The microbial ecology of *Escherichia coli* in the vertebrate gut

Ebenezer Foster-Nyarko[1,2, 3#], Mark J. Pallen[1,4,5]

[1] Quadram Institute Bioscience, Norwich Research Park, Norwich, Norfolk, United Kingdom

[2] Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, United Kingdom

[3] Department of Infection Biology, London School of Hygiene and Tropical Medicine, London, United Kingdom

[4] School of Veterinary Medicine, University of Surrey, Guildford, Surrey, United Kingdom

[5] School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, United Kingdom

[#]**Corresponding author**: Dr Ebenezer Foster-Nyarko, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom

Email: Ebenezer.Foster-Nyarko2@lshtm.ac.uk

Telephone: +44(0)7493550435

ABSTRACT

Escherichia coli has a rich history as biology's "rock star", driving advances across many fields. In the wild, *E. coli* resides innocuously in the gut of humans and animals but is also a versatile pathogen commonly associated with intestinal and extraintestinal infections and antimicrobial resistance—including large foodborne outbreaks such as the one that swept across Europe in 2011, killing fifty-four individuals and causing approximately 4,000 infections and 900 cases of haemolytic uraemic syndrome. Given that most *E. coli* are harmless gut colonisers, an important ecological question plaguing microbiologists is what makes *E. coli* an occasionally devastating pathogen? To address this question requires an enhanced understanding of the ecology of the organism as a commensal. Here, we review how our knowledge of the ecology and within-host diversity of this organism in the vertebrate gut has progressed in the 136 since *E. coli* was first described. We also review current approaches to the study of within-host bacterial diversity. In closing, we discuss some of the outstanding questions yet to be addressed and prospects for future research.

INTRODUCTION

Escherichia coli: **a versatile organism**

"Tout ce qui est vrai pour le Colibacille est vrai pour l'éléphant" (All that is true of *E. coli* is also true of the elephant) [1]. This widely quoted epigram reflects the exalted status of *E. coli* in microbiology as the favourite model organism. Many advances in biology have been driven by studies with *E. coli*, particularly, using the strain designated K-12 and its derivatives [2, 3] (Table 1).

Theodor Escherich, a Bavarian doctor, first described *Escherichia coli* on 14 July 1885 in a lecture to the Society for Morphology and Physiology in Munich [4]. Over fifteen months, Escherich observed and isolated nineteen different bacteria in Otto von Bollinger's bacteriology laboratory [4], including the *Bacterium coli commune* ("the common colon bacillus" [5, 6])—which we now know as *E. coli*—and *Bacterium lactis aërogenes* (now *Klebsiella pneumoniae*). Aided by Christian Gram's new staining technique [3, 7, 8], Escherich described the organism as a Gram-negative bacillus of approximately 1.1-1.5 μm x 2.0-6.0 μm. The name *Escherichia coli* was proposed by Castellani and Chalmers in 1919 to honour Escherich and was officially adopted in 1958 by the Judicial Commission of the International Committee on Systematic Bacteriology. It was subsequently included in the Approved Lists of Bacterial Names in 1980 [7, 9-11].

In the wild, *E. coli* exists as a common resident of the vertebrate gut and non-host associated habitats such as water, soil, manure and food [12-15]. Due to its remarkable metabolic and regulatory abilities, *E. coli* can also survive under prolonged periods of non-growth [16]. Thus, *E. coli* represents a highly versatile species, capable of adapting to many different

ecological habitats [16]. *E. coli* is also a versatile pathogen, eliciting a broad spectrum of

diseases and responsible for at least two million human deaths per year [17]. The organism's

role in intestinal and extraintestinal disease was recognised not long after its discovery [3].

Ørskov and Ørskov note, "Any *E. coli* strain can probably cause invasive disease given the

right opportunities, and *E. coli* has therefore aptly been called an opportunistic pathogen"

[18].

E. coli is a facultative anaerobe, meaning that it can grow in aerobic and anaerobic

conditions. It is the most common aerobe in the lower intestine of mammals [2]; however, it

typically constitutes only 0.1-5% of a gut microbial community comprised of over 500 other

bacterial species [2]. This community is typically dominated by obligate anaerobes,

including members of Firmicutes and Bacteroidetes—which make up at least 90% of the gut

microbial population [2, 19, 20]. Nevertheless, *E. coli* can hold its ground in this highly

competitive, ever-changing niche, existing in a life-long relationship with its host.

E. coli **pathotypes**

Pathogenic *E. coli* are classified into "pathotypes" or "pathovars" [21, 22] (Table 2) based on

several criteria, including the site of infection (e.g., uropathogenic *E. coli* (UPEC), named for

their impact on the urinary tract, extraintestinal pathogenic *E. coli* (ExPEC), which cause

infections in organs outside the gut), host (e.g., avian pathogenic *E. coli* (APEC), named after

infections in avian species), site and host (e.g., neonatal meningitis *E. coli* (NMEC) which

infect the cerebrospinal fluid in new-borns) and pathogenesis (e.g., Shiga-toxigenic *E. coli*

(STEC)).

The pathogenic strains of *E. coli* have acquired specific virulence factors that enable them to adapt to new niches and cause a wide variety of diseases [21, 23]. These include adhesion/colonisation factors, toxins and effectors enabling pathogenic strains to colonise sites such as the urethra and small intestine and affect various fundamental eukaryotic processes [21] (Tables 3 and 4). For example, pathogenic *E. coli* can adhere to host epithelial cells using adhesion molecules known as intimin [24]. Pathogenic *E. coli* are also known to induce inflammation in host cells, resulting in the production of inducible nitric oxide synthase by the host innate immune cells. This results in the accumulation of nitrate, which the pathogenic strains can utilise to produce energy [25]. Uropathogenic *E. coli* strains are equipped with type I fimbriae, *AfA/Dr* adhesins and pyelonephritis-associated pili (PAP) that enable them to colonise and infect the urinary tract. NMEC and sepsis-associated *E. coli* are armed with the K1 polysaccharide capsule that facilitates their evasion from host complement-mediated killing [16, 26]. Differential utilisation of nutrients is another way in which pathogenic strains of *E. coli* can overcome the resistance mounted by the commensal strains. For example, pathogenic EHEC can use the sugars galactose, mannose, ribose and hexuronates as carbon sources, while commensal *E. coli* cannot [27, 28]. Besides, EHEC possesses the *eut* operon, which enables it to assimilate nutrients such as ethanolamine, whereas commensal *E. coli* lack this operon and are therefore unable to utilise ethanolamine [27, 29, 30].

Despite its designation as a separate genus, *Shigella* is classified as an intestinal pathogenic *E. coli* (InPEC) pathotype (InPEC encompasses Adherent Invasive *E. coli* (AIEC) and diarrheagenic *E. coli*). It possesses virulence traits and pathogenicity that closely resemble enteroinvasive *E. coli* (EIEC) and is thus regarded as an EIEC pathotype [31, 32]. Phylogenomic data support the classification of *Shigella* as an *E. coli* pathotype [33, 34],

even though its nomenclature has been retained solely for historical reasons and to avoid confusion in the clinical setting. Chaudhuri and Henderson [35] advocate that the clinical and academic community working on *E. coli/Shigella* adopts a similar approach as was used in the classification of species such as *Salmonella typhimurium, S. enteritidis* and *S. typhi* as *Salmonella enterica* subgroups so that the *Shigella* species are re-designated as subspecies within *E. coli* to avoid their neglect in *E. coli* studies.

Despite the importance of pathotyping to the epidemiology and pathogenesis of strains, the recent discovery of complex hybrid pathotypes has challenged some of the classical pathotype definitions [23], as illustrated in the following examples.

The strain that caused an outbreak of foodborne illness in Germany in 2011 [36], which swept across most of Europe, killing fifty-four individuals and causing approximately 4,000 infections including 900 cases of haemolytic uraemic syndrome (HUS), was, in fact, an EHEC-EAEC hybrid strain [23, 37] belonging to serotype O104:H4 and sequence type ST678. This strain combined virulence characteristics of EHEC (Shiga toxin production) and adherence typical of EAEC strains, despite lacking the Type III secretion and *tir*/intimin system. It also possessed virulence factors commonly found in ExPEC strains, such as yersiniabactin and aerobactin (iron acquisition factors) and demonstrated expanded-spectrum beta-lactamase resistance [37].

A hybrid clone belonging to serotype O80:H2 and sequence type ST301 (clonal complex 165) has emerged in France, Belgium and Switzerland, capable of causing HUS and bacteraemia [38-42]. This strain possesses all the virulence factors typical of EHEC strains, such as intimin, Shiga toxin production and enterohaemolysin, yet it belongs to the

phylogenetic group A, unlike other EHEC strains [39]. It possesses a large plasmid (>100 kb), which encodes a resistance cassette, providing resistance characteristics to a wide range of antimicrobials, including cotrimoxazole, tetracyclines, streptomycin and penicillin [38, 43]. The carriage of large plasmids is commonly associated with ExPEC strains [44, 45].

E. coli are classified into eight major phylogenetic groups (phylotypes or phylogroups) linked with the ecological niche, lifestyle and pathogenic potential of strains [14, 46]. Phylogroups B2, D, F and H appear to be the most basal taxa, with phylogroup E emerging before phylogroups C, B1 and A, which are thought to be the most recently diverged lineages [47, 48]. The most anciently diverged encompass mostly extraintestinal pathogenic strains. In contrast, the most recently diverged lineages span strains that are associated with life-threatening intestinal diseases such as dysentery and haemolytic uraemic syndrome [49, 50]. Hybrid clones exhibiting characteristics of the B2 phylogenetic backbone typical of ExPEC strains and EPEC and STEC attributes have been recently reported in the literature, further blurring the lines of pathotype boundaries that have been used to define *E. coli* pathogens for so long. Isolates belonging to serotype O153:H10 and harbouring *eae* (depicting an atypical EPEC-ExPEC hybrid pathotype) have been detected in meat, poultry farms, human diarrhoeagenic samples and wildlife from northwest Spain [51]. Similarly, isolates belonging to serotype O137:H6 and ST2678 and positive for *eae*, *bfpA* and *stx2f* genes—demonstrating typical EPEC-STEC hybrid pathotype—have been isolated in exotic psittacine birds [52]. In addition, strains belonging to serotype O2:H4 and ST141 and demonstrating uropathogenic traits have been found with some characteristics of the EHEC pathotype [53, 54]. Also, an ST12 strain belonging to serotype O4:H1 was found to harbour the locus of enterocyte effacement pathogenicity island and the bundle forming pili protein (encoded by the *bfp* gene)* typical of the EPEC pathotype but simultaneously causing diarrhoea and bacteraemia

in the same patient [55]. Furthermore, some EPEC isolates that harbour the heat-stable enterotoxin produced by ETEC strains have been described [24].

These examples amply demonstrate that pathotyping is limited in its capacity to adapt to new strains that fail to respect the currently utilised pathotype boundaries. However, it can be argued that the emergence of pathovars that exhibit traits uncharacteristic of their genetic background is not unexpected, given that pathotypes are usually defined based on horizontally acquired elements like plasmids and genomic islands. Thus, while the majority of a pathovar's membership may show similarity in pathogenic phenotype, horizontal genetic exchange would result in an elastic continuum within a pathovar's definition.

A further limitation lies in the use of negative criteria to identify pathogens. For instance, a strain is described as a typical EPEC if it presents the locus of enterocyte effacement but does not produce Shiga toxin. On the other hand, if a strain both lacks Shiga toxin and *bfp,* it is classified as an atypical EPEC [21]. As Robins-Browne *et al.* [56] point out, "characterising pathogens based on their lack of one or more virulence determinants may group several types of distantly related or unrelated bacteria together and cause some distinct pathogenic categories with uncharacterised virulence determinants to be overlooked".

THE TAXONOMY OF *ESCHERICHIA*

In the pre-molecular era, several non-*coli* species were designated under the *Escherichia* genus based on DNA relatedness/hybridisation and overall phenotypic similarity: *E. blattae* (1973) [57], *E. vulneris* (1982) [58], *E. hermannii* (1982) [59] and *E. fergusonii* (1985) [60].

Lawrence *et al.* [61] utilised DNA sequences of the slowly evolving genes *gap* and *ompA* (encoding glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and outer membrane protein 3A respectively), representing conserved genes across eleven species of enteric bacteria, to map the phylogenetic relationships among the above mentioned five species. Their analysis revealed these species to be distantly related, rather than a monophyletic group. Apart from *E. fergusonii*, the other species were more divergent from *E. coli* than *Salmonella*. As *Salmonella* is traditionally considered most closely related to *E. coli*, the results justified the redesignation of *E. blattae*, *E. hermannii* and *E. vulneris* in alternative genera [61].

Subsequently, in 1991, Albert *et al.* [62] isolated a diarrhoeagenic isolate with EPEC-like phenotypic and genetic features from a nine-month old girl with watery diarrhoea in Bangladesh, initially designated *Hafnia alvei.* Subsequent identification and characterisation of five "*H. alvei*-like" isolates by DNA-DNA hybridization, phenotypic characterization and 16S rDNA sequencing led to their redesignation as *E. albertii* [63]. Members of this species were later found to be closely related to *Shigella boydii* serotype 13—a divergent lineage in *Escherichia*; the *E. albertii*/*Shigella boydii* serotype 13 lineage estimated to have diverged from a most-recent common ancestor circa 28 million years ago [64].

Consequently, the taxonomy of the genus *Escherichia* has been modified with the reclassification of *E. hermannii*, *E. blattae* and *E. vulneris* to other genera and the description of five cryptic clades within *Escherichia* [46, 65-67]. Until recently, the genus was comprised of three named species: *E. albertii*, *E. coli*, *E. fergusonii* and five cryptic clades, designated *Escherichia* clades I-V [23]. Of the three named species within the genus, *E. albertii* is the most divergent species, while *E. fergusonii* is closely related to *E. coli sensu stricto* [68].

The name *E. marmotae* has recently been validly published for Clade V [69, 70], although the species is not restricted to marmots [71, 72]. This builds on an earlier suggestion to classify clade V as a novel species and clades III and IV combined as a second novel species, based on digital DNA-DNA association and Average Nucleotide Identity (ANI) data [46] (an ANI of 95% between two genomes is generally taken as the standard for the demarcation of prokaryotic species [67, 73]). Furthermore, cryptic clade I strains possess the trademark virulence traits of *E. coli*, besides their potential to cause human infections [68, 74, 75]. Thus, *E. coli* and *Escherichia* clade I are now designated as *E. coli sensu lato* and the classic *E. coli* (designated phylogroups A-G) as *E. coli sensu stricto* [76]. Recently, Gilroy *et al.* [72] assigned the species name *E. whittamii* to clade II, in honour of the American bacteriologist, Thomas S. Whittam, for his contributions to the study of *E. coli*. van der Putten *et al.* have also recently proposed that the cryptic clades III and IV be named *E. ruysiae* sp. nov [77].

Walk and colleagues [46] estimated the lineages' divergence times that gave rise to each of the cryptic clades using a minimum evolution tree. Based on the assumption that *E. coli* split from *Salmonella enterica* between 100 to 160 million years ago, they estimated that the *Escherichia* lineages shared a common ancestor between 48-75 million years ago. They suggested that *E. albertii*, *E. fergusonii*, *Escherichia* clade II and clade V split between 38-75 million years ago, whereas *E. coli*, *Escherichia* clade I, clade III and clade IV split between 19-31 million years ago [46].

As evidenced by earlier efforts at classifying *E. blattae*, *E. hermanii* and *E. vulneris*, metabolic characteristics such as the utilisation of specific carbon sources or the production/catabolism of certain biochemical compounds are insufficient to delineate strains

into species and species groups accurately. First, multiple genes are often required for the expression of a particular phenotype [61]. Second, convergent evolution in distantly related species confounds the delineation of species based on these phenetic characteristics. Similarly, DNA hybridisation fails to consider the relatedness among congeneric species reliably [61].

The Genome Taxonomy Database Toolkit (GTDB-Tk), a tool for the automatic classification of draft bacterial and archaeal genomes, was published recently [78]. GTDB-Tk inputs genome assemblies in FASTA format and predicts the placements within domains based on identifying a set of 120 bacterial and archaeal marker genes and domain-specific reference trees. Then, species designations are computed using a GTDB reference tree, the relative evolutionary divergence parameter and ANI values. However, GTDB automatically assigns a random alphanumeric designation to novel species, which do not scale well to the increasing number of newly identified novel species and are often confusing and user-unfriendly [72]. To address this gap, Pallen *et al.* [79] recently developed an automated combinatorial approach to creating more than one million Linnaean binomials for Bacteria and Archaea.

