Characterization of *E. coli* from artificially contaminated cheese samples (PT33) using long-read metagenomics

EURL-VTEC –Rome
ANSES, LSAI IdentyPath Plateform
BfR, NRL *E. coli* and National Study Centre for Sequencing in Risk Assessment

PhD project - METADETECT

- Shared project:

German federal institute for risk assessment (BfR, Berlin Germany)

French national agency for food safety (ANSES, Maisons-Alfort France)

- Aim:

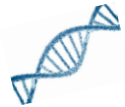
Identification of *eae*-positive STEC with novel metagenomics approach and its application on dairy farms in France and Germany

WP1:	WP2:	WP3:	WP4:
HMW DNA extraction for MinION seq	Testing long-reads potential on artificially contaminated samples	Develop a bio-informatics pipeline	Test on naturally contaminated samples

Jaudou *et al*., 2022

In review

STEC pure culture 

HMW DNA extraction
=> 3 methods (n=102) 

DNA quantification and qualification (> 1 µg)

Library preparation and sequencing
=> MinION (ONT) and MiSeq (Illumina) 

PLOS ONE

RESEARCH ARTICLE

Evaluation of high molecular weight DNA extraction methods for long-read sequencing of Shiga toxin-producing *Escherichia coli*

Sandra Jaudou[1], Mai-Lan Tran[1,2], Fabien Vorimore[2], Patrick Fach[1,2], Sabine Delannoy[1,2]*

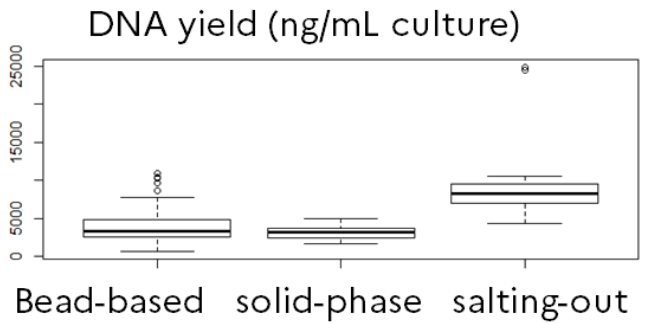
1 Pathogenic E. coli Unit, Laboratory for Food Safety, Anses, Maisons-Alfort, France, **2** IdentyPath Platform, Laboratory for Food Safety, Anses, Maisons-Alfort, France

