Early settlers—which E. coli strains do you not want at birth?

Thomas Secher 1, Camille Brehin 1,2 and Eric Oswald 1,3

1 UMR1220, Institute de Recherche en Santé Digestive, INSERM-INRA-UPS-ENVT, Toulouse, France
2 CHU Toulouse, Hôpital Purpan, Service de Pédiatrie, Toulouse, France.
3 CHU Toulouse, Hôpital Purpan, Service de Bactériologie-Hygiène, Toulouse, France.

Keywords: E. coli; neonatal sepsis, IBD, cancer, allergy

Corresponding author:
Eric Oswald
UMR1220, Institute de Recherche en Santé Digestive, INSERM-INRA-UPS-ENVT, Toulouse, France
eric.oswald@inserm.fr
Phone: +33 (0)5 62 74 45 43
Abstract

The intestinal microbiota exerts vital biological processes throughout the human lifetime and imbalances in its composition have been implicated in both health and disease status. Upon birth, the neonatal gut moves from a barely sterile to a massively colonized environment. The development of the intestinal microbiota during the first year of life is characterized by rapid and important changes in microbial composition, diversity and magnitude. The pioneer bacteria colonizing the postnatal intestinal tract profoundly contribute to the establishment of the host-microbe symbiosis which is essential for health throughout life. *Escherichia coli* is one of the first colonizers of the gut after birth. *E. coli* is a versatile population including harmless commensal, probiotic strains as well as frequently deadly pathogens. The prevalence of the specific phylogenetic B2 group, which encompasses both commensal and extra- or intra-intestinal pathogenic *E. coli* strains, is increasing among *E. coli* strains colonizing infants quickly after birth. 50% of the B2 group strains carry in their genome the *pks* gene cluster encoding the synthesis of a non-ribosomal peptide-polyketide hybrid genotoxin named colibactin.

In this review, we summarize both clinical and experimental evidences associating the recently emerging neonatal B2 *E. coli* population with several pathology and discuss how the expression of colibactin by both normal inhabitants of intestinal microflora and virulent strains may darken the borderline between commensalism and pathogenicity.
Development of the intestinal microbiota-establishing a symbiosis.

The intestinal microbiota is the largest microbial ecosystem of the human microbiome consisting of $3 \times 10^{13}$ bacterial cells and more than a hundred times the number of genes composing the human genome (43). The use of high-throughput sequencing technologies has revealed that the intestinal microbiota is composed of more than 1,000 species (61). Among this very broad community, a consensus has emerged on the composition of a human adult intestinal microbiota with two main phyla that are Firmicutes (mainly represented by the genera Clostridium, Faecalibacterium, Ruminococcus and Lactobacillus) and Bacteroidetes (Bacteroides and Prevotella). Other phyla compose the subdominant microflora, including Proteobacteria (Gammaproteobacteria), Actinobacteria (Bifidobacterium) or Verrucomicrobia (Akkermansia) (53). It was originally assumed that this “core microbiota” would be similarly shared among individuals, but this was not confirmed by experimental evidences that demonstrated a large diversity within individuals and to a greater extent between individuals. These variations are even more acute considering species or clone level (7, 43, 60).

In healthy individuals, the intestinal microbiota coexists in a symbiotic relationship that confers benefits to both the host and the microbes. In fact, the host is able to determine or modify the composition and development of the intestinal microbiota and the microbiota is involved in essential metabolic functions of the host.

This host-intestinal microbiota symbiosis is initiated early in life. The classical pioneer species colonizing the infant gut are facultative anaerobes such as E. coli or other Enterobacteriaceae (38). These early settlers will create an anaerobic environment by consuming the initial oxygen supply, promoting the subsequent colonization by obligate anaerobes such as Bifidobacterium, Clostridium and Bacteroides.

