

Application of a strain-level shotgun metagenomics approach on food samples: resolution of the source of a *Salmonella* food-borne outbreak

Florence E. Buytaers^{1,2}, Assia Saltykova^{1,2}, Wesley Mattheus³, Bavo Verhaegen⁴, Nancy H. C. Roosens¹, Kevin Vanneste¹, Valeska Laisnez⁵, Naïma Hammami⁵, Brigitte Pochet⁶, Vera Cantaert⁶, Kathleen Marchal^{2,7,8}, Sarah Denayer^{4†} and Sigrid C.J. De Keersmaecker^{1*,†}

Abstract

Food-borne outbreak investigation currently relies on the time-consuming and challenging bacterial isolation from food, to be able to link food-derived strains to more easily obtained isolates from infected people. When no food isolate can be obtained, the source of the outbreak cannot be unambiguously determined. Shotgun metagenomics approaches applied to the food samples could circumvent this need for isolation from the suspected source, but require downstream strain-level data analysis to be able to accurately link to the human isolate. Until now, this approach has not yet been applied outside research settings to analyse real food-borne outbreak samples. In September 2019, a *Salmonella* outbreak occurred in a hotel school in Bruges, Belgium, affecting over 200 students and teachers. Following standard procedures, the Belgian National Reference Center for human salmonellosis and the National Reference Laboratory for *Salmonella* in food and feed used conventional analysis based on isolation, serotyping and MLVA (multilocus variable number tandem repeat analysis) comparison, followed by whole-genome sequencing, to confirm the source of the contamination over 2 weeks after receipt of the sample, which was freshly prepared tartar sauce in a meal cooked at the school. Our team used this outbreak as a case study to deliver a proof of concept for a short-read strain-level shotgun metagenomics approach for source tracking. We received two suspect food samples: the full meal and some freshly made tartar sauce served with this meal, requiring the use of raw eggs. After analysis, we could prove, without isolation, that *Salmonella* was present in both samples, and we obtained an inferred genome of a *Salmonella enterica* subsp. *enterica* serovar Enteritidis that could be linked back to the human isolates of the outbreak in a phylogenetic tree. These metagenomics-derived outbreak strains were separated from sporadic cases as well as from another outbreak circulating in Europe at the same time period. This is, to our knowledge, the first *Salmonella* food-borne outbreak investigation uniquely linking the food source using a metagenomics approach and this in a fast time frame.

Received 01 February 2021; Accepted 23 February 2021; Published 07 April 2021

Author affiliations: ¹Transversal Activities in Applied Genomics, Sciensano, Brussels, Belgium; ²Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium; ³National Reference Center for *Salmonella* and *Shigella* spp., Sciensano, Brussels, Belgium; ⁴National Reference Laboratory for *Salmonella* and Food-Borne Infections, Food-Borne Pathogens, Sciensano, Brussels, Belgium; ⁵Agentschap Zorg en Gezondheid, Brussels, Belgium; ⁶Federal Agency for the Security of the Food Chain, Brussels, Belgium; ⁷Department of Information Technology, IDlab, IMEC, Ghent University, Ghent, Belgium; ⁸Department of Genetics, University of Pretoria, Pretoria, South Africa.

***Correspondence:** Sigrid C.J. De Keersmaecker, sigrid.dekeersmaecker@sciensano.be

Keywords: food surveillance; metagenomics; outbreak; *Salmonella*; SNP analysis; strain-level.

Abbreviations: AMR, antimicrobial resistance; ECDC, European Centre for Disease Prevention and Control; EFSA, European Food Safety Authority; EPIS, Epidemic Intelligence Information System; MLST, multilocus sequence typing; MLVA, multilocus variable number tandem repeat analysis; NCBI, National Center for Biotechnology Information; NRC, National Reference Center; NRL, National Reference Laboratory; qPCR, real-time PCR; RASFF, Rapid Alert System for Food and Feed; WGS, whole-genome sequencing.

The NCBI accession numbers for the new sequence data presented in this paper are SAMN15963373–SAMN15963404, SAMN15957185 and SAMN15957186 (see Table S1).

†These authors contributed equally to this work

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Seven supplementary tables are available with the online version of this article.

000547 © 2021 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution NonCommercial License.

DATA SUMMARY

The *Salmonella enterica* subsp. *enterica* serovar Enteritidis isolate sequencing data is described in the supplementary data.

INTRODUCTION

The detection and characterization of pathogens in food aims at avoiding contamination of consumers if carried out as a continuous screening, but also at putting an end to epidemics when consumers have already been infected. According to European Union legislation, typically the analysis of a suspect food sample involved in a food-borne outbreak includes an attempt at obtaining an isolate of the micro-organism, most often by the official control laboratories, such as the National Reference Laboratory (NRL), to further characterize it, e.g. by real-time PCR (qPCR) or whole-genome sequencing (WGS) [1–3]. To unambiguously identify the source of the outbreak, the food contaminant also has to be uniquely linked to the pathogens usually obtained from human cases by the National Reference Center (NRC). This strengthens the assumption on the food source based on epidemiological studies only. However, isolation from food samples is not straightforward nor always successful, as opposed to the human samples, which typically contain higher loads of the pathogen. In these cases, the relatedness to the human isolates cannot be obtained and the outbreak is never resolved to its food source. Indeed, the European Food Safety Authority (EFSA) reported that the causative agent was unknown in 23.8% of outbreaks that occurred in 2018 [4, 5]. In some cases, the wrong foodstuff can even be blamed, leading to huge economic losses in the sector [6]. A novel approach, i.e. shotgun metagenomics, has been investigated in recent years in an attempt to characterize the pathogen but without the need to isolate it from the food matrix [7–10]; therefore, in a possibly shorter time frame and, most importantly, increasing the chance of finding the source of the outbreak. EFSA recently published an opinion on the use of WGS and metagenomics for outbreak investigation, confirming the possibility for typing and source attribution from shotgun metagenomics data, in particular if a draft reconstructed genome of the pathogen at the strain-level can be obtained [11]. Until now, only a few studies have investigated the possibility of achieving strain-level characterization for pathogens in food samples; however, these did not link strains obtained from the food samples to isolates from the human cases, a prerequisite for the trace back of the outbreak [12–15]. We have previously developed such a metagenomics approach to be implemented for food-borne outbreak investigations [16, 17] using artificially contaminated samples, targeting the Shiga toxin-producing *Escherichia coli* (STEC), and we were able to link it back to isolates from humans. This method has, however, not yet been implemented for another pathogen or during a real outbreak.

