STORAGE DEVELOPER CONFERENCE

SDC 23

BY Developers FOR Developers

Creating confidence in preserving digital data in DNA

Dave Landsman, Marthe Colotte, Jacques Bonnet, Chris Takahashi

Data Retention Workgroup
DNA Data Storage Alliance

DNA Data Retention Workgroup - Overview

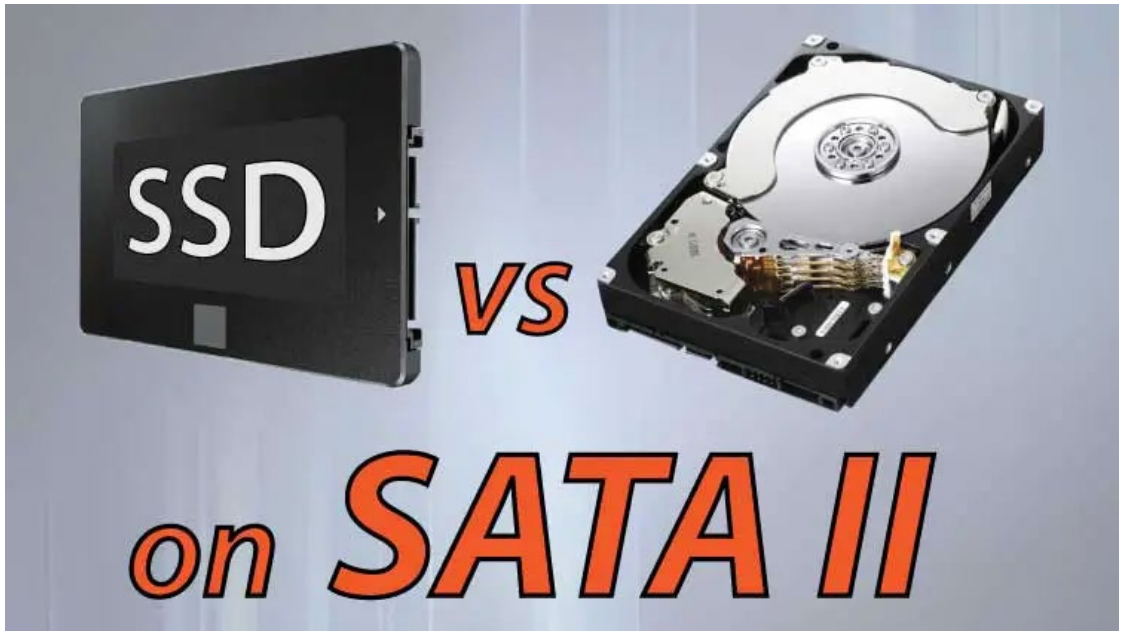
- Overall Mission
  - Enable users of DNA-based storage systems to have confidence that the data they store in DNA can be reliably preserved and recovered
- Considering a number of ways to build this confidence in the ecosystem
- First Spec
  - Stability Evaluation Method for DNA Data Storage Containment Systems

A brief aside: How did the storage industry try to resolve endurance ratings for SSDs

JEDEC JESD 218/219

The SSD Endurance Problem, circa 2010

In 2008-2009, SSDs were still new



And a question was growing re: SSDs:

- I hear SSD's "wear out"; what does that mean and how long does it take?
- How do I compare wear out claims (endurance) between vendors?

JESD 218/219: SSD Reqt's and Endurance Test Method

Establish standard reference use conditions, success criteria, and metric

First we needed a standard metric

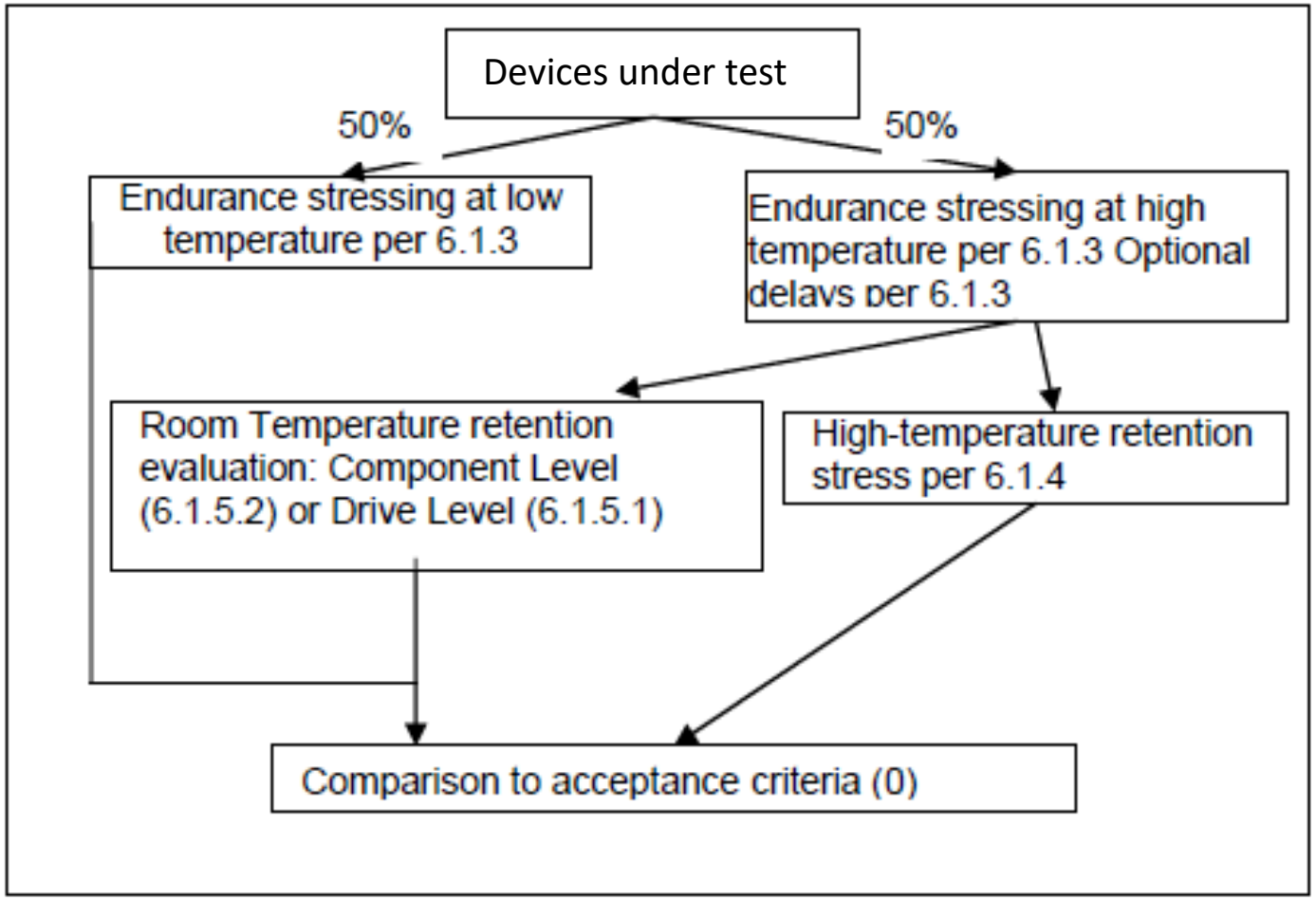
TBW: Terabytes of data that can be written to a drive over its 3-5 year lifetime, assuming a standard set of Reference Use Conditions (Workload, Temp, Duty Cycle) and meeting standard Success Requirements (FFR & UBER)

	Reference Use Conditions Over Drive Lifetime			Success Requirements	
Application Class	**Standard Workload (JESD 219)**	**Active Use Period (power on)**	**Retention Period (power off)**	**Functional Failure Rate (FFR)**	**UBER**
Client	Trace based	40°C 8hrs/day	30°C 1yr	≤3%	10^{-15}
Enterprise	SPC based	55°C 24hrs/day	40°C 3 months	≤3%	10^{-16}

