MIT Libraries | DSpace@MIT

MIT Open Access Articles

Genotoxic Escherichia coli Strains Encoding Colibactin, Cytolethal Distending Toxin, and Cytotoxic Necrotizing Factor in Laboratory Rats

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: Kurnick, Susanna A. et al. "Genotoxic Escherichia coli Strains Encoding Colibactin, Cytolethal Distending Toxin, and Cytotoxic Necrotizing Factor in Laboratory Rats." Comparative Medicine 69, 2 (April 2019): 103-113 © 2019 American Association for Laboratory Animal Science

As Published: http://dx.doi.org/10.30802/aalas-cm-18-000099

Publisher: American Association for Laboratory Animal Science

Persistent URL: https://hdl.handle.net/1721.1/125964

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike

Massachusetts Institute of Technology

1
2
3
4
5
6
7
8
9 **Genotoxic *Escherichia coli* Strains Encoding Colibactin, Cytolethal Distending**
10 **Toxin, and Cytotoxic Necrotizing Factor Colonize Laboratory Rats**
11
12 SA Kurnick[1], A Mannion[1], Y Feng[1], C Madden[1], P Chamberlain[1], JG Fox[1*]
13 [1] Division of Comparative Medicine, Massachusetts Institute of Technology, 77
14 Massachusetts Avenue, 16-825, Cambridge, MA 02139, USA
15
16 Running Title: Genotoxic *E. coli* in rats
17 *Address correspondence to James G. Fox (jgfox@mit.edu)
18
19 Abbreviations: pks, polyketide synthase; cnf, cytotoxic necrotizing factor; cdt,
20 cytolethal distending toxin

21

Abstract

22 (heading line)

23 While many *Escherichia coli* strains are considered commensals in mammals, strains encoding the cyclomodulin genotoxins are associated with clinical and subclinical disease in the urogenital and gastrointestinal tracts, meningitis, and inflammatory disorders. These genotoxins include the polyketide synthase (*pks*) pathogenicity island, cytolethal distending toxin (*cdt*), and hemolysin-associated cytotoxic necrotizing factor (*cnf*). *E. coli* strains are not excluded from rodents housed under specific-pathogen free (SPF) conditions in academic or vendor facilities. This study isolated and characterized genotoxin-encoding *E. coli* from laboratory rats obtained from four different academic institutions and three different vendors. Sixty-nine distinct *E. coli* isolates were cultured from fecal, rectal swab, or extra-intestinal regions of 52 different rats and biochemically characterized. Polymerase chain reaction for cyclomodulin genes and phylogroup was performed on all 69 isolates. Forty five of 69 isolates (65%) were positive for *pks*, 20/69 (29%) were positive for *cdt*, and 4/69 (6%) were positive for *cnf*. Colibactin was the sole genotoxin identifed in 21 of 45 *pks+* isolates (47%), whereas *cdt* or *cnf* was also present in the remaining 24 isolates (53%). *cdt* or *cnf* was never present together or without *pks*. All genotoxin-associated strains were members of pathogen-associated phylogroup B2. Select *E. coli* isolates were characterized by HeLa cell *in vitro* cytotoxicity assays, serotyped, and whole genome sequenced by Illumina MiSeq. All isolates encoding cyclomodulins induced megalocytosis. Serotypes corresponded with vendor origin and cyclomodulin composition, with the *cnf+* serotype representing a known human uropathogen. Whole genome

45 sequencing confirmed the presence of complete *pks*, *cdt*, and hemolysin-*cnf*
46 pathogenicity islands. These findings indicate that genotoxin-encoding *E. coli*
47 colonize laboratory rats from multiple commercial vendors and academic
48 institutions and suggest potential to contribute to clinical disease and introduce
49 confounding variables into experimental rat models.
50

51 **Introduction**

52 *Escherichia coli* is a gram-negative bacillus that colonizes the gastrointestinal
53 tract of humans and animals.[44] While some strains are considered commensals,
54 various intestinal (IPEC) and extra-intestinal pathogenic *E. coli* (ExPEC) pathotypes
55 are associated with a wide range of clinical disease states in the host;[16,39] these
56 strains are responsible for the deaths of more than 2 million humans annually.[65]
57 Specific pathotypes often harbor similar virulence factors and correspond to distinct
58 clinical and histological lesions. Intestinal pathotypes include enteropathogenic *E.*
59 *coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enteroinvasive *E. coli* (EIEC),
60 enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAggEC), diffusely
61 adhering *E. coli* (DAEC), adherent-invasive *E. coli* (AIEC).[65] Extra-intestinal
62 pathotypes include uropathogenic *E. coli* (UPEC) and neonatal-meningitis *E. coli*
63 (NMEC), which have an enhanced ability to translocate through the intestinal
64 epithelium and cause severe clinical disease.

65 Various strains are typically classified into one of the four major phylogenetic
66 groups: A, B1, B2, or D.[10,14,59] Groups B2 and D are often associated with
67 pathogenicity, while fecal strains belonging to groups A and B1 generally lack

68 virulence factors.[20,59] Strains belonging to pathogroup B2 have been isolated from
69 feces of individuals from developed countries with increasing frequency.[50,70]
70 These pathogenic strains encode various combinations of virulence genes
71 and pathogenicity islands which promote invasion and colonization, evasion of host
72 defenses, and damage to host tissues. Associated virulence factors include
73 cytotoxins such as genotoxic cyclomodulins, cytotoxic necrotizing factors (*cnf*),
74 cytolethal distending toxin (*cdt*), and the genotoxin colibactin (*pks*). These virulence
75 factors are known to modulate host cellular differentiation, proliferation, and
76 apoptosis and promote cytopathic effects.[7,19,69]
77 CNF is a 115 kDa cyclomodulin protein that induces cell cycle alterations and
78 cytoskeletal changes by activating rho GTPases, which leads to a variety of aberrant
79 phenotypic effects including micropinocytosis, megalocytosis, and
80 multinucleation.[61] *cnf1* is chromosomally encoded[21] while *cnf2* is plasmid
81 encoded[64]. *cnf*-producing *E. coli* are considered necrotoxigenic (NTEC) and are
82 associated with intestinal, urinary,[21] and meningeal infection of humans.[39] *cnf+ E.*
83 *coli* have previously been isolated from clinically normal and clinically ill ferrets,[47]
84 cats,[23] dogs,[35,67] pigs,[73] birds,[40] and macaques.[22]
85 Cytolethal distending toxins (CDTs) are encoded by three adjacent genes:
86 *cdtA, cdtB,* and *cdtC* that can be either chromosomal or plasmid encoded. [72]
87 All three genes are required for the production of this heat stable exotoxin, which
88 bears considerable homology to DNAse I and causes DNA breaks.[13] CDTs have been
89 classified into subgroups I-V[19] based on variations in amino acid sequences and
90 genomic locations.[32] Various EPEC serotypes carry this cyclomodulin, which, like

91 colibactin, also induces irreversible megalocytosis and G1 or G2 cell cycle arrest.
92 *cdt+ E. coli* have been isolated from healthy and diseased humans as well as cattle,
93 swine, and birds.[32,56]

94 A 54kb polyketide synthase (*pks*) pathogenicity island encodes multiple *clb*
95 genes (nonribosome peptide synthases) that are collectively responsible for
96 colibactin synthesis. The *pks* island was first identified in 2006 in a case of ExPEC-
97 induced neonatal meningitis[55] and is associated with a variety of extra-intestinal
98 infections in humans including bacterial meningitis, septicemia, and infections of the
99 genitourinary tract.[26,51] Additionally, it is associated with increased persistence in
100 the gastrointestinal tract. Colibactin induces double stranded DNA breaks, which
101 leads to chromosomal instability and subsequent promotion of carcinogenesis.[51] In
102 human studies, colibactin-producing *E. coli* are isolated from human colorectal
103 tumors with significantly increased frequency.[7] Furthermore, Cougnoux's group
104 found that *pks+ E. coli* promoted tumor survival by inducing cellular senescence via
105 growth factor secretion.[15] This association is recapitulated in laboratory animal
106 models. Monoassociation of *pks+ E. coli* strain NC101 caused typhlitis[42] and
107 promoted invasive carcinoma in azoxymethane (AOM)-treated interleukin 10
108 knockout (C57BLIL10-/-) mice;[1] these effects were dependent on the presence of
109 the *pks* island. *In vitro* studies have confirmed these findings by demonstrating that
110 colibactin-encoding *E. coli* strains induce significant megalocytosis, double-stranded
111 DNA breaks, phosphorylated γ-H2AX foci,[1] and G2 cell-cycle arrest in eukaryotic
112 cells.[69]

