DNA DATA STORAGE ALLIANCE
A SNIA TECHNOLOGY AFFILIATE

Creating confidence in preserving digital data in synthetic DNA

Dave Landsman (dave.landsman@wdc.com)
Western Digital, DNA Data Storage Alliance

Acknowledgements
Jacques Bonnet, Marthe Colotte, Lee Organick, Chris Takahashi

Library of Congress Seminar
April 2024

DNA Data Storage Alliance

40+ member organizations today

Mission

- Create an interoperable storage ecosystem based on DNA as a data storage medium

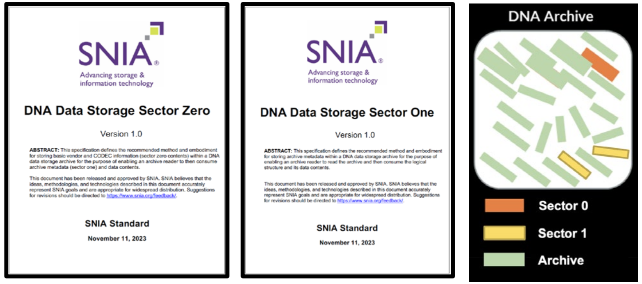
Scope

- Educate the market to create awareness and adoption of DNA data storage
- Develop a DNA data storage technology roadmap to drive R&D and funding
- Develop standards and/or specifications to encourage evolution of the ecosystem

DNA Data Storage Alliance Standards Work

Bootstrap a nascent ecosystem without stifling innovation

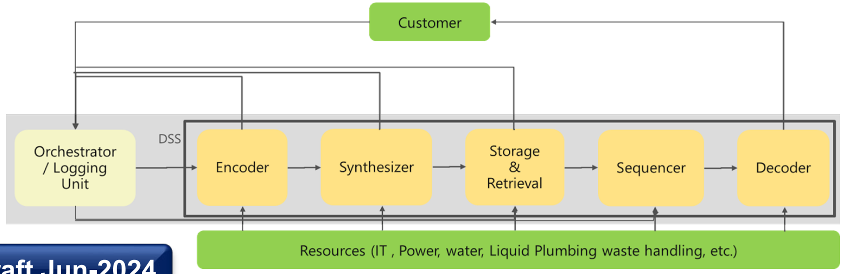
1) DNA Archive Rosetta Stone - MBR



Published

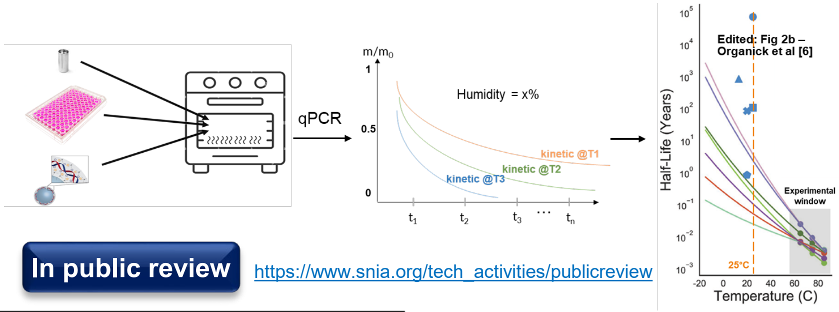
https://www.snia.org/tech_activities/standards/curr_standards

2) Interoperable Interfaces – DNA in the Datacenter



Draft Jun-2024

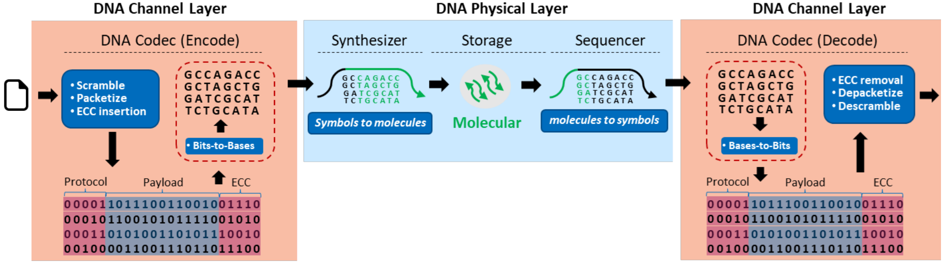
3) Data Stability Evaluation – Compare Half-Life



