University of the Sunshine Coast
Queensland, Australia

USC RESEARCH BANK

http://research.usc.edu.au

This is the author accepted version of the following publication:

Vollmerhausen, T L, Ramos, N, Gundagdu, A, Robinson, W A, Brauner, A, Katouli, M (2011) Population structure and uropathogenic virulence associated genes of faecal Escherichia coli from healthy young and elderly adults. Journal of Medical Microbiology. Vol. 60, No. 5, pp.574-584. DOI:10.1099/jmm.0.027037-0

PERMISSIONS
Permission has been granted by the copyright holder to deposit this author accepted version as Open Access in the USC Research Bank. Open Access research is digital, online and free of charge, and is made possible by the consent of the author or copyright holder.

This is an author manuscript that has been accepted for publication in Journal of Medical Microbiology, copyright Society for General Microbiology, but has not been copy-edited, formatted or proofed. Cite this article as appearing in Journal of Medical Microbiology. This version of the manuscript may not be duplicated or reproduced, other than for personal use or within the rule of 'Fair Use of Copyrighted Materials' (section 17, Title 17, US Code), without permission from the copyright owner, Society for General Microbiology. The Society for General Microbiology disclaims any responsibility or liability for errors or omissions in this version of the manuscript or in any version derived from it by any other parties. The final copy-edited, published article, which is the version of record, can be found at http://jmm.sgmjournals.org, and is freely available without a subscription.

Copyright © 2011 Society for General Microbiology.

Journal of Medical Microbiology Papers in Press. Published February 3, 2011 as doi:10.1099/jmm.0.027037-0

1 # Population structure and uropathogenic virulence-
2 # associated genes of fecal *Escherichia coli* of healthy young
3 # and elderly adults

4

5 Tara L. Vollmerhausen[a], Nubia L. Ramos[a], Aycan Gündoğdu[a], Wayne Robinson[b]

6 Annelie Brauner[c], Mohammad Katouli[a]*

7 *[a]Faculty of Science, Health and Education, University of the Sunshine Coast, Maroochydore*

8 *DC, Queensland 4558, Australia*

9 *[b] School of Environmental Sciences, Charles Sturt University, Thurgoona, NSW, 2641,*

10 *Australia*

11 *[c]Department of Clinical Microbiology, MTC , Karolinska University Hospital, S-171 76*

12 *Stockholm, Sweden*

13 **Subject category:** Pathogenicity and virulence

14 **Running title:** Uropathogenic *E. coli* in healthy adults

15 ***Corresponding author:** Dr Mohammad Katouli, Faculty of the Science, Health and
16 Education, University of the Sunshine Coast, Queensland, Maroochydore DC 4558,
17 Australia.

18 Email: mkatouli@usc.edu.au

19 Phone: (+61) 7 5430 2845

20 Fax: (+61) 7 5430 2887

21

22 **Summary**

23 We investigated population structures of faecal *Escherichia coli* in 30 healthy young (13
24 males and 17 females) aged ranging between 20 and 45 years and 29 elderly (14 females and
25 15 males) aged ranging between 65 and 77 years. In all, 1566 strains were typed with the
26 PhPlate system and grouped into biochemical phenotypes (BPTs). Strains with shared BPTs
27 were further typed using randomly amplified polymorphic DNA analysis. Forty-four percent
28 of the strains were shared between two or more age and gender groups. Elders had a
29 significantly higher ($p < 0.001$) number of BPTs (3.3 ± 0.27 mean $\pm$ standard error) than
30 younger groups (1.82 ± 0.27). Phylogenetic affiliation and virulence-associated genes (VAG)
31 of the strains showed that more than 80% of the strains belonging to dominant types belonged
32 to phylogroups B2 and D. Amongst dominant BPTs, phylogenetic group A was significantly
33 associated with females ($p < 0.0001$), and elders were more likely to carry group D ($p <$
34 0.0124). Elderly males had a higher prevalence of VAGs than young males ($p < 0.0001$) and
35 young females ($p < 0.0005$). We conclude that healthy young adults carry less *E. coli* with
36 uropathogenic properties than elders.

37

38 ## Introduction

39 Urinary tract infection (UTI) is one of the most common diseases of young women
40 worldwide, with 60% of women experiencing at least one episode during their lifetime
41 (Foxman *et al.*, 2000). Young women are more susceptible to UTI than men, with the
42 shortness of the female urethra a predisposing factor to infection (Foxman, 2003). Other
43 factors also increase a woman's likelihood of infection, including sexual activity, oral
44 contraceptives, estrogen deficiency, diabetes, obstructing lesions, and genetic factors such as
45 blood group secretor status (Harrington & Hooton, 2000; Scholes *et al.*, 2000). Among the
46 causative agents of UTI, *Escherichia coli* is responsible for >90% of cases and is almost
47 always found amongst the faecal flora of the same host, although these bacteria may be
48 harboured in the vagina or periurethra (Czaja *et al.*, 2009; Moreno *et al.*, 2008).
49
50 The intestinal niche can harbour a range of commensal and pathogenic *E. coli* strains and has
51 been elucidated as a reservoir of strains capable of causing extraintestinal infection. It is
52 widely accepted that uropathogenic *E. coli* (UPEC) strains largely belong to phylogenetic
53 groups B2, and to a lesser extent group D, and intrinsically harbour a range of virulence
54 genes (Johnson *et al.*, 2005; Moreno *et al.*, 2008; Picard *et al.*, 1999; Sannes *et al.*, 2004).
55 Commensal strains, on the other hand, generally belong to phylogenetic groups A and B1,
56 and are associated less frequently with disease (Bingen *et al.*, 1998; Boyd & Hartl, 1998;
57 Duriez *et al.*, 2001; Johnson & Stell, 2000). All of these groups are commonly isolated from
58 the gut of healthy humans, with groups B2 and D commonly associated with dominant
59 populations (Moreno *et al.*, 2009; Zhang *et al.*, 2002).
60

61 UPEC have virulence-associated genes (VAGs) that facilitate their adherence to the opening
62 of the urinary tract and allow them to ascend to the bladder to cause UTI. Type 1 fimbriae is
63 important for initiating adhesion, although it is the P-pili that is regarded as an essential
64 virulence factor for adhesion of *E. coli* to the uroepithelium (Abraham *et al.*, 2001; Arthur *et*
65 *al.*, 1989; Wright *et al.*, 2007). Nonetheless, these VAGs also have an important role for
66 survival of these strains within the gut, enhancing their colonisation as dominant strains
67 (Tullus *et al.*, 1992; Wold *et al.*, 1992; Wold *et al.*, 1988). We hypothesized that in addition
68 to predisposing factors for UTI mentioned above, the dominant intestinal *E. coli* populations
69 of healthy young and elderly females may be dominantly comprised of UPEC strains than
70 males of the same age groups, making them more vulnerable to UTI. Furthermore, the
71 population structure, diversity and the prevalence of specific VAGs associated with UPEC
72 among *E. coli* from healthy young and elderly adults is not well understood. The aim of this
73 study was therefore to investigate the prevalence of uropathogenic VAGs among faecal *E.*
74 *coli* of healthy young and elderly adults and to determine whether differences in *E. coli*
75 population structures of the gastrointestinal tract could indicate predisposition to UTI.

