#### RESEARCH



## Selection, Comparative Genomics, and Potential Probiotic Features of *Escherichia coli* 5C, a *pks*-Negative Strain Isolated from Healthy Infant Donor Feces

Francesco Di Pierro<sup>1,2,3</sup> · Valeria Sagheddu<sup>4</sup> · Serena Galletti<sup>4</sup> · Alice Casaroli<sup>4</sup> · Edoardo Labrini<sup>4</sup> · Sara Soldi<sup>4</sup> · Massimiliano Cazzaniga<sup>2</sup> · Alexander Bertuccioli<sup>1,5</sup> · Mariarosaria Matera<sup>1,6</sup> · Ilaria Cavecchia<sup>1,7</sup> · Chiara Maria Palazzi<sup>1</sup> · Maria Laura Tanda<sup>8</sup> · Nicola Zerbinati<sup>3</sup>

Accepted: 10 March 2025 © The Author(s) 2025

#### Abstract

Among the emerging issues in probiotic safety, the possible presence of *pks*, a gene cluster synthetizing a genotoxin known as colibactin, is one of the most alarming. Indeed, indigenous *E. coli* strain *pks*-positive are found in 60% of patients with colorectal cancer, and the most widely used *E. coli*-based probiotic, known as *E. coli* Nissle 1917 (DSM 6601), is *pks*-positive. Starting from 25 potential candidates selected by screening 25 infant stool samples, we have selected an *E. coli* strain (named 5C, deposited as LMG S-33222) belonging to the phylotype A and having the serovar O173:H1. Having been previously completely sequenced by our group, we have further characterized this strain, demonstrating that it is (i) devoid of the most known potential pathogenic-related genes, (ii) devoid of possible plasmids, (iii) antibiotic-sensitive according to the EFSA panel, (iv) resistant in gastric and enteric juice, (v) significantly producing acetate, (vi) poorly producing histamine, (vii) endowed with a significant in vitro antipathogenic profile, (viii) promoting a significant in vitro immunological response based on IL-10 and IL-12, and (ix) devoid of the *pks* genes. A comparative genomics versus *E. coli* Nissle 1917 is also provided. Considering that the other two most commonly used *E. coli*-based probiotics (*E. coli* DSM 17252 and *E. coli* A0 34/86) are respectively *pks*-positive and alpha-hemolysin-(*hly*) and cytotoxic necrotizing factor-1-(*cnf1*) positive, this novel strain (*E. coli* 5C) is likely the probiotic *E. coli* strain with the best safety profile available to date for human use.

Keywords LMG S-33222 (5C) · ECP24® · Probiotics · Colibactin · Virulence factors · Mutagenic

## Introduction

### The Concept of Safety for Probiotics

Francesco Di Pierro and Valeria Sagheddu have equally contributed to the paper.

Chiara Maria Palazzi pchiaramaria@gmail.com

- <sup>1</sup> Microbiota International Clinical Society, 10123 Turin, Italy
- <sup>2</sup> Scientific & Research Department, Velleja Research, 20125 Milan, Italy
- <sup>3</sup> Department of Medicine and Technological Innovation, University of Insubria, 21100 Varese, Italy
- <sup>4</sup> AAT-Advanced Analytical Technologies, Fiorenzuola d'Arda, 29017 Piacenza, Italy

The gut microbiota plays a central role in maintaining human health and influences digestive, immune, metabolic, and neurological functions, with probable repercussions also

- <sup>5</sup> Department of Biomolecular Sciences, University of Urbino Carlo Bo, 61122 Urbino, Italy
- <sup>6</sup> Department of Pediatric Emergencies, Misericordia Hospital, 58100 Grosseto, Italy
- <sup>7</sup> Microbiomic Department, Koelliker Hospital, 10134 Turin, Italy
- <sup>8</sup> Endocrine Unit, Department of Medicine and Surgery, University of Insubria, 21100 Varese, Italy

in the oncological field [1-3]. Understanding what affects its composition and diversity and promoting its eubiosis is therefore fundamental [4]. Numerous studies have demonstrated the beneficial effects deriving from the use of probiotics, most frequently attributable to the Bifidobacterium and Lactobacillus species, on human health [5-7]. Although there is consensus on the safety of using known and extensively tested probiotics [8-10], potential adverse events may occur when they are administered to immunocompromised individuals or to subjects with weakened immune systems, such as pediatric patients, transplant recipients, or patients undergoing chemotherapy [11–14]. Sometimes, intestinal discomfort may also occur, which is generally transient and decreases as the microbiota adapts [15]. More rarely, the administration of probiotics may (i) lead to hypersensitivity reactions (skin rashes, itching, and swelling); (ii) alter the effects of treatments by interfering with their efficacy; (iii) be involved in the production and metabolism of histamine (potentially promoting intestinal inflammatory response); or (iv) be associated with the transmission of antibiotic resistance elements or with the production of genotoxic metabolites, such as colibactin [16-23].

### The Probiotic Strain *E. coli* Nissle 1917 and the Cancer-Colibactin Relationship

E. coli Nissle 1917 is a Gram-negative member of the B2 phylogenetic group of Escherichia coli, isolated by Alfred Nissle in 1917 from the stool of a German soldier. Deployed in the Dobruja region for some time, then heavily contaminated by Shigella, this soldier, unlike his comrades, did not develop diarrhea or other intestinal diseases. Dr. A. Nissle hypothesized that he was a carrier of a strain of E. coli capable of direct antagonism towards other possible pathogenic enterobacteria. Indeed, he isolated from a stool sample of him a strain of E. coli that, in laboratory tests, demonstrated antagonistic activity towards other intestinal pathogens [24, 25]. Since then, the strain was formulated as a probiotic supplement and commercialized in Germany (Mutaflor®, Ardeypharm; Germany) and, later, in many other European countries (i. e. EcN®, Cadigroup, Italy). Currently, the strain, hereafter called E. coli Nissle 1917, or simpler Nissle 1917 or EcN, is deposited at the German Collection for Microorganisms and Cell Cultures (Deutsche Sammlung von Mikroorganismen und Zellkulturen, DSMZ), where it has got the designation E. coli DSM 6601 [25]. E. coli Nissle 1917's basic microbiological and molecular features include (i) having common type fimbria (F1A), F1C fimbria, and Curli fimbria (respectively described as involved in biofilm formation and adhesion (F1A), gut and bladder colonization (F1C), and in bacterial aggregation (Curli) [26, 27]; (ii) having siderophores for iron acquisition (anti-inflammatory and/or fitness factors) [26, 28]; (iii) having O6 repeating

units for LPS (a feature commonly found in extra-intestinal pathogenic E. coli strains), K5-type for capsule (described to make serum-resistant the extra-intestinal pathogenic E. coli strains), and H1-type for flagella (responsible for strain motility) [26, 29, 30]; (iv) having two plasmids (pMUT1 and pMUT2) synthetizing microcins M and H47 (active against Salmonella strains) [26, 31, 32]; (v) being strong bile-resistant and colonizing also due to the gene Rfah [26, 33]; and (vi) having the capability of synthetizing and releasing colibactin, a mutagenic substance described to promote colorectal polyposis and cancer, and reducing cancer response to irinotecan [26, 34–39]. As regards the cancer-colibactin relationship, very recent metagenomics-based and population-genomics-based surveys have shown that the prevalence of dominant colibactin-producing lineages of E. coli varies considerably across geographical regions, being higher in countries with a high human development index (HDI) like UK and USA and lower in countries with low HDI, like Pakistan and Bangladesh. This prevalence is strongly associated with the age-standardized incidences of colorectal cancer, bladder cancer, and prostate cancer, suggesting that the degree of colibactin exposure in a population might contribute to the geographical variation of these cancers [40]. Although some of these characteristics reported above from (i) to (vi) may be considered negative for a probiotic, the strain is widely used, especially (i) in the management of patients diagnosed with ulcerative colitis, but in remission, (ii) in childhood gastroenteritis, and (iii) in adult constipation [26, 41–44].

#### Other E. coli Probiotic Strains

E. coli Nissle 1917 is not the only E. coli probiotic currently clinically used. A mixture of E. coli strains, (named G1/2, G3/10, G4/9, G5, G6/7, G8) deposited as DSM 17252 and commercially available as Symbioflor® 2 (SymbioPharm, Germany) is used for treating IBS. Moreover, the strain E. coli A0 34/86, commercially available as Colinfant® New Born (Dyntec, Czech Republic), is mainly used in newborns for preventing infections and atopy [45, 46]. Noteworthy, E. coli A0 34/86, a strain apparently devoid of the pks gene [47], contains at least two potential pathogenic genes (hly, cnf-1) respectively responsible for alpha-hemolysin and cytotoxic necrotizing factor-1 production and release [48]. As regards the E. coli mixture described before (DSM 17252), it is devoid of potential pathogenic genes, but the information about the gene *pks* is not precisely reported [49]. To our knowledge, only the strain E. coli CEC15, isolated from a suckling rat pup and initially developed for probiotic purpose, but never commercialized for human use, is a pks-negative strain [50]. The strain is positive for the hly gene [50]. Starting from these assumptions, we attempted to isolate from infant stool a strain of E. coli that, despite the expression of those fitness factors that characterize this species, presented a safety profile considered acceptable for a probiotic, including the complete absence of the pks gene. We have then made a genome comparison between this newly selected strain and E. coli Nissle 1917. Finally, since we were unable to establish, at least on a bibliographical basis, whether the probiotic registered as DSM 17252 contains the *pks* gene or not, we searched for it via PCR analysis directly in the finished product (Symbioflor® 2).

## **Materials and Methods**

probiotic strains

## Bacterial E. coli Isolation from Human Feces

Fecal samples from 11 healthy children aged between 3 months and 4 years were freshly collected and plated onto TBX plates (Merck, Milan, Italy) and incubated for 24 h at 37 °C under aerobic conditions. Blue colonies, presumably belonging to E. coli species, were isolated and plated for three subculturing steps on TBX agar to ensure their purity. Twenty-five newly isolated strains of E. coli were recovered from the isolation step. The funnel-like approach, presented in Fig. 1, was followed to identify putative new probiotic strains.

## **Species-Specific Identification of New Isolates**

E. coli blue colonies grown on TBX agar were lysed for DNA extraction using the Whatman<sup>TM</sup> CloneSaver<sup>TM</sup> Card System (96-well format) (VWR, Milan, Italy). DNA extraction was performed according to the manufacturer's instructions. Firstly, the identification of E. coli species was

performed by a specific PCR protocol for the 16S rRNA gene using the primers and conditions previously described [51]. The strain E. coli Nissle 1917 was used as a positive control.

## **Pathogenetic Characterization of Isolates**

Before evaluating the antibiotic resistance profile of the 25 newly isolated E. coli strains, we removed those that possessed some genes typically observed in microorganisms considered potentially pathogenic. By multiplex PCR, we analyzed the following genes: shiga-like toxin (stx1 and stx2) [52]; intimin A (eaeA) found in enterohemorrhagic E. coli (EHEC) [53]; enterohemorrhagic E. coli emolysin (EHEC) hlyA) [54, 55]; verocytotoxins 1 (vtx1) and 2 (vtx2), characteristic of verocytotoxin-producing E. coli (VTEC) [56]; intimin (eae) found in enteropathogenic E. coli (EPEC) [57]; heat-stable (estA) and heat-labile enterotoxin (eltA), characteristic of enterotoxigenic E. coli (ETEC) [58, 59]; and invasive plasmid antigen (ipaH), characteristic of enteroinvasive E. coli (EIEC) and Shigella spp. [60, 61]. Genes codifying for *cnf-1* (cytotoxic necrotizing factor 1), *sfa* (S fimbriae), and pap (p- associated-pilus) [62, 63] were investigated following the protocol described by Farshad and colleagues [64].

## **Antibiotic Susceptibility Profiles**

The strains of E. coli not harboring pathogenic genes were tested for their sensitivity to antibiotics according to CLSI (Clinical and Laboratory Standards Institute) M07 [65]. Microdilutions were performed in BBL Mueller Hinton II Broth (Merck, Milan, Italy), at pH  $7.3 \pm 0.1$ . The minimal



inhibitory concentrations (MICs) were determined for 16 antibiotics: ampicillin, penicillin, clindamycin, linezolid (range 0.03 to 16 µg/ml), vancomycin, ciprofloxacin (range 0.25 to 128 µg/mL), neomycin, gentamicin, streptomycin (range 0.5 to 256 µg/ml), kanamycin (range 2 to 1024 µg/ ml), erythromycin, quinupristin-dalfopristin (range 0.016 to 8 µg/ml), tetracycline, chloramphenicol, rifampicin, and trimethoprim (range 0.125 to 64 µg/ml). The antibiotic resistance profiles of strains were evaluated based on the European Food Safety Authority (EFSA) guidelines [66]. To determine the susceptibility or resistance of the newly isolated E. coli strains, we made a comparison between the MIC values determined for each antibiotic and the breakpoints described by EFSA. Furthermore, cefixime susceptibility was considered due to the importance of establishing the probiotics' sensitivity to a cephalosporin. E. coli Nissle 1917 was included as control.

