Genomics and pathotypes of the many faces of Escherichia coli

Geurtsen, Jeroen; de Been, Mark; Weerdenburg, Eveline; Zomer, Aldert; McNally, Alan; Poolman, Jan

DOI:
10.1093/femsre/fuac031

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Geurtsen, J, de Been, M, Weerdenburg, E, Zomer, A, McNally, A & Poolman, J 2022, 'Genomics and pathotypes of the many faces of Escherichia coli', *FEMS Microbiology Reviews*, vol. 46, no. 6, fuac031. https://doi.org/10.1093/femsre/fuac031

FEMS Microbiology Reviews, 1–30
DOI: 10.1093/femsre/fuac031
Advance access publication date: 24 June 2022
Review Article

Genomics and pathotypes of the many faces of *Escherichia coli*

Jeroen Geurtsen[1], Mark de Been[1], Eveline Weerdenburg[1], Aldert Zomer[2], Alan McNally[3], Jan Poolman[1,*]

[1]Janssen Vaccines and Prevention B.V., 2333 Leiden, the Netherlands
[2]Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, 3584 Utrecht, the Netherlands
[3]Institute of Microbiology and Infection, College of Medical and Dental Sciences, University of Birmingham, B15 2TT Birmingham, United Kingdom
*Corresponding author: Bacterial Vaccine Discovery and Early Development, Janssen, PO Box 2048, 2301 CA Leiden, the Netherlands. Tel: +31 71 5197626; E-mail: JPoolman@its.jnj.com
Editor: Wilbert Bittercoli [InPEC, IPEC]; diarrheagenicE.

Abstract

Escherichia coli is the most researched microbial organism in the world. Its varied impact on human health, consisting of commensalism, gastrointestinal disease, or extraintestinal pathologies, has generated a separation of the species into at least eleven pathotypes (also known as pathovars). These are broadly split into two groups, intestinal pathogenic *E. coli* (InPEC) and extraintestinal pathogenic *E. coli* (ExPEC). However, components of *E. coli*'s infinite open accessory genome are horizontally transferred with substantial frequency, creating pathogenic hybrid strains that defy a clear pathotype designation. Here, we take a birds-eye view of the *E. coli* species, characterizing it from historical, clinical, and genetic perspectives. We examine the wide spectrum of human disease caused by *E. coli*, the genome content of the bacterium, and its propensity to acquire, exchange, and maintain antibiotic resistance genes and virulence traits. Our portrayal of the species also discusses elements that have shaped its overall population structure and summarizes the current state of vaccine development targeted at the most frequent *E. coli* pathovars. In our conclusions, we advocate streamlining efforts for clinical reporting of ExPEC, and emphasize the pathogenic potential that exists throughout the entire species.

Keywords: accessory genome, antibiotic resistance, bacterial species, *Escherichia coli* pathotypes, population dynamics, virulence factors

Introduction

In 1885, German-Austrian pediatrician Theodor Escherich identified the *Bacterium coli commune* as the causative agent of childhood diarrhea (Friedmann 2014), and a few decades later the bacterium was officially named after its discoverer—*Escherichia coli* was born. Ironically, genome sequencing of the original strain of Dr. Escherich's laboratory later revealed that he had, in fact, isolated a benign member of the gut microbiota, lacking pathogenicity islands and well-known virulence factors such as the Shiga toxin (Meric *et al.* 2016). Today, *E. coli* is the most popular model organism in microbiological, genetic, and molecular biology research, and, arguably, '*the most intensively studied and best understood organism on the planet*' (Blount 2015). Work on *E. coli* shaped our understanding of the genetic code, DNA transcription, and restriction enzymes (Lehman *et al.* 1958, Crick *et al.* 1961, Linn and Arber 1968), among many other groundbreaking concepts of modern genetics. Sophisticated tools that enable targeted manipulations of its genome have been developed for *E. coli* more than for any other bacterial organism. The bacterium's reliable presence in fecal matter and its ability to persist in a variety of environments outside of a (warm-blooded) host has resulted in its long-standing use as a fecal indicator bacterium for water quality (Gruber *et al.* 2014, World Health Organization 2017). *E. coli* is also utilized as an indicator organism for antimicrobial resistance of Gram-negative bacteria, and many surveillance programs evaluate the antibiogram of this bacterium to determine resistance levels in a given location and time (Moreno *et al.* 2000, European Food Safety Authority 2008, Gekenidis *et al.* 2018).

Unfortunately, *E. coli*'s ubiquitous occurrence and commensalism has resulted in a certain degree of clinical confusion. In addition to its harmless existence as a gut commensal, *E. coli* is a major cause of human disease. The bacterium is the causative agent of a variety of intestinal pathologies such as watery and/or bloody diarrhea, hemolytic uremic syndrome (HUS), and colitis. It also causes extraintestinal diseases such as bacteremia and sepsis, meningitis, and urinary tract infections (UTI), and is one of the most common causes of both healthcare-associated and community-onset invasive bacterial disease. Clinical reporting of the frequency of these *E. coli* diseases can lead to inaccurate estimates by the use of imprecise terminology. The term *Enterobacteriaceae*, which is sometimes utilized in clinical reports, is an example—pathogenic *E. coli* is by far more prevalent than any other member of this bacterial family. Even reporting *E. coli* under its correct species name can lead to confusion since diseases caused by different *E. coli* pathotypes, such as uropathogenic *E. coli* (UPEC), enterotoxigenic *E. coli* (ETEC), and enterohemorrhagic *E. coli* (EHEC), are clinically diverse, affect different tissues, use distinctive mechanisms of pathogenesis, and have markedly dissimilar patterns of transmission (see Table 1 for an overview of currently identified *E. coli* pathotypes and their acronyms).

At least 11 different *E. coli* pathotypes in two categories (intestinal and extraintestinal pathogenic *E. coli*) have been defined (Croxen and Finlay 2010, Sarowska *et al.* 2019, Denamur *et al.*

Received: November 20, 2021. Accepted: June 22, 2022

Table 1. Current *E. coli* pathotype nomenclature and characteristics.

Pathotype [acronym]; synonym [synonym acronym]	Clinical presentation[a]	Typical genetic markers[a]
Intestinal pathogenic E. coli [InPEC, IPEC]; diarrheagenic E. coli [DEC]		
Enteropathogenic *E. coli* [EPEC]	attaching and effacing lesions on surfaces of intestinal epithelial cells; diarrhea, often accompanied by fever, vomiting, and dehydration; most often in children	locus of enterocyte effacement (LEE) that includes intimin (*eae*); bundle-forming pilus (*bfp*, only in typical EPEC, tEPEC; missing in atypical EPEC, aEPEC)
Shiga toxin-producing *E. coli* [STEC]; verocytotoxin-producing *E. coli* [VTEC]; enterohemorrhagic *E. coli* [EHEC][b]	attaching and effacing lesions; diverse pathotype: mild to bloody diarrhea (hemorrhagic colitis) accompanied by fever, abdominal cramping, or vomiting; can lead to hemolytic uremic syndrome (HUS)	Shiga toxin 1 or 2 (*stx1* or *stx2*), LEE (in some but not all lineages)
Enterotoxigenic *E. coli* [ETEC]	mild to severe watery diarrhea, often in children in developing countries or travelers	heat-labile enterotoxin (LT) and/or heat-stable cholera toxin-like enterotoxin (ST); various colonization factors (for example, CFA/I, CS6, CS30, etc.)
Enteroinvasive *E. coli* [EIEC] and *Shigella*	highly invasive; cause shigellosis/bacillary dysentery with profuse diarrhea, fever and potential damage to intestinal walls; may proceed to HUS	invasion plasmid pINV, Shiga toxin, enterotoxin on chromosomal pathogenicity island, lack of flagella
Enteroaggregative *E. coli* [EAEC]	travelers' diarrhea, often watery, sometimes chronic; progression to HUS possible; aggregative adhesion with typical stacked-brick adherence pattern on HEp-2 cells	pAA virulence plasmids containing genes for aggregative adherence fimbriae
Diffuse adhering *E. coli* [DAEC][c]	persistent watery diarrhea in children; diffuse adherence pattern; may be associated with Crohn's disease and ulcerative colitis (Mirsepasi-Lauridsen *et al.* 2019) and cause urinary tract infections	Afa/Dr adhesins
Adherent-invasive *E. coli* [AIEC]	associated with Crohn's disease and ulcerative colitis (Mirsepasi-Lauridsen *et al.* 2019); adhere to and invade epithelial cells; survive and replicate within macrophages (Palmela *et al.* 2018)	lack of consistent virulence factors (O'Brien *et al.* 2017, Rakitina *et al.* 2017, Camprubi-Font *et al.* 2018)
Extraintestinal pathogenic E. coli [ExPEC]		
Uropathogenic *E. coli* [UPEC]	urinary tract infections (including recurrent), cystitis, pyelonephritis, prostatitis (Rudick *et al.* 2011)	no canonical pathotype-specific virulence factors determined for any ExPEC pathotype, including UPEC, NMEC, MAEC, and APEC[d]; factors shown to play a role in virulence include PapGII and PapGIII adhesins in UPEC (Stromberg *et al.* 1990, Sung *et al.* 2001), K1 capsule and pS88 antigen in NMEC (Lemaitre *et al.* 2012), curli fiber, fimbrial adhesin H (FimH), Dr fimbriae and afimbrial adhesins (Afa), P fimbriae, hemolysin A, Pic, Sat, Vat, Cdt, Cnf, antigen 43, and many others
Neonatal meningitis *E. coli* [NMEC, NEMEC]; meningitis-associated *E. coli* [MAEC]	potentially fatal meningitis in newborns	
Avian pathogenic *E. coli* [APEC]	colibacillosis and its manifestations (such as septicemia, cellulitis, etc) in poultry	
Sepsis-associated *E. coli* [SEPEC]	bacteremia/sepsis	
Unspecified ExPEC	Skin and soft tissue infections (Moet *et al.* 2007), ventilator-assisted pneumonia (La Combe *et al.* 2019), others	

[a]Compiled from (Croxen and Finlay 2010, Croxen *et al.* 2013, Sarowska *et al.* 2019), except where explicitly referenced otherwise
[b]Occasionally defined as subset of STEC specifically associated with HUS
[c]Specific afa/dr$^+$ DAEC isolates have also been associated with extraintestinal disease (Usein *et al.* 2001, Le Bouguenec and Servin 2006). Therefore, DAEC bacteria may also belong to the ExPEC category.
[d]Presence of the autotransporter Vat, the yersiniabactin siderophore FyuA, the heme siderophore ChuA, and the fimbrial protein YfcV have been suggested to be predictive of the potential to infect the urinary tract (Spurbeck *et al.* 2012).

2021). Their differentiation is based on both bacterial genetics and the pathologies and phenotypes that specific isolates engender in their host. Unfortunately, these attempts have resulted in a divided view of the species in the medical and research community. Reporting of the bacterium with variable nomenclature arguably resulted in fragmentation of epidemiological data on *E. coli* disease, a tendency that is also observed in public datasets of whole genome sequences, where pathotypes are often missing or applied with inconsistent terminology. This fragmentation may have been one reason why *E. coli* as a species was omitted from the Active Bacterial Core surveillance, established in 1995 by the US Centers of Disease Control (CDC), which aims to survey invasive bacterial pathogens of public health importance (Langley *et al.* 2015)—despite being the number one bacterium that causes bloodstream

infections and sepsis reported in population-based studies (Laupland and Church 2014, Fay et al. 2020, Rhee et al. 2020). It was also originally left out of the selection of bacterial species specifically targeted by the US National Institute of Allergy and Infectious Diseases for their antimicrobial resistance (the ESKAPE group) (Peters et al. 2008)—regardless of being the top species in comprehensive assessments of antimicrobial-resistant bacterial pathogens causing healthcare-associated infections in US hospitals (Weiner et al. 2016, Weiner-Lastinger et al. 2019).

Escherichia coli disease is thus imprecisely tracked. Moreover, only two *E. coli* intestinal pathotypes are on the current list of reportable conditions from the US CDC (https://wwwn.cdc.gov/nndss/conditions/notifiable/2021/); Shiga toxin-producing *E. coli* (STEC) and enteroinvasive *E. coli* (EIEC), together with *Shigella* and carbapenemase-producing carbapenem-resistant *Enterobacteriaceae*. The omission of clinical reporting extends to extraintestinal invasive *E. coli* diseases such as bacteremia causing sepsis, a potentially fatal condition now more commonly caused by *E. coli* than by any other bacterial pathogen (Wilson et al. 2011, Rhee et al. 2020). This may hopefully change in the future—whereas *E. coli* bacteremia is not reportable in the US and most other developed and developing countries, it has been under mandatory reporting in the UK since 2011 (Davies et al. 2017).

The fragmented view of the bacterial species of *E. coli* as a pathogen may have contributed to a level of underestimation of the impact of *E. coli* on human health. The following review aims to illuminate and characterize *E. coli* as a species and to encourage a conscious re-evaluation of the bacterium in medical, scientific, and public health discourse. In our conclusions, we underline the ability of *E. coli* to disseminate virulence factors across the species, building ubiquitous pathogenic potential. We also discuss the benefit of grouping *E. coli* pathotypes into two categories—InPEC and ExPEC, and argue for restricting the ExPEC designation to describe an *E. coli* pathogroup instead of a pathotype. Updated guidelines for clinical reporting would benefit from a sharpened focus on more consistent tracking, which would promote a better understanding of the true magnitude of *E. coli*'s effects on human disease.

Definition of the species—conventions and challenges

Since the late 1980s, microbes are assigned to a common species if their reciprocal, pairwise DNA re-association values are ≥ 70% in DNA–DNA hybridization experiments under standardized conditions, and their melting temperature is ≤ 5°C (Wayne 1988). In addition, microbes with 16S ribosomal RNAs (rRNAs) that are ≤ 98.7% identical are generally considered to be members of different species, as this correlates well with < 70% DNA–DNA similarity (Achtman and Wagner 2008). However, distinct species have been occasionally described with 16S rRNAs that are > 98.7% identical. Such exceptions are often based on the sensible guideline that phenotypic consistency should prevail within members of the same species, and/or to maintain historical species allocations that remain clinically relevant.

Escherichia coli is such an example—*Shigella* represents a separate genus despite their members' close genetic relationship with *E. coli*. *Shigella*, or, more precisely, *S. dysenteriae*, was first isolated and described as causing bacillary dysentery in 1898 by the Japanese scientist Kiyoshi Shiga (Trofa et al. 1999), 13 years after the discovery of *E. coli*. *Shigella* was first listed under its current genus name in *Bergey's Manual of Determinative Bacteriology* in 1930. The bacterium is currently organized into four named species, *S. dysenteriae, S. flexneri, S. boydii,* and *S. sonnei*, each representing a separate serogroup within the genus (*S. sonnei* bacteria all share the same serotype) (Kotloff et al. 2018). *Shigella* strains are highly contagious: as few as ten *S. dysenteriae* type 1, *S. flexneri* or *S. sonnei* cells resulted in disease in humans (DuPont et al. 1989). *Shigella* are characterized by a high level of gene reduction, a likely consequence of their intracellular and host-restricted lifestyle. This pathoadaptive gene reduction includes the silencing of antivirulence loci (Bliven and Maurelli 2012) as well as superfluous gene products or metabolic pathways, resulting in high numbers of pseudogenes (Wei et al. 2003). Many insertion sequences (IS) are scattered throughout the bacterium's genome, and the expansion of IS elements IS1, IS2, IS4, IS600, and IS911 in all four *Shigella* species contributed to the convergent functional gene loss within and between species (Hawkey et al. 2020). *Shigella*'s level of functional gene reduction has also been linked, at least in part, to a genome-wide (rather than pathway-specific) reduction in purifying selection (Hershberg et al. 2007), a consequence of their limited host range (humans and, rarely, primates). *Shigella* strains have repeatedly evolved from different branches of the *E. coli* tree, via independent acquisition of the pINV virulence plasmid (Lan and Reeves 2002, Hershberg et al. 2007), which contains over 100 genes and includes a 30 kb core pathogenicity island necessary for enteroinvasiveness (Buchrieser et al. 2000). In addition to this plasmid, *Shigella* contains several pathogenicity islands and produce toxins, most notably the Shiga toxin encoded by the *stx* locus in members of *S. dysenteriae*.

In the 1940s, enteroinvasive *E. coli* (EIEC) were discovered, which—other than a slightly lower virulence and a higher infectious dose requirement—cause very similar clinical symptoms to *Shigella*. EIEC are biochemically distinguishable from *Shigella* by their higher ability to ferment mucate and utilize serine, xylose, or sodium acetate (Belotserkovsky and Sansonetti 2018). The bacteria share most of *Shigella*'s characteristics, including the lack of motility, the ability to invade epithelial cells, the invasion plasmid pINV, and an intracellular lifestyle. Phylogenetic analyses suggest that *Shigella* and EIEC arose from *E. coli* ancestors independently on multiple occasions (Pupo et al. 2000, Pettengill et al. 2015, Sahl et al. 2015). Three diverse EIEC lineages evolved separately from different lineages of *E. coli* by pINV acquisition and, in some cases, (some of) the *Shigella* pathogenicity islands (Hazen et al. 2016). Consequently, EIEC contain some but not all of *Shigella* virulence factors. Moreover, their level of gene reduction is generally not as advanced as in *Shigella*, and EIEC display metabolic activity that is more similar to *E. coli* (Pettengill et al. 2015, Pasqua et al. 2017). Some pathoadaptive reductions are common between *Shigella* and EIEC, such as the lack of motility and the deletion of *ompT,* a gene that is located on a defective prophage and encodes a protease that interferes with host cell invasion (Nakata et al. 1993). Another example is the silencing of lysine decarboxylase activity through mutations in the *cad* locus, preventing cadaverine synthesis, a molecule that blocks the release of bacteria into the cytoplasm of an infected cell (Bliven and Maurelli 2012).

Based purely on ANI, *E. coli* and *Shigella* combined represent a singular typical species with a whole-genome average nucleotide identity of up to 98% (Jain et al. 2018). Plasmid transfer has also been proven to occur between members of the two genera (Thanh Duy et al. 2020). *Shigella* experts have therefore frequently called for a taxonomic revision of the current status that would group *Shigella* firmly within the genus of *Escherichia*, to promote the understanding of accurate evolutionary relationships of *Shigella* and *E. coli* (Lan and Reeves 2002, Chaudhuri and Henderson 2012), with

Shigella to be classified as EIEC (Pettengill *et al.* 2015). However, despite the purely historical reasons behind the ongoing taxonomic separation of these organisms, their differentiation has served the medical community well—the term shigellosis immediately conveys a disease severity that any *E. coli*-related term currently fails to transmit. Consequently, 'shigellosis' (as opposed to the more precise 'infection with *Shigella* or enteroinvasive *E. coli*') has been a reportable disease in the US since 1944, and the terminology has, to date, resisted modifications. Retaining the four *Shigella* species as 'later heterotypic synonyms of *E. coli*' has also been proposed for the Genome Taxonomy Database (GTDB) since merging of the *Shigella* genus with the *Escherichia* genus resulted in misassignment of the majority of *E. coli* isolates to novel *Escherichia* species formed by the *Shigella* isolates (Parks *et al.* 2021).

In the era of whole-genome sequencing, new approaches to the species definition center on the average nucleotide identity (ANI) of all orthologous genes in complete genome sequences of pairs of strains (Konstantinidis *et al.* 2006). Initial studies suggested that a 95% ANI would correspond to the historical hybridization-based cutoff (Goris *et al.* 2007), a value that has temporarily been adjusted to 92% after interrogation of more than 1000 genomes (Zhang *et al.* 2014), but has since been set back to 95%, after evaluation of over 90 000 prokaryotic genome assemblies (Jain *et al.* 2018). At this threshold, precision of species allocation was 98.7% of all interrogated assemblies, excluding *Shigella* and *E. coli* genomes that were strongly represented in the dataset and displayed consistently higher ANI values in the orthologous genes of their strains. Outliers also included different but very closely related species within the genus of *Mycobacterium*, and *Neisseria* species (Jain *et al.* 2018). However, different ANI or amino acid identity (AAI) computation methods can yield inconsistent measurements. For example, phylogeny inferred after relative evolutionary divergence (RED) normalization from 120 concatenated single-copy ubiquitous proteins on a set of > 90 000 prokaryotic genomes frequently differed in taxonomic assignments from the NCBI database. While most of these inconsistencies were at ranks higher than species, several species assignments were also altered (Parks *et al.* 2018). Therefore, careful normalization may be prudent and species determination should be supported by evolutionary history (Palmer *et al.* 2020).

Complete reconciliation of purely DNA-based species definitions with historical phenotypic assignments may be impossible, not least because recombination events and the resulting transfer of phenotypic traits go some way to blur these lines (Nesbo *et al.* 2006). For *E. coli*, an organism with a wide variety of pathogenic traits and an ability to accept new genetic material in form of plasmids, integrons, transposons, phages, and other mobile elements, this may be particularly true, and discussions are ongoing about the inclusion of additional *Escherichia* clades into the species (Walk 2015). The description of five cryptic atypical *Escherichia* clades based on genetic analysis of 22 house-keeping genes (extended multi-locus sequence typing, eMLST) from isolates that were biochemically indistinguishable from *E. coli* (Walk *et al.* 2009) eventually resulted in three proposed new *Escherichia* species and a suggested inclusion of cryptic clade I as a subspecies of *E. coli* (Liu *et al.* 2015, Walk 2015). The term *E. coli sensu lato* was proposed which encompassed this new subspecies, while *E. coli sensu stricto* would refer to classic *E. coli* strains (Clermont *et al.* 2013). In this review, the term *E. coli* is used as a descriptor of *E. coli sensu stricto.*

Luckily, *E. coli's* eminent recombination rates, despite being much higher than mutation rates, do not override phylogeny (Touchon *et al.* 2009), so that relatively stable phylogenetic groups can be identified within the species. Seven recognized phylogenetic lineages (phylogenetic groups) of *E. coli*, A, B1, B2, C, D, E, and F, are based on PCR patterns and MLST from a total of 17 house-keeping genes (Clermont *et al.* 2013). A subsequent study added an eighth lineage of pathogenic isolates, phylogroup G, as an intermediate between phylogroups F and B2 (Clermont *et al.* 2019). The rapid acquisition of novel genome sequence information reiterated this structure, although pathogenic isolates belonging to additional phylogroups exist (Lu *et al.* 2016, Gonzalez-Alba *et al.* 2019).

The underlying mechanism of how this genomic 'order in disorder' can be retained lies, in part, in preferred integration hotspot regions on the genome where recombination occurs, leaving other 'core' regions of the genome relatively stable (Touchon *et al.* 2009). Another aspect that limits recombination is the need to integrate newly acquired genes into existing metabolic pathways and cellular functions, thereby requiring favorable genetic backgrounds (Touchon *et al.* 2020). Consequently, the phylogenic structure is imprinted on strains occupying different habitats, from the gastrointestinal tract of mammals and birds to host-independent environments such as soil and water. Specific phylogenetic clusters are very frequently isolated from some sources, and rare in others. Isolates from phylogroup B1 are often seen in environmental water samples, for example, and are relatively rare in humans (Touchon *et al.* 2020), while human extraintestinal pathogenic *E. coli* isolates are mostly comprised of strains in phylogenetic groups B2 and D.

The diversity of the species continues to foster discussion on better approaches of taxonomic classification, including suggestions of a designation of *E. coli* as a species complex, with 'ecotype' subclassifications to express some of its phenotypic, genetic, and physiological range (Yu *et al.* 2021).

In an effort to describe part of the variability in diseases and associated virulence factors of *E. coli*, a specific pathotype terminology has been adopted. The currently accepted pathotype scheme is summarized in Table 1 and broadly separates pathogenic *E. coli* into two major categories, based on the primary location of the diseases they cause—(i), intestinal/diarrheagenic; and (ii), extraintestinal. A review of the phylogenetic features of all pathotypes in this classification can be found elsewhere (Denamur *et al.* 2021), and emphasizes the repeated emergence of virulence in the history of the species by acquisition of virulence determinants into specific (phylo)genetic backgrounds.

Distinction of intestinal pathogenic *E. coli* (InPEC) isolates into six major pathotypes has been suggested and reiterated for a few decades (Nataro and Kaper 1998, Kaper *et al.* 2004) (Table 1). Historically, these intestinal pathovars have been classified, determined, and separated from their nonpathogenic counterparts by using tissue culture experiments that included cytotoxicity, attachment, and invasion assays, and animal models (reviewed in detail in (Riley 2020)). The pathogenesis, clinical outcomes, mechanisms of disease and epidemiology of most of these intestinal pathotypes have been extensively and excellently reviewed elsewhere (Kaper *et al.* 2004, Croxen and Finlay 2010, Croxen *et al.* 2013, Mirsepasi-Lauridsen *et al.* 2019). The currently applied InPEC separation into pathovars is often based on expression or presence of specific virulence factors that are directly related to the clinical features of the resultant disease and can be used in diagnostics. The seventh InPEC pathovar, adherent-invasive *E. coli* (AIEC) (Darfeuille-Michaud 2002), was added in the late 1990s and delineates strains that are often associated with inflammatory bowel disease such as ulcerative colitis and Crohn's disease. Members of this pathotype are able to adhere to and invade epithelial in-

testinal cells (Palmela *et al.* 2018), and to survive in macrophages (Shaler *et al.* 2019). Intriguingly, AIEC lack the common virulence factors found in the other InPEC pathotypes, and they are sometimes found in healthy individuals (Shaler *et al.* 2019). Such non-pathogenic AIEC may form a distinct phylogenetic group (Barrios-Villa *et al.* 2020). Lastly, the repertoire of virulence factors and disease associations of ExPEC, an umbrella designation suggested and adopted at the turn of the millennium (Russo and Johnson 2000), as well as the worldwide distribution of their lineages have been studied comprehensively (reviewed in (Dale and Woodford 2015, Manges *et al.* 2019)).

