Can we improve analyses by transforming DNA ?

Luc Blassel, Paul Medvedev, Rayan Chikhi

INSTITUT PASTEUR

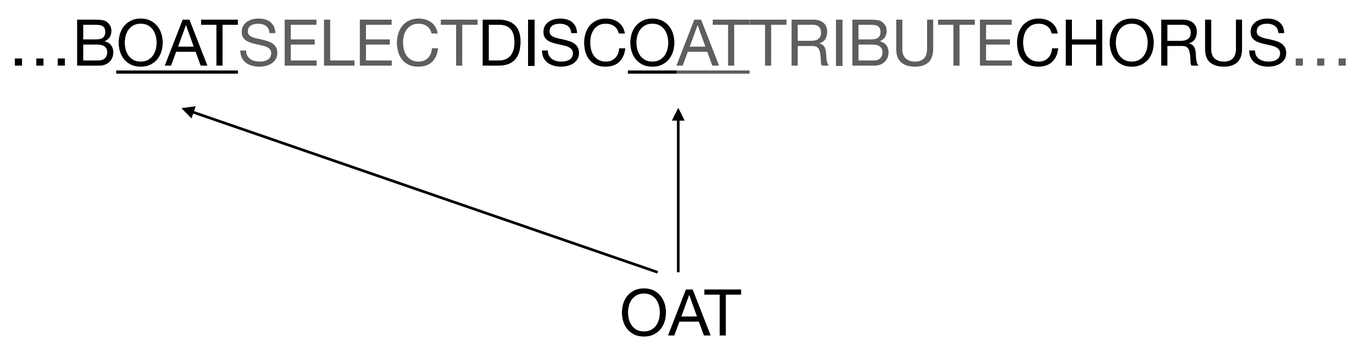
PR[AI]RIE
PaRis Artificial Intelligence Research InstitutE

SORBONNE UNIVERSITÉ

dy

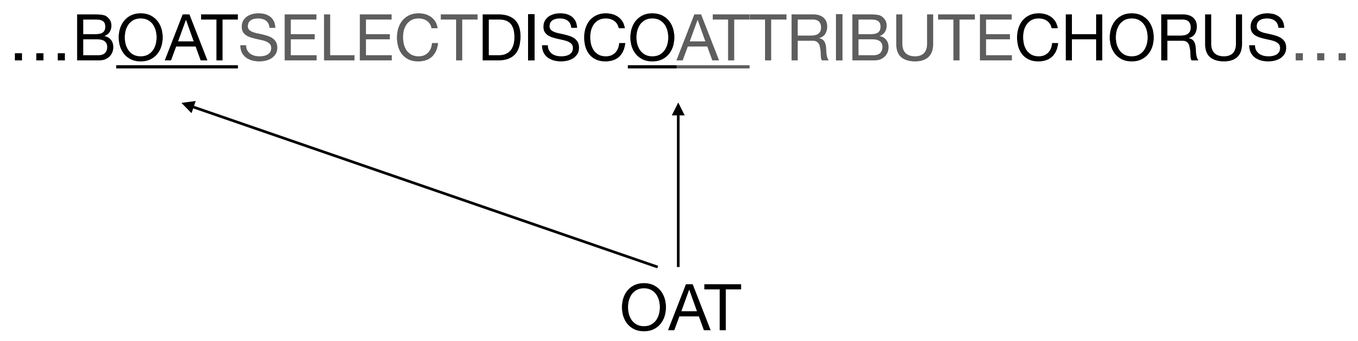
What ?

- **DNA** is like a **sentence** that contains instructions
- Machines start **reading** that sentence at a **random location**
- **Sometimes** the machine makes mistakes
- **Where** does what we **read come from** in the sentence ?



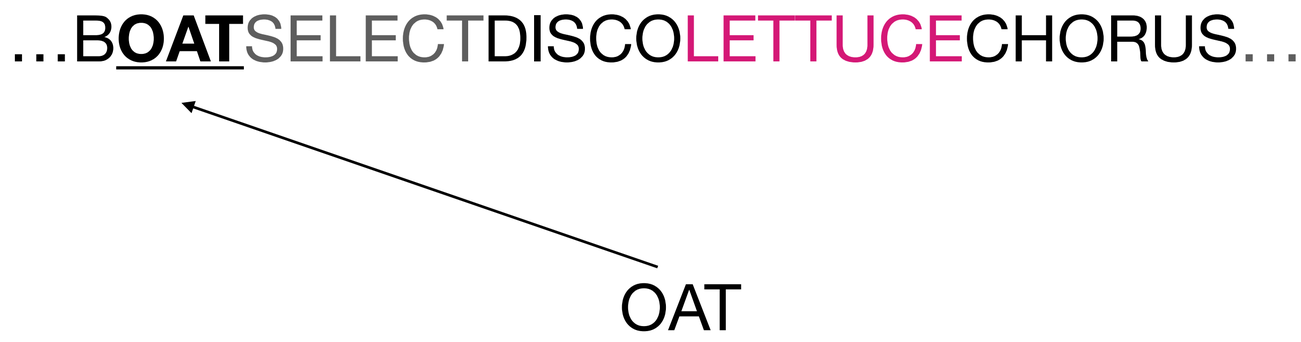
What ?

- Can we **transform the DNA** to **avoid mistakes** ?



What ?

- Can we **transform the DNA** to **avoid mistakes** ?



Why ?

- Finding out **where** a read comes from **is fundamental**
- A **mistake** can lead to **wrong conclusions**
- **Mistakes** are bad !

How ?

1. **Define** what a **transformation** is
2. Figure out **which ones are likely to be good**
3. **Generate** reads for which **we know the truth**
4. **Check if** transforming is **better**

Does it work ?

- We have found **several transformation** that **do better**
- In some cases, the **number of mistakes** is **1000 times lower**
- This is still the case when testing on **different organisms**

What next ?

- Explore a **larger set** of transformations. Either:
  - Find new ways to **guess which transformations are good**
  - **“Learn”** transformations directly with M.L. & A.I.

Camila Duitama Gomez

Fransesco Andreace

Riccardo Vicedomini

Mélanie Bernardini Ridel

Luc Blassel

Téo Lemane

Luca Denti

Yoann Dufresne

Rayan Chikhi

Paul Medvedev

SeqBio Team

INSTITUT PASTEUR

PennState Eberly College of Science

PR[AI]RIE PaRis Artificial Intelligence Research InstitutE

SORBONNE UNIVERSITÉ