



# Pathogenomes of Atypical Non-shigatoxigenic *Escherichia coli* NSF/SF O157:H7/NM: Comprehensive Phylogenomic Analysis Using Closed Genomes

Emmanuel C. Nyong<sup>1,2†</sup>, Sam R. Zaia<sup>1,2†</sup>, Anna Allué-Guardia<sup>1,2</sup>, Armando L. Rodriguez<sup>3</sup>, Zaina Irion-Byrd<sup>1,2</sup>, Sara S. K. Koenig<sup>1,2</sup>, Peter Feng<sup>4‡</sup>, James L. Bono<sup>5</sup> and Mark Eppinger<sup>1,2\*</sup>

## OPEN ACCESS

### Edited by:

Maria Hoffmann,  
Division of Microbiology (FDA), United States

### Reviewed by:

Frederic Auvray,  
Ecole Nationale Vétérinaire de Toulouse, France  
Edward G. Dudley,  
The Pennsylvania State University (PSU), United States

### \*Correspondence:

Mark Eppinger  
mark.eppinger@utsa.edu

†These authors have contributed equally to this work

‡Retired

### Specialty section:

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 04 November 2019

Accepted: 19 March 2020

Published: 15 April 2020

### Citation:

Nyong EC, Zaia SR, Allué-Guardia A, Rodriguez AL, Irion-Byrd Z, Koenig SSK, Feng P, Bono JL and Eppinger M (2020) Pathogenomes of Atypical Non-shigatoxigenic *Escherichia coli* NSF/SF O157:H7/NM: Comprehensive Phylogenomic Analysis Using Closed Genomes. *Front. Microbiol.* 11:619. doi: 10.3389/fmicb.2020.00619

<sup>1</sup> Department of Biology, The University of Texas at San Antonio, San Antonio, TX, United States, <sup>2</sup> South Texas Center for Emerging Infectious Diseases, San Antonio, TX, United States, <sup>3</sup> Research Computing Support Group, The University of Texas at San Antonio, San Antonio, TX, United States, <sup>4</sup> Retired, Rockville, MD, United States, <sup>5</sup> United States Meat Animal Research Center, Agricultural Research Service, United States Department of Agriculture (ARS-USDA), Clay Center, NE, United States

The toxigenic conversion of *Escherichia coli* strains by Shiga toxin-converting (Stx) bacteriophages were prominent and recurring events in the stepwise evolution of enterohemorrhagic *E. coli* (EHEC) O157:H7 from an enteropathogenic (EPEC) O55:H7 ancestor. Atypical, attenuated isolates have been described for both non-sorbitol fermenting (NSF) O157:H7 and SF O157:NM serotypes, which are distinguished by the absence of Stx, the characteristic virulence hallmark of Stx-producing *E. coli* (STEC). Such atypical isolates either never acquired Stx-phages or may have secondarily lost *stx* during the course of infection, isolation, or routine subculture; the latter are commonly referred to as LST (Lost Shiga Toxin)-isolates. In this study we analyzed the genomes of 15 NSF O157:H7 and SF O157:NM strains from North America, Europe, and Asia that are characterized by the absence of *stx*, the virulence hallmark of STEC. The individual genomic basis of the Stx (–) phenotype has remained largely undetermined as the majority of STEC genomes in public genome repositories were generated using short read technology and are in draft stage, posing a major obstacle for the high-resolution whole genome sequence typing (WGST). The application of LRT (long-read technology) sequencing provided us with closed genomes, which proved critical to put the atypical non-shigatoxigenic NSF O157:H7 and SF O157:NM strains into the phylogenomic context of the stepwise evolutionary model. Availability of closed chromosomes for representative Stx (–) NSF O157:H7 and SF O157:NM strains allowed to describe the genomic basis and individual evolutionary trajectories underlying the absence of Stx at high accuracy and resolution. The ability of LRT to recover and accurately assemble plasmids revealed a strong correlation between the strains' featured plasmid genotype and chromosomally inferred clade, which suggests the coevolution of the chromosome and accessory plasmids. The identified ancestral traits

in the pSFO157 plasmid of NSF O157:H7 strain LSU-61 provided additional evidence for its intermediate status. Taken together, these observations highlight the utility of LRTs for advancing our understanding of EHEC O157:H7/NM pathogenome evolution. 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.

**Keywords:** Shiga toxin (Stx) producing *Escherichia coli* (STEC), enterohemorrhagic *E. coli* (EHEC), NSF O157:H7, SF O157:NM, whole genome sequencing and typing (WGST), comparative phylogenomics, LRT (long read technology), SRT (short read technology)

## INTRODUCTION

Production of a potent cytotoxin named Shiga toxin (Stx) (Trofa et al., 1999) is a virulence hallmark of STEC (O'Brien et al., 1983). Within STEC, Enterohemorrhagic *Escherichia coli* (EHEC) is a subset comprised of highly pathogenic strains, such as the O157:H7 serotype. Since its first association with human disease (Wells et al., 1983; Pennington, 2010; Sanjar et al., 2014), this once rare serotype now exists globally, and has become a public health threat for severe and widespread foodborne outbreaks. This particular lineage is the dominant causative agent of STEC outbreaks in the United States between 2007 and 2018<sup>1</sup> (Mead et al., 1999; Perna et al., 2001; Rangel et al., 2005; Ferens and Hovde, 2011), and no vaccines and only a limited arsenal of therapeutic or preventive countermeasures are available (Nguyen and Sperandio, 2012). Similarly, the sorbitol fermenting (SF) O157:NM strains, that carry the H7 antigen genes, but are non-motile (NM), are a recognized pathogen that has caused many severe outbreaks in the EU (Feng et al., 2007). The ability to induce cellular damage leading to disease in humans depends on numerous virulence factors, though the most prominent is Stx, encoded on mobile temperate Stx-converting bacteriophages (Stx-phages) (Huang et al., 1987; Schmidt, 2001; Asadulghani et al., 2009; Kruger and Lucchesi, 2015). Stx<sub>2a</sub> is the most cytopathic subtype produced by this lineage, and has been associated with severe infections in humans (Tesh et al., 1993; Orth et al., 2007; Fuller et al., 2011; Russo et al., 2014). The 70-kDa toxin acts as a protein synthesis inhibitor (Melton-Celsa, 2014), causing cytopathic effects in renal endothelial cells (Konowalchuk et al., 1977; Sandvig, 2001; Tesh, 2010; Bentancor et al., 2013a,b; Lee et al., 2016). Stx is especially toxigenic to renal cells, which express the Stx-receptor globotriaosylceramide or Gb3 (Eaton et al., 2008; Shin et al., 2009; Johannes and Romer, 2010; Mauro and Koudelka, 2011; Melton-Celsa et al., 2012; Nguyen and Sperandio, 2012; Tran et al., 2014; Pacheco et al., 2018). Stx-phages are comprised of highly diverse entities (Miyamoto et al., 1999; Muniesa et al., 2004; Gamage et al., 2006; Smith et al., 2012; Kruger and Lucchesi, 2015; Shaaban et al., 2016) and their genome plasticity has been implicated to directly, or indirectly, alter Stx production (Miyamoto et al., 1999; Wagner et al., 1999; Muniesa et al., 2003; Tyler et al., 2004; Ogura et al., 2015; Yin et al., 2015). The toxigenic conversion through acquisition of Stx-phages or their secondary loss are recurring events in the STEC group (Zhou et al., 2010; Kyle et al., 2012).

In the evolution of O157:H7, the acquisition of Stx-phages by *stx*-negative ancestral progenitors (Ogura et al., 2009; Zhou et al., 2010; Cooper et al., 2014; Dallman et al., 2015; Lorenz et al., 2017) occurred at various stages as described in the stepwise evolution model (Feng et al., 1998; Wick et al., 2005; Leopold et al., 2010; Rump et al., 2011; Jenke et al., 2012; Kyle et al., 2012). Carriage of Stx-phages has far-ranging evolutionary implications in the emergence, persistence, and dissemination of *stx* genes within and across STEC pathovars and prophages make up a significant portion of the EHEC O157:H7/NM chromosome resulting in widespread genetic mosaicism (Ohnishi et al., 2001; Asadulghani et al., 2009; Garcia-Aljaro et al., 2009; Kruger and Lucchesi, 2015). A single strain may carry multiple and different Stx-phages (Rusconi et al., 2016). The loss of Stx-phages can occur via induction, during infection, or even routine culturing of bacterial strains (Bielaszewska et al., 2007; Ferdous et al., 2015). Stx (–) strains have been observed in both NSF O157:H7 and SF O157:NM serogroups (Feng et al., 2001; Mellmann et al., 2005; Friedrich et al., 2007; Bielaszewska et al., 2008; Stephan et al., 2009; Themphachana et al., 2014; Uhlich et al., 2019), and other STEC (Bielaszewska et al., 2007, 2008; Tennant et al., 2009; Bentancor et al., 2010). Stx (–) strains are attenuated and also likely impacted in their overall bacterial fitness. Beyond the anthropomorphic view of Stx as key virulence factor in human STEC pathogenesis, the carriage of Stx-phages has been associated with toxin-dependent killing of predators (King et al., 1999; Eklund et al., 2002; Steinberg and Levin, 2007; Poirier et al., 2008; Lainhart et al., 2009; Los et al., 2011, 2012; Mauro and Koudelka, 2011; Stolfa and Koudelka, 2012), altered acid resistance, motility and metabolism (Robinson et al., 2006; Su et al., 2010; Veses-Garcia et al., 2015; Saile et al., 2016; Berger et al., 2019). However, we note their ability to potentially (re-)acquire Stx through Stx-phage infection (Mellmann et al., 2008). In this study, we applied LRT technology (Koren et al., 2013; Allue-Guardia et al., 2018b) and provide closed genomes and comprehensive phylogenomic analyses for representative Stx (–) NSF O157:H7 and SF O157:NM strains in our collection that tested negative for *stx*, or were cured of their Stx-phage inventory in our laboratory.

## MATERIALS AND METHODS

### Strains and Characterization

A collection of 12 Stx (–) NSF O157:H7 and SF O157:NM strains of global origin was assembled, along with three laboratory-cured

<sup>1</sup><https://www.cdc.gov/ecoli/outbreaks.html>

Stx (–) strains. Strains were isolated from different sources, including clinical cases, the animal reservoir, and produce. Except for SF O157:NM, most O157:H7 strains do not ferment sorbitol due to frameshifts in the *srlA* and *srlE* components of the glucitol/sorbitol-specific phosphotransferase system (Wick et al., 2005). Coded by the *uidA* gene,  $\beta$ -glucuronidase (GUD) is an inducible enzyme produced by most generic and pathogenic *E. coli*, except for O157:H7. *E. coli uidA* features both SNP and frameshift mutations but the loss of GUD production in O157:H7 is caused by a frame-shift mutations in *uidA* (+686 GG) (Feng and Lampel, 1994; Monday et al., 2001). Both of these phenotypes were key components in the development of the stepwise O157:H7 evolution model. For this reason, we specifically selected isolates that either exhibited or were negative in these metabolic phenotypes. Strain-associated metadata can be found in **Supplementary Table S1**. Isolates were cultured on CHROMagar-STE<sub>C</sub> and screened for the O157 antigen with the *E. coli* O157 Latex Test Kit (Thermo Fisher). The presence or absence of *stx* genes was determined by PCR (Scheutz et al., 2012) using GoTaq Green Master Mix (Promega) in a 25  $\mu$ l reaction volume. O157:H7 strains that tested negative for *stx* genes were selected for complete genome sequencing.

## Genome Sequencing, Assembly, and Annotation

Strains were cultured in lysogeny broth (LB) medium (Thermo Fisher Scientific, Asheville, NC, United States) overnight at 37°C in a shaker. Total genomic DNA was extracted using the QIAamp DNA Mini Kit (Qiagen, Inc., Valencia, CA, United States) according to the manufacturer's instructions. Nine isolates were sequenced to closure using long-read PacBio RS II technology (**Supplementary Table S1**). Genomic DNA was sheared into 20-kb fragments using g-TUBE (Covaris, Inc., Woburn, MA, United States). The library was prepared based on the 20-kb PacBio sample preparation protocol and sequenced using P6/C4 chemistry with either one, two or three single-molecule real-time (SMRT) cells with a 240-min collection time. The continuous long-read data were *de novo* assembled using the PacBio hierarchical genome assembly process (HGAP version 3.0) (Chin et al., 2013) with default parameters in SMRT Analysis v2.3.0, including consensus polishing with Quiver (Chin et al., 2013). In addition, some strains were sequenced using the short-read Illumina MiSeq platform (**Supplementary Table S1**). For Illumina sequenced isolates, a paired-end library was prepared using the NxSeq AmpFREE Low DNA Library Kit (Lucigen) with 250-bp read length and sequenced using the MiSeq Reagent kit v2 500-cycle (Illumina), following the manufacturer's guidelines. Illumina fastq reads were trimmed and quality controlled with FastQC (Andrew, 2010) and Trim Galore (Krueger, 2017). Illumina reads were *de novo* assembled with SPAdes in the careful mode, which includes realignment (Bankevich et al., 2012). Resulting contigs were QC-filtered based on size ( $\leq 1$  kb) and coverage ( $\geq 10\times$ ). Illumina reads were further used for error-correction of PacBio-only assemblies with Pilon (Walker et al., 2014). Resulting

assemblies were QCed with QUAST (Gurevich et al., 2013; Mikheenko et al., 2018). The chromosomal and plasmid origins of replication<sup>2</sup>, *oriC* and *repA* (Gao and Zhang, 2008; Luo et al., 2018), were determined prior to annotation through the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova et al., 2016).

## Serotype, Clade, and MLST Classification of Stx (–) NSF O157:H7 and SF O157:NM

To confirm the serotyping results of the *E. coli* O157 Latex Test Kit (Thermo Fisher), *in silico* serotyping was performed as described (Joensen et al., 2015). Clade typing was performed as originally defined by Manning et al. (2008). Clades and subgroups were assigned by *in silico* interrogation of the allelic status of 89 core genome (cg)SNPs in the assembled genomes using a custom workflow on Galaxy (Giardine et al., 2005; Goecks et al., 2010), which is informed by eight definitive polymorphic positions (Riordan et al., 2008; Yokoyama et al., 2012). The multilocus sequence types (MLST) of NSF O157:H7 and SF O157:NM cultures from our strain collection and the genomes of similar strains retrieved from GenBank were determined *in silico* with the Achtman MLST scheme, which determined the Sequence Type (ST) based on alleles in seven housekeeping genes (Wirth et al., 2006). Respective reads and/or assembled genomes were analyzed in MLST 2.0<sup>3</sup> (Larsen et al., 2012) and Ridom SeqSphere+ (Junemann et al., 2013).

## Pathogenome Make-Up and Virulence Complement of Stx (–) NSF O157:H7 and SF O157:NM

The virulence and antibiotic resistance complement was identified using VirulenceFinder<sup>4</sup> (Joensen et al., 2014, 2015), VDFDB (Chen et al., 2016) and ResFinder<sup>5</sup> (Zankari et al., 2012; Kleinheinz et al., 2014). Prophages and plasmids were identified and distinguished from the core genomes using PHASTER (Zhou et al., 2011; Arndt et al., 2016) and PlasmidFinder<sup>6</sup> (Carattoli et al., 2014). Insertion sequence (IS) elements were identified and boundaries further manually curated using ISEScan (Xie and Tang, 2017) and Iceberg (Liu et al., 2019). Genomic islands (GI) were detected with IslandViewer4 (Bertelli et al., 2017, 2018; Bertelli and Brinkman, 2018).

## Comparison of Genome Architectures and Proteome Inventory in NSF O157:H7 and SF O157:NM

To detect major structural changes in the genome architectures, sequences and gene inventories of closed chromosomes and plasmids were compared with BRIG (v0.95) (Alikhan et al., 2011) and by BLASTn/p against the non-redundant NCBI databases

<sup>2</sup><http://origin.tubic.org/Ori-Finder/>

<sup>3</sup><https://cge.cbs.dtu.dk/services/MLST/>

<sup>4</sup><https://cge.cbs.dtu.dk/services/VirulenceFinder/>

<sup>5</sup><https://cge.cbs.dtu.dk/services/ResFinder/>

<sup>6</sup><https://cge.cbs.dtu.dk/services/PlasmidFinder/>

(Altschul et al., 1990; Camacho et al., 2009; Cock et al., 2015). BRIG visualization allowed to distinguish the core chromosome from accessory prophages and other MGEs (Alikhan et al., 2011), and further to catalog subtle polymorphisms among the carried plasmid types. To study the prevalence of the identified virulence gene complement of the core genome and carried pO157 plasmids, we used large-scale BLAST score ratio (LS-BSR) (Eppinger et al., 2010; Sahl et al., 2014; Rusconi et al., 2016) with tBLASTn (Altschul et al., 1990). 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 from VirulenceFinder were visualized as heatmaps with Multiple Experiment Viewer (MeV) (v.4.8) (Saeed et al., 2003).

## Shiga Toxin Status, Stx-Phage Profiling and Visualization of Phage Insertion Sites

To confirm the absence of *stx* observed from the PCR pre-screen, reads and assembled genomes were analyzed *in silico* with VirulenceFinder (Joensen et al., 2014, 2015; Kleinheinz et al., 2014) and VFDB (Chen et al., 2016). The Stx subtypes carried by the *stx*-positive reference isolates EC4115 (*stx2a*, *stx2c*) and EDL933 (*stx1*, *stx2a*) were recorded as described in Scheutz et al. (2012) and Ashton et al. (2015). STEC strains can carry Stx1 and Stx2 toxins that come in different subtypes. In NSF O157:H7 and SF O157:NM host genomes, Stx-phages target preferential loci (Kruger and Lucchesi, 2015; Yang et al., 2020). For the comparative analysis of Shiga toxin-encoding bacteriophage insertion (SBI) sites in Stx (+) and Stx (-) O157:H7, we used intact gene sequences of the NADH quinone oxidoreductase (*wrbA*), the tRNA (*argW*) (Eppinger et al., 2011b), transcriptional regulator (*yehV*) (Yokoyama et al., 2000), exonuclease (*sbcB*) (Ohnishi et al., 2002; Strauch et al., 2008), *yecE* of unknown function (Shringi et al., 2012), and oxidoreductase (Z2577) (Koch et al., 2003; Serra-Moreno et al., 2007). Unoccupied SBI were defined as those strains that showed undisrupted BLASTn alignments when queried (Altschul et al., 1990) against the O157:H7 genome assemblies. Unbiased of established phage insertion sites in O157:H7, boundaries and location of intact, partial or remnant prophages were further identified in PHASTER (Arndt et al., 2016), and genome architectures of chromosomal Stx-phage insertion sites were comprehensively analyzed and visualized using Easyfig (v2.2.2) (Sullivan et al., 2011). For each of the strains, corresponding Stx-phage insertion loci found in *stx*-positive strains EC4115 (*stx2a* at *argW* and *stx2c* at *sbcB*) and EDL933 (*stx2a* at *wrbA*, *stx1* at *yehV*) served as reference. A fragment extended by a 2 kb extension on each side was extracted and compared by BLASTn (Altschul et al., 1990).

