

# Application de la bio informatique au stockage des données

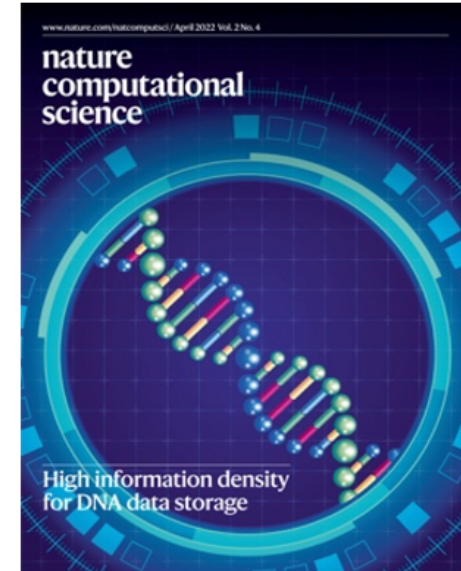
Article :

Titre : “Towards practical and robust DNA-based data archiving using the yin–yang codec system”

Auteurs : Zhi Ping, Shihong Chen, Guangyu Zhou, Xiaoluo Huang, Sha Joe Zhu, Haoling Zhang, Henry H. Lee, Zhaojun Lan, Jie Cui, Tai Chen, Wenwei Zhang, Huanming Yang, Xun Xu, George M. Church and Yue Shen

Date de publication : 25 Avril 2022

Magazine : *nature*



Church, G. (2022). Towards practical and robust DNA-based data archiving using the yin–yang codec system. *Nature*, 2(4), pp.234-242

**I/ Pourquoi stocker sur l'ADN ?**

**II/ Le stockage sur ADN**

**III/ Le Ying-Yang Codec**

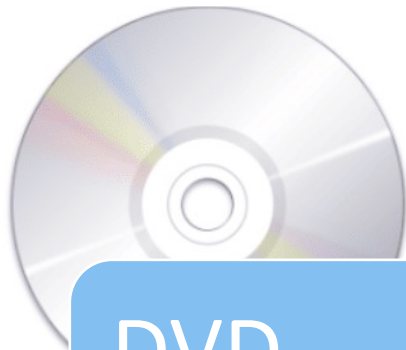
# I/ Pourquoi stocker sur l'ADN ?

## L'évolution du stockage à travers le temps :



RAMAC  
305

- 1955
- 5 Mo



DVD

- 1995
- 5 Go



Carte SD

- 1999
- 64 Go



ADN

- 2013
- 455 EB/g

## I/ Pourquoi stocker sur l'ADN ?

### Les enjeux du stockage relevé par les chercheurs :

- Besoin croissant
- Capacité des stockages traditionnelle insuffisante
- Majorité des données stockées sont inutilisés



