DNA Data Storage
Introduction & Status
Introducing DNA’s benefits / What DNA is and what it is not

- DNA is a new, complementary cold layer in the storage pyramid
- Initially, DNA can be the ideal medium for a 3rd copy (offline, secured, durable)
- It can last thousands of years with the right protection / packaging
- As natural DNA, it will always be readable
- No migration and minimal maintenance costs = attractive TCO
- DNA is a great way to diversify your media, mitigating the risk of a single tape supplier

- DNA is not a tape replacement (Sorry about that!)
- DNA is not going to store all the world’s data in a shoebox
- DNA is not a hot/warm storage medium
- DNA is not cheap (yet)
- DNA is not coming to your nearby data center in the next 2 years
## What we need to build

<table>
<thead>
<tr>
<th>Codec</th>
<th>Synthesis</th>
<th>Storage / Retrieval</th>
<th>Sequencing</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>No vendor lock-in</td>
<td>Synthesize MB</td>
<td>No cooling requirements</td>
<td>Sequence MB</td>
<td>Integration with storage management software</td>
</tr>
<tr>
<td>Error detection / correction</td>
<td>Synthesize GB</td>
<td>Safe &amp; durable medium</td>
<td>Sequence GB</td>
<td>Support object storage and S3 APIs</td>
</tr>
<tr>
<td>Synthesize TB</td>
<td></td>
<td>Easily copied for geographical distribution</td>
<td>Sequence TB</td>
<td>Random Access</td>
</tr>
<tr>
<td>TB &lt; $1000</td>
<td></td>
<td>Compact footprint</td>
<td>TB &lt; $1000</td>
<td>On-Premise deployment</td>
</tr>
<tr>
<td>Throughput of TBs/day</td>
<td></td>
<td>No manual handling</td>
<td>Non-destructive sequencing</td>
<td>Automated storage system (e.g., Tape Library)</td>
</tr>
<tr>
<td>Always accessible</td>
<td></td>
<td>Adjustable &quot;Automation ready&quot; storage container</td>
<td>TBs/day throughput</td>
<td>“Fixity checks”</td>
</tr>
<tr>
<td>(Natural DNA)</td>
<td></td>
<td></td>
<td></td>
<td>Data Center “friendly” deployment</td>
</tr>
<tr>
<td>Enzymatic Synthesis</td>
<td></td>
<td></td>
<td></td>
<td>Data Center “friendly” deployment</td>
</tr>
</tbody>
</table>
## What we have built

<table>
<thead>
<tr>
<th>Codec</th>
<th>Synthesis</th>
<th>Storage / Retrieval</th>
<th>Sequencing</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>No vendor lock-in</td>
<td>Synthesize MB</td>
<td>No cooling requirements</td>
<td>Sequence MB</td>
<td>Integration with storage management software</td>
</tr>
<tr>
<td>Error detection / correction</td>
<td>Synthesize GB*</td>
<td>Safe &amp; durable medium</td>
<td>Sequence GB</td>
<td>Support object storage and S3 APIs</td>
</tr>
<tr>
<td>Synthesize TB</td>
<td>Easily copied for geographical distribution</td>
<td>Sequence TB</td>
<td>Random Access</td>
<td></td>
</tr>
<tr>
<td>TB &lt; $1000</td>
<td>Compact footprint</td>
<td>TB &lt; $1000</td>
<td>On-Premise deployment</td>
<td></td>
</tr>
<tr>
<td>Throughput of TBs/day</td>
<td>No manual handling</td>
<td>Non-destructive sequencing</td>
<td>Automated storage system (e.g., Tape Library)</td>
<td></td>
</tr>
<tr>
<td>Always accessible (Natural DNA)</td>
<td>Adjustable &quot;Automation ready&quot; storage container</td>
<td>TBs/day throughput</td>
<td>&quot;Fixity checks&quot;</td>
<td></td>
</tr>
<tr>
<td>Enzymatic Synthesis*</td>
<td></td>
<td></td>
<td>Data Center “friendly” deployment</td>
<td></td>
</tr>
</tbody>
</table>

* Proof of concept
Solving The Rest of The Puzzle

DNA Data Storage
Technology Progress
Goal is to develop a chip that produces 1 TB of coded DNA

- DNA is synthesized on a chip
  - Use a 2D array of electrochemical reactors to synthesize strands of DNA
  - After synthesis, the DNA is washed into a tube, then amplified, purified, and packaged

- Chip capacity is limited by the array pitch and chip size
  - There is a scaling limit; each reactor needs to produce enough DNA to practically store
  - Given the scaling limit, 1 TB from a chip is the practical limit – otherwise the chip becomes too large

- Twist’s chip capacity roadmap
  - 62.5 GB → 250 GB → 1 TB
  - *We are working on the 62.5 GB chip*

- Synthesis cost drivers
  - At scale, the fixed cost is averaged over many units
  - Variable cost is driven by reagent prices and reaction volume – high density chips yield lowest cost
Packaging

• DNA degrades by oxidation
  – Hermetically packaging DNA leads to a long shelf life
  – The package can be checked periodically for leaks – no leaks, no degradation

• DNA is dense, but packaging needs to be practical
  – Industrial automation required for process steps
  – And tubes that can be laser welded shut

• Barcoded tubes can be packed in arrays
  – Arrays are configurable
  – Array sizes: 96 TB, 384 TB, or 1,536 TB per bio automation spec
Sequencing

- When synthesis is solved, sequencing becomes the most pressing challenge.
- Overall sequencing cost depends on reading frequency.
- Genomic sequencers are a non-starter; orders of magnitude too expensive.
- Multiple groups working on molecular electronics sequencing.
System

- So far, the focus has been on the components as enablers
- We have phenomenal technologies, but not yet a product
- The product would have to integrate seamlessly into the customer workflow
  - Operated by IT team / Offered as a service
  - Integrated with storage management software
  - Data in/out
  - Data Center Ready – Monitoring, Debugging, Alerts, etc.
  - Implement common APIs and File Systems
  - Fixity Check solution
Building the DNA Data Storage ecosystem

History
- Formed on October 12th, 2020 by Illumina, Microsoft, Twist and Western Digital
- More than 50 member organizations (Including leading storage vendors: Seagate, WD, Kioxia, Dell, Quantum, Fujifilm, Fujitsu, Spectra Logic and Microsoft)

Mission
- Create and promote an interoperable storage ecosystem based on DNA as a data storage medium
- Educate the DNA data storage market to create awareness and adoption
- Identify use cases in various markets/industries for the use of DNA data storage
- Develop an industry technology roadmap for DNA data storage
- Foster standards or specifications as needed by ecosystem

https://dnastoragealliance.org/