In public review

https://www.snia.org/tech_activities/publicreview

4) Open-Source Codecs – DNA Channel



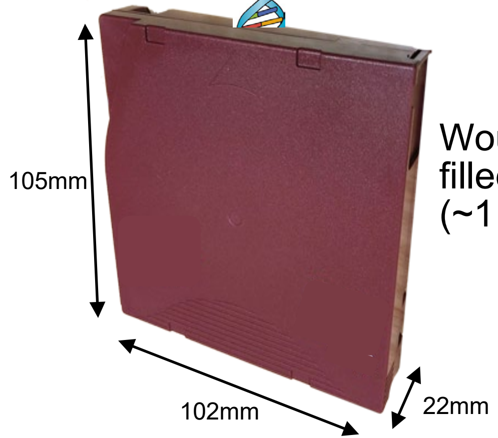
2024

DNA DATA STORAGE ALLIANCE
A SNIA TECHNOLOGY AFFILIATE

Data Stability Evaluation Spec

Why DNA?

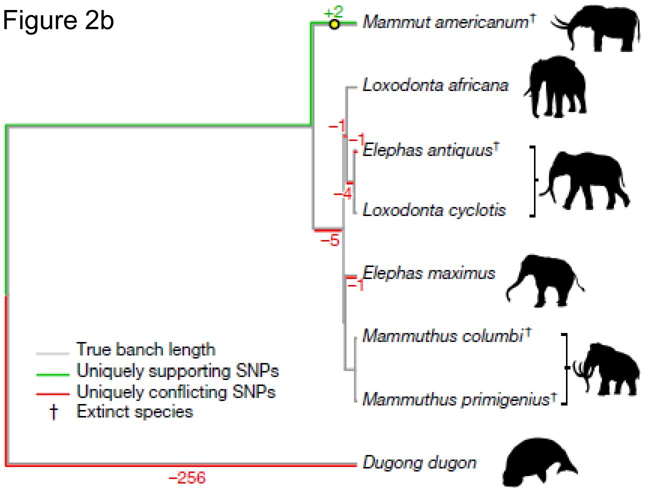
DNA bits are <u>very</u> small, ~1nm^3



Would hold ~2 exabytes if filled with DNA bits (~115,000 LTO9 tapes)

And they are extremely (i.e., 2M+ years) durable and sustainable

Figure 2b

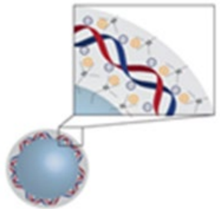


Kjær, K.H., Winther Pedersen, M., De Sanctis, B. et al. A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. Nature 612, 283–291 (2022). https://doi.org/10.1038/s41586-022-05453-y

We don't store digital data in DNA using fossils

Must create trust that manufactured DNA Containment Systems work

- Challenge to creating trust
  - Rating the durability enabled by a DNA Containment System (DCS), and enabling apples-to-apples comparison to another DCS, is not possible without a standard method of conducting media aging experiments that yield standard metrics
- Challenges to creating standard methods/metrics
  - Can a DCS be evaluated independent of the other steps in the DNA data storage pipeline?
  - DNA ages very slowly so we need to believe that an accelerated wear methodology won't skew results