JESD 218/219: SSD Reqt's and Endurance Test Method

Accelerated Wear: Simulate lifetime of wear via temperature acceleration

- Based on NAND flash activation energy assumptions and Arrhenius extrapolation

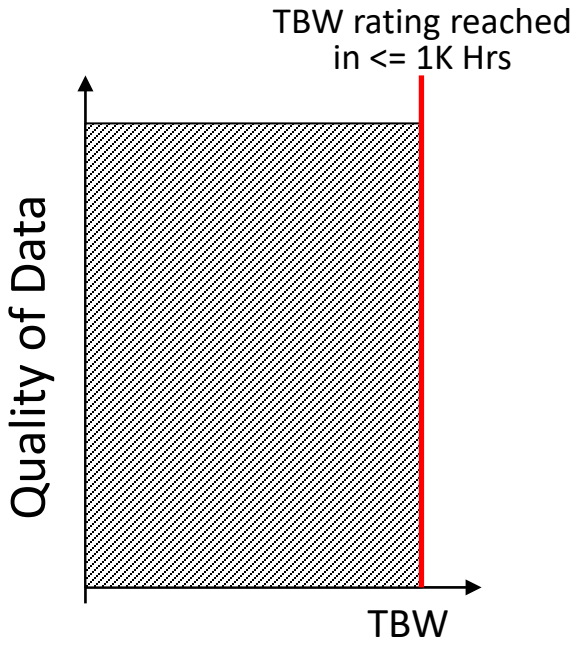


JESD 218/219: SSD Reqt's and Endurance Test Method

Accelerated Wear: Direct Method

When TBW can be reached in <=1K hours

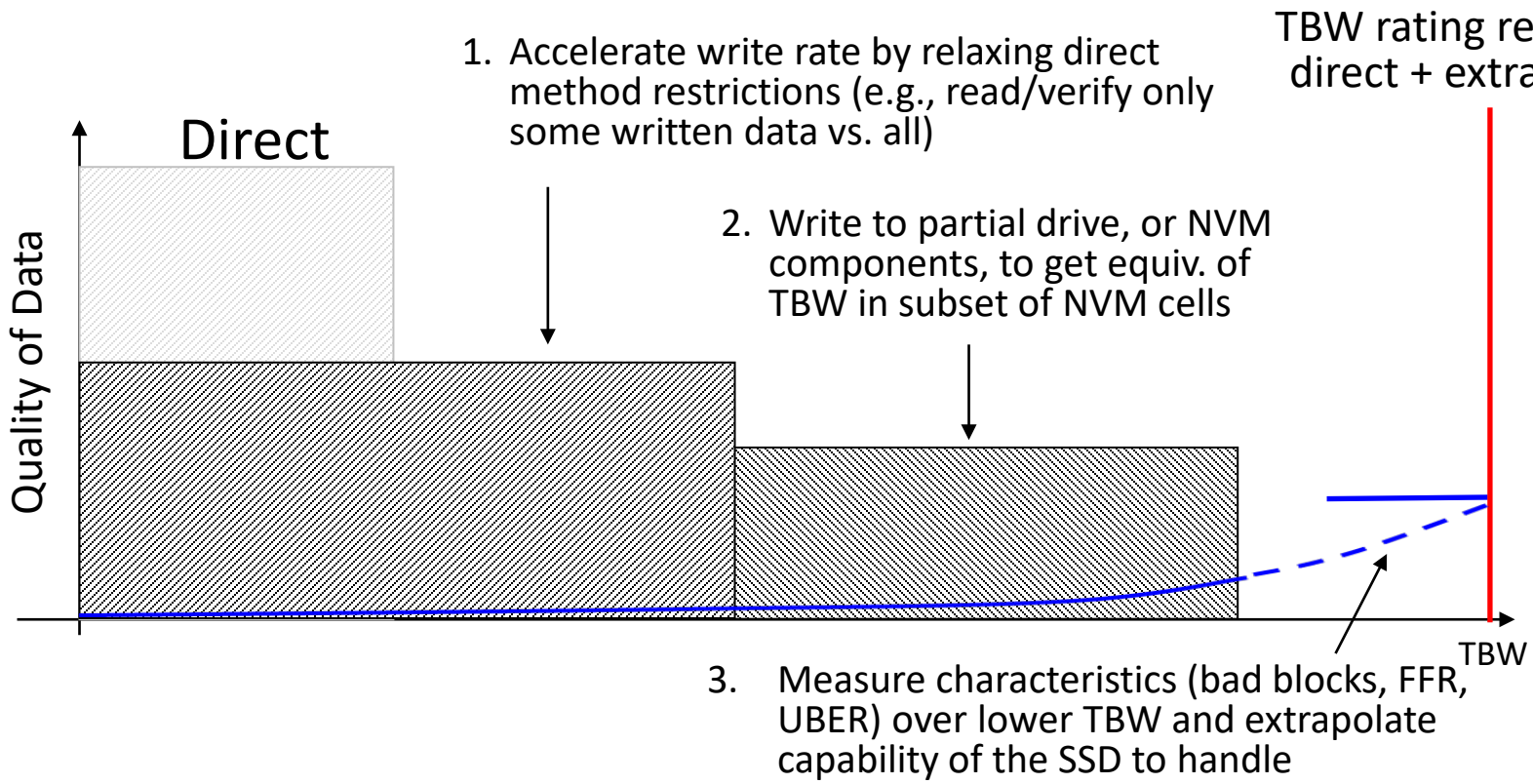
- Whole drive is stressed in normal operation
- Few assumptions: Failures simply counted
- No model-based "predictions" or assumptions