76

77 **Materials and methods**

78 **Subjects.** This study was approved by the University of the Sunshine Coast Human Research
79 Ethics Committee and participation entailed providing informed consent and returning a self-
80 collected fecal swab. Between August and November 2009, a total of 59 healthy community
81 dwelling individuals were registered for this study. These included 17 healthy young females
82 aged ranging between 21 - 45 years (33.2 ± 9.7, mean age ± standard deviation), 13 healthy
83 young males aged between 20 - 45 years (26.5 ± 8.9), 14 healthy elderly females aged
84 between 65 and 77 years (67.9 ± 3.5) and 15 healthy elderly males aged between 65 and 79

85 years (72.1 ± 5.2). Young subjects were recruited after an invitation email to staff and
86 students at the University of the Sunshine Coast community. Elderly subjects were also
87 recruited via the University of the 3rd Age Community Group at University of the Sunshine
88 Coast. Care was taken to recruit subjects that were not from the same house hold but lived in
89 the Sunshine Coast region of Queensland, Australia. None of the subjects had a symptomatic
90 UTI in the preceding six months, nor had taken any antibiotics or commercially available
91 probiotics in the three months prior to participating in the study. In female groups an
92 additional criteria was the lack of known pregnancy.
93

94 **Isolation and confirmation of *E. coli* strains.** Each volunteer returned a self-collected fecal
95 swab inserted into Amies transport media containing charcoal. Swabs were kept at 4 ^{o}C and
96 streaked directly onto MacConkey agar no. 3 and membrane faecal coliform agar (m-FC
97 agar) within 24 hours of collection. Plates were incubated for 24 hours at 37 °C and 28 *E.*
98 *coli*-like colonies (where possible) were randomly selected and tested. This number of
99 colonies provide >90% chance of detecting minor clones amongst the host *E. coli* faecal
100 population as previously described (Schlager *et al.*, 2002). Extraction of chromosomal DNA
101 was done by growing a single colony of the isolates in Luria-Bertani (LB) broth overnight
102 and collecting pellet in 200 µl of sterile Milli-Q water and boiling at 100 °C for 15 mins. In
103 all, 1566 putative *E. coli* strains were collected and subjected to confirmatory tests using PCR
104 amplification of the universal stress protein (*uspA*) gene as previously described (Chen &
105 Griffiths, 1998).
106

107 **Biochemical fingerprinting.** All strains were typed using a high resolution biochemical
108 fingerprinting method (PhPlate system), specifically developed for *E. coli* strains (PhP-RE,

109 Bactus AB). The fingerprinting method was performed according to manufacturer
110 instructions. Briefly, *E. coli* colonies were suspended in the first well of each row containing
111 325 µl of growth medium comprised of 0.011% (wt/vol) bromothymol blue and 0.1%
112 (wt/vol) proteose peptone. Aliquots of 25 µl of bacterial suspensions were transferred into
113 each of the other 11 wells containing 150 µl of growth medium. Plates were then incubated at
114 37 °C and read at intervals of 7, 24 and 48 hours. Images of plates at corresponding times
115 were scanned using a HP Scanjet 4890 scanner. After the final scan, the PhPlate software
116 (PhPWin4.2) was used to create absorbance data from the scanned PhP-RE plate images.
117 After the final reading of the plate images, the mean of the absorbance values from all
118 individual readings was calculated for each reagent, creating the biochemical fingerprint for
119 each isolate (Landgren *et al.*, 2005). Similarity among the isolates was calculated as a
120 correlation coefficient, and clustered according to UPGMA to yield a dendrogram. An
121 identity level of 0.965 was established based on the reproducibility of the system after testing
122 60 isolates in duplicate. Isolates showing similarity to each other above the identity level
123 were regarded as identical and assigned to the same biochemical phenotype (BPT). BPTs
124 containing more than one isolate were called common (C) BPTs and those with one isolate
125 named as single (S) BPTs. Diversity among the isolates was calculated as a diversity index
126 (Di) using Simpson's index of diversity. All data handling, including calculations of
127 correlations and coefficients, diversity indices as well as clustering were performed using the
128 PhPlate software version 4002 (PhPlate AB).
129
130 **Randomly amplified polymorphic DNA analysis.** Strains that were identical in two or more
131 individuals or groups as determined by the biochemical fingerprinting method were further
132 typed by randomly amplified polymorphic DNA (RAPD) PCR. The RAPD analysis was

133 performed with the KG (5' ACACGCACACGGAAGAA 3') primer (Ramos *et al.*, 2010).
134 The PCR reaction was carried out in a 50 µL volume containing 31.3 µl of sterile Milli-Q
135 water, 6 µl of 10x PCR reaction buffer, 3.8 µl of 50mM $MgCl_2$, 1 µl of 10mM
136 deoxynucleoside triphosphates (dNTPs) (Fisher Biotec), 0.4 µl of *Taq* polymerase (Bioline), 6
137 µl of primer (20-pM/µl) (Invitrogen) and 1.5 µl of DNA. PCR reactions were amplified as
138 previously described (Ramos *et al.*, 2010) and products were size separated by
139 electrophoresis in 2.2% agarose stained with ethidium bromide at 90V for 180 min.
140 Molecular weight markers (100bp and 1kb) were used in all of the experiments. Similarity
141 between the banding patterns were compared visually and divided into similarity groups
142 (RAPD types). Strains belonging to the same C-BPT and shared the same RAPD pattern were
143 regarded as common types (CTs) while those with different RAPD patterns but identical BPT
144 or identical RAPD pattern but different BPTs were regarded as single types (STs).
145

