

Salmonella Paratyphi Infection: Use of Nanopore Sequencing as a Vivid Alternative for the Identification of Invading Bacteria

Martin Chmel^{1,2}, Oldřich Bartoš^{2,3}, Ondřej Beran¹, Petr Pajer², Jiří Dresler², Martina Čurdová⁴, Michal Holub¹

¹Department of Infectious Diseases, First Faculty of Medicine, Charles University and Military University Hospital Prague, Prague, Czech Republic;

²Military Health Institute, Military Medical Agency, Prague, Czech Republic;

³Institute of Animal Physiology and Genetics, Laboratory of Fish Genetics, The Czech Academy of Sciences, Liběchov, Czech Republic;

⁴Department of Clinical Microbiology, Military University Hospital Prague, Prague, Czech Republic

Received July 23, 2020; Accepted April 30, 2021.

Key words: Enteric fever – *Salmonella* – Paratyphoid fever – Nanopore sequencing – Pathogen identification

Abstract: In our study we present an overview of the use of Oxford Nanopore Technologies (ONT) sequencing technology on the background of Enteric fever. Unlike traditional methods (e.g., qPCR, serological tests), the nanopore sequencing technology enables virtually real-time data generation and highly accurate pathogen identification and characterization. Blood cultures were obtained from a 48-year-old female patient suffering from a high fever, headache and diarrhea. Nevertheless, both the initial serological tests and stool culture appeared to be negative. Therefore, the bacterial isolate from blood culture was used for nanopore sequencing (ONT). This technique in combination with subsequent bioinformatic analyses allowed for

This study was supported by funding from projects of the Ministry of the Interior of the Czech Republic (VH20172020012: Preparation of the collection of biologically significant toxins with the support of European Biological European Biodefence Laboratory Network), Ministry of Defence of the Czech Republic through a Long-term organization development plan 907930101413 and a project MO1012, and a project of Charles University SVV 260 520.

Mailing Address: Martin Chmel, MSc., Department of Infectious Diseases, First Faculty of Medicine, Charles University and Military University Hospital Prague, U Vojenské nemocnice 1200, 162 02 Prague 6, Czech Republic; Mobile Phone: +420 722 712 999; e-mail: Martin.Chmel@lf1.cuni.cz

prompt identification of the disease-causative agent as *Salmonella enterica* subsp. *enterica* serovar Paratyphi A. The National Reference Laboratory for *Salmonella* (NIPH) independently reported this isolate also as serovar Paratyphi A on the basis of results of biochemical and agglutination tests. Therefore, our results are in concordance with certified standards. Furthermore, the data enabled us to assess some basic questions concerning the comparative genomics, i.e., to describe whether the isolated strain differs from the formerly published ones or not. Quite surprisingly, these results indicate that we have detected a novel and so far, unknown variety of this bacteria.

Introduction

Rapid and unambiguous identification of the pathogenic organism represents a crucial step for the subsequent treatment of patients. However, in many cases this might be complicated, for example by atypical clinical symptoms of a patient. Furthermore, atypical behavior of a pathogen itself may hamper the usual routines as well as the choice of relevant identification techniques. Despite the number of available, well established and certified methodologies and procedures designed for the determination of infectious disease causative agents (Váradi et al., 2017), an atypical behavior of a pathogen could lead to erroneous outcomes.

Within this study we want to focus on Enteric fever, which is a human disease caused by the pathogenic bacteria *Salmonella enterica*. Its clinical identification is usually based on some of the following approaches, or rather, a combination of those approaches: microbiological cultivation methods (Wain et al., 1998); serological tests, e.g., the Widal test (Olopoenia, 2000); PCR based tests (Song et al., 1993; Tennant et al., 2015) and of course the symptomatic manifestation of the disease itself (Matono et al., 2017). The disadvantage of these, let's say, standard approaches represent the fact that the diagnostics is based upon many variables including for example idiosyncratic human reactions, ambiguous clinical manifestation of closely related causative agents (Maskey et al., 2006), as well as the experience and erudition of the responsible physician.

On the other hand, the application of the Third Generation Sequencing (Schadt et al., 2010), Oxford Nanopore Technologies (ONT) sequencing (Magi et al., 2018), offers unique performance, which already plays its role in current, but more importantly, in tomorrow's medical and biological research (Fuselli et al., 2018; Ton et al., 2018) and maybe even in the future daily routines.