THE EMERGENCE OF ANTIMICROBIAL RESISTANCE IN *E. COLI*

E. coli has long been used as an indicator bacterium for monitoring the faecal contamination of food and water and antimicrobial resistance (AMR) in enteric bacteria of animals and humans [80]. For example, *E. coli* has been employed as the model organism in determining the level of AMR in bacteria from people in close contact with food animals, such as those who work at abattoirs and veterinarians—with the observation that such people harbour significantly higher levels of resistant *E. coli* compared to the overall community [81]. The

ability to acquire and transfer AMR traits was recognised in *E. coli* as far back as the 1960s: 26% - 61% and 50% - 76% of human and pig multi-drug resistant *E. coli* strains, respectively were found to transfer resistance to lab strains of *E. coli* K-12 in conjugation experiments [82-88].

According to the World Health Organisation (WHO), the ever-increasing levels of AMR is one of the most significant threats to human health (www.who.int/entity/drugresistance/en), with substantial economic implications [89]. Resistance to antimicrobials represents a daunting challenge to treating many infections, not least those caused by *E. coli*.

As a pathogen, *E. coli* is a prominent cause of urinary tract infections (UTIs), gastroenteritis and bloodstream infections, among others; and as such, antibiotics are frequently applied to treat *E. coli* infections [90-92]. The use (and abuse) of antibiotics in treatment is linked with antibiotic resistance development [93, 94]. Microorganisms exhibit a natural ability to resist antimicrobials' action—a phenomenon referred to as intrinsic resistance [95]. In addition, antimicrobial resistance can arise due to a gene mutation or the acquisition of resistance determinants via horizontal gene transfer [96]. Horizontal gene transfer can take the form of free DNA uptake (transformation), plasmid-mediated transfer of resistance gene traits (conjugation) or via phage-mediated transfer (transduction) [97, 98].

E. coli frequently displays resistance to multiple classes of antimicrobials (MDR)—with observed rates of MDR identified among strains that cause UTI and bacteraemia exceeding 50% [91, 99, 100]. The occurrence of MDR in *E. coli* has been focused on a small number of widely dispersed clones, mediated by extended-spectrum beta-lactamase (ESBL) and carbapenemase-encoding plasmids [101]. Worryingly, the rates of *E. coli* extraintestinal

infections such as bacteraemia and UTIs have been rapidly increasing in recent years due to an increase in the number of antibiotic-resistant infections caused by "superbugs" such as the ST131 clone [99, 102, 103].

Although several strains have been linked with the global emergence and dissemination of MDR in *E. coli* (for example, strains belonging to ST88, ST410, ST648, ST405 and ST73), none equal ST131 in their extent [101]. The MDR ST131 clone belongs to a sublineage designated clade C which arose from two predominantly drug-susceptible clades by acquiring large MDR plasmids which confer ESBL and metallo-beta-lactamases encoding resistance to carbapenems and cephalosporins up to third generation; as well as point mutations leading to fluoroquinolone resistance [104, 105].

McNally *et al.* [102] recently analysed an extensive collection of more than 1,000 ST131 strains. They reported that the acquisition of colonisation and fitness factors and the accumulation of genes encoding dehydrogenase enzymes involved in anaerobic metabolism account, at least in part, for the successful expansion of this clone [102]. These results indicate hyper-resistant clones' ability to outcompete resident non-pathogenic strains of the same species and thus facilitate their long-term colonisation in the gut [101, 106]. The fact that MDR strains such as ST131 are well suited to colonise the gut competitively suggests that antimicrobial resistance traits can be readily transferred from resistant strains to co-colonising susceptible strains, with the potential to hamper the treatment of future infections caused by such strains.

Accumulating data shows that ExPEC are frequently isolated from diseased companion animals and livestock—highlighting the potential for zoonotic as well as anthroponotic

transmission [107-112]. High rates of AMR among *E. coli* isolates from livestock and poultry have been documented, linked with the agricultural use of antimicrobials [113]. The use of antimicrobials as growth promoters in poultry feed was banned in Europe in 2006 [114], with a resultant sharp drop in AMR rates among isolates from livestock and poultry [113]. However, this may be less well controlled in other parts of the world, particularly in low-to-middle income countries. Worryingly, the usage of antimicrobials in developing countries is likely to increase as intensive farming practices are adopted [115].

Several studies have investigated the link between antimicrobial usage and antimicrobial resistance in animals and AMR in humans. In 2015, a systematic review [116] of the published literature to quantify the zoonotic transfer of ESBL-encoding extraintestinal *E. coli* infections found that six studies established the zoonotic transfer of AMR by whole bacterial transmission using molecular methods, in particular, via poultry in the Netherlands. A further thirteen molecular studies suggested the mobile-genetic element-mediated transfer of AMR from resistant organisms in animals to susceptible strains in humans, while four observational epidemiological studies inferred the zoonotic transfer of AMR to humans.

Although the authors cautioned that their conclusions might not be geographically generalisable, it appears that a proportion of human extraintestinal infections caused by ESBL-encoding strains originate from food-producing animals. Further, *E. coli* strains from commercial broilers have recently been reported to share resistance profiles with strains recovered from human extraintestinal infections [117]. This is worrying and highlights the need for further investigation, especially with whole-genome sequence-based studies [93].

E. COLI GENOMICS

Frederick Blattner first conceived the *E. coli* K-12 genome sequence project in 1983 [118];

however, due to funding and technological challenges, it was not until fifteen years later that

the project (which was six-year-long) was finally completed [119]. "*E. coli* was the obvious

choice for a sequencing effort", noted Frederick Neidhardt in [120], since more was known

about the organism than any other [121] and so much information had been gleaned from

studies with *E. coli*, that, as Blattner stated, "Figuring out the microbe's genetic code would

help integrate all those years of study" [120]. Blattner and colleagues submitted the final 2.0

Mb of the 4.6 Mb *E. coli* genome to GenBank on 16 January 1997 [120]. This was closely

followed by the deposition in GenBank of an incomplete genome sequence of the closely

related strain, W3110, seven days later [120] (the complete genome sequence of which was

published in 2006 [121]). Based on atypical codon usage and base composition, Lawrence

and Ochman [122] inferred 18% of the K-12 genome to have arisen from horizontal transfer

(earlier studies had shown that horizontally-acquired genes exhibited atypical codon usage,

base composition and dinucleotide frequencies [123, 124]). This estimate was later revised to

24.5%, using improved methods [125].

Next, the genome sequences of two EHEC O157:H7 isolates were swiftly published,

expanding the number of *E. coli* genomes for comparative studies. First came the EDL933

strain isolated from Michigan ground beef connected to the 1982 multi-state outbreak by

Perna *et al.* [29], then RIMD 0508992—the strain that caused a large outbreak in 1996 in

Sakai city in Osaka prefecture, Japan, involving at least 6,000 schoolchildren [126]—the

latter sequenced by Hayashi *et al.* [127]. Comparisons with K-12 revealed a shared sequence

of 4.1 Mb, representing a common chromosomal 'backbone' of *E. coli*, with the remaining

sequence comprising strain-specific clusters varying in size, encoding putative virulence factors, prophages and prophage-like elements [29, 127]. These analyses provided evidence of extensive horizontal gene transfer, with the description of many "K" and "O" islands—depicting introgressed DNA present only in K-12 but not in O157:H7 or only in O157:H7 respectively. The existence of a shared *E. coli* backbone was confirmed by a three-way comparison of the K-12 and O157:H7 strains with the genomic sequence of the third *E. coli* strain to be completed: the ExPEC strain CFT073 [128]. A surprising finding was that all three strains shared only 39.2% of the combined non-redundant set of proteins. The role of horizontal gene transfer in the pathogenic strains' evolution was also highlighted by the presence of several pathogenicity islands exhibiting atypical codon usage interrupting the common backbone [128].

With the availability of more *E. coli* genome sequences, the core and accessory genome concept was defined to represent a conserved set of roughly 2,200 genes common to all *E. coli* strains and strain-specific sequences respectively [129-131]. The core genomic sequence lends itself as a useful tool for the phylogenomic comparison of isolates, provided the effect of homologous recombination (estimated to affect about a tenth of the *E. coli* core genome [132]) is accounted for [129].

The pan-genome concept

The concept of a "pan-genome" was proposed by Tettelin *et al.* in 2005 [133], during an attempt to utilise genome sequence information from Group B streptococcus to predict proteins that might be exposed on the organism's surface and could be exploited as vaccine candidates [133, 134]. By this concept, the pangenome of each bacterial species is defined by three distinct components: namely, its core genome, representing the genes found in each

isolate of the species; the accessory genome, depicting the genes present in several but not all isolates of the species and strain-specific genes detected in one isolate only. Through genomic comparisons of nineteen GBS isolates, Tettelin and colleagues uncovered the first evidence that closely related isolates differed significantly in their gene content. A single isolate of a particular species was insufficient to capture the species' genome [133, 134].

In a study that compared sixty-one sequenced genomes of *E. coli*, Lukjancenko *et al.* [135] predicted a pangenome comprised of 15,741 gene families, with only 993 (6%) of the gene families present in every genome (the core genome)—indicating an accessory genome of more than 90% in *E. coli*. This equates to an accessory or variable gene content of approximately four-fifths of any given *E. coli* genome [135]. These results corroborate those reported by Rasko *et al.* [130] and others before them [136, 137]. Rasko *et al.* identified a pangenome comprised of more than 13,000 genes via comparison of seventeen *E. coli* reference genome sequences encompassing human commensal and distinct clinical groups of *E. coli* and a core gene set of about 2,200 genes conserved in all isolates. It has become apparent that the more *E. coli* genomes are sequenced and compared, the more the pangenome continues to increase—what has been referred to as an 'open pangenome' [130]—and the core genome shrinks [56]. Thus, the accessory genome content contributes crucial insights which, coupled with single nucleotide polymorphisms (SNPs) in the core genome, can be employed to track the evolutionary history of natural isolates, as has been recently demonstrated by McNally *et al.* and others [138, 139]. The accessory genome content has arisen from repeated gene acquisition and the contemporaneous loss of sequences is thought to account for the distinctions between divergent lineages within the same species [137]. They include genes encoding virulence determinants, bacteriophages, virulence factors and acquired antimicrobial resistance determinants [56].

Whole genome sequencing (WGS) offers many advantages for the diagnosis and understanding of the pathobiology of *E. coli* strains, in that it is possible to predict most of the pathotypes of *E. coli* based on the presence of well-recognised virulence factors, as well as elucidating the full array of virulence factors possessed by individual strains within a particular pathotype. Furthermore, typing schemes that combine several genes within the core genome (e.g., core genome MLST available on platforms such as EnteroBase [140]) and others that would incorporate the accessory genome are expected to become the mainstay of *E. coli* analysis, particularly as the accessibility of whole-genome data continues to increase [56].

THE ECOLOGY OF *E. COLI* IN THE GUT

Microbial ecology is the study of the diversity, distribution and abundance of microorganisms and how microorganisms interact with each other and their environment to generate and preserve such diversities [141]. There are two areas of focus that have encapsulated microbial ecological studies to date: namely, the examination of microbial diversity—'who is there,' i.e., the identification and characterisation and estimation of abundance across a variety of niches; and the study of microbial activity, i.e., what microorganisms are doing and how they are doing it—including their biotic and abiotic interactions and how they impact the observed diversity and the ecosystem.

Members of the gut microbiota have co-diversified with their hosts over millions of years and exist in a mutualistic relationship with their host, in which the gut microbiota carry out vital functions for their hosts and in return occupy a nutrient-rich environment [142, 143]. The gut microbial communities' composition is thought to be influenced by factors such as diet,

physiology of the gut, host phylogeny and diet [144-147]. A healthy gut microbiota plays

critical roles in developing the host immune system and is required for homeostasis in adult

life [148, 149]. For example, the gut microbiota cells help maintain the balance between host

metabolism and the immune system and in the large intestine, metabolise the indigestible

components of the diet [19, 150]. The gut microbiota also detoxify toxic products and serve

as a barrier against the colonisation of opportunistic pathogens (termed as colonisation

resistance) [143, 151]. The mechanisms by which the resident intestinal microbiota elicit

resistance against the colonisation and invasion of pathogens include: The direct competition

for nutrients, the modification of metabolites such as bile salts and short-chain fatty acids that

render them toxic to invading pathogens, the alteration of pH and oxygen tension, the

induction of host antimicrobial peptides, the expression of a dense mucous, IgA and cellular

immunity and direct attacks through the production of bacteriocins or Type IV secretion [25,

152].

Host-directed pathways act to break down complex protein, carbohydrates and fat in ingested

food in eukaryote hosts into simple forms, thereby facilitating the absorption of nutrients. In

addition, members of the microbiota contribute to the catabolic process in producing short-

chain fatty acids utilised by the host and support the growth of the bacterial community

[153]. The members of the microbiota have evolved to maximise the available nutrients such

that under homeostatic conditions, an exogenous species needs to overcome the intense

competition for limited substrates to engraft themselves in the available niches. Resident *E.

coli* use sugars such as fucose, mannose, galactose, ribose, arabinose and N-acetyl

glucosamine in the mucous layer—which are required by pathogenic strains such as *E. coli*

O157—and in this way, mount colonisation resistance to invading pathogens [154]. Besides,

commensal *E. coli* employ the preferential utilisation of nutrients to out-compete exogenous

strains. For example, resistance to EHEC by commensal *E. coli* is due to the competition for

nutrients such as amino acids and organic acids [27, 154-156].

Competition for trace metals also appears to be significant in colonisation resistance, as

demonstrated in murine studies, where the Nissle 1917 *E. coli* strain was found to protect

against *Salmonella* infection by competing for iron [157].

However, intestinal microorganisms constitute a persistent invasion threat, considering their

vast numbers and the large intestinal surface area [149]. Emerging evidence suggests that

dysbiosis—defined as a shift in the composition of the intestinal microbiota and thus an

alteration of the relationship between the host and the gut microbiota—is linked to the

development of various diseases, such as inflammatory bowel disease, obesity, allergy,

autoimmune disease and irritable bowel syndrome [143, 158-164].

Estimates show that anaerobic bacteria outnumber *E. coli* anywhere from 100:1 up to

10,000:1 [165]. The prevalence of *E. coli* in the various hosts they colonise varies widely (0-

100%), being influenced by host characteristics such as body size, microbiota, diet and

digesta retention times [2]. Over 90% of humans carry *E. coli*, while about 25-56% of wild

mammals appear to be colonised by the organism [2, 166-169]. The prevalence rate in

human-associated animals (such as chickens and cats) is estimated to be above 60% [169].

We recently reported a prevalence of 56% from a range of non-human primate species with

varying degrees of contact with humans [170].

In the gut, *E. coli* reside in the mucous covering of the epithelial cells and is shed with the

degraded mucus components into the intestinal lumen and subsequently excreted in faeces

[171, 172]. Human faeces typically contain between 10^2-10^9 colony-forming units (cfu) of *E. coli* per gram [167, 168, 173, 174], while an estimated 10^4-10^6 cfu can be detected in the faeces of domestic animals [174]. Data on the quantity of *E. coli* in stools of wild animals is, however, lacking. As first observed by Escherich [175], *E. coli* is one of the first bacterial colonisers of the infant gut [176], achieving concentrations of up to 10^9 cfu per gram of the stools of infants [167, 168]. Subsequently, anaerobic members of the microbiota expand and dominate the gut [177]. Given that *E. coli* is a facultative anaerobe, its ability to utilise oxygen probably helps create an anaerobic environment favouring the blooming of strict anaerobes [176]. As a gut microbiota member, *E. coli* produces vitamin K and mounts resistance against colonisation by pathogens [178, 179]. Thus, *E. coli* exists in a mutualistic relationship with the human host, although it is mostly described as a commensal [12, 180].

Given that most *E. coli* reside innocuously in the gut, an important ecological question that has been plaguing microbiologists is what makes *E. coli* an occasionally devastating pathogen [169, 180]? To address such questions requires an enhanced understanding of the ecology of the organism as a commensal. However, non-pathogenic *E. coli* have been traditionally underrepresented in ecological studies of this species [180]. More studies exploring the populations of resident or non-pathogenic *E. coli* within and between hosts and how these populations vary over time are needed to shed light on this evolutionary puzzle.

Humans are exposed to *E. coli* through multiple routes [181-190]: such as the consumption of contaminated food and water; through fomites, for example, on bank coins and notes (particularly in the cracks of creased notes) and cell phones; pets and domestic animals and the environment. This high level of exposure is reflected in reports of more than one strain in normal stools [173, 191-197].

Two plausible theories explain the fate of swallowed strains [173, 196, 197]. The displacement theory suggests that newly ingested strains may fail to establish themselves, in which case they are voided out, or if they succeed in establishing themselves, will displace the 'resident' strains present. Alternatively, the dominant-minor strain theory posits that freshly ingested strains do not replace the established strains within the gut but co-exist as minority or transient strains, albeit in small numbers and may be detected from time to time in the stool along with the dominant strain.