## Comparative Phylogenomics Whole Genome Alignment (WGA) Phylogeny

The Stx (-) O157:H7 genomes sequenced in this study along with Stx (-/+) NSF O157:H7 and SF O157:NM genomes downloaded from NCBI GenBank (**Supplementary Table S1**) and Stx (+) O157:H7 reference strain EC4115 (Eppinger et al., 2011b) were used to construct a whole genome-based phylogenetic tree. The phylogeny was inferred from whole genome alignments (WGA) using Mugsy (v1.2.3) (Angiuoli and Salzberg, 2011) and RAXML (v4.0) (Stamatakis, 2014) with 100 bootstrap replicates. The tree was visualized in Geneious 2019 (v1.2.) (Kearse et al., 2012) and decorated with strain-associated metadata in EvolView (v3) (Zhang et al., 2012; He et al., 2016; Subramanian et al., 2019).

## 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 Eppinger et al. (2010, 2011b, 2014), Rusconi et al. (2016), and Hau et al. (2018), which is implemented on the open-source web-based bioinformatics platform Galaxy (Goecks et al., 2010). The chromosomal core was defined as the set of genic and intragenic regions that are not repeated, do not contain prophages, IS elements, GIs, or other mobile genome elements (MGEs), which evolve at different rates and therefore are not indicative of evolutionary relationships. These regions were determined in the designated closed reference *E. coli* O157:H7 strain EC4115 (Eppinger et al., 2011b) as follows: Repeats with NUCmer (v3.22) (Delcher et al., 2003), prophages with PHASTER (Zhou et al., 2011; Arndt et al., 2016), and IS elements with ISFinder (Siguier et al., 2006), ISEScan (v1.7.1) (Xie and Tang, 2017), and ICEberg (Liu et al., 2019). 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 with BWA-MEM (Li and Durbin, 2009) to the designated reference genome EC4115. The resulting alignments were processed with FreeBayes (v1.3.1) (Garrison and Marth, 2012) with the following threshold settings: mapping quality 30, base quality 30, coverage 10, and allelic frequency 0.75. For contig-based discovery, PacBio-only assemblies and Illumina error-corrected PacBio assemblies were aligned to the reference EC4115 chromosome (Eppinger et al., 2011b) using NUCmer, followed by SNP prediction with delta-filter and show-snps distributed with the MUMmer package (Delcher et al., 2003; Marcais et al., 2018). The resulting SNP panels for each of the query genomes were used for further processing; (ii) SNP validation and filtering. We used several SNP curation strategies detailed in our previous works (Eppinger et al., 2011b, 2014; Rusconi et al., 2016). Cataloged 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 described (Myers et al., 2009; Morelli et al., 2010; Eppinger et al., 2014). 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 (Altschul et al., 1990). 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 annotation (Manning et al., 2008; Bono, 2009; Leopold et al., 2010; Rusconi et al., 2016). SNP-matrix tables were manipulated with Query Tabular Tool (Johnson et al., 2018). 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) (Wilgenbusch and Swofford, 2003) with 100 bootstrap replicates. The majority rule consensus SNP tree was visualized in Geneious (Kearse et al., 2012) and decorated with EvolView (Zhang et al., 2012; He et al., 2016; Subramanian et al., 2019). Calculation of the consistency index (CI) in Mesquite (Maddison and Maddison, 2016) for each SNP allowed us to identify parsimony informative SNPs and flag homoplastic SNPs, as described in our previous works (Eppinger et al., 2010, 2011b, 2014; Rusconi et al., 2016; Hau et al., 2018). This strategy was also used for the discovery of plasmid-borne SNPs referenced to closed pO157 and pSFO157 plasmids of strains EC4115 (Eppinger et al., 2011b) and 3072/96 (Brunner et al., 2006), respectively.

## Laboratory Isolation of *stx*-Cured EHEC O157:H7

Therapeutic use of certain antibiotics is known to induce Stx<sub>2a</sub> phages (Huang et al., 1987; Mead and Griffin, 1998; Zhang et al., 2000; Zimmerhackl, 2000; Wong and Brandt, 2002; McGannon et al., 2010; Wong et al., 2012; Amorim et al., 2014; Rahal et al., 2015). Strains were cured of *stx* by inducing the Stx-phages with Mitomycin C (MMC), a potent Stx<sub>2a</sub> phage inducing agent that triggers the RecA-mediated SOS-response (Fuchs et al., 1999; Kimmitt et al., 2000; Los et al., 2009; Imamovic and Muniesa, 2012). Overnight cultures of O157:H7 strains EC4115, PA2, and PA11 were diluted to an OD<sub>600</sub> of 0.05 and grown to an OD<sub>600</sub> of 0.3–0.5 in fresh LB medium at 37°C. MMC, a light sensitive chemical, was added at a final concentration of 0.5 µg/ml and cultures were incubated for 6 h (37°C, 200 rpm) in the dark. To recover clones that lost the *stx*, cultures were serially diluted and plated on LB agar (37°C, 16 h). Plates with 100–300 colonies were selected and the colonies were blotted to a nylon membrane (GE Healthcare Amersham Hybond-N+). Colonies that lost both the *stx1* and *2* genes were detected after colony blot hybridization (Sambrook et al., 2006) with a DIG-labeled *stx*-specific probe using the PCR DIG Synthesis and DNA Labeling and Detection Kits (Roche). The probe consisted of a 255 bp fragment amplified with primers 5'-GGCACTGTCTGAAACTGCTCC-3' and 5'-TCGCCAGTTATCTGACATTCTG-3' (Rajkhowa et al., 2010) following the protocol for the Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific) in a T100 Thermal Cycler (Bio-Rad). The absence of *stx* in the recovered clones was confirmed by PCR using *stx* and insertion site-specific primers using conditions developed for the Shiga toxin-encoding bacteriophage insertion

site (SBI) assay (Shaikh and Tarr, 2003; Besser et al., 2007; Scheutz et al., 2012).

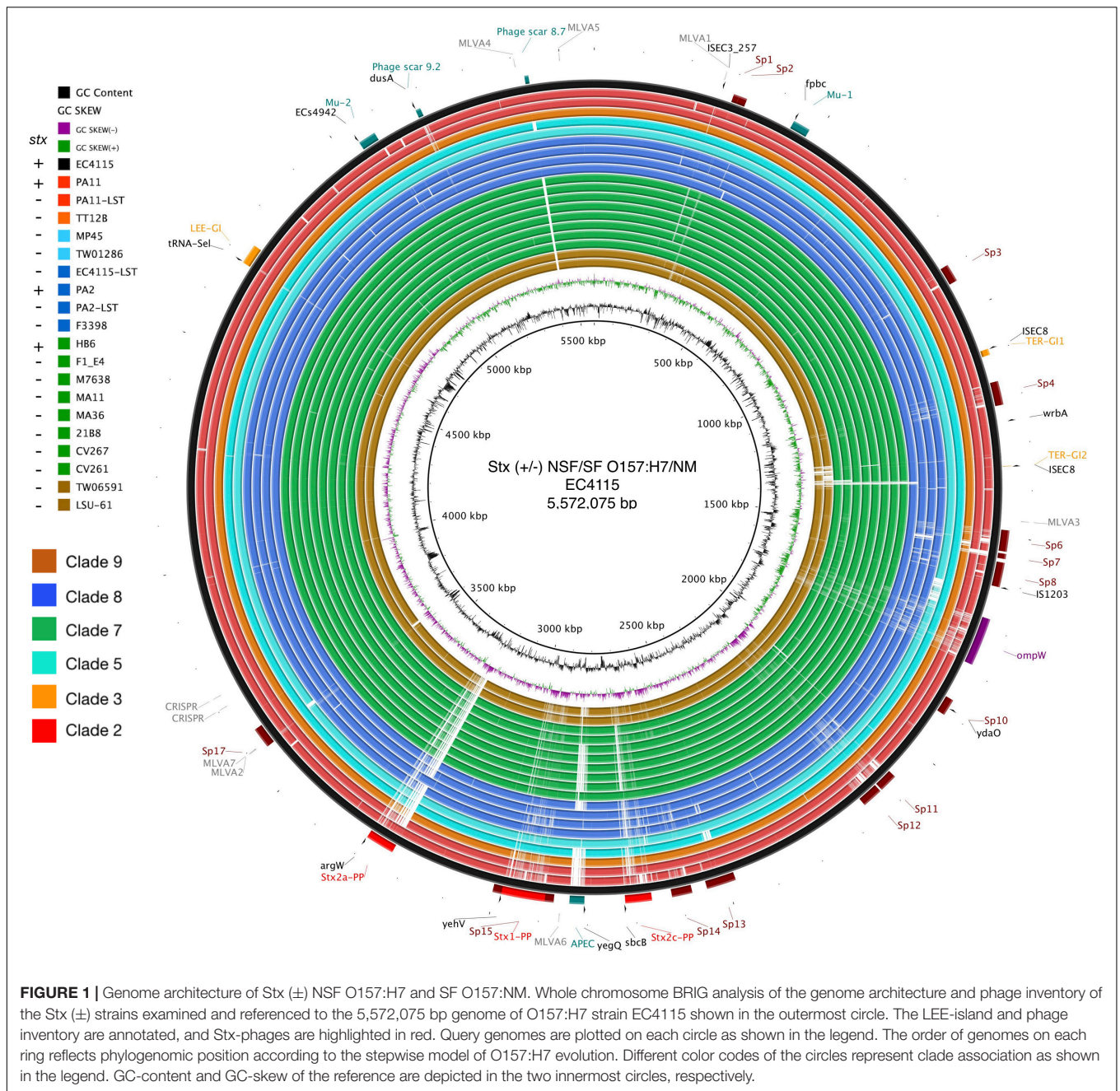
## RESULTS AND DISCUSSION

### LRT Sequencing of *Stx* (–) NSF O157:H7 and SF O157:NM Strains

In this study we sequenced the genomes of 14 NSF O157:H7 strains and one SF O157:H7 that are all characterized by absence of *stx*, the virulence hallmark of STEC. Isolates were analyzed along with 26 other *Stx* (±) NSF O157:H7 and SF O157:NM genomes available from NCBI GenBank (**Supplementary Table S1**). We re-sequenced the genome of SF O157:H7 strain LSU-61 to closure (Dunn et al., 2004), a presumed intermediate (Clade 9.2.36) in the stepwise evolution model (Feng et al., 1998, 2007; Wick et al., 2005; Leopold et al., 2010; Jenke et al., 2012; Kyle et al., 2012), previously only available as draft (Rump et al., 2011). Genomes of EHEC O157:H7/NM are notorious for assembling into fragmented draft genomes by the more commonly used short-read technologies (SRTs) due to the homogeneous nature of lambdoid prophages content and other repeats. Contig breaks in STEC assemblies typically occur within *Stx*-phages producing numerous phage-related contigs, which hinders accurate *Stx*-phage profiling (Asadulghani et al., 2009; Eppinger et al., 2011a,b, 2013; Smith et al., 2012; Sadiq et al., 2014; Yin et al., 2015). In response, we used LRT sequencing (English et al., 2012) on the PacBioRS II platform (Koren et al., 2013; Cooper et al., 2014; Latif et al., 2014; Loman et al., 2015; Cowley et al., 2016; Shaaban et al., 2016; Allue-Guardia et al., 2018a,b), followed by Illumina short read-error correction with Pilon (Walker et al., 2014). The resulting closed (or near-closed) genomes and plasmids (**Supplementary Table S2**) were subjected to high-resolution whole genome sequence typing (WGST), and provided the critical foundation to investigate the genomic basis of the strains' atypical *Stx* (–) status and to determine the phylogenomic position of these atypical *Stx* (–) isolates within the EHEC O157:H7/NM lineage.

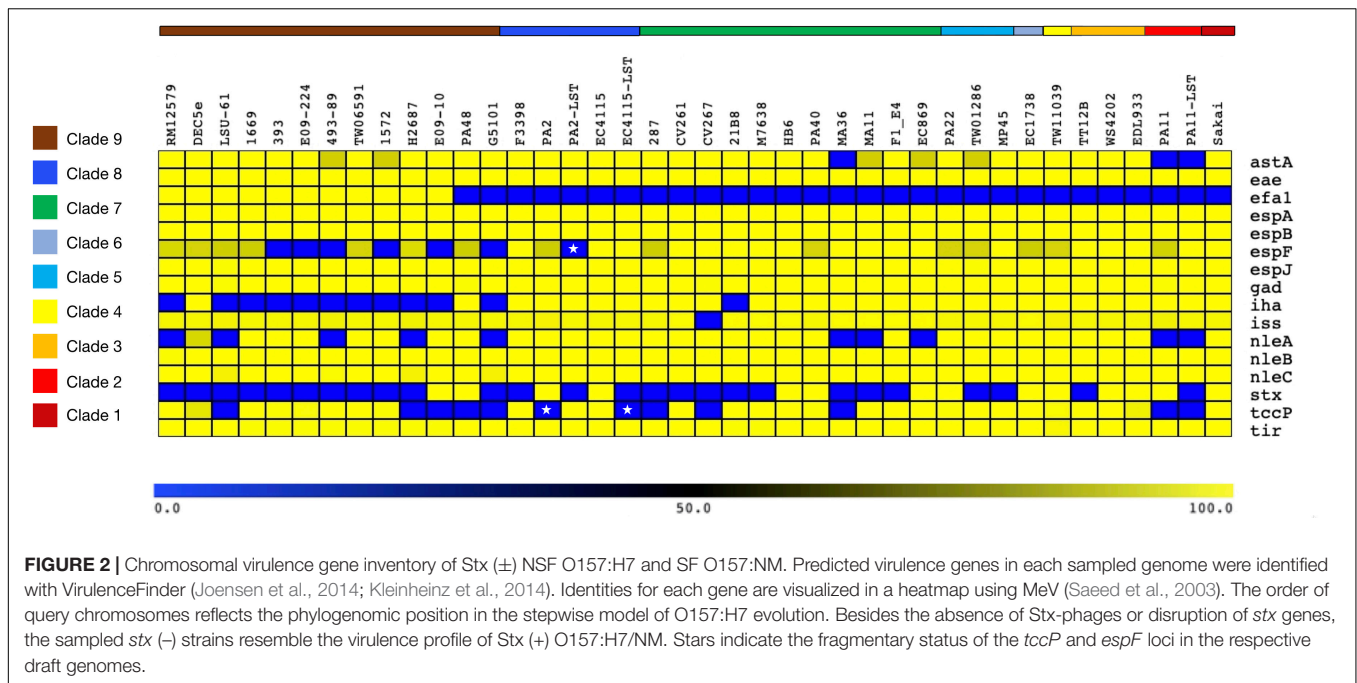
### Pathogenome Architecture and Virulence Profiles of *Stx* (–) NSF O157:H7 and SF O157:NM

Alterations in genome size and architecture in EHEC O157:H7/NM are driven by the individual prophage complement, particularly *Stx*-phages (Asadulghani et al., 2009; Eppinger et al., 2011b; Rusconi et al., 2016; Shaaban et al., 2016) and dynamics of other MGEs (Eppinger et al., 2011b; Yokoyama et al., 2011; Stanton et al., 2014; Toro et al., 2015). We compared the chromosomes of the *Stx* (–) NSF O157:H7 and SF O157:NM strains with BRIG (Alikhan et al., 2011) using *Stx* (+) strain EC4115 (*stx1*–, *stx2a*+, *stx2c*+) as reference (Eppinger et al., 2011b). As evident in **Figure 1**, we observed an overall genome-wide synteny of the chromosomal backbones of *Stx* (+) and (–) strains as has been previously established for this lineage (Eppinger et al., 2011b). The strains belong to six distinct phylogenetic clades and the comparison highlights the



spectrum of genomic variations that can be found in this lineage (Manning et al., 2008). The predicted chromosomal virulence and resistance gene complement was inferred from closed genomes (Figure 2). None of the Stx (-) strains sequenced for this study encoded antibiotic resistance genes. The absence of *stx* determined by PCR-prescreen was confirmed *in silico* for all strains but HB6, which features an altered *stx2c* locus disrupted by IS insertion (Strauch et al., 2008) (Supplementary Figure S1). Apart from the absence of *stx*, these Stx (-) NSF O157:H7 and SF O157:NM strains carried the full repertoire of chromosomal and plasmid virulence determinants found in Stx (+) representatives of this lineage (Levine, 1987; Law, 2000). The *wt* progenitors

and laboratory-cured LST strain pairs are indifferent in their virulence profiles. We note that *tccP*, a phage borne virulence determinant (Garmendia et al., 2004) and *espF* are present in all three *wt*/LST strain pairs, but were found fragmented on several contigs in the respective draft genome. All strains were positive for the locus of enterocyte effacement (LEE) pathogenicity Island (PI) when testing for the presence of *eae* (Yu and Kaper, 1992; Donnenberg et al., 1993; Kaper, 1998) and negative for bundle-forming pilus (*bfp*-), a fimbrial adhesin (Giron et al., 1991) common in EPEC strains. Although the *eae* gene is a virulence trait shared with EPEC, which can carry a variety of *eae* alleles, all Stx (-) NSF O157:H7 and SF O157:NM strains carried



the characteristic  $\gamma$ -intimin allele (McGraw et al., 1999; Blanco et al., 2006) and LEE PI (Perna et al., 1998; Ferdous et al., 2015). Thus, these strains are either progenitors that gave rise to EHEC O157:H7/NM upon acquisition of Stx-phages (Kyle et al., 2012; Schutz et al., 2017), or alternatively are O157:H7/NM strains, which secondarily lost *stx* (Friedrich et al., 2007; Bielaszewska et al., 2008).