The main advantage of the ONT nanopore sequencing is represented by several facts, the sequencing device (i.e., the MinION) is delivered within so called "starter pack" (ONT) basically for free, furthermore, the MinION is a portable USB powered device, which allows its deployment even in field conditions, e.g., literally even in a jungle (Watsa et al., 2019). Nevertheless, its major benefits are the real-time generation of the sequencing data and the read length representing the key parameter for subsequent bioinformatic analyses (Norris et al., 2016).

Within this study we point out that this rapidly evolving techniques have a great potential and maybe, even if not in their current form, represent the future of diagnostic and analytical practice.

Case report

A 48-year-old female was admitted to the Department of Infectious Diseases with a diarrhea following a 10-day travel in India (Delhi and the northern part of India). She had been vaccinated against typhoid fever and cholera three years ago. Initial stool sample culture was negative as well as parasite detection in the stool. The diarrhea was resolved by symptomatic therapy within 14 days. After five days, the patient developed a severe headache, fever with spikes up to 39 °C, chills and myalgias. Even after the following four days these symptoms persisted and the patient visited an emergency department. She had no chronic disease and her medical history included a tonsillectomy in childhood due to recurrent episodes of tonsillitis and meningitis of unknown etiology at the age of ten.

The clinical examination upon admittance was unremarkable other than a temperature of 37.6 °C. Initial laboratory investigations revealed raised liver enzymes and a C-reactive protein (CRP) level of 16 mg/l (normal value < 5 mg/l). Full blood count, serum electrolytes, and creatinine were within normal limits. Chest and abdominal ultrasound showed no significant abnormalities except mild splenomegaly. Malaria blood microscopical and immunochromatographic tests, Widal test, Dengue and Zika serology tests were all negative.

Empirical treatment with doxycycline 100 mg twice daily was initiated after microbiological sampling. However, three blood cultures taken on admission yielded *Salmonella enterica* and therefore the therapy was switched to ceftriaxone 2 g once daily intravenously. Urine and stool cultures were negative for bacterial pathogens. Fever spikes persisted over the next two days on therapy. Therefore, dosing of ceftriaxone was increased to 2 g twice daily intravenously, and metronidazole 500 mg three times daily was added. A repeat abdominal ultrasound test and CT (computed tomography) radiograph revealed no abscess formation.

All symptoms of the patient subsequently resolved and from the 9th day of hospitalisation she remained afebrile. The therapy was switched to co-ampicillin and metronidazole and on the 13th day of hospitalisation the patient was discharged. CRP and liver function tests normalized after 14 days and completion of a 2-week course of antibiotic treatment. After her final check-up in the clinic, she was clinically well and had made a full recovery.

Material and Methods

The initial blood culture isolate of the patient was grown in hemocultivation bottles using the Bact/Alert 3D platform (bioMérieux, Durham, North Carolina, USA) and it was preliminarily identified as *Salmonella* sp. using mass spectrometer IVD MALDI Biotyper/System/MicroflexTM LT/SH System (Bruker Daltonic GmbH,

Bremen, Germany). The identification scores of this technique measured in duplicate were 2.09 and 2.25, respectively. Nevertheless, the strain was sent to the National Reference Laboratory for *Salmonella* of the National Institute of Public Health (NIPH) in Prague for further analysis.

In order to proceed with the ONT nanopore sequencing, the DNA was isolated and purified from the blood culture using High Pure PCR Template Preparation Kit (Roche). Obtained DNA was quantified on the Qubit fluorometer (Invitrogene) and used for sequencing library preparation (SQK-LSK109, ONT). The sequencing itself was performed on the ONT GridIONx5 platform using the R9.4.1 chemistry (Flow Cell). The so called “base-calling”, i.e., the conversion from raw electric-current signal measured by the device to biologically meaningful bases, was performed in real time using the ONT GridION sequencing software version 19.12.2.

Acquired reads were subject to the read classification pipeline, i.e., each single read was compared to a database in order to estimate its taxonomic origin (based upon sequence similarity criteria). The database was built using the NCBI (National Center for Biotechnology Information) archives from Archaea, Bacteria and Viral species with project status “complete genome” in June 2019 and contains over 9,000 unique bacterial genomes/isolates. We used the freely available software Centrifuge for this purpose (Kim et al., 2016), whose outputs were further analyzed and re-interpreted using custom R scripts (R Core Team, 2019), which take advantage of the R packages taxize (Chamberlain and Szöcs, 2013) and data-table (Dowle and Srinivasan, 2019). The classification results were further analysed using the Krona interactive tool (Ondov et al., 2011).

