Faecal source tracking to understand the role of introduced predators and avian species on water quality assessments in the Mākirikiri Reserve, Dannevirke

Marie Moinet and Adrian Cookson

June 2021

Report for National Science Challenges – Our Land and Water

Contract Number: A26518

Output ID Number: 1798

agresearch
āta mātai, mātai whetū

Inquiries or requests to:

Marie Moinet
Marie.Moinet@agresearch.co.nz
Consumer Interface Innovation Centre of Excellence,
AgResearch Ltd, Hopkirk Research Institute,
Cnr University Ave & Library Rd, Private Bag 11008, Palmerston North, New Zealand

Dr Gale Brightwell
Science Team Leader
Food System Integrity Team
Consumer Interface Innovation Centre of Excellence

Contents

1. Executive Summary

Escherichia coli are used as indicators of faecal contamination in water quality assessments, but "naturalised" non-faecal *E. coli* and non-pathogenic *E. coli*-like bacteria can confound routine measurements. Contrary to livestock and human activities, wildlife has not thoroughly been studied as a source of faecal contamination in waterways but could contribute too. We examined the profile of *E. coli* populations in gut contents and faeces from bird and introduced predator species in the Mākirikiri Reserve, Dannevirke, and compared to water, soil, sediment and biofilm samples taken within the reserve to determine whether avian and/or invasive mammal species contribute faecal bacteria in the environment.

E. coli (n=420) were recovered from animal and environmental samples (n=106). Initial characterisation of *E. coli* was using real time PCR targeting the *uid*A gene, and a subset were further typed by sequencing a region of the hypervariable *gnd* gene to generate a specific *gnd* sequence type (gST). These data informed which isolates underwent further phylogenetic analysis using whole genome sequencing. *E. coli* populations from sample enrichments were analysed using metabarcoding and *gnd* amplicon sequencing.

Analyses showed a significant difference in diversity of *E. coli* gSTs between samples of animal and environmental source, with a lower diversity in animal samples, but also that some gSTs were present in both sample types. Two gSTs (535 and 258) were present in 85 and 71% of samples, respectively. Cryptic *Escherichia* species, phenotypically similar to *E. coli*, were isolated and detected by metabarcoding, but only in low abundance.

These data support the hypothesis that introduced predators and avian species contribute to faecal contamination of the environment. The data gathered should allow future comparison of isolates with strains isolated from wildlife.

2. Background

2.1 Context

Escherichia coli are bacteria that live in the gut of warm-blooded animals and birds. When *E. coli* are found in waterways, their presence suggests that faecal contamination has occurred. Regional authorities have a responsibility to undertake water quality assessments to ensure that recreational users and Māori are aware of the risks of potential illness associated with microbes from faecal contamination during swimming or mahinga kai, the harvesting of, for example, eels/tuna or kōura/crayfish. However, recent work has identified *E. coli*-like bacteria that are not able to be distinguished from faecal *E. coli* using the standard water quality monitoring tests (Walk et al. 2009). These 'naturalised' *E. coli*-like bacteria grow and multiply in soil, water and sediment, but are rarely found in faeces (Clermont et al. 2011). Recent discoveries of these 'naturalised' *E. coli*-like bacteria in New Zealand's waterways suggests that the current water quality monitoring test results may cause some monitoring sites to fail when there is no, or very little, faecal contamination (Devane et al. 2020).

E. coli from human and domestic sources as a cause of faecal contamination in the environment has been extensively studied in NZ, but wild animals may also contribute to faecal contamination of waterways. Very few studies have looked at *E. coli* present in NZ wildlife (Devane et al. 2019), and none have undertaken detailed characterization of the different *E. coli* strains found (Sumner et al. 1977; Murphy et al. 2005; Moriarty et al. 2011; Phiri 2015). Genetic markers were developed to detect gull, Canada goose, and duck faecal contamination in water (Green et al. 2012) but data is lacking on invasive mammals, that can be present at high densities and could also contribute to misinterpreting water quality monitoring test results.

In a parent project (Faecal source tracking and the identification of naturalised *Escherichia coli* to assist with establishing water quality and faecal contamination levels), early partnership with Rangitāne O Tamaki nui a Rua and Te Kāuru (Eastern Manawatu River hapū collective) provided an opportunity for site identification and sample collection from the Mākirikiri Stream and Mākirikiri Reserve, Dannevirke, for water quality assessment and faecal source tracking purposes. WGS and phylogenomic analysis of *E. coli* from the Mākirikiri area revealed a clonal type not identified elsewhere. In total, six separate *E. coli* isolates were observed to be no more than 6 nucleotide polymorphisms different from each other and were collected from water, wildfowl faeces, unknown avian faeces, possum faeces (twice), and rat faeces (unpublished data).

2.2 Aims

This proposed extension study aims to examine faecal material from bird and introduced predator species in more detail by examining the profile of *E. coli* populations in these faecal specimens to determine whether avian and/or invasive mammal species contribute faecal bacteria, such as *E. coli* and *E. coli*-like naturalised *Escherichia* species, that impact water quality assessments.

Objectives:

- Whole genome sequencing (WGS) and phylogenomic analysis of at least 75 different *Escherichia coli* and *E. coli*-like naturalised *Escherichia* species isolated from the guts of introduced mammal and faeces of birds
- *E. coli* community profiling using culture-independent methods for at least 30 mammals of a panel of different species
- Comparison with *E. coli* isolates and community profiles obtained from environmental samples collected as part of our current SSIF programme (A Kaupapa approach to food safety)

3. Methods

This study was conducted following the Animal Welfare Act 1999 and the protocol was approved by AgResearch Animal Ethics Committee (AE 15061). Approval for the work to occur in the Mākirikiri Reserve was provided by Tararua District Council (current owner of the land) in conjunction with iwi consultation (Rangitāne o Tamaki nui-a-Rua, and Ngāti Kahungunu.