The origin of early colonizing E. coli strains is diverse. Upon birth, depending on the delivery mode, the infant gut is massively colonized by E. coli originating from the mother’s microbiota. In addition, intra-uterine contamination of the meconium by E. coli resulting from
the translocation of the mother’s gut microbiota has been described (23). Recent radical
evidences also challenged the paradigm of the sterility of the fetus describing a low abundant
but robust placental microbiome, in which *E. coli* is the most abundant specie, and is associated
with remote history of antenatal infection, especially urinary tract infection (1).

From these complex premises, the intestinal microbiota of infants will dynamically
evolve towards a mature and stable consortium by the end of the third year of life (60). Its
development is profoundly influenced by the host and external factors such as the medical
practices (use of antibiotics), diet or geographic origin encountered in the first year of life.
Given the potent interference between the intestinal microbiota and the human immune,
metabolic and cognitive systems, the initiation of the symbiosis is crucial to promote health
later in life (Figure 1).

**E. coli, a versatile gut inhabitant.**

Up to 90% of the human population is colonized with *E. coli*. In adults, while
outnumbered by anaerobic bacteria, *E. coli* remains the predominant aerobic organism in the
intestine. *E. coli* reached density higher than 10^9 cfu per gram of feces, before the expansion of
obligate anaerobes. After 2 years of life, *E. coli* population stabilizes at around 10^7-10^8 cfu in
the colon and to a lesser extent terminal ileum (10^3-10^5 cfu) before gradually decreasing in the
elderly (Figure 1). This peaceful relationship in which one of the partners benefits from the
interaction, whereas the other is neither harmed nor helped is referred as commensalism.
Indeed, commensal *E. coli* uses its host for a constant source nutrient supply, protection,
transport and dissemination.

Besides these harmless commensal strains, *E. coli* family comprises also highly
intestinal (e.g. IPEC for Intestinal Pathogenic *E. coli*) and extra-intestinal pathogenic (e.g.
ExPEC for Extraintestinal Pathogenic *E. coli*) strains responsible for the death of more than
2 million humans per year (47). IPEC was the first *E. coli* group to be involved in human
disease and is divided in six pathovars, defined by the expression of specific virulence factors and the clinical manifestations of the disease, including enteropathogenic *E. coli* (EPEC), enteroheamorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC) and enteroaggregative *E. coli* (EAEC) (27).

*E. coli* strains colonized various locations in the human body. The preferential site of colonization is the large intestine, especially colon and cecum, where commensal strains reside on the mucus layer for which *E. coli* has adapted its metabolism to use it as a source of essential nutrients (8). Due to its impressive phenotypic versatility, *E. coli* can colonize additional ecological niches (including small intestine) for which some strains or clones have evolved and adapted to additional pathogenic lifestyle. Vicinity to host cell is prerequisite for all pathovars, nevertheless each of them has its own intrinsic mechanism to adhere, colonized, invaded the host and evaded to the immune response, explaining some differences in spatial distribution (Figure 1). Thus, considering feces as the single ecological niche to understand the human gut microbiota may be misleading.

Colibactin-producing *E. coli*, the thin red line between commensalism and pathogenesis.

The molecular basis underlying this phenotypic diversity, ranging from commensalism to pathogenicity is a very dynamic genome structure. With a core-genome (genes present in all *E. coli* strains) representing only 40% of total *E. coli*-related genes, *E. coli* genome has been enriched by the acquisition of numerous mobile genetic elements throughout the evolution. This flexible gene pool, including plasmids, phages, genomic islands, encodes for fitness and virulence factors which are responsible for the diverse phenotypes of *E. coli* and its adaptation to various environmental conditions (55). The analysis of the genes present in the core-genome revealed five major phylogenetic groups classified as A, B1, B2, D and E (18), recently updated by Clermont et al. (10). While group A includes mostly non pathogenic strains, group B2 comprises both commensals and pathogenic strains like ExPEC, as well as the newly described...
pathovar of adherent-invasive *E. coli* strains (AIEC) which are isolated from ileon biopsy of Crohn’s disease patients and the probiotic Nissle 1917 strain widely used for the treatment of inflammatory bowel disease (34). The B2 group has early emerged in the phylogenetic *E. coli* tree and exhibits the highest gene diversity (28) (Figure 2).