Since this methodology is based upon analysis of “erroneous” data (Kolmogorov et al., 2019) and thus might be easily doubted, despite the fact that we try to account for this type of error, we use another mostly independent methodology to confirm our results. We thus assembled the sequence data into a draft genomic assembly, which technically results in a consensus sequence, which is at least by an order of magnitude more accurate. This data sequence has been compared to custom BLAST (Zhang et al., 2000) database. This database was built from 1,498 genomic sequences that were downloaded from the NCBI “nucore” database on 2.2.2020. Included were only those sequences that fulfilled all the following conditions: 1) contained the term “*Salmonella enterica*” within the organism name; 2) contained the term “complete genome” within their description (title); and 3) their length was between 3.5 to 10 million base pairs.

Furthermore, in order to obtain some insights about the isolated strain, i.e., some insights about its topological properties or stability (e.g., deletions, duplications, insertions...), we compared our isolate to the reference strain *Salmonella enterica* subsp. *enterica* serovar Paratyphi A str. ATCC 11511 (GenBank accession number NZ_CP019185.1). We mapped all the reads against the reference using the Minimap2 ver. 2.12 (Li, 2018). Some necessary operations, e.g., file format conversions, were conducted with SAMtools ver. 1.9 (Li et al., 2009). Structural

variants (SV) themselves were estimated using the Sniffles ver. 1.0.11 variant-calling tool (Sedlazeck et al., 2018). Inferred SV were visually inspected, i.e., we inspected aligned reads to the reference in the pre-defined presumably SV regions using the CLC Genomics workbench tool ver. 11.0.1 (Qiagen). Genes, or sequences with meaningful open reading frames present within predicted SV regions in the reference strain were further annotated against the NCBI “nr” (protein) database using the Blast2GO tool (Götz et al., 2008), only genes that were not annotated as “hypothetical protein” or “putative ...” were taken into account.

Results and Discussion

In total we acquired 127,161 reads that passed the quality threshold criteria estimated by the base-calling software, i.e., primarily the mean per base quality of a read (Phred quality score > 7). The mean read length equalled 1,527 base pairs (bp) while the longest one achieved 43,235 bp.

Classification

The reads were attempted to be classified with the Centrifuge and acquired classification results were further re-analyzed using three different stringency thresholds. Unlike, e.g., BLAST (Altschul et al., 1990), the Centrifuge seeks only so called “exact match” against the database records and extends it as far as possible within given read (due to computational feasibility of such a demanding task). In result, Centrifuge reports for each read inferred classification to given taxonomic group and length of the match (in base pairs) upon which the estimate was based. Logically, the longer the match, the higher the credibility of the result. Therefore, we re-evaluated the classification results with three thresholds concerning the match length: 23 bp which is the default setting, 50 bp and 75 bp; the latter values were arbitrarily chosen based upon our experience.

We successfully classified 88.93% of reads, which represents quite satisfactory results given the presumable bacterial origin of the sample (DNA). To achieve a 100% rate of classified reads is illusory given the technological, both bioinformatic and the sequencing technology, limitations. While the ONT technology possesses many factual benefits, most importantly the read length and real-time sequencing, still it has some disadvantages compared to the so called “Next Generation Sequencing” (Illumina technology), especially higher error rates (Kolmogorov et al., 2019). This is also another reason why to evaluate the data more strictly than Illumina data.

Therefore, when we applied more stringent criteria as described above, we classified 71.71% and 60.06% of reads, respectively. It is obvious that the numbers logically decrease, nevertheless, these numbers make more sense when we look at individual species (Table 1), i.e., reads classified to the rank of species or higher rank, e.g., subspecies or strain, that were converted to the desired taxonomic rank species, or ignored (e.g., reads classified only to taxonomic rank genus or reads without any classification).