Among food-borne outbreaks occurring in Europe, food contaminations due to *Salmonella* are the second most commonly reported cause of gastrointestinal infections [4]. Salmonellosis are caused by thousands of different serovars, of

Impact Statement

Shotgun metagenomics sequencing is still a relatively new approach, which until now had not yet been applied on food samples in real time to resolve a food-borne outbreak to its food source. This work presents as a case study a *Salmonella* outbreak in a hotel school in Belgium that was analysed with a metagenomics workflow on food in parallel to the conventional outbreak investigation. This allowed us to relate the strains present in the food, analysed through shotgun metagenomics, with isolates from human cases, in a time frame theoretically shorter by at least 1 week than the results from conventional methods. As this is, to the best of our knowledge, the first study presenting the successful use of shotgun metagenomics for the study of contaminated food in a food-borne outbreak investigation, we believe that it will have an important impact on the public health and research community and on the trust in this technology as a reliable, faster and cost-effective alternative. This study also provides a valuable dataset to further explore metagenomic data analysis tools.

which *Salmonella enterica* subsp. *enterica* serovar Enteritidis accounts for over 40% of all infections for which the serovar has been identified. They are most often related to eggs and have been associated with a high proportion of food-borne outbreaks, due to the use of the raw product in several food preparations [4]. The standard protocol for analysing food products potentially contaminated with *Salmonella* according to European Union legislation is to isolate the pathogen through several enrichment and plating steps (ISO 6579:2017 [18]). The isolated strain is then characterized through biochemical and/or serological testing, as well as multilocus variable number tandem repeat analysis (MLVA) to infer phylogeny against a well-characterized background. However, EFSA has now recommended WGS of *Salmonella* isolates, particularly when linked to outbreaks [19]. WGS offers the possibility to study the full genome of the isolate, including potential virulence and antimicrobial-resistance (AMR) genes [20, 21]. It also allows the highest level of precision in relatedness studies based on SNP differences between strains, and allows sporadic bacteria to be distinguished from persistent bacteria in a food-production environment [22, 23]. Using metagenomics, *Salmonella* has thus far only been characterized in faeces [24, 25] or in food after selective concentration of *Salmonella* genomic DNA by immunomagnetic separation [26]. However, food samples contaminated with this species have not yet been tested with an open metagenomics approach in the scope of a real outbreak.

From September 5th 2019 until September 14th 2019, over 200 students and teachers at a hotel and tourism school in Belgium suffered from food poisoning, with symptoms such as abdominal pain, headache, diarrhoea and fever [27, 28].

The outbreak was thoroughly investigated by the local authorities [regional health agency Zorg en Gezondheid, the Federal Agency for the Security of the Food Chain (FASFC) and the NRL (food and feed) and NRC (human)]. Laboratory analyses were conducted on 65 samples obtained from food leftovers and kitchen surfaces, as well as isolates from infected patients. This resulted in the identification of the contamination as being *S. enterica* subsp. *enterica* serovar Enteritidis, found in a meal prepared on September 5th 2019 by students and served in the school restaurant. The meal consisted of fish sticks with mashed potatoes and freshly made tartar sauce. After WGS of isolates from food and human origins, the source of the contamination was established as being the sauce, prepared with raw eggs [27–29]. A rare MLVA profile, i.e. 3-12-5-5-1, was determined for the human and food isolates by the NRC [28]. After disinfection of the kitchen and kitchen equipment, *Salmonella* was not detected anymore in environmental samples and no new cases were recorded. The outbreak was reported through the European Epidemic Intelligence Information System (EPIS) ('Urgent Inquiry' UI-608) and the Rapid Alert System for Food and Feed (RASFF, 2020.3675) and allowed the tracing of this outbreak back to an egg-producing farm in Spain, considered as the source of the contamination [27, 28]. At the same time period (ongoing since 2016), another outbreak was circulating in Europe and was linked to eggs of Polish origin. However, this strain of *S. enterica* subsp. *enterica* serovar Enteritidis was distinct from the isolates from the hotel-school outbreak and was characterized with MLVA profiles 2-9-7-3-2, 2-9-6-3-2, 2-9-10-3-2, 2-10-6-3-2, 2-10-8-3-2 or 2-11-8-3-2 [30, 31].

As this was an ideal case study to apply our previously developed strain-level metagenomics approach on contaminated food samples to be used during a food-borne outbreak, we received from the Belgian NRL, in parallel to the conventional investigation, two samples that were positive for *S. enterica* Enteritidis and linked to the hotel-school outbreak. Both samples were processed with a metagenomics workflow described previously [17]. After short-read sequencing, we conducted data analysis in order to infer the pathogenic strain's genome, characterize it and link it back to the human isolates to resolve the outbreak. The food strain obtained from metagenomics reads was included in a SNP-level phylogenetic tree containing human and food isolates from the hotel-school outbreak, as well as strains related to another outbreak circulating in Europe during the same time period [30, 31] and other sporadic strains that occurred in Belgium in 2019. The time of analysis of such a shotgun metagenomics approach was then compared to the time necessary to elucidate this outbreak with food isolates' data.

METHODS

Sample preparation

Two aliquots of cultured food samples (i.e. a mixture of the meal components and the sauce as a separate component) linked to the outbreak were received from the NRL after a first non-selective enrichment according to ISO 6579 [18]

(i.e. 25 g foodstuff was mixed with 225 ml buffered peptone water and incubated for 18±2 h at 37±1 °C). The sample dish was an aluminium tray with three compartments, one for each component (mashed potatoes, fish stick, tartar sauce). The tartar sauce was tested separately as well, after confirmation that it was the probable source of the contamination. The food enrichments had been tested for the presence of *Salmonella* prior to their selection for this study, using the iQ-Check *Salmonella* II PCR detection kit (Bio-Rad) according to the manufacturer's instructions, and showed positive results (Cq of 18 and 17, respectively) as opposed to the blanks and to other samples collected in the school during the investigation. Aliquots of 4–15 ml of the two cultured food samples were stored in the fridge until metagenomics DNA extraction was carried out.

DNA extraction and qPCR

The sample preparation was carried out according to Buytaers *et al.* [17]. Briefly, 1 ml of the aliquots was centrifuged at 6000 g for 10 min and the cell pellets were used for DNA extraction using a Nucleospin food kit (Macherey-Nagel). In order to confirm the presence of the contaminant (*Salmonella*) in the DNA extracts, a qPCR was performed for the genes *invA* and *rpoD*, according to Barbau-Piednoir *et al.* [32].

Shotgun metagenomics sequencing

The quality and quantity of all DNA extracts were evaluated [17] using the NanoDrop 2000 (Thermo Fisher Scientific), Qubit 3.0 fluorometer (Thermo Fisher Scientific) and 4200 TapeStation (Agilent). All DNA extracts were further processed using the Nextera XT library preparation kit (Illumina) before sequencing on the Illumina MiSeq, generating paired-end 250 bp reads with the reagent kit v3. The samples were sequenced in one run of eight libraries. The number of (paired-end) reads sequenced per metagenomics sample is presented in Table 1. Sequencing metrics were obtained using FastQC version 0.11.7 [33].

