MICROBIAL GENOMICS

RESEARCH ARTICLE

Allué-Guardia *et al.*, *Microbial Genomics* 2022;8:000796

DOI 10.1099/mgen.0.000796

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Pathogenomes and variations in Shiga toxin production among geographically distinct clones of *Escherichia coli* O113:H21

Anna Allué-Guardia[1,2], Sara S. K. Koenig[1,2], Ricardo A. Martinez[1,2], Armando L. Rodriguez[3], Joseph M. Bosilevac[4], Peter Feng†[5] and Mark Eppinger[1,2,*]

Abstract

Infections with globally disseminated Shiga toxin-producing *Escherichia coli* (STEC) of the O113:H21 serotype can progress to severe clinical complications, such as hemolytic uremic syndrome (HUS). Two phylogeographically distinct clonal complexes have been established by multi locus sequence typing (MLST). Infections with ST-820 isolates circulating exclusively in Australia have caused severe human disease, such as HUS. Conversely, ST-223 isolates prevalent in the US and outside Australia seem to rarely cause severe human disease but are frequent contaminants. Following a genomic epidemiology approach, we wanted to gain insights into the underlying cause for this disparity. We examined the plasticity in the genome make-up and Shiga toxin production in a collection of 20 ST-820 and ST-223 strains isolated from produce, the bovine reservoir, and clinical cases. STEC are notorious for assembly into fragmented draft sequences when using short-read sequencing technologies due to the extensive and partly homologous phage complement. The application of long-read technology (LRT) sequencing yielded closed reference chromosomes and plasmids for two representative ST-820 and ST-223 strains. The established high-resolution framework, based on whole genome alignments, single nucleotide polymorphism (SNP)-typing and MLST, includes the chromosomes and plasmids of other publicly available O113:H21 sequences and allowed us to refine the phylogeographical boundaries of ST-820 and ST-223 complex isolates and to further identify a historic non-shigatoxigenic strain from Mexico as a quasi-intermediate. Plasmid comparison revealed strong correlations between the strains' featured pO113 plasmid genotypes and chromosomally inferred ST, which suggests coevolution of the chromosome and virulence plasmids. Our pathogenicity assessment revealed statistically significant differences in the Stx_{2a}-production capabilities of ST-820 as compared to ST-223 strains under RecA-induced Stx phage mobilization, a condition that mimics Stx-phage induction. These observations suggest that ST-820 strains may confer an increased pathogenic potential in line with the strain-associated epidemiological metadata. Still, some of the tested ST-223 cultures sourced from contaminated produce or the bovine reservoir also produced Stx at levels comparable to those of ST-820 isolates, which calls for awareness and for continued surveillance of this lineage.

Received 10 September 2021; Accepted 07 February 2022; Published 08 April 2022

Author affiliations: [1]Department of Molecular Microbiology and Immunology, University of Texas at San Antonio, San Antonio, TX, USA; [2]South Texas Center for Emerging Infectious Diseases (STCEID), San Antonio, TX, USA; [3]University of Texas at San Antonio, Research Computing Support Group, San Antonio, TX, USA; [4]U.S. Department of Agriculture (USDA), Agricultural Research Service (ARS), Roman L. Hruska U.S. Meat Animal Research Center, Clay Center, NE, USA; [5]U.S. Food and Drug Administration (FDA), College Park, MD, USA.

***Correspondence:** Mark Eppinger, mark.eppinger@utsa.edu

Keywords: Shiga toxin (Stx) producing *Escherichia coli* (STEC); O113:H21 serotype; long-read technology (LRT); Whole Genome Sequencing and Typing (WGST); Multilocus Sequence Type (MLST); phylogenomics; genomics epidemiology.

Abbreviations: HUS, hemolytic uremic syndrome; LB, Lysogeny Broth; LRT, long-read technology; MLST, Multilocus Sequence Type; SNP, Single Nucleotide Polymorphism; SRA, Sequence Read Archive; ST, Sequence Type; STEC, Shiga toxin producing *Escherichia coli*; Stx, Shiga toxin; TTP, Thrombotic Thrombocytopenic Purpura; WGST, Whole Genome Sequencing and Typing.

Sequencing datasets for all isolates analysed in this study have been deposited in the Sequence Read Archive (SRA) and the Whole Genome Shotgun Repository at National Center for Biotechnology Information (NCBI). Accessions for reads, and the assembled and annotated genomes are provided in Table S1 [VV00000000, VEOD00000000, VENU00000000, VENT00000000, VENS00000000, VEOC00000000, VENR00000000, VEOF00000000, VEOI00000000, VEOE00000000, VEOH00000000, VEOG00000000, VENY00000000, VENX00000000, VENW00000000, VENZ00000000, VEOB00000000, VEOA00000000, CP045213, CP045214, CP045215, CP045209, CP045210, CP045211, CP045212]; along with accessions and strain-associated metadata of additional STEC O113:H21 strains that have been retrieved from NCBI is support of lineage-scale analyses.

†Retired.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Four supplementary figures and eight supplementary tables are available with the online version of this article.

DATA SUMMARY

The authors confirm all supporting data, code and protocols have been provided within the article or through supplementary data files.

Impact Statement

Shiga toxin-producing *Escherichia coli* (STEC) of serotype O113:H21 are a globally disseminated lineage, which can be partitioned into two major phylogeographical complexes: Sequence Type (ST)–820 strains reside in Australia and have been linked to severe outbreaks of human disease, while ST-223 strains are found outside Australia and are common adulterants, yet with unknown association to human disease. To discern potential differences in the pathogenome make-up and conferred pathogenicity associated with these clonal complexes, we assembled a representative collection of twenty O113:H21 strains. Informed by the sequenced genomes along with recorded Stx-production pathotypes, we were able to refine the phylogenomic and virulence boundaries associated with these two complexes. The established high-resolution framework (based on whole genome alignment, SNP-typing and MLST) allowed us to refine the phylogenomic boundaries between the two complexes and identify a historic non-shigatoxigenic O113:H21 strain from Mexico as quasi-intermediate. Our pathogenicity assessment supports the notion of an increased pathogenic potential of the HUS-associated ST-820 strains consistent with the known strain-associated epidemiological metadata. Insights into the genomic and phenotypic plasticity of STEC on a lineage- and genome-wide scale are foundational to improve and inform risk assessment, biosurveillance, and prevention strategies for STEC.

INTRODUCTION

Shiga toxin (Stx)-producing *Escherichia coli* are notorious for producing a phage-borne toxin [1–5] that is specifically toxigenic towards renal endothelial cells [6–13]. Infections can progress to life-threatening complications, such as hemolytic uremic syndrome (HUS) [14]. Pathogenicity in humans is inexorably linked to the Stx litres produced by individual STEC strains [15, 16]. Hypervirulent clones, as manifested by increased Stx_{2a}-litres [17–24], have been associated with STEC subpopulations through phylogenetic, epidemiological, and phenotypic linkage [4, 5, 11, 19, 20, 25–49]. The specific factors responsible for elevated Stx-production in hypervirulent STEC strains are unknown but presumably modulated by a number of biotic and abiotic triggers [6, 8, 50–54]. Globally disseminated STEC of serotype O113:H21 were first associated with HUS cases in 1983 [14]. The most potent cytopathic toxin subtype, Stx_{2a}, is commonly found in the O113:H21 serotype [55, 56] that lacks the locus of enterocyte effacement (LEE) [57, 58]. Multilocus sequence typing (MLST) has identified two major phylogeographical complexes, where ST-223 strains are found around the world, while ST-820 strains are restricted to Australia [59, 60]. In particular, infections with Australian ST-820 strains have been associated with severe clinical complications [61–63], while ST-223 isolates from the US [55, 64] and elsewhere [65–67] have been rarely associated with severe human disease, even though they are frequent contaminants of produce and cattle [68–70] and can possess virulence traits similar to the clinical O113:H21 strains [59, 60]. There is a dearth of knowledge of the intrinsic genomic make-up and Stx-production associated traits of ST-223 and ST-820 complex isolates that might contribute to the disparity in human disease manifestation. To investigate the plasticity and to discern potential differences in the pathogenome and Stx-production traits, we assembled a collection of O113:H21 strains of global origin including strains from Australia, Asia, Europe and North and South America from clinical cases, cattle and contaminated produce. Following a genomic epidemiology approach, we sequenced a total of 20 O113:H21 strains and analysed their Stx virulence phenotypes alongside 15 other published genomes. Through comprehensive genotypic and phenotypic analyses, we determined the genome make-up and virulence traits associated with Stx-production in these representative ST-223 and ST-820 complex isolates and established a high-resolution phylogenomic framework. Integrating the virulence information to the genome plasticity within this lineage is foundational for improved risk assessment, biosurveillance, and the development of prevention strategies for STEC [71, 72].

METHODS

Bacterial strains analysed in this study

A collection of O113:H21 cultures and genomes, representing the two major complexes ST-223 (#12) and ST-820 (#8), isolated from the bovine and environmental reservoir, produce, and clinical cases were sequenced and analysed in this study. Strain-associated metadata can be found in Table S1 (available in the online version of this article). The sequenced strains include eight Australian strains isolated from either HUS, thrombotic thrombocytopenic purpura (TTP), or dysentery patients, and nine strains from the US recovered from ground beef or spinach [68]. We further included Canadian strains CL-3 (AGTH01000000.1) and TW01391, the latter sequenced in this study, which are clones from different culture repositories; strain TW02918, isolated from

a diarrhoea patient in Thailand, and strain 55 isolated from Uruguayan beef [59, 60, 73]. To support lineage-scale analyses we examined 15 additional O113:H21 genomes retrieved from NCBI GenBank/SRA: eight water-, cattle-, and swine strains from a major produce region in California [55, 64, 74], five US strains from four clinical cases and one from cattle faeces, and a single isolate of unknown source [75, 76]. The set further contains German strain TS18-98 from minced meat and an historic Mexican strain 6182–50 from a clinical diarrhoea case dating back to 1950 [77]. For the purpose of this study, we used strains EH41 and 4 as reference for ST-820 and ST-223, respectively.

Genome sequencing, assembly and annotation

Strains were cultured overnight (o/n) at 37 °C with shaking in lysogeny broth (LB) (Fisher Scientific, Thermo Fisher Scientific, Asheville, NC, USA). Total genomic DNA was extracted from o/n cultures using the QIAamp DNA Mini Kit (Qiagen, Inc., Valencia, CA, USA) according to the manufacturer's instructions, followed by short-read Illumina sequencing on the MiSeq platform. Paired-end libraries were prepared using either the NxSeq AmpFREE Low DNA Library Kit (Lucigen) with 250 bp read length and sequenced with the MiSeq Reagent kit v2 500-cycle (Illumina), or the KAPA HyperPlus DNA kit (Roche) with 300 bp read length and sequenced with the MiSeq Reagent kit v3 600-cycle (Illumina), following the manufacturer's guidelines. Illumina reads were assembled with SPAdes [77]. The genomes of ST-223 and ST-820 complex strains 4 and EH41 were sequenced to closure by complementing Illumina short-reads with Oxford Nanopore long-reads on the MinION platform. Genomic DNA was diluted to a concentration of 1.5 µg in 46 µl of nuclease-free water. The library was prepared using the Nanopore 1D Ligation sequencing kit SQK-LSK108 (R9) with the Native barcoding kit EXP-NBD103 according to the manufacturer's instructions and sequenced on a MinION Mk1B with a SpotON flow cell FLO-MIN107 (R9). Nanopore and Illumina reads were used for hybrid assembly using SPAdes in the careful mode, which includes realignment correction [77]. We further assembled a draft genome of strain 6182–50, for which only reads were available in the NCBI Sequence Read Archive (SRA). Resulting assemblies were QCed with QUAST [78, 79]. The chromosomal and plasmid origins of replication, *oriC* (http://tubic.tju.edu.cn/Ori-Finder/) [80] and *repA*, respectively, were designated as the zero point of the closed EH41 and 4 molecules prior to annotation using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) [81].

MLST classification

MLST typing was performed using targeted and whole genome schemas developed for *E. coli* [82, 83]. We determined the Sequence Type (ST) by applying three different schemas as follows: The ST was inferred *in silico* using the Achtman (7-gene) [84] and Pasteur schemas [85] by examining assembled genomes (https://cge.cbs.dtu.dk/services/MLST/) [86]. Through *in silico* and PCR-based-interrogation [59, 60], we inferred the ST according to the Whittam 7- and 15-gene MLST scheme (http://shigatox.net/new/tools/ecmlst.html) [87]. Sequences of the target genes were then queried against the EcMLST database [88]. For NCBI-retrieved genomes for which cultures were not available, alleles and ST were determined *in silico* by BLASTn comparison against the EcMLST database [88]. Assembled genomes were further genotyped in MLST 2.0 (https://cge.cbs.dtu.dk/services/MLST/) [86] and Ridom SeqSphere+ (v6.0.2) [89] to establish a whole genome (wg) MLST phylogeny.

Comparative phylogenomics

Whole genome alignment (WGA) phylogeny

The 20 O113:H21 genomes sequenced in this study along with 15 genomes downloaded from NCBI GenBank (Table S1) were used to construct a whole genome-based phylogenetic tree. The phylogeny was inferred from WGAs using Mugsy (v1.2.3) [90] and RAxML (v4.0) [91] with 100 bootstrap replicates. The tree was visualized in Geneious Prime (v2021.1.1) [92] and decorated with strain-associated metadata in EvolView [93–95].

Core genome SNP phylogeny

To compute a SNP phylogeny, we used a custom-built core genome (cg) SNP discovery pipeline described in more detail in [96], implemented on the open-source bioinformatics platform Galaxy [97]. The chromosomal core genome was defined as the set of genic and intragenic regions that are not repeated and do not contain phages, IS elements, plasmid regions, genomic islands or other mobile genetic elements, which evolve at different rates and are not indicative of evolutionary relationships. These regions were determined in the designated closed reference O113:H21 strain EH41 as follows: Repeats with NUCmer (v3.22) [98], prophages with PHASTER [99, 100], and IS elements with ISFinder [101], ISEScan (v1.7.1) [102], and ICEberg (v2.0) [103]. The modular pipeline contains the following workflow steps: (***i)* SNP discovery and typing**. When available, Illumina reads were used for read-based SNP discovery. Reads were aligned to the designated reference with BWA-MEM [104]. The resulting alignments were processed with Freebayes (v1.3.1) [105] with the following threshold settings: mapping quality 30, base quality 30, coverage 10, and allelic frequency 0.75. For contig-based discovery, assemblies were aligned to the EH41 reference chromosome using NUCmer followed by SNP prediction with delta-filter and show-snps distributed with the MUMmer package [98, 106]. The resulting SNP panel for each of the query genomes was used for further processing. (***ii)* SNP validation and filtering**. We used several SNP curation strategies detailed in our previous works [96, 107, 108]: catalogued SNPs from each genome were merged into a single SNP panel and SNPs located within identified excluded regions were removed, as well as low quality alignments or

misalignments, non-uniformly distributed regions, and InDels, as previously [108–110]. SNPs were further curated by extracting the surrounding 40 nucleotides (nt) for each predicted SNP in the reference genome, followed by BLASTn of these fragments against the query genomes [111]. SNPs with missing information ('no hits') or multiple hits were filtered out, as well as ambiguous nucleotides. (***iii*) SNP annotation and chromosomal distribution**. Allelic status and chromosomal position of SNPs were recorded. To account for the biological relevance of these point mutations, polymorphisms were classified into genic or intergenic by mapping the SNPs to the reference genome. SNP-matrix tables were manipulated with Query Tabular Tool [112]. In addition, we developed a genotyper tool to provide SNP statistics reporting on the number of individual genotypes in the phylogeny. ***(iv)* SNP phylogeny**. The curated panel of high quality SNPs served as basis for phylogenetic reconstruction by maximum parsimony with PAUP (v4.0a163) [113] with a 100 bootstrap replicates. The majority rule consensus SNP tree was visualized in Geneious Prime (v2021.1.1) [92] and decorated in EvolView (v3) [93–95]. Calculation of the consistency index (CI) in Mesquite (v3.6) [114] for each SNP allowed us to identify parsimony informative SNPs and flag homoplastic SNPs, as previously described [96, 107, 108, 115–117]. For genomes retrieved from NCBI GenBank, where reads were not available, we interrogated the allelic status of the catalogued SNPs in the assembled genomes.