113 The presence of these genotoxins in human *E. coli* isolates is variable;
114 prevalence is dependent on geographic location. In Puerto Rico, Gomez-Moreno *et al*
115 found that of 41 stool samples tested, 8 isolates (20%) tested positive for *pks*.
116 However, only 1 isolate was found to encode *cnf* whereas no isolates encoded *cdt*.
117 The *cnf*-encoding isolate was also *pks+*[28]. Similarly, a group in France found that
118 26% of their 81 patients harbored *pks+ E. coli* strains, 18% were *cnf+*, and 11%
119 were *cdt+*. *cnf* and *cdt* were often associated with *pks*, with a minority of genotoxin-
120 positive strains encoding *cnf* or *cdt* alone.[60] Only 2 isolates (originating from colon
121 cancer patients) were positive for all 3 genotoxins. All *cnf+* strains demonstrated a
122 hemolytic phenotype.[60] In Mexico, a single uropathogenic strain (1/108) was found
123 to encode both *cnf* and *cdt*; *pks* was not evaluated.[45]
124 In research animals, our laboratory demonstrated that 88% of isolates from
125 laboratory mice were colonized with *pks*+ *E. coli* and belonged to pathogen-
126 associated phylogroup B2.[26] Genotoxic *E. coli* have been identified in several other
127 species of laboratory animals: *pks+ E. coli* has also been identified in laboratory
128 macaques and *cnf+ E. coli* in ferrets and nonhuman primates.[47,22]
129 Rats constitute valuable models of both neonatal meningitis (NMEC) and
130 uropathogenic (UPEC) *E. coli* infection. Young rats are commonly used to study
131 systemic dissemination of NMEC K1 infection via the gastrointestinal tract[17,79] and
132 methods of prevention,[82] intestinal barrier permeability,[31] sequelae of bacterial
133 neonatal meningitis.[27] Neonatal rats have recently been used to model maternal to
134 neonatal transmission of *pks+ E. coli*, which resulted in increased rates of intestinal
135 epithelial cell proliferation, apoptosis, and permeability that was transmissible

136 through generations.[58] In addition, numerous studies have utilized rats
137 experimentally infected with *cnf+* uropathogenic *E. coli* in order to study the
138 dissemination and pathogenesis of *E. coli* associated recurrent urinary tract
139 infections, pyelonephritis, and acute kidney injury (AKI).[77,68,62] Potential novel
140 treatments for these conditions such as photodynamic therapy[33] and novel drug
141 delivery methods are also investigated in these experimentally infected rat models.
142 Unfortunately, the prevalence of *pks* and other cyclomodulin positive *E.coli*
143 strains in SPF laboratory rats is currently unknown; vendors typically do not
144 include *E. coli* on their health surveillance reports. Thus, this study focused on
145 determining the comparative prevalence of *pks+, cdt+, and cnf +* isolates from the
146 gastrointestinal tract and several other sites from rats obtained from multiple
147 institutions and vendors. Given previous work regarding prevalence in laboratory
148 mice and its association with urosepsis and meningitis in immunocompromised
149 mice,[26],we hypothesized that the majority of isolates from rats encoded the *pks*
150 genomic island regardless of institution or vendor and asked whether these isolates
151 also encoded *cdt* or *cnf*.
152

153 **<u>Methods</u>**

154 **Animals**

155 A total of 52 different rats from 3 distinct vendors originating from multiple barriers
156 within each vendor facility and ultimately residing at 4 different academic
157 institutions were included in this study. Vendor A rats were housed in institutions
158 W, X, and Z, whereas vendor B supplied rats for institutions Y and Z, and vendor C

159 supplied rats only to institution Z. The most commonly represented strain was
160 Sprague Dawley; 4 animals were Long Evans and 3 were c-fos-lacZ transgenic rats.
161 There was an even distribution of male and female rats. Based on health
162 surveillance reports, all animals were considered specific-pathogen free (SPF). *E.*
163 *coli* was absent from vendor surveillance reports. Samples were collected from
164 2015-2017 and animals ranged in age from 8 weeks to 2 years. Animals were group
165 housed at both the vendors and academic institutions; three out of four academic
166 institutions maintained AAALAC accredited facilities. Rodent chow and water were
167 provided *ad libitum* and housed in polycarbonate cages. All animals were on IACUC
168 approved studies.
169

170 **Culture & Isolation**

171 *E. coli* was isolated from fecal contents, vagina, or nares of clinically normal rats
172 immediately upon delivery to the academic institutions or after being housed in
173 academic facilities. Sixty nine *E.coli* isolates were cultured from fecal/rectal swab
174 (49), vaginal swabs (1), or nares (3). Fecal/rectal samples were collected directly
175 from the rectum of the animals in shipping crates prior to their entrance into the
176 institutional facilities. Fecal pellets or rectal swabs were placed into tubes
177 containing sterile Gram Negative broth (BD) and incubated at 37C overnight. A
178 broth swab was plated onto MacConkey lactose agar plates (Remel) and lactose-
179 positive colonies then plated onto sheep blood agar plates (Remel) based on distinct
180 colony morphologies. The presence or absence of β-hemolysis was noted and

181 recorded; suspect *E. coli* isolates were biochemically characterized using API® 20 E

182 (Biomérieux).

183

184 **DNA Extraction & PCR Amplification**

185 A loop of each of the 69 *E. coli* isolates grown overnight on sheep's blood agar plates

186 was placed in 500 ul of sterile phosphate-buffered saline (PBS) in a microfuge tube

187 and swirled until thoroughly dissolved. Samples were boiled for 10 minutes

188 followed by 10 minutes of centrifugation at 12,000g. The supernatants were used in

189 the PCR reactions. Two sets of primers (*clbA, clbQ*) were used to identify *pks* genes[22].

190 Multiplex PCR was used to amplify *cnf* and *cdt* genes. Five sets of primers for *viaA,*

191 *TSPE4.C2, chuA, svg and uidA* genes were used in multiplex PCR to determine the

192 phylogroup of each isolate.[5,14] The phylogenetic groups were determined based on

193 the PCR gel pattern.

194

195 **Serotyping**

196 Nine *E. coli* isolates chosen from different vendors, barriers, and institutions and

197 representing *pks-/cdt-/cnf–, pks+/cdt-/cnf–, pks+/cdt-/cnf+, pks+/cdt+/cnf-*, and

198 genotypes were submitted to the *E. coli* Reference Center at Penn State University

199 for serotype testing, which included: O and H typing and PCR analyses for heat-

200 labile enterotoxin (*elt*), heat-stabile enterotoxin (*estA* and *estB*), Shiga-type toxin 1

201 and 2 (*stx1* and *stx2*), intimin gamma (*eae*), *cnf1*, and *cnf2*.

202

203 **Cytotoxicity Assay**

204 Control strains included NC101 (*pks+/cdt-/cnf–*) and NC101Δpks (*pks- mutant*),
205 which were gifts from Dr. Chris-tian Jobin. Other control strains included V27 (*+,*
206 *pks+/cdt+/cnf-*, acquired from the E. coli Reference Center), and K12 (triple negative
207 control). Eleven isolates representing all possible combinations of genotype,
208 vendors, and institution were evaluated; these isolates included *pks+, cdt+,* and *cnf+*
209 isolates, triple negative isolates, and isolates from all anatomical locations sampled.
210

211 **Cell culture assay for colibactin cytotoxicity**

212 The cytotoxicity assay was performed as described previously with
213 modifications[26,55]. HeLa S3 cells (ATCC CCL2.2) were grown and maintained in
214 Eagle's Minimum Essential Medium (EMEM, ATCC) containing 10% Fetal Calf Serum
215 (FCS, Sigma) and 1% Antibiotic-Antimycotic (Gibco) at 37 °C with 5% CO_2. 5 × 10^3
216 cells were seeded onto 96-well cell culture plates and incubated at 37 °C with 5%
217 CO_2 for 24 h. Overnight liquid cultures of *E. coli* strains were grown for 2 h at 37 °C
218 and then adjusted to O.D.600 nm in 1% FCS EMEM media to concentrations
219 corresponding to a multiplicity of infection (MOI; the number of bacteria per cell at
220 the onset of infection) of 100. Following inoculation, plates were centrifuged at
221 200 g for 10 min to facilitate bacterial interaction and then incubated at 37 °C with
222 5% CO_2 for 4 h. Cells were then washed with EMEM and replaced with EMEM
223 containing 10% FCS and 200 μg/mL gentamicin (Gibco). Following 72 h incubation,
224 plates were stained with Diff-quick stain (Thermo Scientific). Cells were then
225 inspected under a microscope for confluence and morphological changes. Images

226 were captured with a Zeiss Axiovert-10 microscope using Image Pro-Plus software
227 version 7.0 at 20× magnification.
228

229 **Cell culture assay for sonicate cytotoxicity**

230 Overnight cultures of *E. coli* strains were pelleted by centrifugation at
231 12,000 rpm for 5 min. The pellets were washed in 1 ml of PBS and pelleted again by
232 centrifugation at 12,000 rpm for 5 min. Pellets were re-suspended in 2 ml of PBS
233 and then sonicated on ice using the following program: amplitude: 35; power: 7 W;
234 30 s intervals for a total of 5 min with 1 min breaks between intervals. Sonicate
235 samples were centrifuged at 12000 rpm for 10 min at 4 °C to remove large debris.
236 Supernatant was collected and then filter-sterilized through 0.2 µm filters. Total
237 protein was quantified using the BCA assay (Thermo Fisher Scientific). HeLa cells,
238 5×10^3 were seeded onto 96-well cell culture plates and incubated at 37 °C with 5%
239 CO_2 for 24 h. Cells were treated with 1 or 40 µg/mL total protein of crude bacterial
240 sonicate for 72 h. Cells were stained and microscopically analyzed for confluence
241 and morphological changes as described above.
242