Table 1 – Summary of classification results at the taxonomic rank species

taxID	Species name	Stringency			Drop (%)	Super-kingdom
		1	2	3		
28901	<i>Salmonella enterica</i>	103764	88053	74974	27.75	Bacteria
54736	<i>Salmonella bongori</i>	507	145	58	88.56	Bacteria
562	<i>Escherichia coli</i>	324	180	87	73.15	Bacteria
573	<i>Klebsiella pneumoniae</i>	74	40	26	64.86	Bacteria
545	<i>Citrobacter koseri</i>	55	13	3	94.55	Bacteria
546	<i>Citrobacter freundii</i>	48	12	4	91.67	Bacteria
208962	<i>Escherichia albertii</i>	47	13	2	95.74	Bacteria
550	<i>Enterobacter cloacae</i>	41	10	1	97.56	Bacteria
67825	<i>Citrobacter rodentium</i>	33	8	1	96.97	Bacteria
35703	<i>Citrobacter amalonaticus</i>	32	7	1	96.88	Bacteria
548	<i>Klebsiella aerogenes</i>	32	7	3	90.62	Bacteria
622	<i>Shigella dysenteriae</i>	27	9	5	81.48	Bacteria
61647	<i>Pluralibacter gergoviae</i>	27	4	1	96.30	Bacteria

Per each row we provide: the species taxID; its scientific name; the number of classified reads according to three different Stringency thresholds (details provided in text); the drop of successfully classified reads between the Stringency1 and Stringency3 in percent according to the given species; and the superkingdom of a given species. Only 13 most abundant “putative” species within the sample are presented

A high rate of dropout of classified reads, in a response to more stringent thresholds, belonging to individual species represents an important hint, suggesting that the signal pointing to such species (at the lowest threshold value) represents “noise” in the analysis, i.e., false positives, rather than true positive evidence.

So, while our bioinformatic pipeline clearly pointed out all the positives as well as the negatives of the Centrifuge tool, which indeed represents the core of this pipeline and was not primarily designed to deal with the ONT data, we came to the conclusion that we have identified *Salmonella enterica* as the prevalent DNA donor within our analysis.

Furthermore, based on the Centrifuge classification results, which encompass roughly 4%, i.e., 3,109 individual reads considering the highest threshold value (i.e., exact match of 75 bp), of the classified reads, we were able to classify the bacterial serovar to *Salmonella* Paratyphi A (Figure 1). The reason why only such a sub-portion of reads was able to point directly to this serovar (subspecies) lies in the fact that most of the bacterial species share a set of common genes (Bochkareva et al., 2018), while only some of the genes or even noncoding sequences (features) are able to distinguish between the individual serovar (or even strain).

The data were sufficient for reasonable genome assembly, but they were insufficient for chromosome-level assembly. Therefore, we chose the two longest contigs, 328,615 bp and 233,901 bp, and compared them to the database of all *Salmonella enterica* species (described in Material and Methods). Both contigs

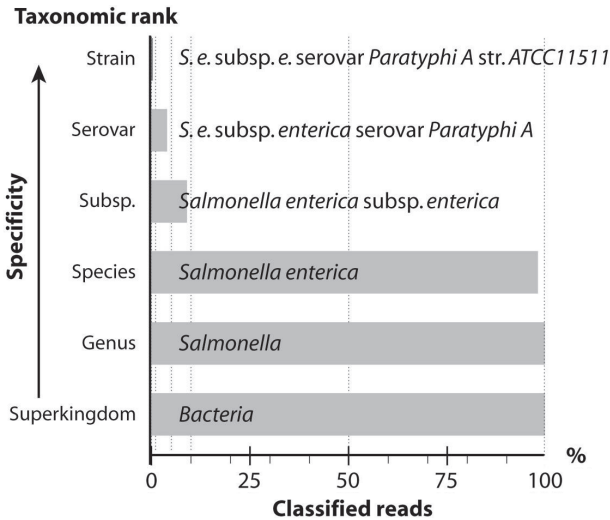


Figure 1 – Barplot shows the proportion of successfully classified reads according to given taxonomic rank. Results of the Centrifuge classification pipeline with the stringency threshold of exact match of 75 bp (i.e., results comparable with the Stringency3 in Table 1; explained in the text); 100% equals 76,367 classified reads. The dotted lines highlight 1%, 5%, 10%, 50% and 100%, respectively.

show highest similarity to *Salmonella* Paratyphi A, which is in concordance with the previous results. However, this methodology has not enough power to distinguish between individual strains, which is in agreement with the fact that this tool was not actually meant as a classification tool, even despite it is often being used in this manner.