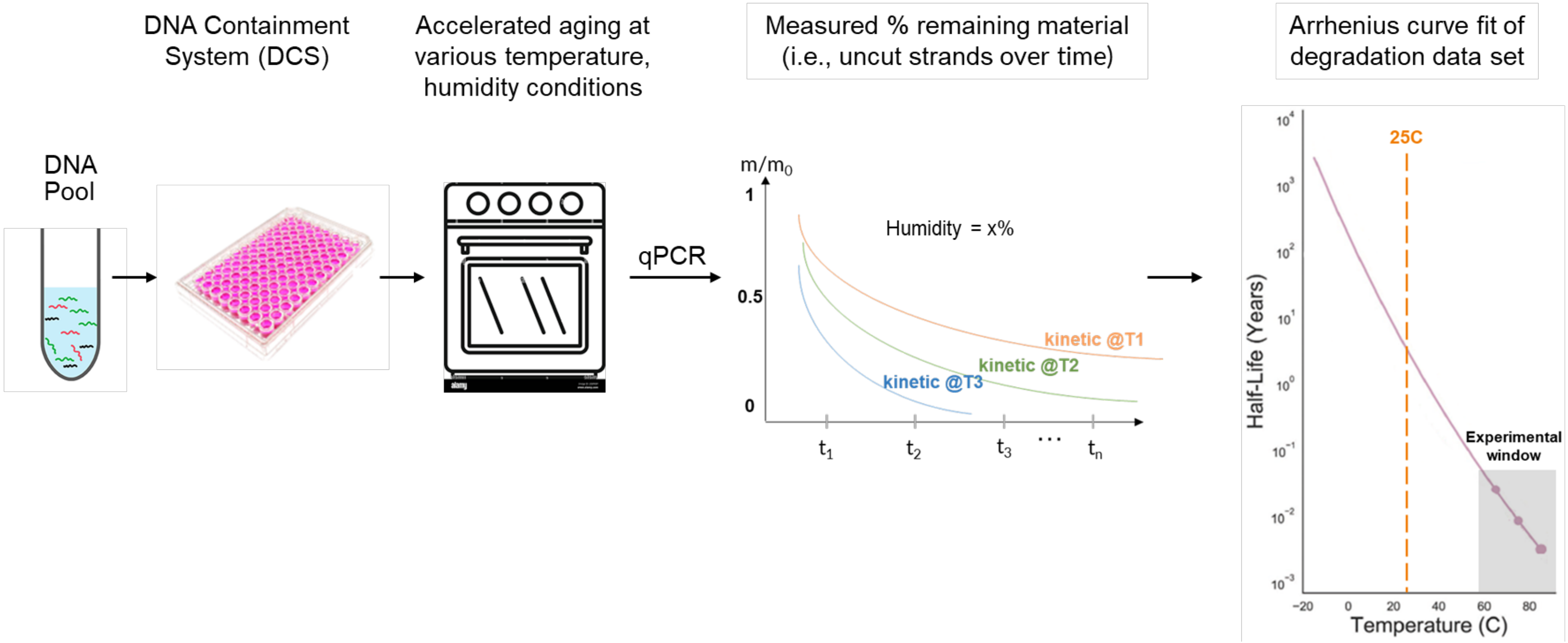


Silica nanoparticles (x years)

Vials (y years)

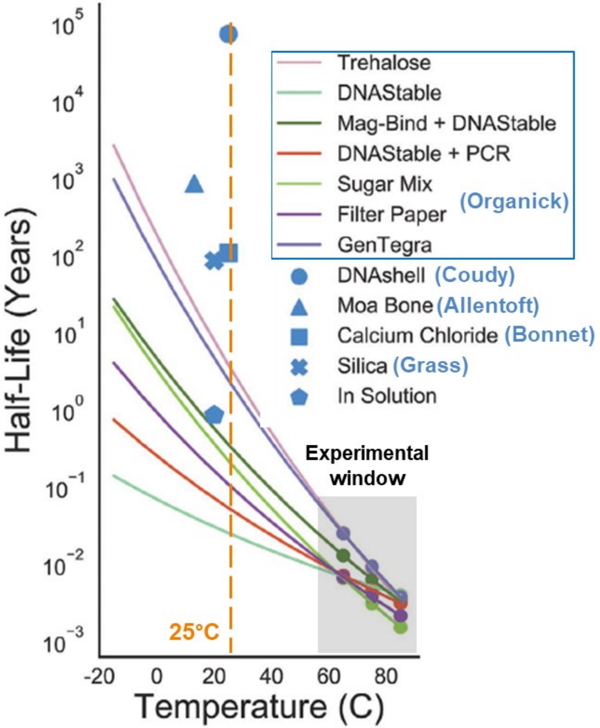
Sealed stainless steel (z years)