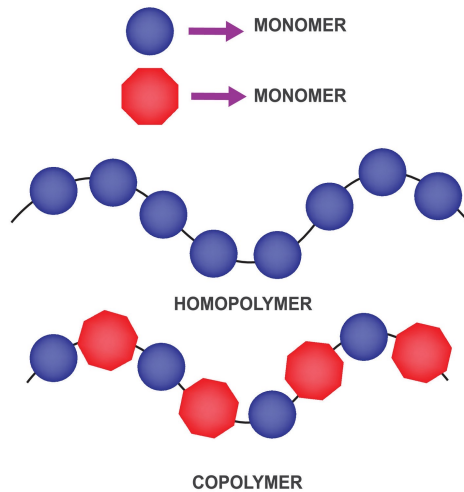
**I/ Pourquoi stocker sur l'ADN ?**

**II/ Le stockage sur ADN**

**III/ Le Ying-Yang Codec**

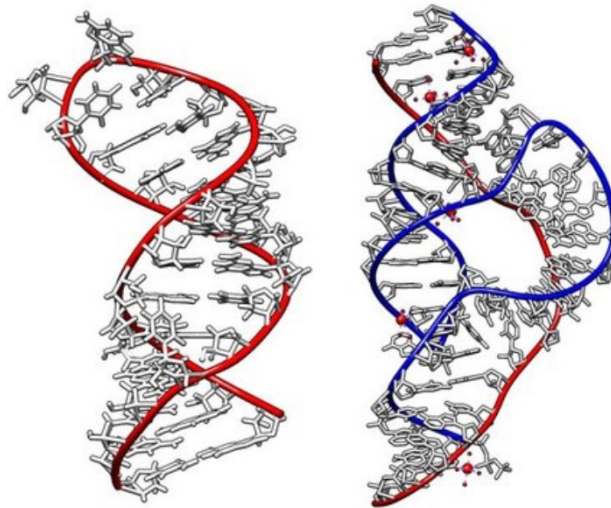
## II/ Le stockage sur ADN

### Les difficultés mises en évidence par l'article :



