

Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC)

Elzbieta Brzuszkiewicz · Andrea Thürmer · Jörg Schuldes · Andreas Leimbach · Heiko Liesegang · Frauke-Dorothee Meyer · Jürgen Boelter · Heiko Petersen · Gerhard Gottschalk · Rolf Daniel

Received: 14 June 2011 / Accepted: 15 June 2011 / Published online: 29 June 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract The genome sequences of two *Escherichia coli* O104:H4 strains derived from two different patients of the 2011 German *E. coli* outbreak were determined. The two analyzed strains were designated *E. coli* GOS1 and GOS2 (German outbreak strain). Both isolates comprise one chromosome of approximately 5.31 Mbp and two putative plasmids. Comparisons of the 5,217 (GOS1) and 5,224 (GOS2) predicted protein-encoding genes with various *E. coli* strains, and a multilocus sequence typing analysis revealed that the isolates were most similar to the entero-

aggregative *E. coli* (EAEC) strain 55989. In addition, one of the putative plasmids of the outbreak strain is similar to pAA-type plasmids of EAEC strains, which contain aggregative adhesion fimbrial operons. The second putative plasmid harbors genes for extended-spectrum β -lactamases. This type of plasmid is widely distributed in pathogenic *E. coli* strains. A significant difference of the *E. coli* GOS1 and GOS2 genomes to those of EAEC strains is the presence of a prophage encoding the Shiga toxin, which is characteristic for enterohemorrhagic *E. coli* (EHEC) strains. The unique combination of genomic features of the German outbreak strain, containing characteristics from pathotypes EAEC and EHEC, suggested that it represents a new pathotype Enter-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC).

Communicated by Erko Stackebrandt.

Elzbieta Brzuszkiewicz, Andrea Thürmer, Jörg Schuldes, Andreas Leimbach, Heiko Liesegang have contributed equally to this article.

Electronic supplementary material The online version of this article (doi:10.1007/s00203-011-0725-6) contains supplementary material, which is available to authorized users.

E. Brzuszkiewicz · A. Thürmer · J. Schuldes · A. Leimbach · H. Liesegang · F.-D. Meyer · G. Gottschalk · R. Daniel
Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen, Grisebachstr. 8, 37077 Göttingen, Germany

R. Daniel (✉)
Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen, Grisebachstr. 8, 37077 Göttingen, Germany
e-mail: rdaniel@gwdg.de

J. Boelter
Roche Diagnostics Deutschland GmbH, Sandhofer Str. 116, 68305 Mannheim, Germany

H. Petersen
Medizinisches Versorgungszentrum für Labormedizin und Humangenetik, Abt. Molekulare Erregerdiagnostik, Bergstraße 14, 20095 Hamburg, Germany

Keywords EHEC outbreak · EAHEC · Genome sequencing · Pathotype · Genome evolution

Introduction

Escherichia coli is a bacterium that is commonly found in the intestine of humans and other mammals. Most *E. coli* strains are harmless commensals. However, some strains such as enterohemorrhagic *E. coli* (EHEC) strains can cause severe food-borne diseases. These pathogens are transmitted to humans primarily through consumption of contaminated drinking water and foods such as raw or undercooked ground meat products, raw milk, and even vegetables (Kaper et al. 2004). In addition, person-to-person transmission is possible. The significance of EHEC as a public health problem was first recognized in 1982, following an outbreak in the United States of America associated with undercooked hamburgers (Kaper et al. 2004).

Infections caused by EHEC may lead to severe diarrhea and hemorrhagic colitis with complications such as microangiopathic hemolytic anemia, thrombocytopenia, and fatal acute renal failure, which are summarized as hemolytic uremic syndrome (HUS) (Karmali et al. 1983, 1985; Law et al. 1992). Ruminants, predominantly cows, are the natural reservoir of EHEC strains (Kaper et al. 2004).

EHEC is known to produce characteristic toxins, which are similar to toxins produced by *Shigella dysenteriae* and are known as verocytotoxins or Shiga toxins (STX) (Kaper et al. 2004; Karch et al. 2005; Tarr et al. 2005). Absorption of these toxins by the bloodstream leads to damage to the kidneys and to HUS. The most significant serogroups among EHEC strains are O26, O103, O111, and O157. *E. coli* O157:H7 is the most important EHEC serotype with respect to public health in North America, the United Kingdom, and Japan (Kaper et al. 2004). Typical EHEC strains produce STX but also encode a LEE (locus of enterocyte effacement) pathogenicity island, which is important for adherence in the colon (Jores et al. 2004). *E. coli* strains that encode a Shiga toxin, but do not contain the LEE pathogenicity island, are designated as STEC (Shiga toxin-producing *E. coli*) strains. Approximately 200 different serogroups of STEC strains are known and more than 100 harbor a virulence potential. Up to 50% of infections with STEC strains are linked to non-O157 serogroups (Kaper et al. 2004).

The EHEC outbreak started in Germany in May 2011 with 3,368 cases including 36 deaths (as of June 14th, 2011, European Centre for Disease Prevention and Control; <http://www.ecdc.europa.eu/en/Pages/home.aspx>). This is the second largest food-borne *E. coli* outbreak in history. The enterohemorrhagic *E. coli* strain O104:H4 was identified as the causative agent of the EHEC infection outbreak. This strain was found in humans before but never as causative agent of an EHEC outbreak (Robert Koch Institute, Berlin, Germany; <http://www.rki.de>). Only one case of infection with strain O104:H4 has been documented in the literature prior to the 2011 outbreak. In this case, the strain was isolated from a 29-year-old Korean woman, who suffered from HUS (Bae et al. 2006).

In this study, we report on the genome sequences of two O104:H4 isolates, which were derived from two patients of the 2011 EHEC outbreak in Germany. The determination of the genomic features of the isolates provides insights into the genomic potential, pathogenicity, and evolution of the O104:H4 strain. Comparison of our *E. coli* O104:H4 genome sequences with that of other pathogenic *E. coli* suggests that strain O104:H4 represents a new *E. coli* pathotype, which we named Enter-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC).