3.1 Study site and fieldwork

The Mākirikiri Reserve is a 15ha remnant of Seventy Mile Bush which formerly spanned from Wairarapa in the south through to southern Hawkes Bay in the north. This bush remnant is of significant cultural significance and is part of the Rangitane o Tamaki nui-a-Rua Treaty Settlement package 2016.

Animal trapping

Four different types of kill-traps were deployed into the Mākirikiri Reserve; 14 Timms traps, 2 Trapinator traps, and 4 A12 GoodNature (mainly targeting possums), and 13 A24 GoodNature traps (targeting rodents, hedgehogs and small mustelids). Traplines were established in the bush and along the Mākirikiri Stream (***Figure 1***). These specific traps were selected as the most appropriate for humane killing while allowing for a reasonably good recovery rate and relatively intact carcasses for *E. coli* isolation (*i.e.* limited damage of distal end of gastro-intestinal tract (GIT) and intact gut contents). Timms traps were baited with apple and a mix of cinnamon and sugar; Trapinator and A12s with a commercial Possum lure; and A24s with a commercial Rat Lure. Traps were monitored at least once every 3 days by local hapū. The number of captures per 100 trap-nights was calculated according to (Cunningham et al. 1996). Carcasses were collected over a period of 2 months, from November 2020 to early January 2021. Traps were subsequently donated to Te Kāuru and local hapū to continue pest control in the Mākirikiri Reserve. Dead animals were collected and stored at +4°C before being brought to the Hopkirk Research Institute for post-mortem sampling of gut contents.

Sample collection

Samples were taken from:

a. the gut contents of introduced predators (hedgehogs, rodents, possums and mustelids) recovered from the traps;
b. environmental samples (water, sediment, soil and periphyton), collected at two different sampling points, along the Mākirikiri stream ('Mākirikiri') and just before the confluence of the Mākirikiri stream with the Mangatera river ('Confluence');
c. avian/mammalian faecal samples found opportunistically in the environment during sampling visits (2 avian and 2 mammalian per sampling site and visit).

These sampling visits for environmental sampling occurred on 7 and 17 November and 4 and 18 December 2020. Of the 16 faeces samples, 9 were of avian origin, and 7 from possums. To prevent confusion with possum gut contents, these possum faeces collected from the environment were described as mammal faeces hereinafter.