In addition to the pathotype and the serogroup classifications of *E. coli*, at least two other classification schemes are frequently applied to facilitate epidemiological studies and scientific discourse about the bacterium. First, serotyping is a classification method first developed in the 1970s based on properties of the 181 O-antigens (components of the surface lipopolysaccharide), 53 H-antigens (indicating the protein content of the bacterium's flagella, often encoded by the *fliC* gene) and the 80 capsule-based K-antigens of *E. coli* (Orskov *et al.* 1977, Whitfield and Roberts 1999, Wang *et al.* 2003, Iguchi *et al.* 2015, DebRoy *et al.* 2016, Liu *et al.* 2019). Second, MLST-derived sequence type (ST) describes the strain variability based on sequence information on seven housekeeping genes (*adk; fumC; gyrB, icd, mdh, purA, recA*). The most comprehensive *E. coli* MLST database is currently hosted by the University of Warwick (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli) and includes over 175 000 strains (Wirth *et al.* 2006).

The clinical distinction of *E. coli* into InPEC, ExPEC, and commensals may be, to a certain extent, mirrored in the bacterium's capability to grow on different substrates. *In silico*, metabolic signatures have been identified from sequencing data that support this general classification. Simulated metabolic phenotypes of 55 *E. coli* isolates on 654 variable growth-supporting nutrient conditions composed of different carbon, nitrogen, phosphorous, and sulfur nutrient sources in aerobic and anaerobic conditions resulted in reliable clustering of the strains based on this broad classification scheme (Monk *et al.* 2013). However, *E. coli* can quickly adapt to novel metabolic conditions and the transfer of a single DNA segment has conferred a new metabolic capability thousands of times in the bacterium's evolutionary history, suggesting a significant degree of mobility and speed in *E. coli*'s adaptation to preferred ecological niches (Pang and Lercher 2019). This capability suggests that metabolic profiles may not reach the necessary level of distinction to be applicable in routine clinical diagnostics, although a supporting role can be envisioned, particularly in view of the extensive pathovar mixing observed in the past few years (explicated in more detail in a subsequent section, below).

The spectrum of *E. coli* diseases in human hosts

Escherichia coli's ability to cause disease is set against a backdrop of its harmless presence in the environment, where it can survive, and perhaps thrive, in soil, water, and sediments, and its benign existence as a commensal. It is a member of the human normal gut flora, a microbial assembly that likely consists of over 1000 species-level phylotypes. Here, *E. coli* is usually asymptomatically carried (Lozupone *et al.* 2012) at low abundance. Although *E. coli*'s exact proportion in the human gut flora may vary from person to person, it is known that members of the *Proteobacteria* phylum collectively represent only about 1% of the healthy gut microbiome (Bradley and Pollard 2017), although this figure can significantly increase in intestinal disorders such as Crohn's Disease (Vester-Andersen *et al.* 2019). *Escherichia coli* is the predominant aerobic bacterium in the human gastrointestinal tract, and present in over 90% of all humans (Tenaillon *et al.* 2010). Particular isolates of *E. coli* are utilized as probiotic treatment of intestinal imbalances (such as in *Mutaflor*, which contains the commensal strain Nissle 1917 (Grozdanov *et al.* 2004), or *Colinfant New Born*, a preparation of an *E. coli* strain with the serological signature O83:K24:H31). These probiotic *E. coli* strains strengthen the mucosal barrier (Ukena *et al.* 2007) with reported benefits in allergy prevention (Lodinova-Zadnikova *et al.* 2003) and suspected therapeutic effects in autoimmune disease (Secher *et al.* 2017). However, the complexity of *E. coli*'s role in human disease is intimated even in these probiotic isolates. As an example, *E. coli* Nissle 1917 has been shown to (at least sometimes) reduce the likelihood of remission when used in treatment regimens of ulcerative colitis (Petersen *et al.* 2014, Mirsepasi-Lauridsen *et al.* 2016).

The spectrum of diseases caused by *E. coli* includes both intestinal and extraintestinal pathologies (Table 1), and a selection of specific prominent loci of possible *E. coli* infections in humans is illustrated in Fig. 1.

Intestinal *E. coli* disease manifestations

The incidence of acute intestinal disease caused by *E. coli* is difficult to assess since diarrhea, one of the major manifestations of intestinal *E. coli* disease, is heavily underreported and, fortunately, self-resolves in most cases without medical intervention. However, studies in children < 5 years of age displaying medium to severe diarrheal symptoms in developing countries identified ETEC (specifically, ETEC producing stable enterotoxin ST) and EPEC as among the most frequent contributors to the disease, and the top agents resulting in an increased risk of death (Kotloff *et al.* 2013). ST-ETEC was also a significant contributor to mild diarrheal symptoms in these cohorts (Kotloff *et al.* 2019). InPEC can cause acute infections displaying as gastroenteritis or colitis, which may become persistent and/or hemorrhagic and progress into hemolytic uremic syndrome (HUS). Based on data gathered between 2000 and 2008, Shiga toxin-producing *E. coli* (including all *E. coli* O157:H7) cause about 265 000 intestinal disease cases per year and over 3600 hospitalizations in the US (Scallan *et al.* 2011). At the same time, bacillary dysentery caused by *Shigella* species and EIEC reached an incidence of about 500 000 cases per year in the US (Scallan *et al.* 2011).

Even within *E. coli* pathotype boundaries, considerable pathogenic variability exists. As an example, an EPEC infection can generate either lethal, or symptomatic, or asymptomatic outcomes, and a comparison of genomes of isolates of these three classes did not reveal a canonical class of genes that characterized these profound pathogenic differences (Hazen *et al.* 2016). Instead, these differences may be expressions of several genomic regions acting together, as well as a response to host factors (Hazen *et al.* 2016). Similarly, a close inspection of genomic features (presence/absence patterns of single genes and k-mer analysis) of *Shigella* and EIEC isolates did not reveal any features predictive of disease severity (Hendriks *et al.* 2020).

In addition to the diseases mentioned above, *E. coli* has been shown to be the overall most frequent Gram-negative bacillus isolated from human intra-abdominal infections (IAIs) (Sartelli *et al.* 2014, Chang *et al.* 2017, Zhang *et al.* 2017, Ponce-de-Leon *et al.* 2018, Yoon *et al.* 2019). These infections include intra-abdominal abscesses, primary, secondary, and tertiary forms of peritonitis, and single-organ infections such as diverticulitis, appendicitis,

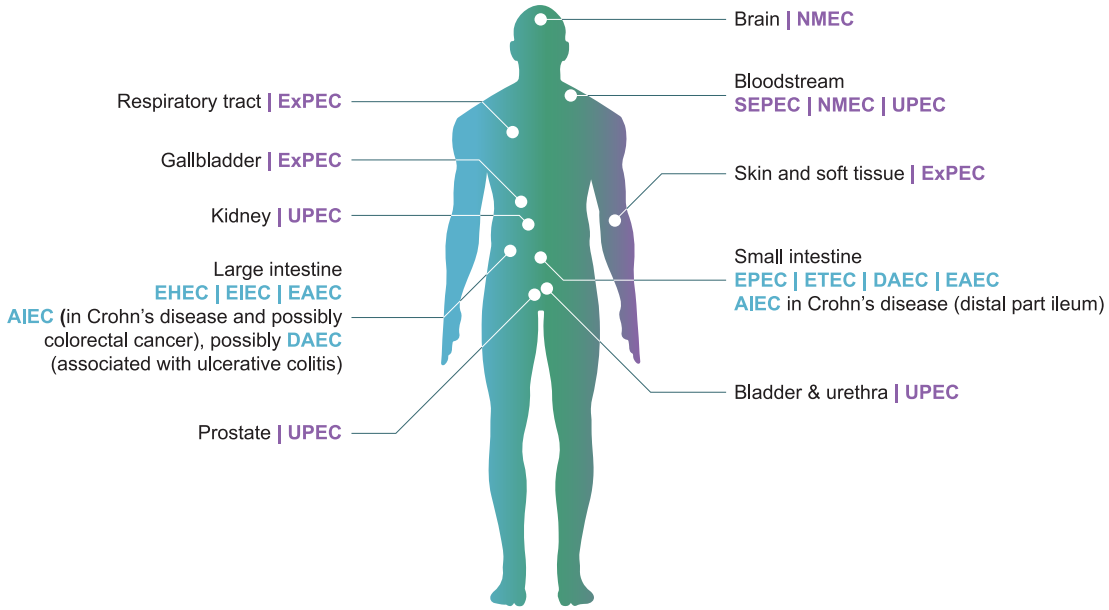

Figure 1. Selection of *E. coli* colonization loci in the human host. Intestinal and extraintestinal pathogenic *E. coli* can colonize and infect many tissue types in the human body. Specific *E. coli* pathotypes are indicated, where known. Intestinal pathogenic *E. coli* (InPEC) are shown in blue font, extraintestinal pathogenic *E. coli* (ExPEC) are in purple. Based on phylogeny and genetic markers, some DAEC bacteria associated with ulcerative colitis and AIEC linked to Crohn's disease may be classified as ExPEC.

cholecystitis, cholangitis, or pancreatitis (Menichetti and Sganga 2009)—and therefore affect the entire gastrointestinal tract. Uncomplicated IAIs are characterized by inflammation of specific organ walls within the gastrointestinal tract without anatomic disruption, whereas complicated IAIs extend beyond the source organ into the peritoneum (Lopez *et al.* 2011). These can occur as postoperative complications in health care settings, or as community-onset maladies. Notably, *E. coli* are most often involved in secondary peritonitis, where microbial contamination has occurred through a perforation or necrosis of a specific gastrointestinal segment (Lopez *et al.* 2011). Complicated IAIs frequently expand from their intestinal origin to the entire body via the bloodstream, causing extraintestinal symptoms and sepsis.

Extraintestinal *E. coli* infections

Escherichia coli leads to extraintestinal disease with high frequency, and is the most reported microorganism to cause hospital-associated disease in the US between 2011 and 2017 (Weiner *et al.* 2016, Weiner-Lastinger *et al.* 2019), a statistic that includes cathether-associated UTIs, surgical site infections, central-line associated bloodstream infections and ventilator-associated pneumonia. Catheter-associated UTIs (CAUTI) represent a major complication in healthcare facilities and are reported about 100 000 times per year in the US (Weiner *et al.* 2016), most of which are caused by UPEC. Between 2015 and 2017, *E. coli* was the second most common cause of bacterial surgical site infections in the US, after *Staphylococcus aureus* (Weiner-Lastinger *et al.* 2019). Interestingly, *E. coli* causing ventilator-associated pneumonia may display lung-specific traits that distinguish them from other known ExPEC pathotypes (La Combe *et al.* 2019).

Escherichia coli is also the most frequent bacterium to cause UTIs (Foxman 2014), a disease that causes about 1% of all ambulatory hospital visits in the US. Unfortunately, UTIs can and often do result in more complicated pathologies that include frequent recurrence and progression to sepsis/bacteremia, potentially life-threathening manifestations of invasive disease. Invasive *E. coli* disease (a.k.a. invasive ExPEC disease, IED) can be defined as an acute illness consistent with systemic bacterial infection, which is microbiologically confirmed either by the isolation and identification of *E. coli* from blood or any other sterile body site, or by the isolation and identification of *E. coli* from urine in patients with urosepsis and no other identifiable source of infection. Notably, *E. coli* is also the most frequent cause of sepsis and bacteremia, and the disease has experienced a markedly increased incidence rate in the past two decades (Poolman and Anderson 2018, Rhee *et al.* 2020). It is estimated at around 600 000 cases per year in North America (Goto and Al-Hasan 2013), particularly affecting the older adult population (Martin *et al.* 2006). In the UK, *E. coli* bacteremia is estimated to occur at an incidence rate of 73.9/100 000 (Abernethy *et al.* 2015), with over two-thirds of bacteremia cases occurring in patients older than 65 years of age. Notably, over 40% of all acute cases originate from UTIs (Bou-Antoun *et al.* 2016). A global analysis of *E. coli* bacteremia incidence in high-income countries included literature published from January 2007 through March 2018 and estimated alarming incidence rates of > 100/100 000 person-years in 55-to-75-year-olds and > 300/100 000 person-years in 75-to-85-year-olds (Bonten *et al.* 2020). Notably, more than a quarter of all reported bacteremia cases were caused by *E. coli*. If treated too late or improperly, bacteremia can progress to sepsis and septic shock and result in fatality in about one out of eight cases (Bonten *et al.* 2020).

Neonatal meningitis-associated *E. coli* (NMEC or NEMEC) causes about 30% of meningitis cases, a disease with a mortality rate of 20%–60% (Ku *et al.* 2015, Bundy and Noor 2019). Furthermore, *E. coli* has been found to cause cellulitis in immunocompromised patients, necrotizing fasciitis, and surgical site infections (Petkovsek *et al.* 2009). *Escherichia coli* is the third most prevalent bacterium identified from skin and soft tissue infections (Moet *et al.* 2007). It can result in cholangitis, cholecystitis (Pitout 2012), and infections of the female reproductive tract (Cook *et al.* 2001). The bacterium has also been implicated in chronic prostatitis disease (Rudick *et al.* 2011), eye infections such as conjunctivitis in immunocom-

promised or post-traumatic patients (Teweldemedhin *et al.* 2017, Ranjith *et al.* 2020), and numerous other inflammatory diseases.

Subacute and chronic diseases associated with *E. coli*

In addition to the acute disease manifestations mentioned above, *E. coli* has also been associated with a number of non-acute pathologies that may be caused, at least in part, by prolonged and/or chronic colonization of the human host by specific *E. coli* pathotypes. These diseases include, first and foremost, colorectal cancer and inflammatory bowel disease (IBD) (Palmela *et al.* 2018, Mirsepasi-Lauridsen *et al.* 2019, Pleguezuelos-Manzano *et al.* 2020).

A possible link between *E. coli* strains producing colibactin from their *pks* locus and colorectal cancer has been proposed for the past twenty years. *Escherichia coli* strains have been isolated with very high frequencies from cancer lesions (Swidsinski *et al.* 1998, Buc *et al.* 2013). The colibactin produced from the *pks* locus alkylates DNA (Wilson *et al.* 2019) and produces double-strand DNA breaks (Nougayrede *et al.* 2006) as well as inter-strand crosslinks (Bossuet-Greif *et al.* 2018). *Pks*-positive AIEC was also shown to induce invasive carcinoma in IL-10-deficient mice (Arthur *et al.* 2012), and a specific mutational signature developed in human intestinal organoids that had been injected with *pks*$^{+}$ wild type *E. coli* but not in those injected with isogenic *pks*$^{-}$ mutant bacteria. This signature was also predominantly found in colorectal cancer genomes (Pleguezuelos-Manzano *et al.* 2020). These findings underscore the ambiguity of using probiotic strains like Nissle 1917, which is *pks*$^{+}$ (Olier *et al.* 2012, Massip *et al.* 2019), in therapeutic regimens against gastrointestinal disease. It may also be noteworthy that the *pks* locus is restricted to, but widely distributed in, *E. coli* isolates of phylogroup B2, which may comprise up to 50% of ExPEC and up to 30% of commensal *E. coli* isolates found in human hosts (Nougayrede *et al.* 2006, Putze *et al.* 2009, Dubois *et al.* 2010).

Further notable diseases increasingly correlated with *E. coli* are the two major forms of IBD, ulcerative colitis (primarily linked to diffuse adherent *E. coli*, DAEC), and Crohn's Disease (primarily associated with adherent invasive *E. coli*, AIEC). Excellent reviews summarize the current knowledge and evidence on the potential involvement of AIEC and DAEC in causation and progression of these diseases (Palmela *et al.* 2018, Mirsepasi-Lauridsen *et al.* 2019). The potential involvement of AIEC in Crohn's Disease is underlined by the finding that the bacterium is present in about 29% of Crohn's patients, a significantly higher proportion compared to healthy individuals (9%) (Nadalian *et al.* 2021). AIEC lack canonical pathotype-specific genetic biomarkers (O'Brien *et al.* 2017, Camprubi-Font *et al.* 2018), although recent advances have helped identify important genetic components that enable the bacterium to establish itself in the human gut, such as its type IV secretion system (Elhenawy *et al.* 2021), and to cause persistent inflammation. The latter is thought to be connected to its ability to metabolize short-chain fatty acids such as propionate, which may also increase its virulence potential (Ormsby *et al.* 2020, Agus *et al.* 2021), and to utilize propanediol (Viladomiu *et al.* 2021). AIEC display a close phylogenetic relationship with and a repertoire of virulence factors similar to (usually non-invasive) ExPEC (Martinez-Medina *et al.* 2009, Nash *et al.* 2010). Similarly, ulcerative colitis-associated DEAC are phylogenetically linked to ExPEC (Mirsepasi-Lauridsen *et al.* 2016, Mirsepasi-Lauridsen *et al.* 2019). Therefore, IBD-related AIEC and DAEC may phenotypically be considered intestinal pathogenic *E. coli*, while genetically being closer related to extraintestinal pathogenic *E. coli*. Moreover, DAEC have also been found to be correlated with extraintestinal pathologies such as urinary tract infections (Usein *et al.* 2001, Le Bouguenec and Servin 2006).

The pathotype concept of *E. coli* had been devised to map *E. coli* isolates onto the landscape of their pathological manifestations. This model greatly improved academic discourse about specific *E. coli* isolates and their genetic relationships, but—in light of increasing reports of hybrid strains that contain characteristics of multiple pathovars (Mariani-Kurkdjian *et al.* 2014, Nyholm *et al.* 2015, Leonard *et al.* 2016, Soysal *et al.* 2016, Lindstedt *et al.* 2018, Bai *et al.* 2019, Gati *et al.* 2019, van Hoek *et al.* 2019, Cointe *et al.* 2020, Díaz-Jiménez *et al.* 2020, Modgil *et al.* 2020, Santos *et al.* 2020, Valiatti *et al.* 2020, Nascimento *et al.* 2021)—remains manifestly inadequate. Consequently, discussions have commenced regarding the future usefulness of the pathotype delineation and its possible replacement with a terminology that is solely founded on genetic information from whole-genome sequencing (Robins-Browne *et al.* 2016). Such deliberations have to include a clear characterization of the genome of the species. How can *E. coli*'s enormous success as a commensal and pathogen be explained genetically? And are there any hints, on a genetic basis, that may enable us to improve clarity when addressing *E. coli* in clinical practice?

Core and accessory genome content of *E. coli*

The *E. coli* core genome and its phylogenetic utility

The assignment of genes into accessory (genes present in some but not all members of the species) and core (ubiquitously present genes) genome in any specific study in the literature depends on strain selection and the cutoff parameters used to determine this categorization. Consistent with its highly variable disease spectrum, the gene cloud of *E. coli* has a vast accessory component (Snipen *et al.* 2009). The number of gene families that are part of the *E. coli* accessory genome continues to grow as more isolates are sequenced, and is characterized as infinite open (Rouli *et al.* 2015). The substantial accessory genome may be rooted in the large overall population size of *E. coli* (Bobay and Ochman 2017), its wide range of ecological niches (Juhas *et al.* 2009), and/or its need for a complex array of genes when competing with other bacterial species.

For estimations of the *E. coli* core genome size, various technologies including microarrays and sequencing have yielded core genome sizes approximating between 1000 and 3000 genes (Fukiya *et al.* 2004, Willenbrock *et al.* 2007, Rasko *et al.* 2008, Chattopadhyay *et al.* 2009, Touchon *et al.* 2009, Lukjancenko *et al.* 2010, Vieira *et al.* 2011, Kaas *et al.* 2012, Land *et al.* 2015). Unlike estimations of the accessory genome size, which are collinearly related to the number of genomes studied, these core genome approximations have remained largely independent from the number of strains investigated. Whereas sequence analysis of 17 *E. coli* strains suggested a core genome of around 2200 gene families (Rasko *et al.* 2008), studies that were performed a decade later on 4400 *E. coli* genomes suggested the core to be represented by a very similar estimate of 2600 genes (Park *et al.* 2019). After interrogation of over 6000 *E. coli* genomes, 1023 genes were considered core at a strict threshold of presence in ≥ 99.9% of studied genomes (Gonzalez-Alba *et al.* 2019).

Many recent studies utilized subsets of available *E. coli* genomes for comparison, and usually arrived at slightly higher rough estimations of core genome scope. As an example, a study investigating over 300 ExPEC and APEC isolates suggested a core genome size of about 3800 genes, with an accessory component of about 13 000 genes (Jorgensen *et al.* 2019). Another study interrogating over 4000 ST131 ExPEC genomes suggested a core of about 3700 genes (Decano and Downing 2019).

The core genome of *E. coli* is, to a large degree, comprised of genes that encode essential housekeeping proteins involved in replication, transcription, and translation, and those that perform basic metabolic functions (Rasko *et al.* 2008, Leimbach *et al.* 2013). It also contains genes involved in transport processes, such as efflux pumps (Teelucksingh *et al.* 2020). Despite the impact of recombination on the frequency and character of single nucleotide polymorphisms (Castillo-Ramirez *et al.* 2011), SNP analysis in these core genes remains of great utility for phylogenetic analyses, and can provide insights into evolutionary forces acting on the organism. A high ratio of synonymous nucleotide substitutions over non-synonymous nucleotide substitutions (dS/dN) has been observed in the *E. coli* core genome, compared to the constituents of the accessory genome, which indicates that the core may have been subjected to substantial purifying selection over time (Bohlin *et al.* 2014). Intriguingly, regulatory regions of core genes have been shown to be influenced by the accessory gene content of each isolate in form of tractable compensatory regulatory mutations that enable maintenance of acquired traits such as antimicrobial resistance (McNally *et al.* 2016).

Phylogenetic studies performed on the core genome of strains from separate *E. coli* pathotypes isolated from a common source, such as the intestine, usually converge on the message that these are interrelated, and that obtained clusters often do not follow pathotype boundaries. These studies, as a rule, also tend to conclude that pathogenic InPEC lineages have evolved multiple times, on different genetic backgrounds, and converged into specific pathotypes (Pupo *et al.* 2000, Hazen *et al.* 2013, Hazen *et al.* 2016, Ingle *et al.* 2016, Denamur *et al.* 2021). In this evolutionary process, commensal *E. coli* strains may function as genetic repositories that can obtain or donate DNA from and to multiple strains, creating pathogenic isolates—and, conversely, pathogenic isolates may revert to commensalism by DNA loss or donation (Rasko *et al.* 2008). This also applies to ExPEC isolates, where phylogenetic separation of appointed ExPEC pathotypes from commensal strains of the same phylogroup is extremely difficult. UPEC isolates, for example, were indistinguishable from commensal strains based on SNP analysis of core genome regions (Nielsen *et al.* 2017). Similarly, separation of APEC strains from intestinal pathotypes remained unsupported by their phylogeny (Dziva *et al.* 2013). Unlike the separation of *E. coli* into different pathotypes based on clinical presentation, genetically (based on core genome phylogeny) the entire *E. coli* species (and *Shigella*) should be considered as one entity.

Components of the *E. coli* accessory genome

The infinite open accessory genome of *E. coli* contains a large proportion of the determinants of its pathogenic capabilities. Among these are genes that confer antibiotic resistance, virulence factors, and (likely) most of the genetic elements that may determine host preferences of the bacterium.

Antibiotic resistance genes are introduced into the bacterial genome by mobile genetic elements such as insertion sequences, transposons and integrons, or maintained as plasmids (extensively reviewed in (Partridge *et al.* 2018)). Such plasmids may and often do carry resistance determinants against a multitude of antibiotics. Resistances of *E. coli* have been described against most antibiotic classes, including (fluoro)quinolones, aminoglycosides, fosfomycin, tetracycline, phenicols, sulfonamides, and trimethoprim (Poirel *et al.* 2018). Consequently, *E. coli* has been included in the World Health Organization's list of critical (top priority) pathogens for research and development of new antibiotics (World Health Organization 2017). Of particular importance is the frequently observed resistance to beta-lactams, which is most often afforded by extended-spectrum beta-lactamases (ESBLs). Notably, the incidence of ESBL-producing *Enterobacteriaceae* increased by more than 50% between 2012 and 2017 in the US (to 57.12 cases per 10 000 hospitalizations), largely because of increased rates of community-onset cases. This change was mostly driven by a surge of ESBL-producing *E. coli* (Jernigan *et al.* 2020). ESBLs include, among others, Ambler class A (Ambler 1980) enzymes (such as $bla_{CTX\text{-}M}$, bla_{TEM}, or bla_{SHV}), and class C enzymes (such as bla_{CMY}), and worldwide, the most common ESBL gene in *E. coli* is $bla_{CTX\text{-}M\text{-}15}$. This variety was acquired on a plasmid and subsequently integrated into the chromosome in some ExPEC ST131 subclade C2 isolates. ST131 is a prominent and globally distributed multidrug-resistant clone associated with bacteremia (Pitout and DeVinney 2017), whose most successful lineages, subclades C1 and C2, diverged from their predecessor in the 1990s (Ludden *et al.* 2020, Pitout and Finn 2020). Notably, ST131 isolates of the same subclade with minimal core genome differences can have very divergent accessory genomes and carry different sets of resistance genes. Moreover, subclades C1 and C2 have different $bla_{CTX\text{-}M}$ profiles, where $bla_{CTX\text{-}M\text{-}15}$ predominates in subclade C2, whereas $bla_{CTX\text{-}M\text{-}27}$ is most often present in isolates of subclade C1 (Decano and Downing 2019). Integration of $bla_{CTX\text{-}M\text{-}15}$ into the genome may contribute to the success of ST131 C2 clones—in an Irish long-term care facility, one such isolate was shown to have replaced a previously hegemonic ST131 C2 clone whose antibiotic resistance had been maintained on a plasmid (Ludden *et al.* 2020).