243 **Draft Genome Sequencing and Comparative Analysis**

244 Genomic DNA was isolated from 7 representative isolates using the
245 MasterPure Complete DNA and RNA Purification Kit (Epicentre) following the
246 manufacturer's protocol for bacterial cell samples. DNA libraries were prepared by
247 the Sequencing Core at the Forsyth Institute (Cambridge, MA) using NextraXT for
248 sequencing of 2x150 paired-end reads by Illumina MiSeq. Raw sequencing reads

249 were decontaminated of adapter sequences and quality trimmed to a Phred quality
250 score (Q) ≥ 10 using BBDuk from the BBMap package version 37.17
251 (http://sourceforge.net/projects/bbmap/). Decontaminated reads were then
252 assembled into contigs with SPAdes[3] and scaffolds with Ragout[41] followed by
253 genome annotation with RAST hosted by PATRIC[2,6,78]. Sequences encoding putative
254 virulence factor and antibiotic resistance genes were identified using
255 VirulenceFinder 1.5[34] and ResFinder 2.1[81] hosted by Center for Genomic
256 Epidemiology. Syntenic relationships of *pks, cdt,* and *hemolysin-cnf* operon genes
257 between genomes were determined with SimpleSynteny.[76]
258
259 **Accession Numbers:** GenBank accession numbers are available in Table 3.
260
261 **<u>Results</u>**
262 **Microbiological Characterization**
263 *E. coli* was isolated from all 52 rats sampled, with all biological sampling locations
264 (rectum, nares, vagina) yielding *E. coli* isolates. Sixty nine *E. coli* isolates were
265 cultured; some animals harbored multiple *E. coli* isolates as determined by distinct
266 API codes and colony morphology. None of the isolates demonstrated hemolysis.
267 Across all vendors and institutions, there was no correlation between API code and
268 genotoxin genotypes. Differences in API codes indicated the ability of the isolates to
269 ferment certain sugars and metabolize specific amino acids. The most common API
270 code, 5144572, was observed in 41/69 isolates, while the second most common was
271 5144552, observed in 23/69 isolates. The major metabolic difference between

272 these codes is that the most common code (5144572) has the ability to ferment
273 sucrose whereas the latter does not. The API code 1144552 appeared in 4 isolates
274 and the API codes 7144472 and 7144572 were observed in single isolates from
275 vendor A. Codes beginning with "1" are lacking lysine and arginine decarboxylase
276 activity. Some correlation was evident between API code and genotype at the
277 individual vendor level, which suggests a certain degree of clonality among isolates
278 from each origin. However, these patterns did not hold when different vendors or
279 institutions were compared. For example, all 9 isolates from vendor B with the API
280 code 5144552 harbored both *pks* and *cdt*, whereas all 11 isolates with the same API
281 code from vendor A were negative for all genotoxins. The 4 isolates with API code
282 1144552 originated from Vendor B rats cultured directly from the shipping crate
283 after arriving at the institution. This API code only occurred in isolates that were
284 *cnf+*.
285
286 **Identification of *pks, cdt, cnf* genes**
287 Conventional PCR for *pks* genes *clbA* and *clbQ* (figure 1A) and multiplex PCR for *cdt*
288 (figure 1B) and *cnf* genes (figure 1C) was performed on all isolates to identify the
289 presence of genotoxin genetic elements. Overall, 45 of 69 (65%) of the total isolates
290 were positive for both *pks* genes; there were no isolates that tested positive for one
291 gene without the other; 20 of 69 (29%) isolates were positive for *cdt* and 4 of 69
292 (6%) isolates were positive for *cnf*. *pks* was the sole genotoxin identifed in 21 of 45
293 *pks+* isolates (47%), whereas *cdt* or *cnf* was also present in the remaining 24
294 isolates (53%). *Cytolethal distending toxin* or *cnf* was never present without *pks* and

295 *cnf* and *cdt* were never present together (Table 1). Roughly half (55%) of the
296 isolates from vendor A animals were positive for *pks* with a 15% minority encoding
297 *cdt* in addition. No vendor A animals tested positive for *cnf*. Conversely, all isolates
298 from vendor B animals were *pks+* and 69% of them encoded *cdt*. Isolates that did
299 not encode *cdt* had *cnf* instead. Thus, all isolates from vendor B were positive for
300 multiple genotoxins (Table 1). Fifty one to eighty percent of isolates from animals
301 arriving at institutions W and X were *pks+,* with a minority of isolates (15-17%)
302 carrying *cdt* in addition. All isolates from institutions Y and Z were *pks+*, with the
303 majority of isolates also harboring *cdt* (64-100%). All rats from vendor C encoded
304 *pks* and *cdt*. Those *E. coli* isolates from institution Y that did not have *cdt* encoded
305 for *cnf* instead (36%) (Table 2).
306
307 **Phylogenetic Analysis**
308 Phylogroup was determined based on the amplification pattern of multiplex PCR for
309 *viaA, TSPE4.C2, chuA, svg and uidA* genes (Figure 2a). The presence of 3 or more
310 bands identifies the isolate as a member of phylogroup B2. All isolates were
311 members of phylogroup B, with 24/69 (35%) of isolates belonging to group B1 and
312 45/69 (65%) of isolates belonging to pathogen-associated phylogroup B2. Only 2
313 isolates that were members of phylogroup B2 did not test positive for any of the
314 cyclomodulins under evaluation. All genotoxin-positive isolates belonged to group
315 B2 (Figure 2b).
316
317 **Serotyping**

318 The most common serotype among isolates was O7:H7; all originated at vendor A,
319 but each isolate originated from rats housed at a different institution (Table 2). Two
320 of these isolates were *pks+* only and the third encoded both *pks* and *cdt*. The next 2
321 most common serotypes were found in duplicate. The two *pks+, cdt+* isolates from
322 vendor B (rats housed at different institutions) were serotype O166:H6. Two triple-
323 negative *E. coli* isolates from vendor A were O179:H8. The *pks+, cdt+ E. coli* isolate
324 from vendor C was OM:H6 and the *pks+, cnf+* isolate from vendor B was O4:H5, a
325 known uropathogen in humans.[57,36] None of the *E. coli* isolates serotyped were
326 positive for *elt, estA, estB, stx1, stx2, eae*, and *cnf2* genes.
327
328 ***In vitro* cytotoxicity of *E. coli* isolates**
329 Cell culture assays were performed to determine if *in vitro* infection or
330 sonicates of representative rat *E. coli* isolates caused cytotoxicity to HeLa cells. A
331 total of 17 isolates encompassing representatives from all institutions, vendors and
332 barriers, anatomical areas of isolation, genotoxin status, and phylogroup were
333 evaluated. Live bacteria were used rather than sonicate as whole cells are required
334 for the complete expression of colibactin.[7] Conversely, CDT and CNF cytotoxicity are
335 only detectable using sonicate preparations. Viable *pks+ E. coli* isolates induced
336 megalocytic cytotoxicity to HeLa cells, indicating contact-dependent colibactin
337 expression (figure 3a). HeLa cells treated with sonicate from *cdt+* or *cnf+ E. coli*
338 isolates also displayed cell body and nuclei enlargement, which are characteristic of
339 these cytotoxin (figure 3b). *E. coli* isolates PCR-negative for *pks* or *cdt* lacked
340 cytotoxicity in their respective sonicate based-cell culture assays. These results

341 indicate rat *E. coli* isolates exhibit cytotoxic *pks*, *cdt* , and *cnf* activity *in vitro*, as their

342 genotypes suggest.

