In vivo multi-dimensional information-keeping in Halobacterium salinarum

Davis, J.1,2, Bisson-Filho, A.3, Kadyrov, D.4, De Kort, T. M.5,1, Biamonte, M. T.6, Thattai, M.7, Thutupalli, S.7,8, Church, G. M.1:

1 Department of Genetics, Blavatnik Institute, Harvard Medical School,  
2 Department of Biology, Massachusetts Institute of Technology  
3 Department of Biology, Rosenstiel Basic Medical Science Research Center, Brandeis University, Waltham, MA 02454.  
4 SkBiolab, Technopark Skolkovo, Skolkovo Innovation center, Moscow 143026, Russia  
5 Biosciences Master’s programme Molecular & Cellular Life Sciences, Faculty of Science, Utrecht University, Utrecht, the Netherlands  
6 Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139  
7 Simons Centre for the Study of Living Machines, National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru 560065, India  
8 International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Bengaluru 560089, India

† Corresponding authors. Jdavis@genetics.med.harvard.edu (J.D.); gchurch@genetics.med.harvard.edu (G.M.C.)

Introduction

Digital data storage in DNA

Manufacture of silicon-based physical structures for digital information-keeping is dependent on shrinking global supplies of high-purity quartz sands. Sources of these raw materials are expected to be exhausted in less than 20 years (1). Increasing demand for data storage has therefore led to a search for suitable alternatives, including digital information storage in DNA. The capacity of DNA to hold digital archives has been under development since the 1980s (2) and a growing community of researchers have pursued this goal (3-18). To date, forms of DNA-encoded information include images, text, voice, music, movies and a computer operating system (13). At present, DNA can store at least 10^18 bytes per mm³, 6 orders of magnitude greater information density than the densest data storage medium currently available (1, 19). A method to store cell lineage data across all 10^19 cells of a mouse, illustrating the power of recording data in DNA in vivo, has recently been reported (18). Several examples of 3-Dimensional (3-D) information-keeping in DNA have also been demonstrated.

3-D encoded DNA

Crystallographer Nadrian Seeman and colleagues pioneered a form of DNA encoding often overlooked by researchers interested in DNA information storage. Seeman’s technique, later called “DNA origami,” relies on the base-pairing properties of DNA’s four nucleotides to encode 3-D information into DNA (20). In following years, Paul Rothemund’s lab refined DNA origami techniques using 7kb strands of DNA from the genome of the M13 virus which are coded to fold into desired shapes with hundreds of smaller complementary “staple” strands (21-22).

More recently, researchers at ETH Zurich and a collaborating Israeli scientist have converted stereolithographic code describing the 3-D figure of a rabbit into 12kb DNA oligonucleotides (23). This 3-D encoded DNA was subsequently encapsulated in silica microbeads (24) and then mixed with thermoplastic 3-D printing materials used to print unstructured triangulated 3-D copies of the described object (rabbit). Published reports refer to this version of 3-D encoded DNA as “DNA-of-Things” (23).

While highly efficient methods have been reported for encoding DNA with digital information (25), “DNA origami” and “DNA-of-things” 3-D DNA encoding methods have been comparatively inefficient, based on in vitro post-processing of relatively large coded molecules (>7 kb DNA) to yield a range of complex surfaces and simple 3-D objects (20-23).
Genes prescribing 3-D structure and function are long-standing biological precedents for 3-D translations of 2-D information. Leaf curvature, for instance, is an example of 3-D outcomes encoded in 2-D patterns of growth (26). Technical advantages of ideally encoded 3-D objects include the ability to acquire, display and manipulate an unlimited number of 2-dimensional (2-D) planes; usefulness for condensing information in graphs, equations, and complex mathematical models, and for representations of architecture, engineering, anatomy, topology, cartography (geography, geology, demographics, and population maps) where individual 3-D models replace multiple 2-D models (one for each view). 3-D shapes as a form of communication and information transfer may also be more readily interpreted by people from any area of the world or from a different era, compatible with the goals of long term data storage.

Limitations of DNA-encoded information carriers

Most efforts to use DNA as a data storage medium have involved inserting data either into “naked,” molecular DNA or into DNA contained in laboratory strains of E. coli. Between 1999 and 2012, all attempts at recording digital information into DNA were recorded in vivo (27).