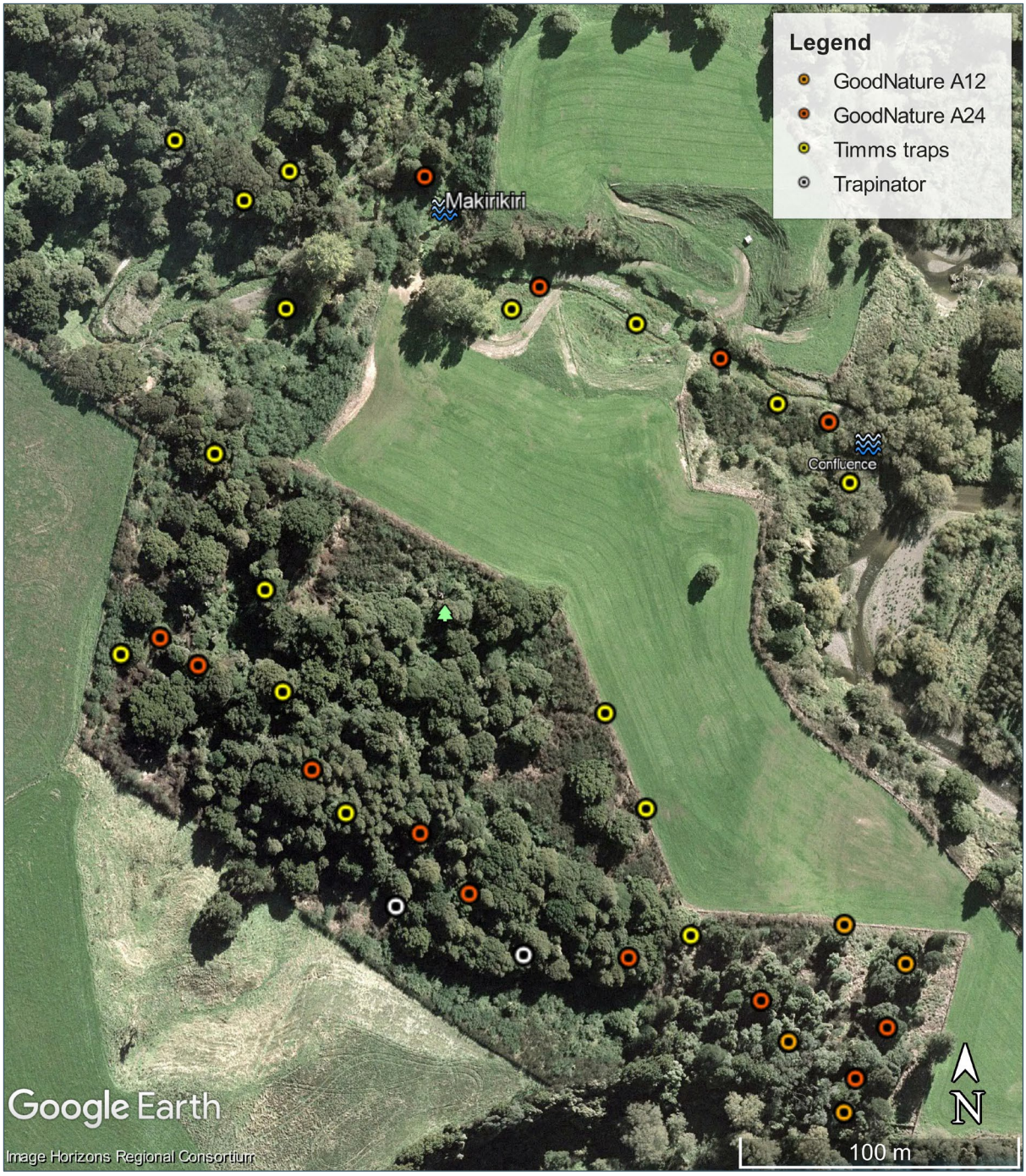

Figure 1: *Trap line within the Mākirikiri reserve, Dannevirke, and localisation of the two environmental sampling points, along the Mākirikiri Stream and just before the confluence with the Mangatera River.*

3.2 Laboratory and data analyses

E. coli enumeration

At each sampling site, 100 mL of water were used for enumeration of coliforms and *E. coli* using the Colilert® Quanti-Tray 2000® method (IDEXX, New Zealand). Also, 35 mL of EC broth was added to the dry sponge swab (periphyton) and stomached for 60s. A 1:10 dilution of this mix (10 mL into 90 mL of sterile RO water) was used for enumeration of coliforms and *E. coli*.

E. coli enrichment and sample-level DNA extraction

Enrichments were made for water, sediment and soil, biofilm (periphyton), faeces and GIT contents. For water, 100 mL was filtered through a 45 µm membrane filter and added to EC broth (10 mL). For sediment and soil samples, 1 g was added to EC broth (9 mL). For biofilm samples the stomached dry sponge swab and EC broth mix (10 mL) was used. For faeces and GIT contents, an Amies swab or a weighed sample was mixed into EC broth (swab into 10 mL or 250 mg into 24.75 mL). After 18-21h, 1 mL of enrichment (stationary incubation at 35°C) was first centrifuged at 13000g for 1min, the supernatant discarded and the pellet resuspended in 1 mL of sterile PBS, then centrifuged again, the supernatant discarded, and the pellet resuspended in 1 mL of MQ water and incubated at 100°C for 10 min before the boiled lysate to be stored at -20°C. From the same enrichment 1 mL was mixed with 450 µL glycerol and stored at -80°C.

E. coli isolation and isolate-level DNA extraction

For each of the different sample types, 10 µL of enrichment were plated onto CHROMagar™ ECC plates — a medium selective for *E. coli* (growing blue) and coliforms (growing mauve) — and incubated overnight at 35°C. After incubation up to 4 well-spaced blue colonies were sub-cultured onto MacConkey agar plates and incubated at 35°C for 18h. For each of the 4 subcultures, one or two well-spaced colonies were removed from the purity plate and resuspended in 400 µL MQ water, briefly vortexed and incubated at 100°C for 10 min before the boiled lysate to be stored at -20°C. Also, for glycerol stock, 6 to 8 well-spaced colonies were removed from a purity plate, resuspended in a cryovial containing 1000 µL BHI (Brain Heart Infusion) broth plus 450 µL glycerol, and stored at -80°C.

uidA qPCR

A real-time quantitative PCR targeting the *E. coli*- and *Shigella*-specific β-glucuronidase gene *uidA* was conducted on DNA extracted from each isolate to distinguish *E. coli* and *Shigella* bacteria from *E. coli*-like bacteria (cryptic clades of *Escherichia*). Primers used and PCR conditions were as in (Anklam et al. 2012). Reactions with a Ct ≥ 35 were deemed negative and associated isolates were putatively considered as cryptic clades pending confirmation by WGS.

Isolate *gnd* sequence typing

Two isolates per sample were submitted to *gnd* PCR and amplicon sequencing. PCR was performed with KAPA HiFi HotStart ReadyMix in a total volume of 20 µL using primers 2*gnd*F 5′-TCYATYATGCCWGGYGGVCAGAAAGAAG (*gnd* coordinates 415 to 442) and 2*gnd*R 5′-CATCAACCARGTAKTTACCSTCTTCATC (*gnd* coordinates 754 to 726) at a final concentration of 0.3 µM and 1 µL of isolate boiled lysate as template. The PCR cycle consists of a single denaturing step at 95°C for 3 minutes followed by 30 cycles of 98°C for 20 seconds, 63°C for 30 seconds and 72°C for 30 seconds and a final elongation step of 72°C for 5 minutes followed by a 12°C hold step using a T100™ Thermal Cycler (Bio-Rad, Auckland, New Zealand). To ensure there was a major PCR product at approximately 340 bp, 2 µL of amplicons were run on a 2% (w/v) agarose gel stained with RedSafe (Custom Science, Auckland, New Zealand) for 40min at 90V, and subsequently visualised using UV illumination in Gel Doc 1000 System (Bio-Rad Laboratories Inc., Hercules, CA, USA). Remaining amplicons were purified using the QIAquick PCR purification kit (Qiagen, Bio-strategy, Auckland, New Zealand) and their DNA concentration quantified by spectrophotometry using NanoDrop (ThermoFisher Scientific, Auckland, New Zealand). Samples had concentration adjusted to 5 ng/nL and were subsequently submitted to Sanger sequencing at Macrogen (Seoul, South Korea).

gnd metabarcoding

Metabarcoding targeting the partial *gnd* allele amplicon generated from 0.8 mL extracted DNA from each separate enrichment was performed as described previously (Cookson et al. 2017) with subsequent analysis of MiSeq illumina sequence reads using the packages dada2 (Callahan et al. 2016) and phyloseq (McMurdie et al. 2013) in R version 4.0.4 (2021-02-15) and using the gndDb database as reference (Cookson et al. 2019) for gST assignment. A Student's T-test was used to assess the difference in alpha-diversity between groups.

