

DNA Writing Improvements for Data Storage

Devin Leake

CSO

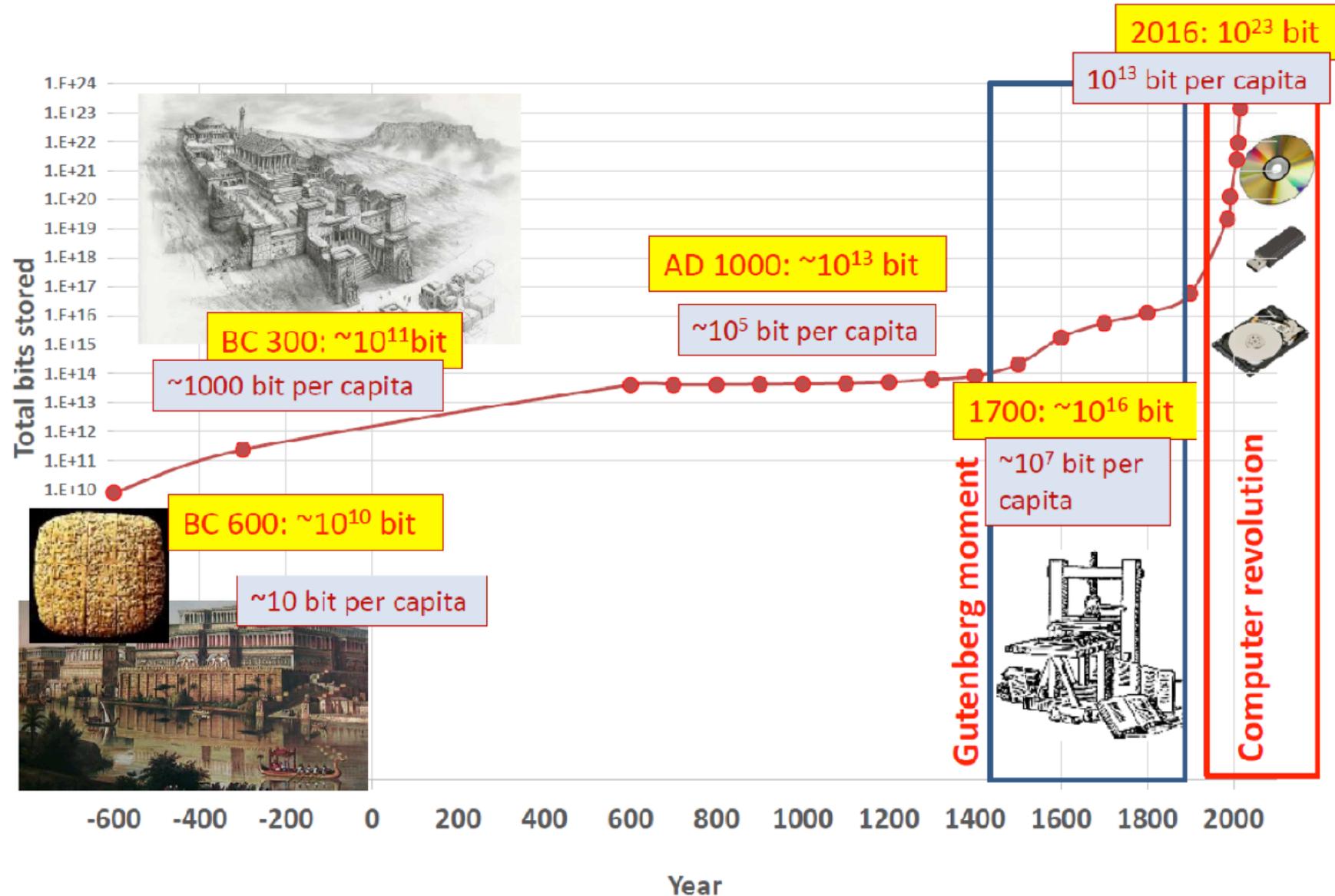
Devin_at_catalogdna_dot_com



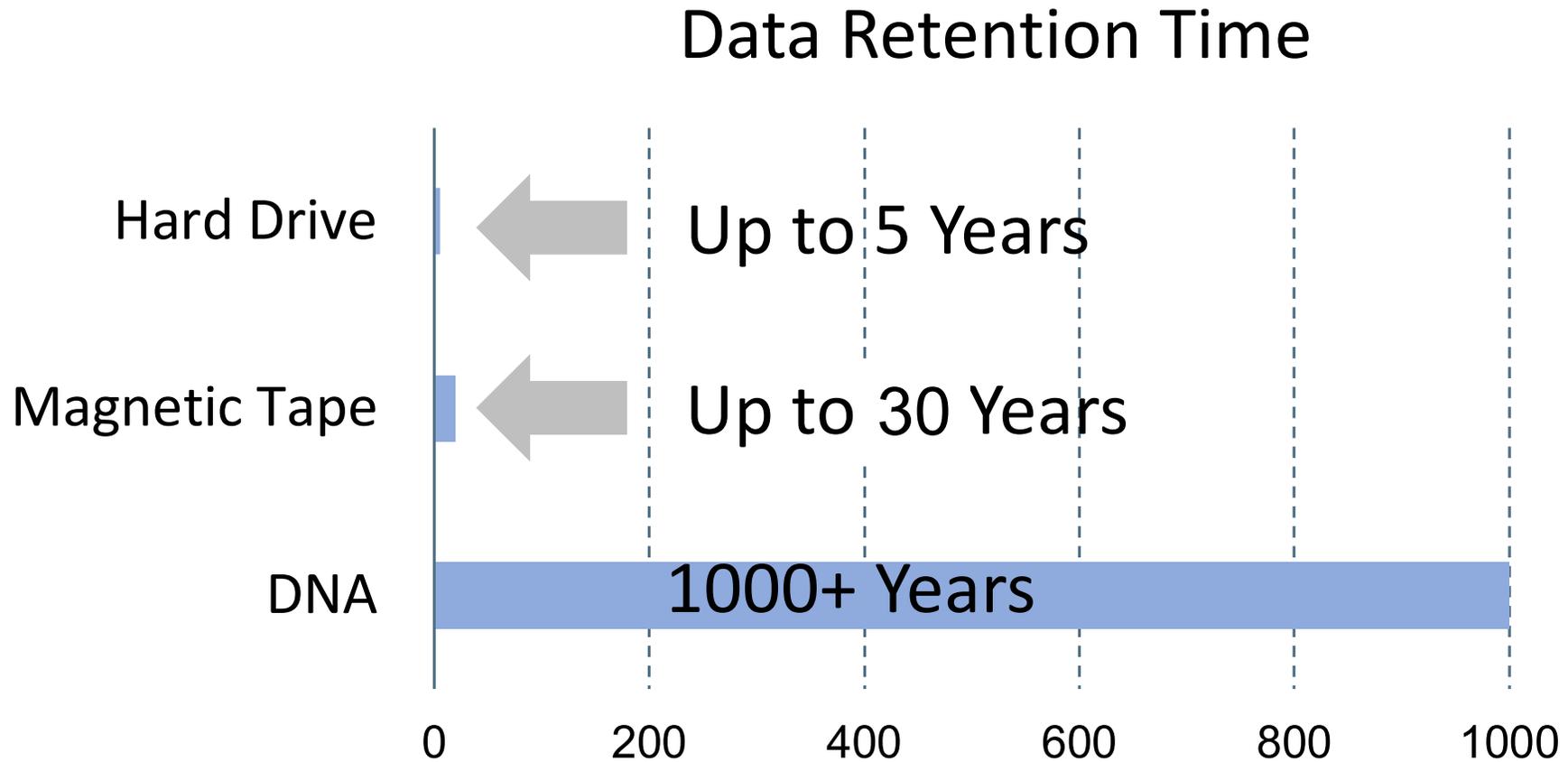
CATALOG
INFINITE DATA ARCHIVES

Data generated per capita is growing

Summary Report: Technology Working Group Meeting on future DNA synthesis technologies (September 14, 2017, Arlington, VA)