After 2012, a shift to in vitro DNA archives was based on the notion that in vivo DNA data storage is unlikely to become a viable alternative for mainstream digital data storage. Two reasons are given for this: first, that while bacterial cells are smaller than most other known cells, and approximately 5 orders of magnitude smaller than microchips currently in use for data storage, in vivo DNA data storage has lower overall information capacity owing to the comparatively large size of bacterial cells when compared with the molecular scale of DNA; and second, that there has been a tendency to avoid highly complex tasks of creating stable modifications and/or additions to natural DNA within living cells (27).

Proponents of in vitro molecular DNA data storage claim that “naked” DNA digital archives could remain intact for hundreds to thousands of years when kept at reasonable temperatures and isolated from light and humidity (27). Longer term survivability of in vitro DNA would also require disposition in sealed, sterile environments shielded from radiation sources. If DNA is left unprotected, it is likely to be digested by enzymes commonly present in natural environments and promptly consumed by microorganisms (28). DNA is an inherently unstable substance, subject to damage by metabolic and hydrolytic processes including oxidative damage, depurination, depyrimidination, and cytosine deamination (29-33). DNA can also accumulate damage from environmental factors including UV light, ionizing radiation, exposure to genotoxic materials, and freeze-thaw cycles that can cause mechanical shearing (29, 32-36). Maintenance of environments stabilizing naked DNA is therefore essential for reducing information loss.

The process of making stable modifications and additions to natural DNA has always been part of the promise of molecular biology. In vivo DNA digital archives could take advantage of cellular machinery to protect and repair DNA (37-39) and to be conveniently and economically reproduced with little or no human intervention. As is the case with the DNA recovered from the past, the most secure and long-lasting way to store information in DNA may be to have it carried by an organism. Distribution and population numbers of cells could help to compensate for both lower information density and incidental data-parity errors in individual cells when compared with in vitro DNA information stores.

Nevertheless, laboratory domesticated strains such as of E. coli are particularly vulnerable. Bacterial strains that have become “workhorses” of molecular biology are – by both accident and intention – unlikely to survive outside of the laboratory (40). Despite these inherent weaknesses, negligent introduction of modified laboratory organisms into the environment may have the potential to damage natural ecosystems. Physical containment of laboratory organisms has therefore become augmented with biocontainment strategies that focus on engineered biological safeguards (41-42) to prevent recombinant organisms from surviving outside of controlled laboratory environments.

These vulnerabilities are inconsistent with qualities desired for DNA data storage. In addition to high information density, new digital storage structures will have to be robust enough to withstand environmental conditions for long periods of time and demand little at-rest energy cost (27).

**Figure 1:** (A) Pink salts with embedded Halobacterium salinarum crystalized from hypersaline culture (B) Phase-contrast images of Halobacterium salinarum cells entrapped between salt crystals in brine from dehydrated hypersaline media. Scale bars represent 5 µm.

**Halobacterium salinarum**

Extremophilic organisms may be more resilient and versatile information-keepers than either naked DNA or laboratory strains of E. coli. Here, the example of Halobacterium salinarum (Hsal), a halophilic archaeon, is taken as representative of this abundant and diverse group.

Hsal is not found in fresh water environments. Osmotic pressure will cause the cells of most halophilic organisms (including Hsal) to burst when immersed in fresh water, but there are many more saline environments in which halophilic organisms can thrive than there are fresh water environments to threaten them. 97.5% of water on Earth is salt water. Just 2.5% is fresh, and only 0.3% of Earth's fresh water is in liquid form on the surface (43-45). In addition to the ability of Hsal to survive in the broadest possible range of available aqueous environments, it can also survive an impressive range of environmental extremes.

Hsal is known to be extremely radiation resistant with reports of its chromosomes having been fragmented and then reassembled after very high...
Spontaneous mutation of DNA in cells serves the purpose of long term storage of immutable digital information. Information-keeping in halophiles could help to solve this problem. In stasis, halophiles can undergo processes of mutation and DNA damage and repair that actively reproducing cells would be expected to encounter much more frequently in other natural environments. Here, we aimed to study whether Hsal can in fact be used for data storage across long periods of time.

