T and in the DNA, groups of three consecutive bases, each with one of four forms of side chain, code for 20 amino acids, the building blocks of proteins. Series of triads thus code for series of amino acids and hence for different proteins.

There is a rule that a C can pair only with a G in the complementary chain and an A with T. This rule specifies the order of bases in one chain once the order in the other is fixed and this is what enables either chain to create a fresh complementary chain when the chains separate in cell division.

Digital data, which consists of series of 1's and 0's, can be coded in a similar fashion in a chain that can fit into the DNA structure. Methods have been developed to synthesise these chains and there is a technique to snip a DNA molecule at a particular spot to insert the portion that codes the digital data. DNA in a living cell can then hold the data record, which would get replicated every time the cell divides. Reading the sequence of units in the DNA, or DNA sequencing, is now well developed and the digital record can hence be recovered.

The trouble, however, is with this last part of the process and the trouble is that the whole DNA needs to be sequenced before the digital record part can be extracted. And then, the whole record has to be read to locate a portion, which is of interest. Early computer records, which were created in spools of magnetic tape, had this same feature, of being placed, one after the other, along the length of the tape. Particular records were identified by a "header" and an end marker, or by segments of the tape, but the tape had to be run from the beginning, till the required record was found.




Running through the length of a spool of tape consumed power and time and large part of computing time was spent in the "sequential search" for the many items of data that could be required even for a small computation. A great development was with the floppy disk and the hard disk, which could be divided into tracks and sectors and a record could be directly accessed by its address, or known location on the disk. The disk is kept spinning, to rapidly scan the sectors and radial detectors can pick up data from the different tracks hundreds of times a second. This method of access to data is known as random access and was probably a larger step in increasing computer speeds than improvements in the processors.

What the researchers at Washington report in Nature is the equivalent in DNA record, of enabling location of a specific portion without having to sequence the whole DNA-like molecule. There are other issues too, that limit the value of the DNA for large-scale digital recording, the authors say. One such is the frequency of errors when the record is first written in. The way this is handled is by creating multiple copies of the data to write, so that there is "redundancy".

This method, however, consumes resources in creating, and ideally verifying the copies and the overheads of the process. In computer storage, there are devices, like the use of a "check digit" or coding that enables recovery till more than a certain number of errors have occurred. The Washington researchers report an improved method of coding that substantially reduces the extent of redundancy, and hence the complexity of data preparation and writing effort that is required.

The main improvement, however, is in the random, or non-sequential access to records that has been made possible. This was done by creating a library of DNA stretches called "primers", which were made with sufficient mutual differences so that they could be readily told apart. The digital data to be recorded is first prepared with a degree of redundancy built in and formed into distinct segments according to a scheme. The segments are then converted into DNA sequences and a "primer", taken from the library is attached to both ends of each sequence. This unique primer is the feature that would allow random access to a particular record, from a soup of the segments of all the data recorded. The DNA sequences are then put together as DNA strands and can be dried and preserved.

Retrieving data involves "rehydrating" the DNA material and carrying out sequencing of the bits of DNA. A four-stage process is employed to filter out dissimilar strands of DNA and the portion of interest is separated by a process of iteration.

The method was tried out with 200 MB of data of different kinds and the entire data could be recovered. The paper says that the most of the DNA sequences had to be read just five times. "This is half as much as the minimum coverage ever reported in decoding digital data from DNA," the paper says.

The paper notes that as data storage in DNA needs synthetic DNA there is need for this industry to rack up to meet data demands. This should be possible, as the quality demands of data applications are not as high as those of the life sciences. While DNA storage, because of its long-term durability, could be interesting even at the current stage, increasing throughput and reduction of cost is expected, the paper says.

The writer can be contacted at response@simplescience.in

PLUS POINTS

### All are welcome

Humans will actually be very happy to meet aliens, according to new research.

Hollywood might have led us to expect that an alien invasion would be greeted with tanks and guns but some of the first serious research into how people would feel about meeting extraterrestrials shows that we would be far more positive than you might think.

To explore how people might react, scientists had them read news reports about a number of potential alien discoveries. They then found out how they were feeling after doing so — exploring whether they seemed excited or fearful -- using a special software programme that could analyse people's language.

Their reactions were very positive, the researchers found. Subjects said that they would be excited even if the life was primitive or happening elsewhere — as any encounter with extraterrestrials is likely to actually be.

Further work had people describe how they thought microbial life would be found on another planet, and how they would react. Again, the software found that people were more positive about that news than negative.

And additional research looked at one of the most famous pieces of news about the potential of alien life in recent weeks. Scientists had people read about the suggestion that a rock flying past Earth — known as Oumuamua — could in fact be an alien spacecraft, in news that sounded like something out of science fiction.