DNA offers advantages of extended lifetime and stability



DNA has significant storage capacity

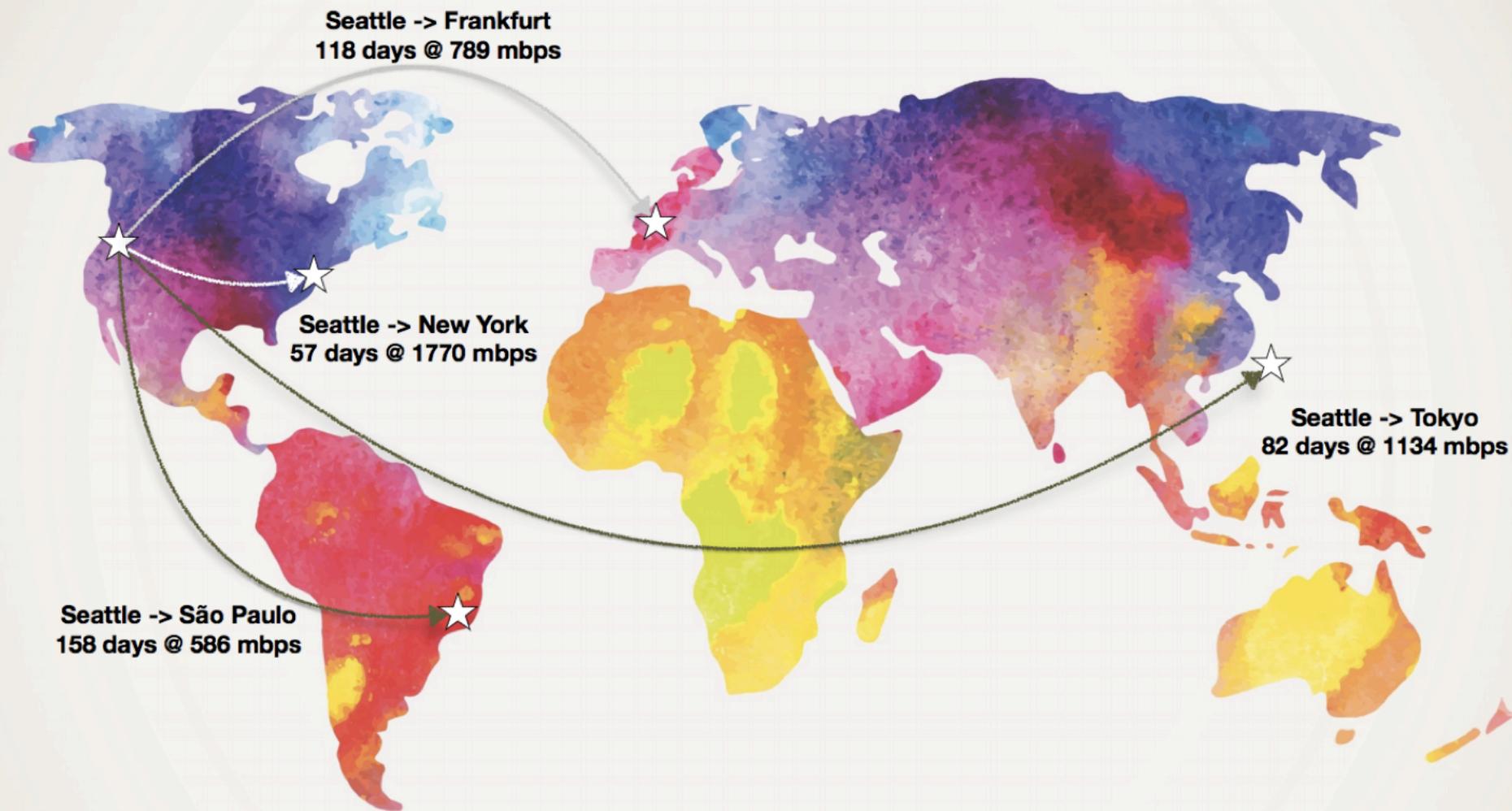
- Constraints of media are obstacles for increasing storage density
- Innovations improving traditional media increase cost

Medium	Information Density
Hard Disk	10 Terabits / cm ³
DRAM	10 Terabits / cm ³
Flash Memory	10 Petabits / cm ³
DNA	10 Exabits / cm ³



Time to Transfer One Petabyte Worldwide

DNA
provides
path to
replication
and
portability



DNA is exciting but still limited by cost and speed

- 200MB of digital information
- 13,000,000 DNA molecules
- 150 nucleotides per oligo