In current medical practice, carbapenems are often used to treat ESBL-resistant bacteria (Rodriguez-Bano *et al.* 2018). Therefore, the emergence of carbapenem-hydrolyzing β-lactamases (such as bla_{OXA}, bla_{VIM}, bla_{NDM}, or bla_{KPC}) in *E. coli* populations is immensely alarming. Furthermore, plasmid-mediated resistance to polymyxins, such as colistin, has emerged in *E. coli* in form of its *mcr* genes (encoding phosphoethanolamine transferase, which alters the lipid A moiety of LPS (Poirel *et al.* 2017)).

A plethora of virulence factors have been identified in the *E. coli* accessory genome (Chen *et al.* 2016, Wattam *et al.* 2017, Liu *et al.* 2019). *E. coli* virulence factors encode a large variety of functions that participate in adherence, invasion, toxicity, delivery of effector molecules, and actin-based motility. These factors also play roles in defensive mechanisms, combating phagocytosis and proteolysis, conferring serum resistance, increasing resistance to stress, preventing antigenic recognition by the host, increasing iron and magnesium acquisition, and regulating other virulence determinants. Unlike antibiotic resistance genes, specific virulence factors have been associated with specific pathotypes of InPEC and have historically been used to define InPEC pathotypes (Jesser and Levy 2020). As an example, ETEC are defined by their diarrheagenic heat-labile (LT) and/or heat-stable (ST) enterotoxins, or by the presence of a large virulence plasmid that includes the colonization factor antigen I (or related factors)—and the heat-stable enterotoxin is known to contribute to an elevated severity of diarrheal disease in children < 5 years of age (Kotloff *et al.* 2013). However, expression levels of the ETEC heat-stable enterotoxin

may vary dramatically between different clinical isolates (Hazen *et al.* 2019). At least 30 colonization factors have been attributed to and characterized in ETEC isolates (Nada *et al.* 2011, Sahl *et al.* 2017, Hazen *et al.* 2019), although some isolates do not contain any of the known colonization factors. There is also considerable redundancy for some of these factors—strains may contain and express none, one, or more of these (Turner *et al.* 2006). Finally, as mentioned earlier, whole-genome sequencing continues to detect strains with hybrid complements of virulence determinants, illustrating that these bacteria exhibit a dynamic and ongoing admixture of potential virulence genes (Fleckenstein and Kuhlmann 2019).

In ExPEC isolates, the distribution of virulence factors across pathovars cannot easily be utilized in typing schemes, as canonical pathovar-specific factors have not (yet) been characterized. Moreover, for ExPEC, amount and type of virulence factors correlates only poorly with pathology. Some virulence factors are found in various pathotypes able to cause extraintestinal infections (i.e. UPEC, NMEC, SEPEC, and APEC) and in AIEC, whereas others seem specific to (but not ubiquitous in) just one of these pathotypes. No mutually exclusive genetic signatures have been identified that can be applied to differentiate these pathotypes from each other, from commensals, and from InPEC. In fact, only minor differences in the virulence factor complement exist between the probiotic commensal Nissle 1917 and the UPEC strain CFT073 (Dale and Woodford 2015). While studies have confirmed a higher level of virulence factor repertoire in UPEC, compared to commensal isolates (Nielsen *et al.* 2017, Pompilio *et al.* 2018), reliable factors that separate these two classes have not been identified. However, an imperfect signature can still be usefully applied. One existing scheme considers the presence of at least two of the five following virulence factors as indicative for likely ExPEC status: P fimbriae (*papA/H*, *papC*), S and F1C fimbriae (*sfa/focDE*), Dr antigen-specific adhesin (*afa/dra*), aerobactin *iutA*, and group 2 capsules *kpsMII* (Johnson *et al.* 2003). In a panel of ST131 isolates, this 'molecular definition of ExPEC' was the only multivariable combination identified to be significantly positively correlated with experimental disease outcome in a murine sepsis model (Merino *et al.* 2020).

Scientists are still unable to pinpoint a genetic marker set that is predictive of a specific pathogenic outcome, such as uropathogenesis (Schreiber *et al.* 2017), but have identified sets of genes with fitness phenotypes in murine UTI models (Subashchandrabose and Mobley 2015, Shea *et al.* 2020). Unfortunately, a linear correlation of ExPEC virulence potential (measured both by the presence of virulence factors and the behavior in animal models) with observed virulence in humans was not corroborated by scientific studies (Tourret and Denamur 2016). This observation could be explained by the concept of virulence being a potential by-product of commensalism where 'genes that enable *E. coli* to thrive within the intestinal tract [may] overlap with those that are responsible for the UPEC phenotype' (Tourret and Denamur 2016). Consequently, the association of virulence factors with ExPEC *E. coli* pathogenesis and pathotype designation is volatile. However, pangenome-wide association studies were recently able to identify a gene set enriched in APEC that exhibited 79 disease-associated variants, enabling computational prediction of APEC disease in random forest models with 73% accuracy (Mageiros *et al.* 2021). Application of similar and more refined approaches to other ExPEC pathotypes may finally uncover more robust pathotype-specific disease-predicting genetic elements.

Finally, identification of core and/or accessory genes that determine host-specificity in the vast genome resources available for *E. coli* has proven to be difficult. As an example, ExPEC ST131 epidemiological data show excessive rarity among environmental and veterinary isolates and a strong preference towards a human host, but phylogenetic analyses of variations in the core, regulatory and accessory genome regions suggest the strains' ability to move easily between species (McNally *et al.* 2016, Pitout and DeVinney 2017). Furthermore, a study that compared avian pathogenic *E. coli* (APEC) of ST95 with human ST95 isolates found that these strains were phylogenetically interconnected (Jorgensen *et al.* 2019). However, virulence factor combinations, phylogroups, and other genomic features may still predict, with some level of confidence, host tropism of *E. coli*. Machine learning algorithms are being developed and tested to determine the zoonotic potential and likely host of bacterial isolates (Lupolova *et al.* 2021), and in one such study, identification of bovine versus human sources was obtained with 83% accuracy for interrogated *E. coli* strains (Lupolova *et al.* 2017). Key markers included genes involved in the assembly of the K capsular subunits, a putative TonB-dependent receptor, preferentially associated with human hosts, and an L-rhamnose transporter and a toxic polypeptide HokE, preferentially associated with bovine hosts (Lupolova *et al.* 2017). The current availability of many more completed *E. coli* genome sequences, correctly annotated and complete with reliable metadata, may increase the accuracy of this or similar computational approaches. In addition, genes are likely to play unique roles in virulence in different hosts, which may leave host-specific *expression* footprints. Investigations into host-specific virulence gene requirements of *E. coli* utilized transposon-directed insertion site sequencing in an ExPEC strain of avian origin capable of causing disease in both avian and mammalian hosts. These studies revealed numerous such host-specific virulence genes (Zhang *et al.* 2019), including genes involved in biotin synthesis. Unraveling the determining factors of host preferences of the different clones and pathotypes of *E. coli* will inform better strategies to prevent its inter-species transmission and provide novel avenues to fight its ongoing success as a human pathogen.

Pathogenic strains with hybrid characteristics

Clinically relevant hybrids of *E. coli* pathotypes are constantly and increasingly discovered and characterized, and are very likely to continue to emerge and expand the current set of recognized *E. coli* pathotypes (reviewed in (Braz *et al.* 2020, Santos *et al.* 2020)). These hybrids are characterized by the presence of virulence genes that are characteristic of more than one *E. coli* pathotype, and are often identified via PCR or whole genome sequence analysis. The prominent and deadly *E. coli* outbreak in 2011 in Germany was caused by an EAEC/STEC hybrid strain, which was generated when an O104:H4 enteroaggregative *E. coli* clone acquired the Shiga-toxin-encoding lambda-like prophage, a hallmark of Shiga-toxin–producing *E. coli* (Rasko *et al.* 2011). More recent examples include the characterization of STEC/ETEC hybrids in humans (Nyholm *et al.* 2015, Leonard *et al.* 2016, Bai *et al.* 2019), swine (where they constituted 8.1% of investigated strains obtained from Spanish industrial pig farms) (García-Meniño *et al.* 2021), and wild deer (Lauzi *et al.* 2021). In addition, EPEC/ETEC hybrids were found in humans (Dutta *et al.* 2015, Hazen *et al.* 2017, Hazen *et al.* 2017), and STEC/EPEC hybrids in pigeons (van Hoek *et al.* 2019). However, *E. coli* pathovar hybrids are not restricted to mixes of InPEC pathovars—InPEC/ExPEC hybrids that contain specific hallmarks of EPEC, UPEC, NMEC and APEC are extremely prevalent in the fecal flora of humans (Lindstedt *et al.* 2018), and STEC/UPEC hybrids causing both diarrhea and urinary tract infections have also been characterized and sequenced (Gati *et al.* 2019). A recently

emerged EHEC strain (O80:H2) that produces Shiga toxin and contains determinants of a plasmid characteristic of ExPEC isolates causes hemolytic uremic syndrome with bacteremia and poses a new challenge in Europe (Mariani-Kurkdjian *et al.* 2014, Soysal *et al.* 2016, Cointe *et al.* 2020). Furthermore, a subclade of ExPEC strains causing recurrent urinary tract infections and bacteremia was recently described that acquired plasmids characteristic of the EAEC pathotype (Boll *et al.* 2020, Mandomando *et al.* 2020). In Brazil, 9/452 strains causing urinary tract infections (about 2%) were shown to contain hybrid characteristics with either EAEC or EPEC (Nascimento *et al.* 2021), while 20% of 163 strains from UTI patients in India contained virulence factors characteristic of EAEC (Modgil *et al.* 2020). A pathogenic EAEC/UPEC hybrid was detected in a panel of 67 investigated strains from UTI patients in Australia (Li *et al.* 2020). These and more examples underline the notion that the pathotype terminology currently adopted for *E. coli* may no longer be able to capture the entire spectrum of the genetic and pathogenic properties of contemporary *E. coli* isolates, and that these hybrids may further complicate clinical communication, diagnosis and treatment of *E. coli*-related disease.

Overall, the constantly occurring genetic exchange between *E. coli* isolates means that genetic markers can be transferred onto different backgrounds, resulting in new combinations of virulence determinants that may manifest in altered pathogenic properties of the isolate. This is exemplified by a reported transfer event of a multidrug-resistant plasmid between *S. sonnei* and commensal *E. coli* in the intestinal tract of the same human host (Thanh Duy *et al.* 2020), and the acquisition of Shiga toxin variants via highly dynamic Stx-encoding bacteriophages by strains with different STEC backgrounds (Ogura *et al.* 2017, Byrne *et al.* 2018). Clearly, the separation of commensal, extraintestinal pathogenic, and intestinal pathogenic *E. coli* cannot be viewed as entirely definitive and static. The surprising inability to phylogenetically and genetically reliably separate ExPEC from commensal or intestinal isolates hints at factors at play that are likely to be host- or environment-related. In addition, it requires the existence of a broad and pliable population of *E. coli*, that maintains a readily accessed diversity in its genetic determinants, under variable conditions. How is such a population maintained?

Factors affecting *E. coli* population dynamics

Maintaining the genetic diversity within *E. coli* via negative frequency-dependent selection

Frequency-dependent selection is an evolutionary concept that is over a century old. In modern genetic context, it simply means that the fitness of a phenotype or genotype depends on the overall frequency of that trait in a given population (Clarke and O'Donald 1964). It can be used to describe the internal forces that help maintain the observed genetic diversity within a species, such as *E. coli*. Why do so many accessory genes exist within the species? How does *E. coli* ensure its survival after significant perturbations, such as exposure to broad-spectrum antibiotics?

In negative frequency-dependent selection (NFDS), fitness decreases as the frequency of the specific trait increases within the population. NFDS can explain the existence of genetic diversity within a bacterial population in an ecological niche (Weeks and Hoffmann 2008, Takeuchi *et al.* 2015), as illustrated for pathogenic *E. coli* in Fig. 2. The model is perhaps best explained by the necessity to maintain serological diversity within a pathogen population, for example, through O antigen variability, to ensure evasion from evolving host defense mechanisms and/or phages. Rare O antigens provide the most advantage to the bacterial population by imparting the highest chance of not being a target for phage or host immune defenses, increasing fitness of those strains in a mixed environment. In *E. coli*, an example of this concept is illustrated by the ascent of ExPEC clone O25B:H4-ST131. A 2017 study applied NFDS to explain the observed frequencies of lineages in a large collection of ExPEC causing bloodstream infections in England between 2003 and 2012. The study confirmed that the introduction and presence of antibiotic resistance genes alone was not necessarily a good predictor of clinical success of a given *E. coli* lineage—the globally disseminated multi-drug-resistant lineage ST131 failed to become dominant in the whole population, and the highly susceptible ST73 lineage remained the most successful lineage over the years (Kallonen *et al.* 2017). NFDS provided the best explanation for the observed flux in frequencies of ExPEC sequence types. NFDS is also one possible mechanism that allows *E. coli* to maintain carriage of low-fitness cost plasmids bearing antibiotic resistance genes, even when there is no selective pressure (Dimitriu *et al.* 2019), although variable fitness effects that encompass both positive and negative net growth in different strains upon acquisition of resistance plasmids have also been observed (Alonso-Del Valle *et al.* 2021, Dunn *et al.* 2021).

An extensive population genomic analysis of the ExPEC ST131 clone found that NFDS using accessory gene loci present at intermediate frequencies successfully replicated the fluctuations observed in the ExPEC population in England from 2001–2012 (McNally *et al.* 2019). The study also found that ST131 clade A and clade B/C strains are exposed to different plasmid and phage pools, suggesting existence in different ecological niches. Notably, genes exclusive to clade B/C encompassed those that encode specific metabolic functions, including five dehydrogenase enzymes involved in anaerobic metabolism. Genes involved in anaerobic metabolism had an unusually high degree of allelic variability in clade B/C, suggesting a unique selection pressure involving competition for oxygen. The importance of these specific ecological niches in *E. coli* evolution has also been highlighted in a study that focused on commensal and environmental isolates. Here, *E. coli* evolution was also found to be strongly influenced by the environmental habitats, suggesting that acquisition of genes and mobile elements from environmental bacteria is an important mechanism in *E. coli* genome diversification. Moreover, the bacterial gene repertoires were found to evolve so quickly through high turnover of mobile genetic elements (phages, transposons, plasmids) that strains with near-identical core genomes can display a whole array of different phenotypes (Touchon *et al.* 2020).

How is this diversity maintained? NFDS likely enables *E. coli* to maintain population diversity after targeted control measures such as vaccinations—an important consideration in a bacterial species that includes commensal constituents with benefits to human health. Studies on the effect of vaccinations against *Streptococcus pneumoniae*, a bacterial pathogen that is often carried asymptomatically, were able to provide insights into the consequences of vaccines on the bacterial population diversity. Post-vaccine selection was found to form pneumococcal populations that were similar in gene frequencies, and alterations resulted in little overall effect on accessory genome composition (Croucher *et al.* 2013). In the best-fitting model of the observed population changes, relatively strong NFDS was found to act on a few hundred accessory genes that are present in 5%–95% of the given bacterial population (Corander *et al.* 2017), including antibiotic resistance determinants, restriction/modification systems, bacteriocins, nutrient import mechanisms, and genes encoding immunogenic structures. The diversity of sequence clusters within non-

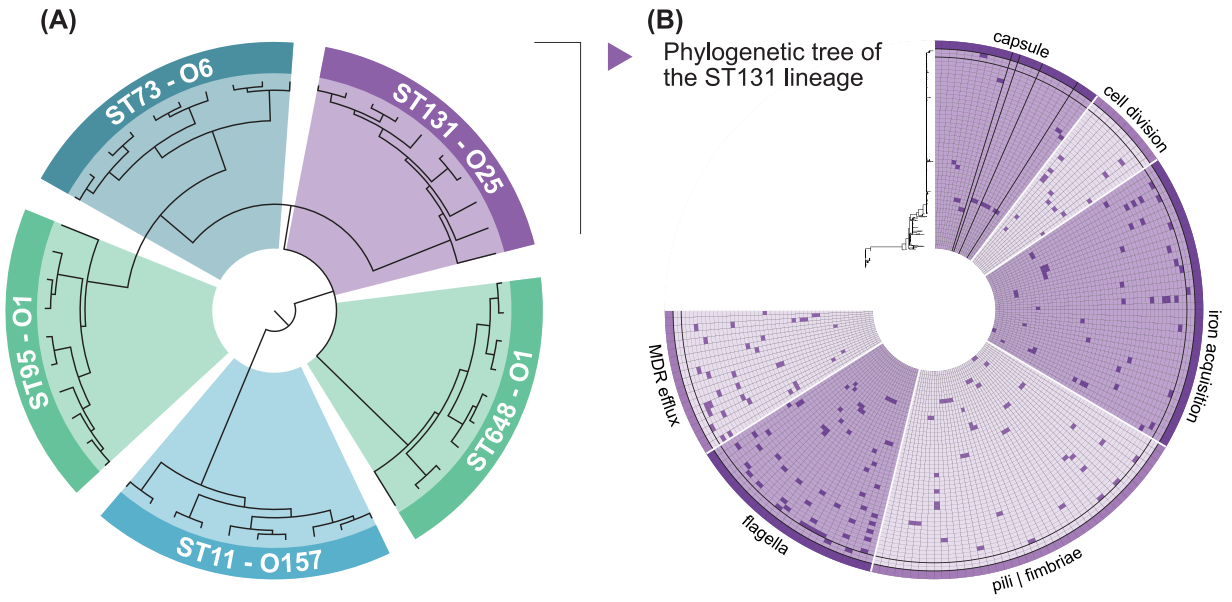

Figure 2. Schematic examples of Negative Frequency Dependent Selection (NFDS) in *E. coli* at (**A**), species population level, and (**B**), clonal lineage level. NFDS generates population diversity by maintaining uncommon phenotypes without maximizing population fitness in the current environment. **A.** Schematic phylogeny of key clonal lineages of *E. coli*. Every lineage is composed of highly successful pathogens, but each lineage is intimately associated with a single O-serotype. Note that ST73 and ST95 share an identical human ecology and yet have different O-serotypes—therefore, a successful immune response against O6 (present in ST73) would prevent ST73 from colonising, but not ST95 and other similar *E. coli* lineages. The presence of O1 in both ST95 and ST648 illustrates that an O-serotype is not restricted to a single lineage but can appear across the species. **B.** Presence of unique alleles of exemplar genes in the *E. coli* ST131 lineage. Each column denotes a gene, with its functional pathway indicated above. Unique alleles are represented by solid blocks, and are distributed across the entire phylogenetic tree at low frequency. These alleles help maintain diversity in the population and allow 'bet-hedging' within the lineage, should they confer a fitness advantage in a changing environment. If these alleles were present at a high frequency across the lineage, their benefit to the lineage would be lost.

vaccine type pneumococci temporarily spiked after introduction of vaccination, and returned back to baseline within four years (Mitchell *et al.* 2019), perhaps involving elevated recombination rates (Hanage *et al.* 2009). This stabilization was apparent both on sequence cluster scale and in gene content, and newer models were able to maintain their accuracy when running simulations that included asymmetric recombination effects (favoring gene loss over gene acquisitions), and were therefore modeling real-life evolutionary processes more accurately (Harrow *et al.* 2021).

Antibiotic use

Antibiotic resistance rates of *E. coli* pathotypes against numerous antibiotics in geographically diverse locations are rising. But does antibiotic use in human medical treatment regimens and/or in animal husbandry cause or increase resistance incidence in *E. coli*? It looks that way. An investigation of patterns of antimicrobial resistance amongst atypical EPEC strains isolated from young children in seven developing countries found that 65% of isolates were resistant to three or more drug classes (Ingle *et al.* 2018), and that resistance prevalence correlated with antibiotic use. Importantly, statistically relevant rises in resistance against ciprofloxacin due to mutations in the quinolone resistance-determining regions *gyrA* and *parC* were found to be specific to geographical locations where this antibiotic was used. However, no such correlation was observed for horizontally acquired *dfr* genes that confer resistance to trimethoprim. Scientists are also investigating whether the extensive use of antibiotics in animal husbandry contributes significantly to the rising resistance rates of circulating *E. coli* strains in human disease. A population-based modeling study shows that non-human sources can be attributed to the presence of ESBL and AmpC genes in the human population, with food, companion animals, and farm animals contributing about 30% to the spread of these genes within humans (Mughini-Gras *et al.* 2019). The use of antibiotics in animal husbandry increases the prevalence of drug resistance in the animals' microbiota, including *E. coli* (Pesciaroli *et al.* 2020, Innes *et al.* 2021), and sporadic transmission of *E. coli* isolates carrying antibiotic resistance genes from non-human sources to humans has frequently been reported (Li *et al.* 2019, Toombs-Ruane *et al.* 2020, van Hoek *et al.* 2020)). Taken together, these clues point to a sizable impact of antibiotic use on the development and maintenance of antibiotic resistance in *E. coli*.

Because of the enormous influence on treatment regimens and disease outcome, the development of antibiotic resistance in *E. coli* in various animal and human populations has been the focus of intense scientific research. Excellent reviews summarize the current state of play, and stress the high prevalence of antibiotic resistance in *E. coli* (see, for example, (Chong *et al.* 2018, Poirel *et al.* 2018)), including recent reports of high resistance rates of commensal *E. coli* strains in residential communities of low and middle income countries (Nji *et al.* 2021). Other excellent reviews shed light on the resistance rates of STEC (Mir and Kudva 2019), UPEC (Asadi Karam *et al.* 2019), or ExPEC (Le Page *et al.* 2019). 'Superbug' *E. coli* with extensive drug resistance have also been reported. One contained 32 different antibiotic resistance genes, with a lone remaining susceptibility to carbapenems (Zeng *et al.* 2019), whereas another one had amassed 68 resistance determinants, rendering the bacterium resistant to 16 different antibiotics currently in regular use, including fourth-generation cephalosporins and carbapenems (Zhang *et al.* 2019). Affected by the selective pressures from widespread use of antimicrobials, an increased maintenance of multiple resistances now occurs in some *E. coli* clones.

The emergence of plasmid-borne resistance against colistin (aka polymyxin B) conferred by the *mcr* gene is of particular importance because polymyxins represent the 'last frontier' of

antibiotics—they are currently used to treat bacterial infections that do not respond to any of the other classes of antibiotics. The first identification of plasmid-born colistin resistance in *E. coli* was reported in 2016 (Liu *et al.* 2016), and was found in China in about 1/5 of investigated chicken and pork retail meat or pigs at slaughter. Colistin has been used for decades in veterinary medicine and as a growth promoter, particularly in pigs (Kempf *et al.* 2016). Research is ongoing as to its exact mechanism of spread. *Mcr-1*-carrying isolates from clinical infections and feces of patients in a Chinese hospital and healthy volunteers in the same hospital were not closely related (Shen *et al.* 2018), and most published *mcr-1* plasmids belong to just three incompatibility groups (Matamoros *et al.* 2017, Elbediwi *et al.* 2019, Migura-Garcia *et al.* 2019). Clonally related *E. coli* isolates found in different continents were carrying the *mcr-1* gene on different plasmid backbones, suggesting a worldwide dissemination of *mcr-1* driven mainly by highly promiscuous plasmids. Moreover, there is little fitness cost of *mcr* plasmids in *E. coli*, suggesting uncomplicated maintenance of the vector within the bacterial population (Choi *et al.* 2020). At least 10 different *mcr* genes, with multiple variants, have so far been characterized (Wang *et al.* 2020, Hussein *et al.* 2021). To date, *mcr*-mediated colistin resistance is still relatively rare in *E. coli*, but it has been detected worldwide, with increasing rates—a recent study detected *mcr* genes in 80/219 surveyed ESBL-resistant *E. coli* strains present in 20 European volunteers traveling in Laos (Kantele *et al.* 2021). In addition to plasmid-encoded colistin resistance, the phenotype is also generated by chromosomal mutations in LPS production genes and two-component systems PhoPQ, PmrAB (Poirel *et al.* 2018, Taati Moghadam *et al.* 2021), and BasRS, the latter of which has been observed in *E. coli* isolates causing bacteremia (Janssen *et al.* 2020).

A second focus rests on carbapenemase-producing *E. coli*, which are now found regularly in hospital settings (Logan and Weinstein 2017, Liu *et al.* 2018, Moghnieh *et al.* 2018, Zong *et al.* 2018, Tian *et al.* 2020). The importance of tracking efforts to study their dissemination is underscored by the finding that infections with carbapenemase-producing *Enterobacteriaceae* are associated with higher mortality rates (Martin *et al.* 2018), and that 5%–10% of these carbapenem-resistant *Enterobacteriaceae* (CRE) have been found to be community-associated in the US (Kelly *et al.* 2017). Currently, five classes of carbapenemases are recognized (Durante-Mangoni *et al.* 2019), all of which have already been found in *E. coli*. Interestingly, carbapenemase and ESBL determinants are often found on separate plasmids in *E. coli*, while they are often co-transferred on the same mobile element in *Klebsiella pneumoniae* (Dunn *et al.* 2019).

A population genomics analysis of the distribution of 2172 antimicrobial-resistance genes in 4022 *E. coli* genomes identified antimicrobial combinations that are less likely to be found together, and may provide guidance for future antibiotic treatment regimens against multi-resistant *E. coli* (Goldstone and Smith 2017). Novel 'evolvable antibiotics' strategies or combination therapies that include antimicrobial peptides, probiotics, prebiotics, enzymes, vaccines, or phage therapy may also address the problem of increasing resistance.

Intra-species and inter-species competition

The population of *E. coli* is heavily influenced by the fabric of its biotic environment. Some of the genetic determinants that help *E. coli* navigate competitors and counteract host attacks represent promising targets for medical intervention, including lipopolysaccharide components, fimbriae epitopes, iron chelators, toxins, factors that prevent maturation of a host's immune system, and O-antigens, among others (Mobley and Alteri 2015, Mellata *et al.* 2016, Poolman and Wacker 2016, Mirhoseini *et al.* 2018, Barry *et al.* 2019, Duan *et al.* 2020). Many of these molecules function as virulence factors without which the bacterium is unable to cause infections. How do they contribute to maintaining *E. coli*'s population in adverse environments?