Isolate data

Sequencing data from *S. enterica* subsp. *enterica* serovar Enteritidis isolates (see Table S1, available with the online version of this article) included data from five isolates of the hotel-school outbreak from food origin (the leftover meal and the three components of this meal that were all probably contaminated through spreading of the sauce between the compartments, and a chicken-based meal consumed on September 24th 2019 at the hotel school that was probably contaminated in the rubbish bin) and from five isolates from human origin linked to the hotel-school outbreak, obtained following conventional methods [34]. These 10 isolates showed the same MLVA profile. As background for the phylogenetic analysis, data were also included from isolates linked to the still ongoing Polish outbreak [30, 31], presenting distinct MLVA profiles, i.e. seven Belgian isolates from food origin, five Belgian isolates from human origin and four isolates from public databases representing the different outbreak clusters defined by the Public Health England SNP pipeline described in an outbreak

Table 1. Quality metrics of the metagenomics sequencing and metagenomics assemblies

Metric	Sauce	Meal
Sequencing metrics		
Total reads	2653700	4857796
Sequences flagged as poor quality	0	0
Sequence length	35–251	35–251
G+C (mol%)	49	47
Mean quality score	35.83	36.1
Median quality score	30	31
Strain assembly metrics*		
No. of contigs	78	75
Largest contig	325096	325086
Total length	4703829	4704090
G+C (mol%)	52.13	52.13
N50	106626	12874
Mean coverage	93.9	88.35
Median coverage	73.5	65.5

*Statistics based on contigs of size ≥ 500 bp.

assessment from the European Centre for Disease Prevention and Control (ECDC) and EFSA [31], supplemented with ten isolates of human origin from Belgian sporadic cases from 2019, also presenting a different MLVA profile to the one of the hotel-school outbreak.

Data analysis

The metagenomics sequencing data were analysed through the workflow presented by Buytaers *et al.* [17]: after trimming, a taxonomic classification of all reads to the genus level was performed using Kraken2 [35] (same databases as previously described [17]) in order to obtain an overview of the taxa present in the sample. The taxonomic classification results from Kraken2 [35] were verified using the online tools PathogenFinder (designed for isolate WGS) [36] using the model created for all bacteria, as well as CCMetagen [37] used with the National Center for Biotechnology Information (NCBI) nucleotide database. Then, a strain-level read classification was performed using Sigma [38] on a database of 787 complete genome assemblies of *Salmonella* (all serovars) from NCBI (list available upon request), using the default parameters as described by Saltykova *et al.* [16] to obtain the reads of the pathogenic strain, as *Salmonella* was the only pathogen detected after analysis of the taxonomic classification results. These reads as well as the sequencing reads from all isolates were assembled using SPAdes 3.13.0 [39]. Quality metrics from the assemblies (Table 1) were obtained using QUAST version 5.0.2 [40]. All assemblies from isolates and metagenomics samples were then typed (serovar prediction)

using the online *Salmonella In Silico* Typing Resource (SISTR) [41] and the presence of AMR genes was detected using BLAST 2.6.0 on the ResFinder database [42], with a minimum identity threshold of 90% and a minimum length of 60% for metagenomics assemblies, and 90% minimum identity and minimum length for isolate assemblies [43]. The parameters were lowered for the metagenomics assemblies compared to the parameters (90% gene coverage and 90% nucleotide identity) chosen for the study of isolates, considering the lower depth obtained with metagenomics sequencing. For phylogenetic analysis, SNP calling was carried out on the classified (unassembled) reads as previously described [16], with *S. enterica* subsp. *enterica* serovar Enteritidis strain EC20120200 (*Enterobacteria*) as a reference (GenBank accession no. CP007434.2). Maximum-likelihood substitution model selection and phylogenetic tree inference were done with MEGA [44], using the NNI (nearest-neighbour-interchange) heuristic method, keeping all informative sites and using a bootstrap method with 100 replicates. The model selected to build the phylogenetic tree was that of Tamura and Nei [45]. iTOL [46] was used for the representation of the tree, with the percentage of the reference genome covered annotated on each branch.

RESULTS

Taxonomic classification of the metagenomics samples

Two food samples (meal and sauce component) that could be related to the outbreak after a first screening (culture and qPCR) were tested using a shotgun metagenomics approach in parallel to the conventional outbreak investigation carried out at the NRL. After culture-based enrichment of the food matrices, the DNA was extracted and sequenced. The reads obtained were then taxonomically classified to determine the genera that were present in the food matrices.

Only bacteria could be detected in both samples (89 and 96% of the sequenced reads for the meal and the sauce, respectively), although the meal consisted of fish, mashed potatoes and sauce, and the sauce was made with fresh eggs. This was expected as the latter species (fish, potato and chicken) are not represented in the taxonomic databases used and, therefore, should be part of the unclassified section of the reads (Fig. 1). The same bacterial genera were detected in both matrices albeit at different relative abundances, except for *Streptococcus*, which was only present in the meal sample. The consensus in detected bacterial genera was to be anticipated since the sauce was sampled from the meal. *Salmonella*, the genus implicated in the outbreak, was detected at a high percentage in both matrices (70% in the sauce, 40% in the meal). This is consistent with the qPCR detection of the *Salmonella*-specific *invA* and *rpoD* genes in the DNA extracts of both samples (Table S2). However, other detected genera like *Escherichia*, *Bacillus*, *Klebsiella* or *Streptococcus* may also represent pathogenic species. Therefore, in an attempt to use the taxonomic classification as an agnostic tool to identify the causative food-borne pathogen,