343

344 **Draft Genome Sequencing and Comparative Analysis**

345 The draft genome sequences of seven representative rat *E. coli* isolates were

346 obtained for comparative analysis of the *pks*, *cdt*, and *cnf* genes as well as for

347 identification of other virulence factor and antibiotic resistance genes. The rat *E.*

348 *coli* isolate genomes have similar genome sizes, G+C% contents, and protein and

349 RNA genes as the representative *pks+ E. coli* strains IHE3034 and NC101, as

350 summarized in Table 3. Homologous genes for all *pks* genes were identified in the

351 rat *E. coli* isolates and showed identical synteny to IHE3034 and NC101. Compared

352 to IHE3034, all PKS genes from the rat *E. coli* isolates had ≥98% sequence coverage

353 and identity, except the *clbJ* and *clbK* genes from isolate S15 had ~90% and ~45%

354 sequence coverage, respectively. Further analysis of the *clbJ* and *clbK* genes from

355 isolate S15 suggests they could be expressed as a hybridized gene (see

356 supplementary results/discussion). Cytolethal distending toxin genes were detected

357 in 3 out of 7 genomes. All 3 genomes had complete tripartite *cdt* holotoxin island

358 including *cdtA, cdtB*, and *cdtC*. The *cnf* gene was intact, but the adjacent hemolysin

359 operon demonstrated an insertional event that interrupted the *hlyA* gene (see

360 supplementary results/discussion). None of these *cnf+* isolates were hemolytic. All

361 rat isolates encoding cyclomodulin genes induced megalocytosis in HeLa cells

362 (figure 3), indicating cyclomodulin expression. Other virulence factor genes were

363 also identified in the rat *E. coli* isolate genomes and included toxins (*astA*, *cdtABC*,

364 *pic*, *vat*), bacteriocin synthesis genes (*cba*, *celb*, *cma*, *mchB*, *mchC*, *mchF*, *mcmA*),
365 nutrient/survival factors (*gad*, *iroN*, *iss*), and adherence (*lpfA, sfaS*) (table 3). Gene
366 sequences for cell cycle inhibiting factor (*cif*) were not detected in any of the 7
367 genomes. Antibiotic resistance genes were also not detected in any of the rat *E. coli*
368 isolates. The genomic results suggest the rat *E. coli* isolates encode *pks* gene islands,
369 *cdt, cnf* and other virulence genes that endow them with pathogenic potential.
370

371 **Discussion**

372 As a major commensal organism of the human and animal intestinal tract, a
373 thorough understanding of *E. coli* is warranted in both humans and animals. A shift
374 in genetic makeup of these *E. coli* colonizing the gut from phylogroups A and B1 to
375 pathogen-associated phylogroups B2 and D has occurred in recent years in
376 industrialized countries; this shift affects both humans and animals.[50,70] Colibactin
377 production induces double stranded DNA breaks, activation of the DNA damage
378 response, and subsequent genomic instability in the mammalian host. Senescence,
379 cell death, and carcinogenesis are associated with colonization of *pks+ E. coli* strains.
380 Similarly, *cdt* encodes a DNAse genotoxin that causes single and double-stranded
381 DNA breaks which results in increased mutagenesis; this cyclomodulin has been
382 detected in *E. coli* isolated from proximal and distal colon cancer tissues from
383 human patients.[7,60] Cytotoxic necrotizing factor is a third cyclomodulin that is
384 known to induce cell cycle disturbances and abnormal cytoskeletal effects.
385 There is a paucity of information regarding the *E. coli* status of laboratory
386 rats and the variability of genotoxin expressing *E. coli* in animals from different

387 vendors and institutions. This is the first report to our knowledge to characterize
388 colibactin, *cdt*, and *cnf* presence in unmanipulated laboratory rats. In this study, we
389 demonstrated significant variability in the prevalence of *pks+, cdt+,* and *cnf+ E. coli*
390 across multiple, commonly used vendors and 4 separate academic institutions.
391 Overall, the majority of isolates (65%) were *pks+* and members of phylogroup B2
392 (65%). There was a strong association of genotoxin-positive strains with
393 phylogroup B2, as is the case in human isolates.[19] Surprisingly, as *cdt* and *cnf* were
394 not identified in *E. coli* colonizing mice,[26] cyclomodulins *cdt* and *cnf* were present in
395 laboratory rats; 29% of rat *E. coli* isolates carried *cdt*, whereas only 6% carried *cnf*.
396 This is in contrast to our hypothesis and available human surveys, where *cnf*-
397 positive *E. coli* (39.5%) is isolated much more commonly than *cdt*-positive *E. coli* (1-
398 6%).[7] Cytolethal distending toxin and *cnf* were never present in the *E. coli* strains
399 without colibactin or in strains with each other. The co-association of *pks* and *cdt* in
400 some *E. coli* strains suggests mechanisms that potentiate genotoxicity, although *pks*
401 and *cdt* are not commonly identified within the same human *E. coli* isolate.[25,28]
402 Double-positive isolates (*pks+/cnf+)* have been characterized from both healthy
403 humans and urosepsis patients.[19] This is in contrast to surveys in humans and other
404 laboratory animals where *cnf* is occasionally present in colibactin-negative
405 isolates.[47,22,60] Many previous studies have shown a correlation between *cnf* and
406 hemolysis,[43,47,49] which is consistent with the proximity of the hemolysin to the *cnf*
407 gene. Interestingly, none of the *cnf+* isolates from laboratory rats demonstrated
408 hemolysis due to an insertion event in the *hylA* gene. All *cnf+ E. coli* strains isolates
409 were isolated from vendor B rats.

410 The results of both the whole cell and sonicate cytotoxicity assays correlated
411 with the presence or absence of *pks*, *cdt,* and *cnf*. As cell contact is required for
412 colibactin cytotoxicity,[69] HeLa cell death and megalocytosis was due to *cdt* or *cnf* in
413 the sonicate assay. Genotoxin-negative *E. coli* isolates produced results that were
414 indistinguishable from those of the non-pathogenic strain K12, which suggests
415 attenuated pathogenicity due to lack of genotoxins. While only 55% of *E. coli* isolates
416 from vendor A encoded the *pks* island, 100% of isolates from vendor B were *pks+.*
417 Institutional *pks+ E. coli* prevalence in rats was consistent with reported rat vendor
418 usage and origin, with vendor A institutions having lower *E. coli* genotoxin
419 prevalence in rats versus rats housed in vendor B institutions. In addition, serotype
420 patterns tended to correlate with vendor origin rather than institution (Table 1).
421 This pattern underscores that genotoxin-positive *E. coli* efficiently colonize and
422 likely persist in the bowel throughout life;[65] these strains likely colonize rats at the
423 vendors and inhabit the alimentary tract of the rats for the duration of their studies
424 performed at destination institutions.

425 Comparative genomic analysis of the *pks* islands revealed that *clbJ* and *clbK*
426 genes from isolate S15 have ~90% and ~45% sequence coverage, respectively,
427 compared to IHE3034. The *clbJ* gene appears to be missing 624 bp at the 3' end
428 including the stop codon, but retains two nonribosomal peptide synthetase (NRPS)
429 modules. *clbK* appears to lack 3,540 bp at the 5' end including a start codon and the
430 PKS module, but retains the NRPS module and the oxidase domain. Further analysis
431 of the putative *clbJ* and *clbK* genes shows their sequences overlap by 1,480 bp in the
432 genome, suggesting they are not expressed as separate genes and instead form a

433 single, continuous gene sequence. When the *clbJ* start codon is used as the position
434 of the open reading frame, the predicted sequence is translated into a 2,240 amino
435 acid product (7,323 bp) that includes the *clbJ* and *clbK* sequences and terminates at
436 the *clbK* stop codon. This suggests the putative *clbJ* and *clbK* sequences may be
437 transcribed and translated into a hybridized protein (designated *clbJK*-hybrid). The
438 predicted *clbJK*-hybrid protein would contain two NPRS modules as well as an
439 oxidase domain (Figure 4a). A BLAST search found identical *clbJK*-hybrid gene
440 sequences in the three other genomes: neonatal meningitis-causing *E. coli* Strain
441 NMEC O18 (GenBank: CP007275), *Klebsiella pneumoniae* str. Kp52.145 (GenBank:
442 FO834906), and *K. pneumoniae* subsp. *pneumoniae* strain KPNIH32 (GenBank:
443 CP009775). This indicates other *E. coli* and *K. pneumoniae* strains have a putative
444 *clbJK*-hybrid sequence instead of separate *clbJ* and *clbK* genes in their PKS islands.
445 Isolate S15 still induced megalocytosis to HeLa cells, indicating cytotoxic colibactin
446 expression despite having a putative *clbJK*-hybrid gene. Whether *E. coli* Strain
447 NMEC O18 and the two other *K. pneumoniae* strains also exhibit colibactin
448 cytotoxicity has not been reported.

449 The *clb* genes encoded on the PKS island constitute an “assembly line” of
450 enzymes that produce pre-colibactin and colibactin metabolites by complex and
451 incompletely defined biosynthetic pathways. Furthermore, these metabolites can be
452 formed or modified by the *clb* enzymes via alterative pathways, leading to a large
453 structural diversity of molecules that has not been entirely catalogued. In particular,
454 recent reports have indicated the PKS module in *clbK* can be biochemically bypassed
455 to yield an alternative pre-colibactin metabolite with unknown cytopathogenic

456 properties.[74,83] This alternative pathway still requires the NRPS modules and
457 oxidase activity from *clbJ* and *clbK*. The putative *clbJK*-hybrid gene detected in
458 isolate S15 is predicted to contain two NRPS modules and an oxidase domain, but
459 lacks the PKS module from *clbK*. As a result, it may be possible for the putative
460 *clbJK*-hybrid gene to synthesize pre-colibactin metabolites in analogous fashion to
461 the alternate scheme mentioned above.