Typical DNA Data Stability Evaluation Experiment



What did the research say about durability?

- Possible to, in general, extend the durability of DNA media by using various additives and containment systems
- Containment systems which shield DNA from atmosphere preserve molecular stability for long periods at “room temp” (i.e., 20C-25C)



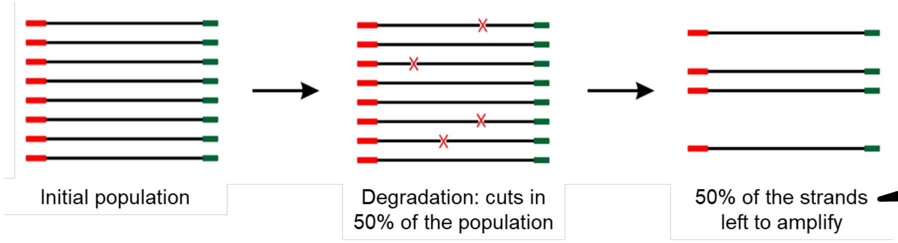
Edited from figure 2b, Organick et al [6]

Preseveration Category	Preservation Substrate/Method	
Chemical encapsulation	Encapsulation in salts	[12, 16]
	Degradable Polymer Microcapsules	[32]
	Cationic Diblock Copolymer	[33]
Physical encapsulation	Silica nanoparticles	[1]
	Stainless steel capsules	[3, 7, 30]
	Magnetic silica nanoparticles	[13]
Inclusion in a matrix	DNAstable	[1, 21]
	Gentegra DNA	[1, 22]
	Pullulan	[14]
	Silk	[15]
	composite nucleic acid-polymer fibers	[34]
	300K matrix inclusion	[25]
Absorption on paper	FTA paper	[1, 23, 24]
	Chitosan treated paper	[17]
Dehydration on solid supports	Capillaries	[20]
	Glass	[26, 27]
	Tube walls	[28]
Dissolution in liquid salts	Imidazolium cations	[18]
	Imidazolium cations	[19]
Living organism	yeast genome	[36]
	Ecoli genome	[37]
	yeast cells	[38]
	Bacteria	[29]
DNA beads	Magnetic Bead Spherical Nucleic Acid Microstructure	[35]

Source: Bonnet, J., Colotte, M.

What did the research say about measuring durability?

- The cited studies measured broken strands (i.e., strands which have at least one break in the sugar-phosphate chain) as a function of concentration of remaining intact DNA



Only intact strands can be amplified

Half-Life: Time for concentration of initial intact strands to decline by 50%

Several studies examined intact strands for data retention

- Read errors did not appear to be affected by the DCS (i.e., preservation method), nor by any particular sequence of bases
- The encoded data in DNA strands that survived accelerated wear aging with no strand breaks appeared recoverable