The availability of resources, particularly oxygen, iron, and energy, has profound effects on the population dynamics of the organism, and *E. coli* has developed sophisticated ways to ensure its survival in an environment stacked with bacterial competitors. Perhaps the best-known direct defense strategy is its ability to produce microcins (Telhig *et al.* 2020), small (<10 kDa) thermostable peptides, and colicins, toxic proteins that are 40–80 kDa in size. According to ADP3, an online antimicrobial peptide database (Wang *et al.* 2016), more than a third of *E. coli* isolates are microcinogenic, a much higher percentage than any other enterobacterial genus cohabitating in the human gut. Microcins are not activated by the SOS response and play a vital role in regulating microbial communities in niches such as the human gut by blocking vital functions in target cells. They primarily act on phylogenetically closely related organisms (reviewed in (Baquero *et al.* 2019)). Expression of colicins, on the other hand, is activated by the SOS response. Colicins enter the cells of colicin-sensitive members of the species via specific outer membrane complexes (Tol or Ton) and then exert their lethal action either by forming pores in the inner membrane of bacteria or via nuclease activity on genomic DNA, ribosomal RNA, or transport RNA (Cascales *et al.* 2007, Duche and Houot 2019).

Aside from this direct killing mechanism, how does *E. coli* compete for nutrients? Iron is a classic growth-limiting nutrient of most free-living bacteria. Consequently, *E. coli* has developed a plethora of ways to import iron from the environment, either directly as free iron ions or as part of molecular complexes, or via bacterial chelators, termed siderophores, that bind iron ions and are then shuttled through dedicated transport channels (Robinson *et al.* 2018). Reduced ferrous iron (Fe(II)) is imported directly by at least five transport systems in *E. coli*, encoded by the *feo, sit, mnt, efe*, and *zup* loci (Kammler *et al.* 1993, Makui *et al.* 2000, Grass *et al.* 2005, Cao *et al.* 2007). Bound iron can similarly be transported by a variety of systems that include Chu and Hma, which transport heme complexes, the Fec system that imports ferric citrate, and the Fhu transporter that can provide entry for a variety of ferrous complexes, including the siderophore aerobactin (Robinson *et al.* 2018). Import of unique *fhu* alleles via recombination represents a key evolutionary event in the formation of the emergent ST410 carbapenem-resistant *E. coli* B4/H24RxC MDR clone and results in enhanced fitness of the strain in iron-limited conditions (Feng *et al.* 2019). There are at least four *E. coli* siderophores (enterobactin, aerobactin, salmochelin and yersiniabactin) with specific roles in iron uptake. Among these, yersiniabactin was shown to be the strongest predictor for extraintestinal virulence in a mouse model of sepsis (Galardini *et al.* 2020), but may also activate host responses such as autophagy in Crohn's Disease (Dalmasso *et al.* 2021). Escherichelin, a byproduct of the yersiniabactin biosynthetic pathway in *E. coli*, has been identified as an inhibitor of iron sequestration by *Pseudomonas* and its relatives (Mislin *et al.* 2006, Robinson *et al.* 2018).

Competition for oxygen by *Enterobacteriaceae*, including *E. coli* in the human gut, is one mechanism for how a healthy microflora protects humans from *Salmonella* infections (Litvak *et al.* 2019). The importance of oxygen availability and respirational flexibility of *E. coli* has been underscored by the finding that genes in-

volved in anaerobic metabolism exhibit a higher allelic diversity in a successful *E. coli* ST131 clade than ST131 isolates of a less successful clade (McNally *et al.* 2019). In the gastrointestinal murine environment, the bacterium can quickly adapt to competitors by introducing mutations into genes involved in anaerobic respiration, a strategy that is not apparent in competitor-free gnotobiotic mice, where *E. coli* favors amino acid metabolism instead (Barroso-Batista *et al.* 2020). *E. coli*'s regulation of virulence factors and biofilm formation is tightly controlled by oxygen levels via FNR, the global regulator and central player in the adaptation of *E. coli* to variable oxygen conditions during infection (Crofts *et al.* 2018). Furthermore, cytochrome bd-I respiratory oxidase, encoded by *cydAB,* has been identified as the main contributor to nitric oxide tolerance and host colonization under microaerobic conditions (Shepherd *et al.* 2016). This enzyme has subsequently come under more scrutiny, since it is the most abundantly expressed respiratory complex in biofilm communities. Remarkably, the majority of *E. coli* in biofilms expresses this enzyme, while only about 15% of cells do so in planktonic state. Such heterogeneous expression of respiratory enzymes may ensure respiratory plasticity of *E. coli* across different host niches with variable levels of oxygenation (Beebout *et al.* 2019). Moreover, a homolog of cytochrome bd-I respiratory oxidase encoded by *appBCX* has been shown to provide a fitness advantage to *E. coli* in an inflamed environment by being able to respire the oxygen converted from reactive oxygen species such as H_2O_2 produced by the inflamed epithelium (Chanin *et al.* 2020). Finally, biofilm formation itself may be tightly regulated by oxygen levels: biofilm production by a panel of urine-associated *E. coli* isolates was reduced under all sub-atmospheric levels of oxygen, except at 4% oxygen, the level predominating in urine (Eberly *et al.* 2017).

As a generalist bacterium, *E. coli* displays generous metabolic flexibility, which allows it to thrive and compete in varied environments, including nutritionally limited milieus like the human urinary tract, where it can rapidly multiply (Reitzer and Zimmern 2019). As an example, *E. coli* can metabolize multiple types of sugars, sorbitol and galactosides, and rapidly regulate expression of the required enzymes (Maltby *et al.* 2013, Mann *et al.* 2017). This ability is essential when competing against other bacteria in the human gut (Conway and Cohen 2015) as well as during adaptations to harsher environments. Practically all of the thousands of metabolic innovations that *E. coli* has undergone during its evolution can each be transferred in a single horizontal transfer event (Pang and Lercher 2019). Other strategies that different pathotypes of *E. coli* may utilize in their battle against other commensals or pathogens that reside in or close to their ecological niche include capabilities to reduce threonine under hypoxic conditions (Letoffe *et al.* 2017), a specific capability of EHEC to assimilate aspartate (Bertin *et al.* 2018), and degradation of ethanolamine in the presence of bile salts by EIEC (Delmas *et al.* 2019).

In addition to interacting with microbial competitors, *E. coli* must also find ways to neutralize or avoid defense mechanisms of their hosts. Commensal gut bacteria are excluded from invading host tissues and reside above and in the outer mucus layer coating the epithelial cells of the gastrointestinal tract. The host has developed sophisticated ways to hyposensitize and downregulate its innate immune system responses to the presence of these commensals when these exist in their regular niche (Tanoue *et al.* 2010). However, the situation differs profoundly when pathogenic *E. coli* enter the equation, which cause epithelial barrier disruption leading to bacterial invasion of underlying tissues, and migration to the circulatory system. Strategies used to circumvent the host immune response to pathogenic invasion involves evasion of innate immunity and complement killing leading to serum resistance (Abreu and Barbosa 2017), downregulation of the inflammatory response, for example by interfering with the NFkB and MAPK pathways (Santos and Finlay 2015, de Jong and Alto 2018, Zhou *et al.* 2019), and escape from phagocytosis (Santos and Finlay 2015). In UPEC, the host's repertoire of attack mechanisms and the bacterial countermeasures result in a battle whose outcome scenarios seesaw between infection and host health (Olson and Hunstad 2016, Welch 2016, Schwab *et al.* 2017). Production of toxins such as hemolysins, adoption of an intracellular life stage, suppression of neutrophil induction via upregulation of the host's indoleamine 2,3-dioxygenase (an enzyme involved in tryptophan catabolism), and filamentation all contribute to the ability of UPEC to withstand the host's immune responses and play a role in bacterial survival. All these mechanisms are designed to increase bacterial transmission to novel hosts, but not, in essence, to cause host death.

A repertoire of virulence factors is key for *E. coli* pathogenicity and survival in the human host. However, close relatives of very infectious pathogenic isolates may still persist as harmless gut commensals, despite the presence of an array of virulence factors in their genetic footprint. It is remarkable that even today, the processes that dictate when an isolate of the most researched bacterial species on earth becomes pathogenic are not entirely understood.

Computational prediction of phenotypes and emergent lineages

The rise and global distribution of the multi-resistant ExPEC clone O25B:H4-ST131, which arose from a single progenitor strain and emerged at the beginning of the new millennium, caught scientists and healthcare personnel by surprise. It was reported as a broadly spreading clone in 2008 (Coque *et al.* 2008, Nicolas-Chanoine *et al.* 2008), and has disseminated worldwide at remarkable speed (Rogers *et al.* 2011). Among the reasons for its success may have been the fortunate confluence of the acquisition of antibiotic resistance genes (such as $bla_{CTX\text{-}M\text{-}15}$), the presence of competent virulence traits, and a serological novelty afforded by O25B that prevented a high level of pre-existing population immunity against this particular antigen. New strains are constantly identified that may be on the same distribution trajectory, such as ST648, a generalist *E. coli* clone with high virulence potential and a supreme ability to form biofilms, (Schaufler *et al.* 2019), and MS8345, a strain of the ST95 lineage that is very closely related to the highly virulent O45:K1:H4 clone associated with neonatal meningitis and carries a multitude of antibiotic resistances in addition to a plasmid linked to bacteremia (Forde *et al.* 2018). Consequently, one of the most desired goals in bacterial evolutionary biology is the ability to predict successful clones of bacterial pathogens before they become prevalent. Predictive algorithms might be getting closer to achieving this goal.

Computational modeling of bacterial behavior has made quantum leaps in recent years. One of the most relevant areas, and arguably one of the more straightforward applications, concerns the prediction of antibiotic resistance from whole-genome sequencing data (Su *et al.* 2019, Macesic *et al.* 2020, Van Camp *et al.* 2020). Notably, for some bacterial pathogens, genomic information may also be obtained in sufficient quality directly from clinical samples (Bradley *et al.* 2015), obviating the need for bacterial culture. However, in contemporary whole-genome sequencing for antimicrobial susceptibility testing, culture of the pathogen is currently (still) associated with more accurate downstream predictions, and

may remain the gold standard for computational resistance determinations for some time. A number of databases with specific focus on antibiotic resistance genes exist that aid detection of relevant information in genomic data (Xavier *et al.* 2016, Alcock *et al.* 2020, Doster *et al.* 2020, Sayers *et al.* 2020). An important advantage for the use of sequence-based technologies in antibiotic resistance determinations is the potential to be *ex post facto* mined for information on either newly identified genotypic variations or even for entirely unrelated genomic interrogations. As an example, genomic *E. coli* sequencing data obtained since 2009 in Germany were retroactively interrogated to detect previously unnoticed instances of colistin resistance in four isolates (Falgenhauer *et al.* 2016). Other scientific explorations investigated the development of antibiotic resistance during laboratory evolution under stress conditions. The phenotypic states of *E. coli* undergoing such an experiment werecoli pathotype nomenclature and analyzed and used to build a Random Forest model to identify genes that had high-feature importance in the resulting antibiotic resistance development (Maeda *et al.* 2020).

In other computational applications, metabolic consequences of genome-wide genetic alterations can be monitored in *E. coli* with increased accuracy. As examples, gene-metabolome association determinations were performed using > 3800 single-gene deletion mutants from the KEIO collection (Baba *et al.* 2006) and relative concentrations of > 7000 intracellular metabolite ions (Fuhrer *et al.* 2017). In more complex applications of computer-based modeling approaches, the time course of *E. coli* growth transitions (biomass accumulation and transcription patterns) in response to carbon upshifts and downshifts were accurately predicted with a flux-controlled regulation model that did not require knowledge on kinetics of the underlying molecular mechanisms (Erickson *et al.* 2017).

Machine-learning algorithms have been in use for genome assembly and characterization, including strain identification, genome annotation, detection of virulence and antimicrobial resistance genes, for the past decade (reviewed in (Vilne *et al.* 2019)). Progress in systems biology and computational algorithms is also increasingly applied to studies of more direct clinical relevance. Despite prior failures to correlate overall disease severity from genetic information in *E. coli* pathotypes such as EIEC or EPEC, machine-learning algorithms were employed to predict clinical outcomes of STEC infections from accessory genome information with some, though not great, accuracy (0.75) (Njage *et al.* 2019). In these algorithms, genetic predictors of poorer STEC clinical outcomes included genetic determinants for functions such as initial attachment to the host cell, LEE regulation, plasmid or genomic island maintenance, conjugative plasmid transfer and formation of sex pili, and transport across the cell envelope (Njage *et al.* 2019). Similarly, the pathogenic potential of STEC strains was very accurately predicted by an SVM-based machine-learning algorithm after evaluation of whole-genome datasets from more than 2500 isolates (Im *et al.* 2021), with > 98% of the 2292 pathogenic strains and > 96% of the 354 environmental ('non-pathogenic') strains being correctly classified. Here, few of the top 20 genes with the highest predictive values had known functions; among those that did were the antitermination protein Q, the effector protein EspFu, and the toxin LsoA (Im *et al.* 2021). Host genetic information can be similarly utilized for *E. coli* disease outcome prediction, and algorithms based on single-cell RNA sequencing data have been used for this purpose, by associating transcription signatures with specific immune cell types (Bossel Ben-Moshe *et al.* 2019). If both types of data—the measurable transcriptional reaction of the host immune system and the genetic makeup of the pathogen—can be combined in one platform, the accuracy of outcome prediction of *E. coli* infections may reach levels acceptable for routine clinical application.

Arguably, a different quality of computational algorithms is needed to forecast upcoming pathogenic bacterial strains; one that combines evolutionary models (perhaps including considerations on mixed populations) with the observation that an equilibrium of gene frequencies in the accessory genome within strains of a bacterial population is quickly maintained after strong specific disturbances (Croucher *et al.* 2013, Cremers *et al.* 2015). Sophisticated algorithms may then be able to model the effects of disturbances of the population (caused by antibiotic treatments, vaccinations, or the rise of a specific strain) and their consequences for the strain composition of the population after such a disruption. Precise genomic information on the existing strain population (pre-disturbance) would allow, through analysis of the accessory genome content of each strain, to pinpoint candidates whose ascendance would bring about the desired equilibrium. One such simplified model assumes NFDS to be the primary evolutionary force (and currently ignores alternative evasive mechanisms such as recombination). This model was able to (retroactively) correctly predict the impact of administration of a pneumococcal conjugate vaccine on strain frequency in the post-vaccine population of *S. pneumoniae* (Azarian *et al.* 2020). Refining such approaches and combining them with flux analyses may eventually result in computational models that can successfully delineate outcomes of human interventions on the composition of bacterial populations. These outcomes also depend heavily on the quality, scope, and target span of the interventions. Arguably, the most consequential and far-reaching of these are efficacious vaccines. What types of formulations have already been licensed, and what other vaccines are currently developed against *E. coli*?

Development of *E. coli* vaccines to protect against human disease

Given the importance and prevalence of *E. coli* in human disease, it is perhaps surprising that there is still no licensed vaccine for human applications to battle any intestinal pathovars of this bacterium. Current vaccine development efforts employ a variety of strategies that include both whole-cell and acellular formulations (Fig. 3).

Licensed whole-cell vaccine formulations currently only exist against a single ExPEC pathovar, UPEC, which are discussed in more detail in the section centered on vaccines to prevent urinary tract infections, below. Attenuated whole-cell approaches are also pursued for other pathovars of the bacterium, with some success (such as ACE527, a combination of three genetically attenuated and engineered ETEC strains (Harro *et al.* 2019)). But designing acellular vaccines against multiple pathovars has proven to be challenging. The search for 'universal targets' within the species has not revealed many potential candidates: the use of reverse vaccinology using thousands of draft and complete *E. coli* genomes in combination with proteomic investigations identified only four conserved surface-associated candidate genes that may represent broad-coverage vaccine antigens (Moriel *et al.* 2016). The variability of *E. coli* virulence factors, antigenic molecules, and disease manifestations dictates that a vaccine against any targeted pathovar has to either contain several components that can provide additive protection or be composed of a multi-serotype formulation in case of a dominant targeted virulence factor. *E. coli*'s mostly be-

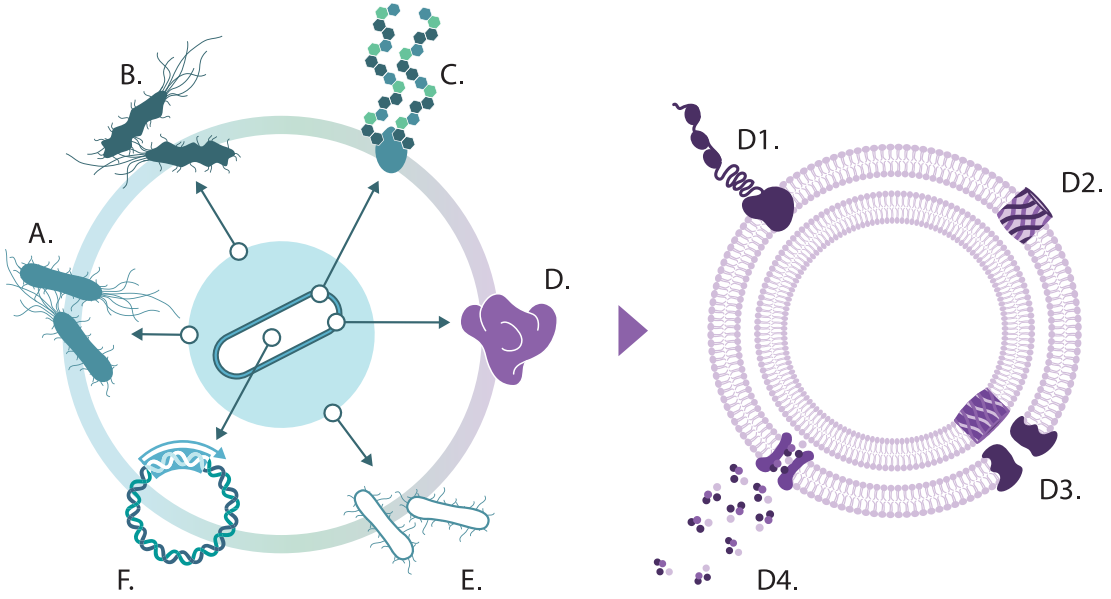

Figure 3. Selection of *E. coli* vaccine types in (pre-)clinical development. **A.** Live attenuated whole-cell vaccines (such as attenuated and engineered ETEC isolates (Harro *et al.* 2019), recombinant *Lactococci* (Sagi *et al.* 2020); or recombinant *Salmonella* (Oliveira *et al.* 2012)); **B.** Inactivated whole-cell formulations (for example, ETVAX (Qadri *et al.* 2020) and the commercially available Urovac); **C.** Glycoconjugates based on *E. coli* O antigen polysaccharides (for example, ExPEC4V (Frenck *et al.* 2019) and its expansion in current phase 3 clinical trials, ExPEC9V); **D.** Protein subunit vaccines, such as **D1.** Proteins involved in bacterial adherence to host cells (for example, CFA antigens (Maciel *et al.* 2019), fimbrial tip adhesin (Riddle *et al.* 2020), FimH (Eldridge *et al.* 2021), and SslE (Naili *et al.* 2019)), **D2.** Proteins with functions in iron acquisition (for example, IutA, Hma (Forsyth *et al.* 2020), and FyuA (Brumbaugh *et al.* 2015)), **D3.** Proteins required for immune evasion (for example, LpxR (Rojas-Lopez *et al.* 2019)); **D4.** Toxoid fusion proteins (such as CFA/ST/LT fusions (Nandre *et al.* 2018) and Stx fusions (Cai *et al.* 2011)); **E.** Bacterial ghosts/outer membrane vehicles (Cai *et al.* 2015, Leitner *et al.* 2015); **F.** DNA-based vaccines using expression vectors (Riquelme-Neira *et al.* 2015).

nign presence as a commensal in the human gut and the mild-to-moderate severity of most *E. coli* disease manifestations may also contribute to a lack of recognition of the bacterium as a dangerous pathogen, impeding public and commercial enthusiasm towards vaccine development.

Nevertheless, a large number of protective antigens against specific pathotypes of *E. coli*, such as the ETEC colonization factor antigens and heat-labile toxin (Zhang and Sack 2012), components and effectors of the EPEC type III secretion system (Horne *et al.* 2002), Shiga toxin from EHEC (Garcia-Angulo *et al.* 2013), and UPEC fimbrial adhesins and siderophore receptors (Moriel and Schembri 2013, Mobley and Alteri 2015), among others, have been investigated for vaccine development. However, even within pathovar boundaries, immunological heterogeneity makes development of broadly protective vaccines challenging. As an example, the heat-stable enterotoxin ST of ETEC remained resistant to vaccine development, although being a conserved antigen, due to its conformational similarity to the human peptide hormones guanylin and uroguanylin (Taxt *et al.* 2014). A search for common antigens characteristic of separate *E. coli* pathovars found that phylogenetic clades often drive these pathovar-associated virulence factors. Potential 'pathovar-specific' targets largely belong to one of four functional gene classes: toxins, genes involved in bacterial adherence to host cells, genes with a role in iron acquisition, and genes required for immune evasion (Clark and Maresso 2021). A detailed review of vaccine development efforts for each of these *E. coli* pathovars is beyond the scope of this review, but the following few paragraphs provide a brief overview of the current status of the field for the most prominent pathovars of the species.

Vaccine preparations against ETEC

ETEC is the most common bacterial cause of diarrhea in young children living in low- and middle-income endemic countries and in travelers visiting these areas. The long search for effective oral or transcutaneous ETEC vaccine candidates has been excellently detailed elsewhere (Zhang and Sack 2012). Five major classes of ETEC candidate vaccine strategies are recognized: attenuated whole cells, outer membrane vehicles (OMVs), vaccines based on toxins, formulations built using a bacterial autotransporter, and adhesin-based constructs (Rojas-Lopez *et al.* 2018). To date, several vaccines against ETEC are in various stages of clinical trials (reviewed in (Khalil *et al.* 2021)). The most advanced of these is an inactivated whole-cell vaccine, ETVAX, which consists of a mixture of four *E. coli* strains that each overproduce a specific adhesin (CFA/I, CS3, CS5, or CS6), and a hybrid subunit of heat-labile enterotoxin combined with the structurally closely related cholera toxin (Holmgren *et al.* 2013). In a phase I trial, ETVAX was given to Swedish adults with or without a double mutant heat-labile toxin as adjuvant, where it induced statistically significant secretory IgA and IgA antibody responses (Lundgren *et al.* 2014). The vaccine subsequently performed well in phase 1/2 trials in hundreds of Bangladeshi children, a cohort whose age approximates the envisioned vaccine target population, with age-dependent IgA responses against all or subsets of the adhesins (Qadri *et al.* 2020). A similar live-attenuated formulation of three ETEC strains, ACE527, was shown to provide strong but partial protection when co-administered with a mucosal adjuvant, the double-mutant heat-labile toxin (dmLT) of ETEC (Harro *et al.* 2019). Alternative cellular vaccine formulations in development include live attenuated *Shigella* bacteria that either express the ETEC heat-

labile enterotoxin subunit B and a detoxified version of the heat-stable enterotoxin (Harutyunyan *et al.* 2020) or the ETEC colonization factor antigen-1 (CFA/1) and the heat labile-enterotoxin (LTb) (Medeiros *et al.* 2020).

More recent ETEC subunit vaccine formulations include candidates primarily designed to prevent bacterial adhesion, either by utilizing a combination of class five fimbriae and colonization factor antigens (Maciel *et al.* 2019, Rollenhagen *et al.* 2019, Riddle *et al.* 2020) or by toxoid fusion proteins generated with a structure-based multiepitope-fusion-antigen (MEFA) platform (Duan *et al.* 2017, Nandre *et al.* 2017, Nandre *et al.* 2018, Duan *et al.* 2020). Pursued strategies further include immunization with secretory IgA raised against colonization factor antigens, with moderate success in non-human primate models (Stoppato *et al.* 2020). However, since not all ETEC express colonization factor antigens, other vaccine candidates are still actively being explored. Among the most promising candidates are the serine protease EatA, the adhesin EtpA, and the metalloprotease YghJ (Chakraborty *et al.* 2018, Kuhlmann *et al.* 2021).

EHEC and EPEC vaccines

Despite the prominence of EHEC O157:H7, there is no licensed vaccine against this pathovar for use in humans. However, fully licensed vaccines targeting O157:H7 *E. coli* are available for use in cattle. As an example, *E. coli* O157 SRP, currently distributed by Vaxxinova International, is an iron sequestering formulation that is administered in three subcutaneous jabs. The vaccine is effective (Thomson *et al.* 2009) but not widely used in feedlots—the bacterium, while causing serious disease in humans, stays predominantly asymptomatic in cattle (Gansheroff and O'Brien 2000).

While no human trial is currently registered to battle EHEC disease, many research efforts have been undertaken to identify suitable vaccine technologies and targets to battle this pathovar. These are described in more detail elsewhere (Rojas-Lopez *et al.* 2018), and include attenuated bacteria, bacterial ghosts (empty bacterial cell envelopes (Langemann *et al.* 2010)), vaccines based on the shiga toxin, and multiple others that often include presentations of intimin (Eae, an adhesin), EspA (a filament), the type 3 secretion system secretin EscC, and/or the translocated intimin receptor Tir. Chimeric proteins are also frequently utilized and explored as vaccine components, such as a bivalent antigen of EspA (the filamentous tip of a type III secretion system) fused with Tir-M (the central part of Tir) (Lin *et al.* 2017).

