



OPEN

## Nanopore metatranscriptomics reveals cryptic catfish species as potential *Shigella flexneri* vectors in Kenya

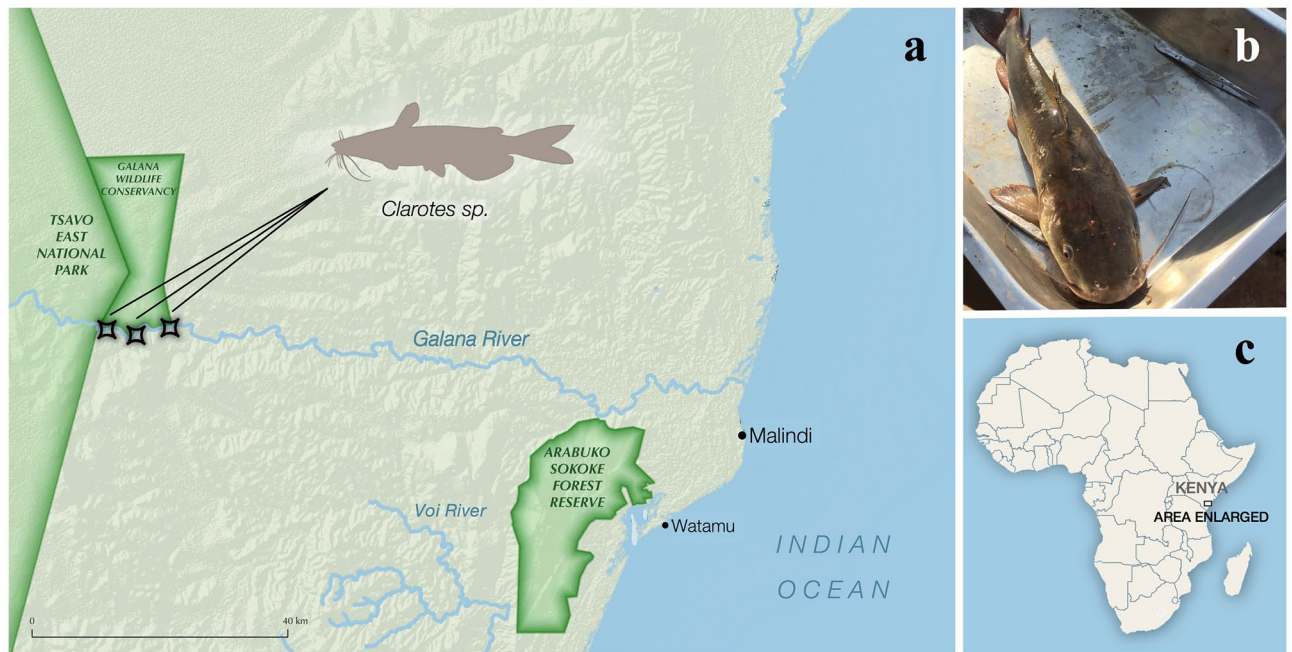
Andrew J. Tighe<sup>1,2</sup>✉, Sean Grayson<sup>1</sup>, John Byrne<sup>1</sup>, Sanni Hintikka<sup>1</sup>, Lisa Jessen<sup>3</sup>, Jake Dempsey<sup>1</sup>, Lauren Browne<sup>1</sup>, Mary Kelly-Quinn<sup>3</sup>, Bernerd Fulanda<sup>4</sup>, Neil M. Ruane<sup>5</sup> & Jens Carlsson<sup>1</sup>

Bacteria in the *Shigella* genus remain a major cause of dysentery in sub-Saharan Africa, and annually cause an estimated 600,000 deaths worldwide. Being spread by contaminated food and water, this study highlights how wild caught food, in the form of freshwater catfish, can act as vectors for *Shigella flexneri* in Southern Kenya. A metatranscriptomic approach was used to identify the presence of *Shigella flexneri* in the catfish which had been caught for consumption from the Galana river. The use of nanopore sequencing was shown to be a simple and effective method to highlight the presence of *Shigella flexneri* and could represent a potential new tool in the detection and prevention of this deadly pathogen. Rather than the presence/absence results of more traditional testing methods, the use of metatranscriptomics highlighted how primarily one SOS response gene was being transcribed, suggesting the bacteria may be dormant in the catfish. Additionally, *COI* sequencing of the vector catfish revealed they likely represent a cryptic species. Morphological assignment suggested the fish were widehead catfish *Clarotes laticeps*, which range across Africa, but the *COI* sequences from the Kenyan fish are distinctly different from *C. laticeps* sequenced in West Africa.

Zoonotic pathogens pose an ongoing threat to human society, as acutely highlighted with SARs-COV2, the causative agent of Covid-19<sup>1</sup>. Generally, zoonotic viruses originate in mammalian or avian hosts, with no known zoonotic viruses originating in fish<sup>2</sup>. However, fish can be sources of a number of bacterial and parasitic diseases<sup>3,4</sup>. The primary pathogens from fish that pose a risk to humans are bacterial, with pathogens such as *Salmonella* sp., *Campylobacter* sp., *Escherichia coli*, *Listeria monocytogenes* and *Yersinia* sp. being responsible for most foodborne outbreaks from fish worldwide<sup>5</sup>. It is likely also that undescribed zoonotic pathogens are circulating in wild fish, with developing countries in the tropics predicted to be the highest risk areas for future zoonoses to emerge<sup>6</sup>. Therefore in this paper, a pilot study for an early warning system using shotgun sequencing approach was used to determine whether any pathogens could be identified from wild caught widehead catfish *Clarotes laticeps* found in the Galana river in South East Kenya. *Clarotes laticeps* are an important protein source for local communities and are routinely harvested directly from the river. The species is a tropical, freshwater catfish in the family *Claroteidae*<sup>7</sup>, which is present in all Nilo-Sudanese basins<sup>8</sup> in addition to East Africa and the Nile<sup>9</sup>.