Results

Long term survival

While experiments performed over thousands of years are obviously impractical, to investigate further the hypothesis that Hsal cells can survive for long periods of time when entrapped in salt crystals, we started a liquid culture until late-exponential growth (O.D. ~1.0) stage and then split it into two identical aliquots. The first aliquot was used for genomic DNA extraction, while the second aliquot was left to evaporate within the sterile environment of a laminar hood. After 5 days, all liquid media was transformed into salt crystals with pink color tones (Figure 1A). Salt crystals were then distributed in sterile test tubes, sealed with parafilm and stored at room temperature. For the past three years, speckles of salt crystals from one of these samples were regularly tested for growth in fresh liquid media every six months. Notably, in every test performed to this point, cells reached mid-exponential growth (O.D. ~0.5) in less than 24 hours after inoculation. Figure 1B shows Hsal cells entrapped between salt crystals in brine from dehydrated hypersaline media.

3-D information-keeping in E. coli

To bypass one of the main limitations for in vivo DNA storage and increase its cache capacity, we determined to expand the number of dimensions of information that a given linear DNA could carry. We utilized highly efficient vector encoding to precisely describe a 3-D double-helix object using information encoded into only 46 DNA bases (Figure 2A). From there, a traditional paper folding pattern was used to yield the 3-D double-helix (Figure 2B).

Creasing patterns from origami are 2-D diagrams whose lines represent folds one must perform to transform a piece of flat paper into a 3-D origami shape (69). In origami folding using these patterns, there are two types of creases indicating the type of fold: “mountain” creases and “valley” creases. In diagrams, these two types of creases are often indicated using solid and dashed lines or using two distinct colors. In more mathematical terminology, crease patterns are planar graphs with labeled edges in which there are at most two labels (70). Our double-helix figure coding method accounts for the 2-D origami folding pattern as well as for “valley” and “mountain” creases.

The origami double-helix crease pattern was interpreted as a 94-bit lattice vector encoding 11 segments, taking advantage of the spatial periodicity of the pattern. This same method can be used to encode larger files (Supplementary Material 1). The asymptotic size of this encoding, in bits, is:

\[ C + S^2(2^{ceil(log2(x))} + 2^{ceil(log2(y))} + 1) \]

where, ‘S’ is the number of segments (each associated with defined start and end points, with one bit encoding the folding sense, positive or negative); ‘x’ and ‘y’ give horizontal and vertical lattice resolutions; ‘C’ is a constant which depends on the maximum possible values of ‘S’, ‘x’, ‘y’. The "1" is one bit to encode positive or negative fold. This encoding must first be decoded into the corresponding 2-D origami pattern, and subsequently folded to yield the desired 3-D structure (Figure 3). DNA manipulation and E. coli transformations were performed according to the method of Sambrook and Russell (71). Genetic constructions were generated by conventional enzymatic restriction and ligation of inserts into vectors using competent E. coli MG1655 as a primary recipient. The locus was sequenced from E. coli and reconstructed into the 3-D double-helix using the decoded crease pattern.
3-D information-keeping in Halobacterium salinarum

In fall 2018, we were offered an opportunity to contribute to the immortality-themed 5th Ural Industrial Biennale (72-73), a symposium and exhibition that opened in September, 2019 in Yekaterinburg, Russia. We decided to insert 3-D digital information into Hsal to serve as an extremely enduring human-sourced informational archive to compliment the Ural Biennale theme. Content for the Hsal archive consisted of two 3-D objects drawn from Russian folklore about an immortal evil wizard known as “Koschei the Deathless” (Коссе́й бесстра́шный)*. Based on these legends, the project was organized to insert 3-D models of a needle and an egg into the Hsal genome. We coded folding patterns of 3-D needle and spatial coordinates for a 3-D egg using a combination of highly efficient sparse row matrix encoding techniques (74-76) and degenerate elements inspired by “DNA Supercode” (77-80), a DNA-encoding scheme composed in the 1990s.