A homolog of the lipid A deacylase LpxR has been identified by reverse vaccinology on the EHEC O157:H7 EDL933 type strain to be a highly conserved immunogenic target that was successfully tested in a native OMV of an LpxR-overproducing strain and reduced EHEC intestinal bacterial colonization in mice (Rojas-Lopez *et al.* 2019). LomW, a putative outer membrane protein encoded in bacteriophage BP-933 W, has also recently been explored as potential target. The protein (in addition to EscC) was covalently linked to gold nanoparticles, and subcutaneous delivery to mice was able to reduce EHEC intestinal colonization 3 days post-challenge, although the effect did not remain statistically significant by day 6 (Sanchez-Villamil *et al.* 2019). In a variant application of nanoparticle technology, chitosan nanoparticles have been explored as delivery vehicles for EHEC and ETEC vaccine components Stx2 and LT, respectively (Ghaffari Marandi *et al.* 2019).

A construct of O111 polysaccharide conjugated either to cytochrome C or to EtxB (a recombinant B subunit of the heat-labile toxin LT) was able to engender antibodies in rabbits that inhibited the adhesion of O111 *E. coli* (Andrade *et al.* 2014). The use of glycoconjugates as vaccine candidates has been pursued against various other *E. coli* serotypes (Ma *et al.* 2019), including, most prominently, O157, where a conjugate with the maltose binding protein (MBP) elicited both humoral and Th_1-biased cellular responses in mice (Ma *et al.* 2014).

Vaccines to prevent urinary tract infections

Moderately successful whole-cell vaccine formulations against UPEC, the primary causative agent of urinary tract infections, exist. Uro-Vaxom (OM Pharma), first registered in 1987, consists of extracts of 18 uropathogenic *E. coli* strains prevalent in UTIs. The vaccine's efficacy in preventing UTIs, including recurrent UTIs, has been investigated in many studies at the end of the 20th century, with most showing protective effects against both disease manifestations (Tammen 1990, Schulman *et al.* 1993, Magasi *et al.* 1994). However, data robustness and trial designs have been questioned (Taha Neto *et al.* 2016), with a noted lack of long-term follow-up data beyond 6 months. More recent investigations support the product's efficacy in preventing recurrent UTIs (Bauer *et al.* 2005, Brodie *et al.* 2020), with an estimated reduction in disease reoccurrence by 34% in 12 months. Solco-Urovac (Valeant Pharmaceuticals) consists of whole-cell, heat-killed uropathogens that include six prominent *E. coli* strains, *Proteus mirabilis, Morganella morganii, Enterococcus faecalis* and *Klebsiella pneumoniae*. While showing efficacy for short-term protection, the product did not provide significant long-term protection from UTI, although *E. coli*-generated UTI occurred at significantly reduced frequency in the treated cohort compared to the one taking placebo (Hopkins *et al.* 2007, Brumbaugh and Mobley 2012). Similar cocktails have been introduced to regional markets, including Strovac (O'Brien *et al.* 2016), Urvakol, and Urostim (Mobley and Alteri 2015), displaying limited effectiveness. Finally, Uromune (Q-Pharma), a sublingual spray, is currently pre-licensed in phase 3 development and contains inactivated *E. coli, Klebsiella pneumoniae, Proteus vulgaris* and *Enterococcus faecalis*. Initial data suggest acceptable efficacy (Yang and Foley 2018, Ramirez Sevilla *et al.* 2019, Nickel *et al.* 2020), and the formulation is now scientifically explored as part of a novel therapy that also targets candidiasis (Martin-Cruz *et al.* 2020).

In addition to these bacterial cocktails (aka immunoactive agents (Magistro and Stief 2019)), intense research is currently aimed at improved non-cellular vaccine designs against UPEC. An adjuvanted UPEC-directed vaccine preparation using FimH, the adhesin protein on type 1 pili of *E. coli* necessary to colonize bladders and to form intracellular biofilms in this organ (Langermann *et al.* 1997, Martinez *et al.* 2000, Wright *et al.* 2007), exhibited a good safety profile in phase 1 trials (Eldridge *et al.* 2021). However, the development of this FimH-based *E. coli* vaccine has experienced setbacks in the past—after disappointing phase II clinical trials with women volunteers, MedImmune, Inc. announced the discontinuation of further development of the vaccine in 2003, citing lack of efficacy (Brumbaugh and Mobley 2012).

Other vaccines targeting various UPEC colonization factors are still in the preclinical stages, including formulations targeting toxins, fimbriae, and adhesins (Asadi Karam *et al.* 2019). One specific avenue to highlight is the role of iron sequesters. Iron acquisition was validated as a target for vaccination against UTI-causing UPEC in a study outlining the path towards a multivalent vaccine, where genome and proteome information from UPEC strain CFT073 was used to identify candidates that were surface localized, showed induction during growth in human urine, displayed high *in vivo* expression in experimentally infected mice

and in women with UTI, were immunogenic, and had pathogen-specificity. Four of these (IutA, IreA, FyuA, and Hma, all predicted to be involved in iron acquisition) provided significant protection from experimental infection in CBA/J mice (Mobley and Alteri 2015). Latest vaccine construct iterations center on IutA and Hma, two of the four initially investigated UPEC siderophores (Forsyth *et al.* 2020). Previously, a vaccine based on the yersiniabactin receptor FyuA was shown to be a protective vaccine target against *E. coli*-mediated pyelonephritis in a murine model of UTI (Brumbaugh *et al.* 2013, Brumbaugh *et al.* 2015). We eagerly await the maturation of these and more candidate formulations to human clinical trials.

Vaccines targeting invasive disease

Given the increasing and alarming numbers of cases of and deaths from invasive bacterial disease (bacteremia, leading to sepsis), with *E. coli* as the #1 bacterial agent (Martin *et al.* 2006, Poolman and Anderson 2018, Rhee *et al.* 2020), a vaccine that can prevent this most severe form of *E. coli* disease is highly sought after. The incidence is particularly high in the elderly, where it exceeds 300 per 100 000 people-years in adults 80 years and older, with an overall fatality rate of 12% across all ages (Bonten *et al.* 2020). For a few decades, the bacterium's O-antigen has been deemed a worthy target for exploration. O-antigen is a target of the serum's opsonophagocytic killing activity, and thereby provides a protective mechanism also exploited in other polysaccharide-based bacterial vaccines (Cross *et al.* 1994, Poolman and Wacker 2016). *E. coli* vaccines based on this premise have now been tested in clinical trials. One such glycoconjugate vaccine candidate, ExPEC4V, targets the O antigens of *E. coli* lipopolysaccharide of the four most common ExPEC serotypes observed in bacteremic infections, O1A, O2, O6A, and O25B, with a global coverage of around 40%-50% of bacteremia caused by *E. coli*. Phase 1 and phase 2 trials of this vaccine showed an excellent safety profile and good immunogenicity, with robust and functional opsonophagocytic antibody responses against all ExPEC serotypes included in the vaccine formulation, in all age groups (Huttner *et al.* 2017, Frenck *et al.* 2019). An expanded formulation, covering a broader range of O-serotypes represents approximately 58%–73% of infecting ExPEC in subjects undergoing prostate needle biopsy (Saade *et al.* 2020). While clinical phase 1/2a trials are ongoing (NCT03819049), phase 3 trials with a 9-valent vaccine (ExPEC9V) have recently started (NCT04899336), focusing on adults 60 years and older.

Conclusions

E. coli is a diverse and widely distributed species that can live commensally in the human gastrointestinal environment, persist outside of hosts in water and soil, and cause disease both in- and outside of a human or animal host's digestive tract. It has proven to be of immense utility as a model organism in research, where it has enabled the discovery of many groundbreaking mechanisms and biological processes. It is modified and harnessed in bioreactors to produce biological therapies such as vaccine antigens, insulin, blood clotting factors, and other compounds of industrial importance (such as enzymes, vitamins, biofuels). It leaves a pronounced footprint on human existence, but its ubiquitous presence and the frequency with which it causes disease elevates it to one of the most formidable microbial pathogens of mankind.

Despite the long history of *E. coli* as the subject of intensive research, the bacterial characteristics that engender the variable outcomes of infection in human hosts have not been fully elucidated. Much progress has been made, however. The sequence, functions and regulation of hundreds of *E. coli* virulence factors have been characterized, which further our understanding of how *E. coli* causes disease and how it is able to subvert the host's immune responses. Nevertheless, to this date, a whole-genome sequence of an *E. coli* isolate, when analyzed without prior knowledge of associated metadata, may not reveal the range of potential pathologies that the bacterium may cause in the host. The presence or absence of specific virulence factors does not always reveal the pathogenic capabilities and manifestations of an *E. coli* isolate. No clear genetic bacterial signatures determining *E. coli*-associated extraintestinal disease have (yet?) been identified, despite the plethora of factors known to be involved in the various pathogenic traits of ExPEC isolates. This may partly be a result of the relatively low level of curated labels of clinical manifestation in the existing large whole-genome sequence data sets. The fact that ExPEC are heavily enriched for phylogroup B2 and D isolates suggests the existence of genetic determinants of their inherent capability to cause invasive disease. Improvements in machine learning algorithms may help in future efforts to extricate these, and to identify the pathogenic potential of a specific *E. coli* isolate. However, alternative factors, from environmental circumstances and host genetics to the surrounding microbiota and population immunity, certainly also play a role in disease initiation and outcome. The bacterium's exact localization within the host, the interplay of the virulence factors with the host cells, and the host's multiple defense mechanisms likely all contribute to the clinical manifestation of an *E. coli* infection.

Importantly, analysis of an *E. coli* genome sequence also fails to reveal its *future* potential as a human pathogen. The vastness of the *E. coli* mobilome and the propensity of the bacterium to accept and transfer genetic information within different strains of the species, and between closely related species, means that the concept of separate *E. coli* pathotypes with defined clinical manifestations is not able to capture the bacterium's true versatility. This is evidenced by the frequent identification of strains with mixed genetic heritage and assortments of virulence determinants (Braz *et al.* 2020, Santos *et al.* 2020). It may also mean that pathogenic potential exists throughout the entire species of *E. coli*. Under certain conditions, *E. coli* can accept free foreign DNA (Baur *et al.* 1996, Tsen *et al.* 2002, Woegerbauer *et al.* 2002, Hasegawa *et al.* 2018), and the large footprint of horizontal gene transfer in *E. coli*'s phylogenetic history delineates its capability to adapt and change. *E. coli*'s capacity to accommodate phage, transposons, integrons, and plasmids and its formidable ability to integrate new genetic elements into existing transcriptional and translational regulatory mechanisms means that Jekyll can turn into Hyde at any turn (but may also revert to Jekyll thereafter). The bacterium's capability to transform the pathogenic profile of individual isolates by exchanging DNA poses a challenge that can perhaps best be met by viewing the entire species as a reservoir for human disease.

Over the past century, the concept of a bacterial species has evolved, and species boundaries have now been established based on whole-genome information. It is also clear that phylogeny does not unambiguously support taxonomical separation of the different pathotypes of *E. coli* (Pupo *et al.* 2000, Dziva *et al.* 2013, Hazen *et al.* 2016, Ingle *et al.* 2016, Nielsen *et al.* 2017, Denamur *et al.* 2021). Nevertheless, the pathotype concept has been extremely useful in medical, biological, and molecular research communication between scientists and experts in the field. However, based on the phylogenetic evidence, scientists have suggested abandoning the phenotypic pathotype terminology in favor of whole genome-

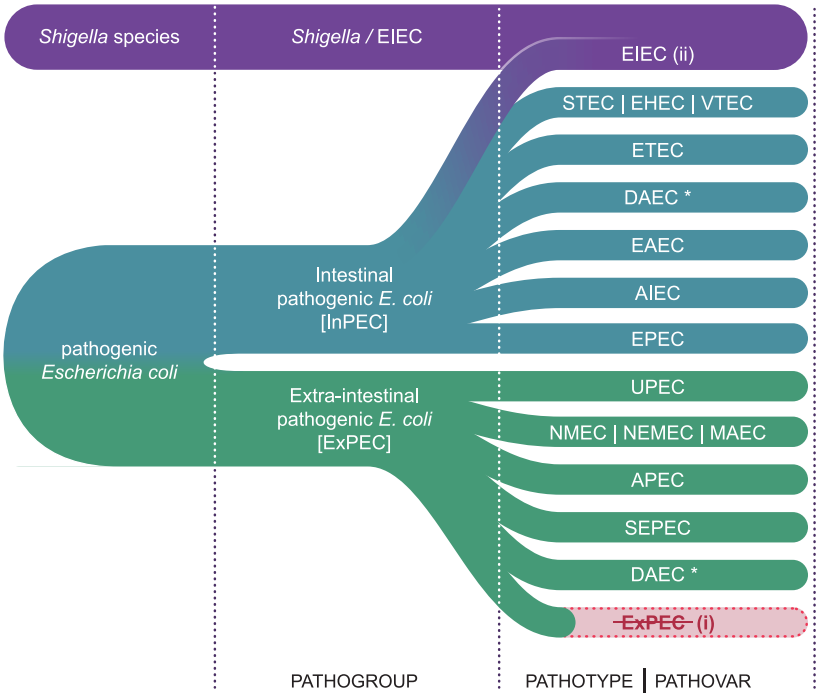

Figure 4. Suggested modifications for clinical reporting of pathogenic *E. coli* causing acute disease. Proposed alterations include (i), an exclusive use of the ExPEC acronym to delineate the *E. coli* pathogroup; and (ii), a potential categorization of pathotype EIEC with *Shigella*. For medical reporting of ExPEC, universal use of the pathogroup term rather than a separation into pathotypes may improve tracking consistency. For scientific and expert discourse, the focus may either stay on *E. coli* pathotypes or may migrate to a terminology that is based on whole-genome analyses. Novel ExPEC pathotype designations may improve precision of future scientific communication on ExPEC diseases where no current pathotype has yet been defined. See Table 1 for explanation of pathotype abbreviations. *Pathotype is associated with both intestinal and extraintestinal disease.

based concepts (Robins-Browne *et al.* 2016). This would be certain to increase specificity and precision in communications among scientists.

Unfortunately, such a genome-based terminology may be difficult to implement in routine clinical reporting, where the separation of *E. coli* into the currently used 11 different pathotypes already results in inconsistent tracking of the bacterium. The challenge to implement the most accurate reporting of *E. coli* disease while minimizing inconsistencies will be difficult to overcome with an even more resolved typing scheme that (at least in the near future) may also be too costly to be widely employed in developing economies. Arguably, if the major separations in disease outcomes of the *E. coli* species can be expressed by a simplified and intuitive terminology, then this would improve clarity for the medical community when discussing the impact of *E. coli* on human health. But what would such a terminology look like, and how can it reasonably be applied without loss of clinically important information? A discussion of these issues should continue where clinical and scientific experts each should carry equal weight.

In the specific case of ExPEC, a pathogroup that encompasses the most common causes of some of the most prevalent infectious diseases worldwide (UTIs and bacteremia/sepsis), a reduction of terms to be used in clinical reporting may be worth considering. In ExPEC, neither absence/presence patterns of specific virulence determinants nor other genomic features in the bacterium's core or accessory genome canonically define existing pathotypes. Arguably, rather than reporting UTIs, neonatal meningitis, or bacteremia under separate pathotype designations, a replacement by a unified umbrella term such as the ExPEC pathogroup may better communicate the underlying pathogen to medical practitioners and the public. Reporting all extraintestinal *E. coli* infections as caused by ExPEC may foster more awareness of the true impact of this pathogroup on human health. Confoundingly, the ExPEC acronym is currently used in scientific reports as both a path*ogroup* moniker and a path*otype* designation. The latter is usually utilized when no other pre-defined ExPEC pathotype terminology is available. This presently applies to *E. coli* isolates causing skin and soft tissue infections or ventilator-assisted pneumonia, among others. It may be worth considering a modified pathotype term (such as 'unspecified extraintestinal *E. coli*'), if these strains are discussed within the ExPEC pathogroup, and use the term ExPEC exclusively for clear delineation of the pathogroup (Fig. 4).

Regarding the separation of *Shigella* from *E. coli*, use of its historical name persists despite a near-unanimous agreement of its correct positioning within the species of *E. coli*. This has arguably created more understanding and acceptance within the public and medical practitioners than *E. coli* as a species. Rather than rename *Shigella* as EIEC, as suggested (Pettengill *et al.* 2015), perhaps EIEC should be viewed as a separate pathogroup combined with *Shigella* (Fig. 4). This integration is already applied in clinical reporting when a culture-independent clinical diagnostics test is unable to differentiate the two.

Given *E. coli*'s status as a significant public health threat, targeted research efforts to understand and combat this pathogenic species should be boosted. The search for novel antibiotics has slowed down substantially in the past decade, and this could be rectified by economic incentivization of targeted R&D activities, which should be based on product innovation and performance but unrelated to the volume of its future use (Harbarth *et al.* 2015,

Theuretzbacher *et al.* 2017). In the meantime, global implementation of antibiotic stewardship programs should help to slow the global rise in antimicrobial resistance in pathogenic bacteria such as *E. coli* (World Health Organization 2015, Barlam *et al.* 2016). Alternative therapy approaches should be furthered to keep pathogenic *E. coli* at bay. This may include probiotics, nutraceuticals, immunomodulating agents, small compounds such as pilicides, phage therapies, and vaccines (Loubet *et al.* 2020). Existing whole-cell vaccine products against pathogenic *E. coli* and ongoing explorations of acellular vaccine formulations directed against the more prominent pathotypes of the species foster hope that we will be able to better control the spread of *E. coli*'s pathogenic isolates.

Strategies to prevent *E. coli* pathologies without affecting the commensal *E. coli* population that is beneficial to the human host may be sustained by governmental incentives to non-profit researchers, by supporting academic and commercial joint ventures, and by raising more awareness to the substantial impact of the bacterium on human health. Separating pathogenic *E. coli* from commensal isolates remains one of the most important and most difficult tasks in medical research. To that end, the complex relationship between intestinal colonization and disease potential of *E. coli* isolates needs to be further characterized. As an example, long-term intestinal colonization by the highly successful pathogenic ExPEC lineage ST131 is fairly low and associated with specific risk factors (Morales Barroso *et al.* 2018, Torres *et al.* 2018). Elimination of such virulent strains from the intestinal repertoire without disturbing the commensal *E.coli* population would be very desirable. Longitudinal population-based colonization studies as well as detailed bacterial adherence analyses can provide more insight and help to achieve this goal.

The species of *E. coli* has been harnessed for many years in laboratories and bioreactors—a continued focus on the bacterium as a human pathogen may help to reign in its pathogenic potential.

Acknowledgements

Aldert Zomer is an employee of Utrecht University, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, in Utrecht, The Netherlands, providing services for Janssen Vaccines and Prevention B.V., Leiden, the Netherlands. Jan Poolman, Jeroen Geurtsen, Eveline Weerdenburg, and Mark de Been are employees of Janssen Vaccines and Prevention B.V. (former Crucell Holland BV), part of the Janssen Pharmaceutical Companies of Johnson and Johnson, Department of Bacterial Vaccines Discovery and Early Development, in Leiden, The Netherlands.

Conflict of interest. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in this review, apart from those disclosed.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Writing and editorial support was funded by Janssen Vaccines and Prevention B.V. and provided by Steffen Porwollik and Maria Thompson (APEX Think Corporation, Sheridan, WY, USA).

References

Abernethy JK, Johnson AP, Guy R *et al.* Thirty day all-cause mortality in patients with *Escherichia coli* bacteraemia in England. *Clin Microbiol Infect* 2015;**21**:251 e251–258.

Abreu AG, Barbosa AS How *Escherichia coli* circumvent complement-mediated killing. *Front Immunol* 2017;**8**:452.

Achtman M, Wagner M Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol* 2008;**6**:431–40.

Agus A, Richard D, Faïs T *et al.* Propionate catabolism by CD-associated adherent-invasive *E. coli* counteracts its anti-inflammatory effect. *Gut Microbes* 2021;**13**:1–18.

Alcock BP, Raphenya AR, Lau TTY *et al.* CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020;**48**:D517–25.

Alonso-Del Valle A, León-Sampedro R, Rodríguez-Beltrán J *et al.* Variability of plasmid fitness effects contributes to plasmid persistence in bacterial communities. *Nat Commun* 2021;**12**:2653.

Ambler RP The structure of beta-lactamases. *Philos Trans R Soc Lond B Biol Sci* 1980;**289**:321–31.

Andrade GR, New RR, Sant'Anna OA *et al.* A universal polysaccharide conjugated vaccine against O111 *E . coli*. *Hum Vaccin Immunother* 2014;**10**:2864–74.

Arthur JC, Perez-Chanona E, Muhlbauer M *et al.* Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* 2012;**338**:120–3.

Asadi Karam MR, Habibi M, Bouzari S Urinary tract infection: pathogenicity, antibiotic resistance and development of effective vaccines against uropathogenic *Escherichia coli*. *Mol Immunol* 2019;**108**:56–67.

Azarian T, Martinez PP, Arnold BJ *et al.* Frequency-dependent selection can forecast evolution in *Streptococcus pneumoniae*. *PLoS Biol* 2020;**18**:e3000878.

Baba T, Ara T, Hasegawa M *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2006;**2**:2006 0008.

Bai X, Zhang J, Ambikan A *et al.* Molecular characterization and comparative genomics of clinical hybrid shiga toxin-producing and enterotoxigenic *Escherichia coli* (STEC/ETEC) strains in Sweden. *Sci Rep* 2019;**9**:5619.

Baquero F, Lanza VF, Baquero MR *et al.* Microcins in enterobacteriaceae: peptide antimicrobials in the eco-active intestinal chemosphere. *Front Microbiol* 2019;**10**:2261.

Barlam TF, Cosgrove SE, Abbo LM *et al.* Implementing an antibiotic stewardship program: guidelines by the infectious diseases society of America and the society for healthcare epidemiology of America. *Clin Infect Dis* 2016;**62**:e51–77.

Barrios-Villa E, Martinez de la Pena CF, Lozano-Zarain P *et al.* Comparative genomics of a subset of adherent/invasive *Escherichia coli* strains isolated from individuals without inflammatory bowel disease. *Genomics* 2020;**112**:1813–20.

Barroso-Batista J, Pedro MF, Sales-Dias J *et al.* Specific eco-evolutionary contexts in the mouse gut reveal *Escherichia coli* metabolic versatility. *Curr Biol* 2020;**30**:1049–62 e1047.

Barry E, Cassels F, Riddle M *et al.* Vaccines against *Shigella* and enterotoxigenic *Escherichia coli*: a summary of the 2018 VASE conference. *Vaccine* 2019;**37**:4768–74.

Bauer HW, Alloussi S, Egger G *et al.* A long-term, multicenter, double-blind study of an *Escherichia coli* extract (OM-89) in female patients with recurrent urinary tract infections. *Eur Urol* 2005;**47**:542–8; discussion 548.

Baur B, Hanselmann K, Schlimme W *et al.* Genetic transformation in freshwater: *Escherichia coli* is able to develop natural competence. *Appl Environ Microbiol* 1996;**62**:3673–8.

Beebout CJ, Eberly AR, Werby SH *et al.*. Respiratory heterogeneityshapes biofilm formation and host colonization in uropathogenic *Escherichia coli mBio* 2019;**10**: e02400–18.

Belotserkovsky I, Sansonetti PJ *Shigella* and enteroinvasive *Escherichia coli*. *Curr Top Microbiol Immunol* 2018;**416**:1–26.

Bertin Y, Segura A, Jubelin G *et al.* Aspartate metabolism is involved in the maintenance of enterohaemorrhagic *Escherichia coli* O157:H7 in bovine intestinal content. *Environ Microbiol* 2018;**20**:4473–85.

Bliven KA, Maurelli AT Antivirulence genes: insights into pathogen evolution through gene loss. *Infect Immun* 2012;**80**:4061–70.

Blount ZD The unexhausted potential of *E . coli*. *Elife* 2015;**4**:e05826.

Bobay LM, Ochman H The evolution of bacterial genome architecture. *Front Genet* 2017;**8**:72.

Bohlin J, Brynildsrud OB, Sekse C *et al.* An evolutionary analysis of genome expansion and pathogenicity in *Escherichia coli*. *BMC Genomics* 2014;**15**:882.

Boll EJ, Overballe-Petersen S, Hasman H *et al.* Emergence of enteroaggregative *Escherichia coli* within the ST131 lineage as a cause of extraintestinal infections. *mBio* 2020;**11**:e00353–20.

Bonten M, Johnson JR, van den Biggelaar AHJ *et al.* Epidemiology of *Escherichia coli* bacteremia: a systematic literature review. *Clin Infect Dis* 2021;**72** (7):1211–9.

Bossel Ben-Moshe N, Hen-Avivi S, Levitin N *et al.* Predicting bacterial infection outcomes using single cell RNA-sequencing analysis of human immune cells. *Nat Commun* 2019;**10**:3266.

Bossuet-Greif N, Vignard J, Taieb F *et al.* The colibactin genotoxin generates DNA interstrand cross-links in infected cells. *mBio* 2018;**9**:e02393–17.

Bou-Antoun S, Davies J, Guy R *et al.* Descriptive epidemiology of *Escherichia coli* bacteraemia in England, April 2012 to March 2014. *Eurosurveillance* 2016;**21**:30329.

Bradley P, Gordon NC, Walker TM *et al.* Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat Commun* 2015;**6**:10063.

Bradley PH, Pollard KS Proteobacteria explain significant functional variability in the human gut microbiome. *Microbiome* 2017; **5**:36.

Braz VS, Melchior K, Moreira CG *Escherichia coli* as a multifaceted pathogenic and versatile bacterium. *Front Cell Infect Microbiol* 2020;**10**:548492.

Brodie A, El-Taji O, Jour I *et al.* A retrospective study of immunotherapy treatment with Uro-Vaxom (OM-89(R)) for prophylaxis of recurrent urinary tract infections. *Curr Urol* 2020;**14**: 130–4.

Brumbaugh AR, Mobley HL Preventing urinary tract infection: progress toward an effective *Escherichia coli* vaccine. *Expert Rev Vaccines* 2012;**11**:663–76.

Brumbaugh AR, Smith SN, Mobley HL Immunization with the yersiniabactin receptor, FyuA, protects against pyelonephritis in a murine model of urinary tract infection. *Infect Immun* 2013;**81**:3309–16.