A number of bacterial pathogens are known to be associated with the consumption of catfish species (both farmed and wild caught), such as *Salmonella*, *Campylobacter*, *E. coli*, *L. monocytogenes*, *Staphylococcus aureus* and *Vibrio*<sup>3</sup>. Some bacteria, such as *Salmonella*<sup>10–12</sup> and *Campylobacter*<sup>13</sup> are thought to be introduced to aquaculture ponds through the faeces of birds and wildlife, and faecal matter from ruminants is known to contaminate ponds with pathogenic *E. coli*<sup>3</sup>. Additionally, *Salmonella* has been detected in wild catfish<sup>14</sup> and *Grimontia hollisae* (formerly known as *Vibrio hollisae*) was implicated as the cause of septicaemia in a man who had eaten wild caught catfish in the USA<sup>15</sup>. In Sub-Saharan Africa, wild caught fish have been shown to be sources of bacterial enteropathogens, with one example in Central African Republic specifically linking the presence of Shiga toxin producing *E. coli* in wild fish and river water to the upstream slaughter of zebu cattle *Bos indicus*<sup>16</sup>.

<sup>1</sup>Area 52 Research Group, School of Biology and Environmental Science/Earth Institute/University College Dublin, Dublin 4, Ireland. <sup>2</sup>Fish Health Unit, Marine Institute, Oranmore, Co., Galway, Ireland. <sup>3</sup>School of Biology and Environmental Science/Earth Institute, University College Dublin, Dublin 4, Ireland. <sup>4</sup>Department of Biological Sciences, Pwani University, Kilifi, Kenya. <sup>5</sup>Fisheries Ecosystems Advisory Services, Marine Institute, Oranmore, Co., Galway H91 R673, Ireland. ✉email: andrew.tighe@ucdconnect.ie



**Figure 1.** Locations where samples were collected. (A) Map showing where the fish were collected in South East Kenya. (B) Photograph of one of the widehead catfish provided by the scouts. (C) Map highlighting the location of the region shown in (A) on the African continent. Map generated using QGIS3.6 (<https://qgis.org/en/site/fusers/download.html>).

Enterotoxigenic *E. coli* in addition to *Shigella* spp. are responsible for the majority of bacterial caused diarrhoeal episodes in children under 5 years in Sub-Saharan Africa<sup>17</sup>. Similar to *E. coli*, *Shigella* spp. are also known to be introduced to water sources from anthropogenic sources such as untreated sewage and agricultural runoff<sup>18</sup>. Outbreaks of shigellosis due to *Shigella* spp. from contaminated water are known throughout Sub-Saharan Africa<sup>19</sup>, in addition to Asia<sup>18,20</sup>. Additionally the presence of antimicrobial resistance genes has been detected in *Shigella* spp. in waterways, likely due to pollution in water sources from industry, veterinary medications and human medical treatment<sup>18</sup>.

The detection of bacterial pathogens, both from fish and other sources, has increasingly been based on the use of polymerase chain reaction (PCR) based approaches over the past two decades<sup>21</sup>. One significant drawback of this approach is that it can only be used to test for known pathogens, with multiple presence/absence tests needed to check for all potential pathogens. Second and third generation sequencing platforms such as Illumina's MiSeq and Oxford nanopore's MinION, however, allow for shotgun sequencing approaches which can be used to sequence all DNA present in a sample, potentially revealing most microbial life present in one test, saving time and allowing for the rapid discovery of novel disease causing pathogens. Third generation nanopore sequencing has been used in a number of shotgun metagenomic studies to identify pathogens, such as diagnosing the cause of infections in orthopaedic implants<sup>22</sup>, identifying oral bacteriophages<sup>23</sup> and characterising the microbiomes of preterm human babies<sup>24</sup>. The approach has also been used to identify waterborne pathogens, with Reddington et al.<sup>25</sup> showing how wastewater influents increased the abundance of *Arcobacter* when compared to cleaner parts of the Havelse river in Denmark.

In this study, a shotgun transcriptomics approach was used in an attempt to identify pathogens found in wild widehead catfish in the Galana river in South-East Kenya. In addition to shotgun sequencing, the *cytochrome oxidase I* (COI) gene of the fish was sequenced using Sanger sequencing, in order to confirm the species of catfish found in the river. Although the widehead catfish in the Galana river is currently classified as *C. laticeps*, there remains taxonomic uncertainty as to whether this population of catfish are in fact *C. laticeps*<sup>26</sup>.

## Methods

**Sample collection and sequencing.** Four catfish samples were collected between the 7th and 13th of August, 2019 from three points along the Galana river (Fig. 1). The deceased catfish, which had been caught using a baited hook and line, were gifted by local scouts and samples were taken from remains prior to preparation as food. The catfish sampled were identified as widehead catfish based on morphological characteristics. Fish were dissected in a sterile metal tray, with utensils sterilised by ethanol and flamed between samples. Single muscle and heart samples were taken from each fish, but in the process of dissecting the fish the intestines were significantly nicked, meaning the heart samples were likely contaminated with intestinal material. Tissue samples were placed in RNAlater (Invitrogen, USA) in 2 mL O-ring tubes and kept at ambient temperature in the field for up to eleven days, and later stored at  $-80^{\circ}\text{C}$ .