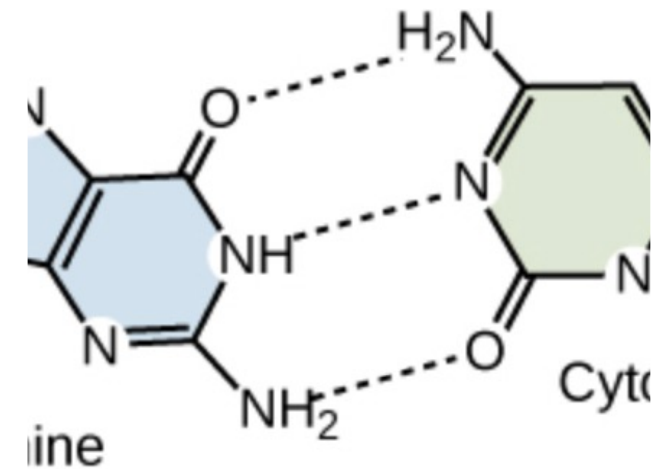
Un homopolymère

Présence d'homopolymères



Hélices

Les structures secondaire



3 liaisons hydrogène

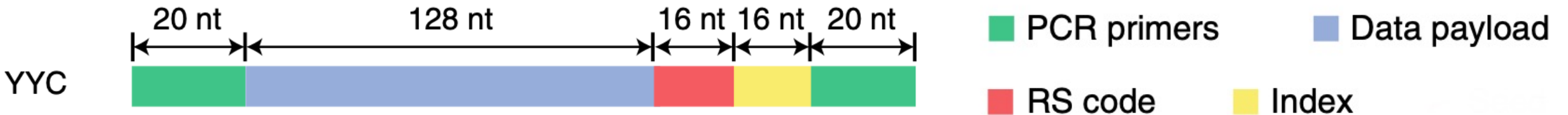
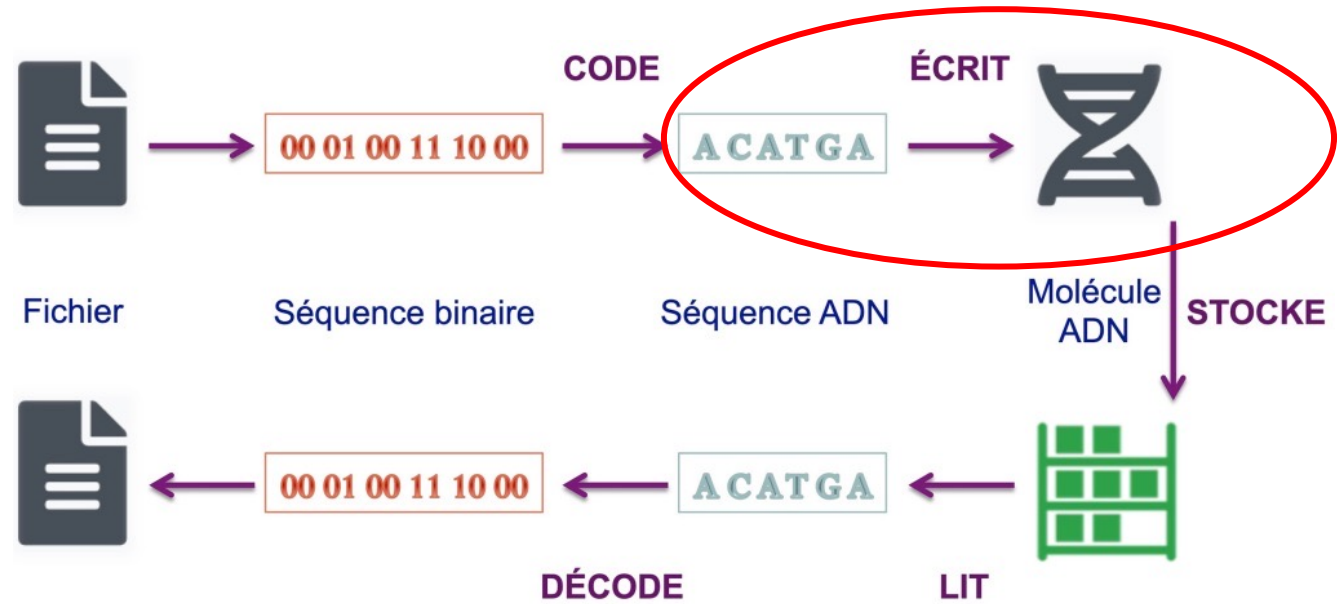
Le taux en nucléotide G et C



