Journal of Biomedical Informatics 166 (2025) 104841

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Bioinformatic challenges in metagenomic next generation sequencing data analysis while unravelling a case of uncommon campylobacteriosis

Rok Kogoj [a], Martin Bosilj [a], Andraž Celar Šturm [a], Misa Korva [a], Katja Strašek Smrdel [a], Eva Kvas [a], Mateja Pirš [a], Lidija Lepen [b], Tina Triglav [a,*]

[a] Institute of Microbiology and Immunology, Faculty of Medicine, University of Ljubljana, Zaloška 4, 1000 Ljubljana, Slovenia
[b] Department of Infectious Diseases, University Medical Centre Ljubljana, Japljeva 2, 1000 Ljubljana, Slovenia

ARTICLE INFO

Keywords:
Metagenomics
NGS
16S rRNA metabarcoding
Campylobacter
Bioinformatics

ABSTRACT

Objective: This study aimed to employ advanced bioinformatics and modern sequencing approaches to solve a diagnostic problem of persistent *Campylobacter* spp. molecular detection yet negative culture results from four consecutive stool samples of a previously healthy patient with newly diagnosed selective IgA deficiency and prolonged diarrhoea.

Methods: Metagenomic next-generation sequencing (mNGS) based on short-paired end reads with basic bioinformatic read classification analysis was used at first. Due to ambiguous results, advanced bioinformatics involving contigs construction and classification, reference genome mappings and reads filtering with BBSplit, additionally coupled with metagenomic long-reads sequencing and Full-length 16S rRNA metabarcoding were employed to further elucidate the results. Virulence factors were analysed using the Prokka Genome Annotation tool. Modified classical bacteriology methods were finally used for further clarification.

Results: Short-pair end reads analysis identified several *Campylobacter* species in all four samples. After advanced bioinformatic approaches were applied, candidatus *C. infans* was suspected as the putative pathogen. This result was further supported by metagenomic long-reads sequencing and Full-length 16S rRNA metabarcoding. Nevertheless, after modifying the culture conditions based on mNGS results, a mixed culture of candidatus *C. infans* and *C. ureolyticus* was obtained. Sequencing of the mixed culture resulted in an 87.48% and 73.47% genome coverage of candidatus *C. infans* and *C. ureolyticus*, respectively. In the candidatus *C. infans* genome more virulence factors hits were found than in the *C. ureolyticus* genome thus supporting the first as the most probable cause of symptoms.

Conclusion: This study shows the pivotal role and strengths of mNGS in unravelling an unusual case of diarrhoea and demonstrates how mNGS can guide established microbiological methods to improve on current limitations. However, it also emphasises the need for careful interpretation of sequencing data, particularly for closely related bacterial species from clinical samples that are known to support complex microbial communities.

1. Introduction

Within the *Campylobacter* genus, *Campylobacter jejuni* and *Campylobacter coli* are considered to be the most frequent causative agents of human bacterial gastrointestinal infections world-wide [1]. Globally, they account for more than 95% of campylobacteriosis cases in developed countries [2]. which also applies to Slovenia. Data from 2023 shows, that *C. jejuni* and *C. coli* represented 77% and 12% of reported campylobacteriosis cases, respectively (National Institute of Public Health data available in Slovenian at https://nijz.si/publikacije/creve sne-nalezljive-bolezni-in-zoonoze-v-sloveniji-v-letu-2023/ Accessed on: 2.02.2025). However, this number is biased due to both clinical indications for stool cultivation and bacterial stool culture conditions optimized for these two thermophilic species [3]. With advances in culture independent tests (CIDT) and innovative culture technologies, other members in the *Campylobacter* genus are also gaining recognition as human pathogens [4].

The awareness of emerging *Campylobacter* species is resulting in growing evidence of abscess formation, bacteremia, gastroenteritis, meningitis, periodontitis, pneumonia, prosthetic infection,

---

* Corresponding author.
E-mail address: tina.triglav@mf.uni-lj.si (T. Triglav).

https://doi.org/10.1016/j.jbi.2025.104841
Received 16 March 2025; Received in revised form 25 April 2025; Accepted 26 April 2025

postinfectious complications (Guillian-Barré syndrome, inflammatory bowel disease, reactive arthritis), and even cancer cases associated with non-jejuni/coli *Campylobacter* species, *including C. concisus* [5], *C. fetus* [6], *C. gracilis* [7]), candidatus *C. infans* [8], *C. insulaenigrae* [9], *C. lari* [10], *C. mucosalis* [11], *C. rectus* [12], *C. showae* [13,14], *C. ureolyticus* [15], *C. upsalensis* [16] and, *C. volucris.* [17] Although these cases are much rarer than infections with *C. jejuni* or *C. coli* and are more likely to occur in low- to middle-income countries [18,19], in immunocompromised patients, the elderly or pregnant women [20], their importance should not be neglected. In symptomatic patients with repeatedly positive *Campylobacter* spp. targeted CIDT and negative culture, the use of metagenomic Next Generation Sequencing (mNGS) could be considered. mNGS overcomes limitations of CIDTs by providing unbiased broad-range detection and subsequent bioinformatic identification of nucleic acids and the possibility to identify previously uncharacterized microbes or novel pathogens [21], however it is not without limitations.

This study describes the pivotal role of mNGS in resolving a case of persistent non-*C. jejuni*/non-*C. coli* campylobacteriosis in a patient with selective IgA deficiency, suspected due to repeatedly positive *Campylobacter* spp. PCR test from diarrheic stool with unsuccessful cultivation. Three next generation sequencing technology approaches, metagenomic Illumina sequencing (ILL), Oxford Nanopore Technologies (ONT) metagenomic sequencing and targeted Full-length 16S rRNA Metabarcoding were used to identify a possible pathogen directly from clinical samples and also to demonstrate the advantages and challenges of each method [22]. Furthermore, we show how mNGS results led to modifications of the conventional cultivation protocol, which finally enabled us to obtain a bacterial culture, which was eventually characterized as two different species.