It found that people were very excited about that possibility, despite the fact it could mean aliens were headed towards Earth.

The Independent

### Little nurses

African Matabele ants dress the wounds of comrades injured during hunting raids and nurse them back to health, according to an "astonishing" discovery reported last week.

After collecting their wounded from the battlefield and carrying them back home, nestmates become medics, massing around patients for "intense licking" of open wounds, according to a study in the journal, Proceedings of the Royal Society B. This behaviour reduces the fatality rate from about 80 per cent of injured soldiers to a mere 10 percent, researchers observed.

The study claimed to be the first to show such nursing behaviour in any non-human animal. "This is not conducted through self-medication, as is known in many animals, but rather through treatment by nestmates which, through intense licking of the wound, are likely able to prevent an infection," said study co-author Erik Frank.

He contributed to the research when he was at the Julius Maximilian University of Wuerzburg in Germany, and continues his work at the University of Lausanne in Switzerland.

Matabeles, one of the world's largest ant species, are fierce warriors and attack even humans with their ferocious bite. Named after Southern Africa's feared Matabele warrior tribe, the insects hunt termites bigger than themselves, attacking their feeding sites in column formations of 200 to 600 individuals. This hunting method causes many ants to get hurt, often having their legs bitten off by termite soldiers.

In the aftermath of fighting, while some of the ants return home with their dead termite prey, others scuttle around the battlefield looking for injured colleagues. "After the battle, injured ants call for help with pheromones," a chemical SOS signal produced in a special gland, said Frank. Rescuers use their strong jaws to pick up the wounded and drag them back to the nest for treatment.

Astoundingly, warriors that are too severely injured — missing five of their six legs, for example — signal rescuers not to bother picking them up. Unlike peers that are less seriously hurt and lie still to make their saviours' job easier, terminally-wounded ants lash out and struggle until rescuers give up and move on.

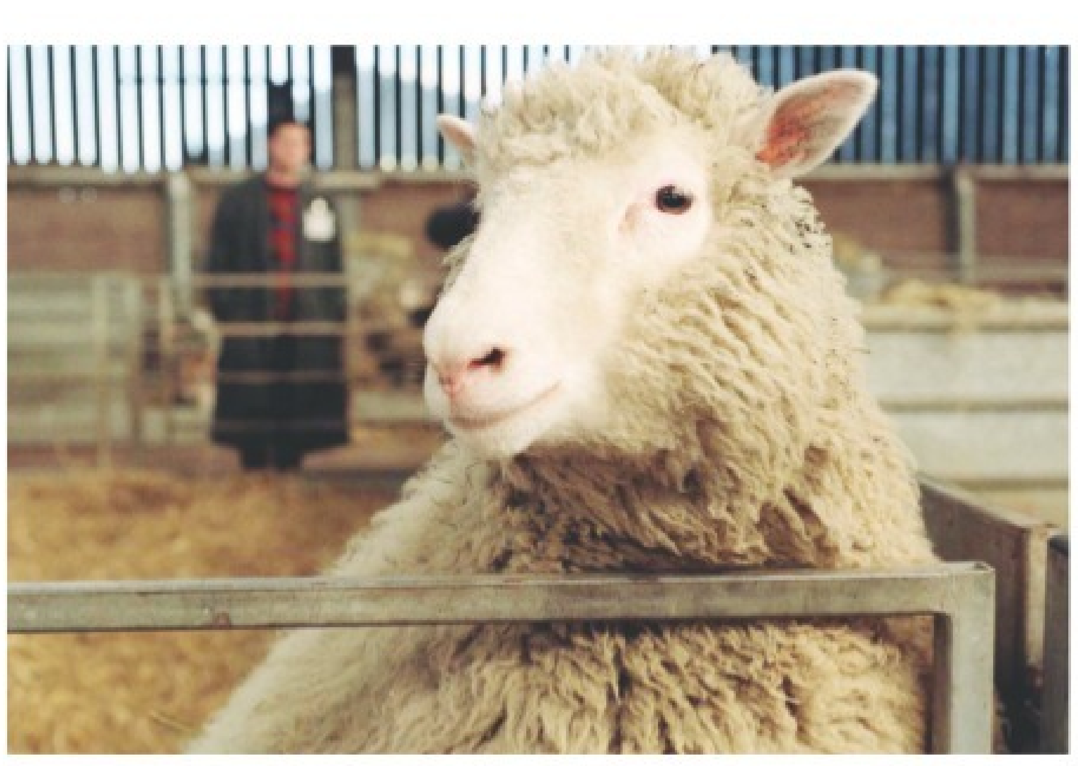
The Straits Times/ANN

# With the power to shock

## Cloning is now being used to produce livestock and even replicate lost pets but its implications still concern people, 15 years down the line from Dolly the sheep

JOSH GABBATISS

Fifteen years have passed since Dolly the sheep was euthanised after developing a lung disease and severe arthritis. She had lived a life in the spotlight and was revealed to the world in 1996 as the first mammal ever to be cloned from another individual's body cell.