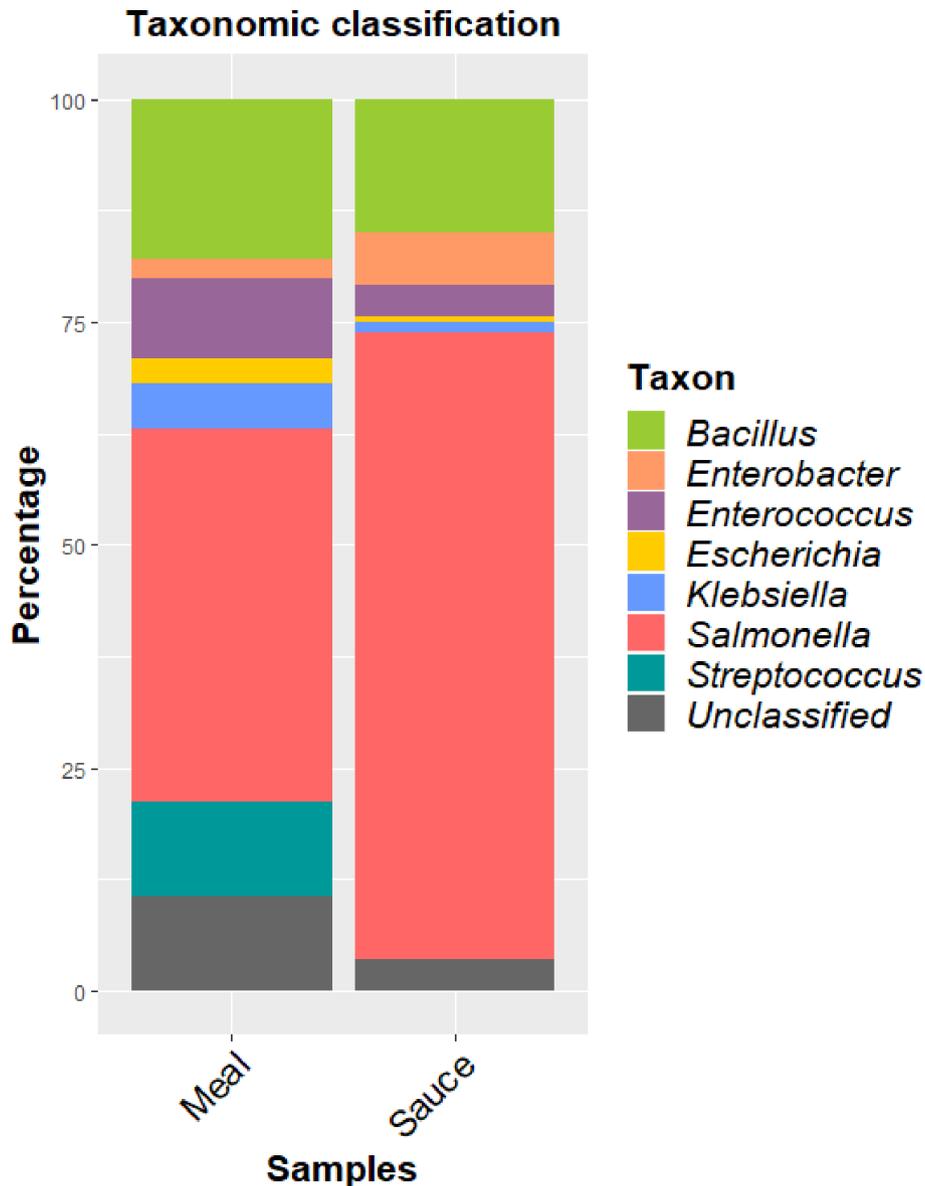


Fig. 1. Percentages of reads classified to the genus level using a taxonomic classification tool (Kraken2) from metagenomics samples (full meal and sauce) with in-house databases of mammals, archaea, bacteria, fungi, human, protozoa and viruses. Red represents the proportion of '*Salmonella*' in the samples. The reads that could not be classified to the genus level for mammals, archaea, bacteria, fungi, human, protozoa or viruses are represented in grey.

two other data analysis tools were used to determine the presence of a pathogen in the sample (CCMetagen and PathogenFinder). CCMetagen and PathogenFinder identified *S. enterica* as the main or only pathogen in the two samples (the results are shown in Table S3) after analysis based on KMA sequence alignments on the NCBI nucleotide database (CCMetagen) or prediction of pathogenicity based on the detection of groups of genes associated with human pathogenic bacteria (PathogenFinder). The output of the three different tools used, based on different bioinformatics approaches, confirmed that *Salmonella* was

considered as the only pathogen meriting further investigation in this study.

***Salmonella* strain inference from metagenomics samples and *in silico* typing**

Obtaining strains from the metagenomics reads is necessary to mimic the recovery and characterization of an isolate with conventional methods. This was done for each metagenomic sample following a previously reported metagenomics strain-level analysis pipeline [16, 17]. After classification of the reads to a database of *Salmonella* genomes, 1843873 and

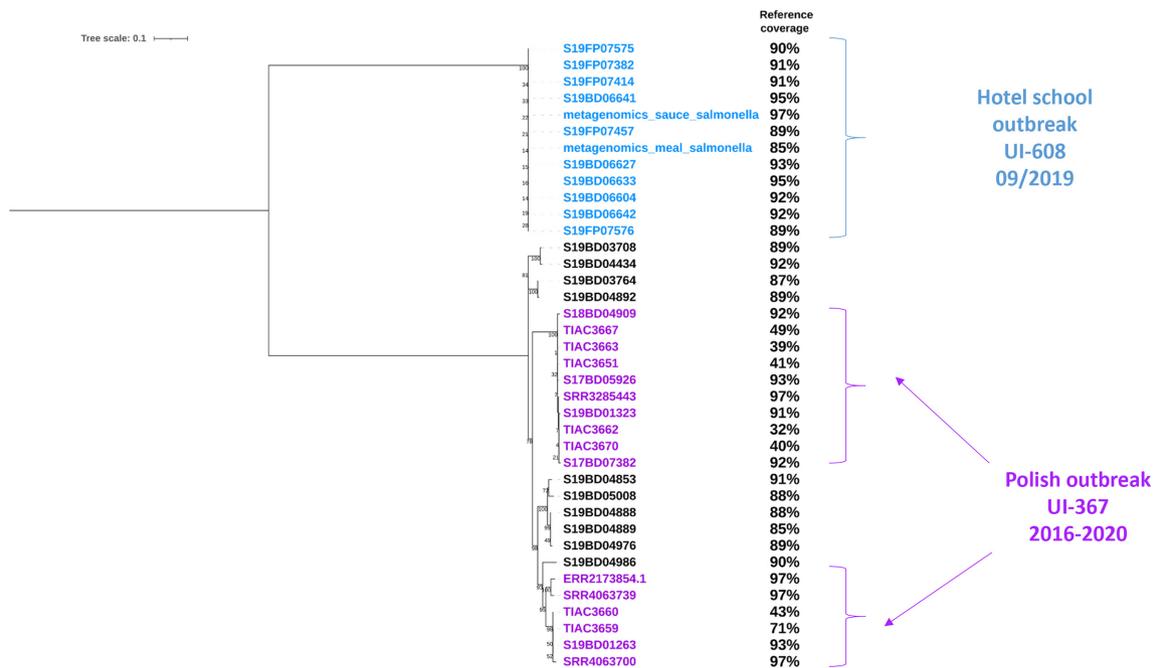


Fig. 2. SNP-based phylogenetic tree representing the isolates and metagenomics-derived strains from food samples linked to the hotel-school outbreak (UI-608, in blue) in the global context of *S. enterica* subsp. *enterica* serovar Enteritidis circulating in Belgium and in Europe during the same time period. Isolates linked to the Polish outbreak (UI-367) are indicated in purple, and isolates from sporadic cases in Belgium in 2019 in black. Percentage of the reference genome covered is presented on the side of each branch. Bar, nucleotide substitutions per 100 nucleotide sites. Node values represent bootstrap support values.

1 618 032 reads were classified as ASM303203v1 [*S. enterica* subsp. *enterica* serovar Enteritidis (enterobacteria), RefSeq accession no. GCF_003032035.1], respectively, for the meal and the sauce (Table S4). This represents 38% of the total sequenced reads for the meal and 61% of the reads for the sauce. Less than 7000 reads (<0.5% of the total reads) were classified to other *Salmonella* genomes for both samples, indicating that most probably only one strain of this species was present in the sample and that the reads assigned to ASM303203v1 correspond to that strain.