Indications of relationship between molecular stability and data retention

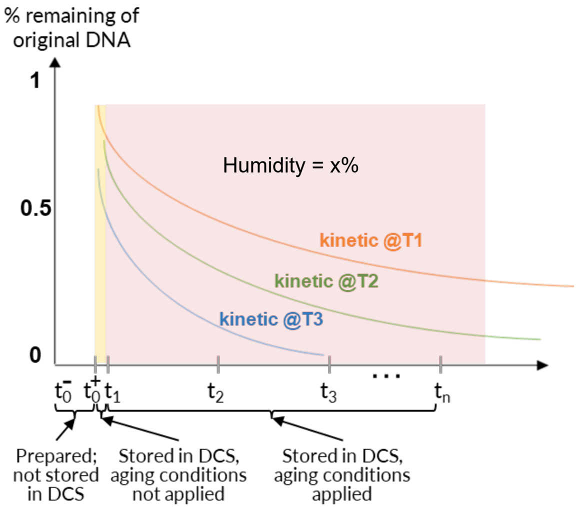
Conclusion

Gaining confidence into how well a DCS preserves data is possible by measuring how well it prevents DNA molecular breakdown, independent of synthesis/retrieval/sequencing.

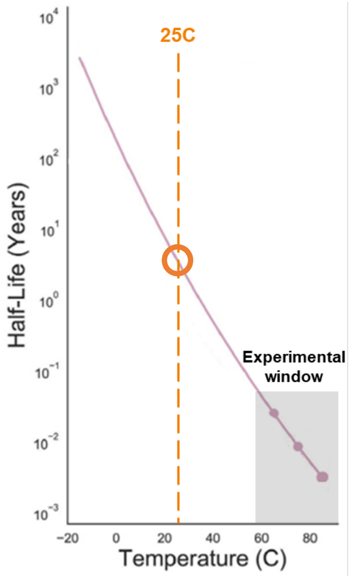
Summary

Standardizing an aging protocol and lifetime metric enables meaningful vendor independent comparisons of durability claims for different DCSs

1. Apply aging conditions (temp/humidity points) to media in a DCS, and count chain breaks as a concentration of intact DNA over time



