Host species-specific metabolic fingerprint database of enterococci and *Escherichia coli* and its application to identify the sources of fecal contamination in surface waters

***W. AHMED, R. NELLER and M. KATOULI**

Institute of Sustainability, Health and Regional Engagement, University of the Sunshine Coast, Maroochydore DC 4558, Queensland, Australia.

Running title: Tracking the sources of fecal contamination.

*Corresponding Author.
Faculty of Science, University of the Sunshine Coast, Maroochydore DC 4558, Queensland, Australia. Phone: (+61) 07 5479 0875. Fax: (+61) 07 5430 2887. E-mail: shuhat@yahoo.com

Abstract

A metabolic fingerprint database of enterococci and *Escherichia coli* from 10 host groups of animals was developed to trace the sources of fecal contamination in surface waters. In all, 526 enterococci biochemical phenotypes (BPTs) and 530 *E. coli* BPTs were obtained from 4057 entercocci and 3728 *E. coli* isolates tested. Of these, 231 enterococci BPTs and 257 *E. coli* BPTs were found in multiple host groups. The remaining 295 enterococci BPTs and 273 *E. coli* BPTs were unique to individual host groups. The database was used to trace the sources of fecal contamination in a local creek. The mean diversity of enterococci (Di= 0.76 ± 0.05) and *E. coli* (Di= 0.88 ± 0.04) was high (maximum 1) in water samples indicating diverse sources of fecal contamination. Overall, 71% of enterococci BPTs and 67% of *E. coli* BPTs from water samples were identified as human and animal sources. Altogether, 248 enterococci BPTs and 282 *E.coli* BPTs were found in water samples. Among enterococci, 26 (10%) BPTs were identical to humans and 152 BPTs (61%) were identical to animals (animal-BPTs). Among *E. coli*, 36 (13%) BPTs were identical to humans and 151 (54%) BPTs were identical to animals. Of the animal-BPTs, 101 (66%) enterococci BPTs and 93 (62%) *E. coli* BPTs were also unique to individual animal groups. On the basis of these unique enterococci BPTs, chickens contributed 14% of contamination followed by humans (10%), dogs (7%) and horses (6%). For *E. coli*, humans contributed 13% of contamination followed by ducks (9%), cattle (7%) and chickens (6%). The developed metabolic fingerprint database was able to distinguish between human and animal sources as well as among animal species in the studied catchment.

Introduction

Surface water is frequently contaminated with fecal bacteria. Non-point sources (NPS) such as domestic and wild animal defecation (7, 19), malfunctioning septic trenches (2, 19, 25), storm water drainage and urban runoff (25, 34) and/or point sources (PS) such as industrial effluents and municipal wastes (40), are known to be potential sources of such contamination. It has been reported that various human enteric pathogens such as *Salmonella* spp., *Shigella* spp. (13) and hepatitis A (7, 23, 36, 42) have been found in surface waters as a result of human fecal contamination. Defecation from domestic animals may further contribute pathogens such as *Escherichia coli* 0157:H7 and *Cryptosporidium* spp. (13, 15, 36, 44). Identification of major sources of fecal bacteria therefore, whether human or animal, is necessary for improved management of surface water quality and the minimization of public health risks associated with such contamination.

Fecal coliforms have been widely used as an indicator of the microbiological quality of surface and ground waters (16, 19, 22, 44). This group of bacteria is commonly found in the gastrointestinal tracts of all warm-blooded animals (21, 35, 52). However, the value of fecal coliforms as an indicator has recently been questioned, because these bacteria can also derive from various sources such as soil, agricultural run-off, composted animals, decaying vegetation and industrial processes (13, 16, 28). Instead, it has been suggested that *E. coli* and enterococci are much better indicators of fecal contamination as these bacteria colonize in the gut of humans and other warm-blooded animals (8, 39). *E. coli* is widely accepted as a potential fecal indicator bacterium because they are not normally pathogenic, easy to detect and culture, and are found at concentrations much higher than other pathogens in surface

waters (50). Fecal streptococci are also considered ideal fecal indicator bacteria because of their ability to survive in the natural environment for lengthy periods (16, 22, 27, 48). However, it has also been noted that the sole presence of these bacteria in surface waters does not provide definitive information regarding their possible source(s) (22, 29, 34, 52).

In recent years, several methods, collectively known as bacterial or microbial source tracking (BST or MST) methods, have been developed to distinguish the various sources of animal and/or human fecal contamination (35, 52). These methods include ribotyping (4, 11, 18, 19, 41), pulsed-field gel electrophoresis (46, 47), ribosomal genetic markers (9, 10), repetitive DNA sequences (12, 13), carbon source utilization (17) and antibiotic resistance profiles (22, 40, 53, 54) of fecal indicator bacteria. Chemical methods such as the detection of caffeine (44) and fecal sterols analysis (33) have also been used to detect the source(s) of fecal contamination in surface waters. Most of these methods are based on the hypothesis that phenotypic or genotypic characteristics of specific strains are associated with specific animals (6, 20, 25, 34). On the basis of this hypothesis, a fingerprint database (i.e. phenotypic or genotypic profiles) of strains from known sources has been developed to predict the source(s) of unknown environmental isolates (45, 52). The advantages and disadvantages of these methods have been discussed in various studies (34, 35, 44). For instance, genotypic methods, although highly discriminatory, can be laborious and/or expensive for ecological studies where a large number of isolates need to be tested (19, 30, 38). The most commonly used phenotypic method, the antibiotic resistance profiles, can be used to test a large number of isolates within a short time and is rather inexpensive. However, it is known that antibiotic resistance genes can be lost from or gained by bacteria

under certain conditions (14, 44). In addition, this method does not provide information about fecal indicator bacteria that are not resistant to antibiotics, but are derived from different animal species. Chemical methods such as caffeine/pharmaceuticals or fecal sterols detection require stringent sampling and can be expensive. In addition, it has also been reported that chemical methods are not sensitive enough to detect recent pollution (44).

A biochemical fingerprinting method known as the PhPlate system (PhPlate AB, Stockholm, Sweden) has been reported and used in many epidemiological and ecological studies (2, 30, 31, 51). It measures the kinetics of bacterial metabolism in micro titer plates. For each bacterial isolate, it yields a biochemical fingerprint made of several quantitative data, which are used with the PhPlate software to calculate the level of similarity between the tested isolates. This system has a high discriminatory ability and reproducibility (30, 31, 32), and is shown to be comparable with many genotypic methods in comparative studies (30, 31). The PhPlate system is simple to use and can be applied to studies involving large numbers of isolates, and is therefore an excellent tool for studying the diversity and persistence of fecal indicator bacteria in surface waters (24, 30, 31, 51). In this study we used the PhPlate system to characterize two fecal indicator bacteria, enterococci and *E. coli* from different host groups (i.e. animal species) to develop a metabolic fingerprint database to identify the source(s) of fecal contamination in a local creek.