JESD 218/219: SSD Reqt's and Endurance Test Method

Accelerated Wear: Extrapolation Methods

When TBW cannot be reached in <=1K hours



- Visibility into higher TBW
- Tradeoff in data quality, number of assumptions

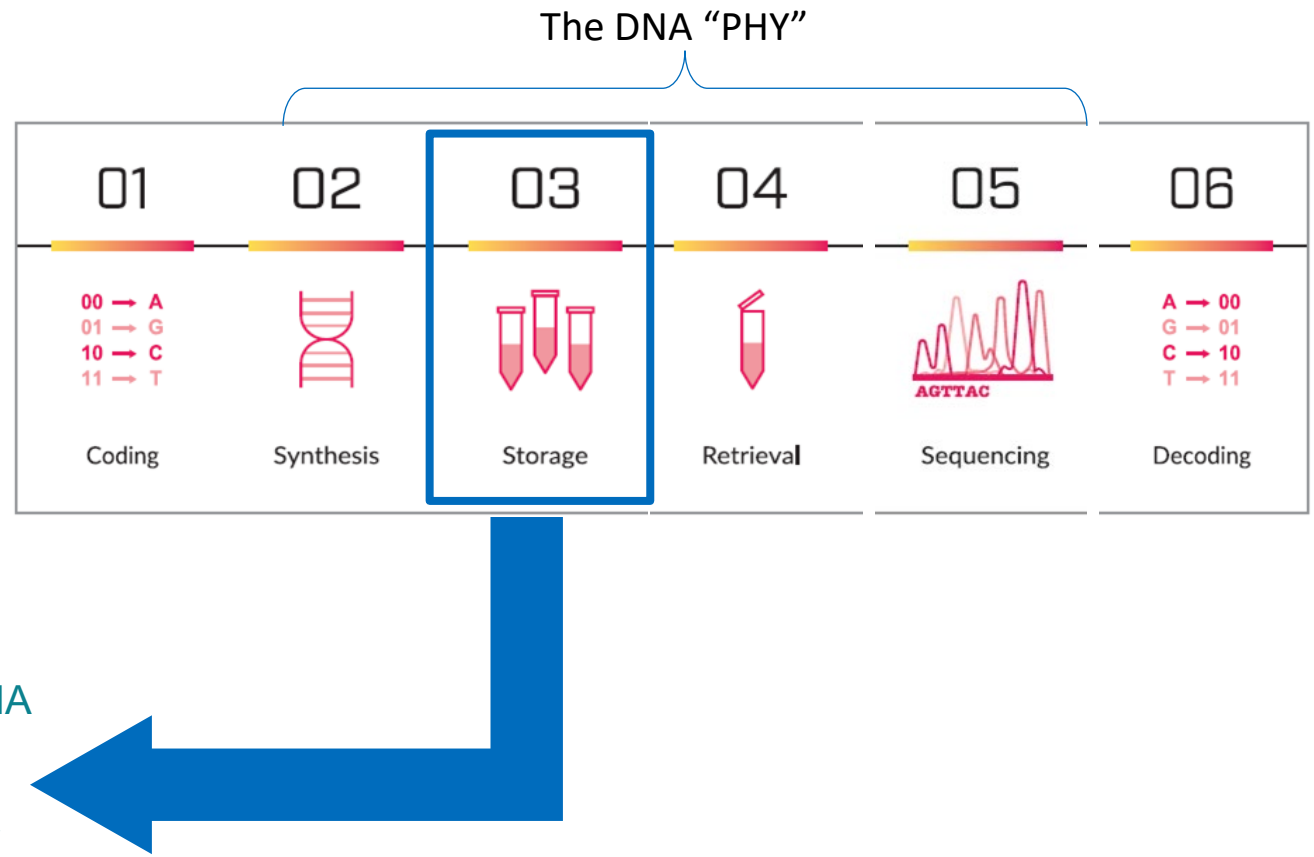
Comparing drives to each other

- If we run drives from different vendors using same assumed lifetime conditions and well defined wear acceleration, then if Drive A TBW = x and Drive B TBW = y, we know the two numbers are comparable.

So how does comparing SSD endurance and data retention relate to doing something similar w/ DNA?

Comparing SSD and DNA Data Storage cases

- With DNA, there is no integrated storage device
  - Writer (Synthesis) is separate from Reader (Sequencing)
- If one takes a use case/workload approach, as with SSD, the results must be evaluated end-to-end
  - Today, no standard implementation of the DNA "PHY" (Synthesis, Retrieval, Storage, Sequencing)
  - DNA encoding and decoding, anticipating and mitigating PHY errors, is still PHY-specific
- Well encapsulated DNA degrades very slowly
  - No "direct methods" will be able to count enough errors; Arrhenius curve extrapolation is a given
- We picked molecular stability rating as first goal:
  - Define accelerated wear methodology to characterize DNA molecular breakdown rate
  - Enables DNA containment/preservation system vendor claims to be compared, apples-to-apples, independent of coding/synthesis/retrieval/sequencing/decoding.

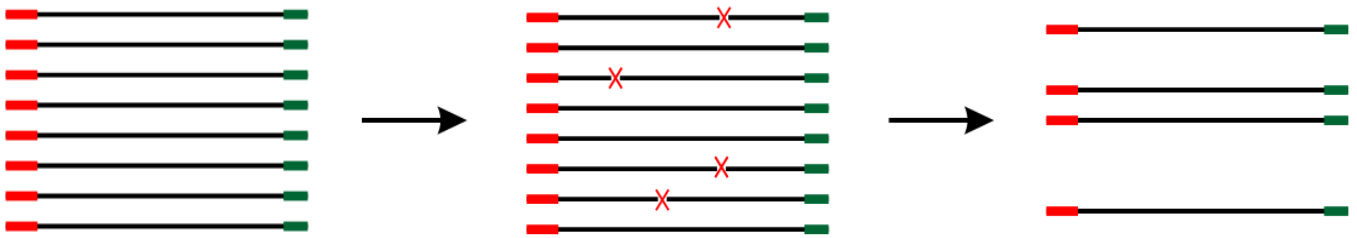


Our premises

- If a strand survives storage with no chain breakage (i.e., it is able to be amplified) the data stored in it will be recoverable
- Detection of chain breakage can be monitored and measured (e.g., by qPCR), since a broken chain can no longer be amplified
- Errors caused by storage conditions can be considered independent of errors caused in the rest of the DNA data storage pipeline

➢ And thus: We can make apples-to-apples comparisons between DNA data storage preservation methods by characterizing half-life of molecular breakdown in storage, as manifested by chain breaks

  ➢ That is, for a particular preservation method, how long till only 50% of the chains remain to be amplified?



initial population

degradation:
cuts in 50 % of the
population

50% of the strands
left to amplify

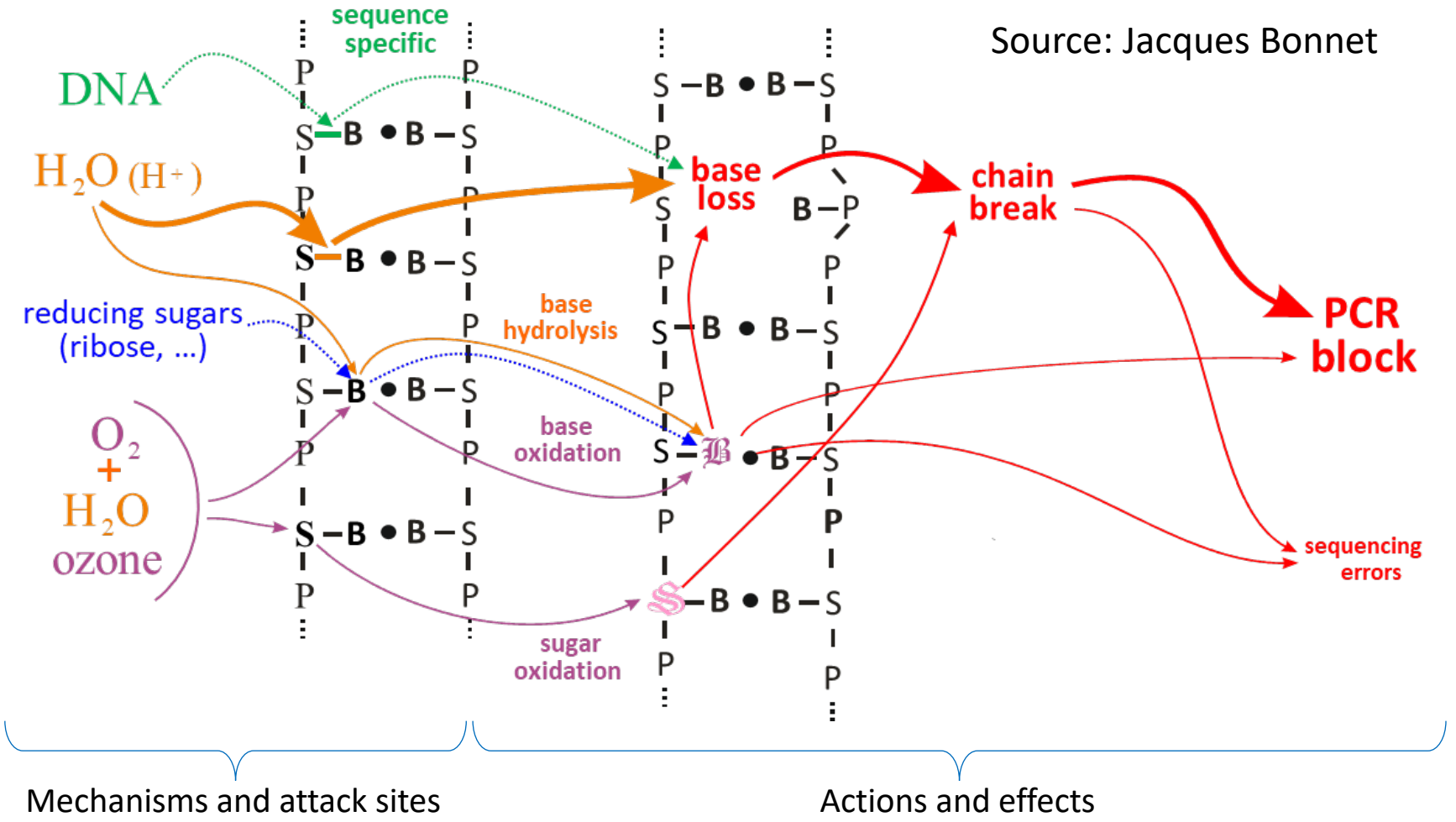
DNA Stability in Storage

Mechanisms of Molecular Breakdown

Main, in vitro, non-enzymatic, DNA degradation factors

(State diagram, 'ish)