Results

General features of *E. coli* GOS1 and GOS2 genome sequences

The genome sequences of two *E. coli* O104:H4 strains derived from two different patients, a 75-year-old woman and 48-year-old man, from the 2011 German EHEC outbreak were determined using 454 pyrosequencing technology (Margulies et al. 2005). The two analyzed strains were designated *E. coli* GOS1 and GOS2 (German outbreak strain). PCR-based detection of four specific marker genes (*stx2*, *terD*, *rfb0104*, and *fliC* H4) confirmed that both were O104:H4 strains (Fig. S1). The general genomic features of the genomes of *E. coli* GOS1 and GOS2 are presented together with features of already sequenced and selected *E. coli* reference genomes in Table S1. The assembly of the draft genomes of *E. coli* GOS1 and GOS2 yielded 171 and 204 large contigs, respectively (Table 1). The estimated genome size of both isolates is 5.31 Mbp. In addition, a total of 5,217 (GOS1) and 5,224 (GOS2) protein-encoding genes were predicted.

Genome comparison of GOS1 and GOS2 with selected *E. coli* genomes

Sequence alignment of *E. coli* GOS1 and GOS2 genome sequences using the MUMmer software tool (Kurtz et al. 2003) revealed 99.9% identity of both sequences. We could not find a single-nucleotide polymorphism when we compared the draft genomes of *E. coli* GOS1 and GOS2 by employing the GS Mapper Reference software (Roche 454, Branford, USA). Thus, as these isolates derived from patients showing different gender and age, it appears that the genome of *E. coli* O104:H4 is stable during its infection in different hosts. This assumption was supported by comparison of the *E. coli* GOS1 and GOS2 genomes with the three

Table 1 Assembly data of the *Escherichia coli* GOS1 and GOS2 genome sequences

	<i>E. coli</i> GOS1	<i>E. coli</i> GOS2
Genome size (Mbp)	5.31	5.31
GC content (%)	50.6	50.6
Coverage	24-fold	21-fold
Number of large contigs (>500 bp)	171	204
Average contig size (kbp)	30.99	25.96
N50 contig size (kbp)	109.54	88
Largest contig size (kbp)	337.55	247.7
Q40 value (%)	99.41	99.42

The genomes of *E. coli* GOS1 and *E. coli* GOS2 were assembled de novo from 349,788 and 311,478 shotgun reads, respectively, by employing the Roche Newbler assembly software

other available draft genome sequences of *E. coli* O104:H4 isolates derived from the German outbreak. The sequence identities of *E. coli* GOS1 to the genome sequences of *E. coli* O104:H4 isolates TY-2482 (Beijing Genomics Institute, China), LB226692 (Life Technologies, Germany; University of Münster, Germany), and H112180280 (Health Protection Agency, Cambridge, United Kingdom) were 99.8, 99.5, and 99.9%, respectively. Taking into account the overall high similarity of all five genome sequences and the different sequencing approaches used, we assume that the recorded differences of the genome sequences are mainly due to sequencing errors and not to changes within the genome of the different isolates. In addition, as all analyzed chromosomal *E. coli* sequences share synteny over the whole chromosome length, we could align chromosomal contigs of all available sequences of the German outbreak to the chromosome of EAEC 55989 and obtain the contig order for the genomes of *E. coli* GOS1 and GOS2 (Fig. S2).

Comparison of the complete gene content of *E. coli* GOS1 and GOS2 with selected *E. coli* genomes showed that the chromosome of both isolates is most similar to that of the entero-aggregative *E. coli* (EAEC) strain 55989 (Fig. S2). *E. coli* strain 55989 was originally isolated from the diarrheagenic stools of an HIV-positive adult suffering from persistent watery diarrhea (Mossoro et al. 2002). Genome wide BiBag comparisons revealed a set of 4,606 (GOS1) and 4,607 (GOS2) orthologous genes that are shared by at least one chromosome of the selected reference *E. coli* strains (Table S1). Among the remaining 611 (GOS1) and 617 (GOS2) genes 122 and 211, respectively, genes were orthologous to genes located on plasmids.

Comparisons of the *E. coli* GOS1 and GOS2 chromosomes with those of EAEC 55989 and EHEC O157:H7 Sakai using the Artemis comparison tool (Carver et al. 2005) revealed that the chromosomal backbone of the German outbreak strain is different from that of typical *E. coli* EHEC or EAEC strain. Most important differences are the lack of the LEE pathogenicity island and the presence of a Stx-phage in the genomes of *E. coli* GOS1 and GOS2 (Fig. 1).

A multilocus sequence typing (MLST) analysis of seven housekeeping genes *adhA*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA* of the two *E. coli* isolates GOS1 and GOS2 was done according to Wirth et al. (2006). *E. coli* GOS1 and GOS2 share the same sequence for all seven genes. By interrogation of the Achtman's MLST scheme database (Wirth et al. 2006), the outbreak strain could be assigned to the sequence type 678 (ST678) complex (*adhA* 6, *fumC* 6, *gyrB* 5, *icd* 136, *mdh* 9, *purA* 7, *recA* 7). This complex belongs to the ECOR ancestral group B1, which is a very heterogeneous group with respect to included pathotypes (Tenailon et al. 2010). The group B1 includes non-O157, EHEC, ETEC, and commensal *E. coli* strains. In addition, EAEC strain 55989 is grouped in B1. A Maximum Likelihood tree of completely sequenced *E. coli* genomes confirmed the close relationship of the German outbreak strain to EAEC 55989 (Fig. 2).