The four heart samples were homogenised using a TissueLyser (Qiagen), and RNA was then extracted using an QIAamp Viral RNA Mini Kit (Qiagen) via a QiaCube extraction machine (Qiagen). The RNA was converted

Species	Genbank Assembly Accession
<i>Candida dubliniensis</i>	GCF_000026945.1
<i>Acinetobacter baumannii</i>	GCF_002116935.1
<i>Aquicella siphonis</i>	GCF_902459484.1
<i>Arcobacter cibarius</i>	GCA_013372265.1
<i>Chryseobacterium lactis</i>	GCA_003815875.1
<i>Clostridium perfringens</i>	GCA_000013285.1
<i>Cutibacterium acnes</i>	GCA_000008345.1
<i>Escherichia coli</i>	GCA_904425475.1
<i>Enterococcus faecium</i>	GCA_010120755.1
<i>Flavobacterium gilvum</i>	GCA_001761465.1
<i>Klebsiella pneumoniae</i>	GCA_000240185.2
<i>Metabacillus litoralis</i>	GCA_003667825.1
<i>Moraxella osloensis</i>	GCA_001553955.1
<i>Nodularia spumigena</i>	GCA_000340565.3
<i>Pseudomonas synxantha</i>	GCA_003851495.1
<i>Salmonella enterica</i>	GCA_001558355.2
<i>Sphingobacterium daejeonense</i>	GCA_901472535.1
<i>Sphingomonas paucimobilis</i>	GCA_003314795.2
<i>Staphylococcus aureus</i>	GCA_000013425.1
<i>Vibrio parahaemolyticus</i>	GCA_004006515.1
<i>Shigella flexneri</i>	GCA_013391345.1
<i>Escherichia coli</i>	GCA_000183345.1
<i>Sugiyamaella lignohabitans</i>	GCA_001640025.2

**Table 1.** Microbial genomes included in the reference database. List of species genomes downloaded from Genbank based on the results generated from WIMP

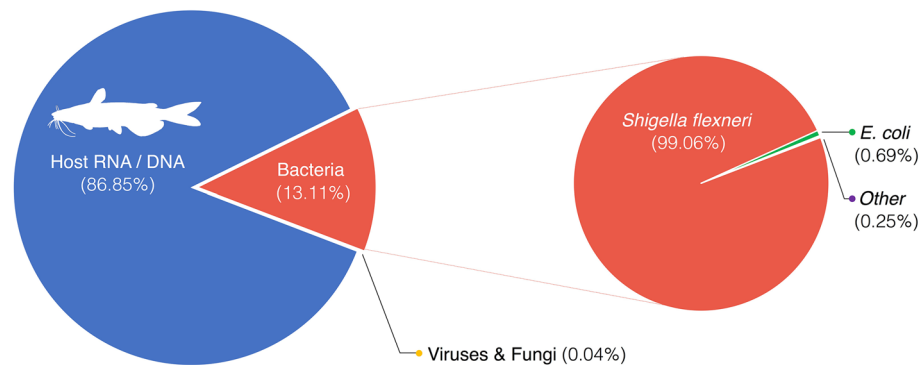
to cDNA using Superscript II Reverse Transcriptase (Invitrogen, USA) following the manufacturer's guidelines using 8 µl of RNA and 1 µl random hexamer primers per reaction.

The concentration of the RNA/DNA samples for four samples was determined using a BioDrop µLite v1.0.4 (Thermo Fisher Scientific) and 7.5 µl of each sample was barcoded using 2.5 µl of Fragmentation Mix RB01-04 from Oxford Nanopore's Rapid Barcoding Sequencing kit (SQK-RBK004, Oxford Nanopore Technologies). Barcoded samples were then equimolar pooled and sequencing adapters were ligated to the DNA library using 4 µl of Rapid Adapter reagent (RAP, Oxford Nanopore Technologies). The barcoded library was loaded onto a MinION flow cell Mk1 R9.4.1 (Oxford Nanopore Technologies) and run via MinKNOW software v.3.6.5 (Oxford Nanopore Technologies) without real-time base-calling for 68 h on a MinION sequencer (Oxford Nanopore Technologies).

**Nanopore data analysis.** Guppy v.3.2.10 (Oxford Nanopore Technologies) was used to base-call the output from the MinION sequencing run. Porechop (<https://github.com/rrwick/Porechop>) was used to remove the adaptor sequences from the MinION sequence data. Porechop was also originally used to sort samples by barcodes, but as 215,028 reads were left unassigned (89.44% of reads) and due to the low read depth it was decided to pool the reads for all four catfish samples for further analysis.