DNA purification, library preparation and whole genome sequencing

DNA extractions and library preparations for whole genome sequencing (WGS) was undertaken using standard methods. WGS was undertaken by Novagene Limited (Beijing, China) using the Illumina HiSeq paired-end v4 platform (2 x 125 bp).

4. Results and Discussion

4.1 Trapping

There were 1523.5 corrected trap-nights (successful traps and traps sprung but empty were counted as set for half of the associated nights), and 3.8 captures per 100 trap-nights. Overall, 49 possums, 5 ship rats, 3 hedgehogs and one ferret were captured and sampled.

4.2 E. coli enumeration

E coli counts obtained from water and biofilm samples are presented in **Figure 2**. The geometric mean *E. coli* MPN were 317 [95% CI 285 – 352] at Mākirikiri sampling point and 487 [326 – 727] at the Confluence in water, and 423 [300 – 596] at Mākirikiri sampling point and 562 [198 – 1596] at the Confluence in biofilm samples.

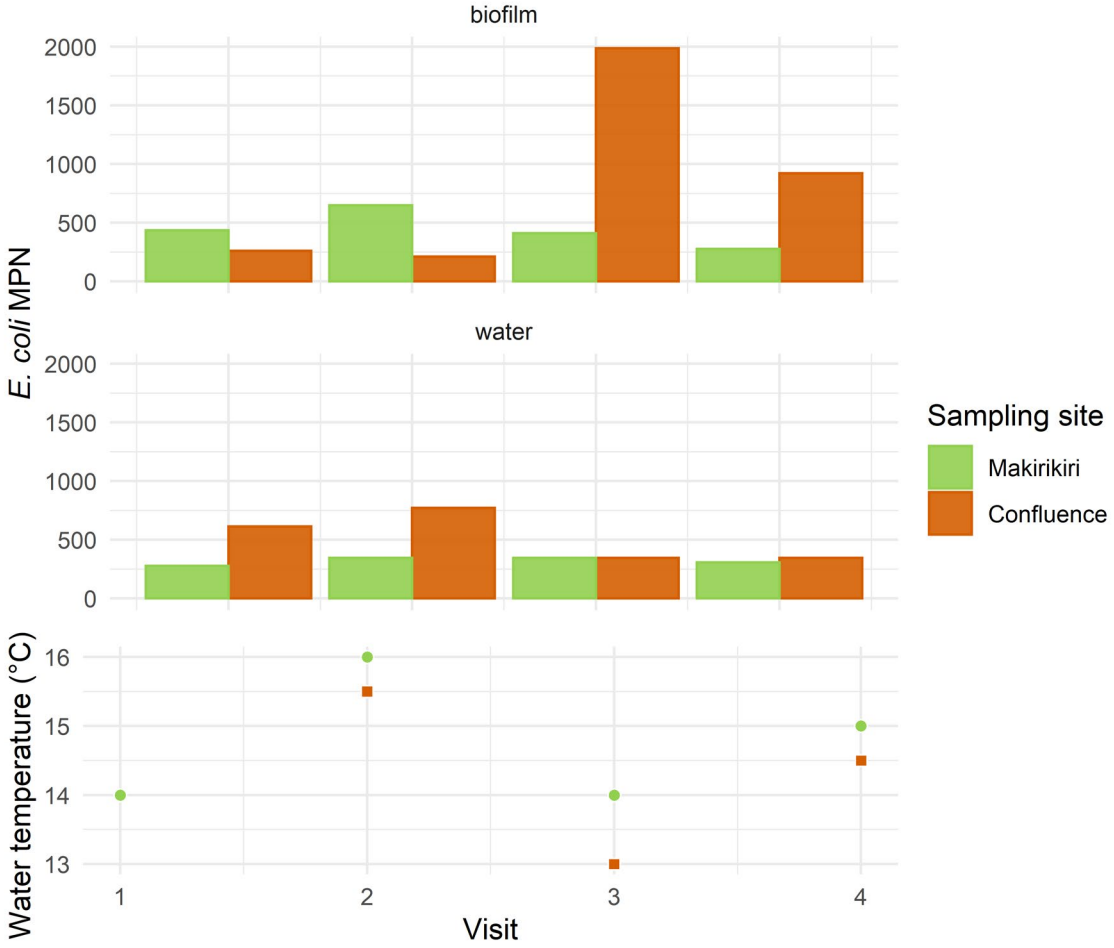

Figure 2: *Most probable number of Escherichia coli per 100mL of water / per cm² of biofilm on stone and water temperature at the two sampling sites and four visit occasions.*

4.3 E coli isolates diversity

A total of 420 isolates of *E. coli* were obtained by culture-based methods. Of them, 207 were further analysed using *gnd* sequencing to determine their respective gSTs, and 102 for WGS (Table 1).

Table 1: *Number of isolates obtained, gnd-sequence typed and submitted to Whole genome Sequencing (WGS) per sample type*

Sample type[1]	Isolated[2]	*gnd* PCR[3]	WGS
Water	32	16	7
Sediment	32	16	5
Soil	28	14	6
Biofilm	32	15	5
Faeces	64	32	13
Gut content	232	100	66

[1] Faeces were found in the environment, gut content was sampled in trapped animals during necropsy; [2] 4 isolates per sample, except one soil sample for which no isolates were obtained; [3] up to 2 isolates tested per sample.