#### Whole Genome Sequencing

DNA extraction, library preparation, and in silico genome analysis were performed on the strains according to methods elsewhere described [67-76]. As also previously reported [77], for the strain E. coli 5C, a total amount of ~3.6 M  $(2 \times 300 \text{ bp})$  reads was generated by Illumina sequencing. After quality filtering and adapter stripping, ~3.3 M (~92.3%) high quality sequences remained, accounting for ~869 M bases. After filtering, ~95.7% of remaining reads showed an average Phred quality score of Q30. Filtered reads were then assembled in contigs with SPAdes. Contigs shorter than 500 bp and below  $2 \times coverage$  were discarded. Read mapping showed a mean coverage of  $182.59 \times$  and a GC content of 58.39%. After inspecting the BLAST4 results (vs. NCBI database), 40 contigs (97.56%) were assigned to Escherichia sp. and considered in the follow-up analysis. Genome assembly evaluation resulted in 40 high-quality contigs, with a total size of 4,712,575 bp and a final GC content of ~ 50.79%. The largest contig was 509,219 bp, and the assembly N50 was 335,055. Genome quality assessment displayed a completeness of ~99.93% while using 1207 single-copy orthologous genes and no significant evidence of contaminant contigs. As regards the strain 8C, a total amount of ~2.1 M ( $2 \times 300$  bp) reads was generated by Illumina sequencing. After quality filtering and adapter stripping, ~1.9 M (~92.8%) high quality sequences remained, accounting for ~ 505 M bases. After filtering, ~ 96% of remaining reads showed an average Phred quality score of Q30. Filtered reads were then assembled in contigs with SPAdes. Contigs shorter than 500 bp and below 2 x coverage were discarded. Read mapping showed a mean coverage of 100×and a GC content of 50.43%. After inspecting the BLAST results (versus NCBI database), 83 contigs (96.51%) were assigned to Escherichia sp. and considered in the follow-up analysis. Genome assembly evaluation resulted in 83 high-quality contigs, with a total size of 4,936,625 bp and a final GC content of ~50.57%. The largest contig was 440,212 bp, and the assembly N50 was 147,936. Genome quality assessment displayed a completeness of ~99.97% while using 1173 single-copy orthologous genes and no significant evidence of contaminant contigs.

### **In Silico pks Detection**

The putative presence of genes related to colibactin production was ascertained using a BLAST and mapping approach, respectively. The sequence of target genes was extracted from the annotated genome of E. coli Nissle 1917 (NZ\_CP058217). For this purpose, a Python script was written to parse the GenBank file of the complete E. coli Nissle 1917 genome, and the genes belonging to the pks island were found in the following order: *clbA*, *clbR*, *clbB*, clbC, clbD, clbE, clbF, clbG, clbH, clbI, clbJ, clbK, clbL, clbM, clbN, clbO, clbP, clbQ, and clbS. Based on the output reference FASTA file, two complementary analyses were conducted, first adopting a BLAST approach and second through read mapping. More in detail, the BLAST search was carried out by building a BLAST database with pks island gene sequences, and a local alignment was performed using previously annotated genes from E. coli 5C and E. coli 8C strains as queries. BLAST tables were then filtered in R using an identity threshold > 80% and coverage > 70%[78]. As a complementary analysis, filtered reads of *E. coli* 5C and 8C strain genomes were mapped against the target genes. More in detail, reads were mapped using BBMap with a minimum read identity of 99% and excluding reads with secondary alignment [79].

## Variant Calling and Comparative Bioinformatic Analysis of *E. coli* 5C Genome and *E. coli* Nissle 1917

Substitutions and insertions/deletions of *E. coli* strain 5C genome versus *E. coli* Nissle 1917 genome as reference were detected using Snippy [80]. Five genome sequences of *E. coli* Nissle strain were downloaded from NCBI (GCF\_000333215: 4514, GCF\_000714595: 4547, GCF\_003546975: 4506, GCF\_019967895: 4505, GCF\_021559835: 4502). Genomes were sequenced between 2013 and 2022. The FASTA file of assembled genomes included plasmid sequences, which were kept in the analysis since they can harbor AMR elements. Assembled genomic contigs of 5C strain were obtained from GCF\_039944155.1 [81].

The presence of putative AMR genes was ascertained using ABRicate [82] against the following databases: NCBI AMRFinderPlus [83], CARD [84], Resfinder [85], ARG-ANNOT [86], MEGARES [87], and EcOH [88]. The databases used for identifying virulence factors were VFDB [89] and E. coli VF [90]. ABRicate was used with default parameters, and only genes with at least 80% identity and 80% coverage were considered as legitimate hits. AMR and VF data were imported in R [91], processed with the dplyr package [92] and plotted using the ggplot2 package [93]. A pangenome analysis of E. coli 5C and E. coli Nissle 1917 was conducted with PPanGGOLiN [94]. Coding sequences (CDS) were parsed with Prodigal [95], tRNAs were found using ARAGORN [96], and rRNAs were annotated using the Infernal command-line tools coupled with the HMM of the bacterial and archaeal rRNAs downloaded from RFAM [97]. Then, the CDS overlapping any RNA genes were deleted, as they are usually false positive calls [98]. All proteins were clustered using MMseqs2 [99] according to 80% identity and 80% coverage; gene families were assigned to "persistent," "shell," or "cloud" partitions and saved.

#### Prophages Detection in E. coli Strains

The PHASTEST tool (Phage Search Tool for *Escherichia coli*), a freely accessible [100] and user-friendly online tool that can identify and classify prophages and other phage-related sequences in bacterial genomes, was used to detect phage sequences in the *E. coli* genomes [101].

## **Rep-PCR Analysis**

Rep-PCR is a molecular technique applied to genomic DNA that allows the genetic profile of bacterial strains belonging to the same species to be compared. We performed Rep-PCR reactions using BoxAR1 (5'- CTACGGCAAGGCGAC CTGACG-3') using thermal cycles and PCR conditions following the method described by Estrada et al. [102]. The genomic DNA of both *E. coli* Nissle 1917 and *E. coli* 5C was analyzed by Rep-PCR using the (GTG)5 (5'-GTGGTG GTGGTGGTG-3') primer and the REP1R-Dt/REP2-Dt couple of primers (REP1R-Dt: 5'-IIINCGNCGNATCNGGC-3' and REP2-Dt: 5'- NCGNCTTATCNGGCCTAC-3') [103] in agreement with the thermal protocol previously described [104]. We loaded Rep-PCR amplification products on agarose gel 2.5% (weight/volume).

#### Serotyping

To address the *E. coli* 5C serotyping determination, we first run the ECTyper script for the in silico prediction [105, 106]. Default parameters were used to complete the analysis. To validate the result obtained, we also made the serotyping determination at the Statens Serum Institut (Copenhagen, Denmark) according to a method previously described [107].

#### Phylotyping

To address the *E. coli* 5C phylotyping determination, we adopt a previously described quadruplex method that exploits aligning the sequences available in Genbank for the *chuA*, *yjaA*, tspE4.*c*2, *arpA*, and *trpA* genes [108].

#### **Resistance to Digestion Process**

The strain of *E. coli* 5C was cultured on BHI (Brain Heart Infusion) broth (Kairosafe, Trieste, Italy). To evaluate its tolerance to simulated gastric juice (SGJ), the broth culture was centrifuged, the supernatant was discarded, and the pellet was resuspended in 10 ml of sterile SGJ ( $pH=1.8/3.4\pm0.1$ ). The ingredients (expressed as mM) contained in the SGJ composition are sodium taurocholate (0.08), phospholipids (0.02), sodium (34), and chlorine (59). The so-prepared tube was incubated at 37 °C; aliquots of 1 ml were removed and serially diluted at three different time points (T0, T30, and T60 min) for determination of total viable counts. The medium chosen for the viable counts was the BHI agar. All plates were incubated for 24 h at 37 °C under aerobic conditions.

Results are expressed as two percentages using the formula  $P = (\mu/M) \times 100$ , where *P* is the percentage of resistance of the microbial strain to the simulated gastric juice;  $\mu$ is the count of live *E. coli* cells in the test sample (in log10) after 30 or 60 (respectively T30 and T60) min of incubation at 37 °C with the simulated gastric juice, and nM is the count of *E. coli* (in log10) at the time of preparation of the cell suspension (T0).

We then performed the tolerance test in simulated intestinal juice (SIJ;  $pH = 8 \pm 0.1$ ) after 240 and 360 min of contact. The resistance assessment consists of viable counts performed at different sampling times: at time zero T0 and after 240 (T240) and 360 (T360) min of incubation at 37 °C. The strains under analysis were grown in BHI broth for 24 h at 37 °C under aerobic conditions; then, the cultures were centrifuged, the supernatants discarded, and the bacterial pellets resuspended in 10 ml of SIJ. One ml of the initial suspensions was serially diluted at time zero and at subsequent experimental times T240 and T360. The ingredients (expressed as g/L) contained in the SIJ composition are pancreatin (1), ox-bile (3), and sodium chloride (9). Calculations for intestinal survival are expressed as those for gastric resistance. E. coli Nissle strain was used as a reference control. All assays were performed in triplicates.

#### Adhesion of E. coli to Human Cell Lines

*E. coli* 5C adhesion to human epithelial cell line HT29-MTX derived from the human colon was assessed. Briefly, the HT29-MTX cell line was routinely cultured in High Glucose DMEM (Dulbecco's modified eagle medium) supplemented with 10% FBS (fetal bovine serum) and antibiotic at 37 °C with 5% CO2 (all sera and media Euroclone, Pero, Milan, Italy). Before the adhesion assay, cells were rinsed with Hank's balanced salt solution, trypsinized, counted, and seeded in a 24-well plate  $(2.5 \times 10^5)$ cell/well). The seeded 24-well plate was incubated for 48 h at 37 °C with 5% CO2 until cells reached the confluence. The day before the test, E. coli strains were inoculated in BHI broth and cultured for 24 h at 37 °C under an aerobic atmosphere. On the day of the adhesion test, the wells seeded with cell lines were checked for confluence (85% or more) and washed with Hank's balanced salt solution. Meanwhile, the strains were prepared for the adhesion assay by washing with sterile distilled water and resuspended in 1% FBS-supplemented DMEM medium. The so-obtained inocula, prepared at a concentration of  $1 \times 10^{6}$  CFU/ml, were serially diluted and plated to determine the viable count of microorganisms. Human cells and bacteria were co-incubated for 60 min at 37 °C with 5% CO2, with a multiplicity of infection (MOI) between microorganisms and human cells in the ratio 5:1. Following incubation, the medium was removed from the infected eukaryotic cells, and the monolayer was washed to discard unbonded bacteria. The monolayer was trypsinized, and the suspension of E. coli and human cell debris was serially diluted, plated on BHI agar, and incubated for 24 h at 37 °C under aerobic conditions. Percentages of adhesion were calculated considering viable raw count data with the formula  $P = (\mu/M) \times 100$ , where P represents the adhesion percentage of E. coli to the HT29-MTX cell line,  $\mu$ represents the viable count of analyzed strains bonded to the human HT29-MTX cell line expressed as a logarithmic value, and M represents the viable count of analyzed strains transformed as a logarithmic value of the inoculum. E. coli strain Nissle 1917 was used as a reference control. Tests were assayed in triplicates.

#### **Histamine Release**

In order to promote the enzyme induction before the screening test, *E. coli* 5C and *E. coli* Nissle 1917 were subcultured 5 times in Nutrient Standard (Millipore 1,07882; Merck, Darmstadt, Germany) containing 0.1% of histidine monohydrochloride as a precursor amino acid, supplemented with 0.005% of pyridoxal-5-phosphate [109]. After 5 days, overnight cultures were centrifuged at 2500 rpm for 15 min, and supernatants were collected. Then, supernatants were analyzed through HPLC. Supernatants were dosed for histamine content three times. Positive histamine-spiked control and blank were included in the analysis.

#### **Assessment of Acetate Production**

Acetic acid production was determined for *E. coli* 5C and *E. coli* Nissle 1917 by a colorimetric commercial kit supplied by Megazyme (Michigan, USA) following the manufacturer's instruction. The acetate release in the supernatants was quantified after overnight incubation at 37 °C in Nutrient Broth (Liofilchem, Teramo, Italy) under aerobic and anaerobic atmosphere. The positive control included in the kit and the blank sample were run alongside the samples.

#### E. coli Antipathogenic Activity

E. coli strains (5C and Nissle 1917), E. faecium (ATCC 19434), E. faecalis (ATCC 19433), E. cloacae (ATCC 13047), K. pneumoniae (ATCC 25955), K. aerogenes (ATCC 13048), C. sakazakii (ATCC 29544), S. enterica (serovar Abony NCTC 6017), and S. enterica (serovar typhimurium DSM 5569) were grown in LAPTg (Lactose, Tryptone, Peptone, Tween 80, glucose) broth and incubated aerobically for 24 h at 37 °C. The activity of E. coli strains against pathogens was assessed as previously reported [26] by means of coculture in LAPTg medium. Fifty microliter of an overnight culture of the putative probiotic E. coli strain and one pathogen culture were incubated in LAPTg medium for an overnight period. One milliliter of these suspensions was serially diluted and plated on Brilliance<sup>™</sup> UTI Clarity (Oxoid, Basingstoke, UK), except for S. enterica strains that were plated on MacConkey agar (Oxoid, Basingstoke, UK) and incubated for 24 h at 37 °C under aerobic conditions. Pathogens were also incubated without the presence of the putative probiotic strain as a reference undisturbed control. An ANOVA followed by Dunnett's T3 multiple comparison test versus the pathogen alone was performed (p < 0.05).

## Assessment of the Immunomodulatory Properties E. coli Strains

Human peripheral blood mononuclear cells (hPBMCs; Lonza, Basel, Swisse) were stimulated with viable cells of each strain (*E. coli* 5C and *E. coli* Nissle 1917) for 24 h at an MOI ratio of 1:10 (hPBMCs: bacteria). Upon arrival, following the manufacturer's instructions, the cells were immediately thawed following the specific protocol and were plated in a 24-multiwell-plate at the concentration of  $1 \times 10^6$ cell/ml in RPMI (Roswell Park Memorial Institute) medium added with 10% heat-inactivated FBS, 2 mM L-glutamine, and 50 µg/ml gentamicin. Cells were left undisturbed for 24 h; then, we started the experiment. We evaluated the immunomodulation potential of the strains by stimulating hPBMCs with probiotic viable cells for 24 h in basal and inflamed (LPS-triggered) conditions (0.1 µg/ml of LPS). At the end of this period, supernatants were collected and used for cytokine quantification. IL-12p70 was quantified to evaluate the pro-inflammatory potential of the strains, whereas IL-10 was used as an anti-inflammatory marker. Each condition was tested in duplicate to obtain a biological replicate. For both IL-12p70 and IL-10 quantification induced by the novel isolated strains, an ANOVA followed by Dunnett's T3 multiple comparison test versus unstimulated cells was performed (p < 0.05 versus negative control). A *t*-test between the two strains (*E. coli* 5C versus *E. coli* Nissle 1917) was also performed.