Brumbaugh AR, Smith SN, Subashchandrabose S *et al.* Blocking yersiniabactin import attenuates extraintestinal pathogenic *Escherichia coli* in cystitis and pyelonephritis and represents a novel target to prevent urinary tract infection. *Infect Immun* 2015;**83**:1443–50.

Buc E, Dubois D, Sauvanet P *et al.* High prevalence of mucosa-associated *E. coli* producing cyclomodulin and genotoxin in colon cancer. *PLoS One* 2013;**8**:e56964.

Buchrieser C, Glaser P, Rusniok C *et al.* The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Mol Microbiol* 2000;**38**:760–71.

Bundy LM, Noor A *Neonatal meningitis. StatPearls [Internet]*. Treasure Island (FL). https://www.ncbi.nlm.nih.gov/books/NBK532264/

Byrne L, Dallman TJ, Adams N *et al.* Highly pathogenic clone of shiga toxin-producing *Escherichia coli* O157:H7, England and Wales. *Emerg Infect Dis* 2018;**24**:2303–8.

Cai K, Tu W, Liu Y *et al.* Novel fusion antigen displayed-bacterial ghosts vaccine candidate against infection of *Escherichia coli* O157:H7. *Sci Rep* 2015;**5**:17479.

Cai K, Gao X, Li T *et al.* Enhanced immunogenicity of a novel Stx2Am-Stx1B fusion protein in a mice model of enterohemorrhagic *Escherichia coli* O157:H7 infection. *Vaccine* 2011;**29**:946–52.

Camprubi-Font C, Lopez-Siles M, Ferrer-Guixeras M *et al.* Comparative genomics reveals new single-nucleotide polymorphisms that can assist in identification of adherent-invasive *Escherichia coli*. *Sci Rep* 2018;**8**:2695.

Cao J, Woodhall MR, Alvarez J *et al.* EfeUOB (YcdNOB) is a tripartite, acid-induced and CpxAR-regulated, low-pH Fe2+ transporter that is cryptic in *Escherichia coli* K-12 but functional in *E. coli* O157:H7. *Mol Microbiol* 2007;**65**:857–75.

Cascales E, Buchanan SK, Duche D *et al.* Colicin biology. *Microbiol Mol Biol Rev* 2007;**71**:158–229.

Castillo-Ramirez S, Harris SR, Holden MT *et al.* The impact of recombination on dN/dS within recently emerged bacterial clones. *PLoS Pathog* 2011;**7**:e1002129.

Chakraborty S, Randall A, Vickers TJ *et al.* Human experimental challenge with enterotoxigenic *Escherichia coli* elicits immune responses to canonical and novel antigens relevant to vaccine development. *J Infect Dis* 2018;**218**:1436–46.

Chang YT, Coombs G, Ling T *et al.* Epidemiology and trends in the antibiotic susceptibilities of Gram-negative bacilli isolated from patients with intra-abdominal infections in the Asia-Pacific region, 2010-2013. *Int J Antimicrob Agents* 2017;**49**:734–9.

Chanin RB, Winter MG, Spiga L *et al.* Epithelial-derived reactive oxygen species enable AppBCX-mediated aerobic respiration of *Escherichia coli* during intestinal inflammation. *Cell Host Microbe* 2020;**28**:780–788 e785.

Chattopadhyay S, Weissman SJ, Minin VN *et al.* High frequency of hotspot mutations in core genes of *Escherichia coli* due to short-term positive selection. *Proc Natl Acad Sci* 2009;**106**:12412–7.

Chaudhuri RR, Henderson IR The evolution of the *Escherichia coli* phylogeny. *Infect Genet Evol* 2012;**12**:214–26.

Chen L, Zheng D, Liu B *et al.*VFDB 2016: hierarchical and refined dataset for big data analysis–10 years on. *Nucleic Acids Res* 2016;**44**:D694–697.

Choi Y, Lee JY, Lee H *et al.* Comparison of fitness cost and virulence in chromosome- and plasmid-mediated colistin-resistant *Escherichia coli*. *Front Microbiol* 2020;**11**:798.

Chong Y, Shimoda S, Shimono N Current epidemiology, genetic evolution and clinical impact of extended-spectrum beta-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae*. *Infect Genet Evol* 2018;**61**:185–8.

Clark JR, Maresso AM Comparative pathogenomics of *Escherichia coli*: polyvalent vaccine target identification through virulome analysis. *Infect Immun* 2021;**89**:e0011521.

Clarke B, O'Donald P Frequency-dependent selection. *Heredity* 1964;**19**:201–6.

Clermont O, Christenson JK, Denamur E *et al.* The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Repor* 2013;**5**:58–65.

Clermont O, Dixit OVA, Vangchhia B *et al.* Characterization and rapid identification of phylogroup G in *Escherichia coli*, a lineage with high virulence and antibiotic resistance potential. *Environ Microbiol* 2019;**21**:3107–17.

Cointe A, Birgy A, Bridier-Nahmias A *et al. Escherichia coli* O80 hybrid pathotype strains producing shiga toxin and ESBL: molecular characterization and potential therapeutic options. *J Antimicrob Chemother* 2020;**75**:537–42.

Conway T, Cohen PS Commensal and pathogenic *Escherichia coli* metabolism in the gut. *Microbiol Spectrum* 2015;**3**: doi: 10.1128/microbiolspec.MBP-0006-2014.

Cook SW, Hammill HA, Hull RA Virulence factors of *Escherichia coli* isolated from female reproductive tract infections and neonatal sepsis. *Infect Dis Obstet Gynecol* 2001;**9**:203–7.

Coque TM, Novais A, Carattoli A *et al.* Dissemination of clonally related *Escherichia coli* strains expressing extended-spectrum beta-lactamase CTX-M-15. *Emerg Infect Dis* 2008;**14**:195–200.

Corander J, Fraser C, Gutmann MU *et al.* Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. *Nat Ecol Evol* 2017;**1**:1950–60.

Cremers AJ, Mobegi FM, de Jonge MI *et al.* The post-vaccine microevolution of invasive *Streptococcus pneumoniae*. *Sci Rep* 2015;**5**:14952.

Crick FH, Barnett L, Brenner S *et al.* General nature of the genetic code for proteins. *Nature* 1961;**192**:1227–32.

Crofts AA, Giovanetti SM, Rubin EJ *et al.* Enterotoxigenic *E. coli* virulence gene regulation in human infections. *Proc Natl Acad Sci* 2018;**115**:E8968–76.

Cross A, Artenstein A, Que J *et al.* Safety and immunogenicity of a polyvalent *Escherichia coli* vaccine in human volunteers. *J Infect Dis* 1994;**170**:834–40.

Croucher NJ, Finkelstein JA, Pelton SI *et al.* Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat Genet* 2013;**45**:656–63.

Croxen MA, Finlay BB Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol* 2010;**8**:26–38.

Croxen MA, Law RJ, Scholz R *et al.* Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin Microbiol Rev* 2013;**26**:822–80.

Dale AP, Woodford N Extra-intestinal pathogenic *Escherichia coli* (ExPEC): disease, carriage and clones. *J Infect* 2015;**71**:615–26.

Dalmasso G, Nguyen HTT, Faïs T *et al.* Yersiniabactin siderophore of Crohn's disease-associated adherent-invasive *Escherichia coli* is involved in autophagy activation in host cells. *Int J Mol Sci* 2021;**22**.

Darfeuille-Michaud A Adherent-invasive *Escherichia coli*: a putative new *E. coli* pathotype associated with crohn's disease. *Int J Med Microbiol* 2002;**292**:185–93.

Davies J, Johnson AP, Hope R Identifying hospital-onset *Escherichia coli* bacteraemia cases from english mandatory surveillance: the case for applying a two-day post-admission rule. *J Hosp Infect* 2017;**97**:207–11.

de Jong MF, Alto NM Cooperative immune suppression by *Escherichia coli* and *Shigella* effector proteins. *Infect Immun* 2018;**86**: e00560–17

DebRoy C, Fratamico PM, Yan X *et al.* Comparison of O-antigen gene clusters of all O-serogroups of *Escherichia coli* and proposal for adopting a new nomenclature for O-typing. *PLoS One* 2016;**11**:e0147434.

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

Delmas J, Gibold L, Fais T *et al.* Metabolic adaptation of adherent-invasive *Escherichia coli* to exposure to bile salts. *Sci Rep* 2019;**9**:2175.

Denamur E, Clermont O, Bonacorsi S *et al.* The population genetics of pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2021;**19**:37–54.

Díaz-Jiménez D, García-Meniño I, Herrera A *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**:doi: 10.3390/antibiotics9040192.

Dimitriu T, Medaney F, Amanatidou E *et al.* Negative frequency dependent selection on plasmid carriage and low fitness costs maintain extended spectrum beta-lactamases in *Escherichia coli*. *Sci Rep* 2019;**9**:17211.

Doster E, Lakin SM, Dean CJ *et al.* MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Res* 2020;**48**:D561–9.

Duan Q, Lee KH, Nandre RM *et al.* MEFA (multiepitope fusion antigen)-Novel technology for structural vaccinology, proof from computational and empirical immunogenicity characterization of an enterotoxigenic *Escherichia coli* (ETEC) adhesin MEFA. *J Vacc Vaccin* 2017;**08**:doi: 10.4172/2157-7560.1000367.

Duan Q, Wu W, Pang S *et al.* Coimmunization with two enterotoxigenic *Escherichia coli* (ETEC) fimbrial multiepitope fusion antigens induces the production of neutralizing antibodies against five ETEC fimbriae (F4, F5, F6, F18, and F41). *Appl Environ Microbiol* 2020;**86**:e00217–20.

Dubois D, Delmas J, Cady A *et al.* Cyclomodulins in urosepsis strains of *Escherichia coli*. *J Clin Microbiol* 2010;**48**:2122–9.

Duche D, Houot L Similarities and differences between colicin and filamentous phage uptake by bacterial cells. *EcoSal Plus* 2019;**8**:doi: 10.1128/ecosalplus.ESP-0030-2018.

Dunn S, Carrilero L, Brockhurst M *et al.* Limited and strain-specific transcriptional and growth responses to acquisition of a multidrug resistance plasmid in genetically diverse *Escherichia coli* lineages. *mSystems* 2021;**6**:e00083–21.

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.

DuPont HL, Levine MM, Hornick RB *et al.* Inoculum size in shigellosis and implications for expected mode of transmission. *J Infect Dis* 1989; **159**:1126–8.

Durante-Mangoni E, Andini R, Zampino R Management of carbapenem resistant Enterobacteriaceae infections. *Clin Microbiol Infect* 2019;**25**:943–950.

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

Dziva F, Hauser H, Connor TR *et al.* Sequencing and functional annotation of avian pathogenic *Escherichia coli* serogroup O78 strains reveal the evolution of *E. coli* lineages pathogenic for poultry via distinct mechanisms. *Infect Immun* 2013;**81**:838–49.

Eberly AR, Floyd KA, Beebout CJ *et al.* Biofilm formation by uropathogenic *Escherichia coli* is favored under oxygen conditions that mimic the bladder environment. *Int J Mol Sci* 2017;**18**:2077.

Elbediwi M, Li Y, Paudyal N *et al.* Global burden of colistin-resistant bacteria: mobilized colistin resistance genes study (1980-2018). *Microorganisms* 2019;**7**:461.

Eldridge GR, Hughey H, Rosenberger L *et al.* Safety and immunogenicity of an adjuvanted *Escherichia coli* adhesin vaccine in healthy women with and without histories of recurrent urinary tract infections: results from a first-in-human phase 1 study. *Hum Vaccin Immunother* 2021;**17**:1262–70.

Elhenawy W, Hordienko S, Gould S *et al.* High-throughput fitness screening and transcriptomics identify a role for a type IV secretion system in the pathogenesis of Crohn's disease-associated *Escherichia coli*. *Nat Commun* 2021;**12**:2032.

Erickson DW, Schink SJ, Patsalo V *et al.* A global resource allocation strategy governs growth transition kinetics of *Escherichia coli*. *Nature* 2017;**551**:119–23.

European Food Safety Authority Report from the task force on zoonoses data collection including guidance for harmonized monitoring and reporting of antimicrobial resistance in commensal *Escherichia coli* and Enterococcus spp. from food animals. *EFSA Journal* 2008;**6**:141r.

Falgenhauer L, Waezsada SE, Yao Y *et al.* Colistin resistance gene *mcr-1* in extended-spectrum beta-lactamase-producing and carbapenemase-producing Gram-negative bacteria in Germany. *Lancet Infect Dis* 2016;**16**:282–3.

Fay K, Sapiano MRP, Gokhale R *et al.* Assessment of health care exposures and outcomes in adult patients with sepsis and septic shock. *JAMA Network Open* 2020;**3**:e206004.

Feng Y, Liu L, Lin J *et al.* Key evolutionary events in the emergence of a globally disseminated, carbapenem resistant clone in the *Escherichia coli* ST410 lineage. *Commun Biol* 2019;**2**:322.

Fleckenstein JM, Kuhlmann FM Enterotoxigenic *Escherichia coli* infections. *Curr Infect Dis Rep* 2019;**21**:9.

Forde BM, Zowawi HM, Harris PNA *et al.* Discovery of *mcr-1*-mediated colistin resistance in a highly virulent *Escherichia coli* lineage. *mSphere* 2018;**3**:e00486–18.

Forsyth VS, Himpsl SD, Smith SN *et al.* Optimization of an experimental vaccine to prevent *Escherichia coli* urinary tract infection. *mBio* 2020;**11**:e00555–20.

Foxman B. Urinary tract infection syndromes: occurrence, recurrence, bacteriology, risk factors, and disease burden. *Infect Dis Clin North Am* 2014;**28**:1–13.

Frenck RW, Ervin J, Chu L *et al.* Safety and immunogenicity of a vaccine for extra-intestinal pathogenic *Escherichia coli* (ESTELLA): a phase 2 randomised controlled trial. *Lancet Infect Dis* 2019;**19**:631–40.

Friedmann HC.Escherich and *Escherichia*. *EcoSal Plus* 2014;**6**:doi: 10.1128/ecosalplus.ESP-0025-2013.

Fuhrer T, Zampieri M, Sevin DC *et al.* Genomewide landscape of gene-metabolome associations in *Escherichia coli*. *Mol Syst Biol* 2017;**13**:907.

Fukiya S, Mizoguchi H, Tobe T *et al.* Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* strains revealed by comparative genomic hybridization microarray. *J Bacteriol* 2004;**186**:3911–21.

Galardini M, Clermont O, Baron A *et al.* Major role of iron uptake systems in the intrinsic extra-intestinal virulence of the genus *Escherichia* revealed by a genome-wide association study. *PLos Genet* 2020;**16**:e1009065.

Gansheroff LJ, O'Brien AD *Escherichia coli* O157:H7 in beef cattle presented for slaughter in the U.S.: higher prevalence rates than previously estimated. *Proc Natl Acad Sci* 2000;**97**:2959–61.

Garcia-Angulo VA, Kalita A, Torres AG Advances in the development of enterohemorrhagic *Escherichia coli* vaccines using murine models of infection. *Vaccine* 2013;**31**:3229–35.

García-Meniño I, García V, Alonso MP *et al.* Clones of enterotoxigenic and shiga toxin-producing *Escherichia coli* implicated in swine enteric colibacillosis in Spain and rates of antibiotic resistance. *Vet Microbiol* 2021;**252**:108924.

Gati NS, Middendorf-Bauchart B, Bletz S *et al.* Origin and *evolution of hybrid* shiga toxin-producing *and u*ropathogenic *Escherichia coli strains of sequence type* 141. *J Clin Microbiol* 2019;**58**:e01309–19.

Gekenidis MT, Qi W, Hummerjohann J *et al.* Antibiotic-resistant indicator bacteria in irrigation water: high prevalence of extended-spectrum beta-lactamase (ESBL)-producing *Escherichia coli*. *PLoS One* 2018;**13**:e0207857.

Ghaffari Marandi BH, Zolfaghari MR, Kazemi R *et al.* Immunization against *Vibrio cholerae*, ETEC, and EHEC with chitosan nanoparticle containing LSC chimeric protein. *Microb Pathog* 2019;**134**:103600.

Goldstone RJ, Smith DGE A population genomics approach to exploiting the accessory 'resistome' of *Escherichia coli*. *Microb Genom* 2017;**3**:e000108.

Gonzalez-Alba JM, Baquero F, Canton R *et al.* Stratified reconstruction of ancestral *Escherichia coli* diversification. *BMC Genomics* 2019;**20**:936.

Goris J, Konstantinidis KT, Klappenbach JA *et al.* DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007;**57**:81–91.

Goto M, Al-Hasan MN Overall burden of bloodstream infection and nosocomial bloodstream infection in North America and Europe. *Clin Microbiol Infect* 2013;**19**:501–9.

Grass G, Franke S, Taudte N *et al.* The metal permease ZupT from *Escherichia coli* is a transporter with a broad substrate spectrum. *J Bacteriol* 2005;**187**:1604–11.

Grozdanov L, Raasch C, Schulze J *et al.* Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle 1917. *J Bacteriol* 2004;**186**:5432–41.

Gruber JS, Ercumen A, Colford JM Coliform bacteria as indicators of diarrheal risk in household drinking water: systematic review and meta-analysis. *PLoS One* 2014;**9**:e107429.

Hanage WP, Fraser C, Tang J *et al.* Hyper-recombination, diversity, and antibiotic resistance in *Pneumococcu s*. *Science* 2009;**324**:1454–7.

Harbarth S, Theuretzbacher U, Hackett J *et al.* Antibiotic research and development: business as usual? *J Antimicrob Chemother* 2015;**70**:1604–7.

Harro C, Louis Bourgeois A, Sack D *et al.* Live attenuated enterotoxigenic *Escherichia coli* (ETEC) vaccine with dmLT adjuvant protects human volunteers against virulent experimental ETEC challenge. *Vaccine* 2019;**37**:1978–86.

Harrow GL, Lees JA, Hanage WP *et al.* Negative frequency-dependent selection and asymmetrical transformation stabilise multi-strain bacterial population structures. *ISME J* 2021;**15**:1523–38.

Harutyunyan S, Neuhauser I, Mayer A *et al.* Characterization of ShigETEC, a novel live attenuated combined vaccine against *shigella*e and ETEC. *Vaccines (Basel)* 2020;**8**:689.

Hasegawa H, Suzuki E, Maeda S Horizontal plasmid transfer by transformation in *Escherichia coli*: environmental factors and possible mechanisms. *Front Microbiol* 2018;**9**:2365.

Hawkey J, Monk JM, Billman-Jacobe H *et al.* Impact of insertion sequences on convergent evolution of *Shigella* species. *PLos Genet* 2020;**16**:e1008931.

Hazen TH, Daugherty SC, Shetty AC *et al.* Transcriptional variation ofdiverse enteropathogenic Escherichia coli isolates under virulenceinducingconditions. *mSystems* 2017;**2**:e00024–17.

Hazen TH, Sahl JW, Fraser CM *et al.* Refining the pathovar paradigm via phylogenomics of the attaching and effacing *Escherichia coli*. *Proc Natl Acad Sci* 2013;**110**:12810–5.

Hazen TH, Leonard SR, Lampel KA *et al.* Investigating the relatedness of enteroinvasive *Escherichia coli* to other *E. coli* and *Shigella* isolates by using comparative genomics. *Infect Immun* 2016;**84**:2362–71.

Hazen TH, Michalski J, Luo Q *et al.* Comparative genomics and transcriptomics of *Escherichia coli* isolates carrying virulence factors of both enteropathogenic and enterotoxigenic *E . coli*. *Sci Rep* 2017;**7**:3513.

Hazen TH, Nagaraj S, Sen S *et al.* Genome and functional characterization of colonization factor antigen I- and CS6-encoding heat-stable enterotoxin-only enterotoxigenic *Escherichia coli* reveals lineage and geographic variation. *mSystems* 2019;**4**:e00329–18

Hazen TH, Donnenberg MS, Panchalingam S *et al.* Genomic diversity of EPEC associated with clinical presentations of differing severity. *Nat Microbiol* 2016;**1**:15014.

Hendriks ACA, Reubsaet FAG, Kooistra-Smid A *et al.* Genome-wide association studies of *Shigella* spp. and enteroinvasive *Escherichia coli* isolates demonstrate an absence of genetic markers for prediction of disease severity. *BMC Genomics* 2020;**21**:138.

Hershberg R, Tang H, Petrov DA Reduced selection leads to accelerated gene loss in *Shigella*. *Genome Biol* 2007;**8**:R164.

Holmgren J, Bourgeois L, Carlin N *et al.* Development and preclinical evaluation of safety and immunogenicity of an oral ETEC vaccine containing inactivated *E. coli* bacteria overexpressing colonization factors CFA/I, CS3, CS5 and CS6 combined with a hybrid LT/CT b subunit antigen, administered alone and together with dmLT adjuvant. *Vaccine* 2013;**31**:2457–64.

Hopkins WJ, Elkahwaji J, Beierle LM *et al.* Vaginal mucosal vaccine for recurrent urinary tract infections in women: results of a phase 2 clinical trial. *J Urol* 2007;**177**:1349–53; quiz 1591.

Horne C, Vallance BA, Deng W *et al.* Current progress in enteropathogenic and enterohemorrhagic *Escherichia coli* vaccines. *Expert Rev Vaccines* 2002;**1**:483–93.

Hussein NH, Al-Kadmy IMS, Taha BM *et al.* Mobilized colistin resistance (*mcr*) genes from 1 to 10: a comprehensive review. *Mol Biol Rep* 2021;**48**:2897–907.

Huttner A, Hatz C, van den Dobbelsteen G *et al.* Safety, immunogenicity, and preliminary clinical efficacy of a vaccine against extraintestinal pathogenic *Escherichia coli* in women with a history of recurrent urinary tract infection: a randomised, single-blind, placebo-controlled phase 1b trial. *Lancet Infect Dis* 2017;**17**:528–37.

Iguchi A, Iyoda S, Kikuchi T *et al.* A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res* 2015;**22**:101–7.

Im H, Hwang SH, Kim BS *et al.* Pathogenic potential assessment of the shiga toxin-producing *Escherichia coli* by a source attribution-considered machine learning model. *Proc Natl Acad Sci* 2021;**118**:e2018877118.

Ingle DJ, Levine MM, Kotloff KL *et al.* Dynamics of antimicrobial resistance in intestinal *Escherichia coli* from children in community settings in South Asia and Sub-Saharan Africa. *Nat Microbiol* 2018;**3**:1063–73.

Ingle DJ, Tauschek M, Edwards DJ *et al.* Evolution of atypical enteropathogenic *E. coli* by repeated acquisition of LEE pathogenicity island variants. *Nat Microbiol* 2016;**1**:15010.

Innes GK, Nachman KE, Abraham AG *et al.* Contamination of retail meat samples with multidrug-resistant organisms in relation to organic and conventional production and processing: a cross-sectional analysis of data from the United States National Antimicrobial Resistance Monitoring system, 2012-2017. *Environ Health Perspect* 2021;**129**:57004.

Jain C, Rodriguez RL, Phillippy AM *et al.* High throughput ANI analysis of 90 K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;**9**:5114.

Janssen AB, Bartholomew TL, Marciszewska NP *et al.* Nonclonal emergence of colistin resistance associated with mutations in the BasRS two-component system in *Escherichia coli* bloodstream isolates. *mSphere* 2020;**5**:e00143–20.

Jernigan JA, Hatfield KM, Wolford H *et al.* Multidrug-resistant bacterial infections in U.S. hospitalized patients, 2012-2017. *N Engl J Med* 2020;**382**:1309–19.

Jesser KJ, Levy K Updates on defining and detecting diarrheagenic *Escherichia coli* pathotypes. *Curr Opin Infect Dis* 2020;**33**:372–80.

Johnson JR, Murray AC, Gajewski A *et al.* Isolation and molecular characterization of nalidixic acid-resistant extraintestinal pathogenic *Escherichia coli* from retail chicken products. *Antimicrob Agents Chemother* 2003;**47**:2161–8.

Jorgensen SL, Stegger M, Kudirkiene E *et al.* Diversity and population overlap between avian and human *Escherichia coli* belonging to sequence type 95. *mSphere* 2019;**4**: e00333-18.

Juhas M, van der Meer JR, Gaillard M *et al.* Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* 2009;**33**:376–93.

Kaas RS, Friis C, Ussery DW *et al.* Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genomics* 2012;**13**:577.

Kallonen T, Brodrick HJ, Harris SR *et al.* Systematic longitudinal survey of invasive *Escherichia coli* in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. *Genome Res* 2017;**27**:1437–49.

Kammler M, Schon C, Hantke K Characterization of the ferrous iron uptake system of *Escherichia coli*. *J Bacteriol* 1993;**175**:6212–9.

Kantele A, Kuenzli E, Dunn SJ *et al.* Dynamics of intestinal multidrug-resistant bacteria colonisation contracted by visitors to a high-endemic setting: a prospective, daily, real-time sampling study. *Lancet Microbe* 2021;**2**:e151–8.

Kaper JB, Nataro JP, Mobley HL Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2004;**2**:123–40.

Kelly AM, Mathema B, Larson EL Carbapenem-resistant Enterobacteriaceae in the community: a scoping review. *Int J Antimicrob Agents* 2017;**50**:127–34.

Kempf I, Jouy E, Chauvin C Colistin use and colistin resistance in bacteria from animals. *Int J Antimicrob Agents* 2016;**48**:598–606.

Khalil I, Walker R, Porter CK *et al.* Enterotoxigenic *Escherichia coli* (ETEC) vaccines: priority activities to enable product development, licensure, and global access. *Vaccine* 2021;**39**: 4266–4277.

Konstantinidis KT, Ramette A, Tiedje JM The bacterial species definition in the genomic era. *Philos Trans Royal Soc B: Biolog Sci* 2006;**361**:1929–40.