Up to 30% of the B2 group *E. coli* strains carry in their genome the *pks* gene cluster responsible for the synthesis of non ribosomal peptide-polyketide hybrid(s) named colibactin (34) (Figure 2). The pattern of distribution of the *pks* island reveals a complete restriction to the phylogenetic group B2. The 54kb *pks* island clusters the *clb* genes encoding a multi-modular complex leading to the synthesis in the cytosol of an inactive colibactin containing a prodrug motif that is subsequently hydrolyzed in the periplasm to generate the activated colibactin. The recent characterization of intermediate biosynthesis products revealed that activated colibactin is composed of an electrophilic “warhead” that may be involved in DNA-double strand breaks (56) (Figure 2). Indeed, acute infection of eukaryotic cells with colibactin-producing B2 *E. coli* strains induces massive DNA double-strand breaks, promotes the DNA-damage response (DDR) which results into the cell cycle arrest and *in fine* to apoptosis (34). Thus, colibactin was referred as a genotoxin. Cells that survive low-dose infection can repair their DNA owing to the DDR machinery and resume their cell cycle (12). However, DNA repair can be incomplete and trigger chromosomal abnormalities, genomic instability and increased mutation frequency (12). In addition, milder infectious dose induced persistent DNA damage and hallmarks of senescence (13, 50) (Figure 2).

Large-scale epidemiological studies have revealed a profound shift into the prevalence of *E. coli* in the human population. Indeed, the prevalence of the B2 group, including colibactin-producing strains, is increasing among *E. coli* strains colonizing humans from developed countries (31, 54), including infants (35) (Figure 3), while A and B1 groups remain predominant in developing countries (15). Socioeconomic factors, including diet, level of hygiene, medical treatments are presumably the main factors accounting for this major
modification in the phylogenetic group distribution of *E. coli* population. Carriage of the *pks* island has been estimated in the population from developed countries and revealed that up to 25% of humans are colonized by a *pks*+ B2 *E. coli* (17, 24, 42). Up to 18% of Swedish infants (36) and 15% of French infants are colonized at birth by a *pks*+ B2 *E. coli* (39). In addition, carriage of the *pks* island was associated with an increased persistence of neonatal acquired B2 *E. coli* strain (36).

*Colibactin and host health—the unexpected contributions of E. coli.*

The intriguing production of the genotoxin colibactin by normal inhabitants of the intestinal microbiota as well as virulent strains of *E. coli* during the critical neonatal period alters our current vision of commensalism and pathogenicity and confused the existing definition of the microbial carrier state. This prompts us to investigate the biological roles of colibactin production during health and disease (Figure 3).