In order to first assess the microbial organisms present, the fastq reads were processed using the What's In My Pot (WIMP) workflow on Oxford Nanopore's EPI2ME platform (<https://epi2me.nanoporetech.com/>, Oxford Nanopore Technologies). Following the WIMP analysis, a reference fasta database was generated using an existing genome for North American yellow catfish *Tachysurus fulvidraco* (Genbank Assembly Accession: GCA\_003724035.1) to act as a proxy for the host genome. Additionally a genome for the parasite *Schistosoma haematobium* (Genbank Assembly Accession: GCA\_000699445.2), the 12,642 existing viral sequences from ray-finned fish *Actinopterygii* hosts available on the NCBI virus database (<https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/>) and the genomes for the 23 bacterial and fungal species identified through WIMP (Table 1) were added to the reference fasta file. The fastq reads were mapped against the reference fasta database using the NanoPipe web server (<http://www.bioinformatics.uni-muenster.de/tools/nanopipe2/index.hbi?>)<sup>27</sup>. The consensus sequences for mapped reads were then subject to a BLASTn search (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), and consensus sequences which did not correspond to any existing sequences were removed from the final result.

**Sequencing and analysing the COI of the host.** A separate DNA extraction was performed on the catfish muscle samples collected, with samples being incubated at 56 °C for 2 h in a mix of 12 µl of proteinase K (20 mg/ml) and 400 µl of 10% Chelex solution, followed by 15 min at 99 °C. Samples were then centrifuged for 1 min at max speed (20,817 G) and 150 µl of DNA supernatant was placed in a new 1.5 mL Eppendorf tube. A



**Figure 2.** Taxon of the 59,316 reads which were successfully mapped. Host RNA is based on reads that mapped to the genome of yellow catfish. The species breakdown of the reads which mapped to bacterial sequences is also given.

653 bp amplicon of the *COI* gene was amplified using the primer pair FishF2 (5'- TGT AAA ACG ACG GCC AGT CGA CTA ATC ATA AAG ATA TCG GCA C -3')<sup>28</sup> and FishR2 (5'- CAG GAA ACA GCT ATG ACA CTT CAG GGT GAC CGA AGA ATC AGA A -3')<sup>28</sup>. A 25 µl master mix was prepared in a UV-sterilized hood consisting of 3.125 µl of Buffer (Kapa Biosystems), 1.25 µl of 10 mM dNTPs (Invitrogen), 1.25 µl of each primer (Integrated DNA Technologies), 0.125 µl Taq polymerase (Kapa Biosystems), 2.5 µl of Bovine Serum Albumin (BSA)(20 mg/ml)(Thermo Scientific), 14.5 µl of dd H<sub>2</sub>O and 1 µl of DNA. PCR conditions were as follows: initiation at 95 °C for 5 min, followed by 35 cycles of 95 °C for 30 s, 50 °C for 1 min and 72 °C for 1 min, with a final extension for 10 min at 72 °C. Samples which showed a clear band on a safeview stained agarose gel after electrophoresis were then subject to enzymatic clean-up prior to sequencing using a mixture of 0.05 µl ExoI, 0.1 µl TSAP and 4.85 µl H<sub>2</sub>O, added to 20 µl of PCR product and then incubated at 37 °C for 15 min followed by 80 °C for 15 min.

Samples were subsequently sent for commercial Sanger sequencing (Macrogen). Forward and reverse strands were aligned using Geneious version 10.2.3<sup>29</sup>, and the consensus sequences were initially analysed using BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Following this the corresponding aligned sequence from the top 100 BLAST hits were downloaded in fasta format and added to an alignment containing the four sequences for the Galana widehead catfish. The alignment file was then uploaded to the IQ-TREE web server for model selection (<http://iqtree.cibiv.univie.ac.at/>),<sup>30</sup> and the web server of the RAXML program (<https://raxml-ng.vital-it.ch/>)<sup>31</sup> was used to generate an ML tree, with 100 bootstrap replicates performed.

## Results

**MinION sequence analysis.** Following adaptor removal, 240,429 reads remained, of which 68,908 were successfully mapped to one of the reference sequences using NanoPipe<sup>27</sup>. Of the 68,908 mapped reads, 59,849 mapped to the proxy host (Fig. 2), 9031 mapped to bacterial sequences, 22 mapped to fungal sequences, and 6 mapped to viral sequences. The majority of the 9031 bacterial reads (99.06%, 8,938 reads) (Fig. 2) mapped to the genome of *Shigella flexneri*, specifically to a predicted response regulator (Genbank Accession: CP045522.1, position 4,265,154–4,265,537) which shares protein homology with the two component response regulator gene *DpiA* in *Shigella boydii*<sup>32</sup>. Further investigation of the viral and fungal mapped reads revealed these to be ambiguous, and no bacteriophages were detected.

**Host *COI* analysis.** Clean, unambiguous Sanger sequences were generated for all four widehead catfish samples, with all samples sharing the same haplotype. A BLAST search of the consensus sequence revealed the closest match to be a sequence for *Bathybagrus tetranema* (Accession number: HG803463)<sup>33</sup> at 93.75% similarity. The consensus sequence matched 91.89% to an existing *Clarotes laticeps* on Genbank (Accession number: HG803491)<sup>33</sup>.

The maximum likelihood tree (Fig. 3) showed that whilst the host catfish sequence is not closely related to any existing catfish sequences, it does cluster within the family *Claroteidae*. The Galana widehead catfish show a deep split from the widehead catfish sequenced from Nigeria (Fig. 3), and show a basal split near the base of the *Claroteidae* cluster.