Statement of significance

Problem or Issue:	Metagenomic data analysis of bacteria in clinical samples containing complex microbial communities.
What Already Known:	Bioinformatic mNGS data analysis based on short pair-end reads can be challenging for the analysis of closely related bacteria.
What this Paper Adds:	This paper highlights the usefulness of mNGS in solving a complex diagnostic case and its role in guiding modifications of cultivation protocols. The paper illustrates the advantages and challenges of using three different next generation sequencing technologies and shows how advanced bioinformatic tools can refine basic classification results, but also demonstrates that improvements are needed to obtain more reliable mNGS results for the identification of closely related bacteria.
Who would benefit from the knowledge in this paper:	Bioinformaticians and microbiologists involved in bacteriology metagenomic NGS data analysis from clinical samples that can sustain complex bacterial communities and clinicians using clinical metagenomics in patient management.

2. Methods

2.1. Patient samples and clinical background

A 45-year-old male patient with no known comorbidities and no regular medication at the time of presentation was examined due to prolonged intermittent diarrhoea, that begun in August 2022. At his initial examination the basic blood laboratory tests were normal. His first stool sample tested positive for *Giardia duodenalis*. A few days after discontinuation of treatment with metronidazole, relapse of diarrhoea occurred. Four consecutive stool samples (Sample 1, 2, 3 and 4) collected in week 10, 14, 27 and 29, respectively, were sent for gastroenteritis multiplex PCR testing. All samples tested positive for *Campylobacter* spp. and negative for *Giardia*. The patient was treated twice with azithromycin with no significant improvement. An abdominal ultrasound detected mild hepatic steatosis. Screening for hepatitis B / C and HIV was negative. With additional diagnostic workup, a selective IgA deficiency was confirmed. After the check-up in October 2023 due to persistent symptoms, the patient was hesitant to start the proposed doxycycline trial based on mixed culture susceptibility test result. In December the patient underwent an esophagogastroduodenoscopy and colonoscopy. Histopathological results revealed *Helicobacter pylori* infection and sessile serrated lesion in the right colon. *H. pylori* eradication with proton-pump inhibitor, amoxicillin and clarithromycin was initiated in January 2024 and was prescribed in the same regimen again in March 2024 due to a positive posttreatment urea breath test. Concomitantly with the treatment, diarrhoea resolved as well. Screening for hepatitis B / C and HIV infection was performed again in June and August 2024, which detected the antibodies to hepatitis B core antigen and surface antigen, indicating recovery from asymptomatic infection. Although the patient denied gastrointestinal symptoms at the last examination in September 2024, he agreed to submit another stool sample. Bacterial syndrome-based multiplex PCR was performed, which yielded positive result for *Campylobacter* spp. with further unsuccessful bacterial cultivation.

The first four consecutive clinical stool samples, PCR-positive for *Campylobacter* spp. and a later obtained *Campylobacter* spp. bacterial culture were included in the study.

2.2. Routine syndrome-based molecular gastroenteritis diagnostic approach

Multiplex real-time PCR (rt-PCR)-based detection of gastrointestinal pathogens from faecal samples was performed from total nucleic acids isolated on a MagNA Pure 24 System using the MagNA Pure Total Nucleic Acid Isolation Kit (Roche, Germany) according to the manufacturer's instructions. *Campylobacter* spp.*, Salmonella* sp.*, Yersinia enterocolitica, Shigella* sp. / EIEC, *Aeromonas* sp.*,* Shiga toxin genes stx1 and stx2, group A rotaviruses, adenoviruses species F, genogroup I and II noroviruses, human astroviruses, human sapoviruses, *Cryptosporidium* spp. and *Giardia duodenalis* are simultaneously detected in four parallel reactions using the LightMix Modular Assays GastroPanel (Roche, Germany). Assays are performed according to the manufacturer's instructions on a Light Cycler 480 II instrument (Roche, Germany). All rt-PCR-positive samples for bacterial targets are always additionally plated onto selected media.

2.3. Standard and expanded Routine Campylobacter cultivation protocol

The standard cultivation protocol consists of plating a faecal sample onto a biplate medium consisting of Karmali agar (Oxoid, UK) and a chromogenic CHROMagarTM *Campylobacter* agar (CHROMagar, France), incubated at 42 °C in microaerophilic atmosphere (Anoxomat system, MART Microbiology BV, Netherlands) for 48 h.

For rt-PCR-positive / standard culture-negative samples cultivation is expanded and consists of sample filtration and cultivation on non-selective enriched media using a modified Cape Town protocol [18,23]. Upon observation of suspicious colonies, MALDI-TOF is used for identification and appropriate antibiotic susceptibility testing (AST) is performed according to EUCAST guidelines [24].

2.4. Metagenomic next generation sequencing

NGS was performed directly from all four clinical samples and later again from bacterial culture. Libraries were prepared with no pre-treatment and, in parallel, with DNAze pre-treatment and Sequence-Independent, Single-Primer-Amplification (SISPA) RNA enrichment. Both ILL short paired end reads and ONT long reads sequencing technologies were employed for metagenomic analysis of clinical samples and nearly complete genome sequence construction from the cultured bacterium. ILL sequencing was performed on a NextSeq 550 instrument (Illumina, USA) loaded with the NextSeq 500/550 HighOutput Kit v2.5 (300 cycles) (Illumina, USA) after NexteraXT (Illumina, USA) library preparation all following the manufacturer's recommendations. For ONT sequencing, the Native Barcoding Kit 24 (V14) (Oxford Nanopore Technologies, UK) was used following the standard protocol. Sequencing was performed on a PromethION R10.4.1 flow cell (Oxford Nanopore Technologies, UK) on a P2 Solo instrument (Oxford Nanopore Technologies, UK). For both systems, the library preparation procedure was quality controlled for concentration and fragmentation using the Qubit dsDNA High Sensitivity Assay on a Qubit 3.0 (Thermo Fisher Scientific, USA) and, the Agilent HS DNA Kit on a Bioanalyzer 2100 (Agilent Technologies, USA), respectively.

2.5. Full-length 16S rRNA Metabarcoding analysis

Full-length 16S rRNA Metabarcoding amplicons were prepared following the Rapid sequencing amplicons – 16S barcoding protocol (Oxford Nanopore Technologies, UK) and sequenced on a MinION R9.4.1 flow cell (Oxford Nanopore Technologies, UK) on a GridION (Oxford Nanopore Technologies, UK) instrument.