Scaling up DNA data storage and random access retrieval

Lee Organick¹, Siena Dumas Ang², Yuan-Jyue Chen², Randolph Lopez³, Sergey Yekhanin², Konstantin Makarychev², Miklos Z. Racz², Govinda Kamath², Parikshit Gopalan², Bichlien Nguyen², Christopher Takahashi¹, Sharon Newman¹, Hsing-Yeh Parker², Cyrus Rashtchian², Kendall Stewart¹, Gagan Gupta², Robert Carlson², John Mulligan², Douglas Carmean², Georg Seelig^{1,4}, Luis Ceze¹, Karin Strauss²

¹ Computer Science and Engineering Department, University of Washington, Seattle, Washington, 98195

² Microsoft Research, Redmond, Washington, 98052

³ Bioengineering Department, University of Washington, Seattle, Washington, 98195

⁴ Electrical Engineering Department, University of Washington, Seattle, Washington, 98195

Assuming \$0.0005/base ➤ cost would be ~\$1M



DNA is an exciting alternative but still limited by cost and speed

- 200MB of digital information
- 13,000,000 DNA molecules
- 150 nucleotides per oligo

Scaling up DNA data storage and random access retrieval

Lee Organick¹, Siena Dumas Ang², Yuan-Jyue Chen², Randolph Lopez³, Sergey Yekhanin², Konstantin Makarychev², Miklos Z. Racz², Govinda Kamath², Parikshit Gopalan², Bichlien Nguyen², Christopher Takahashi¹, Sharon Newman¹, Hsing-Yeh Parker², Cyrus Rashtchian², Kendall Stewart¹, Gagan Gupta², Robert Carlson², John Mulligan², Douglas Carmean², Georg Seelig^{1,4}, Luis Ceze¹, Karin Strauss²

¹ Computer Science and Engineering Department, University of Washington, Seattle, Washington, 98195

² Microsoft Research, Redmond, Washington, 98052

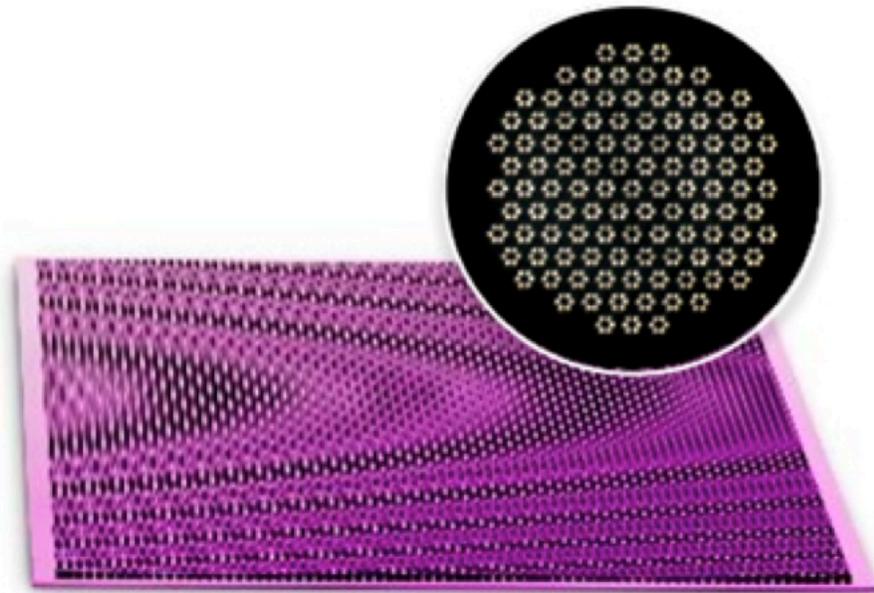
³ Bioengineering Department, University of Washington, Seattle, Washington, 98195

⁴ Electrical Engineering Department, University of Washington, Seattle, Washington, 98195

Assuming 360 second cycles ➤ encoding speed is ~3.7 kB/s



Can phosphoramidite chemistry speeds be increased and cost reduced?