Experimental studies with the Nissle 1917 strain have shown that not all strains are equal in their propensity to establish themselves following immigration [198]. Evidence suggests that repeated exposure may facilitate immigration and the establishment of strains [108, 199]. Host-related factors appear to contribute to the pattern of phylogroup distribution in the gut. In human studies, phylogroups E and F have been found to encourage a more heterogeneous population of *E. coli* following their establishment in the gut. Conversely, the establishment of phylogroup B2 strains in the gut appears to limit the immigration and establishment of other phylogroups [200]. Also, locale-specific differences in the distribution of *E. coli* phylotypes have been reported, linked with differences in climate and diet [201, 202]. For instance, strains belonging to phylogroups A and B1 are often prevalent in stools sourced from people living in developing countries. On the other hand, stools from people in developed countries are frequently dominated by *E. coli* belonging to phylogroups D and B2 [203, 204].

The co-existence of multiple strains in a single host raises an essential question about the factors that govern residency in the gut, i.e., how incoming strains overcome the colonisation

resistance posed by the existing *E. coli* population. The current body of evidence suggests the following.

Freter's successful competition hypothesis. Freter theorised that successful colonisation occurs due to successful competition for nutrients [205-207]. Accordingly, the gut microbiota composition is determined by several limiting substrates, which different members of the microbiota can utilise with variable efficacy. Conway and colleagues demonstrated this principle in *E. coli* strains (*E. coli* strains HS and Nissle 1917 vs *E. coli* O157:H7) using carbohydrate metabolism in a mouse model [154, 208]. This theory is in tandem with Gause's exclusion principle, which precludes two organisms' co-existence if they share the same limiting resource [209]. Iron competition appears to influence the colonisation of resident strains, as demonstrated by studies that found strains lacking siderophore genes to have a reduced ability to establish themselves in mouse models, compared to wild-type strains. Conversely, resident strains in the human gut have been found to encode siderophores [210], signalling their potential contribution to successful colonisation. The mechanisms by which two organisms exclude each other from a particular niche might be through either direct (for example, through bacteriocin production or phage to damage or kill competitors) or indirect (for example, through passive resource utilisation) competition [207, 211, 212].

Efficient utilisers. Some colonisers thrive because they utilise available nutrients much more efficiently than others who use the same nutrients. In murine studies, mice fed with *E. coli* MG1655, non-motile *flhD* mutants were found to persist in stools collected three days post-feeding and dominate the population (roughly 90%) by day 15. The mutants colonised better than the wild-type parent strain and grew in the caecal mucus faster than the wild type counterparts [213, 214]. Further analyses using high-throughput genomic approaches

revealed that the *flhD* mutants possessed an enhanced ability to oxidise several carbon sources because the loss of FlhD conferred an increased expression of genes involved in carbon and energy metabolism. (The *flhDC* operon encodes the FlhD4C2 regulatory complex, which has been shown to negatively regulate the genes involved in galactose transport and the citric acid cycle while positively regulating the genes involved in ribose transport [213, 215]).

Restaurant hypothesis. Leatham-Jensen *et al.* [216] have described how polysaccharide-degrading anaerobes break down polysaccharides into sugars, which they serve to *E. coli* cells within a shared biofilm [217]. This is an example of 'syntrophy'—an association where one organism feeds on the nutritional products of another [218]—as has been observed between *Bacteroides ovatus* and *B. thetaiotaomicron* [219]. The biofilms that feed these *E. coli* strains are referred to as "restaurants". Many such restaurants are thought to exist, proposed to comprise a mix of different commensal strains, each serving different nutrients to the commensal *E. coli* strains residing therein [208].

Different nutritional requirements. This hypothesis suggests that pathogenic strains may utilise other nutrients than those of the commensal residents. For example, it has been shown that *E. coli* strain HS*, E. coli* Nissle 1917*, E. coli* MG1655 and *E. coli* EDL933 utilise unique metabolic niches in the mouse intestine [27, 154, 208]. *E. coli* HS uses six out of the twelve sugars available in the mucous layer, namely, arabinose, galactose, gluconate, lactose, ribose and N-acetylglucosamine. *E. coli* MG1655 uses five sugars, namely, fucose, arabinose, gluconate, N-acetylneuraminate and N-acetylglucosamine. *E. coli* Nissle 1917, on the other hand, utilises the following seven sugars: fucose, galactose, arabinose, N-acetylglucosamine, gluconate, N-acetylneuraminate and mannose. Lastly, *E. coli* EDL933 also uses a unique

selection of carbon sources: galactose, arabinose, hexuronates, N-acetylglucosamine, mannose, sucrose, and ribose.

Different biogeographical niches. Pathogenic strains of *E. coli* may colonise different biogeographic regions of the gut from commensal strains. Scientists from the University of Oklahoma have shown that *E. coli* EDL933, a strain of enterohaemorrhagic *E. coli* O157:H7 colonises a different niche in the mouse intestine to human commensal strains [154, 156]. In the bovine host (where *E. coli* O157:H7 commonly colonizes the gut innocuously), studies in 12-month-old naturally colonised steer revealed a unique tropism of *E. coli* strain O157:H7 for the rectal mucosa adjoining the recto-anal junction [220].

Hierarchical nutrient utilisation. Here, closely related bacterial species may co-colonise the gut by hierarchical utilisation of similar nutrients, as has been demonstrated between two generalists, *B. ovatus* and *B. thetaiotaomicron* [219]. Thus, direct competition for substrates is excluded by each species utilising the available glycans with differing priorities (i.e., order and speed of consumption).

Exploiting colonisation resistance for therapy

Antibiotic treatment has significantly reduced mortality and morbidity associated with potentially life-threatening diseases and, thus, saved millions of lives [221]. However, broad-spectrum antibiotics rarely target pathogens alone and concomitantly result in deleterious effects on the commensal bacterial populations, resulting in increased susceptibility to infections due to alteration of the host microbiota (dysbiosis) [222-226]. This problem is compounded by the fact that many pathogens have become increasingly antimicrobial resistant. Consequently, the administration of live bacteria (probiotics) to restabilise the

altered microbiota and restore the colonisation resistance conferred by the resident microbiota has increasingly been the focus of intense research [221, 222, 224, 227]. This heightened interest has been fuelled by the success of faecal transplantation in patients suffering from diarrhoea associated with *Clostridiodes difficile* infection [228-230].

Probiotics are defined as living microbes of human origin that can colonise host sites such as the gut and oropharynx when ingested in adequate quantities and thus deliver benefits to the host [227, 231]. Probiotic strains are attractive as preventive measures against gut infections. Bacterial candidates that have been proposed to promote high-level colonisation resistance to infection include members of the lactate-producing genera, e.g., *Lactobacillus* [232] and *Bifidobacteria* [233]—although evidence of their effectiveness in health in humans, reducing infections or fostering longevity is sparse [221].

E. coli Nissle 1917 (also known as Mutaflor) is one of the most extensively researched probiotic strains globally [234]. Alfred Nissle observed that this strain ameliorates ulcerative colitis [235], following his isolation of the strain in 1917 from the stool of a soldier who did not suffer from diarrhoea like his comrades did during the First World War [236, 237]. The available data shows that Mutaflor possesses many traits that favour its suitability as a probiotic strain. For example, the strain forms a biofilm well and, in this way, outcompetes numerous pathogenic and non-pathogenic strains such as EPEC and ETEC [238]. The complete genome sequence information revealed that Mutaflor contains several fitness traits such as adhesion factors and an array of iron uptake systems that enable it to outcompete other bacteria and block the adherence and invasion of pathogenic strains [239-241]. The strain also boosts intestinal barrier function and protects against epithelial disruption by EPEC [242]. Moreover, *E. coli* Nissle 1917 induces the release of beta-defensin-2 (an

antimicrobial peptide) from human epithelial cells, thus eliciting a broad antimicrobial response against both Gram-positive and Gram-negative bacteria and fungi and viruses and, thus, inhibits invasion and colonisation of other bacteria [243].

Accumulating evidence suggests that probiotics comprised of a combination of strains are effective, owing to the symbiosis among strains—including those from different genera [244]. An example is the VSL#3 multi-strain probiotic consortium—comprising *Eubacterium faecium, Streptococcus thermophilus, Lactobacillus acidophilus*, *Bifidobacterium breve*, *Bifidobacterium infantis, Lactobacillus delbrueckii subspecies bulgaricus*, *Bifidobacterium longum*, *Lactobacillus plantarum* and *Lactobacillus casei*—which has been shown to treat ulcerative colitis effectively [244, 245]. Researchers working with *E. coli* EDL 933 found a combination of two strains—*E. coli* Nissle 1917 and *E. coli* HS, both promising probiotic candidates—can overlap nearly all the nutrient requirements of *E. coli* EDL 933 [246]. However, the same combination of commensals could not prevent colonisation by uropathogenic strain *E. coli* CFT073 and the enteropathogenic strain *E. coli* E2348/69. Different pathogenic strains thus occupy distinct nutrient niches within the gut microbiome.

As interest in probiotics heightens, there is the need to understand the mechanics of how potential probiotic strains compete or co-exist with the commensal bacteria in the gut. We do not yet fully understand how probiotic strains interact with the usual residents of the gut. Previous attempts at unravelling the interactions between pathogens and commensals in the gut have been mainly carried out using well-characterised reference strains within in-vitro environments. It will be desirable to test these hypotheses with the 'real world' strains in the natural habitat (i.e., the gut).

WITHIN HOST BACTERIAL DIVERSITY OF *E. COLI*

As previously discussed, continuous exposure to *E. coli* via multiple routes contributes to the turnover of strains in the vertebrate gut, thus resulting in considerable diversity among the population of *E. coli* that exists within a single individual and between different hosts. Besides immigration events, the diversity of *E. coli* is enriched by within-host evolution events [247].

Members of a bacterial community undergo changes as they interact with each other and their hosts or environment. The sources of novel variation that account for within-host evolution may arise from the following.

Point mutations. These involve the substitution, insertion or deletion of a single nucleotide. Point mutations represent the smallest evolution unit [248]. Reeves *et al.* reported a within-host point mutation of approximately 1 per year per genome (corresponding to 1.44×10^{-7} accumulation rate per site per year) for *E. coli* [185], while Lee *et al.* estimate a 2.54×10^{-10} mutation rate per site per generation [249].

Insertions and deletions (indels) and genomic rearrangements [466, 467]. Phylogenetic analysis of the O104:H4 strain that caused the large outbreak of gastroenteritis and haemolytic uraemic syndrome in Germany in 2011 revealed that the acquisition of a Shiga toxin 2-encoding prophage and an extended spectrum beta-lactamase CTX-M-15-encoding plasmid contributed to its emergence [468-471]. Similarly, the acquisition of virulence gene-encoding plasmids into many ancestral *Shigella* spp. was crucial to the evolution of *Shigella* as human pathogens [53].

Bacterial genomes can also undergo rearrangements during DNA recombination, replication or error-prone DNA repair [250, 251]. Genomic rearrangements alter chromosomes or large chromosomal regions. They involve the processes of deletion, duplication, insertion, inversion or translocation. Homologous recombination leads to the reassortment of genes between chromosomal pairs; however, the genome's arrangement remains unchanged. Other forms of recombination may result in rearrangements of genomic DNA. DNA rearrangements contribute to gene expression and function and may contribute to genetic diversity in *E. coli* [251, 252]. In long-term evolution experiments involving twelve *E. coli* populations propagated for more than 25 years in glucose-restricted environments, large-scale genomic rearrangements were found to occur in the same loci in at least two populations and contributed to genomic plasticity [253]. Moreover, most rearrangements were observed to confer higher fitness, highlighting the significant influence of large-scale rearrangements in genomic evolution.

Other factors that shape within-host diversity include genetic drifts: random processes by which allelic frequencies change over time due to birth and death of individuals within the population [254, 255], as well as by both purifying and diversifying selection [247].

Recombination and horizontal gene transfer (HGT). HGT involves the uptake of extracellular DNA (transformation), cell-to-cell transfer of genetic material via surface appendages (conjugation) or viral import (transduction) from genetically distant relatives. The acquisition of genetic sequences from unrelated organisms via HGT results in faster diversification of the genome compared to point mutation alone [256]—consequently termed 'evolution in quantum leaps' [257]. Fragments of the chromosomal genome can be replaced with homologous sequences from another cell through homologous recombination, which

plays a significant role in the evolution of *E. coli* [258]. In particular, the presence of mixed

infection facilitates homologous recombination by providing material for import in the

chromosome—as has been exemplified in *H. pylori*, where evolution is noted to accelerate up

to a 100-fold in the presence of mixed infections [259-261]. Thus, homologous

recombination is a strong driver of within-host evolution. Alternatively, non-homologous

sequences can be gained and incorporated into the genome—a phenomenon that is

compensated by genome degradation [262].

Hypermutators. Besides spontaneous mutations, genetic variation can arise due to the

breakdown of DNA repair mechanisms such as the mismatch repair system, producing

bacteria with decreased replication fidelity [263]. In *E. coli*, the loss of mismatch repair is

known to increase mutation rates by up to 200-fold [264]. Lee *et al.* [249] investigated the

accumulation of spontaneous mutations over thousands of generations in wild-type *E. coli*

compared to a mismatch repair-defective variant with the *mutL* gene encoding MutL, one of

the key enzymes required for mismatch repair, deleted. In their experiment, the mutant

(MutL⁻) strain showed a 233-fold increase in mutation rate to Nal^R (resistance to nalidixic

acid) compared to the wild type.

In particular, hypermutation is advantageous during infection when the ability to adapt

quickly can facilitate evasion of host immune defences and antimicrobial therapy [263].

Hypermutator strains have been identified during an MRSA outbreak in a neonatal intensive

care unit [265] and among *S. aureus* isolates from cows suffering from benign forms of

mastitis [266]. Strikingly, in one study where cystic fibrosis patients with chronic respiratory

problems were followed up over thirty-eight years, nearly half of them harboured

hypermutator *Pseudomonas aeruginosa* isolates [267]. Of note, the accumulation of SNPs via

hypermutation can hamper the analysis of transmission and result in inaccurate conclusions [263]. In particular, the use of SNP thresholds to identify and delineate transmission lines could potentially exclude hypermutator strains or the inclusion of such strains may bias results [263].

Investigating within-host bacterial diversity

To study within-host bacterial evolution in detail requires the application of whole-genome sequencing to multiple clinical samples taken from an individual host (these samples may be collected longitudinally or simultaneously from a single body site or several sites) or multiple isolates or genomes derived from a single clinical sample.

Sampling a single colony does not sufficiently capture the *E. coli* strain diversity in stools, as established by several studies [173, 191, 192, 194]. Lidin-Janson *et al.* [268] estimated that sampling five colonies enables detecting dominant genotypes at a >99% probability. At the same time, Schlager *et al.* [269] showed that sampling twenty-eight colonies provides a >90% chance of detecting minor genotypes in a stool sample.

In its simplest form, two genomes or more can be compared to each other by counting the number of positions where they differ. Within-host diversity is determined by comparing two or more genomes from the same host [270]. Comparing two genomes from different hosts lends insight into chains of transmission. These approaches have been successfully applied to several organisms beside *E. coli* [99, 100] [261, 271-278].

Genotyping multiple isolates per host can also facilitate the identification of multiply infected individuals, representing the carriage of bacterial sub-populations descended from a distinct founder strain. Multiple concurrent infections are clinically crucial for the evolution of pathogens, particularly in facilitating recombination between divergent strains as discussed above and hence, the emergence of novel genotypes [279]. Clinically, multiple infections are known to affect disease progression. For example, in HIV infection, the disease is reported to accelerate in the presence of dual infections [280]. Two plausible scenarios could explain the origin of multiple infections. In the first instance, two distinct strains could be transmitted simultaneously from an external source that harbours sufficient strain diversity. Alternatively, the two strains could have arisen from separate sources at different times.

Studies of the within-host diversity of *E. coli*

Studies of the genetic within-host diversity of *E. coli* in the human gut date back to well over a hundred years ago and have involved a wide variety of methodologies—encompassing both microbiologic and molecular techniques. The earliest studies involved the use of serotyping and subsequently multilocus enzyme electrophoresis. Later studies employed multilocus sequence typing and WGS to characterise strains and infer within-host diversity and evolution of strains. Below, we present a summary of studies investigating the within-host diversity of *E. coli* since 1890.

Pre-PCR era

The first evidence of temporal changes in the population of *E. coli* in the human gut was documented by Smith in 1899 [281]. In this pioneering study (reviewed in [282]), Smith analysed forty-eight isolates obtained from four separate cultures of the stools of a single normal infant and reported a high degree of temporal antigenic stability. Subsequently,

Totsuka [283] (reviewed in [282] and [180]) self-collected and cultured his own stool once a week for twelve weeks and recovered up to 32 isolates per culture, yielding a total of 332 isolates. Upon agglutinating these against antisera which he had prepared against isolates obtained from the first two cultures, he provided evidence of gradual shifts from one antigenic type to the other over the course of time in the intestine of one individual.

