



## Article

# Genomic and Phenotypic Characterization of Shiga Toxin-Producing *Escherichia albertii* Strains Isolated from Wild Birds in a Major Agricultural Region in California

Michelle Qiu Carter <sup>1,\*</sup>, Beatriz Quiñones <sup>1</sup>, Xiaohua He <sup>2</sup>, Antares Pham <sup>1</sup>, Diana Carychao <sup>1</sup>, Michael B. Cooley <sup>1</sup>, Chien-Chi Lo <sup>3</sup>, Patrick S. G. Chain <sup>3</sup>, Rebecca L. Lindsey <sup>4</sup> and James L. Bono <sup>5</sup>

- <sup>1</sup> Produce Safety and Microbiology Research Unit, U.S. Department of Agriculture, Western Regional Research Center, 800 Buchanan Street, Albany, CA 94710, USA; beatriz.quinones@usda.gov (B.Q.); apdienc@gmail.com (A.P.); diana.carychao@usda.gov (D.C.); masdpcooley@gmail.com (M.B.C.)
- <sup>2</sup> Foodborne Toxin Detection and Prevention Research Unit, U.S. Department of Agriculture, Western Regional Research Center, Albany, CA 94710, USA; xiaohua.he@usda.gov
- <sup>3</sup> Biosecurity and Public Health Group, U.S. Department of Energy, Los Alamos National Laboratory, Santa Fe, NM 87545, USA; chienchi@lanl.gov (C.-C.L.); pchain@lanl.gov (P.S.G.C.)
- <sup>4</sup> Enteric Diseases Laboratory Branch, Centers for Disease Control and Prevention, Atlanta, GA 30329, USA; wmi1@cdc.gov
- <sup>5</sup> Meat Safety and Quality Research Unit, U.S. Department of Agriculture, U.S. Meat Animal Research Center, Clay Center, NE 68933, USA; jim.bono@usda.gov
- \* Correspondence: michelle.carter@usda.gov; Tel.: +1-(510)-559-6180; Fax: +1-(510)-559-6162



**Citation:** Carter, M.Q.; Quiñones, B.; He, X.; Pham, A.; Carychao, D.; Cooley, M.B.; Lo, C.-C.; Chain, P.S.G.; Lindsey, R.L.; Bono, J.L. Genomic and Phenotypic Characterization of Shiga Toxin-Producing *Escherichia albertii* Strains Isolated from Wild Birds in a Major Agricultural Region in California. *Microorganisms* **2023**, *11*, 2803. <https://doi.org/10.3390/microorganisms1112803>

Academic Editor: Kim Stanford

Received: 16 September 2023

Revised: 2 November 2023

Accepted: 16 November 2023

Published: 18 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

*Escherichia albertii*, an emerging human and avian pathogen, was first isolated from Bangladeshi children with diarrhea in 1991 and classified initially as *Hafnia alveri* [1]. In 2003, these isolates were reclassified as *E. albertii* based on their biochemical properties, DNA hybridizations, and 16S rRNA gene sequences [2]. *E. albertii* is the most divergent lineage in the genus *Escherichia* and is closely related to *Shigella boydii* based on the extended multilocus sequence typing (MLST) and phylogenetic analyses [3,4]. As an avian pathogen, *E. albertii* was the cause of the large-scale mortality of redpoll finches (*Carduelis Flammula*) in Alaska and contributed to several mortality incidents of wild finches (*Fringillidae*) in Scotland [5,6]. In humans, *E. albertii* causes diarrhea, abdominal pain, and high fever,

although bacteremia and extraintestinal infections were also reported [7,8]. Sporadic outbreaks of *E. albertii* infections have been reported worldwide, particularly in Japan [9–14]. Transmission of *E. albertii* is thought to occur mainly via contaminated food or water. Although *E. albertii* has been isolated from various domestic and wild animals, poultry and wild birds are considered the primary reservoirs [15].

One of the characteristics of *E. albertii* pathogenesis is the formation of attaching-and-effacing (A/E) lesions on the surfaces of intestinal epithelial cells, similar to other attaching-and-effacing pathogens, including enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), and *Citrobacter rodentium* [16]. Similarly, genes responsible for formation of A/E lesions in *E. albertii* are located on a locus of enterocyte effacement (LEE) island, including the intimin gene *eae* and the gene (*tir*) encoding the translocated intimin receptor protein. Additionally, LEE in *E. albertii* carries the genes involved in the biogenesis of the type three secretion system (T3SS) apparatus and several T3SS effector genes [17]. Interestingly, some *E. albertii* strains harbor a genomic island (GI) encoding a second T3SS (ETT2), a cryptic T3SS distantly related to the *Salmonella* T3SS [18]. The function of ETT2 was originally thought to encode a secretion system for delivery of effectors to host cells, as it carries genes encoding T3SS components. However, recent studies demonstrated that rather than secretion of effectors, ETT2 is involved in the secretion of flagellar proteins, fimbrial proteins, and cell surface proteins and contributes to bacterial motility and resistance to detergents and hydrophobic antibiotics [19]. The genes *cdtABC*, encoding the cytolethal distending toxin (CDT), are conserved in *E. albertii*. CDT is composed of three subunits. The CdtA and CdtC subunits function together to mediate the binding of the holotoxin to eukaryotic cells and the subsequent delivery of CdtB into the host cell. CdtB possesses DNase I activity and induces host cell cycle arrest and, eventually, apoptosis via DNA damage [20,21]. Shiga toxin genes have been detected in a subset of *E. albertii* strains, in which subtype *stx<sub>2f</sub>* is common, with one exception of an *stx<sub>2a</sub>*-positive clinical isolate in Norway [22]. Possession of both *eae* and *stx<sub>2a</sub>* may increase the risk of developing hemolytic uremic syndrome (HUS) during *E. albertii* infections, a clinical signature of EHEC infections [23]. Other reported virulence genes include *paa*, encoding the porcine attaching-effacing-associated protein that plays a role in the pathogenesis of EPEC and enterotoxigenic *E. coli* (ETEC) [24].

Recently, *E. albertii* strains isolated from infected humans, environmental samples, poultry, and healthy and diseased wild birds have been reported worldwide [9,12,25,26]. Genomic characterization is desired, as such a study will provide a comprehensive picture of virulence gene repertoire as well as uncover valuable information about disease ecology and public health associated with this pathogen. Additionally, considering the long-distance migration of wild birds, a deeper understanding of the role of the avian reservoirs in *E. albertii*'s dispersal will provide valuable information for the development of effective measures for controlling the transmission of this human pathogen. In our previous study that assessed the prevalence of Shiga toxin-producing *E. coli* (STEC) in a leafy greens-growing region in Central California [27], *E. albertii* strains were isolated from many wild birds near this agricultural area. In this study, we performed comparative genomics and pathogenomic analyses of ten wild bird *E. albertii* strains with a set of *E. albertii* clinical strains isolated from 1954–2014 in the United States. We further performed comparative phenotypic analyses with several STEC and EHEC strains to determine Stx production and Vero cell-based cytotoxicity. Our study uncovered novel virulence determinants in avian and clinical *E. albertii* strains and revealed that wild birds, especially American crow (*Corvus brachyrhynchos*) and brown-headed cowbird (*Molothrus ater*), carried Shiga toxin-producing *E. albertii* (STEA) strains that have an increased potential to cause severe diseases in humans.

## 2. Materials and Methods

### 2.1. Bacterial Strains and Growth Media

*E. coli* and *E. albertii* strains and their sources are listed in Table 1. Environmental strains were isolated by cloacal swab from captured birds near a major agricultural region in California, as described previously [28]. Wildlife sampling at all locations was approved under a set of California Department of Fish and Game (CDFG) Scientific Collection Permits issued to USDA Wildlife Services and CDFG personnel contracted to collect the samples and ship to USDA in Albany, California. The wildlife sampling was conducted through a contract with state and federal wildlife agencies using their standard protocols. The strains were grown routinely in Luria-Bertani (LB) broth (10 g tryptone, 5 g yeast extract, and 5 g NaCl per liter).