Plasmids

We identified two genes encoding plasmid replication proteins in each dataset (GOS1, RGOS01291, and RGOS00376; GOS2, RGOT04762, and RGOT01786). Therefore, it is assumed that the outbreak strain harbors at least two extrachromosomal replicons. In order to identify the potential plasmid-encoded proteins, our sequence data were mapped on several reference plasmids (Table S2). A total of 169 potential plasmid-located genes were thereby identified. Further data analysis revealed the presence of a putative plasmid in *E. coli* GOS1 and GOS2, which is almost identical to the pEC_Bactec plasmid (Fig. 3). Contigs from our data spanned over 90% of the total pEC_Bactec plasmid length (84,221 bp out of 92,970 bp). Small contigs coding only for transposases or insertion elements were not included in the analysis. The reconstructed plasmids of *E. coli* GOS1 and GOS2 consist of only three contigs (Fig. 3). The resistance genes TEM-1 and CTX-M-15 are located on this plasmid. Extended-spectrum beta-lactamases (ESBLs) such as TEM-1 and

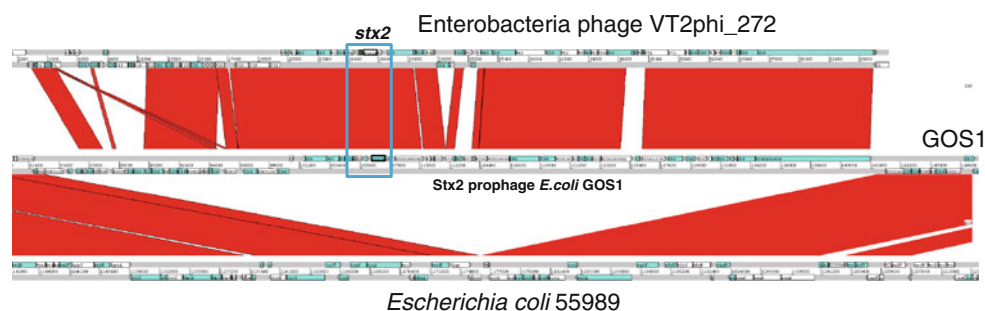


Fig. 1 Comparisons of enterobacteria phage VT2phi_272 with the corresponding genomic region of *E. coli* GOS1 and *E. coli* strain 55989. Analysis was performed by employing the ACT software tool

(Sanger Institute, <http://www.sanger.ac.uk>). The relationship between each pair of sequences are depicted. Similar coding sequences are indicated by red-colored lines. The *stx* genes are boxed

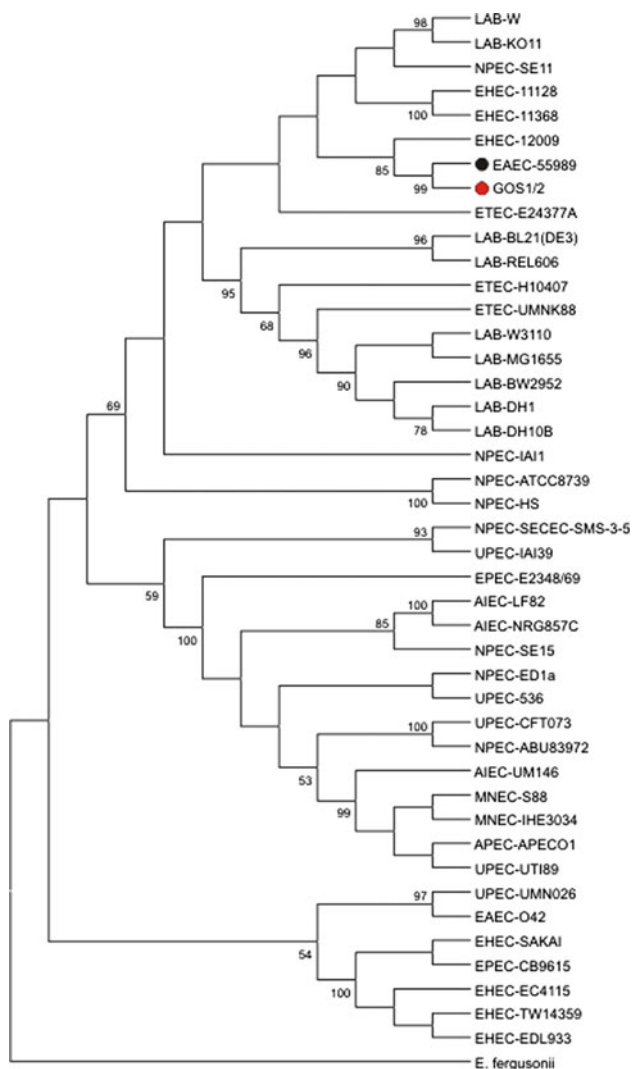


Fig. 2 Phylogenetic analysis of completely sequenced *E. coli* strains based on multilocus sequence typing. The phylogenetic analysis was conducted with MEGA 5.05 (Tamura et al. 2011). The resulting Maximum Likelihood tree illustrates the close relationship of the German outbreak strain (red dot) to EAEC 55989 (black dot). The pathotype of each *E. coli* strain is indicated in front of the strain name (see below for abbreviations). Bootstrap values were calculated from 100 resamplings. Bootstrap values below 50 were not shown. The following *E. coli* strains were used in the analysis: entero-aggregative *E. coli* (EAEC) 042 (FN554766), uropathogenic *E. coli* (UPEC) 536 (CP000247), EAEC 55989 (CU928145), commensal non-pathogenic *E. coli* (NPEC) ABU83972 (CP001671), avian pathogenic *E. coli* (APEC) O1 (CP000468), lab B strain BL21(DE3) (AM946981), lab B strain REL606 (CP000819), industrial production strain KO11 (CP002516), enteropathogenic *E. coli* (EPEC) CB9615 (CP001846), UPEC CFT073 (AE014075), EPEC E2348/69 (FM180568), enterotoxigenic *E. coli* (ETEC) E24377A (CP000800), commensal ED1a (CU928162), ETEC H10407 (FN649414), commensal HS (CP000802), commensal IAI1 (CU928160), UPEC IAI39 (CU928164), meningitis-associated *E. coli* (MNEC) IHE3034 (CP001969), commensal strain K-12 substrain ATCC 8739/Crooks (CP000946), lab strain K-12 substrain BW2952 (CP001396), lab strain K-12 substrain DH1 (CP001637), lab strain K-12 substrain DH10B (CP000948), lab strain K-12 substrain MG1655 (U00096), lab strain K-12 substrain W3110 (AP009048), adherent-invasive *E. coli* (AIEC) LF82 (CU651637), AIEC NRG 857C (CP001855), EHEC O103:H2 12009 (AP010958), EHEC O111:H- 11128 (AP010960), EHEC O157:H7 EC4115 (CP001164), EHEC O157:H7 EDL933 (AE005174), EHEC O157:H7 Sakai (BA000007), EHEC O157:H7 TW14359 (CP001368), EHEC O26:H11 11368 (AP010953), MNEC S88 (CU928161), commensal SE11 (AP009240), commensal SE15 (AP009378), environmental strain SECEC SMS-3-5 (CP000970), AIEC UM146 (CP002167), UPEC UMN026 (CU928163), porcine ETEC UMNK88 (CP002729), UPEC UTI89 (CP000243), and lab strain W (CP002185). *Escherichia fergusonii* ATCC 35469 was used as outgroup (CU928158)

CTX-M-15 are the most prevalent secondary beta-lactamases among clinical isolates of *Enterobacteriaceae* worldwide (Livermore 1995). ESBLs are a group of β -lactamases, which share the ability to hydrolyze third-generation cephalosporins and aztreonam (Paterson and Bonomo 2005).