In a similar study, Wallick and Stuart [282] isolated 650 *E. coli* strains from one individual over the course of 14 months and described four distinct serotypes that persisted for a month at a time—including co-occurrence of dominant and minority strains. They also inferred household transmission based on antigenic identity (serotype).

Sears, Brownlee and Uchiyama [196, 197] investigated changes in *E. coli* serotypes over time in five adults sampled over three and thirty months. They were the first to coin the term "residents" and "transients" to describe strains of *E. coli* "which establish themselves firmly and continue to multiply over extended periods of time" and those that persisted for only a few weeks at most respectively. For the first time, they also investigated the production of bacteriocins by resident clones to ward off the invasion of other clones.

Next, Robinet [284, 285] confirmed the periodic fluctuations of antigenic types in six healthy individuals followed up for six months and concluded that host antibody production does not explain the turnover of resident strains of *E. coli* in the gut. Robinet was involved with further work investigating bacteriocin production by the resident clones by testing the activity of strains collected in one month against those recovered the previous month. These studies concluded that the turnover of strains was lower when the resident strains produced colicins.

Shooter *et al.* [194] investigated the serotype dynamics of *E. coli* isolated from nine adults over three months. They combined H- and O-antigen testing for the first time and observed that several strains with an identical O antigen, in fact, displayed different H antigens. Following on from this, Selander and colleagues [286, 287] sampled *E. coli* from a single individual for eleven months and, using multilocus enzyme electrophoresis, provided seminal insight on the genetic structure of the resident and transient clones in the healthy adult gut. Importantly, they documented the loss and acquisition of plasmids in identical clones over time and the gain and loss of novel clones over the study period as essential contributors to the generation of *E. coli* diversity in the gut.

Finally, Tzabar and colleagues [288] explored the dynamics of *E. coli* carriage in a group of individuals residing at the British Antarctic Survey research base for twenty-six weeks. Using MLEE, they provided evidence of strain-sharing among the members of the research station—although several isolates were lost to viability, hampering the study conclusions on the clonal turnover over time.

PCR era

Several studies [108, 199, 289] have explored the diversity of *E. coli* among adults, children and companion animals living in the same household over time. Key findings include the detection of clones that appeared to persist for four to forty-five weeks, as well as documenting evidence of strain sharing among members of a household and their pets. Shared strains between humans and companion animals were found to be of phylogroups D and B2, with an ST73 strain (phylogroup B2) causing UTI in a family dog.

Martinson *et al.* [290] studied eight healthy adults via biweekly sampling for six months to two years and observed that the resident clones often belonged to phylogroups A, B2 and F, including multiple resident clones persisting for more than a year in one individual.

Some studies have pointed to the possible exchange of plasmid-borne antimicrobial resistance genes between co-colonising strains within a single individual. First, Stoesser and colleagues [291] applied multiple colony sampling (sixteen colonies from each of eight faecal samples) followed by whole-genome sequencing to investigate the transmission of *E. coli* strains harbouring expanded-spectrum beta-lactamase genes collected in Cambodia. They reported substantial core- and accessory genome diversity, with a median of four STs recovered per individual. Remarkably, different clones from a single individual tended to share the same bla_{CTX-M} variant, and identical clones were found with different bla_{CTX-M} variants. Significant accessory genome diversity was also observed within and between clones—highlighting the utility of multiple colony sampling and whole-genome sampling in the analysis of *E. coli* within-host diversity.

Second, Knudsen *et al.* [292, 293] collected *E. coli* (and other enterobacteria) from the stools of children with cystic fibrosis or cancer along with matched healthy controls over a nine-month period, picking up to five colonies per faecal sample. Children with cystic fibrosis or cancer received antibiotic treatment over the course of the study. Of the isolated *E. coli* (90% of children with cystic fibrosis, 93% of children with cancer and 94% of healthy controls), the prevalence of antibiotic-resistant enterobacteria did not significantly differ between the children who received antibiotics and the healthy controls at the start and at the end of the study, suggesting that the level of antibiotic resistance they observed arose from the community. As a follow-up study, seven isolates collected from three consecutive samples

collected from one child with cystic fibrosis was characterised by WGS. Here, three distinct strains were detected, and the different strains were found to harbour Inc1 plasmids encoding $bla_{CTX-M-1}$. However, the plasmids from the three different strains were found to differ by only a few SNPs and varied with limited regions, suggesting recombination events. This study documented the horizontal transfer of $bla_{CTX-M-1}$-harbouring plasmids within a single individual.

Stegger *et al.* [294] analysed twenty *E. coli* isolates from each of nine urine samples collected from nine women who presented with UTI at a general practice in Zealand, Denmark, from which a single clone was detected in eight out of the nine samples. The authors then selected a total of forty isolates belonging to the same clone: ten each from two urine samples and two rectal swabs isolates from two healthy individuals (collected in a previous study [295]) and investigated the intra-clonal diversity of *E. coli* among the commensal and uropathogenic strains. A low intra-clonal diversity was observed for each clone in both the commensal and pathogenic strains (0-2 non-synonymous SNPs). The authors reached an interesting conclusion, stating that "sampling of one colony would be enough for surveillance, outbreak investigations and clonal evolution", although there is overwhelming evidence to support the opposite. Evidence from their own study showed that among other clones, a variation in gene content of 2-15 genes was detected for all clones—which would have been unnoticeable had they isolated only a single colony from each sample.

We recently surveyed *E. coli* in healthy children (less than five years) from a rural Gambian community and assessed the relative contribution of independent immigration and establishment of strains versus within-host evolution to the generation of within-host diversity in the healthy gut [296]. Our results showed a dominance of independent

immigration and establishment of strains, with several strains harbouring putative virulence factors commonly associated with Enteropathogenic and Enteroaggregative *E. coli*—thus adding to the body of evidence indicating a blurring of the lines of demarcation between resident strains and their pathogenic variants, while also adding significantly to the number of whole-genome sequences from commensal *E. coli* isolates, particularly from sub-Saharan Africa.

Prospective studies of travellers to low-income countries have provided vital insight into the acquisition and loss of strains, particularly, MDR clones. In a recent report [296], Kantele *et al.* investigated the dynamics of intestinal carriage of MDR bacteria during travel to a low-income country via daily stool sampling for the duration of their visit—unlike most previous studies, which only involved before and after samples. The study involved participants attending a medical course in Vientiane, Laos, together but staying at three different hotels, visiting several restaurants together or in small groups, and participating in daily rounds at the local hospitals. Remarkably, all the study participants acquired ESBL-producing bacteria at least at one point during the study—dominated by *E. coli*, with 95% of the participants acquiring between two to seven strains and multiple transient colonisation events, with 20% of the participants sharing some common strains over the study period.

In contrast to the abundance of studies in humans, few studies have investigated the dynamics of *E. coli* residency in non-human mammals. In a temporal survey of *E. coli* residency in fifty-four mountain brushtail possums sampled on four occasions over the course of a year, Blyton *et al.* reported phylogroup B2 to dominate the resident *E. coli* population [297]. This study concluded that *E. coli* was rapidly gained and lost among these non-human mammals, with just 36% of resident strains recovered at one timepoint recovered at the subsequent

timepoint. Recent data from our lab involving a range of diurnal monkey species in the Gambia also showed a dominance of *E. coli* belonging to phylogroup B2 [170], including a prevalence of well-characterised lineages associated with human extra-intestinal infections or antimicrobial resistance. However, the strains that occurred in these monkeys were genetically divergent from those sourced from humans, with little antimicrobial resistance among the simian isolates. Several monkeys were colonised by two or more strains, with the evidence of strain-sharing among monkeys sampled from the same park.

Data from poultry further support the observation that co-colonisation with multiple strains is prevalent in non-human vertebrate hosts, reflecting the findings of previous studies in humans. Li *et al.* [298] sampled >100 colonies per caecum sample from nine mcr-1-positive broiler chickens from three provinces in China (a total of 962 *E. coli* isolates). A high rate of co-colonisation (three to nine STs per chicken) was observed, with several birds harbouring one to five Inc type plasmids encoding mrc-1—this high level of heterogeneity facilitating the transmission of mcr-1 among these chickens. The authors concluded that the "gut is a 'melting pot' for active horizontal transfer of the mcr-1 gene". We recently investigated the within-host diversity of *E. coli* from backyard poultry (ten backyard chickens and nine guinea fowl) drawn from rural Gambia, and found similarly high colonisation with multiple strains in non-commercial birds [299]. Worryingly, a dominance of ST155—a well characterised multi-drug resistant clone—was observed in poultry *E. coli* isolates from Africa and South America. However, the isolates from these backyard birds carried few ESBL genotypes compared to what was found among other isolates from the African continent [299].

In summary, the body of evidence shows that colonisation with multiple strains is common in the vertebrate gut, with the rapid gain and loss of strains contributing to the within-host

diversity of *E. coli*. At the same time, some clones can establish themselves in the gut for extended periods. As evidenced by the existing data, this level of diversity positions the gut as a hot spot for the active horizontal transfer of AMR and virulence traits within and between strains. Other vertebrate hosts may carry lineages well-established in humans-- indicating bacteria that have probably existed in the specific host species for long periods and passed down through vertical transmission. The direct contribution of non-human vertebrate hosts to the spread of AMR and virulence traits in humans remains to be established, despite the prevalence of isolates with similar profiles as their close relatives circulating in humans. Significant gaps remain in our understanding of the ecology of this familiar ally and foe of the vertebrate host, key examples of which are outlined below.

CONCLUSIONS AND PERSPECTIVES

It has been said: "All cell biologists have at least two cells of interest: the one they are studying and *E. coli"* [300]. Yet, in the 136 years since *E. coli* was first described, most studies on the ecology of *E. coli* have been biased to pathogenic strains. Only a handful of studies have focused on the diversity of *E. coli* populations in the healthy human gut, particularly in the post-PCR era. Even fewer studies have explored the within-host diversity of non-pathogenic *E. coli* in non-human vertebrates in the post-PCR period, leaving much to be learned in terms of the resident populations of *E. coli* in the healthy vertebrate gut. However, such studies are vital to our understanding of the evolution of *E. coli* in the health and disease of both humans and animals.

An exploration of the commensal population of *E. coli* in non-human vertebrate hosts such as poultry and non-human primates is timely and relevant for several reasons. First, emerging infections are often linked with pathogens (such as *E. coli*) that inhabit both humans and non-

human vertebrates and can cross the species barrier, potentially causing zoonotic as well as anthroponotic infections. Understanding the diversity within the commensal population would provide vital insights into the evolution of pathogens and antimicrobial resistance within this ecological context. Secondly, understanding the within-host evolution of *E. coli* during health may also inform new therapeutic approaches that exploit our understanding of the gut microbiome, such as faecal microbiome transplantation, probiotic development and the control of foodborne diseases.

In particular, *E. coli* diversity studies based on whole-genome sequence data from sub-Saharan Africa are scarce. Very few studies have examined the population of *E. coli* among healthy individuals in this setting, particularly in healthy children. However, given the increased exposure to the environment, unsanitary conditions and proximity to animals especially in rural areas, such studies are likely to yield crucial insights into the dynamics of strain turnover and residency of *E. coli* in the gut, ultimately facilitating our understanding of infection by this organism.

Furthermore, the bulk of existing surveys of *E. coli* in poultry have been biased towards the surveillance of antimicrobial resistance in commercial birds; thus, we know very little of the dynamics of colonisation and the diversity of *E. coli* in this niche. Also, backyard birds, which are kept in proximity to humans, have been largely neglected in ecological studies. However, given the frequent contact with humans, studies involving backyard poultry might yield critical insights into the potential for the dissemination of strains and antimicrobial resistance between the two host species.

The direct contribution of human-habituated animals to *E. coli* transmission remains unknown. To address this would require future studies incorporating samples from whole households and animals living in proximity. The inclusion of non-human primate species in future studies of resident *E. coli* diversity would elucidate the full scale of the impact of human encroachment on natural habitats. The recent advances in Illumina sequencing platforms twinned with the potential of long-read sequencing (such as Pacific Biosciences sequencing and Oxford Nanopore technology) provide a timely opportunity to apply these approaches to these lines of study.

Outstanding questions that need addressing in future studies include: What factors govern co-existence or competition between the different *E. coli* strains co-colonising the vertebrate gut? Do they exploit different micro-geographical or nutrient niches in the gut? What is the importance, if any, of flagellar motility? How far do within-species antibacterial factors (e.g., colicins and type VI secretions systems) play a role in competitive exclusion? Can differences in the distribution of *E. coli* strains *in vivo* be accounted for by behaviours *in vitro* (e.g., competitive growth under different nutrient conditions)?

Biolog assays (Biolog Inc., Hayward, California, USA) are one way to screen for the key metabolic differences between different strains that may account for their co-existence or competitive exclusion within the gut. Candidate strains could then be selected for onward in-vitro competition assays via individual and competitive fitness experiments and indirect antagonism assays as have been used by Durso *et al.* to assess fitness differences and competition between commensal strains and O157:H7 strains [301]. Applikon Biotechnology (https://www.applikon-biotechnology.com/en/products/cultivation-systems/micro-matrix/) have designed mini-bioreactors that can be used for multiple competition assays in parallel,

using 24-well microtiter plates with built-in controls for pH, temperature and dissolved oxygen. Such experiments may shed light on the physiological and biochemical properties underlying the co-colonisation of different strains in the vertebrate gut.

REFERENCES

1.   **Friedmann HC**. From "butyribacterium" to "*E. coli*": an essay on unity in biochemistry. *Perspect Biol Med* 2004;47(1):47-66.

2.   **Tenaillon O, Skurnik D, Picard B *et al.*** The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* 2010;8(3):207-217.

3.   **Friedmann H**. Escherich and *Escherichia*. *EcoSal Plus* 2014. Reprinted from *Adv Appl Microbiol* 2014;60:133-196.

4.   **Escherich T**. The intestinal bacteria of the neonate and breast-fed infant. 1884. *Rev Infect Dis* 1988;10(6):1220-1225.

5.   **Shulman ST, Friedmann HC, Sims RH**. Theodor Escherich: the first pediatric infectious diseases physician? *Clin Infect Dis* 2007;45(8):1025-1029.

6.   **Escherich T, Pfaundler M**. *Bacterium coli commune*. In: **Kolle W, Wassermann A** (eds). Handbuch der pathogenen mikroorganismen, vol 2. Verlag von Gustav Fischer, Jena 1903; pp. 334–474.

7.   **Bettelheim KA**. Commemoration of the publication 100 years ago of the papers by Dr. Th. Escherich in which are described for the first time the organisms that bear his name. *Zentralbl Bakteriol Mikrobiol Hyg A* 1986;261(3):255-265.

8.   **Escherich T**. Ueber die bacterien des milchkotes. *Æerztliches Intelligenz-Blatt, Münchner Medicinische Wochenschrift* 1888;32:243.

9.   **Castellani A, Chalmers A**. **J**. Manual of tropical medicine (3rd ed.). London: Baillière, Tindall and Cox; 1919; p. 941.

10.   **Judicial Commission of the International Committee on Bacterial Nomenclature**. Conservation of the family name *Enterobacteriaceae,* of the name of the type genus, and designation of the type species. Opinion No. 15. *Int Bull Bacteriol Nomencl Taxon* 1958;8:73–74.

11. **Skerman V, McGowan, V., Sneath, P**. Approved lists of bacterial names. *Int J Syst Bacteriol* 1980;30:225–420.

12. **Blount ZD**. The unexhausted potential of *E. coli. Elife* 2015;4.

13. **van Elsas JD, Semenov AV, Costa R *et al.*** Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *ISME J* 2011;5(2):173-183.

14. **Alm EW, Walk, ST, Gordon, DM**. The niche of *Escherichia coli.* In: **Walk ST and Feng, PCH** (eds). Population Genetics of Bacteria: A tribute to Thomas S. Whittam. ASM Press; Washington, DC. 2011; Chapter 6 pp. 67-89.

15. **Macfarlane GT, Macfarlane S**. Human colonic microbiota: ecology, physiology and metabolic potential of intestinal bacteria. *Scand J Gastroenterol Suppl* 1997;222:3-9.

16. **Leimbach A, Hacker J, Dobrindt U**. *E. coli* as an all-rounder: The thin line between commensalism and pathogenicity. In: **Dobrindt U, Hacker JH, Svanborg C** (eds). *Between Pathogenicity and Commensalism*. Springer Berlin Heidelberg; 2013. pp. 3-32.

17. **Kosek M, Bern C, Guerrant RL**. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *Bull World Health Organ* 2003;81(3):197-204.

18. **Orskov F, Orskov I**. *Escherichia coli* serotyping and disease in man and animals. *Can J Microbiol* 1992;38(7):699-704.

19. **Bäckhed F, Ley RE, Sonnenburg JL *et al.*** Host-bacterial mutualism in the human intestine. *Science* 2005;307(5717):1915-1920.