*Neonatal sepsis and neonatal meningitis.* Neonatal sepsis is an extra-intestinal infection targeting infants before 1 month of age with an incidence estimated to be ~0.77 per 1,000 births and with 11% of mortality (reviewed by (52)). ExPEC are the second most frequent etiologic agents (after Group B Streptococci) accounting for 24% of the entire sepsis episodes and the most common etiology of sepsis in preterm infants (51). The incidence of *E. coli*-induced neonatal sepsis may be explained by the fact that *E. coli* can contaminate infants at or just before delivery. Data from experimental models revealed that after a successful asymptomatic colonization of the gastrointestinal tract, septicemic *E. coli* have the ability to translocate through the neonatal intestinal barrier and enter the blood circulation via the mesenteric lymph node causing severe systemic infection with potential manifestations in the central nervous system (40, 52). Indeed neonatal bacterial meningitis is a common consequence of neonatal bacteremia (22) associated with high mortality and morbidity. *E. coli* is also the main etiologic
agent of meningitis (20) especially in neonates which is the population at greatest risk (3). The molecular characterization of neonatal septicemic *E. coli* isolates revealed that the phylogroup B2 was the most prevalent (46, 59). Specific *E. coli* virulence factors have been associated with septicemic *E. coli* isolates, including the expression of K1 capsule, adhesion and invasion molecules (21). Interestingly, the most virulent septicemic *E. coli* strains cluster in the B2 phylogroup which include colibactin-producing strains (46). The contribution of colibactin in the virulence of septicemic *E. coli* strains has been investigated in murine models of sepsis. In the recently emerged clinical ExPEC ST131 clone, the production of colibactin was significantly correlated to its improved virulence trait as compared to other ExPEC isolates (26). The prototypal septicemic B2 *E. coli* K1 strain (25), isolated from a neonatal meningitis patient, has been shown to express the *pks* island and produce colibactin (34). Our team recently revealed that colibactin is a virulence factor for ExPEC, by showing that mice infected with *E. coli* K1 presented profound lymphopenia and altered survival rates as compared to mice infected with colibactin-deficient ExPEC (29). The contribution of colibactin to *E. coli* K1 virulence was recently confirmed in a neonatal rat sepsis model in which colibactin-producing strain was shown to better colonize the gastrointestinal tract, translocate to the blood compartment and cause invasive, lethal disease as compared to rat infected with strain defective for essential genes involved in colibactin production (32). A recent report has provided evidences that UPEC (another ExPEC pathovar), that may originate from the birth canal, could be also a risk factor for the development of necrotizing enterocolitis, the most common gastrointestinal emergency occurring in neonates (58). The involvement of colibactin in the pathogenesis of this infection remains to be determined.

These data suggest that the *pks* island, which was primarily identified in an neonatal meningitis isolate (36), is associated with the virulence of ExPEC strains. The production of colibactin may act as a classical virulence factor increasing the ability to translocate through the
intestinal barrier and to survive in the systemic compartment but also as a fitness factor, improving colonization of the gastro-intestinal tract.

**Intestinal barrier homeostasis.** The gastrointestinal tract is a complex interface between the external environment and the host. The intestinal epithelial barrier is involved in the regulation of the interactions between the gut microbiota and the gut associated immune system. Given the potency of colibactin-producing ExPEC to cross the intestinal barrier, our team recently investigated whether early gut colonization by colibactin-producing B2 *E. coli* is involved in the intestinal barrier homeostasis. Using a novel experimental model of neonatal gut colonization, we demonstrated that colibactin-producing commensal *E. coli* acquired from mother induced DNA damage in the enterocytes of the progeny and was associated with an enhanced ability of these strains to translocate through the intestinal epithelium (39, 49). At adulthood, the consequences of genotoxic stress observed after colibactin *in vitro* intoxication (34) were recapitulated *in vivo*, with an increased occurrence of crypt fission and anaphase bridges in enterocytes detected in individuals colonized by colibactin-producing *E. coli*. It has been widely accepted that the interaction between the early gut microbiota and the host is crucial for the proper development of immune functions by targeting various cell types (19). Interestingly, we observed, in animals early colonized by colibactin-producing B2 *E. coli*, a decreased of secretory epithelial cells (39) and regulatory T cells (49) populations. At a functional level, this was correlated to an impaired intestinal permeability and oral tolerance (39, 49).