**Table 1.** Genomic characteristics of *E. coli* and *E. albertii* strains used in this study.

Strains <sup>1</sup>	Sources/ Year of Isolation	Chromosome (bp)/ GenBank Accession #	Plasmids (bp)/ GenBank Accession #	Sequence Type <sup>2</sup> (ST)	<i>stx</i> Genes	References
<i>E. coli</i>						
EDL933	Ground beef/1982	5,528,445/ AE005174.2	92,077/AF074613.1	11	<i>stx</i> <sub>1a</sub> + <i>stx</i> <sub>2a</sub>	[29,30]
MG1655	Human/1922	4,641,652/ U00096.3	None	10	None	[31]
<i>E. albertii</i>						
RM9973	American crow ( <i>Corvus brachyrhyn- chos</i> )/2009	4,648,335/ CP043271.1	p1, 105,082/CP043273.1; p2, 102,351/CP043272.1	ST6057 *	<i>stx</i> <sub>2f</sub>	This study
RM9974	American crow ( <i>Corvus brachyrhyn- chos</i> )/2009	4,863,808/ CP043266.1	p1, 132,007/CP043270.1; p2, 99,091/CP043269.1; p3, 96,650/CP043268.1; p4, 27,507/CP043267.1	ST2682	<i>stx</i> <sub>2f</sub>	This study
RM9976	American crow ( <i>Corvus brachyrhyn- chos</i> )/2009	4,976,140/ CP043262.1	p1, 200,52/CP043265.1; p2, 85,807/CP043264.1; p3, 84,014/CP043263.1	ST8692	<i>stx</i> <sub>2f</sub>	This study
RM10507	Brown-headed cowbird ( <i>Molothrus ater</i> )/2009	5,007,500/ CP043258.1	p1, 102,034/CP043261.1; p2, 99,002/CP043260.1; p3, 42,515/CP043259.1	ST10004 *	<i>stx</i> <sub>2f</sub>	This study
RM10705	Brown-headed cowbird ( <i>Molothrus ater</i> )/2009	4,904,046/ CP045690.1	p1, 117,701/CP045688.1; p2, 102,214/CP045689.1	ST9018	<i>stx</i> <sub>2f</sub>	This study
RM15112	Oregon Junco ( <i>Junco hyemalis</i> )/2011	4,712,580/ CP043254.1	p1, 136,729/CP043257.1; p2, 97,372/CP043256.1; p3, 62,266/CP043255.1	ST7971	<i>stx</i> <sub>2f</sub>	This study
RM15113	Oregon Junco ( <i>Junco hyemalis</i> )/2011	4,712,547/ CP043250.1	p1, 136,729/CP043253.1; p2, 97,374/CP043252.1; p3, 62,269/CP043251.1	ST7971	<i>stx</i> <sub>2f</sub>	This study
RM15114	Oregon Junco ( <i>Junco hyemalis</i> )/2011	4,712,362/ CP043246.1	p1, 136,747/CP043249.1; p2, 97,369/CP043248.1; p3, 62,259/CP043247.1	ST7971	<i>stx</i> <sub>2f</sub>	This study
RM15115	White-Breasted Nuthatch ( <i>Sitta carolinensis</i> )/2011	4,712,540/ CP043242.1	p1, 136,727/CP043245.1; p2, 97,365/CP043244.1; p3, 62,256/CP043243.1	ST7971	<i>stx</i> <sub>2f</sub>	This study

**Table 1.** Cont.

Strains <sup>1</sup>	Sources/ Year of Isolation	Chromosome (bp)/ GenBank Accession #	Plasmids (bp)/ GenBank Accession #	Sequence Type <sup>2</sup> (ST)	stx Genes	References
RM15116	Oregon Junco ( <i>Junco hyemalis</i> )/2011	4,712,500/ CP043238.1	p1, 136,731/CP043241.1; p2, 97,363/CP043240.1; p3, 62,266/CP043239.1	ST7971	<i>stx2f</i>	This study
2014C-4356	Chicken Carcass/2009	4,852,165/ CP024282.1	p1, 40,461/CP024283.1; p2, 59,626/CP024284.1; p3, 127,606/CP024285.1; p4, 113,727/CP024286.1; p5, 124,142/CP024287.1; p6, 19,118/CP024288.1	ST7415	None	[32]
05-3106	Human/2005	4,719,735/ CP030778.2	P1, 56,603/CP030779.2; P2, 80,632/CP030780.2	ST4619	None	[33]
07-3866	Human/2007	4,940,006/ CP030781.1	104,269/CP030782.1	ST7414	None	[33]
54-2045 (NCTC 9362)	Human/1954	4,551,125/ CP034213.1	40,180/CP034214.1	ST413	None	[33]
2010C-3449	Human/2010	4,923,641/ CP034212.1	None	ST5390	None	[33]
2011C-4180	Human/2011	4,790,629/ CP126912	p1, 48,807/CP126913 p2, 98,642/CP126914 p3, 151,845/CP126915	ST2681	<i>stx2f</i>	This study
2012EL-1823B	Human/2012	4,809,821/ CP030783.2	p1, 100,347/CP030784.2; p2, 81,130/CP030785.2; p3, 105,846/CP030786.2	ST5983	<i>stx2f</i>	[33]
2013C-4143	Human/2013	4,659,709/ CP030787.2	None	ST7960	None	[33]
2014C-4015	Human/2014	4,623,903/ CP034166.1	p1, 63,809/CP034165.1; p2, 96,264/CP034164.1; p3, 136,645/CP034167.1	ST5992	<i>stx2f</i>	[33]
2014EL-1348	Human/2014	4,662,779/ CP126908	p1, 62,344/CP126909 p2, 96,264/CP126910 p3, 136,695/CP126911	ST5990	<i>stx2f</i>	This study

<sup>1</sup> The source and isolation year for strain MG1655 is based on the information available for its parental strain K-12 as described previously [34]. <sup>2</sup> The sequence type (ST) was determined using MLST 2.0 service at the Center for Genomic Epidemiology with the *Escherichia coli* # 1 configuration [35]. \* The closest ST.

## 2.2. Genome Sequencing and Annotation

The genomes of the ten avian *E. albertii* strains were sequenced using an RS II instrument (Pacific Biosciences, Menlo Park, CA, USA) as previously reported [36]. Briefly, bacterial DNA was extracted from the mid-exponential phase cultures grown in LB broth and resuspended in Qiagen Elution Buffer (QIAGEN, Germantown, MD, USA) for genome sequencing. Genomic libraries were prepared according to the PacBio (Pacific Biosciences) 20 Kb library standard protocol, and DNA sequencing was performed with P6/C4 sequencing chemistry and 360 min data collection protocol. The sequence reads were filtered with a PreAssembler Filter prior to de novo assembly with RS\_HGAP\_Assembly v.3.

To obtain complete genome sequences of the two clinical strains that only have draft genome sequences (Table 1), bacterial DNA was extracted from exponential phase cultures grown in LB broth as previously reported [37]. DNA (10 µg) was sheared to a 30 Kb target fragment length using g-TUBEs (Covaris, Woburn, MA, USA) and concentrated with 0.45× volume AMPure PB beads (Pacific Biosciences). Sheared and concentrated DNA (5 µg) was used to make PacBio sequencing libraries using the SMRTbell Prep Kit 3.0 and barcoded using the SMRTbell barcoded adapter plate 3.0. The Sequel II binding kit 3.2 and Sequel II sequencing plate 2.0 were used to run the library on a Sequel IIe system with the application HiFi reads and a 30 h movie time with a 6 h pre-extension. PacBio reads were assembled using Microbial Genome Analysis in SMRT analysis v 10.1 and contigs imported into Geneious Prime® (Dotmatics, Bishop Stortford, UK or Boston, MA,

USA). The overlapping sequences on the ends of the contigs were removed from the 5' and 3' ends to generate circularized chromosomes and plasmids. Closed chromosomes were reoriented using Ori-Finder 2 [38], and all subsequent genomes were oriented to the same start position. The closed chromosome and plasmids were manually polished by mapping Illumina and PacBio reads to the chromosome and known plasmids using Geneious mapper. Unused reads were de novo assembled using the Geneious assembler for small plasmid identification. All 12 complete genome sequences were submitted to GenBank and annotated with the NCBI Prokaryotic Genome Annotation Pipeline [39] (Table 1).