To generate DNA sequences with minimal lengths that encode the 3-D geometric information of the needle and egg shapes, we used techniques from both the ancient art of origami and modern data-compression algorithms. In particular, origami crease patterns were employed to compress the 3-D geometric information into 2-D diagrams. Although in general it is difficult, in the complexity-theoretic sense, to generate a 3-D origami shape from its crease pattern, the needle and egg crease patterns used in this work are simple enough that the folding instructions follow immediately from the diagrams (Figure 4) (81-82). This dimensionality reduction via crease pattern is achieved by the fact that the structure of the patterns effectively encodes the folding instructions required to transform a flat sheet of paper into a 3-D object. Typically, in origami crease patterns, two types of lines are drawn to indicate which direction the fold should be performed, called “mountain” creases and “valley” creases (Figure 5).

To describe which vertices are connected to each other, it is standard to use a matrix—a square table of numbers—data structure. The matrix describing the connectivity of the vertices is called an “adjacency matrix.” We created adjacency matrices for each of our encoded 3-D figures (Figure 6). An adjacency matrix for an N-vertex diagram is an N by N (square) matrix whose (i, j) entry is one if vertex i is connected to vertex j and zero otherwise. For most crease patterns, each vertex in the diagram is not connected to every other vertex and the adjacency matrix will have many zeros. This means that for most reasonable objects, the adjacency matrix associated with the crease pattern will be a what is termed a “spare matrix.” For visual clarity, instead of writing the numbers 0 and 1, we plot a black square if the value in the matrix is 1 and plot a white square if the value is 0.

* Koschei is said to be “deathless” because he is supposed to have hidden his evil soul in the tip of a needle which he then concealed in an egg, then in a duck, then in a rabbit, and so on, and then to have buried these concentric concealments in a chest that was buried under a tree on a mythical oceanic island (Buyan) that would appear and disappear with alternating tides.
plausibly hold an unlimited number of dimensions – and "DNA origami" and "DNA of Things" methods, which are both in vitro only techniques, intensive in the size and number of prerequisite DNA molecules, and limited to 3 spatial dimensions. The 4-D hypercube was encoded via specification of the locations of its vertices together with its adjacency matrix.

This protocol was repeated to clone a separate Hsal line with DNA coding for the 4-D hypercube figure***.

**Digitally encoded Halobacterium salinarum embedded in mineral salts**
Sterile brines made from various fossil mineral salts were inoculated with 3-D encoded Hsal and then recrystallized over a 3-day period in a series of workshops conducted at Ural Federal University Department of Experimental Biology and Biotechnology (Yekaterinburg, Russia), as part of the Ural Biennale.

**Discussion**

**Prospects for 3-D language**
While information written into sets of phonetic characters require readers to be fluent in particular languages, information written into sets of 3-D objects merely requires mathematical understandings needed to reconstruct 3-D objects from code. Some of the oldest systems of written language (e.g., cuneiform, hieroglyphics) operated by encoding information into 2-D pictures. Picture-writing systems may be useful as a first step to interpret information initially composed in the form of spoken language into more universally translatable sets of 3-D objects. Number sets could be written as sets of multifaceted polyhedra, while abstract ideas might be written as sets of contrasting, comparative, or metaphorical 3-D figures. Whether or not the nuance and poetry of language might be transcribed into 3-D figure-sets can be compared with questions about whether or not expressions in the form of painting might also be expressed in the form of sculpture.

**Legacy and burdens of transcendance**
It may be short-sighted to think about digitally encoded DNA merely as a means to provide relatively near-term solutions to 21st-century problems in data storage. Biological information repositories can now be expected to survive for much longer than the span of time that Homo sapiens sapiens is likely to persist as a species.

Fossil records indicate that the average lifespan of mammalian species is roughly on the order of 1 million years (87-89). In the case of closely related hominids, Neanderthals (Homo sapiens neanderthalensis) survived for only 200,000-300,000 years (90-91), while Homo erectus survived for about 1.6 million (92-97). Scholars debate forecasts of the survivability of Homo sapiens for many reasons but setting aside the possibility that humanity will shortly engage in catastrophic thermonuclear exchange, predictions made over the past few decades suggest that Homo sapiens sapiens will survive for anywhere from ~600 to 7.8 million years (98-100).