20. **Claesson MJ, O'Sullivan O, Wang Q *et al.*** Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLos One* 2009;4(8):e6669.

21. **Kaper JB, Nataro JP, Mobley HL**. Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2004;2(2):123-140.

22. **Croxen MA, Finlay BB**. Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol* 2010;8(1):26-38.

23. **Denamur E, Clermont O, Bonacorsi S *et al.*** The population genetics of pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2020.

24. **Dutta S, Pazhani GP, Nataro JP, Ramamurthy T**. Heterogenic virulence in a diarrheagenic *Escherichia coli*: evidence for an EPEC expressing heat-labile toxin of ETEC. *Int J Med Microbiol* 2015;305(1):47-54.

25. **Kamada N, Chen GY, Inohara N, Núñez G**. Control of pathogens and pathobionts by the gut microbiota. *Nat Immunol* 2013;14(7):685-690.

26. **Pitout JD**. Extraintestinal pathogenic *Escherichia coli*: A combination of virulence with antibiotic resistance. *Front Microbiol* 2012;3:9.

27. **Fabich AJ, Jones SA, Chowdhury FZ, Cernosek A, Anderson A et al.** Comparison of carbon nutrition for pathogenic and commensal *Escherichia coli* strains in the mouse intestine. *Infect Immun* 2008;76(3):1143-1152.

28. **Le Bouguénec C, Schouler C**. Sugar metabolism, an additional virulence factor in enterobacteria. *Int J Med Microbiol* 2011;301(1):1-6.

29. **Perna NT, Plunkett G, Burland V, Mau B, Glasner JD et al.** Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 2001;409(6819):529-533.

30. **Bertin Y, Girardeau JP, Chaucheyras-Durand F, Lyan B, Pujos-Guillot E et al.** Enterohaemorrhagic *Escherichia coli* gains a competitive advantage by using ethanolamine as a nitrogen source in the bovine intestinal content. *Environ Microbiol* 2011;13(2):365-377.

31. **Escobar-Páramo P, Giudicelli C, Parsot C, Denamur E**. The evolutionary history of *Shigella* and enteroinvasive *Escherichia coli* revised. *J Mol Evol* 2003;57(2):140-148.

32. **Pupo GM, Lan R, Reeves PR**. Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc Natl Acad Sci U S A* 2000;97(19):10567-10572.

33. **Lan R, Alles MC, Donohoe K, Martinez MB, Reeves PR**. Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. *Infect Immun* 2004;72(9):5080-5088.

34. **The HC, Thanh DP, Holt KE, Thomson NR, Baker S**. The genomic signatures of *Shigella* evolution, adaptation and geographical spread. *Nat Rev Microbiol* 2016;14(4):235-250.

35. **Chaudhuri RR, Henderson IR**. The evolution of the *Escherichia coli* phylogeny. *Infect Genet Evol* 2012;12(2):214-226.

36. **Frank C, Werber D, Cramer JP, Askar M, Faber M et al.** Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *N Engl J Med* 2011;365(19):1771-1780.

37. **Karch H, Denamur E, Dobrindt U, Finlay BB, Hengge R et al.** The enemy within us: lessons from the 2011 European *Escherichia coli* O104:H4 outbreak. *EMBO Mol Med* 2012;4(9):841-848.

38. **Soysal N, Mariani-Kurkdjian P, Smail Y, Liguori S, Gouali M et al.** Enterohemorrhagic *Escherichia coli* hybrid pathotype O80:H2 as a new therapeutic challenge. *Emerg Infect Dis* 2016;22(9):1604-1612.

39. **Mariani-Kurkdjian P, Lemaître C, Bidet P, Perez D, Boggini L et al.** Haemolytic-uraemic syndrome with bacteraemia caused by a new hybrid *Escherichia coli* pathotype. *New Microbes New Infect* 2014;2(4):127-131.

40. **Thiry D, Saulmont M, Takaki S, De Rauw K, Duprez JN et al.** Enteropathogenic *Escherichia coli* O80:H2 in young calves with diarrhea, Belgium. *Emerg Infect Dis* 2017;23(12):2093-2095.

41. **De Rauw K, Thiry D, Caljon B, Saulmont M, Mainil J et al.** Characteristics of Shiga toxin producing- and enteropathogenic *Escherichia coli* of the emerging serotype O80:H2 isolated from humans and diarrhoeic calves in Belgium. *Clin Microbiol Infect* 2019;25(1):111.e115-111.e118.

42. **Nüesch-Inderbinen M, Cernela N, Wüthrich D, Egli A, Stephan R**. Genetic characterization of Shiga toxin producing *Escherichia coli* belonging to the emerging hybrid pathotype O80:H2 isolated from humans 2010-2017 in Switzerland. *Int J Med Microbiol* 2018;308(5):534-538.

43. **Cointe A, Birgy A, Bridier-Nahmias A, Mariani-Kurkdjian P, Walewski V et al.** *Escherichia coli* O80 hybrid pathotype strains producing Shiga toxin and ESBL: molecular characterization and potential therapeutic options. *J Antimicrob Chemother* 2020;75(3):537-542.

44. **Nicholson BA, West AC, Mangiamele P, Barbieri N, Wannemuehler Y et al.** Genetic characterization of ExPEC-like virulence plasmids among a subset of NMEC. *PLOS ONE* 2016;11(1):e0147757.

45. **Peigne C, Bidet P, Mahjoub-Messai F, Plainvert C, Barbe V et al.** The plasmid of *Escherichia coli* strain S88 (O45:K1:H7) that causes neonatal meningitis is closely related to avian pathogenic *E. coli* plasmids and is associated with high-level bacteremia in a neonatal rat meningitis model. *Infect Immun* 2009;77(6):2272-2284.

46.   **Walk ST, Alm EW, Gordon DM, Ram JL, Toranzos GA et al.** Cryptic lineages of the genus *Escherichia*. *Applied and environmental microbiology* 2009;75(20):6534-6544.

47.   **Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O**. ClermonTyping: an easy-to-use and accurate in silico method for Escherichia genus strain phylotyping. *Microbial genomics* 2018;4(7):e000192.

48.   **Lu S, Jin D, Wu S, Yang J, Lan R et al.** Insights into the evolution of pathogenicity of *Escherichia coli* from genomic analysis of intestinal *E. coli* of *Marmota himalayana* in Qinghai-Tibet plateau of China. *Emerg Microbes Infect* 2016;5(12):e122.

49.   **Escobar-Páramo P, Clermont O, Blanc-Potard AB, Bui H, Le Bouguénec C et al.** A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli. Mol Biol Evol* 2004;21(6):1085-1094.

50.   **Picard B, Garcia JS, Gouriou S, Duriez P, Brahimi N et al.** The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect Immun* 1999;67(2):546-553.

51.   **Díaz-Jiménez D, García-Meniño I, Herrera A, García V, López-Beceiro AM et al.** Genomic characterization of *Escherichia coli* isolates belonging to a new hybrid aEPEC/ExPEC pathotype O153:H10-A-ST10 *eae*-beta1 occurred in meat, poultry, wildlife and human diarrheagenic samples. *Antibiotics (Basel)* 2020;9(4).

52.   **Gioia-Di Chiacchio RM, Cunha MPV, de Sá LRM *et al.*** Novel hybrid of typical enteropathogenic *Escherichia coli* and Shiga-toxin-producing *E. coli* (tEPEC/STEC) emerging from pet birds. *Front Microbiol* 2018;9:2975.

53. **Gati NS, Middendorf-Bauchart B, Bletz S *et al.*** Origin and evolution of hybrid Shiga toxin-producing and uropathogenic *Escherichia coli* strains of sequence type 141. *J Clin Microbiol* 2019;58(1).

54. **Bielaszewska M, Schiller R, Lammers L *et al.*** Heteropathogenic virulence and phylogeny reveal phased pathogenic metamorphosis in *Escherichia coli* O2:H6. *EMBO Mol Med* 2014;6(3):347-357.

55. **Kessler R, Nisa S, Hazen TH *et al.*** Diarrhea, bacteremia and multiorgan dysfunction due to an extraintestinal pathogenic *Escherichia coli* strain with enteropathogenic *E. coli* genes. *Pathog Dis* 2015;73(8):ftv076.

56. **Robins-Browne RM, Holt KE, Ingle DJ *et al.*** Are *Escherichia coli* pathotypes still relevant in the era of whole-genome sequencing? *Frontiers in Cellular and Infection Microbiology* 2016;6:141.

57. **Burgess NR, McDermott SN, Whiting J**. Aerobic bacteria occurring in the hind-gut of the cockroach, *Blatta orientalis*. *J Hyg (Lond)* 1973;71(1):1-7.

58. **Brenner DJ, McWhorter AC, Knutson JK *et al.*** *Escherichia vulneris*: a new species of *Enterobacteriaceae* associated with human wounds. *J Clin Microbiol* 1982;15(6):1133-1140.

59. **Brenner DJ, Davis BR, Steigerwalt AG *et al.*** Atypical biogroups of *Escherichia coli* found in clinical specimens and description of *Escherichia hermannii* sp. nov. *J Clin Microbiol* 1982;15(4):703-713.

60. **Farmer JJ, Fanning GR, Davis BR *et al.*** *Escherichia fergusonii* and *Enterobacter taylorae*, two new species of *Enterobacteriaceae* isolated from clinical specimens. *J Clin Microbiol* 1985;21(1):77-81.

61. **Lawrence JG, Ochman H, Hartl DL**. Molecular and evolutionary relationships among enteric bacteria. *J Gen Microbiol* 1991;137(8):1911-1921.

62. **Albert MJ, Alam K, Islam M et al.** *Hafnia alvei*, a probable cause of diarrhea in humans. *Infect Immun* 1991;59(4):1507-1513.

63. **Huys G, Cnockaert M, Janda JM et al.** *Escherichia albertii* sp. nov., a diarrhoeagenic species isolated from stool specimens of Bangladeshi children. *Int J Syst Evol Microbiol* 2003;53(Pt 3):807-810.

64. **Hyma KE, Lacher DW, Nelson AM et al.** Evolutionary genetics of a new pathogenic *Escherichia* species: *Escherichia albertii* and related *Shigella boydii* strains. *J Bacteriol* 2005;187(2):619-628.

65. **Priest FG, Barker M**. Gram-negative bacteria associated with brewery yeasts: reclassification of *Obesumbacterium proteus* biogroup 2 as *Shimwellia pseudoproteus* gen. nov., sp. nov., and transfer of *Escherichia blattae* to *Shimwellia blattae* comb. nov. *Int J Syst Evol Microbiol* 2010;60(Pt 4):828-833.

66. **Hata H, Natori T, Mizuno T et al.** Phylogenetics of family *Enterobacteriacea*e and proposal to reclassify *Escherichia hermannii* and *Salmonella subterranea* as *Atlantibacter hermannii* and *Atlantibacter subterranea* gen. nov., comb. nov. *Microbiol Immunol* 2016;60(5):303-311.

67. **Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S**. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9(1):5114.

68. **Clermont O, Gordon DM, Brisse S, Walk ST, Denamur E**. Characterization of the cryptic *Escherichia* lineages: rapid identification and prevalence. *Environ Microbiol* 2011;13(9):2468-2477.

69. **Liu S, Jin D, Lan R, Wang Y, Meng Q et al.** *Escherichia marmotae* sp. nov., isolated from faeces of *Marmota himalayana*. *Int J Syst Evol Microbiol* 2015;65(7):2130-2134.

70. **Liu S, Feng J, Pu J, Xu X, Lu S et al.** Genomic and molecular characterisation of *Escherichia marmotae* from wild rodents in Qinghai-Tibet plateau as a potential pathogen. *Sci Rep* 2019;9(1):10619.

71. **Ocejo M, Tello M, Oporto B, Lavín JL, Hurtado A**. Draft genome sequence of *Escherichia marmotae* E690, Isolated from beef cattle. *Microbiol Resour Announc* 2020;9(32).

72. **Gilroy R, Ravi A, Getino M, Pursley I, Horton DL et al.** Extensive microbial diversity within the chicken gut microbiome revealed by metagenomics and culture. *PeerJ* 2021;9:e10941.

73. **Richter M, Rosselló-Móra R**. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 2009;106(45):19126-19131.

74. **Steinsland H, Lacher DW, Sommerfelt H, Whittam TS**. Ancestral lineages of human enterotoxigenic *Escherichia coli*. *J Clin Microbiol* 2010;48(8):2916-2924.

75. **Ingle DJ, Clermont O, Skurnik D, Denamur E, Walk ST et al.** Biofilm formation by and thermal niche and virulence characteristics of *Escherichia* spp. *Appl Environ Microbiol* 2011;77(8):2695-2700.

76. **Clermont O, Christenson JK, Denamur E, Gordon DM**. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 2013;5(1):58-65.

77. **van der Putten BCL, Matamoros S, Mende DR, Scholl ER, Consortium C et al.** sp. nov., a novel Gram-stain-negative bacterium, isolated from a faecal sample of an international traveller. *Int J Syst Evol Microbiol* 2021;71(2).

78. **Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH**. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 2019.

79. **Pallen MJ, Telatin A, Oren A**. The Next million names for Archaea and Bacteria. *Trends Microbiol* 2020.

80. **Teuber M**. Spread of antibiotic resistance with food-borne pathogens. *Cell Mol Life Sci* 1999;56(9-10):755-763.

81. **Bongers JH, Franssen F, Elbers AR, Tielen MJ**. Antimicrobial resistance of *Escherichia coli* isolates from the faecal flora of veterinarians with different professional specialties. *Vet Q* 1995;17(4):146-149.

82. **Datta N**. Drug resistance and R factors in the bowel bacteria of London patients before and after admission to hospital. *Br Med J* 1969;2(5654):407-411.

83. **Guinée P, Ugueto N, van Leeuwen N**. Escherichia coli with resistance factors in vegetarians, babies, and nonvegetarians. *Appl Microbiol* 1970;20(4):531-535.

84. **Linton KB, Lee PA, Richmond MH, Gillespie WA, Rowland AJ et al.** Antibiotic resistance and transmissible R-factors in the intestinal coliform flora of healthy adults and children in an urban and a rural community. *J Hyg (Lond)* 1972;70(1):99-104.

85. **Gyles CL, Palchaudhuri S, Maas WK**. Naturally occurring plasmid carrying genes for enterotoxin production and drug resistance. *Science* 1977;198(4313):198-199.

86. **Davies M, Stewart PR**. Transferable drug resistance in man and animals: genetic relationship between R-plasmids in enteric bacteria from man and domestic pets. *Aust Vet J* 1978;54(11):507-512.

87. **Bourque R, Lallier R, Larivière S**. Influence of oral antibiotics on resistance and enterotoxigenicity of *Escherichia coli*. *Can J Comp Med* 1980;44(1):101-108.

88. **Nijsten R, London N, van den Bogaard A, Stobberingh E**. In-vitro transfer of antibiotic resistance between faecal *Escherichia coli* strains isolated from pig farmers and pigs. *J Antimicrob Chemother* 1996;37(6):1141-1154.

89. **Brown L, Langelier C, Reid MJ, Rutishauser RL, Strnad L**. Antimicrobial

resistance: A call to action! *Clin Infect Dis* 2017;64(1):106-107.

90. **Riley LW**. Distinguishing pathovars from nonpathovars: *Escherichia coli*. *Microbiol*

Spectr 2020;8(4).

91. **de Kraker ME, Jarlier V, Monen JC, Heuer OE, van de Sande N et al.** The

changing epidemiology of bacteraemias in Europe: trends from the European

Antimicrobial Resistance Surveillance System. *Clin Microbiol Infect* 2013;19(9):860-

868.

92. **Foxman B**. The epidemiology of urinary tract infection. *Nat Rev Urol*

2010;7(12):653-660.

93. **Webber MA, Piddock, L. J. V.** Antibiotic Resistance in *Escherichia coli*. *Frontiers*

in Antimicrobial Resistance: a Tribute to Stuart B Levy. Washington, D. C.: ASM

Press; 2005. pp. 374-386.

94. **Ventola CL**. The antibiotic resistance crisis: part 1: causes and threats. *P T*

2015;40(4):277-283.

95. **Cheng G, Dai M, Ahmed S, Hao H, Wang X et al.** Antimicrobial drugs in fghting

against antimicrobial resistance. *Front Microbiol* 2016;7:470.

96. **de la Cruz F, Davies J**. Horizontal gene transfer and the origin of species: lessons

from bacteria. *Trends Microbiol* 2000;8(3):128-133.

97. **Lee JW**. Protocol measuring horizontal gene transfer from algae to non-

photosynthetic organisms. *MethodsX* 2019;6:1564-1574.

98. **Burmeister AR**. Horizontal Gene Transfer. *Evol Med Public Health*

2015;2015(1):193-194.