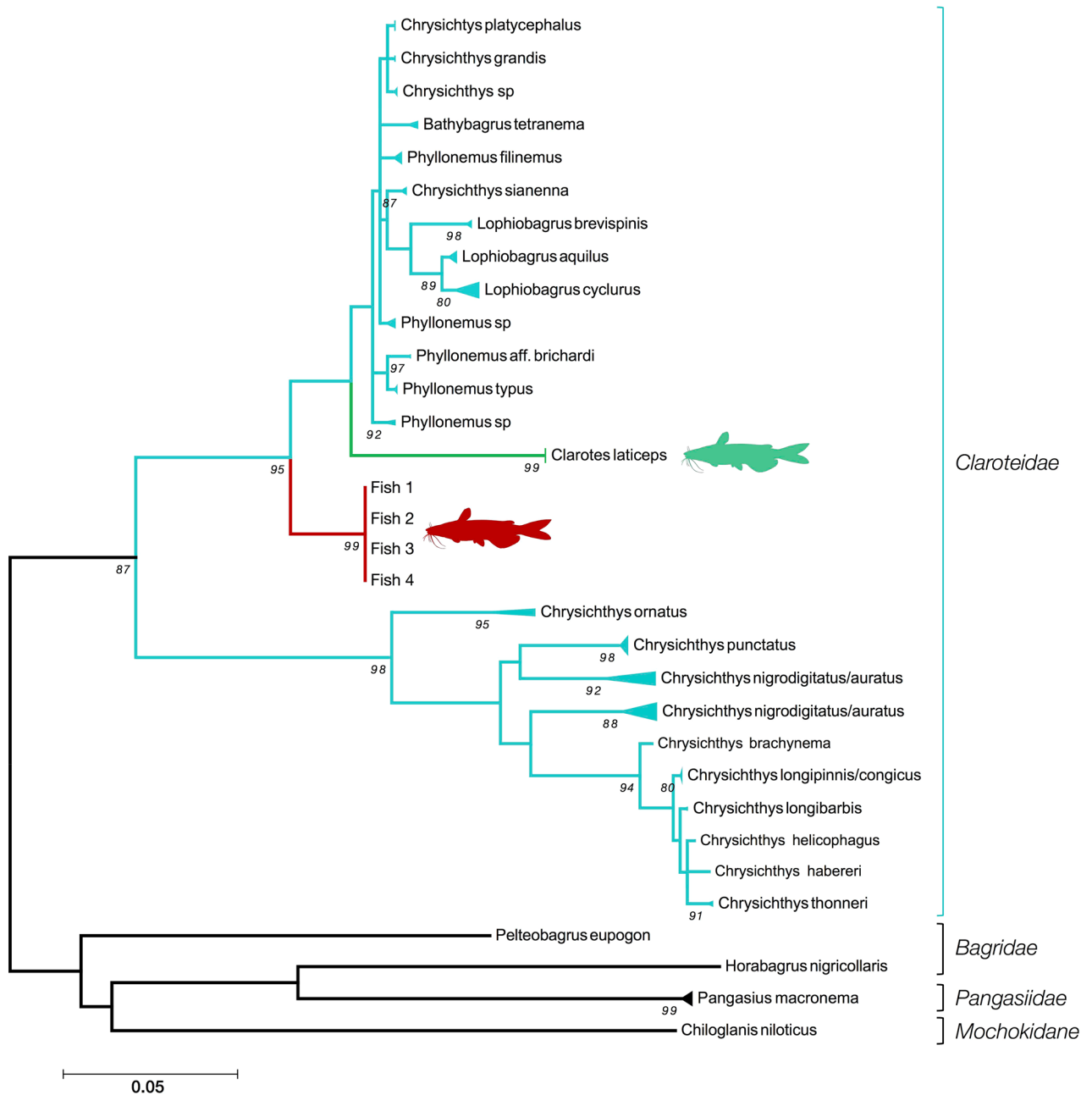
## Discussion

Metagenomic approaches using second generation sequencing platforms have been used previously to identify foodborne pathogens directly from a source animal<sup>34</sup>, however it would appear that this study represents the first time a metagenomic approach using nanopore sequencing has been used to detect a foodborne bacterial pathogen directly from an animal. Previous studies have shown the effectiveness of MinION's nanopore sequencer as a tool for sequencing foodborne bacterial pathogens<sup>35–37</sup>, but these studies have relied on bacteria that were isolated from a food source and then cultured prior to sequencing. It is estimated that approximately 99% of prokaryotes are unable to be cultured in the laboratory with currently available methods<sup>38</sup>, which highlights the importance of developing diagnostic workflows which are culture independent, such as the approach outlined in this study.

### Location where widehead catfish were caught:

 | Galana river, Kenya

 | Amambra river, Nigeria



**Figure 3.** The ML tree generated using the catfish COI sequences. Highlighted in red are the COI sequences of the widehead catfish collected from the Galana river. Highlighted in green are COI sequences corresponding to widehead catfish collected in the Amambra river, Nigeria.

The identification of the presence of *Shigella flexneri* in the catfish in this study highlights the potential of nanopore metagenomics as a relatively simple and effective way to detect human pathogens in one test. The causative agent of human shigellosis, *Shigella* is a genus of gram-negative bacteria that cause diarrhoea and dysentery, being a major cause of moderate-to-severe diarrhoea in sub-Saharan Africa<sup>39</sup>. Annually, *Shigella* is estimated to cause 600,000 deaths from 80 to 165 million cases worldwide<sup>40</sup>. In the developing world, *S. flexneri* is the predominant cause of shigellosis, with *S. sonnei* being the predominant strain in industrialised countries<sup>40</sup>. The