- Degradation events leading to PCR Block are predominantly due to precursor events which cause base loss (generally through depurination), leading to chain breakage
- H2O is involved in nearly all of these degradation events directly, or indirectly



Storage methods used/described in literature

- Drying is the most prevalent mechanism
- Methods offering full protection from atmosphere (i.e., relative humidity) yield very long stability
- New methods being explored all the time

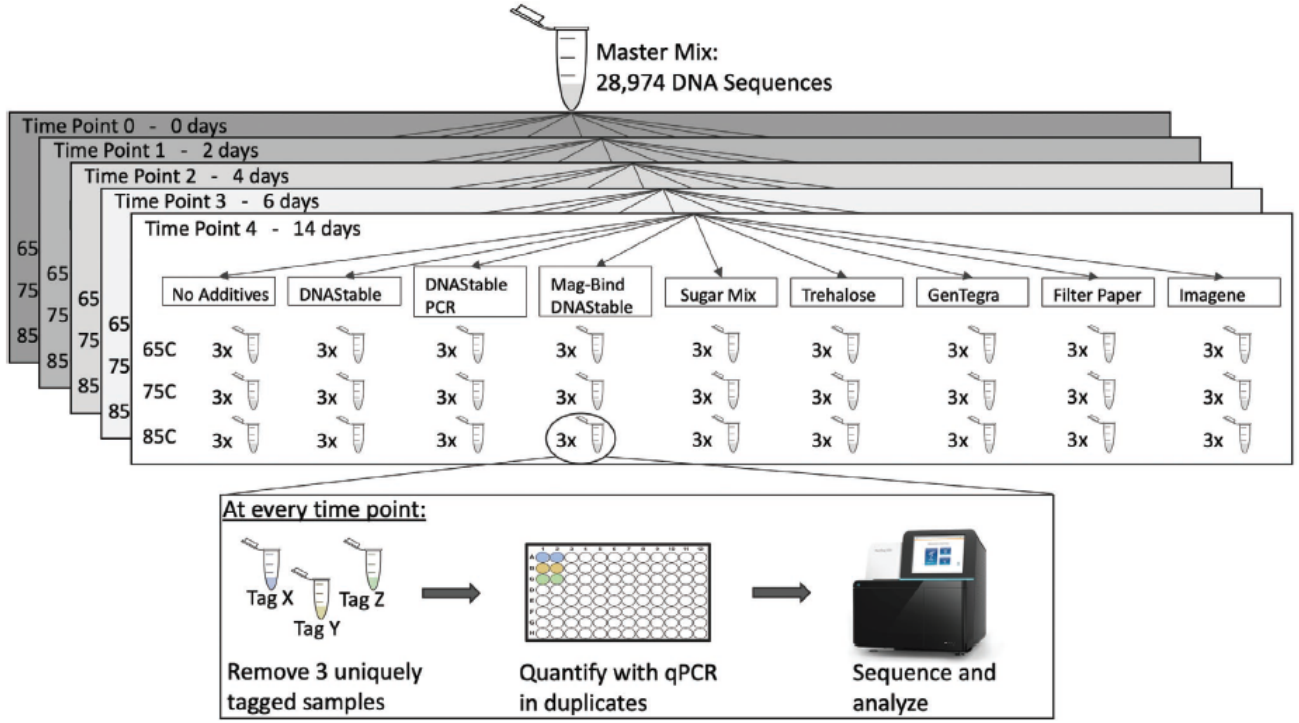
Principle	Procedure		Drying	Complete protection from atmosphere	Stability estimation
Chemical encapsulation	Moha bones	[8]			
	Fox teeth	[9]			
	Encapsulation in salts	[12, 16]	✓		Accelerated aging
Physical encapsulation	Silica nanoparticles	[1]		✓	Arrhenius
	Stainless steel capsules	[3]	✓	✓	Arrhenius
	Magnetic silica nanoparticles	[13]		✓	Accelerated aging
Inclusion in a matrix	DNAstable	[1, 21]	✓		Arrhenius
	Gentegra DNA	[1, 22]	✓		
	Pullulan	[14]	✓		
	Silk	[15]	✓		
	300K matrix inclusion	[25]	✓		
Absorption on paper	FTA paper	[1, 23, 24]	✓		Arrhenius
	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	Bacteria	[29]			

Source: Jacques Bonnet, Marthe Colotte

Some studies on DNA degradation

Grass et al 2015 [1], Organick et al 2021 [6], Coudy et al 2021 [7]

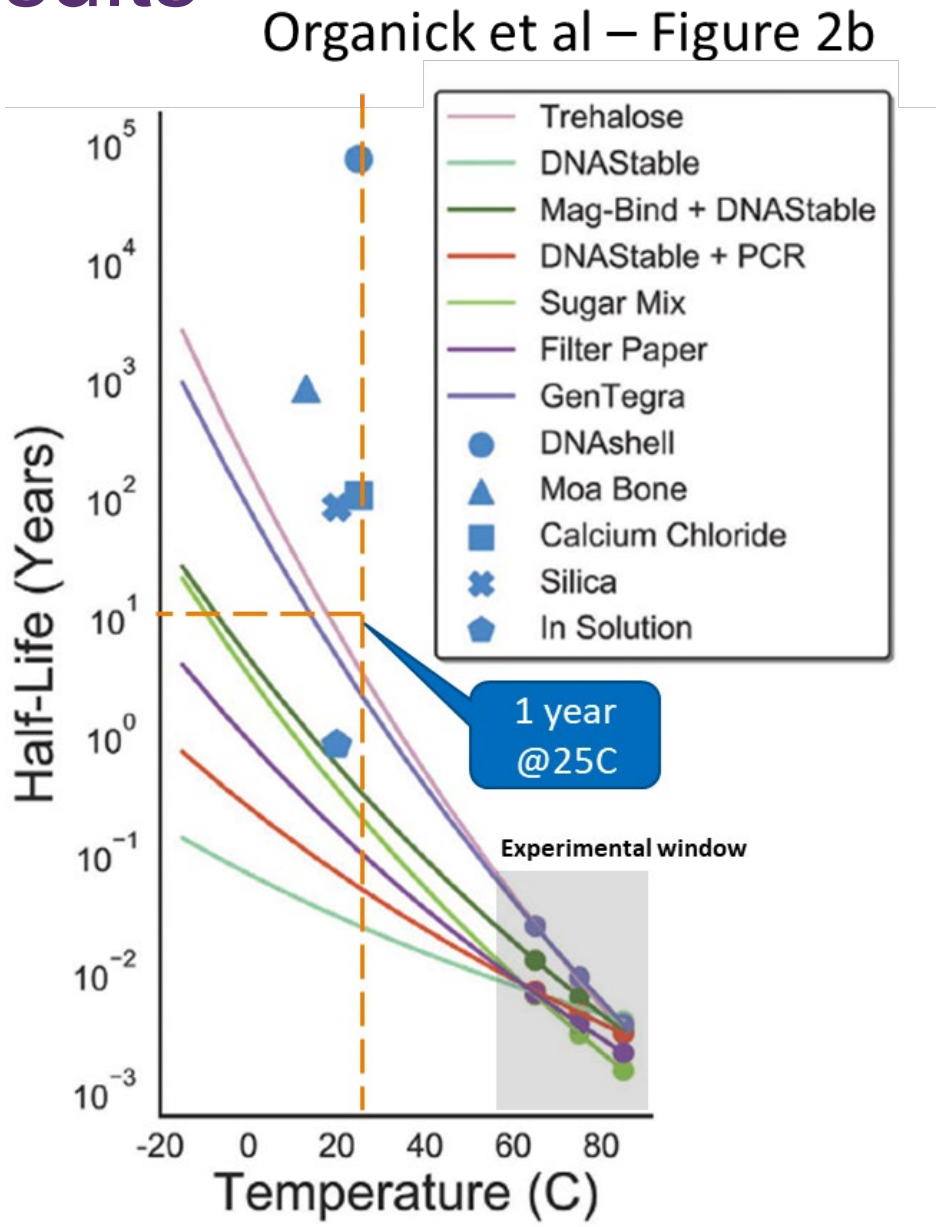
- Same general method
  - Accelerated wear
  - Control for time, temp and humidity
  - Grass and Organick encoded digital data in synthetic DNA and did sequence analysis after retrieval from storage