Collectively these data suggest that the early gut colonization with genotoxic commensal strains from the newly dominant B2 *E. coli* group contributes to defective intestinal homeostasis and oral tolerance that could facilitate the development of deregulated immune-mediated diseases at adulthood. This hampers an unambiguous definition of commensal versus pathogenic *E. coli*. 
Colorectal cancer (CRC). Since the identification of the genotoxic nature of colibactin, we hypothesized that this metabolite may confer carcinogenic properties to *E. coli* B2 strains. Long-term gut colonization with such genotoxic *E. coli* strains may have an impact on tumor development and colorectal cancer. CRC is the fourth most frequent cancer, with more than one million cases per year and the third most deadly worldwide. Tumors tend to develop more frequently in the distal part of the large intestine (including descending colon and rectum), as compared to more proximal regions (44). Recent studies have demonstrated that modification in either the diversity or the magnitude of the intestinal microbiota, like dysbiosis observed in inflammatory bowel disease (IBD), may be associated with CRC (reviewed by (48)). This suggests that commensal inhabitants may be involved in the promotion of carcinogenesis. However, few carcinogenic bacteria have identified so far. Our team and others revealed that the consequences of infection by genotoxic *E. coli* is complex. Surviving cells exposed to colibactin presented chromosome aberrations, genomic instability and increased gene mutation frequency (12), lesions often seen in CRC. In addition, we demonstrated that intoxicated cells exhibited hallmarks of cellular senescence which was accompanied by an increased production of pro-inflammatory mediators, ROS and proteases. This senescent-associated secretory pattern (SASP) supported the *in vitro* growth of tumor cells (50). This work has been extended by Cougnoux et al., showing that colibactin-producing *E. coli* were able to enhance tumor growth in xenografts and CRC mouse models through the induction of cellular senescence and SASP (11). These data revealed a novel mechanism in understanding colon carcinogenesis in which commensal colibactin-producing *E. coli* can induce senescent cells to produce SASP-mediators which in response may propel tumor growth. Nevertheless, the exact relevance of colibactin-producing *E. coli* in CRC *in vivo* remains unknown. Metagenomics analyzes revealed an expansion of *Proteobacteria* (especially *Enterobacteriaceae*, family from which *E. coli* is originating) in the intestinal microbiota of IBD and CRC patients (reviewed by (33)). These adherent invasive *E.
coli (AIEC) were abundantly detected on colonic and ileal mucosal surface of IBD and CRC patients (14, 30). Interestingly, these AIEC associated with colorectal tumors are predominantly from the B2 phylogroup (6, 45). According to the different studies, 50-60% of them express the \textit{pks} island and produce colibactin (2, 4, 6, 30, 41, 45). The presence of the \textit{pks} island is associated with the expression of additional pro-carcinogenic factors, facilitating bacterial translocation or angiogenesis (41) leading to a worsened tumor prognosis (4).

These findings indicate that colibactin-producing B2 \textit{E. coli} could be a critical cofactor in colon carcinogenesis fueling tumor growth through the modulation of its microenvironment (13).
Concluding remarks – Colibactin at the intersection between microbiota metabolome and the host.

Rapid advances in sequencing technologies over the past decade encouraged researchers to assess the link between intestinal microbiota and human health. The extensive access to additional microbial genome sequencing data has allowed the discovery of novel biosynthetic gene clusters encoding for unknown bacterial molecules, especially coming from the secondary metabolism (9). Indeed the importance of specialized metabolites produced by commensal microorganisms in human health has recently emerged (16).

This intestinal microbiome, unlike the human genome, is highly plastic and can be altered by factors such as diet, drugs, antibiotics, pre or probiotics and other microbial-derived products. Alterations in the composition of the intestinal microbiota can therefore alter the intestinal microbiome and its metabolic capabilities finally affecting the host health. However, it can be speculated that the composition of the microbial community or its gene content can be exempted from any modification by environmental exposure while the gene expression may be affected. Thus it is becoming critical not only to identify the type of bacteria which are composing the intestinal microbiota but also to understand what they can produce. A wide range of small molecules, including colibactin, has already been isolated from so called “commensal bacteria”. Despite large efforts from both biologists and chemists to understand its chemical structure and the molecular basis for its biosynthesis and biological activity, colibactin has not yet been isolated but its activity may arise from an array of molecules rather than a specific single one (57).

Considering the fact that colibactin(s) exerts potent effects on host immunity (29, 49), intestinal permeability (39), virulence (32), colorectal cancer (2) and intriguingly is involved in the probiotic activity of *E. coli* Nissle 1917 (37), a better characterization of colibactin biological activities, and to a larger extent to microbiota-derived metabolites, chemistry and function would be of particular interest for understanding the effects of the microbiota on
human health at steady-state and during disease. This will lead to the design of tailored therapeutic strategy to suppress or enhance specific microbial communities according to their production of beneficial or deleterious metabolites. This may help us to reconsider the innoxiousness of intestinal E. coli strains bacteria which were seen as peaceful commensal whereas they could be the enemy within us.