2.6. Bioinformatics

Initial bioinformatics steps were performed as described previously [25]. As mapping reference genomes *C. coli* (NZ_CP046317.1), *C. concisus* (NZ_CP012541.1), *C. curvus* (NZ_CP053826.1), *C. fetus* (NZ_CP059443.1), *C. gracilis* (CP012196.1), *C. hominis* (CP000776.1), *C. hyointestinalis* (CP053828.1), candidatus *C. infans* (NZ_CP049075.1), *C. jejuni* (NZ_LR134359.1, NC_002163.1), *C. lanienae* (CP015578.1), *C. lari* (NC_012039.1), *C. mucosalis* (NZ_CP053831.1), *C. rectus* (NZ_CP012543.1), *C. showae* (NZ_CP012544.1), *C. sputorum* (CP019684.1), *C. upsaliensis* (NZ_OU701459.1), *C. ureolyticus* (NZ_CP053832.1) were chosen based on reads classification results and literature search. Additionally, BBSplit (v39.01) was used for non-unique short PE reads filtering [26].

The Prokka Genome Annotation tool (v1.11.1, https://github.com/tseemann/prokka, Accessed on: 19.02.2024) was used for identification and annotation of coding regions in the two obtained genomes from bacterial culture and ABRicate (v1.0.0, https://github.com/tseemann/abricate, Accessed on: 19.02.2024) tool for virulence factors search.

2.7. Modified Campylobacter cultivation and antibiotic susceptibility testing protocol

Based on mNGS results, the expanded cultivation protocol was further modified with a prolonged incubation of blood agar plates in microaerophilic atmosphere at 37 °C for 5 days [8]. AST was performed according to EUCAST and CA-SFM guidelines [24].

3. Results

3.1. Metagenomic next generation sequencing from stool samples

No putative RNA pathogen was found in the DNAze, SISPA pretreated variant of the samples. In all four directly sequenced samples members of the *Campylobacterales* order were found. ILL reads classified as many campylobacter species, with candidatus *C. infans* showing 78%, 35%, 20% and 5% abundance among *Campylobacterales* in the four tested samples, respectively. Similarly, in the first three samples, *de novo* assembled contigs classified as candidatus *C. infans* amounted to 93%, 74% and 33% of all *Campylobacterales* contigs, respectively. In the last sample no *de novo* contigs corresponding to candidatus *C. infans* were found. Mapping of ILL reads to selected *Campylobacter* reference genomes resulted in candidatus *C. infans* genome coverage being the highest in all samples except for the last. In the first sample a 44.6% (45,959 reads) coverage of the reference candidatus *C. infans* genome was found. A seemingly equal number of reads mapped also to all other reference sequences showing genome coverage between 2.6% (*C. jejuni*) and 6.1% (*C. hominis*). After BBSplit was used to remove non-unique short PE reads, mapping revealed an almost absolute drop in the number of reads and coverage for all reference sequences except candidatus *C. infans* (Table 1).

From the same samples mNGS ONT sequencing and amplicon enriched, full-length 16S rRNA Metabarcoding was performed. Classification of metagenomically obtained ONT reads showed candidatus *C. infans* in 77.8%, 66.4% and 74.2% of all *Campylobacterales* reads in the first three samples, respectively. In the *Campylobacterales* reads of the last sample, 1.2% (1/80 reads) were classified as candidatus *C. infans*, while other 79 reads were classified as *Campylobacter* spp. Classification of ONT *de novo* assembled contigs resulted in *Campylobacter* hits only in samples one and three. In both, solely candidatus *C. infans* was detected. ONT reads were mapped to the same reference genomes as ILL reads with the best result revealing a total of 4,646 reads, scattered across the whole of the candidatus *C. infans* reference genome, amounting to a 77.8% coverage in the first sample (Table 1).

Full-length 16S rRNA Metabarcoding showed a similar result than previous methods. In the four samples 16S rRNA *Campylobacter* reads classified as candidatus *C. infans* in 80%, 89%, 82% and, 21%, respectively.

3.2. Modified cultivation approach and Campylobacter species identification from culture

Based on mNGS results, the standard cultivation process has been further modified and suspected *Campylobacter* spp. colonies were observed. Initially, MALDI-TOF identification was inconclusive, yet the best hit was *Campylobacter* spp. We performed several passages to increase bacterial mass and perform AST using gradient diffusion testing (Biomerieux, Marcy l'Etoile, France) for ampicillin, ertapenem, gentamicin, ciprofloxacin, tetracycline and erythromycin (256 mg/L) on MH-F agar. We noted a double inhibition zone for ampicillin (0.016 and 1.5 mg/L) and tetracycline (0.5 and 2 mg/L). After performing several subcultivations to clarify the reason for double inhibition zones, MALDI-TOF identification resulted in *C. ureolyticus* (score: 2.16). Initial and subcultivated colonies were both tested with the *Campylobacter* spp. rt-PCR: initial colonies were rt-PCR positive, while the pure (*C. ureolyticus*) colonies were rt-PCR negative. At the end, isolation of initial colonies in pure culture was unsuccessful as the bacteria were no longer viable; pure culture of *C. ureolyticus* had ampicillin MIC 0.016 mg/L and tetracycline MIC 2 mg/L.

The initial colonies were analysed using mNGS. The great majority of ILL reads (84.7%) belonged to *C. ureolyticus* and 10.8% to candidatus *C. infans*. The rest of ILL reads classified to 32 more *Campylobacter* species. By unique reads mapping, for candidatus *C. infans* 87.1% coverage of the genome with a 67.4 mean depth, while for *C. ureolyticus* 73.0% genome coverage with a 395.3 mean depth was observed (Table 2).