462 All 3 *cdt* genes in the *cdt* island are intact and conserved among isolates
463 (Figure 4b). While the *cnf* island itself was intact, the hemolysin *hlyA* gene was
464 disrupted by an approximately 500 bp insertion consisting of insertion element IS1
465 protein InsB, which is the most common transposase in the *E. coli* genome (Figure
466 4c). Transposable IS1 elements have been reported to disrupt other portions of the
467 hemolysin operon.[9]

468 Aside from *pks*, *cdt*, and *cnf*, other virulence factor genes were identified in
469 the rat *E. coli* isolates that are known to enhance colonization/survival and promote
470 disease in the host. Glutamate decarboxylase (*gad*) and increased serum
471 survival/bor protein precursor (*iss*) promote survival in the host by neutralizing
472 stomach acid during oral transmission[4,29,66] and by promoting resistance against
473 host complement protein,[37,46,52] respectively, while enterobactin siderophore
474 receptor protein (*iroN*) allows uptake of the essential nutrient iron into the
475 pathogen.[12,24,39] Long polar fimbriae (*lpfA*) and s-fimbriae minor subunit (*sfaS*) are
476 both adhesion factors for colonization of host epithelial cells.[38,48,63,71] Colicin B (*cba*),
477 colicin E (*celb*), colicin M (*cma*), and microcin H47 (*mchB*, *mchC*, *mchF*, *mcmA*) are
478 bacteriocins produced by pathogenic *E. coli* strains that target and kill susceptible

479 bacteria.[8,11,53,54] As a result, bacteriocin-producers may have competitive
480 advantages in niches with scare essential nutrients like iron. Enteroaggregative *E.*
481 *coli* heat-stable enterotoxin 1 (*astA*) is a cytotoxin that actives guanylyl cyclase in
482 the gastrointestinal epithelium resulting in ion secretion that contributes to watery
483 diarrheal disease.[16,39,75] Protease involved in intestinal colonization (*pic*) and
484 vacuolating autotransporter toxin (*vat*) are both serine protease autotransporters of
485 *Enterobacteriaceae* (SPATE) that degrade the mucous barrier to facilitate invasion[18]
486 and cause intracellular vacuolation,[18,30] respectively. Of particular interest,
487 cytolethal distending toxin (CDT) genes were identified in genomic sequence of
488 three isolates (S11, 14, S15).

489 The presence or absence of cyclomodulin genotoxins in laboratory rats may
490 have unintended impacts on experimental results and repeatability across
491 institutions. As *E. coli* is not included on vendor surveillance reports, rats from
492 various institutions may have vastly different gastrointestinal microbiota, producing
493 inherent variability in results and conclusions. Genotoxic *E. coli* colonizing rats
494 arriving from vendors may interfere with studies of experimental *E. coli* infection.
495 This is especially relevant as neonatal rats are an extremely popular model of *E. coli*
496 K1 infection and sequelae,[79] in which the K1 capsule protects the bacteria from the
497 host's immune response. This strain is another early colonizer of the neonatal GI
498 tract that can translocate from lumen to blood. These rats are used to characterize
499 changes in oxidative responses following *E. coli* inoculation throughout life,[27] track
500 vertical transmission of *pks+ E. coli* from mothers to offspring,[58] and to evaluate the
501 efficacy of a variety of antimicrobial agents against genotoxic *E. coli* infection.[31,80,82]

502 If genotoxic *E. coli* species are present at the initiation of these and other
503 relevant studies, comparisons between sham and experimental groups may be
504 erroneous. Additionally, the possibility of zoonotic transfer from rats to humans
505 should not be overlooked, especially given that O4:H5 *E. coli* isolated from rats in
506 this study are associated with urosepsis in humans.[35,57] This possibility emphasizes
507 the importance of proper hygiene and personal protective equipment, even in
508 seemingly low risk areas. Together, the identification of virulence factor genes from
509 genotoxin-encoding rat *E. coli* isolates suggests these pathobionts have the potential
510 to cause clinical or subclinical disease in rats and significantly confound rat research
511 models.
512
513 Acknowledgements: We would like to thank Alyssa Pappa for her excellent
514 assistance preparing the manuscript.
515
516 Competing interests: The authors declare that they have no competing interests.
517
518 Ethics approval and consent to participate in work involving animals was approved
519 by the Massachusetts Institute of Technology's Committee on Animal Care and Use
520 Office.
521
522 Funding Research reported in this publication was supported by The National
523 Institutes of Health under Award Numbers T32OD010978, R01OD01141 and
524 National Institute of Environmental Health Sciences under Award P30ES0022109.

525 The content is solely the responsibility of the authors and does not necessarily

526 represent the official views of the National Institutes of Health.

527

528

529

530

531

532 References

533

534 1. **Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan
535 TJ, Campbell BJ, Abujamel T, Dogan B, Rogers AB, et al.** Intestinal
536 inflammation targets cancer-inducing activity of the microbiota. Science.
537 2012;**338**:120–123.
538 2. **Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, Formsma K,
539 Gerdes S, Glass EM, Kubal M, et al.** The RAST Server: Rapid annotations
540 using subsystems technology. BMC Genomics. 2008;**9**.
541 3. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS,
542 Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al.** SPAdes: A New
543 Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.
544 Journal of Computational Biology [Internet]. 2012;**19**:455–477. Available
545 from: http://online.liebertpub.com/doi/abs/10.1089/cmb.2012.0021
546 4. **Bergholz TM, Tarr CL, Christensen LM, Betting DJ, Whittam TS**. Recent
547 gene conversions between duplicated glutamate decarboxylase genes (*gadA*
548 and *gadB*) in pathogenic *Escherichia coli*. Molecular Biology and Evolution.
549 2007;**24**:2323–2333.
550 5. **Bidet P, Metais A, Mahjoub-Messai F, Durand L, Dehem M, Aujard Y,
551 Bingen E, Nassif X, Bonacorsi S**. Detection and identification by PCR of a
552 highly virulent phylogenetic subgroup among extraintestinal pathogenic
553 *Escherichia coli* B2 strains. Applied and Environmental Microbiology.
554 2007;**73**:2373–2377.
555 6. **Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R,
556 Overbeek R, Parrello B, Pusch GD, et al.** RASTtk: A modular and extensible
557 implementation of the RAST algorithm for building custom annotation
558 pipelines and annotating batches of genomes. Scientific Reports. 2015;**5**:8365.
559 7. **Buc E, Dubois D, Sauvanet P, Raisch J, Delmas J, Darfeuille-Michaud A,
560 Pezet D, Bonnet R**. High Prevalence of Mucosa-Associated *E. coli* Producing
561 Cyclomodulin and Genotoxin in Colon Cancer. PLoS ONE. 2013;**8**:e56964.
562 8. **Budič M, Rijavec M, Petkovšek Ž, Žgur-Bertok D**. *Escherichia coli*
563 bacteriocins: Antimicrobial efficacy and prevalence among isolates from
564 patients with bacteraemia. PLoS ONE. 2011;**6**.
565 9. **Burgos Y**. Common origin of plasmid encoded alpha-hemolysin genes in

566 *Escherichia coli*. BMC Microbiology. 2010;**10**.
567 10. **Carlos C, Pires MM, Stoppe NC, Hachich EM, Sato MIZ, Gomes TAT, Amaral**
568 **LA, Ottoboni LMM**. *Escherichia coli* phylogenetic group determination and its
569 application in the identification of the major animal source of fecal
570 contamination. BMC Microbiology. 2010;**10**:161.
571 11. **Cascales E, Buchanan SK, Duche D, Kleanthous C, Lloubes R, Postle K,**
572 **Riley M, Slatin S, Cavard D**. Colicin Biology. Microbiology and Molecular
573 Biology Reviews [Internet]. 2007;**71**:158–229. Available from:
574 http://mmbr.asm.org/cgi/doi/10.1128/MMBR.00036-06
575 12. **Caza M, Kronstad JW**. Shared and distinct mechanisms of iron acquisition by
576 bacterial and fungal pathogens of humans. Frontiers in Cellular and Infection
577 Microbiology [Internet]. 2013;**3**. Available from:
578 http://journal.frontiersin.org/article/10.3389/fcimb.2013.00080/abstract
579 13. **Chumduri C, Gurumurthy RK, Zietlow R, Meyer TF**. Subversion of host
580 genome integrity by bacterial pathogens. Nature Publishing Group [Internet].
581 2016;**17**:659–673. Available from: http://dx.doi.org/10.1038/nrm.2016.100
582 14. **Clermont O, Bonacorsi S, Bingen E**. Rapid and simple determination of the
583 *Escherichia coli* phylogenetic group. Applied and Environmental Microbiology.
584 2000;**66**:4555–4558.
585 15. **Cougnoux A, Dalmasso G, Martinez R, Buc E, Delmas J, Gibold L, Sauvanet**
586 **P, Darcha C, Déchelotte P, Bonnet M, et al.** Bacterial genotoxin colibactin
587 promotes colon tumour growth by inducing a senescence-associated
588 secretory phenotype. Gut. 2014;**63**:1932–1942.
589 16. **Croxen MA, Finlay BB**. Molecular mechanisms of *Escherichia coli*
590 pathogenicity. Nature Reviews Microbiology. 2010;**8**:26–38.
591 17. **Dalgakiran F, Witcomb LA, McCarthy AJ, Birchenough GMH, Taylor PW**.
592 Non-invasive model of neuropathogenic *Escherichia coli* infection in the
593 neonatal rat. Journal of visualized experiments : JoVE [Internet]. 2014;e52018.
594 Available from: http://www.jove.com/video/52018/non-invasive-model-
595 neuropathogenic-escherichia-coli-infection
596 18. **Dautin N**. Serine protease autotransporters of *Enterobacteriaceae* (SPATEs):
597 Biogenesis and function. Toxins. 2010;**2**:1179–1206.
598 19. **Dubois D, Delmas J, Cady A, Robin F, Sivignon A, Oswald E, Bonnet R**.
599 Cyclomodulins in urosepsis strains of *Escherichia coli*. Journal of Clinical
600 Microbiology. 2010;**48**:2122–2129.
601 20. **Escobar-Páramo P, Menac'h A Le, Gall T Le, Amorin C, Gouriou S, Picard B,**
602 **Skurnik D, Denamur E**. Identification of forces shaping the commensal
603 *Escherichia coli* genetic structure by comparing animal and human isolates.
604 Environmental Microbiology. 2006;**8**:1975–1984.
605 21. **Falbo V, Pace T, Picci L, Pizzi E, Caprioli A**. Isolation and nucleotide
606 sequence of the gene encoding cytotoxic necrotizing factor 1 of *Escherichia*
607 *coli*. Infection and Immunity. 1993;**61**:4909–4914.
608 22. **Feng Y, Mannion A, Madden CM, Swennes AG, Townes C, Byrd C, Marini**
609 **RP, Fox JG**. Cytotoxic *Escherichia coli* strains encoding colibactin and
610 cytotoxic necrotizing factor (CNF) colonize laboratory macaques. Gut
611 Pathogens [Internet]. 2017;**9**:1–15. Available from:

612 https://doi.org/10.1186/s13099-017-0220-y
613 23. **Féria C, Machado J, Correia JD, Gonçalves J, Gaastra W**. Virulence genes
614 and P fimbriae PapA subunit diversity in canine and feline uropathogenic
615 *Escherichia coli*. Veterinary Microbiology. 2001;**82**:81–89.
616 24. **Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ**. Urinary tract
617 infections: Epidemiology, mechanisms of infection and treatment options.
618 Nature Reviews Microbiology. 2015;**13**:269–284.
619 25. **Gagnaire A, Nadel B, Raoult D, Neefjes J, Gorvel JP**. Collateral damage:
620 Insights into bacterial mechanisms that predispose host cells to cancer.
621 Nature Reviews Microbiology. 2017;**15**:1–20.
622 26. **García A, Mannion A, Feng Y, Madden CM, Bakthavatchalu V, Shen Z, Ge Z,**
623 **Fox JG**. Cytotoxic *Escherichia coli* strains encoding colibactin colonize
624 laboratory mice. Microbes and Infection. 2016;**18**:777–786.
625 27. **Giridharan V V, Simões LR, Dagostin VS, Generoso JS, Rezin GT,**
626 **Florentino D, Muniz JP, Collodel A, Petronilho F, Quevedo J, et al.**
627 Temporal changes of oxidative stress markers in *Escherichia coli* K1-induced
628 experimental meningitis in a neonatal rat model. Neuroscience Letters
629 [Internet]. 2017;**653**:288–295. Available from:
630 http://dx.doi.org/10.1016/j.neulet.2017.06.002
631 28. **Gómez-Moreno R, Robledo IE, Baerga-Ortiz A**. Direct Detection and
632 Quantification of Bacterial Genes Associated with Inflammation in DNA
633 Isolated from Stool. Advances in Microbiology [Internet]. 2014;**04**:1065–1075.
634 Available from:
635 http://www.scirp.org/journal/doi.aspx?DOI=10.4236/aim.2014.415117
636 29. **Grant MA, Weagant SD, Feng P**. Glutamate Decarboxylase Genes as a
637 Prescreening Marker for Detection of Pathogenic *Escherichia coli* Groups.
638 Applied and Environmental Microbiology. 2001;**67**:3110–3114.
639 30. **Guignot J, Segura A, Tran Van Nhieu G**. The Serine Protease EspC from
640 Enteropathogenic *Escherichia coli* Regulates Pore Formation and Cytotoxicity
641 Mediated by the Type III Secretion System. PLoS Pathogens. 2015;**11**.
642 31. **He X, Zeng Q, Puthiyakunnon S, Zeng Z, Yang W, Qiu J, Du L, Boddu S, Wu**
643 **T, Cai D, et al.** *Lactobacillus rhamnosus* GG supernatant enhance neonatal
644 resistance to systemic Escherichia coli K1 infection by accelerating
645 development of intestinal defense. Scientific Reports [Internet]. 2017;**7**:1–14.
646 Available from: http://dx.doi.org/10.1038/srep43305
647 32. **Hinenoya A, Shima K, Asakura M, Nishimura K, Tsukamoto T, Ooka T,**
648 **Hayashi T, Ramamurthy T, Faruque SM, Yamasaki S**. Molecular
649 characterization of cytolethal distending toxin gene-positive *Escherichia coli*
650 from healthy cattle and swine in Nara, Japan. BMC Microbiology. 2014;**14**.
651 33. **Huang Y-Y, Wintner A, Seed PC, Brauns T, Gelfand JA, Hamblin MR**.
652 Antimicrobial photodynamic therapy mediated by methylene blue and
653 potassium iodide to treat urinary tract infection in a female rat model.
654 Scientific Reports. 2018;**8**:7257.
655 34. **Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM,**
656 **Aarestrup FM**. Real-time whole-genome sequencing for routine typing,
657 surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. Journal

658 of Clinical Microbiology. 2014;**52**:1501–1510.
659 35. **Johnson JR, Kaster N, Kuskowski MA, Ling G V.** Identification of
660 urovirulence traits in *Escherichia coli* by comparison of urinary and rectal *E.*
661 *coli* isolates from dogs with urinary tract infection. Journal of Clinical
662 Microbiology. 2003;**41**:337–345.
663 36. **Johnson JR, Stell AL, Delavari P, Murray AC, Kuskowski M, Gaastra W**.
664 Phylogenetic and Pathotypic Similarities between *Escherichia coli* Isolates
665 from Urinary Tract Infections in Dogs and Extraintestinal Infections in
666 Humans. J Infect Dis [Internet]. 2001;**183**:897–906. Available from:
667 http://jid.oxfordjournals.org/content/183/6/897.abstract\nhttp://www.ncb
668 i.nlm.nih.gov/pubmed/11237806
669 37. **Johnson TJ, Wannemuehler YM, Nolan LK**. Evolution of the *iss* gene in
670 *Escherichia coli*. Applied and Environmental Microbiology. 2008;**74**:2360–
671 2369.
672 38. **Jordan DM, Cornick N, Torres AG, Dean-Nystrom EA, Kaper JB, Moon HW**.
673 Long polar fimbriae contribute to colonization by *Escherichia coli* O157:H7 in
674 vivo. Infection and Immunity. 2004;**72**:6168–6171.
675 39. **Kaper J, Nataro J, Mobley H**. Pathogenic *Escherichia coli*. Nature Reviews
676 Microbiology. 2004;**2**:;123–140.
677 40. **Knöbl T, Micke Moreno A, Paixo R, Gomes TAT, Vieira MAM, Silva Leite D**
678 **Da, Blanco JE, Ferreira AJP**. Prevalence of avian pathogenic *Escherichia coli*
679 (APEC) clone harboring *sfa* gene in Brazil. The Scientific World Journal.
680 2012;**2012**:437342.
681 41. **Kolmogorov M, Raney B, Paten B, Pham S**. Ragout - A reference-assisted
682 assembly tool for bacterial genomes. Bioinformatics. 2014;**30**:i302–i309.
683 42. **Krautkramer KA, Kreznar JH, Romano KA, Vivas EI, Barrett-Wilt GA,**
684 **Rabaglia ME, Keller MP, Attie AD, Rey FE, Denu JM**. Diet-Microbiota
685 Interactions Mediate Global Epigenetic Programming in Multiple Host Tissues.
686 Molecular cell [Internet]. 2016;**64**:982–992. Available from:
687 http://www.ncbi.nlm.nih.gov/pubmed/27889451
688 43. **Landraud L, Gauthier M, Fosse T, Boquet P**. Frequency of *Escherichia coli*
689 strains producing the cytotoxic necrotizing factor (CNF1) in nosocomial
690 urinary tract infections. Letters in Applied Microbiology. 2000;**30**:213–216.
691 44. **Leser TD, Molbak L**. Better living through microbial action: The benefits of
692 the mammalian gastrointestinal microbiota on the host. Environmental
693 Microbiology. 2009;**11**:2194–2206.
694 45. **López-Banda DA, Carrillo-Casas EM, Leyva-Leyva M, Orozco-Hoyuela G,**
695 **Manjarrez-Hernández ÁH, Arroyo-Escalante S, Moncada-Barrón D,**
696 **Villanueva-Recillas S, Xicohtencatl-Cortes J, Hernández-Castro R**.
697 Identification of virulence factors genes in *Escherichia coli* isolates from
698 women with urinary tract infection in Mexico. BioMed Research International.
699 2014: 959206.
700 46. **Lynne AM, Skyberg JA, Logue CM, Nolan LK**. Detection of Iss and Bor on the
701 surface of *Escherichia coli*. Journal of Applied Microbiology. 2007;**102**: 660–
702 666.
703 47. **Marini RP, Taylor NS, Liang AY, Knox KA, Peña JA, Schauer DB, Fox JG**.