Materials and Methods

Host groups sampling

Ten host groups were sampled between July 2003 and August 2004. These groups included horses, cattle, sheep, pigs, ducks, chickens, deer, kangaroos, dogs and humans (via septic tanks). For each group of farm animals we initially collected 5 fecal samples from 5 individuals within a farm. Up to 32 isolates of both enterococci and *E. coli* were tested from each sample (i.e. each animal) to determine the diversity of these indicator bacteria. Based on the low diversity (0.41 ± 0.09 for enterococci and 0.53 ± 0.11 for *E. coli*) (minimum 0 and maximum 1) obtained from this assessment, sampling was extended to include multiple (up to 20 farms where possible) farms for each group of farm animals. All septic tanks tested for were within the 100m distance of the creek (see figure 1). However, for farm animals we collected samples from as many farms as we had access in the studied catchment. In addition, samples were also collected from other catchments within the same geographical area. At each farm, up to 3 animals were sampled and from each animal up to 12 isolates were tested. A total number of 234 samples were collected from horses (38 samples), cattle (54 samples), sheep (28 samples), pigs (32 samples), chickens (36 samples) and ducks (46 samples). All samples were collected from fresh feces of individual animals with sterile swabs and inserted into Amies transport medium (Interpath, Melbourne, Australia), transported to the laboratory and tested within 6h. Dog samples (47 samples) were collected from two city dog parks on 8 occasions. Deer samples (25 samples) were collected from a local deer sanctuary park and kangaroo (20 samples) samples were collected from the University of the Sunshine Coast where a large number of kangaroos roam.

137 Human samples were collected from the outlet of 39 septic tanks using sterile swabs. Swabs
138 were then inserted into Amies transport medium (Interpath), transported on ice to the
139 laboratory and tested within 6h.
140
141 **Isolation of enterococci and *E. coli***
142 All fecal samples were streaked on m-enterococcus (Difco, USA) and chromogenic *E.*
143 *coli*/coliform (Oxoid, UK) agar plates and were incubated at 37ºC for 24 h (for *E. coli*) and
144 48 h (for enterococci). This chromogenic medium allows specific detection of *E. coli*
145 through substrate cleavage by the enzyme glucuronidase and formation of purple colonies,
146 which are different from other fecal coliforms (rose/pink colonies). All enterococci isolates
147 were tested for esculin hydrolysis on bile esculin agar (Oxoid) to confirm their identification
148 (5) before being tested for biochemical fingerprint with the PhPlate system.
149
150 **Biochemical fingerprinting with the PhPlate system**
151
152 The principle of the biochemical fingerprinting with the PhPlate system has been described
153 previously (2, 37). This method uses quantitative measurements of the kinetics of several
154 biochemical reactions of bacteria in microtiter plates with dehydrated substrates (26, 37).
155 The typing reagents used in this method are specifically chosen for different groups of
156 bacteria to give an optimal discriminatory power and reproducibility (37). For each bacterial
157 isolate, it yields a biochemical fingerprint made of several quantitative data, which are used
158 with the PhPlate software to calculate the level of similarity between the tested isolates.
159 Prepared microtiter plates contain 11 different substrates in each row and allow testing of 8

160 isolates per plate. In this study, we used two types of plates specifically developed for typing
161 of *E. coli* (PhP-RE plates) and enterococci strains (PhP-RF plates). The 11 substrates used
162 for enterococci and *E.coli* have been described before (24, 30, 51). The growth medium for
163 PhP-RF contains 0.2% (w/v) proteose peptone (Oxoid), 0.05% (w/v) yeast extract (Oxoid),
164 and 0.5% (w/v) NaCl, and 0.011% (w/v) bromothymol blue and for *E. coli* it contained
165 0.1% (w/v) proteose peptone, and 0.011% (w/v) bromothymol blue according to the
166 manufacturer instructions.
167 From each sample up to 12 single and isolated colony were randomly selected with sterile
168 tooth picks directly from the chromogenic coliform/*E.coli* agar plates (for *E.coli*) and from
169 the bile esculin agar (for enterococci) and suspended into the first well of each row
170 containing only 350µl of growth medium. Using a multi-channel pipette, aliquots of 25µl of
171 bacterial suspension were transferred into each of the other 11 wells containing 150µl
172 growth medium. Plates were then incubated at 37°C and A_{620} was measured at 7, 24 and 48
173 h for *E.coli* and at 16, 40 and 64h for enterococci using a micro plate reader (Lab-systems
174 Multiskan, Finland). After the final reading the mean value for all three readings was
175 calculated for each isolate (biochemical fingerprint). Similarities between the isolates were
176 calculated as correlation coefficients and clustered according to the un-weighted pair group
177 method (UPGMA) with arithmetic averages (49). An identity (ID) level of 0.965 was
178 established based on the reproducibility of the system after testing 20 isolates in duplicate.
179 Isolates with similarity higher than the ID-level were regarded as identical and assigned to
180 similar biochemical phenotypes (BPTs). BPTs with identical isolates were called common
181 (C-BPT) and those with one isolate were called single (S-BPT). All S-BPTs and
182 representative isolates of each C-BPT were transferred to McConkey agar (Oxoid) for purity

183 and further tested for indole production and citrate before they were saved on tryptic soy
184 broth (Oxoid) with 15% (v/v) glycerol at –80^{o}C.
185
186 The phenotypic diversity among the isolates was measured with Simpson's index of
187 diversity (Di) (4). Di in the present study depends on isolates distribution into different
188 BPTs. Diversity is high (maximum 1) for a population consisting of different BPTs and is
189 low (minimum 0) if the population consisting of few BPTs. The phenotypic similarity
190 between different bacterial populations in two or more samples was calculated as population
191 similarity coefficient. The population similarity coefficient calculates the proportion of
192 isolates that are identical in two or more compared bacterial populations (29). It is high
193 (maximum 1) if two populations contain similar BPTs, and is low (minimum 0) if the
194 population contains different BPTs. Clustering of population similarity coefficients was also
195 performed according to the UPGMA method. All data handling, including optical readings,
196 calculations of correlations and coefficients, diversity indexes, population similarity values
197 as well as clustering and printing dendrograms, was performed using the PhPlate software
198 version 4001 (PhPlate system, PhPlate AB, Stockholm).
199
200 **Database development**
201 In developing the database, we categorized the BPTs into two distinct types, unique (UQ)
202 and shared (SH) BPTs on the basis of their occurrence in host groups. The UQ-BPTs are
203 those BPTs that are specific to a single host group, whereas SH-BPTs were found in
204 multiple host groups. To achieve this, all BPTs obtained from each animal were compared
205 with other animals within a host group. If identical, a representative of identical BPTs, as

well as all non-identical BPTs, were initially saved in the database and regarded as total-BPTs for each host group of animals. Further, total-BPTs from each host group were cross-referenced with those of others to calculate the occurrence of BPTs among different host groups. For instance, if a BPT from a host group (e.g. horse) was identical to a BPT from another (e.g. sheep), this BPT was regarded as SH-BPT (i.e. "shared") between two host groups. If a BPT from a host group was not identical to any other groups, it was regarded as UQ-BPT (i.e. "unique").