Kotloff KL, Riddle MS, Platts-Mills JA *et al.* Shigellosis. *Lancet North Am Ed* 2018;**391**:801–12.

Kotloff KL, Nasrin D, Blackwelder WC *et al.* The incidence, aetiology, and adverse clinical consequences of less severe diarrhoeal episodes among infants and children residing in low-income and middle-income countries: a 12-month case-control study as a follow-on to the global enteric multicenter study (GEMS). *Lancet Glob Health* 2019;**7**:e568–84.

Kotloff KL, Nataro JP, Blackwelder WC *et al.* Burden and aetiology of diarrhoeal disease in infants and young children in devel-

oping countries (the global enteric multicenter study, GEMS): a prospective, case-control study. *Lancet North Am Ed* 2013;**382**: 209–22.

Ku LC, Boggess KA, Cohen-Wolkowiez M Bacterial meningitis in infants. *Clin Perinatol* 2015;**42**:29–45, vii-viii.

Kuhlmann FM, Laine RO, Afrin S *et al.* Contribution of noncanonical antigens to virulence and adaptive immunity in human infection with enterotoxigenic *E. coli*. *Infect Immun* 2021;**89**: e00041–21.

La Combe B, Clermont O, Messika J *et al.* Pneumonia-specific *Escherichia coli* with distinct phylogenetic and virulence profiles, France, 2012-2014. *Emerg Infect Dis* 2019;**25**:710–8.

Lan R, Reeves PR *Escherichia coli* in disguise: molecular origins of *Shigell a*. *Microbes Infect* 2002;**4**:1125–32.

Land M, Hauser L, Jun SR *et al.* Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics* 2015;**15**:141–61.

Langemann T, Koller VJ, Muhammad A *et al.* The bacterial ghost platform system: production and applications. *Bioengineered Bugs* 2010;**1**:326–36.

Langermann S, Palaszynski S, Barnhart M *et al.* Prevention of mucosal *Escherichia coli* infection by fimh-adhesin-based systemic vaccination. *Science* 1997;**276**:607–11.

Langley G, Schaffner W, Farley MM *et al.* Twenty years of active bacterial core surveillance. *Emerg Infect Dis* 2015;**21**:1520–8.

Laupland KB, Church DL Population-based epidemiology and microbiology of community-onset bloodstream infections. *Clin Microbiol Rev* 2014;**27**:647–64.

Lauzi S, Luzzago C, Chiani P *et al.* Free-ranging red deer (*Cervus elaphus*) as carriers of potentially zoonotic shiga toxin-producing *Escherichia coli*. *Transbound Emerg Dis* 2022;**69**:1902–11.

Le Bouguenec C, Servin AL Diffusely adherent *Escherichia coli* strains expressing Afa/Dr adhesins (Afa/Dr DAEC): hitherto unrecognized pathogens. *FEMS Microbiol Lett* 2006;**256**:185–94.

Le Page S, Dubourg G, Baron SA *et al.* No global increase in resistance to antibiotics: a snapshot of resistance from 2001 to 2016 in Marseille, France. *Eur J Clin Microbiol Infect Dis* 2019;**38**:395–407.

Lehman IR, Bessman MJ, Simms ES *et al.* Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from *Escherichia coli*. *J Biol Chem* 1958;**233**:163–70.

Leimbach A, Hacker J, Dobrindt U *E. coli* as an all-rounder: the thin line between commensalism and pathogenicity. *Curr Top Microbiol Immunol* 2013;**358**:3–32.

Leitner DR, Lichtenegger S, Temel P *et al.* A combined vaccine approach against *Vibrio cholerae* and ETEC based on outer membrane vesicles. *Front Microbiol* 2015;**6**:823.

Lemaitre C, Bidet P, Bingen E *et al.* Transcriptional analysis of the *Escherichia coli* Colv-Ia plasmid pS88 during growth in human serum and urine. *BMC Microbiol* 2012;**12**:115.

Leonard SR, Mammel MK, Rasko DA *et al.* Hybrid shiga toxin-producing and enterotoxigenic *Escherichia* sp. cryptic lineage 1 strain 7v harbors a hybrid plasmid. *Appl Environ Microbiol* 2016;**82**:4309–19.

Letoffe S, Chalabaev S, Dugay J *et al.* Biofilm microenvironment induces a widespread adaptive amino-acid fermentation pathway conferring strong fitness advantage in *Escherichia coli*. *PLos Genet* 2017;**13**:e1006800.

Li D, Reid CJ, Kudinha T *et al.* Genomic analysis of trimethoprim-resistant extraintestinal pathogenic *Escherichia coli* and recurrent urinary tract infections. *Microb Genom* 2020;**6**:mgen000475.

Li J, Bi Z, Ma S *et al.* Inter-host transmission of carbapenemase-producing *Escherichia coli* among humans and backyard animals. *Environ Health Perspect* 2019;**127**:107009.

Lin R, Zhu B, Zhang Y *et al.* Intranasal immunization with novel EspA-Tir-M fusion protein induces protective immunity against enterohemorrhagic *Escherichia coli* O157:H7 challenge in mice. *Microb Pathog* 2017;**105**:19–24.

Lindstedt BA, Finton MD, Porcellato D *et al.* High frequency of hybrid *Escherichia coli* strains with combined intestinal pathogenic *Escherichia coli* (IPEC) and extraintestinal pathogenic *Escherichia coli* (ExPEC) virulence factors isolated from human faecal samples. *BMC Infect Dis* 2018;**18**:544.

Linn S, Arber W Host specificity of DNA produced by *Escherichia coli*, x. In vitro restriction of phage fd replicative form. *Proc Natl Acad Sci* 1968;**59**:1300–6.

Litvak Y, Mon KKZ, Nguyen H *et al.* Commensal Enterobacteriaceae protect against *Salmonella* colonization through oxygen competition. *Cell Host and Microbe* 2019;**25**:128–139 e125.

Liu B, Zheng D, Jin Q *et al.* VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res* 2019;**47**:D687–92.

Liu B, Furevi A, Perepelov AV *et al.* Structure and genetics of Escherichiacoli Oantigens *FEMS Microbiol Rev* 2020;**44**:655–83.

Liu L, Feng Y, McNally A *et al. blaNDM-21*, a new variant of *blaNDM* in an *Escherichia coli* clinical isolate carrying *blaCTX-M-55* and *rmtB*. *J Antimicrob Chemother* 2018;**73**:2336–9.

Liu S, Jin D, Lan R *et al. Escherichia marmotae* sp. nov., isolated from faeces of *Marmota himalayana*. *Int J Syst Evol Microbiol* 2015;**65**: 2130–4.

Liu YY, Wang Y, Walsh TR *et al.* Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *Lancet Infect Dis* 2016;**16**:161–8.

Lodinova-Zadnikova R, Cukrowska B, Tlaskalova-Hogenova H Oral administration of probiotic *Escherichia coli* after birth reduces frequency of allergies and repeated infections later in life (after 10 and 20 years). *Intern Arch Allerg Immunol* 2003;**131**:209–11.

Logan LK, Weinstein RA The epidemiology of carbapenem-resistant Enterobacteriaceae: the impact and evolution of a global menace. *J Infect Dis* 2017;**215**:S28–36.

Lopez N, Kobayashi L, Coimbra R A comprehensive review of abdominal infections. *World J Emerg Surg* 2011;**6**:7.

Loubet P, Ranfaing J, Dinh A *et al.* Alternative therapeutic options to antibiotics for the treatment of urinary tract infections. *Front Microbiol* 2020;**11**:1509.

Lozupone CA, Stombaugh JI, Gordon JI *et al.* Diversity, stability and resilience of the human gut microbiota. *Nature* 2012;**489**: 220–30.

Lu S, Jin D, Wu S *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 Microb Infect* 2016;**5**:e122.

Ludden C, Decano AG, Jamrozy D *et al.* Genomic surveillance of *Escherichia coli* ST131 identifies local expansion and serial replacement of subclones. *Microbial Genomics* 2020;**6**: e000352.

Lukjancenko O, Wassenaar TM, Ussery DW Comparison of 61 sequenced *Escherichia coli* genomes. *Microb Ecol* 2010;**60**:708–20.

Lundgren A, Bourgeois L, Carlin N *et al.* Safety and immunogenicity of an improved oral inactivated multivalent enterotoxigenic *Escherichia coli* (ETEC) vaccine administered alone and together with dmLT adjuvant in a double-blind, randomized, placebo-controlled phase i study. *Vaccine* 2014;**32**:7077–84.

Lupolova N, Chalka A, Gally DL Predicting host association for shiga toxin-producing *E. coli* serogroups by machine learning. *Methods Mol Biol* 2021;**2291**:99–117.

Lupolova N, Dallman TJ, Holden NJ *et al.* Patchy promiscuity: machine learning applied to predict the host specificity of *Salmonella enterica* and *Escherichia coli*. *Microb Genom* 2017;**3**:e000135.

Ma Z, Zhang H, Li L *et al.* Direct cloning of bacterial surface polysaccharide gene cluster for one-step production of glycoconjugate vaccine. *ACS Infect Dis* 2019;**5**:74–78.

Ma Z, Zhang H, Shang W *et al.* Glycoconjugate vaccine containing *Escherichia coli* O157:H7 O-antigen linked with maltose-binding protein elicits humoral and cellular responses. *PLoS One* 2014;**9**:e105215.

Macesic N, Bear Don't Walk OI, Pe'er I *et al.* Predicting phenotypic polymyxin resistance in *Klebsiella pneumoniae* through machine learning analysis of genomic data. *mSystems* 2020;**5**: e00656-19.

Maciel M, Bauer D, Baudier RL *et al.* Intradermal or sublingual delivery and heat-labile enterotoxin proteins shape immunologic responses to a CFA/I fimbria-derived subunit antigen vaccine against enterotoxigenic *Escherichia coli*. *Infect Immun* 2019;**87**.

Maeda T, Iwasawa J, Kotani H *et al.* High-throughput laboratory evolution reveals evolutionary constraints in *Escherichia coli*. *Nat Commun* 2020;**11**:5970.

Magasi P, Panovics J, Illes A *et al.* Uro-Vaxom and the management of recurrent urinary tract infection in adults: a randomized multicenter double-blind trial. *Eur Urol* 1994;**26**:137–40.

Mageiros L, Méric G, Bayliss SC *et al.* Genome evolution and the emergence of pathogenicity in avian *Escherichia coli*. *Nat Commun* 2021;**12**:765.

Magistro G, Stief CG Vaccine development for urinary tract infections: where do we stand? *Euro Urol Focus* 2019;**5**:39–41.

Makui H, Roig E, Cole ST *et al.* Identification of the *Escherichia coli* K-12 Nramp orthologue (MntH) as a selective divalent metal ion transporter. *Mol Microbiol* 2000;**35**:1065–78.

Maltby R, Leatham-Jensen MP, Gibson T *et al.* 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**:e53957.

Mandomando I, Vubil D, Boisen N *et al.* *Escherichia coli* ST131 clones harbouring AggR and AAF/V fimbriae causing bacteremia in Mozambican children: emergence of new variant of *fimH27* subclone. *PLoS NeglTrop Dis* 2020;**14**:e0008274.

Manges AR, Geum HM, Guo A *et al.* Global extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages. *Clin Microbiol Rev* 2019;**32**: e00135-18.

Mann R, Mediati DG, Duggin IG *et al.* Metabolic adaptations of uropathogenic *E. coli* in the urinary tract. *Front Cell Infect Microbiol* 2017;**7**:241.

Mariani-Kurkdjian P, Lemaitre C, Bidet P *et al.* Haemolytic-uraemic syndrome with bacteraemia caused by a new hybrid *Escherichia coli* pathotype. *New Microbes New Infect* 2014;**2**:127–31.

Martin-Cruz L, Sevilla-Ortega C, Benito-Villalvilla C *et al.* A combination of polybacterial MV140 and *Candida albicans* V132 as a potential novel trained immunity-based vaccine for genitourinary tract infections. *Front Immunol* 2020;**11**:612269.

Martin A, Fahrbach K, Zhao Q *et al.* Association between carbapenem resistance and mortality among adult, hospitalized patients with serious infections due to Enterobacteriaceae: results of a systematic literature review and meta-analysis. *Open Forum Infectious Diseases* 2018;**5**:ofy150.

Martin GS, Mannino DM, Moss M The effect of age on the development and outcome of adult sepsis. *Crit Care Med* 2006;**34**:15–21.

Martinez-Medina M, Mora A, Blanco M *et al.* Similarity and divergence among adherent-invasive *Escherichia coli* and extraintestinal pathogenic *E. coli* strains. *J Clin Microbiol* 2009;**47**:3968–79.

Martinez JJ, Mulvey MA, Schilling JD *et al.* Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. *EMBO J* 2000;**19**:2803–12.

Massip C, Branchu P, Bossuet-Greif N *et al.* Deciphering the interplay between the genotoxic and probiotic activities of *Escherichia coli* Nissle 1917. *PLoS Pathog* 2019;**15**:e1008029.

Matamoros S, van Hattem JM, Arcilla MS *et al.* Global phylogenetic analysis of *Escherichia coli* and plasmids carrying the *mcr-1* gene indicates bacterial diversity but plasmid restriction. *Sci Rep* 2017;**7**:15364.

McNally A, Kallonen T, Connor C *et al.* Diversification of colonization factors in a multidrug-resistant *Escherichia coli* lineage evolving under negative frequency-dependent selection. *mBio* 2019;**10**:e00644–19.

McNally A, Oren Y, Kelly D *et al.* Combined analysis of variation in core, accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. *PLos Genet* 2016;**12**:e1006280.

Medeiros P, Bolick DT, Ledwaba SE *et al.* A bivalent vaccine confers immunogenicity and protection against *Shigella flexneri* and enterotoxigenic *Escherichia coli* infections in mice. *NPJ Vaccines* 2020;**5**:30.

Mellata M, Mitchell NM, Schodel F *et al.* Novel vaccine antigen combinations elicit protective immune responses against *Escherichia coli* sepsis. *Vaccine* 2016;**34**:656–62.

Menichetti F, Sganga G Definition and classification of intra-abdominal infections. *J Chemother* 2009;**21**:3–4.

Meric G, Hitchings MD, Pascoe B *et al.* From Escherich to the *Escherichia coli* genome. *Lancet Infect Dis* 2016;**16**:634–6.

Merino I, Porter SB, Johnston B *et al.* Molecularly defined extraintestinal pathogenic *Escherichia coli* status predicts virulence in a murine sepsis model better than does virotype, individual virulence genes, or clonal subset among *E. coli* ST131 isolates. *Virulence* 2020;**11**:327–36.

Migura-Garcia L, Gonzalez-Lopez JJ, Martinez-Urtaza J *et al.* *mcr*-colistin resistance genes mobilized by IncX4, incHI2, and incI2 plasmids in *Escherichia coli* of pigs and white stork in Spain. *Front Microbiol* 2019;**10**:3072.

Mir RA, Kudva IT Antibiotic-resistant shiga toxin-producing *Escherichia coli*: an overview of prevalence and intervention strategies. *Zoonoses Public Health* 2019;**66**:1–13.

Mirhoseini A, Amani J, Nazarian S Review on pathogenicity mechanism of enterotoxigenic *Escherichia coli* and vaccines against it. *Microb Pathog* 2018;**117**:162–9.

Mirsepasi-Lauridsen HC, Vallance BA, Krogfelt KA *et al.* *Escherichia coli* pathobionts associated with inflammatory bowel disease. *Clin Microbiol Rev* 2019;**32**:e00060–18.

Mirsepasi-Lauridsen HC, Halkjaer SI, Mortensen EM *et al.* Extraintestinal pathogenic *Escherichia coli* are associated with intestinal inflammation in patients with ulcerative colitis. *Sci Rep* 2016;**6**:31152.

Mislin GL, Hoegy F, Cobessi D *et al.* Binding properties of pyochelin and structurally related molecules to FptA of *Pseudomonas aeruginosa*. *J Mol Biol* 2006;**357**:1437–48.

Mitchell PK, Azarian T, Croucher NJ *et al.* Population genomics of pneumococcal carriage in Massachusetts children following introduction of PCV-13. *Microb Genom* 2019;**5**.e000252

Mobley HL, Alteri CJ Development of a vaccine against *Escherichia coli* urinary tract infections. *Pathogens* 2015;**5**: e5010001.

Modgil V, Kaur H, Mohan B *et al.* Molecular, phylogenetic and antibiotic resistance analysis of enteroaggregative *Escherichia coli*/uropathogenic *Escherichia coli* hybrid genotypes causing urinary tract infections. *Indian J Med Microbiol* 2020;**38**:421–9.

Moet GJ, Jones RN, Biedenbach DJ *et al.* Contemporary causes of skin and soft tissue infections in North America, Latin America, and Europe: report from the SENTRY antimicrobial surveillance program (1998-2004). *Diagn Microbiol Infect Dis* 2007;**57**: 7–13.

Moghnieh RA, Kanafani ZA, Tabaja HZ *et al.* Epidemiology of common resistant bacterial pathogens in the countries of the Arab league. *Lancet Infect Dis* 2018;**18**:e379–94.

Monk JM, Charusanti P, Aziz RK *et al.* Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci* 2013;**110**:20338–43.

Morales Barroso I, Lopez-Cerero L, Navarro MD *et al.* Intestinal colonization due to *Escherichia coli* ST131: risk factors and prevalence. *Antimicr Resist Infect Control* 2018;**7**:135.

Moreno MA, Dominguez L, Teshager T *et al.* Antibiotic resistance monitoring: the Spanish programme. The VAV network. Red de vigilancia de resistencias antibioticas en bacterias de origen veterinario. *Int J Antimicrob Agents* 2000;**14**:285–90.

Moriel DG, Schembri MA Vaccination approaches for the prevention of urinary tract infection. *Curr Pharm Biotechnol* 2013;**14**:967–74.

Moriel DG, Tan L, Goh KG *et al.* A novel protective vaccine antigen from the core *Escherichia coli* genome. *mSphere* 2016;**1**:e00326–16.

Mughini-Gras L, Dorado-Garcia A, van Duijkeren E *et al.* Attributable sources of community-acquired carriage of *Escherichia coli* containing beta-lactam antibiotic resistance genes: a population-based modelling study. *The Lancet Planetary Health* 2019;**3**: e357–69.

Nada RA, Shaheen HI, Khalil SB *et al.* Discovery and phylogenetic analysis of novel members of class b enterotoxigenic *Escherichia coli* adhesive fimbriae. *J Clin Microbiol* 2011;**49**:1403–10.

Nadalian B, Yadegar A, Houri H *et al.* Prevalence of the pathobiont adherent-invasive *Escherichia coli* and inflammatory bowel disease: a systematic review and meta-analysis. *J Gastroenterol Hepatol* 2021;**36**:852–63.

Naili I, Vinot J, Baudner BC *et al.* Mixed mucosal-parenteral immunizations with the broadly conserved pathogenic *Escherichia coli* antigen SslE induce a robust mucosal and systemic immunity without affecting the murine intestinal microbiota. *Vaccine* 2019;**37**:314–24.

Nakata N, Tobe T, Fukuda I *et al.* The absence of a surface protease, OmpT, determines the intercellular spreading ability of *Shigella*: the relationship between the *ompT* and *kcpA* loci. *Mol Microbiol* 1993;**9**:459–68.

Nandre R, Ruan X, Lu T *et al.* Enterotoxigenic *Escherichia coli* adhesin-toxoid multiepitope fusion antigen CFA/I/II/IV-3xSTaN12S-mnLTG192G/L211A-derived antibodies inhibit adherence of seven adhesins, neutralize enterotoxicity of LT and STa toxins, and protect piglets against diarrhea. *Infect Immun* 2018;**86**:e00550–17.

Nandre RM, Duan Q, Wang Y *et al.* Passive antibodies derived from intramuscularly immunized toxoid fusion 3xSTaN12S-dmLT protect against STa+ enterotoxigenic *Escherichia coli* (ETEC) diarrhea in a pig model. *Vaccine* 2017;**35**:552–6.

Nascimento JAS, Santos FF, Valiatti TB *et al.* Frequency and diversity of hybrid *Escherichia coli* strains isolated from urinary tract infections. *Microorganisms* 2021;**9**: e9040693.

Nash JH, Villegas A, Kropinski AM *et al.* Genome sequence of adherent-invasive *Escherichia coli* and comparative genomic analysis with other *E. coli* pathotypes. *BMC Genomics* 2010;**11**:667.

Nataro JP, Kaper JB Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* 1998;**11**:142–201.

Nesbo CL, Dlutek M, Doolittle WF Recombination in *Thermotoga*: implications for species concepts and biogeography. *Genetics* 2006;**172**:759–69.

Nickel JC, Saz-Leal P, Doiron RC Could sublingual vaccination be a viable option for the prevention of recurrent urinary tract infection in canada? A systematic review of the current literature and plans for the future. *Canad Urolog Associat J* 2020;**14**:281–7.

Nicolas-Chanoine MH, Blanco J, Leflon-Guibout V *et al.* Intercontinental emergence of *Escherichia coli* clone O25:H4-ST131 producing CTX-M-15. *J Antimicrob Chemother* 2008;**61**:273–81.

Nielsen KL, Stegger M, Kiil K *et al.* Whole-genome comparison of urinary pathogenic *Escherichia coli* and faecal isolates of UTI patients and healthy controls. *Int J Med Microbiol* 2017;**307**:497–507.

Njage PMK, Leekitcharoenphon P, Hald T Improving hazard characterization in microbial risk assessment using next generation sequencing data and machine learning: predicting clinical outcomes in shigatoxigenic *Escherichia coli*. *Int J Food Microbiol* 2019;**292**:72–82.

Nji E, Kazibwe J, Hambridge T *et al.* High prevalence of antibiotic resistance in commensal *Escherichia coli* from healthy human sources in community settings. *Sci Rep* 2021;**11**:3372.

Nougayrede JP, Homburg S, Taieb F *et al.* *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* 2006;**313**: 848–51.

Nyholm O, Halkilahti J, Wiklund G *et al.* Comparative genomics and characterization of hybrid shigatoxigenic and enterotoxigenic *Escherichia coli* (STEC/ETEC) strains. *PLoS One* 2015;**10**:e0135936.

O'Brien CL, Bringer MA, Holt KE *et al.* Comparative genomics of Crohn's disease-associated adherent-invasive *Escherichia coli*. *Gut* 2017;**66**:1382–9.

O'Brien VP, Hannan TJ, Nielsen HV *et al.* Drug and vaccine development for the treatment and prevention of urinary tract infections. *Microbiol Spectr* 2016;**4**: doi: 10.1128/microbiolspec.UTI-0013-2012.

Ogura Y, Gotoh Y, Itoh T *et al.* Population structure of *Escherichia coli* O26 : H11 with recent and repeated *stx2* acquisition in multiple lineages. *Microb Genom* 2017;**3**: e000141.

Olier M, Marcq I, Salvador-Cartier C *et al.* Genotoxicity of *Escherichia coli* Nissle 1917 strain cannot be dissociated from its probiotic activity. *Gut Microbes* 2012;**3**:501–9.

Oliveira AF, Cardoso SA, Almeida FB *et al.* Oral immunization with attenuated *Salmonella* vaccine expressing *Escherichia coli* O157:H7 intimin gamma triggers both systemic and mucosal humoral immunity in mice. *Microbiol Immunol* 2012;**56**:513–22.

Olson PD, Hunstad DA Subversion of host innate immunity by uropathogenic *Escherichia coli*. *Pathogens* 2016;**5**:2.

Ormsby MJ, Johnson SA, Carpena N *et al.* Propionic acid promotes the virulent phenotype of Crohn's disease-associated adherent-invasive *Escherichia coli*. *Cell Rep* 2020;**30**:2297–2305.e2295.

Orskov I, Orskov F, Jann B *et al.* Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriol Rev* 1977;**41**: 667–710.

Palmela C, Chevarin C, Xu Z *et al.* Adherent-invasive *Escherichia coli* in inflammatory bowel disease. *Gut* 2018;**67**:574–87.

Palmer M, Steenkamp ET, Blom J *et al.* All ANIs are not created equal: implications for prokaryotic species boundaries and integration of ANIs into polyphasic taxonomy. *Int J Syst Evol Microbiol* 2020;**70**:2937–48.

Pang TY, Lercher MJ Each of 3,323 metabolic innovations in the evolution of *E. coli* arose through the horizontal transfer of a single DNA segment. *Proc Natl Acad Sci* 2019;**116**:187–92.

Park SC, Lee K, Kim YO *et al.* Large-scale genomics reveals the genetic characteristics of seven species and importance of phylo-

genetic distance for estimating pan-genome size. *Front Microbiol* 2019;**10**:834.

Parks DH, Chuvochina M, Reeves PR *et al.* Reclassification of *Shigella* species as later heterotypic synonyms of *Escherichia coli* in the genome taxonomy database. *bioRxiv* 2021.2009.2022.461432.

Parks DH, Chuvochina M, Waite DW *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;**36**:996–1004.

Partridge SR, Kwong SM, Firth N *et al.* Mobile genetic elements associated with antimicrobial resistance. *Clin Microbiol Rev* 2018;**31**:e00088–17.

Pasqua M, Michelacci V, Di Martino ML *et al.* The intriguing evolutionary journey of enteroinvasive *E. coli* (EIEC) toward pathogenicity. *Front Microbiol* 2017;**8**:2390.

Pesciaroli M, Magistrali CF, Filippini G *et al.* Antibiotic-resistant commensal *Escherichia coli* are less frequently isolated from poultry raised using non-conventional management systems than from conventional broiler. *Int J Food Microbiol* 2020;**314**: 108391.

Peters NK, Dixon DM, Holland SM *et al.* The research agenda of the National Institute of Allergy and Infectious Diseases for antimicrobial resistance. *J Infect Dis* 2008;**197**:1087–93.