146 **Phylogenetic grouping and virulence-associated genes.** All confirmed *E. coli* strains were
147 tested for their phylogenetic groups (A, B1, B2 and D) using a triplex PCR method for *chu*A,
148 *yja*A and TSPE4.C2 as previously described (Clermont *et al.*, 2000). The presence of *ibeA*
149 gene among Clermont phylogenetic group D (i.e. *chuA* and TSPE4.C2 positive and *yjaA*
150 negative) was also investigated and those harbouring *ibeA* genes were regarded as belonging
151 to phylogenetic group B2 as described before (Gordon *et al*., 2008).
152

153 Strains were also tested for 12 VAGs known to be associated with UTI. These included
154 adhesin genes for P-pili (*papAH*, *papEF*, *papC*, *papG* allele II and *papG* allele III) and type 1
155 fimbriae (*fimH*); toxins genes α-haemolysin (*hlyA*) and cytotoxic necrotizing factor 1 (*cnf1*);
156 siderophores genes ferric aerobactin receptor (*iutA*) and a catecholate siderophore ($iroN_{E.coli}$);

157 K1 capsule gene (*kpsMT* K1) and group II capsular polysaccharide synthesis e.g. K1, K5, and

158 K12 (*kpsMT* II). Genes were grouped into four multiplex primer sets and two uniplex sets in

159 a modified method to that previously described (Johnson & Stell, 2000), with appropriate

160 positive and negative controls.

161

162 The multiplex PCR reactions consisted of 2.5 µl 10x reaction buffer, 5 µl 2 mM dNTPs, 2.5

163 µl 25 mM $MgCl_2$, 0.3 µl of each primer from 50-pM/µl stock solution for *papAH*, *papEF*,

164 *fimH*, *iutA*, *papG* allele III, *kpsMT* K1,and *papG* allele II and III (0.15 µl for *kpsMT* II, *papC*,

165 and *cnf1)*, 0.1 µl Hotstar *Taq* polymerase (Qiagen) (0.2 µl for *papG* allele II and III and

166 *cnf1*), 2 µl DNA, and sterile Milli-Q water to make up 25 µl reaction volume. Amplification

167 used the following cycles: denaturation for 15min at 95 °C; 25 cycles at 94 °C for 30 s, 63 °C

168 for 30 s, 68 °C for 3 mins; and final extension of 72 °C for 10 mins.

169

170 The uniplex PCR reactions consisted of 2.5 µl 10x reaction buffer (3.0 µl for $iroN_{E.coli}$), 5 µl

171 2 mM dNTPs, 1.0 µl 50 mM $MgCl_2$, 0.3 µl of each primer from 50-pM/µl stock solution for

172 *papG* allele II (0.4 µl for $iroN_{E.coli}$), 0.15 µl *Taq* polymerase (Bioline) (0.1 µl for $iroN_{E.coli}$), 2

173 µl DNA, and sterile Milli-Q water to make up 25 µl reaction volume (28 µl for $iroN_{E.coli}$).

174 Amplification used the following cycles: denaturation for 4 min at 95 °C; 25 cycles at 94 °C

175 for 30 s, 63 °C for 30 s, 68 °C for 3 mins (1 min for $iroN_{E.coli}$); and final extension of 72 °C

176 for 10 mins. All PCR products were size separated by electrophoresis in 2% agarose strained

177 with ethidium bromide at 90V for 90 min.

178

179 **Definitions.** C-BPTs containing more than 80% of the tested isolates in each subject were

180 regarded as dominant BPTs. Some individuals harboured two C-BPTs. To evaluate whether

181 these two C-BPTs were true dominant populations and in order to do a quality control of the
182 ability of the biochemical fingerprinting to diversify strains from each individual we
183 confirmed the presence of two populations in these individual with RAPD PCR (available as
184 supplementary data in the online journal). Under these conditions both BPTs/RAPD-PCR
185 were regarded as dominant if their population sizes did not differ from each other by more
186 15% and together their population size exceeded 90% of total isolates tested in each subject.
187 All other BPTs were regarded as minor BPTs.
188
189 A VAG score was first established for each BPT. Each VAG was given a score of one
190 (adjusted for multiple detection of *pap* genes i.e. *papAH*, *papC*, *papEF*, *papG* allele II and III
191 and capsule genes i.e. *kpsMT* II and *kpsMT* K1). VAG score for shared BPTs was the VAG
192 score for each BPT multiplied by the number of isolates it represented, divided by the total
193 number of isolates represented in each shared BPT/RAPD type.
194
195 **Statistical analysis.** Factorial analysis of variance was used to compare diversity and
196 abundance of BPTs between age and gender groups and residual analyses was used to
197 investigate the assumptions of normality and homogeneity of variances. A square root
198 transformation was applied to meet the assumption of homogeneity of variances for the
199 number of BPTs. The numbers of VAGs per dominant and minor BPT's were compared
200 using the Wilcoxon Rank Sum test with a normal approximation correction for large sample
201 sizes.
202 A log-linear model was used to compare the distribution of dominant and minor BPTs
203 between the phylogenetic groups and gender and age categories. When there was a significant

204 interaction of effects, we partitioned the analysis and applied Benjamini–Hochberg method to
205 moderate the false discovery rate (Waite & Campbell, 2006).
206
207 **Results**

208 **BPT diversity amongst groups.** Of the 1566 putative *E. coli* strains, 1541 were confirmed as
209 *E. coli* using the *uspA* gene. Biochemical fingerprinting of these isolates yielded 112 C-BPTs
210 comprising of 1502 isolates and 39 S-BPTs. The number of BPTs found in each subject
211 ranged from 1.7 ± 0.9 in young males to 3.8 ± 2.0 in elderly males (Table 1). There was a
212 significant difference between elderly (3.3 ± 0.27 mean ± standard error) and younger (1.82 ±
213 0.27) subjects ($F_{agegroup}$=12.3, df = 1,55, p < 0.001). There were however, no differences when
214 the numbers of BPTs in two genders were compared. Similarly, the diversity of *E. coli* strains
215 found among elderly (0.35 ± 0.046) subjects was significantly higher than those found in
216 younger subjects (0.20 ± 0.045) ($F_{agegroup}$= 4.97, df = 1.55, p<0.05) (Table 1). Again, there
217 was no difference between the diversity of *E. coli* strains among genders.
218