Whole genome MLST derived phylogeny of STEC O113:H21

Classification results from the wgMLST analysis in Ridom SeqSphere+ [89] were used to construct a minimum spanning tree (MST) for alleles present in all isolates with default settings. The resulting tree was decorated with strain-specific metadata, including the respective allele status of the 7-gene and 15-gene Whittam MLST classification [87] inferred from EcMLST [88].

Pathogenome make-up of STEC O113:H21

The virulence complement was identified using VirulenceFinder (https://cge.cbs.dtu.dk/services/VirulenceFinder/) [118, 119] and VDFDB [120]. The resistome was analysed with the Comprehensive Antibiotic Resistance Database (CARD) (https://card.mcmaster.ca/home) [121], ARG-ANNOT [122] and ResFinder (https://cge.cbs.dtu.dk/services/ResFinder/) [123, 124]. Prophages were distinguished from the core genomes using PHASTER [99, 100]. Plasmid replicon and relaxase types and conjugation potential was determined with PlasmidFinder (https://cge.cbs.dtu.dk/services/PlasmidFinder/) [125] and MOB-suite [126].

Stx-bacteriophage profiling and visualization

Boundaries and locations of intact, partial, or remnant prophages were identified using PHASTER [100]. For Stx-bacteriophages the stx-subtypes and insertion sites were recorded as described in (Scheutz *et al.*, 2012; Ashton *et al.*, 2015). The *in silico* delineated *stx* subtypes for the O113:H21 strains sequenced in this study, were confirmed by PCR according to (Scheutz *et al.*, 2012) using GoTaq Green Master Mix (Promega) in a 25 µl reaction volume. The Stx-bacteriophages of closed genomes were compared and visualized using Geneious Prime (v2021.1.1) [92] and Easyfig (v2.2.2) [127].

Shiga toxin subtyping

To confirm the stx_{2d} allele subtype *in silico*, complete stx_2 genes were translated and aligned using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) [128, 129]. We confirmed the presence of the activatable property of Stx_{2d} located at the C-terminal end of the A_2 subunit (KSQSLYTTGE). Amino acid sequences of the B-subunit of Stx_{2d} are distinguished from Stx_{2a} by serine (S) and glutamic acid (E) along with the N-terminal END-motif motif [130, 131]. Virulence genes were identified with VirulenceFinder [118, 124, 132] and VFDB [120].

Mobile genome element (MGEs) and InDels

Insertion sequence elements (IS) were identified and classified with Iceberg [103], and ISEScan [102] in Galaxy [133]. ICE/IME regions were determined in ICEfinder (https://bioinfo-mml.sjtu.edu.cn/ICEfinder/index.php) [103]. Genomic islands (GI) were detected with IslandViewer4 [134–136]. InDels between the closed chromosomes of strains EH41 and 4 were identified in Geneious Prime (v2021.1.1) [92].

Comparison of genome architectures and distribution of virulence genes

Architectures and gene inventories of closed chromosomes and virulence plasmids were comprehensively analysed and visualized with Blast Ring Image Generator (BRIG) [137]. Chromosomal, phage-and plasmid-borne pathogenicity genes were recorded with Virulence Finder [118, 124, 132], PHASTER [100] and by tBLASTn [111] query of a list of established O113:H21 virulence determinants [56, 60]. To study the prevalence of identified virulence gene complement of the core and carried plasmids in all sampled closed and draft genomes, we used Large-Scale BLAST Score Ratio (LS-BSR) [96, 138, 139] with tBLASTn [111]. We compared both the overall proteome inventory and the prevalence of the combined virulence factor complement identified in this strain set. For each of the proteins, a BLASTp raw score was obtained for the alignment against itself (reference bit score) and the most similar protein (query bit score) in each of the genomes. The BSR value was calculated by dividing the query bit score by the reference bit score, resulting in a BSR value between 0.0 and 1.0. Proteins with a normalized BSR of <0.4 were not considered homologous. The resulting BSR matrix or alternatively the percent

identities were visualized as heatmaps with Multiple Experiment Viewer (MeV) (v4.8) [140] and Morpheus (https://software.broadinstitute.org/morpheus).

Stx-production pathotypes

The Stx_2-production phenotypes of the cultures were determined by recording the Stx litres through Enzyme-Linked ImmunoSorbent Assay (ELISA) under both spontaneous and Mitomycin C (MMC)-induced conditions. Overnight (o/n) cultures were diluted to an OD_{600} of 0.05 and grown to an OD_{600} of 0.3–0.5 in fresh LB media at 37 °C. At this stage, cultures were split and incubated at 37 °C for 6 h under non-induced and induced (MMC: 0.5 µg ml^{-1}) conditions. Toxin production was measured after harvesting 5 ml of each culture for parallel processing. To lyse bacterial cells and release produced Stx, induced cultures were treated with polymyxin B (6 mg ml^{-1} 37 °C, 10 min). Supernatants were collected after centrifugation (3500 r.p.m., 10 min), filtered through 0.22 µm low protein-binding membrane filters (Millex-GP; Millipore) and diluted to measurable concentrations. Stx_2-production was measured using the Premier EHEC kit (Meridian Bioscience, Cincinnati, OH) following the manufacturer's instructions. Titres were calculated using a standard curve generated from serial dilutions of purified Stx_{2a} (BEI resources, NR-4478). Statistical significance was determined using Prism v9.0.1 (GraphPad Software, San Diego, CA). A two-way ANOVA with Sidak's multiple comparisons test was used to compare non-induced vs. MMC-induced conditions for each strain. A one-way ANOVA with Tukey's multiple comparisons test was used to compare Stx-production by ST-223 and ST-820 groups under (both non-induced and MMC-induced conditions.

RESULTS

Whole genome sequencing of a global collection of 20 O113:H21 strains

For this study, we sequenced and analysed the pathogenomes and Stx-production traits of 20 O113:H21 isolates sourced from the bovine reservoir, produce, or clinical cases (Table S1); along with 15 O113:H21 genomes retrieved from NCBI GenBank, which added four closed and 11 draft genomes to our dataset [64, 75, 76]. The strains in our collection represent the two major phylogeographical complexes: Australian ST-820 and ST-223 strains from the US and elsewhere including subvariants (Table S2) (Reid *et al.*, 2000). To establish a refined high-resolution phylogenomic framework for this lineage, we sequenced the genomes of the designated reference strains EH41 and 4 to closure using the Nanopore LRT platform. These strains served as high quality representative genomes for the ST-820 and ST-223 complex respectively, while the other strains were sequenced using Illumina technology to draft stage, yielding between 41 and 131 contigs (Table S1). The predicted genome sizes in this set range from 4.9 to 5.1 MB with an average GC-content of 50.8%. The chromosomes of strains EH41 (ST-820) and 4 (ST-223) are: 5040503 and 4907913 bp in size with 4978/5054 coding sequences (CDS), 89/94 tRNAs and 22/22 rRNAs, respectively. The application of LRT allowed us to also recover and accurately assemble the plasmids of strains 4 and EH41 (Table S1). MLST types, as inferred from the 7- and 15-gene Whittam schemes, grouped the isolates into two major phylogeographical complexes (Table S2) [59, 60]: ST-820 strains, a distinct clonal group found in Australia, and ST-223 strains found in the US [55, 64] and elsewhere in the world [65–67]. However, two US strains from ground beef, 16 and 53, were classified as ST-820, characteristic for Australian strains. A likely explanation is that the US imports lean boneless beef trimmings from different countries as ingredients for ground beef manufacturing. Strain 53, was isolated in the US from such trim that had been imported from Australia [73]. Although direct production records are lacking, strain 16 was isolated from ground beef later produced by the same supplier of the imported beef trim samples [68], thus strain 16 too, likely originated in Australia. All but two non-Australian strains are of ST-223, while clinical isolates 2013 C-3181 and 10 are classified as ST-846 and ST-234, respectively.

Comprehensive analyses of pathochromosome architectures, phages and plasmids

To initially assess the degree of chromosomal plasticity within the O113:H21 serotype we compared the closed chromosomes of ST-820 and ST-223 reference strains EH41 and 4 along with four closed genomes retrieved from GenBank (2014C-4135 (ST-223), 00–3076 (ST-223), 2013 C-8131 (ST-846) [76] and RM10466 (ST-223) [74]. The pairwise identity of ST-820 strain EH41 to strains 4 (ST-223) and 2013 C-8131 (ST-846) is 79.6 and 66.6%, respectively. The chromosomal architectures of closed and draft genomes were compared using strain EH41 (ST-820) as the designated reference. As evident in Fig. 1, we observed a largely genome-wide synteny of the chromosomal backbone disrupted by multiple mobile genome elements (MGE), which are major contributors of O113:H21 genome diversification [141]. Catalogued prophages and their respective length, chromosomal insertion sites, GC-content, and number of predicted coding sequences (CDS) are listed in Table S3. In the six closed genomes that were available for this study, prophages account for 5.5–6.1% of the total chromosome and their dynamic acquisition and secondary loss contributes to the genome size variation observed in the analysed strains (Table S1). We further identified a highly plastic region (HPR) spanning about 200 kb that features multiple InDels associated with MGEs. A comparison referenced to strain 4 (ST-223) can be found in Fig. S1. We note that the phylogeographical separation of ST-223 and ST-820 complex isolates is mirrored in the HPR genomic organization and composition. The genomes of the closed ST-223 strains 4, 00–3076, and RM10466 are distinguished by deletions not present in EH41 (ST-820). This particular

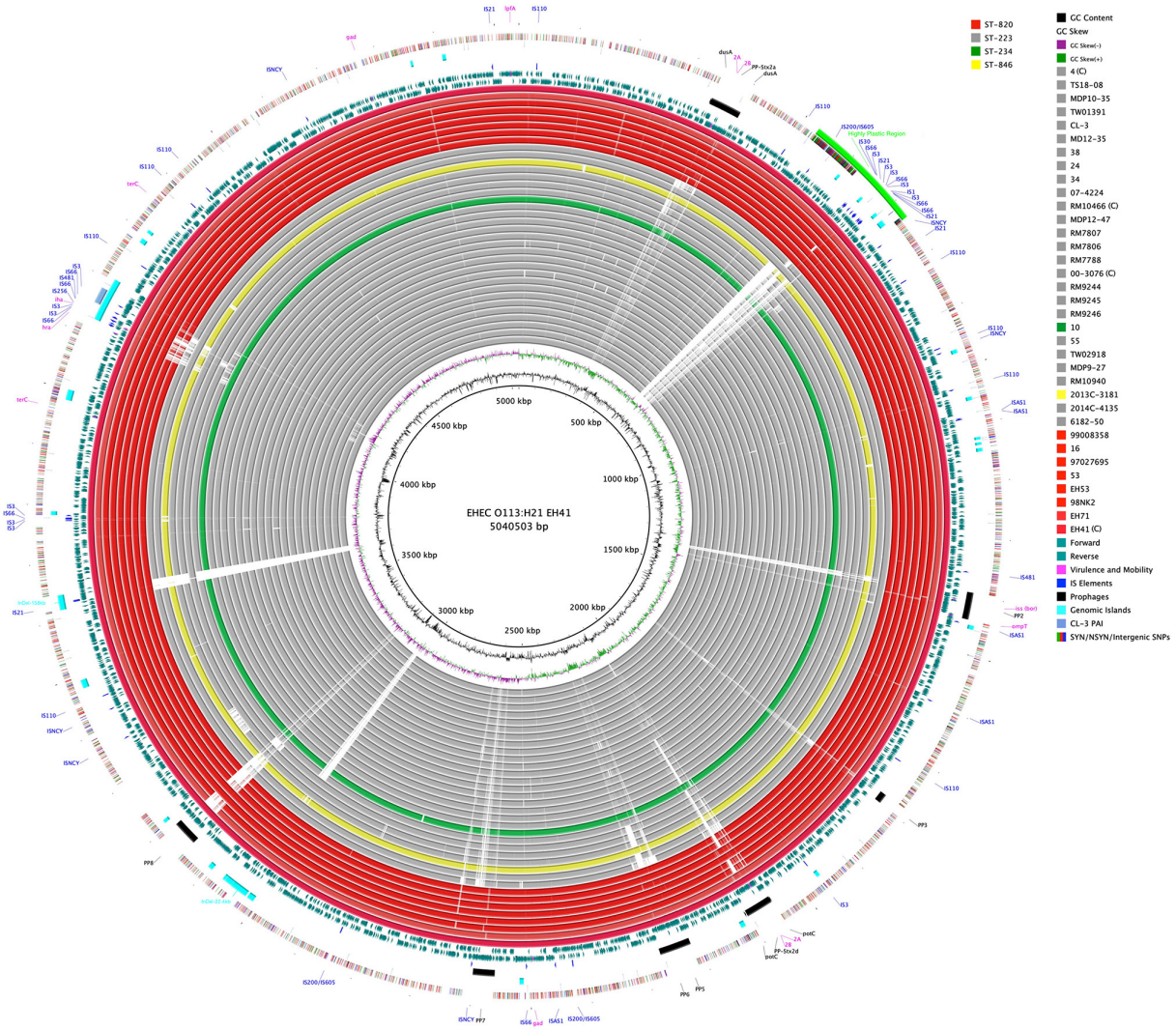
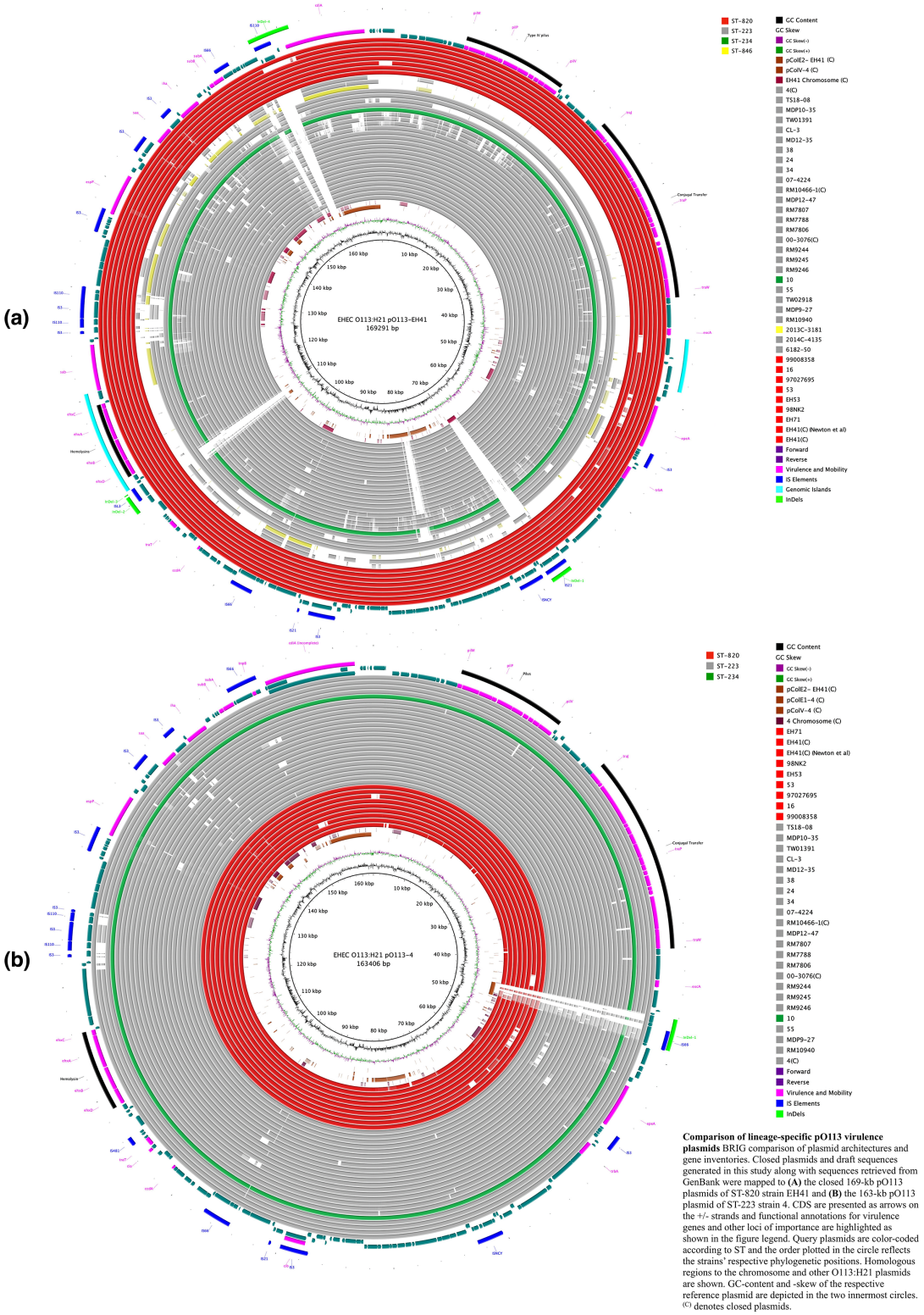