## Pks Gene Cluster Detection in Finished Products Through PCR

To confirm the absence of the pks gene in E. coli A0 34/86 and to evaluate its possible presence in DSM 17252, we have respectively analyzed the bacterial cells contained in Colinfant® New Born and the mixture of E. coli strains (G1/2, G3/10, G4/9, G5, G6/7, G8) contained in Symbioflor® 2 through PCR, according to the method described previously [110]. We have used as analytical references the strains E. coli Nissle 1917 (positive control) and E. coli 5C (negative control). The pks gene cluster was searched using primers targeting four genes (*clbA*, *clbB*, *clbN*, and *clbQ*) [107]. After incubating overnight, the strains Nissle 1917 and 5C were resuspended in 20 µl of Microlysis Plus (Microzone, Stourbridge, UK) and lysed according to the manufacturer's instructions. One ml of Colinfant® New Born and 1 ml of Symbioflor<sup>®</sup> 2 were centrifuged, the supernatants decanted, and the pellets lysed through the Microlysis Plus protocol. The PCR amplification was performed as previously described [111].

## Results

### Species-Specific Identification, Pathogenetic Characterization and Antibiotic Sensitivity Profiles

Stool samples were collected from infant healthy volunteers, and after the isolation process, 25 strains of presumptive *E. coli* were recovered. All of them were confirmed to belong to *E. coli* species by species-specific PCR [50]. All isolates were tested for the presence of 13 potential virulence factor genes, as described in the "Materials and Methods" section. As shown in Table 1, only two strains (namely 5C and 8C) out of the 25 initial isolates were positively selected, being completely devoid of these virulence factor genes, which were, on the opposite, retrieved in the other 23 strains. As shown in Table 2, both these two strains, as well as the strain Nissle 1917, were found to be safe according to the EFSA probiotics guidelines concerning their antibiotic-susceptibility profile. Also, versus cefixime, a representative of the

cephalosporin class not included in the EFSA panel, the strains resulted to be susceptible according to a set cut-off recently demonstrated [112].

## Genome Analysis, Comparative Genomics, and REP-PCR

According to the Achtman MLST (multi-locus sequence typing) scheme, analysis demonstrated that all (strains 5C, 8C, and Nissle 1917) genomes belonged to the E. coli species, with ST73 [40] being the most prevalent sequence types (STs) identified in Nissle 1917 and not present in strains 5C and 8C. Functional genetic annotation revealed that neither antibiotic resistance genes nor toxin-related genes were detected in prophage sequences observed in the three strains. A detailed review of the MLST analysis performed on the genomes of the three strains, including putative virulence factors, efflux pumps, and plasmids is shown in Online Resource 1. To better identify and classify prophages and other phage-related sequences in bacterial genomes, we have used the PHAST-EST tool [100, 101]. As regards E. coli 5C, we have found three complete sequences: PHAGE Escher HK639 (NCBI Accession: NC\_016158), PHAGE\_Entero\_HK629 (NCBI Accession: NC\_019711), and PHAGE\_Entero\_fiAA91 (NCBI Accession: NC 022750). As regards E. coli 8C, we found one intact and two questionable, since incomplete, sequences. They are respectively phage\_Entero\_DE3 (NCBI Accession: NC\_042057), phage\_Salmon\_118970\_sal3 (NCBI Accession: NC 031940), and phage Entero lambda (NCBI Accession: NC\_001416). As regards E. coli Nissle 1917, the analysis performed on GCF000333215 demonstrated a too short sequence to be analyzed by the PHASTEST tool. The analysis on GCF000714595 demonstrated two intact prophage sequences: phage\_Entero\_lambda (NCBI Accession: NC 001416), also found in strain 8C, and phage Entero c 1 (NCBI Accession: NC\_019706). Both correspond to bacteriophages harmless for humans, specifically infecting only E. coli strains. The analysis on GCF003546975 demonstrated two intact and one incomplete sequences, respectively, phage\_ Entero\_DE3 (NCBI Accession: NC\_042057), also found in strain 8C; phage\_Entero\_c\_1 (NCBI Accession: NC\_019706), also shown by GCF000714595; and Phage\_Gifsy\_1 (NCBI Accession: NC\_010392), a lambda-like phage affecting the Salmonella genus. The analysis performed on GCF019967895 found two complete sequences: phage\_Entero\_lambda (NCBI Accession: NC\_001416) and phage\_Entero\_c\_1 (NCBI Accession: NC\_019706). Both the analysis performed on GCF021559835 and GCF043228065 demonstrated two complete and one incomplete sequences: phage Entero lambda (NCBI Accession: NC\_001416), phage\_Entero\_c\_1 (NCBI Accession: NC\_019706), and phage\_Gifsy\_1 (NCBI Accession: NC\_010392). A detailed review performed on the PHASTEST tool is shown in Online Resource 2.

		Typical genes of pathogenic related microorganisms											
	stx1	stx2	eaeA	hlyA	vtx1	vtx2	eae	estA	eltA	ipaH	cnf-1	sfa	рар
E. coli 1C	-	+	-	+	-	-	-	+	-	+	-	+	+
E. coli 2C	-	-	-	+	-	-	-	-	-	-	-	+	+
E. coli 3C	-	+	-	-	-	-	-	-	-	-	-	-	-
E. coli 4C	-	+	-	-	-	-	-	-	-	-	-	-	-
E. coli 5C	-	-	_	-	-	-	-	-	-	-	-	-	_
E. coli 6C	-	+	-	-	-	-	-	-	-	+	-	-	-
E. coli 7C	-	+	-	-	-	-	-	-	-	+	-	-	-
E. coli 8C	-	-	-	-	-	-	-	-	-	-	-	-	_
E. coli 9C	-	+	-	-	-	-	-	-	-	+	-	-	-
E. coli 10C	-	+	-	+	-	-	-	-	-	-	-	+	+
E. coli 11C	-	+	-	-	-	-	-	-	-	-	-	-	-
E. coli 12C	-	-	-	-	-	-	+	-	-	-	-	-	-
E. coli 13C	-	-	-	+	-	-	-	+	-	-	-	-	-
E. coli 14C	-	-	-	-	-	-	-	+	-	I	I	-	-
E. coli 15C	-	-	-	-	-	-	-	-	-	I	I	-	-
E. coli 16C	-	-	+	-	-	-	-	-	-	I	I	-	-
E. coli 17C	-	-	+	-	-	-	-	-	-	I	I	-	-
E. coli 18C	-	-	+	-	-	-	-	-	-	I	I	-	-
E. coli 19C	-	-	-	+	-	-	-	+	-	-	-	+	+
E. coli 20C	-	+	-	-	-	-	-	-	-	-	-	-	-
E. coli 21C	-	-	-	+	-	-	-	+	-	-	-	-	-
E. coli 22C	-	+	-	-	-	-	-	-	-	+	-	-	-
E. coli 23C	-	+	-	+	-	-	-	-	-	-	-	+	+
E. coli 24C	-	-	-	+	-	-	-	-	-	-	-	+	+
E. coli 25C	-	-	-	-	-	-	-	-	+	-	-	-	-

 Table 1
 Potential pathogenic related genes screening

+: positive amplification fragment; -: negative amplification band. *stx1* and *stx2*: shiga toxin, bacterial toxins inhibiting cellular protein synthesis; *eaeA*: intimin, a toxin causing attaching-effacing lesions in the gut mucosa found in enterohemorrhagic *E. coli* (EHEC); *hlyA*: an emolysin found in EHEC *E. coli*; *vtx1* and *vtx2*: verocytotoxins found in verocytotoxin-producing *E. coli* (VTEC); *eae*: intimin found in enteropathogenic *E. coli* (EHEC); *thuA*: an emolysin found in enterohemorrhagic *E. coli* (ETEC); *eae*: intimin found in enteropathogenic *E. coli* (ETEC); *estA* and *eltA*: heat-stable and heat-labile enterotoxins, characteristic of enterotoxigenic *E. coli* (ETEC); *ipaH*: invasive plasmid antigen, characteristic of enteroinvasive *E. coli* (EIEC) and *Shigella* spp.; *cnf-1*: cytotoxic necrotizing factor-1 found in uropathogenic *E. coli* (UPEC); *sfa*: *S-fimbriae* found in UPEC *E. coli*; *pap*: p-associated-pilus, a virulence factor associated with pyelonephritis (for appropriate references, see "Materials and Methods" section)

Main results of the genome analysis of the two strains of *E. coli* 5C and 8C are summarized in Table 3. As also previously reported [77], for the strain *E. coli* 5C, gene prediction and annotation identified 4664 genes, among which 4357 were coding sequences. Moreover, no genes belonging to the *pks* island for colibactin production nor plasmids were identified [77]. As regards the strain *E. coli* 8C, gene prediction and annotation identified 4998 genes, among which 4602 were coding sequences. As for strain 5C, also for strain 8C, no genes belonging to the *pks* island for colibactin production were identified. In strain 8C, two contigs were identified as plasmid sequences. The presence of plasmids was also confirmed by BLAST search [113]. Plasmid contigs were found to host putative virulence factors 
 Table 2
 Antibiotic resistance

 profiles of strains 5C and 8C, in
 comparison with *E. coli* Nissle

 1917
 1917

MIC value (µg/ml)										
	E. coli Nissle	E. coli 5C	E. coli 8C	EFSA cut-off						
Gentamicin	1	2	1	2						
Kanamycin	8	4	8	8						
Streptomycin	16	16	8	16						
Neomycin	2	2	2	nr						
Tetracycline	4	8	8	8						
Erythromycin	>8	> 8	>8	nr						
Clindamycin	>16	>16	>16	nr						
Chloramphenicol	8	8	16	nr						
Ampicillin	4	4	4	8						
Penicillin	>16	>16	>16	nr						
Vancomycin	>128	>128	>128	nr						
Quinupristin-dalfopristin	>8	> 8	>8	nr						
Linezolid	>16	>16	>16	nr						
Trimethoprim	1	16	2	nr						
Ciprofloxacin	< 0.25	< 0.25	< 0.25	0,06						
Rifampicin	16	32	32	nr						
	E. coli Nissle	E. coli 5C	E. coli 8C	Set cut-off*						
Cefixime	0.25	0.25	0.5	1						

nr not requested

\*According to references no. 108

Table 3Main features derivedfrom the genomic analysis of *E.*coli strains 5C and 8C

d <i>E</i> .		Total size* (bp)	GC content* (%)	Genes	Coding sequences	pks	Plasmids number	Virulence factors on plasmids
	E. coli 5C	4,712,575	50.79	4664	4357	Negative	0	/
	E. coli 8C	4,936,625	50.57	4998	4602	Negative	2	Positive

\*Data reported in the "Materials and Methods" section

(Online Resource 3). Due to this feature, the E. coli 8C isolate was discarded. Differently, the whole draft sequence of E. coli strain 5C was deposited in Genbank with the BioProject no. PRJNA1114436, and the strain was safely deposited at the LMG-BCCM under the LMG S-33222 number [77]. The entire genome shotgun project of E. coli 5C has been submitted to NCBI with the BioProject accession number PRJNA1114436 and the corresponding Gen-Bank accession number JBDPNK000000000.1. The raw reads are available in the Sequence Read Archive under accession number SRR29289048, and the assembly can be accessed using the accession number GCA\_039944155.1. Phylogenetic placement analysis performed on genes *fumC* and fimH, both in strain 5C and in strain Nissle 1917, used as control (Online Resource 4), confirmed that the strain E. coli 5C belongs to the E. coli species [114]. In detail, the strain 5C harbors the alleles fumC95 and fimH121, and the strain Nissle 1917, fumC24 and fimH30. Noteworthy, this last allele found in strain Nissle 1917 is described as a possible pathogenic feature found in antibiotic-resistant E. coli [115]. The in silico evaluation of substitutions and insertions/deletions in the genome of E. coli strain 5C, versus the genome of E. coli Nissle 1917 used as reference, demonstrated a total amount of 82,332 variants detected. In detail, comparative analysis between the strain 5C and the strain Nissle 1917 shows substantial similarity in terms of antimicrobial resistance genes, where most of these putative resistances are species-specific elements attributable to the E. coli non-susceptibility to antibiotics commonly effective against Gram-positive bacteria (Online Resource 5) [26]. Similarly, the number of putative virulence factors, to be considered elements of potential bacterial fitness, is superimposable between the strain 5C and the strain Nissle 1917. Indeed, both strains harbor genes related to adhesins, iron uptake, and flagellar biosynthesis and motility, which are important for bacterial survival and spread, but that are not necessarily related to virulence and/or pathogenicity (Online Resource 6). As shown in the Venn diagram superimposing the 5C and the complete Nissle 1917 genomes (Online Resource 7), the number of coding sequences (CDS) common to both strains' genomes was 3555. The number of CDS unique to the 5C strain was 744. Genes specific to strain E. coli 5C were further analyzed, and functional annotation was performed as described [97]. The number of corresponding proteins that found a match in the eggNOG database was 687. The Clusters of Orthologous Groups (COG) categories summarizing the proteins unique to strain E. coli 5C covered a wide variety of metabolic functions, as shown in Online Resource 8. Many fitness factors seem to be apparently specific for the E. coli strain 5C (Online Resource 9), which demonstrates a certain uniqueness (Online Resource 10). A list of genetic features characterizing the strain E. coli 5C, in comparison with those already known and reported for the strain E. coli Nissle 1917 [26], is reported in Table 4. As regards to adhesion molecules (types of fimbria) [26] E. coli 5C is positive for the fim (common type I fimbria; F1A) and csgA (Curli fimbria) genes, but negative for the foc (F1C fimbria), sfa (S-type fimbria), pap (P-type fimbria), and cfa I/ II (colonizing factor antigens I and II) genes (absence of sfa and *pap* genes is also reported in Table 1). As regards the gene rfaH, a transcriptional regulation factor enhancing the colonizing capability of E. coli strains and their capability to resist bile salts [26, 33], the strain E. coli 5C was found to have it. Regarding the genes of the most important toxins possibly found in E. coli [26], like α-Hemolysin (hly), CNF I (cytotoxic necrotizing factor I; cnf I), H-LT (heat-labile enterotoxin; etx), H-ST (heat-stable enterotoxin; est), and shiga toxins (stx), some of them also reported in Table 1, they were all not found in strain E. coli 5C. As regards iron acquisition (siderophores) [26], we have evaluated the following genes: ent (Enterobactin), iro (Salmochelin), iuclaer (Aerobactin), ybt (Yersiniabactin), and chu (Hemin uptake system). They were all absent in E. coli 5C. Finally, as regards the citrate-dependent iron acquisition capability, both the strain E. coli 5C and Nissle 1917 demonstrated the same positive superimposable gene pattern (Online Resource 11). As regards to bacteriocin genes, the strain Nissle 1917, besides the already described two microcins, known as M and H47 [31, 32], demonstrates the presence