### 2.3. Pangenome Analyses

The pangenome of *E. albertii* was calculated using Roary (V3.13.0) as previously reported [40]. Briefly, the annotated Coding DNA Sequences (CDSs) were extracted from each genome and converted to protein sequences. To identify homologous proteins, all-against-all comparison was performed using BLASTP with 95% as the minimum percentage identity. The resulting core genes were aligned using MAFFT (V7.471), and a maximum-likelihood tree was constructed in IQ-TREE (version 2.1.2) using the best-fit model GTR+F+I, as selected by ModelFinder [41,42]. To identify strains carrying more common genes, an approximate-maximum-likelihood tree of the binary accessory genes was constructed in FastTree (v2.1.10) with the default settings (Jukes-Cantor + CAT model or GTR+CAT model) as described previously [40]. The pangenome-wide association analysis was performed using Scoary (v1.6.16) to reveal host-associated genes and/or strain-specific genes.

### 2.4. Identification of Virulence Genes and Pathogenicity Islands (PAIs)

*E. coli* virulence genes listed in the Virulence Factor DataBase (VFDB) as of May 2023 were used as queries of BLASTn to search the genomes examined in this study (Table 1). The BLASTn was performed in Geneious Prime® with a threshold of 65% for gene coverage and 70% or 25% for sequence identity at the nucleotide or amino acid level, respectively. BLASTn searches were first performed to identify homologs of a genomic island (GI) or a PAI in all genomes using Geneious Prime® as previously reported [40,43]. When a complete GI or PAI was not detected, all CDSs encoded by the query GI or PAI were used to search the genome of the testing strain by BLASTP. The association of the T3SS effector genes and toxin genes was assessed by first calculating the median difference for each examined gene between the clinical and wild bird strains. Following this initial comparison, a list of genes that were different between the two isolation sources by greater than two standard deviations were subsequently selected and evaluated with the Fisher's exact test with the Bonferroni correction for multiple comparison using R version 4.3.0. as detailed in previous reports [44,45].

### 2.5. Comparative Analysis of Stx-Prophages

The complete genome sequence of each strain was submitted to PHASTER [46] for identification of prophage and prophage-like elements. The putative integration sites were initially identified by PHASTER and confirmed in Geneious Prime® using "Find Repeats" in the defined chromosomal regions. The Stx-prophage genomes were extracted from the corresponding bacterial genomes. The Stx-prophage genomes were aligned using Clustal Omega in Geneious Prime®. A neighbor-joining consensus tree was constructed with the following parameters: Genetic Distance Model, Jukes-Cantor; Resampling Method, bootstrap; and number of replicates, 10,000.

### 2.6. Quantification of Shiga Toxin by ELISA

The amount of Stx in each bacterial supernatant was quantified by ELISA, as described previously [47]. Briefly, a 96-well plate was coated at 4 °C overnight with a monoclonal antibody against Stx2 at a concentration of 5 µg/mL, followed by incubation with 3% BSA at 37 °C for 1 h. The Stx2f was detected using biotinylated mouse monoclonal antibody

Stx2f-4 [48] in conjugation with streptavidin-HRP at a concentration of 0.2 µg/mL (Invitrogen, Waltham, MA, USA). The plates were developed using SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific, Waltham, MA, USA), and the luminescence was measured using the Victor-3 plate reader (Perkin-Elmer, Shelton, CT, USA). The concentration of Stx2f in each sample was calculated using the standard curves generated with purified Stx2f.

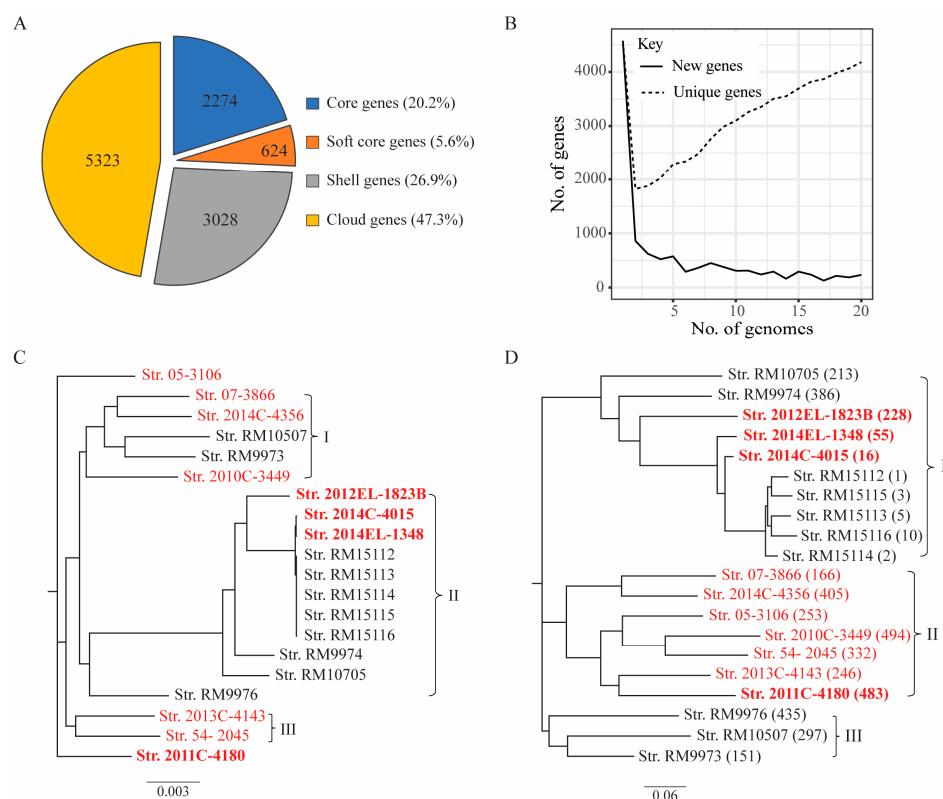
### 2.7. Cytotoxicity Assay

The Stx activity from each strain was determined with a mammalian Vero cell line, Vero-d2EGFP with a destabilized variant ( $t_{1/2} = 2$  h) of the enhanced green fluorescent protein (EGFP) [49]. Briefly, Vero-d2EGFP cell intoxications were performed after incubation at 37 °C in a 5% CO<sub>2</sub> humidified incubator with Ham's F-12 Nutrient medium (Life Technologies) containing tenfold dilutions of cell-free culture supernatants from each bacterial strain. After 20 h, the Vero-d2EGFP cells were briefly rinsed with 1X phosphate-buffered saline, and the EGFP fluorescence was measured using a BioTek Synergy HT Multi-Detection Microplate Reader (Agilent Technologies, Santa Clara, CA, USA), as described in a previous study [49]. Statistical analysis of the cytotoxicity levels among the tested strains was performed with R software version 4.3.0 (R Core Team 2022).

## 3. Results

### 3.1. Pangenome of *E. albertii*

Pangenome analyses of the 20 *E. albertii* strains revealed a total of 11,249 genes, among which 2274 genes were present in all genomes (core genes, 20.2%) (Figure 1A). The accessory genes included 3028 genes present in more than 3 but less than 19 genomes (shell genes, 26.9%) and 5325 genes present in less than 3 genomes (cloud genes, 47.3%), implying that *E. albertii* carried a large number of dispensable genes. This large accessory genome showed the great genetic diversity in this species, which was further supported by the observation that the number of unique genes increased when the number of genomes increased (Figure 1B). Core gene-based phylogeny analysis placed the strains in three clades and two singletons but failed to separate clinical strains from the non-clinical strains (Figure 1C). The largest clade (Clade II) contained three *stx<sub>2f</sub>*-positive clinical strains and eight wild bird strains. The other *stx<sub>2f</sub>*-positive clinical strain (2011C-4180) formed a singleton. The second-largest clade contained two *stx*-negative clinical strains, the chicken carcass strain (*stx*-negative), and two wild bird strains that both harbored the *stx<sub>2f</sub>* genes. The relatedness of clinical and avian strains was further evaluated with a tree constructed based on the presence of the accessory genes (Figure 1D). Three clades were revealed, among which the largest one (Clade I) contained the three *stx<sub>2f</sub>*-positive clinical strains and seven wild bird strains. The second clade contained all other clinical strains and the chicken carcass strain, while the third clade contained the rest of the three wild bird strains. Therefore, the three *stx<sub>2f</sub>*-positive clinical strains, 2012EL-1823B, 2014C-4015, and 2014EL-1348, appeared to be phylogenetically highly related to a clonal group of wild bird strains belonging to ST7971 (RM15112-RM15116) and also shared more accessory genes with this group of STEA strains than any other STEA strains examined in this study (Table 1).