Dolly the sheep

In Dolly's case, that was a single mammary gland cell from an adult sheep. According to Dr Ian Wilmut, the scientist who led the cloning research team, the sheep earned her name because they "couldn't think of a more impressive pair of glands than Dolly Parton's".

Dolly's death, like her life, was controversial. Normally sheep live for around 10 years but Dolly had only managed six. This raised questions about the long-term health of clones, and added fuel to the fire of those who considered cloning to be unethical. Upon the initial announcement of Dolly's birth, the press went into overdrive describing the "furious debate" in the scientific community the discovery had ignited. Many suggested it meant human cloning was inevitable.

At the time, a Princeton University biologist, Dr Lee Silver, told The New York Times it was "unbelievable". "It basically means that there are no limits. It means all of science fiction is true. They said it could never be done and now here it is, done before the year 2000," he said.

Yet in many ways, since Dolly's birth and subsequent death, cloning has become normalised. In recent years, champion horses have been replicated in a bid to — in the words of US cloning company ViaGen — "allow breeders to better leverage their most exceptional animals". Beloved cats and dogs have been cloned by owners who cannot bear to let go of their favourite pets, while "cloning factories" in China are being used to produce the best livestock in large quantities.

Despite all this, when the cloned monkeys Zhong Zhong and Hua Hua were revealed to the world in January, the furore around cloning was rekindled. While some focused on the treatment of the captive animals themselves, the main concern for many was once again the potential for a "slippery slope" — would cloned monkeys mean cloned humans are right around the corner?

"People are worried about applications in humans, but I think that's always nonsense," Professor Robin Lovell-Badge, a cloning expert at the Francis Crick Institute, told The Independent, "Given how inefficient and definitely unsafe it is, you would be crazy to even try it."

The monkeys were produced using the same technique used to produce Dolly all those years ago — somatic cell nuclear transfer. It involves taking the nucleus, which contains the genetic information, from a donor egg cell, and replacing it with one from another individual's cell. Following an electric shock to jump start cell division, the embryo will develop into a clone of the animal that donated the genetic material.

According to Professor Lovell-Badge, despite the 15 years that have elapsed, this technique still leaves a lot to be desired. "It works well enough for the commercial companies and the experiments that have been done, so not many people are devoting the effort to try and improve the method," he said.

While recent research suggests the scare stories about Dolly's ill health resulting from cloning were unfounded, the inefficiencies of the cloning process itself still make it a difficult procedure for many to swallow. The number of failed attempts required to successfully produce the cloned monkeys hammers home the serious ethical issues that still come with this practice.

The cloned macaques, Zhong Zhong and Hua Hua

"There were two born that died shortly afterwards that were abnormal," said professor Lovell-Badge, explaining his concerns about the recent study. "If they had shown pictures of those instead of the two cuddly ones that looked OK, then the response might have been quite different."

In fact, far from there being a total free-for-all on cloning, even the researchers currently working in this area are still keen to emphasise the care that must be taken when using this technology. "Just like nuclear power and artificial intelligence, cloning technology is also a double-edged sword," said Qiang Sun, one of the scientists responsible for the cloned monkeys, as he and his colleagues assured Chinese state media that they had "no plan to clone humans".

Monkeys are undoubtedly genetically closer to humans than sheep are, but as Zhong Zhong and Hua Hua were introduced to the world, scientists were once again keen to emphasise the development did not mean human clones were next.

The key point that Lovell-Badge emphasises is that cloning would not create exact copies of existing humans as people might imagine. Environmental factors such as upbringing would interact with the developing child and result in someone very different. "There has been no good justification for doing it," he said, "You're not going to recreate a lost loved one, and that's why I have been so critical of the dog cloning — because they are essentially fooling pet owners into thinking they can recreate the lost pet."

As for the practical applications of cloning, there are a number of suggestions. The research team behind the cloned monkeys said they wanted to use genetically identical primates to study diseases in humans. On the more ambitious end of the spectrum are plans to resurrect extinct species like woolly mammoths and Tasmanian tigers using cloning technology.

However, the main future use of cloning is likely to be producing better livestock, particularly in combination with newer genome editing techniques. "It's a very important experimental tool — a route to genetically modifying animals, particularly large animals like cows, sheep and pigs," said Lovell-Badge.

Having manipulated cells to produce the desired genetic outcome and no unwanted changes, scientists can then clone them. "That's the advantage of cloning procedure — it allows you more precisely to get exactly what you want."

The Independent