



HAL
open science

Storing the declaration of human rights on one data DNA molecule

Gouenou Coatrieux, Chloé Berton, Elsa Dupraz, Belaid Hamoum, Thomas Derrien, Yann Audic, Dominique Lavenier, Jacques Nicolas, Julien Leblanc, Olivier Boullé, et al.

► To cite this version:

Gouenou Coatrieux, Chloé Berton, Elsa Dupraz, Belaid Hamoum, Thomas Derrien, et al.. Storing the declaration of human rights on one data DNA molecule. CominLabs day 2022, Oct 2022, Rennes, France. pp.1-1. hal-03817852

HAL Id: hal-03817852

<https://hal.science/hal-03817852>

Submitted on 17 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

ABSTRACT

Today, the **community consensus** to store information on DNA is to use **short single strand DNA (ssDNA) molecules**. This approach has some limitations: encoding constraint, DNA stability, recovering DNA, sequencing technology, etc. To overcome them, we chose to store information on **long double-strand DNA (dsDNA) molecules**.

Our demonstration consists of storing the first articles of the declaration of human rights (4,2ko text document) on a single DNA molecule.

Declaration of Human Right storage features	Standard approach	dnarXiv approach
DNA structure	Single-strand (ssDNA)	Double-strand (dsDNA)
DNA strand length	100-200 nucleotides (nt)	25,000 base pairs (bp)
DNA molecules number	210 - 420 molecules	1 molecule
Information density	0,81 bits/nt	1,36 bits/nt
Error correction method	Yes	No
Sequencing technology	Illumina: short read	Nanopore: long read