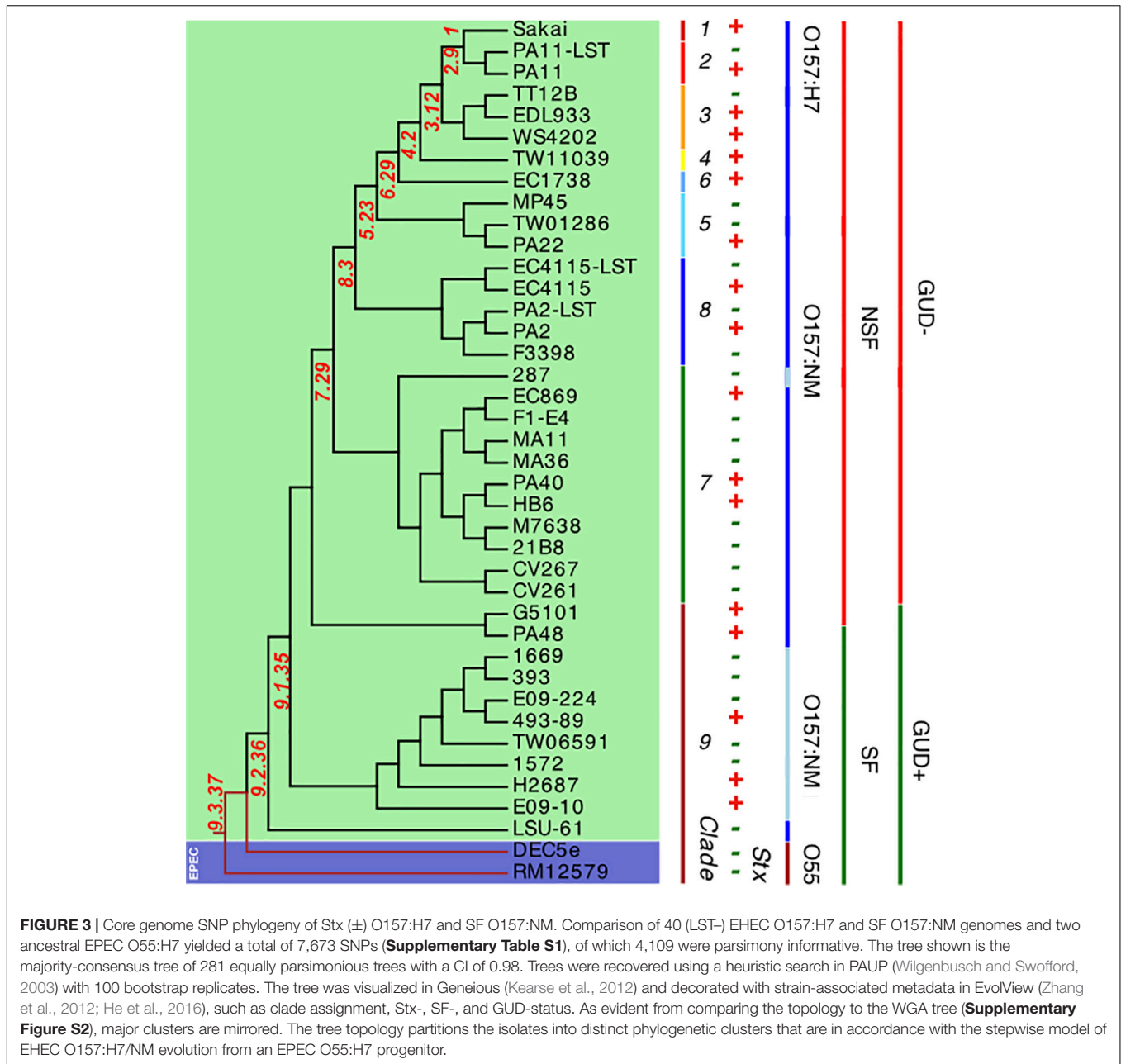
### Comparative Phylogenomics

*In silico* clade typing (Manning et al., 2008) assigned the Stx (–) NSF O157:H7 and SF O157:NM strains into five distinct clades (3, 5, 7, 8, and 9), indicative of evolutionary independent *stx* loss (Lacher, 2011) (Supplementary Table S1). To place the strains in the broader context of the O157:H7 stepwise evolutionary model (Feng et al., 1998; Wick et al., 2005; Leopold et al., 2010; Rump et al., 2011; Jenke et al., 2012; Kyle et al., 2012), we established a robust phylogenomic framework by constructing phylogenetic hypotheses based on WGA and *de novo* SNP discovery including representative *stx*-positive NSF O157:H7 and SF O157:NM strains (Eppinger et al., 2011b; Rusconi et al., 2016). The phylogenies show that the NM-phenotypes arose through two independent evolutionary events (Figure 3 and Supplementary Figure S2). In all analyzed SF O157:NM clade 9 isolates the NM-phenotype (Feng et al., 1996; Barkocy-Gallagher et al., 2001) is caused by a 12-bp deletion in *flhC*, the master regulator of flagellar biosynthesis (Monday et al., 2004). However, the non-motile NSF O157:NM clade 7 strain EPEC\_287 (Ferdous et al., 2015) did not have this particular mutation. Instead, it features a strain-specific nsSNP in the flagellar hook length control gene *fliK* at position 616 (CCG > TCG) resulting in a (206; P > S) transition from non-polar proline to polar serine (Supplementary Table S3). Such a P > S transition, yet at another genic position, has previously been linked to altered flagellar

protein secretion (Uchida et al., 2016). Further structural analysis of FliK in *E. coli* including diverse bacteria suggests functional conservation at this particular position (Marchler-Bauer et al., 2017). To determine the genetic relationships of the individual isolates at a higher level of phylogenetic accuracy and resolution, we performed a cgSNP analysis using a custom developed SNP discovery and validation pipeline described in detail in our previous works (Eppinger et al., 2010, 2011b; Rusconi et al., 2016; Hau et al., 2018). Comparison of the chromosomes yielded a total of 7,673 high-quality SNPs, of which 4,109 were parsimony informative. The resulting maximum parsimony (MP) tree using PAUP (v4a163) (Wilgenbusch and Swofford, 2003) with 100 bootstrap replicates shows bootstrap supports greater than 90 for the majority of nodes (Figure 3). As evident in both the WGA- and SNP-phylogenies, the tree topology corroborates with the phylogenetic clade placement and mirrors the general understanding of the stepwise evolutionary model of O157:H7 from an EPEC O55:H7 progenitor (Feng et al., 1998; Wick et al., 2005; Leopold et al., 2009, 2010; Rump et al., 2011; Jenke et al., 2012; Kyle et al., 2012). The phylogenomic position of LSU-61 is consistent with its proposed intermediate status between O55:H7 and the divergent NSF O157:H7 and SF O157:NM branches.

### Comprehensive Analysis of Stx-Phage Occupation Status in Stx (±) SF O157:NM and NSF O157:H7

Long-read technology provided complete prophage sequences and chromosomal context and was instrumental for accurate Stx-phage profiling (Smith et al., 2012; Yin et al., 2015; Shaaban et al., 2016; Gonzalez-Escalona et al., 2019) (Figure 1). Several Stx-phages integration sites have been established in O157:H7 (Kruger and Lucchesi, 2015). Stx<sub>2a</sub>-phages preferentially target



the NAD(P)H dehydrogenase *wrbA* and tRNA *argW* genes, both of which can be simultaneously occupied (Eppinger et al., 2011b; Rusconi et al., 2016). Stx<sub>1</sub>- and Stx<sub>2c</sub>-prophages preferentially target the NADH quinone oxidoreductase *yehV* (Yokoyama et al., 2000) and transcriptional regulator *sbcB* (Strauch et al., 2008), respectively (Figure 1), though other insertion sites are known. The boundaries and location of Stx-phages were identified by PHASTER (Arndt et al., 2016), unbiased from documented EHEC phages insertions sites, and loci comprehensively analyzed and visualized in Easyfig (Sullivan et al., 2011). To represent the evolved and ancestral Stx-phage insertion states as per the stepwise model of evolution (Feng et al., 1998; Wick et al., 2005; Leopold et al., 2010; Rump et al., 2011; Jenke et al., 2012;

Kyle et al., 2012), we used *stx*-positive strains EC4115 (*stx2a* at *argW*, *stx2c* at *sbcB*) (Eppinger et al., 2011b) and EDL933 (*stx1* at *yehV*) (Perna et al., 2001) and *stx*-negative EPEC O55:H7 strain RM12579 (Kyle et al., 2012) as references. As shown in Supplementary Figures S3–S6, a fragment corresponding to these insertion loci, extended by 2 kb on each side, was compared by BLASTn (Altschul et al., 1990). If a locus is found unoccupied in Stx (–) NSF O157:H7 or O157:NM strains, it may be indicative of secondary loss of the Stx-phage or *stx* locus (Schmidt et al., 1999; Feng et al., 2001; Koitabashi et al., 2006; Wetzel and Lejeune, 2007; Bettelheim, 2008; Rump et al., 2011; Jenke et al., 2012), or alternatively this insertion site was never targeted by Stx-phages. BLASTn comparisons of the complete

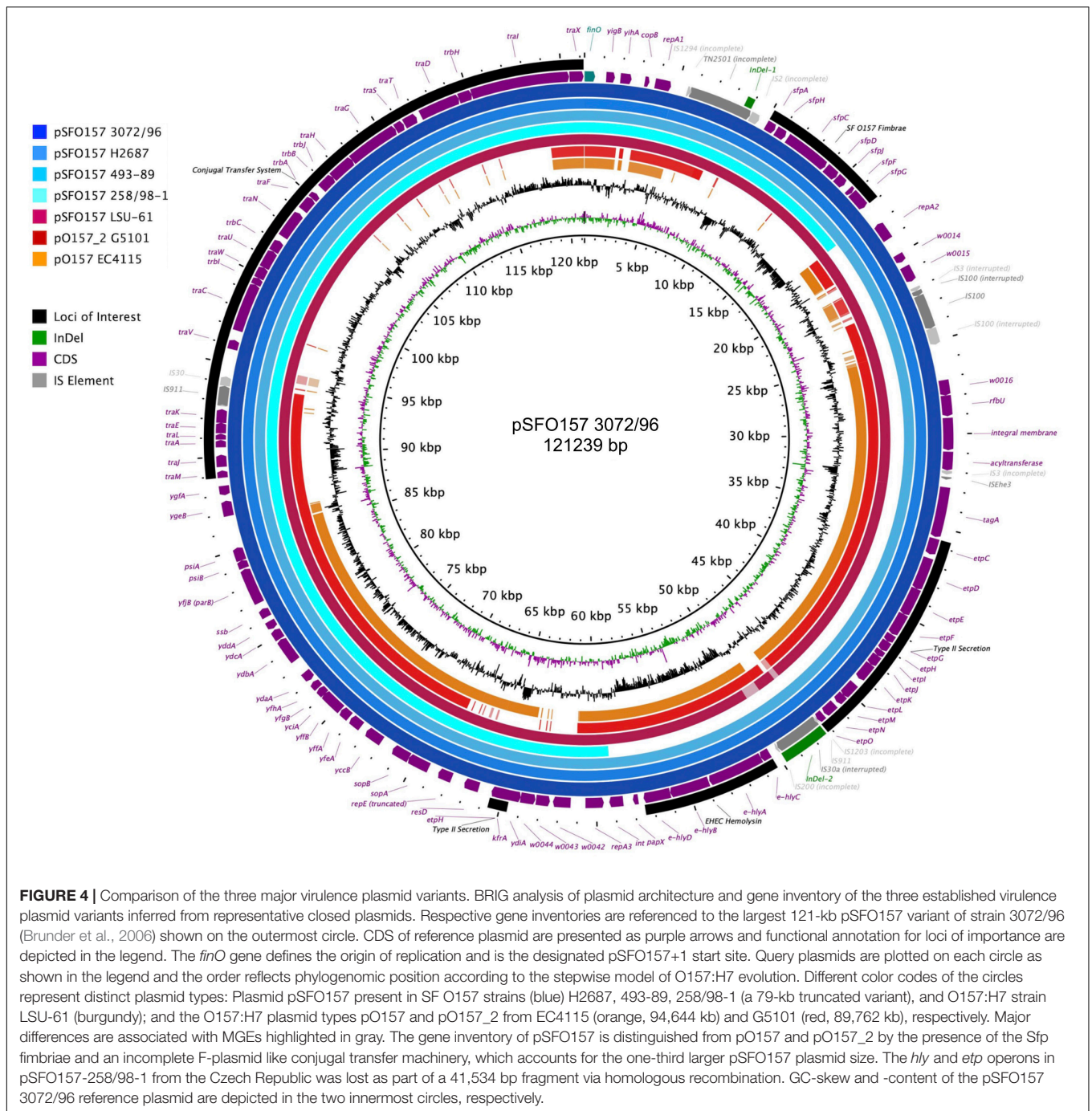


Stx-phage genomes visualized in Easyfig (Sullivan et al., 2011) show largely syntenic and conserved Stx-phage architectures. As previously mentioned, strain HB6 showed a false negative *stx* PCR reaction. Genomic analysis revealed that *stx* in HB6 is disrupted by insertion sequence IS629, which affected the binding of the generic *stx* PCR primer, and probably renders the *stx* gene non-functional (**Supplementary Figure S1**) (Strauch et al., 2008). As evident in **Supplementary Figures S3–S6** all Stx (–) strains examined are devoid of complete Stx-phages. The laboratory-cured LST strains showed complete loss of Stx-phages at respective Stx-phage insertion sites when compared to the *wt* strains [EC4115: *stx2a* (*argW*), *stx2c* (*sbcB*); PA2: *stx2a* (*argW*); PA11: *stx2a* (*wrbA*)]. The *sbcB* locus (**Supplementary Figure S4**) of strain CV261 is occupied by a phage showing homology to the 5' and 3' regions of the EC4115 Stx<sub>2c</sub>-phage, which may suggest a deletion of the *stx*-containing region, as observed in the 2006 Spinach outbreak isolates (Eppinger et al., 2011b). By contrast, phages were detected in the majority of sampled strains occupying the *yehV* locus (**Supplementary Figure S5**). Comparison to the reference Stx<sub>1</sub>-phage of EDL933 shows partial homology but also extended dissimilar regions. It is thus unclear whether these phages are related to Stx-phages that lost *stx* or are part of the *stx*-negative phage complement of O157:H7/NM. Our profiling of phage insertion loci in closed genomes of Stx (–) strains allowed us to identify different and mechanistically unrelated scenarios, from complete absence of Stx-phages to a more confined loss or disruption of the *stx* locus by IS629 (Eppinger et al., 2011b). This element plays a major role in shaping the STEC population structure (Ooka et al., 2009; Yokoyama et al., 2011; Stanton et al., 2014; Toro et al., 2015) including Stx-phage diversification (Eppinger et al., 2011b; Sanjar et al., 2015; Yin et al., 2015; Rusconi et al., 2016).

## Plasmid Inventory of Stx (–) SF O157:NM and NSF O157:H7 Strains

Virulence plasmids play an important role in O157:H7/NM pathogenicity (Johnson and Nolan, 2009; Lim et al., 2010a,b). LRT sequencing enabled us to close plasmids for a detailed comparison of architectures and gene inventories (Gonzalez-Escalona et al., 2019). In the majority (8/12) of Stx (–) strains we detected plasmids (**Supplementary Table S2**). The lineage-specific pO157 plasmid (Burland et al., 1998) was present in seven strains while the larger variant pSFO157, characteristic for SF O157:NM (Brunner et al., 2006; Rump et al., 2012), was found in NSF O157:H7 strain LSU-61 (Dunn et al., 2004) (**Figure 4**). The latter also carries a small 37-kb plasmid with partial homology to pEC4115, described by our group as a distinguishing feature of the 2006 spinach outbreak isolates from Maine (Eppinger et al., 2011b). To support the comprehensive analysis of recovered plasmids in the context of O157:H7/NM evolution (Feng et al., 1998; Wick et al., 2005; Leopold et al., 2010; Rump et al., 2011; Jenke et al., 2012; Kyle et al., 2012), we included genomes of respective plasmid types downloaded from NCBI GenBank, focusing on closed plasmids (Hayashi et al., 2001; Perna et al., 2001; Brunner et al., 2006; Kulasekara et al., 2009; Eppinger et al., 2011b; Rump et al., 2012; Latif et al., 2014;

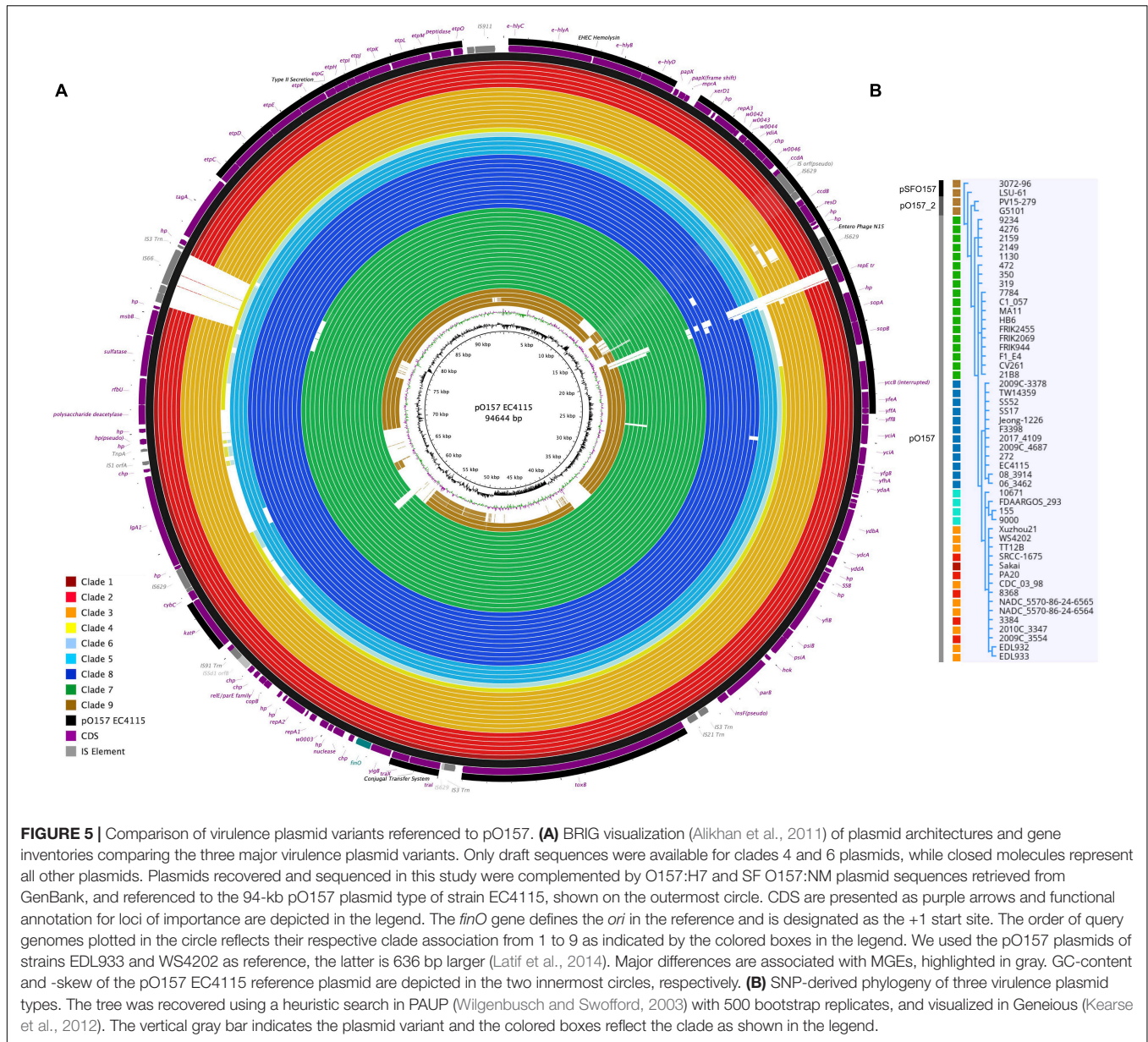
Ferdous et al., 2015; Cowley et al., 2016; Fellner et al., 2016; Bauwens et al., 2017). Altogether, the analyzed plasmid set represents the phylogenetic diversity that can be found in O157:H7 as delineated from the clade assignment that evaluates plasticity only in chromosomal markers (Manning et al., 2008) (**Supplementary Table S1**). The three major plasmid types were individually analyzed by BLASTn/BLASTp (Camacho et al., 2009; Cock et al., 2015) and visualized in BRIG (Alikhan et al., 2011). The largest 121-kb pSFO157 plasmid is characteristic of SF O157:NM (Ferdous et al., 2015) strains isolated mostly in Europe (Brunner et al., 2006; Eklund et al., 2006; Friedrich et al., 2007; Buvens et al., 2009). Further, in strain G5101 (GUD+, s7/ST-65), we found the 94-kb pO157 plasmid of O157:H7 (Burland et al., 1998; Makino et al., 1998; Lim et al., 2010b; Eppinger et al., 2011b) and a distinct, slightly smaller 89,762-kb variant, termed pO157\_2, which is highly related to the other two but not ancestral to pO157 (Rump et al., 2012) (**Figures 4–6**). We further detected a larger variant (+4,629 kb) of pO157\_2 in strain PV15-279 (Ogura et al., 2018), which is larger than the one in G5101 due to the insertion of three mobile elements (IS66, and 2× copies of IS629), while G5101 codes for an additional transposase (**Supplementary Figure S7C**). Virulence determinants shared by these plasmids include enterohemolysin (*ehxA*) (Schmidt et al., 1995; Bielaszewska et al., 2013) and a type II secretion system (Bielaszewska et al., 2009) encoded by the *hlyCABD* and *etp* operons, respectively (**Figures 4–6**). Both of the latter operons were absent in pSFO157 of strain 258/98-1 from Czechia, as it was truncated (79 kb) due to a 41,534-bp deletion caused by homologous recombination, resulting in the loss of *hly* and *etp* operons (Bauwens et al., 2017). Consistent with the fact that SFO157 strains do not produce enterohemolysin. Complete pSFO157 sequences have been published from SF O157:NM strains isolated from Germany, Czechia, and Scotland, and include strains 3072/96 (Brunner et al., 2006), 258/98-1 (Bauwens et al., 2017), 493-89 (Karch et al., 1993; Rump et al., 2011), and H2687 (Rump et al., 2011), though these strains are prevalent in other European countries (Eklund et al., 2006; Friedrich et al., 2007; Buvens et al., 2009). Here we compare the architecture and content of these European-sourced pSFO157 plasmids to the sole North American representative detected in strain LSU-61. This plasmid was determined at 119,161 bp, encoding for 153 CDS. When compared to the 3072/96 pSFO157 reference plasmid (Brunner et al., 2006), we identified two InDel regions that account for the plasmid size difference of 2,078 bp, both of which are associated with IS elements (**Supplementary Figure S7**). Similar to the role of phages in microevolution and emergence of sublineages (Ooka et al., 2009; Eppinger et al., 2011b; Yokoyama et al., 2011; Stanton et al., 2014; Sanjar et al., 2015; Toro et al., 2015; Yin et al., 2015; Rusconi et al., 2016), mobile elements are also major drivers of plasmid diversification and are responsible for the observed differences in pSFO157 plasmid size. InDel-1 is located within the boundaries of transposable element Tn2501 (Michiels et al., 1987) present in the LSU-61 and 3072/96 plasmids, though the LSU-61 variant is 400 bp larger (**Supplementary Figure S7A**). The likely intact 3,815 bp transposon of LSU-61 shares a 100% coverage and 99.5% sequence similarity to respective loci on bovine *E. coli* plasmids