The *uidA* gene was not detected by PCR in 32/420 isolates: 13 isolates from soil (from 4 different samples), 10 from gut contents (from 5 different samples), 5 from biofilm (from 2 different samples) and 4 from faeces samples (from 1 faecal sample). Among them, 18 had a gST assigned and 8 different gSTs were assigned: gST541 (from 1 soil and 2 different possums gut contents), gST587 (from a soil and a biofilm sample), gST537 (from a biofilm and a ship rat gut content), gST540 and gST543 (both from soil samples), gST258 and gST591 (both from possum gut contents), and gST548 (from an avian faeces sample).

Out of the 143 isolates from animal sources that underwent *gnd* amplification and sequencing, 36 different gSTs were assigned. The most frequent were gST258 and gST535 (38 and 30 isolates, 47.6% of animal isolates).

Out of 64 isolates from environmental samples, 41 different gSTs were assigned. The three most frequent gSTs were gST152 (5 isolates, 7.8% of environmental isolates), gST587 and gST535 (respectively 4 isolates, 6.3% of environmental isolates).

The gST258 was isolated from 1 sediment, 3 different avian faecal samples, 22 different possums, 1 ship rat, 1 hedgehog and 1 ferret. gST535 was detected in 2 different water samples and 1 soil sample (at both sampling sites but 3 different occasions), and in 3 mammal faecal samples, 17 possums and 1 ship rat.

WGS was undertaken on 102 isolates. To provide information on the potential role of mammalian pests on environmental contamination, highly abundant *E. coli* gST258 and gST535 were targeted for high resolution phylogenetic analysis, including 30 gST258 found in possum, ship rat, hedgehog, ferret, avian faeces and sediment samples, 24 gST535 found in possum, ship rat, mammal faeces, soil and water samples, 13 cryptic *Escherichia* clade isolates found in possum, ship rat, avian faeces, biofilm and soil samples and other gSTs found in several different sample types. At the date of writing of the present report, WGS data has not yet been received and is therefore not presented.

4.4 *E. coli* community composition

After trimming and filtering, a total of 3,307,917 reads and 580 unique amplicon sequence variants (ASV) were obtained from the 106 samples of water, sediment, soil, biofilm, faeces, and gut contents. Almost half ASVs (283/580, 49%) could be assigned to existing gSTs already present in the gndDb database (Cookson et al. 2019).

Reads obtained from the gut contents of possums represented 33% of the total, and reads from samples of animal source (*i.e.* gut contents and faeces) represented 61.8% of the reads. Biofilms had the highest number of ASVs assigned on average, and gut contents the lowest (Table 2).

Table 2: *Number of reads of* gnd *sequences obtained by metabarcoding per sample type, percentage of the total number of reads and mean number of gST per sample*

Sample type	Reads	% of total	mean number of gSTs[1] +/- SD
Water	391,466	11.8%	87.5 +/- 14
Sediment	248,259	7.5%	55.6 +/- 31
Soil	305,236	9.2%	13.5 +/- 6.5
Biofilm	318,629	9.6%	131.5 +/- 34.9
Faeces	519,372	15.7%	16.2 +/- 11.3
Avian	*335,45*	*10.1%*	*21.4 +/- 10.3*
Mammal	*183,936*	*5.6%*	*9.4 +/- 9.1*
Gut content	1,524,955	46.1%	7.1 +/- 8.8

[1] gSTs with less than 10 reads not included

Despite the higher number of reads obtained from possums compared to other sample types, the alpha diversity was higher in environmental samples than in samples of animal source (Figure 3). The Simpson and Shannon diversity indices were both significantly lower in faeces and gut samples than in environmental samples (P-values <0.0001).

Overall, gST535 and gST258 were the most frequently encountered gSTs in all type of samples (Table 3). The former was detected (with ≥10 reads) in 85% (90/106) of samples and the latter in 71% (75/106). A graph showing the repartition of the 20 most abundant gSTs (in terms of number of reads) is available in Appendix.

Table 3: *Number (Percentage) of samples with at least 10 reads assigned to each of the most frequent gSTs encountered in samples of animal or environmental source.*

Most frequent gSTs	in Animal samples	in Environmental samples	Overall
gST535	62 (84%)	28 (88%)	90 (85%)
gST258	55 (74%)	20 (63%)	75 (71%)
gST522	12 (16%)	26 (81%)	38 (36%)
gST152	11 (15%)	22 (69%)	33 (31%)
gST514	11 (15%)	22 (69%)	33 (31%)
gST308	14 (19%)	17 (53%)	31 (29%)
gST587	13 (18%)	15 (47%)	28 (26%)
gST231	5 (7%)	22 (69%)	27 (25%)

Samples with less than 10 reads for a given gST were not included in the calculation.