ENCODING

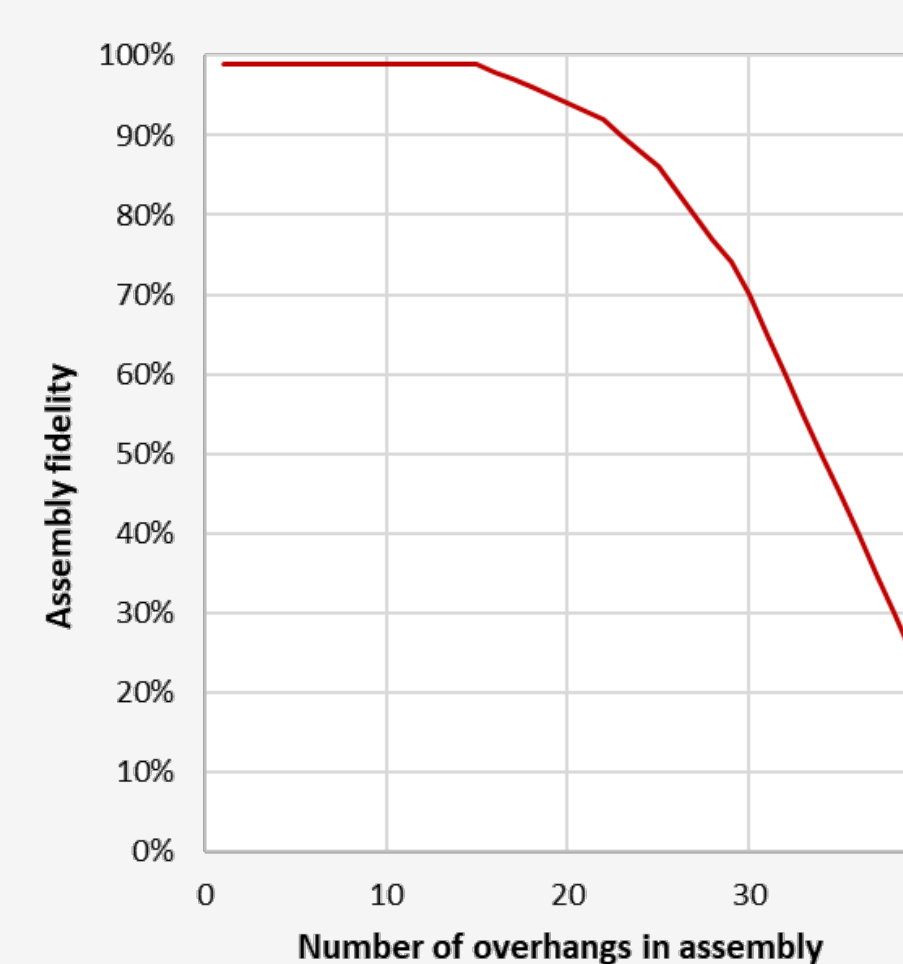
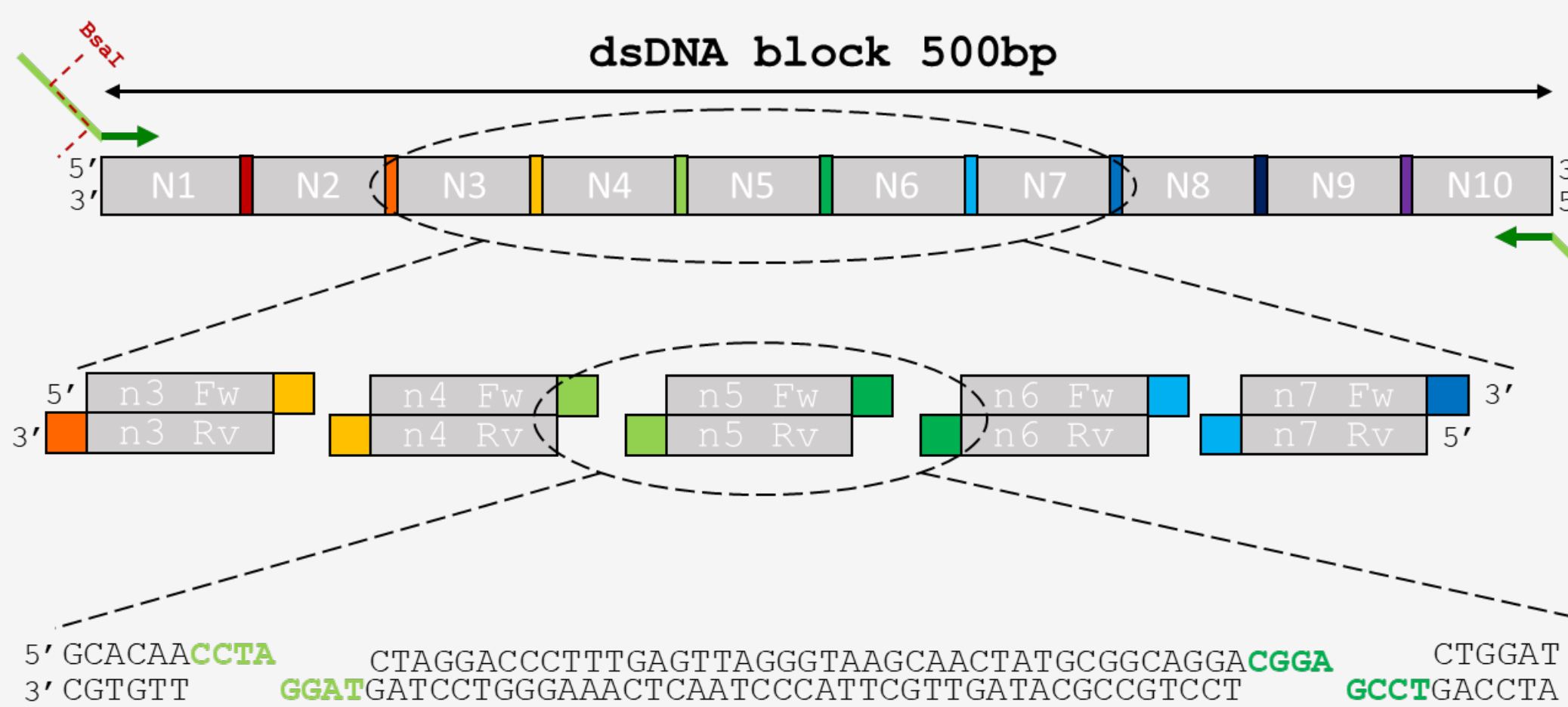
1 octet: 0 0 0 0 1 1 0 1