Fig. 1. Chromosome architecture of STEC O113:H21 BRIG comparison of O113:H21 genomes referenced to the 5040503 bp chromosome of ST-820 strain EH41. CDS are presented as arrows on the +/-strands, and functional annotations for virulence genes and other loci of importance are highlighted as shown in the figure legend. Query genomes are color-coded according to ST and the order plotted in the circle reflects their respective phylogenetic positions. Chromosomal synteny is disrupted by multiple prophages and other MGEs. GC-content and GC-skew of the EH41 chromosome are depicted in the two innermost circles. (C) denotes closed chromosomes.

pattern was also seen in the remainder of analysed ST-223 and ST-820 draft genomes with the notable exception of ST-223 strains 6182–50 and 2014 C-4135. The lineage-specific virulence plasmid, pO113, was detected in the assemblies of all but four strains (ST-223 strains 6182–50, 2014 C-4135, TW02918 and ST-846 strain 2013 C-3181) [141] (Fig. 2). The four closed plasmids shows a highly conserved and syntenic plasmid backbone without any major structural differences featuring a 99% nucleotide identity over the entire plasmid length [111, 142]. Our analysis identified several InDels when compared to the closed ST-reference pO113 plasmids of strains EH41 and 4 (Fig. 2), which are associated with mobile genetic elements (Fig. S2). The genetic relatedness of strains is also reflected in the plasticity found within the lineage-specific virulence plasmid (Fig. 2). We note here that we identified several plasmid-borne signatures unique to ST-820/ST-223 that can be utilized for ST genotyping in analogy to chromosomal MLST markers defining the ST. Colicins are produced by and are toxic to some *E. coli* strains to reduce competition from other phylogenetically related microbial strains [143], including certain O113:H21 strains [143, 144]. Both closed ST-reference EH41 and 4 strains are colicinogenic and code for colicins E2 and E1, respectively encoded on relative small plasmids [143] (Fig. 3). Strain 4 further carries a 65782 bp colicin V plasmid pColV-4 [145]. BLASTn of its sequence [111, 142] against the non-redundant (nr) NCBI database revealed local similarity to O113:H21 plasmid pRM10466-2 (Fig. 3c) and other phylogenetically diverse *E. coli* plasmids. Plasmid profiling of the other genomes identified a number of plasmid replicons, which provides a testament to the considerable plasticity of the plasmid types carried in this



Comparison of lineage-specific pO113 virulence plasmids BRIG comparison of plasmid architectures and gene inventories. Closed plasmids and draft sequences generated in this study along with sequences retrieved from GenBank were mapped to **(A)** the closed 169-kb pO113 plasmids of ST-820 strain EH41 and **(B)** the 163-kb pO113 plasmid of ST-223 strain 4. CDS are presented as arrows on the +/- strands and functional annotations for virulence genes and other loci of importance are highlighted as shown in the figure legend. Query plasmids are color-coded according to ST and the order plotted in the circle reflects the strains' respective phylogenetic positions. Homologous regions to the chromosome and other O113:H21 plasmids are shown. GC-content and -skew of the respective reference plasmid are depicted in the two innermost circles. (C) denotes closed plasmids.

Fig. 2. Comparison of lineage-specific pO113 virulence plasmids BRIG comparison of plasmid architectures and gene inventories. Closed plasmids and draft sequences generated in this study along with sequences retrieved from GenBank were mapped to (a) the closed 169 kb pO113 plasmids of ST-820 strain EH41 and (b) the 163 kb pO113 plasmid of ST-223 strain 4. CDS are presented as arrows on the +/-strands and functional annotations for virulence genes and other loci of importance are highlighted as shown in the figure legend. Query plasmids are color-coded according to ST and the order plotted in the circle reflects the strains' respective phylogenetic positions. Homologous regions to the chromosome and other O113:H21 plasmids are shown. GC-content and -skew of the respective reference plasmid are depicted in the two innermost circles. (C) denotes closed plasmids.

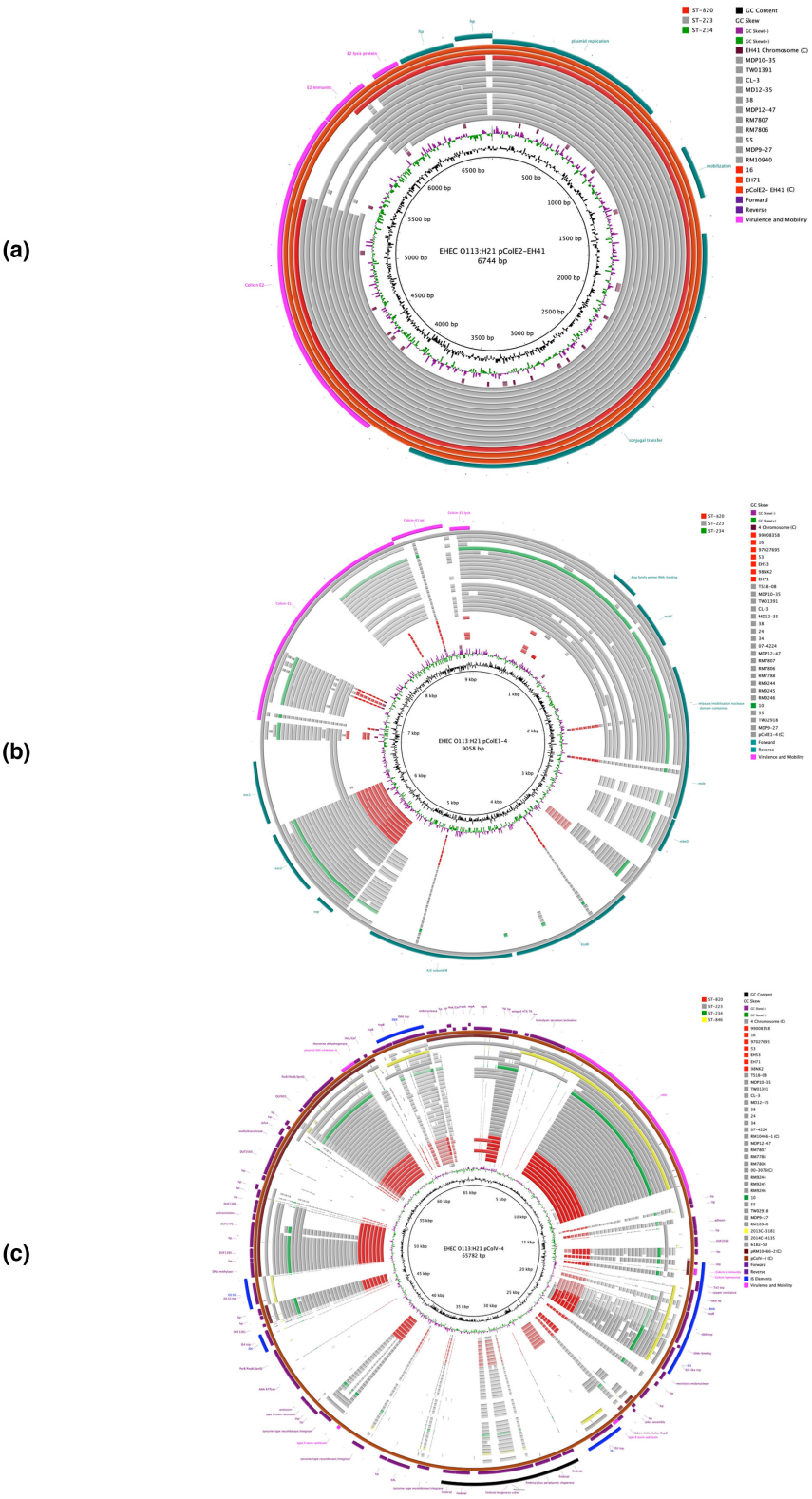


Fig. 3. Colicin plasmids in closed reference strains EH41 (ST-820) and 4 (ST-223) BRIG comparison of colicin plasmid architectures and gene inventories of (a) pColE2-EH41, (b) pColE1-4 and (c) pColV-4. CDS are presented as arrows on the +/-strands and functional annotations for virulence genes and other loci of importance are highlighted. Query plasmids are color-coded according to ST and the order plotted in the circle reflects the strains' respective phylogenetic positions. Homologous regions to the chromosome and other O113:H21 plasmids are shown. GC-content and -skew of the reference plasmid are depicted in the two innermost circles. (C) denotes closed plasmids.

lineage. Further details on replicon subtypes and plasmid content can be found in Table S4, although the draft status of these genomes did not allow us to fully reconstruct their plasmid inventories.

High-resolution phylogenomic framework for STEC O113:H21

To investigate the phylogenomic boundaries between ST-820 and ST-223 and to determine the relationships of individual strains, we established a refined high-resolution phylogenomic framework for the O113:H21 lineage [60, 67] that was inferred from expanded wgMLST, WGA, and SNPs. We computed an MLST phylogeny based on the Whittam schema (15-gene) by inferring the allele status in Ridom SeqSphere+ and assigning allele profiles in the EcMLST database. (Fig. 4a) [88, 89]. Within ST-223 and ST-820 complex isolates we detected further allelic variations when applying both the Whittam and 7/15-gene MLST schemas, which is detailed in Table S2. Five strains (RM7788, RM7806, RM7807, MDP12-47 and 55) are classified as ST-223v3, distinguished by alleles *mdh8*^{C124T} or *mdh318*^{G526T} [59]. The wgMLST phylogeny, decorated with the delineated ST-types, mirrors the phylogeographical separation observed in the MST tree, but provides higher resolution (Fig. 4b). Two ST-223 strains, 6182–50 from Mexico [77] and 2014 C-4135 from the US, are positioned in the phylogenetic boundary between the major ST-223 and ST-820 complexes, suggesting that these isolates might occupy a quasi-intermediate position. To further investigate this question, we constructed phylogenetic hypotheses that are based on whole genome alignments (WGA) and *de novo* SNP discovery [96]. The resulting WGA tree topology (Fig. 5) corroborates with phylogenetic placement of strains by wgMLST (Fig. 4). Core genome (cg) SNPs in the 35 O113:H21 genomes yielded a total of 4398 high-quality SNPs (Table S5). The constructed maximum parsimony (MP) tree shows bootstrap supports greater than 80% for the majority of nodes (Fig. 6), and its topology mirrors the generated wgMLST- and WGA-phylogenies (Figs 4 and 5). In the catalogued SNP panel, we identified 1979 parsimony informative SNPs and delineated a total of 77 clades and clusters differentiating individual SNP genotypes (Table S6). These genotypes represent about three times the number of tree nodes (#25) in the tree, which is attributed to the high number of terminal strain-specific SNPs. Overall SNPs were found dispersed throughout the EH41 reference chromosome; however, we noted an elevated SNP density (Fig. S3) in the previously identified region of high plasticity (Fig. 1). The ST-223 and ST-820 complex isolates are distinguished by their particular SNP pattern within this highly plastic region. The majority of genes in this area feature relatively high SNP numbers considering gene length and fulfil diverse functions, such as in metabolic pathways, substrate utilization, and stress response (Table S5), which may imply a potential role in the phylogeographical diversification of O113:H21. The majority of genic SNPs of this locus are synonymous (88.4%) that may suggest evolutionary pressure. However, a total of 52 genes in this region feature non-synonymous SNPs (Table S8). Among these are the translocation and assembly module (*tam*) for secretion of adhesins [146] and metabolic operons, such as LaAscorbate dissimilation (*ula*) [147], *ytf* and trehalose (*tre),* along with the *fim* operon. The latter four loci were previously identified as SNP hotspots in extraintestinal pathogenic *E. coli* of ST-131 [148]. Such accumulation of regionally localized SNPs may indicate mutational hotspots [109] or, alternatively, suggests a potential site of recombination. As a result of these scenarios, the phylogenetic signal can be in conflict with the signal from clonally inherited regions. Maximum parsimony (MP) provides a homoplasy metrics as indicator of accuracy and also a basis to identify potential recombination events [149, 150]. The calculated overall consistency index (CI) [114] of the tree is 0.99, and we found no evidence that SNPs in this HPR region are biassed towards homoplastic (CI=<1) or multiallelic SNPs (Table S7), which are not confined to a particular region and found scattered throughout the chromosome. Homoplasy may simply be the result of random nucleotide substitutions over time and thus may not require an evolutionary explanation. The SNP signature within the HPR allows distinguishing ST-223 from ST-820 complex isolates (Fig. 1) with the notable exception of ST-223 strains 6182–50 and 2014 C-4135. Their architecture and SNP profile in this particular region rather resembles the ST-820 complex (Table S5). In fact, when using the Achtman MLST scheme (Table S2), both strains grouped with the Australian clonal complex, a finding that is in agreement with the strain's supposed quasi-intermediate phylogenetic position (Figs 4–6). The phylogeographical separation into the two major ST complexes is also evident in the comparison of the proteome inventories by hierarchical average linkage clustering (Fig. S4).

