OligoArchive – making DNA storage reality

Abstract

With Zetabytes of data in need to be stored for long-term archival already today, the EC funded project OligoArchive (https://oligoarchive.eu) will develop the fundamental technologies needed to make DNA storage a reality. Running for three years with a €3M budget, the project brings together all experts across Europe to build an end-to end prototype for DNA storage. The resulting technology will make DNA storage comparable to today's archival storage in terms of speed but will make long-term data archival more cost-effective.

DNA Storage Background

The demand for analysing archival data coupled with the need to retain data for regulatory compliance has resulted in a rapid increase in the amount of archival data stored by industry and academia. As data generation far outpaces the rate of improvement in traditional storage technology, finding a new storage media that can store at very low cost such archival data is pivotal.

Synthetic DNA is one such storage media that has recently received attention due to its high density and durability. DNA possesses four key properties that make it relevant for archival storage. First, it is an extremely dense three-dimensional storage medium which can store 455 Exabytes of data in 1 gram; in contrast, a 3.5" hard disk can store 10TB and weighs 600 grams today. Second, DNA can last several centuries even in harsh storage environments; traditional storage technology (e.g. HDD and tape) have lifetimes of five and thirty years. Third, it is very easy, quick, and cheap to perform in-vitro replication of DNA; for tape and HDD it takes hours or days for copying large Exabyte-sized archives due to their limited bandwidth. Fourth, given the demand for digital data storage, we will soon run out of silicon, while DNA is abundantly available to cover our storage needs.

Turning the Vision into Reality

The vision of this project is to develop an end-to-end prototype for DNA storage which enables storing and analysing arbitrary information in DNA. We will carry out the research to develop the fundamental building blocks needed in a series of scientific breakthroughs:

- Near-molecule data analysis: to develop approaches to analyse data stored in DNA using biomolecular techniques directly in storage. The approaches will be faster and more energy efficient compared to traditional computers thanks to the unprecedented parallelism.
- Accelerated sequencing: speeding up reading data from DNA storage by developing novel sequencing techniques for DNA storage.
- Optimal encoding for different types of data: novel, tuneable error correction for different types of data (e.g., imaging data can tolerate some errors whereas text cannot).
- Synthesis: novel synthesis methods to cheaply write data to DNA for storage (which can tolerate small imprecisions as opposed to DNA used for biological purposes).
- End-to-end automation: automatic translation from binary data to DNA, synthesis, data analysis, selective retrieval and sequencing to read back to binary information, all based on robotic equipment.

These building blocks enable us to build the first efficient end-to-end prototype for DNA storage comparable to today's archival storage in terms of speed but more cost-effective.

Consortium

The consortium brings together all expertise needed from across Europe:

Imperial College London (UK) – error correction and encoding documents in DNA, DNA computing and overall project lead

Université Nice/CNRS (France) – encoding images in DNA and accelerated decoding for DNA storage

Eurecom (France) – encoding databases in DNA

Helixworks (Ireland) – novel, cost-efficient synthesis techniques for data storage

Contact & More Info

Thomas Heinis t.heinis@imperial.ac.uk

Website: https://oligoarchive.eu