Organick - Figure 1

Overview of degradation results

- From Organick et al; they also included results from other studies (blue shapes)
  - Grass et al [1] – Silica
  - Allentoft et al [9] – Moa Bone
  - Bonnet et al [3] – Calcium Chloride
  - Coudy et al [7] – DNA Shell

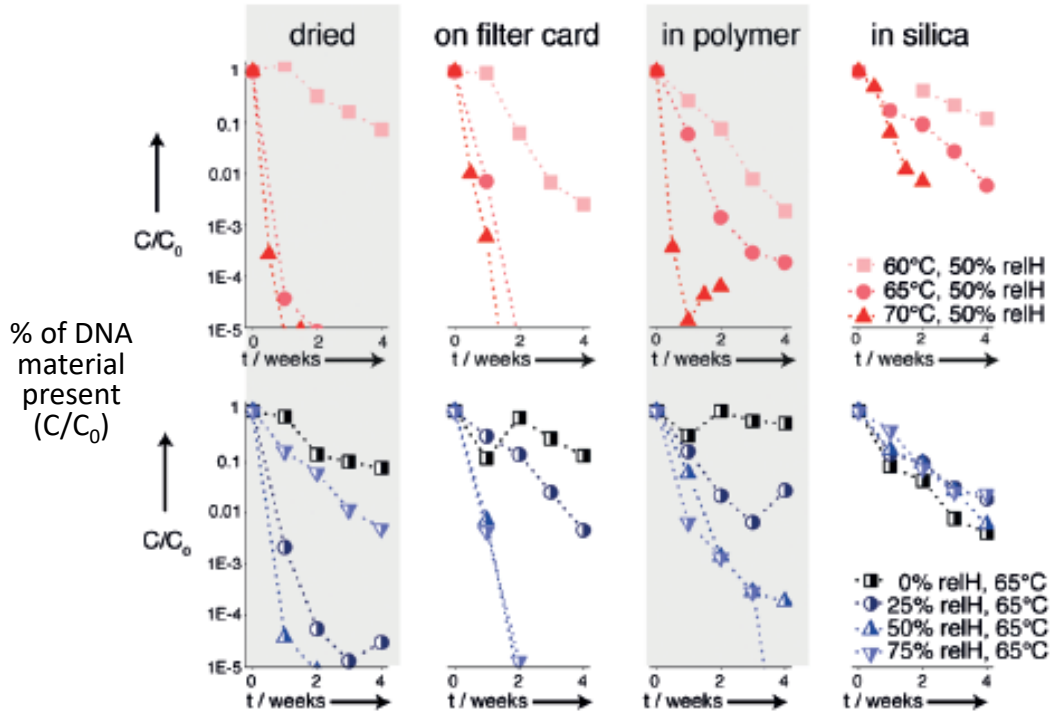


Grass et al [1]

Can we recover all uncut strands?