Consecutively, a sequence-based characterization can be performed on the reads of each inferred strain, corresponding to the characterization of the isolate with conventional methods. The reads were assembled (Table 1) and then typed *in silico*. The results (Table S5) confirmed that the strains obtained are indeed *S. enterica* subsp. *enterica* serovar Enteritidis, based on O- and H-type prediction (serogroup D1, H1 g, m, H2-), multilocus sequence typing (MLST) clustering (ST11) and matches of their closest public genome. When comparing to the *in silico* typing of sequenced isolates from food and human origin from the outbreak (Table S5), the results were identical except for the detection of all 330 whole-genome MLST alleles in the isolates and 329 identical alleles in the metagenomics-based strains (one allele present partially). Other isolates obtained from the NRC, the NRL and from another outbreak circulating in Europe (not related to the hotel-school outbreak) were typed with the same tool. These were also defined as

S. enterica subsp. *enterica* serovar Enteritidis, but were related to other genomes from public databases (Table S5).

The presence of AMR genes was also investigated in the assembled contigs of the metagenomics-based strains (Table S6), to follow the analysis that is usually performed on isolates (using the technique of microdilutions in broth), but then at the genotype level. The locus *aac(6')-Iaa_1*, linked to resistance to aminoglycoside due to a chromosomally encoded aminoglycoside acetyltransferase, was detected in all strains from the hotel-school outbreak, including strains derived from metagenomics sequencing, as well as all non-outbreak-related strains included in this study with 96.35% identity and 100% coverage (Table S6). The prevalence of this gene in *S. enterica* WGS from NCBI is 29% [47]. No other AMR genes were detected in any strain.

Metagenomics-based trace back investigation of the outbreak to its food source

Finally, in order to relate cases from food and human origins, the MLVA profiles can be compared with traditional methods, but EFSA now recommends WGS of *Salmonella* isolates and uses core-genome MLST in data sharing platforms such as EPIS. In our analysis, all isolates and metagenomics-derived strains were compared using SNP calling and reconstruction of a phylogenetic tree (Fig. 2). SNP calling offers the possibility of comparing the full genome and is considered more suited to use for

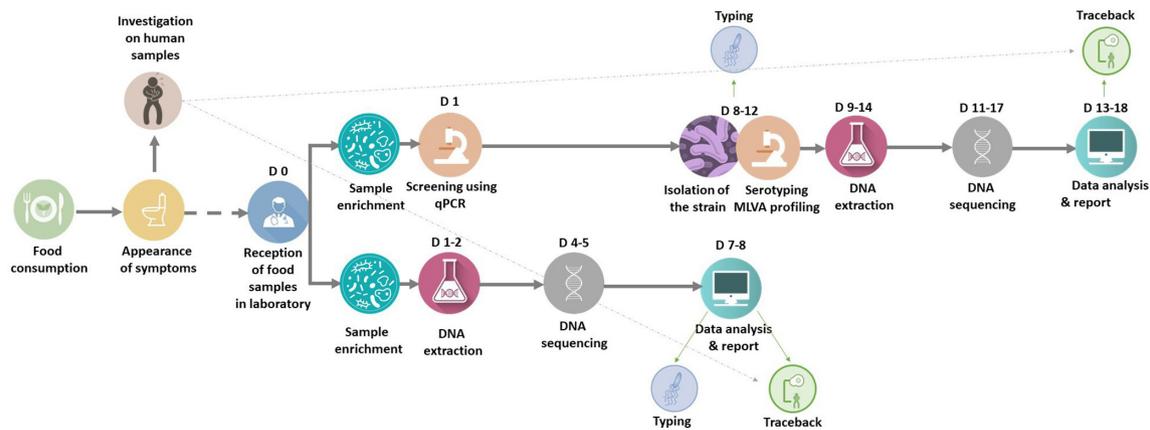


Fig. 3. Comparison of theoretical processing time for the conventional approach (upper level) and the shotgun metagenomics approach (lower level) for *Salmonella*-contaminated food samples from receipt of the samples to strain typing and trace back between human and food strains. A range of days (D x-y) accounts for a range of duration of some laboratory analyses, which can vary due to the presence of technicians during weekends, success in the isolation process or cost-effectiveness (start of the sequencing run with sufficient samples).

metagenomics-derived strains [16]. The cluster corresponding to the hotel-school outbreak (represented in blue in Fig. 2) includes the isolates from patients and suspicious food vehicles obtained by the NRC and NRL, as well as the two inferred strains obtained from direct sequencing of two food samples (suspect meal and sauce) using a shotgun metagenomics approach. The breadth of coverage of the reference genome for the two reconstructed strains from metagenomics samples is 97 and 85% for the sauce and the meal, respectively. These values are in the same range as the values obtained for the isolates of the same outbreak. All strains of the hotel-school outbreak cluster, including the strains from the metagenomics samples, have 0 SNP differences per million genomic positions (Table S7). Other *S. enterica* subsp. *enterica* serovar Enteritidis circulating in Europe at the same time period, including isolates linked to an outbreak of Polish origin that started in 2016 but was still ongoing (shown in purple in Fig. 2), were included in the analysis, and could be separated both from the isolates and the metagenomics strains from the hotel-school outbreak.

Timing for a conventional and a metagenomics-based approach to resolve outbreak investigation to the food source

A schematic representation of the theoretical timeline of the conventional analysis conducted at the NRL on food samples, in parallel to the investigation on human samples conducted at the NRC, is presented in Fig. 3 (upper line). After receipt of the samples, the confirmation of the presence of *Salmonella* in the food is first conducted with qPCR on the food matrices, then normally isolates are obtained after approximately 1 week (if isolates can be produced from the food samples), and characterized for serotype and MLVA profile. Once the MLVA profile is confirmed to be identical to the one detected in the patients' isolates, the DNA of the food isolates is extracted for WGS analysis. At the Belgian NRL, the

serotyping and MLVA profile of the food isolates, if obtained, are currently prerequisites before sequencing, to prove that the strains have a high chance of being linked to the outbreak, as only outbreak cases are eligible for obtaining budget and priority for WGS. Notably, the isolates from human origin are most often already characterized at that stage as they are detected and isolated most often more easily and earlier in the investigation process. Together with library preparation, the sequencing takes approximately 4 days. The sequencing typically occurs 2 to 3 weeks after receipt of the samples depending on the isolation time, the time necessary to gather sufficient isolates to be cost-efficient for multiplexing in a single sequencing run, and to perform the sequencing run. Data analysis is then conducted, followed by sharing of the information, with national and international instances (in this case: RASFF 2019.3675 on October 16th 2019 and EPIS UI-608 updated on October 24th 2019 with the NGS data). In this outbreak, it allowed determination of the source of the contamination as an egg-producing farm in Spain and detection of 13 related human cases from France and 2 human cases in both the Netherlands and the UK [27]. In the same time period, an outbreak was reported in the Netherlands involving eggs originating from Spain (RASFF 2019.3069, UI-601). However, the strains of *S. enterica* Enteritidis had distinct MLVA profiles, 2-11-7-3-2, 3-10-5-4-1, 2-10-7-3-2, 3-11-5-4-1, and 170 core-genome MLST allelic differences from our outbreak strain. The UK also reported an outbreak linked to eggs (RASFF 2019.1412, UI-602), but again no link with the Belgian outbreak strain was established. The WGS data of these strains were not publicly available and, therefore, could not be added to the phylogenetic analysis in this study.