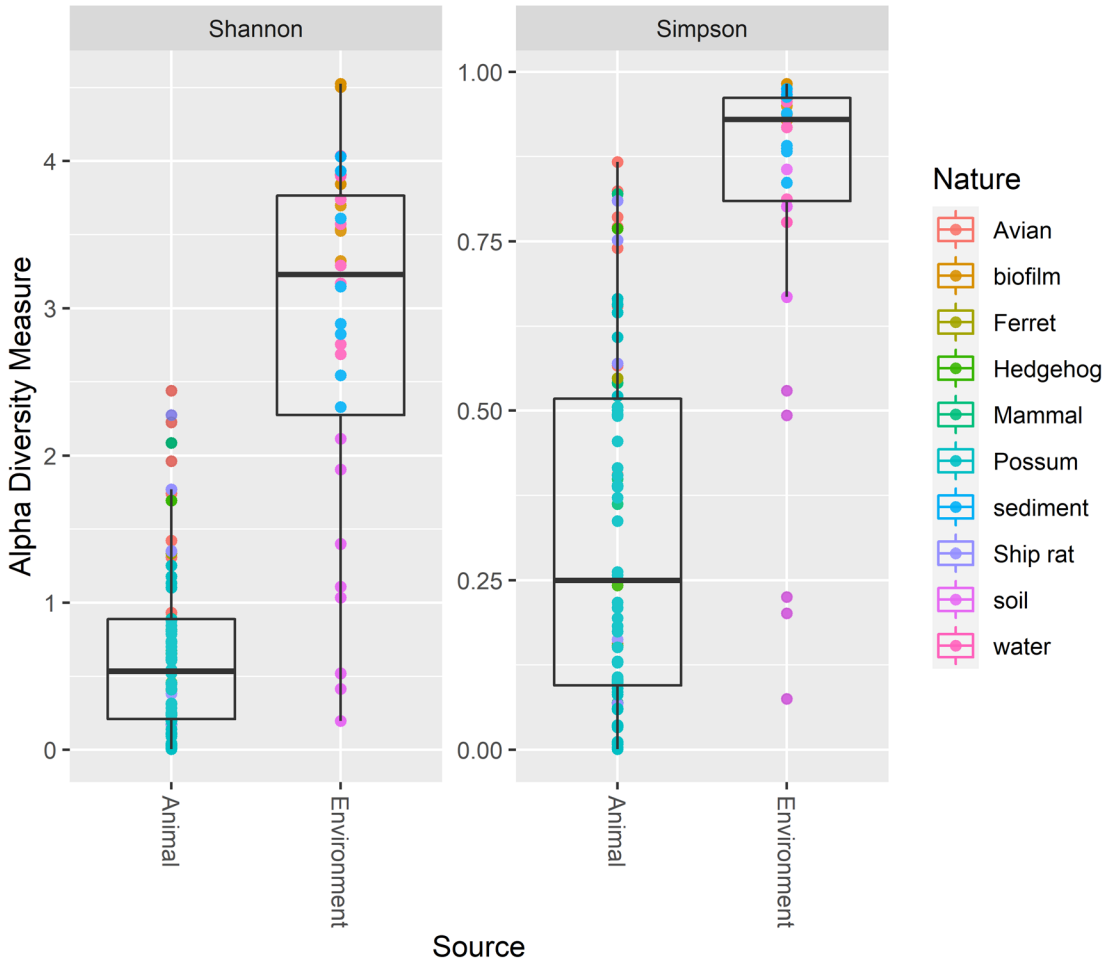

Figure 3: *Alpha diversity of environment and animal samples. Alpha diversity was measured by Shannon and Simpson, diversity indexes. Statistical testing showed a difference for both measures (P-values <0.0001)*

The communities of *E. coli* associated with animal samples and soil were similar, whereas the communities associated with water, sediment, and biofilms showed important variation, but were not distinct from each other (Figure 4).