Comprehensive analyses of Stx-status and virulence genes

The prevalence of the identified chromosomal, phage-and plasmid-borne virulence genes revealed a considerable plasticity in the strains' individual virulence complement, though we did not detect a clear-cut virulence profile boundary that would allow to distinguish ST-223 from ST-820 complex isolates (Table S8). Carriage of Stx-prophages is a virulence hallmark of STEC; the Mexican strain 6182–50 however is distinguished from all the other analysed strains by its *stx*-negative status featuring intact Stx-phage insertion sites in its draft genome. Such atypical STEC-like isolates have been described in a number of lineages. Such strains either never acquired Stx-phages or may have secondarily lost *stx* during the course of infection, isolation or routine subculture [117]. The remainder of strains carry up to three Stx-phages of different or the same *stx*-subtypes (Fig. 7) featuring suballeles stx_{1a}, stx_{2a}, stx_{2c} and/or stx_{2d} (Table S8) [130]. The major stx_2 allele was found in 33 strains, and two strains carry stx_1 alone (2013 C-3181) or in combination with stx_{2c} (TW02918). Except for strain 07–4224, for which only a draft sequence was available [75], the stx_2 suballele could be unambiguously identified either by *in silico* analysis or PCR interrogation. In 74% of the 35 strains analysed we detected the most cytotoxic subtype stx_{2a} [151];

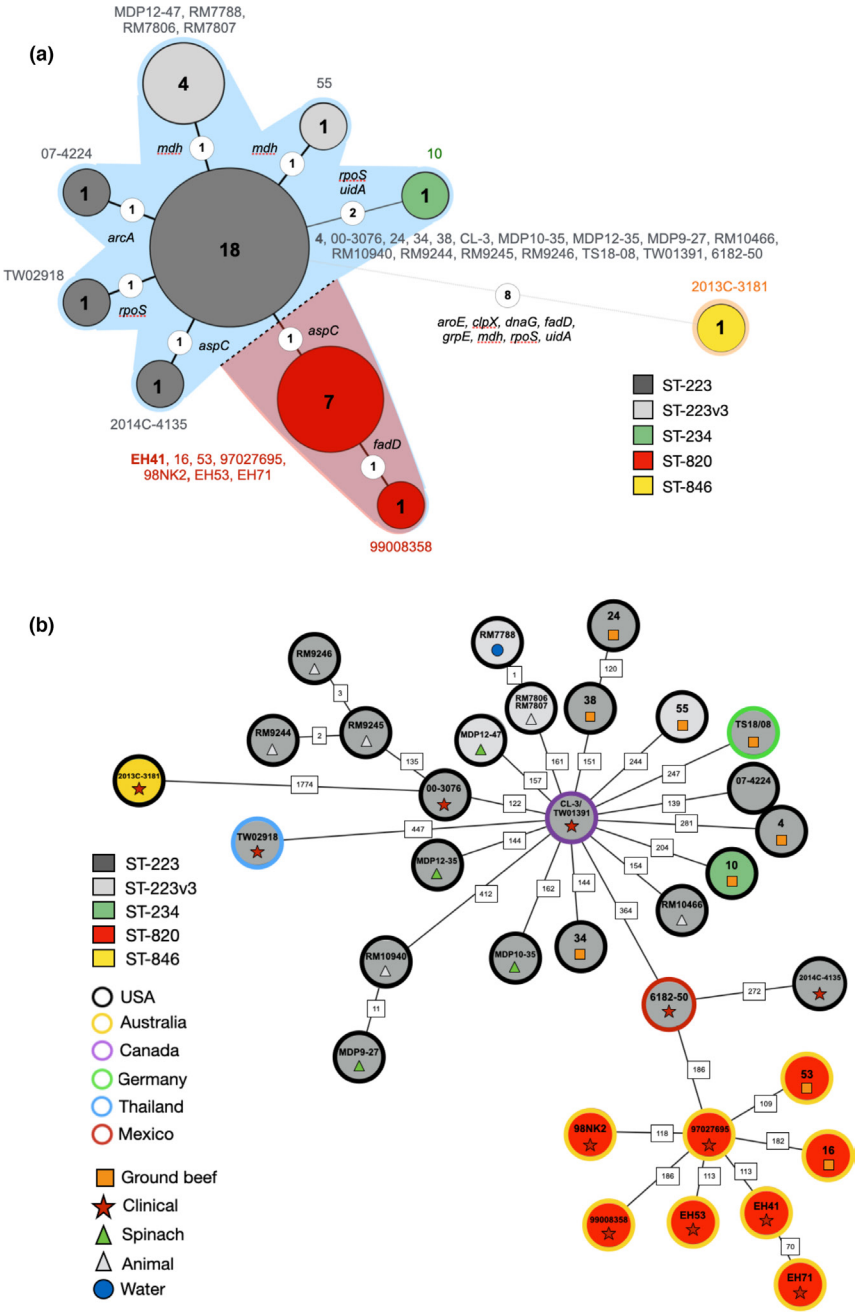


Fig. 4. Minimum Spanning Tree based on MLST analyses. (a) The MLST phylogeny for 35 O113:H21 strains based on the 15-gene Whittam schema was constructed in Ridom SeqSphere+ [89]. Circle size corresponds to the number of isolates with the same allele status (Table S2). Numbers on connecting branches show allelic differences between strains. (b) wgMLST-based phylogeny. The closed chromosome of isolate EH41 (ST-820) was provided as a seed for wgMLST analysis in Ridom SeqSphere+ [89]. The shared gene inventory of the 35 sampled O113:H21 genomes was determined at 3559 genes. The established wgMLST phylogeny clearly reflects the phylogenomic, as well as geographical separation of Australian ST-820 from ST-223 isolates found elsewhere and provides further evidence for a quasi-intermediate state of the Mexican strain 6182–50, positioned between the two major complexes. Distance values represent the number of genes with differing allele status in the network. Circle colours denote ST-classification according to the 7-gene Whittam MLST scheme [87]. Stars indicate the clonal status of strains TW01391 and CL-3, which were derived from the same repository and, as expected, are found clustered [59].

a comparison of Stx_{2a}-phages is shown in Fig. 8. This suballele is present in all eight sampled Australian ST-820 strains but absent in seven of the 35 ST-223/234/846 strains we examined. Stx_{2a} is carried either alone (13/35), or in combination with another copy of Stx_{2a} (7/35), Stx_{2c} (1/35) or Stx_{2d} (5/35). Strain RM10466 is the only isolate that carries three Stx-prophages (Stx_{2a} x2 and Stx_{2d}). We further noticed that all but one of the clinical HUS isolates are *stx_{2a}*-positive, with the exception of

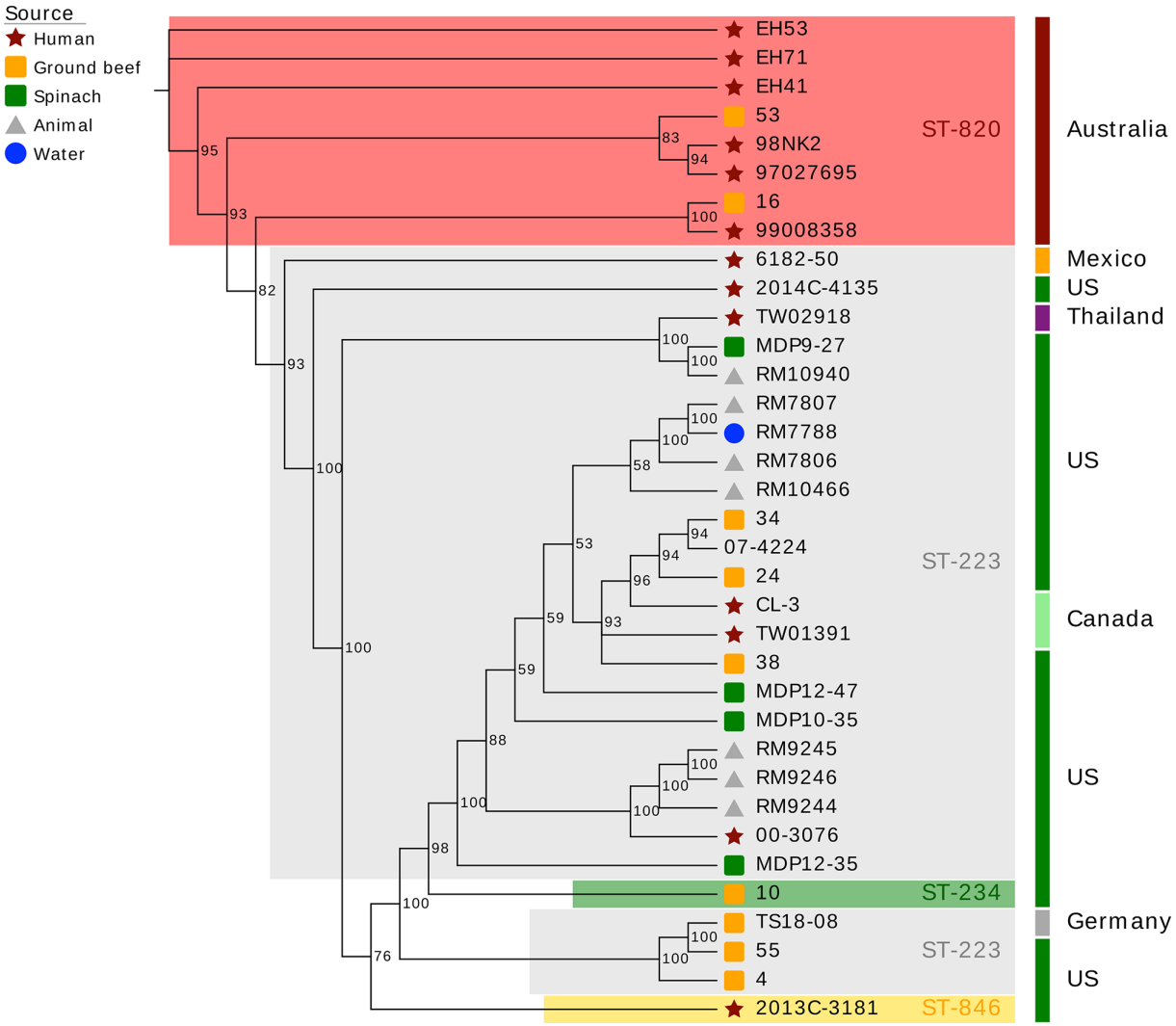


Fig. 5. Whole genome alignment phylogeny. Genomes of a total of 35 isolates, comprised of 20 strains sequenced in this study along with 15 strains obtained from NCBI GenBank, were aligned with Mugsy [90]. The phylogenetic tree was inferred using RAxML [91] with 100 bootstrap replicates and decorated with bootstrap support and other strain-associated metadata in EvolView [93, 95]. The tree topology partitions the isolates into two major phylogeographical complexes, separating Australian ST-820 from ST-223 strains found in the US and elsewhere, and further, suggests a quasi-intermediate position of the Mexican ST-223 strain 6182–50.

ST-223 strain TW01391 (*stx*$_{2c}$, *stx*$_{2d}$). In the analysed strain set, Stx-phages are inserted at six different sites, some of which are established phage target sites in STEC (Table S8): Phage Stx$_{1a}$_2013 C-3181 is inserted into *wrbA*, as described before [24, 40, 152, 153]. Four Stx$_2$-phages (Stx$_{2a}$_4, Stx$_{2a-1}$_00–3076, Stx$_{2a-2}$_RM10466, and Stx$_{2d}$_EH41) disrupt the spermidine uptake operon *potAB-CD* [154, 155]. In analogy, *potC* was previously found to be occupied by O113:H21 Stx$_{2a}$-phages [64], an O2:H25 Stx$_{2g}$-phage [3, 153], as well as other O157:H7 Stx-phages [4, 107]. Stx$_2$-phages in EH41 (Stx$_{2a}$) and 2014 C-4135 (Stx$_{2d}$) are both inserted at *dusA*. This tRNA-dihydrouridine synthase also carries a phage scar in O157:H7 strain EC4115 [107]. Other Stx$_2$-phages are located between two transporters (Stx$_{2d}$_RM10466), downstream of the HTH-type transcriptional repressor *ycgE* (Stx$_{2a-2}$_00–3076), or the BAX-inhibitor *yccA* gene (Stx$_{2a}$_RM10466), the latter associated with the Sp4-prophage in O157:H7 [4, 107, 156]. Phages Stx$_{2a}$_EH41 and Stx$_{2d}$_2014 C-4135 feature different *stx*-alleles though show overall structural similarities and are both inserted into the tRNA-dihydrouridine synthase (*dusA*). We noted that the four phages inserted at *potC* (Stx$_{2a}$_4, Stx$_{2a-2}$_RM10466, and Stx$_{2a-1}$_00–3076 and Stx$_{2d}$_EH41) all feature the same highly homologous integrase type. Taken together, our data did not show an association between Stx-phage subtypes and insertion sites [153], but may suggest that the sequence specificity of the respective integrase type determines chromosomal phage location, independent of its featured *stx*$_2$-suballele [157, 158]. Stx$_2$-phages were likely subjected to recombination events that may have led to an exchange of the *stx*-allele, in analogy to a potential antiterminator shuffling [107], while other phage characteristics

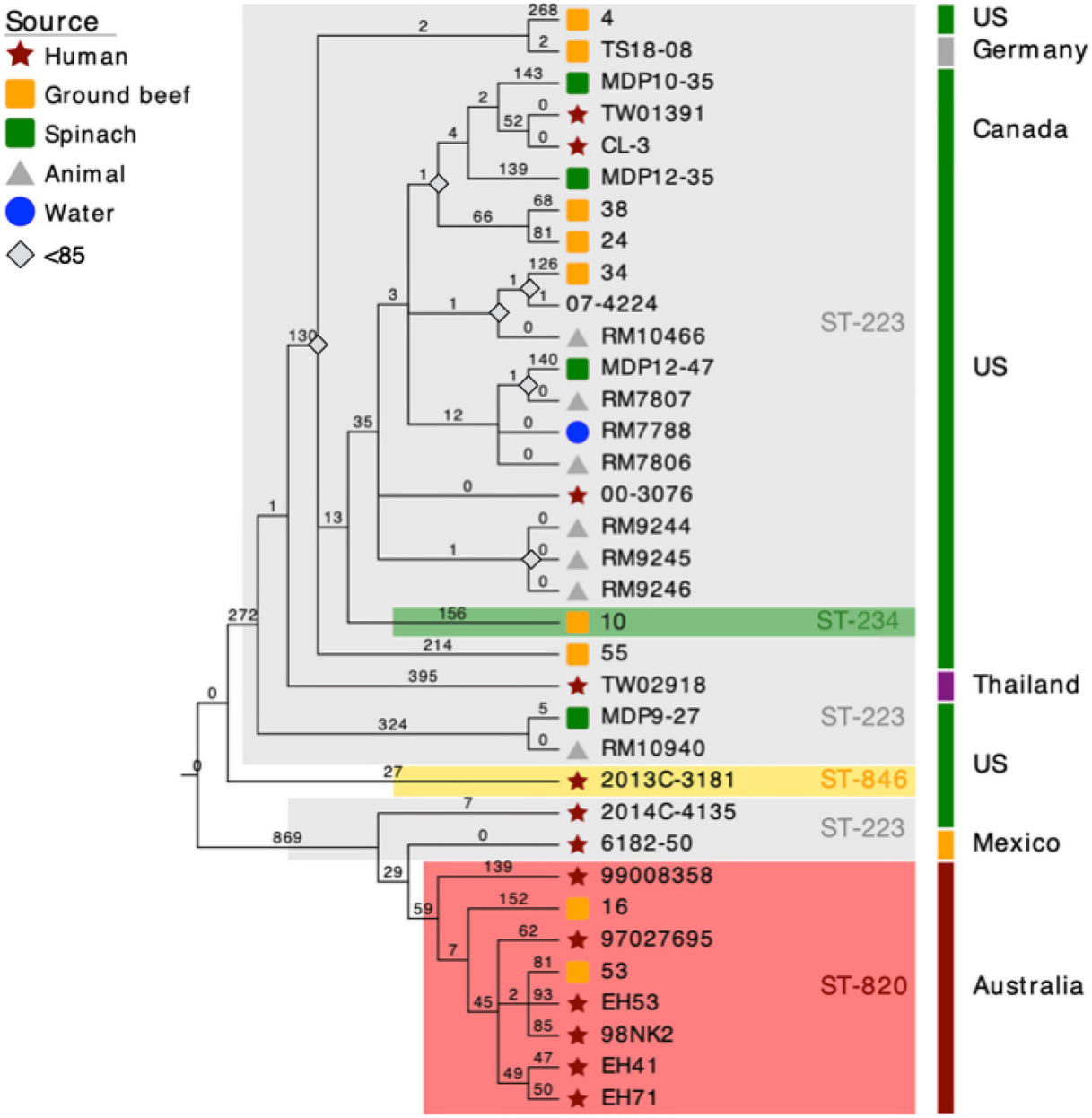


Fig. 6. Maximum parsimony cgSNP-phylogeny. Comparison of 35 genomes yielded a total of 4398 SNPs of which 1979 were parsimony informative. The tree shown is a majority-consensus tree of 3010 equally parsimonious trees with a CI of 0.99, decorated with strain-associated metadata in EvolView [93, 95]. Trees were recovered using a heuristic search in PAUP (v4.0a163) [113] with 1000 bootstrap replicates. Bootstrap supports below 85 as well as numbers of separating SNPs are shown. The tree topology mirrors the established MLST- and WGA-phylogenies, separating ST-820 from ST-223 complex isolates.

were maintained, and resulted in an overall mosaic-like phage composition, as observed in other LEE-negative STEC [153]. The prevalence of *stx* subtypes (and associated phages), does not necessarily reflect the core genome relationships (Figs 4–6), as a strain's MGEs inventory is also shaped by environmental niche, independent of Stx-phage acquisition, rather than by a common evolutionary history [159, 160].

