



**HAL**  
open science

## **RandomRead : a sequence-read simulator program for metagenomic shotgun**

Guillaume Gricourt, Mélissa N'Debi, Vanessa Demontant, Anais Nguyen-Goument, Abdelrazak Aissat, Paul-Louis Woerther, Christophe Rodriguez

### ► **To cite this version:**

Guillaume Gricourt, Mélissa N'Debi, Vanessa Demontant, Anais Nguyen-Goument, Abdelrazak Aissat, et al.. RandomRead : a sequence-read simulator program for metagenomic shotgun. JOBIM, Jul 2019, Nantes, France. <hal-04603334>

**HAL Id: hal-04603334**

**<https://hal.science/hal-04603334v1>**

Submitted on 6 Jun 2024

**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.



HAL Authorization

# RandomRead : a sequence-read simulator program for metagenomic shotgun

Guillaume Gricourt<sup>1,2</sup>, Mélissa N'Debi<sup>1,2</sup>, Vanessa Demontant<sup>1,2</sup>, Anais Nguyen-Goument<sup>1</sup>, Abdelrazak Aissat<sup>1,2,4</sup>, Paul-Louis Woerther<sup>3</sup> and Christophe Rodriguez<sup>1,2,3</sup>

1 NGS platform AP-HP - IMRB Institute, Henri Mondor Hospital - University Paris-Est, 94000, Créteil, France

2 Inserm U955, Henri Mondor Hospital - University Paris-Est, 94000, Créteil, France

3 Department of Microbiology AP-HP, Henri Mondor Hospital - University Paris-Est, 94000, Créteil, France

4 Department of Genetics, Henri Mondor Hospital - University Paris-Est, 94000, Créteil, France

## Introduction

Evaluation of the performances of any analysis software used for Next Generation Sequencing is required in medical diagnosis through the Norma NF EN ISO 15189 (Schlaberg et al, 2017).

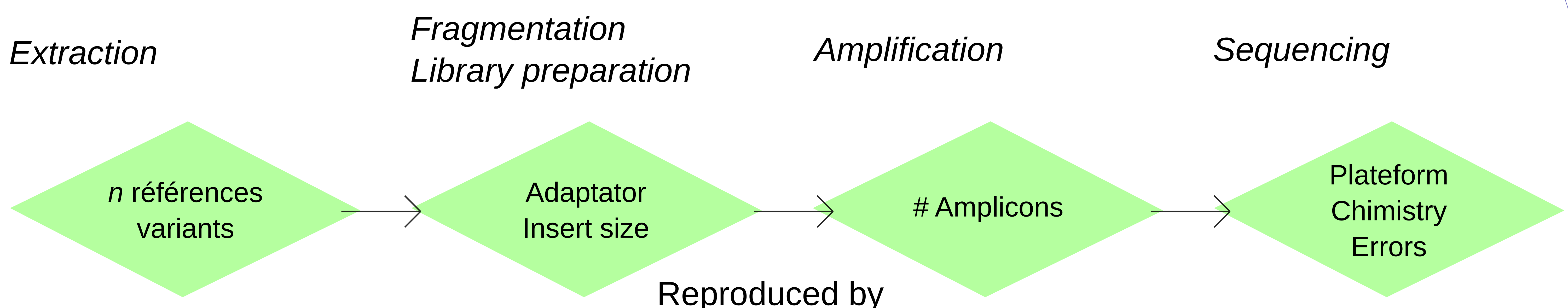
To satisfy these requirements different datasets were created in silico to simulate real sequences produced by an experimental protocol. It allow to evaluate the performance of the analysis software of our project of shotgun metagenomic.

## Aim

- Create a software able to simulate sequences like the experimental protocol
- Evaluate specifics performances of MetaMIC®, analysis software of shotgun metagenomic

## Materials and Methods

Steps simulation to reproduce experimental protocol (Escalona et al, 2016)



### RandomRead

written in C++ with Htslib (Li et al, 2009). Generate  $\approx 9800$  sequences/seconds.

## Results

### 120 simulated datasets

| 10 species by microorganisms | Nucleic acids | Concentration of microorganisms |
|------------------------------|---------------|---------------------------------|
| virus                        | RNA           | Low                             |
| bacteria                     |               |                                 |
| fungi                        | DNA           | High                            |

Evaluate performances of MetaMIC

- sensibility
- specificity
- reproductibility.

### Cost comparison between real/simulated datasets

|             | Real              | Simulated |
|-------------|-------------------|-----------|
| Time (Unit) | Months            | Hours     |
| Cost (€)    | $\approx 30\ 000$ | —         |

## Conclusion

- Saving time and money
- Evaluate specifics performances of MetaMIC® used for accreditation to the norma ISO EN NF 15189
- Test software for future updates