**Figure 1.** Pangenome of *E. albertii* strains. (A) Core and accessory genes of *E. albertii*. The numbers of genes in the core and accessory genomes were calculated using Roary (V3.13.0), as detailed in the Materials and Methods section. The core genes refer to all genes shared by all input genomes ( $n = 20$ ); the soft-core genes refer to all genes present in all 19 input genomes; the shell genes refer to all genes present in at least 3 but less than 19 input genomes; the cloud genes refer to all genes present in less than 3 input genomes. (B) Change in the total number of new genes and unique genes as genomes are added in random orders. As the number of genomes used for the pangenome analyses increased, the total number of new genes (genes not found in the previously compared genomes) decreased, while the total number of unique genes (genes unique to individual strains) increased. (C) Core genome-based phylogeny of the *E. albertii* strains. The Roary matrix contained a total of 11,249 gene clusters. Strain names in red are clinical strains, and strain names in red bold are *stx2f*-positive clinical isolates. (D) The accessory genome-based strain relatedness. The number in the parentheses represents the number of strain-specific genes. Strain names in red are clinical strains, and strain names in red bold are *stx2f*-positive clinical strains.

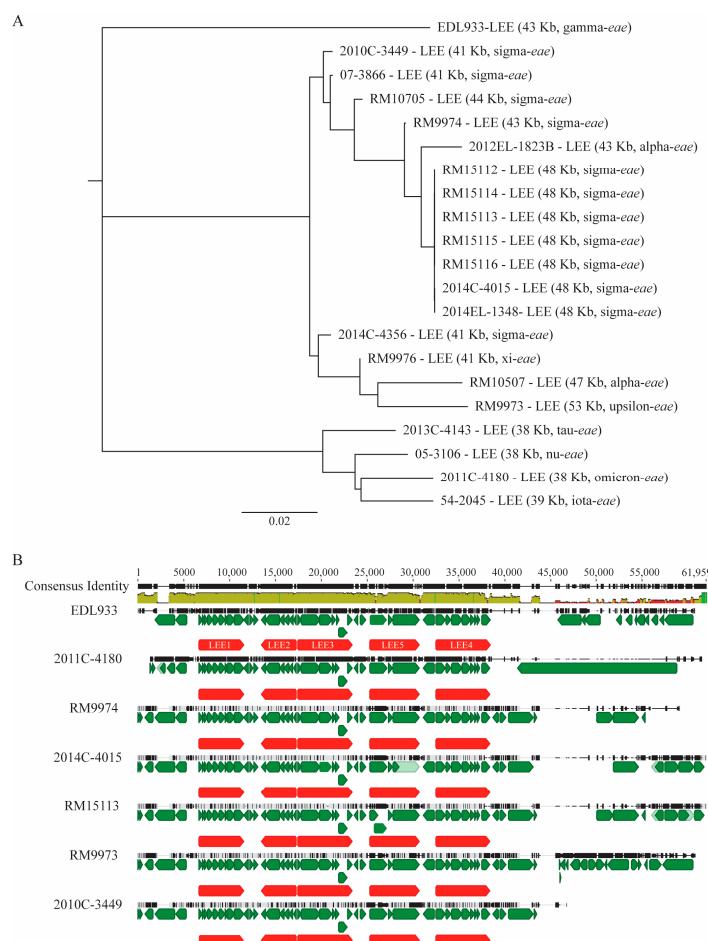
### 3.2. Analyses of PAIs in *E. albertii*

Seven PAIs, including the high pathogenicity island (HPI), locus of adhesion and autoaggregation (LAA), LEE, locus of proteolysis activity (LPA), OI-122, OI-57, and the subtilase-encoding pathogenicity island (SE-PAI) as described previously [40,43], and the gene clusters encoding the type six secretion system (T6SS) were examined in detail in *E. albertii* strains. Complete or partial homologs of LEE, OI-122, HPI, and T6SS were detected in *E. albertii*, while homologs for others were not identified.

#### 3.2.1. LEE

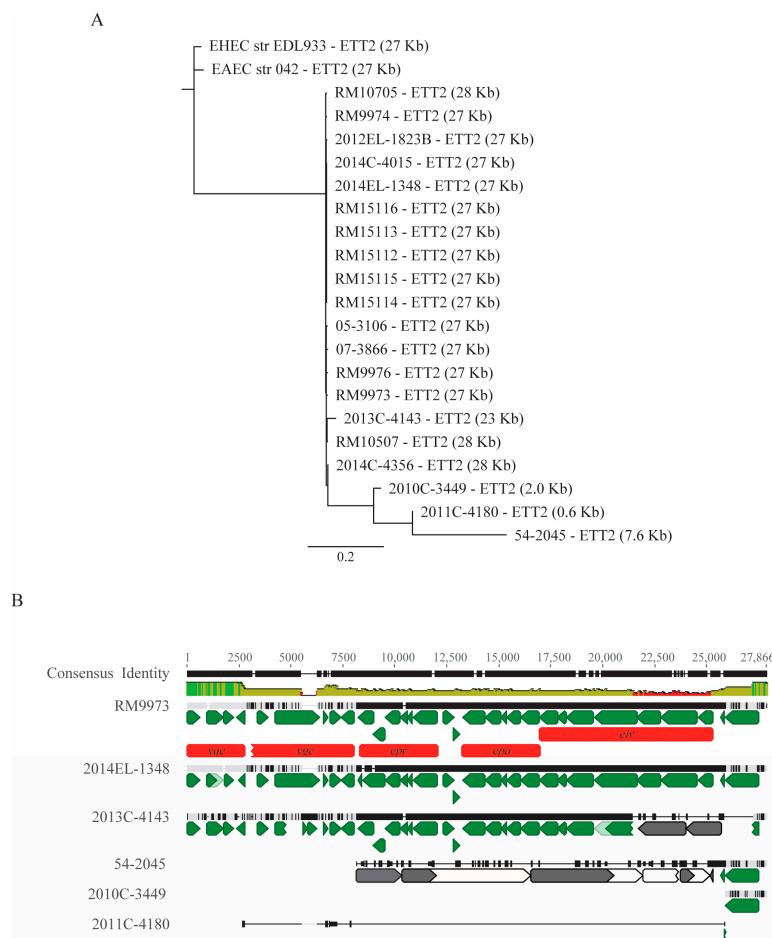
Regardless of the isolation source, all *E. albertii* strains examined carried an LEE, with a size ranging from 37,521 bp to 52,723 bp. Sequence alignments of LEEs revealed three clusters (Figure 2A). Cluster I contained LEEs from seven *stx2f*-positive wild bird strains and five clinical strains, among which three were *stx2f*-positive. All strains within this cluster carried an intimin sigma subtype gene except for the clinical strain 2012EL-1823, which encoded an intimin alpha subtype. Cluster II contained LEEs from three wild bird

strains and the chicken carcass isolate 2014C-4356. The intimin subtypes encoded by the strains within this cluster included alpha, sigma, upsilon, and xi. Cluster III contained four clinical strains, including an *stx2f*-positive strain and encoded intimin subtypes of alpha, iota, omicron, and tau. Unlike the LEE in the EHEC strain EDL933, all LEEs in the *E. albertii* strains were inserted downstream of the tRNA gene *pheU*. However, all five polycistronic operons (LEE1–LEE5) defined in the EDL933 LEE were conserved in the LEEs of the *E. albertii* strains examined (Figure 2B). Variation in the LEE was mostly detected at the two junction regions, especially in the region between the LEE4 and the LEE integration site. The accessory genes located within these variable regions included mainly the mobile genetic element (MGE)-related genes, T3SS effector genes, transcriptional regulator genes, and heavy metal resistance genes.



**Figure 2.** Sequence analyses of LEE in *E. albertii*. **(A)** Phylogeny of *E. albertii* LEEs. The LEEs in the *E. albertii* strains were identified by BLASTn search of a database containing all genomes examined in this study using DNA sequence of EDL933 LEE as a query in Geneious Prime®. The sequences of the LEEs were extracted from corresponding genomes and aligned using Clustal Omega alignment in Geneious Prime®. A neighbor-joining consensus tree was constructed using the LEE in strain EDL933 as an outgroup with the following parameters: Genetic Distance Model, Jukes-Cantor; Resampling tree method: Bootstrap; Number of Replicates: 10,000; Support Threshold: 50%. The length of each LEE is indicated in parentheses. Subtypes of intimin gene *eae* were identified by BLASTn search of a database with 19 subtypes of *eae*. **(B)** Alignment of representative *E. albertii* LEEs. Representative *E. albertii* LEEs were aligned with the EDL933 LEE using Clustal Omega method. Green triangles represent the annotated CDSs, and red blocks represent LEE polycistronic operons. The identity of the consensus sequence is color-coded (green: 100% identity; brown: at least 30% and under 100% identity; and red: below 30% identity).