## II/ Le stockage sur ADN

### Comment synthétiser ?

- Synthétisation par une entreprise
- Segmentation par bloc
- Chaque oligonucléotide est composé de plusieurs séquences

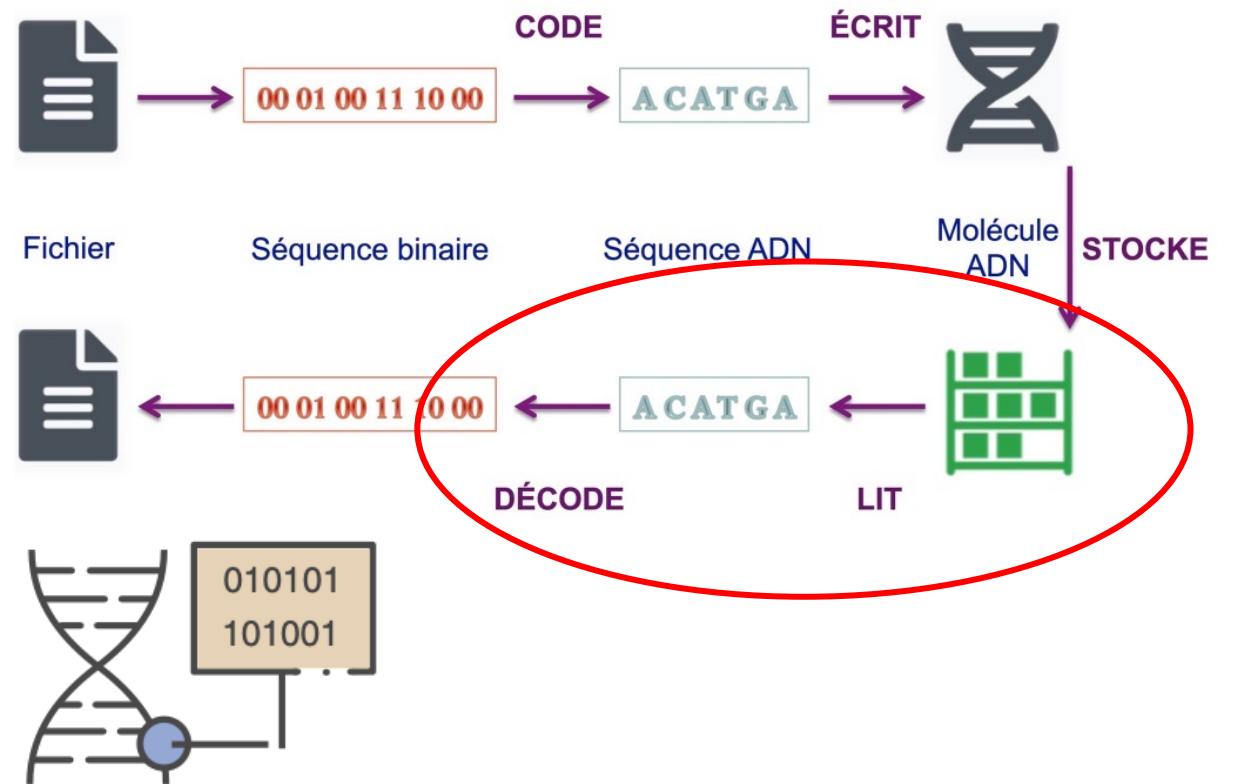




## II/ Le stockage sur ADN

### Comment lire ?

- Amplification par PCR
- Séquençage électronique
- L'ordinateur décode la chaîne de nucléotides séquencé en binaire



**I/ Pourquoi stocker sur l'ADN ?**

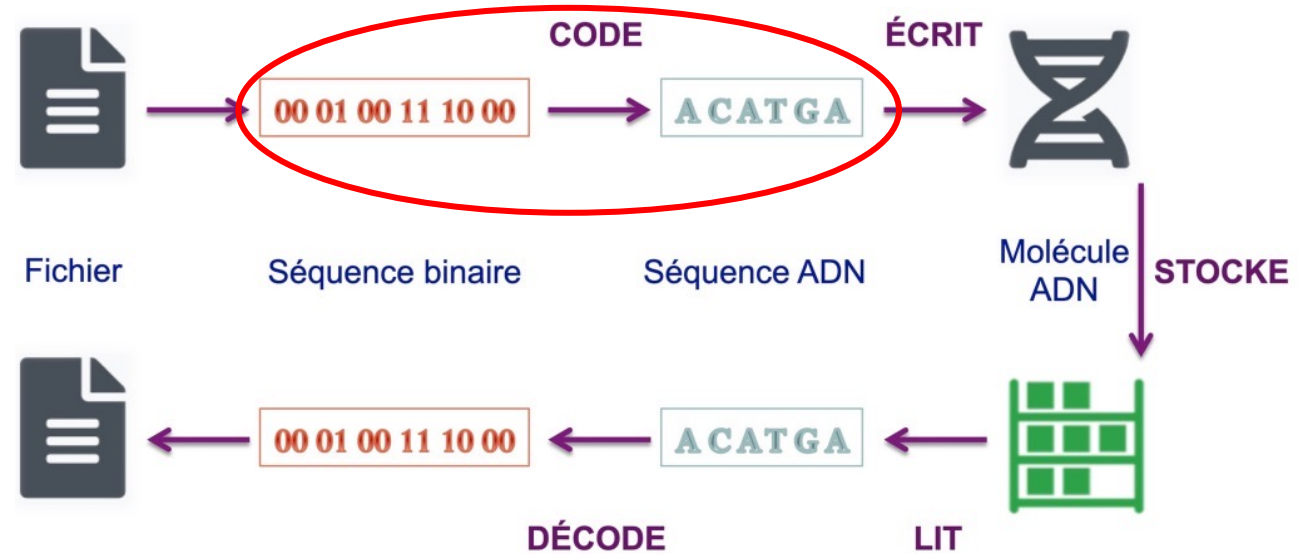
**II/ Le stockage sur ADN**

**III/ Le Ying-Yang Codec**

### III/ Le Ying-Yang Codec

#### En quoi consiste-il ?

- Encoder une suite de bits en une succession de nucléotides stable
- Appairer deux entiers de même longueur de bits