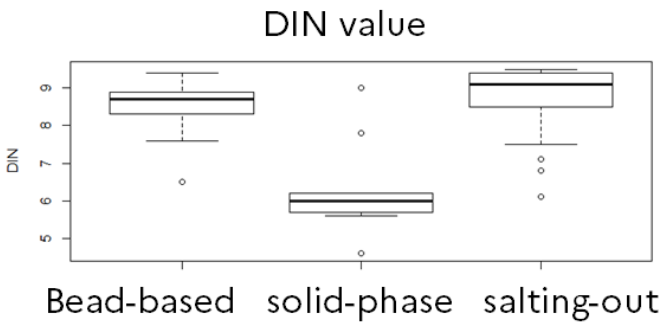
* sabine.delannoy@anses.fr

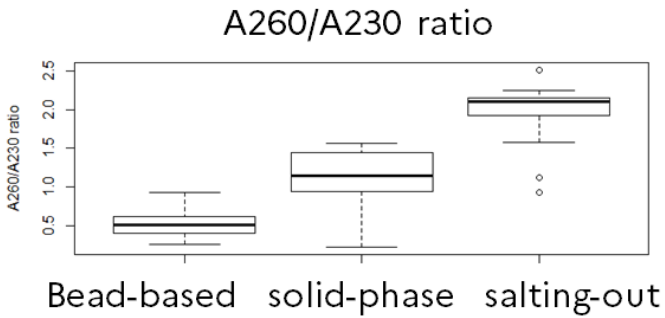
Abstract

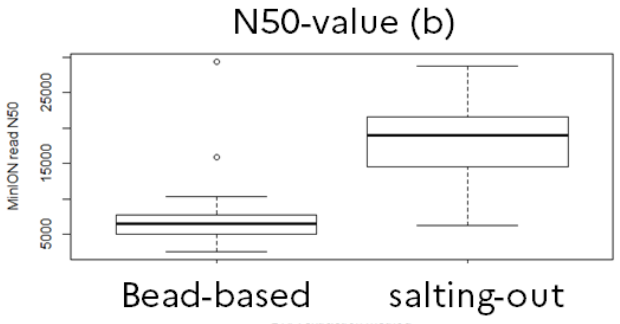
Next generation sequencing has become essential for pathogen characterization and typing. The most popular second generation sequencing technique produces data of high quality with very low error rates and high depths. One major drawback of this technique is the short reads. Indeed, short-read sequencing data of Shiga toxin-producing *Escherichia coli* (STEC) are difficult to assemble because of the presence of numerous mobile genetic elements (MGEs), which contain repeated elements. The resulting draft assemblies are often highly fragmented, which results in a loss of information, especially concerning MGEs or

HMW DNA extraction









High quantity of HMW DNA fragments with salting-out

Metagenomics STEC identification from raw milk

Artificial contamination of STEC-negative raw milk
500 CFU/mL, 50 CFU/mL and 5 CFU/mL
(n=3, 4 and 5)

↓

Raw milk enrichment
(1:10 BPW, 37°C + acriflavine* 18-20h) * 12mg/L

↓

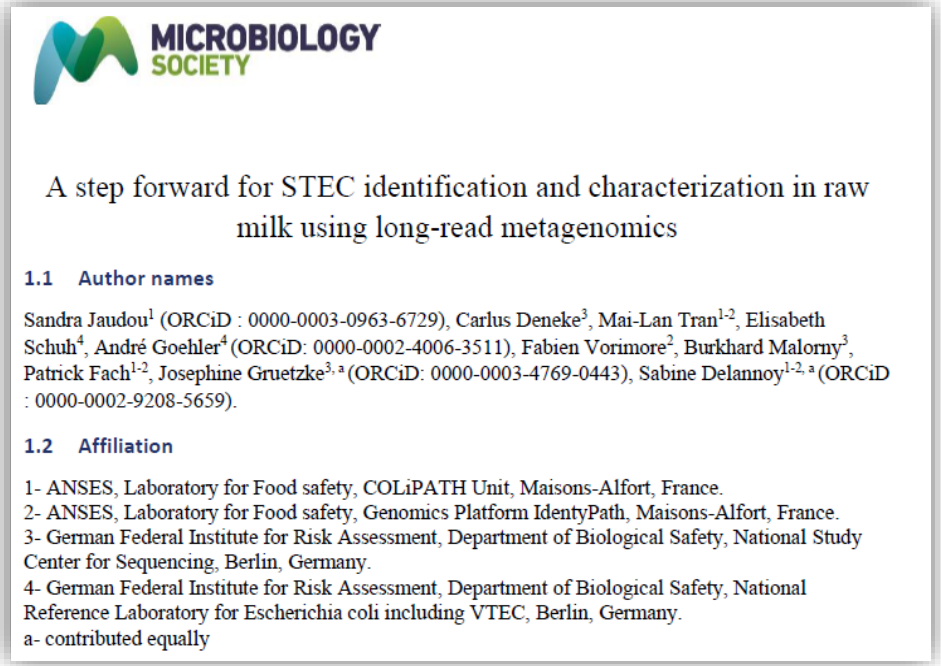
HMW DNA extraction (Salting-out)

↓

MinION library preparation and sequencing

↓

Guppy basecaller + STECmetadetectorþ



MICROBIOLOGY SOCIETY

A step forward for STEC identification and characterization in raw milk using long-read metagenomics