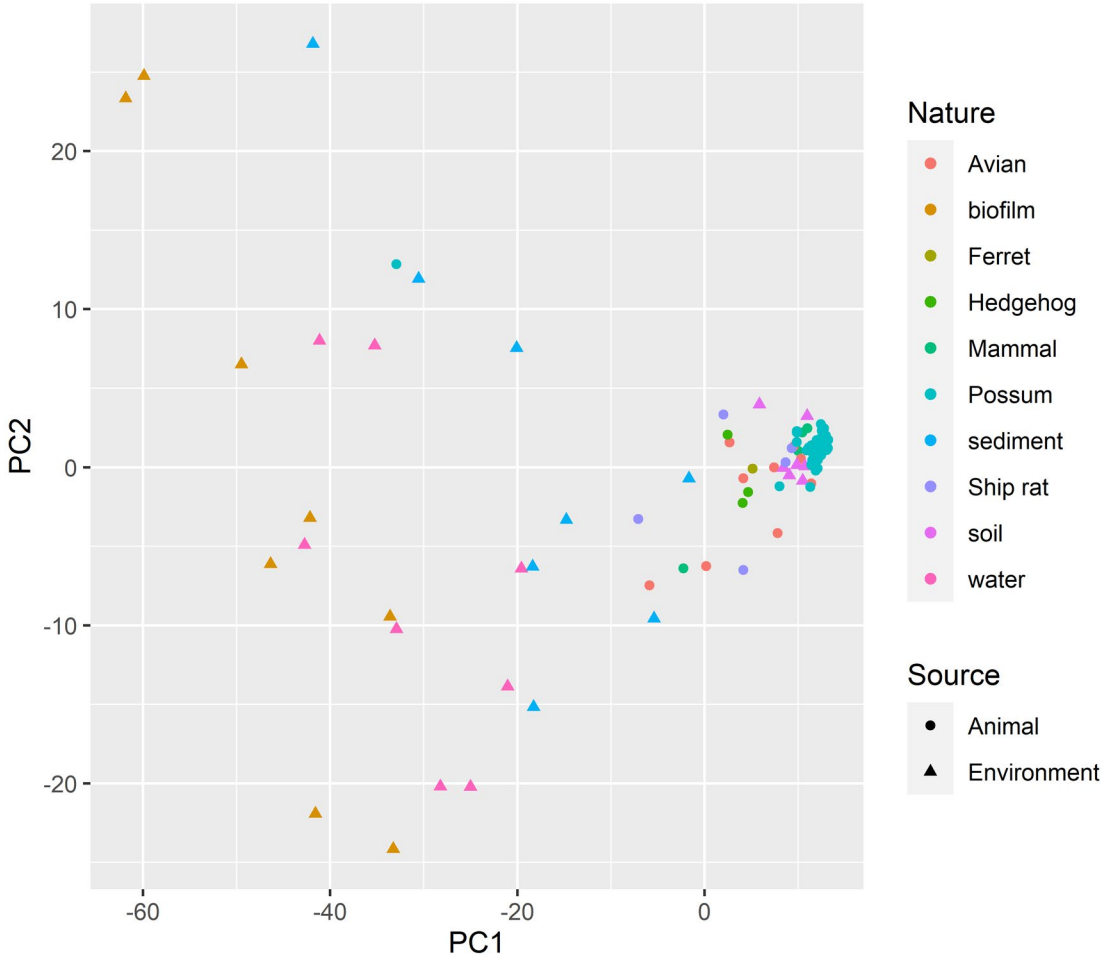

Figure 4: *Principal Component Analysis of* gnd *metagenomic data of* Escherichia coli *in animal and environmental samples from the Mākirkiri reserve, Dannevirke. Each data point represents an individual sample.*

5. Discussion

The present study investigated and compared *E. coli* isolates and community profiles obtained from environmental and faecal samples or gut contents of introduced predators and avian species living in the vicinity. Our aim was to determine whether introduced predators and/or avian species are a source of *E. coli* and naturalised *Escherichia* species.

All water samples taken had counts greater than 260 MPN of *E. coli*/100 mL and two were greater than 550 *E. coli*/100mL. Although the number of samples was not sufficient to provide reasonable statistical power in testing for compliance with water quality standards (minimum of 20 samples) (Ministry for the Environment 2003), they indicated that, at least at the time of sampling, the Mākirikiri Stream was exposed to faecal contamination.

The two methods used to investigate the composition of *E. coli* communities represent an imperfect image of the initial composition of *E. coli* communities in the samples. The composition of *E. coli* communities in different samples was assessed at the sample level, using metabarcoding and Illumina MiSeq, and, at the isolate level, using Sanger sequencing on up to 2 isolates per sample. Both methods targeted the same sequence of the *gnd* gene and used similar primers. The enrichment step that was used before metabarcoding and isolation can favour the growth of some strains over others and confound the results. Similarly, the *gnd* PCR amplification and sequencing reactions rely on a set of degenerated primers (Cookson et al. 2017) that may amplify more efficiently some gSTs over others, therefore confounding the results in terms of reads abundance (Appendix). For this reason, results were mostly interpreted in terms of presence/absence of a gST across all samples.

Despite this limitation, both methods led to similar observations. The community profiles of *E. coli* as measured by *gnd* metabarcoding revealed a significantly higher diversity of *E. coli* gSTs in environmental samples *vs.* animal samples, and a difference in community composition, but also that some gSTs were shared between animal and environmental samples. The same observation was noted with the variety of isolates obtained from both type of samples. These results support the hypothesis that introduced predators and avian species contribute at least in part to faecal contamination of the environment.

The *uidA* PCR allowed for the detection of 12/106 samples with putative cryptic clades isolates. One of them was typed as gST258 (a sequence type not described in a cryptic clade) and it is likely this sample was a false negative. The other *uidA*-negative isolates with a gST available all had a sequence type previously described in a clade V *Escherichia*. They were isolated from a variety of samples of soil, biofilm, possum, rat gut contents and avian faeces. The same gSTs were found with metabarcoding, but in low abundance (in terms of reads), and in only a small proportion of samples, especially for samples of animal source. Interestingly, despite reads obtained from avian faeces representing only 16% of the reads from an animal source, they represented 85% of the reads attributed to gSTs previously described in cryptic clades (data not shown). The contribution of invasive mammals to *E. coli*-like naturalised *Escherichia* species in the environment appears limited from the data gathered in this study, but the results in bird faeces warrants further investigations on the role of birds as a source of *E. coli*-like naturalised *Escherichia* species.

6. Acknowledgements

We would like to thank Te Kāuru Eastern Manawatu River Hapū Collective and local hapū involved in the trap monitoring and carcass removal, and Tararua District Council staff and Rose Collis (PhD student, AgResearch) that helped transport the carcasses to Hopkirk Research Institute.

We also thank Darren Peters (GoodNature) and Lisa Whittle (Ruahine Whio Collective) for their advice and help with trap placement and James C. Russell (University of Auckland) and Bruce Warbuton (Landcare Research) for their advice on the choice of traps.

Many thanks to Richard Muirhead (AgResearch) and our ESR, NIWA and Massey University partners for their helpful input into this work, especially Jonathan Marshall for help with Principal Component Analysis R code. Overall funding for this work was through Our Land & Water National Science Challenge Extension Funding.

Picture on cover from Andrew Mercer (www.baldwhiteguy.co.nz), CC BY-SA 4.0, via Wikimedia Commons.