**FIGURE 4 |** Comparison of the three major virulence plasmid variants. BRIG analysis of plasmid architecture and gene inventory of the three established virulence plasmid variants inferred from representative closed plasmids. Respective gene inventories are referenced to the largest 121-kb pSFO157 variant of strain 3072/96 (Brunner et al., 2006) shown on the outermost circle. CDS of reference plasmid are presented as purple arrows and functional annotation for loci of importance are depicted in the legend. The *finO* gene defines the origin of replication and is the designated pSFO157+1 start site. Query plasmids are plotted on each circle as shown in the legend and the order reflects phylogenomic position according to the stepwise model of O157:H7 evolution. Different color codes of the circles represent distinct plasmid types: Plasmid pSFO157 present in SF O157 strains (blue) H2687, 493-89, 258/98-1 (a 79-kb truncated variant), and O157:H7 strain LSU-61 (burgundy); and the O157:H7 plasmid types pO157 and pO157\_2 from EC4115 (orange, 94,644 kb) and G5101 (red, 89,762 kb), respectively. Major differences are associated with MGEs highlighted in gray. The gene inventory of pSFO157 is distinguished from pO157 and pO157\_2 by the presence of the Sfp fimbriae and an incomplete F-plasmid like conjugal transfer machinery, which accounts for the one-third larger pSFO157 plasmid size. The *hly* and *etp* operons in pSFO157-258/98-1 from the Czech Republic was lost as part of a 41,534 bp fragment via homologous recombination. GC-skew and -content of the pSFO157 3072/96 reference plasmid are depicted in the two innermost circles, respectively.

of strains GB089-pCFSAN004181P (CP012499) and a serogroup O168 strain 09-00049-pCFSAN004180G (CP012500), while in 3072/96 this InDel results in an N-terminal truncated transposase and recombinase. Altogether this suggests an ancestral status of LSU-61 and secondary decay in the pSFO157 plasmids from 3072/96, 493/89, and H2687, provided additional evidence for its intermediate status (Feng et al., 1998). To the contrary, LSU-61 lacks a nested mobile element of 2,080 bp comprised of IS30/IS911 (Supplementary Figure S7B). Plasticity in the transfer machinery accounts for the observed variation in

plasmid size of pSFO157, pO157, and pO157\_2 (Fernandez-Lopez et al., 2016). The pSFO157 plasmid found in SF O157:NM strains resembles most closely the conjugal transfer region of the F-plasmid (Frost et al., 1994), with deletions in the conjugal transfer machinery (*traP*, *trbD* *trbG*, and *traV*) mediated by IS insertions. All plasmids share the F-plasmid leading region (*yccB-parB*). However, compared to the ancestral pSFO157 state, both pO157 and pO157\_2 secondarily lost parts of the extensive transfer machinery and flanking regions, yet to a different extent: pO157 lacks the entire transfer region between *parB*

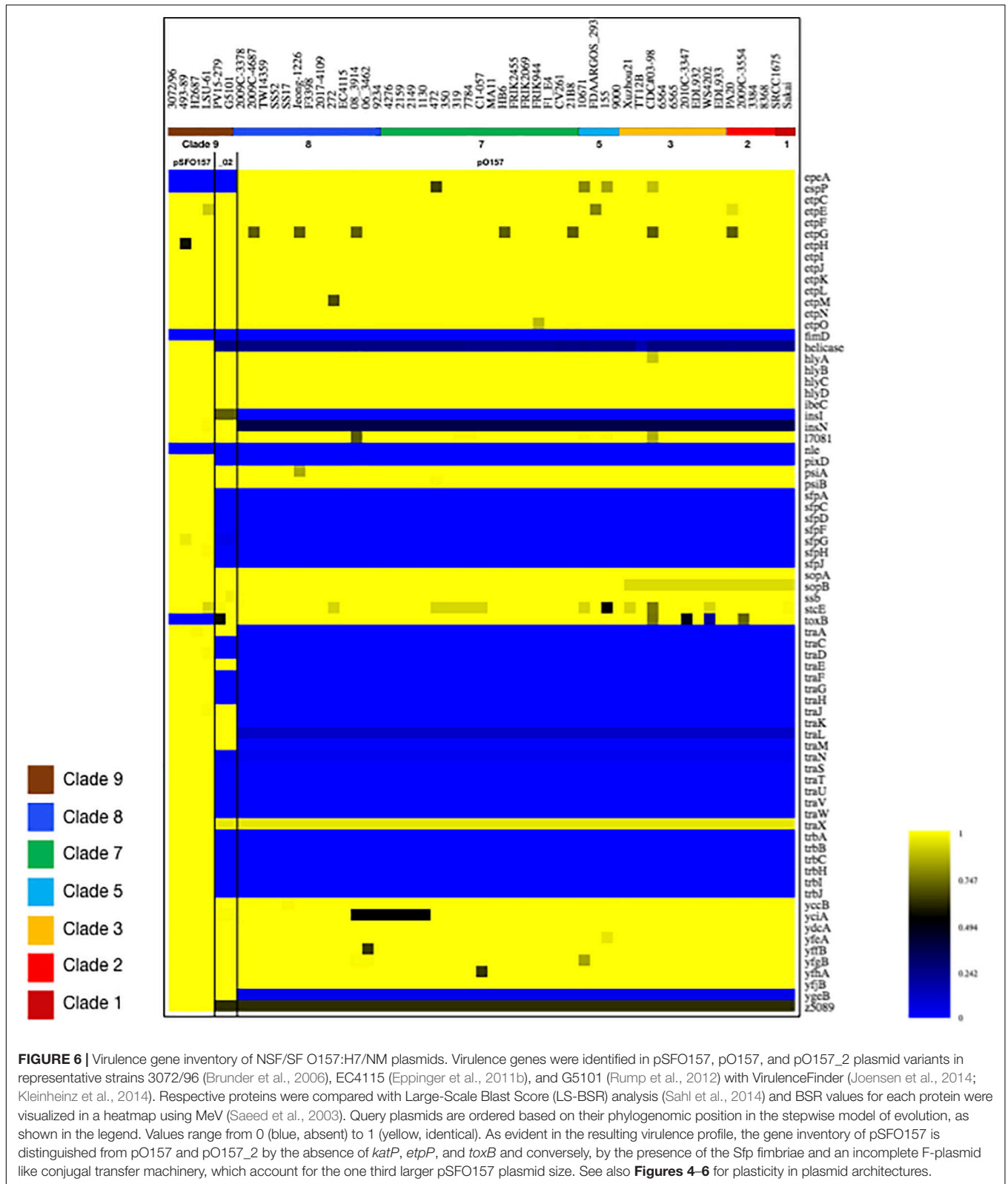


and *traX*, while pO157\_2 retains transfer genes (*traM-K*). The corresponding region in both pO157 and pO157\_2 carries *tox*B (Tatsuno et al., 2001) nested between IS elements (IS3, IS21, and IS629) (Figure 6). Both pO157/pO157\_2 also feature a truncated *traI* locus neighboring *traX*. Plasmid pSFO157 is further distinguished from pO157 and pO157\_2 by the *Sfp* fimbriae encoded by the *sfpAHCDJFG* operon (Brunner et al., 2001), which functions as adhesin in SFO157 (Musken et al., 2008) (Figure 6) and is also present in O165:H25/NM strains (Bielaszewska et al., 2009). In contrast, pSFO157 lacks the catalase peroxidase KatP (Brunner et al., 1996) and the serine protease EspP (Brunner et al., 1997), present in both pO157\_2 and pO157. In the truncated pSFO157 variant 258/98-1 the deletion of a 41,534 bp region via homologous recombination resulted in the loss of *hly* and *etp* operons (Bauwens et al., 2017). Referenced based SNP

discovery using SF O157:NM strain 3072/96 (Brunner et al., 2006) identified a total of 38 SNPs within pSFO157-type plasmids. LSU-61 features 34 strain-specific SNPs on its plasmid, which corroborates with its distinct phylogenetic position (Figure 3 and Supplementary Figure S2). The majority of SNPs previously reported by Rump et al. (2012) were confirmed, though three were rejected during QC (Supplementary Table S6).

### Correlation of Detected Plasmid Genotypes and Stepwise Evolutionary Model

The application of LRT sequencing allowed us to accurately assemble six pO157- and one pSFO157-type plasmid to closure that complement the relatively few closed O157:H7/NM

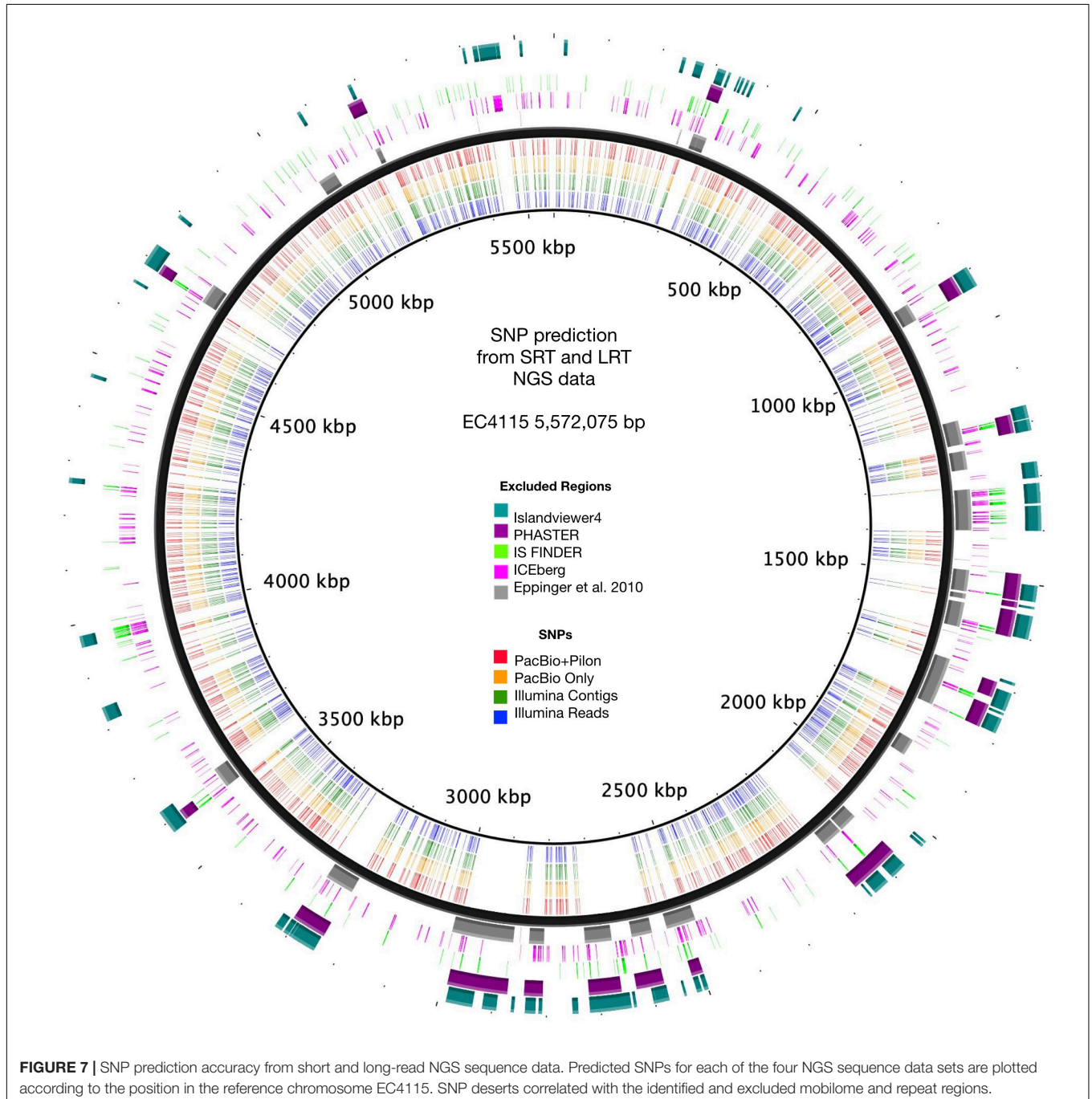


virulence plasmids deposited in NCBI GenBank. This sequence base opened an avenue to investigate whether pSFO157 and pO157 plasmids and the chromosomes evolve in parallel and

how the individual plasmid genotypes relate to the stepwise model of evolution (Feng et al., 1998; Wick et al., 2005; Leopold et al., 2010; Rump et al., 2011; Jenke et al., 2012;

Kyle et al., 2012). This plasmid set now encompasses a wide phylogenetic space as delineated from chromosomal based clade assignment (Manning et al., 2008) (**Supplementary Table S1**). In **Figure 5**, we plotted the closed plasmid types according to their phylogenetic position in the stepwise emergence of O157:H7 (**Supplementary Table S1**). As evident in **Figure 5**, we observed a stable evolutionary relationship between the host chromosome delineated clade and respective carried plasmid genotype. We note here that although these plasmids carry different portions of the conjugal transfer machinery, neither

type is considered to be transferable on its own due to the absence of a functional *traB* gene. This locus is critical for F pilus assembly and is either truncated (41%) in pSFO157 and pO157\_2 or absent in pO157 (Kim et al., 1993; Fernandez-Lopez et al., 2016). Comprehensive analysis of these on average 94-kb plasmids (**Figure 5**) showed that InDels are mostly associated with mobile insertion elements, but also other loci dispersed throughout the plasmid (e.g., pseudogene *insF-hp*), some of which are clade-defining signatures in plasmids (e.g., clade 3). IS elements have been utilized for chromosomal and also Stx-phage

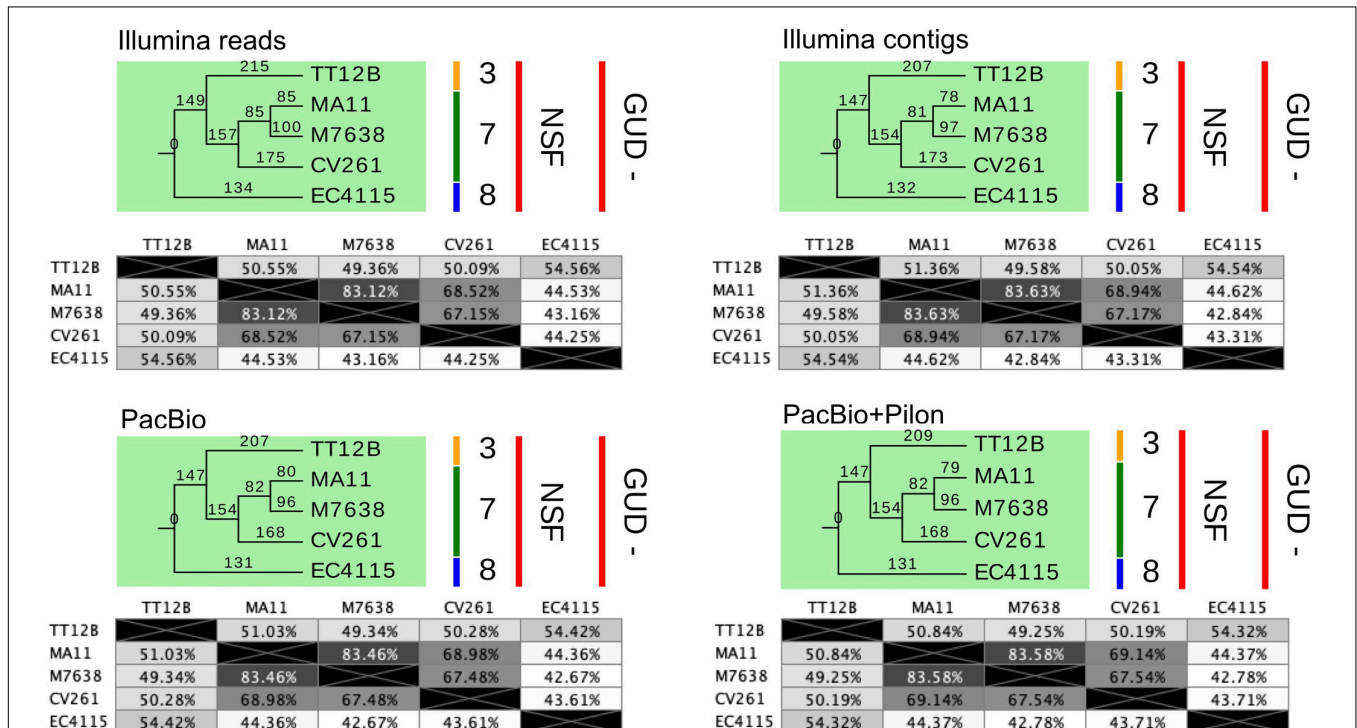


subtyping in O157:H7 (Eppinger et al., 2011b; Yokoyama et al., 2011; Stanton et al., 2014; Toro et al., 2015). The mosaic-like composition of the large pO157 virulence plasmid types, as reflected by carriage of a number of replication-associated genes and mobile genetic elements driving plasmid diversification, clearly suggests complex evolutionary origins resulting in distinct coding capacity and architectures. In this study, we identified ancestral and derived plasmid characteristics that opens the avenue for refined assay development in support of established chromosomal typing schemas (Rusconi and Eppinger, 2016). The pSFO157, pO157, and pO157\_2 plasmids are related, though the latter was found to be non-ancestral to pO157 (Rump et al., 2012) (Figures 4, 5). To further elucidate the genetic relationships of the three distinct pSFO157, pO157, and pO157\_2 variants, we performed SNP discovery in 54 representative plasmids. The tree was inferred from 87 SNPs, of which 37 are parsimony informative (Supplementary Table S7). The resulting topology places the plasmid variants at distinct phylogenetic positions, and further clearly shows that the overall plasmids cluster in accordance to their chromosomally inferred clade type (Figure 5B). Some of these SNP signatures were found to be clade specific (Supplementary Table S7). Altogether, our genotypic data from this limited sample suggests a stable relationship between the bacterial chromosome and carried virulence plasmids in NSF O157:H7 and SF O157:NM strains. We

also investigated the prevalence, gene inventory and structural organization of pEC4115 by BRIG comparison (Alikhan et al., 2011) to related *Escherichia* and *Salmonella* plasmids, as determined by BLASTn-inferred nucleotide sequence similarities (Altschul et al., 1990). We detected this plasmid in LSU-61 and strains (FDAARGOS293, FRIK2455, SS17, 144 and 155) (Supplementary Figure S8). This relatively small 37,452 bp plasmid was sequenced previously by our group and identified as a distinguishing feature of the 2006 O157:H7 spinach outbreak isolates from Maine (Eppinger et al., 2011b). Noteworthy, it carries a number of conjugal transfer genes; however, these genes are unrelated to the *tra* machinery on pSFO157, and to remnant systems in pO157 and pO157\_2.