faecal-oral route is the primary way by which *Shigella* spreads, with transmission also documented via contaminated food, drinking water and flies<sup>41</sup>. The *S. flexneri* gene that the majority of the reads mapped to, *DpiA*, has been shown to interfere with plasmid maintenance when overexpressed, inducing the SOS response<sup>42</sup>, which is a bacterial response that promotes dormancy in unfavourable environmental conditions<sup>43</sup>. The overabundance of the *DpiA* gene could imply that the bacteria were in a dormant state, but as other genes known to be involved in the SOS response were not detected with this approach, this cannot be confirmed. It may be the case that *DpiA* is expressed at much higher levels than the other genes in the SOS response, and a deeper sequencing effort might reveal further genes being expressed. Additionally the overexpression of *DpiA* has been shown to be part of the bacterial defence following exposure to  $\beta$ -lactam antibiotics in *E. coli*<sup>44</sup>, which could imply the presence of antimicrobial resistance as seen with *Shigella* spp. in other water sources<sup>18</sup>.

It is unclear how the catfish became carriers of *S. flexneri*, so it can only be speculated how the bacteria may have been introduced to the fish based on the known biology of the pathogen. Although the *Shigella* RNA was extracted from the heart tissue, we believe in fact it was present in the intestinal tract of the catfish, as whilst the fish were dissected in a sterilised tray, clean gloves were worn and utensils were flamed before each dissection, in the process of initially opening the fish the intestines were significantly nicked, meaning the heart samples were likely contaminated with DNA from material within the intestines. The primary hosts of *S. flexneri* are humans and primates<sup>45</sup>, but it has also been detected in rabbits<sup>46</sup>, cattle<sup>47</sup>, pigs<sup>48</sup> and chicken<sup>45</sup>. Previous studies have also shown that bacterial pathogens can be introduced to fish in water bodies via the faeces of cattle, birds and wild animals<sup>3,10–13,16</sup>. This could imply that *S. flexneri* was introduced to the river either via the large herds of cattle which are brought to the river to drink, or through the various wild animals that use the river, with species such as hippopotamus and elephant known to be sources of other bacterial diseases such as anthrax, brucellosis, tetanus and salmonellosis (hippopotamus)<sup>49</sup> and tuberculosis (elephant)<sup>50</sup>, and subsequently ingested in sediment by the bottom feeding catfish. In addition to the potential animal sources listed, a number of villages located approximately 150 km upstream of the sampling site, on the western side of Tsavo East National Park, are suspected to be primary sources of untreated sewage into the Galana river (John Byrne, University College Dublin, pers. comm.).

The sequence analysis of the host catfish *COI* also revealed that it was not the species *Clarotes laticeps* as suggested based on morphology, but the phylogeny does suggest the fish belongs in the family *Claroteidae*. The species *C. laticeps* is found naturally in West Africa, the Nilo-Sudanese basin and East Africa<sup>8,26,51,52</sup>, and the other *C. laticeps* sequence included in the analysis (Fig. 3) was collected in the Amambra river in Nigeria<sup>53</sup> which is ~4000 km away from the Galana river, and the two rivers are separated by at least two major watersheds. According to both Okeyo<sup>54</sup> and Seegers et al.<sup>26</sup>, and based on morphology, the only member of the family *Claroteidae* found in the Galana river is *C. laticeps*, however Seegers et al.<sup>26</sup> added the caveat “the taxonomic status of the Kenyan populations is uncertain and needs detailed study”. This would suggest that the fish examined in this study were in fact the fish that to date have been classified as *C. laticeps* based on morphology, but now the genetic data have revealed that this population of catfish are likely a cryptic species. Based on these findings we tentatively suggest a new species name *Clarotes kambare*, or Kenyan widehead catfish, with “kambare” being the Swahili word for catfish, although further in-depth morphological and genetic work will be needed to clarify the unique taxonomic status of this population of fish. Further studies are also required to determine the full geographic location of this species and its conservation status.

In conclusion, this study has highlighted how a shotgun metatranscriptomic approach can be used to identify human pathogens in wild caught fish, and how it may highlight the transmission potential of enteropathogens in non-host animals. Further research is needed going forward to explore the potential benefits of this approach over the presence/absence results of conventional PCR assays or 16S sequencing, as while we did highlight the presence of one gene involved in the bacterial SOS response, there are multiple genes involved which were not detected. Furthermore, the potential identification of the host as a cryptic species highlights the need for further populations of wild harvested fish to be characterised genetically, as unknowingly managing a species complex as one species may lead to severe threats to local cryptic species, in addition to masking potential disease risks in the cryptic species.

## Data availability

The raw fastq sequences generated during the MinION sequencing are available on NCBI GenBank under the accession number PRJNA785244, and the fasta sequences for the catfish *COI* sequences are available under the GenBank accession numbers OM176588—OM176591.