Surface water sampling

Water samples were collected from the Eudlo Creek, a sub-catchment of the Maroochy River in South-East of Queensland, Australia (Figure 1). The total area of this largely rural sub-catchment is approx. 7980 ha, of which > 85% is not serviced by a centralized sewer system. The creek is approx. 8 km in length and has been reported by the Environmental Protection Agency (EPA) and Waterwatch (a community-based organization) to be contaminated with fecal bacteria and nitrates. Possible sources of contamination include intensive animal farms and approx. 1600 conventional septic systems (2).

Samples were collected from up to 5 sites across the Eudlo sub-catchment (see Figure 1) during November 2003 to December 2003 and during August 2004 to September 2004 on 7 different occasions. In all, 27 samples were collected and were tested in triplicate. Water samples were collected in 500ml sterile bottles from 30cm below the water surface and transported on ice to the laboratory and tested within 6 h. The membrane filtration method was used to process all the water samples (3). Different dilutions of water samples were filtered through a 0.45µm pore size membranes (Millipore, USA) and placed on

chromogenic *E. coli*/coliform (Oxoid) and m-enterococcus agar plates (Difco) and the plates were incubated at 37^{o}C for 24 h (for *E. coli*) and 48 h (for enterococci). After incubation, from each water sample up to 40 (where possible) enterococci and *E. coli* isolates (where possible) were typed with the PhPlate system as described earlier.

Statistical analysis

Mann Whitney's non-parametric test was used to determine the significant difference between the mean number of enterococci BPTs and *E. coli* BPTs found in all host groups. In addition, this test was also performed on the overall diversity of enterococci and *E. coli* from all host groups.

Results

A total number of 4057 enterococci and 3728 *E. coli* isolates were typed from 10 host groups. Within each host group, different BPTs were found some of which, were identical. Representatives of the identical BPTs and the non-identical BPTs were initially included in the database and regarded as total-BPTs found in each host group. Applying this approach, a total number of 526 BPTs of enterococci and 530 BPTs of *E. coli* were obtained from all host groups. Table 1 shows the number of isolates tested and the number of total- BPTs found in each host group. For enterococci, the ratio of BPTs over the number of total isolates tested from each host group ranged from 7.3% (for sheep) to 18.7% (for horse) yielding a mean value of 13.9 ± 4.0 for all host groups. With *E. coli* this ratio ranged from 8.2% (for sheep) to 17% (for ducks), yielding a mean value of 14.4 ± 2.5 (Table 1). The mean number of total enterococci and *E. coli* BPTs found in all host groups did not differ significantly (p=0.97).

252 The mean diversity of both enterococci and *E. coli* within each host group ranged from
253 0.41± 0.38 (for sheep) to 0.75± 0.25 (for horses) and from 0.44 ± 0.27 (for sheep) to 0.85±
254 0.07 (for deer) respectively (Table 2). However, the overall diversity of both indicator
255 bacteria (0.6 ± 0.1 for enterococci vs. 0.65 ± 0.1 for *E. coli*) did not differ significantly
256 (p=0.36).
257
258 **Unique and shared BPTs**
259 When we compared the total-BPTs of all host groups with each other, it was found that
260 certain BPTs were specific to individual host groups. These BPTs were referred to as UQ-
261 BPTs. For enterococci, the range of UQ-BPTs among host groups varied from 7 (in sheep)
262 to 66 (in humans). For *E. coli*, this figure was 6 (in kangaroos) and 69 (in humans) (Table
263 3). The mean percentage of total UQ-BPTs among enterococci and *E. coli* was 56% and
264 51% respectively. Certain BPTs were also found in multiple host groups and they were
265 referred to as shared SH-BPTs. For instance, of the 76 enterococci total-BPTs found in
266 horses, 54 were only found in horses (i.e. UQ-BPTs), whereas 22 were found not only in
267 horses but also in other host groups (SH-BPTs).
268
269 For enterococci, the range of SH-BPTs among host groups varied from 14 (in sheep) to 33
270 (in chickens) and for *E. coli*, these figures were 13 (in kangaroos) and 37 (in ducks) (Table
271 3). Therefore, a total number of 295 enterococci BPTs and 273 *E. coli* BPTs occurred only
272 once in the database while 231 BPTs for enterococci and 257 BPTs for *E. coli* occurred in
273 multiple host groups. All BPTs (i.e. UQ or SH-BPTs) from animal groups that were not
274 found in humans were collectively categorized as animal-BPTs. The animal-BPTs consisted

275 of 432 enterococci BPTs and 438 *E. coli* BPTs of which, 229 (53%) enterococci BPTs and
276 204 (47%) *E.coli* BPTs were UQ-BPTs (see Tables 1 and 3).
277
278 **Tracking the source of contamination in Eudlo Creek**
279 A total number of 27 water samples were collected from 5 sites along the Eudlo creek
280 mainstream (Figure 1). From each water sample, up to 40 enterococci and *E. coli* isolates
281 (where possible) were typed and compared with the database. The mean diversity of
282 enterococci (Di= 0.76 ± 0.05) and *E. coli* (Di= 0.88 ± 0.04) was generally high (maximum
283 1) in water samples indicating diverse sources of these bacteria. A total number of 791
284 enterococci isolates (248 total-BPTs) were tested from water samples, of which, 26 BPTs
285 (10%) were only found in humans (i.e. UQ-BPTs) and 152 BPTs (61%) belonged to animals
286 (i.e. animal-BPTs) tested in this study (Table 4). Of the 550 *E. coli* isolates (282 total BPTs)
287 tested from the same water samples, 36 BPTs (13%) were of human origin and 151 BPTs
288 (54%) belonged to animals tested (Table 4). The remaining 70 enterococci BPTs and 95 *E.*
289 *coli* BPTs, either belonged to BPTs shared between humans and animals (28 enterococci
290 BPTs and 23 *E. coli* BPTs), or did not match the database and were therefore regarded as
291 unknown BPTs (Table 4).
292 Comparison of total-BPTs found in water samples over the entire sampling period with the
293 database showed that 61% enterococci and 54% *E. coli* BPTs were identical to animal-
294 BPTs, and that some were also unique to individual animal group. Distribution of UQ-BPTs
295 among animal species ranged between 0% (deer) to 13% (chicken) for enterococci and 0%
296 (deer) to 8% (ducks) for *E. coli*. Ten percent of enterococci UQ- BPTs and 13% of *E.coli*
297 UQ-BPTs found in water samples were identical to humans.