1.1 Author names

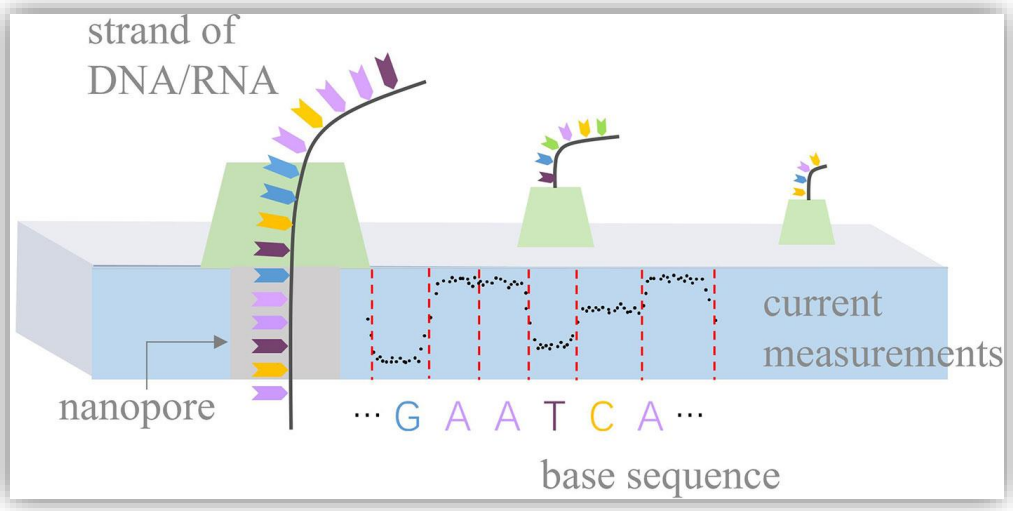
Sandra Jaudou[1] (ORCiD : 0000-0003-0963-6729), Carlus Deneke[3], Mai-Lan Tran[1-2], Elisabeth Schuh[4], André Goehler[4] (ORCiD: 0000-0002-4006-3511), Fabien Vorimore[2], Burkhard Malorny[3], Patrick Fach[1-2], Josephine Gruetzke[3, a] (ORCiD: 0000-0003-4769-0443), Sabine Delannoy[1-2, a] (ORCiD : 0000-0002-9208-5659).

1.2 Affiliation

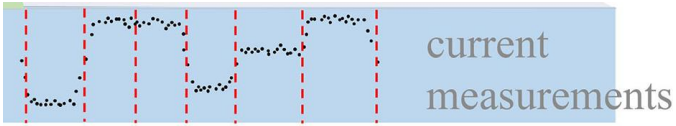
1- ANSES, Laboratory for Food safety, COLiPATH Unit, Maisons-Alfort, France.
2- ANSES, Laboratory for Food safety, Genomics Platform IdentyPath, Maisons-Alfort, France.
3- German Federal Institute for Risk Assessment, Department of Biological Safety, National Study Center for Sequencing, Berlin, Germany.
4- German Federal Institute for Risk Assessment, Department of Biological Safety, National Reference Laboratory for Escherichia coli including VTEC, Berlin, Germany.
a- contributed equally