704 Characterization of hemolytic *Escherichia coli* strains in ferrets: Recognition of
705 candidate virulence factor CNF1. Journal of Clinical Microbiology.
706 2004;**42**:5904–5908.
707 48. **Marre R, Kreft B, Hacker J**. Genetically engineered S and F1C fimbriae differ
708 in their contribution to adherence of *Escherichia coli* to cultured renal tubular
709 cells. Infection and Immunity. 1990;**58**:3434–3437.
710 49. **Martin HR, Taylor NS, Buckley EM, Marini RP, Patterson MM, Fox JG**.
711 Characterization of cytotoxic necrotizing factor 1-producing *Escherichia coli*
712 strains from faeces of healthy macaques. Journal of Medical Microbiology.
713 2009;**58**:1354–1358.
714 50. **Massot M, Daubié AS, Clermont O, Jauréguy F, Couffignal C, Dahbi G, Mora**
715 **A, Blanco J, Branger C, Mentré F, et al.** Phylogenetic, virulence and antibiotic
716 resistance characteristics of commensal strain populations of *Escherichia coli*
717 from community subjects in the paris area in 2010 and evolution over 30
718 years. Microbiology (United Kingdom). 2016;**162**:642–650.
719 51. **McCarthy AJ, Martin P, Cloup E, Stabler RA, Oswald E, Taylor PW**. The
720 genotoxin colibactin is a determinant of virulence in *Escherichia coli* K1
721 experimental neonatal systemic infection. Infection and Immunity. 2015;**83**:
722 3704–3711.
723 52. **Miajlovic H, Smith SG**. Bacterial self-defence: How *Escherichia coli* evades
724 serum killing. FEMS Microbiology Letters. 2014;**354**:1–9.
725 53. **Micenková L, Bosák J, Štaudová B, Kohoutová D, Čejková D, Woznicová V,**
726 **Vrba M, Ševčíková A, Bureš J, Šmajs D**. Microcin determinants are
727 associated with B2 phylogroup of human fecal *Escherichia coli* isolates.
728 Microbiology Open. 2016;**5**: 490–498.
729 54. **Micenková L, Štaudová B, Bosák J, Mikalová L, Littnerová S, Vrba M,**
730 **Ševčíková A, Woznicová V, Šmajs D**. Bacteriocin-encoding genes and ExPEC
731 virulence determinants are associated in human fecal *Escherichia coli* strains.
732 BMC Microbiology. 2014;**14**.
733 55. **Nougayrede J-P**. *Escherichia coli* Induces DNA Double-Strand Breaks in
734 Eukaryotic Cells. Science [Internet]. 2006;**313**:848–851. Available from:
735 http://www.sciencemag.org/cgi/doi/10.1126/science.1127059
736 56. **Pandey M, Khan A, Das SC, Sarkar B, Kahali S, Chakraborty S,**
737 **Chattopadhyay S, Yamasaki S, Takeda Y, Nair GB, et al.** Association of
738 cytolethal distending toxin locus *cdtB* with enteropathogenic *Escherichia coli*
739 isolated from patients with acute diarrhea in Calcutta, India. J Clin Microbiol.
740 2003;**41**:5277–5281.
741 57. **Paula Coelho C de, Motta PD, Petrillo M, Oliveira Iovine R de, Dalboni LC,**
742 **Santana FR, Correia MSF, Casarin RCV, Carvalho VM, Bonamin LV**.
743 Homeopathic medicine Cantharis modulates uropathogenic *E. coli* (UPEC)-
744 induced cystitis in susceptible mice. Cytokine. 2017;**92**:103–109.
745 58. **Payros D, Secher T, Boury M, Brehin C, Ménard S, Salvadorcartier C,**
746 **Cuevas-Ramos G, Watrin C, Marcq I, Nougayrède JP, et al.** Maternally
747 acquired genotoxic *Escherichia coli* alters offspring's intestinal homeostasis.
748 Gut Microbes. 2014;**5**:313–325.
749 59. **Picard B, Garcia JS, Gouriou S, Duriez P, Brahimi N, Bingen E, Elion J,**

750 **Denamur E**. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection? Infection and Immunity. 1999;**67**:546–553.

752 60. **Raisch J, Buc E, Bonnet M, Sauvanet P, Vazeille E, Vallée A de, Déchelotte P, Darcha C, Pezet D, Bonnet R, et al.** Colon cancer-associated B2 *Escherichia coli* colonize gut mucosa and promote cell proliferation. World Journal of Gastroenterology. 2014;**20**: 6560–6572.

756 61. **Rosadi F, Fiorentini C, Fabbri A**. Bacterial protein toxins in human cancers. Pathogens and Disease. 2016;**74**: 1–11.

758 62. **Rosenberg S, Horowitz R, Coppenhagen-Glazer S, Pizov G, Elia A, Gofrit ON, Ginsburg I, Pode D**. Intravesical administration of green tea extract attenuates the inflammatory response of bacterial cystitis - A rat model. BJU International. 2014;**114**: 601–607.

762 63. **Ross BN, Rojas-Lopez M, Cieza RJ, McWilliams BD, Torres AG**. The role of long polar fimbriae in *Escherichia coli* O104:H4 adhesion and colonization. PLoS ONE. 2015;**10**.

765 64. **Rycke J De, González EA, Blanco J, Oswald E, Blanco M, Boivin R**. Evidence for two types of cytotoxic necrotizing factor in human and animal clinical isolates of *Escherichia coli*. Journal of clinical microbiology [Internet]. 1990;**28**:694–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=267779&tool=pmcentrez&rendertype=abstract

771 65. **Secher T, Brehin C, Oswald E**. Early settlers: which *E. coli* strains do you not want at birth? American Journal of Physiology - Gastrointestinal and Liver Physiology [Internet]. 2016;**311**: G123–G129. Available from: http://ajpgi.physiology.org/lookup/doi/10.1152/ajpgi.00091.2016

775 66. **Shin S, Castanie-Cornet MP, Foster JW, Crawford JA, Brinkley C, Kaper JB**. An activator of glutamate decarboxylase genes regulates the expression of enteropathogenic *Escherichia coli* virulence genes through control of the plasmid-encoded regulator, per. Molecular Microbiology. 2001;**41**: 1133–1150.

780 67. **Siqueira AK, Ribeiro MG, Leite D da S, Tiba MR, Moura C de, Lopes MD, Prestes NC, Salerno T, Silva AV da.** Virulence factors in *Escherichia coli* strains isolated from urinary tract infection and pyometra cases and from feces of healthy dogs. Research in Veterinary Science. 2009;**86**: 206–210.

784 68. **Skowron B, Baranowska A, Kaszuba-Zwoińska J, Więcek G, Malska-Woźniak A, Heczko P, Strus M**. Experimental model for acute kidney injury caused by uropathogenic *Escherichia coli*. Postepy higieny i medycyny doswiadczalnej (Online). 2017;**71**: 520–9.

788 69. **Taieb F, Petit C, Nougayrède J, Oswald E**. The Enterobacterial Genotoxins: Cytolethal Distending Toxin and Colibactin. EcoSal Plus [Internet]. 2016;**7**. Available from: http://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosalplus.ESP-0008-2016

793 70. **Tenaillon O, Skurnik D, Picard B, Denamur E**. The population genetics of commensal *Escherichia coli*. Nature Reviews Microbiology. 2010;**8**:207–217.