5 bases: α β α α β
A G G T C

α : $\begin{cases} 0 \mapsto A \\ 01 \mapsto G \\ 10 \mapsto T \\ 11 \mapsto C \end{cases}$ β : $\begin{cases} 0 \mapsto A \\ 1 \mapsto T \end{cases}$ if previous base $\in \{G,C\}$
 $\begin{cases} 0 \mapsto G \\ 1 \mapsto C \end{cases}$ if previous base $\in \{A,T\}$

- Input : **binary sequence**
- Code respecting **biological constraints**
- Output : **dna sequence**

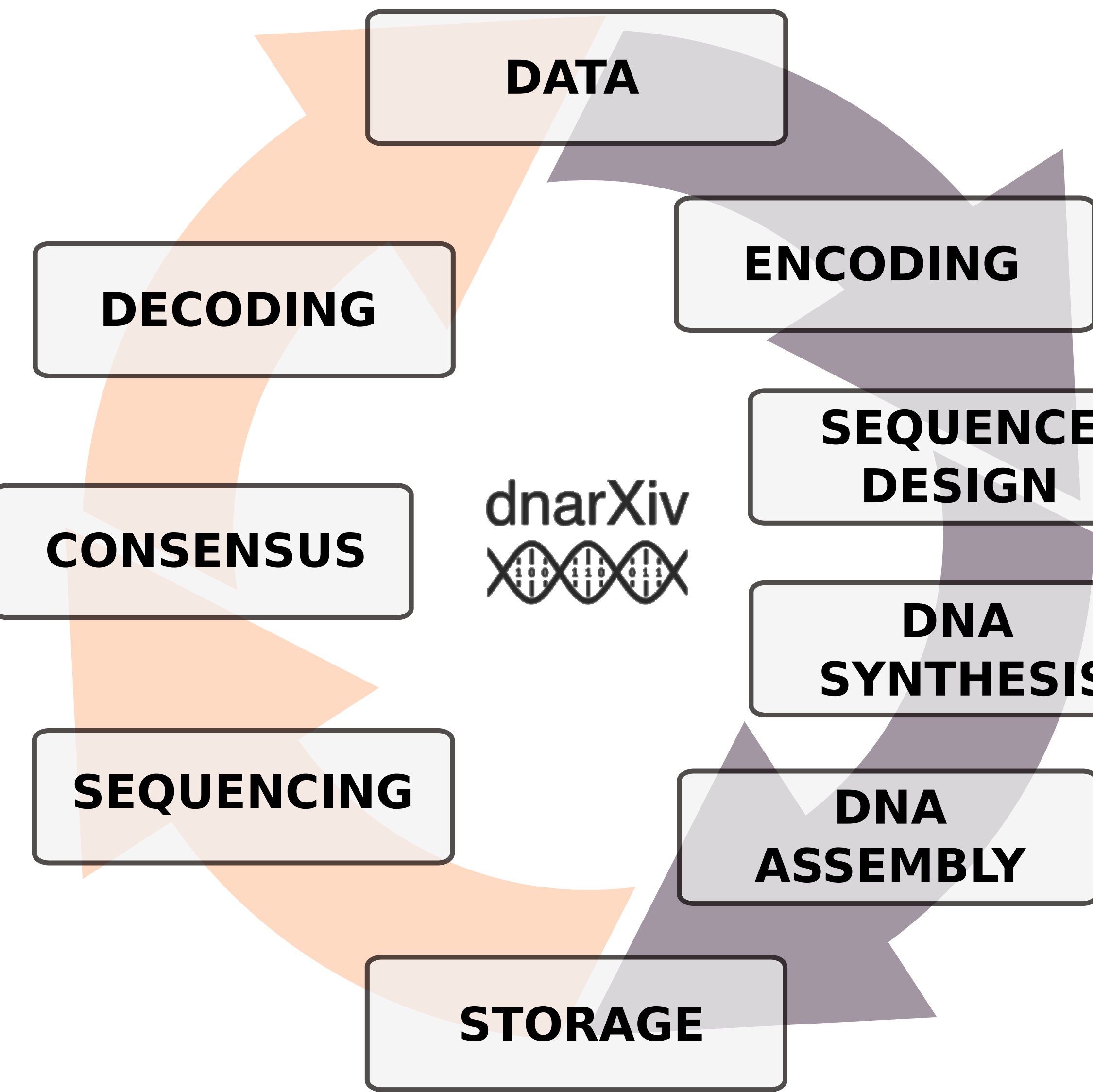
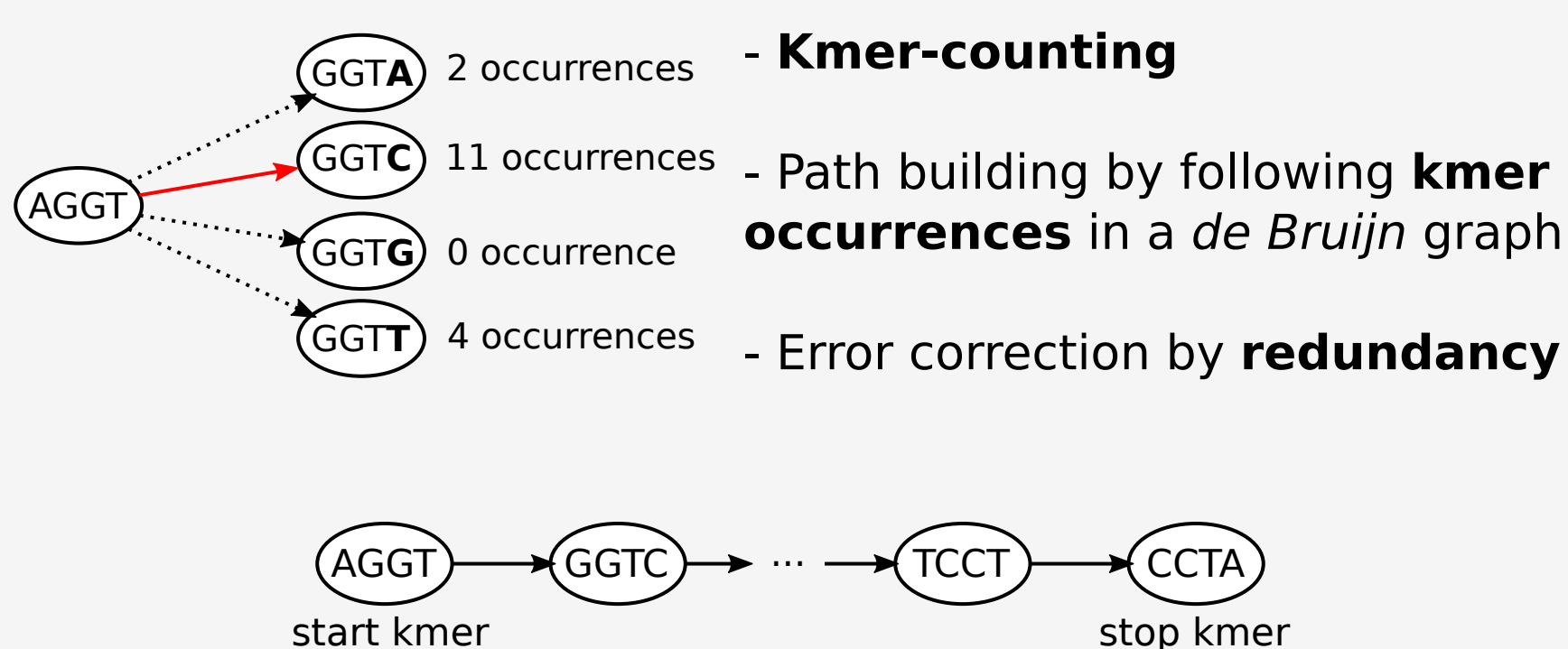
SEQUENCE DESIGN