3.3. Genome annotation and virulence factors analysis

In total 679 genes were identified and annotated with an additional 819 hypothetical proteins of unknown function for candidatus *C. infans* and 625 genes with an additional 799 hypothetical proteins of unknown function for *C. ureolyticus*. Both genomes shared 443 genes. For candidatus *C. infans*, 23 *C. jejuni* related genes associated with virulence were detected using a threshold of minimum 60% coverage and 60% identity. From these, 17 are associated with motility, 4 with immune modulation and, 1 with adherence. Under same conditions only 1 *C. jejuni* related gene associated with virulence (adherence) was detected in the *C. ureolyticus* genome.

Table 1.
Reads mapping results and genome coverage percentages for non-unique and unique ILL reads and ONT reads to selected *Campylobacter* species reference genomes.

Organism (Reference Acc. No.)	Sample 1 (week 10) (*Campylobater* spp. Cp = 21.3)			Sample 2 (week 14) (*Campylobater* spp. Cp = 26.0)			Sample 3 (week 27) (*Campylobater* spp. Cp = 23.2)			Sample 4 (week 29) (*Campylobater* spp. Cp = 32.2)		
	ILL reads		ONT reads	ILL reads		ONT reads	ILL reads		ONT reads	ILL reads		ONT reads
	non-unique	unique		non-unique	unique		non-unique	unique		non-unique	unique	
	coverage (reads)	coverage (reads)	coverage (reads)	coverage (reads)	coverage (reads)	coverage (reads)	coverage (reads)	coverage (reads)	coverage (reads)	coverage (reads)	coverage (reads)	coverage (reads)
cand. *C. infans* (NZ_CP049075.1)	44.6 % (45,959)	40.8 % (14,688)	77.8 % (4,646)	7.7 % (5,024)	6.4 % (1,305)	3.9 % (177)	3.8 % (5,132)	2.4 % (501)	19.3 % (617)	1.4 % (3,821)	0.2 % (52)	0.3 % (12)
C. hominis (CP000776.1)	6.1 % (36,778)	0.1 % (424)	0.5 % (195)	1.5 % (4,509)	<0.1 % (1)	<0.1 % (2)	1.9 % (5,652)	<0.1 % (11)	0.1 % (15)	1.7 % (5,305)	<0.1 % (43)	0.3 % (10)
C. ureolyticus (NZ_CP053832.1)	5.8 % (38,937)	0.1 % (179)	0.3 % (14)	1.1 % (3,436)	<0.1 % (2)	0.0 % (0)	1.6 % (4,700)	<0.1 % (2)	<0.1 % (1)	1.6 % (5,042)	<0.1 % (8)	0.1 % (3)
C. sputorum (CP019684.1)	5.8 % (34,090)	0.1 % (122)	0.4 % (22)	1.1 % (3,106)	<0.1 % (5)	0.0 % (0)	1.6 % (4,463)	<0.1 % (10)	<0.1 % (12)	1.7 % (4,652)	<0.1 % (7)	<0.1 % (1)
C. lari (NC_012039.1)	5.4 % (30,427)	0.1 % (170)	0.3 % (29)	1.0 % (2,659)	0.0 % (0)	0.0 % (0)	1.5 % (4,534)	<0.1 % (3)	0.1 % (3)	1.4 % (4,281)	<0.1 % (8)	0.2 % (8)
C. upsaliensis (NZ_OU701459.1)	5.3 % (31,604)	0.3 % (274)	0.7 % (26)	1.3 % (3,241)	<0.1 % (10)	0.0 % (0)	1.5 % (3,813)	<0.1 % (5)	0.2 % (8)	1.3 % (3,877)	<0.1 % (6)	0.0 % (0)
C. coli (NZ_CP046317.1)	4.9 % (29,934)	<0.1 % (297)	0.1 % (16)	0.9 % (2,609)	<0.1 % (5)	0.0 % (0)	1.4 % (3,736)	<0.1 % (1)	<0.1 % (1)	1.3 % (3,865)	<0.1 % (2)	0.0 % (0)
C. concisus (NZ_CP012541.1)	4.7 % (37,150)	0.2 % (287)	1.0 % (30)	1.2 % (4,039)	<0.1 % (32)	0.0 % (0)	1.3 % (4,436)	<0.1 % (12)	<0.1 % (1)	1.3 % (4,125)	0.0 % (0)	0.0 % (0)
C. gracilis (CP012196.1)	4.6 % (38,006)	0.2 % (337)	0.3 % (34)	2.7 % (13,773)	0.1 % (201)	0.2 % (3)	2.0 % (7,688)	<0.1 % (84)	0.2 % (15)	1.4 % (4,646)	<0.1 % (46)	<0.1 % (5)
C. lanienae (CP015578.1)	4.4 % (27,708)	0.3 % (1,501)	0.3 % (486)	1.0 % (2,634)	0.2 % (227)	0.1 % (23)	1.3 % (3,296)	0.2 % (226)	0.3 % (102)	1.3 % (3,895)	0.2 % (250)	0.2 % (46)
C. fetus (NZ_CP059443.1)	4.4 % (28,524)	<0.1 % (128)	0.1 % (16)	1.0 % (2,994)	<0.1 % (2)	0.0 % (0)	1.4 % (3,801)	<0.1 % (2)	<0.1 % (5)	1.2 % (3,691)	<0.1 % (39)	<0.1 % (6)
C. curvus (NZ_CP053826.1)	4.3 % (33,831)	<0.1 % (228)	0.3 % (13)	2.0 % (8,119)	<0.1 % (135)	0.0 % (0)	1.6 % (5,816)	<0.1 % (53)	<0.1 % (1)	1.2 % (3,999)	<0.1 % (18)	0.1 % (1)
C. mucosalis (NZ_CP053831.1)	4.3 % (30,406)	<0.1 % (288)	0.3 % (54)	1.0 % (2,800)	<0.1 % (2)	0.0 % (0)	1.2 % (3,236)	<0.1 % (7)	0.1 % (1)	1.1 % (3,231)	<0.1 % (7)	0.1 % (1)
C. hyointestinalis (CP053828.1)	4.3 % (30,027)	<0.1 % (120)	0.2 % (4)	1.0 % (3,013)	<0.1 % (1)	<0.1 % (1)	1.3 % (3,843)	<0.1 % (5)	0.2 % (2)	1.2 % (4,143)	<0.1 % (8)	0.1 % (2)
C. rectus (NZ_CP012543.1)	3.7 % (33,759)	<0.1 % (182)	0.4 % (14)	1.9 % (9,756)	0.1 % (131)	<0.1 % (1)	1.5 % (6,377)	<0.1 % (55)	<0.1 % (1)	1.1 % (4,387)	<0.1 % (32)	<0.1 % (2)
C. showae (NZ_CP012544.1)	3.6 % (27,140)	0.1 % (311)	0.6 % (47)	1.8 % (7,766)	0.1 % (100)	0.0 % (0)	1.4 % (4,652)	<0.1 % (59)	0.1 % (3)	1.1 % (3,415)	<0.1 % (14)	0.0 % (0)
C. jejuni subsp. *doylei* (NZ_LR134359.1)	2.7 % (18,807)	<0.1 % (150)	0.6 % (42)	0.5 % (1,502)	0.0 % (0)	0.0 % (0)	0.7 % (2,128)	<0.1 % (13)	<0.1 % (2)	0.7 % (2,532)	<0.1 % (16)	0.2 % (3)
C. jejuni subsp. *jejuni* (NC_002163.1)	2.6 % (16,697)	<0.1 % (79)	0.7 % (43)	0.5 % (1,701)	0.0 % (0)	0.0 % (0)	0.7 % (2,527)	0.0 % (0)	0.1 % (5)	0.7 % (2,642)	0.0 % (0)	0.2 % (4)