219 **Phylogenetic groups.** A total of 63 dominant BPTs were identified among all 59 subjects
220 (four subjects had two dominant BPTs). Of these, 80% belonged to phylogenetic groups B2
221 and D (Table 2). The prevalence of phylogentic group was dependant on age and gender (χ^2
222 = 8.19, df = 3, p < 0.0423), partitioning the analysis indicated that gender was significantly
223 associated with phylogenetic group A (χ^2 = 16, df = 1, p < 0.0001) (Table 2) and that only
224 females carried the *E. coli* belonging to this phylogenetic group. Elderly were more likely to
225 be carriers of *E. coli* belonging to phylogenetic group D than young groups (χ^2 = 6.25, df = 1,
226 p < 0.0124) (Table 2). There were no differences for phylogenetic groups B1 or B2 between
227 different age and gender groups. Elderly subjects had significantly more minor BPTs than

228 younger subjects ($\chi^2 = 121.0$, df = 1, $p < 0.0001$) and phlylogenetic group D had significantly
229 more minor BPTs than the other groups ($\chi^2 = 118.5$, df = 3, $p < 0.0001$). The tendency for
230 *E. coli* to belong to group D was even more pronounced in elderly subjects ($\chi^2 = 31.5$, df = 1,
231 $p < 0.0001$).

232

233 **Distribution of strains among individuals.** Biochemical fingerprinting identified 63
234 dominant BPTs and 88 minor BPTs. These 151 BPTs were further typed using RAPD-PCR
235 to confirm the clonality of the isolates by two typing methods (i.e. BPT/RAPD types). It was
236 found that 50 BPTs, representing 44% of isolates (n=674) that shared between two or more
237 age and gender groups, belonged to 18 BPT/RAPD types (Fig. 1). These BPT/RAPD types
238 were named as common types (CTs). Of these, CT 5 and CT 14 constituted more than 24 %
239 of the shared isolates and all belonged to phylogenetic groups B2 and D. These CTs
240 contained strains with the highest VAG scores (Fig. 1).

241

242 **Virulence traits.** Comparison of dominant and minor BPTs found dominant BPTs (2.48 ±
243 0.185) had a significantly ($z = 2.25$, n = 63, 88, $p = 0.0244$) higher VAG scores than minor
244 ones (1.98 ± 0.137). There was a significant difference in the prevalence between the age
245 groups ($\chi^2 = 22.13$, df = 1, $p < 0.0001$) and some VAGs occurred more frequently than others
246 ($\chi^2 = 250.23$, df = 12, $p < 0.0001$) (Table 3). There was a barely not significant interaction
247 between age and gender of subjects ($\chi^2 = 2.97$, df = 1, $p < 0.10$) (Table 3). A follow up
248 analysis found that elderly males were significantly higher in the prevalence of VAGs than
249 either young males ($p < 0.0001$) or young females ($p < 0.0005$) and elderly females were
250 higher than young males ($p < 0.05$).

251

252 Discussion

253 In this study we used a high resolution biochemical fingerprinting specifically developed for
254 typing of *E. coli* strains. The system has been shown to be as powerful as other molecular
255 typing methods such as RAPD-PCR (Ramos *et al.*, 2010) and ERIC-PCR (Ansaruzzaman *et*
256 *al.*, 2000) for typing of *E. coli* strains. From each participant up to 28 colonies were tested,
257 providing greater than 90% chance of detecting minor clones amongst the host *E. coli* fecal
258 population (Schlager *et al.*, 2002). Based on this number of isolates, we found that elderly
259 males had greater diversity of *E. coli* strains than both young males and females. Increased
260 diversity amongst dominant bacterial species has been observed amongst healthy elderly
261 subjects (Saunier & Doré, 2002). One possible reason could be changes in the mucosal
262 surfaces amongst elderly subjects as also suggested by others (Ouwehand *et al.*, 1999). It has
263 also been postulated that mucous production amongst elderly subjects differs than those of
264 young adults, containing more carbohydrates and less protein (Ouwehand *et al.*, 1999).
265 Furthermore, decreased mucous production amongst elderly subjects may expose more
266 receptors on the gut epithelium as postulated before (Ouwehand *et al.*, 1999).

267

268 All subject groups, except for young males, carried populations of *E. coli* belonging to
269 phylogenetic group D as the major gut population. Phylogenetic group B2 was the second
270 dominant group for elders. Young females carried strains belonging to phylogenetic group D
271 as their largest population size, followed closely by group A, as found previously (Moreno *et*
272 *al.*, 2009). Amongst dominant types belonging to phylogenetic group A, young females had
273 the largest population, which intrinsically carry less extraintestinal virulence factors than
274 groups B2 and D (Clermont *et al.*, 2000; Johnson *et al.*, 2001). Similarly, young males and to

275 a lesser extent, young females also carried strains belonging to B1 group which is also less
276 associated with UTI. These results suggest that faecal dominance of an *E. coli* strain in
277 healthy young females is not a predicting factor of UTI. Similar to this finding, Schlager *et*
278 *al.* (2003) identified amongst girls aged 3 – 6 years, dominant faecal clones with P-pili were
279 less likely to be found in the urinary tract than minor clones. Interestingly, young females in
280 our study carried less *E. coli* with uropathogenic properties and lower VAG scores, although
281 this was not a significant result due to the low sample size. This may indicate that other
282 factors, such as shortness of urethra, sexual activity or oral contraceptives, are more
283 important than the structure of dominant *E. coli* populations of young females for their
284 susceptibility to UTI. Capsule antigens were the only genes found more in females, although
285 this was not a significant difference in this study, a larger sample size is needed to determine
286 whether females carry more capsular antigen K1 than males. This was an interesting
287 observation as the K1 antigen is a key VAG in extraintestinal pathogenic *E. coli* associated
288 with neonatal meningitis (Öhman *et al.*, 1995) and perhaps higher faecal K1 strains in this
289 group is involved in contamination and infection of neonates after birth.