Build 500bp dsDNA block with ssDNA oligonucleotides:

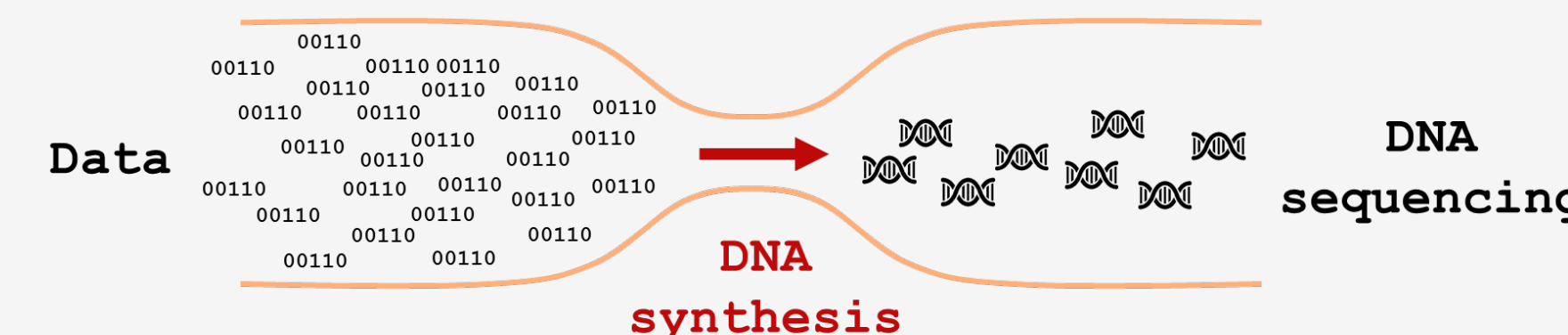
- **84,8%** "Payload": encoded data.
- **7,2%** "Overhangs": 9 specific 3' overhangs of 4nt.
- **8,0%** "PCR primers": block amplification + add specific 3' overhangs through **BsaI** restriction sites.