Table 2

Mapping results of ILL reads from the *Campylobacter* culture obtained after culture conditions modification without and with BBSplit filtering of non-unique reads.

Organism (reference Acc. No.)	non-unique				unique			
	Reads [No.]	Bases [No.]	Coverage [%]	Meandepth [No.]	Reads [No.]	Bases [No.]	Coverage [%]	Meandepth [No.]
cand. *C. infans* (NZ_CP049075.1)	924,655	1,534,796	87.48	71.71	898,178	1,527,673	87.07	67.36
C. ureolyticus (NZ_CP053832.1)	5,791,521	1,337,450	73.47	422.52	5,652,202	1,329,151	73.02	395.31
C. hominis (CP000776.1)	217,401	37,495	2.19	11.53	134,007	30,314	1.77	8.64
C. upsaliensis (NZ_OU701459.1)	20,874	17,826	1.13	0.73	5,907	13,366	0.84	0.47
C. sputorum (CP019684.1)	34,551	11,657	0.68	0.71	4,947	3,505	0.20	0.22
C. jejuni subsp. *jejuni* (NC_002163.1)	8,845	7,401	0.45	0.19	1,155	4,855	0.30	0.08
C. jejuni subsp. *doylei* (NZ_LR134359.1)	14,111	7,925	0.42	0.18	446	4,235	0.22	0.02
C. lanienae (CP015578.1)	24,119	5,988	0.38	0.43	2,668	1,442	0.09	0.09
C. hyointestinalis (CP053828.1)	18,581	6,616	0.37	0.26	661	2,485	0.14	0.03
C. coli (NZ_CP046317.1)	13,778	5,794	0.35	0.20	460	2,364	0.14	0.03
C. lari (NC_012039.1)	12,747	5,198	0.34	0.17	325	1,013	0.07	0.01
C. mucosalis (NZ_CP053831.1)	22,852	5,694	0.32	0.27	259	1,144	0.06	0.01
C. concisus (NZ_CP012541.1)	22,072	5,683	0.31	0.33	1,549	1,417	0.08	0.05
C. fetus (NZ_CP059443.1)	20,166	4,946	0.28	0.24	511	832	0.05	0.01
C. curvus (NZ_CP053826.1)	7,481	2,902	0.15	0.09	173	672	0.03	0.00
C. gracilis (CP012196.1)	10,949	2,266	0.10	0.12	198	299	0.01	0.00
C. rectus (NZ_CP012543.1)	4,938	1,861	0.07	0.04	0	0	0.00	0.00
C. showae (NZ_CP012544.1)	3,200	1,374	0.07	0.03	0	0	0.00	0.00

4. Discussion

In this study an in depth molecular and traditional microbiological investigation of a prolonged diarrhoea case in an IgA deficient patient is demonstrated. Repeatedly positive *Campylobacter* spp. rt-PCR and consistently negative culture lead to the assumption that some other *Campylobacter* species might be involved. Three NGS approaches were selected for the identification of the most probable causative agent: metagenomic ILL sequencing, metagenomic ONT sequencing and Full-length 16S rRNA Metabarcoding. Metagenomic sequencing was obviously chosen due to its unbiased nature of nucleic acid detection [21], however three approaches were ultimately used in order to exploit their individual strengths and circumvent respective limitations [22] in order to increase the confidence in the final result.

Initial mNGS ILL results excluded other pathogens and indicated the presence of more than one C*ampylobacter* species in the stool samples. A previous study from 2022 by Parker et al. detected up to 6 *Campylobacter* species in stool samples from children [27]. Similarly, the study from Bian et al. also showed multiple *Campylobacter* species infections in infants [28]. However, in these studies the number of putatively coexisting species is much lower than in our case. Under default classification parameters, ILL reads might result in multiple hits, especially if they correspond to conserved genomic regions. Therefore, contigs were assembled *de novo* and used for classification analysis. The number of detected *Campylobacter* species decreased considerably. Genome mapping results supported candidatus *C. infans* as the most likely pathogen, but a non-negligible number of reads mapped to all other reference genomes used. The visualisation of mapped reads to selected reference genomes, with the exception of candidatus *C. infans*, showed localised mapping in discrete towers of reads, probably representing conserved genomic regions. Therefore, additional bioinformatics were used to filter out ambiguous reads and map only the remaining (unique) reads. Using this approach, almost no reads mapped to other species than candidatus *C. infans*. Furthermore, the percentage drop for candidatus *C. infans* was similar to the percentage of mapped reads to other *Campylobacter* spp. prior to the usage of additional bioinformatic polishing, indicating that candidatus *C. infans* was the most probable cause of diarrhoea and other detected *Campylobacter* spp. were a bioinformatic anomaly.