2. Do an Arrhenius curve fit on the collected data to extrapolate half-life of the media for that DCS at 25°C



Data Stability Evaluation Method for DNA Data Storage Containment Systems

https://www.snia.org/tech_activities/publicreview

References

1. Grass RN, Heckel R, Puddu M, Paunescu D, Stark WJ. Robust Chemical Peservation of Digital Information on DNA in Silica with Error-Correcting Codes. Angew Chem Int Ed Engl. 2015; 54(8): 2552-2555.
2. Lindahl T, Nyberg B. Rate of depurination of native deoxyribonucleic acid. Biochemistry. 1972; 11(19): 3610-3618
3. Bonnet J, Colotte M, Coudy D, Couallier V, Portier J, Morin B, et al. Chain and conformation stability of solid-state DNA: implications for room temperature storage. Nucleic Acids Res. 2010; 38(5): 1531-1546
4. Anchordoquy TJ, Molina MC. Frontiers in Clinical Research. Preservation of DNA. Cell Preservation Technology. 2007; 5(4): 180-188
5. Molina MC, Anchordoquy TJ. Degradation of lyophilized lipid/DNA complexes during storage: The role of lipid and reactive oxygen species. Biochim Biophys Acta- Biomembr. 2008; 1778(10): 2119-2126
6. Organick L, Nguyen BH, McAmis R, Chen WD, Kohll AX, Ang SD, et al. An Empirical Comparison of Preservation Methods for Synthetic DNA Data Storage. Small Methods. 2021; 5(5): e2001094
7. Coudy D, Colotte M, Luis A, Tuffet S, Bonnet J. Long term conservation of DNA at ambient temperature. Implications for DNA data storage. PLoS One. 2021; 16(11): e0259868
8. Wandeler, P. Patterns of nuclear DNA degeneration over time: a case study in historic teeth samples, Mol. Ecol. 12 (2003) 1087–1093.
9. Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, et al. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proc Biol Sci. 2012; 279(1748): 4724-4733.
10. Colotte M, Couallier V, Tuffet S, Bonnet J. Simultaneous assessment of average fragment size and amount in minute samples of degraded DNA. Anal Biochem. 2009; 388345–347
11. Kohll AX, Antkowiak PL, Chen WD, Nguyen BH, Stark WJ, Ceze L, et al. Stabilizing synthetic DNA for long-term data storage with earth alkaline salts. Chem Commun (Camb). 2020
12. Antkowiak PL, Koch J, Rzepka P, Nguyen BH, Strauss K, Stark WJ, et al. Anhydrous calcium phosphate crystals stabilize DNA for dry storage. Chem Commun (Camb). 2022.
13. Strauss, K et al 2021 US2021291134 (A1) Silica Encapsulated DNA on Magnetic Nanoparticles.
14. Jahanshahi-Anbuhi S et al. Pullulan encapsulation of labile biomolecules to give stable bioassay tablets. Angew Chem Int Ed Engl. 2014; 53(24): 6155–6158
15. Liu Y, et al. DNA preservation in silk. Biomater Sci. 2017.
16. Strauss, K. 2019 US2019390194 (A1) High-density DNA storage with salt.
17. Moon, W.-C. US2007254294 (A1) 2007 Method for Storing Dna by Using Chitosan, and Products Using the Methods.
18. Khavani, M. Application of amino acid ionic liquids for increasing the stability of DNA in long term storage, J Biomol Struct Dyn 1-15 2022
19. Hiroyuki O, et al US2007196826 A1 Solvent For Dissolving Nucleic Acid, Nucleic Acid-Containing Solution And Method Of Preserving Nucleic Acid
20. Mohanty, P.S. et al AU2020244494B2 Capillary assisted vitrification processes and devices.
21. Whitney, S. E., et al 2014 US 2014/0065627Al compositions and methods for biological sample storage.
22. Hogan, M. et alUS2017145477 (A1) Matrices and media for storage and stabilization of biomolecules.
23. Horton, J. K. et al US20170151545A1 Oligonucleotide data storage on solid supports
24. Fomovskaia, G. and et. Al WO0062023 (A1) FTA-coated media for use as a molecular diagnostic tool.
25. https://300k.bio/ accessed on 2024-03-27
26. Newman, S., et al High density DNA data storage library via dehydration with digital microfluidic retrieval. Nature communications 1706 10 (1) 2019
27. Provisional patent application (application number 62/812,521) filed on March 1, 2019 by the University of Washington
28. Trapmann, S. et al, Development of a novel approach for the production of dried genomic DNA for use as standards for qualitative PCR testing of food-borne pathogens. , Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement 695-699 9 (11) 2004
29. Wong, P.C. US2006024811 (A1), Storing data encoded DNA in living organisms.
30. Colotte, M. et al., Adverse Effect of Air Exposure on the Stability of DNA Stored at Room Temperature. Biopreserv Biobank. 2011; 9(1): 47–50
31. Molina, M et al, Degradation of lyophilized lipid/DNA complexes during storage: The role of lipid and reactive oxygen species, https://doi.org/10.1016/j.bbamem.2008.04.003
32. ACS Nano 63-9 1 (1) 2007 A general approach for DNA encapsulation in degradable polymer microcapsules. Zelikin, A. N., A. L. Becker, A. P. Johnston, K. L. Wark, F. Turatti and F. Caruso
33. Langmuir 34-42 21 (1) 2005 Encapsulation of DNA by cationic diblock copolymer vesicles. Korobko, A. V., W. Jesse and J. R. van der Maarel
34. Mater Today Bio 100900 24 2024 DNA data storage in electrospun and melt-electrowritten composite nucleic acid-polymer fibers. Soukarie, D., L. Nocete, A. M. Bittner and I. Santiago
35. ACS Synthetic Biology 2023 Magnetic Bead Spherical Nucleic Acid Microstructure for Reliable DNA Preservation and Repeated DNA Reading. Shen, P., X. Qu, Q. Ge, T. Huang, Q. Sun and Z. Lu
36. Adv Sci (Weinh) e2305921 2024 "Cell Disk" DNA Storage System Capable of Random Reading and Rewriting. Hou, Z., W. Qiang, X. Wang, X. Chen, X. Hu, X. Han, W. Shen, B. Zhang, P. Xing, W. Shi, J. Dai, X. Huang and G. Zhao
37. National Science Review 8 (5) 2021 An artificial chromosome for data storage. Chen, W., M. Han, J. Zhou, Q. Ge, P. Wang, X. Zhang, S. Zhu, L. Song and Y. Yuan
38. Nature Computational Science 2022 Towards practical and robust DNA-based data archiving using the yin–yang codec system. Ping, Z., S. Chen, G. Zhou, X. Huang, S. J. Zhu, H. Zhang, H. H. Lee, Z. Lan, J. Cui, T. Chen, W. Zhang, H. Yang, X. Xu, G. M. Church and Y. Shen