This timeline was compared to that of a metagenomics-based analysis of the food samples. DNA from the meal and the sauce was extracted from a small fraction of the cultured food matrices for subsequent metagenomics analysis after

suspicion of the contamination with qPCR (not necessary for a metagenomics-only workflow). From the time of the DNA extraction, depending on the availability of a sequencing instrument and the preparation of the libraries, the sequenced reads could be obtained in a minimum of 4 days (Fig. 3, lower line). Thereafter, a taxonomic classification was obtained in a few minutes and, after 1 day, a pathogenic strain was obtained and fully typed. In less than a week after receipt of the samples in the laboratory, the pathogen was fully described and related to other cases from the outbreak (from food and human origin) in a phylogenetic tree. This corresponds already to the mean time necessary to only obtain an isolate from food in routine analysis, if obtained, with no information about relatedness of the cases at that stage of the conventional analysis. Indeed, in the conventional analysis, obtaining a food isolate is a prerequisite for performing the molecular analysis, including WGS, to be able to determine relatedness.

DISCUSSION

We deliver in the present study a proof of concept for the shotgun metagenomics approach on food samples previously developed on food samples artificially spiked with STEC (Shiga toxin-producing *E. coli*) [17] to resolve a *Salmonella* outbreak in Belgium up to the food source. We described the analysis of an outbreak that affected over 200 students and teachers at a hotel school in Belgium, using a strain-level shotgun metagenomics-based approach in parallel to the investigation based on WGS of isolates performed by the NRL and NRC. Two suspect samples of leftovers of the meal and the tartar sauce included in this dish were analysed with a shotgun metagenomics workflow, in a relatively very short time frame, and the pathogenic strain was inferred from the sequenced metagenomics reads and characterized as a *S. enterica* subsp. *enterica* serovar Enteritidis that was related with 0 SNP differences to the isolates of human origin from the same outbreak. Therefore, the outbreak could be resolved, i.e. source attribution, using metagenomics data for the food samples. As this was a proof of concept, isolates were also obtained and characterized from the food samples through conventional analysis, and were also related to the metagenomics strains with 0 SNP differences, as a validation of the obtained results. Moreover, the outbreak cluster was placed in a global perspective of the situation of salmonellosis in Belgium and Europe using a phylogenetic tree including other strains circulating at the same time period.

The timing of an outbreak investigation is a critical factor to limit the propagation of the contamination. Shotgun metagenomics is an alternative to the conventional approaches circumventing the need for isolation, which is time-consuming and most importantly not always achievable in routine analysis. This study showed the potential of metagenomics to be used during outbreak investigations on food samples for obtaining the same level of information as from food isolates, in a time frame reduced by over 1 week. Moreover, this constitutes a pathogen-agnostic approach dependent on a non-selective enrichment, which allows the detection of the pathogenic

strains (here *Salmonella*) and the characterization of this contaminant without prior knowledge on the species or the number of different species and/or strains present in the sample [17], in contrast to conventional methods where the assumption of the species to test for is based on the symptoms of the patients. Therefore, this metagenomics approach is also advantageous in case of a limited quantity of food leftovers, because no choice for best fit symptoms-pathogen should be made as for conventional methods. Hence, this approach can potentially increase the range of pathogens detected in a mixed sample, and help reduce even more the economic burden of such food-borne pathogens, as was already stated for WGS of isolates [48]. Our approach still relies on the isolation of the pathogen from the human samples and is not a stand-alone metagenomics approach. As the bacterial load is generally higher in human samples, isolation is not reported as a challenge in these matrices. Moreover, the isolation in the human samples is often not a limiting factor for the timing of food-borne outbreak investigation, as these samples are often obtained before the food samples in the case of outbreaks. Nevertheless, metagenomics studies of stool samples, included during outbreaks, have been published previously [25, 49, 50], and such an approach could be performed in parallel to the one we present, in the corresponding institution (NRC). However, this would represent a higher cost and the sequencing of human DNA might lead to ethical and privacy issues, in particular in Europe.

At a national scale, the typing data of food and human isolates are shared between the NRL and NRC, and matches are reported at the European level, i.e. EFSA and ECDC [27]. No shared database is publicly available at the moment and access to this data or the samples must go through contact between both national entities. Communication concerning human health at the international level for outbreaks in Europe is done through the use of a communication platform and data sharing between public-health experts, by 'Urgent Inquiries' at the EPIS platform. For food safety, communications are done by the competent authorities through the RASFF system. These tools were used in the hotel-school outbreak investigation and helped to trace back and link the outbreak to eggs originating from Spain and other human cases in France, the UK and the Netherlands [27]. However, for confidentiality reasons, these data were not made publicly available and, therefore, could not be included in our presented phylogenetic tree. Our study highlights that access to scientific data, including both raw WGS data and processed results, from public-health and food-safety authorities at both the national and international level will help to strengthen analyses on international outbreaks such as the one presented in this study, and consequently should be considered in the line of data sharing systems that have proven their efficiency.

The shotgun metagenomics approach has proven its potential for outbreak investigation through studies like this one, yet additional research could help with the actual further implementation of this method in routine settings. First, the culture of the food matrix as currently specified in the ISO (International Organization for Standardization) method could

be adapted to suit a larger number of species concurrently for pathogen-agnostic metagenomics studies. Second, the optimal quality-control metrics for metagenomic sequencing have not yet been established, in contrast to ongoing efforts for WGS of isolates (e.g. ISO/DIS 23418 [51]). In the current analysis, eight metagenomic food samples (six were not related to this study) were multiplexed in a single MiSeq run, with a relatively high cost per sample as a result. This allowed achievement of a sequencing depth of >85× for the single detected *Salmonella* strain for both metagenomic samples, which is comparable to values typically achieved for isolates and is more than sufficient for the reconstruction of the pathogen's genome. This indicates that, in the future, sequencing of a higher number of samples simultaneously can be attempted, lowering the cost. The observed coverage is, however, much higher than in our previous work, where multiplexing of 12 minced meat samples resulted in sequencing depths between 0.9× and 10× for detected *E. coli* strain(s) [17]. Leonard *et al.* [12, 13] reported that multiplexing of 12 enriched spinach samples yielded coverages between 5× and 145× for an *E. coli* reference genome, with 4 samples having coverages less than 30×. Therefore, the minimal required sequencing depth will likely differ for each sample type, and will depend on biological factors such as the initial load of contamination or the efficiency of the enrichment procedure, and the expected number of bacterial strains. Generally, we have observed that coverages of over 5–10× can be sufficient for detection of virulence genes and phylogenetic placement of bacterial strains in case reference-based assembly is used [16]. However, there is a need to precisely establish the reliability of the strain characterization and subtyping results obtained using data of different sequencing depth. Third, user-friendly pipelines need to be developed to be used directly in the laboratory by non-expert bioinformaticians. Moreover, bioinformatics taxonomic identification tools should be further tested and improved, so that different tools, each with their advantages and limitations, provide the same results, and to avoid misclassifications [52]. However, the focus of this study was not to present a benchmarking of bioinformatics tools for strain-level shotgun metagenomics, but rather a proof of concept based on previously developed bioinformatics methodologies [16, 17]. Other approaches and tools might still improve the results (accuracy, speed of analysis) and could be evaluated in further studies [53, 54]. This confirms the need for studies such as this one to produce data to make benchmarking analyses possible or help in the design of new tools. Another perspective for the implementation of this method in routine analysis is the reduction of the analysis cost. As elaborated above, shotgun metagenomics analyses imply runs with a very limited number of samples on Illumina sequencers in order to maximize the sequencing depth. Other sequencing devices as manufactured, for instance, by Oxford Nanopore Technologies offer real-time long-read sequencing of one sample at a time, at a low price if using the Flongle flow cell. Such fast sequencing could also further reduce the turnaround time of a metagenomics-based outbreak investigation [55]. However, its applicability for strain-level characterization in complex samples remains to be demonstrated.