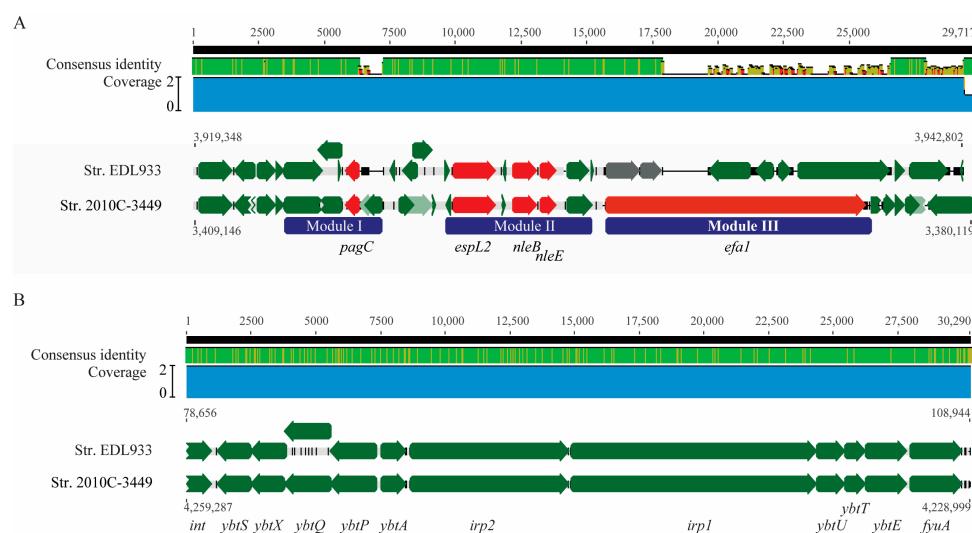
Besides LEE, some *E. coli* strains carry homologs of ETT2 (*E. coli* Type Three Secretion System 2), a 29 Kb PAI exhibiting sequence similarity with the *Salmonella* pathogenicity island 1 (SPI-1) [18]. Homologs of a complete ETT2 were identified in all ten *E. albertii* wild bird strains (Figure 3A). The majority of genes involved in biogenesis of T3SS apparatus were present on the ETT2s in the wild bird strains. A complete ETT2 was also detected in five clinical strains and the chicken isolate, 2014C-4356, while an incomplete ETT2 was present in the other strains (Figure 3A). In the clinical strain 2011C-4180, none of the ETT2-encoded genes were detected. There was only a 623 bp intergenic DNA fragment bordered by the tRNA gene *glyU* and an IS1 family transposase gene, while in clinical strain 2013C-4143, deletions in multiple ETT2 genes were observed, resulting in an incomplete ETT2 (~23 Kb) carrying various deletions in the *eiv* gene cluster (Figure 3B). Similar to strain EDL933, all ETT2s in the *E. albertii* were inserted downstream of tRNA gene *glyU*.



**Figure 3.** Sequence analyses of ETT2 in *E. albertii*. **(A)** Phylogeny of *E. albertii* ETT2s. The ETT2s in *E. albertii* strains were identified by BLASTn search of a database containing all genomes examined in this study using DNA sequence of EDL933 ETT2 as a query in Geneious Prime®. The sequences of the ETT2s were extracted from corresponding genomes and aligned using Clustal Omega alignment, and a neighbor-joining consensus tree was constructed using EDL933 ETT2 as an outgroup with the following parameters: Genetic Distance Model, Jukes-Cantor; Resampling tree method: Bootstrap; Number of Replicates: 10,000; Support Threshold: 50%. **(B)** Alignment of representative *E. albertii* ETT2s. Six representative *E. albertii* ETT2s were aligned using Clustal Omega method. Green triangles represent the annotated CDSs, and red blocks represent genes with annotated functions. Grey triangles represent hypothetical genes. The triangles with an overlapping smaller triangle represent genes carrying mutations that result in a premature stop codon. The identity of the consensus sequence is color-coded (green: 100% identity; brown: at least 30% and under 100% identity; and red: below 30% identity).

### 3.2.2. OI-122

The OI-122 in strain EDL933 is composed of three virulence gene modules: module I carries the virulence gene *pagC*, module II carries non-LEE effector genes *espL2*, *nleB*, and *nleE*, and module III carries a gene encoding the lymphostatin Efa1/LifA protein (Figure 4A). A complete OI-122 homolog was detected in clinical strain 2010C-3449 and inserted immediately downstream of tRNA gene *pheV*. All virulence genes located on the EDL933 OI-122 were present on the OI-122 of the *E. albertii* strain 2010C-3449. In other *E. albertii* strains, although no homologs of OI-122 were detected, some of the OI-122 virulence genes were detected, such as *nleB* and *nleE* in the clinical strains 2012EL-1823B and 2014C-4014 and in the wild bird strains RM9973, RM9974, RM10507, and RM10705; and *espL2*, *nleB*, and *nleE* in the clinical strains 07-3866, 2014C-4015, and 2014EL-1348 and in the five wild bird strains belonging to ST7971 (Strains RM15112-RM15116).



**Figure 4.** Sequence analyses of OI-122 and HPI in *E. albertii*. (A) Pairwise comparison of OI-122 in strains EDL933 and 2010C-3449. The sequences were aligned using Clustal Omega in Geneious Prime®. Green arrows represent genes, while red arrows represent virulence genes located on three modules (purple blocks). (B) Pairwise comparison of HPI in strains EDL933 and 2010C-3449. The sequences were aligned using Clustal Omega in Geneious Prime®. The identity of the consensus sequence is color-coded (green: 100% identity; brown: at least 30% and under 100% identity; and red: below 30% identity).

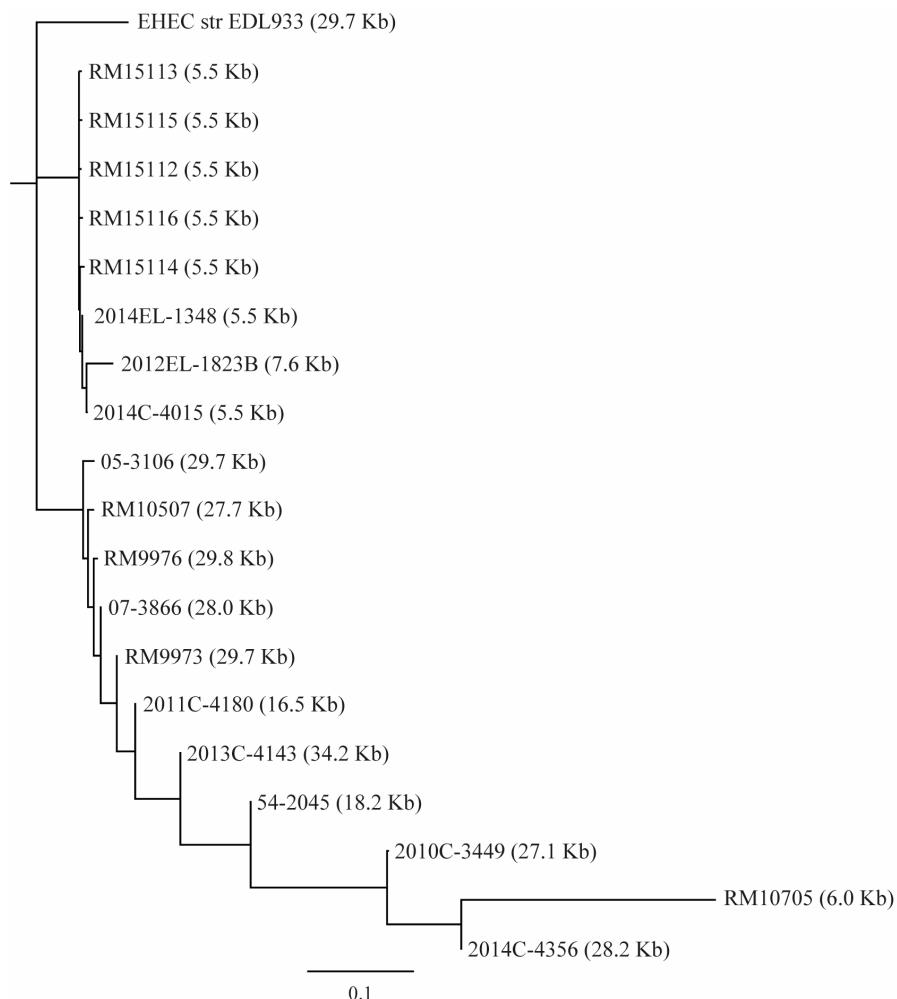
### 3.2.3. HPI

HPI, originally identified in *Yersinia* species [50], is involved in iron storage and uptake and has been detected in diverse STEC strains [51]. Like the OI-122, a complete homolog of HPI was only detected in the clinical strain 2010C-3449, which exhibited >99% sequence identity to the *Yersinia* HPI (Figure 4B). In the other *E. albertii* strains, homologs to any of the HPI genes were not detected.