A significant number of genes mapped to the plasmids p042 and 55989p, which are typical for EAEC strains (Fig. 4a; Table S2) (Touchon et al. 2009; Chaudhuri et al. 2010). The plasmids of GOS1 and GOS2 share a set of 46 genes with EAEC plasmid 55989p (Table S2) including the aggregative adhesion operon *aat* and the regulator *aggR*. Additionally, the toxin-antitoxin system *ccd* and the replication protein RepFIB were found. However, genes encoding for aggregative adherent fimbriae (AAF), a primary virulence factor of EAEC strains (Kaper et al. 2004), are different from the 55989p variant. Mapping *E. coli* GOS1 and GOS2 data on the second reference plasmid p042 showed also a significant number of homologous

proteins (Fig. 4b; Table S2). Many potential virulence factors are shared with p042 plasmid such as the AAF (*agg3*) operon and the serine protease *pet*. *Pet* is secreted by many EAEC strains and exhibits enterotoxic activity (Navarro-García et al. 1998).

Phage analysis

We could identify 336 prophage-encoding genes for GOS1 and 334 for GOS2 (Tables S3, S4). The key virulence factor of EHEC, STX, is encoded on a lambda-like bacteriophage, the Stx-phage. Acquisition of this phage was a key step in the evolution of EHEC from EPEC (Reid et al. 2000). A Stx-phage is present in the outbreak strain (Fig. 1). This phage shows high identity to the *stx2*-containing enterobacteria phage VT2phi_272 from *E. coli* O157:H7 strain 71074 (HQ424691). The GOS1 Stx-phage consists of 66 encoding genes and is identical to the GOS2 Stx-phage (Tables S3, S4). In addition to the Stx-phage, 70 prophage-encoding genes (Tables S3, S4) that are not present in *E. coli* 55989 could be identified in the genome of *E. coli* GOS1. These genes have high similarity to STX-producing prophages and also to the other

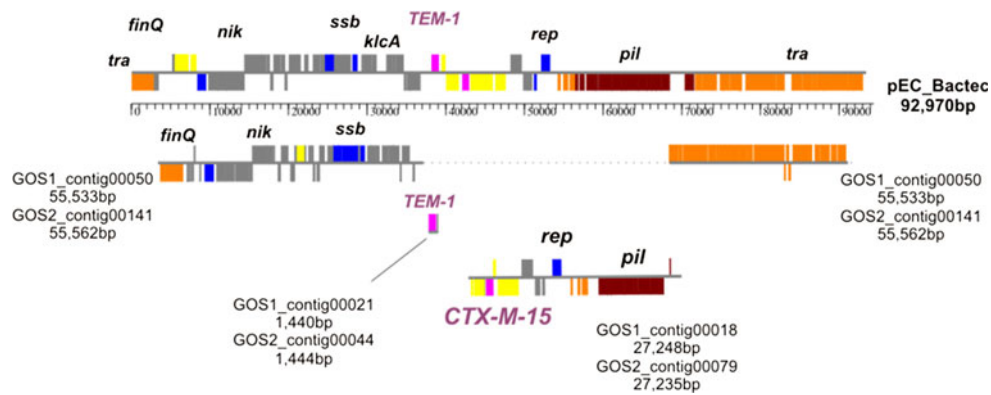
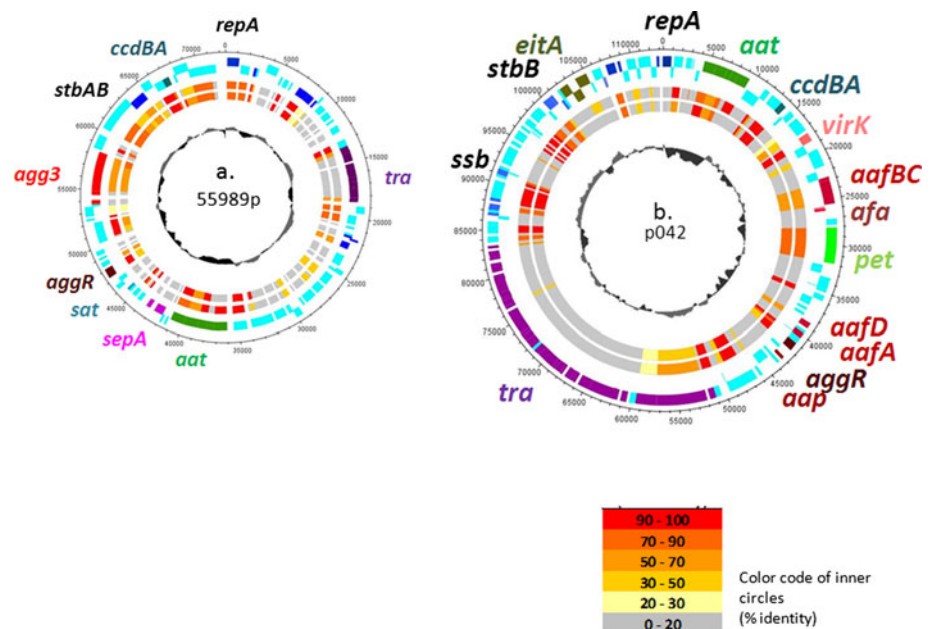


Fig. 3 Linear comparison of *E. coli* pEC_Bactec plasmid with corresponding GOS1 and GOS2 contigs. The top map represents the pEC_Bactec plasmid (GU371927.1), the resistance genes are highlighted in pink, IS-elements/transposases in yellow, plasmid replication/

stabilization genes in blue, the tra operon in orange, pil operon in brown, and remaining genes in gray. The scale is in base pairs. All maps were done with GenVision software (<http://www.dnastar.com/t-products-genvision.aspx>)

Fig. 4 Comparison of GOS1 and GOS2 genes with two different pAA-type plasmids. The two outermost rings represent maps of **a** 55989p and **b** p042 from strain *E. coli* 55989 and *E. coli* 042, respectively. Virulence factors and selected important genes are highlighted and colored. The second and the third rings represent presence (colored) or absence (gray) of GOS1 and GOS2 orthologs. The inner rings represent the GC contents of the plasmids



above-mentioned phage in the outbreak strain, but lack *stx2AB* (Fig. S3).