### Comparison of Accuracy From NGS Short- and Long-Read Technology

As evident from our analysis of the Stx-occupation status, closed genomes are key for detailed comparison of structural and genetic polymorphisms (Shaaban et al., 2016). We further used the Illumina and PacBio data set to investigate the SNP accuracy as inferred from NGS short- and long-read data. Using read- and contig-based workflows in Galaxy for reference-based SNP discovery against EC4115 (clade 8.32) (Rusconi et al., 2016), we comprehensively analyzed SNP prediction accuracy in clade



**FIGURE 8 |** SNP statistics from NGS short- and long-read sequence data. The trees shown are the majority-consensus trees of equally parsimonious trees inferred from all four NGS sequence data sets. Trees were recovered using a heuristic search in PAUP (Wilgenbusch and Swofford, 2003) with 500 bootstrap replicates and each of the computed trees features an overall CI value > =0.997. The numbers in the pair-wise distance matrices represent percent identity of bases, which are identical. The trees were visualized in Geneious (Kearse et al., 2012) and decorated with strain-associated metadata in EvolView (Zhang et al., 2012; He et al., 2016), such as clade assignment and metabolic properties. As evident when comparing the tree topologies and statistics, all NGS datasets provide a robust basis for phylogenetic studies and carry the identical phylogenetic information.

7.29 strains CV261, MA11, MA7638, and clade 3.12 strain TT12B (Feng et al., 2001). The SNP discovery was performed on four NGS data sets: Illumina reads, Illumina assemblies, PacBio assemblies, and in addition Illumina error-corrected PacBio assemblies in Pilon (**Supplementary Table S1**) (Walker et al., 2014). The total number of predicted unique SNPs inferred from NGS short- and long-read sequence data is 1117 with a range between 1064 (PacBio assembly) and 1096 (Illumina reads) with a standard deviation of 12.97 (**Supplementary Table S4** and **Figure 7**). The effect on coding capacity of predicted SNPs is summarized in **Supplementary Table S5**. The Pilon error-corrected PacBio assemblies show only minimal changes when compared to the PacBio only assemblies with a deviation in predicted SNP numbers of 1.41. As evidenced in **Figure 8**, all four analyzed NGS datasets provide a robust basis for phylogenetic studies and all trees transport the identical phylogenetic information. The CI index can be an indication of a highly accurate tree and was examined for each individual SNP in the four NGS datasets (Eppinger et al., 2014). The CI for SNP prediction from Illumina and error-corrected PacBio assemblies is 1. As shown in **Supplementary Table S4**, a single SNP in the PacBio only assembly has a CI < 1, while three SNPs with CI < 1 were recorded from discovery in Illumina reads. Among the latter, three SNPs were called in Illumina reads only, thus the remaining 40 SNPs only predicted from Illumina reads are considered high-quality SNPs with a CI of 1. Overall, our data suggest that while Illumina read data can provide additional and phylogenetically relevant SNPs, error correction of long-reads by Illumina reads is optional.

## CONCLUSION

Taken our findings together, this study highlights the utility of LRT for advancing our understanding of the EHEC O157:H7/NM pathogenome evolution. The generated long-read data was critical to put these atypical Stx (–) strain into the phylogenomic context of the stepwise evolutionary model for O157:H7/NM (Feng et al., 1998; Wick et al., 2005; Leopold et al., 2009, 2010; Jenke et al., 2012; Kyle et al., 2012). Availability of closed genomes recovered through LRT allowed us to describe the underlying genomic basis and evolutionary scenarios for the absence, acquisition and loss of *stx*. Analysis of the closed virulence plasmids showed a strong correlation between plasmid genotypes and chromosomally inferred clades (Manning et al., 2008), which may indicate coevolution of the chromosome and carried plasmids. Accessory plasmids may get lost during laboratory cultivation or are often only recovered in fragments by SRT. Typing efforts are thus often focused on stable chromosomal markers (Yang et al., 2004; Manning et al., 2008); however, our data suggest that plasmid information is valuable to complement chromosomal markers and further refine the stepwise model. Insight into the strain-to-strain plasticity in the pathogenome and pathotypes of Stx (–) and (+) NSF/SF O157:H7/NM strains is foundational to improve risk assessment, biosurveillance and prevention strategies (Eppinger and Cebula, 2015; Werber and Scheutz, 2019).

## DATA AVAILABILITY STATEMENT

The sequence data sets generated and analyzed in this study have been deposited in the Short Read Archive (SRA) and the Whole Genome Shotgun Repository at NCBI. Accessions for reads, annotated genomes and plasmids together with strain-associated metadata are provided in **Supplementary Tables S1, S2**.

## ETHICS STATEMENT

An ethical review process for the bacterial strains sequenced and analyzed in this study was not required. The animal strains 21B8, F1-E4, HB6, MP45, CV261, CV267, and M7638, sequenced in this study, were all taken at facilities that are privately owned and permission to collect freshly deposited bovine-, respectively, swine feces, from the pen surfaces was granted by the owners. Isolation history and strain-associated metadata for strains that were (re-) sequenced for this study, such as TT12B and LSU-61, as well as the wt strains from which the Stx-cured strains EC4115-LST, PA2-LST, and PA11-LST are derived, have been previously published.

## AUTHOR CONTRIBUTIONS

ME conceived and designed the experiments. EN, SZ, AA-G, ZI-B, SK, PF, JB, and ME analyzed the data. PF contributed strain material. JB provided PacBio sequence data. AR provided computational support. ME and SK wrote the manuscript.

## FUNDING

This work was supported by the South Texas Center for Emerging Infectious Diseases (STCEID) and the Department of Biology at the University of Texas at San Antonio; and received funding through the National Institute of Allergy and Infectious Diseases of the National Institutes of Health [SC2AI120941]; and the United States Department of Homeland Security [2014-ST-062-000058] to ME. The contributions of JB were funded by the Agricultural Research Service of the United States Department of Agriculture (ARS-USDA). 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 United States Meat Animal Research Center is an equal opportunity employer and provider.

## ACKNOWLEDGMENTS

This work received computational support from UTSA's HPC cluster Shamu, operated by the Office of Information Technology. We would like to thank Sandy Fryda-Bradley (USDA) and Kristin Kuhn (USDA) for providing technical assistance.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.00619/full#supplementary-material>

**FIGURE S1** | Comparison of Stx<sub>2c</sub>-prophages in strains EC4115 and HB6. The boundaries of the Stx<sub>2c</sub>-phage inserted at the *sbcB* locus were identified in strain HB6 using PHASTER (Arndt et al., 2016). Phage architecture and gene inventories were compared to the corresponding locus in Stx<sub>2c</sub>-positive strain EC4115 (Eppinger et al., 2011b) by BLASTn and visualized in Easyfig (Sullivan et al., 2011). Comparison revealed disruption of the *stx* locus in HB6 by IS629 insertion. Arrows represent predicted CDSs and gray shaded regions connect homologous sequences. Core chromosome genes flanking the phage are colored in teal. Functional annotations of predicted phage genes are shown in the legend.

**FIGURE S2** | Whole genome phylogeny of Stx (±) NSF O157:H7 and SF O157:NM. Genomes of 40 Stx (–) NSF O157:H7 and SF O157:NM and two progenitor EPEC O55:H7 strains, 15 of which were sequenced for this study, were aligned with Mugsy (Angiuoli and Salzberg, 2011). The phylogenetic tree with representative strains for all nine established phylogenetic clades was inferred using RAxML with a 100 bootstrap replicates (Rump et al., 2011). The majority consensus tree was visualized in Geneious (Kearse et al., 2012) and decorated with strain-associated metadata in Evolview (Zhang et al., 2012; He et al., 2016). The tree topology partitions the isolates into distinct phylogenetic clusters that corroborate with the stepwise model of O157:H7 evolution from an EPEC O55:H7 progenitor. Stx (–) strains examined clustered with NSF O157:H7 and SF O157:NM, indicative of their evolutionary independent Stx-phage or *stx* loss.

**FIGURE S3** | Comparison of Stx-phage occupation status at the *wrbA* locus. Easyfig comparison (Sullivan et al., 2011) of *wrbA*, a preferred target locus for Stx<sub>2a</sub>-phage insertion. A fragment extended by 2 kb on each side of corresponding *wrbA* loci was extracted and compared by BLASTn (Altschul et al., 1990). Comparison of the reference Stx<sub>2a</sub>-phage locus in strain EDL933 (Perna et al., 2001) and sampled Stx (–) O157:H7 strains reveals unoccupied *wrbA* loci. Arrows represent predicted CDSs and gray shaded regions connect homologous sequences. Core chromosome genes flanking the phage are colored in teal. Functional annotations of predicted phage genes are shown in the legend.

**FIGURE S4** | Comparison of Stx-phage occupation status at the *sbcB* locus. Easyfig comparison (Sullivan et al., 2011) of *sbcB*, a preferred target locus for Stx<sub>2c</sub>-phage insertion. A fragment extended by 2 kb on each side of the corresponding *sbcB* loci was extracted and compared by BLASTn (Altschul et al., 1990). Comparison between the Stx<sub>2c</sub>-phage locus in EC4115 (Eppinger et al., 2011b) and sampled Stx (–) strains reveals unoccupied *sbcB* loci with the notable exception of strain CV261 that carried phage remnants. Arrows represent predicted CDSs and gray shaded regions connect homologous sequences. Core chromosome genes flanking the phage are colored in teal. Functional annotations of predicted phage genes are shown in the legend.

**FIGURE S5** | Comparison of Stx-prophage occupation status at the *yehV* locus. Easyfig comparison (Sullivan et al., 2011) of *yehV*, a preferred target locus for Stx<sub>1</sub>-phage insertion. A fragment extended by 2 kb on each side of the corresponding *yehV* loci was extracted and compared by BLASTn (Altschul et al., 1990). Comparison between the reference Stx<sub>1</sub>-phage locus in EDL933

(Perna et al., 2001) and sampled Stx (–) strains reveals prophage sequences at *yehV* for the majority of strains, while this locus is unoccupied in TT12B and LSU-61. Arrows represent predicted CDSs and gray shaded regions connect homologous sequences. Core chromosome genes flanking the phage are colored in teal. Functional annotations of predicted phage genes are shown in the legend.

**FIGURE S6** | Comparison of Stx-prophage occupation status at the *argW* locus. Easyfig comparison (Sullivan et al., 2011) of *argW*, a preferred target locus for Stx<sub>2a</sub>-phage insertion. A fragment extended by 2 kb on each side of the corresponding *argW* loci was extracted and compared by BLASTn (Altschul et al., 1990). Comparison between the reference Stx<sub>2a</sub>-phage locus in EC4115 (Eppinger et al., 2011b) and sampled Stx (–) strains reveals unoccupied *argW* loci. Arrows represent predicted CDSs and gray shaded regions connect homologous sequences. Core chromosome genes flanking the phage are colored in teal. Functional annotations of predicted phage genes are shown in the legend.

**FIGURE S7** | Length polymorphisms in plasmid pSFO157 of LSU-61 and 3072/96. Plasmid architecture and gene inventories were compared with Mauve (Darling et al., 2010), and respective annotations mapped in Geneious (Maddison and Maddison, 2016). IS elements are a major driver and hotspots for pSFO157 and pO157\_2 plasmid diversification. The pSFO157 plasmid of strains LSU-61 and 3072/96 differs by 2,078 bp, (A) InDel-1 is located within the boundaries of a Tn2501 element resembling the 400 bp larger ancestral variant in LSU-61. (B) To the contrary, LSU-61 lacks a nested composite IS30/IS911 element of 2,080 bp. (C) InDels in pO157\_2 of strains G5101 and PV15-279, all of which are associated with insertion sequences.

**FIGURE S8** | Comparison of pEC4115 with phylogenetically related plasmids. BRIG analysis of the plasmid architecture and gene inventory of pEC4115 and related *Escherichia* and *Salmonella* plasmids, as determined by BLASTn inferred nucleotide sequence similarities (Altschul et al., 1990). Respective gene inventories are referenced to the 37,452 bp pEC4115 plasmid in the outermost circle. CDS are shown as purple arrows and the functional annotation of loci of importance are depicted in the legend. The order of plasmids on each ring reflects the sequence similarity of query plasmids as reported by BLASTn and circle colors identify plasmid source as shown in the legend. Depicted plasmids show a considerable range in plasmid size from 34,276 bp (*E. coli* O103 FH121) to 65,419 bp (*Salmonella enterica*). Only plasmids from SS17, and human fecal isolates O181:H49 2014C-3097 and O103 FH121 show homology to the pEC4115 origin of replication. Noteworthy, the pEC4115 plasmid also codes conjugal transfer genes. GC-skew and GC-content of the pEC4115 reference plasmid are depicted in the two innermost circles, respectively.

**TABLE S1** | Strain associated metadata of isolates investigated in this study.

**TABLE S2** | Plasmids and associated metadata of isolates investigated in this study.

**TABLE S3** | Core genome SNPs in investigated Stx (±) NSF O157:H7 and SF O157:NM.

**TABLE S4** | SNP predictions from NGS short and long-read sequence data.

**TABLE S5** | SNP statistics from NGS short and long-read sequence data.

**TABLE S6** | SNPs in pSFO157-variant plasmids referenced to strain 3072/96.

**TABLE S7** | SNPs in pO157-variant plasmids referenced to strain EC4115.

## REFERENCES

- Alikhan, N. F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genom.* 12:402. doi: 10.1186/1471-2164-12-402
- Allue-Guardia, A., Echazarreta, M., Koenig, S. S. K., Klose, K. E., and Eppinger, M. (2018a). Closed genome sequence of *Vibrio cholerae* O1 El tor inaba strain A1552. *Microbiol. Resour. Anounc.* 6:e00098-18. doi: 10.1128/genomeA.00098-18
- Allue-Guardia, A., Nyong, E. C., Koenig, S. S., Vargas, S. M., Bono, J., and Eppinger, M. (2018b). Closed genome sequence of *Escherichia coli* K-12 group strain C600. *Microbiol. Resour. Anounc.* 2:e1052-18. doi: 10.1128/MRA.01052-18
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Amorim, J. H., Del Cogliano, M. E., Fernandez-Brando, R. J., Bilen, M. F., Jesus, M. R., Luiz, W. B., et al. (2014). Role of bacteriophages in STEC infections: new implications for the design of prophylactic and treatment approaches. *F1000Res.* 3:74. doi: 10.12688/f1000research.3718.2
- Andrew, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed October 6, 2010).