Figure 2a

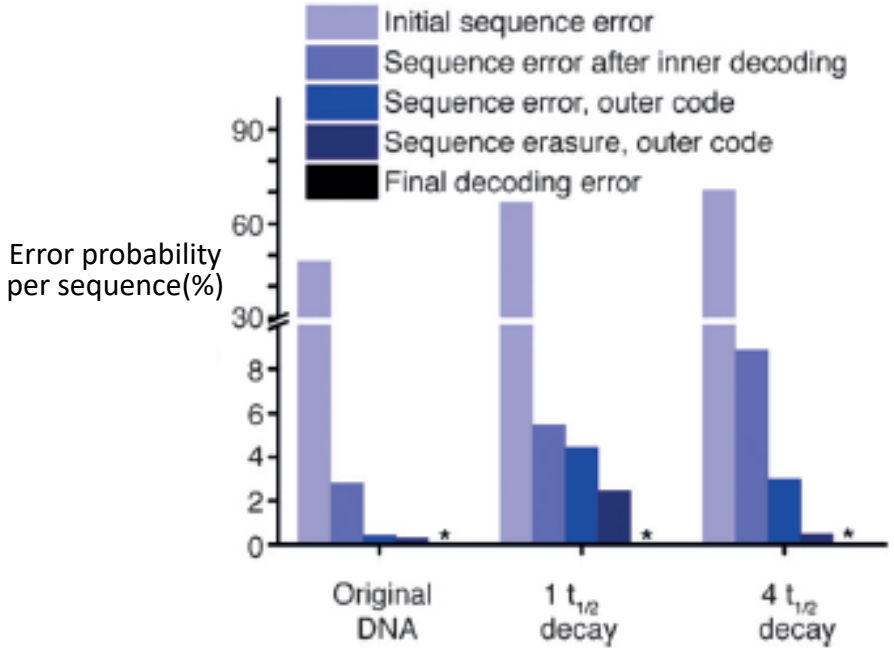
Degradation kinetics of dry DNA storage



- Clear advantages for Silica encapsulation

Figure 3

Recovering original data from silica substrate

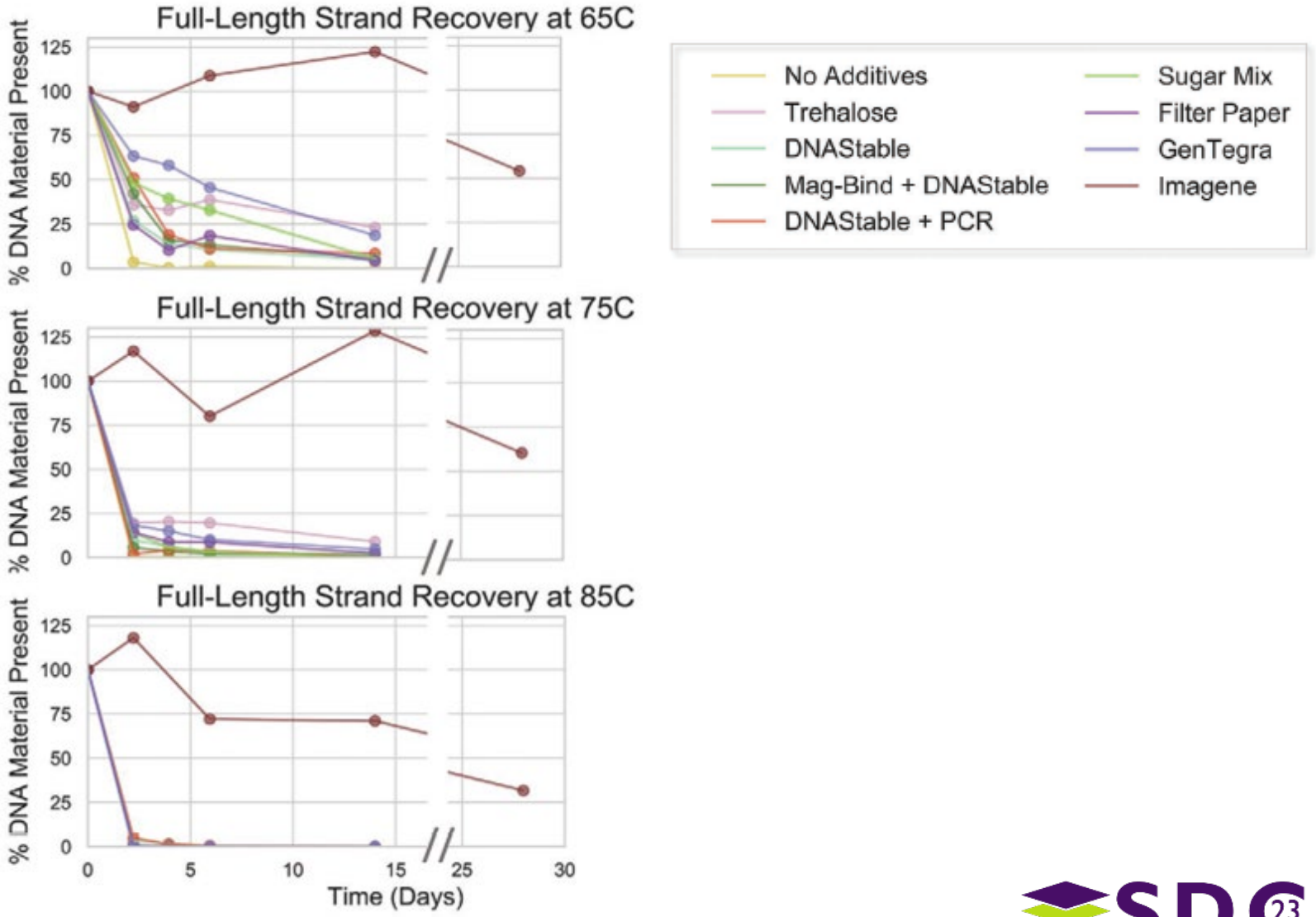


- Strands which were not cut during storage remained undamaged enough to enable full data recovery

Organick et al [6]

- Added substantially to the list of substrates and methods evaluated for durably storing DNA
  - Accelerated at three temps
  - Humidity held at 50%
- Begins setting stage for a standard evaluation methodology

Figure 2a

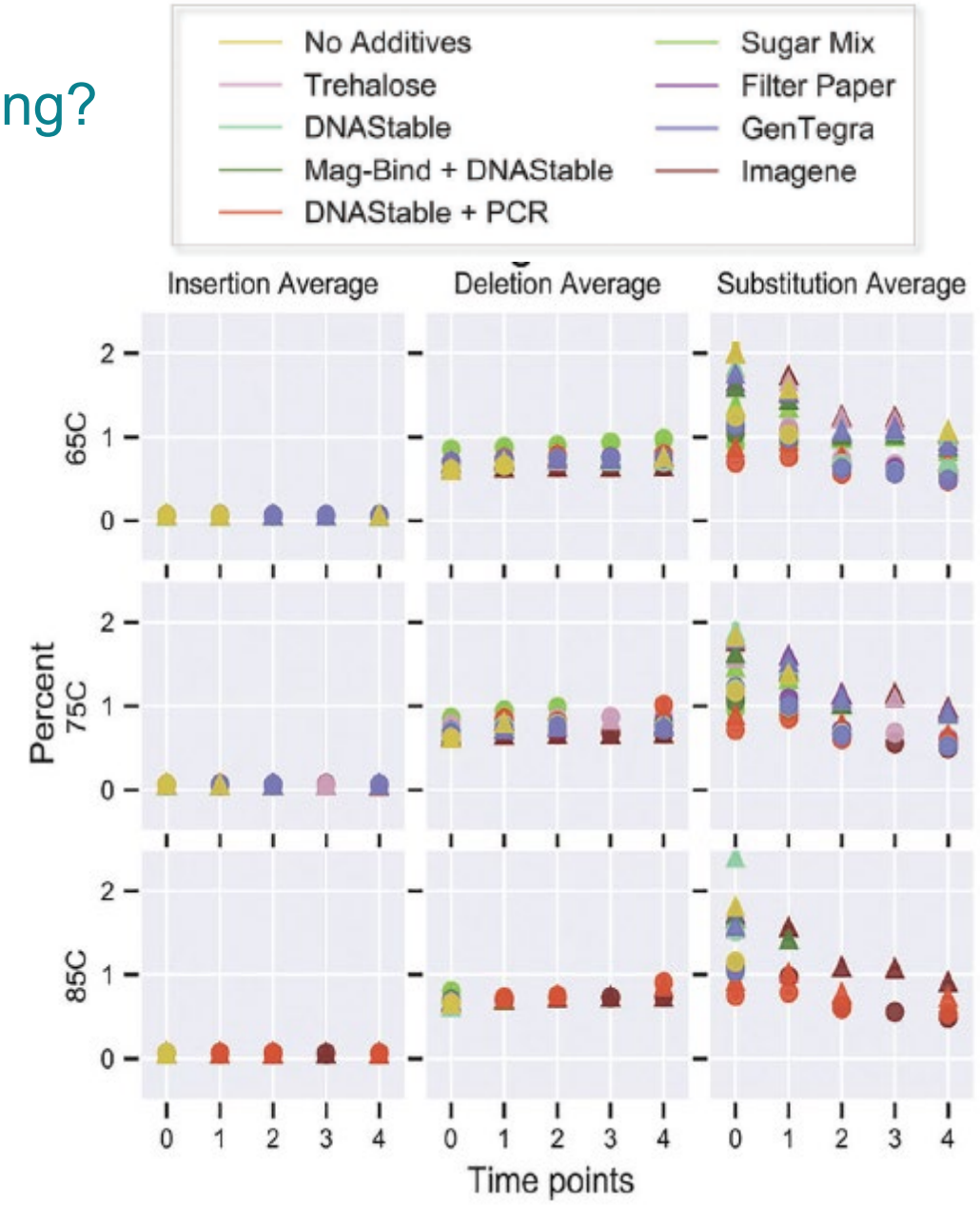


Organick et al [6]

Do degradation errors in storage affect sequencing?

- Minimal variation in error rates across methods, temperatures and time points, and even substitutions, which show most variance, show this variance before any aging
  - Aa authors concluded, “Suggests that insertion, deletion, and substitution errors are independent from the storage method”

Figure 4

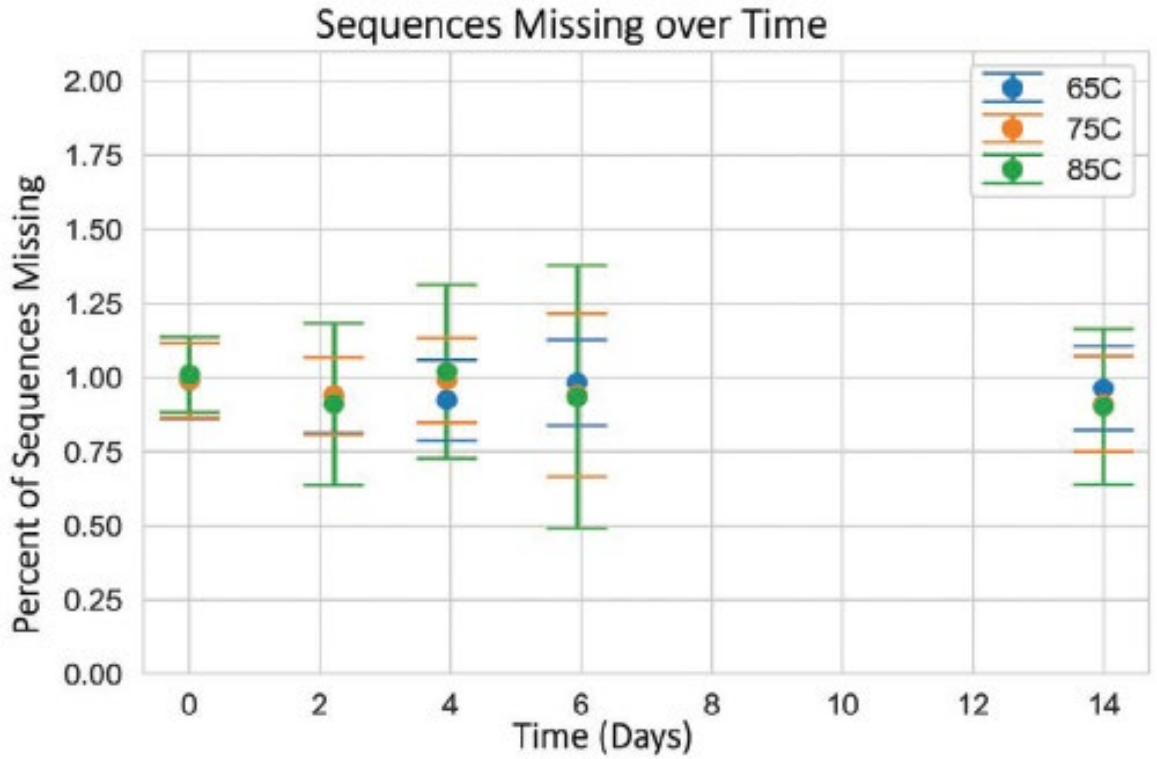


Organick et al [6]

Do specific sequences cause storage errors?