Both, these results are in agreement with the NIPH certified laboratory results. The serovar *Salmonella* Paratyphi A was confirmed by biochemical tests and agglutination tests. According to the exact strain identification we used the results provided by the Centrifuge, which classified roughly 0.3% of reads, i.e., 260, to *Salmonella* Paratyphi A str. ATCC 11511. Despite the fact these numbers seem to be low, they represent a reasonable result. Nevertheless, the exact confirmation of such a deep classification could only be attained by a combination of both ONT and Illumina sequencing, and a set of bioinformatic analyses (i.e., phylogenomics and comparative genomics).

Exploratory genomic analysis

Beyond the classification of the disease causative pathogen, we were also able to provide an exploratory analysis of the strain characteristics, i.e., to test whether it possesses some significant structural variation (SV) compared to the reference strain of *Salmonella* Paratyphi A.

Quite surprisingly, we have identified two major deletion events, one encompasses region 1,848 bp long and comprises 5 well annotated genes (protein sequences) while the other one encompasses a region roughly 43,746 bp long which in the reference strain harbours 40 well annotated genes (according to criteria described in Material and Methods).

Despite the fact that bacterial strains are, with respect to genomic evolution and gene content, probably much more plastic (Bochkareva et al., 2018) than for example vertebrates; compare loss of 45 most likely functional genes with the massive impact on the phenotype and even viability caused by simple variation in the copy number of certain blocks of genes, i.e., chromosomal aneuploidy (Hassold et al., 1996).

Still, it is questionable whether such significant loss of part of a likely ancestral genome might be neglected as an example of the bacterial plasticity and variability, or whether it would rather deserve careful description and status of evolutionary independent and novel entity (Simpson, 1951). Nevertheless, these and other questions fall beyond the scope of this study and would require further research before they could be seriously answered.

Conclusion

Oxford Nanopore Technologies sequencing provides a new and easy-to-use tool for both rapid and unambiguous identification of causative agents of infectious diseases. In this study, we successfully demonstrate application of this new technology for the purposes of identification of the invasive pathogenic organism. The initial bacteria culture was cultivated from the patient blood sample. Subsequently, we sequenced the DNA content of the culture using the Nanopore (ONT) sequencing platform. Then, we analysed the data by two methodologically independent bioinformatic pipelines, which allowed us to conclude that the disease causative agent was *Salmonella* Paratyphi A. These results were in a perfect agreement with the classification obtained for this isolate from the Czech National Reference Laboratory for *Salmonella* in Prague.

In comparison to the mass spectrometry identification technique (mentioned in Material and Methods), which might be considered a standard method nowadays, nanopore sequencing is utterly different. Mass spectrometry identification based on proteins analysis possesses few drawbacks in its limits, e.g., the impossibility to detect synonymous mutations, deletions etc., which might be essential for precise identification. Nanopore sequencing overcomes these limits by working directly with bare DNA of causative agents.

In addition, this technique allows not only for the identification of the pathogen, but it also allows the exploration of its genomic architecture. We demonstrate this point of view as well as its importance.

References

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215(3)**, 403–410.
- Bochkareva, O. O., Moroz, E. V., Davydov, I. I., Gelfand, M. S. (2018) Genome rearrangements and selection in multi-chromosome bacteria *Burkholderia* spp. *BMC Genomics* **19(1)**, 1–17.
- Chamberlain, S. A., Szöcs, E. (2013) Taxize: Taxonomic search and retrieval in R. *F1000Res.* **2**, 191.