**Fig. 2** *E. coli* strains genomic fingerprinting. 1,5,9: Marker 200 pb; 2: E. coli Nissle 1917 with BoxAR1 primer; 3: *E. coli* 5C with BoxAR1 primer; 6: *E. coli* Nissle 1917 with GTG5 primer; 7: *E. coli* 5C with GTG5 primer; 10: *E. coli* Nissle 1917 with Rep1,2 Dt primer; 11: *E. coli* 5C with Rep1,2 Dt primer; 4,8,12: Negative Template Control

of one colicin and one bottromycin [116], both not phenotypically expressed. The last gene, bottromycin, is genetically present, but likely not phenotypically expressed, also in the strain 5C. It is possible that the bottromycin gene is interrupted in both strains (Online Resource 12). Lastly, as shown in Fig. 2, the *E. coli* 5C genomic DNA was amplified through REP-PCR, and the *E. coli* Nissle 1917 was used as control. The molecular fingerprinting of the two strains was different for all the REP-PCR reactions, indicating their different genomic patterns.

 Table 4. Genetic features concerning adhesion and colonization of the gut and bladder (light grey), toxins release (white) and iron acquisition properties (grey) derived from the genomic analysis of *E. coli* strains 5C (in comparison with *E. coli* Nissle 1917)

	fim	csgA	foc	sfa	pap	cfa I/II	<i>rfaH</i>	hly	cnf I	etx	est	stx	ent	iro	iuc/aer	ybt	chu
E. coli 5C	+	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
Nissle 1917	+	+	+	-	-	-	+	-	-	-	-	-	+	+	+	+	+

*fim*: common type 1 fimbria; *csgA*: Curli fimbria; *foc*: F1C fimbria; *sfa*: S-type fimbria; *pap*: P-type fimbria; *cfa* I/II (colonization factor antigens I/II); *rfaH*: a transcriptional antiterminator gene; *hly*: emolysin; *cnf-1*: cytotoxic necrotizing factor-1; *etx*: heat-labile enterotoxin; *est*: heat-stable enterotoxin; *stx*: shiga toxin; *ent*: enterobactin; *ivc*/*aer*: aerobactin; *ybt*: yersiniabactin; *chu*: hemin uptake system. Data concerning the strain *E. coli* Nissle 1917 are both described in ref. [26] and further confirmed by us by gene alignment

	SGJ (% ± SD) pH 3.4	SGJ (%±SD) pH 1.8	SIJ (% ± SD) pH 8	Adhesion $(\% \pm SD)$	Histamine (ppm±SD)	Acetic acid (g/L $\pm$ SD) aero- bic atmosphere	Acetic acid (g/L $\pm$ SD) anaer- obic atmosphere
<i>E. coli</i> Nissle 1917	$96.3 \pm 4.5$	35.3±8.2	97±2.8	58±5	4±2.2	$0.362 \pm 0.05$	$0.373 \pm 0.04$
E. coli 5C	$99.7 \pm 3.7$	$67.9 \pm 7.4$	$98.3 \pm 1.1$	84±4	$19 \pm 3.6$	$0.366 \pm 0.06$	$0.384 \pm 0.03$

Table 5 Simulated gastric and intestinal resistances, histamine and acetate release for E. coli Nissle 1917 and E. coli 5C

SGJ simulated gastric juice,; SIJ simulated intestinal juice, ppm parts per million

#### Serotyping and Phylotyping

The serotyping determination of strain *E. coli* 5C, performed with the ECTyper script for the in silico prediction, further validated by the SerotypeFinder for whole-genome sequencing-based O and H typing, gold standard for *E. coli* serotyping [117], confirmed the presence of the O173 antigen (LPS) and the H1 antigen (flagella). The K antigen (capsule) does not correspond to any of the known ones. The *E. coli* 5C has therefore the following serovar: O173:KNT:H1 (NT-no typable, Online Resource 13). The serovar for *E. coli* Nissle 1917 is O6:K5:H1 [26]. As regards phylotyping, among the five genes considered in the quadruplex method [108], *E. coli* 5C matches only with the *ArpA* gene, therefore belonging to the phylogroup A (Online Resource 14).

### Simulated Gastric and Intestinal Resistances, Histamine release, and Acetic Acid Production

Some probiotic properties of the E. coli strain 5C, like the gastric and intestinal resistances, the histamine release, and the acetic acid production, are reported in Table 5. The strain 5C and the strain Nissle 1917 similarly survived both in the presence of simulated gastric juice (SGJ) and intestinal juices (SIJ) at pH 3.4, but strain 5C displayed higher tolerance at a lower pH, like 1.8. The adhesive properties of the strain 5C seemed to be more pronounced if compared to the control strain (Nissle 1917). Furthermore, the release of histamine was slightly increased for the strain 5C strain even if at a very low level and not significantly different from Nissle 1917. The acetate production is comparable between the two strains and seemed not to be affected by the presence of oxygen. The resistance of the newly isolated E. coli 5C to the digestion process is comparable to that of the strain Nissle 1917.

#### E. coli 5C Antipathogenic Activity

To evaluate the anti-pathogens' potential role exerted by the strain *E. coli* 5C, we have tested it against strains representative of *K. pneumoniae* (ATCC 25955), *K.*  aerogenes (ATCC 13048), E. cloacae (ATCC 13047), C. sakazakii (ATCC 29544), E. faecium (ATCC 19434), E. faecalis (ATCC 19433), S. enterica (serovar Abony NCTC 6017), and S. enterica (serovar typhimurium DSM 5569). As a reference, we have used the strain E. coli Nissle 1917. According to the results shown in Fig. 3, the strain E. coli 5C displayed a significant antipathogenic activity against K. pneumoniae (Fig. 3A), E. cloacae (Fig. 3C), and the two serovars of S. enterica (Fig. 3G, H). The same performances were registered for the reference E. coli Nissle 1917.

#### E. coli Strain Immunomodulation Properties

To evaluate the potential ability of the strain *E. coli* 5C in modulating the immune response, we stimulated hPB-MCs with this strain for 24 h and quantified the secreted cytokines IL-12p70 and IL-10 by ELISA technique. *E. coli* Nissle 1917 was used as a control strain. In basal conditions, the two strains exerted similar immune modulatory roles (Fig. 4A, B), inducing the release of significant levels of IL-12p70 and IL-10 cytokines but without significant differences. In LPS-triggered conditions (Fig. 4C), the strain *E. coli* 5C induced the release of IL-12p70 in a comparable manner to that induced by the control strain. As shown in Fig. 4D, the strain *E. coli* 5C was able to induce the release of the anti-inflammatory IL-10 cytokine even in these LPS-triggered conditions, but the effect seems to be lower than the one yielded by the control strain.

#### **Pks Gene Cluster Detection in DSM 17252**

We have then analyzed the mixture of *E. coli* strains (G1/2, G3/10, G4/9, G5, G6/7, and G8) contained in Symbioflor® 2 and the probiotic contained in Colinfant® New Born through PCR to highlight the possible presence of the *pks* gene. The analysis has revealed that the *E. coli* mixture is positive for colibactin (Online Resource 15). The low intensity of the signal observed in the electrophoretic run is due to the low bacterial load, relative to the live and viable bacteria formulated in the mixture and found in the finished product. The manufacturer declares, in fact, to inactivate some of the strains in the mixture, making them non-viable. As expected,



Fig. 3 Antipathogenic activity exerted by the strain *E. coli* 5C compared to *E. coli* Nissle 1917 against selected pathogens. Black bars represented the undisturbed pathogen growth, while light and dark

strain *E. coli* Nissle 1917 also resulted to be positive, while *E. coli* A0 34/86 and *E. coli* 5C resulted to be negative for *pks* detection.

## Discussion

Although the first *E. coli* ever isolated (by T. Escherich in 1885, today known as strain NCTC86) was a non-pathogenic strain [118], and although the species is abundantly present in most newborns and around 1% in 90% of healthy adults [119, 120], most of what we know about this species comes from studies performed on pathogenic strains, and little is known about which biotic and abiotic factors influence the presence and distribution of non-pathogenic strains in humans. Most non-pathogenic *E. coli* present in the human intestine belong to phylogroups A, B2, and F [121]. Despite

gray bars displayed the effect of the co-culture of probiotic and pathogen strains. One-way ANOVA with multiple comparisons test versus control: \*p < 0.05 and \*\*p < 0.01

understandable emerging issues in probiotic safety, among probiotics used for both generally healthy consumers and in clinical settings, a few belong to the *E. coli* species [122].

#### The Issue of Colibactin-Releasing E. coli Strains

The most investigated *E. coli* probiotic is the strain Nissle 1917 [123]. This strain has been used for over a century with few reports of adverse events reported in the literature [124]. However, like other members of the B2 phylogenetic group, Nissle 1917 contains the *pks* gene cluster that encodes a series of enzymes that produce a DNA alkylating and double-strand break-inducing genotoxin known as colibactin [23]. Concern arises because, in several mouse models, other colibactin-producing *E. coli* have been observed to promote the development of colorectal cancer [125–131], and human organoids exposed to certain *pks*-positive *E. coli* exhibit colibactin-dependent mutational signatures also found in human colorectal cancers

Fig. 4 Immune modulatory responses in basal (A, B) and in LPS-triggered (C, D) conditions exerted by *E. coli* 5C (versus *E. coli* Nissle 1917). A, B: One-way ANOVA with multiple comparisons test versus control: \*p < 0.05 and \*\*p < 0.01. *t*-test between the two strains does not reach statistical significance. C, D: One-way ANOVA with multiple comparisons test versus LPS-stimulated cells: \*p < 0.05and \*\*p < 0.01. *t*-test between 5C and Nissle strains: # < 0.05



[35]. Nissle 1917 was initially thought not to be genotoxic [131]. However, colibactin was then detected in Nissle 1917 supernatants, and cell lines exposed to such supernatants were found to have clear DNA mutations [132]. In germ-free mice administered with Nissle 1917, intestinal epithelial cells showed evidence of DNA damage [132], which was not observed when mice were inoculated with variants of Nissle 1917 carrying mutations in *clbA* or *clbP*, genes encoded within the *pks* cluster and which are directly involved in colibactin synthesis [47, 133]. To the best of our knowledge, there are no data linking Nissle 1917 with colorectal cancer in humans. Since the risk still exists, a "solution" suggested by many authors is the removal of the

*pks* island [134]. To date, such a modified strain of Nissle 1917 for human use has not yet been produced.

## Selection of a Probiotic *E. coli* Strains Not-Releasing Colibactin

Another chance is to select a new strain of *E. coli*, possibly with similar probiotic features and devoid of the *pks* cluster. With this aim in 2020, we have started selecting from infant stool a new strain of *E. coli*. We decided to start from newborn stool (no older than 6 months of life) because, in that fecal context, it is easier to find a significant number of *E. coli* strains. In healthy newborns, *E. coli* can account, on

average, for 15% of the entire gut microbiota and up to 40% in some subjects [135]. We then decided to start from those newborns that had not undergone antibiotic therapy to avoid finding potentially antibiotic-resistant bacteria. Of course, newborn feces do not guarantee finding pks-negative E. coli strains. Scientific literature has shown that between 27 and 33% of E. coli detectable in the feces of European newborns are pks-positive [136–138]. However, in consideration of the greater possibility of finding E. coli strains in newborns compared to adults [119, 120], we decided to opt for newborns as a possible source of E. coli. As shown in the "Results" section, from an initial list of 25 potential candidates, on the basis of the results obtained by screening them for a wide pattern of virulence factors, we have first selected two strains, E. coli 5C and 8C (both pks-negative and antibiotic-sensitive), preferring then the strain 5C one as the strain 8C was found to contain a plasmid carrying potential virulence factors.