- Total # of sequences found missing during sequencing (across all methods, all time points, all 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”)
- The “no-storage bias” finding is further supported by finding in study that individual sequences missing at a particular timepoint had > 90% probability of reappearing in other time points and being successfully sequenced

Figure 5a

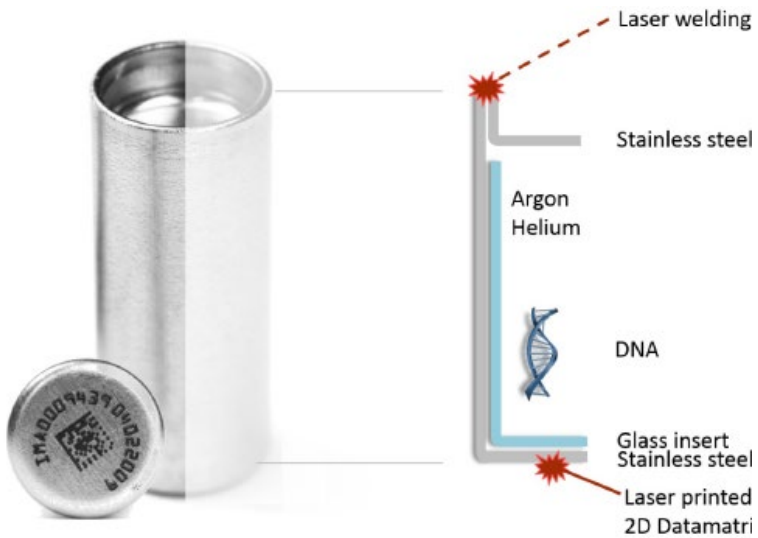


“If the total number of sequences missing increased after the pre-aging time point 0, we could hypothesize that there was some sequence-dependent degradation as more-vulnerable sequences degraded. However, we observed no difference in the number of sequences missing across all time points. This suggests that sequence loss is stochastic across all storage methods.”

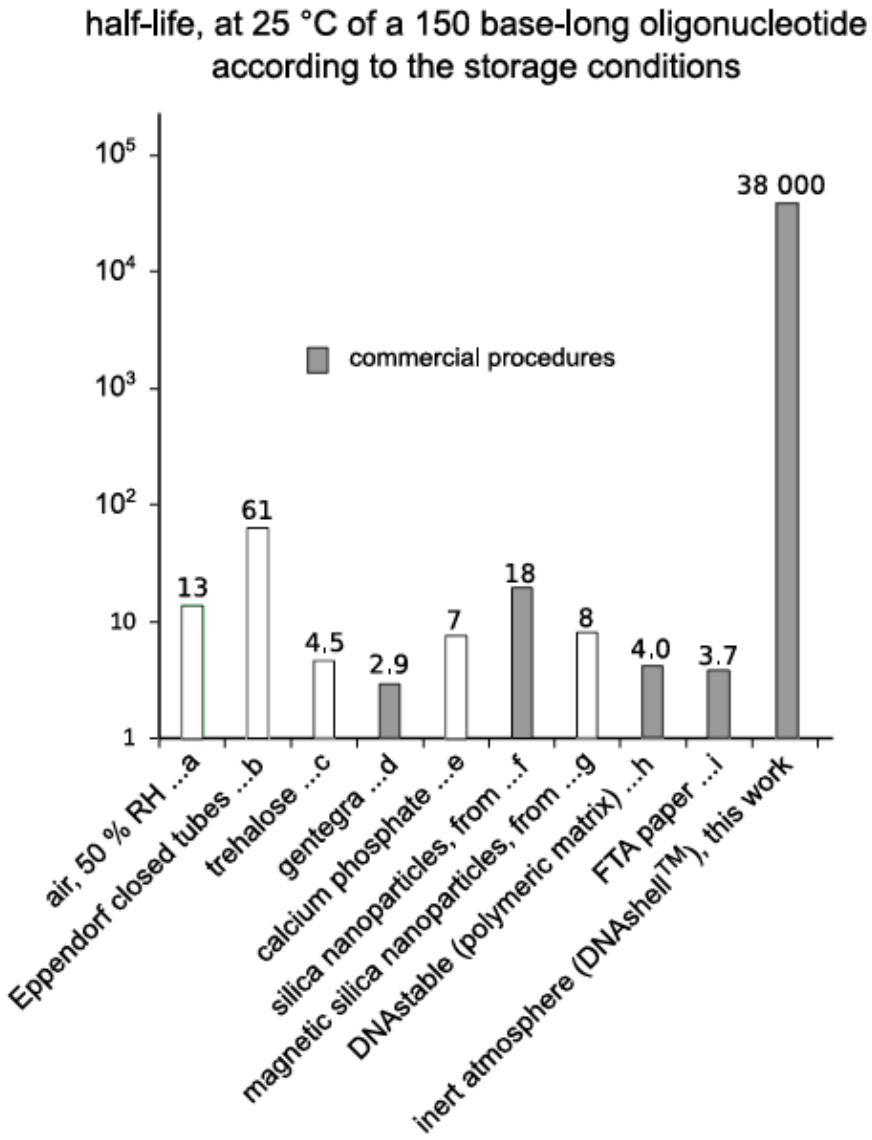
DNA sealed in inert atmosphere

Coudy et al [7]

- Built on previous studies; all measurements done under air or at controlled 50% relative humidity
  - Reinforced that methods which completely seal media from atmosphere yield very high durability



www.imagene.eu



Conclusions

- The data in strands that survived storage with no chain breaks was recoverable from those surviving strands
- Storage phase errors were shown to be independent of errors caused in other phases of the DNA data storage pipeline
- There was no sequence dependent degradation (i.e., storage bias) across preservation methods

➢Leading us to conclude that

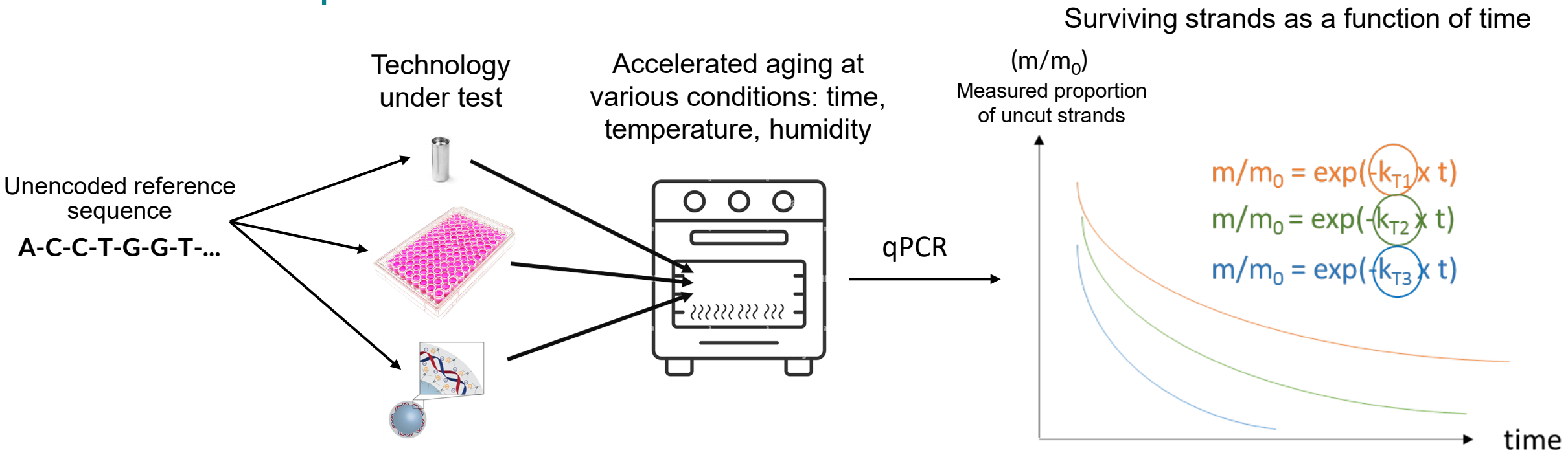
- We can make meaningful apples-to-apples comparisons between DNA data storage preservation methods by characterizing the half-life of molecular breakdown in storage, as manifested by chain breaks