Petersen AM, Mirsepasi H, Halkjaer SI *et al.* Ciprofloxacin and probiotic *Escherichia coli* Nissle add-on treatment in active ulcerative colitis: a double-blind randomized placebo controlled clinical trial. *J Crohn's Colitis* 2014;**8**:1498–505.

Petkovsek Z, Elersic K, Gubina M *et al.* Virulence potential of *Escherichia coli* isolates from skin and soft tissue infections. *J Clin Microbiol* 2009;**47**:1811–7.

Pettengill EA, Pettengill JB, Binet R Phylogenetic analyses of *Shigella* and enteroinvasive *Escherichia coli* for the identification of molecular epidemiological markers: whole-genome comparative analysis does not support distinct genera designation. *Frontiers in Microbiology* 2015;**6**:1573.

Pitout JD Extraintestinal pathogenic *Escherichia coli*: a combination of virulence with antibiotic resistance. *Frontiers in Microbiology* 2012;**3**:9.

Pitout JD, DeVinney R. *F1000 Research* 2017;**6**:F1000 Faculty Rev–195.

Pitout JDD, Finn TJ The evolutionary puzzle of *Escherichia coli* ST131. *Infect Genet Evol* 2020;**81**:104265.

Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A *et al.* Mutational signature in colorectal cancer caused by genotoxic *pks(+) E . coli*. *Nature* 2020;**580**:269–73.

Poirel L, Jayol A, Nordmann P Polymyxins: antibacterial activity, susceptibility testing, and resistance mechanisms encoded by plasmids or chromosomes. *Clin Microbiol Rev* 2017;**30**:557–96.

Poirel L, Madec JY, Lupo A *et al.* Antimicrobial resistance in *Escherichia coli*. *Microbiol Spect* 2018;**6**: ARBA-0026-2017.

Pompilio A, Crocetta V, Savini V *et al.* Phylogenetic relationships, biofilm formation, motility, antibiotic resistance and extended virulence genotypes among *Escherichia coli* strains from women with community-onset primitive acute pyelonephritis. *PLoS One* 2018;**13**:e0196260.

Ponce-de-Leon A, Rodriguez-Noriega E, Morfin-Otero R *et al.* Antimicrobial susceptibility of Gram-negative bacilli isolated from intra-abdominal and urinary-tract infections in Mexico from 2009 to 2015: results from the study for monitoring antimicrobial resistance trends (SMART). *PLoS One* 2018;**13**:e0198621.

Poolman JT, Wacker M Extraintestinal pathogenic *Escherichia coli*, a common human pathogen: challenges for vaccine development and progress in the field. *J Infect Dis* 2016;**213**:6–13.

Poolman JT, Anderson AS *Escherichia coli* and *Staphylococcus aureus*: leading bacterial pathogens of healthcare associated infections and bacteremia in older-age populations. *Expert Rev Vaccines* 2018;**17**:607–18.

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* 2000;**97**:10567–72.

Putze J, Hennequin C, Nougayrede JP *et al.* Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. *Infect Immun* 2009;**77**:4696–703.

Qadri F, Akhtar M, Bhuiyan TR *et al.* Safety and immunogenicity of the oral, inactivated, enterotoxigenic *Escherichia coli* vaccine ETVAX in Bangladeshi children and infants: a double-blind, randomised, placebo-controlled phase 1/2 trial. *Lancet Infect Dis* 2020;**20**:208–19.

Rakitina DV, Manolov AI, Kanygina AV *et al.* Genome analysis of *E. coli* isolated from Crohn's disease patients. *BMC Genomics* 2017;**18**:544.

Ramirez Sevilla C, Gomez Lanza E, Manzanera JL *et al.* Active immunoprophyilaxis with Uromune(R) decreases the recurrence of urinary tract infections at three and six months after treatment without relevant secondary effects. *BMC Infectious Diseases* 2019;**19**:901.

Ranjith K, SaiAbhilash CR, Sai Prashanthi G *et al.* Phylogenetic grouping of human ocular *Escherichia coli* based on whole-genome sequence analysis. *Microorganisms* 2020;**8**:422.

Rasko DA, Rosovitz MJ, Myers GS *et al.* The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol* 2008;**190**:6881–93.

Rasko DA, Webster DR, Sahl JW *et al.* Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in germany. *N Engl J Med* 2011;**365**:709–17.

Reitzer L, Zimmern P Rapid growth and metabolism of uropathogenic *Escherichia coli* in relation to urine composition. *Clin Microbiol Rev* 2019 ;**33**:e00101–19.

Rhee C, Kadri SS, Dekker JP *et al.* Prevalence of antibiotic-resistant pathogens in culture-proven sepsis and outcomes associated with inadequate and broad-spectrum empiric antibiotic use. *JAMA Network Open* 2019;**3**:e202899.

Riddle MS, Maciel M, Porter CK *et al.* A first in human clinical trial assessing the safety and immunogenicity of transcutaneously delivered enterotoxigenic *Escherichia coli* fimbrial tip adhesin with heat-labile enterotoxin with mutation R192G. *Vaccine* 2020;**38**:7040–8.

Riley LW Distinguishing pathovars from nonpathovars: *Escherichia coli*. *Microbiol Spect* 2020;**8**: AME-0014-2020.

Riquelme-Neira R, Rivera A, Saez D *et al.* Vaccination with DNA encoding truncated enterohemorrhagic *Escherichia coli* (EHEC) factor for adherence-1 gene (*efa-1*') confers protective immunity to mice infected with *E. coli* O157:H7. *Front Cell Infect Microbiol* 2015;**5**: 104.

Robins-Browne RM, Holt KE, Ingle DJ *et al.* Are *Escherichia coli* pathotypes still relevant in the era of whole-genome sequencing? *Front Cell Infect Microbiol* 2016;**6**:141.

Robinson AE, Heffernan JR, Henderson JP The iron hand of uropathogenic *Escherichia coli*: the role of transition metal control in virulence. *Fut Microbiol* 2018;**13**:745–56.

Rodriguez-Bano J, Gutierrez-Gutierrez B, Machuca I *et al.* Treatment of infections caused by extended-spectrum-beta-lactamase-, AmpC-, and carbapenemase-producing Enterobacteriaceae. *Clin Microbiol Rev* 2018;**31**.

Rogers BA, Sidjabat HE, Paterson DL *Escherichia coli* O25b-ST131: a pandemic, multiresistant, community-associated strain. *J Antimicrob Chemother* 2011;**66**:1–14.

Rojas-Lopez M, Monterio R, Pizza M *et al.* Intestinal pathogenic *Escherichia coli*: insights for vaccine development. *Front Microbiol* 2018;**9**:440.

Rojas-Lopez M, Martinelli M, Brandi V *et al.* Identification of lipid A deacylase as a novel, highly conserved and protective antigen against enterohemorrhagic *Escherichia coli*. *Sci Rep* 2019;**9**: 17014.

Rollenhagen JE, Jones F, Hall E *et al.* Establishment, *validation, and application of a new world primate model of enterotoxigenic Escherichia coli disease for vaccine development*. *Infect Immun* 2019;**87**.

Rouli L, Merhej V, Fournier PE *et al.* The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microb New Infect* 2015;**7**:72–85.

Rudick CN, Berry RE, Johnson JR *et al.* Uropathogenic *Escherichia coli* induces chronic pelvic pain. *Infect Immun* 2011;**79**:628–35.

Russo TA, Johnson JR Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: exPEC. *J Infect Dis* 2000;**181**:1753–4.

Saade E, Gravenstein S, Donskey CJ *et al.* Characterization of *Escherichia coli* isolates potentially covered by ExPEC4V and ExPEC10V, that were collected from post-transrectal ultrasound-guided prostate needle biopsy invasive urinary tract and bloodstream infections. *Vaccine* 2020;**38**:5100–4.

Sagi S, Konduru B, Parida M Heterologous expression of intimin and IpaB fusion protein in *Lactococcus lactis* and its mucosal delivery elicit protection against pathogenicity of *Escherichia coli* O157 and *Shigella flexneri* in a murine model. *Int Immunopharmacol* 2020;**85**:106617.

Sahl JW, Sistrunk JR, Baby NI *et al.* Insights into enterotoxigenic *Escherichia coli* diversity in Bangladesh utilizing genomic epidemiology. *Sci Rep* 2017;**7**:3402.

Sahl JW, Morris CR, Emberger J *et al.* Defining the phylogenomics of *Shigella* species: a pathway to diagnostics. *J Clin Microbiol* 2015;**53**:951–60.

Sanchez-Villamil JI, Tapia D, Torres AG Development of a gold nanoparticle vaccine against enterohemorrhagic *Escherichia coli* O157:H7. *mBio* 2019;**10**:e01869–19.

Santos ACM, Santos FF, Silva RM *et al.* Diversity of hybrid- and heteropathogenic *Escherichia coli* and their potential implication in more severe diseases. *Front Cell Infect Microbiol* 2020;**10**:339.

Santos AS, Finlay BB Bringing down the host: enteropathogenic and enterohaemorrhagic *Escherichia coli* effector-mediated subversion of host innate immune pathways. *Cell Microbiol* 2015;**17**:318–32.

Sarowska J, Futoma-Koloch B, Jama-Kmiecik A *et al.* Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: recent reports. *Gut Pathogens* 2019;**11**:10.

Sartelli M, Catena F, Ansaloni L *et al.* Complicated intra-abdominal infections worldwide: the definitive data of the CIAOW study. *World J Emerg Surg* 2014;**9**:37.

Sayers EW, Beck J, Brister JR *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2020;**48**:D9–D16.

Scallan E, Hoekstra RM, Angulo FJ *et al.* Foodborne illness acquired in the United States–major pathogens. *Emerg Infect Dis* 2011;**17**:7–15.

Schaufler K, Semmler T, Wieler LH *et al.* Genomic and functional analysis of emerging virulent and multi-drug resistant *E. coli* lineage ST648. *Antimicrob Agents Chemother* 2019;**63**:e00243–19.

Schreiber HL, Conover MS, Chou WC *et al.* Bacterial virulence phenotypesof Escherichia coli and host susceptibility determine riskfor urinary tract infections. *Sci Transl Med* 2017;**9**:eaaf1283.

Schulman CC, Corbusier A, Michiels H *et al.* Oral immunotherapy of recurrent urinary tract infections: a double-blind placebo-controlled multicenter study. *J Urol* 1993;**150**:917–21.

Schwab S, Jobin K, Kurts C Urinary tract infection: recent insight into the evolutionary arms race between uropathogenic *Escherichia coli* and our immune system. *Nephrology Dialysis Transplantation* 2017;**32**:1977–83.

Secher T, Kassem S, Benamar M *et al.* Oral administration of the probiotic strain *Escherichia coli* Nissle 1917 reduces susceptibility to neuroinflammation and repairs experimental autoimmune encephalomyelitis-induced intestinal barrier dysfunction. *Front Immunol* 2017;**8**:1096.

Shaler CR, Elhenawy W, Coombes BK The unique lifestyle of Crohn's disease-associated adherent-invasive *Escherichia coli*. *J Mol Biol* 2019;**431**:2970–81.

Shea AE, Marzoa J, Himpsl SD *et al. Escherichia coli* CFT073 fitness factors during urinary tract infection: identification using an ordered transposon library. *Appl Environ Microbiol* 2020;**86**:e00691–20.

Shen Y, Wu Z, Wang Y *et al.* Heterogeneous and flexible transmission of *mcr-1* in hospital-associated *Escherichia coli*. *mBio* 2018;**9**:e00943–18.

Shepherd M, Achard ME, Idris A *et al.* The cytochrome bd-I respiratory oxidase augments survival of multidrug-resistant *Escherichia coli* during infection. *Sci Rep* 2016;**6**:35285.

Snipen L, Almoy T, Ussery DW Microbial comparative pan-genomics using binomial mixture models. *BMC Genomics* 2009;**10**: 385.

Soysal N, Mariani-Kurkdjian P, Smail Y *et al.* Enterohemorrhagic *Escherichia coli* hybrid pathotype O80:H2 as a new therapeutic challenge. *Emerg Infect Dis* 2016;**22**:1604–12.

Spurbeck RR, Dinh PC, Walk ST *et al. Escherichia coli* isolates that carry *vat, fyuA, chuA*, and *yfcV* efficiently colonize the urinary tract. *Infect Immun* 2012;**80**:4115–22.

Stoppato M, Gaspar C, Regeimbal J *et al.* Oral administration of an anti-CfaE secretory IgA antibody protects against enterotoxigenic *Escherichia coli* diarrheal disease in a nonhuman primate model. *Vaccine* 2020;**38**:2333–9.

Stromberg N, Marklund BI, Lund B *et al.* Host-specificity of uropathogenic *Escherichia coli* depends on differences in binding specificity to gal alpha 1-4Gal-containing isoreceptors. *EMBO J* 1990;**9**:2001–10.

Su M, Satola SW, Read TD Genome-based prediction of bacterial antibiotic resistance. *J Clin Microbiol* 2019;**57**:e01405–18.

Subashchandrabose S, Mobley HLT Virulence and fitness determinants of uropathogenic *Escherichia coli*. *Microbiol Spect* 2015;**3**:doi: 10.1128/microbiolspec.UTI-0015-2012.

Sung MA, Fleming K, Chen HA *et al.* The solution structure of PapGII from uropathogenic *Escherichia coli* and its recognition of glycolipid receptors. *EMBO Rep* 2001;**2**:621–7.

Swidsinski A, Khilkin M, Kerjaschki D *et al.* Association between intraepithelial *Escherichia coli* and colorectal cancer. *Gastroenterology* 1998;**115**:281–6.

Taati Moghadam M, Mirzaei M, Fazel Tehrani Moghaddam M *et al.* The challenge of global emergence of novel colistin-resistant Escherichiacoli ST131. *Microb Drug Resist* 2021;**27**:1513–24.

Taha Neto KA, Nogueira Castilho L, Reis LO Oral vaccine (OM-89) in the recurrent urinary tract infection prophylaxis: a realistic systematic review with meta-analysis. *Actas Urológicas Españolas* 2016;**40**:203–8.

Takeuchi N, Cordero OX, Koonin EV *et al.* Gene-specific selective sweeps in bacteria and archaea caused by negative frequency-dependent selection. *BMC Biol* 2015;**13**:20.

Tammen H Immunobiotherapy with Uro-Vaxom in recurrent urinary tract infection. The German urinary tract infection study group. *Br J Urol* 1990;**65**:6–9.

Tanoue T, Umesaki Y, Honda K Immune responses to gut microbiota-commensals and pathogens. *Gut Microbes* 2010;**1**:224–33.

Taxt AM, Diaz Y, Bacle A *et al.* Characterization of immunological cross-reactivity between enterotoxigenic *Escherichia coli* heat-stable toxin and human guanylin and uroguanylin. *Infect Immun* 2014;**82**:2913–22.

Teelucksingh T, Thompson LK, Cox G The evolutionary conservation of *Escherichia coli* drug efflux pumps supports physiological functions. *J Bacteriol* 2020;**202**:e00367–20.

Telhig S, Ben Said L, Zirah S *et al.* Bacteriocins to thwart bacterial resistance in Gram negative bacteria. *Front Microbiol* 2020;**11**:586433.

Tenaillon O, Skurnik D, Picard B *et al.* The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* 2010;**8**:207–17.

Teweldemedhin M, Gebreyesus H, Atsbaha AH *et al.* Bacterial profile of ocular infections: a systematic review. *BMC Ophthalmol* 2017;**17**:212.

Thanh Duy P, Thi Nguyen TN, Vu Thuy D *et al.* Commensal *Escherichia coli* are a reservoir for the transfer of XDR plasmids into epidemic fluoroquinolone-resistant *Shigella sonnei*. *Nat Microbiol* 2020;**5**:256–64.

Theuretzbacher U, Ardal C, Harbarth S Linking sustainable use policies to novel economic incentives to stimulate antibiotic research and development. *Infect Dis Rep* 2017;**9**:6836.

Thomson DU, Loneragan GH, Thornton AB *et al.* Use of a siderophore receptor and porin proteins-based vaccine to control the burden of *Escherichia coli* O157:H7 in feedlot cattle. *Foodborne Pathog Dis* 2009;**6**:871–7.

Tian X, Zheng X, Sun Y *et al.* Molecular mechanisms and epidemiology of carbapenem-resistant *Escherichia coli* isolated from Chinese patients during 2002-2017. *Infect Drug Resist* 2020;**13**:501–12.

Toombs-Ruane LJ, Benschop J, French NP *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**:e01613–20.

Torres E, Lopez-Cerero L, Morales I *et al.* Prevalence and transmission dynamics of *Escherichia coli* ST131 among contacts of infected community and hospitalized patients. *Clin Microbiol Infect* 2018;**24**:618–23.

Touchon M, Perrin A, de Sousa JAM *et al.* Phylogenetic background and habitat drive the genetic diversification of *Escherichia coli*. *PLos Genet* 2020;**16**:e1008866.

Touchon M, Hoede C, Tenaillon O *et al.* Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLos Genet* 2009;**5**:e1000344.

Tourret J, Denamur E Population phylogenomics of extraintestinal pathogenic *Escherichia coli*. *Microbiol Spect* 2016;**4**. doi: 10.1128/microbiolspec.UTI-0010-2012.

Trofa AF, Ueno-Olsen H, Oiwa R *et al.* Dr. Kiyoshi Shiga: discoverer of the dysentery bacillus. *Clin Infect Dis* 1999;**29**:1303–6.

Tsen SD, Fang SS, Chen MJ *et al.* Natural plasmid transformation in *Escherichia coli*. *J Biomed Sci* 2002;**9**:246–52.

Turner SM, Scott-Tucker A, Cooper LM *et al.* Weapons of mass destruction: virulence factors of the global killer enterotoxigenic *Escherichia coli*. *FEMS Microbiol Lett* 2006;**263**:10–20.

Ukena SN, Singh A, Dringenberg U *et al.* Probiotic *Escherichia coli* Nissle 1917 inhibits leaky gut by enhancing mucosal integrity. *PLoS One* 2007;**2**:e1308.

Usein CR, Damian M, Tatu-Chitoiu D *et al.* Prevalence of virulence genes in *Escherichia coli* strains isolated from Romanian adult urinary tract infection cases. *J Cell Mol Med* 2001;**5**:303–10.

Valiatti TB, Santos FF, Santos ACM *et al.* Genetic and virulence characteristics of a hybrid atypical enteropathogenic and uropathogenic *Escherichia coli* (aEPEC/UPEC) strain. *Front Cell Infect Microbiol* 2020;**10**:492.

Van Camp PJ, Haslam DB, Porollo A Prediction of antimicrobial resistance in Gram-negative bacteria from whole-genome sequencing data. *Front Microbiol* 2020;**11**:1013.

van Hoek A, Dierikx C, Bosch T *et al.* Transmission of ESBL-producing *Escherichia coli* between broilers and humans on broiler farms. *J Antimicrob Chemother* 2020;**75**:543–9.

van Hoek A, van Veldhuizen JNJ, Friesema I *et al.* Comparative genomics reveals a lack of evidence for pigeons as a main source of *stx2f*-carrying *Escherichia coli* causing disease in humans and the common existence of hybrid shiga toxin-producing and enteropathogenic *E. coli* pathotypes. *BMC Genomics* 2019;**20**:271.

Vester-Andersen MK, Mirsepasi-Lauridsen HC, Prosberg MV *et al.* Increased abundance of proteobacteria in aggressive Crohn's disease seven years after diagnosis. *Sci Rep* 2019;**9**:13473.

Vieira G, Sabarly V, Bourguignon PY *et al.* Core and panmetabolism in *Escherichia coli*. *J Bacteriol* 2011;**193**:1461–72.

Viladomiu M, Metz ML, Lima SF *et al.* Adherent-invasive *E. coli* metabolism of propanediol in crohn's disease regulates phagocytes to drive intestinal inflammation. *Cell Host and Microbe* 2021;**29**:607–619.e608.

Vilne B, Meistere I, Grantina-Ievina L *et al.* Machine learning approaches for epidemiological investigations of food-borne disease outbreaks. *Front Microbiol* 2019;**10**:1722.

Walk ST The "cryptic" *Escherichia*. *EcoSal Plus* 2015;**6**: doi: 10.1128/ecosalplus.ESP–0002-2015.

Walk ST, Alm EW, Gordon DM *et al.* Cryptic lineages of the genus *Escherichia*. *Appl Environ Microbiol* 2009;**75**:6534–44.

Wang C, Feng Y, Liu L *et al.* Identification of novel mobile colistin resistance gene *mcr-10*. *Emerg Microb Infect* 2020;**9**:508–16.

Wang G, Li X, Wang Z APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res* 2016;**44**:D1087–1093.

Wang L, Rothemund D, Curd H *et al.* Species-wide variation in the *Escherichia coli* flagellin (H-antigen) gene. *J Bacteriol* 2003;**185**:2936–43.

Wattam AR, Davis JJ, Assaf R *et al.* Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res* 2017;**45**:D535–42.

Wayne LG International committee on systematic bacteriology announcement of the report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *J Appl Bacteriol* 1988;**64**:283–4.

Weeks AR, Hoffmann AA Frequency-dependent selection maintains clonal diversity in an asexual organism. *Proc Natl Acad Sci* 2008;**105**:17872–7.

Wei J, Goldberg MB, Burland V *et al.* Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect Immun* 2003;**71**:2775–86.

Weiner-Lastinger LM, Abner S, Edwards JR *et al.* Antimicrobial-resistant pathogens associated with adult healthcare-associated infections: summary of data reported to the National Healthcare Safety Network, 2015-2017. *Infect Control Hosp Epidemiol* 1–18.

Weiner LM, Webb AK, Limbago B *et al.* Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the National Healthcare Safety Net-

work at the Centers for Disease Control and Prevention, 2011-2014. *Infect Control Hosp Epidemiol* 2016;**37**:1288–301.

Welch RA Uropathogenic *Escherichia coli*-associated exotoxins. *Microbiol Spect* 2016;**4**: doi:10.1128/microbiolspec.UTI-0011-2012.

Whitfield C, Roberts IS Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Mol Microbiol* 1999;**31**:1307–19.

Willenbrock H, Hallin PF, Wassenaar TM *et al.* Characterization of probiotic *Escherichia coli* isolates with a novel pan-genome microarray. *Genome Biol* 2007;**8**:R267.

Wilson J, Elgohari S, Livermore DM *et al.* Trends among pathogens reported as causing bacteraemia in England, 2004-2008. *Clin Microbiol Infect* 2011;**17**:451–8.

Wilson MR, Jiang Y, Villalta PW *et al.* The human gut bacterial genotoxin colibactin alkylates DNA. *Science* 2019;**363**:eaar7785.

Wirth T, Falush D, Lan R *et al.* Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* 2006;**60**:1136–51.

Woegerbauer M, Jenni B, Thalhammer F *et al.* Natural genetic transformation of clinical isolates of *Escherichia coli* in urine and water. *Appl Environ Microbiol* 2002;**68**:440–3.

World Health Organization Global action plan on antimicrobial resistance. 2016. https://www.who.int/publications/i/item/9789241509763.

World Health Organization(Global priority list of antibiotic resistant bacteria to guide research, discovery, and development of new antibiotics. 2017. https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed.

World Health Organization Guidelines for drinking-water quality: fourth edition incorporating the first addendum. Geneva. 2017. https://www.who.int/publications/i/item/9789241549950.

Wright KJ, Seed PC, Hultgren SJ Development of intracellular bacterial communities of uropathogenic *Escherichia coli* depends on type 1 pili. *Cell Microbiol* 2007;**9**:2230–41.

Xavier BB, Das AJ, Cochrane G *et al.* Consolidating and exploring antibiotic resistance gene data resources. *J Clin Microbiol* 2016;**54**:851–9.

Yang B, Foley S First experience in the UK of treating women with recurrent urinary tract infections with the bacterial vaccine Uromune((r)). *BJU Int* 2018;**121**:289–92.

Yoon YK, Kim J, Moon C *et al.* Antimicrobial susceptibility of microorganisms isolated from patients with intraabdominal infection in Korea: a multicenter study. *J Korean Med Sci* 2019;**34**:e309.

Yu D, Banting G, Neumann N A review of the taxonomy, genetics and biology of the genus *Escherichia* and the type species *Escherichia coli*. *Can J Microbiol* 2021;**67**:553–71.

Zeng X, Chi X, Ho BT *et al.* Comparative genome analysis of an extensively drug-resistant isolate of avian sequence type 167 *Escherichia coli* strain Sanji with novel *in silico* serotype O89b:H9. *mSystems* 2019;**4**:e00242–18.

Zhang H, Chen X, Nolan LK *et al.* Identification of host adaptation genes in extraintestinal pathogenic *Escherichia coli* during infection in different hosts. *Infect Immun* 2019;**87**.

Zhang H, Yang Q, Liao K *et al.* Update of incidence and antimicrobial susceptibility trends of *Escherichia coli* and *Klebsiella pneumoniae* isolates from Chinese intra-abdominal infection patients. *BMC Infect Dis* 2017;**17**:776.

Zhang W, Sack DA Progress and hurdles in the development of vaccines against enterotoxigenic *Escherichia coli* in humans. *Expert Rev Vaccines* 2012;**11**:677–94.

Zhang W, Du P, Zheng H *et al.* Whole-genome sequence comparison as a method for improving bacterial species definition. *J Gen Appl Microbiol* 2014;**60**:75–78.

Zhang X, Xiao S, Jiang X *et al.* Genomic characterization of *Escherichia coli* LCT-EC001, an extremely multidrug-resistant strain with an amazing number of resistance genes. *Gut Pathogens* 2019;**11**:25.

Zhou R, Chen Z, Hao D *et al.* Enterohemorrhagic *Escherichia coli* Tir inhibits TAK1 activation and mediates immune evasion. *Emerg Microb Infect* 2019;**8**:734–48.

Zong Z, Fenn S, Connor C *et al.* Complete genomic characterization of two *Escherichia coli* lineages responsible for a cluster of carbapenem-resistant infections in a Chinese hospital. *J Antimicrob Chemother* 2018;**73**:2340–6.