290

291 Comparison of the *E. coli* types among the four groups showed a high proportion, 44 % of
292 the *E. coli* tested in this study, were common among the groups with two major groups
293 accounting for more than 24 % of the shared populations. The fact that these CTs belonged to
294 pathogenic groups B2 and D and had the highest VAG scores suggests that certain *E. coli*
295 clones of the dominant flora of the gut of healthy individuals are better equipped to cause
296 UTI. Alternatively, these strains may be better adapted for the intestinal niche (Diard *et al.*,
297 2010). Contrary to our expectation, we found that clones with increased virulence factors
298 were more prevalent among elderly males than young females. It has to be noted that in our

299 study we tested *E. coli* from a small number of individuals (between 13 and 17 individuals in
300 each group) and therefore confirmation of such finding requires testing more individuals.
301
302 In this study, elderly males had a significantly higher prevalence of the tested VAGs.
303 Contrary to our initial hypothesis that young females might have highly virulent dominant *E.*
304 *coli* as a predisposition to UTI, elderly groups, especially males, showed significantly higher
305 prevalence of VAGs. This finding not only indicates that dominant *E. coli* populations of
306 healthy young females may not be the main cause of UTI in this group, but helps to explain
307 the reported higher prevalence of extraintestinal infections such as UTI and septicaemia in
308 hospitalised elderly patients, especially males, as more virulent strains were carried by these
309 groups (Kaye, 1980).
310
311 Our results showed that dominant *E. coli* strains have more VAGs than minor strains as
312 previously reported (Moreno *et al.*, 2009). Apart from their role in adhering to the urinary
313 tract cells, P-pili have also been shown to bind to the intestinal tract which enhances their
314 colonisation of the gut (Tullus *et al.*, 1992; Wold *et al.*, 1992; Wold *et al.*, 1988).
315 Pathogenicity-associated islands which carry many extraintestinal virulence factors have been
316 shown to enhance fitness of these strains to survive in the intestinal tract, with the ability to
317 cause extraintestinal infections being a coincidental by-product (Diard *et al.*, 2010). Our
318 unpublished data also indicates that *E. coli* strains carrying these adhesins bind with almost
319 the same rate to Caco-2 cells (an intestinal epithelial cell line) as they do to the renal cell line
320 A-489 cells (Katouli M). These finding explain why dominant BPTs in our study carried
321 more pap genes than minor ones.
322

323 Zhang *et al.* (2002) found P-pili amongst young women (aged 18 – 39 years) were strongly
324 associated with phylogenetic groups B2 and D. However group B2 had two distinct
325 subgroups with differing levels of pathogenicity. These researchers suggested that healthy
326 adults are capable of carrying B2 strains however with less virulent subclasses than UTI
327 isolates (Zhang *et al.*, 2002). In our study, B2 was the second major phylogenetic group after
328 group D (except in young adults) suggesting that the B2 strains may be a subset of pathogenic
329 strains that only increase in number and become dominant when there is a major upset of the
330 colonic *E. coli* leading to UTI. Furthermore, these results indicate that under healthy
331 conditions, and with regard to *E. coli* population structure of the gastrointestinal tract, elderly
332 females may be at a greater risk for UTI from their faecal intestinal flora than younger
333 females. Despite this, other factors such as sexual activity increase the likelihood of UTI
334 amongst young women (Hooton *et al*., 1996). Additionally, healthy males may serve as a
335 reservoir of uropathogens transmitted directly via sexual transmission (Foxman *et al*., 1997).
336 The low prevalence of uropathogenic virulence factors amongst young women in our study is
337 somewhat contradictory to previous reports of females with UTI which have found the
338 pathogenic urinary strains are often associated with dominant faecal strains (Moreno *et al.*,
339 2008). These results suggest that a dramatic change may occur in the intestinal *E. coli*
340 populations of young females before developing UTI. If this hypothesis is correct, then we
341 postulate that a sub-set of the intestinal *E. coli* populations, probably a minor clone, in young
342 females may find an opportunity to increase in population size, colonise the opening of the
343 urinary tract and become a short-term dominant clone during the course of UTI.
344

345 In conclusion, our findings indicate that the *E. coli* population of healthy young adults have a
346 lower prevalence of VAG than elderly males, and to a lesser extent, elderly females. Despite

347 this, uropathogenic strains are commonly reported as the dominant clones amongst the faecal
348 population in young women with UTI. We postulate that in young females a subset of *E. coli*
349 populations may increase in size to become a short-term dominant flora during the course of
350 UTI. We also conclude that the presence and dominance of VAGs in *E. coli* strains amongst
351 elderly males and females may be a risk factor for developing extraintestinal infections in
352 these groups.
353

354 **References**

355

356 **Abraham, S. N., Shin, J. S. & Malaviya, R. (2001).** Type 1 fimbriated *Escherichia coli* -
357 Mast cell interactions in cystitis. *J Infect Dis* **183**, S51- S55.
358 **Ansaruzzaman, M., Albert, M. J., Nahar, S., Byun, R., Katouli, M., Kühn, I. & Möllby,**
359 **R. (2000).** Clonal groups of enteropathogenic *Escherichia coli* isolated in case-control studies
360 of diarrhoea in Bangladesh. *J Med Microbiol* **49**, 177-185.
361 **Arthur, M., Johnson, C. E., Rubin, R. H., Arbeit, R. D., Campanelli, C., Kim, C.,**
362 **Steinbach, S., Agarwal, M., Wilkinson, R. & Goldstein, R. (1989).** Molecular
363 epidemiology of adhesin and hemolysin virulence factors among uropathogenic *Escherichia*
364 *coli*. *Infect Immun* **57**, 303-313.
365 **Bingen, E., Picard, B., Brahimi, N., Mathy, S., Desjardins, P., Elion, J. & Denamur, E.**
366 **(1998).** Phylogenetic analysis of *Escherichia coli* strains causing neonatal meningitis
367 suggests horizontal transfer from a predominant pool of highly virulent B2 group strains. *J*
368 *Infect Dis* **177**, 642-650.
369 **Boyd, E. F. & Hartl, D. L. (1998).** Chromosomal regions specific to pathogenic isolates of
370 *Escherichia coli* have phylogenetically clustered distribution. *J Bacteriol* **180**, 1159-1165.
371 **Chen, J. & Griffiths, M. W. (1998).** PCR differentiation of *Escherichia coli* from other
372 Gram-negative bacteria using primers derived from the nucleotide sequences flanking the
373 gene encoding the universal stress protein. *Lett Appl Microbiol* **27**, 369-371.
374 **Clermont, O., Bonacorsi, S. & Bingen, E. (2000).** Rapid and simple determination of the
375 *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* **66**, 4555-4558.
376 **Czaja, C. A., Stamm, W. E., Stapleton, A. E., Roberts, P. L., Hawn, T. R., Scholes, D.,**
377 **Samadpour, M., Hultgren, S. J. & Hooton, T. M. (2009).** Prospective cohort study of

