Limitations of sequence typing for isolate inclusion in outbreak investigations

K. Jackowska [1], D. Frampton [1], J.F. Peden [1], N.J. Saunders [1]. [1]Genpax - London (United Kingdom)

Introduction

- Campylobacteriosis, the leading cause of gastrointestinal disease in the EU, is primarily caused by two species: *Campylobacter jejuni* and *Campylobacter coli*.

- Correct detection of outbreaks is key to infection control.

- MLST and cgMLST are commonly used to identify clusters for further analysis.

- Here, we assess their performance against an openly scalable solution comparing strains at SNV resolution to determine their sensitivity and specificity for likely transmission-linked strains, using data from two recent publications. (Hsu *et al.,* Harrison *et al.*)

Methods

- In total, 3762 readsets were processed through the Genpax IDEM platform.

- All pairwise SNP distances for 844 *C. jejuni* and 2918 *C. coli* isolates were determined using the platform.

- Allele difference groups, as well as cgST and MLST were defined and derived for each sample in the two publications. (Hsu *et al.,* Harrison *et al.*)

- The results were combined and compared.

Key

AD0	Allele difference of 0
AD5	Allele difference of 5
AD10	Allele difference of 10
AD25	Allele difference of 25
AD100	Allele difference of 100
AD200	Allele difference of 200
cgST	Core genome sequence type
MLST	Multi-locus sequence typing

Results - Overview

- Of the pairs which were linked by IDEM at a distance of ≤10 SNV, 80/318 (25%) of *C. jejuni* and 1374/2920 (47%) of *C. coli* isolate pairs were not clustered by cgMLST at an allelic difference of 10 (AD10); representing thresholds used for outbreak detection.

- Using a higher threshold of AD25, a portion (21/318 (7%)) of the ≤10 SNP *C. jejuni* isolates remained ungrouped.

- Strikingly, 4/39 (10%) *C. jejuni* and 47/117 (40%) *C. coli* 0-SNV isolates are also not grouped at AD10.

- Examination of reported sequence types (ST) showed that 11/318 (3%) of *C. jejuni* and 200/2920 (7%) *C. coli* sample pairs were identified as different sequence types (STs) while differing by ≤ 10 SNVs.

- In contrast, samples belonging to the same ST had a median pairwise SNV distance of 568 (IQR: 75-1961) for *C. jejuni* and 95 (IQR: 66-153) for *C. coli* respectively.

Campylobacter jejuni results

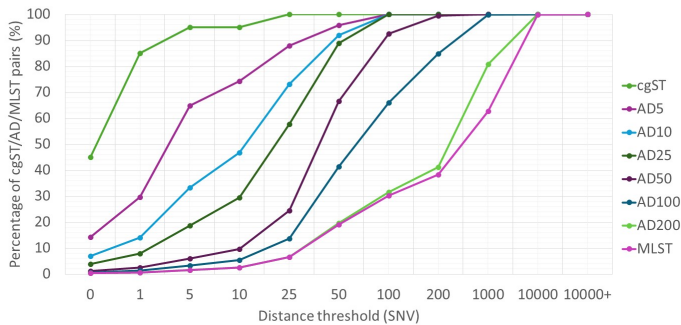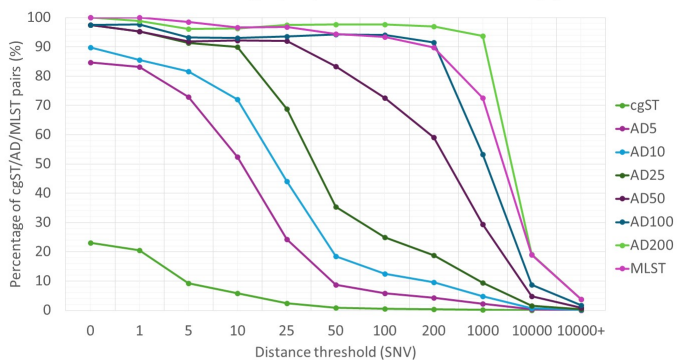**Figure 1**. Distribution of cgST/AD/MLST pairs at each SNV distance threshold for *C. jejuni*

Figure 2. Percentage of pairs at each SNV distance threshold that have the same cgST/AD/MLST classification for *C. jejuni*