### 3.2.4. T6SS

A BLASTn search of the three T6SS GIs, T6SS-1 from the EAEC strain 42 (~33 Kb), T6SS-2 from strain EDL933 (~30 Kb), and T6SS-3 from StxEAAC strain 2011C-3493 (~18 Kb) revealed homologs of T6SS-2 in the majority of *E. albertii* strains (Figure 5). A complete or a nearly complete T6SS-2 was detected in the clinical strains 05-3106, 07-3866, 2010C-3449, 2013C-4143, the chicken strain 2014C-4356, and the wild bird strains RM9973, RM9976, and RM10507. An incomplete T6SS-2 was detected in other *E. albertii* strains except in strain RM9974, where the corresponding chromosome site was occupied by an ISAs1 family transposases gene (locus tag, FYK18\_08020). Sequence analysis of T6SS-2 revealed two large clades, one containing the five wild bird strains carrying ST7971 (strains RM15112-RM15116)

and three *stx<sub>2f</sub>*-positive clinical strains 2012EL-1823B, 2014C-4015, and 2014EL-1345. The T6SS-2 GIs in this clade were all incomplete (5.5 or 7.6 Kb), only containing genes that encode the T6SS tip protein VgrG and the large type IV secretion protein Rhs. The T6SS-2 GIs in Clade II exhibited greater sequence diversity compared with the T6SS genes in Clade I, as demonstrated by the multiple branches in Clade II (Figure 5). A majority of the strains in this clade carried a complete T6SS gene cluster.



**Figure 5.** T6SSs in *E. albertii* strains. Phylogenetic analyses of T6SS-2. Sequences of *E. albertii* T6SS gene clusters and the T6SS gene cluster from EHEC strain EDL933 were aligned using Clustal Omega in Geneious Prime® software Version 2023.2.1. A consensus tree was constructed using Geneious Tree Builder with the following settings: Genetic Distance Model: Jukes-Cantor; Tree Build Method: Neighbor-Joining; Outgroup: EDL933 T6SS; Resampling Method: Bootstrap; Random Seed: 326,544; Number of Replicates: 10,000. The length in parentheses represents the length of DNA fragment carrying T6SS genes. Strain RM9974 was not included in the analyses due to absence of T6SS gene cluster.

### 3.3. Detection of *E. coli* virulence Genes in *E. albertii*

A total of 55 genes encoding T3SS effectors reported mainly in STEC and EPEC were retrieved from VFDB and used as queries to search a database containing all *E. albertii* strains examined in this study (Supplementary Table S1). The number of detected T3SS effector genes in *E. albertii* strains varied from 17 to 43 in clinical strains and from 28 to 39 in wild bird strains. The most common T3SS effector genes detected were *nleG* (nine alleles), *nleH* (two alleles), *nleF*, *map*, *espM* (two alleles), *espO1* (two alleles), *espH*, *espZ*, *espB*, *espG*, and *espC*. The genes *espM2*, *nleC*, *nleG-1*, *nleG-2*, *nleG5-1*, and *nleG6-1* appeared to be associated with the wild bird strains.

A total of 30 *E. coli* genes encoding toxins or transporters involved in the toxin delivery were further examined in *E. albertii* strains (Supplementary Table S2). Among the eight genes encoding serine protease autotransporter toxins, homologs of *espC* and *espI* were detected in three clinical strains but not in any of the wild bird strains. In contrast, a gene encoding the vacuolating autotransporter toxin Vat was detected in all of the wild bird strains and in the majority of the clinical strains. The toxin B gene was also detected in all of the wild bird strains but only in seven of the clinical strains. The Shiga toxin genes, encoding the Stx<sub>2f</sub> subtype, were detected in all wild bird strains but only in four clinical strains. The cytolethal distending toxin gene *cdtABC* was detected in all *E. albertii* strains examined. Interestingly, multiple *cdtABC* genes were present in all *stx<sub>2f</sub>*-positive strains, while only a single copy was detected in all *stx<sub>2f</sub>*-negative strains. Genome mapping revealed that all wild bird strains and three clinical strains harbored two copies of *cdtABC*, and one of them was associated with a prophage (Supplementary Table S2). In the clinical strain 2012EL-1823B, there were three copies of *cdtABC* genes, and two of them were on the prophage genomes. Mutations were detected in the *cdtB* of the clinical strains 54-2045, 2011C-4180, and 2012EL-1823 and in the *cdtC* of the wild bird strain RM15114.

### 3.4. *Stx2f*-Converting Prophages and Production of *Stx2f* under Non-Inducing Condition

The genome size of *Stx2f*-prophages in *E. albertii* strains varied from 42 Kb to 58 Kb. Sequence alignment revealed three clusters, among which the largest cluster contained three clinical and seven wild bird strains (Figure 6A). The *Stx2f*-prophages in this group were all inserted within the transfer messenger RNA (tmRNA) gene *ssrA*. The *Stx2f*-prophages in the wild bird strains RM9973 and RM9974 were highly similar and were inserted within a hypothetical gene located upstream of the *aspS* gene (Figure 6A). The *Stx2f*-prophage in the wild bird strain RM9976 was inserted downstream of the tRNA gene *serT*. The *Stx2f*-prophage in the clinical strain 2011C-4180 displayed the highest sequence similarity with the *Stx2f*-prophage in strain RM9976 but was inserted within the *ssrA* gene, like the majority of *stx<sub>2f</sub>*-positive strains examined in this study.