**Assembly of Needle/Egg and Hypercube oligonucleotides**
As repeats in the Needle/Egg and the Hypercube DNA sequences were predicted to cause issues during synthesis of the DNA oligos, both sequences were divided into segments, with Egg-Needle sequences being split into three segments and the Hypercube sequence into two. Each segment was amplified with BbsI recognition sites and elongated with random sequences until the complexities of the segments were sufficiently reduced (Table S1). The augmented sequences were ordered as gBlocks (Integrated DNA Technologies). The segments were subsequently assembled by Golden Gate assembly with the BbsI-HF restriction enzyme (New England Biolabs) and T7 ligase (New England Biolabs). The resulting products were amplified by PCR using Q5 High-Fidelity Master Mix (New England Biolabs) and the amplicons (Table S2) were purified with a PCR Purification kit (Qiagen). Constructs were verified using Sanger sequencing.

**Transformations with Halobacterium salinarum**
The Egg-Needle and hypercube constructs were inserted into Hsal NRC-1 (ATCC700922) main chromosome at the ura3 locus by double-crossover method transformation (83), with modifications. Instead of using standard counterselection with ura3, we selected transformants by introducing a MevR resistance cassette by a 4-piece Gibson reaction (84) with the following fragments: ura3 upstream region (primers oHS273 and oHV172, amplified from Hsal NRC-1 genomic DNA); Egg-Needle fragment (obtained as described above); MevR cassette (primers oHV173 and oHS278, amplified from pNBKO7 plasmid (85-86); ura3 downstream region (primers oHS279 and oHS280, amplified from Hsal NRC-1 genomic DNA). Primer sequences are listed in Table S1. The assembled fragments were then directly transformed by the standard PEG transformation method (83). Transformants (Figure 7) were then screened by PCR and the inserted sequence was confirmed by Sanger sequencing. A locus map of the final construct is represented in Figure 7.
Humanity has inherited special awareness of mortality that may be unique among members of the animal kingdom. Presumably, this special knowledge serves as inspiration to make lasting contributions to the legacy that each individual human being may leave behind. This same knowledge can also inspire questions about what bequests our species as a whole might leave for a far deeper future.

**Cosmological time-scales**

It is safe to assume that testaments written into salt-embedded microorganisms could persist over timescales that would be required for interstellar transits. The intelligibility of 3-D language might also be conveniently adapted for that purpose. But these ideas are overshadowed by the futility of communications over enormous spans that separate stars in our region of the galaxy.

Currently at a distance of 100 astronomical units (AU) and traveling outward at about 2.54 AU per year, *Pioneer 10* is one of the fastest ballistic objects ever sent into motion by human beings. Traveling at about 12.13 km/s relative to the sun, the *Pioneer 10* spacecraft is approximately 271,000 AU from *Proxima Centauri* (the nearest star) (101).

If *Pioneer 10* was targeted on *Proxima Centauri* (which it is not), it would not arrive for about 106,693 years. For context, behaviorally modern humans (possessing behavioral and cognitive traits such as abstract thinking, depth of planning, uses of symbolism, and blade-making technology) are thought to have been in existence for only about 50,000 years (102) while *Homo sapiens* is thought to have existed as a species distinct from other hominids for about 200,000 years (103).

It seems safe to assume that by the time any physical message could arrive in the neighborhood of a star more remote than the sun’s nearest companions, *Homo sapiens* will either be extinct or, will have evolved into another species.

We do not know how many unique star systems, all at greater distances than *Proxima Centauri*, would need to be targeted to reach a planet where life has developed, much less a planet with sentient life having the capacity to intercept and understand any message humanity had decided to transmit. Despite sophisticated searches for signals from others, the chances that humanity will ever find itself communicating with another intelligent species somewhere else in the universe are slim to none (104).

Earth will likely outlast humanity and remain habitable to other DNA-based organisms. Whether or not future terrestrial life evolves into organisms that are remotely like human beings, if we can manage to commit the legacy of our science, culture and civilization into forms of data storage that will outlast *Homo sapiens*, we may enable communications with beings who could turn out to be the only other communicable species in the cosmos, right here on Earth.

**Swansong**

The 3-D encoded “Koschei” *Hsal* has a poetic part in the form of coded figures that conjure an ancient myth about immortality. It is also first use of a medium that can reach into an indefinite future and as such, it begs the question about what serious, information-dense messages could also be projected across such extreme periods of time.