CONSENSUS



DNA SYNTHESIS

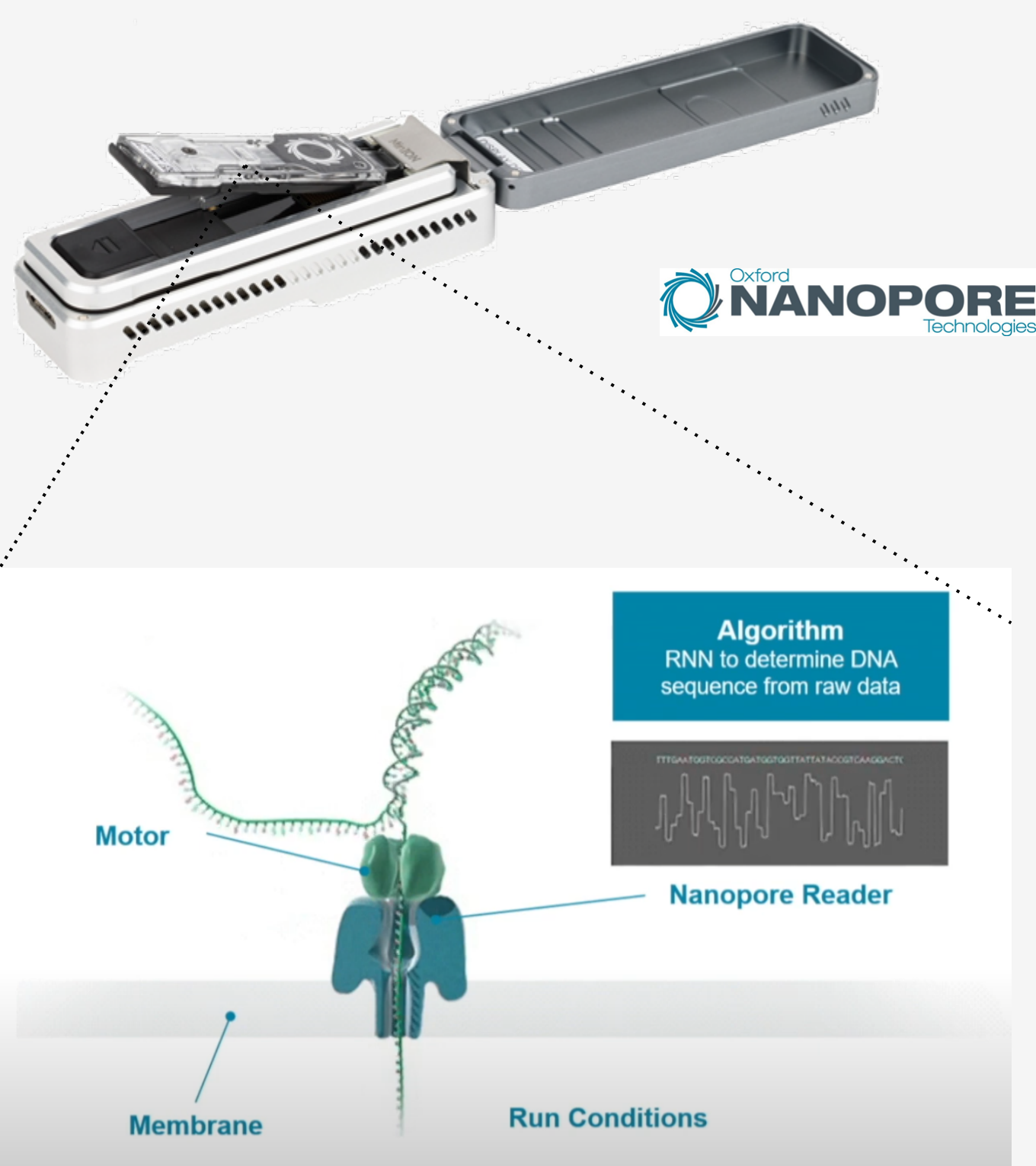
Pipeline BOTTLENECK



- **Chemical synthesis** by phosphoramidite method
- **High cost**: 0,123€/base
- Oligonucleotide length: **350nt**
- **Error rate**: negligible
- Synthesis speed: **100 to 200 seconds/nucleotide**
- In development: **enzymatic synthesis**

DNA SEQUENCING

- Read length: **up to 1Mb**
- **Accuracy: 99%**
- **Real time** sequencing
- **Portable**
- **Similar cost to Illumina sequencing**
- **In progress**: real-time selective sequencing and detection of chemical modification



Nanopore reader: raw signal (fast5)
Basecalling: converting files in DNA sequence (fastq)

STORAGE

In vitro: DNA molecules are simply stored at **room temperature** in a **buffer** (Tris-EDTA pH8).

In vitro DNA storage		
Features	Short ssDNA	Long dsDNA
Storage density*	+	++
Random access capacity°	+	+++
Duplication capacity**	-	+
Storage stability	+	++
Security & stenography	+	++

* Storage density: DNA molecules/m³.
° Random access: access to a specific file without decoding the entire storage.
** Duplication: create complete copy of DNA storage.

DNA ASSEMBLY