> 1 Million Oligos → 9,600 Genes



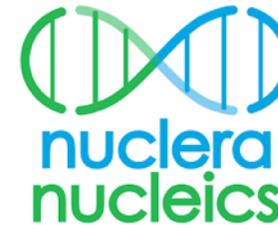
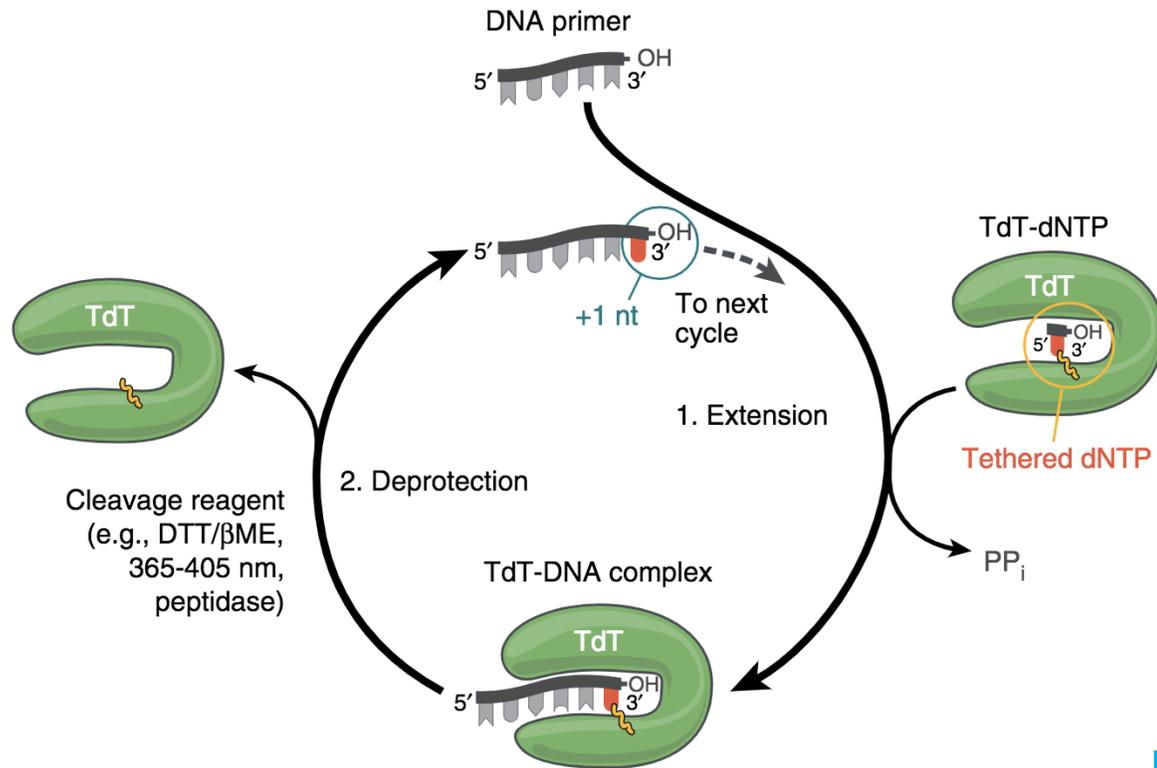
Agilent Technologies