Acknowledgements and financial support

T.S and C.B are supported by the Agence Nationale de la Recherche (ANR-GUI-AAP-06).

We are grateful to Dr. Jean-Philippe Nougayrede for critical reading of this manuscript and many helpful suggestions.
References


16


Figure 1. Modes and sites of *Escherichia coli* colonization. (1) The development of the host-microbiota symbiosis is driven by both developmental and environmental factors especially in early life. The prenatal environment harbors a low abundant microbiota which, combined after birth with early colonizing bacteria from the mother and the environment, will generate signals driving the post-natal maturation of the host-microbiota symbiosis. *E. coli* massively colonized infant gut within hours after birth promoting a favorable environment for the subsequent colonization by anaerobic bacteria which will form the dominant microflora in adults. (2) *Escherichia coli* colonize various sites in the human body. Besides commensal *E. coli* (comEC) and Enteropathogenic *E. coli* (EPEC) living within both small and large intestines, enterotoxigenic *E. coli* (ETEC) colonize the small intestine and cause diarrhoea, while enterohaemorrhagic *E. coli* (EHEC) and adherent and invasive *E. coli* (AIEC) cause disease in the colon; enteroaggregative *E. coli* (EAEC) can colonize both compartments. Septicaemia, meningitis and UTI can occur with infections by Extra-intestinal *E. coli* (ExPEC) comprising neonatal meningitis *E. coli* (NMEC) and Urinary pathogenic *E. coli* (UPEC) which can cross the intestinal barrier into the bloodstream and finally to the central nervous system and the urinary tract.

Figure 2. Colibactin prevalence, biosynthesis and mode of action. (1) Phylogenic tree of a selection of complete *E. coli* genomes showing the ECOR phylogroups and the pathotypes producing colibactin (highlighted in red). (2) Prevalence of the main *E. coli* phylogroups in France between 1980 and 2010 (adapted from (31, 54). (3) B2 *E. coli* harboring the *pks* island produce a peptide-polyketide hybrid(s) named colibactin that is synthetized as a prodrug and need a maturation in the periplasm before its excretion. (4) Colibactin putative chemical structure revealed and electrophilic cluster that may alkylate DNA leading to DNA-double strand breaks (DSB) in eukaryotic cells. (5) The amount of DSB will determine the fate of the intoxicated cells (adapted from (5, 56).

Figure 3. Schematic of colibactin contributions to host health alteration. Expression of *pks* island by *E. coli* enhances their property to colonize host gut (1), to alter intestinal permeability and increase bacterial translocation from the gut lumen (2). Altered immune homeostasis and inflammatory response induced by *pks*+ *E. coli* promotes tumor growth (3) and impairs oral tolerance (4). Systemic dissemination of *pks*+ *E. coli* subsequently lead to sepsis (5).
1-year old Adult
Subdominant gut microbiota
Escherichia coli

10^4 - 10^6 CFU/g feces

Foetus Neonate
~ 10^9 CFU/g feces

Placental bacteria
Pioneer gut bacteria
Escherichia coli

Escherichia coli
Low abundancy

~ 10^8 CFU/g feces

Aero-Anaerobic bacteria

Microbiota diversity

Figure 1

CENTRAL NERVOUS SYSTEM

PERIPHERAL NERVOS SYSTEM

SMALL INTESTINE

COLON

BLOODSTREAM

ExPEC

NMEC

Loose mucus

Outer loose mucus

Inner mucus
Figure 3

- E. coli pks+
- Intestinal translocation
- Intestinal permeability
- Decreased Tregs population
- Altered oral tolerance
- Secretory cell phenotype alteration
- Enhanced gut colonization
- Enhanced gut colonization
- Genotoxicity Gene mutation
- Tumor growth
- Inflammation SASP
- Intestinal translocation
- Intestinal permeability
- Lymph drainage
- Blood drainage
- Systemic dissemination
- Sepsis