MinION data analysis

MinION sequencing



fast5



Base-calling

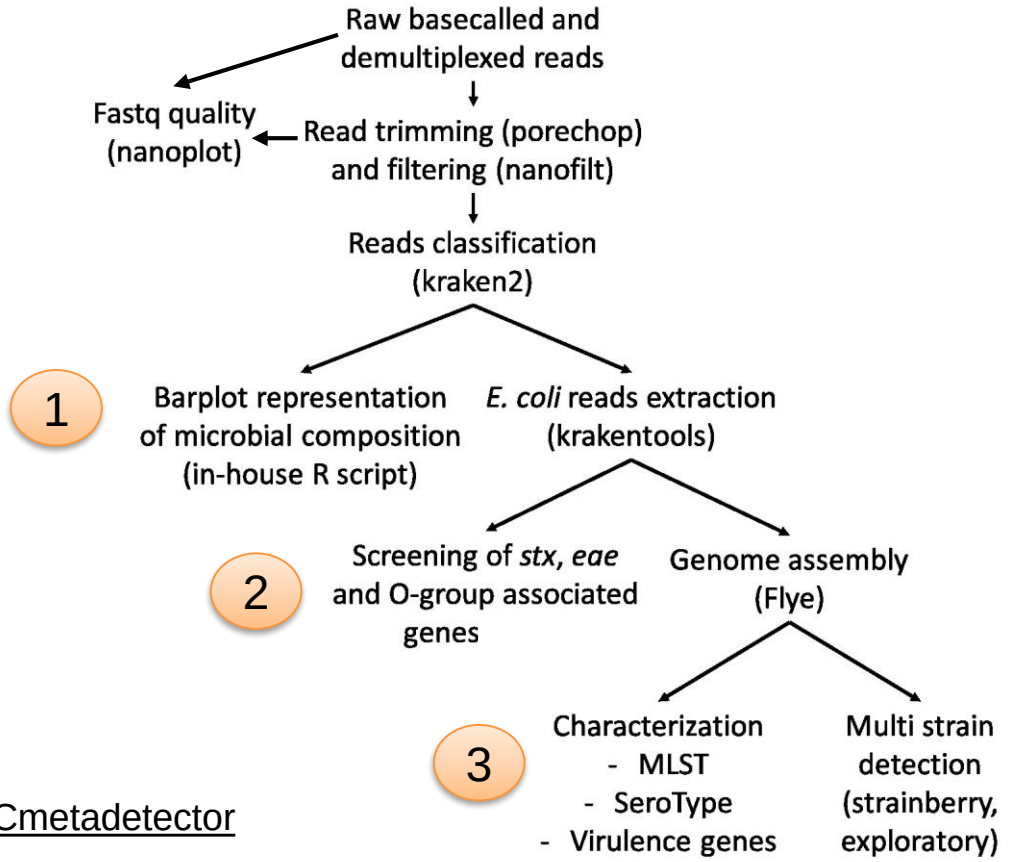
… G A A T C A …

fastq/a

base sequence

Zeng *et al.*, 2020

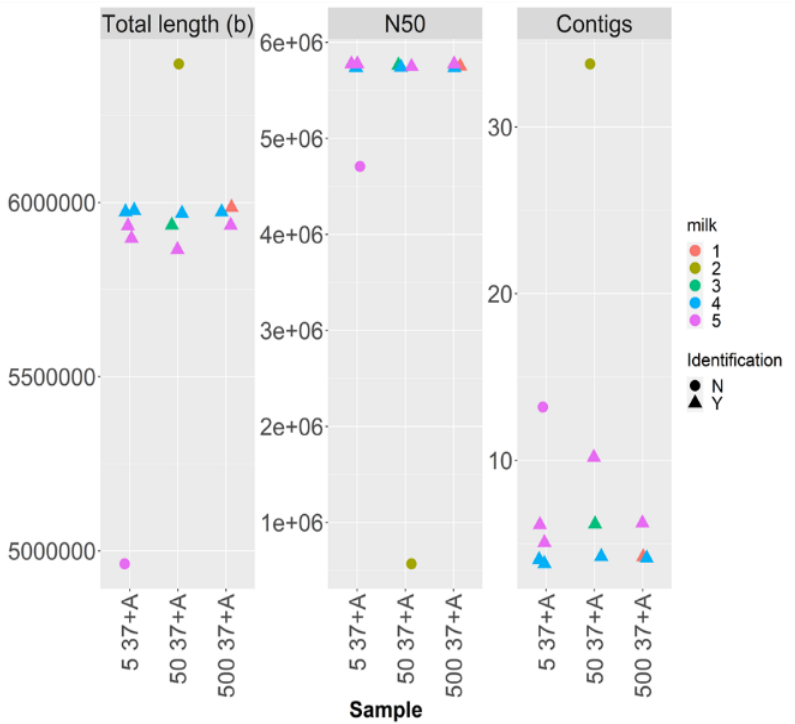
STECmetadetector



https://gitlab.com/bfr_bioinformatics/STECmetadetector

Optimized workflow

1. Optimized enrichment condition: 37°C with acriflavine
2. STECmetadetector pipeline to facilite data analysis
3. Characterization of *eae*-positive STEC O26 from 5 CFU/mL
4. Additional *E. coli* background flora may be an issue