Legends about the swan who sings a last, final lament were proverbial by the 3rd century BCE (105-106) and reiterated many times in later Western poetry and art. But any ideas we may entertain about contacting beings far across space and time are complicated by another problem, one perhaps even more daunting than the distance between stars.

Understanding what intelligence is in the first place is prerequisite to reliably imagining any other intelligence or, to say it another way, “You have to reveal yourself to yourself before you can reveal yourself to anyone else.” This is what Aristotle considered to be a principal element of human tragedy (“Recognition and Reversal” in *Poetics*) (107), and yet, it is probably the most important reason we have to continue the search. Any messages we may wish to preserve for the “Other” will have to attempt to answer the questions “Is this who we are? Is this what we know?”

While “DNA origami” techniques may have useful applications, 3-D objects rendered as digitally-encoded DNA are more efficiently described as strictly mathematical constructs rather than as exotic, secondary molecular conformations.

Extremophiles can be more robust and long-lasting information-carriers than earlier repositories for DNA-based digital archives. Indeed, techniques currently exist that can be used to transfer digital information into the DNA of *Halobacterium salinarum*. Furthermore, useful applications of digital archives encoded into extremophile DNA are not limited to routine data storage systems. Other plausible applications include interstellar messaging and legacy terrestrial, lunar and planetary archives.

**Acknowledgments**

The authors are indebted to art-and-science curator, Olga Vad, whose initial offering brought a large part of this work into the realm of possibility, and to Yashas Shetty, whose enthusiasm brought several of us together for the first time, and whose insights into alternative methods for encoding 3-D objects helped to start the ball rolling. We are grateful to Irina Kiselyova, head of Department of Experimental Biology and Biotechnology at Ural Federal University, Russia for the opportunity to host a series of our experiments with extremophilic archaea; and to Ido Bachelet, for his advice and support. We also acknowledge Gabriel Filsinger, Eswar Iyer, Srividya Chandramouli, Erkin Kuru, and Henry Lee for time spent reviewing the manuscript.

**References**

[Art Journal 55 (1996) This entire issue (Ellen Levy, ed.) was devoted to *Art and the Genetic Code*, representing an early example of an art journal addressing issues of art and genetics]
(4) Bancroft, C. Long-term storage of information in DNA. *Science* 293, 1763-1765 (10 August 2001)
(9) Allenberg, M. Rotstein, O. D. An improved Huffman coding method for archiving text, images, and music characters in DNA. *Biotechniques* 47, 747-754 (Sept. 2009)
on storage in synthesized DNA.

(11) Goldman, N., Bertone, P., Chen, S., Dessimoz, C., Leproust, E. M., Sipos, B. and Birney, E. Towards practical, high-capacity, low mainte-
nance information storage in synthesized DNA. Nature 494, 77-80, (07 Feb. 2013)

(12) Bornholt, J., Lopez, R., Carmane, D., Ceze, L., Seelig, G., Strauss, K. A DNA-based archival storage system. Proceedings, 21st Interna-
tional Conference on Architectural Support for Programming Lan-
guages and Operating Systems (ASPLOS); Published by ACM – Asso-
ciation for Computing Machinery; https://www.microsoft.com/en-
us/research/publication/dna-based-archival-storage-system (April 2016)

(13) Erlich, Y. & Zielinski, D. DNA Fountain enables a robust and efficient

2016).

(15) Shipman, S., Nivala, J., Macklis, J.D., Church, G. M. CRISPR-Cas en-
coding of a digital movie into the genomes of a population of living

(16) Choi, Y., Ryu, T., Lee, A. C., Choi, H., Lee, H., Park, J., Song, S., Kim,
S., Kim, H., Park, W., Kwon, S. High Information capacity DNA-based
data storage with augmented encoding characteristics using degen-
erate bases. Nature Scientific Reports 9, (29 April 2019)

(17) Lee, H. H., Kalhor, R., Naveen, G., Bolot, J., Church, G. M. Termina-
tor-free template-independent enzymatic DNA synthesis for digital in-
formation storage Nature Communications 10, (2019) [https://www.nature.com/articles/s41467-019-10258-1]