# Detection of Prophage Regions by the PHASTEST Tool

Prophages analysis in E. coli strains should be considered an essential aspect when characterizing a potential probiotic. In strain E. coli 5C, we have found only three complete sequences corresponding to phage Escher HK639, phage Entero\_HK629, and phage\_Entero\_fiAA91. The first, the phage HK639, is a double-stranded DNA virus that infects E. coli. It is classified as a member of the Caudoviricetes class, specifically within the unclassified Caudoviricetes order. It does not encode any known toxins or virulence factors that could harm humans [139]. The second, the phage HK629, is also considered not dangerous for humans or animals. It is not inherently present in all E. coli strains, and it can integrate into the bacterial genome under certain conditions [140]. The third, the phage fiAA91, is not dangerous for humans or animals as well, showing similar features as HK629 [141]. According to a recent depiction of human gut-derived prophages, useful in providing a substantial collection of reference sequences for forthcoming human gut phageome-related investigations and potentially enabling better risk assessment of prophage dissemination [142], E. coli is one of the bacterial species with the higher prevalence of prophages. Indeed, 97.0% of E. coli strains harbor complete prophages, and 87.0% of E. coli strains contain prophage regions. This suggests that E. coli is particularly prone to prophage integration, which aligns with the observation of multiple prophages in a single E. coli genome. For instance, E. coli strain M0110 was found to harbor as many as 18 complete prophages. This indicates that E. coli genomes can accommodate multiple prophages, which is consistent with the presence of the three phages in the genome of the strain E. coli 5C. Nevertheless, some prophages could have the ability to integrate across different bacterial genera, although this is relatively rare ( $\sim 4\%$ of prophages) [142]. However, the presence of multiple prophages in a single E. coli genome suggests that these phages may have a narrow host range, specifically targeting E. coli. This is also supported by the evidence that the considered phages, found in strain 5C, are all associated with the family Enterobacteriaceae to which E. coli belongs. Moreover, the prophages could potentially contribute to the bacterium's adaptability, pathogenicity, or resistance to environmental stresses. This is particularly relevant given that E. coli is a common gut bacterium endowed with many fitness factors. Lastly, the nature of prophages is dynamic. They can switch between lysogenic (integrated) and lytic (active) states. The presence of multiple prophages in a single E. coli genome could indicate that these phages are in a lysogenic state, being integrated into the bacterial chromosome and replicating along with the host. However, under certain conditions (e.g., stress), these prophages could be induced into the lytic cycle, leading to the production of phage particles, potentially causing lysis of the bacterial cell.

## Comparison Between the Main Features of *E. coli* Strains 5C and Nissle 1917

The direct genetic comparison between the *E. coli* strains, 5C and Nissle 1917, highlights that they are quite similar (according to the reference databases, the two strains are completely overlapping by 66.9%; Online Resource 4) but with some differences since, always according to the available databases, the strain E. coli 5C shows genomic uniqueness by 14.0% (Online Resource 4). They belong to different phylotypes, A-type for strain 5C and B2-type for Nissle 1917. Recent findings have shown that phylotypes A and B2 are the most distant from each other phylogenetically among E. coli sensu stricto [107]; that strains responsible for extra-intestinal infection were far more likely to be members of phylogroups B2 than A [143, 144]; and that the B2 phylotype is likely the colibactin-positive most virulent phylogroup among E. coli [145]. As regards serovar, E. coli Nissle 1917 is characterized by an LPS-type, O6, commonly found in extra-intestinal pathogenic E. coli [29, 30, 146]. Moreover, only E. coli Nissle 1917 is endowed with F1C fimbria described to allow bacteria to adhere to the urinary tract [147]. Indeed, when transurethrally inoculated into the bladder of female rats, the Nissle 1917 strain determines the presence of colonies both in the bladder and in the renal tissue [26]. In terms of potential probiotic activity, they show superimposable features. Both are antibiotic-sensible, at least according to the antibiotic panel established by EFSA authorities; both similarly resist in simulated gastric and intestinal juice, and both have a similar capability of histamine and acetate release. Both share a similar pattern of anti-pathogenic activity and immunological properties. The only minimal differences observed between the two strains concern a slightly better gastrointestinal resistance of *E. coli* 5C and a slightly better IL-10 response generated in human cells by the presence of *E. coli* 1917. Noteworthy is the difference between the two strains with regard to the presence of the *pks* genetic island.

#### E. coli 5C, a Putative Safe Probiotic

As shown in our study, the *pks* genetic island is absent in the E. coli 5C but present and productive in the E. coli Nissle 1917 [39]. Considering that the E. coli-based probiotic mixture Symbioflor® 2 is pks-positive and Colinfant® New Born is hly- and cnf-1-positive, the strain E. coli 5C could be considered one of the most, if not the most, safe E. coli-based probiotics available for human use. This could also have a significant impact in the oncology field. Colorectal cancer (CRC) is the third most common cancer worldwide. Approximately 2 million people are diagnosed with CRC each year, of which 50% succumb to the disease [148]. Studies on the gut microbiota have recently provided strong evidence of the involvement of intestinal oncobacteria such as F. nucleatum, enterotoxigenic B. fragilis, and pathogenic E. coli strains [149–152]. The negative role of the latter is due (also) to their ability to produce colibactin through the pks gene. These pkspositive E. coli strains have been detected in approximately 20% of healthy individuals, 40% of patients with inflammatory bowel disease, and 60% of patients with CRC and are therefore associated with the inflamed and neoplastic mucosa, where they can promote both the onset and progression of CRC [126, 153, 154]. Noteworthy, human intestinal organoids injected with pks-positive E. coli, including the strain E. coli Nissle 1917, revealed that colibactin promotes a distinct mutational signature that is also detected in around 12% of patients with CRC [39], particularly in patients diagnosed at a young age [35]. Very recent studies indicate that, in contrast to extremely pathogenic strains like E. coli 11G5, E. coli Nissle 1917 is less oncogenic but may not be considered harmless [155]. This difference in oncogenicity can be explained by observing the actual lower adhesion capabilities to the intestinal epithelium shown by the E. coli Nissle 1917 strain when compared to that of the more oncogenic strains [155]. The idea that a probiotic has the genetic capability to produce colibactin is worrisome, and the safe use of E. coli Nissle 1917 as a probiotic remains very uncertain, also considering the potential for strain variation due to evolutionary changes and/or horizontal gene transfer and recombination events that could increase its adhesion capability. In our opinion, *pks*-negative strains should therefore always be preferred when considering E. coli-based probiotics. Recently, this approach has been preferred by some physicians who have safely and successfully treated patients recurrently suffering from intestinal discomfort following colonoscopy with the strain *E. coli* 5C [156]. With this same application, and with the brand ECP24®, a clinical trial on thousands of patients is being planned and will probably start in 2025.

## Conclusion

Starting from 25 possible candidates, our study allowed us to select an E. coli strain (5C; deposited as LMG S-33222) with fitness and probiotic characteristics similar to those of the E. coli Nissle 1917, but with a main substantial difference: the absence of the pks gene. Overall, its characteristics perhaps make it the progenitor of a long series of possible future E. coli strains capable of creating competition in the host with those indigenous pks-positive E. coli strains, for which a significant correlation between a higher incidence of CRC and urinary tract tumors and their geographical presence has been recently demonstrated [40]. It is worth remembering that cancer research suggests that colibactin causes DNA damage early after exposure [157, 158]. High-resolution shotgun metagenomic screening of newborns in the UK has shown that pks-island-carrying E. coli of the B2 phylotype frequently colonize as early as the first 15 days of life [159]. This observation suggests that colibactin-induced DNA damage may accumulate at a young age, thus contributing to the recently reported increased incidence of early-onset colorectal cancer. Indeed, pks-positive CRC occurs mainly in young people [39]. Perhaps an innovative approach should be considered to prevent, above all, neonatal but also adult, exposure to colibactin in order to reduce the risk of developing colibactin-associated cancers.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12602-025-10522-5.

Acknowledgements We thank Dr. Livio Antonielli, biostatisticians, for the important work done.

Author Contribution FDP, VS, EL, SG,SS Conceptualization; FDP, VS, SG, AC, EL, SS Methodology; FDP, SG, AC, VS Validation; FDP, VS Formal analysis; FDP, SG, AC, VS Investigation; FDP, VS, EL, AC, SS Resources; FDP, VS, SG, AC, SS Data curation; FDP, VS Writing–original draft preparation; FDP, VS, EL, SS, MC, AB, MM, IC, CMP, MLT, NZ Writing–review and editing; FDP, VS, SG, AC, MC, AB, MLT, NZ Visualization; FDP, SS, NZ Supervision; FDP, SS, NZ Project administration.

**Data Availability** No datasets were generated or analysed during the current study.

#### Declarations

**Competing interests** FDP is the Scientific & Research Director in Pharmextracta SpA. IC, MC and AB are Pharmextracta scientific advisers. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

## References

- Ogunrinola GA, Oyewale JO, Oshamika OO, Olasehinde GI (2020) The human microbiome and its impacts on health. Int J Microbiol 2020:8045646. https://doi.org/10.1155/2020/8045646
- Manor O, Dai CL, Kornilov SA, Smith B, Price ND, Lovejoy JC, Gibbons SM, Magis AT (2020) Health and disease markers correlate with gut microbiome composition across thousands of people. Nat Commun 11(1):5206. https://doi.org/10.1038/ s41467-020-18871-1
- Yang Q, Wang B, Zheng Q, Li H, Meng X, Zhou F, Zhang L (2023) A review of gut microbiota-derived metabolites in tumor progression and cancer therapy. Adv Sci (Weinh) 10(15):e2207366. https://doi.org/10.1002/advs.202207366
- Beam A, Clinger E, Hao L (2021) Effect of diet and dietary components on the composition of the gut microbiota. Nutrients 13(8):2795. https://doi.org/10.3390/nu13082795
- Hemarajata P, Versalovic J (2021) Effects of probiotics on gut microbiota: mechanisms of intestinal immunomodulation and neuromodulation. Therap Adv Gastroenterol 6(1):39–51
- Wang X, Zhang P, Zhang X (2021) Probiotics regulate gut microbiota: an effective method to improve immunity. Molecules 26(19):6076. https://doi.org/10.1177/1756283X12459294
- Bodke H, Jogdand S (2022) Role of probiotics in human health. Cureus 14(11):e31313. https://doi.org/10.7759/cureus.31313
- Salemi R, Vivarelli S, Ricci D, Scillato M, Santagati M, Gattuso G, Falzone L, Libra M (2023) Lactobacillus rhamnosus GG cellfree supernatant as a novel anti-cancer adjuvant. J Transl Med 21(1):195. https://doi.org/10.1186/s12967-023-04036-3
- Abdelhamid AG, Ss E-M, El-Dougdoug NK (2019) Probiotic lactobacillus and bifidobacterium strains possess safety characteristics, antiviral activities and host adherence factors revealed by genome mining. Epma J 10(4):337–350. https://doi.org/10. 1007/s13167-019-00184-z
- Kim MJ, Ku S, Kim SY, Lee HH, Jin H, Kang S, Li R, Johnston TV, Park MS, Ji GE (2018) Safety evaluations of bifidobacterium bifidum BGN4 and Bifidobacterium longum BORI. Int J Mol Sci 19(5):1422. https://doi.org/10.3390/ijms19051422
- Redman MG, Ward EJ, Phillips RS (2014) The efficacy and safety of probiotics in people with cancer: a systematic review. Ann Oncol 25(10):1919–1929. https://doi.org/10.1093/annonc/mdu106
- Katkowska M, Garbacz K, Kusiak A (2021) Probiotics: should all patients take them? Microorganisms 9(12):2620. https://doi. org/10.3390/microorganisms9122620
- Oggioni MR, Pozzi G, Valensin PE, Galieni P, Bigazzi C (1998) Recurrent septicemia in an immunocompromised patient due to probiotic strains of Bacillus subtilis. J Clin Microbiol 36(1):325– 326. https://doi.org/10.1128/JCM.36.1.325-326.1998
- Mikucka A, Deptuła A, Bogiel T, Chmielarczyk A, Nurczyńska E, Gospodarek-Komkowska E (2022) Bacteremia caused by probiotic strains of Lacticaseibacillus rhamnosus-case studies

highlighting the need for careful thought before using microbes for health benefits. Pathogens 11(9):977. https://doi.org/10.3390/ pathogens11090977