Stx_{2a}-production by ST-820 and ST-223 complex isolates

Indeed, the most potent cytopathic toxin subtype Stx_{2a} is commonly found in the O113:H21 serotype [55]. Using Stx_2-ELISA, we assessed toxin production under both spontaneous and Stx-phage mobilizing conditions (Fig. 9), which allowed for cross-comparison of ST-820 and ST-223 Stx_{2a}-production pathotypes (Table S1). We recorded toxin production under both

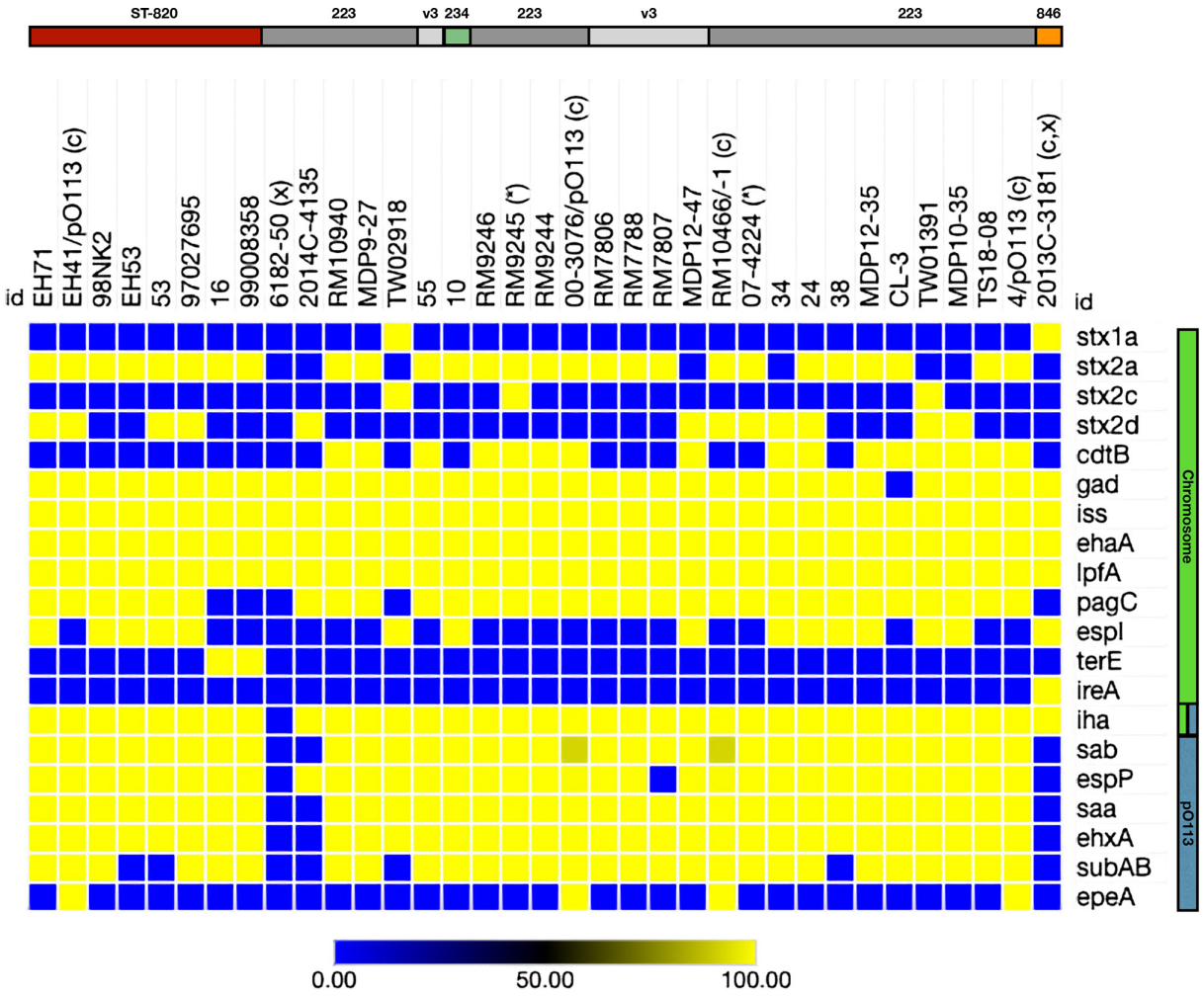


Fig. 7. Virulence gene inventory. The heatmap is based on the nucleotide percentage identity of virulence genes. Prevalence was determined by BLASTn query of these loci with a minimum 60% length requirement. Percentage identities for each gene are visualized in a heatmap showing their prevalence and distribution using Morpheus (https://software.broadinstitute.org/morpheus). The order of strains on top reflects their inferred phylogenomic position. The coloured bar on the left indicates the coding molecule. [(C)] denotes a closed genome; [(x)] pO113 not detected or sequence not available; and [(*)] fragmented *stx*-loci in the draft genomes of strains 07–4224 and RM9245 prevented determination of the suballele *in silico*. None of the 35 analysed strains carry antibiotic resistance genes [123], an observation in agreement with previous studies [209, 210]. Details on prevalence can be found in Table S8.

non-induced and Stx-phage mobilizing conditions (Fig. 9). Induction efficiency of the Stx-phages is positively correlated to Stx-production [161–166], and widely used as a means to assess an STEC strains' individual Stx-conferred pathogenic potential (Fig. 9) [24, 25, 28, 35, 39, 153, 167–172]. As expected, the Stx_2-cytotoxicity in ST-820 and ST-223 cultures was exacerbated by MMC-treatment, which triggers the RecA-mediated SOS response by causing DNA damage that leads to lytic Stx_2 phage activation and ultimately toxin production [3] (Fig. 9). The amounts of Stx in the biological replicates were reproducible and showed a strong correlation to the known strain-associated epidemiological metadata. The Stx-levels produced under non-induced conditions were indiscriminate between ST-820 and ST-223 strains, however, Stx2 litres recorded under MMC-induction were found to be elevated at statistically significant levels in ST-820 complex isolates, and thus may increase pathogenicity. Strain EH53 is an outlier amongst ST-820 strains showing low Stx-production under both induced and non-induced conditions (Fig. 9). The integrity of the Stx_{2a}-phage seems unaltered, though its fragmented status on three contigs prevents a more detailed analysis (data not shown) (Table S3). Within ST-223, we identified strains MDP9-27, 10, and TW02918, the latter isolated from a patient with diarrhoea, with significantly increased Stx-levels comparable to those recorded for ST-820 (Fig. 9). Strains with these genotypes could be potentially pathogenic to humans, which raises concerns should such isolates enter the human food chain. The actual disease outcome, however, cannot be predicted. Stx-production is a definite factor in causing disease, though disease severity is a result of complex interactions between STEC, the host microbiota [8, 50, 52, 173–178], and the age and immunogenetics of the infected patient, with higher incidence of HUS

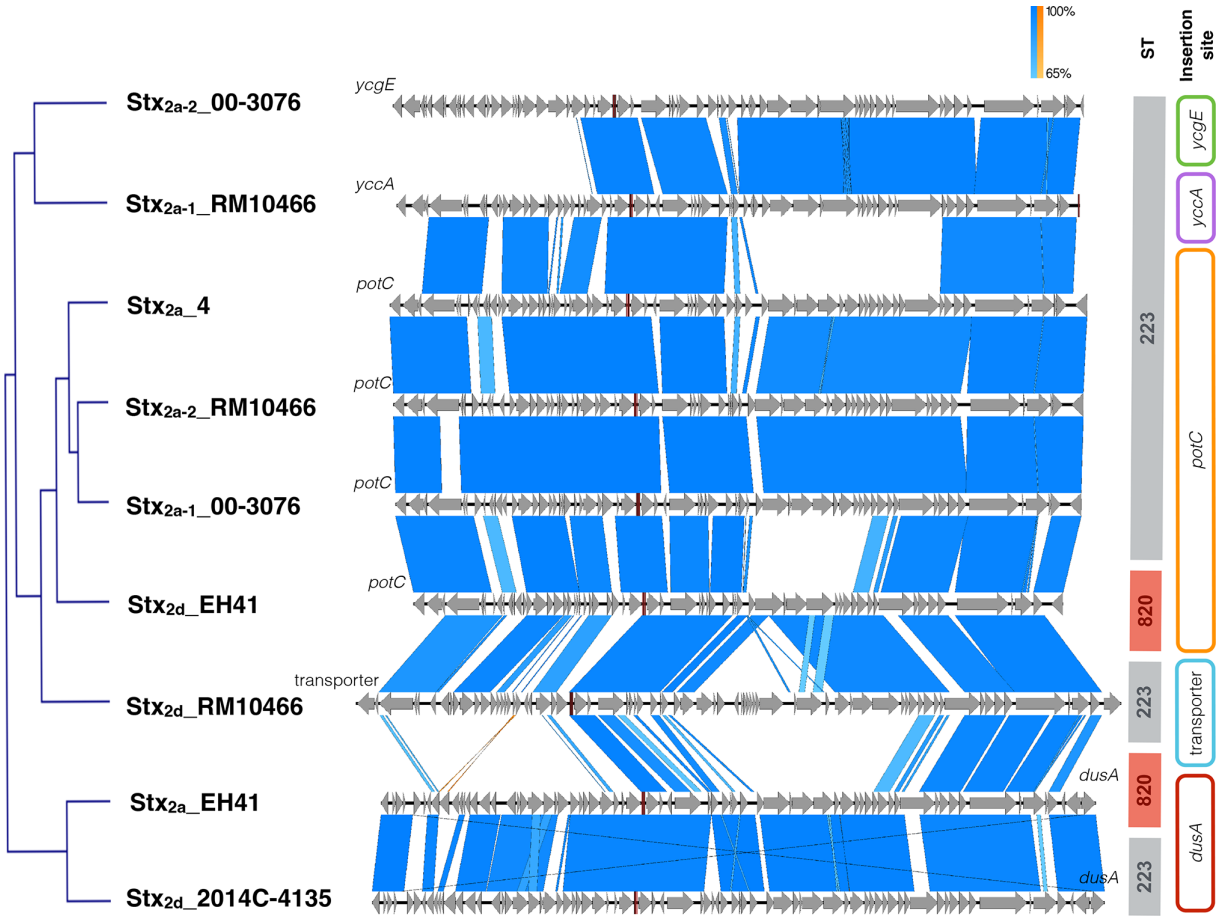


Fig. 8. Stx_2-prophage inventory. Stx_2-phage architectures were compared by BLASTn and visualized in Easyfig [127]. Phage CDS are depicted as grey arrows. Blue homologous blocks indicate unidirectional sequence similarity. The *stx*-suballele along with the respective chromosomal insertion and MLST of the bacterial host strain is shown.

typically in the young and elderly [179–181]. All these factors may independently or in combination impact Stx-production and ultimately the outcomes of severe pathogenesis in humans [182].

DISCUSSION

This study serves as basis to identify genomic signatures associated with hypervirulent Stx-producing, HUS-causing subpopulations in STEC O113:H21. Integrating genomic and virulence information following genome-wide association studies (GWAS) principles [183–185] opens the avenue for improved typing, biosurveillance and risk assessment and may allow the identification of circulating hypervirulent STEC subpopulations [19, 32, 42, 186]. Insights into the pathogenome make-up and associated virulence traits is foundational to understand the evolutionary trajectories of STEC lineages [187–189], which in the case of O113:H21, resulted in the phylogeographical diversification in the ST-820 and ST-223 complexes. The established phylogenomic framework allowed us to refine the genome and virulence boundaries between these two major complexes and identified a quasi-intermediate position for the historic non-shigatoxigenic ST-223 strain 6182–50 from Mexico. Since plasmids may get lost during laboratory cultivation or often recovered only in fragments, typing efforts are mostly focused on stable chromosomal markers. However, our data suggest that plasmid information may be used to further refine the phylogenetic model. Plasmid comparison revealed a correlation between the strains' pO113 plasmid genotype and chromosomally inferred sequence type, which suggests the coevolution of the chromosome and accessory plasmids. In accordance with the epidemiological metadata, we further describe significant differences in Stx_{2a}-production capabilities of ST-223 and ST-820 complex strains. The latter may possess an increased pathogenic potential (Fig. 9). This is particularly evident under MMC-induction, a condition that mimics the Stx-phage induction [190–195]. In this context, we note that the definition of pathogenic potential is often skewed by anthropogenic biases, such as fitness factors that are not, per se, accounted for in the virulence inventory and may allow pathogens to access the production foods for human consumption