THANK YOU

dnastoragealliance.org
Twitter: @DnaDataStorage
LinkedIn: @dna-data-storage-alliance
Email: info@dnastoragealliance.org

BACKUP

Grass et al [1]

Can we recover data in uncut strands that survive accelerated wear?

Study observations

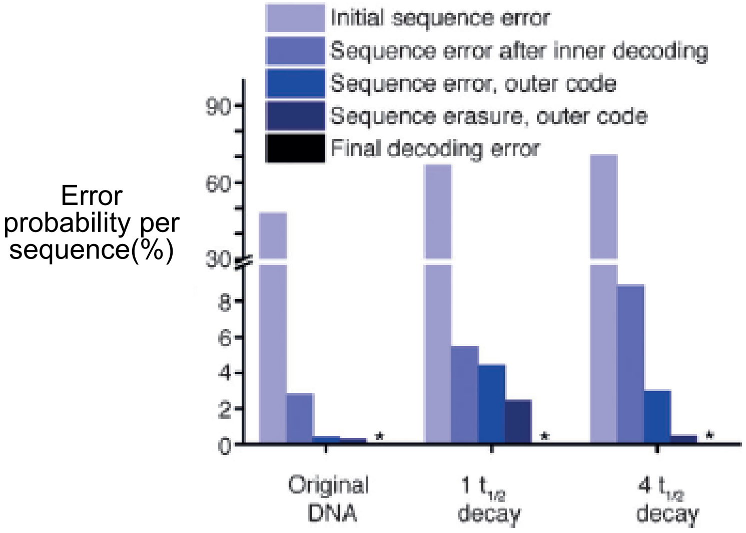
- "Inner and outer code of the error correcting scheme had to correct significantly more errors than in the non-heat-treated sample, [but] in both cases the original information could be recovered without final error"

Answer: Yes

- Enough strands survived temperatures used in accelerated wear to validate high temperature stress method

Figure 3

Recovering original data from silica substrate



Organick et al [6]

Do read errors vary due to preservation method?