378 microbial and inflammatory events immediately preceding *Escherichia coli* recurrent urinary
379 tract infection in women. *J Infect Dis* **200**, 528-536.
380 **Diard, M., Garry, L., Selva, M., Mosser, T., Denamur, E. & Matic, I. (2010).**
381 Pathogenicity-associated islands in extraintestinal pathogenic *Escherichia coli* are fitness
382 elements involved in intestinal colonization. *J Bacteriol* **192**, 48885-4893.
383 **Duriez, P., Clermont, O., Bonacorsi, S., Bingen, E., Chaventré, A., Elion, J., Picard, B.**
384 **& Denamur, E. (2001).** Commensal *Escherichia coli* isolates are phylogenetically
385 distributed among geographically distinct human populations. *Microbiology* **147**, 1671-1676.
386 **Foxman, B. (2003).** Epidemiology of urinary tract infections: incidence, morbidity, and
387 economic costs. *Dm-Dis Mon* **49**, 53-70.
388 **Foxman, B., Barlow, R., D'Arcy, H., Gillespie, B. & Sobel, J. D. (2000).** Urinary tract
389 infection: self-reported incidence and associated costs. *Ann Epidemiol* **10**, 509-515.
390 **Foxman, B., Zhang, L., Tallman, P., Andree, B. C., Geiger, A. M., Koopman, J. S.,**
391 **Gillespie, B. W., Palin, K. A., Sobel, J. D., Rode, C. K., Bloch, C. A. & Marrs, C. F.**
392 **(1997).** Transmission of uropathogens between sex partners. *J Infect Dis* **175**, 989-992.
393 **Gordon, D. M., Clermont, O., Tolley, H., & Denamur, E. (2008).** Assigning *Escherichia*
394 *coli* strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex
395 method. *Environ Microbiol* **10**, 2484-2496.
396 **Harrington, R. D. & Hooton, T. M. (2000).** Urinary tract infection risk factors and gender. *J*
397 *Gend Specif Med* **3**, 27-34.
398 **Hooton, T. M., Scholes, D., Hughes, J. P., Winter, C., Roberts, P. L., Stapleton, A. E.,**
399 **Stergachis, A. & Stamm, W. E. (1996).** A prospective study of risk factors for symptomatic
400 urinary tract infection in young women. *N Engl J Med* **335**, 468-474.

401 **Johnson, J. R., Delavari, P., Kuskowski, M. & Stell, A. L. (2001).** Phylogenetic
402 distribution of extraintestinal virulence-associated traits in *Escherichia coli*. *J Infect Dis* **183**,
403 78-88.
404 **Johnson, J. R., Owens, K., Gajewski, A. & Kuskowski, M. A. (2005).** Bacterial
405 characteristics in relation to clinical source of *Escherichia coli* isolates from women with
406 acute cystitis or pyelonephritis and uninfected women. *J Clin Microbiol* **43**, 6064-6072.
407 **Johnson, J. R. & Stell, A. L. (2000).** Extended virulence genotypes of *Escherichia coli*
408 strains from patients with urosepsis in relation to phylogeny and host compromise. *J Infect*
409 *Dis* **181**, 261-272.
410 **Kaye, D. (1980).** Urinary tract infections in the elderly. *Bull N Y Acad Med* **56**, 209-220.
411 **Landgren, M., Odén, H., Kühn, I., Österlund, A. & Kahlmeter, G. (2005).** Diversity
412 among 2481 *Escherichia coli* from women with community-acquired lower urinary tract
413 infections in 17 countries. *J Antimicrob Chemother* **55**, 928-937.
414 **Moreno, E., Andreu, A., Pigrau, C., Kuskowski, M. A., Johnson, J. R. & Prats, G.**
415 **(2008).** Relationship between *Escherichia coli* strains causing acute cystitis in women and the
416 fecal *E. coli* population of the host. *J Clin Microbiol* **46**, 2529-2534.
417 **Moreno, E., Johnson, J. R., Pérez, T., Prats, G., Kuskowski, M. A. & Andreu, A. (2009).**
418 Structure and urovirulence characteristics of the fecal *Escherichia coli* population among
419 healthy women. *Microbes Infect* **11**, 274-280.
420 **Öhman, L., Tullus, K., Katouli, M., Burman, L. & Stendahl, O. (1995).** Correlation
421 between susceptibility of infants to infection and interaction with neutrophils of *Escherichia*
422 *coli* strains causing neonatal and infantile septicemia. *J Infect Dis* **171**, 128-133.

423 **Ouwehand, A. C., Isolauri, E., Kirjavainen, P. V. & Salminen, S. J. (1999).** Adhesion of
424 four *Bifidobacterium* strains to human intestinal mucus from subjects in different age groups.
425 *FEMS Microbiol Lett* **172**, 61-64.

426 **Picard, B., Garcia, J. S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E., Elion, J. &**
427 **Denamur, E. (1999).** The link between phylogeny and virulence in *Escherichia coli*
428 extraintestinal infection. *Infect Immun* **67**, 546-553.

429 **Ramos, N. L., Saayman, M. L., Chapman, T. A., Tucker, J. R., Smith, H. V., Faoagali,**
430 **J., Chin, J. C., Brauner, A. & Katouli, M. (2010).** Genetic relatedness and virulence gene
431 profiles of *Escherichia coli* strains isolated form septicaemic and uroseptic patients. *Eur J*
432 *Clin Microbiol Infect Dis* **29**, 15-23.

433 **Sannes, M. R., Kuskowski, M. A., Owens, K., Gajewski, A. & Johnson, J. R. (2004).**
434 Virulence factor profiles and phylogenetic background of *Escherichia coli* isolates from
435 veterans with bacteremia and uninfected control subjects. *J Infect Dis* **190**, 2121-2128.

436 **Saunier, K. & Doré, J. (2002).** Gastrointestinal tract and the elderly: functional food, gut
437 microflora and healthy ageing. *Dig Liver Dis* **34**, S19-S24.

438 **Schlager, T. A., Hendley, J. O., Bell, A. L. & Whittam, T. S. (2002).** Clonal diversity of
439 *Escherichia coli* colonizing stools and urinary tracts of young girls. *Infect Immun* **70**, 1225-
440 1229.

441 **Schlager, T. A., Whittam, T. S., Hendley, J. O., Bhang, J. L., Wobbe, C. L. & Stapleton,**
442 **A. (2003).** Variation in frequency of the virulence-factor gene in *Escherichia coli* clones
443 colonizing the stools and urinary tracts of healthy prepubertal girls. *J Infect Dis* **188**, 1059-
444 1064.