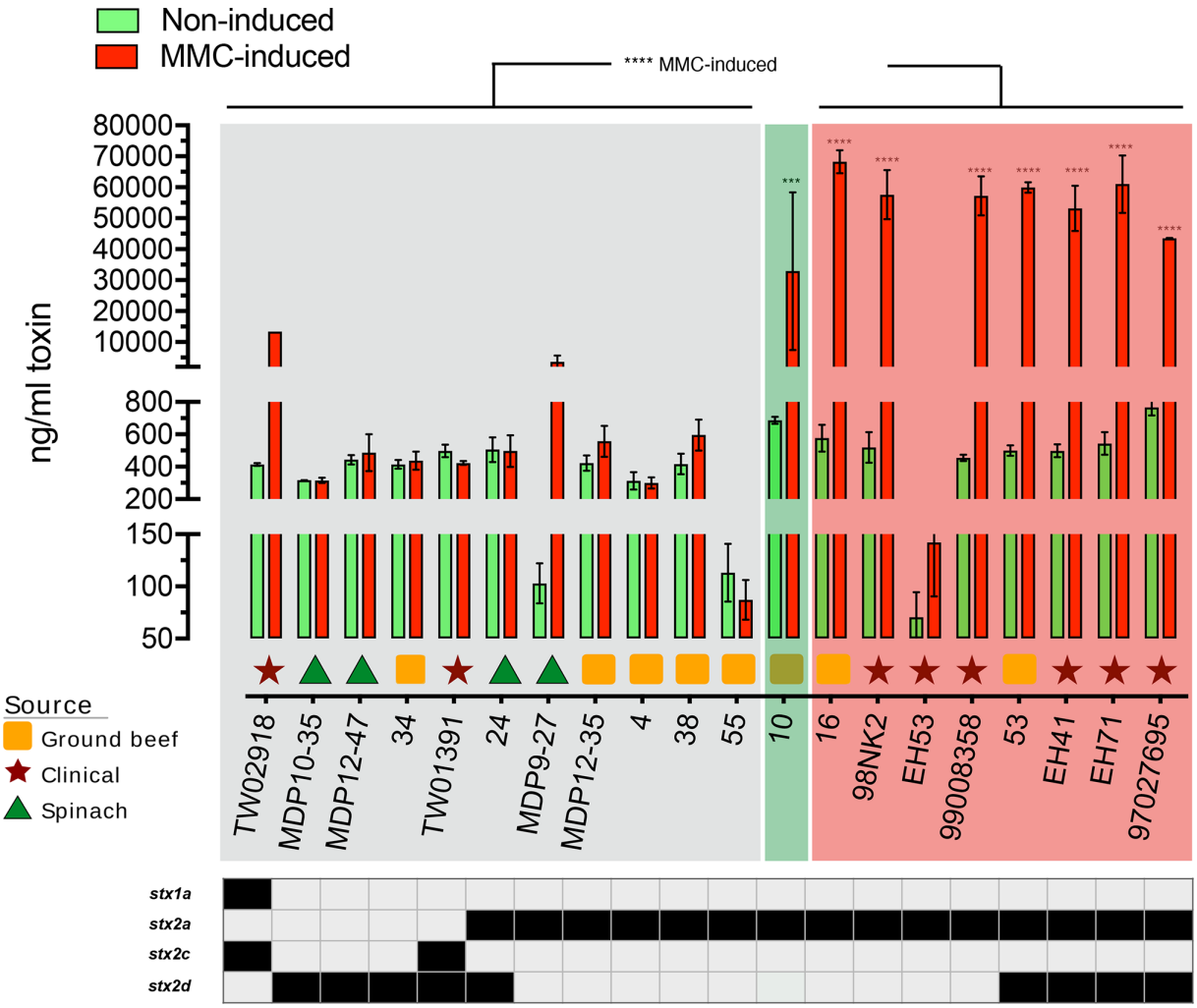


Fig. 9. Variability in Stx-production. Stx_2-litres of ST-820 and ST-223 cultures were determined by ELISA under non-induced and MMC-induced conditions. A two-way ANOVA with Sidak's multiple comparisons test was used to compare non-induced vs. MMC-induced conditions for each O113:H21 strain. A one-way ANOVA with Tukey's multiple comparisons test was used to compare ST groups for both non-induced and MMC-induced conditions. Statistical significance for $N=2$ experiments is reported as $^{*}P<0.05$; $^{**}P<0.005$; $^{***}P<0.0005$; $^{****}P<0.00005$. The Stx levels produced under non-induced conditions by strains from both clonal groups were indiscriminate; however, Stx2 litres recorded under MMC-induced conditions were found to be elevated at statistically significant levels in ST-820 complex isolates, suggesting increased pathogenicity.

[196, 197]. Our lineage-scale study did not identify a distinct virulence gene profile associated with either ST. However, the noted complex-specific trends in the *in vitro* Stx_{2a}-production pathotypes may suggest alterations in the underlying dynamic regulatory networks [198–200], which is the focus of our ongoing research. The O113:H21 lineage is not one of the so called 'Big Six' non-O157 serogroups of high-risk concern [201], and at present, O113:H21 strains have not caused major outbreaks in the US. Still, some of the tested ST-223 strains from contaminated produce or bovine reservoir, showed Stx-production levels comparable to ST-820 isolates and are likely capable of causing disease in humans. Taken together, our findings call for increased awareness and continued surveillance of this serotype. Clearly, further research is required to elucidate how the individual O113:H21 genotype relates to human disease by taking into account pathogenicity traits that were not examined, such as biofilm formation, adherence and invasiveness [66, 198, 199, 202–206], production of non-Stx toxins [207, 208] and other putative STEC virulence factors.

Funding information
Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Numbers [SC1GM135110] and the US Department of Homeland Security [2014-ST-062–000058] to ME. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the US Department of Homeland Security. The contributions of JMB and PF were part of their official duties at the Agricultural Research Service of the U.S. Department of Agriculture (ARS-USDA)

and the U.S. Food and Drug Administration (FDA), respectively. The mention of a trade name, proprietary product, or specific equipment does not constitute a guarantee or warranty by the USDA and does not imply approval to the exclusion of other products that might be suitable. The USDA is an equal opportunity employer and provider.

Acknowledgements
This work received support by the South Texas Centre of Emerging Infectious Diseases (STCEID), the Department of Biology and the High-Performance Computing Cluster operated by University Technology Solutions (UTS) at the University of Texas at San Antonio.

Author contributions
Conceived and designed the study: M.E. Analysed the data: A.A.G., S.S.K., R.A.M., J.M.B., P.F. and M.E. Chromosome and plasmid comparisons: M.E. and R.A.M. Contributed strain material: J.M.B. and P.F. Provided computational support: A.L.R. Wrote the manuscript: M.E. and S.S.K. All authors read and approved the manuscript.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Sperandio V, Hovde CJ. *Enterohemorrhagic Escherichia coli and other Shiga toxin-producing E. coli*. Washington, DC: ASM Press; 2015.
2. Zuppi M, Tozzoli R, Chiani P, Quiros P, Martinez-Velazquez A, *et al*. Investigation on the evolution of shiga toxin-converting phages based on whole genome sequencing. *Front Microbiol* 2020;11:1472.
3. Kruger A, Lucchesi PM. Shiga toxins and stx phages: highly diverse entities. *Microbiology* 2015;161:451–462.
4. Asadulghani M, Ogura Y, Ooka T, Itoh T, Sawaguchi A, *et al*. The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *PLoS Pathog* 2009;5:e1000408.
5. Huang A, Friesen J, Brunton JL. Characterization of a bacteriophage that carries the genes for production of Shiga-like toxin 1 in *Escherichia coli*. *J Bacteriol* 1987;169:4308–4312.
6. Mauro SA, Koudelka GB. Shiga toxin: expression, distribution, and its role in the environment. *Toxins (Basel)* 2011;3:608–625.
7. Melton-Celsa A, Mohawk K, Teel L, O'Brien A. Pathogenesis of shiga-toxin producing *Escherichia coli*. *Curr Top Microbiol Immunol* 2012;357:67–103.
8. Nguyen Y, Sperandio V. Enterohemorrhagic E. coli (EHEC) pathogenesis. *Front Cell Infect Microbiol* 2012;2:90.
9. Johannes L, Romer W. Shiga toxins-from cell biology to biomedical applications. *Nat Rev Microbiol* 2010;8:105–116.
10. Tran SL, Billoud L, Lewis SB, Phillips AD, Schuller S. Shiga toxin production and translocation during microaerobic human colonic infection with shiga toxin-producing *E. coli* O157:H7 and O104:H4. *Cell Microbiol* 2014;16:1255–1266.
11. Eaton KA, Friedman DI, Francis GJ, Tyler JS, Young VB, *et al*. Pathogenesis of renal disease due to enterohemorrhagic *Escherichia coli* in germ-free mice. *Infect Immun* 2008;76:3054–3063.
12. Shin I-S, Ishii S, Shin J-S, Sung K-I, Park B-S, *et al*. Globotriaosylceramide (Gb3) content in HeLa cells is correlated to Shiga toxin-induced cytotoxicity and Gb3 synthase expression. *BMB Rep* 2009;42:310–314.
13. Pacheco AR, Lazarus JE, Sit B, Schmieder S, Lencer WI, *et al*. CRISPR Screen Reveals that EHEC's T3SS and shiga toxin rely on shared host factors for infection. *mBio* 2018;9:e01003-18.
14. Karmali MA, Petric M, Lim C, Fleming PC, Steele BT. *Escherichia coli* cytotoxin, haemolytic-uraemic syndrome, and haemorrhagic colitis. *Lancet* 1983;2:1299–1300.
15. Donohue-Rolfe A, Kondova I, Oswald S, Hutto D, Tzipori S. *Escherichia coli* O157:H7 strains that express Shiga toxin (Stx) 2 alone are more neurotropic for gnotobiotic piglets than are isotypes producing only Stx1 or both Stx1 and Stx2. *J Infect Dis* 2000;181:1825–1829.
16. Russo LM, Melton-Celsa AR, O'Brien AD. Shiga toxin (Stx) type 1a reduces the oral toxicity of Stx type 2a. *J Infect Dis* 2016;213:1271–1279.
17. Amigo N, Mercado E, Bentancor A, Singh P, Vilte D, *et al*. Clade 8 and Clade 6 strains of *Escherichia coli* O157:H7 from cattle in argentina have hypervirulent-like phenotypes. *PLoS One* 2015;10:e0127710.
18. Iyoda S, Manning SD, Seto K, Kimata K, Isobe J, *et al*. Phylogenetic clades 6 and 8 of enterohemorrhagic *Escherichia coli* O157:H7 with particular stx subtypes are more frequently found in isolates from hemolytic uremic syndrome patients than from asymptomatic carriers. *Open Forum Infect Dis* 2014;1:fu061.
19. Manning SD, Motiwala AS, Springman AC, Qi W, Lacher DW, *et al*. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *Proc Natl Acad Sci U S A* 2008;105:4868–4873.
20. Vanaja SK, Springman AC, Besser TE, Whittam TS, Manning SD. Differential expression of virulence and stress fitness genes between *Escherichia coli* O157:H7 strains with clinical or bovine-biased genotypes. *Appl Environ Microbiol* 2010;76:60–68.
21. Mellor GE, Sim EM, Barlow RS, D'Astek BA, Galli L, *et al*. Phylogenetically related Argentinean and Australian *Escherichia coli* O157 isolates are distinguished by virulence clades and alternative Shiga toxin 1 and 2 prophages. *Appl Environ Microbiol* 2012;78:4724–4731.
22. Strachan NJ, Rotariu O, Lopes B, MacRae M, Fairley S, *et al*. Whole genome sequencing demonstrates that geographic variation of *Escherichia coli* O157 genotypes dominates host association. *Sci Rep* 2015;5:14145.
23. Pearce MC, Chase-Topping ME, McKendrick IJ, Mellor DJ, Locking ME, *et al*. Temporal and spatial patterns of bovine *Escherichia coli* O157 prevalence and comparison of temporal changes in the patterns of phage types associated with bovine shedding and human *E. coli* O157 cases in Scotland between 1998-2000 and 2002-2004. *BMC Microbiol* 2009;9:276.
24. Ogura Y, Mondal SI, Islam MR, Mako T, Arisawa K, *et al*. The Shiga toxin 2 production level in enterohemorrhagic *Escherichia coli* O157:H7 is correlated with the subtypes of toxin-encoding phage. *Sci Rep* 2015;5.
25. Neupane M, Abu-Ali GS, Mitra A, Lacher DW, Manning SD, *et al*. Shiga toxin 2 overexpression in *Escherichia coli* O157:H7 strains associated with severe human disease. *Microb Pathog* 2011;51:466–470.
26. Wagner PL, Acheson DW, Waldor MK. Isogenic lysogens of diverse shiga toxin 2-encoding bacteriophages produce markedly different amounts of shiga toxin. *Infect Immun* 1999;67:6710–6714.
27. de Sablet T, Bertin Y, Vareille M, Girardeau J-P, Garrivier A, *et al*. Differential expression of stx2 variants in Shiga toxin-producing *Escherichia coli* belonging to seropathotypes A and C. *Microbiology (Reading)* 2008;154:176–186.
28. Kulasekara BR, Jacobs M, Zhou Y, Wu Z, Sims E, *et al*. Analysis of the genome of the *Escherichia coli* O157:H7 2006 spinach-associated outbreak isolate indicates candidate genes that may enhance virulence. *Infect Immun* 2009;77:3713–3721.
29. Baker DR, Moxley RA, Steele MB, Lejeune JT, Christopher-Hennings J, *et al*. Differences in virulence among *Escherichia coli* O157:H7 strains isolated from humans during disease

outbreaks and from healthy cattle. *Appl Environ Microbiol* 2007;73:7338–7346.