Resistance

EHEC O157:H7 strains resist the highly toxic tellurium oxyanion, tellurite (Tel) (Zadik et al. 1993; Taylor et al. 2002; Bielaszewska et al. 2005; Orth et al. 2007). Tellurite resistance (TelR) of EHEC O157:H7 is encoded by the chromosomal *terZABCDEF* gene cluster (Taylor et al. 2002; Bielaszewska et al. 2005), which is highly homologous to the *ter* cluster on plasmid R478 of *Serratia marcescens* (Whelan et al. 1995; Taylor et al. 2002). TelR is a common, but not obligatory, feature of EHEC O157:H7 strains, as tellurite-susceptible *E. coli* O157:H7 strains have been isolated in North America (Taylor et al. 2002)

and Europe (Bielaszewska et al. 2005). We identified all proteins of the *terZABCDEF* operon in the outbreak strain (ORFs RGOS02836 to RGOS02842).

In addition, the German outbreak strain could bear a mercuric resistance plasmid, as in many bacteria resistance to mercury is associated with a plasmid (Smith 1967; Novick and Roth 1968; Summers and Silver 1972; Kondo et al. 1974). Correspondingly, the predicted proteins involved in mercury resistance were located all on one contig (GOS1_contig00023). These genes encode the putative mercuric ion transport proteins MerT, MerP, and MerC (RGOS00392, RGOS00393, and RGOS00394, respectively), the corresponding transcriptional regulators MerR (RGOS00391) and MerD (RGOS00396), and mercuric ion reductase MerA (RGOS00395). In addition to genes involved in mercuric resistance and tellurium resistance, we

have predicted and annotated many genes involved in antibiotic resistance such as putative gene-encoding chloramphenicol (RGO00056), tetracycline (RGOS00387, RGOS00388), or streptomycin resistance (RGOS00359).

Discussion

Chromosomes and plasmids

The chromosomes of the *E. coli* isolates GOS1 and GOS2 are most similar to the chromosome of EAEC strain 55989 isolated in Africa over a decade ago. EAEC strains are the most recently emerged *E. coli* intestinal pathotype and the second most common agent of traveler's diarrhea (Huang et al. 2006). EAEC pathogenesis is thought to involve three primary steps. First, the bacteria adhere to the intestinal mucosa using aggregative adherent fimbriae (AAF). Second, these fimbriae cause autoaggregative adhesion, by which the bacteria adhere to each other in a 'stacked-brick' configuration producing a mucous-mediated biofilm on the enterocyte surface. Third, the bacteria release toxins that affect the inflammatory response, intestinal secretion, and mucosal cytotoxicity. Aspects of each of these steps involve plasmid-encoded traits but also chromosomal-encoded virulence factors (Kaper et al. 2004).

In addition to the chromosomal similarity, *E. coli* GOS1 and GOS2 share with EAEC strain 55989 part of the EAEC plasmid 55989p. This plasmid carries the AAF operon *aat* and the regulator *aggR*. Nevertheless, a different aggregative adhesion fimbrial complement was present in our strains. The AAF operon is usually localized on an approximately 100-kb plasmid, termed the "pAA plasmid" (Nataro et al. 1987). Four genetically distinct allelic variants of AAF have been identified previously, AAF/I from EAEC strain 17-2 (Nataro et al. 1992), AAF/II from strain O42 (Nataro et al. 1995), AAF/III from strain 55989 (Bernier et al. 2002), and Hda from strain C1010-00 (Boisen et al. 2008). All the identified AAF allelic types appear to be plasmid encoded, and most of the analyzed strains possess only a single AAF allelic type (Harrington et al. 2006). The outbreak strain is no exception and seems to contain the relatively rare AAF/I locus of EAEC. Additionally, the *ipd* gene encoding an extracellular serine protease and the gene encoding serine protease *Pet* were found in the German outbreak strain. Usually, these virulence factors are localized next to the AAF operon on the pAA plasmid. Another virulence feature, the *aatPABCD* operon (dispersin secretion locus), is a plasmid-borne characteristic of EAEC strains. This operon is also present in the genome of the German outbreak strain.

Two RepA proteins were found in the German outbreak strain. This suggests that this strain harbors at least two

plasmids. In addition to the pAA-like plasmid, we identified contigs showing high similarity to the previously described plasmids pEC_Bactec, pCVM29188_101, and pEK204 (Fricke et al. 2009; Woodford et al. 2009; Smet et al. 2010). These plasmids encode the extended-spectrum β -lactamases blaCTX-M and blaTEM-1.

Evolution: horizontal gene transfer (HGT)

Escherichia coli virulence factors such as enterotoxins, invasion factors, adhesion factors, or Shiga toxins can be encoded by several mobile genetic elements, including transposons (Tn), plasmids, bacteriophages, or pathogenicity islands (e.g., LEE island). Bacterial plasmids play a key role in a variety of traits like drug resistance, virulence, and the metabolism of rare substrates under specific conditions (Actis et al. 1999). Plasmids are able to mobilize these traits between different strains and thus play an important role in horizontal gene transfer. The analyses indicate that a number of horizontal gene transfer events took place to create the genome of the German outbreak strain. This strain probably originated from an EAEC pathotype, which is suggested by the missing LEE island and the high similarity of the genome to the genome of EAEC strain 55989. In contrast to the EAEC strains, the German outbreak strain has acquired the Stx-phage, which is typical for EHEC strains (Fig. 1).