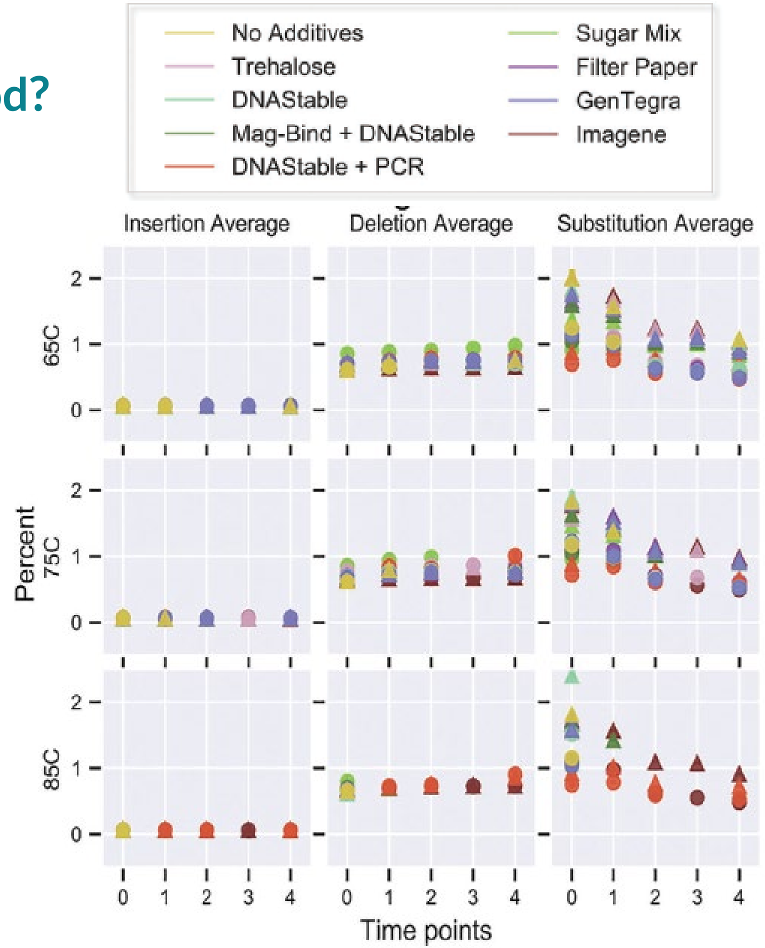
Study observations:

- Minimal (< 1%) variation in error rates across preservation methods, temps, and time points
  - Even substitutions, which show most variance, show this variance before any aging begins
- No one preservation method showed consistently more or fewer errors than any other method across different temperatures and time points
  - Suggests insertion, deletion, substitution errors are independent of storage method

Answer: No

- For the purposes of evaluating and comparing a DCS, errors introduced/corrected by synthesis, retrieval, and sequencing can be ignored

Figure 4 – Observed error rates



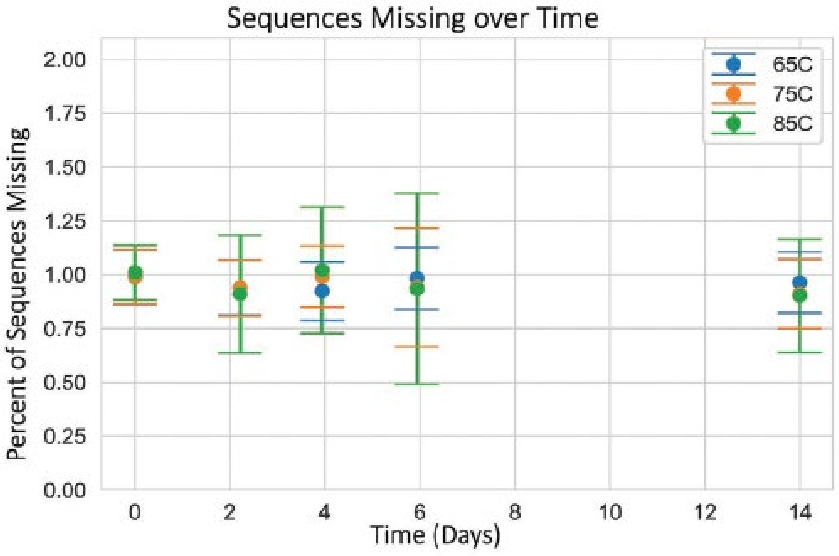
Organick et al [6]

Do certain sequences cause read errors with specific preservation methods?

Study Observations

- Total # of sequences found missing during sequencing (across all methods, time points, temperatures) were analyzed for sequence loss
  - Total # missing sequences did not increase over Time 0, indicating no sequence dependent degradation caused by preservation method (i.e., no "storage bias")
  - This finding reinforced by further finding that individual sequences missing at a particular timepoint had > 90% probability of reappearing and being successfully sequenced later

Figure 5a



Answer: No

- Further reinforces that one can define a standard stability evaluation methodology that is independent of the effects of synthesis, retrieval, and sequencing