30. Zhang Y, Laing C, Zhang Z, Hallewell J, You C, *et al*. Lineage and host source are both correlated with levels of Shiga toxin 2 production by *Escherichia coli* O157:H7 strains. *Appl Environ Microbiol* 2010;76:474–482.
31. Koitabashi T, Vuddhakul V, Radu S, Morigaki T, Asai N, *et al*. Genetic characterization of *Escherichia coli* O157: H7/- strains carrying the stx_2 gene but not producing Shiga toxin 2. *Microbiol Immunol* 2006;50:135–148.
32. Dowd SE, Williams JB. Comparison of Shiga-like toxin II expression between two genetically diverse lineages of *Escherichia coli* O157:H7. *J Food Prot* 2008;71:1673–1678.
33. Baker DR, Moxley RA, Francis DH. Variation in virulence in the gnotobiotic pig model of O157:H7 *Escherichia coli* strains of bovine and human origin. *Adv Exp Med Biol* 1997;412:53–58.
34. Ziebell K, Steele M, Zhang Y, Benson A, Taboada EN, *et al*. Genotypic characterization and prevalence of virulence factors among Canadian *Escherichia coli* O157:H7 strains. *Appl Environ Microbiol* 2008;74:4314–4323.
35. Abu-Ali GS, Ouellette LM, Henderson ST, Whittam TS, Manning SD. Differences in adherence and virulence gene expression between two outbreak strains of enterohaemorrhagic *Escherichia coli* O157: H7. *Microbiology* 2010;156:408–419.
36. Zhou Z, Li X, Liu B, Beutin L, Xu J, *et al*. Derivation of *Escherichia coli* O157:H7 from its O55:H7 precursor. *PLoS One* 2010;5:e8700.
37. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, *et al*. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med* 2011;365:709–717.
38. Yang Z, Kovar J, Kim J, Nietfeldt J, Smith DR, *et al*. Identification of common subpopulations of non-sorbitol-fermenting, beta-glucuronidase-negative *Escherichia coli* O157:H7 from bovine production environments and human clinical samples. *Appl Environ Microbiol* 2004;70:6846–6854.
39. Zhang X, McDaniel AD, Wolf LE, Keusch GT, Waldor MK, *et al*. Quinolone antibiotics induce Shiga toxin-encoding bacteriophages, toxin production, and death in mice. *J Infect Dis* 2000;181:664–670.
40. Yin S, Rusconi B, Sanjar F, Goswami K, Xiaoli L, *et al*. *Escherichia coli* O157:H7 strains harbor at least three distinct sequence types of Shiga toxin 2a-converting phages. *BMC Genomics* 2015;16:733.
41. Besser TE, Shaikh N, Holt NJ, Tarr PI, Konkel ME, *et al*. Greater diversity of Shiga toxin-encoding bacteriophage insertion sites among *Escherichia coli* O157:H7 isolates from cattle than in those from humans. *Appl Environ Microbiol* 2007;73:671–679.
42. Elhadidy M, Elkhatib WF, Piérard D, De Reu K, Heyndrickx M. Model-based clustering of *Escherichia coli* O157:H7 genotypes and their potential association with clinical outcome in human infections. *Diagn Microbiol Infect Dis* 2015;83:198–202.
43. Whitworth J, Zhang Y, Bono J, Pleydell E, French N, *et al*. Diverse genetic markers concordantly identify bovine origin *Escherichia coli* O157 genotypes underrepresented in human disease. *Appl Environ Microbiol* 2010;76:361–365.
44. Whitworth JH, Fegan N, Keller J, Gobius KS, Bono JL, *et al*. International comparison of clinical, bovine, and environmental *Escherichia coli* O157 isolates on the basis of Shiga toxin-encoding bacteriophage insertion site genotypes. *Appl Environ Microbiol* 2008;74:7447–7450.
45. Shringi S, Schmidt C, Katherine K, Brayton KA, Hancock DD, *et al*. Carriage of stx2a differentiates clinical and bovine-biased strains of *Escherichia coli* O157. *PLoS One* 2012;7:e51572.
46. Shringi S, García A, Lahmers KK, Potter KA, Muthupalani S, *et al*. Differential virulence of clinical and bovine-biased enterohemorrhagic *Escherichia coli* O157:H7 genotypes in piglet and Dutch belted rabbit models. *Infect Immun* 2012;80:369–380.
47. Ahmad A, Zurek L. Evaluation of the anti-terminator Q933 gene as a marker for *Escherichia coli* O157:H7 with high Shiga toxin production. *Curr Microbiol* 2006;53:324–328.
48. Elhadidy M, Alvarez-Ordonez A. Diversity of survival patterns among *Escherichia coli* O157:H7 genotypes subjected to food-related stress conditions. *Front Microbiol* 2016;7:322.
49. Byrne L, Dallman TJ, Adams N, Mikhail AFW, McCarthy N, *et al*. Highly pathogenic clone of shiga toxin-producing *Escherichia coli* O157:H7, England and Wales. *Emerg Infect Dis* 2018;24:2303–2308.
50. Pruimboom-Brees IM, Morgan TW, Ackermann MR, Nystrom ED, Samuel JE, *et al*. Cattle lack vascular receptors for *Escherichia coli* O157:H7 Shiga toxins. *Proc Natl Acad Sci U S A* 2000;97:10325–10329.
51. Flockhart AF, Tree JJ, Xu X, Karpiyevich M, McAteer SP, *et al*. Identification of a novel prophage regulator in *Escherichia coli* controlling the expression of type III secretion. *Mol Microbiol* 2012;83:208–223.
52. Goswami K, Chen C, Xiaoli L, Eaton KA, Dudley EG. Coculture of *Escherichia coli* O157:H7 with a Nonpathogenic *E. coli* strain increases toxin production and virulence in a germfree mouse model. *Infect Immun* 2015;83:4185–4193.
53. Pacheco AR, Curtis MM, Ritchie JM, Munera D, Waldor MK, *et al*. Fucose sensing regulates bacterial intestinal colonization. *Nature* 2012;492:113–117.
54. Pacheco AR, Sperandio V. Shiga toxin in enterohemorrhagic *E.coli*: regulation and novel anti-virulence strategies. *Front Cell Infect Microbiol* 2012;2:81.
55. Silva CJ, Lee BG, Yambao JC, Erickson-Beltran ML, Quiñones B. Using nanospray liquid chromatography and mass spectrometry to quantitate shiga toxin production in environmental *Escherichia coli* recovered from a major produce production region in California. *J Agric Food Chem* 2019;67:1554–1562.
56. González-Escalona N, Kase JA. Virulence gene profiles and phylogeny of Shiga toxin-positive *Escherichia coli* strains isolated from FDA regulated foods during 2010-2017. *PLoS One* 2019;14:e0214620.
57. Paton AW, Woodrow MC, Doyle RM, Lanser JA, Paton JC. Molecular characterization of a Shiga toxigenic *Escherichia coli* O113:H21 strain lacking eae responsible for a cluster of cases of hemolytic-uremic syndrome. *J Clin Microbiol* 1999;37:3357–3361.
58. Bondì R, Chiani P, Michelacci V, Minelli F, Caprioli A, *et al*. The gene *tia*, harbored by the subtilase-encoding pathogenicity Island, is involved in the ability of locus of enterocyte effacement-negative shiga toxin-producing *Escherichia coli* strains to invade monolayers of epithelial cells. *Infect Immun* 2017;85:e00613-17.
59. Feng P, Delannoy S, Lacher DW, Bosilevac JM, Fach P. Characterization and virulence potential of serogroup O113 shiga toxin-producing *Escherichia coli* strains isolated from beef and cattle in the United States. *J Food Prot* 2017;80:383–391.
60. Feng PC, Delannoy S, Lacher DW, Dos Santos LF, Beutin L, *et al*. Genetic diversity and virulence potential of shiga toxin-producing *Escherichia coli* O113:H21 strains isolated from clinical, environmental, and food sources. *Appl Environ Microbiol* 2014;80:4757–4763.
61. Paton AW, Paton JC. Molecular characterization of the locus encoding biosynthesis of the lipopolysaccharide O antigen of *Escherichia coli* serotype O113. *Infect Immun* 1999;67:5930–5937.
62. Goldwater PN, Giles N, Bettelheim KA. An unusual case of micro-angiopathic haemolytic anaemia associated with enterohaemorrhagic *Escherichia coli* O113:H21 infection, a verocytotoxin-2/shiga toxin-2 producing serotype. *J Infect* 1998;37:302–304.
63. Newton HJ, Sloan J, Bulach DM, Seemann T, Allison CC, *et al*. Shiga toxin-producing *Escherichia coli* strains negative for locus of enterocyte effacement. *Emerg Infect Dis* 2009;15:372–380.
64. Quiñones B, Yambao JC, Lee BG. Draft genome sequences of *Escherichia coli* O113:H21 strains recovered from a major produce production region in California. *Genome Announc* 2017;5:e01203-17.
65. G M Gonzalez A, M F Cerqueira A. Shiga toxin-producing *Escherichia coli* in the animal reservoir and food in Brazil. *J Appl Microbiol* 2020;128:1568–1582.

66. Dos Santos LF, Biscola FT, Goncalves EM, Guth BE. Biofilm formation, invasiveness and colicinogeny in locus of enterocyte and effacement negative O113:H21 Shigatoxigenic *Escherichia coli*. *J Appl Microbiol* 2017;122:1101–1109.
67. dos Santos LF, Irino K, Vaz TM, Guth BE. Set of virulence genes and genetic relatedness of O113: H21 *Escherichia coli* strains isolated from the animal reservoir and human infections in Brazil. *J Med Microbiol* 2010;59:634–640.
68. Bosilevac JM, Koohmaraie M. Prevalence and characterization of non-O157 shiga toxin-producing *Escherichia coli* isolates from commercial ground beef in the United States. *Appl Environ Microbiol* 2011;77:2103–2112.
69. Irino K, Kato MAMF, Vaz TMI, Ramos II, Souza MAC, *et al*. Serotypes and virulence markers of Shiga toxin-producing *Escherichia coli* (STEC) isolated from dairy cattle in São Paulo State, Brazil. *Vet Microbiol* 2005;105:29–36.
70. Galli L, Miliwebsky E, Irino K, Leotta G, Rivas M. Virulence profile comparison between LEE-negative Shiga toxin-producing *Escherichia coli* (STEC) strains isolated from cattle and humans. *Vet Microbiol* 2010;143:307–313.
71. Sadiq M, Hazen T, Rasko D, Eppinger M. Enterohemorrhagic *Escherichia coli* genomics: past, present, and future. In: SperandioV and Hovde CJ (eds). *Enterohemorrhagic Escherichia Coli and Other Shiga Toxin Producing E Coli*. Washington, DC: ASM Press; 2015. pp. 55–71.
72. Franz E, Delaquis P, Morabito S, Beutin L, Gobius K, *et al*. Exploiting the explosion of information associated with whole genome sequencing to tackle Shiga toxin-producing *Escherichia coli* (STEC) in global food production systems. *Int J Food Microbiol* 2014;187:57–72.
73. Bosilevac JM, Guerini MN, Brichta-Harhay DM, Arthur TM, Koohmaraie M. Microbiological characterization of imported and domestic boneless beef trim used for ground beef. *J Food Prot* 2007;70:440–449.
74. Parker CT, Cooper KK, Huynh S, Smith TP, Bono JL, *et al*. Genome sequences of eight shiga toxin-producing *Escherichia coli* strains isolated from a produce-growing region in California. *Microbiol Resour Announc* 2018;7:e00807-18.
75. Trees E, Strockbine N, Changayil S, Ranganathan S, Zhao K, *et al*. Genome sequences of 228 shiga toxin-producing *Escherichia coli* Isolates and 12 isolates representing other diarrheagenic *E. coli* pathotypes. *Genome Announc* 2014;2:e00718-14.
76. Patel PN, Lindsey RL, Garcia-Toledo L, Rowe LA, Batra D, *et al*. High-quality whole-genome sequences for 77 shiga toxin-producing *Escherichia coli* strains generated with PacBio sequencing. *Genome Announc* 2018;6:e00391-18.
77. Orskov I, Orskov F, Jann B. Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriol Rev* 1977;41:667–710.
78. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
79. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 2018;34:i142–i150.
80. Gao F, Zhang CT. Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. *BMC Bioinformatics* 2008;9:79.
81. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, *et al*. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016;44:6614–6624.
82. Foley SL, Lynne AM, Nayak R. Molecular typing methodologies for microbial source tracking and epidemiological investigations of Gram-negative bacterial foodborne pathogens. *Infect Genet Evol* 2009;9:430–440.
83. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Agama Study G, *et al*. The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny and *Escherichia* core genomic diversity. *Genome Res* 2019.
84. Wirth T, Falush D, Lan R, Colles F, Mensa P, *et al*. Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* 2006;60:1136–1151.
85. Jaureguy F, Landraud L, Passet V, Diancourt L, Frapy E, *et al*. Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics* 2008;9:560.
86. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, *et al*. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 2012;50:1355–1361.
87. Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS. Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* 2000;406:64–67.
88. Weihong Qi DWL. *EcMLST: an Online Database for Multi Locus Sequence Typing of Pathogenic Escherichia coli. Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference (CSB 2004)*. Stanford, CA: IEEE; 2004.
89. Junemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, *et al*. Updating benchtop sequencing performance comparison. *Nat Biotechnol* 2013;31:294–296.
90. Angiuoli SV, Salzberg SL. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* 2011;27:334–342.
91. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
92. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, *et al*. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012;28:1647–1649.
93. Zhang H, Gao S, Lercher MJ, Hu S. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res* 2012;40:W569-72.
94. Subramanian B, Gao S, Lercher MJ, Hu S, Chen WH. Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res* 2019.
95. He Z, Zhang H, Gao S, Lercher MJ, Chen WH, *et al*. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res* 2016;44:W236-241.
96. Rusconi B, Sanjar F, Koenig SS, Mammel MK, Tarr PI, *et al*. Whole genome sequencing for genomics-guided investigations of *Escherichia coli* O157:H7 outbreaks. *Front Microbiol* 2016;7:985.
97. Goecks J, Nekrutenko A, Taylor J, Galaxy T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 2010;11:R86.
98. Delcher AL, Salzberg SL, Phillippy AM. Using mummer to identify similar regions in large sequence sets. *Curr Protoc Bioinformatics* 2003;10:13.
99. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. *Nucleic Acids Res* 2011;39:W347-352.
100. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, *et al*. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;44:W16-21.
101. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 2006;34:D32-36.
102. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 2017;33:3340–3347.
103. Liu M, Li X, Xie Y, Bi D, Sun J, *et al*. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res* 2019;47:D660–D665.
104. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–1760.
105. Garrison E, Marth G. (n.d.) Haplotype-based variant detection from short-read sequencing. *ArXiv e-prints*

106. Marcais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, *et al*. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput Biol* 2018;14:e1005944.
107. Eppinger M, Mammel MK, Leclerc JE, Ravel J, Cebula TA. Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. *Proc Natl Acad Sci U S A* 2011;108:20142–20147.
108. Eppinger M, Pearson T, Koenig SS, Pearson O, Hicks N, *et al*. Genomic epidemiology of the haitian cholera outbreak: a single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. *mBio* 2014;5:e01721.
109. Myers GS, Mathews SA, Eppinger M, Mitchell C, O'Brien KK, *et al*. Evidence that human *Chlamydia pneumoniae* was zoonotically acquired. *J Bacteriol* 2009;191:7225–7233.
110. Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, *et al*. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet* 2010;42:1140–1143.
111. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
112. Johnson JE, Kumar P, Easterly C, Esler M, Mehta S, *et al*. Improve your Galaxy text life: The Query Tabular Tool. *F1000Res* 2018;7:1604.
113. Wilgenbusch JC, Swofford D. Inferring evolutionary trees with PAUP*. *Curr Protoc Bioinformatics* 2003;6:4.
114. Maddison WP, Maddison DR. *Mesquite: A Modular System for Evolutionary Analysis*. 2016.
115. Eppinger M, Worsham PL, Nikolich MP, Riley DR, Sebastian Y, *et al*. Genome sequence of the deep-rooted Yersinia pestis strain Angola reveals new insights into the evolution and pangenome of the plague bacterium. *J Bacteriol* 2010;192:1685–1699.
116. Hau SJ, Allue-Guardia A, Rusconi B, Haan JS, Davies PR, *et al*. Single nucleotide polymorphism analysis indicates genetic distinction and reduced diversity of swine-associated methicillin resistant *Staphylococcus aureus* (MRSA) ST5 isolates compared to clinical MRSA ST5 isolates. *Front Microbiol* 2078;2018:9.
117. Nyong EC, Zaia SR, Allué-Guardia A, Rodriguez AL, Irion-Byrd Z, *et al*. Pathogenomes of atypical non-shigatoxigenic *Escherichia coli* NSF/SF O157:H7/NM: comprehensive phylogenomic analysis using closed genomes. *Front Microbiol* 2020;11:619.
118. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, *et al*. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J Clin Microbiol* 2014;52:1501–1510.
119. Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J Clin Microbiol* 2015;53:2410–2426.
120. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. *Nucleic Acids Res* 2016;44:D694-7.
121. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, *et al*. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020;48:D517–D525.
122. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, *et al*. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* 2014;58:212–220.
123. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, *et al*. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012;67:2640–2644.
124. Kleinheinz KA, Joensen KG, Larsen MV. Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and *E. coli* virulence genes in bacteriophage and prophage nucleotide sequences. *Bacteriophage* 2014;4:e27943.
125. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, *et al*. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;58:3895–3903.
126. Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb Genom* 2018;4.
127. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics* 2011;27:1009–1010.
128. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;7:539.
129. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* 2018;27:135–145.
130. Scheutz F, Teel LD, Beutin L, Pierard D, Buvens G, *et al*. Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *J Clin Microbiol* 2012;50:2951–2963.
131. Sanchez S, Llorente MT, Herrera-Leon L, Ramiro R, Nebreda S, *et al*. Mucus-activatable shiga toxin genotype stx2d in *Escherichia coli* O157:H7. *Emerg Infect Dis* 2017;23:1431–1433.
132. Malberg Tetzschner AM, Johnson JR, Johnston BD, Lund O, Scheutz F, *et al*. *In Silico* Genotyping of *Escherichia coli* Isolates for Extraintestinal Virulence Genes by Use of Whole-Genome Sequencing Data. *J Clin Microbiol* 2020;58:e01269-01220.
133. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, *et al*. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 2018;46:W537–W544.
134. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, *et al*. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res* 2017;45:W30–W35.
135. Bertelli C, Brinkman FSL. Improved genomic island predictions with IslandPath-DIMOB. *Bioinformatics* 2018;34:2161–2167.
136. Bertelli C, Tilley KE, Brinkman FSL. Microbial genomic island discovery, visualization and analysis. *Brief Bioinform* 2018.
137. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 2011;12:402.
138. Sahl JW, Caporaso JG, Rasko DA, Keim P. The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. *PeerJ* 2014;2:e332.
139. Eppinger M, Worsham PL, Nikolich MP, Riley DR, Sebastian Y, *et al*. Genome sequence of the deep-rooted *Yersinia pestis* strain Angola reveals new insights into the evolution and pangenome of the plague bacterium. *J Bacteriol* 2010;192:1685–1699.
140. Saeed AI, Sharov V, White J, Li J, Liang W, *et al*. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 2003;34:374–378.
141. Delannoy S, Mariani-Kurkdjian P, Webb HE, Bonacorsi S, Fach P. The mobilome; a major contributor to *Escherichia coli* stx2-positive O26:H11 strains intra-serotype diversity. *Front Microbiol* 2017;8:1625.
142. Cock PJA, Chilton JM, Gruning B, Johnson JE, Soranzo N. NCBI BLAST plus integrated into galaxy. *Gigascience* 2015;4.
143. Cascales E, Buchanan SK, Duche D, Kleanthous C, Lloubes R, *et al*. Colicin biology. *Microbiol Mol Biol Rev* 2007;71:158–229.
144. Vriezen JA, Valliere M, Riley MA. The evolution of reduced microbial killing. *Genome Biol Evol* 2009;1:400–408.
145. Zhang LH, Fath MJ, Mahanty HK, Tai PC, Kolter R. Genetic analysis of the colicin V secretion pathway. *Genetics* 1995;141:25–32.
146. Selkrig J, Mosbahi K, Webb CT, Belousoff MJ, Perry AJ, *et al*. Discovery of an archetypal protein transport system in bacterial outer membranes. *Nat Struct Mol Biol* 2012;19:510-S501.
147. Campos E, Baldoma L, Aguilar J, Badia J. Regulation of expression of the divergent ulaG and ulaABCDEF operons involved in LaAscorbate dissimilation in *Escherichia coli*. *J Bacteriol* 2004;186:1720–1728.