(18) Kalhor, R., Kalhor, K., Mejia, L., Leeper, K., Graveline, A., Mali, P.,
Church, G. M Developmental barcoding of whole mouse via homing

(19) Rutten, M. G. T. A., Vaandragter, F. W., Eleman, J. A. A. W. & Nolte,
R. J. M. Encoding information into polymers. Nature Reviews Chem-

(20) Chen, J. & Seeman, N. C. Synthesis from DNA of a molecule with the

(21) Rothemund, P. W. K. Folding DNA to create nanoscale shapes and

(22) Rothemund, P. W. K. & Andersen, E. N Nanotechnology: the im-

(23) Koch, J., Gantenbein, S., Masania, K., Wendelin, J. S., Erlich, Y.,
Grass, R. N. A DNA-of things storage architecture to create materials
with embedded memory. Nature Biotechnology 38, 39-43 (Jan. 2020;
Epub 09 Dec. 2019)

(24) Stark, W. J. Robust chemical preservation of digital information in sil-
ica with error-correction codes. Angewandte Chemie (International ed
English) 54, 2552-2555 (04 Feb. 2015)

(25) Choi, Y., Ryu, T., Lee, A. C., Choi, H., Lee, H., Park, J., Song, S., Kim,
S., Kim, H., Park, W., Kwon, S. High information capacity DNA-based
data storage with augmented encoding characteristics using degen-
erate bases. Nature Scientific Reports 9, (29 April 2019)

(26) Nath, U., Crawford, B., C., W., Carpenter, R., Coen, E. Genetic control
of surface curvature Science 28, 1404-1407 (28 Feb., 2003)

(27) Ceze, L., Nivala, J., Strauss, K. Molecular digital data storage using

(28) Clark, D. P., Pazdernik N. J. Transformation, uptake of naked DNA. In
Molecular Biology (Second Edition/Academic Cell Update) Academic
Press (Elsevier) 641-646 (2013)

(29) Lindahl, T. Instability and decay of the primary structure of DNA Na-

(30) De Bont R, van Larebeke N. Endogenous DNA damage in humans: a

(31) Lindahl T, Nyberg B. Rate of depurination of native deoxyribonucleic

(32) Shikama, K. Effect of freezing and thawing on the stability of double
helix of DNA. Nature 207, 529-530 (July 1965)

(33) Röder, B., Frühwirth, K., Vogl, C., Wagner, M., Rossmanith, P. Impact
of long-term storage on stability of standard DNA for nucleic acid-
based methods Journal of Clinical Microbiology 45, 4260-4262 (Nov.
2010)

(34) Wondergem, J. (ed.) Radiation biology: a handbook for teachers and
students. Vienna: International Vienna Atomic Energy Agency pp. 26,
72 (Mar. 2010)

(35) Häder D-P, Sinha RP. Solar ultraviolet radiation-induced DNA damage
in aquatic organisms: potential environmental impact. Mutation Re-
search. 571, 2005; 221–233

(36) Rastogi, R. P., Richa, Kumar, A., Tyagi, M. B., Sinha, R. P. Molecular
Mechanisms of ultraviolet radiation-Induced DNA damage and repair.

(37) Wigly, D. B. Bacterial DNA repair: recent insights into the mechanism
of RecB, AddAB and AdnAB. Nature Reviews Microbiology 11, 9-13
(Dec. 2013)


(39) Yi, C., He, C. DNA repair by reversal of DNA damage. Cold Spring
Harbor Perspectives in Biology. 5 (Jan. 2013)

(40) Fux, C. A., Shirliff, M., Stoodley, P., Costerton, J. W. Can laboratory
reference strains mirror “real world” pathogenesis? Trends in Micro-
biology 13, 58-32 (01 Feb. 2005)

(41) Mandell, D., Lajoie, M., Mee, M., Takeuchi, R., Kuznetsov, G., Norville,
J. E., Gregg, C. J., Stoddard, B. L., Church, G. M. Bioccontainment of
generically modified organisms by synthetic protein design. Na-
ture 518, 55–60 (05 Feb 2015; Epub 21 Jan. 2015)