298

299 To identify whether there is a fundamental difference between the population of both fecal
300 indicator bacteria from humans and animals, we also performed a population similarity
301 comparison between all BPTs found in humans and animals. The results indicated that the
302 mean similarity among enterococci (0.27 ± 0.1) and *E.coli* (0.34 ± 0.06) populations
303 between different animal groups was significantly higher (p=0.003 for enterococci and p =
304 0.001 for *E.coli*) than the mean similarity between humans and animals (i.e. 0.16 ± 0.03 for
305 enterococci and 0.98 ± 0.02 for *E. coli*) (Figure 2).

306

307 Discussion

308 Identification of potential source(s) of fecal contamination in surface waters requires a
309 method that is capable of distinguishing between human and animal sources. Ideally, the
310 method should also be sensitive enough to discriminate different animal species. In recent
311 years, several genotypic and phenotypic methods have been developed to trace the sources
312 of fecal contamination in surface waters by typing fecal indicator bacteria (9-11, 13, 19, 22,
313 40, 41, 43, 46, 47, 53, 54). In this study, we used a biochemical fingerprinting method to
314 develop a host-specific metabolic fingerprint database of two recommended fecal indicator
315 bacteria, enterococci and *E. coli* (1, 44) to trace the sources of fecal contamination in surface
316 waters.

317 It is known that the size and representativeness of the database are two important factors for
318 determining the source(s) of fecal contamination (44, 52). To date, most of the genotypic
319 and phenotypic host origin databases are based on testing up to 500 isolates of either

320 enterococci or *E. coli* (11-13, 19, 34, 40, 44, 52). In addition, these isolates were usually
321 collected from a small number of samples, which may not sufficiently represent the diverse
322 fecal indicator bacteria found in different host groups. In developing the metabolic
323 fingerprint database we specifically focused on two important factors. One was the number
324 of isolates to be tested from each animal species, and the other was how well these numbers
325 represent the diversity of indicator bacteria among the animal species. To address this
326 question, we initially tested 160 isolates of both fecal indicator bacteria from 5 randomly
327 chosen individuals of the same species of animals within a farm (data not shown). This
328 comparison showed that animals of same species within a farm carry many identical BPTs,
329 which could be explained by their frequent contact with each other or dietary similarity, and
330 therefore sharing a common bacterial population (19, 20, 26). However, we obtained a better
331 diversity when we compared animals of the same species from one farm with those of
332 another within a radius of around 20km both inside and outside of the studied catchment.
333 For this reason, we reduced the number of samples up to 3 individuals of same species from
334 each farm and increased the number of farms up to 20 where possible within and outside the
335 study area. Using this strategy, we tested an average of 400 isolates of both fecal indicator
336 bacteria from 10 host groups yielding a total number of 4087 enterococci and 3728 *E. coli*
337 isolates from different farms or locations. Based on this experience, however, we suggest
338 that the emphasis should be on focused testing of more individual animals (preferably from
339 different farms) rather than testing more isolates from each individual, so as to obtain
340 diverse phenotypes or genotypes of known sources.
341

342 The comparison of total-BPTs in each host group with others showed that many identical
343 BPTs were shared in multiple host groups. Bacteria are ubiquitous in the environment and
344 can be found transitionally in many animal species simultaneously. Similar shared
345 fingerprints (ribotypes) have also been reported among different host groups in other studies
346 (19, 20, 34). However, in our study, the percentage of shared BPTs among host groups was
347 quite high. This is due to the fact that we not only tested a large number of isolates from
348 each host group but also tested a wide range of host groups and therefore found more shared
349 BPTs among host groups. In contrast, a number of BPTs were also specific to individual
350 host group and therefore regarded as unique (UQ) BPTs. A recent molecular-based study
351 (25) defined unique genotypes on the basis of specificity to individual host group rather than
352 comparing these genotypes to those found in other host groups. However, in our study, we
353 defined UQ-BPTs as those BPTs that occurred once only in each host group after comparing
354 with all other total-BPTs found in other host groups. The number of UQ-BPTs in our study,
355 varied among different host groups. Some host groups (i.e. sheep, deer and kangaroos)
356 contained a smaller number of UQ-BPTs than others. This may be explained by the fact that
357 a smaller number of samples tested from these host groups from limited locations and
358 therefore our sampling effort could not capture the diversity found among these host groups.
359 Nevertheless, we found that these UQ-BPTs can be used as specific fingerprints to pinpoint
360 the sources of fecal contamination in surface waters. In contrast, some SH-BPTs were found
361 in two or more animal species including humans. For instance, we found 28 BPTs of
362 enetrococci and 23 BPTs of *E. coli*, which were occurred in both human and animal host
363 groups. These BPTs could not be used to distinguish the various sources of fecal
364 contamination and were excluded from our database. However, we also found that certain

365 SH-BPTs, though found among different animal species, were not found in humans and
366 could therefore be categorized as animal-BPTs (including UQ- and SH-BPTs among 9 host
367 groups of animals).

368 To evaluate the ability of our database to identify the sources of fecal contamination, we
369 collected a total number of 27 water samples from 5 different sites along the Eudlo creek
370 mainstream on several occasions and compared the total-BPTs of both fecal indicator
371 bacteria isolated from these sites with our database. Ten percent of enterococci BPTs and
372 13% of *E. coli* BPTs were identified as human UQ-BPTs. It should be noted that, in our
373 study, human samples were obtained from septic tanks rather than fresh human fecal
374 samples and therefore, some unique and/or shared strains may have not survived in the
375 septic tanks and therefore not detected. Of the animal-BPTs, 101 (66%) of enterococci BPTs
376 and 93 (62%) of *E. coli* BPTs were unique to individual host groups. On the basis of UQ-
377 BPTs for enterococci, chickens contributed 13% of contamination followed by humans
378 (13%). For *E. coli*, humans contributed 13% followed by ducks (9%). Both the enetrococci
379 and *E. coli* databases was in close agreement in terms of identifying the sources of
380 contamination (i.e. for humans 10% enterococci and 13% *E. coli*; for cattle 6% enterococci
381 and 7% *E. coli* were identified from same water samples) though it was not quite consistent
382 for certain host groups (i.e. for chickens 14% enterococci and 6% *E. coli*). However,
383 interestingly total-BPTs from deer were not identical to those found in the water samples
384 which can be explained by the fact that deer are normally kept in a sanctuary, which
385 restricted their access to nearby creeks and in addition, our study area did not contain any
386 wild deer. On the basis of this, we conclude that both fecal indicator bacteria can be used
387 alone or in combination with each other, which provide a much better insight regarding the