- Angiuoli, S. V., and Salzberg, S. L. (2011). Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* 27, 334–342. doi: 10.1093/bioinformatics/btq665
- Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387
- Asadulghani, M., Ogura, Y., Ooka, T., Itoh, T., Sawaguchi, A., Iguchi, A., et al. (2009). The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *PLoS Pathog.* 5:e1000408. doi: 10.1371/journal.ppat.1000408
- Ashton, P. M., Perry, N., Ellis, R., Petrovska, L., Wain, J., Grant, K. A., et al. (2015). Insight into Shiga toxin genes encoded by *Escherichia coli* O157 from whole genome sequencing. *PeerJ* 3:e739. doi: 10.7717/peerj.739
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Barkocy-Gallagher, G. A., Arthur, T. M., Siragusa, G. R., Keen, J. E., Elder, R. O., Laegreid, W. W., et al. (2001). Genotypic analyses of *Escherichia coli* O157:H7 and O157 nonmotile isolates recovered from beef cattle and carcasses at processing plants in the Midwestern states of the United States. *Appl. Environ. Microbiol.* 67, 3810–3818. doi: 10.1128/aem.67.9.3810-3818.2001
- Bauwens, A., Marejkova, M., Middendorf-Bauchart, B., Prager, R., Kossow, A., Zhang, W., et al. (2017). Sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:H(-) isolates from czech patients with novel plasmid composition not previously seen in german isolates. *Appl. Environ. Microbiol.* 83:e01454-17. doi: 10.1128/AEM.01454-17
- Bentancor, A., Vilte, D. A., Rumi, M. V., Carbonari, C. C., Chinen, I., Larzabal, M., et al. (2010). Characterization of non-Shiga toxin-producing *Escherichia coli* O157 strains isolated from dogs. *Revis. Argent. Microbiol.* 42, 46–48.
- Bentancor, L. V., Bilen, M. F., Mejias, M. P., Fernandez-Brando, R. J., Panek, C. A., Ramos, M. V., et al. (2013a). Functional capacity of Shiga-toxin promoter sequences in eukaryotic cells. *PLoS One* 8:e57128. doi: 10.1371/journal.pone.0057128
- Bentancor, L. V., Mejias, M. P., Pinto, A., Bilen, M. F., Meiss, R., Rodriguez-Galan, M. C., et al. (2013b). Promoter sequence of Shiga toxin 2 (Stx2) is recognized in vivo, leading to production of biologically active Stx2. *mBio* 4:e00501-13. doi: 10.1128/mBio.00501-13
- Berger, P., Kouzel, I. U., Berger, M., Haarmann, N., Dobrindt, U., Koudelka, G. B., et al. (2019). Carriage of Shiga toxin phage profoundly affects *Escherichia coli* gene expression and carbon source utilization. *BMC Genom.* 20:504. doi: 10.1186/s12864-019-5892-x
- Bertelli, C., and Brinkman, F. S. L. (2018). Improved genomic island predictions with IslandPath-DIMOB. *Bioinformatics* 34, 2161–2167. doi: 10.1093/bioinformatics/bty095
- Bertelli, C., Laird, M. R., Williams, K. P., Simon Fraser University Research Computing Group, Lau, B. Y., Hoad, G., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45, W30–W35. doi: 10.1093/nar/gkx343
- Bertelli, C., Tilley, K. E., and Brinkman, F. S. L. (2018). Microbial genomic island discovery, visualization and analysis. *Brief Bioinform.* 20, 1685–1698. doi: 10.1093/bib/bby042
- Besser, T. E., Shaikh, N., Holt, N. J., Tarr, P. I., Konkel, M. E., Malik-Kale, P., et al. (2007). 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.* 73, 671–679. doi: 10.1128/AEM.01035-06
- Bettelheim, K. (2008). Re: Isolation of *Escherichia coli* O157:H7 strains that do not produce Shiga toxin from bovine, avian and environmental sources. *Lett. Appl. Microbiol.* 45, 504–507. doi: 10.1111/j.1472-765X.2007.02300.x
- Bielaszewska, M., Middendorf, B., Kock, R., Friedrich, A. W., Fruth, A., Karch, H., et al. (2008). Shiga toxin-negative attaching and effacing *Escherichia coli*: distinct clinical associations with bacterial phylogeny and virulence traits and inferred in-host pathogen evolution. *Clin. Infect. Dis.* 47, 208–217. doi: 10.1086/589245
- Bielaszewska, M., Prager, R., Kock, R., Mellmann, A., Zhang, W., Tschape, H., et al. (2007). Shiga toxin gene loss and transfer in vitro and in vivo during enterohemorrhagic *Escherichia coli* O26 infection in humans. *Appl. Environ. Microbiol.* 73, 3144–3150. doi: 10.1128/AEM.02937-06
- Bielaszewska, M., Prager, R., Vandivinit, L., Musken, A., Mellmann, A., Holt, N. J., et al. (2009). Detection and characterization of the fimbrial sfp cluster in enterohemorrhagic *Escherichia coli* O165:H25/NM isolates from humans and cattle. *Appl. Environ. Microbiol.* 75, 64–71. doi: 10.1128/AEM.01815-08
- Bielaszewska, M., Ruter, C., Kunsmann, L., Greune, L., Bauwens, A., Zhang, W., et al. (2013). Enterohemorrhagic *Escherichia coli* hemolysin employs outer membrane vesicles to target mitochondria and cause endothelial and epithelial apoptosis. *PLoS Pathog.* 9:e1003797. doi: 10.1371/journal.ppat.1003797
- Blanco, M., Blanco, J. E., Dahbi, G., Mora, A., Alonso, M. P., Varela, G., et al. (2006). Typing of intimin (eae) genes from enteropathogenic *Escherichia coli* (EPEC) isolated from children with diarrhoea in Montevideo, Uruguay: identification of two novel intimin variants (muB and xiR/beta2B). *J. Med. Microbiol.* 55(Pt 9), 1165–1174. doi: 10.1099/jmm.0.46518-0
- Bono, J. L. (2009). Genotyping *Escherichia coli* O157:H7 for its ability to cause disease in humans. *Curr. Protoc. Microbiol.* Chapter 5, Unit 5A 3. doi: 10.1002/9780471729259.mc05a03s14
- Brunder, W., Karch, H., and Schmidt, H. (2006). Complete sequence of the large virulence plasmid pSFO157 of the sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:H- strain 3072/96. *Int. J. Med. Microbiol.* 296, 467–474. doi: 10.1016/j.jimm.2006.05.005
- Brunder, W., Khan, A. S., Hacker, J., and Karch, H. (2001). Novel type of fimbriae encoded by the large plasmid of sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:H(-). *Infect. Immun.* 69, 4447–4457. doi: 10.1128/IAI.69.7.4447-4457.2001
- Brunder, W., Schmidt, H., and Karch, H. (1996). KatP, a novel catalase-peroxidase encoded by the large plasmid of enterohaemorrhagic *Escherichia coli* O157:H7. *Microbiology* 142(Pt 11), 3305–3315. doi: 10.1099/13500872-142-11-3305
- Brunder, W., Schmidt, H., and Karch, H. (1997). EspP, a novel extracellular serine protease of enterohaemorrhagic *Escherichia coli* O157:H7 cleaves human coagulation factor V. *Mol. Microbiol.* 24, 767–778. doi: 10.1046/j.1365-2958.1997.3871751.x
- Burland, V., Shao, Y., Perna, N. T., Plunkett, G., Sofia, H. J., and Blattner, F. R. (1998). The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. *Nucleic Acids Res.* 26, 4196–4204. doi: 10.1093/nar/26.18.4196
- Buvsen, G., Pierard, D., Hachimi-Idrissi, S., and Lauwers, S. (2009). First sorbitol-fermenting verocytotoxin-producing *Escherichia coli* O157: H- isolated in Belgium. *Acta Clin. Belg.* 64, 59–64. doi: 10.1179/acb.2009.011
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Carattoli, A., Zankari, E., Garcia-Fernandez, A., Voldby Larsen, M., Lund, O., Villa, L., et al. (2014). In silico detection and typing of plasmids using Plasmid Finder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903. doi: 10.1128/AAC.02412-14
- Chen, L., Zheng, D., Liu, B., Yang, J., and Jin, Q. (2016). VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* 44, D694–D697. doi: 10.1093/nar/gkv1239
- Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10, 563–569. doi: 10.1038/nmeth.2474
- Cock, P. J. A., Chilton, J. M., Gruning, B., Johnson, J. E., and Soranzo, N. (2015). NCBI BLAST plus integrated into galaxy. *Gigascience* 4:39. doi: 10.1186/s13742-015-0080-7
- Cooper, K. K., Mandrell, R. E., Louie, J. W., Korlach, J., Clark, T. A., Parker, C. T., et al. (2014). Comparative genomics of enterohemorrhagic *Escherichia coli* O145:H28 demonstrates a common evolutionary lineage with *Escherichia coli* O157:H7. *BMC Genomics* 15:17. doi: 10.1186/1471-2164-15-17
- Cowley, L. A., Dallman, T. J., Fitzgerald, S., Irvine, N., Rooney, P. J., McAteer, S. P., et al. (2016). Short-term evolution of Shiga toxin-producing *Escherichia coli* O157:H7 between two food-borne outbreaks. *Microbial. Genomics* 2:e000084. doi: 10.1099/mgen.0.000084
- Dallman, T. J., Ashton, P. M., Byrne, L., Perry, N. T., Petrovska, L., Ellis, R. J., et al. (2015). Applying phylogenomics to understand the emergence of Shiga-toxin-producing *Escherichia coli* O157:H7 strains causing severe human disease in the UK. *Microbial. Genomics* 1:e000029. doi: 10.1099/mgen.0.000029

- Darling, A. E., Mau, B., and Perna, N. T. (2010). Progressive mauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi: 10.1371/journal.pone.0011147.s005
- Delcher, A. L., Salzberg, S. L., and Phillippy, A. M. (2003). Using MUMmer to identify similar regions in large sequence sets. *Curr. Protoc. Bioinformatics* Chapter 10, Unit 10.3. doi: 10.1002/0471250953.bi1003s00
- Donnenberg, M. S., Tzipori, S., McKee, M. L., O'Brien, A. D., Alroy, J., and Kaper, J. B. (1993). The role of the *eae* gene of enterohemorrhagic *Escherichia coli* in intimate attachment in vitro and in a porcine model. *J. Clin. Invest.* 92, 1418–1424. doi: 10.1172/JCI116718
- Dunn, J. R., Keen, J. E., Moreland, D., and Alex, T. (2004). Prevalence of *Escherichia coli* O157:H7 in white-tailed deer from Louisiana. *J. Wildl. Dis.* 40, 361–365. doi: 10.7589/0090-3558-40.2.361
- Eaton, K. A., Friedman, D. I., Francis, G. J., Tyler, J. S., Young, V. B., Haeger, J., et al. (2008). Pathogenesis of renal disease due to enterohemorrhagic *Escherichia coli* in germ-free mice. *Infect. Immun.* 76, 3054–3063. doi: 10.1128/IAI.01626-07
- Eklund, M., Bielaszewska, M., Nakari, U. M., Karch, H., and Siitonen, A. (2006). Molecular and phenotypic profiling of sorbitol-fermenting *Escherichia coli* O157:H- human isolates from Finland. *Clin. Microbiol. Infect.* 12, 634–641. doi: 10.1111/j.1469-0691.2006.01478.x
- Eklund, M., Leino, K., and Siitonen, A. (2002). Clinical *Escherichia coli* strains carrying *stx* genes: *stx* variants and *stx*-positive virulence profiles. *J. Clin. Microbiol.* 40, 4585–4593. doi: 10.1128/jcm.40.12.4585-4593.2002
- English, A. C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., et al. (2012). Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* 7:e47768. doi: 10.1371/journal.pone.0047768
- Eppinger, M., and Cebula, T. A. (2015). Future perspectives, applications and challenges of genomic epidemiology studies for food-borne pathogens: a case study of Enterohemorrhagic *Escherichia coli* (EHEC) of the O157:H7 serotype. *Gut Microbes* 6, 194–201. doi: 10.4161/19490976.2014.969979
- Eppinger, M., Daugherty, S., Agrawal, S., Galens, K., Sengamalay, N., Sadzewicz, L., et al. (2013). Whole-genome draft sequences of 26 enterohemorrhagic *Escherichia coli* O157:H7 Strains. *Genome Announc* 1:e0013412. doi: 10.1128/genomeA.00134-12
- Eppinger, M., Mammel, M. K., Leclerc, J. E., Ravel, J., and Cebula, T. A. (2011b). Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20142–20147. doi: 10.1073/pnas.1107176108
- Eppinger, M., Mammel, M. K., Leclerc, J. E., Ravel, J., and Cebula, T. A. (2011a). Genome signatures of *Escherichia coli* O157:H7 isolates from the bovine host reservoir. *Appl. Environ. Microbiol.* 77, 2916–2925. doi: 10.1128/AEM.02554-10
- Eppinger, M., Pearson, T., Koenig, S. S., Pearson, O., Hicks, N., Agrawal, S., et al. (2014). Genomic epidemiology of the Haitian cholera outbreak: a single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. *mBio* 5:e01721. doi: 10.1128/mBio.01721-14
- Eppinger, M., Worsham, P. L., Nikolich, M. P., Riley, D. R., Sebastian, Y., Mou, S., et al. (2010). Genome sequence of the deep-rooted *Yersinia pestis* strain Angola reveals new insights into the evolution and pangene of the plague bacterium. *J. Bacteriol.* 192, 1685–1699. doi: 10.1128/JB.01518-09
- Fellner, L., Huptas, C., Simon, S., Muhlig, A., Scherer, S., and Neuhaus, K. (2016). Draft genome sequences of three European laboratory derivatives from enterohemorrhagic *Escherichia coli* O157:H7 strain EDL933, including two plasmids. *Genome Announc.* 4:e1331-15. doi: 10.1128/genomeA.01331-15
- Feng, P., Dey, M., Abe, A., and Takeda, T. (2001). Isogenic strain of *Escherichia coli* O157:H7 that has lost both Shiga toxin 1 and 2 genes. *Clin. Diagn. Lab. Immunol.* 8, 711–717. doi: 10.1128/CDLI.8.4.711-717.2001
- Feng, P., Fields, P. I., Swaminathan, B., and Whittam, T. S. (1996). Characterization of nonmotile variants of *Escherichia coli* O157 and other serotypes by using an anti-flagellin monoclonal antibody. *J. Clin. Microbiol.* 34, 2856–2859. doi: 10.1128/jcm.34.11.2856-2859.1996
- Feng, P., and Lampel, K. A. (1994). Genetic analysis of *uidA* expression in enterohaemorrhagic *Escherichia coli* serotype O157:H7. *Microbiology* 140(Pt 8), 2101–2107. doi: 10.1099/13500872-140-8-2101
- Feng, P., Lampel, K. A., Karch, H., and Whittam, T. S. (1998). Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. *J. Infect. Dis.* 177, 1750–1753.
- Feng, P. C., Monday, S. R., Lacher, D. W., Allison, L., Siitonen, A., Keys, C., et al. (2007). Genetic diversity among clonal lineages within *Escherichia coli* O157:H7 stepwise evolutionary model. *Emerg. Infect. Dis.* 13, 1701–1706. doi: 10.3201/eid1311.070381
- Ferdous, M., Zhou, K., Mellmann, A., Morabito, S., Croughs, P. D., de Boer, R. F., et al. (2015). Is shiga toxin-negative *Escherichia coli* O157:H7 enteropathogenic or enterohemorrhagic *Escherichia coli*? Comprehensive molecular analysis using whole-genome sequencing. *J. Clin. Microbiol.* 53, 3530–3538. doi: 10.1128/JCM.01899-15
- Ferens, W. A., and Hovde, C. J. (2011). *Escherichia coli* O157:H7: animal reservoir and sources of human infection. *Foodborne. Pathog. Dis.* 8, 465–487. doi: 10.1089/fpd.2010.0673
- Fernandez-Lopez, R., de Toro, M., Moncalian, G., Garcillan-Barcia, M. P., and de la Cruz, F. (2016). Comparative genomics of the conjugation region of F-like plasmids: five shades of F. *Front. Mol. Biosci.* 3:71. doi: 10.3389/fmolb.2016.00071
- Friedrich, A. W., Zhang, W., Bielaszewska, M., Mellmann, A., Kock, R., Fruth, A., et al. (2007). Prevalence, virulence profiles, and clinical significance of Shiga toxin-negative variants of enterohemorrhagic *Escherichia coli* O157 infection in humans. *Clin. Infect. Dis.* 45, 39–45. doi: 10.1086/518573
- Frost, L. S., Ippen-Ihler, K., and Skurray, R. A. (1994). Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* 58, 162–210. doi: 10.1128/membr.58.2.162-210.1994
- Fuchs, S., Muhldorfer, I., Donohue-Rolfé, A., Kerenyi, M., Emody, L., Alexiev, R., et al. (1999). Influence of RecA on in vivo virulence and Shiga toxin 2 production in *Escherichia coli* pathogens. *Microb. Pathog.* 27, 13–23. doi: 10.1006/mpat.1999.0279
- Fuller, C. A., Pellino, C. A., Flagler, M. J., Strasser, J. E., and Weiss, A. A. (2011). Shiga toxin subtypes display dramatic differences in potency. *Infect. Immun.* 79, 1329–1337. doi: 10.1128/IAI.01182-10
- Gamage, S. D., Patton, A. K., Strasser, J. E., Chalk, C. L., and Weiss, A. A. (2006). Commensal bacteria influence *Escherichia coli* O157:H7 persistence and Shiga toxin production in the mouse intestine. *Infect. Immun.* 74, 1977–1983. doi: 10.1128/IAI.74.3.1977-1983.2006
- Gao, F., and Zhang, C. T. (2008). Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. *BMC Bioinformatics* 9:79. doi: 10.1186/1471-2105-9-79
- Garcia-Aljaro, C., Muniesa, M., Jofre, J., and Blanch, A. R. (2009). Genotypic and phenotypic diversity among induced, *stx2*-carrying bacteriophages from environmental *Escherichia coli* strains. *Appl. Environ. Microbiol.* 75, 329–336. doi: 10.1128/AEM.01367-08
- Garmendia, J., Phillips, A. D., Carlier, M. F., Chong, Y., Schuller, S., Marches, O., et al. (2004). TccP is an enterohaemorrhagic *Escherichia coli* O157:H7 type III effector protein that couples Tir to the actin-cytoskeleton. *Cell Microbiol.* 6, 1167–1183. doi: 10.1111/j.1462-5822.2004.00459.x
- Garrison, E., and Marth, G. (2012). *Haplotype-Based Variant Detection From Short-Read Sequencing*. *ArXiv e-prints [Online]*, 1207. Available online at: <http://adsabs.harvard.edu/abs/2012arXiv1207.3907G> (accessed July 1, 2012).
- Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., et al. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* 15, 1451–1455. doi: 10.1101/gr.4086505
- Giron, J. A., Ho, A. S., and Schoolnik, G. K. (1991). An inducible bundle-forming pilus of enteropathogenic *Escherichia coli*. *Science* 254, 710–713. doi: 10.1126/science.1683004
- Goecks, J., Nekrutenko, A., Taylor, J., and Galaxy, T. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11:R86. doi: 10.1186/gb-2010-11-8-r86
- Gonzalez-Escalona, N., Allard, M. A., Brown, E. W., Sharma, S., and Hoffmann, M. (2019). Nanopore sequencing for fast determination of plasmids, phages, virulence markers, and antimicrobial resistance genes in Shiga toxin-producing *Escherichia coli*. *PLoS One* 14:e0220494. doi: 10.1371/journal.pone.0220494
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Hau, S. J., Allue-Guardia, A., Rusconi, B., Haan, J. S., Davies, P. R., Frana, T. S., et al. (2018). 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.* 9:2078. doi: 10.3389/fmicb.2018.02078

- Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., et al. (2001). Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* 8, 11–22. doi: 10.1093/dnares/8.1.11
- He, Z., Zhang, H., Gao, S., Lercher, M. J., Chen, W. H., and Hu, S. (2016). Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res.* 44, W236–W241. doi: 10.1093/nar/gkw370
- Huang, A., Friesen, J., and Brunton, J. L. (1987). Characterization of a bacteriophage that carries the genes for production of Shiga-like toxin 1 in *Escherichia coli*. *J. Bacteriol.* 169, 4308–4312. doi: 10.1128/jb.169.9.4308-4312.1987
- Imamovic, L., and Muniesa, M. (2012). Characterizing RecA-independent induction of Shiga toxin2-encoding phages by EDTA treatment. *PLoS One* 7:e32393. doi: 10.1371/journal.pone.0032393
- Jenke, C., Leopold, S. R., Weniger, T., Rothganger, J., Harmsen, D., Karch, H., et al. (2012). Identification of intermediate in evolutionary model of enterohemorrhagic *Escherichia coli* O157. *Emerg. Infect. Dis.* 18, 582–588. doi: 10.3201/eid1804.111414
- Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510. doi: 10.1128/JCM.03617-13
- Joensen, K. G., Tetzschner, A. M., Iguchi, A., Aarestrup, F. M., and Scheutz, F. (2015). Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J. Clin. Microbiol.* 53, 2410–2426. doi: 10.1128/JCM.00008-15
- Johannes, L., and Romer, W. (2010). Shiga toxins—from cell biology to biomedical applications. *Nat. Rev. Microbiol.* 8, 105–116. doi: 10.1038/nrmicro2279
- Johnson, J. E., Kumar, P., Easterly, C., Esler, M., Mehta, S., Eschenlauer, A. C., et al. (2018). Improve your galaxy text life: the query tabular tool. *F1000Res.* 7:1604. doi: 10.12688/f1000research.16450.2
- Johnson, T. J., and Nolan, L. K. (2009). Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* 73, 750–774. doi: 10.1128/MMBR.00015-09
- Junemann, S., Sedlazeck, F. J., Prior, K., Albersmeier, A., John, U., Kalinowski, J., et al. (2013). Updating benchtop sequencing performance comparison. *Nat. Biotechnol.* 31, 294–296. doi: 10.1038/nbt.2522
- Kaper, J. B. (1998). The locus of enterocyte effacement pathogenicity island of Shiga toxin-producing *Escherichia coli* O157:H7 and other attaching and effacing *E. coli*. *Jpn. J. Med. Sci. Biol.* 51(Suppl.), S101–S107.
- Karch, H., Bohm, H., Schmidt, H., Gunzer, F., Aleksic, S., and Heesemann, J. (1993). Clonal structure and pathogenicity of Shiga-like toxin-producing, sorbitol-fermenting *Escherichia coli* O157:H. *J. Clin. Microbiol.* 31, 1200–1205. doi: 10.1128/jcm.31.5.1200-1205.1993
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199
- Kim, S. R., Funayama, N., and Komano, T. (1993). Nucleotide sequence and characterization of the traABC region of Inc11 plasmid R64. *J. Bacteriol.* 175, 5035–5042. doi: 10.1128/jb.175.16.5035-5042.1993
- Kimmit, P. T., Harwood, C. R., and Barer, M. R. (2000). Toxin gene expression by shiga toxin-producing *Escherichia coli*: the role of antibiotics and the bacterial SOS response. *Emerg. Infect. Dis.* 6, 458–465. doi: 10.3201/eid0605.000503
- King, A. J., Sundaram, S., Cendoroglo, M., Acheson, D. W., and Keusch, G. T. (1999). Shiga toxin induces superoxide production in polymorphonuclear cells with subsequent impairment of phagocytosis and responsiveness to phorbol esters. *J. Infect. Dis.* 179, 503–507. doi: 10.1086/314579
- Kleinheinz, K. A., Joensen, K. G., and Larsen, M. V. (2014). Applying the res finder and virulence finder web-services for easy identification of acquired antibiotic resistance and *E. coli* virulence genes in bacteriophage and prophage nucleotide sequences. *Bacteriophage* 4:e27943. doi: 10.4161/bact.27943
- Koch, C., Hertwig, S., and Appel, B. (2003). Nucleotide sequence of the integration site of the temperate bacteriophage 6220, which carries the Shiga toxin gene stx(1ox3). *J. Bacteriol.* 185, 6463–6466. doi: 10.1128/jb.185.21.6463-6466.2003
- Koitaishi, T., Vuddhakul, V., Radu, S., Morigaki, T., Asai, N., Nakaguchi, Y., et al. (2006). Genetic characterization of *Escherichia coli* O157: H7/- strains carrying the stx2 gene but not producing Shiga toxin 2. *Microbiol. Immunol.* 50, 135–148. doi: JST.JSTAGE/mandi/50.135 [pii]
- Konowalchuk, J., Speirs, J. I., and Stavric, S. (1977). Vero response to a cytotoxin of *Escherichia coli*. *Infect. Immun.* 18, 775–779. doi: 10.1128/iai.18.3.775-779.1977
- Koren, S., Harhay, G. P., Smith, T. P., Bono, J. L., Harhay, D. M., McVey, S. D., et al. (2013). Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biol.* 14:R101. doi: 10.1186/gb-2013-14-9-r101
- Krueger, F. (2017). *Trim Galore*. Available online at: [http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) (accessed April 11, 2019).
- Kruger, A., and Lucchesi, P. M. (2015). Shiga toxins and stx phages: highly diverse entities. *Microbiology* 161(Pt 3), 451–462. doi: 10.1099/mic.0.000003
- Kulasekara, B. R., Jacobs, M., Zhou, Y., Wu, Z., Sims, E., Saenphimmachak, C., et al. (2009). Analysis of the genome of the *Escherichia coli* O157:H7 2006 spinach-associated outbreak isolate indicates candidate genes that may enhance virulence. *Infect. Immun.* 77, 3713–3721. doi: 10.1128/IAI.00198-09
- Kyle, J. L., Cummings, C. A., Parker, C. T., Quinones, B., Vatta, P., Newton, E., et al. (2012). *Escherichia coli* serotype O55:H7 diversity supports parallel acquisition of bacteriophage at Shiga toxin phage insertion sites during evolution of the O157:H7 lineage. *J. Bacteriol.* 194, 1885–1896. doi: 10.1128/JB.00120-12
- Lacher, D. W. (2011). “The evolutionary model of *Escherichia coli* O157:H7,” in *Population Genetics of Bacteria: A Tribute to Thomas S. Whittam*, eds S. T. Walk and P. C. H. Feng (Washington, DC: ASM Press), 227–239. doi: 10.1128/9781555817114.ch13
- Lainhart, W., Stolfa, G., and Koudelka, G. B. (2009). Shiga toxin as a bacterial defense against a eukaryotic predator, *Tetrahymena thermophila*. *J. Bacteriol.* 191, 5116–5122. doi: 10.1128/JB.00508-09
- Larsen, M. V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., et al. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* 50, 1355–1361. doi: 10.1128/JCM.06094-11
- Latif, H., Li, H. J., Charusanti, P., Palsson, B. O., and Aziz, R. K. (2014). A gapless, unambiguous genome sequence of the enterohemorrhagic *Escherichia coli* O157:H7 strain EDL933. *Genome Announc* 2:e00821-14. doi: 10.1128/genomeA.00821-14
- Law, D. (2000). Virulence factors of *Escherichia coli* O157 and other Shiga toxin-producing *E. coli*. *J. Appl. Microbiol.* 88, 729–745. doi: 10.1046/j.1365-2672.2000.01031.x
- Lee, M. S., Koo, S., Jeong, D. G., and Tesh, V. L. (2016). Shiga toxins as multi-functional proteins: induction of host cellular stress responses, role in pathogenesis and therapeutic applications. *Toxins* 8:777. doi: 10.3390/toxins8030077
- Leopold, S. R., Magrini, V., Holt, N. J., Shaikh, N., Mardis, E. R., Cagno, J., et al. (2009). A precise reconstruction of the emergence and constrained radiations of *Escherichia coli* O157 portrayed by backbone concatenomic analysis. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8713–8718. doi: 10.1073/pnas.0812949106
- Leopold, S. R., Shaikh, N., and Tarr, P. I. (2010). Further evidence of constrained radiation in the evolution of pathogenic *Escherichia coli* O157:H7. *Infect. Genet. Evol.* 10, 1282–1285. doi: 10.1016/j.meegid.2010.07.021
- Levine, M. M. (1987). *Escherichia coli* that cause diarrhea – enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. *J. Infect. Dis.* 155, 377–389. doi: 10.1093/infdis/155.3.377
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Lim, J. Y., La, H. J., Sheng, H., Forney, L. J., and Hovde, C. J. (2010a). Influence of plasmid pO157 on *Escherichia coli* O157:H7 Sakai biofilm formation. *Appl. Environ. Microbiol.* 76, 963–966. doi: 10.1128/AEM.01068-09
- Lim, J. Y., Yoon, J., and Hovde, C. J. (2010b). A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. *J. Microbiol. Biotechnol.* 20, 5–14. doi: 10.4014/jmb.0908.08007
- Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., et al. (2019). ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* 47, D660–D665. doi: 10.1093/nar/gky1123
- Loman, N. J., Quick, J., and Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* 12, 733–735. doi: 10.1038/nmeth.3444
- Lorenz, S. C., Gonzalez-Escalona, N., Kotewicz, M. L., Fischer, M., and Kase, J. A. (2017). Genome sequencing and comparative genomics of enterohemorrhagic *Escherichia coli* O145:H25 and O145:H28 reveal distinct evolutionary paths

- and marked variations in traits associated with virulence & colonization. *BMC Microbiol.* 17:183. doi: 10.1186/s12866-017-1094-3
- Los, J. M., Los, M., Wegrzyn, A., and Wegrzyn, G. (2012). Altruism of Shiga toxin-producing *Escherichia coli*: recent hypothesis versus experimental results. *Front. Cell Infect. Microbiol.* 2:166. doi: 10.3389/fcimb.2012.00166
- Los, J. M., Los, M., and Wegrzyn, G. (2011). Bacteriophages carrying Shiga toxin genes: genomic variations, detection and potential treatment of pathogenic bacteria. *Future Microbiol.* 6, 909–924. doi: 10.2217/FMB.11.70
- Los, J. M., Los, M., Wegrzyn, G., and Wegrzyn, A. (2009). Differential efficiency of induction of various lambdoid prophages responsible for production of Shiga toxins in response to different induction agents. *Microb. Pathog.* 47, 289–298. doi: 10.1016/j.micpath.2009.09.006
- Luo, H., Quan, C. L., Peng, C., and Gao, F. (2018). Recent development of Ori-Finder system and DoriC database for microbial replication origins. *Brief Bioinform.* 20, 1114–1124. doi: 10.1093/bib/bbx174
- Maddison, W. P., and Maddison, D. R. (2016). *Mesquite: A Modular System for Evolutionary Analysis*. Available at: <http://mesquiteproject.org> (accessed April 11, 2019).
- Makino, K., Ishii, K., Yasunaga, T., Hattori, M., Yokoyama, K., Yutsudo, C. H., et al. (1998). Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic *Escherichia coli* O157:H7 derived from Sakai outbreak. *DNA Res.* 5, 1–9. doi: 10.1093/dnares/5.1.1
- Manning, S. D., Motiwala, A. S., Springman, A. C., Qi, W., Lacher, D. W., Ouellette, L. M., et al. (2008). Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4868–4873. doi: 10.1073/pnas.0710834105
- Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., and Zimin, A. (2018). MUMmer4: a fast and versatile genome alignment system. *PLoS Comput. Biol.* 14:e1005944. doi: 10.1371/journal.pcbi.1005944
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45, D200–D203. doi: 10.1093/nar/gkw1129
- Mauro, S. A., and Koudelka, G. B. (2011). Shiga toxin: expression, distribution, and its role in the environment. *Toxins (Basel)* 3, 608–625. doi: 10.3390/toxins3060608
- McGannon, C. M., Fuller, C. A., and Weiss, A. A. (2010). Different classes of antibiotics differentially influence shiga toxin production. *Antimicrob. Agents Chemother.* 54, 3790–3798. doi: 10.1128/AAC.01783-09
- McGraw, E. A., Li, J., Selander, R. K., and Whittam, T. S. (1999). Molecular evolution and mosaic structure of alpha, beta, and gamma intimins of pathogenic *Escherichia coli*. *Mol. Biol. Evol.* 16, 12–22. doi: 10.1093/oxfordjournals.molbev.a026032
- Mead, P. S., and Griffin, P. M. (1998). *Escherichia coli* O157:H7. *Lancet* 352, 1207–1212. doi: 10.1016/S0140-6736(98)01267-7
- Mead, P. S., Slutsker, L., Dietz, V., McCaig, L. F., Bresee, J. S., Shapiro, C., et al. (1999). Food-related illness and death in the United States. *Emerg. Infect. Dis.* 5, 607–625. doi: 10.3201/eid0505.990502
- Mellmann, A., Bielaszewska, M., Zimmerhackl, L. B., Prager, R., Harmsen, D., Tschape, H., et al. (2005). Enterohemorrhagic *Escherichia coli* in human infection: in vivo evolution of a bacterial pathogen. *Clin. Infect. Dis.* 41, 785–792. doi: 10.1086/432722
- Mellmann, A., Lu, S., Karch, H., Xu, J. G., Harmsen, D., Schmidt, M. A., et al. (2008). Recycling of Shiga toxin 2 genes in sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:NM. *Appl. Environ. Microbiol.* 74, 67–72. doi: 10.1128/AEM.01906-07
- Melton-Celsa, A., Mohawk, K., Teel, L., and O'Brien, A. (2012). Pathogenesis of Shiga-toxin producing *Escherichia coli*. *Curr. Top. Microbiol. Immunol.* 357, 67–103. doi: 10.1007/82\_2011\_176
- Melton-Celsa, A. R. (2014). Shiga toxin (Stx) classification, structure, and function. *Microbiol. Spectr.* 2:EHEC–0024–2013. doi: 10.1128/microbiolspec.EHEC-0024-2013
- Michiels, T., Cornelis, G., Ellis, K., and Grinstead, J. (1987). Tn2501, a component of the lactose transposon Tn951, is an example of a new category of class II transposable elements. *J. Bacteriol.* 169, 624–631. doi: 10.1128/jb.169.2.624-631.1987
- Mikheenko, A., Pribelski, A., Saveliev, V., Antipov, D., and Gurevich, A. (2018). Versatile genome assembly evaluation with QUASt-LG. *Bioinformatics* 34, i142–i150. doi: 10.1093/bioinformatics/bty266
- Miyamoto, H., Nakai, W., Yajima, N., Fujibayashi, A., Higuchi, T., Sato, K., et al. (1999). Sequence analysis of Stx2-converting phage VT2-Sa shows a great divergence in early regulation and replication regions. *DNA Res.* 6, 235–240. doi: 10.1093/dnares/6.4.235
- Monday, S. R., Minnich, S. A., and Feng, P. C. H. (2004). A 12-base-pair deletion in the flagellar master control gene *flhC* causes nonmotility of the pathogenic German sorbitol-fermenting *Escherichia coli* O157 : H- strains. *J. Bacteriol.* 186, 2319–2327. doi: 10.1128/Jb.186.8.2319-2327.2004
- Monday, S. R., Whittam, T. S., and Feng, P. C. (2001). Genetic and evolutionary analysis of mutations in the *gusA* gene that cause the absence of beta-glucuronidase activity in *Escherichia coli* O157:H7. *J. Infect. Dis.* 184, 918–921. doi: 10.1086/323154
- Morelli, G., Song, Y., Mazzoni, C. J., Eppinger, M., Roumagnac, P., Wagner, D. M., et al. (2010). *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat. Genet.* 42, 1140–1143. doi: 10.1038/ng.705
- Muniesa, M., Blanco, J. E., De Simon, M., Serra-Moreno, R., Blanch, A. R., and Jofre, J. (2004). Diversity of *stx2* converting bacteriophages induced from Shiga-toxin-producing *Escherichia coli* strains isolated from cattle. *Microbiology* 150(Pt 9), 2959–2971. doi: 10.1099/mic.0.27188-0
- Muniesa, M., de Simon, M., Prats, G., Ferrer, D., Panella, H., and Jofre, J. (2003). Shiga toxin 2-converting bacteriophages associated with clonal variability in *Escherichia coli* O157:H7 strains of human origin isolated from a single outbreak. *Infect. Immun.* 71, 4554–4562. doi: 10.1128/iai.71.8.4554-4562.2003
- Musken, A., Bielaszewska, M., Greune, L., Schweppe, C. H., Muthing, J., Schmidt, H., et al. (2008). Anaerobic conditions promote expression of Sfp fimbriae and adherence of sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:NM to human intestinal epithelial cells. *Appl. Environ. Microbiol.* 74, 1087–1093. doi: 10.1128/AEM.02496-07
- Myers, G. S., Mathews, S. A., Eppinger, M., Mitchell, C., O'Brien, K. K., White, O. R., et al. (2009). Evidence that human *Chlamydia pneumoniae* was zoonotically acquired. *J. Bacteriol.* 191, 7225–7233. doi: 10.1128/JB.00746-09
- Nguyen, Y., and Sperandio, V. (2012). Enterohemorrhagic *E. coli* (EHEC) pathogenesis. *Front. Cell Infect. Microbiol.* 2:90. doi: 10.3389/fcimb.2012.00090
- O'Brien, A. D., Lively, T. A., Chang, T. W., and Gorbach, S. L. (1983). Purification of *Shigella dysenteriae* 1 (Shiga)-like toxin from *Escherichia coli* O157:H7 strain associated with haemorrhagic colitis. *Lancet* 2:573. doi: 10.1016/s0140-6736(83)90601-3
- Ogura, Y., Mondal, S. I., Islam, M. R., Mako, T., Arisawa, K., Katsura, K., et al. (2015). The Shiga toxin 2 production level in enterohemorrhagic *Escherichia coli* O157:H7 is correlated with the subtypes of toxin-encoding phage. *Sci. Rep.* 5:16663. doi: 10.1038/srep16663
- Ogura, Y., Ooka, T., Iguchi, A., Toh, H., Asadulghani, M., Oshima, K., et al. (2009). Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17939–17944. doi: 10.1073/pnas.0903585106
- Ogura, Y., Seto, K., Morimoto, Y., Nakamura, K., Sato, M. P., Gotoh, Y., et al. (2018). Genomic Characterization of beta-glucuronidase-positive *Escherichia coli* O157:H7 producing Stx2a. *Emerg. Infect. Dis.* 24, 2219–2227. doi: 10.3201/eid2412.180404
- Ohnishi, M., Kurokawa, K., and Hayashi, T. (2001). Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? *Trends Microbiol.* 9, 481–485. doi: 10.1016/s0966-842x(01)02173-4
- Ohnishi, M., Terajima, J., Kurokawa, K., Nakayama, K., Murata, T., Tamura, K., et al. (2002). Genomic diversity of enterohemorrhagic *Escherichia coli* O157 revealed by whole genome PCR scanning. *Proc. Natl. Acad. Sci. U.S.A.* 99, 17043–17048. doi: 10.1073/pnas.262441699
- Ooka, T., Ogura, Y., Asadulghani, M., Ohnishi, M., Nakayama, K., Terajima, J., et al. (2009). Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in *Escherichia coli* O157 genomes. *Genome Res.* 19, 1809–1816. doi: 10.1101/gr.089615.108
- Orth, D., Grif, K., Khan, A. B., Naim, A., Dierich, M. P., and Wurzner, R. (2007). The Shiga toxin genotype rather than the amount of Shiga toxin or