Campylobacter coli results

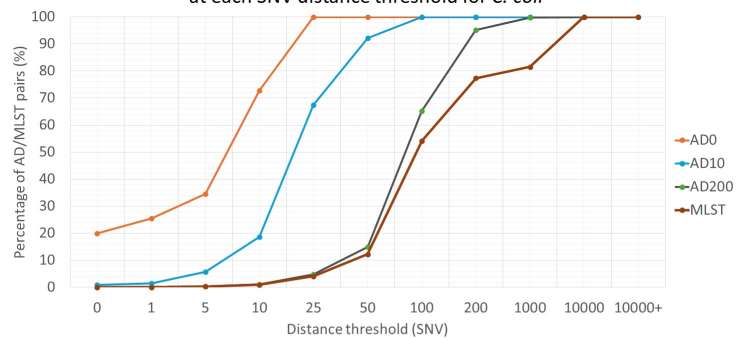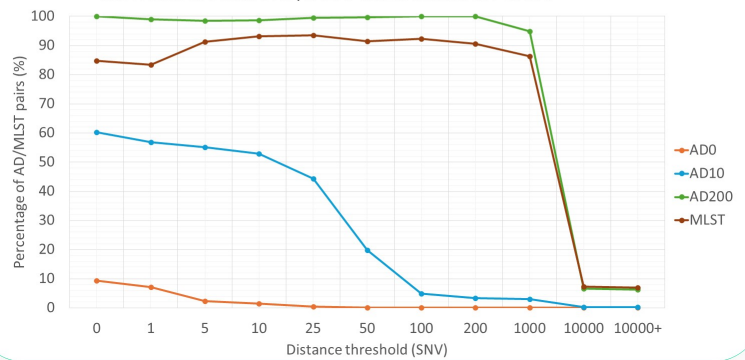**Figure 3**. Distribution of AD/MLAST pairs at each SNV distance threshold for *C. coli*

Figure 4. Percentage of pairs at each SNV distance threshold that have the same AD/MLST classification for *C. coli*

Conclusions

- Both MLST and cgMLST under- and over-predict sample linkage in both published studies (generate both false negative and false positive results).

- Across the two studies, *211 pairs of isolates ≤10 SNV distance identified by IDEM do not share the same sequence type* (ST) and **1466 pairs do not share the same cgMLST cluster at AD10**, which would normally exclude them from being identified as potential members of outbreaks and subsequent more detailed comparisons and analysis.

- In addition, of the 6% of isolate pairs identified as the same ST (300959/4514920), only 1% (3039) fall within a distance of 10 SNVs. The majority (99%) of same ST pairs have a greater SNV distance than would normally be considered to indicate outbreak / transmission connection.

- This performance indicates that Sequence Typing (MLST or cgMLST/cgST) is not an optimal first stage analysis for the detection and investigation of transmission-linked strains and outbreaks in these species.

- This may reflect the combined effects of issues inherent to the underlying methodology in the context of the highly recombining and panmictic nature of the genus. Similar analyses seem warranted in other species.

References

1. Hsu C–H, Harrison L, Mukherjee S, Strain E, McDermott P, Zhang Q, Zhao S. Core Genome Multilocus Sequence Typing for Food Animal Source Attribution of Human *Campylobacter jejuni* Infections. Pathogens. 2020;9(7):532. https://doi.org/10.3390/pathogens9070532

2. Harrison L, Mukherjee S, Hsu C–H, et al. Core genome MLST for source attribution of *Campylobacter coli*. Frontiers. June 21, 2021. Accessed March 19, 2024. https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2021.703890/full.

Check out our website:

SPECIES COVERED

Our prioritized species address the most common Healthcare Associated Infections and Food Safety Species.

Species Name	
Acinetobacter baumannii	HD
Campylobacter coli	HD
Campylobacter jejuni	HD
Campylobacter lari	HD
Citrobacter freundii	SD
Clostridiodes difficile	HD
Corynebacterium diphtheria complex	SD
Cronobacter sakazakii	SD
***Enterobacter cloacae* complex*:** • *Enterobacter asburiae* • *Enterobacter cloacae* • *Enterobacter hormaechei* • *Enterobacter kobei* • *Enterobacter ludwigii* • *Enterobacter roggenkampii*	HD
Enterococcus faecalis	HD
Entercoccus faecium	HD
Escherichia coli	HD
Haemophilus influenzae	SD
Klebsiella aerogenes	SD
Klebsiella oxytoca	HD

Species Name	
Klebsiella pneumoniae	HD
Klebsiella quasipneumoniae	HD
Klebsiella variicola	HD
Legionella pneumophila	SD
Listeria monocytogenes	HD
Mycobacterium tuberculosis	HD
Neisseria gonorrhoeae	SD
Neisseria lactamica	SD
Neisseria meningitidis	SD
Pseudomonas aeruginosa	HD
Salmonella enterica (inc. *typhi*)	HD
Serratia marcescens	HD
Shigella species: • *Shigella boydii (S1 & S3)* • *Shigella dysenteriae (S1 & S3)* • *Shigella flexneri* • *Shigella sonnei*	HD
Staphylococcus aureus	HD
Staphylococcus epidermidis	SD
Stenotrophomonas maltophilia	SD
Streptococcus pneumoniae	SD
Vibrio cholera	SD

HD = High definition; SD = Standard definition