(42) Castanon, O., Smith, C. J., Khoshkakhlagh, P., Ferreira, R., Guell, M.,
Said, K., Ramazan, Y., Dysart, M., Wang, S., Thompson, D., Myllykal-
lio, H., Church, G. M. CRISPR-mediated bioccontainment
https://cto/qRUp6M6rV#bioRxiv (04 Feb 2020)

(43) Shiklomanov, L. A. World fresh water resources In: Gleick P. H. (ed.),

(44) Where is Earth’s water? United States Geological Survey
https://web.archive.org/web/20131214091601/http://ga.wa-
ter.usgs.gov/edu/earthtwerewater.html

(45) Eakins, B. W., Sharman, G. F. volumes of the world’s oceans from
ETOPO1 Boulder: NOAA National Geophysical Data Center (2010)
[ETOPO1 is a 1 arc-minute global relief model of Earth's surface
that integrates land topography and ocean bathymetry]

(46) https://science.nasa.gov/science-news/at-
nasa/2004/10sep_radmicrobe/

(47) DeVeaux, L. C., Müller, J. A., Smith, J., Petriko, J., Wells, D. P., Das-
sarma, S. Extremely Radiation-resistant mutants of a halophilic ar-
chaean with increased single-stranded DNA-binding protein (RPA)

(48) Thayer, D. W., Boyd, G. Elimination of Escherichia coli O157:H7 in
meats by gamma irradiation. Applied Environmental Microbiology 59,
1030-1034 (Apr. 1993)

(49) Hildenbrand, C., Stock, T., Lange, C., Rother, M., Soppa, J. Genome
copy numbers and gene conversion in mangenochogenic archaea. Jour-
nal of Bacteriology 18, 734-743 (Nov. 2010)

(50) Weider, G., Leuko, S., Stan-Lotter, H. Survival and growth of Halo-
bacterium sp. NRC-1 following incubation at -15°C, freezing or freeze-
drying, and the protective effect of cations. Proceedings of the Third
European Workshop on Exo-Astrobiology, 18 - 20, Madrid, Spain.
Ed.: Harris, R. A., Ouwehand, L. ESA SP-545 311-312 (November 2003)


(53) Cocker, J. A., DasSarma, P., Kumar, J., Müller, J. A., DasSarma, S. Transcriptional profiling of the model archaeon Halobacterium sp. NRC-1: responses to changes in salinity and temperature Saline Systems **3**; http://www.salinesystems.org/content/3/1/6 (25 July 2007)


(67) Vreeland, H; Rosenzweig, W D; Lowenstein, T; Satterfield, C; Ven- tosa, A (December 2006). Fatty acid and DNA analyses of Permian bacteria isolated from ancient salt crystals reveal differences with their modern relatives. *Extremophiles* **10**, 71–78 (Dec. 2006)


(76) https://en.wikipedia.org/wiki/Sparse_matrix


(85) Busch, C., R. DNA mismatch repair and response to oxidative stress in the extremely halophilic archaeon *Halobacterium sp*. strain NRC-1 [Dissertation submitted to the Faculty of the Graduate School of the University of Maryland, College Park, in partial fulfillment of the requirements for the degree of Doctor of Philosophy 2008] [first reference to pNBK07]


(87) Wilson, E. O ., *Half Earth: our planet’s fight for life* (First edition.). New York: Liveright Publishing Corporation, a division of W.W. Norton & Company. (07 March 2016) [*History Redefined* Chapter 16: “The average span across all groups combined appears to be (very roughly) a million years.”]


(89) The current mass extinction [https://www.pbs.org/wgbh/evolution/library/03/2/1_032_04.html]

(91) Alper, J., Rethinking Neanderthals Smithsonian (June 2003)


(97) https://www.nhm.ac.uk/discover/homo-erectus-our-ancient-ancestor.html

(98) Benétreau-Dupin, Y., (ed.) Doomsday argument San Francisco State University (2019); https://philpapers.org/browse/doomsday-argument


(100) Hawking, S., Tencent WE Summit 2017 keynote: https://www.youtube.com/watch?v=U-hcSLya0_w [ca. 2017 Stephen Hawking predicted extinction of Homo sapiens in 583 years.]


(104) Scharf, C., Even if the Milky Way is teeming with spacefaring aliens, we should not be surprised that Earth Remains unvisited Scientific American 322, 32-39 (Jan. 2020)