In 2019, the EFSA published an opinion on the use of metagenomics for outbreak investigation [11], describing the possibilities offered by an isolation-free method. However, at that time, metagenomics had not yet been used to resolve a food-borne outbreak investigation to its food source and was considered as experimental. Moreover, it was considered technically challenging to obtain a draft genome of the pathogenic strain in order to assign particular genetic determinants to the causative agent. This study has shown that a *Salmonella* outbreak caused by a complex food matrix could be resolved to strain resolution using shotgun metagenomics, in a shorter time frame than needed for isolation of the strain, paving the way for future studies to use this method outside the experimental scope and to support the EFSA opinion.

Funding information

The research that yielded these results was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contract RF 17/6316 StEQIDEMIC.be and by Sciensano through the contract RP Be READY.

Acknowledgements

We would like to thank all collaborators from the agency Zorg en Gezondheid, the Federal Agency for the Security of the Food Chain and the Sciensano epidemiology department who participated in this outbreak investigation. We are grateful to the technicians from the NRL and NRC of *Salmonella* (Sciensano) for their work on the bacterial isolates and the food enrichments, as well as Stefan Hoffman, Maud Delvoeye and Els Vandermassen from Transversal Activities in Applied Genomics (Sciensano) for their involvement in the sequencing runs. Finally, we acknowledge Mathieu Gand for sharing his expertise about *Salmonella* and serotyping.

Author contributions

Conceptualization: F. E. B., S. D., K. M., S. C. J. D. K. Methodology: F. E. B., S. D., S. C. J. D. K. Software: F. E. B., A. S., K. V. Validation: F. E. B., W. M., B. V., K. V., S. D., S. C. J. D. K. Formal Analysis: F. E. B., A. S., S. C. J. D. K. Investigation: F. E. B., W. M., B. V., V. L., N. H., B. P., V. C., S. D. Resources: W. M., B. V., N. H. C. R., K. V., V. L., N. H., B. P., V. C., S. D. Data Curation: F. E. B., A. S. Writing – Original Draft Preparation: F. E. B., S. D., S. C. J. D. K. Writing – Review and Editing: F. E. B., A. S., W. M., B. V., N. H. C. R., K. V., V. L., N. H., B. P., V. C., K. M., S. D., S. C. J. D. K. Visualization: F. E. B. Supervision: K. M., S. C. J. D. K. Project Administration: S. C. J. D. K. Funding: N. H. C. R., S. C. J. D. K.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. ECDC, EFSA. EFSA and ECDC technical report on the collection and analysis of whole genome sequencing data from food-borne pathogens and other relevant microorganisms isolated from human, animal, food, feed and food/feed environmental samples in the joint ECDC-EFSA molecular typing database. EFSA Support Publ [Internet]. 2019;16(5). Available from: <http://doi.wiley.com/10.2903/sp.efsa.2019.EN-1337>.
2. Naravaneni R, Jamil K. Rapid detection of food-borne pathogens by using molecular techniques. *J Med Microbiol* 2005;54:51–54.
3. European Union. Commission regulation (EC) no 2073/2005 of 15 November 2005 on microbiological criteria for foodstuffs. *Off J Eur Union*;2005:32.
4. EFSA, ECDC. The European Union one health 2018 zoonoses report. *EFSA J* 2019;17:5926.
5. Sala C, Mordhorst H, Grütze J, Brinkmann A, Petersen TN *et al.* Metagenomics-based proficiency test of smoked salmon spiked with a mock community. *Microorganisms* 2020;8:1861.