7. Appendix

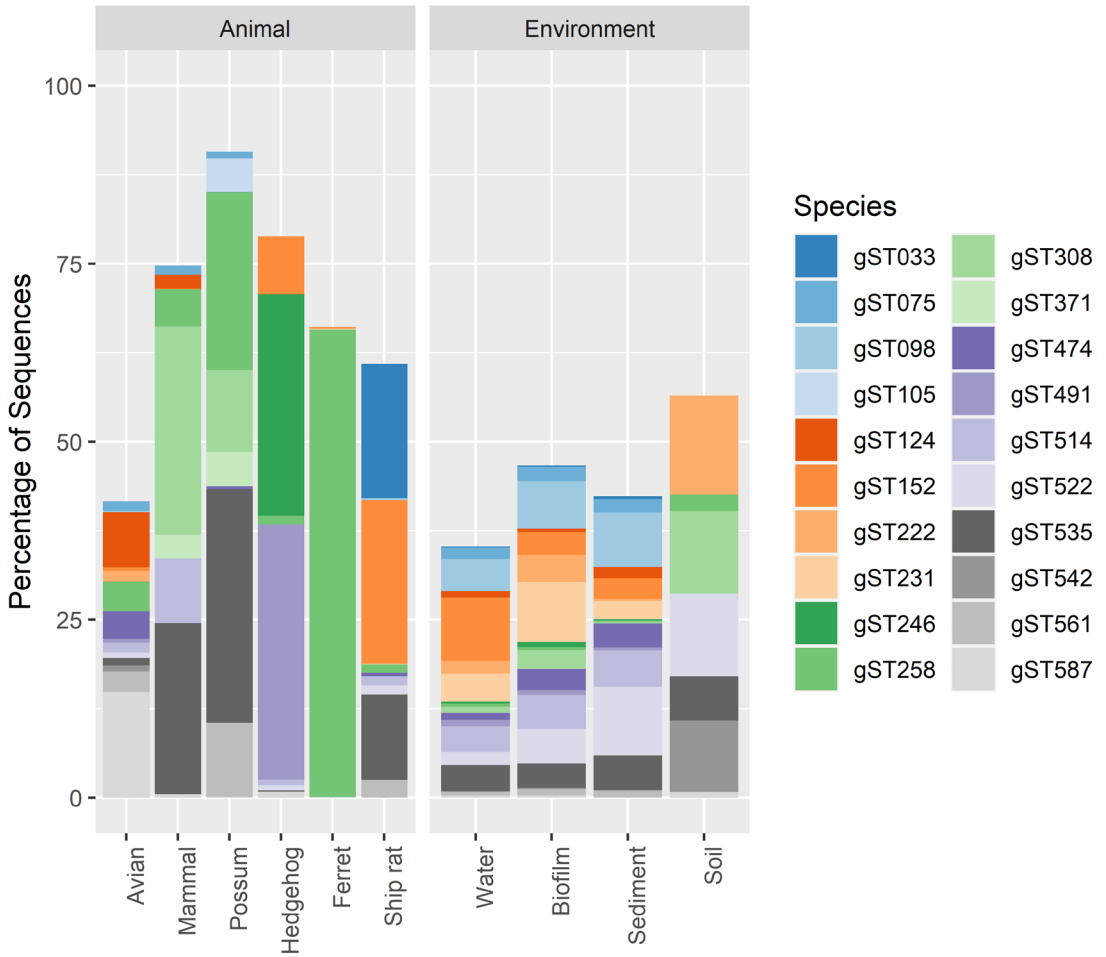

Figure 5: Composition of Escherichia coli communities associated with animal and/or environmental sources sampled at the Mākirikiri reserve, Dannevirke. Only the 20 most abundant gSTs are shown.

8. References

Anklam KS, Kanankege KS, Gonzales TK, Kaspar CW, Döpfer D 2012. Rapid and reliable detection of Shiga toxin-producing *Escherichia coli* by real-time multiplex PCR. J Food Prot 75: 643-50.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13: 581-583.

Clermont O, Gordon DM, Brisse S, Walk ST, Denamur E 2011. Characterization of the cryptic *Escherichia* lineages: rapid identification and prevalence. Environ Microbiol 13: 2468-77.

Cookson AL, Biggs PJ, Marshall JCet al. 2017. Culture independent analysis using *gnd* as a target gene to assess *Escherichia coli* diversity and community structure. Sci Rep 7: 841.

Cookson AL, Lacher DW, Scheutz Fet al. 2019. gndDb, a Database of Partial *gnd* Sequences To Assist with Analysis of *Escherichia coli* Communities Using High-Throughput Sequencing. Microbiol Resour Announc 8.

Cunningham DM, Moors PJ 1996. Guide to the identification and collection of New Zealand rodents. 3rd ed. Wellington, N.Z., Department of Conservation. 23 p.

Devane ML, Gilpin B, Moriarty E 2019. The sources of “natural” microorganisms in streams. The Institute of Environmental Science and Research Ltd., Christchurch, New Zealand.

Devane ML, Moriarty E, Weaver L, Cookson A, Gilpin B 2020. Fecal indicator bacteria from environmental sources; strategies for identification to improve water quality monitoring. Water Research 185: 116204.

Green HC, Dick LK, Gilpin B, Samadpour M, Field KG 2012. Genetic markers for rapid PCR-based identification of gull, Canada goose, duck, and chicken fecal contamination in water. Applied and Environmental Microbiology 78: 503-510.

McMurdie PJ, Holmes S 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8: e61217.

Ministry for the Environment 2003. Microbiological water quality guidelines for marine and freshwater recreational areas. Ministry for the Environment (MfE) and Ministry of Health (MoH), Wellington, New Zealand. ISBN: 0478240910. 155 p.

Moriarty EM, Karki N, MacKenzie M, Sinton LW, Wood DR, Gilpin BJ 2011. Faecal indicators and pathogens in selected New Zealand waterfowl. New Zealand Journal of Marine and Freshwater Research 45: 679-688.

Murphy J, Devane ML, Robson B, Gilpin BJ 2005. Genotypic characterization of bacteria cultured from duck faeces. J Appl Microbiol 99: 301-9.

Phiri BJ 2015. Estimating the public health risk associated with drinking water in New Zealand. Unpublished Master's thesis thesis, Massey University. xvi, 244 pages p.

Sumner JL, Perry IR, Reay CA 1977. Microbiology of New Zealand feral venison. Journal of the Science of Food and Agriculture 28: 829-832.

Walk ST, Alm EW, Gordon DM et al. 2009. Cryptic Lineages of the Genus *Escherichia*. Applied and Environmental Microbiology 75: 6534-6544.