Received: 18 January 2022; Accepted: 20 July 2022

Published online: 16 August 2022

## References

- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020).
- Warren, C. J. & Sawyer, S. L. How host genetics dictates successful viral zoonosis. *PLoS Biol.* **17**, e3000217 (2019).
- McCoy, E. et al. Foodborne agents associated with the consumption of aquaculture catfish. *J. Food Prot.* **74**, 500–516 (2011).
- Mukwabi, D. M., Otieno, S. O., Okemo, P. O., Odour, R. O. & Agwanda, B. Parasites infesting Nile tilapia grown in aquaculture systems in Kenya. *Livest. Res. Rural Dev.* **31**, 2–6 (2019).
- Novoslavskij, A. et al. Major foodborne pathogens in fish and fish products: A review. *Ann. Microbiol.* **66**, 1–15 (2016).
- Allen, T. et al. Global hotspots and correlates of emerging zoonotic diseases. *Nat. Commun.* **8**, 1–10 (2017).
- Riede, K. Global register of migratory species: From global to regional scales: Final report of the R&D-Projekt 808 05 081. *Federal Agency for Nature Conservation*. (2004)

8. Risch, L. M. Claroteidae. In *The Fresh and Brackish Water Fishes of West Africa* (eds Paugy, D. *et al.*) 60–96 (Royal Museum for Central Africa, Tervuren, 2003).
9. Risch, L. M. Bagridae. In *Check-List of the Freshwater Fishes of Africa (CLOFFA)* (ed. Daget, J., Gosse, J. P., & Thys van den Audenaerde, D. F. E.) Vol. 2, 2–35. (ISNB, Brussels; MRAC, Tervuren; and ORSTOM, Paris, 1986)
10. Berg, R. W. & Anderson, A. W. Salmonellae and Edwardsiella tarda in gull feces: A source of contamination in fish processing plants. *Appl. Microbiol.* **24**, 501–503 (1972).
11. Mikaelian, I., Daignault, D., Duval, M. C. & Martineau, D. Salmonella infection in wild birds from Quebec. *Can. Vet. J.* **38**, 385 (1997).
12. Koonse, B., Burkhardt, W. III., Chirtel, S. & Hoskin, G. P. Salmonella and the sanitary quality of aquacultured shrimp. *J. Food Prot.* **68**, 2527–2532 (2005).
13. Jones, K. Campylobacters in water, sewage and the environment. *J. Appl. Microbiol.* **90**, 68S–79S (2001).
14. Wyatt, L. E., Nickelson, R. & Vanderzant, C. Occurrence and control of Salmonella in freshwater catfish. *J. Food Sci.* **44**, 1067–1073 (1979).
15. Lowry, P. W., McFarland, L. M. & Threefoot, H. K. Vibrio hollisae septicemia after consumption of catfish. *J. Infect. Dis.* **154**, 730–731 (1986).
16. Tuyet, D. T. N. *et al.* Enteropathogenic Escherichia coli O157 in Bangui and N’Goila, Central African Republic: A brief report. *Am. J. Trop. Med. Hyg.* **75**, 513–515 (2006).
17. Kotloff, K. L. *et al.* The incidence, aetiology, and adverse clinical consequences of less severe diarrhoeal episodes among infants and children residing in low-income and middle-income countries: a 12-month case-control study as a follow-on to the Global Enteric Multicenter Study (GEMS). *Lancet Glob. Health* **7**, e568–e584 (2019).
18. Shahin, K., Bouzari, M., Wang, R. & Khorasgani, M. R. Distribution of antimicrobial resistance genes and integrons among Shigella spp. isolated from water sources. *J. Glob. Antimicrob. Resist.* **19**, 122–128 (2019).
19. Kinge, C. W. & Mbewe, M. Characterisation of Shigella species isolated from river catchments in the North West province of South Africa. *S. Afr. J. Sci.* **106**, 1–4 (2010).
20. He, F. *et al.* Shigellosis outbreak associated with contaminated well water in a rural elementary school: Sichuan Province, China, June 7–16, 2009. *PLoS ONE*, **7**, e47239 (2012).
21. Austin, B. Methods for the diagnosis of bacterial fish diseases. *Mar. Life Sci. Technol.* **1**, 41–49 (2019).
22. Noone, J. C., Helmersen, K., Leegaard, T. M., Skråmm, I. & Aamot, H. V. Rapid diagnostics of orthopaedic-implant-associated infections using nanopore shotgun metagenomic sequencing on tissue biopsies. *Microorganisms* **9**, 97 (2021).
23. Yahara, K. *et al.* Long-read metagenomics using PromethION uncovers oral bacteriophages and their interaction with host bacteria. *Nat. Commun.* **12**, 1–12 (2021).
24. Leggett, R. M. *et al.* Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. *Nat. Microbiol.* **5**, 430–442 (2020).
25. Reddington, K. *et al.* Metagenomic analysis of planktonic riverine microbial consortia using nanopore sequencing reveals insight into river microbe taxonomy and function. *GigaScience* **9**, 1–12 (2020).
26. Seegers, L., De Vos, L. & Okeyo, D. O. Annotated checklist of the freshwater fishes of Kenya (excluding the lacustrine haplochromines from Lake Victoria). *J. East Afr. Nat. Hist.* **92**, 11–47 (2003).
27. Shabardina, V. *et al.* NanoPipe—a web server for nanopore MinION sequencing data analysis. *GigaScience* **8**, giy169 (2019).
28. Ivanova, N. V., Zemlak, T. S., Hanner, R. H. & Hebert, P. D. Universal primer cocktails for fish DNA barcoding. *Mol. Ecol. Notes* **7**, 544–548 (2007).
29. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
30. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A. & Jermin, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
31. Stamatakis, A., Hoover, P. & Rougemont, J. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* **57**, 758–771 (2008).
32. Pao, G. M. & Saier, M. H. Response regulators of bacterial signal transduction systems: Selective domain shuffling during evolution. *J. Mol. Evol.* **40**, 136–154 (1995).
33. Peart, C. R., Bills, R., Wilkinson, M. & Day, J. J. Nocturnal claroteine catfishes reveal dual colonisation but a single radiation in Lake Tanganyika. *Mol. Phylogenet. Evol.* **73**, 119–128 (2014).
34. Grütze, J. *et al.* Fishing in the soup—pathogen detection in food safety using metabarcoding and metagenomic sequencing. *Front. Microbiol.* **10**, 1805 (2019).
35. Taylor, T. L. *et al.* Rapid, multiplexed, whole genome and plasmid sequencing of foodborne pathogens using long-read nanopore technology. *Sci. Rep.* **9**, 1–11 (2019).
36. Sadek, M. *et al.* First genomic characterization of blaVIM-1 and mcr-9-coharbouring Enterobacter hormaechei isolated from food of animal origin. *Pathogens* **9**, 687 (2020).
37. Yang, M. *et al.* Direct metatranscriptome RNA-seq and multiplex RT-PCR amplicon sequencing on Nanopore MinION—promising strategies for multiplex identification of viable pathogens in food. *Front. Microbiol.* **11**, 514 (2020).
38. Schloss, P. D. & Handelsman, J. Metagenomics for studying unculturable microorganisms: Cutting the Gordian knot. *Genome Biol.* **6**, 1–4 (2005).
39. Mani, S., Wierzbza, T. & Walker, R. I. Status of vaccine research and development for Shigella. *Vaccine* **34**, 2887–2894 (2016).
40. Bowen, A. Chapter 3: Infectious diseases related to travel. In *The Yellow Book: Health Information for International Travel* (CDC, Florida, 2016).
41. Ashkenazi, S. Shigella infections in children: New insights. In *Seminars in Pediatric Infectious Diseases* 246–252 (WB Saunders, Philadelphia, 2004).
42. Miller, C., Ingmer, H., Thomsen, L. E., Skarstad, K. & Cohen, S. N. DpiA binding to the replication origin of Escherichia coli plasmids and chromosomes destabilizes plasmid inheritance and induces the bacterial SOS response. *J. Bacteriol.* **185**, 6025–6031 (2003).
43. Žgur-Bertok, D. DNA damage repair and bacterial pathogens. *PLoS Pathog.* **9**, e1003711 (2013).
44. Memar, M. Y. *et al.* The central role of the SOS DNA repair system in antibiotics resistance: A new target for a new infectious treatment strategy. *Life Sci.* **262**, 118562 (2020).
45. Shi, R. *et al.* Pathogenicity of Shigella in chickens. *PLoS ONE* **9**, e100264 (2014).
46. Jianjun, J. Isolation and identification of rabbits Shigella dysenteriae in a large scale warren. *J. Anhui Agric. Sci.* **33**, 1666 (2005).
47. Zhu, Z., Cao, M., Zhou, X., Li, B. & Zhang, J. Epidemic characterization and molecular genotyping of Shigella flexneri isolated from calves with diarrhea in Northwest China. *Antimicrob. Resist. Infect. Control* **6**, 1–11 (2017).
48. Maurelli, A. T. *et al.* Shigella infection as observed in the experimentally inoculated domestic pig, Sus scrofa domestica. *Microbial Pathog.* **25**, 189–196 (1998).
49. Dudley, J. P. *et al.* Carnivory in the common hippopotamus Hippopotamus amphibius: Implications for the ecology and epidemiology of anthrax in African landscapes. *Mamm. Rev.* **46**, 191–203 (2016).
50. Murphree, R., Warkentin, J. V., Dunn, J. R., Schaffner, W. & Jones, T. F. Elephant-to-human transmission of tuberculosis, 2009. *Emerg. Infect. Dis.* **17**, 366 (2011).
51. Kapetsky, J. M., & Petr, T. Status of African reservoir fisheries. *FAO* (1984).