388 contributing sources. Using this PhPlate system and the same indicator bacteria, we have
389 recently shown that combination of both fecal bacteria provide a better understanding of the
390 sources of human contamination in surface waters through failed septic systems (2).
391 Certain BPTs of both fecal indicator bacteria found in water samples did not match our
392 database. This may be due to the fact that either our database was not large enough to
393 capture the diversity of these indicator bacteria, or that these unknown BPTs might have
394 originated from other non-point sources or a combination of both. It has been suggested that
395 a library size up to 40,000 isolates may be needed to capture the genetic diversity present
396 among *E. coli* (25). It is also recommended that database should ideally be developed from
397 the animal species residing in the study area, as they more likely to contribute fecal
398 contamination to surface waters in study area (25). In our study, although we tried to collect
399 samples from as many farms as possible within and outside the studied area, we did not have
400 access to all farms in the studied catchment. Another important factor that has to be
401 considered is that the number and the types of animals within a study area may vary over
402 time due to agricultural practices and/or animal migration (25) and therefore it may not be
403 possible to include samples from all animals reside in a study area. This will restrict the
404 ability of a database to trace the sources of contamination from all animals within a
405 watershed. In addition, it is known that geographical variability exists among indicator
406 bacteria (19), which limit the efficiency of a database to identify unknown environmental
407 isolates when these bacteria are collected from another geographical area. Therefore, it has
408 to be noted that the universal use of such host species-specific database which is developed
409 for a limited geographical area should be interpret with care.
410

411 In conclusion, we developed a host-specific metabolic fingerprint database of two fecal
412 indicator bacteria and used that successfully to trace the sources of fecal contamination in
413 the studied creek. The database was capable of identifying the sources of more than 65% of
414 fecal bacteria in the studied creek. We also found that while this system could differentiate
415 between humans and animal sources of fecal contamination, it was also capable of further
416 differentiating between animal species and therefore can be used as a potential tool to trace
417 the sources of fecal contamination in the subject catchment.

418

419 ## Acknowledgment

420 We thank Anthony Weston, Sandra Hipwood and Daniel Morgan for their assistance in
421 collecting samples. We also thank the Maroochy Shire Council for their cooperation. This
422 study was financially supported by an internal grant URG 02/4 from the University of the
423 Sunshine Coast.

References

1. **Abbot, S., B. Caughley, and G. Scott.** 1993. Evaluation of Enterolert® for the enumeration of enterococci in the marine environment. N. Z. J. Mar. Freshw. Res. **32**:505-513.

2. **Ahmed, W., R. Neller, and M. Katouli.** 2004. Evidence of septic systems failure determined by a bacterial biochemical fingerprinting method. J. Appl. Microbiol. 98:910-920.doi:10.1111/j.1365-2672.2004.02522.x.

3. **Anon.** 1997. Water quality- Detection and Enumeration of Coliform organisms, thermo tolerant Coliform Organism and presumptive *Escherichia coli;* Part 1: Membrane Filtration Method. I.S.O 9308/1 Geneva: ISO.

4. **Atlas, R.** 1984. Use of microbial diversity measurements to assess environmental stress, p. 540-545. *In* M.J. Klug and C.A. Reddy (ed.), Current perspective in microbial ecology. American society for Microbiology, Washington, D.C.

5. **American Public Health Association.** 1995. Standard method for the examination of water and wastewater. 19th ed. American Public Health Association, Inc., Washington, D. C.

6. **Amor, K., D. E. Heinrichs, E. Frirdich, K. Ziebell, R. P. Johnson, and C. Whitfield.** 2000. Distribution of core oligosaccharide types in lipopolysaccharides from *Escherichia coli.* Infect. Immun. **68:**1116-1124.

7. **Barnes, D., and D. M. Gordon.** 2004. Coliform dynamics and the implications for source tracking. Environ. Microbiol. **6**: 501-509.

8. **Baudisöva, D.** 1997. Evaluation of *Escherichia coli* as the main indicator of fecal pollution. Wat. Sci. Tech. **35:** 333-356.

9. **Bernhard, A. E., and K. G. Field.** 2000a. A PCR assay to discriminate human and animal feces on the basis of the host difference in *Bacteroides-Prevotella* genes encoding 16S rRNA. Appl. Environ. Microbiol. **66:** 4571-4574.

10. **Bernhard, A. E., and K. G. Field.** 2000b. Identification of non-point sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. Appl. Environ. Microbiol. **66:**1587-1594.

11. **Carson, S. A., B. L. Shear, M. R. Ellersieck, A. M. H. A. Afsaw.** 2001. Identification of Faecal *Escherichia coli* from humans and animals by Ribotyping. Appl. Environ. Microbiol. **67:**1503-1507.

12. **Carson, S. A., B. L. Shear, M. R. Ellersieck, and J. D. Schnell.** 2003. Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal Escherichia coli from humans and animals. Appl. Environ. Microbiol. **69:**1836-1839.

13. **Dombek, P.E, L. K. Johnson, M. B. Brown, and M. J. Sadowsky.** 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Appl. Environ. Microbiol. **66:**2572-2577.

14. **Freter, R.** 1984. Factors affecting conjugal plasmid transfer in natural bacterial communities. P-105-104. *In*. M.J. Klug and C.A. Reddy (ed.) Current perspectives in microbial ecology, Washington, D.C.

15. **Fujioka, R. S.** 2002. Microbial indicators of water quality, p-234-243. *In* C. J. Hurst, R. L. Crawford, G. R. Knudsen, K. J. McInereny, and L. D. Stetzenbach (ed.), Manual of environmental microbiology, 2nd ed. ASM Press, Washington, D.C.

16. **Hagedorn, C., S. L. Robinson, J. R. Filtz, S. M. Grubbs, T.A. Angier, and R. B. Reneau, Jr.** 1999. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. Appl. Environ. Microbiol. **65:**5522-5531.

17. **Hagedorn, C., J.B. Crozier, K.A. Mentz, A.M. Booth, A.K. Graves, N.J. Nelson and R.B. Reneau Jr.** 2002. Carbon source utilization profiles as a method to identify sources of faecal pollution in water. J. Appl. Microbiol. **94:**792-799

18. **Hartel, P. G., W. I. Segars, N. J Stern, J. Steiner, and A. Buchan.** 1999. Ribotyping to determine host origin of *E. coli* isolates in different water samples, p.377-382. *In.* D.S. Olsen and J.P. Potyondy (ed.), Wild land hydrology. Technical Publication Series TPS-99-3. American Water Resource Association, Herndon, Va.

19. **Hartel, P. G., J. D. summer, J. L. Hill, J. C. Collins, J. A. Entry, and W. I. Segers.** 2002. Geographic variability of *Escherichia coli* isolates from animals in Idaho and Georgia. J. Environ. Qual. **31:**1273-1278.