Another feature of the new outbreak strain is the acquisition of plasmid-encoded drug resistances. The strain has acquired a plasmid sharing high similarity with the plasmids pEC_Bactec, pCVM29188_10, and pEK204. The origin of this plasmid remains unclear, since the extended-spectrum β -lactamases (ESBLs) CTX-M and TEM-1 resistances seem to be located on a Tn3-type transposon that has been widely spread among enteric bacteria.

To conclude, *E. coli* O104:H4 possesses a Stx-phage typical for EHEC strains but is missing the characteristic LEE island. In addition to the high overall genome sequence similarity to EAEC strains, it harbors an AAF operon, which is a distinguishing feature for EAEC strains. The German outbreak strain harbors a unique combination of EHEC and EAEC genomic features (Fig. 5). These data suggest a new *E. coli* pathotype EAHEC that has EHEC and EAEC ancestors.

Materials and methods

Sample preparation and DNA extraction

The two *E. coli* O104:H4 isolates GOS1 and GOS2 were derived from stool samples of two different patients of the 2011 German outbreak. *E. coli* GOS1 and GOS2 were

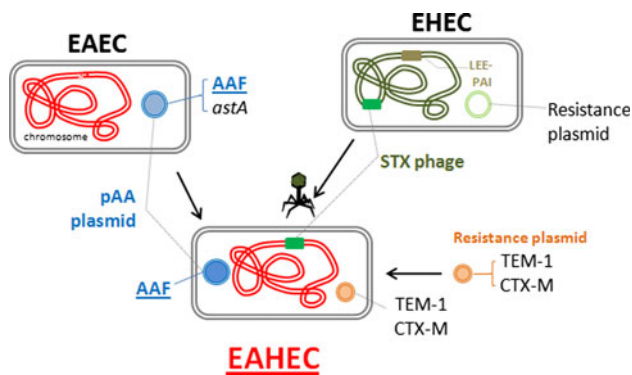


Fig. 5 Proposed scheme of the origin of the new *E. coli* pathotype—EAHEC

recovered from a 75-year-old woman and a 48-year-old man, respectively. To isolate these strains, stool samples were plated on Brilliance™ ESBL Agar plates (Oxoid, Wesel, Germany) and incubated for 24 h at 37°C. Initially, the *E. coli* O104:H4 strains were identified by the ability to produce STX2. For this purpose, the LightMix® kits *E. coli* EHEC Stx1 and Stx2 were applied as recommended by the manufacturer (TIB MOLBIOL, Berlin, Germany). A colony of each strain from the thereby recovered positive strains, *E. coli* GOS1 and GOS1, was grown in 4 ml EHEC-direct-media (Heipha Diagnostics, Eppenheim, Germany) overnight at 37°C. To isolate genomic DNA, the cultures were pelleted (5 min, 2,000g), resuspended in 1 ml S.T.A.R. Buffer (Roche, Molecular Diagnostics, Rotkreuz, Switzerland), and incubated for 5 min at 95°C. Subsequently, the suspension was subjected to centrifugation for 1 min at 1,100g. The cell-free supernatant (500 µl) was used for the preparation of the genomic DNA by employing the High Pure 16 System Viral Nucleic Acid kit as recommended by the manufacturer (Roche Applied Science, Mannheim, Germany). The resulting DNA solution (260 ng/µl) was used for further analysis.

To confirm that *E. coli* isolates GOS1 and GOS2 were O104:H4 serotype, a PCR-based detection of four specific marker genes (*stx2*, *terD*, *rfbO104*, and *fliC* H4) was performed according to the PCR typing scheme by the group of Prof. Karch at the National Consulting Laboratory on HUS at the University of Münster (see http://www.ehec.org/pdf/Laborinfo_01062011.pdf, 2011) with slight adaptations. Briefly, the PCR reaction mixture (25 µl) contained 2.5 µl tenfold reaction buffer (Bioline, Luckenwalde, Germany), 0.2 mM of each of the four deoxynucleoside triphosphates, 1.5 mM MgCl₂, 0.2 µM of each of the primers, 1 U of BIO-X-ACT™ DNA Polymerase (Bioline), and 100 ng of isolated genomic DNA as template. The *stx2*, *terD*, *rfbO104*, and *fliC* H4 were amplified with the following set of primers: *stx2*, 5'-ATCCTATTCCCGGGAGTTTACG-3' and 5'-GCGTCATCGTATACAC

AGGAGC-3'; *terD*, 5'-AGTAAAGCAGCTCCGTCAA T-3' and 5'-CCGAACAGCATGGCAGTCT-3'; *rfbO104*, 5'-TGAAGTATTTTTAGGATGG-3' and 5'-AGAACC TCACTCAAATTATG-3'; and *fliC* H4, 5'-GGCGAA ACTGACGGCTGCTG-3' and 5'-GCACCAACAGTT ACCGCCGC-3'. The following thermal cycling scheme was used: initial denaturation at 94°C for 5 min, 30 cycles of denaturation at 94°C for 45 s, annealing at 55°C (*stx2*, *terD*, *rfbO104*) or 63°C (*fliC* H4) for 45 s, and extension at 72°C for 60 s (*stx2*, *terD*, *rfbO104*) or 30 s (*fliC* H4) followed by a final extension period at 72°C for 5 min. Subsequently, PCR products were separated by agarose gel electrophoresis (1.5% gels) and analyzed. The analysis revealed that all four marker genes were present in *E. coli* isolates GOS1 and GOS2 in the expected sizes (Fig. S1).

Sequencing and assembly

The isolated DNA from both strains was used to create 454-shotgun libraries following the GS Rapid library protocol (Roche 454, Branford, USA). The resulting two 454 DNA libraries were sequenced with the Genome Sequencer FLX (Roche 454) using Titanium chemistry. For sequencing of each sample, 1.5 medium lanes of a Titanium picotiter plate were used. A total of 349,788 and 311,478 shotgun reads were achieved for *E. coli* GOS1 and *E. coli* GOS2, respectively. Reads were assembled de novo using the Roche Newbler assembly software 2.3 (Roche 454) (Table 1).