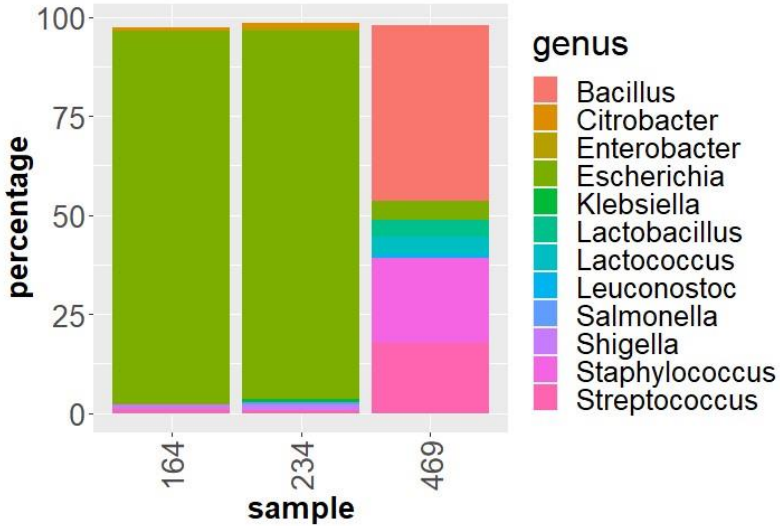


Flye assembly metrics obtained from extracted E. coli reads. STEC artificially contaminated cow raw milk at final concentration of 500 CFU.mL^{-1} (n=3; 500), 50 CFU.mL^{-1} (n=4, 50) or 5 CFU.mL^{-1} (n=5, 5) and enriched at 37°C with acriflavine (37+A).

1- Sequencing data: taxonomic assignment

Sample	DNA concentration ($ng.\mu L^{-1}$)	Percentage of *Escherichia* (%)	Generated data (Mb seq)
164	202	93	289,8
234	256	94,43	251
469	16,1+14,3	4,85	0,14

Barplot representing major genus sequenced per sample (kraken2)



- Sample 469: Too low DNA for MinION sequencing, negative control?
- >90 % *Escherichia* reads in samples 164 and 234

2- Sequencing data: O-group and virulence factors

- Mapping of reads on O-group database (minimap2)
- Virulence genes detection (abricate)

Sample	Detected O-groups (depth)	*eae*	*stx* (depth)	*ipaH, aggR, aaiC, STh, STp, LT*
469	O104 (3) / O78 (2)	na	*stx1* (1)	na
164	O78 (141) / O104 (5)	na	*stx1c* (48)	na
234	O104 (197) / O78 (2)	na	na	*aggR*, *aaiC*

Reads not confirmed using qPCR

3- Sequencing data: assembly of *E. coli* reads

- *E. coli* reads extraction
- Flye assembly

Sample	Total assembly length (Mb)	# of contigs	Largest contig (kb)	ST	Serotype*	Stx Sub-type	Pathotype
469	1.5	250	20.9	-	-	Na	Na
164	5.7	21	893.9	3101	O78:H4	1c	STEC
234	5.3	39	844	678	O104:H4	Na	EAEC