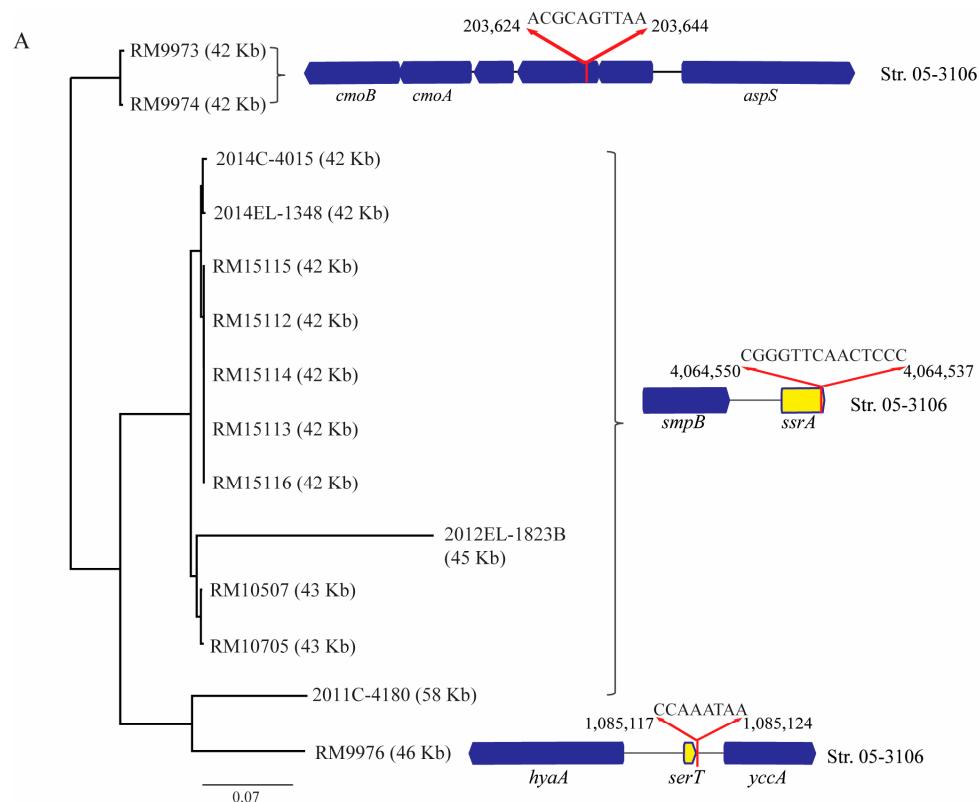
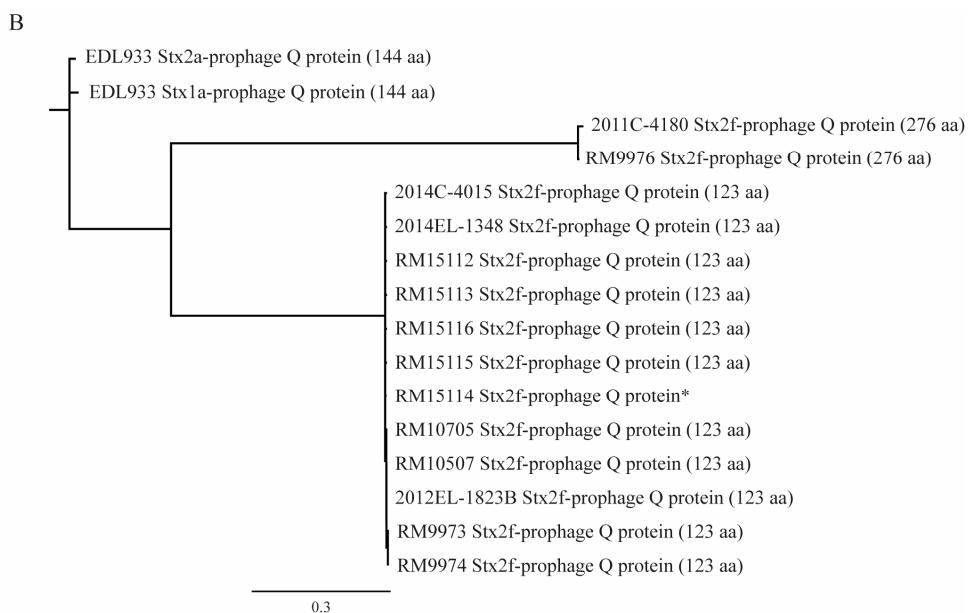


Figure 6. Cont.



**Figure 6.** Sequence analyses of Stx2f-converting prophages in *E. albertii*. (A) Relatedness and chromosomal locations of Stx2f prophages. The genome size of each prophage is indicated in parentheses. The chromosomal insertion sites of Stx2f prophage are indicated with the corresponding positions in the clinical strain 05-3106. The putative integration sites are indicated by the red arrows. Yellow blocks refer to tmRNA and tRNA genes. Blue blocks refer to annotated genes. (B) Sequence analyses of antitermination protein Q genes located upstream of *stx2f* coding sequences. The coding sequences of antitermination protein Q were aligned using Clustal Omega. A consensus tree was generated using neighbor-joining method with Jukes-Cantor for genetic distance model and rooted with the Q protein encoded by the Stx2a-prophage in EHEC strain EDL933. The Q protein size is indicated in parentheses. An asterisk indicates a truncated Q protein.

Examination of the DNA sequences upstream of *stx2f* coding sequences in *E. albertii* did not identify the putative late promoter  $P_{R'}$  or the antitermination Q protein binding site *qut*. The majority of STEA strains carried a 123 aa antitermination Q protein, which exhibited about 50% identity with the Q protein (144 aa) encoded by the Stx2a-prophage in strain EDL933 (Figure 6B). A larger antitermination Q protein (276 aa) was revealed in the clinical strain 2011C-4180 and in the wild bird strain RM9976, exhibiting only about 11% identity with the Q protein in strain EDL933.

Spontaneous production of Stx2f in *E. albertii* was further evaluated, and the results showed a large variation in Stx2f production among the STEA strains grown in rich growth medium to the stationary phase. The greatest production was observed in wild bird strain RM9974 (28.5 ng/mL), followed by wild bird strain RM9973 (5.6 ng/mL), and the clinical strain 2011C-4180 (3.3 ng/mL). Spontaneous production of Stx2f was in the range of 0.1–1.0 ng/μL in the majority of *E. albertii* strains and undetectable in the clinical strain 2012EL-1823B (Table 2). Similarly, production of Stx2f in the three STEC strains under non-inducing conditions varied greatly. In strain RM7007 from pigeon and strain RM10064 from water, the Stx2f was below 5.0 ng/mL, while in another isolate from water, RM16643, Stx2f production was not statistically different from the *stx*-negative strains (Table 2).

**Table 2.** Production of Stx2f and cytotoxicity in *E. albertii*.

Strains (Serotype) <sup>1</sup>	Sources	stx Subtype	Stx2f (ng/mL)	Cytotoxicity <sup>2</sup>
<i>E. albertii</i>				
RM9973	American crow ( <i>Corvus brachyrhynchos</i> )	stx <sub>2f</sub>	5.6 ± 0.3	++
RM9974	American crow ( <i>Corvus brachyrhynchos</i> )	stx <sub>2f</sub>	28.5 ± 1.6	++++
RM9976	American crow ( <i>Corvus brachyrhynchos</i> )	stx <sub>2f</sub>	0.8 ± 0.2	+
RM10507	Brown-headed cowbird ( <i>Molothrus ater</i> )	stx <sub>2f</sub>	0.1 ± 0.0	+
RM10705	Brown-headed cowbird ( <i>Molothrus ater</i> )	stx <sub>2f</sub>	1.1 ± 0.1	+++
RM15112	Oregon Junco ( <i>Junco hyemalis</i> )	stx <sub>2f</sub>	0.6 ± 0.0	++
RM15113	Oregon Junco ( <i>Junco hyemalis</i> )	stx <sub>2f</sub>	0.3 ± 0.3	+++
RM15114	Oregon Junco ( <i>Junco hyemalis</i> )	stx <sub>2f</sub>	0.4 ± 0.1	++
RM15115	Oregon Junco ( <i>Junco hyemalis</i> )	stx <sub>2f</sub>	0.2 ± 0.0	++
RM15116	White-Breasted Nuthatch ( <i>Sitta carolinensis</i> )	stx <sub>2f</sub>	0.6 ± 0.2	++
2014C-4356	Chicken Carcass	stx negative	0.1 ± 0.1	-
2011C-4180	Human	stx <sub>2f</sub>	3.3 ± 0.4	++
2012EL-1823B	Human	stx <sub>2f</sub>	0.0 ± 0.0	-
2014EL-1348	Human	stx <sub>2f</sub>	0.3 ± 0.1	+
2014C-4015	Human	stx <sub>2f</sub>	0.8 ± 0.6	+
05-3106	Human	stx negative	0.0 ± 0.0	-
07-3866	Human	stx negative	0.0 ± 0.0	-
54-2045	Human	stx negative	0.0 ± 0.0	-
2010C-3449	Human	stx negative	0.1 ± 0.0	-
2013C-4243	Human	stx negative	0.0 ± 0.0	-
<i>E. coli</i>				
RM7007	Pigeon ( <i>Columba livia</i> )	stx <sub>2f</sub>	3.9 ± 0.1	+
RM10064	Water	stx <sub>2f</sub>	1.2 ± 0.1	++
RM16643	Water	stx <sub>2f</sub>	0.0 ± 0.0	-
RM5034 (K-12 strain ATCC#29425)	Human	stx negative	0.0 ± 0.0	-
RM4876 (O157:H7)	Water	stx negative	NA	-
RM1273 (O157:H7)	Human	stx negative	NA	-
EDL933 (O157:H7)	Ground beef	stx <sub>1a</sub> + stx <sub>2a</sub>	NA	+++
RM6013 (O157:H7)	Human	stx <sub>2a</sub>	NA	+++
RM12238 (O145:H28)	Human	stx <sub>2a</sub>	NA	+++

<sup>1</sup> Strain RM6013 is linked to the 2006 spinach-associated outbreak in the U.S.; strain RM12238 is linked to a 2010 romaine lettuce-associated outbreak in the U.S. <sup>2</sup> The cytotoxicity of each strain is classified by the relative EGFP fluorescence at the time of reading. ++++, 0–20%; ++, 20–40%; +, 40–60%; -, 60–80%; -, >80%. The %EGFP and the cytotoxicity are inversely proportional.