Gene prediction and annotation

Gene prediction was performed with Glimmer3 (Delcher et al. 2007). Automatic gene annotation was done by transferring annotations from orthologous genes of reference strains (Table S1) available at the EMBL database. Orthologous genes were identified as described previously by bidirectional BLAST comparisons (Schmeisser et al. 2009). Proteins without orthologs in the reference strains were annotated according to their best BLAST hits to the SwissProt subset of the UniProt Database (Jain et al. 2009, <http://www.uniprot.org>). Sequence data of isolates GOS1 and GOS2 are publicly available and can be downloaded from the Göttingen Genomics Laboratory website (<ftp://134.76.70.117>; UserID: EAHEC_GOS; Password: EAHEC_GOS).

Genome analysis

In order to analyze the presence of prophage regions, the Prophage Finder software has been employed (<http://131.210.201.64/~phage/ProphageFinder.php>). This web application provides a quick prediction of prophage loci in

prokaryotic genome sequences based on BLASTX comparisons to predicted prophage sequences. The contig order of the *E. coli* GOS1 and GOS2 draft genomes was obtained by comparison to the reference genome of *E. coli* strain 55989 using the Mauve Multiple Genome Alignment software (Darling et al. 2010).

Whole genome sequence alignments of the different *E. coli* O104:H4 isolates (GOS1, GOS2, TY-2482, LB226692, H112180280) were done with the MUMmer software tool (Kurtz et al. 2003). Single-nucleotide polymorphism (SNP) analyses were performed using the GS Reference Mapper Software tool (Roche 454). SNPs were filtered using the following criteria: 100% variation frequency, a minimum of tenfold depth within the variation, the variation is located outside a homopolymer region, and each nucleotide exchange is located at least 100 bp offwards a contig end. For whole genome comparison, the BiBag software tool (Bidirectional BLAST for the identification of bacterial pan and core genomes, Göttingen Genomics Laboratory, Germany) was applied. Visualization of genomic, plasmid, and phage region comparisons was done with the programs Artemis (Rutherford et al. 2000), ACT (Carver et al. 2005), and DNAPlotter (Carver et al. 2009) from the Sanger Institute (<http://www.sanger.ac.uk/>).

Phylogenetic analysis based on MLST

The phylogenetic tree was calculated according to the Achtman MLST scheme (Wirth et al. 2006), which includes sequences of seven housekeeping genes *adhA*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*. The alleles for these genes were extracted from *E. coli* GOS1 and GOS2, and 42 completely sequenced *E. coli* strains. Sequences of the seven housekeeping genes were concatenated, and an alignment was calculated with ClustalW included in MEGA 5.05 (Tamura et al. 2011). The tree was calculated with the Maximum Likelihood method based on the Tamura-Nei model (Tamura and Nei 1993). The bootstrap consensus tree was inferred from 100 replicates. Tree calculation and drawing were done with the software MEGA 5.05 (Tamura et al. 2011). The alleles of the seven housekeeping genes from *Escherichia fergusonii* ATCC 35469 were used as outgroup.

Acknowledgments We thank Sascha Dietrich for bioinformatic support.

Conflict of interest The authors declare no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Actis LA, Tolmasky ME, Crosa JH (1999) Bacterial plasmids: replication of extrachromosomal genetic elements encoding resistance to antimicrobial compounds. *Front Biosci* 4:D43–D62
- Bae WK, Lee YK, Cho MS, Ma SK, Kim SW, Kim NH, Choi KC (2006) A case of hemolytic uremic syndrome caused by *Escherichia coli* O104:H4. *Yonsei Med J* 47:473–479
- Bernier C, Gounon P, Le Bouguenec C (2002) Identification of an aggregative adhesion fimbria (AAF) type III-encoding operon in enteroaggregative *Escherichia coli* as a sensitive probe for detecting the AAF encoding operon family. *Infect Immun* 70:4302–4311
- Bielaszewska M, Tarr PI, Karch H, Zhang W, Mathys W (2005) Phenotypic and molecular analysis of tellurite resistance among enterohemorrhagic *Escherichia coli* O157:H7 and sorbitol-fermenting O157:NM clinical isolates. *J Clin Microbiol* 43:452–454
- Boisen N, Struve C, Scheutz F, Krogfelt KA, Nataro JP (2008) New adhesin of enteroaggregative *Escherichia coli* related to the Afa/Dr/AAF family. *Infect Immun* 76:3281–3292
- Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J (2005) ACT: the artemis comparison tool. *Bioinformatics* 21:3422–3423
- Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J (2009) DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25:119–120
- Chaudhuri RR, Sebahia M, Hobman JL, Webber MA, Leyton DL, Goldberg MD, Cunningham AF, Scott-Tucker A, Ferguson PR, Thomas CM, Frankel G, Tang CM, Dudley EG, Roberts IS, Rasko DA, Pallen MJ, Parkhill J, Nataro JP, Thomson NR, Henderson IR (2010) Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain 042. *PLoS ONE* 5:e8801
- Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss, and rearrangement. *PLoS ONE* 5:e11147
- Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679
- Fricke WF, McDermott PF, Mammel MK, Zhao S, Johnson TJ, Rasko DA, Fedorka-Cray PJ, Pedrosa A, Whichard JM, Leclerc JE, White DG, Cebula TA, Ravel J (2009) Antimicrobial resistance-conferring plasmids with similarity to virulence plasmids from avian pathogenic *Escherichia coli* strains in *Salmonella enterica* serovar Kentucky isolates from poultry. *Appl Environ Microbiol* 75:5963–5971
- Harrington SM, Dudley EG, Nataro JP (2006) Pathogenesis of enteroaggregative *Escherichia coli* infection. *FEMS Microbiol Lett* 254:12–18
- Huang DB, Mohanty A, DuPont HL, Okhuysen PC, Chiang T (2006) A review of an emerging enteric pathogen: enteroaggregative *Escherichia coli*. *J Med Microbiol* 55:1303–1311
- Jain E, Bairoch A, Duvaud S, Phan I, Redaschi N, Suzek BE, Martin MJ, McGarvey P, Gasteiger E (2009) Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics* 10:136
- Jores J, Rumer L, Wieler LH (2004) Impact of the locus of enterocyte effacement pathogenicity island on the evolution of pathogenic *Escherichia coli*. *Int J Med Microbiol* 294:103–113
- Kaper JB, Nataro JP, Mobley HL (2004) Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2:123–140
- Karch H, Tarr PI, Bielaszewska M (2005) Enterohaemorrhagic *Escherichia coli* in human medicine. *Int J Med Microbiol* 295:405–418