445 **Scholes, D., Hooton, T. M., Roberts, P. L., Stapleton, A. E., Gupta, K. & Stamm, W. E.**
446 **(2000).** Risk factors for recurrent urinary tract infection in young women. *J Infect Dis* **182**,
447 1177-1182.
448 **Tullus, K., Kühn, I., Ørskov, I., Ørskov, F. & Möllby, R. (1992).** The importance of P and
449 type 1 fimbriae for the persistence of *Escherichia coli* in the human gut. *Epidemiol Infect*
450 **108**, 415-421.
451 **Waite, T. A. & Campbell, L. G. (2006)** Controlling the false discovery rate and increasing
452 statistical power in ecological studies. *Ecoscience* **13**, 439-442.
453 **Wold, A. E., Caugant, D. A., Lidin-Janson, G., deMan, P. & Svanborg, C. (1992).**
454 Resident colonic *Escherichia coli* strains frequently display uropathogenic characteristics. *J*
455 *Infect Dis* **165**, 46-52.
456 **Wold, A. E., Thorssén, M., Hull, S. & SvanborgEdén, C. (1988).** Attachment of
457 *Escherichia coli* via mannose- or Galα1→ 4Galβ- containing receptors to human colonic
458 epithelial cells. *Infect Immun* **56**, 2531-2537.
459 **Wright, K. J., Seed, P. C. & Hultgren, S. J. (2007).** Development of intracellular bacterial
460 communitities of uropathogenic *Escherichia coli* depends on type 1 pili. *Cell Microbiol* **9**,
461 2230-2241.
462 **Zhang, L., Foxman, B. & Marrs, C. (2002).** Both urinary and rectal *Escherichia coli*
463 isolates are dominated by strains of phylogenetic group B2. *J Clin Microbiol* **40**, 3951-3955.
464

465 **Legend to Figure 1**

466 An unweighted pair group method with arithmetic mean (UPGMA) dendrogram showing
467 comparison between 50 *Escherichia coli* types shared between two or more individual
468 volunteers of different age and gender groups as determine by biochemical fingerprinting and
469 Randomly Amplified Polymorphic DNA (RAPD). The 18 common types (CT) are shown as
470 CT 1 – CT 18. RAPD pattern of each CT are also shown. Strains that were highly similar but
471 showed separate RAPD patterns were designated as sub-types using letters. Percentage of the
472 shared population and virulence-associated gene (VAG) scores are also shown. PGG:
473 Phylogenetic groups; A/G: age and gender group (YF: Young Female; YM: Young Male; EF:
474 Elderly Female; and EM: Elderly Male).

475 **Table 1.** Descriptive data of participants and the number of biochemical phenotypes (BPTs) of *Escherichia coli* isolates found in each

476 age and gender group.

Parameters	Young females (range)	Elderly females (range)	Young males (range)	Elderly males (range)
No. of participants	17	14	13	15
Mean age	33.2 ± 9.7 (21-45)	67.9 ± 3.5 (65-77)	26.5 ± 8.9 (20-45)	72.1 ± 5.2 (65-79)
No. of isolates tested/person	25.3 ± 5.6 (10-28)	25.7 ± 4.1 (15-28)	27.4 ± 1.9 (21-28)	25.8 ± 5.7 (24-28)
No. of BPT /person	1.9 ± 1.0 (1-4)	2.8 ± 1.7 (1-7)	1.7 ± 0.9 (1-3)	3.8 ± 2.0 (1-7)
Diversity index	0.22 ± 0.24 (0.0-0.62)	0.26 ± 0.22 (0.00-0.49)	0.18 ± 0.23 (0.00-0.61)	0.43 ± 0.28 (0.00-0.80)
No. of dominant BPTs	19 *	14	14 *	16 *
No. of minor BPTs	14	25	8	41

477 * Two dominant BPTs were found in two young females, one young male and one elderly male

478

479 **Table 2.** Distribution of different phylogenetic groups and VAG score amongst dominant (n= 63) and minor (n= 88) *Escherichia coli*
480 types found in 59 healthy adults: young females (n=16), elderly females (n=14), young males (n=13), and elderly males (n=16). Four
481 participants carried two dominant types.

	Number of *E. coli* types (Average VAG score ± standard deviation)				
Phylogenetic groups	Young females (n=33)	Elderly females (n=39)	Young males (n=22)	Elderly males (n=57)	Total (n=151)
Dominant Types					
A	6 (1.8 ± 1.5)	2 (2.5 ± 1.7)	0	0	8 (2 ± 1.3)
B1	1 (1.0)	0	3 (1.3 ± 0.6)	0	4 (1.3 ± 0.5)
B2	5 (3.6 ± 1.8)	5 (4.0 ± 1.0)	6 (3.7 ± 1.2)	6 (4.2 ± 2.1)	22 (3.9 ± 1.5)
D	7 (2.3 ± 1.7)	7 (2.0 ± 1.5)	5 (3.2 ± 2.2)	10 (3.0 ± 1.9)	29 (2.6 ± 1.8)
Minor Types					
A	4 (2.3 ± 2.5)	2 (2.5 ± 0.7)	1 (1.0)	8 (1.9 ± 1.4)	15 (2.0 ± 1.6)
B1	2 (1.0 ± 1.0)	5 (1.4 ± 0.9)	0	5 (1.2 ± 0.4)	12 (1.3 ± 0.6)
B2	3 (3.3 ± 1.5)	5 (3.2 ± 1.1)	2 (4.5 ± 0.7)	11 (2.8 ± 1.9)	21 (3.1 ± 1.6)
D	5 (1.8 ± 1.1)	13 (2.7 ± 1.2)	5 (1.6 ± 1.3)	17 (1.7 ± 1.4)	40 (2.0 ± 1.3)