### III/ Le Ying-Yang Codec

#### Comment fonctionne-il ?

- Le nucléotide précédent doit être pris en compte

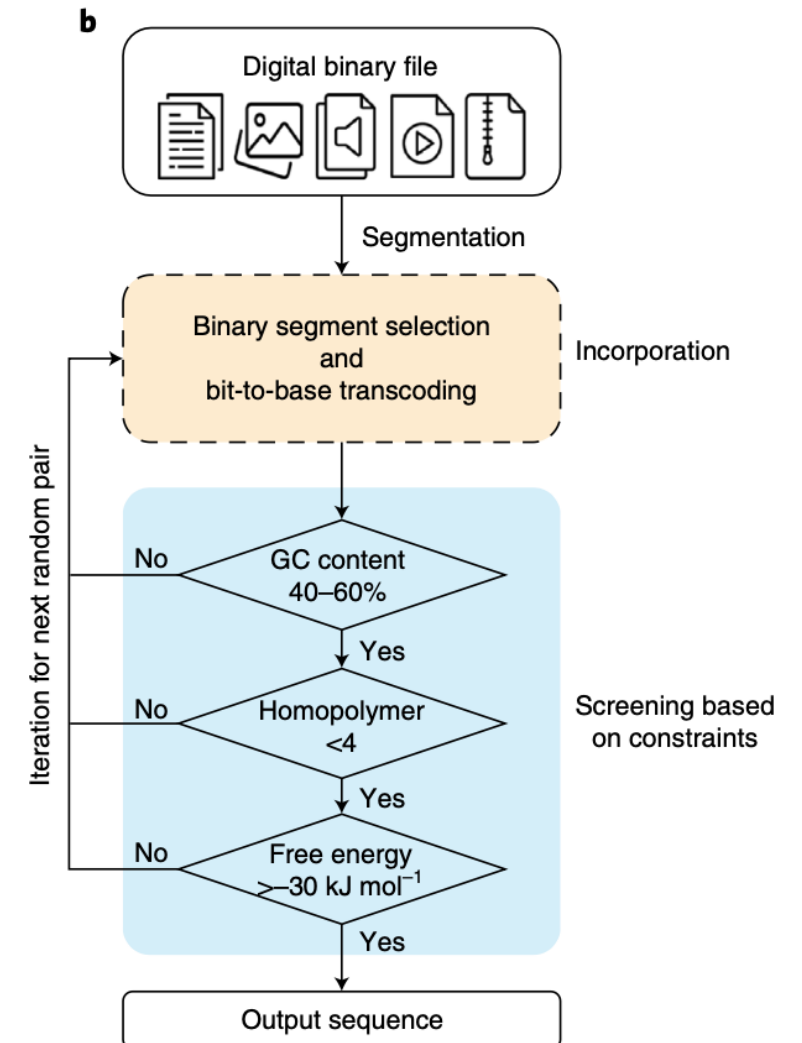
Binary Digits	
0	1
A or T	C or G

'Yang' RULE

Previous Nucleotide

	Binary Digits	
	0	1
A	A or G	C or T
T	C or T	A or G
G	A or G	C or T
C	C or T	A or G