- the cytotoxicity of Shiga toxin in vitro correlates with the appearance of the hemolytic uremic syndrome. *Diagn. Microbiol. Infect. Dis.* 59, 235–242. doi: 10.1016/j.diagmicrobio.2007.04.013
- Pacheco, A. R., Lazarus, J. E., Sit, B., Schmieder, S., Lencer, W. I., Blondel, C. J., et al. (2018). CRISPR screen reveals that EHEC's T3SS and Shiga toxin rely on shared host factors for infection. *mBio* 9:e01003-18. doi: 10.1128/mBio.01003-18
- Pennington, H. (2010). *Escherichia coli* O157. *Lancet* 376, 1428–1435. doi: 10.1016/S0140-6736(10)60963-4
- Perna, N. T., Mayhew, G. F., Posfai, G., Elliott, S., Donnenberg, M. S., Kaper, J. B., et al. (1998). Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect. Immun.* 66, 3810–3817. doi: 10.1128/iai.66.8.3810-3817.1998
- Perna, N. T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., et al. (2001). Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7. *Nature* 409, 529–533. doi: 10.1038/35054089
- Poirier, K., Faucher, S. P., Beland, M., Brousseau, R., Gannon, V., Martin, C., et al. (2008). *Escherichia coli* O157:H7 survives within human macrophages: global gene expression profile and involvement of the Shiga toxins. *Infect. Immun.* 76, 4814–4822. doi: 10.1128/IAI.00446-08
- Rahal, E. A., Fadlallah, S. M., Nassar, F. J., Kazzi, N., and Matar, G. M. (2015). Approaches to treatment of emerging Shiga toxin-producing *Escherichia coli* infections highlighting the O104:H4 serotype. *Front. Cell Infect. Microbiol.* 5:24. doi: 10.3389/fcimb.2015.00024
- Rajkhowa, S., Scaria, J., Garcia, D. L., Musser, K. A., Akey, B. L., and Chang, Y. F. (2010). Analysis of *Escherichia coli* O157 clinical isolates by multilocus sequence typing. *BMC Res. Notes* 3:343. doi: 10.1186/1756-0500-3-343
- Rangel, J. M., Sparling, P. H., Crowe, C., Griffin, P. M., and Swerdlow, D. L. (2005). Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982–2002. *Emerg. Infect. Dis.* 11, 603–609.
- Riordan, J. T., Viswanath, S. B., Manning, S. D., and Whittam, T. S. (2008). Genetic differentiation of *Escherichia coli* O157:H7 clades associated with human disease by real-time PCR. *J. Clin. Microbiol.* 46, 2070–2073. doi: 10.1128/JCM.00203-08
- Robinson, C. M., Sinclair, J. F., Smith, M. J., and O'Brien, A. D. (2006). Shiga toxin of enterohemorrhagic *Escherichia coli* type O157:H7 promotes intestinal colonization. *Proc. Natl. Acad. Sci. U.S.A.* 103, 9667–9672. doi: 10.1073/pnas.0602359103
- Rump, L. V., Meng, J., Strain, E. A., Cao, G., Allard, M. W., and Gonzalez-Escalona, N. (2012). Complete DNA sequence analysis of enterohemorrhagic *Escherichia coli* plasmid pO157\_2 in beta-glucuronidase-positive *E. coli* O157:H7 reveals a novel evolutionary path. *J. Bacteriol.* 194, 3457–3463. doi: 10.1128/JB.00197-12
- Rump, L. V., Strain, E. A., Cao, G., Allard, M. W., Fischer, M., Brown, E. W., et al. (2011). Draft genome sequences of six *Escherichia coli* isolates from the stepwise model of emergence of *Escherichia coli* O157:H7. *J. Bacteriol.* 193, 2058–2059. doi: 10.1128/JB.00118-11
- Rusconi, B., and Eppinger, M. (2016). “Whole genome sequence typing for strategies for enterohemorrhagic *Escherichia coli* of the O157:H7 serotype,” in *Microbial Bioresources*, eds V. K. Kumar, D. Thangadurai, G. D. Sharma, and R. Gaur (Wallingford: Commonwealth Agricultural Bureaux International (CABI)), 800.
- Rusconi, B., Sanjar, F., Koenig, S. S., Mammel, M. K., Tarr, P. I., and Eppinger, M. (2016). Whole genome sequencing for genomics-guided investigations of *Escherichia coli* O157:h7 outbreaks. *Front. Microbiol.* 7:985. doi: 10.3389/fmicb.2016.00985
- Russo, L. M., Melton-Celsa, A. R., Smith, M. A., Smith, M. J., and O'Brien, A. D. (2014). Oral intoxication of mice with Shiga toxin type 2a (Stx2a) and protection by anti-Stx2a monoclonal antibody 11E10. *Infect. Immun.* 82, 1213–1221. doi: 10.1128/IAI.01264-13
- Sadiq, S. M., Hazen, T. H., Rasko, D. A., and Eppinger, M. (2014). EHEC genomics: past, present, and future. *Microbiol. Spectr.* 2:EHEC–0020–2013. doi: 10.1128/microbiolspec.EHEC-0020-2013
- Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., et al. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34, 374–378. doi: 10.2144/03342mt01
- Sahl, J. W., Caporaso, J. G., Rasko, D. A., and Keim, P. (2014). The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. *PeerJ* 2:e332. doi: 10.7717/peerj.332
- Saile, N., Voigt, A., Kessler, S., Stressler, T., Klumpp, J., Fischer, L., et al. (2016). *Escherichia coli* O157:H7 Strain EDL933 harbors multiple functional prophage-associated genes necessary for the utilization of 5-N-Acetyl-O-Acetyl neuraminic acid as a growth substrate. *Appl. Environ. Microbiol.* 82, 5940–5950. doi: 10.1128/AEM.01671-16
- Sambrook, J., Russell, D. W., and Sambrook, J. (2006). *The Condensed Protocols From Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Sandvig, K. (2001). Shiga toxins. *Toxicon* 39, 1629–1635.
- Sanjar, F., Hazen, T. H., Shah, S. M., Koenig, S. S., Agrawal, S., Daugherty, S., et al. (2014). Genome sequence of *Escherichia coli* O157:H7 Strain 2886-75, associated with the first reported case of human infection in the United States. *Genome Announc* 2:e01120-13. doi: 10.1128/genomeA.01120-13
- Sanjar, F., Rusconi, B., Hazen, T. H., Koenig, S. S., Mammel, M. K., Feng, P. C., et al. (2015). Characterization of the pathogenome and phylogenomic classification of enteropathogenic *Escherichia coli* of the O157:non-H7 serotypes. *Pathog. Dis.* 73:ftv033. doi: 10.1093/femspd/ftv033
- Scheutz, F., Teel, L. D., Beutin, L., Pierard, D., Buvens, G., Karch, H., et al. (2012). Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *J. Clin. Microbiol.* 50, 2951–2963. doi: 10.1128/JCM.00860-12
- Schmidt, H. (2001). Shiga-toxin-converting bacteriophages. *Res. Microbiol.* 152, 687–695. doi: 10.1016/s0923-2508(01)01249-9
- Schmidt, H., Beutin, L., and Karch, H. (1995). Molecular analysis of the plasmid-encoded hemolysin of *Escherichia coli* O157:H7 strain EDL 933. *Infect. Immun.* 63, 1055–1061. doi: 10.1128/iai.63.3.1055-1061.1995
- Schmidt, H., Scheef, J., Huppertz, H. I., Frosch, M., and Karch, H. (1999). *Escherichia coli* O157:H7 and O157:H(-) strains that do not produce Shiga toxin: phenotypic and genetic characterization of isolates associated with diarrhea and hemolytic-uremic syndrome. *J. Clin. Microbiol.* 37, 3491–3496. doi: 10.1128/jcm.37.11.3491-3496.1999
- Schutz, K., Cowley, L. A., Shaaban, S., Carroll, A., McNamara, E., Gally, D. L., et al. (2017). Evolutionary context of non-sorbitol-fermenting shiga toxin-producing *Escherichia coli* O55:H7. *Emerg. Infect. Dis.* 23, 1966–1973. doi: 10.3201/eid2312.170628
- Serra-Moreno, R., Jofre, J., and Muniesa, M. (2007). Insertion site occupancy by stx2 bacteriophages depends on the locus availability of the host strain chromosome. *J. Bacteriol.* 189, 6645–6654. doi: 10.1128/JB.00466-07
- Shaaban, S., Cowley, L. A., McAteer, S. P., Jenkins, C., Dallman, T. J., Bono, J. L., et al. (2016). Evolution of a zoonotic pathogen: investigating prophage diversity in enterohaemorrhagic *Escherichia coli* O157 by long-read sequencing. *Microb. Genom.* 2:e000096. doi: 10.1099/mgen.0.000096
- Shaikh, N., and Tarr, P. I. (2003). *Escherichia coli* O157:H7 Shiga toxin-encoding bacteriophages: integrations, excisions, truncations, and evolutionary implications. *J. Bacteriol.* 185, 3596–3605. doi: 10.1128/jb.185.12.3596-3605.2003
- Shin, I. S., Ishii, S., Shin, J. S., Sung, K. I., Park, B. S., Jang, H. Y., et al. (2009). Globotriaosylceramide (Gb3) content in HeLa cells is correlated to Shiga toxin-induced cytotoxicity and Gb3 synthase expression. *BMB Rep.* 42, 310–314. doi: 10.5483/bmbrep.2009.42.5.310
- Shringi, S., Schmidt, C., Katherine, K., Brayton, K. A., Hancock, D. D., and Besser, T. E. (2012). Carriage of stx2a differentiates clinical and bovine-biased strains of *Escherichia coli* O157. *PLoS One* 7:e51572. doi: 10.1371/journal.pone.0051572
- Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–D36. doi: 10.1093/nar/gkj014
- Smith, D. L., Rooks, D. J., Fogg, P. C., Darby, A. C., Thomson, N. R., McCarthy, A. J., et al. (2012). Comparative genomics of Shiga toxin encoding bacteriophages. *BMC Genomics* 13:311. doi: 10.1186/1471-2164-13-311
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stanton, E., Park, D., Dopfer, D., Ivanek, R., and Kaspar, C. W. (2014). Phylogenetic characterization of *Escherichia coli* O157 : H7 based on IS629 distribution and Shiga toxin genotype. *Microbiology* 160(Pt 3), 502–513. doi: 10.1099/mic.0.073437-0
- Steinberg, K. M., and Levin, B. R. (2007). Grazing protozoa and the evolution of the *Escherichia coli* O157:H7 Shiga toxin-encoding prophage. *Proc. Biol. Sci.* 274, 1921–1929. doi: 10.1098/rspb.2007.0245

- Stephan, R., Zhang, W., Bielaszewska, M., Mellmann, A., and Karch, H. (2009). Phenotypic and genotypic traits of Shiga toxin-negative *E. coli* O157:H7/H(-) bovine and porcine strains. *Foodborne Pathog. Dis.* 6, 235–243. doi: 10.1089/fpd.2008.0205
- Stolfa, G., and Koudelka, G. B. (2012). Entry and killing of *Tetrahymena thermophila* by bacterially produced Shiga toxin. *mBio* 4:e00416-12. doi: 10.1128/mBio.00416-12
- Strauch, E., Hammerl, J. A., Konietzny, A., Schneiker-Bekel, S., Arnold, W., Goesmann, A., et al. (2008). Bacteriophage 2851 is a prototype phage for dissemination of the Shiga toxin variant gene 2c in *Escherichia coli* O157:H7. *Infect. Immun.* 76, 5466–5477. doi: 10.1128/IAI.00875-08
- Su, L. K., Lu, C. P., Wang, Y., Cao, D. M., Sun, J. H., and Yan, Y. X. (2010). [Lysogenic infection of a Shiga toxin 2-converting bacteriophage changes host gene expression, enhances host acid resistance and motility]. *Mol. Biol. (Mosk.)* 44, 60–73.
- Subramanian, B., Gao, S., Lercher, M. J., Hu, S., and Chen, W. H. (2019). Evolvview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* 47, W270–W275. doi: 10.1093/nar/gkz357
- Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039
- Tatsuno, I., Horie, M., Abe, H., Miki, T., Makino, K., Shinagawa, H., et al. (2001). toxB gene on pO157 of enterohemorrhagic *Escherichia coli* O157:H7 is required for full epithelial cell adherence phenotype. *Infect. Immun.* 69, 6660–6669. doi: 10.1128/IAI.69.11.6660-6669.2001
- Tatusova, T., DiCuccio, M., Badretdin, A., Chetverin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624. doi: 10.1093/nar/gkw569
- Tennant, S. M., Tauschek, M., Azzopardi, K., Bigham, A., Bennett-Wood, V., Hartland, E. L., et al. (2009). Characterisation of atypical enteropathogenic *E. coli* strains of clinical origin. *BMC Microbiol.* 9:117. doi: 10.1186/1471-2180-9-117
- Tesh, V. L. (2010). Induction of apoptosis by Shiga toxins. *Future Microbiol.* 5, 431–453. doi: 10.2217/fmb.10.4
- Tesh, V. L., Burriss, J. A., Owens, J. W., Gordon, V. M., Wadolkowski, E. A., O'Brien, A. D., et al. (1993). Comparison of the relative toxicities of Shiga-like toxins type I and type II for mice. *Infect. Immun.* 61, 3392–3402. doi: 10.1128/iai.61.8.3392-3402.1993
- Themphachana, M., Nakaguchi, Y., Nishibuchi, M., Seto, K., Rattanachay, P., Singkhamanan, K., et al. (2014). First report in Thailand of a stx-negative *Escherichia coli* O157 strain from a patient with diarrhea. *Southeast Asian J. Trop. Med. Public Health* 45, 881–889.
- Toro, M., Rump, L. V., Cao, G., Meng, J., Brown, E. W., and Gonzalez-Escalona, N. (2015). Simultaneous presence of insertion sequence-excision enhancer (IEE) and insertion sequence IS629 correlates with increased diversity and virulence in Shiga-toxin producing *Escherichia coli* (STEC). *J. Clin. Microbiol.* 53, 3466–3473. doi: 10.1128/JCM.01349-15
- Tran, S. L., Billoud, L., Lewis, S. B., Phillips, A. D., and Schuller, S. (2014). Shiga toxin production and translocation during microaerobic human colonic infection with Shiga toxin-producing *E. coli* O157:H7 and O104:H4. *Cell Microbiol.* 16, 1255–1266. doi: 10.1111/cmi.12281
- Trofa, A. F., Ueno-Olsen, H., Oiwa, R., and Yoshikawa, M. (1999). Dr. Kiyoshi Shiga: discoverer of the dysentery *Bacillus*. *Clin. Infect. Dis.* 29, 1303–1306. doi: 10.1086/313437
- Tyler, J. S., Mills, M. J., and Friedman, D. I. (2004). The operator and early promoter region of the Shiga toxin type 2-encoding bacteriophage 933W and control of toxin expression. *J. Bacteriol.* 186, 7670–7679. doi: 10.1128/JB.186.22.7670-7679.2004
- Uchida, K., Dono, K., and Aizawa, S. (2016). Identification of the key sequence in the FliK C-terminal domain for substrate specificity switching in the flagellar protein secretion. *J. Bacteriol.* 198, 410–415. doi: 10.1128/JB.00712-15
- Uhlich, G. A., Paoli, G. C., Zhang, X., and Andreozzi, E. (2019). Whole-genome sequence of *Escherichia coli* serotype o157:h7 strain ATCC 43888. *Microbiol. Resour. Announc.* 8:e00906-19. doi: 10.1128/MRA.00906-19
- Veses-Garcia, M., Liu, X., Rigden, D. J., Kenny, J. G., McCarthy, A. J., and Allison, H. E. (2015). Transcriptomic analysis of shiga-toxigenic bacteriophage carriage reveals a profound regulatory effect on acid resistance in *Escherichia coli*. *Appl. Environ. Microbiol.* 81, 8118–8125. doi: 10.1128/AEM.02034-15
- Wagner, P. L., Acheson, D. W., and Waldor, M. K. (1999). Isogenic lysogens of diverse shiga toxin 2-encoding bacteriophages produce markedly different amounts of shiga toxin. *Infect. Immun.* 67, 6710–6714.
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963
- Wells, J. G., Davis, B. R., Wachsmuth, I. K., Riley, L. W., Remis, R. S., Sokolow, R., et al. (1983). Laboratory investigation of hemorrhagic colitis outbreaks associated with a rare *Escherichia coli* serotype. *J. Clin. Microbiol.* 18, 512–520. doi: 10.1128/jcm.18.3.512-520.1983
- Werber, D., and Scheutz, F. (2019). The importance of integrating genetic strain information for managing cases of Shiga toxin-producing *E. coli* infection. *Epidemiol. Infect.* 147:e264. doi: 10.1017/S0950268819001602
- Wetzel, A. N., and Lejeune, J. T. (2007). Isolation of *Escherichia coli* O157:H7 strains that do not produce Shiga toxin from bovine, avian and environmental sources. *Lett. Appl. Microbiol.* 45, 504–507. doi: 10.1111/j.1472-765X.2007.02228.x
- Wick, L. M., Qi, W., Lacher, D. W., and Whittam, T. S. (2005). Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7. *J. Bacteriol.* 187, 1783–1791. doi: 10.1128/JB.187.5.1783-1791.2005
- Wilgenbusch, J. C., and Swofford, D. (2003). Inferring evolutionary trees with PAUP\*. *Curr. Protoc. Bioinformatics* Chapter 6, Unit 6.4. doi: 10.1002/0471250953.bi0604s00
- Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., et al. (2006). Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol. Microbiol.* 60, 1136–1151. doi: 10.1111/j.1365-2958.2006.05172.x
- Wong, C. S., and Brandt, J. R. (2002). Risk of hemolytic uremic syndrome from antibiotic treatment of *Escherichia coli* O157:H7 colitis. *JAMA* 288, 3111; author reply 3112.
- Wong, C. S., Mooney, J. C., Brandt, J. R., Staples, A. O., Jelacic, S., Boster, D. R., et al. (2012). Risk factors for the hemolytic uremic syndrome in children infected with *Escherichia coli* O157:H7: a multivariable analysis. *Clin. Infect. Dis.* 55, 33–41. doi: 10.1093/cid/cis299
- Xie, Z., and Tang, H. (2017). ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 33, 3340–3347. doi: 10.1093/bioinformatics/btx433
- Yang, X., Bai, X., Zhang, J., Sun, H., Fu, S., Fan, R., et al. (2020). *Escherichia coli* strains producing a novel Shiga toxin 2 subtype circulate in China. *Int. J. Med. Microbiol.* 310:151377. doi: 10.1016/j.ijmm.2019.151377
- Yang, Z., Kovar, J., Kim, J., Niefeldt, J., Smith, D. R., Moxley, R. A., et al. (2004). 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.* 70, 6846–6854. doi: 10.1128/AEM.70.11.6846-6854.2004
- Yin, S., Rusconi, B., Sanjar, F., Goswami, K., Xiaoli, L. Z., Eppinger, M., et al. (2015). *Escherichia coli* O157:H7 strains harbor at least three distinct sequence types of Shiga toxin 2a-converting phages. *BMC Genomics* 16:733. doi: 10.1186/s12864-015-1934-1
- Yokoyama, E., Hashimoto, R., Etoh, Y., Ichihara, S., Horikawa, K., and Uchimura, M. (2011). Biased distribution of IS629 among strains in different lineages of enterohemorrhagic *Escherichia coli* serovar O157. *Infect. Genet. Evol.* 11, 78–82. doi: 10.1016/j.meegid.2010.10.007
- Yokoyama, E., Hirai, S., Hashimoto, R., and Uchimura, M. (2012). Clade analysis of enterohemorrhagic *Escherichia coli* serotype O157:H7/H- strains and hierarchy of their phylogenetic relationships. *Infect. Genet. Evol.* 12, 1724–1728. doi: 10.1016/j.meegid.2012.07.003
- Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C. H., et al. (2000). Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157:H7 strain derived from the Sakai outbreak. *Gene* 258, 127–139. doi: S0378-1119(00)00416-9
- Yu, J., and Kaper, J. B. (1992). Cloning and characterization of the eae gene of enterohaemorrhagic *Escherichia coli* O157:H7. *Mol. Microbiol.* 6, 411–417.

- Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi: 10.1093/jac/dks261
- Zhang, H., Gao, S., Lercher, M. J., Hu, S., and Chen, W. H. (2012). EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res.* 40, W569–W572. doi: 10.1093/nar/gks576
- Zhang, X., McDaniel, A. D., Wolf, L. E., Keusch, G. T., Waldor, M. K., and Acheson, D. W. (2000). Quinolone antibiotics induce Shiga toxin-encoding bacteriophages, toxin production, and death in mice. *J. Infect. Dis.* 181, 664–670. doi: 10.1086/315239
- Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39, W347–W352. doi: 10.1093/nar/gkr485
- Zhou, Z., Li, X., Liu, B., Beutin, L., Xu, J., Ren, Y., et al. (2010). Derivation of *Escherichia coli* O157:H7 from its O55:H7 precursor. *PLoS One* 5:e8700. doi: 10.1371/journal.pone.0008700
- Zimmerhackl, L. B. (2000). E. coli, antibiotics, and the hemolytic-uremic syndrome. *N. Engl. J. Med.* 342, 1990–1991. doi: 10.1056/NEJM200006293422611

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with one of the authors ME.

Copyright © 2020 Nyong, Zaia, Allué-Guardia, Rodriguez, Irion-Byrd, Koenig, Feng, Bono and Eppinger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.