99. **Croxall G, Hale J, Weston V, Manning G, Cheetham P et al.** Molecular

epidemiology of extraintestinal pathogenic *Escherichia coli* isolates from a regional

cohort of elderly patients highlights the prevalence of ST131 strains with increased antimicrobial resistance in both community and hospital care settings. *J Antimicrob Chemother* 2011;66(11):2501-2508.

100. **Alhashash F, Weston V, Diggle M, McNally A**. Multidrug-resistant *Escherichia coli* bacteremia. *Emerg Infect Dis* 2013;19(10):1699-1701.

101. **Dunn SJ, Connor C, McNally A**. The evolution and transmission of multi-drug resistant *Escherichia coli* and *Klebsiella pneumoniae:* the complexity of clones and plasmids. *Curr Opin Microbiol* 2019;51:51-56.

102. **McNally A, Kallonen T, Connor C, Abudahab K, Aanensen DM et al.** Diversification of colonization factors in a multidrug-resistant *Escherichia coli* lineage evolving under negative frequency-dependent selection. *mBio* 2019;10(2).

103. **Banerjee R, Johnson JR**. A new clone sweeps clean: the enigmatic emergence of *Escherichia coli* sequence type 131. *Antimicrob Agents Chemother* 2014;58(9):4997-5004.

104. **Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE et al.** Evolutionary history of the global emergence of the *Escherichia coli* epidemic clone ST131. *mBio* 2016;7(2):e02162.

105. **Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW et al.** Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of *Escherichia coli* ST131. *mBio* 2016;7(2):e00347-00316.

106. **Benz F, Huisman JS, Bakkeren E, Herter JA, Stadler T et al.** Plasmid- and strain-specific factors drive variation in ESBL-plasmid spread in vitro and in vivo. *ISME J* 2021;15(3):862-878.

107. **Zogg AL, Zurfluh K, Schmitt S, Nuesch-Inderbinen M, Stephan R**. Antimicrobial resistance, multilocus sequence types and virulence profiles of ESBL producing and

non-ESBL producing uropathogenic *Escherichia coli* isolated from cats and dogs in Switzerland. *Veterinary microbiology* 2018;216:79-84.

108. **Johnson JR, Clabots C, Kuskowski MA**. Multiple-host sharing, long-term persistence, and virulence of *Escherichia coli* clones from human and animal household members. *J Clin Microbiol* 2008;46(12):4078-4082.

109. **Johnson JR, Owens K, Gajewski A, Clabots C**. *Escherichia coli* colonization patterns among human household members and pets, with attention to acute urinary tract infection. *J Infect Dis* 2008;197(2):218-224.

110. **Johnson JR, Miller S, Johnston B, Clabots C, Debroy C**. Sharing of *Escherichia coli* sequence type ST131 and other multidrug-resistant and Urovirulent *E. coli* strains among dogs and cats within a household. *J Clin Microbiol* 2009;47(11):3721-3725.

111. **Achtman M, Heuzenroeder M, Kusecek B, Ochman H, Caugant D et al.** Clonal analysis of *Escherichia coli* O2:K1 isolated from diseased humans and animals. *Infection and immunity* 1986;51(1):268-276.

112. **Ewers C, Grobbel M, Stamm I, Kopp PA, Diehl I et al.** Emergence of human pandemic O25:H4-ST131 CTX-M-15 extended-spectrum-β-lactamase-producing *Escherichia coli* among companion animals. *Journal of Antimicrobial Chemotherapy* 2010;65(4):651-660.

113. **Bennani H, Mateus A, Mays N, Eastmure E, Stärk KDC et al.** Overview of evidence of antimicrobial use and antimicrobial resistance in the food chain. *Antibiotics (Basel)* 2020;9(2).

114. **Castanon JIR**. History of the use of antibiotic as growth promoters in European poultry feeds. *Poultry science* 2007;86(11):2466-2471.

115. **Van Boeckel TP, Brower C, Gilbert M, Grenfell BT, Levin SA et al.** Global trends in antimicrobial use in food animals. *Proceedings of the National Academy of Sciences* 2015;112(18):5649.

116. **Lazarus B, Paterson DL, Mollinger JL, Rogers BA**. Do human extraintestinal *Escherichia coli* infections resistant to expanded-spectrum cephalosporins originate from food-producing animals? A systematic review. *Clin Infect Dis* 2015;60(3):439-452.

117. **Borges CA, Tarlton NJ, Riley LW**. *Escherichia coli* from commercial broiler and backyard chickens share sequence types, antimicrobial resistance profiles, and resistance genes with human extraintestinal pathogenic *Escherichia coli*. *Foodborne Pathog Dis* 2019;16(12):813-822.

118. **Blattner FR**. Biological frontiers. *Science* 1983;222(4625):719-720.

119. **Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V et al.** The complete genome sequence of *Escherichia coli* K-12. *Science* 1997;277(5331):1453-1462.

120. **Pennisi E**. Laboratory workhorse decoded. *Science* 1997;277(5331):1432-1434.

121. **Hayashi K, Morooka N, Yamamoto Y, Fujita K, Isono K et al.** Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol Syst Biol* 2006;2:2006.0007.

122. **Lawrence JG, Ochman H**. Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci U S A* 1998;95(16):9413-9417.

123. **Muto A, Osawa S**. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl Acad Sci U S A* 1987;84(1):166-169.

124. **Lawrence JG, Ochman H**. Amelioration of bacterial genomes: rates of change and exchange. *J Mol Evol* 1997;44(4):383-397.

125. **Lawrence JG, Ochman H**. Reconciling the many faces of lateral gene transfer. *Trends Microbiol* 2002;10(1):1-4.

126. **Watanabe H, Wada A, Inagaki Y, Itoh K, Tamura K**. Outbreaks of enterohaemorrhagic *Escherichia coli* O157:H7 infection by two different genotype strains in Japan, 1996. *Lancet* 1996;348(9030):831-832.

127. **Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K et al.** Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* 2001;8(1):11-22.

128. **Welch RA, Burland V, Plunkett G, Redford P, Roesch P et al.** Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* 2002;99(26):17020-17024.

129. **Chaudhuri RR, Sebaihia M, Hobman JL, Webber MA, Leyton DL et al.** Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain 042. *PLoS One* 2010;5(1):e8801.

130. **Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF et al.** The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol* 2008;190(20):6881-6893.

131. **Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S et al.** Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLOS Genetics* 2009;5(1):e1000344.

132. **Mau B, Glasner JD, Darling AE, Perna NT**. Genome-wide detection and analysis of homologous recombination among sequenced strains of *Escherichia coli*. *Genome Biol* 2006;7(5):R44.

133. **Tettelin H, Masignani V, Cieslewicz MJ *et al.*** Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome". *Proc Natl Acad Sci USA* 2005;102(39):13950-13955.

134. **Tettelin H, Medini D**. The pangenome: diversity, dynamics and evolution of genomes. 2020. Springer; Cham. 2020.

135. **Lukjancenko O, Wassenaar TM, Ussery DW**. Comparison of 61 sequenced *Escherichia coli* genomes. *Microb Ecol* 2010;60(4):708-720.

136. **Fukiya S, Mizoguchi H, Tobe T, Mori H**. Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* strains revealed by comparative genomic hybridization microarray. *J Bacteriol* 2004;186(12):3911-3921.

137. **Ochman H, Jones IB**. Evolutionary dynamics of full genome content in *Escherichia coli*. *EMBO J* 2000;19(24):6637-6643.

138. **Decano AG, Downing T**. An *Escherichia coli* ST131 pangenome atlas reveals population structure and evolution across 4,071 isolates. *Sci Rep* 2019;9(1):17394.

139. **McNally A, Oren Y, Kelly D, Pascoe B, Dunn S et al.** Combined analysis of variation in cre, accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. *PLoS Genet* 2016;12(9):e1006280.

140. **Zhou Z, Alikhan NF, Mohamed K, Fan Y, Achtman M *et al.*** The EnteroBase user's guide, with case studies on *Salmonella* transmission, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome Res* 2020;30(1):138-152.

141. **Xu J**. Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol Ecol* 2006;15(7):1713-1731.

142. **Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G et al.** Host-gut microbiota metabolic interactions. *Science* 2012;336(6086):1262-1267.

143. **Round JL, Mazmanian SK**. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* 2009;9(5):313-323.

144. **Muegge BD, Kuczynski J, Knights D, Clemente JC, González A et al.** Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* 2011;332(6032):970-974.

145. **Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR et al.** Evolution of mammals and their gut microbes. *Science* 2008;320(5883):1647-1651.

146. **Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY et al.** Linking long-term dietary patterns with gut microbial enterotypes. *Science* 2011;334(6052):105-108.

147. **David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE et al.** Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 2014;505(7484):559-563.

148. **Flint HJ**. The impact of nutrition on the human microbiome. *Nutr Rev* 2012;70 Suppl 1:S10-13.

149. **Hooper LV, Macpherson AJ**. Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nat Rev Immunol* 2010;10(3):159-169.

150. **Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R et al.** The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med* 2009;1(6):6ra14.

151. **Claus SP, Guillou H, Ellero-Simatos S**. The gut microbiota: a major player in the toxicity of environmental pollutants? *NPJ Biofilms Microbiomes* 2016;2:16003.

152. **Brown AC, Valiere A**. Probiotics and medical nutrition therapy. *Nutr Clin Care* 2004;7(2):56-68.

153. **Sorbara MT, Pamer EG**. Interbacterial mechanisms of colonization resistance and the strategies pathogens use to overcome them. *Mucosal Immunol* 2019;12(1):1-9.

154. **Maltby R, Leatham-Jensen MP, Gibson T, Cohen PS, Conway T**. Nutritional basis for colonization resistance by human commensal *Escherichia coli* strains HS and Nissle 1917 against *E. coli* O157:H7 in the mouse intestine. *PLoS One* 2013;8(1):e53957.

155. **Chang DE, Smalley DJ, Tucker DL, Leatham MP, Norris WE et al.** Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proc Natl Acad Sci U S A* 2004;101(19):7427-7432.

156. **Leatham MP, Banerjee S, Autieri SM, Mercado-Lubo R, Conway T et al.** Precolonized human commensal *Escherichia coli* strains serve as a barrier to *E. coli* O157:H7 growth in the streptomycin-treated mouse intestine. *Infection and immunity* 2009;77(7):2876-2886.

157. **Deriu E, Liu JZ, Pezeshki M, Edwards RA, Ochoa RJ et al.** Probiotic bacteria reduce *Salmonella typhimurium* intestinal colonization by competing for iron. *Cell Host Microbe* 2013;14(1):26-37.

158. **Tanaka M, Nakayama J**. Development of the gut microbiota in infancy and its impact on health in later life. *Allergol Int* 2017;66(4):515-522.

159. **Quince C, Ijaz UZ, Loman N, Eren AM, Saulnier D et al.** Extensive modulation of the fecal metagenome in children with Crohn's disease during exclusive enteral nutrition. *Am J Gastroenterol* 2015;110(12):1718-1729; quiz 1730.

160. **Codling C, O'Mahony L, Shanahan F, Quigley EM, Marchesi JR**. A molecular analysis of fecal and mucosal bacterial communities in irritable bowel syndrome. *Dig Dis Sci* 2010;55(2):392-397.

161. **Miyoshi M, Ogawa A, Higurashi S, Kadooka Y**. Anti-obesity effect of *Lactobacillus gasseri* SBT2055 accompanied by inhibition of pro-inflammatory gene

expression in the visceral adipose tissue in diet-induced obese mice. *Eur J Nutr* 2014;53(2):599-606.

162.   **Kang JH, Yun SI, Park MH, Park JH, Jeong SY et al.** Anti-obesity effect of *Lactobacillus gasseri* BNR17 in high-sucrose diet-induced obese mice. *PLoS One* 2013;8(1):e54617.

163.   **Kadooka Y, Sato M, Imaizumi K, Ogawa A, Ikuyama K et al.** Regulation of abdominal adiposity by probiotics (*Lactobacillus gasseri* SBT2055) in adults with obese tendencies in a randomized controlled trial. *Eur J Clin Nutr* 2010;64(6):636-643.

164.   **Lynch SV, Boushey HA**. The microbiome and development of allergic disease. *Curr Opin Allergy Clin Immunol* 2016;16(2):165-171.

165.   **Berg RD**. The indigenous gastrointestinal microflora. *Trends Microbiol* 1996;4(11):430-435.

166.   **Gordon DM, Cowling A**. The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology (Reading)* 2003;149(Pt 12):3575-3586.

167.   **Mitsuoka T, Ohno K, Benno Y, Suzuki K, Namba K**. [The fecal flora of man. IV. Communication: Comparison of the newly developed method with the old conventional method for the analysis of intestinal flora (author's transl)]. *Zentralbl Bakteriol Orig A* 1976;234(2):219-233.

168.   **Penders J, Thijs C, Vink C, Stelma FF, Snijders B et al.** Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* 2006;118(2):511-521.

169. **Lescat M, Clermont O, Woerther PL, Glodt J, Dion S et al.** Commensal *Escherichia coli* strains in Guiana reveal a high genetic diversity with host-dependant population structure. *Environ Microbiol Rep* 2013;5(1):49-57.

170. **Foster-Nyarko E, Alikhan NF, Ravi A *et al.*** Genomic diversity of *Escherichia coli* from non-human primates in the Gambia. *Microb Genom* 2020;6(9).

171. **Poulsen LK, Lan F, Kristensen CS, Hobolth P, Molin S et al.** Spatial distribution of *Escherichia coli* in the mouse large intestine inferred from rRNA in situ hybridization. *Infect Immun* 1994;62(11):5191-5194.

172. **Poulsen LK, Licht TR, Rang C, Krogfelt KA, Molin S**. Physiological state of *Escherichia coli* BJ4 growing in the large intestines of streptomycin-treated mice. *J Bacteriol* 1995;177(20):5840-5845.

173. **Bettelheim KA, Faiers M, Shooter RA**. Serotypes of *Escherichia coli* in normal stools. *Lancet* 1972;2(7789):1223-1224.

174. **Slanetz LW, Bartley CH**. Numbers of enterococci in water, sewage, and feces determined by the membrane filter technique with an improved medium. *J Bacteriol* 1957;74(5):591-595.

175. **Escherich T**. Die darmbakterien des däuglings und ihre beziehungen zur physiologie der verdauung. Verlag von Ferdinand Enke, Stuttgart 1886.

176. **Mueller NT, Bakacs E, Combellick J, Grigoryan Z, Dominguez-Bello MG**. The infant microbiome development: mom matters. *Trends Mol Med* 2015;21(2):109-117.

177. **Syed SA, Abrams GD, Freter R**. Efficiency of various intestinal bacteria in assuming normal functions of enteric flora after association with germ-free mice. *Infect Immun* 1970;2(4):376-386.

178. **Richter TKS, Michalski JM, Zanetti L, Tennant SM, Chen WH et al.** Responses of the human gut. *mSystems* 2018;3(4).

179. **Suvarna K, Stevenson D, Meganathan R, Hudspeth ME**. Menaquinone (vitamin K2) biosynthesis: localization and characterization of the *menA* gene from *Escherichia coli*. *J Bacteriol* 1998;180(10):2782-2787.

180. **Martinson JNV, Walk ST**. Residency in the gut of healthy human adults. *EcoSal Plus* 2020;9(1).

181. **Abia ALK, Ubomba-Jaswa E**. Dirty money on holy ground: Isolation of potentially pathogenic bacteria and fungi on money collected from church offerings. *Iran J Public Health* 2019;48(5):849-857.

182. **Akoachere JF, Gaelle N, Dilonga HM, Nkuo-Akenji TK**. Public health implications of contamination of Franc CFA (XAF) circulating in Buea (Cameroon) with drug resistant pathogens. *BMC Res Notes* 2014;7:16.

183. **Nwankwo EO, Ekwunife N, Mofolorunsho KC**. Nosocomial pathogens associated with the mobile phones of healthcare workers in a hospital in Anyigba, Kogi state, Nigeria. *J Epidemiol Glob Health* 2014;4(2):135-140.

184. **Pal S, Juyal D, Adekhandi S, Sharma M, Prakash R et al.** Mobile phones: Reservoirs for the transmission of nosocomial pathogens. *Adv Biomed Res* 2015;4:144.

185. **Reeves PR, Liu B, Zhou Z, Li D, Guo D et al.** Rates of mutation and host transmission for an *Escherichia coli* clone over 3 years. *PloS One* 2011;6(10):e26907-e26907.

186. **Derakhshandeh A, Eraghi V, Boroojeni AM, Niaki MA, Zare S et al.** Virulence factors, antibiotic resistance genes and genetic relatedness of commensal *Escherichia coli* isolates from dogs and their owners. *Microb Pathog* 2018;116:241-245.

187. **Toombs-Ruane LJ, Benschop J, French NP, Biggs PJ, Midwinter AC *et al.***
Carriage of extended-spectrum-beta-lactamase- and AmpC beta-lactamase-producing

Escherichia coli strains from humans and pets in the same households. *Appl Environ Microbiol* 2020;86(24).