20. **Hartel, P. G., J. D. Summer, W. I. Segars.** 2003. Deer diet affects ribotype diversity of *Escherichia coli* for bacterial source tracking. Water. Res. **37:**3263-3268.

21. **Harwood, V. J., B. D. Parrish and V. Wagner.** 1999. Isolation of fecal coliform bacteria from the diamondback terrapin (*Malaclemys terrapain centrata*) Appl. Environ. Microbiol. **65:**865-867.

22. **Harwood, V. J., J. Whitlock, and V. Withington.** 2000. Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of fecal contamination in tropical waters. Appl. Environ. Microbiol. **66:**3698-3704.

23. **Hurst, C.J.** 1997. Overview of water microbiology as it relates to public health. p. 133-135. *In*. C. J. Hurst *et al.* (ed.) Manual of environmental microbiology. American Society for Microbiology, Washington, D.C.

514

515 24. **Iverson, A., I. Kühn, A. Franklin, and R. Möllby.** 2002. High prevalence of
516 vancomycin resistant enterococci in Swedish sewage. Appl. Environ Microbiol. **68:**2838-
517 2842.

518

519 25. **Johnson, L. K., M.B. Brown., E. A. Carruthers., J. A. Ferguson., P. E. Dombek,**
520 **and M. J. Sadowsky.** 2004. Sample size, library composition, and genotypic diversity
521 among natural populations if Escherichia coli from different animals influence accuracy of
522 determining sources of fecal pollution. Appl. Environ. Microbiol. **70**: 4478-4485.

523

524 26. **Katouli, M., A. Lund., P. Wallgren., I. Kühn., O. Söderlind, and R. Möllby.** 1997.
525 Metabolic fingerprinting and fermentative capacity of the intestinal flora of pigs during pre-
526 and post-weaning periods. J. Appl. Microbiol. **83:** 147-154.

527

528 27. **Kibbey, H. J., C. Hagedorn, and E. L. McCoy.** 1978. Use of fecal streptococci as
529 indicators of pollution in soil. Appl. Environ. Microbiol. **35:**711-717

530

531 28. **Kreader, C. A.** 1995. Design and evaluation of bacteroides DNA probes for the specific
532 detection of human fecal pollution. Appl. Environ. Microbiol. **61:** 1171-1179.

533

534 29. **Kühn, I., G., Allestam, T. A. Strenstrom, and R. Möllby.** 1991. Biochemical
535 fingerprinting of water coliform bacteria-a new method for measuring the phenotypic

536 diversity and for comparing different bacterial populations. Appl. Environ. Microbial. **57:**
537 3171-3177.
538
539 30. **Kühn, I., Katouli, M., Wallgren, P., Söderlind, O. and Möllby, R.** 1995. Biochemical
540 fingerprinting as a tool to study the diversity and stability of intestinal microfloras.
541 Microecol. Therap. **23:**140-148.
542
543 31. **Kühn, I., G. Allestam, M. Engdhal, and T. A. Stenström.** 1997a. Biochemical
544 fingerprinting of coliform bacterial populations-comparisons between polluted river water
545 and factory effluents. Wat. Sci. Tech. **35:**343-350.
546
547 32. **Kühn, I., J. Albert, M. Ansaruzzaman, S. A. Alabi., N. A. Bhuiyanand, G. Huys,**
548 **M. S. Islam, P. Jansen, K. Kersters, P. K. B. Neogi and R. Möllby.** 1997b.
549 Characterization of aeromonas sp. isolated from humans with diarrhea, from healthy
550 controls, and from in surface water in Bangladesh. J. Clin. Microbiol. **35:** 369-373.
551
552 33. **Leeming R., A. Ball., N. Ashbolt, and P. D. Nicholas.** 1996. Using faecal sterols from
553 humans and animals to distinguish faecal pollution in receiving waters. Wat. Res. **30:** 2893-
554 2900.
555
556 34. **McLellan, S. L.** 2004. Genetic diversity of *Escherichia coli* isolated from urban rivers
557 and beach water. Appl. Environ. Microbiol. 2004. **70**: 4658-4665.
558

559 35. **Meays, C. L., K. Broersma, R. Nordin, and A. Majumder.** 2004. Source tracking
560 fecal bacteria in water: a critical review of current methods. J. Environ. Manag. **73:**71-79
561
562 36. **Moe, C.** 2002. Waterborne transmission of infectious agents, p.184-204. *In* C. J. Hurst,
563 R. L. Crawford, G. R. Knudsen, K. J. McInereny, and L. D. Stetzenbach (ed.), Manual of
564 environmental microbiology, 2nd ed. ASM Press, Washington, D.C.
565
566 37. **Möllby, R., I. Kühn, and M. Katouli.** 1993. Computerized biochemical fingerprinting-
567 A new tool for typing of bacteria. Rev. Med. Microbiol. l4:231-241.
568
569 38. **Olive, D. M and P. Bean.** 1999. Principles and applications of methods for DNA-based
570 typing of microbial organisms. J. Clin. Microbiol. **37:**1661-1669.
571
572 39. **Ørskov, F.** 1984. Genus I. Escherichia Castellani and Chalmers 1919, 941^{AL}. p. 420-
573 423. *In* N. R. Kreig and J.G. Holt (ed.) Bergey's manual of systematic bacteriology. Vol. 1.
574 Williams and Wilkins, Baltimore, MD.
575
576 40. **Parveen, S., R. L. Murphree, L. Edmiston, C. W. Kaspar, K. M. Portier, and M. L.**
577 **Tamplin.** 1997. Association of multiple-antibiotic-resistance profiles with point and non-
578 point sources of *Escherichia coli* in Apalachicola Bay. Appl. Environ. Microbiol. **63:**2607-
579 2612.
580

581 41. **Parveen, S., K. M. Portier, K. Robinson, L. Edmiston, and M. Tamplin.** 1999.
582 Discrimination analysis of ribotype profiles of *Escherichia coli* for differentiating human
583 and nonhuman sources of fecal pollution. Appl. Environ. Microbiol. **65:**3142-3147.
584
585 42. **Puech, M. C., J. M. McAnulty, M. Lesjak, N. Shaw, L. Heron, and J. M. Watson.**
586 2001. A statewide outbreak of cryptosporidiosis in New South Wales associated with
587 swimming at public pools. Epidemiol. Infect. **126:**389-396.
588
589 43. **Samadpour, M., and N. Checowitz.** 1995. Little Soos Creek microbial source tracking:
590 a survey. University of Washington Department of environmental Health, Seattle.
591
592 44. **Scott, T. M., J. B. Rose, T. M. Jenkins, S. R. Farrah, and J. Lukasik.** 2002.
593 Microbial source tracking: current methodology and future directions. Appl. Environ.
594 Microbiol. **68:**5796-5803.
595
596 45. **Simpson, J.M., J.W. Santo Domingo, and D.J. Reasoner.** 2003. Microbial source
597 tracking: state of the science. Environ. Sci. Technol. **36:**5280-5288.
598
599 46. **Simmons, G. M., Jr., S. A. Herbain, and C. M. James.** 1995. Managing non-point
600 fecal coliforms sources to tidal inlets. Wat. Res. Upd. **1995:**64-74.
601

