Improving the genome reference assembly for *E.coli* CFT073 and *in silico* detection of potential antibiotic resistance genes

Guo, Yan, CDER/DARS; Xu, Lei, CDER/DARS; Samarth, Deepti, CDER/DARS; Rouse, Rodney, CDER/DARS, Weaver, James, CDER/DARS, Hyland, Paula, CDER/DARS

Abstract

A better understanding of whether antimicrobial resistant (AMR) genotype can predict phenotype has the potential to lead to new concepts of diagnostics. CFT073 is a commonly used uropathogenic *E. coli* (UPEC) isolate and model for *in vitro* antimicrobial studies; however, the strain lacks a complete genome. We generated a high-quality reference genome using long-read sequencing and used *in silico* analyses to evaluate residual resistance. We identified 54 genotypes associated with resistance to 23 different antibiotic classes. Comparing genotype to phenotype for specific antibiotics suggested that CFT073 genotypes do not always correlate with phenotype. Expansion of phenotype-genotype data for CFT073 susceptibility and resistance in the public domain is needed to establish accurate genotype to phenotype prediction for future research.

Introduction

Whole-genome sequencing (WGS) is positioned to become an essential tool in the control of antibiotic resistance. Most bacterial reference genomes for clinical isolates are generated from short-read sequence data which contain sequencing errors or gaps and therefore may be incomplete[1]. To identify antibiotic-resistant genes (ARGs) and/or mechanisms, an accurate and complete reference genome is required. Long-read sequencing can generate an accurate *de novo* assembled reference genome for genotyping of antibiotic resistance in bacteria.

Materials and Methods

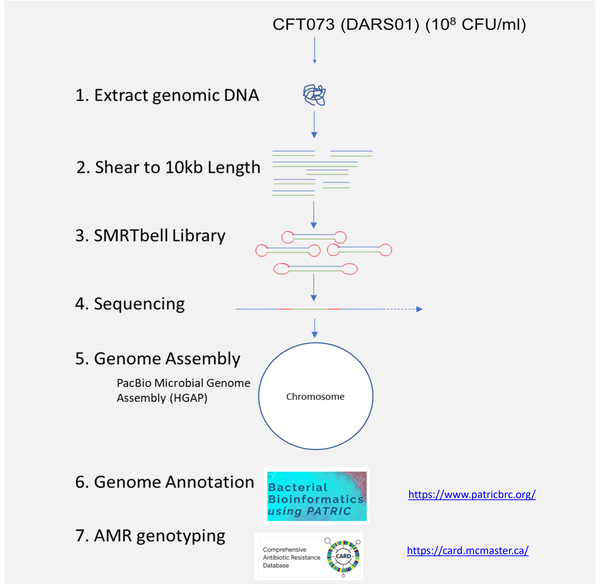

Figure 1. Sequencing, reference genome assembly and AMR gene analysis

Results

Sequence Info	
Contigs	1
Genome Length	5.23Mb
GC Content	50.48

Genomic Features		
Genome Feature	AE014075	CFT073
Coding genes	5161	5113 ↓
Repeat_regions	134	137 ↑
tRNA	89	91 ↑
rRNA	21	22 ↑

Table 1. Comparison of sequence information and functional annotation between the NCBI reference AE014075 and CFT073/DARS01

CFT073/DARS01 scored 100% completeness by BUSCO.

Results

Figure 1. *In silico* identification of AMR genes in CFT073 using CARD (https://card.mcmaster.ca/)

AMR Genes	
Database	CFT073
CARD	54

Largest number of hits were observed for Penam & Fluoquinolone

Drug Class	No. Hits
Penam	25
Fluoroquinolone antibiotic	24
Tetracycline antibiotic	18
Cephalosporin	17
Rifamycin antibiotic	11
Aminocoumarin antibiotic	10
Phenicol antibiotic	10
Peptide antibiotic	9
Triclosan	9
Glycylcycline	9
Aminoglycoside antibiotic	8
Cephamycin	8
Macrolide antibiotic	6
Diaminopyrimidine antibiotic	5
Elfamycin antibiotic	3
Acridine dye	3
Carbapenem	3
Monobactam	2
Fosfomycin	2
Nucleoside antibiotic	1
Nitroimidazole antibiotic	1
Rhodamine	1
Benzalkonium chloride	1

Ciprofloxacin

AST Results	*Phenotype Interpretation**		*CARD Result*	*Genotype Interpretation*
MIC*(ug/ml)	US	EU	ARG Names	
0.03	Susceptible	Susceptible	None	Susceptible

Phenotypic interpretation is consistent for both US and EU standard.

Genotypic result is negative.

*: CLSI and EUCAST breakpoints = 0.25 ug/ml

Ampicillin

AST Results	*Phenotype Interpretation***		*CARD Result*	*Genotype Interpretation*
MIC(ug/ml)	US	EU	ARG Names	
4	Susceptible	Susceptible	AmpC beta-lactamase AmpH beta-lactamase	Resistance

Phenotypic interpretation is consistent for both US and EU standard.

Both genes belong to class C beta-lactamase.
Prevalence in E. coli is 62%.
Preferentially hydrolyzes cephalosporin which belongs to a different group of beta-lactam antibiotics.

**: CLSI and EUCAST breakpoints = 8 ug/ml

Susceptible; yes · Resistant; Unlikely · Resistant; Unclear

Genotype is correlated with phenotype · Genotype is not correlated with phenotype

Figure 2. CFT073 Genotype:phenotype correlation Ciprofloxacin and Ampicillin. Phenotype data was described previously[2].

Results

Fosfomycin

AST Results	*Phenotype Interpretation****		*CARD Result*	*Genotype Interpretation*
MIC(ug/ml)	US	EU	ARG Names	
64	Susceptible	Resistance	MdtG GlpT E448K	Resistance

Phenotypic interpretation could be geographical specific.
EU just changed its breakpoint to 8ug/ml on Jan. 1st, 2021.
Clinical practices are different in US and EU.
MIC number should be used for tracking and comparing susceptibility changes cross clinical labs and clinical studies.

***: CLSI breakpoints = 64 ug/ml
EUCAST breakpoints = 8 ug/ml

Frequency of MdtG and GlpT (E448K) sequence changes in *E. coli* is 51% and 40%, respectively.
In contrast, Fosfomycin resistance in clinical *E.coli* isolates is reported to be only 3.4%-11%
Collectively, these data suggest that genotype prediction of potential resistance to Fosfomycin is not reliable.
A lack of phenotype-genotype data (e.g. only 13 datasets for *E.coli* are available in NCBI) limits potential for further analysis

Susceptible; yes · Resistant; Unlikely · Resistant; Unclear

Genotype is correlated with phenotype · Genotype is not correlated with phenotype

Figure 3. CFT073 genotype:phenotype correlation for Fosfomycin. Phenotype data was described previously[2].

Summary & Conclusions

- For the first time, we have produced an accurate *de novo* genome reference of *E. coli* CFT073 using long-read sequencing.
- *In silico* analysis can identify potential AMR genes and facilitate phenotype-genotype correlation to predict antibiotic susceptibility or resistance.
- For CFT073, genotyping alone may not be reliable for certain antibiotics to predict antibiotic susceptibility or resistance.
- Expansion of phenotype-genotype data for CFT073 susceptibility and resistance would greatly improve predicting phenotype from genotype for future research.

References

(1) Ouellette, M., Bhattacharya, A.*(2020)*. *EMBO Rep* 21(4): e50249
(2) Garimella, N., et al. (2020). Int J Antimicrob Agents ***55****(4): 105861.*