- 15. Dore MP, Bibbò S, Fresi G, Bassotti G, Pes GM (2019) Side effects associated with probiotic use in adult patients with inflammatory bowel disease: a systematic review and Meta-analysis of randomized controlled trials. Nutrients 11(12):2913. https://doi. org/10.3390/nu11122913
- Martín-Muñoz MF, Fortuni M, Caminoa M, Belver T, Quirce S, Caballero T (2012) Anaphylactic reaction to probiotics. Cow's milk and hen's egg allergens in probiotic compounds. Pediatr Allergy Immunol 23(8):778–784. https://doi.org/10.1111/j.1399-3038.2012.01338.x
- Hwang JB, Kang KJ, Kang YN, Kim AS (2009) Probiotic gastrointestinal allergic reaction caused by saccharomyces boulardii. Ann Allergy Asthma Immunol 103(1):87–88. https://doi.org/10. 1016/S1081-1206(10)60154-8
- Onofrei MI, Ghiciuc CM, Luca CM, Postolache P, Sapaniuc C, Enache Leonte G, Rosu FM (2023) Optimization of therapy and the risk of probiotic use during antibiotherapy in septic critically Ill patients: a narrative review. Medicina (Kaunas) 59(3):478. https://doi.org/10.3390/medicina59030478
- Suez J, Zmora N, Segal E, Elinav E (2019) The pros, cons, and many unknowns of probiotics. Nat Med 25(5):716–729. https:// doi.org/10.1038/s41591-019-0439-x
- Gao C, Major A, Rendon D, Lugo M, Jackson V, Shi Z, Mori-Akiyama Y, Versalovic J (2015) Histamine H2 receptor-mediated suppression of intestinal inflammation by probiotic lactobacillus reuteri. mBio 6(6):e01358-15. https://doi.org/10.1128/mBio. 01358-15
- Tóth AG, Csabai I, Judge MF, Maróti G, Becsei Á, Spisák S, Solymosi N (2021) Mobile antimicrobial resistance genes in probiotics. Antibiot (Basel) 10(11):1287. https://doi.org/10.3390/ antibiotics10111287
- Montassier E, Valdés-Mas R, Batard E, Zmora N, Dori-Bachash M, Suez J, Elinav E (2021) Probiotics impact the antibiotic resistance gene reservoir along the human GI tract in a person-specific and antibiotic-dependent manner. Nat Microbiol 6(8):1043– 1054. https://doi.org/10.1038/s41564-021-00920-0
- Nougayrède JP, Chagneau CV, Motta JP, Bossuet-Greif N, Belloy M, Taieb F, Gratadoux JJ, Thomas M, Langella P, Oswald E (2021) A toxic friend: genotoxic and mutagenic activity of the probiotic strain Escherichia coli nissle 1917. mSphere 6(4):e0062421. https://doi.org/10.1128/mSphere.00624-21
- 24. Nissle A (1925) Weiteres uber Grundlagen und Praxis der Mutaflorbehandlung. Deut Med Wochenschr 44:1809–1813
- Sonnenborn U (2016) Escherichia coli strain Nissle 1917-from bench to bedside and back: history of a special Escherichia coli strain with probiotic properties. FEMS Microbiol Lett 363(19):fnw212. https://doi.org/10.1093/femsle/fnw212
- Sonnenborn U, Schulze J (2009) The non-pathogenic Escherichia coli strain Nissle 1917 – features of a versatile probiotic. Microbial Ecology in Health and Disease 21:122–158. https://doi.org/ 10.3109/08910600903444267
- Barnhart MM, Chapman MR (2006) Curli biogenesis and function. Annu Rev Microbiol 60:131–147. https://doi.org/10.1146/ annurev.micro.60.080805
- Große C, Scherer J, Koch D, Otto M, Taudte N, Grass G (2006) A new ferrous iron-uptake transporter, EfeU (YcdN), from Escherichia coli. Mol Microbiol 62:120–131. https://doi.org/10.1111/j. 1365-2958.2006.05326.x
- 29. Matsumoto T, Mikamo H, Ohge H, Yanagihara K, Weerdenburg E, Go O, Spiessens B, van Geet G, van den Hoven T, Momose A, Hagiwara Y, Nakayama Y, Poolman J, Geurtsen J, Kaku M (2022) Distribution of extraintestinal pathogenic Escherichia coli

O-serotypes and antibiotic resistance in blood isolates collected from patients in a surveillance study in Japan. J Infect Chemother 28(11):1445–1451. https://doi.org/10.1016/j.jiac.2022.07.001

- 30. Weerdenburg E, Davies T, Morrow B, Zomer AL, Hermans P, Go O, Spiessens B, van den Hoven T, van Geet G, Aitabi M, DebRoy C, Dudley EG, Bonten M, Poolman J, Geurtsen J (2023) Global distribution of O serotypes and antibiotic resistance in extraintestinal pathogenic Escherichia coli collected from the blood of patients with bacteremia across multiple surveillance studies. Clin Infect Dis 76(3):e1236–e1243. https://doi.org/10. 1093/cid/ciac421
- Vassiliadis G, Destoumieux-Garzón D, Lombard C, Rebuffat S, Peduzzi J (2010) Isolation and characterization of two members of the siderophore-microcin family, microcins M and H47. Antimicrob Agents Chemother 54(1):288–297. https://doi.org/ 10.1128/AAC.00744-09
- 32. Patzer SI, Baquero MR, Bravo D, Moreno F, Hantke K (2003) The colicin G, H and X determinants encode microcins M and H47, which might utilize the catecholate siderophore receptors FepA, Cir. Fiu and IroN Microbiology (Reading) 149(Pt 9):2557–2570. https://doi.org/10.1099/mic.0.26396-0
- Nagy G, Dobrindt U, Grozdanov L, Hacker J, Emody L (2005) Transcriptional regulation through RfaH contributes to intestinal colonization by Escherichia coli. FEMS Microbiol Lett 244(1):173–180. https://doi.org/10.1016/j.femsle.2005.01.038
- Arthur JC (2020) Microbiota and colorectal cancer: colibactin makes its mark. Nat Rev Gastroenterol Hepatol 17(6):317–318. https://doi.org/10.1038/s41575-020-0303-y
- 35. Pleguezuelos-Manzano C, Puschhhof J, Rosendah Huber A, van Hoeck A, Wood HM, Nomburg J, Gurjao C, Manders F, Dalmasso G, Stege PB, Paganelli FL, Geurts MH, Beumer J, Mizutani T, Miao Y, van der Linden R, van der Elst S, Genomics England Research Consortium, Garcia KC, Top J, Willems RJL, Giannakis MH, Bonnet R, Quirke P, Meyerson M, Cuppen E, van Boxtel H, Clevers H (2020) Mutational signature in colorectal cancer caused by genotoxic pks+ E coli. Nature 580(7802):269–273. https://doi.org/10.1038/ s41586-020-2080-8
- Chen B, Ramazzotti D, Heide T, Spiteri I, Fernandez-Mateos J, James C, Magnani L, Graham TA, Sottoriva A (2023) Contribution of pks+ *E. coli* mutations to colorectal carcinogenesis. Nat Commun 14(1):7827. https://doi.org/10.1038/ s41467-023-43329-5
- 37. Sogari A, Rovera E, Grasso G, Mariella E, Reilly NM, Lamba S, Mauri G, Durinikova E, Vitiello PP, Lorenzato A, Avolio M, Piumatti E, Bonoldi E, Aquilano MC, Arena S, Sartore-Bianchi A, Siena S, Trusolino L, Donalisio M, Russo M, Di Nicolantonio F, Lembo D, Bardelli A (2024) Tolerance to colibactin correlates with homologous recombination proficiency and resistance to irinotecan in colorectal cancer cells. Cell Rep Med 5(2):101376. https://doi.org/10.1016/j.xcrm.2023.101376
- 38. Terlouw D, Boot A, Ducarmon QR, Nooij S, Suerink M, van Leerdam ME, van Egmond D, Tops CM, Zwittink RD, Ruano D, Langers AMJ, Nielsen M, van Wezel T, Morreau H (2024) Enrichment of colibactin-associated mutational signatures in unexplained colorectal polyposis patients. BMC Cancer 24(1):104. https://doi.org/10.1186/s12885-024-11849-y
- 39. Rosendahl Huber A, Pleguezuelos-Manzano C, Puschhof J, Ubels J, Boot C, Saftien A, Verheul M, Trabut LT, Groenen N, van Roosmalen M, Ouyang KS, Wood H, Quirke P, Meijer G, Cuppen E, Clevers H, van Boxtel R (2024) Improved detection of colibactin-induced mutations by genotoxic E. coli in organoids and colorectal cancer. Cancer Cell 42(3):487-496.e6. https://doi.org/10.1016/j.ccell.2024.02.009
- Mäklin T, Taira A, Arredondo-Alonso S, Shao Y, Stratton MR, Lawley TD, Aaltonen LA, Corander J (2024) Geographical

variation in the incidence of colorectal cancer and urinary tract cancer is associated with population exposure to colibactinproducing Escherichia coli. Lancet Microbe Volume 0, Issue 0, 101015.

- 41. Kruis W, Fric P, Pokrotnieks J, Lukás M, Fixa B, Kascák M, Kamm MA, Weismueller J, Beglinger C, Stolte M, Wolff C, Schulze J (2004) Maintaining remission of ulcerative colitis with the probiotic Escherichia coli Nissle 1917 is as effective as with standard mesalazine. Gut 53(11):1617–1623. https://doi.org/10. 1136/gut.2003.037747
- 42. Krammer HJ, Kämper H, von Bünau R, Zieseniss E, Stange C, Schlieger F, Clever I, Schulze J (2006) Probiotische Arzneimitteltherapie mit *E. coli* Stamm Nissle 1917 (EcN): Ergebnisse einer prospektiven Datenerhebung mit 3,807 Patienten [Probiotic drug therapy with *E. coli* strain Nissle 1917 (EcN): results of a prospective study of the records of 3,807 patients]. Z Gastroenterol 44(8):651–656. https://doi.org/10.1055/s-2006-926909
- 43. Henker J, Laass M, Blokhin BM, Bolbot YK, Maydannik VG, Elze M, Wolff C, Schulze J (2007) The probiotic Escherichia coli strain Nissle 1917 (EcN) stops acute diarrhoea in infants and toddlers. Eur J Pediatr 166(4):311–318. https://doi.org/10.1007/ s00431-007-0419-x
- 44. Henker J, Laass MW, Blokhin BM, Maydannik VG, Bolbot YK, Elze M, Wolff C, Schreiner A, Schulze J (2008) Probiotic Escherichia coli Nissle 1917 versus placebo for treating diarrhea of greater than 4 days duration in infants and toddlers. Pediatr Infect Dis J 27(6):494–499. https://doi.org/10.1097/INF.0b013 e318169034c
- Enck P, Zimmermann K, Menke G, Klosterhalfen S (2009) Randomized controlled treatment trial of irritable bowel syndrome with a probiotic E.-coli preparation (DSM17252) compared to placebo. Z Gastroenterol 47(2):209–14. https://doi.org/10.1055/ s-2008-1027702
- 46. Micenková L, Bosák J, Smatana S, Novotný A, Budinská E, Šmajs D (2020) Administration of the probiotic Escherichia coli strain A0 34/86 resulted in a stable colonization of the human intestine during the first year of life. Probiotics Antimicrob Proteins 12(2):343–350. https://doi.org/10.1007/s12602-019-09548-3
- 47. Fedrová P, Hrala M, Tom N, Micenková L, Nascimento AMA, Bosák J, Šmajs D. Complete genome sequences of five Escherichia coli strains with probiotic attributes. Microbiol Resour Announc. 2023 Sep 19;12(9):e0036323. https://doi.org/10.1128/ MRA.00363-23. Epub 2023 Aug 7. PMID: 37548468; PMCID: PMC10508162.
- Sheshko V, Hejnova J, Rehakova Z, Sinkora J, Faldyna M, Alexa P, Felsberg J, Nemcova R, Bomba A, Sebo P (2006) HlyA knock out yields a safer Escherichia coli A0 34/86 variant with unaffected colonization capacity in piglets. FEMS Immunol Med Microbiol 48(2):257–266. https://doi.org/10.1111/j.1574-695X. 2006.00140.x
- Zschüttig A, Auerbach C, Meltke S, Eichhorn C, Brandt M, Blom J, Goesmann A, Jarek M, Scharfe M, Zimmermann K, Wassenaar TM, Gunzer F (2015) Complete sequence of probiotic Symbioflor 2 Escherichia coli strain G3/10 and draft sequences of Symbioflor 2 *E. coli* strains G1/2, G4/9, G5, G6/7, and G8. Genome Announc 3(2):e01330-14. https://doi.org/10.1128/genomeA.01330-14
- 50. da Silva TF, Glória RA, de Sousa TJ, Americo MF, Freitas ADS, Viana MVC, de Jesus LCL, da Silva Prado LC, Daniel N, Ménard O, Cochet MF, Dupont D, Jardin J, Borges AD, Fernandes SOA, Cardoso VN, Brenig B, Ferreira E, Profeta R, Aburjaile FF, de Carvalho RDO, Langella P, Le Loir Y, Cherbuy C, Jan G, Azevedo V, Guédon É (2023) Comprehensive probiogenomics analysis of the commensal Escherichia coli CEC15 as a potential

probiotic strain. BMC Microbiol 23(1):364. https://doi.org/10. 1186/s12866-023-03112-4