Metagenomic ONT sequencing was additionally used in order to obtain biological data that would improve resolution especially for repetitive, complex and conserved genomic regions as shown previously *in vitro* [29] and *in silico* [30]. This approach further confirmed initial results as it also showed candidatus *C. infans* in the highest abundances among *Campylobacter* reads in Samples 1 through 3 (77.8%, 66.4%, 74.2%) and also amounted to a higher portion of covered genome in Sample 1 (77.8%).

As a final complementary method, Full-length 16S rRNA Metabarcoding was chosen as it is a targeted approach, specifically tailored to analyse microbial communities and, widely used in microbiome studies [31]. Candidatus *C. infans* was again the most abundantly detected *Camylobacter* species.

Altogether, mNGS, additional bioinformatic analyses as well as Full-length 16S rRNA Metabarcoding all seemed to point to candidatus *C. infans* as the probable cause of the patient's gastrointestinal symptoms. Candidatus *C. infans* was first described and identified as a potential pathogen in 2020 in the study by Bian et al. on breastfed infants with diarrhoea. The authors detected candidatus *C. infans* three times more frequently in exclusively breastfed than in nonbreastfed diseased infants. In one extreme case candidatus *C. infans* composed 83% of the total faecal microbiome, thus strengthening its role as a pathogen [28]. Soon after the molecular characterisation of this novel *Campylobacter* species, its isolation was also achieved from a patient with travel-related chronic diarrhoea. However, in this case multiple *Campylobacter* species were identified during the 15 months of observation, raising concern about the causative relation between candidatus *C. infans* and symptoms [8]. Recent research from Europe, USA and Peru in 2021 and 2022 have either further strengthened the role of candidatus *C. infans* as a pathogen [32] or showed evidence of no association with disease [33]. Due to conflicting results in the published literature, a modified culture protocol was employed which would enable the growth of candidatus *C. infans*. A culture was obtained, however, not as a single *Campylobacter* species. Initially MALDI-TOF did not yield an identification above genus level while diagnostic *Campylobacter* spp. rt-PCR was positive. Subsequently we performed several additional passages in order to obtain a pure culture of candidatus *C. infans*, however *C. ureolyticus* was in the end identified by MALDI-TOF. We believe there was initially only a small amount of *C. ureolyticus* present in the sample which has subsequently overgrown a slower growing candidatus *C. infans*. mNGS was performed from the culture and indeed both species were bioinformatically identified. *C. ureolyticus* has also been previously shown as a potential pathogen [34], however, after performing genome annotation and virulence factors analysis coupled with the relative abundance of reads in the stool samples we still believe candidatus *C. infans* to be the more likely cause of symptoms. Despite the fact that several microbiological investigation techniques singled out candidatus

C. infans as the potential causative agent of prolonged diarrhoea, a comprehensive analysis which also takes into account detailed clinical data is needed and the causal relation between detected bacteria and patient's gastrointestinal symptoms cannot be confirmed. It could only represent colonization of the intestinal tract or potentially pathogenic bacteria, when additional infections or a weakened immune system are present (as *H. pylori* and hepatitis B infection in addition to selective IgA immunodeficiency in our case).

Although this case is based on the detection of a rare *Campylobacter* species, similar situations can be expected when analysing sequencing data from clinical samples containing complex bacterial communities. This is especially important when several species within a genus can be present simultaneously, but not all of them are pathogenic. Due to high genome similarity within a genus, respective species can either be overlooked (falsely negative) or overrepresented (falsely positive) if only read classification is used to analyse metagenomic data, as clearly demonstrated in this work. Although the wet–lab part of the procedure would probably need adjustments to better suit specific cases, most parts of the described dry–lab approach (classification of reads and *de novo* contigs, filtering of ambiguous reads, genome annotation and virulence factors identification) can be generalized and used by researchers focusing on metagenomic identification of bacteria to the species level in complex samples. The only specific difference is the selection of reference genomes for reads mapping, which has to be done on a case-by-case basis, classification results and other available data in order to achieve the most accurate metagenomic result.

5. Conclusion

In conclusion, this study shows the pivotal role of mNGS in solving a specific diagnostic problem, but also the individual limitations of different approaches when applied for the analysis of closely related bacteria. Our results show how mNGS coupled with modern bioinformatic tools, when used appropriately, can refine basic classification results to a higher resolution, reducing the possibility of misidentification. As shown in this work, mNGS findings can modify established microbiological methods which in turn enables the isolation and definite characterisation of a bacterial species that would normally remain unidentified. In spite of remaining limitations when using mNGS to identify putative bacterial pathogens, as also shown by our results, mNGS offers a truly unique opportunity to solve complicated diagnostic cases.

Ethical statement.

The study protocol conformed to the Declaration of Helsinki, Oviedo Convention on Human Rights and Biomedicine, and Slovenian Code on Medical Deontology. Due to the retrospective nature of the mNGS analysis which had no direct impact on the patient management, the need for ethics approval was waived.

The patient signed a written consent allowing to use and publish of his medical data.

CRediT authorship contribution statement

Rok Kogoj: Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Martin Bosilj:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis. **Andraž Celar Šturm:** Investigation, Formal analysis. **Misa Korva:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Katja Strašek Smrdel:** Writing – original draft, Investigation. **Eva Kvas:** Writing – original draft, Investigation. **Mateja Pirš:** Writing – review & editing, Validation, Methodology, Investigation. **Lidija Lepen:** Writing – review & editing, Resources, Investigation. **Tina Triglav:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to express their utmost gratitude to Doroteja Vlaj for her excellent mNGS wet-laboratory work and all the personnel involved in patient care.

Funding:

This work has been partially supported by the Slovenian Research and Innovation Agency (grant P3-0083). This publication reflects only the authors' view. The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbi.2025.104841.

References

[1] M.D. Kirk, S.M. Pires, R.E. Black, M. Caipo, J.A. Crump, B. Devleesschauwer, D. Döpfer, A. Fazil, C.L. Fischer-Walker, T. Hald, A.J. Hall, K.H. Keddy, R.J. Lake, C.F. Lanata, P.R. Torgerson, A.H. Havelaar, F.J. Angulo, World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and Viral Diseases, 2010: A Data Synthesis, PLoS Med. 12 (12) (2015) e1001921, https://doi.org/10.1371/journal.pmed.1001921.