52. Eccles, D. H. FAO species identification sheets for fishery purposes. In *Field Guide to the Freshwater Fishes of Tanzania* (FAO, Rome, 1992).
53. Nwani, C. D. *et al.* DNA barcoding discriminates freshwater fishes from southeastern Nigeria and provides river system-level phylogeographic resolution within some species. *Mitochondrial DNA* **22**, 43–51 (2011).
54. Okeyo, D. O. Updating names, distribution and ecology of riverine fish of Kenya in the Athi-Galana-Sabaki River drainage system, Naga. *ICLARM Q.* **21**, 44–53 (1998).

### Acknowledgements

This research was funded under the Marine Institute Cullen fellowship (grant number: CF/17/04/01). The project was conducted in collaboration with Pwani University, Kenya, within the MOA between Pwani University and University College Dublin, Ireland. All samples were collected and transferred to the laboratories in UCD Ireland in accordance with local laws (S029/2019). Special thanks to Friends from Ireland for facilitating the fieldwork, to the Galana Wildlife Conservancy for allowing us to collect samples and to Kenya Wildlife Service for overall watch in the field. Thanks also to Dr Graham Hughes and Dr Edward Feil who gave valuable feedback on this study when it was a PhD chapter. We thank also the other members of the Area 52 Research Group and the members of the Fish Health Unit in the Marine Institute.

### Author contributions

A.J.T., B.F., N.M.R. and J.C. devised the study. A.J.T., J.B., S.H., S.G., L.J., J.D., L.B. and M.K.Q performed the field work. B.F. facilitated the field work and sampling. A.J.T. and S.G. performed the lab work and analysis, with additional support from B.F., N.M.R. and J.C. A.J.T drafted the manuscript, figures and tables. All authors contributed to writing and improving each manuscript draft and all approved the final submitted draft.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to A.J.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022