602 47. **Simmons, G. M., Jr., and S. A. Herbain.** 1998. Potential sources of *Escherichia coli*
603 (*E. coli*) to Children's Pool in La Jolla, California. Final report for the City of San Diego
604 and the County of San Diego Dep. of Environmental Health, San Diego, CA.
605
606 48. **Sinton, L.W., A.M. Donnison, and C.M. Hastie.** 1993. Faecal streptococci as faecal
607 pollution indicators: a review. Part II. Sanitary significance, survival and use. N.Z. J. Mar.
608 Freshwater Res. **27:**117-137
609
610 49. **Sneath, P.H. A., and R. R. Sokal.** 1973. Numerical taxonomy: the principles and
611 practice of numerical classification. W.H. Freeman, San Francisco, CA.
612
613 50. **Toranzos, G.A., and G. A. McFeters.** 1997. Detection of indicator microorganisms in
614 environmental fresh waters and drinking waters, p. 184-194. *In* C. J. Hurst, G. R. Knudsen,
615 M. J. McInerney, L. D. Stetzenbach, and M. V. Walter (ed.), Manual of environmental
616 microbiology. American Society for Microbiology, Washington, D. C.
617
618 51. **Vilanova, X., A. Manero, M. Cerdà Cuéller and A. R. Blanch.** 2002. The effect of a
619 sewage treatment plant effluent on the faecal coliforms and enterococci populations of the
620 reception river waters. J. Appl. Microbiol. **92:**210-214.
621
622 52. **Whitlock, J. R., D.T. Jones, and V. J. Harwood.** 2002. Identification of the sources of
623 fecal coliforms in an urban watershed using antibiotic resistance analysis. Water Res.
624 **36:**4273-4282.

53. **Wiggins, B. A.** 1996. Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters. Appl. Environ. Microbiol. **62:**3997-4002.

54. **Wiggins**, **B. A, R. W. Andrews, R. A. Conway, C. L. Corr, E. J. Dobratz, D. P. Dougherty, J. R. Eppard, S. R. Knupp, M. C. Limjoco, J. M. Mettenburg, J. M. Rinehardt, J. Sonsino, R. L. Torrijos, and M. E. Zimmerman.** 1999. Use of antibiotic resistance analysis to identify non-point sources of fecal pollution. Appl. Environ. Microbiol. **65:**3483-3486.

TABLE 1. Number of samples tested from each host group and the number of total-BPTs found.

Host groups	No. of samples	No. of isolates tested		No. of total-BPTs found (% over isolates)	
		Enterococci	*E. coli*	Enterococci	*E. coli*
Human:	56	1072	621	94 (8.8)	92 (14.8)
Animals:					
Horses	38	407	407	76 (18.7)	60 (14.7)
Dogs	47	404	408	49 (12.1)	64 (15.7)
Ducks	46	408	404	58 (14.2)	69 (17)
Cattle	55	411	401	47 (11.4)	53 (13.2)
Chicken	36	408	408	74 (18.1)	59 (14.5)
Pigs	32	312	400	54 (17.3)	53 (13.3)
Sheep	27	287	367	21 (7.3)	30 (8.2)
Deer	25	204	200	28 (13.7)	31 (15.5)
Kangaroos	20	144	112	25 (17.4)	19 (17.)
Total	382	4057	3728	526 (13.9 ± 4)*	530 (14.4 ± 2.5)*

*Mean and standard deviation

TABLE 2. The mean diversity of fecal indicator bacteria in host groups.

Host groups	Enterococci	*E. coli*
	Mean Di	Mean Di
Human	0.50 ± 0.30	0.50 ± 0.30
Animals		
Horses	0.75 ± 0.25[a1]	0.63 ± 0.26[b1]
Dogs	0.45± 0.32	0.57 ± 0.27
Ducks	0.72 ± 0.23	0.77 ± 0.22
Cattle	0.54 ± 0.34	0.53 ± 0.28
Chicken	0.72 ± 0.26[a2]	0.82 ± 0.18[b2]
Pigs	0.68 ± 0.28	0.73 ± 0.24
Sheep	0.41 ± 0.38	0.44 ± 0.27
Deer	0.59 ± 0.32[a3]	0.85 ± 0.07[b3]
Kangaroos	0.64 ± 0.20	0.72 ± 0.14

P= < 0.2 for *a*1 vs. *b*1 and *a2* vs. *b*2; P= < 0.005 for *a*3 vs. *b*3. Di: Diversity index.

TABLE 3. Number of unique (UQ) and shared (SH) biochemical phenotypes (BPTs) in host groups.

Host sources	No. of UQ-BPTs[a] (% over total BPTs)		No. of SH-BPTs[b] (% over total BPTs)	
	Enterococci	*E. coli*	Enterococci	*E. coli*
Human:	66 (70)	69(75)	28(30)	23(25)
Animals:				
Horses	54 (71)	32(53)	22(29)	28(47)
Dogs	24 (49)	32(50)	25(51)	32(50)
Ducks	29 (50)	32(46)	29(50)	37(54)
Cattle	23 (49)	24(45)	24(51)	29(55)
Chicken	41 (55)	33(56)	33(45)	26(44)
Pigs	28 (52)	25(47)	26(48)	28(53)
Sheep	7 (33)	11(37)	14((67)	19(63)
Deer	13 (46)	9(29)	15(54)	22(71)
Kangaroos	10 (40)	6(32)	15(60)	13(68)
Total	295 (56)	273(51)	231(44)	257(49)

[a] Identical BPTs within each host group are not included. [b] BPTs found in multiple host groups.