- 51. Matsuki T, Watanabe K, Fujimoto J, Takada T, Tanaka R (2004) Use of 16S rRNA gene-targeted group-specific primers for realtime PCR analysis of predominant bacteria in human feces. Appl Environ Microbiol 70(12):7220–7228. https://doi.org/10.1128/ AEM.70.12.7220-7228.2004
- Melton-Celsa AR (2014). Shiga Toxin (Stx) Classification, structure, and function. Microbiol Spectr 2(4):EHEC-0024–2013. https://doi.org/10.1128/microbiolspec.EHEC-0024-2013.
- 53. Agin TS, Cantey JR, Boedeker EC, Wolf MK (1996) Characterization of the eaeA gene from rabbit enteropathogenic Escherichia coli strain RDEC-1 and comparison to other eaeA genes from bacteria that cause attaching-effacing lesions. FEMS Microbiol Lett 144(2–3):249–258. https://doi.org/10.1111/j.1574-6968. 1996.tb08538.x
- 54. Johnsen N, Hamilton ADM, Greve AS, Christensen MG, Therkildsen JR, Wehmöller J, Skals M, Praetorius HA (2019) α-Haemolysin production, as a single factor, causes fulminant sepsis in a model of Escherichia coli-induced bacteraemia. Cell Microbiol 21(6):e13017. https://doi.org/10.1111/cmi.13017
- 55. Fagan PK, Hornitzky MA, Bettelheim KA, Djordjevic SP (1999) Detection of shiga-like toxin (stx1 and stx2), intimin (eaeA), and enterohemorrhagic Escherichia coli (EHEC) hemolysin (EHEC hlyA) genes in animal feces by multiplex PCR. Appl Environ Microbiol 65(2):868–872. https://doi.org/10.1128/AEM.65.2. 868-872.1999
- Messens W, Bolton D, Frankel G, Liebana E, Mclauchlin J, Morabito S, Oswald E, Threlfall EJ (2015) Defining pathogenic verocytotoxin-producing Escherichia coli (VTEC) from cases of human infection in the European Union, 2007–2010. Epidemiol Infect 143(8):1652–1661. https://doi.org/10.1017/S095026881 400137X
- 57. Jerse AE, Kaper JB (1991) The eae gene of enteropathogenic Escherichia coli encodes a 94-kilodalton membrane protein, the expression of which is influenced by the EAF plasmid. Infect Immun 59(12):4302–4309. https://doi.org/10.1128/iai.59.12. 4302-4309.1991
- Joffré E, von Mentzer A, Svennerholm AM, Sjöling Å (2016) Identification of new heat-stable (STa) enterotoxin allele variants produced by human enterotoxigenic Escherichia coli (ETEC). Int J Med Microbiol 306(7):586–594. https://doi.org/10.1016/j. ijmm.2016.05.016
- Joffré E, Sjöling Å (2016) The LT1 and LT2 variants of the enterotoxigenic Escherichia coli (ETEC) heat-labile toxin (LT) are associated with major ETEC lineages. Gut Microbes 7(1):75–81. https://doi.org/10.1080/19490976.2015.1127480.Erratum.In:doi: 10.1128/JB.02050-14
- 60. Vu DT, Sethabutr O, Von Seidlein L, Tran VT, Do GC, Bui TC, Le HT, Lee H, Houng HS, Hale TL, Clemens JD, Mason C, Dang DT (2004) Detection of Shigella by a PCR assay targeting the ipaH gene suggests increased prevalence of shigellosis in Nha Trang. Vietnam J Clin Microbiol 42(5):2031–2035. https://doi. org/10.1128/JCM.42.5.2031-2035.2004
- Persson S, Olsen KE, Scheutz F, Krogfelt KA, Gerner-Smidt P (2007) A method for fast and simple detection of major diarrhoeagenic Escherichia coli in the routine diagnostic laboratory. Clin Microbiol Infect 13(5):516–524. https://doi.org/10.1111/j. 1469-0691.2007.01692.x
- Boquet P (2001) The cytotoxic necrotizing factor 1 (CNF1) from Escherichia coli. Toxicon 39(11):1673–1680. https://doi.org/10. 1016/s0041-0101(01)00154-4
- 63. Rahdar M, Rashki A, Miri HR, Rashki Ghalehnoo M (2015) Detection of pap, sfa, afa, foc, and fim adhesin-encoding operons in uropathogenic Escherichia coli isolates collected from

patients with urinary tract infection. Jundishapur J Microbiol 8(8):e22647. https://doi.org/10.5812/jjm.22647

- 64. Farshad S, Emamghoraishi F, Japoni A (2010) Association of virulent genes hly, sfa, cnf-1 and pap with antibiotic sensitivity in Escherichia coli strains isolated from children with community-acquired UTI. Iran Red Crescent Med J 12:33–37
- Clinical and Laboratory Standards Institute 2009. Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically; approved standard, 8th ed. M07-A8 CLSI, Wayne, PA.
- 66. European Food Safety Authority (EFSA) (2012) Panel on additives and products or substances used in animal feed (FEEDAP) guidance on the assessment of bacterial susceptibility to antimicrobials of human and veterinary importance. EFSA J 10:2740– 2750. https://doi.org/10.2903/j.efsa.2012.2740
- 67. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455–477. https://doi.org/10.1089/cmb.2012.0021
- Okonechnikov K, Conesa A, García-Alcalde F (2016) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32(2):292–294. https://doi.org/ 10.1093/bioinformatics/btv566
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/ 1471-2105-10-421
- Laetsch DR, Blaxter ML (2017) BlobTools: interrogation of genome assemblies. F1000Research 6:1287. https://doi.org/10. 12688/f1000research.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–1075. https://doi.org/10.1093/bioinformatics/btt086
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25(7):1043–1055. https://doi.org/10.1101/gr.186072.114
- Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH (2019) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36(6):1925–1927. https://doi. org/10.1093/bioinformatics/btz848
- Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14):2068–2069. https://doi.org/10.1093/bioin formatics/btu153
- Schwengers O, Barth P, Falgenhauer L, Hain T, Chakraborty T, Goesmann A (2020) Platon: identification and characterization of bacterial plasmid contigs in short-read draft assemblies exploiting protein sequence-based replicon distribution scores. Microb Genom 6(10):mgen000398. https://doi.org/10.1099/mgen.0. 000398
- Olson ND, Morrow JB (2012) DNA extract characterization process for microbial detection methods development and validation. BMC Res Notes 5:668. https://doi.org/10.1186/1756-0500-5-668
- 77. Di Pierro F, Zerbinati N, Guasti L, Cazzaniga M, Bertuccioli A, Palazzi CM, Labrini E, Sagheddu V, Soldi S (2024) Draft genome sequence of non-pathogenic Escherichia coli 5C LMG S-33222, isolated from healthy donor feces. Microbiol Resour Announc 13(10):e0058024. https://doi.org/10.1128/mra. 00580-24
- 78. Wickham H, Averick M, Bryan J, Chang W, D'Agostino McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen LT, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan

D, Wilke C, Woo K, Yutani H (2019) Welcome to the Tidyverse JOSS 4(43):1686. https://doi.org/10.21105/joss.01686

- Bushnell B, Rood J, Singer E (2017) BBMerge accurate paired shotgun read merging via overlap. PLoS ONE 12(10):e0185056. https://doi.org/10.1371/journal.pone.0185056
- Seemann, T (2020) Snippy. Github https://github.com/tseemann/ snippy
- Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast allin-one FASTQ preprocessor. Bioinformatics 34(17):i884–i890. https://doi.org/10.1093/bioinformatics/bty560
- Seemann, T. ABRicate. (2023). https://github.com/tseemann/ abricate
- 83. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, Hoffmann M, Pettengill JB, Prasad AB, Tillman GE, Tyson GH, Klimke W (2021) AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep 11(1):12728. https://doi.org/10.1038/ s41598-021-91456-0
- 84. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen AV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran HK, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG (2020) CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48(D1):D517–D525. https://doi.org/10.1093/nar/gkz935
- Florensa AF, Kaas RS, Clausen PTLC, Aytan-Aktug D, Aarestrup FM (2022) ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. Microb Genom 8(1):000748. https://doi.org/10.1099/mgen.0. 000748
- Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, Rolain JM (2014) ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 58(1):212–220. https:// doi.org/10.1128/AAC.01310-13
- 87. Bonin N, Doster E, Worley H, Pinnell LJ, Bravo JE, Ferm P, Marini S, Prosperi M, Noyes N, Morley PS, Boucher C (2023) MEGARes and AMR++, v3.0: an updated comprehensive database of antimicrobial resistance determinants and an improved software pipeline for classification using high-throughput sequencing. Nucleic Acids Res 51(D1):D744–D752. https://doi. org/10.1093/nar/gkac1047
- Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE (2014) SRST2: rapid genomic surveillance for public health and hospital microbiology labs. Genome Med 6(11):90. https://doi.org/10.1186/s13073-014-0090-6
- Liu B, Zheng D, Jin Q, Chen L, Yang J (2019) VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res 47(D1):D687–D692. https://doi.org/ 10.1093/nar/gky1080
- GitHub phac-nml/ecoli\_vf: curated virulence factors for Escherichia coli. https://github.com/phac-nml/ecoli\_vf.
- 91. The Comprehensive R Archive Network. https://cran.r-project. org/.
- Wickham, H., Francois, R., Henry, L., and Müller, K. (2023), "dplyr: a grammar of data manipulation," R Package Version 0.7.4.
- 93. Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, Dunnington D. (2023). RStudio Ggplot2: create elegant data visualisations using The grammar of graphics; 2021. URL

https://CRAN.R-project.org/package=ggplot2. R package version, 3(5), 233.

- 94. Gautreau G, Bazin A, Gachet M, Planel R, Burlot L, Dubois M, Perrin A, Médigue C, Calteau A, Cruveiller S, Matias C, Ambroise C, Rocha EPC, Vallenet D (2020) PPanGGOLiN: depicting microbial diversity via a partitioned pangenome graph. PLoS Comput Biol 16(3):e1007732. https://doi.org/10.1371/journal.pcbi.1007732. (Erratum.In:PLoSComputBiol.2021D ec10;17(12):e1009687.doi:10.1371/journal.pcbi.1009687)
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https:// doi.org/10.1186/1471-2105-11-119
- Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32(1):11–16. https://doi.org/10.1093/nar/gkh152
- 97. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR (2003) Rfam: an RNA family database. Nucleic Acids Res 31(1):439–441. https://doi.org/10.1093/nar/gkg006
- Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47(D1):D309-D314. https://doi.org/10.1093/nar/gky1085.
- Steinegger M, Söding J (2018) Clustering huge protein sequence sets in linear time. Nat Commun 9(1):2542. https://doi.org/10. 1038/s41467-018-04964-5
- 100. https://phast.wishartlab.com/ visited in March 1st, 2025.
- 101. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, Stothard P, Gautam V (2023) PHASTEST: faster than PHASTER, better than PHAST. Nucleic Acids Res 51(W1):W443–W450. https://doi.org/10.1093/nar/gkad382
- 102. Estrada-de Los Santos P, Vacaseydel-Aceves NB, Martínez-Aguilar L, Cruz-Hernández MA, Mendoza-Herrera A, Caballero-Mellado J (2011) Cupriavidus and Burkholderia species associated with agricultural plants that grow in alkaline soils. J Microbiol 49(6):867–876. https://doi.org/10.1007/ s12275-011-1127-9
- 103. Versalovic J, Koeuth T, Lupski JR (1991) Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. Nucleic Acids Res 19(24):6823–6831. https://doi.org/10.1093/nar/19.24.6823
- 104. Gevers D, Huys G, Swings J (2001) Applicability of rep-PCR fingerprinting for identification of Lactobacillus species. FEMS Microbiol Lett 205(1):31–36. https://doi.org/10.1111/j.1574-6968.2001.tb10921.x
- 105. https://github.com/phac-nml/ecoli\_serotyping; last access in November 16th, 2024.
- 106. Bessonov K, Laing C, Robertson J, Yong I, Ziebell K, Gannon VPJ, Nichani A, Arya G, Nash JHE, Christianson S (2021) ECTyper: in silico Escherichia coli serotype and species prediction from raw and assembled whole-genome sequence data. Microb Genom 7(12):000728. https://doi.org/10.1099/mgen.0. 000728
- 107. Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F (2015) Rapid and easy in silico serotyping of Escherichia coli isolates by use of whole-genome sequencing data. J Clin Microbiol 53(8):2410–2426. https://doi.org/10.1128/JCM.00008-15
- Clermont O, Christenson JK, Denamur E, Gordon DM (2013) The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. Environ Microbiol Rep 5(1):58–65. https://doi.org/10.1111/ 1758-2229.12019
- Bover-Cid S, Holzapfel WH (1999) Improved screening procedure for biogenic amine production by lactic acid bacteria. Int

J Food Microbiol 53(1):33–41. https://doi.org/10.1016/s0168-1605(99)00152-x

- 110. Shi Q, Quan J, Lan P, Huang D, Zhou J, Jiang Y, Yu Y (2020) Prevalence and characteristics of pks gene cluster harbouring Klebsiella pneumoniae from bloodstream infection in China. Epidemiol Infect 148:e69. https://doi.org/10.1017/S095026882 0000655
- 111. Luo C, Chen Y, Hu X, Chen S, Lin Y, Liu X, Yang B (2023) Genetic and functional analysis of the pks gene in clinical Klebsiella pneumoniae isolates. Microbiol Spectr 11(4):e0017423. https://doi.org/10.1128/spectrum.00174-23
- 112. Zaragoza G, Pérez-Vázquez M, Villar-Gómara L, González-Prieto A, Oteo-Iglesias J, Alós JI (2024) Community emergence of cefixime-resistant Escherichia coli belonging to ST12 with chromosomal AmpC hyperproduction. Antibiotics (Basel) 13(3):218. https://doi.org/10.3390/antibiotics13030218
- 113. Li D, Li P, Peng M, Zhao X, Jiang X, Wang D, Yuan Y, Guo Q, Wang M, Xu X, Wang M (2022) Transmission barrier of the blaKPC plasmid mediated by type I restriction-modification systems in Escherichia coli. J Antimicrob Chemother 77(4):952–956. https://doi.org/10.1093/jac/dkab489
- 114. Oliveira GS, Lentz SA, Wink PL, Martins AF (2023) Molecular typing of mcr-1 Escherichia coli isolates from pigs and farm environment based on fumC and fimH alleles. Future Microbiol 18:481–488. https://doi.org/10.2217/fmb-2022-0173
- 115. Duggett N, Ellington MJ, Hopkins KL, Ellaby N, Randall L, Lemma F, Teale C, Anjum MF (2021) Detection in livestock of the human pandemic Escherichia coli ST131 fimH30(R) clone carrying blaCTX-M-27. J Antimicrob Chemother 76(1):263– 265. https://doi.org/10.1093/jac/dkaa407
- 116. Franz L, Kazmaier U, Truman AW, Koehnke J (2021) Bottromycins - biosynthesis, synthesis and activity. Nat Prod Rep 38(9):1659–1683. https://doi.org/10.1039/d0np00097c
- 117. Fratamico PM, DebRoy C, Liu Y, Needleman DS, Baranzoni GM, Feng P (2016) Advances in molecular serotyping and subtyping of Escherichia coli. Front Microbiol 7:644. https://doi.org/ 10.3389/fmicb.2016.00644
- Méric G, Hitchings MD, Pascoe B, Sheppard SK (2016) From Escherich to the Escherichia coli genome. Lancet Infect Dis 16:634–636. https://doi.org/10.1016/S1473-3099(16)30066-4
- Tenaillon O, Skurnik D, Picard B, Denamur E (2010) The population genetics of commensal Escherichia coli. Nat Rev Microbiol 8:207–217. https://doi.org/10.1038/nrmicro2298
- Mueller NT, Bakacs E, Combellick J, Grigoryan Z, Dominguez-Bello MG (2015) The infant microbiome development: mom matters. Trends Mol Med 21:109–117. https://doi.org/10.1016/j. molmed.2014.12.002
- 121. Martinson JNV, Walk ST (2020) Escherichia coli residency in the gut of healthy human adults. EcoSal Plus 9(1):10. https://doi.org/10.1128/ecosalplus.ESP-0003-2020
- 122. Merenstein D, Pot B, Leyer G, Ouwehand AC, Preidis GA, Elkins CA, Hill C, Lewis ZT, Shane AL, Zmora N, Petrova MI, Collado MC, Morelli L, Montoya GA, Szajewska H, Tancredi DJ, Sanders ME (2023) Emerging issues in probiotic safety: 2023 perspectives. Gut Microbes 15(1):2185034. https://doi.org/10. 1080/19490976.2023.2185034
- de Vrese M, Schrezenmeir J (2008) Probiotics, prebiotics, and synbiotics. Adv Biochem Eng Biotechnol 111:1–66. https://doi. org/10.1007/10\_2008\_097
- 124. Guenther K, Straube E, Pfister W, Guenther A, Huebler A (2010) Sever sepsis after probiotic treatment with Escherichia coli NISSLE 1917. Pediatr Infect Dis J 29(2):188–189. https://doi. org/10.1097/INF.0b013e3181c36eb9
- 125. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan TJ, Campbell BJ, Abujamel T, Dogan B, Rogers