Spec Proposal

Stability Evaluation Method for DNA Data Storage Containment Systems

DNA media degradation methodology

General Principles

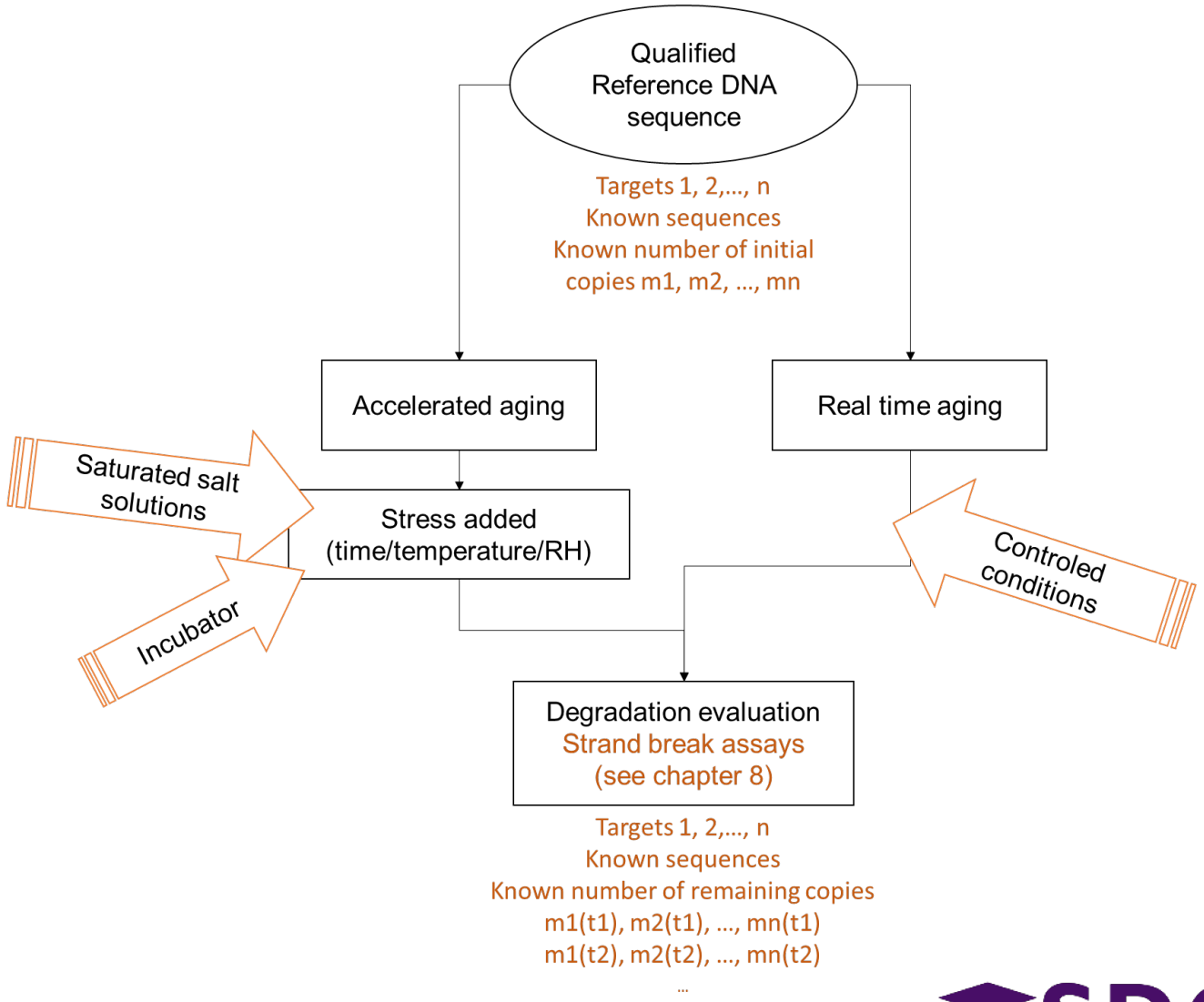


Metrics

- Half-life at 25C
- Fraction of intact strands (m/m0)
- Number of cuts / second / nucleotide

Draft flow for spec, and challenges/discussions

- Setting criteria for reference DNA sequence
- Activation energy varies between methods and we still don't know all the reasons
- Proposed spec captures most of the current methods
  - New methods constantly being researched which may require evolution of the spec



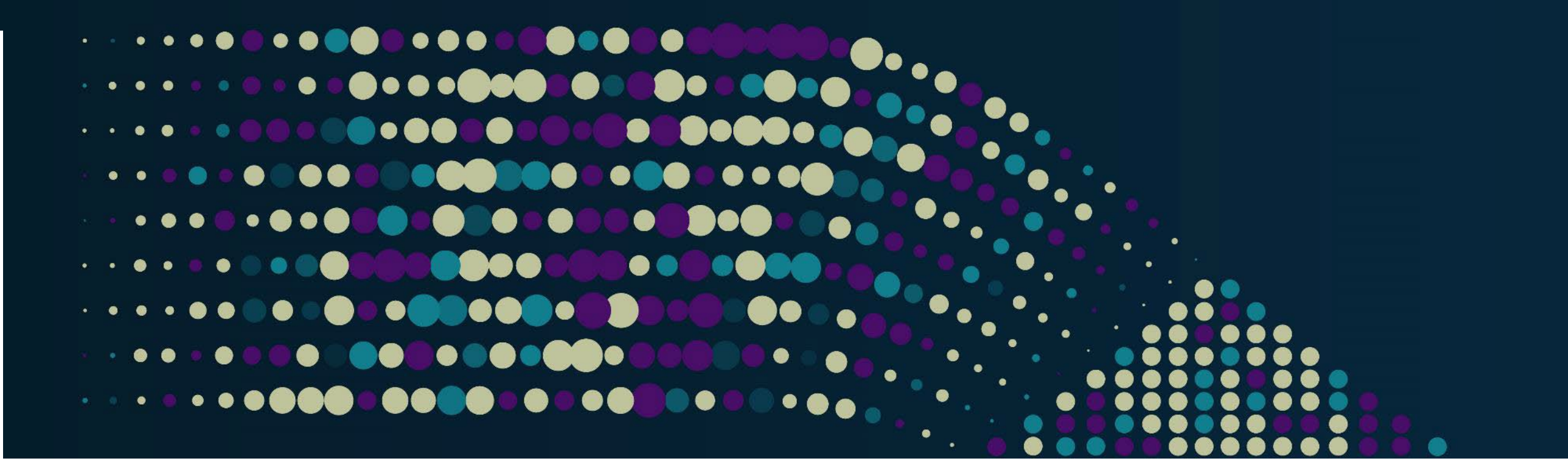
Bibliography

1. Grass RN, Heckel R, Puddu M, Paunescu D, Stark WJ. Robust Chemical Preservation 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 2022-07-19
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 Simultaneous assessment of average fragment size and amount in minute samples of degraded DNA. *Anal Biochem 345–347 388 2009*

Come Join Us

DNA Data Storage Alliance
Data Retention Workgroup



Please take a moment to rate this session.

Your feedback is important to us.