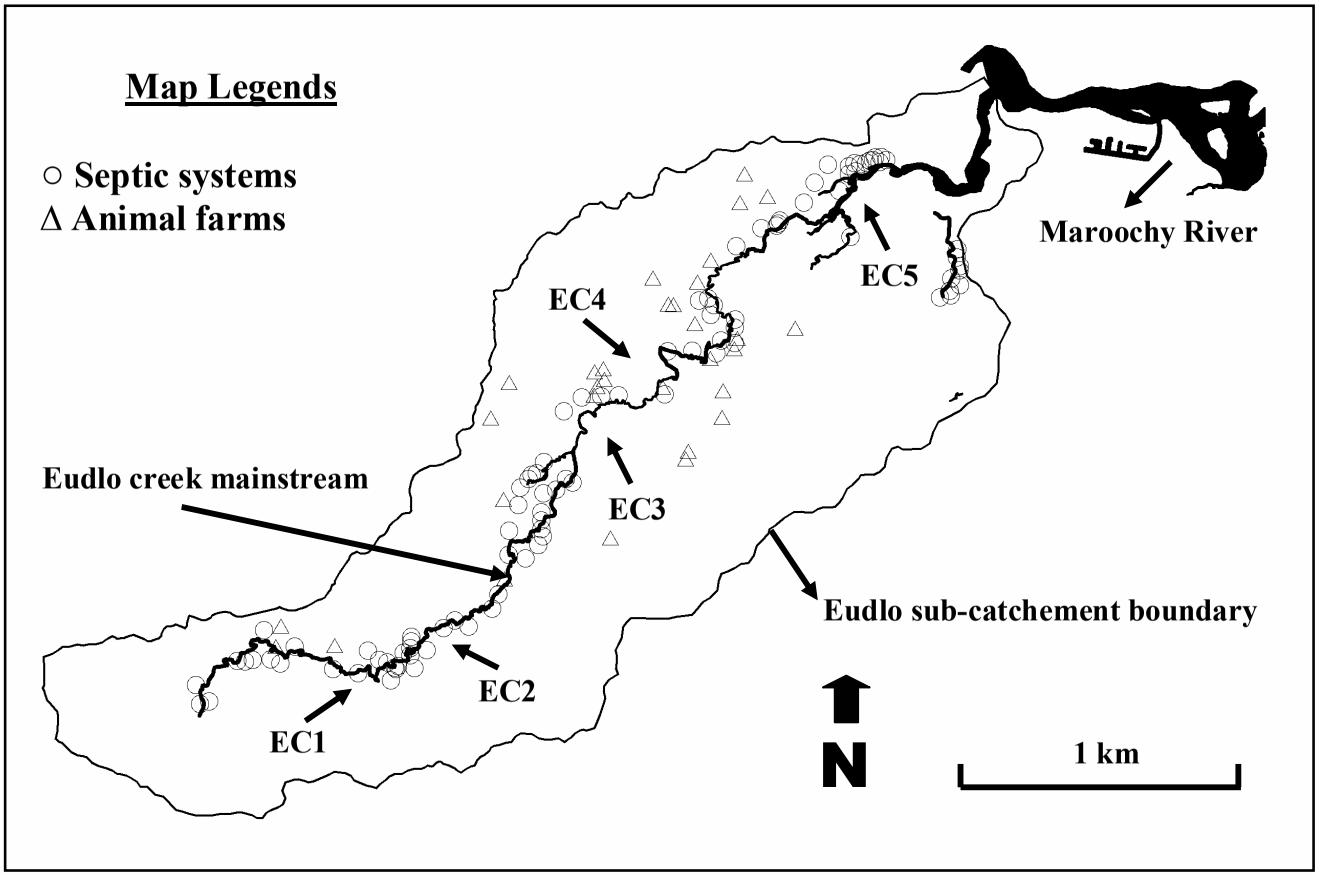
TABLE 4: Comparison of BPTs from water samples with the database.

Sampling occasion	Sampling sites	No. of isolates tested (no. of total-BPTs found)		No. of total-BPTs identical to database				Unknown BPTs	
				Human UQ-BPTs		Animal-BPTS			
		Enterococci	*E. coli*	Enterococci	*E. coli*	Enterococci	*E. coli*	Enterococci	*E. coli*
1	EC1	32 (10)	32 (12)	2	1	6	5	2	6
	EC2	38 (15)	25 (15)	1	4	9	11	5	0
	EC3	39 (16)	14 (11)	1	1	13	8	2	2
2	EC1	29 (9)	23 (14)	1	3	5	11	3	0
	EC2	39 (13)	22 (12)	1	2	6	8	6	2
	EC3	40 (10)	65 (26)	2	4	7	17	1	5
3	EC1	38 (12)	19 (10)	2	2	8	5	2	3
	EC2	36 (14)	21 (14)	3	3	8	11	3	0
	EC3	39 (17)	23 (8)	3	-	9	6	5	2
4	EC1	22 (9)	18 (10)	1	1	6	5	2	4
	EC2	23 (10)	19 (12)	-	-	8	5	2	7
	EC3	23 (8)	20 (12)	-	2	3	5	5	5
	EC4	23 (6)	17 (13)	-	3	3	5	3	5
	EC5	23 (10)	10(16)	-	-	6	4	4	6
5	EC1	23 (7)	7 (5)	1	-	4	4	2	1
	EC2	23 (7)	14 (6)	1	-	4	2	2	4
	EC3	23 (7)	11 (8)	1	-	5	3	1	5
	EC4	23 (8)	-	-	-	6	-	2	-
	EC5	23 (7)	13 (11)	-	2	4	4	3	5
6	EC1	23 (9)	21 (10)	-	2	7	4	2	4
	EC2	23 (6)	29 (18)	1	2	4	4	1	12
	EC3	23 (6)	7 (5)	-	-	5	2	1	3
	EC4	23 (8)	-	1	-	5	-	2	-
	EC5	23 (6)	11 (8)	-	1	3	3	3	4
7	EC1	39 (4)	33 (10)	2	2	2	3	-	5
	EC2	39 (7)	33 (7)	1	-	3	6	3	1
	EC3	39 (7)	37 (15)	1	1	3	10	3	4
Total	N=27	791 (248)	550 (282)	26	36	152	151	70	95

Legend to figures

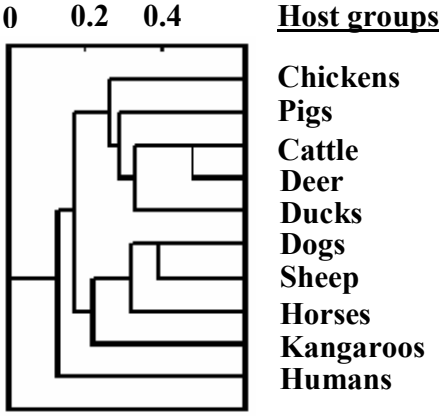
Fig. 1: Sampling sites (EC1-EC5) on Eudlo creek mainstream. Conventional septic systems (○) within 50 m distance of the creek and animal farms (Δ).

Fig. 2: A UPGMA dendrogram of population similarity (Sp-value) of enterococci (a) and *E. coli* (b) populations from all host groups.


Map Legends
○ Septic systems
Δ Animal farms
Maroochy River
EC5
EC4
EC3
EC2
EC1
Eudlo creek mainstream
Eudlo sub-catchement boundary
N
1 km

Population similarity

a.



Population similarity

b.