- Dowle, M., Srinivasan A. (2019) *Extension of “data.frame” (R package data.table version 1.12.8)*. Available at: <https://CRAN.R-project.org/package=data.table>
- Fuselli, S., Baptista, R. P., Panziera, A., Magi, A., Guglielmi, S., Tonin, R., Benazzo, A., Bauzer, L. G., Mazzoni, C. J., Bertorelle, G. (2018) A new hybrid approach for MHC genotyping: High-throughput NGS and long read MinION nanopore sequencing, with application to the non-model vertebrate Alpine chamois (*Rupicapra rupicapra*). *Heredity* **121(4)**, 293–303.
- Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., Conesa, A. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36(10)**, 3420–3435.
- Hassold, T., Abruzzo, M., Adkins, K., Griffin, D., Merrill, M., Millie, E., Saker, D., Shen, J., Zaragoza, M. (1996) Human aneuploidy: Incidence, origin, and etiology. *Environ. Mol. Mutagen.* **28(3)**, 167–175.
- Kim, D., Song, L., Breitwieser, F. P., Salzberg, S. L. (2016) Centrifuge: Rapid and sensitive classification of metagenomic sequences. *Genome Res.* **26(12)**, 1721–1729.
- Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P. A. (2019) Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37(5)**, 540–546.
- Li, H. (2018) Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34(18)**, 3094–3100.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25(16)**, 2078–2079.
- Magi, A., Semeraro, R., Mingrino, A., Giusti, B., D’Aurizio, R. (2018) Nanopore sequencing data analysis: State of the art, applications and challenges. *Brief. Bioinform.* **19(6)**, 1256–1272.
- Maskey, A. P., Day, J. N., Phung, Q. T., Thwaites, G. E., Campbell, J. I., Zimmerman, M., Farrar, J. J., Basnyat, B. (2006) *Salmonella enterica* serovar Paratyphi A and *S. enterica* serovar Typhi cause indistinguishable clinical syndromes in Kathmandu, Nepal. *Clin. Infect. Dis.* **42(9)**, 1247–1253.
- Matono, T., Kutsuna, S., Kato, Y., Katanami, Y., Yamamoto, K., Takeshita, N., Hayakawa, K., Kanagawa, S., Kaku, M., Ohmagari, N. (2017) Role of classic signs as diagnostic predictors for enteric fever among returned travellers: Relative bradycardia and eosinopenia. *PLoS One* **12(6)**, e0179814.
- Norris, A. L., Workman, R. E., Fan, Y., Eshleman, J. R., Timp, W. (2016) Nanopore sequencing detects structural variants in cancer. *Cancer Biol. Ther.* **17(3)**, 246–253.
- Olopoenia, L. A. (2000) Classic methods revisited: Widal agglutination test – 100 years later: Still plagued by controversy. *Postgrad. Med. J.* **76(892)**, 80–84.
- Ondov, B. D., Bergman, N. H., Phillippy, A. M. (2011) Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* **12**, 385.
- R Core Team (2019) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/>
- Schadt, E. E., Turner, S., Kasarskis, A. (2010) A window into third-generation sequencing. *Hum. Mol. Genet.* **19(R2)**, R227–R240.
- Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., Schatz, M. C. (2018) Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **15(6)**, 461–468.
- Simpson, G. G. (1951) The species concept. *Evolution* **5**, 285–298.
- Song, J. H., Cho, H., Park, M. Y., Na, D. S., Moon, H. B., Pai, C. H. (1993) Detection of *Salmonella typhi* in the blood of patients with typhoid fever by polymerase chain reaction. *J. Clin. Microbiol.* **31(6)**, 1439–1443.
- Tennant, S. M., Toema, D., Qamar, F., Iqbal, N., Boyd, M. A., Marshall, J. M., Blackwelder, W. C., Wu, Y., Quadri, F., Khan, A., Azif, F., Ahmad, K., Kalam, A., Asif, E., Qureshi, S., Khan, E., Zaidi, A. K.,

- Levine, M. M. (2015) Detection of typhoidal and paratyphoidal *Salmonella* in blood by real-time polymerase chain reaction. *Clin. Infect. Dis.* **61**, S241–S250 (Suppl. 4).
- Ton, K. N. T., Cree, S. L., Gronert-Sum, S. J., Merriman, T. R., Stamp, L. K., Kennedy, M. A. (2018) Multiplexed nanopore sequencing of HLA-B locus in Māori and Pacific Island samples. *Front. Genet.* **9**, 152.
- Váradi, L., Luo, J. L., Hibbs, D. E., Perry, J. D., Anderson, R. J., Orenga, S., Groundwater, P. W. (2017) Methods for the detection and identification of pathogenic bacteria: past, present, and future. *Chem. Soc. Rev.* **46(16)**, 4818–4832.
- Wain, J., Diep, T. S., Ho, V. A., Walsh, A. M., Hoa, N. T. T., Parry, C. M., White, N. J. (1998) Quantitation of bacteria in blood of typhoid fever patients and relationship between counts and clinical features, transmissibility, and antibiotic resistance. *J. Clin. Microbiol.* **36(6)**, 1683–1687.
- Watsa, M., Erkenwick, G. A., Pomerantz, A., Prost, S. (2019) Genomics in the jungle: Using portable sequencing as a teaching tool in field courses. *bioRxiv*.
- Zhang, Z., Schwartz, S., Wagner, L., Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7(1–2)**, 203–214.