AB, Rhodes JM, Stintzi A, Simpson KW, Hansen JJ, Keku TO, Fodor AA, Jobin C (2012) Intestinal inflammation targets cancerinducing activity of the microbiota. Science 338(6103):120–123. https://doi.org/10.1126/science.1224820

- 126. Arthur JC, Gharaibeh RZ, Mühlbauer M, Perez-Chanona E, Uronis JM, McCafferty J, Fodor AA, Jobin C (2014) Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer. Nat Commun 5:4724. https:// doi.org/10.1038/ncomms5724
- 127. Tomkovich S, Yang Y, Winglee K, Gauthier J, Mühlbauer M, Sun X, Mohamadzadeh M, Liu X, Martin P, Wang GP, Oswald E, Fodor AA, Jobin C (2017) Locoregional effects of microbiota in a preclinical model of colon carcinogenesis. Cancer Res 77(10):2620–2632. https://doi.org/10.1158/0008-5472. CAN-16-3472
- 128. Cougnoux A, Dalmasso G, Martinez R, Buc E, Delmas J, Gibold L, Sauvanet P, Darcha C, Déchelotte P, Bonnet M, Pezet D, Wodrich H, Darfeuille-Michaud A, Bonnet R (2014) Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype. Gut 63(12):1932–1942. https://doi.org/10.1136/gutjnl-2013-305257
- 129. Dalmasso G, Cougnoux A, Delmas J, Darfeuille-Michaud A, Bonnet R (2014) The bacterial genotoxin colibactin promotes colon tumor growth by modifying the tumor microenvironment. Gut Microbes 5:675–680. https://doi.org/10.4161/19490976. 2014.969989
- 130. Bonnet M, Buc E, Sauvanet P, Darcha C, Dubois D, Pereira B, Déchelotte P, Bonnet R, Pezet D, Darfeuille-Michaud A (2014) Colonization of the human gut by *E. coli* and colorectal cancer risk. Clin Cancer Res 20(4):859–67. https://doi.org/10.1158/ 1078-0432.CCR-13-1343
- 131. Dubbert S, Klinkert B, Schimiczek M, Wassenaar TM, Bünau RV (2020) No genotoxicity is detectable for Escherichia coli strain Nissle 1917 by standard in vitro and in vivo tests. Eur J Microbiol Immunol (Bp) 10(1):11–19. https://doi.org/10.1556/ 1886.2019.00025
- 132. Dubbert S, von Bünau R (2021) A Probiotic Friend mSphere 6(6):e0085621. https://doi.org/10.1128/mSphere.00856-21
- 133. Martin P, Marcq I, Magistro G, Penary M, Garcie C, Payros D, Boury M, Olier M, Nougayrède JP, Audebert M, Chalut C, Schubert S, Oswald E (2013) Interplay between siderophores and colibactin genotoxin biosynthetic pathways in Escherichia coli. PLoS Pathog 9(7):e1003437. https://doi.org/10.1371/journal.ppat.1003437
- 134. Lynch JP, Goers L, Lesser CF (2022) Emerging strategies for engineering Escherichia coli Nissle 1917-based therapeutics. Trends Pharmacol Sci 43(9):772–786. https://doi.org/10. 1016/j.tips.2022.02.002
- 135. Mancabelli L, Tarracchini C, Milani C, Lugli GA, Fontana F, Turroni F, van Sinderen D, Ventura M (2020) Multi-population cohort meta-analysis of human intestinal microbiota in early life reveals the existence of infant community state types (ICSTs). Comput Struct Biotechnol J 15(18):2480–2493. https://doi.org/10.1016/j.csbj.2020.08.028
- 136. Nowrouzian F, Hesselmar B, Saalman R, Strannegard IL, Aberg N, Wold AE, Adlerberth I (2003) Escherichia coli in infants' intestinal microflora: colonization rate, strain turnover, and virulence gene carriage. Pediatr Res 54(1):8–14. https:// doi.org/10.1203/01.PDR.0000069843.20655.EE
- 137. Faïs T, Delmas J, Barnich N, Bonnet R, Dalmasso G (2018) Colibactin: more than a new bacterial toxin. Toxins (Basel) 10(4):151. https://doi.org/10.3390/toxins10040151
- 138. Secher T, Brehin C, Oswald E (2016) Early settlers: which E. coli strains do you not want at birth? Am J Physiol Gastrointest Liver Physiol 311(1):G123-9. https://doi.org/10.1152/ ajpgi.00091.2016

- 139. Neumann B, Lippmann N, Wendt S, Karlas T, Lübbert C, Werner G, Pfeifer Y, Schuster CF (2022) Recurrent bacteremia with a hypermucoviscous Escherichia coli isolated from a patient with perihilar cholangiocarcinoma: insights from a comprehensive genome-based analysis. Ann Clin Microbiol Antimicrob 21(1):28. https://doi.org/10.1186/s12941-022-00521-7
- 140. Naha A, Cameron TA, Margolin W (2025) A predicted helixturn-helix core is critical for bacteriophage Kil peptide to disrupt Escherichia coli cell division. Antibiotics (Basel) 14(1):52. https://doi.org/10.3390/antibiotics14010052
- 141. Qi Z, Meng B, Wei X, Li X, Peng H, Li Y, Feng Q, Huang Y, Zhang Q, Xu X, Zhao H, Yang X, Wang C, Zhao X (2022) Identification and characterization of P2-like bacteriophages of Yersinia pestis. Virus Res 322:198934. https://doi.org/10. 1016/j.virusres.2022.198934
- 142. Pei Z, Liu Y, Chen Y, Pan T, Sun X, Wang H, Ross RP, Lu W, Chen W (2024) A universe of human gut-derived bacterial prophages: unveiling the hidden viral players in intestinal microecology. Gut Microbes 16(1):2309684. https://doi.org/ 10.1080/19490976.2024.2309684
- 143. Picard B, Garcia JS, Gouriou S, Duriez P, Brahimi N, Bingen E, Elion J, Denamur E (1999) The link between phylogeny and virulence in Escherichia coli extraintestinal infection. Infect Immun 67(2):546–553. https://doi.org/10.1128/IAI.67.2.546-553.1999
- 144. Johnson JR, Stell AL (2020) Extended virulence genotypes of Escherichia coli strains from patients with urosepsis in relation to phylogeny and host compromise. J Infect Dis 181(1):261–272. https://doi.org/10.1086/315217
- 145. Tariq H, Noreen Z, Ahmad A, Khan L, Ali M, Malik M, Javed A, Rasheed F, Fatima A, Kocagoz T, Sezerman U, Bokhari H (2022) Colibactin possessing *E. coli* isolates in association with colorectal cancer and their genetic diversity among Pakistani population. PLoS One 17(11):e0262662. https://doi.org/10.1371/journ al.pone.0262662
- 146. Rosenberg S, Bonten M, Haazen W, Spiessens B, Abbanat D, Go O, Wagenlehner F, Shore N, Hagiwara Y, Ibarra de Palacios P, Geurtsen J, Hermans P, Poolman J (2021) Epidemiology and O-serotypes of extraintestinal pathogenic Escherichia coli disease in patients undergoing transrectal ultrasound prostate biopsy: a prospective multicenter study. J Urol 205(3):826–832. https://doi.org/10.1097/JU.000000000001425
- 147. Bäckhed F, Alsén B, Roche N, Angström J, von Euler A, Breimer ME, Westerlund-Wikström B, Teneberg S, Richter-Dahlfors A (2002) Identification of target tissue glycosphingolipid receptors for uropathogenic, F1C-fimbriated Escherichia coli and its role in mucosal inflammation. J Biol Chem 277(20):18198–18205. https://doi.org/10.1074/jbc.M111640200
- 148. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, Vignat J, Ferlay J, Murphy N, Bray F (2023) Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. Gut 72(2):338–344. https://doi.org/10.1136/gutjnl-2022-327736
- Brennan CA, Garrett WS (2019) Fusobacterium nucleatum symbiont, opportunist and oncobacterium. Nat Rev Microbiol 17(3):156–166. https://doi.org/10.1038/s41579-018-0129-6
- 150. O'Keefe SJ (2016) Diet, microorganisms and their metabolites, and colon cancer. Nat Rev Gastroenterol Hepatol 13(12):691– 706. https://doi.org/10.1038/nrgastro.2016.165
- 151. Wong SH, Yu J (2019) Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. Nat Rev

Gastroenterol Hepatol 16(11):690–704. https://doi.org/10.1038/ s41575-019-0209-8

- 152. Zepeda-Rivera M, Minot SS, Bouzek H, Wu H, Blanco-Míguez A, Manghi P, Jones DS, LaCourse KD, Wu Y, McMahon EF, Park SN, Lim YK, Kempchinsky AG, Willis AD, Cotton SL, Yost SC, Sicinska E, Kook JK, Dewhirst FE, Segata N, Bullman S, Johnston CD (2024) A distinct Fusobacterium nucleatum clade dominates the colorectal cancer niche. Nature 628(8007):424– 432. https://doi.org/10.1038/s41586-024-07182-w
- 153. Buc E, Dubois D, Sauvanet P, Raisch J, Delmas J, Darfeuille-Michaud A, Pezet D, Bonnet R (2013) High prevalence of mucosa-associated E coli producing cyclomodulin and genotoxin in colon cancer. PLoS One 8(2):e56964. https://doi.org/10.1371/ journal.pone.0056964
- 154. Dejea CM, Fathi P, Craig JM, Boleij A, Taddese R, Geis AL, Wu X, DeStefano Shields CE, Hechenbleikner EM, Huso DL, Anders RA, Giardiello FM, Wick EC, Wang H, Wu S, Pardoll DM, Housseau F, Sears CL (2018) Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. Science 359(6375):592–597. https://doi.org/10.1126/ science.aah3648
- 155. Jans M, Kolata M, Blancke G, D'Hondt A, Gräf C, Ciers M, Sze M, Thiran A, Petta I, Andries V, Verbandt S, Shokry E, Sumpton D, Vande Voorde J, Berx G, Tejpar S, van Loo G, Iliev ID, Remaut H, Vereecke L (2024) Colibactin-driven colon cancer requires adhesin-mediated epithelial binding. Nature 635(8038):472–480. https://doi.org/10.1038/ s41586-024-08135-z
- 156. Di Pierro F, Cazzaniga M, Bertuccioli A, Palazzi CM, Cavecchia I, Matera M, Zerbinati N (2024) *E. coli* 5C: a probiotic effective for preventing post-colonoscopy gastrointestinal discomfort? Minerva Gastroenterol (Torino). https://doi.org/10.23736/S2724-5985.24. 03772-0.
- 157. Lee-Six H, Olafsson S, Ellis P, Osborne RJ, Sanders MA, Moore L, Georgakopoulos N, Torrente F, Noorani A, Goddard M, Robinson P, Coorens THH, O'Neill L, Alder C, Wang J, Fitzgerald RC, Zilbauer M, Coleman N, Saeb-Parsy K, Martincorena I, Campbell PJ, Stratton MR (2019) The landscape of somatic mutation in normal colorectal epithelial cells. Nature 574(7779):532–537. https://doi.org/10.1038/s41586-019-1672-7
- 158. Robinson PS, Coorens THH, Palles C, Mitchell E, Abascal F, Olafsson S, Lee BCH, Lawson ARJ, Lee-Six H, Moore L, Sanders MA, Hewinson J, Martin L, Pinna CMA, Galavotti S, Rahbari R, Campbell PJ, Martincorena I, Tomlinson I, Stratton MR (2021) Increased somatic mutation burdens in normal human cells due to defective DNA polymerases. Nat Genet 53(10):1434–1442. https://doi.org/10.1038/s41588-021-00930-y
- 159. Mäklin T, Thorpe HA, Pöntinen AK, Gladstone RA, Shao Y, Pesonen M, McNally A, Johnsen PJ, Samuelsen Ø, Lawley TD, Honkela A, Corander J (2022) Strong pathogen competition in neonatal gut colonisation. Nat Commun 13(1):7417. https://doi. org/10.1038/s41467-022-35178-5

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.