795 71. **Torres AG, Blanco M, Valenzuela P, Slater TM, Patel SD, Dahbi G, López C,**

796 **Barriga XF, Blanco JE, Gomes TAT, et al.** Genes related to long polar
797 fimbriae of pathogenic *Escherichia coli* strains as reliable markers to identify
798 virulent isolates. Journal of Clinical Microbiology. 2009;**47**: 2442–2451.
799 72. **Tóth I, Hérault F, Beutin L, Oswald E.** Production of cytolethal distending
800 toxins by pathogenic *Escherichia coli* strains isolated from human and animal
801 sources: Establishment of the existence of a new *cdt* variant (type IV). Journal
802 of Clinical Microbiology. 2003;**41**: 4285–4291.
803 73. **Tóth I, Oswald E, Mainil JG, Awad-Masalmeh M, Nagy B**. Characterization of
804 intestinal *cnf1+ Escherichia coli* from weaned pigs. International journal of
805 medical microbiology : IJMM [Internet]. 2000;**290**: 539–42. Available from:
806 http://www.ncbi.nlm.nih.gov/pubmed/11100828\nhttp://www.sciencedire
807 ct.com/science/article/pii/S1438422100800193
808 74. **Trautman EP, Healy AR, Shine EE, Herzon SB, Crawford JM**. Domain-
809 Targeted Metabolomics Delineates the Heterocycle Assembly Steps of
810 Colibactin Biosynthesis. Journal of the American Chemical Society. 2017;**139**:
811 4195–4201.
812 75. **Turner SM, Scott-Tucker A, Cooper LM, Henderson IR**. Weapons of mass
813 destruction: Virulence factors of the global killer enterotoxigenic *Escherichia*
814 *coli*. FEMS Microbiology Letters. 2006;**263**: 10–20.
815 76. **Veltri D, Wight MM, Crouch JA**. SimpleSynteny: a web-based tool for
816 visualization of microsynteny across multiple species. Nucleic acids research.
817 2016;**44**: W41–W45.
818 77. **Vysakh A, Suma D, Jayesh K, Jyothis M, Latha MS**. The influence of tissue
819 antioxidant enzymes and inflammatory cascade in pathology of cystitis: An
820 experimental rat model. Microbial Pathogenesis. 2017;**113**: 102–106.
821 78. **Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N,**
822 **Dietrich EM, Disz T, Gabbard JL, et al.** Improvements to PATRIC, the all-
823 bacterial bioinformatics database and analysis resource center. Nucleic Acids
824 Research. 2017;**45**: D535–D542.
825 79. **Witcomb LA, Collins JW, McCarthy AJ, Frankel G, Taylor PW**.
826 Bioluminescent imaging reveals novel patterns of colonization and invasion in
827 systemic *Escherichia coli* K1 experimental infection in the neonatal rat.
828 Infection and Immunity. 2015;**83**: 4528–4540.
829 80. **Yoon J, Chul HC, Park PB, Min J**. Efficacy of Yeast ' Vacuoles as Antimicrobial
830 Agents to *Escherichia coli* Bacteremia in Rat. Current Microbiology [Internet].
831 2017;**74**: 22–27. Available from: "http://dx.doi.org/10.1007/s00284-016-
832 1146-1
833 81. **Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O,**
834 **Aarestrup FM, Larsen MV**. Identification of acquired antimicrobial resistance
835 genes. Journal of Antimicrobial Chemotherapy. 2012;**67**: 2640–2644.
836 82. **Zeng Q, He X, Puthiyakunnon S, Xiao H, Gong Z, Boddu S, Chen L, Tian H,**
837 **Huang SH, Cao H**. Probiotic mixture golden Bifido prevents neonatal
838 *Escherichia coli* K1 translocation via enhancing intestinal defense. Frontiers in
839 Microbiology. 2017;**8**:1–13.
840 83. **Zha L, Wilson MR, Brotherton CA, Balskus EP**. Characterization of
841 Polyketide Synthase Machinery from the *pks* Island Facilitates Isolation of a

842 Candidate Precolibactin. ACS Chemical Biology. 2016;**11**: 1287–1295.
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859

Table 1. Distribution of genotoxin prevalence by vendor origin and institution destination.

Vendor A	Total	Institution W	Institution X	Institution Z
Total pks+ E.coli	29/53 (55%)	4/5 (80%)	24/47 (51%)	1/1
Total cdt+ E.coli	8/53 (15%)	0/5 (0%)	8/47 (17%)	0/1
Total cnf + E. coli	0/53 (0%)	0/5 (0%)	0/47 (0%)	0/1
pks-/ cdt-/ cnf-	24	1	23	0
pks-/ cdt+/cnf-	0	0	0	0
pks-/cdt-/cnf+	0	0	0	0
pks+/ cdt-/cnf -	21	*4*	*16*	1
pks+/ cdt+/cnf-	8	*0*	*8*	0
pks+/ cdt-/cnf+	0	*0*	*0*	0

Vendor B	Total	Institution Y	Institution Z
Total pks+ E.coli	13/13(100%)	11/11 (100%)	2/2 (100%)
Total cdt+ E.coli	9/13 (69%)	7/11 (64%)	2/2 (100%)
Total cnf + E. coli	4/13 (31%)	4/11 (36%)	0/2 (0%)
pks-/ cdt-/ cnf-	0	0	0
pks-/ cdt+/cnf-	0	0	0
pks-/cdt-/cnf+	0	0	0
pks+/ cdt-/cnf -	0	*0*	*0*
pks+/ cdt+/cnf-	9	*7*	*2*
pks+/ cdt-/cnf+	4	*4*	*0*

Vendor C	Institution Z
Total pks+ E.coli	3/3(100%)
Total cdt+ E.coli	3/3 (100%)
Total cnf + E. coli	0/3 (0%)
pks-/ cdt-/ cnf-	0
pks-/ cdt+/cnf-	0
pks-/cdt-/cnf+	0
pks+/ cdt-/cnf -	0
pks+/ cdt+/cnf-	3
pks+/ cdt-/cnf+	0

All Isolates	Total
Total pks+ E. coli	45/69 (65%)
Total cdt+ E. coli	20/69 (29%)
Total cnf + E. coli	4/69 (6%)
pks-/ cdt-/ cnf-	24
pks-/ cdt+/cnf-	0
pks-/cdt-/cnf+	0
pks+/ cdt-/cnf -	21
pks+/ cdt+/cnf-	20
pks+/ cdt-/cnf+	4

Table 2: Serotype and virulence factors testing results of *E. coli* isolates from rats

Sample#	O-type	H-type	Vendor	Institution	elt	estA	estB	stx1	stx2	eae	cnf1	cnf2
S1: pks-, cdt-, cnf-	179	8	A	W	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
S16: pks-, cdt-, cnf-	179	8	A	X	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
S5: pks+, cdt-, cnf-	7	7	A	Z	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
S2: pks+, cdt-, cnf-	7	7	A	W	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
S14: pks+, cdt+, cnf-	7	7	A	X	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
S4: pks+, cdt+, cnf-	166	6	B	Z	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
S8: pks+, cdt+, cnf-	166	6	B	Y	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
S7: pks+, cdt+, cnf-	M	6	C	Z	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
S9: pks+, cdt-, cnf+	4	5	B	Y	NEG	NEG	NEG	NEG	NEG	NEG	POS	NEG

Table 3. Novel rat *E. coli* genomes have similar statistics as pathogenic, *PKS*-encoding *E. coli* strains IHE3034 and NC101. Virulence factor genes for toxins, survival factors, and adhesions were identified in the rat *E. coli* genomes. *astA*: EAST-1 heat-stable toxin; *cba*: Colicin B; *cdtABC*: Cytolethal distending toxin subunits A, B,C ; *celb*: Endonuclease colicin E2; *cma*: Colicin M; *gad*: Glutamate decarboxylase; *iroN*: Enterobactin siderophore receptor protein; *iss*: Increased serum survival; *lpfA*: Long polar fimbriae; *mchB*: Microcin H47 part of colicin H; *mchC*: MchC protein; *mchF*: ABC transporter protein MchF; *mcmA*: Microcin M part of colicin H; *pic*: serine protease autotransporters of Enterobacteriaceae (SPATE); *PKS*: polyketide synthetase (colibactin); *sfaS*: S-fimbriae minor subunit; *vat*: vacuolating autotransporter toxin

Strain	Isolation Source	Genome Length (bp)	Contigs	G+C% Content	Protein Coding Sequences (CDS)	tRNA	rRNA	Virulence Factors Genes	GenBank Accession
S11	Research rat	5201802	32	49.72	5149	81	10	*cdtABC, gad, iroN, iss, mchB, mchC, mchF, mcmA, pic, PKS, vat*	NHYT00000000
S14		5208467	37	49.64	5153	81	10		NHYQ00000000
S12		5092914	14	50.04	4995	77	11	*iroN, iss, PKS, sfaS, vat*	NHYS00000000
S13		5296109	47	48.57	5101	73	9	*astA, cba, cma, gad, lpfA, pic, PKS*	NHYR00000000
S15		5248403	58	47.71	5078	79	8	*celb, gad, iss, PKS*	NHYP00000000
S16		5623575	140	45.33	5509	70	3	*cdtABC, celb, gad, iss, PKS*	NHYO00000000
S17		5139109	40	49.47	5022	76	9	*cnf1, gad, iroN, mchB, mchC, mchF, mcmA, PKS, vat*	QLVH00000000
IHE3034	Human neonatal meningitis	5108383	1 (complete genome)	50.70	5045	97	22	*gad, iroN, iss, PKS, sfaS, vat, cdtABC*	CP001969.1
NC101	Research mouse	5021144	27	50.57	4917	72	4	*gad, iroN, iss, PKS, sfaS, vat*	AEFA00000000.1
K-12 substr. DH10B	Human non-pathogenic	4686137	1 (complete genome)	50.80	4606	87	14	*gad, iss*	CP000948.1