* O and H sequences located on different contigs

3- Virulence and AMR genes

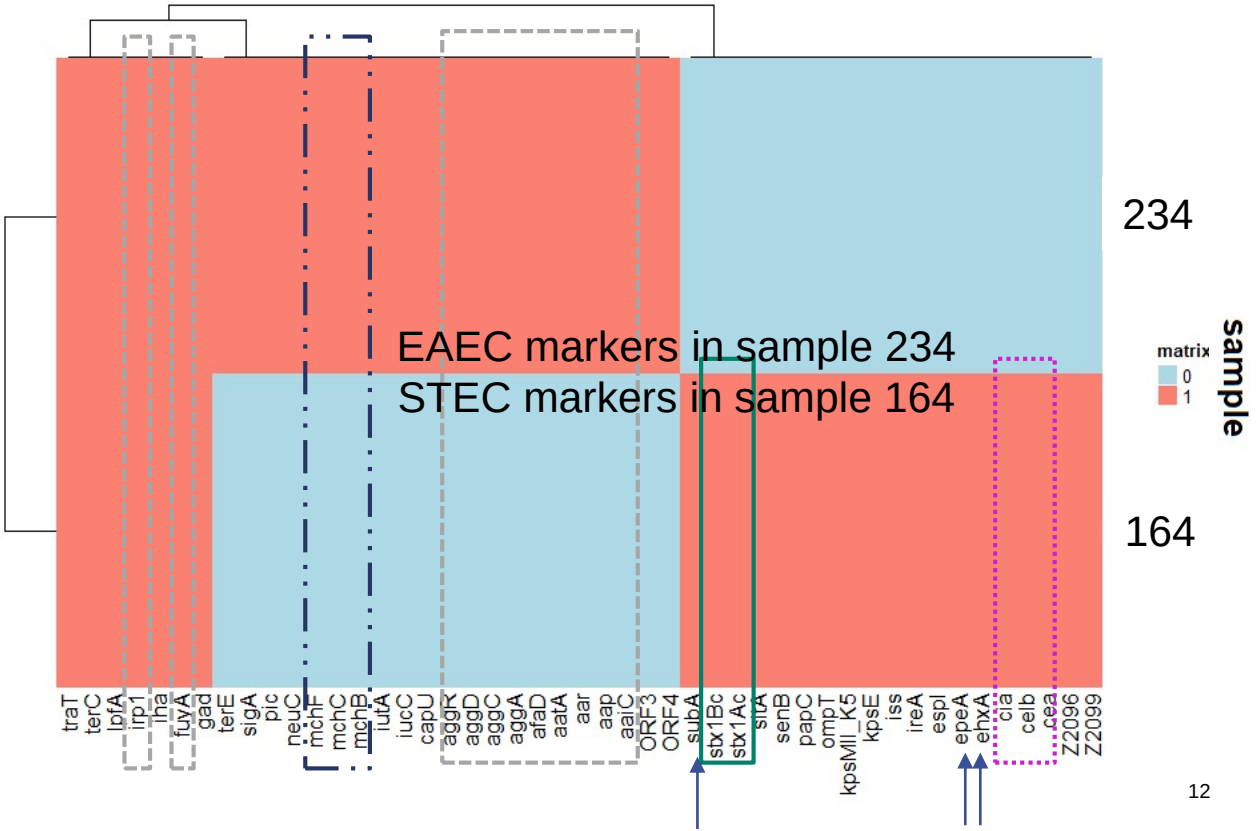
Enteroaggregative *E. coli* specific markers

STEC specific markers

Colicin genes

Microcin genes

→ pO113-like plasmid



Interpretation of long-read metagenomic data

- Sample 164 contains an STEC strain of serotype O78:H4 (O and H genes not on the same contig), *stx*1c, **ST 3101**, colicin genes.
- Sample 234 contains an EAEC strain of O-group O104 and H-type H4 (O and H genes not on the same contig) and harbors *aggR* and *aaiC* genes, **ST 678**, microcin genes.
- Sample 469 contained no *E. coli* strain, negative control?

ANSES:
Sabine DELANNOY
Patrick FACH
Fabien VORIMORE
Mai-Lan TRAN

BfR:
Josephine GRUETZKE
Burkhard MALORNY
Carlus DENEKE
Elisabeth SCHUH
Andre GOEHLER

ISS:
Stefano MORABITO
Valeria MICHELACCI
Federica GIGLIUCCI
And the ISS team

Thank you for your attention