[2] J.P. Butzler, Campylobacter, from obscurity to celebrity, Clin Microbiol Infect. 10 (10) (2004) 868–876, https://doi.org/10.1111/j.1469-0691.2004.00983.x. PMID: 15373879.

[3] F.C. Campylobacter, Clin Lab Med. 35 (2) (2015) 289–298, https://doi.org/10.1016/j.cll.2015.03.001. PMID: 26004643.

[4] S.M. Man, The clinical importance of emerging Campylobacter species, Nat Rev Gastroenterol Hepatol. 8 (12) (2011) 669–685, https://doi.org/10.1038/nrgastro.2011.191. PMID: 22025030.

[5] N.O. Kaakoush, H.M. Mitchell, Campylobacter concisus - A new player in intestinal disease, Front Cell Infect Microbiol. 3 (2) (2012) 4, https://doi.org/10.3389/fcimb.2012.00004. PMID: 22919596; PMCID: PMC3417403.

[6] C.T. Lynch, C. Buttimer, L. Epping, J. O'Connor, N. Walsh, C. McCarthy, D. O'Brien, C. Vaughan, T. Semmler, D. Bolton, A. Coffey, B. Lucey, Phenotypic and genetic analyses of two Campylobacter fetus isolates from a patient with relapsed prosthetic valve endocarditis, Pathog Dis. (2022), https://doi.org/10.1093/femspd/ftab055. Jan 7;79(9):ftab055. PMID: 34962980.

[7] T. Shinha, Fatal bacteremia caused by Campylobacter gracilis, United States. Emerg Infect Dis. 21 (6) (2015) 1084–1085, https://doi.org/10.3201/eid2106.142043. PMID: 25988682; PMCID: PMC4451896.

[8] J. Flipse, B. Duim, J.A. Wallinga, L.R.H. de Wijkersloot h, L.V. Graaf-van Bloois, A. J. Timmerman, A.L. Zomer, K.T. Veldman, J.A. Wagenaar, P. Bloembergen, A Case of Persistent Diarrhea in a Man with the Molecular Detection of Various Campylobacter species and the First Isolation of candidatus Campylobacter infans, Pathogens. 9 (12) (2020) 1003, https://doi.org/10.3390/pathogens9121003. PMID: 33265947; PMCID: PMC7761484.

[9] M. Kyotani, T. Kenzaka, H. Akita, S. Arakawa, Campylobacter insulaenigrae bacteremia with meningitis: a case report, BMC Infect Dis. 21 (1) (2021) 633, https://doi.org/10.1186/s12879-021-06353-8. PMID: 34210285; PMCID: PMC8252270.

[10] Patrick ME, Henao OL, Robinson T, Geissler AL, Cronquist A, Hanna S, Hurd S, Medalla F, Pruckler J, Mahon BE. Features of illnesses caused by five species of Campylobacter, Foodborne Diseases Active Surveillance Network (FoodNet) - 2010-2015. Epidemiol Infect. 2018 Jan;146(1):1-10. doi: 10.1017/S0950268817002370. Epub 2017 Dec 14. PMID: 29237513; PMCID: PMC9134565.

[11] N. Figura, P. Guglielmetti, A. Zanchi, N. Partini, D. Armellini, P.F. Bayeli, M. Bugnoli, S. Verdiani, Two cases of Campylobacter mucosalis enteritis in children, J Clin Microbiol. 31 (3) (1993) 727–728, https://doi.org/10.1128/jcm.31.3.727-728.1993. PMID: 8458973; PMCID: PMC262855.

[12] T. Ogata, T. Urata, D. Nemoto, S. Hitomi, Thoracic empyema caused by Campylobacter rectus, J Infect Chemother. 23 (3) (2017) 185–188, https://doi.org/10.1016/j.jiac.2016.08.013. Epub 2016 Sep 25 PMID: 27681234.

[13] Y. Etoh, F.E. Dewhirst, B.J. Paster, A. Yamamoto, N. Goto, Campylobacter showae sp. nov., isolated from the human oral cavity, Int J Syst Bacteriol. 43 (4) (1993) 631–639, https://doi.org/10.1099/00207713-43-4-631. PMID: 7694633.

[14] R.L. Warren, D.J. Freeman, S. Pleasance, P. Watson, R.A. Moore, K. Cochrane, E. Allen-Vercoe, R.A. Holt, Co-occurrence of anaerobic bacteria in colorectal

carcinomas, Microbiome. 1 (1) (2013) 16, https://doi.org/10.1186/2049-2618-1-16. PMID: 24450771; PMCID: PMC3971631.

[15] S. Bullman, D. Corcoran, J. O'Leary, B. Lucey, D. Byrne, R.D. Sleator, Campylobacter ureolyticus: an emerging gastrointestinal pathogen? FEMS Immunol Med Microbiol. 61 (2) (2011) 228–230, https://doi.org/10.1111/j.1574-695X.2010.00760.x. Epub 2010 Dec 6 PMID: 21320172.

[16] Y. Ohkoshi, T. Sato, H. Murabayashi, K. Sakai, Y. Takakuwa, Y. Fukushima, C. Nakajima, Y. Suzuki, S.I. Yokota, Campylobacter upsaliensis isolated from a giant hepatic cyst, J Infect Chemother. 26 (7) (2020) 752–755, https://doi.org/10.1016/j.jiac.2020.02.015. Epub 2020 Mar 19 PMID: 32199791.

[17] Kweon OJ, Lim YK, Yoo B, Kim HR, Kim TH, Lee MK. First Case Report of Campylobacter volucris Bacteremia in an Immunocompromised Patient. J Clin Microbiol. 2015 Jun;53(6):1976-8. doi: 10.1128/JCM.00442-15. Epub 2015 Apr 1. PMID: 25832303; PMCID: PMC4432047.

[18] A.J. Lastovica, Emerging Campylobacter spp.: The tip of the iceberg, Clinical Microbology Newsletter 28 (2006) 49–56.