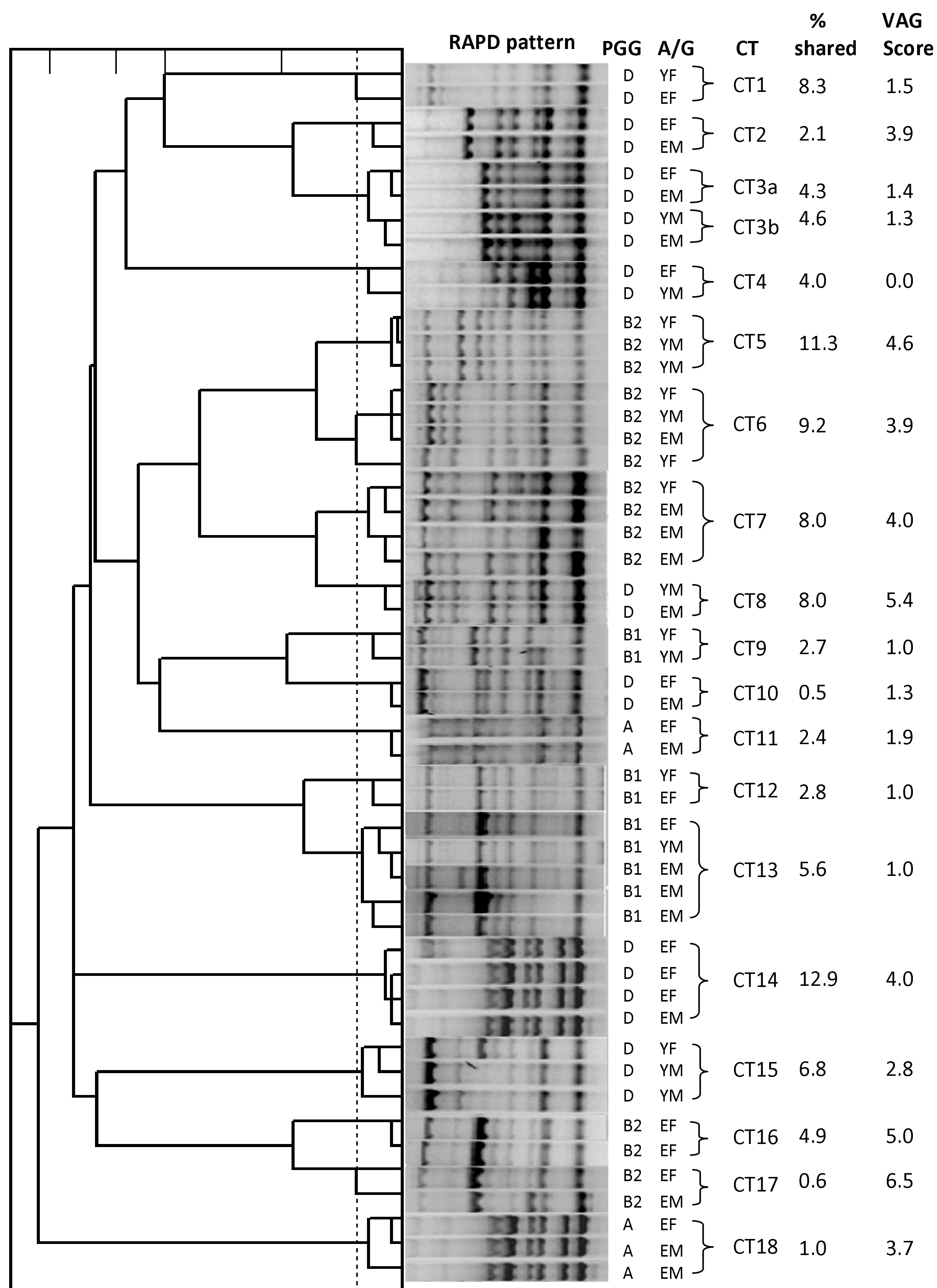
482

483

484 **Table 3.** Prevalence of virulence-associated genes (VAGs) among 151 biochemical phenotypys (BPTs; 63 dominant and 88 minor) found in all 4
485 groups of age and gender studied. These include 33 BPTs found in young females, 39 BPTs found in elderly females, 22 BPTs found in young
486 males and 57 BPTs found in elderly males.

487
488

VAG tested	No. of dominant BPTs (%)	No. of minor BPTs (%)	No. of young female BPTs (%)	No. of elderly female BPTs (%)	No. of young male BPTs (%)	No. of elderly male BPTs (%)
Adhesins						
papAH	14 (22)	7 (8)	6 (18)	4 (10)	5 (23)	6 (11)
papEF	21 (33)	19 (22)	7 (21)	11 (28)	6 (27)	16 (28)
papC	20 (32)	13 (15)	7 (21)	8 (21)	6 (27)	12 (21)
papG allele II	11 (18)	7 (8)	1 (3)	7 (18)	3 (14)	7 (12)
papG allele III	7 (11)	4 (5)	2 (6)	2 (5)	2 (9)	5 (9)
papG allele II and III	16 (25)	10 (11)	5 (15)	6 (15)	4 (18)	11 (19)
fimH	59 (94)	85 (97)	32 (97)	38 (97)	20 (91)	54 (95)
Toxins						
hlyA	9 (14)	5 (6)	3 (9)	3 (8)	1 (5)	7 (12)
cnf1	11 (18)	7 (8)	3 (9)	4 (10)	4 (18)	7 (12)
Siderophores						
iutA	18 (29)	13 (15)	6 (18)	10 (26)	6 (27)	9 (16)
iroN $_{E.coli}$	25 (40)	18 (21)	9 (9)	8 (21)	11 (50)	15 (26)
Protectins						
kpsMT K1	20 (32)	22 (25)	10 (30)	19 (49)	5 (23)	8 (14)
kpsMT II	35 (56)	45 (51)	13 (39)	28 (72)	12 (55)	27 (47)



RAPD pattern
PGG
A/G
CT
% shared
VAG Score
D YF
D EF
CT1 8.3 1.5
D EF
D EM
CT2 2.1 3.9
D EF
D EM
CT3a 4.3 1.4
D YM
D EM
CT3b 4.6 1.3
D EF
D YM
CT4 4.0 0.0
B2 YF
B2 YM
B2 YM
CT5 11.3 4.6
B2 YF
B2 YM
B2 EM
B2 YF
CT6 9.2 3.9
B2 YF
B2 EM
B2 EM
B2 EM
CT7 8.0 4.0
D YM
D EM
CT8 8.0 5.4
B1 YF
B1 YM
CT9 2.7 1.0
D EF
D EM
CT10 0.5 1.3
A EF
A EM
CT11 2.4 1.9
B1 YF
B1 EF
CT12 2.8 1.0
B1 EF
B1 YM
B1 EM
B1 EM
B1 EM
CT13 5.6 1.0
D EF
D EF
D EF
D EM
CT14 12.9 4.0
D YF
D YM
D YM
CT15 6.8 2.8
B2 EF
B2 EF
CT16 4.9 5.0
B2 EF
B2 EM
CT17 0.6 6.5
A EF
A EM
A EM
CT18 1.0 3.7