148. Clark G, Paszkiewicz K, Hale J, Weston V, Constantinidou C, *et al*. Genomic analysis uncovers a phenotypically diverse but genetically homogeneous *Escherichia coli* ST131 clone circulating in unrelated urinary tract infections. *J Antimicrob Chemother* 2012;67:868–877.
149. Driebe EM, Sahl JW, Roe C, Bowers JR, Schupp JM, *et al*. Using whole genome analysis to examine recombination across diverse sequence types of *Staphylococcus aureus*. *PLoS One* 2015;10:e0130955.
150. Bruen TC, Bryant D. Parsimony via consensus. *Syst Biol* 2008;57:251–256.
151. Hauser JR, Atitkar RR, Petro CD, Lindsey RL, Strockbine N, *et al*. The Virulence of *Escherichia coli* O157:H7 isolates in mice depends on shiga toxin type 2a (Stx2a)-induction and high levels of Stx2a in stool. *Front Cell Infect Microbiol* 2020;10:62.
152. Ogura Y, Ooka T, Terajima J, Nougayrède J-P, *et al*. Extensive genomic diversity and selective conservation of virulence-determinants in enterohemorrhagic *Escherichia coli* strains of O157 and non-O157 serotypes. *Genome Biol* 2007;8:R138.
153. Steyert SR, Sahl JW, Fraser CM, Teel LD, Scheutz F, *et al*. Comparative genomics and stx phage characterization of LEE-negative Shiga toxin-producing *Escherichia coli*. *Front Cell Infect Microbiol* 2012;2:133.
154. Higashi K, Sakamaki Y, Herai E, Demizu R, Uemura T, *et al*. Identification and functions of amino acid residues in PotB and PotC involved in spermidine uptake activity. *J Biol Chem* 2010;285:39061–39069.
155. Igarashi K, Kashiwagi K. Polyamine transport in bacteria and yeast. *Biochem J* 1999;344:633–642.
156. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, *et al*. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* 2001;8:11–22.
157. Groth AC, Calos MP. Phage integrases: biology and applications. *J Mol Biol* 2004;335:667–678.
158. Serra-Moreno R, Jofre J, Muniesa M. Insertion site occupancy by stx2 bacteriophages depends on the locus availability of the host strain chromosome. *J Bacteriol* 2007;189:6645–6654.
159. Ravi A, Estensmo ELF, Abee-Lund TML, Foley SL, Allgaier B, *et al*. Association of the gut microbiota mobilome with hospital location and birth weight in preterm infants. *Pediatr Res* 2017;82:829–838.
160. Cowley LA, Dallman TJ, Jenkins C, Sheppard SK. Phage predation shapes the population structure of shiga-toxigenic Escherichia coli O157:H7 in the UK: an evolutionary perspective. *Front Genet* 2019;10:763.
161. Del Cogliano ME, Pinto A, Goldstein J, Zotta E, Ochoa F, *et al*. Relevance of bacteriophage 933W in the development of hemolytic uremic syndrome (HUS). *Front Microbiol* 2018;9:3104.
162. Tyler JS, Beeri K, Reynolds JL, Alteri CJ, Skinner KG, *et al*. Prophage induction is enhanced and required for renal disease and lethality in an EHEC mouse model. *PLoS Pathog* 2013;9:e1003236.
163. Shimizu T, Ohta Y, Tsutsuki H, Noda M. Construction of a novel bioluminescent reporter system for investigating Shiga toxin expression of enterohemorrhagic *Escherichia coli*. *Gene* 2011;478:1–10.
164. Loś JM, Loś M, Wegrzyn G, Wegrzyn A. Differential efficiency of induction of various lambdoid prophages responsible for production of Shiga toxins in response to different induction agents. *Microb Pathog* 2009;47:289–298.
165. Muniesa M, Blanco JE, De Simon M, Serra-Moreno R, Blanch AR, *et al*. Diversity of stx2 converting bacteriophages induced from Shiga-toxin-producing *Escherichia coli* strains isolated from cattle. *Microbiology* 2004;150:2959–2971.
166. Balasubramanian S, Osburne MS, BrinJones H, Tai AK, Leong JM. Prophage induction, but not production of phage particles, is required for lethal disease in a microbiome-replete murine model of enterohemorrhagic *E. coli* infection. *PLoS Pathog* 2019;15:e1007494.
167. Abu-Ali GS, Ouellette LM, Henderson ST, Lacher DW, Riordan JT, *et al*. Increased adherence and expression of virulence genes in a lineage of *Escherichia coli* O157:H7 commonly associated with human infections. *PLoS One* 2010;5:e10167.
168. Bielaszewska M, Idelevich EA, Zhang W, Bauwens A, Schaumburg F, *et al*. Effects of antibiotics on Shiga toxin 2 production and bacteriophage induction by epidemic *Escherichia coli* O104:H4 strain. *Antimicrob Agents Chemother* 2012;56:3277–3282.
169. McAdams HH, Shapiro L. Circuit simulation of genetic networks. *Science* 1995;269:650–656.
170. Matsushiro A, Sato K, Miyamoto H, Yamamura T, Honda T. Induction of prophages of enterohemorrhagic *Escherichia coli* O157:H7 with norfloxacin. *J Bacteriol* 1999;181:2257–2260.
171. Wagner PL, Livny J, Neely MN, Acheson DW, Friedman DI, *et al*. Bacteriophage control of Shiga toxin 1 production and release by *Escherichia coli*. *Mol Microbiol* 2002;44:957–970.
172. Karch H, Schmidt H, Janetzki-Mittmann C, Scheef J, Kroger M. Shiga toxins even when different are encoded at identical positions in the genomes of related temperate bacteriophages. *Mol Gen Genet* 1999;262:600–607.
173. Iversen H, L' Abée-Lund TM, Aspholm M, Arnesen LPS, Lindbäck T. Commensal *E. coli* stx2 lysogens produce high levels of phages after spontaneous prophage induction. *Front Cell Infect Microbiol* 2015;5.
174. Gamage SD, Patton AK, Strasser JE, Chalk CL, Weiss AA. Commensal bacteria influence *Escherichia coli* O157:H7 persistence and Shiga toxin production in the mouse intestine. *Infect Immun* 2006;74:1977–1983.
175. Figler HM, Dudley EG. The interplay of *Escherichia coli* O157:H7 and commensal *E. coli*: the importance of strain-level identification. *Expert Rev Gastroenterol Hepatol* 2016;10:415–417.
176. Baumler AJ, Sperandio V. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* 2016;535:85–93.
177. Smith DL, James CE, Sergeant MJ, Yaxian Y, Saunders JR, *et al*. Short-tailed stx phages exploit the conserved YaeT protein to disseminate Shiga toxin genes among enterobacteria. *J Bacteriol* 2007;189:7223–7233.
178. Richter TKS, Michalski JM, Zanetti L, Tennant SM, Chen WH, *et al*. Responses of the human gut *Escherichia coli* population to pathogen and antibiotic disturbances. *mSystems* 2018;3.
179. Wong CS, Jelacic S, Habeeb RL, Watkins SL, Tarr PI. The risk of the hemolytic-uremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 infections. *N Engl J Med* 2000;342:1930–1936.
180. Dundas S, Todd WT, Stewart AI, Murdoch PS, Chaudhuri AK, *et al*. The central Scotland *Escherichia coli* O157:H7 outbreak: risk factors for the hemolytic uremic syndrome and death among hospitalized patients. *Clin Infect Dis* 2001;33:923–931.
181. Gould LH, Demma L, Jones TF, Hurd S, Vugia DJ, *et al*. Hemolytic uremic syndrome and death in persons with *Escherichia coli* O157:H7 infection, foodborne diseases active surveillance network sites, 2000-2006. *Clin Infect Dis* 2009;49:1480–1485.
182. Foster DB. Modulation of the enterohemorrhagic *E-coli* virulence program through the human gastrointestinal tract. *Virulence* 2013;4:315–323.
183. Dutilh BE, Backus L, Edwards RA, Wels M, Bayjanov JR, *et al*. Explaining microbial phenotypes on a genomic scale: GWAS for microbes. *Brief Funct Genomics* 2013;12:366–380.
184. Falush D. Toward the use of genomics to study microevolutionary change in bacteria. *PLoS Genet* 2009;5:e1000627.
185. Falush D, Bowden R. Genome-wide association mapping in bacteria? *Trends Microbiol* 2006;14:353–355.
186. Bono JL. Genotyping *Escherichia coli* O157:H7 for its ability to cause disease in humans. *Curr Protoc Microbiol* 2009;5A:3.

187. Werber D, Scheutz F. The importance of integrating genetic strain information for managing cases of shiga toxin-producing *E. coli* infection. *Epidemiol Infect* 2019;147:e264.

188. Sadiq SM, Hazen TH, Rasko DA, Eppinger M. EHEC genomics: past, present, and future. microbiol spectr. *Microbiol Spectr* 2014;2.

189. Eppinger M, Daugherty S, Agrawal S, Galens K, Sengamalay N, *et al*. Whole-genome draft sequences of 26 enterohemorrhagic escherichia coli O157:H7 strains. *Genome Announc* 2013;1:e0013412.

190. Balasubramanian S, Osburne MS, BrinJones H, Tai AK, Leong JM. Prophage induction, but not production of phage particles, is required for lethal disease in a microbiome-replete murine model of enterohemorrhagic *E. coli* infection. *PLOS Pathog* 2019;15:e1007494.

191. Hernandez-Doria JD, Sperandio V. Bacteriophage transcription factor Cro regulates virulence gene expression in enterohemorrhagic *Escherichia coli*. *Cell Host Microbe* 2018;23:607–617.

192. Hu J, Ye H, Wang S, Wang J, Han D. Prophage activation in the intestine: insights into functions and possible applications. *Front Microbiol* 2021;12:3930.

193. Kimmitt PT, Harwood CR, Barer MR. Toxin gene expression by shiga toxin-producing *Escherichia coli*: the role of antibiotics and the bacterial SOS response. *Emerg Infect Dis* 2000;6:458–465.

194. Rodríguez-Rubio L, Haarmann N, Schwidder M, Muniesa M, Schmidt H. Bacteriophages of shiga toxin-producing *Escherichia coli* and their contribution to pathogenicity. *Pathogens* 2021;10.

195. Tyler JS, Beeri K, Reynolds JL, Alteri CJ, Skinner KG, *et al*. Prophage induction is enhanced and required for renal disease and lethality in an EHEC mouse model. *PLOS Pathog* 2013;9:e1003236.

196. Boyer RR, Sumner SS, Williams RC, Pierson MD, Popham DL, *et al*. Influence of curli expression by *Escherichia coli* 0157:H7 on the cell's overall hydrophobicity, charge, and ability to attach to lettuce. *J Food Prot* 2007;70:1339–1345.

197. Macarisin D, Patel J, Bauchan G, Giron JA, Sharma VK. Role of curli and cellulose expression in adherence of *Escherichia coli* O157:H7 to spinach leaves. *Foodborne Pathog Dis* 2012;9:160–167.

198. Bando SY, Iamashita P, Guth BE, Dos Santos LF, Fujita A, *et al*. A hemolytic-uremic syndrome-associated strain O113:H21 Shiga toxin-producing *Escherichia coli* specifically expresses A transcriptional module containing dicA and is related to gene network dysregulation in Caco-2 cells. *PLoS One* 2017;12:e0189613.

199. Bando SY, Iamashita P, Silva FN, Costa L da F, Abe CM, *et al*. Dynamic gene network analysis of Caco-2 cell response to shiga toxin-producing *Escherichia coli*-associated hemolytic-uremic syndrome. *Microorganisms* 2019;7:E195.

200. Sy BM, Lan R, Tree JJ. Early termination of the Shiga toxin transcript generates a regulatory small RNA. *Proc Natl Acad Sci U S A* 2020;117:25055–25065.

201. Kiel M, Sagory-Zalkind P, Miganeh C, Stork C, Leimbach A, *et al*. Identification of novel biomarkers for priority serotypes of shiga toxin-producing *Escherichia coli* and the development of multiplex PCR for their detection. *Front Microbiol* 2018;9:1321.

202. Dytoc MT, Ismaili A, Philpott DJ, Soni R, Brunton JL, *et al*. Distinct binding properties of eaeA-negative verocytotoxin-producing *Escherichia coli* of serotype O113:H21. *Infect Immun* 1994;62:3494–3505.

203. Herold S, Paton JC, Paton AW. Sab, a novel autotransporter of locus of enterocyte effacement-negative shiga-toxigenic *Escherichia coli* O113:H21, contributes to adherence and biofilm formation. *Infect Immun* 2009;77:3234–3243.

204. Luck SN, Badea L, Bennett-Wood V, Robins-Browne R, Hartland EL. Contribution of FliC to epithelial cell invasion by enterohemorrhagic *Escherichia coli* O113:H21. *Infect Immun* 2006;74:6999–7004.

205. Rogers TJ, Paton JC, Wang H, Talbot UM, Paton AW. Reduced virulence of an fliC mutant of Shiga-toxigenic *Escherichia coli* O113:H21. *Infect Immun* 2006;74:1962–1966.

206. Rogers TJ, Thorpe CM, Paton AW, Paton JC. Role of lipid rafts and flagellin in invasion of colonic epithelial cells by Shiga-toxigenic *Escherichia coli* O113:H21. *Infect Immun* 2012;80:2858–2867.

207. Gerhardt E, Masso M, Paton AW, Paton JC, Zotta E, *et al*. Inhibition of water absorption and selective damage to human colonic mucosa are induced by subtilase cytotoxin produced by *Escherichia coli* O113:H21. *Infect Immun* 2013;81:2931–2937.

208. Seyahian EA, Oltra G, Ochoa F, Melendi S, Hermes R, *et al*. Systemic effects of Subtilase cytotoxin produced by *Escherichia coli* O113:H21. *Toxicon* 2017;127:49–55.

209. Santos E, Castro VS, Cunha-Neto A, Santos LFD, Vallim DC, *et al*. *Escherichia coli* O26 and O113:H21 on Carcasses and beef from a slaughterhouse located in Mato Grosso, Brazil. *Foodborne Pathog Dis* 2018;15:653–659.

210. Sanso AM, Bustamante AV, Kruger A, Cadona JS, Alfaro R, *et al*. Molecular epidemiology of Shiga toxin-producing O113:H21 isolates from cattle and meat. *Zoonoses Public Health* 2018;65:569–577.