[19] R. François, P.P. Yori, S. Rouhani, M. Siguas Salas, M. Paredes Olortegui, D. Rengifo Trigoso, N. Pisanic, R. Burga, R. Meza, G. Meza Sanchez, M.J. Gregory, E.R. Houpt, J.A. Platts-Mills, M.N. Kosek, The other Campylobacters: Not innocent bystanders in endemic diarrhea and dysentery in children in low-income settings, PLoS Negl Trop Dis. 12 (2) (2018) e0006200, https://doi.org/10.1371/journal.pntd.0006200. PMID: 29415075; PMCID: PMC5819825.

[20] M.B. Skirrow, M.J. Blaser, Clinical Aspects of Campylobacter Infection, in: I. Nachamkin, M.J. Blaser (Eds.), Campylobacter, American Society for Microbiology, Washington DC, 2000, pp. 69–88.

[21] S.C. Forster, B.O. Anonye, N. Kumar, B.A. Neville, M.D. Stares, D. Goulding, et al., Culturing of "unculturable" human microbiota reveals novel taxa and extensive sporulation, Nature 533 (2016) 543–546, https://doi.org/10.1038/nature17645.

[22] M. Ben Khedher, K. Ghedira, J.M. Rolain, R. Ruimy, O. Croce, Application and Challenge of 3rd Generation Sequencing for Clinical Bacterial Studies, Int J Mol Sci. 23 (3) (2022) 1395, https://doi.org/10.3390/ijms23031395. PMID: 35163319; PMCID: PMC8835973.

[23] Nielsen, Hans Linde, Jørgen Engberg, Tove Ejlertsen, and Henrik Nielsen. 'Comparison of Polycarbonate and Cellulose Acetate Membrane Filters for Isolation of Campylobacter Concisus from Stool Samples'. Diagnostic Microbiology and Infectious Disease 76, no. 4 (August 2013): 549–50. https://doi.org/10.1016/j.diagmicrobio.2013.05.002.

[24] 'EUCAST V_13.0_Breakpoint_Tables'. EUCAST, 2023. https://www.eucast.org/clinical_breakpoints/.

[25] Bosilj M, Suljič A, Zakotnik S, Slunečko J, Kogoj R, Korva M. MetaAll: integrative bioinformatics workflow for analysing clinical metagenomic data. Brief Bioinform. 2024 Sep 23;25(6):bbae597. doi: 10.1093/bib/bbae597. PMID: 39550223; PMCID: PMC11568877.

[26] Bushnell, B. BBMap: A Fast, Accurate, Splice-Aware Aligner. in (2014).

[27] C.T. Parker, F. Schiaffino, S. Huynh, M. Paredes Olortegui, P. Peñataro Yori, P. F. Garcia Bardales, T. Pinedo Vasquez, G.E. Curico Huansi, K. Manzanares Villanueva, W.V. Shapiama Lopez, K.K. Cooper, M.N. Kosek, Shotgun metagenomics of fecal samples from children in Peru reveals frequent complex co-infections with multiple Campylobacter species, PLoS Negl. Trop. Dis. 16 (10) (2022) e0010815, https://doi.org/10.1371/journal.pntd.0010815.

[28] Bian, X., Garber, J. M., Cooper, K. K., Huynh, S., Jones, J., Mills, M. K., Rafala, D., Nasrin, D., Kotloff, K. L., Parker, C. T., Tennant, S. M., Miller, W. G., & Szymanski, C. M. (2020). Campylobacter Abundance in Breastfed Infants and Identification of a New Species in the Global Enterics Multicenter Study. mSphere, 5(1), e00735-19. https://doi.org/10.1128/mSphere.00735-19.

[29] W. Zhao, W. Zeng, B. Pang, M. Luo, Y. Peng, J. Xu, B. Kan, Z. Li, X. Lu, Oxford nanopore long-read sequencing enables the generation of complete bacterial and plasmid genomes without short-read sequencing, Front Microbiol. 15 (14) (2023) 1179966, https://doi.org/10.3389/fmicb.2023.1179966. PMID: 37256057; PMCID: PMC10225699.

[30] W.S. Pearman, N.E. Freed, O.K. Silander, Testing the advantages and disadvantages of short- and long- read eukaryotic metagenomics using simulated reads, BMC Bioinf. 21 (1) (2020) 220, https://doi.org/10.1186/s12859-020-3528-4. PMID: 32471343; PMCID: PMC7257156.

[31] Noecker C, McNally CP, Eng A, Borenstein E. High-resolution characterization of the human microbiome. Transl Res. 2017 Jan;179:7-23. doi: 10.1016/j.trsl.2016.07.012. Epub 2016 Jul 25. PMID: 27513210; PMCID: PMC5164958.

[32] B. Duim, B.L. van der Graaf-van, A. Timmerman, J.A. Wagenaar, J. Flipse, J. Wallinga, P. Bloembergen, W.G. Miller, A.L. Zomer, Complete Genome Sequence of a Clinical Campylobacter Isolate Identical to a Novel Campylobacter Species, Microbiol Resour Announc. 10 (7) (2021) e00721–e00820, https://doi.org/10.1128/MRA.00721-20. PMID: 33602730; PMCID: PMC7892663.

[33] P.F. Garcia Bardales, F. Schiaffino, S. Huynh, M. Paredes Olortegui, P. Peñataro Yori, T. Pinedo Vasquez, K. Manzanares Villanueva, G.E. Curico Huansi, W. V. Shapiama Lopez, K.K. Cooper, C.T. Parker, M.N. Kosek, "Candidatus Campylobacter infans" detection is not associated with diarrhea in children under the age of 2 in Peru, PLoS Negl. Trop. Dis. 16 (10) (2022) e0010869, https://doi.org/10.1371/journal.pntd.0010869.

[34] O'Donovan D, Corcoran GD, Lucey B, Sleator RD. Campylobacter ureolyticus: a portrait of the pathogen. Virulence. 2014 May 15;5(4):498-506. doi: 10.4161/viru.28776. Epub 2014 Apr 9. PMID: 24717836; PMCID: PMC4063811.