GenScript
Make Research Easy

Current technology is still at least 4 orders of magnitude too slow and costly

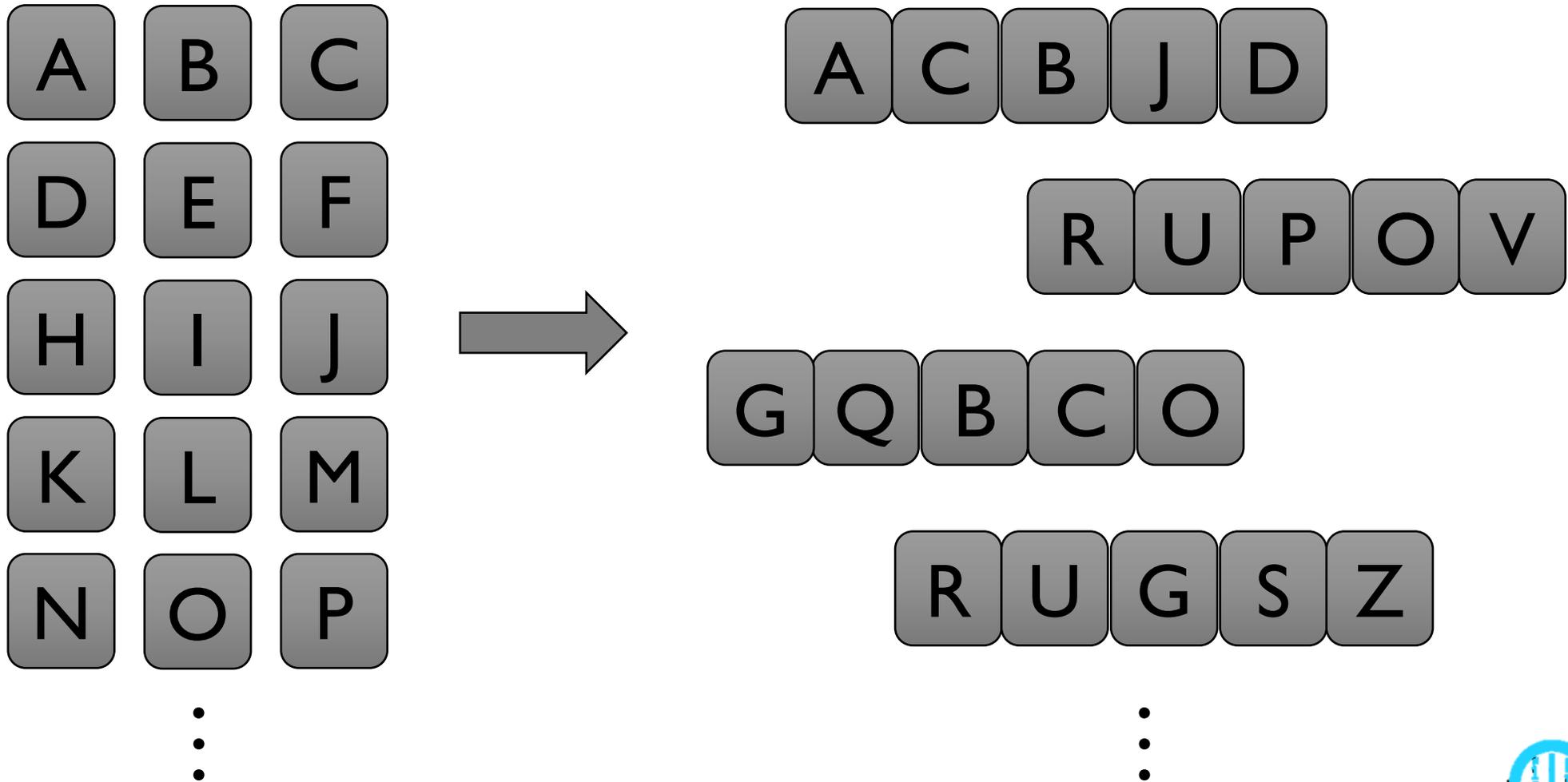
Can enzymatic synthesis be developed to be faster and cheaper?



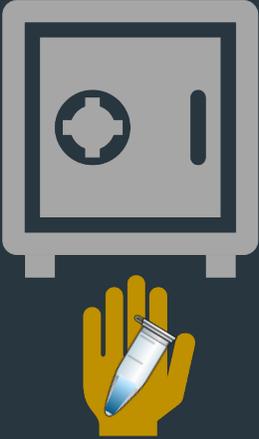
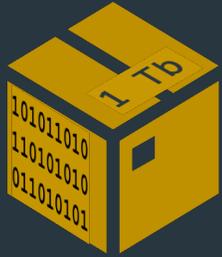
Palluk, et.al. Nature Biotechnology 2018

Development is still needed to produce oligos using enzymatic synthesis

Permutation scheme to generate entropy



Next step is developing an end-to-end workflow



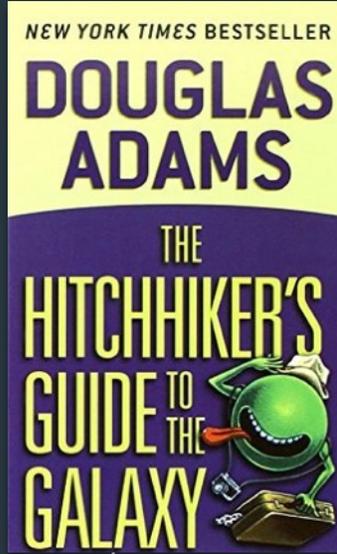
The Road Not Taken

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear;
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.



Generating 500 Billion
DNA constructs
per day!

Encoding capacity:
(bits / day)

(~1.5 GB/s)

Pb

(~1.5 MB/s)

Tb

Gb

Mb

Kb

2017

2018

2019

2022





CATALOG



Infinite Data Archives