- Karmali MA, Steele BT, Petric M, Lim C (1983) Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. *Lancet* 1:619–620
- Karmali MA, Petric M, Lim C, Fleming PC, Arbus GS, Lior H (1985) The association between idiopathic hemolytic-uremic syndrome and infection by verotoxin-producing *Escherichia coli*. *J Infect Dis* 151:775–782
- Kondo I, Ishikawa T, Nakahara H (1974) Mercury and cadmium resistances mediated by the penicillinase plasmid in *Staphylococcus aureus*. *J Bacteriol* 117:1–7
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL (2003) Versatile and open software for comparing large genomes. *Genome Biol* 5:R12
- Law D, Ganguli LA, Donohue-Rolfe A, Acheson DW (1992) Detection by ELISA of low numbers of Shiga-like, toxin-producing *Escherichia coli* in mixed cultures after growth in the presence of mitomycin C. *J Med Microbiol* 36:198–202
- Livermore DM (1995) Beta-lactamases in laboratory and clinical resistance. *Clin Microbiol Rev* 8:557–584
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380
- Mossoro C, Glaziou P, Yassibanda S, Lan NT, Bekondi C, Minssart P, Bernier C, Le Bouguéne C, Germani YH (2002) Chronic diarrhea, hemorrhagic colitis, and hemolytic-uremic syndrome associated with HEp-2 adherent *Escherichia coli* in adults infected with human immunodeficiency virus in Bangui, Central African Republic. *J Clin Microbiol* 40:3086–3088
- Nataro JP, Kaper JB, Robins-Browne R, Prado V, Vial P, Levine MM (1987) Patterns of adherence of diarrheagenic *Escherichia coli* to HEp-2 cells. *Pediatr Infect Dis J* 6:829–831
- Nataro JP, Deng Y, Maneval DR, German AL, Martin WC, Levine MM (1992) Aggregative adherence fimbriae I of enteroaggregative *Escherichia coli* mediate adherence to HEp-2 cells and hemagglutination of human erythrocytes. *Infect Immun* 60:2297–2304
- Nataro JP, Deng Y, Cookson S, Cravioto A, Savarino SJ, Guers LD, Levine MM, Tacket CO (1995) Heterogeneity of enteroaggregative *Escherichia coli* virulence demonstrated in volunteers. *J Infect Dis* 171:465–468
- Navarro-García F, Eslava C, Villaseca JM, López-Revilla R, Czeckulin JR, Srinivas S, Nataro JP, Cravioto A (1998) In vitro effects of a high-molecular weight heat-labile enterotoxin from enteroaggregative *Escherichia coli*. *Infect Immun* 66:3149–3154
- Novick RP, Roth C (1968) Plasmid-linked resistance to inorganic salts in *Staphylococcus aureus*. *J Bacteriol* 95:1335–1342
- Orth D, Grif K, Dierich MP, Würzner R (2007) Variability in tellurite resistance and the *ter* gene cluster among Shiga toxin-producing *Escherichia coli* isolated from humans, animals and food. *Res Microbiol* 158:105–111
- Paterson DL, Bonomo RA (2005) Extended-spectrum beta-lactamases: a clinical update. *Clin Microbiol Rev* 18:657–686
- Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS (2000) Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* 406:64–67
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B (2000) ACT: the artemis comparison tool. *Bioinformatics* 16:944–945
- Schmeisser C, Liesegang H, Krysiak D, Bakkou N, Le Quére A, Wollherr A, Heinemeyer I, Morgenstern B, Pommerening-Röser A, Flores M, Palacios R, Brenner S, Gottschalk G, Schmitz RA, Broughton WJ, Perret X, Strittmatter AW, Streit WR (2009) *Rhizobium* sp. strain NGR234 possesses a remarkable number of secretion systems. *Appl Environ Microbiol* 75:4035–4045
- Smet A, Van Nieuwerburgh F, Vandekerckhove TT, Martel A, Deforce D, Butaye P, Haesebrouck F (2010) Complete nucleotide sequence of CTX-M-15-plasmids from clinical *Escherichia coli* isolates: insertional events of transposons and insertion sequences. *PLoS ONE* 5:e11202
- Smith DH (1967) R factors mediate resistance to mercury, nickel and cobalt. *Science* 156:1114–1116
- Summers AO, Silver S (1972) Mercury resistance in plasmid-bearing strains of *Escherichia coli*. *J Bacteriol* 112:1228–1236
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol and Evol* 10:512–526
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. doi:10.1093/molbev/msr121
- Tarr PI, Gordon CA, Chandler WL (2005) Shiga toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *Lancet* 365:1073–1086
- Taylor DE, Rooker M, Keelan M, Ng LK, Martin I, Perna NT, Burland NT, Blattner FR (2002) Genomic variability of *O* islands encoding tellurite resistance in enterohemorrhagic *Escherichia coli* O157:H7 isolates. *J Bacteriol* 184:4690–4698
- Tenaillon O, Skurnik D, Picard B, Denamur E (2010) The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* 8:207–217
- Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard M, Dossat C, Karoui ME, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguéne C, Lescat M, Mangenot S, Martinez-Jéhanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf CS, Schneider D, Tournet J, Vacherie B, Vallenet D, Médigue C, Rocha EP, Denamur E (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 5:e1000344
- Whelan KF, Collieran E, Taylor DE (1995) Phage inhibition, colicin resistance, and tellurite resistance are encoded by a single cluster of genes on the IncHI2 plasmid R478. *J Bacteriol* 177:5016–5027
- Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M (2006) Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* 60:1136–1151
- Woodford N, Carattoli A, Karisik E, Underwood A, Ellington MJ, Livermore DM (2009) Complete nucleotide sequences of plasmids pEK204, pEK499, and pEK516, encoding CTX-M enzymes in three major *Escherichia coli* lineages from the United Kingdom, all belonging to the international O25:H4-ST131 clone. *Antimicrob Agents Chemother* 53:4472–4482
- Zadik PM, Chapman PA, Siddons CA (1993) Use of tellurite for the selection of verocytotoxigenic *Escherichia coli* O157. *J Med Microbiol* 39:155–158