188. **Ercumen A, Pickering AJ, Kwong LH, Arnold BF, Parvez SM *et al.*** Animal feces contribute to domestic fecal contamination: Evidence from *E. coli* measured in water, hands, food, flies, and soil in Bangladesh. *Environ Sci Technol* 2017;51(15):8725-8734.

189. **Aijuka M, Santiago AE, Girón JA, Nataro JP, Buys EM**. Enteroaggregative *Escherichia coli* is the predominant diarrheagenic *E. coli* pathotype among irrigation water and food sources in South Africa. *Int J Food Microbiol* 2018;278:44-51.

190. **Yar DD**. Bacterial contaminants and antibiogram of Ghana paper currency notes in circulation and their associated health risks in Asante-Mampong, Ghana. *Int J Microbiol* 2020;2020:8833757.

191. **Dixit OVA, O'Brien CL, Pavli P, Gordon DM**. Within-host evolution versus immigration as a determinant of *Escherichia coli* diversity in the human gastrointestinal tract. *Environ Microbiol* 2018;20(3):993-1001.

192. **Richter TKS, Hazen TH, Lam D, Coles CL, Seidman JC *et al.*** Temporal variability of *Escherichia coli* diversity in the gastrointestinal tracts of Tanzanian children with and without exposure to antibiotics. mSphere 2018;3(6).

193. **Chen SL, Wu M, Henderson JP, Hooton TM, Hibbing ME et al.** Genomic diversity and fitness of *E. coli* strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. *Sci Transl Med* 2013;5(184):184ra160.

194. **Shooter RA, Bettleheim KA, Lennox-King SM, O'Farrell S**. Escherichia coli serotypes in the faeces of healthy adults over a period of several months. *J Hyg (Lond)* 1977;78(1):95-98.

195. **Wiedemann B, Habermann R, Knothe H, Ihrig L**. [Stablility of *Escherichia coli*-flora in healthy men. 3. Occurrence of permanent and transitory strains in infants]. *Arch Hyg Bakteriol* 1971;154(6):581-589.

196. **Sears HJ, Brownlee I, Uchiyama JK**. Persistence of individual strains of *Escherichia coli* in the intestinal tract of man. *J Bacteriol* 1950;59(2):293-301.

197. **Sears HJ, Brownlee I**. Further observations on the persistence of individual strains of *Escherichia coli* in the intestinal tract of man. *J Bacteriol* 1952;63(1):47-57.

198. **Tannock GW, Tiong IS, Priest P, Munro K, Taylor C et al.** Testing probiotic strain *Escherichia coli* Nissle 1917 (Mutaflor) for its ability to reduce carriage of multidrug-resistant E. coli by elderly residents in long-term care facilities. *J Med Microbiol* 2011;60(Pt 3):366-370.

199. **Johnson JR, Clabots C**. Sharing of virulent *Escherichia coli* clones among household members of a woman with acute cystitis. *Clin Infect Dis* 2006;43(10):e101-108.

200. **Gordon DM, O'Brien CL, Pavli P**. *Escherichia coli* diversity in the lower intestinal tract of humans. *Environ Microbiol Rep* 2015;7(4):642-648.

201. **Escobar-Páramo P, Grenet K, Le Menac'h A, Rode L, Salgado E et al.** Large-scale population structure of human commensal *Escherichia coli* isolates. *Appl Environ Microbiol* 2004;70(9):5698-5700.

202. **Duriez P, Clermont O, Bonacorsi S, Bingen E, Chaventré A et al.** Commensal *Escherichia coli* isolates are phylogenetically distributed among geographically distinct human populations. *Microbiology* 2001;147(Pt 6):1671-1676.

203. **Skurnik D, Bonnet D, Bernède-Bauduin C, Michel R, Guette C et al.** Characteristics of human intestinal *Escherichia coli* with changing environments. *Environ Microbiol* 2008;10(8):2132-2137.

204. **Massot M, Daubié AS, Clermont O, Jauréguy F, Couffignal C et al.** Phylogenetic, virulence and antibiotic resistance characteristics of commensal strain populations of *Escherichia coli* from community subjects in the Paris area in 2010 and evolution over 30 years. *Microbiology* 2016;162(4):642-650.

205. **Freter R, Brickner H, Botney M, Cleven D, Aranki A**. Mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora. *Infection and immunity* 1983;39(2):676-685.

206. **Freter R, Brickner H, Fekete J, Vickerman MM, Carey KE**. Survival and implantation of *Escherichia coli* in the intestinal tract. *Infect Immun* 1983;39(2):686-703.

207. **Pereira FC, Berry D**. Microbial nutrient niches in the gut. *Environ Microbiol* 2017;19(4):1366-1378.

208. **Conway T, Cohen PS**. Commensal and pathogenic *Escherichia coli* metabolism in the gut. *Microbiol Spectr* 2015;3(3).

209. **Hardin G**. The competitive exclusion principle. *Science* 1960;131(3409):1292-1297.

210. **Nowrouzian F, Adlerberth I, Wold AE**. P fimbriae, capsule and aerobactin characterize colonic resident *Escherichia coli. Epidemiol Infect* 2001;126(1):11-18.

211. **Granato ET, Meiller-Legrand TA, Foster KR**. The evolution and ecology of bacterial warfare. *Curr Biol* 2019;29(11):R521-R537.

212. **Ghoul M, Mitri S**. The ecology and evolution of microbial competition. *Trends Microbiol* 2016;24(10):833-845.

213. **Fabich AJ, Leatham MP, Grissom JE, Wiley G, Lai H et al.** Genotype and phenotypes of an intestine-adapted *Escherichia coli* K-12 mutant selected by animal passage for superior colonization. *Infection and immunity* 2011;79(6):2430-2439.

214. **Leatham MP, Stevenson SJ, Gauger EJ, Krogfelt KA, Lins JJ et al.** Mouse intestine selects nonmotile flhDC mutants of *Escherichia coli* MG1655 with increased colonizing ability and better utilization of carbon sources. *Infect Immun* 2005;73(12):8039-8049.

215. **Gauger EJ, Leatham MP, Mercado-Lubo R, Laux DC, Conway T et al.** Role of motility and the flhDC Operon in *Escherichia coli* MG1655 colonization of the mouse intestine. *Infect Immun* 2007;75(7):3315-3324.

216. **Leatham-Jensen MP, Frimodt-Moller J, Adediran J, Mokszycki ME, Banner ME et al.** The streptomycin-treated mouse intestine selects *Escherichia coli* envZ missense mutants that interact with dense and diverse intestinal microbiota. *Infection and immunity* 2012;80(5):1716-1727.

217. **Ng KM, Ferreyra JA, Higginbottom SK, Lynch JB, Kashyap PC et al.** Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* 2013;502(7469):96-99.

218. **Dolfing J**. Syntrophy in microbial fuel cells. *The ISME journal* 2014;8(1):4-5.

219. **Tuncil YE, Xiao Y, Porter NT, Reuhs BL, Martens EC *et al.*** Reciprocal prioritization to dietary glycans by gut bacteria in a competitive environment promotes stable coexistence. *mBio* 2017;8(5).

220. **Naylor SW, Low JC, Besser TE, Mahajan A, Gunn GJ et al.** Lymphoid follicle-dense mucosa at the terminal rectum is the principal site of colonization of enterohemorrhagic *Escherichia coli* O157:H7 in the bovine host. *Infect Immun* 2003;71(3):1505-1512.

221. **Pamer EG**. Resurrecting the intestinal microbiota to combat antibiotic-resistant pathogens. *Science* 2016;352(6285):535-538.

222. **Buffie CG, Pamer EG**. Microbiota-mediated colonization resistance against intestinal pathogens. *Nat Rev Immunol* 2013;13(11):790-801.

223. **Ravi A, Halstead FD, Bamford A, Casey A, Thomson NM et al.** Loss of microbial diversity and pathogen domination of the gut microbiota in critically ill patients. *Microb Genom* 2019;5(9).

224. **Buffie CG, Jarchum I, Equinda M, Lipuma L, Gobourne A et al.** Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to *Clostridium difficile*-induced colitis. *Infect Immun* 2012;80(1):62-73.

225. **Dethlefsen L, Huse S, Sogin ML, Relman DA**. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* 2008;6(11):e280.

226. **Ribeiro CFA, Silveira GGOS, Cândido ES, Cardoso MH, Espínola Carvalho CM et al.** Effects of antibiotic treatment on gut microbiota and how to overcome its negative impacts on human health. *ACS Infect Dis* 2020;6(10):2544-2559.

227. **Morrow LE, Wischmeyer P**. Blurred lines: Dysbiosis and probiotics in the ICU. *Chest* 2017;151(2):492-499.

228. **Gough E, Shaikh H, Manges AR**. Systematic review of intestinal microbiota transplantation (fecal bacteriotherapy) for recurrent *Clostridium difficile* infection. *Clin Infect Dis* 2011;53(10):994-1002.

229. **Manges AR, Steiner TS, Wright AJ**. Fecal microbiota transplantation for the intestinal decolonization of extensively antimicrobial-resistant opportunistic pathogens: a review. *Infect Dis (Lond)* 2016;48(8):587-592.

230. **Youngster I, Russell GH, Pindar C, Ziv-Baran T, Sauk J et al.** Oral, capsulized, frozen fecal microbiota transplantation for relapsing *Clostridium difficile* infection. *JAMA* 2014;312(17):1772-1778.

231. **Hill C, Guarner F, Reid G, Gibson GR, Merenstein DJ et al.** Expert consensus document. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nat Rev Gastroenterol Hepatol* 2014;11(8):506-514.

232. **Gorbach SL**. Probiotics and gastrointestinal health. *Am J Gastroenterol* 2000;95(1 Suppl):S2-4.

233. **Sivan A, Corrales L, Hubert N, Williams JB, Aquino-Michaels K et al.** Commensal *Bifidobacterium* promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* 2015;350(6264):1084-1089.

234. **Jacobi CA, Malfertheiner P**. *Escherichia coli* Nissle 1917 (Mutaflor): new insights into an old probiotic bacterium. *Dig Dis* 2011;29(6):600-607.

235. **Nissle A**. [Mutaflor and its medical significance]. *Z Klin Med* 1951;2(3-4):68.

236. **Nissle A**. [Explanations of the significance of colonic dysbacteria & the mechanism of action of *E. coli* therapy (mutaflor)]. *Medizinische* 1959;4(21):1017-1022.

237. **Nissle A.** [Old and new experiences on therapeutic successes by restoration of the colonic flora with mutaflor in gastrointestinal diseases]. *Med Welt* 1961;29-30:1519-1523.

238. **Hancock V, Dahl M, Klemm P**. Probiotic *Escherichia coli* strain Nissle 1917 outcompetes intestinal pathogens during biofilm formation. *J Med Microbiol* 2010;59(Pt 4):392-399.

239. **Lodinová-Zádniková R, Sonnenborn U**. Effect of preventive administration of a nonpathogenic *Escherichia coli* strain on the colonization of the intestine with microbial pathogens in newborn infants. *Biol Neonate* 1997;71(4):224-232.

240. **Altenhoefer A, Oswald S, Sonnenborn U, Enders C, Schulze J et al.** The probiotic *Escherichia coli* strain Nissle 1917 interferes with invasion of human intestinal epithelial cells by different enteroinvasive bacterial pathogens. *FEMS Immunol Med Microbiol* 2004;40(3):223-229.

241. **Boudeau J, Glasser AL, Julien S, Colombel JF, Darfeuille-Michaud A**. Inhibitory effect of probiotic *Escherichia coli* strain Nissle 1917 on adhesion to and invasion of intestinal epithelial cells by adherent-invasive E. coli strains isolated from patients with Crohn's disease. *Aliment Pharmacol Ther* 2003;18(1):45-56.

242. **Zyrek AA, Cichon C, Helms S, Enders C, Sonnenborn U et al.** Molecular mechanisms underlying the probiotic effects of *Escherichia coli* Nissle 1917 involve ZO-2 and PKCzeta redistribution resulting in tight junction and epithelial barrier repair. *Cell Microbiol* 2007;9(3):804-816.

243. **Wehkamp J, Harder J, Wehkamp K, Wehkamp-von Meissner B, Schlee M et al.** NF-kappaB- and AP-1-mediated induction of human beta defensin-2 in intestinal epithelial cells by *Escherichia coli* Nissle 1917: a novel effect of a probiotic bacterium. *Infect Immun* 2004;72(10):5750-5758.

244. **Timmerman HM, Niers LE, Ridwan BU, Koning CJ, Mulder L et al.** Design of a multispecies probiotic mixture to prevent infectious complications in critically ill patients. *Clin Nutr* 2007;26(4):450-459.

245. **Venturi A, Gionchetti P, Rizzello F, Johansson R, Zucconi E et al.** Impact on the composition of the faecal flora by a new probiotic preparation: preliminary data on

maintenance treatment of patients with ulcerative colitis. *Aliment Pharmacol Ther* 1999;13(8):1103-1108.

246. **Meador JP, Caldwell ME, Cohen PS, Conway T**. *Escherichia coli* pathotypes occupy distinct niches in the mouse intestine. *Infection and immunity* 2014;82(5):1931-1938.

247. **Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ**. Within-host evolution of bacterial pathogens. *Nat Rev Microbiol* 2016;14(3):150-162.

248. **Bryant J, Chewapreecha C, Bentley SD**. Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences. *Future Microbiol* 2012;7(11):1283-1296.

249. **Lee H, Popodi E, Tang H, Foster PL**. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci U S A* 2012;109(41):E2774-2783.

250. **Rodgers K, McVey M**. Error-prone repair of DNA double-strand breaks. *Journal of cellular physiology* 2016;231(1):15-24.

251. **Periwal V, Scaria V**. Insights into structural variations and genome rearrangements in prokaryotic genomes. *Bioinformatics (Oxford, England)* 2015;31(1):1-9.

252. **Iguchi A, Iyoda S, Terajima J, Watanabe H, Osawa R**. Spontaneous recombination between homologous prophage regions causes large-scale inversions within the *Escherichia coli* O157:H7 chromosome. *Gene* 2006;372:199-207.

253. **Raeside C, Gaffé J, Deatherage DE, Tenaillon O, Briska AM et al.** Large chromosomal rearrangements during a long-term evolution experiment with *Escherichia coli*. *mBio* 2014;5(5):e01377-01314.

254. **Charlesworth B**. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 2009;10(3):195-205.

255. **Kuo CH, Moran NA, Ochman H**. The consequences of genetic drift for bacterial genome complexity. *Genome Res* 2009;19(8):1450-1454.

256. **Ochman H, Lawrence JG, Groisman EA**. Lateral gene transfer and the nature of bacterial innovation. *Nature* 2000;405(6784):299-304.

257. **Hacker J, Carniel E**. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep* 2001;2(5):376-381.

258. **Didelot X, Méric G, Falush D, Darling AE**. Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*. *BMC Genomics* 2012;13:256.

259. **Cao Q, Didelot X, Wu Z, Li Z, He L et al.** Progressive genomic convergence of two *Helicobacter pylori* strains during mixed infection of a patient with chronic gastritis. *Gut* 2015;64(4):554-561.

260. **Kennemann L, Didelot X, Aebischer T, Kuhn S, Drescher B et al.** *Helicobacter pylori* genome evolution during human infection. *Proc Natl Acad Sci U S A* 2011;108(12):5033-5038.

261. **Didelot X, Nell S, Yang I, Woltemate S, van der Merwe S et al.** Genomic evolution and transmission of *Helicobacter pylori* in two South African families. *Proceedings of the National Academy of Sciences of the United States of America* 2013;110(34):13880-13885.

262. **Andersson JO, Andersson SG**. Insights into the evolutionary process of genome degradation. *Curr Opin Genet Dev* 1999;9(6):664-671.

263. **Mather AE, Harris SR**. Playing fast and loose with mutation. *Nat Rev Microbiol* 2013;11(12):822.

264. **Marinus MG**. DNA methylation and mutator genes in *Escherichia coli* K-12. *Mutat Res* 2010;705(2):71-76.

265. **Köser CU, Holden MTG, Ellington MJ, Cartwright EJP, Brown NM *et al.*** Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *NEJM* 2012;366(24):2267-2275.

266. **Wang S, Wu C, Shen J, Wu Y, Wang Y**. Hypermutable *Staphylococcus aureus* strains present at high frequency in subclinical bovine mastitis isolates are associated with the development of antibiotic resistance. *Vet Microbiol* 2013;165(3-4):410-415.

267. **Marvig RL, Johansen HK, Molin S, Jelsbak L**. Genome analysis of a transmissible lineage of *Pseudomonas aeruginosa r*eveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genet* 2013;9(9):e1003741.