6. **European Commission.** *Lessons Learned from the 2011 Outbreak of Shiga Toxin-Producing Escherichia coli (STEC) O104:H4 in Sprouted Seeds, Commission Staff Working Document* (https://ec.europa.eu/food/sites/food/files/safety/docs/biosafety-crisis-cswd_lessons_learned_en.pdf). Brussels: European Commission; 2011.
7. Kovac J, den Bakker H, Carroll LM, Wiedmann M. Precision food safety: a systems approach to food safety facilitated by genomics tools. *Trends Analyt Chem* 2017;96:52–61.
8. Carleton HA, Besser J, Williams-Newkirk AJ, Huang A, Trees E et al. Metagenomic approaches for public health surveillance of foodborne infections: opportunities and challenges. *Foodborne Pathog Dis* 2019;16:474–479.
9. Höper D, Mettenleiter TC, Beer M. Metagenomic approaches to identifying infectious agents. *Rev Sci Tech* 2016;35:83–93.
10. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet* 2018;19:9–20.
11. **EFSA Panel on Biological Hazards (EFSA BIOHAZ Panel), Koutsoumanis K, Allende A, Alvarez-Ordóñez A, Bolton D et al.** Whole genome sequencing and metagenomics for outbreak investigation, source attribution and risk assessment of food-borne microorganisms. *EFSA J* 2019;17:e05898.
12. Leonard SR, Mammel MK, Lacher DW, Elkins CA. Application of metagenomic sequencing to food safety: detection of Shiga toxin-producing *Escherichia coli* on fresh bagged spinach. *Appl Environ Microbiol* 2015;81:8183–8191.
13. Leonard SR, Mammel MK, Lacher DW, Elkins CA. Strain-level discrimination of Shiga toxin-producing *Escherichia coli* in spinach using metagenomic sequencing. *PLoS One* 2016;11:e0167870.
14. Walsh AM, Crispie F, Daari K, O'Sullivan O, Martin JC et al. Strain-level metagenomic analysis of the fermented dairy beverage nunu highlights potential food safety risks. *Appl Environ Microbiol* 2017;83:e01144-17.
15. Yang X, Noyes NR, Doster E, Martin JN, Linke LM et al. Use of metagenomic shotgun sequencing technology to detect foodborne pathogens within the microbiome of the beef production chain. *Appl Environ Microbiol* 2016;82:2433–2443.
16. Saltykova A, Buytaers FE, Denayer S, Verhaegen B, Piérard D et al. Strain-level metagenomic data analysis of enriched in vitro and in silico spiked food samples: paving the way towards a culture-free foodborne outbreak investigation using STEC as a case study. *Int J Mol Sci* 2020;21:5688.
17. Buytaers FE, Saltykova A, Denayer S, Verhaegen B, Vanneste K et al. A practical method to implement strain-level metagenomics-based foodborne outbreak investigation and source tracking in routine. *Microorganisms* 2020;8:1191.
18. **ISO.** ISO 6579-1:2017 Microbiology of the Food Chain – Horizontal Method for the Detection, Enumeration and Serotyping of *Salmonella* – Part 1: Detection of *Salmonella* spp. Geneva: International Organization for Standardization; 2017.
19. **EFSA Panel on Biological Hazards (BIOHAZ).** Scientific opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance: part 1 (evaluation of methods and applications). *EFSA J* 2013;11:3502.
20. **EFSA.** *Use of Whole Genome Sequencing (WGS) of Food-borne Pathogens for Public Health Protection*. Luxembourg: EFSA; 2014.
21. Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M et al. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST subcommittee. *Clin Microbiol Infect* 2017;23:2–22.
22. Tang S, Orsi RH, Luo H, Ge C, Zhang G et al. Assessment and comparison of molecular subtyping and characterization methods for *Salmonella*. *Front Microbiol* 2019;10:1591.
23. Franz E, Gras LM, Dallman T. Significance of whole genome sequencing for surveillance, source attribution and microbial risk assessment of foodborne pathogens. *Curr Opin Food Sci* 2016;8:74–79.
24. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods* 2015;12:733–735.
25. Huang AD, Luo C, Pena-Gonzalez A, Weigand MR, Tarr CL et al. Metagenomics of two severe foodborne outbreaks provides diagnostic signatures and signs of coinfection not attainable by traditional methods. *Appl Environ Microbiol* 2017;83:e02577-16.
26. Hyeon J-Y, Li S, Mann DA, Zhang S, Li Z et al. Quasimetagenomics-based and real-time sequencing-aided detection and subtyping of *Salmonella enterica* from food samples. *Appl Environ Microbiol* 2018;84:e02340-17.
27. **Sciensano.** *Voedselvergiftigen in België Jaaroverzicht 2019* (https://www.sciensano.be/sites/default/files/jaarverslagboekje_vti2019_nl2020.pdf). Brussels: Sciensano; 2020.
28. **Sciensano.** *Centre National de Référence Salmonella & Shigella Rapport Annuel 2019* (https://nrchm.wiv-isp.be/fr/centres_ref_lab0/salmonella_et_shigella_spp/Rapports/Salmonella+Shigella2019.pdf). Brussels: Sciensano; 2020.
29. **AFSCA.** Communiqué de presse de l'Agence Régionale "Zorg en Gezondheid" et de l'Agence Fédérale pour la Sécurité de la Chaîne Alimentaire: résultats de l'enquête sur le foyer de salmonelles l'école hôtelière Spermalie Bruges. Brussels: AFSCA; 2019. Available at: <http://www.afsca.be/professionnels/publications/presse/2019/2019-09-23b.asp>
30. Pijnacker R, Dallman TJ, Tijmsa ASL, Hawkins G, Larkin L et al. An international outbreak of *Salmonella enterica* serotype Enteritidis linked to eggs from Poland: a microbiological and epidemiological study. *Lancet Infect Dis* 2019;19:778–786.
31. **ECDC, EFSA.** *Joint ECDC-EFSA Rapid Outbreak Assessment: Multi-Country Outbreak of Salmonella enteritidis Infections Linked to Eggs, fourth update*. Solna, Parma: ECDC, EFSA; 2020.
32. Barbau-Piednoir E, Bertrand S, Mahillon J, Roosens NH, Böttel-dorn N. SYBR®Green qPCR *Salmonella* detection system allowing discrimination at the genus, species and subspecies levels. *Appl Microbiol Biotechnol* 2013;97:9811–9824.
33. Andrews S. FastQC: a quality control tool for high throughput sequence data; 2010. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
34. **EFSA Panel on Biological Hazards (BIOHAZ).** Scientific opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance: part 2 (surveillance and data management activities). *EFSA J* 2014;12:3784.
35. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;20:257.
36. Cosentino S, Voldby Larsen M, Møller Aarestrup F, Lund O. PathogenFinder – distinguishing friend from foe using bacterial whole genome sequence data. *PLoS One* 2013;8:e77302.
37. Marcelino VR, Clausen PTLC, Buchmann JP, Wille M, Iredell JR et al. CCMetagen: comprehensive and accurate identification of eukaryotes and prokaryotes in metagenomic data. *Genome Biol* 2020;21:103.
38. Ahn TH, Chai J, Pan C. Sigma: strain-level inference of genomes from metagenomic analysis for biosurveillance. *Bioinformatics* 2015;31:170–177.
39. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
40. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
41. Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VPJ et al. The *Salmonella in silico* typing resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft *Salmonella* genome assemblies. *PLoS One* 2016;11:e0147101.
42. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012;67:2640–2644.

43. Bogaerts B, Winand R, Fu Q, Van Braekel J, Ceysens PJ *et al.* Validation of a bioinformatics workflow for routine analysis of whole-genome sequencing data and related challenges for pathogen typing in a European national reference center: *Neisseria meningitidis* as a proof-of-concept. *Front Microbiol* 2019;10:362.
44. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018;35:1547–1549.
45. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 1993;10:512–526.
46. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 2016;44:W242–W245.
47. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA *et al.* The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother* 2013;57:3348–3357.
48. Jain S, Mukhopadhyay K, Thomassin PJ. An economic analysis of *Salmonella* detection in fresh produce, poultry, and eggs using whole genome sequencing technology in Canada. *Food Res Int* 2019;116:802–809.
49. Loman NJ, Constantinidou C, Christner M, Rohde H, Chan JZ-M *et al.* A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxicogenic *Escherichia coli* O104:H4. *JAMA* 2013;309:1502–1510.
50. Quick J, Ashton P, Calus S, Chatt C, Gossain S *et al.* Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol* 2015;16:114.
51. ISO. ISO/DIS 23418 Microbiology of the Food Chain – Whole Genome Sequencing for Typing and Genomic Characterization of Foodborne Bacteria – General Requirements and Guidance. Geneva: International Organization for Standardization; 2021.
52. Marcelino VR, Holmes EC, Sorrell TC. The use of taxon-specific reference databases compromises metagenomic classification. *BMC Genomics* 2020;21:184.
53. Seeman T. snippy: fast bacterial variant calling from NGS reads; 2015. <https://github.com/tseemann/snippy>
54. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;37:1530–1534.
55. Juul S, Izquierdo F, Hurst A, Dai X, Wright A. What's in my pot? Real-time species identification on the MinION. *bioRxiv* 2015:030742.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.