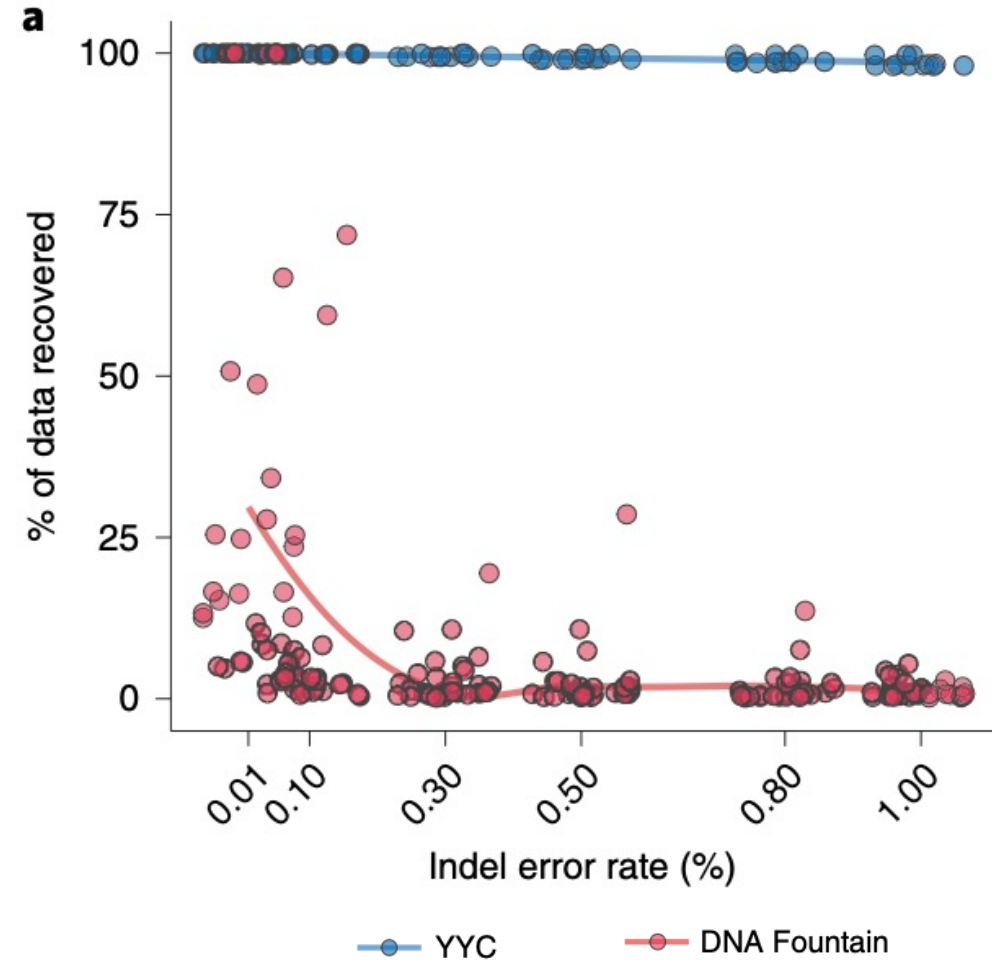
'Yin' RULE



### III/ Le Ying-Yang Codec

#### Quel est l'intérêt du YYC ?

- Comparaison par rapport au DNA Fountain
- YYC plus robuste que le DNA Fountain



### Limites du Yin-Yang Codec et du stockage sur ADN

- Efficacité
- Onéreux en argent

### Contestation de l'article

- Huang, Yiyi. (2024). Fake data, lies and traps in paper "Towards practical and robust DNA-based data archiving using the yin–yang codec system".

**6 How a filthy garbage paper like YYC got published**

# Bibliographie :

- Boutaud, A.-S. (2021). Stockage de données : la révolution sur ADN. *Journal du CNRS*, 23 novembre 2021.
- Huang, Yiyi. (2024). Fake data, lies and traps in paper "Towards practical and robust DNA-based data archiving using the yin–yang codec system".
- IBM RAMAC 305. (2023, décembre 21). *Wikipédia, l'encyclopédie libre*.
- *gorodenkoff*. (2019). [Photographie de centre de données]. Référence de la photo : 1148233863.10/05/19.
- Neogi,S.(2022).[Image vectorielle de types de polymères]. Numéro de la ressource : 2235639453. 8/12/22.
- Murat, P. [Scientific Figure]. *ResearchGate*.
- Madprime. DNA chemical structure [Image]. Wikimedia Commons.
- GitHub repository (<https://github.com/ntpz870817/DNA-storage-YYC>)