268. **Lidin-Janson G, Kaijser B, Lincoln K, Olling S, Wedel H**. The homogeneity of the faecal coliform flora of normal school-girls, characterized by serological and biochemical properties. *Med Microbiol Immunol* 1978;164(4):247-253.

269. **Schlager TA, Hendley JO, Bell AL, Whittam TS**. Clonal diversity of *Escherichia coli* colonizing stools and urinary tracts of young girls. *Infect Immun* 2002;70(3):1225-1229.

270. **Golubchik T, Batty EM, Miller RR, Farr H, Young BC et al.** Within-host evolution of *Staphylococcus aureus* during asymptomatic aarriage. *PloS one* 2013;8(5):e61319.

271. **Hall MD, Holden MT, Srisomang P, Mahavanakul W, Wuthiekanun V et al.** Improved characterisation of MRSA transmission using within-host bacterial sequence diversity. *Elife* 2019;8.

272. **Worby CJ, Lipsitch M, Hanage WP**. Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. *PLoS Comput Biol* 2014;10(3):e1003549.

273.   **Ley SD, de Vos M, Van Rie A, Warren RM**. Deciphering within-host microevolution of *Mycobacterium tuberculosis* through whole-genome sequencing: the phenotypic impact and way forward. *Microbiol Mol Biol Rev* 2019;83(2).

274.   **Ssengooba W, de Jong BC, Joloba ML, Cobelens FG, Meehan CJ**. Whole genome sequencing reveals mycobacterial microevolution among concurrent isolates from sputum and blood in HIV infected TB patients. *BMC Infect Dis* 2016;16:371.

275.   **Trauner A, Liu Q, Via LE, Liu X, Ruan X et al.** The within-host population dynamics of *Mycobacterium tuberculosis* vary with treatment efficacy. *Genome Biol* 2017;18(1):71.

276.   **Herranz M, Pole I, Ozere I, Chiner-Oms Á, Martínez-Lirola M *et al.*** *Mycobacterium tuberculosis* acquires limited genetic diversity in prolonged infections, reactivations and transmissions involving multiple hosts. *Front Microbiol* 2017;8:2661.

277.   **Kay GL, Sergeant MJ, Zhou Z, Chan JZ, Millard A et al.** Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nature communications* 2015;6:6717.

278.   **Halachev MR, Chan JZM, Constantinidou CI, Cumley N, Bradley C et al.** Genomic epidemiology of a protracted hospital outbreak caused by multidrug-resistant *Acinetobacter baumannii* in Birmingham, England. *Genome Medicine* 2014;6(11):70.

279.   **Wymant C, Hall M, Ratmann O, Bonsall D, Golubchik T et al.** PHYLOSCANNER: Inferring Transmission from within- and between-host pathogen genetic diversity. *Mol Biol Evol* 2018;35(3):719-733.

280. **Cornelissen M, Pasternak AO, Grijsen ML, Zorgdrager F, Bakker M et al.** HIV-1 dual infection is associated with faster CD4+ T-cell decline in a cohort of men with primary HIV infection. *Clin Infect Dis* 2012;54(4):539-547.

281. **Smith HL**. Zur kenntnis der colibacillen des säuglingsstuhles. *Zentr Bakt Parasitenk I Orig* 1899;25:689-693.

282. **Wallick H, Stuart CA**. Antigenic relationships of *Escherichia coli* isolated from one individual. *J Bacteriol* 1943;45(2):121-126.

283. **Totsuka, K**. Studien über *Bacterium coli*. *Z Hyg Infektionskrankh* 1902;45:115-124.

284. **Robinet HG**. Relationship of host antibody to fluctuations of *Escherichia coli* serotypes in the human intestine. *J Bacteriol* 1962;84:896-901.

285. **Branche WC, Young VM, Robinet HG, Massey ED**. Effect of colicine production on *Escherichia coli* in the normal human intestine. *Proc Soc Exp Biol Med* 1963;114:198-201.

286. **Caugant DA, Levin BR, Selander RK**. Genetic diversity and temporal variation in the E. coli population of a human host. *Genetics* 1981;98(3):467-490.

287. **Selander RK, Levin BR**. Genetic diversity and structure in *Escherichia coli* populations. *Science* 1980;210(4469):545-547.

288. **Tzabar Y, Pennington TH**. The population structure and transmission of *Escherichia coli* in an isolated human community; studies on an Antarctic base. *Epidemiol Infect* 1991;107(3):537-542.

289. **Damborg P, Nielsen SS, Guardabassi L**. *Escherichia coli* shedding patterns in humans and dogs: insights into within-household transmission of phylotypes associated with urinary tract infections. *Epidemiol Infect* 2009;137(10):1457-1464.

290. **Martinson JNV, Pinkham NV, Peters GW, Cho H, Heng J et al.** Rethinking gut microbiome residency and the Enterobacteriaceae in healthy human adults. *The ISME Journal* 2019.

291. **Stoesser N, Sheppard AE, Moore CE, Golubchik T, Parry CM et al.** Extensive within-host diversity in fecally carried extended-spectrum-beta-lactamase-producing *Escherichia coli* isolates: Implications for transmission analyses. *J Clin Microbiol* 2015;53(7):2122-2131.

292. **Knudsen PK, Brandtzaeg P, Høiby EA, Bohlin J, Samuelsen Ø et al.** Impact of extensive antibiotic treatment on faecal carriage of antibiotic-resistant enterobacteria in children in a low resistance prevalence setting. *PLoS One* 2017;12(11):e0187618.

293. **Knudsen PK, Gammelsrud KW, Alfsnes K, Steinbakk M, Abrahamsen TG** *et al.* Transfer of a *bla*$_{CTX-M-1}$-carrying plasmid between different *Escherichia coli* strains within the human gut explored by whole genome sequencing analyses. *Sci Rep* 2018;8(1):280.

294. **Stegger M, Leihof RF, Baig S, Sieber RN, Thingholm KR et al.** A snapshot of diversity: Intraclonal variation of *Escherichia coli* clones as commensals and pathogens. *Int J Med Microbiol* 2020;310(3):151401.

295. **Nielsen KL, Dynesen P, Larsen P, Frimodt-Møller N**. Faecal *Escherichia coli* from patients with *E. coli* urinary tract infection and healthy controls who have never had a urinary tract infection. *J Med Microbiol* 2014;63(Pt 4):582-589.

296. **Foster-Nyarko E, Alikhan NF, Ikumapayi UN, Sarwar G, Okoi C et al**. Genomic diversity of E*scherichia coli* from healthy children in rural Gambia *PeerJ* 2021;9:e10572.

ORIGINAL UNEDITED MANUSCRIPT

297. **Blyton MD, Banks SC, Peakall R, Gordon DM**. High temporal variability in commensal *Escherichia coli* strain communities of a herbivorous marsupial. *Environ Microbiol* 2013;15(8):2162-2172.

298. **Li SP, Tan J, Yang X, Ma C, Jiang L**. Niche and fitness differences determine invasion success and impact in laboratory bacterial communities. *Isme j* 2019;13(2):402-412.

299. **Foster-Nyarko E, Alikhan NF, Ravi A, Thomson NM, Jarju S** *et al.* Genomic diversity of *Escherichia coli* from chickens and guinea fowl in the Gambia. *Microb Genom* 2020.

300. **Neidhardt FC, Eingraham JL, Low KB, Magasanik B, Schaechter M & Umbarger HE** (eds). *Escherichia coli* and *Salmonella typhimurium: Cellular and Molecular Biology.* ASM Press; Washington, DC. 1996.

301. **Durso LM, Smith D, Hutkins RW**. Measurements of fitness and competition in commensal *Escherichia coli* and *E. coli* O157:H7 strains. *Appl Environ Microbiol* 2004;70(11):6466-6472.

Table 1: *The contribution of studies using E. coli across various fields.*

Field(s)	Contribution of studies with *E. coli*	Reference(s)
Molecular biology, physiology and genetics	The explanation of the genetic code	[298]
	DNA replication	[299]
	Transcription	[300]
	The life cycle of bacterial viruses	[301, 302]
	The elucidation of the molecular basis of antibiotic tolerance	[303]
	The discovery of restriction enzymes	[304, 305]
	Swarming motility	[306, 307]
	Gene regulation	[308]
	Elucidation of the structure and function of ATP synthase	[309]
Pharmaceuticals	The synthesis of recombinant proteins in vivo, such as insulin, which is used to treat millions of people with diabetes worldwide	[50]
	The synthesis of several other biopharmaceuticals, such as human interferon-β, interleukin-2, human growth hormone and human blood clotting factors	[310, 311]
Evolution	*E. coli* is the model organism of choice in experimental evolution studies; for example, in Lenski's long-term evolution experiment, on-going since February 1988 and spanning over 60,000 generations	[312]
	The demonstration of the stochastic nature of mutations	[313, 314]
	Mapping the trajectory of long-term fitness	[315]
	The elucidation of how sexual recombination influences adaptation	[316]
	Insights into predator-prey interactions	[317]
	The evolution of a novel trait, aerobic citrate utilization	[318]
Genetic engineering and biotechnology	The development of recombinant DNA techniques and molecular cloning, the production of biofuels and industrial chemicals such as phenol	[319]
	Mannitol production	[320]
	Ethanol production	[321, 322]

Table 2: *Classification and examples of the pathotypes of E. coli.*

Pathotype/pathovar	Virulence mechanism(s)	Host range	Reference
InPEC	Locus of enterocyte effacement; pathogenicity island 1	Humans, all mammals	[51]
InPEC	Small fimbrial adhesins; toxins; transcriptional activator gene; aggregative adhesion	Humans	[51]
InPEC	Shiga toxin or verotoxins; afimbrial and fimbrial adhesins	Humans, piglets	[51]
InPEC	Heat labile and heat-stable enterotoxins	Humans, ruminants, pigs, dogs	[51]
InPEC	Invasion and multiplication in enterocytes	Humans, primates	[51]
InPEC	Adhesins	Humans, animals	[51]
InPEC	Shiga toxin	Humans	[323]
InPEC	Adherent invasive phenotype	Humans and animals	[51]
ExPEC	Adhesins, secretion and iron uptake systems, increased serum survival and cytotoxic proteins	Birds	[324]
ExPEC	Fimbrial adhesins; siderophores, resistance to complement	Humans, animals (especially dogs and cats)	[325]
ExPEC	Iron acquisition systems, degradation of interferon-gamma and cleavage of the human defensin LL-37	Humans	[44]
ExPEC	Fimbrial adhesins; siderophores; resistance to complement	All mammals and birds (especially poultry)	[325]
ExPEC	Unknown	Animals	[326]
ExPEC	Type II, IV and VI secretion systems, long polar fimbriae (lpfA) and iron acquisition	Animals	[327]
ExPEC	Cytotoxic Necrotizing Factors 1 or 2 and α haemolysin; fimbrial and/or afimbrial adhesins; siderophores; resistance to complement	Humans, animals and ruminants	[325]

InPEC, Intestinal pathogenic *E. coli*; ExPEC, Extraintestinal pathogenic *E. coli*; EPEC, Enteropathogenic *E. coli;* EAEC, Enteroaggregative *E. coli;* EHEC*,* Enterohaemorrhagic *E. coli;* ETEC*,* Enterotoxigenic *E. coli;* EIEC*,* Enteroinvasive *E. coli;* DAEC, Diffusely adherent *E. coli;* STEAEC*,* Shiga-toxigenic Enteraggregative *E. coli;* AIEC*,* Adherent invasive *E. coli*; APEC*,* Avian pathogenic *E. coli;* UPEC, Uropathogenic *E. coli;* NMEC, Neonatal meningitis *E. coli;* SePEC, Human sepsis-associated *E. coli;* MPEC*,* Mammary pathogenic *E. coli;* EnPEC*,* Endometrial pathogenic *E. coli;* NTEC, Necrotoxigenic *E. coli.*

Table 3: *Colonisation and fitness factors (Adapted from [21]).*

Virulence factor	Pathotype/pathovar	Effect(s)
icsA/virG	EIEC	Nucleates actin filaments
Intimin	EPEC/EHEC	Adhesin, inducing TH1 response
Dr adhesins	DAEC/UPEC	Adhesin, binds to decay-accelerating factor and activates phosphatidylinositol 3-kinase, induces MHC class I chain-related gene A
P (*Pap*) fimbriae	UPEC	Adhesin, also induces cytokine expression
Colonisation factor antigens	ETEC	Adhesin
S fimbriae	UPEC/NMEC	Adhesin
Bundle-forming pili (BFP)	EPEC	Type IV pili
Aggregative adherence fimbriae	EAEC	Adhesin
paa	EPEC/EHEC	Adhesin
toxB	EHEC	Adhesin
Efa-1/LifA	EHEC	Adhesin
Long polar filaments	EHEC/EPEC	Adhesin
saa	EHEC	Adhesin
ompA	NMEC/EHEC	Adhesin
Curli	Various	Adhesin, binds to fibronectin
ibeA/B/C	NMEC	Stimulates invasion
aslA	NMEC	Stimulates invasion
Dispersin	EAEC	Stimulates colonisation; facilitates mucous penetration
K antigen capsules	MNEC	Antiphagocytic activity
Aerobactin	EIEC	Siderophore, iron acquisition
Yersiniabactin	Various	Siderophore, iron acquisition
ireA	UPEC	Siderophore, iron acquisition
iroN	UPEC	Siderophore, iron acquisition
chu (*shu*)	EIEC/UPEC/NMEC	Siderophore, iron acquisition
Flagellin	All	Motility, inducing cytokine expression through Toll-like receptors
Liposaccharides	All	Inducing cytokine expression through Toll-like receptors

Please see Table 2 for the full names of the pathovars/pathotypes.

Table 4: *E. coli toxins and effectors (Adapted from [21]).*

Virulence factor	Pathotype/pathovar	Toxin class	Effect(s)
Heat-labile enterotoxin	ETEC	AB subunit/type II effector	ADP ribosylates and activation of adenylate cyclase, leading to ion secretion
Shiga toxin	EHEC	AB subunit	Depurination of rRNA, inhibiting protein synthesis and inducing apoptosis
Cytolethal distending toxin	Various	ABC subunit	DNase activity, blocks mitosin in G2/M phase
Shigella enterotoxin 1	EAEC/ EIEC	AB subunit	Ion secretion
Urease	EHEC	ABC subunit	Cleaves urea to NH_3 and CO_3
EspC	EPEC	Autotransporter	Serine protease, cleavage of coagulation
EspP	EHEC	Autotransporter	Serine protease, cleavage of coagulation factor V
Haemoglobin-binding protease	ExPEC, APEC	Autotransporter	Degradation of haemoglobin to release haem/iron
pet	EAEC	Autotransporter	Serine protease; ion secretion and cytotoxicity
pic	UPEC, EAEC, EIEC	Autotransporter	Protease/mucinase
sat	UPEC	Autotransporter	Vacuolation
sepA	EIEC	Autotransporter	Serine protease
sigA	EIEC	Autotransporter	Ion secretion
Cycle-inhibiting factor	EPEC, EHEC	Type III effector	Blocks mitosis in G2/M phase, resulting in the inactivation of cdk1
espF	EPEC, EHEC	Type III effector	Opens tight junctions and reduces apoptosis
espH	EPEC, EHEC	Type III effector	Modulates filopodia and pedestal formation
map	EPEC, EHEC	Type III effector	Disrupts mitochondrial membrane potential
tir	EPEC, EHEC	Type III effector	Nucleates cytoskeletal proteins, loss of microvilli and GAP-like activity
ipaA	EIEC	Type III effector	Actin depolymerisation
ipaB	EIEC	Type III effector	Apoptosis, Interleukin-1 release and membrane insertion
ipaC	EIEC	Type III effector	Actin polymerisation
ipaH	EIEC	Type III effector	Modulation of inflammation
ipgD	EIEC	Type III effector	Inositol 4-phosphatase and membrane blebbing
VirA	EIEC	Type III effector	Microtubule destabilisation and membrane ruffling
stcE	EHEC	Type II effector	Cleavage of C1-esterase inhibitor and disruption of the complement cascade
hlyA	UPEC	Repeats-in-toxin (RTX) toxin	Cell lysis
ehx	EHEC	RTX toxin	Cell lysis
Cytotoxic necrotising factors (1 and 2)	NMEC, UPEC, NTEC	-	Altered cytoskeleton and necrosis
LifA/Efa	EPEC, EHEC	-	Inhibits lymphocyte activation and adhesion
Shigella enterotoxin 2	EIEC, ETEC		Ion secretion
Heat-stable enterotoxin a	ETEC	Heat-stable enterotoxins	Activating guanylate cyclase, leading to ion secretion
Heat-stable enterotoxin b	ETEC	Heat-stable enterotoxins	Ion secretion via increasing intracellular calcium
Enteroaggregative *E. coli* heat-stable enterotoxin	Various	Heat-stable enterotoxin	Activating guanylate cyclase, leading to ion secretion

Please see Table 2 for the full names of the pathovars/pathotypes.