### 3.5. Cytotoxicity in *E. albertii* Strains

The Stx-mediated cytotoxicity of the STEA strains was assessed by using the mammalian host Vero-d2EGFP cells and bacterial cell-free culture supernatants. In this Vero cell-based assay, incubation with a high concentration of active Stx resulted in low levels of fluorescence in the Vero-d2EGFP cells, and the detected fluorescence levels in the examined wild bird STEA strains were compared to those of the stx-negative *E. coli* strains, stx<sub>2f</sub>-positive *E. coli* strains, and STEC outbreak strains associated with HUS in infected humans (Table 2). For the ten wild bird STEA strains, the Vero-d2EGFP-based cytotoxicity was all significantly higher than the stx-negative *E. coli* strains (One-way ANOVA,  $p = 0.0005$ ); compared with the stx<sub>2f</sub>-positive *E. coli* strains, strains RM9973, RM9974, RM10705, RM15112, RM15113, and RM15115 exhibited significantly higher cytotoxicity (One-way ANOVA,  $p < 0.05$ ). Although One-Way ANOVA revealed no significant difference in the average of fluorescence values between the wild bird STEA strains and the three STEC outbreak strains (EDL933, RM6013, and RM12238), pairwise strain comparison suggested that strains RM9976, RM10507, and RM15116 displayed lower cytotoxicity than the three outbreak strains (Dunnett's test,  $p < 0.05$ ). Although the average fluorescence values detected for the wild bird strains RM9974 and RM10705 were lower than for any of the three STEC outbreak strains (EDL933, RM6013, and RM12238), the differences were not statistically sig-

nificant (Dunnet's test,  $p = 0.0605$ ). When examining the clinical STEA, strains 2011C-4180, 2014C-4015, and 2014EL-1348 exhibited higher cytotoxicity than those of *stx<sub>2f</sub>*-negative *E. albertii* strains (paired *t*-test,  $p < 0.05$ ). There was no significant difference in the average of fluorescence values between the strain 2012-EL-1823B (*stx<sub>2f</sub>*-positive) and any of the *stx<sub>2f</sub>*-negative *E. albertii* strains (Table 2).

#### 4. Discussion

*E. albertii* was recognized almost 20 years ago to be closely related to strains of *Shigella boydii* and estimated to have separated from an *E. coli*-like ancestor about 28 million years ago [3]. Recent genomic-based studies revealed remarkable genetic diversity within this species [9,12,17,25], providing supportive evidence for the evolutionary radiation of this species in exploiting a wide range of ecological niches. Consistently, our comparative genomic analyses of 20 complete *E. albertii* genomes revealed a large pangenome and a very small core genome, supporting that, like *E. coli*, *E. albertii* holds an open pangenome and evolves more rapidly compared to the species with closed pangenesomes [52]. Among the ten wild bird strains, four Oregon junco strains and one white-breasted nuthatch strain carry the same sequence type (ST7971) and are phylogenetically closely related to clinical strains 2014C-4015 (ST5992), 2014EL-1348 (ST5990), and 2012EL-1823 (ST5983) (Supplementary Figure S1), consistent with the core gene-based phylogeny (Figure 1). Therefore, these strains are within the Clade II lineage, according to a previous report on the phylogenetic diversity and population structure of *E. albertii* [53]. The other wild bird strains that are phylogenetically highly related to the above clinical strains include one American crow strain (RM9974, ST2682) and one brown-headed cowbird strain (RM10705, ST9018) since both core genome- and MLST-based phylogenetic analyses placed both strains in the same cluster as the above three clinical strains (Figure 1 and Supplementary Figure S1).

A search of Enterobase revealed that strains belonging to ST7971 have been isolated previously from spinach (BioSample number: SAMN12230424) and stool samples (BioSample: SAMN08199346), and both were reported as *E. coli* isolates. ST5969, ST5983, ST5992, and ST7833 differ from ST7971 with one locus. Strains belonging to one of the above four STs include strains recovered from kale leaves (BioSample: SAMN04422427), human blood (BioSample: SAMN10875686), and a diarrheal patient in the United Kingdom (Strain 367208). A search of Enterobase revealed 12 strains belonging to ST2682, same as the American crow strain RM9974, among which a majority of these were from wild birds isolated in Japan, Australia, and the United Kingdom. Of the two U.S. strains carrying ST2682, one was isolated from vegetable (BioSample: SAMN04429912) while the other one was a clinical isolate (BioSample: SAMN32772508). Strains belonging to ST9018, same as the brown-headed cowbird strain RM10705, included a river water isolate (BioSample: SAMN08797203) and a municipal water isolate (BioSample: SAMN08797189) reported by Agriculture and Agri-Food Canada and a clinical strain (BioSample: SAMN12100513) reported by the CDC. The sequence type classified for the American crow strain RM9973, ST6057, was also classified for a clinical *E. albertii* isolate reported in 2016 (BioSample: SAMD00006218). The presence of the clinically relevant *E. albertii* genotypes recovered from various environmental samples may suggest diverse sources of potential clinical infections.

Current knowledge about the prevalence, epidemiology, and pathogenesis of *E. albertii* is limited due to historical misclassification of this species as *E. coli*, *Shigella boydii*, or *H. alvei* [7,8]. The best-known virulence determinants in *E. albertii* are located on the LEE, similar to the LEEs in EPEC/EHEC, a PAI containing genes (*eae* and *tir*) responsible for the initial adherence of pathogen cells to the host epithelial cell surfaces as well as the genes encoding T3SS apparatus proteins and effectors. Like EPEC/EHEC strains, *E. albertii* appears to have acquired an LEE via horizontal gene transfer since the GC contents of the LEEs (38.6–40.7%) examined in this study were all much lower than the average GC contents of the corresponding chromosomes (49.5–50.1%). All LEEs in the *E. albertii* strains examined were uniformly integrated next to the tRNA gene *phelU*, and this observation was consistent with a previous report stating that the *phelU* locus is a preferable integration

site for LEE in *E. albertii* [7]. Furthermore, diverse intimin subtypes were detected. The dominant subtype was sigma, present in seven wild bird strains and five clinical strains. Other subtypes included alpha, upsilon, and xi in the wild bird strains and alpha, iota, nu, omicron, and tau in the clinical strains. As previously documented, intimin-mediated pathogen adherence is the first step in the formation of A/E lesions during the pathogenesis of attaching-and-effacing *E. coli* (AEEC) strains [54]. Additionally, intimin is thought to be responsible for host specificity and tissue cell tropism because EHEC strains harboring different intimin subtypes exhibited different colonization patterns [55]. It is unknown if intimin subtype sigma confers *E. albertii* with any host colonization specificity, such as birds. Currently, there are nearly 40 intimin subtypes reported, and more studies are needed to understand the prevalence, host/tissue specificity, and clinical relevance of other intimin subtypes.

The other well-known virulence factor in *E. albertii* is CDT, a tripartite holotoxin that induces eukaryotic cell cycle arrest and eventually leads to apoptosis [56]. Production of CDT is thought to be associated with increased persistence, invasion, and disease severity in various bacterial pathogens [57,58]. Because multiple copies of *cdt* genes were detected in all *stx<sub>2f</sub>*-positive *E. albertii* strains, we further examined the chromosome locations and the relatedness of the *cdt* operons in the examined bacterial strains. Interestingly, in all except the wild bird strain RM9976, one set of *cdtABC* was located on the *Stx2f*-prophage genome. In strain RM9976, the *cdtABC* genes were on a prophage (chromosome location 1,103,419–1,169,082) but not the *Stx2f*-prophage. In *E. coli*, CDT has been classified into five subtypes based on the sequence variations in *cdtB*, although a published study suggested that *cdtB*-II was associated with *E. albertii* [56,59]. Sequence comparison of the *E. albertii cdtB* genes with the *E. coli cdtB* I-V subtypes revealed that all prophage-encoded *cdtB* genes were grouped with the *cdtB*-I while others were grouped with *cdtB*-II or *cdtB*-III/V (clinical strains 2013C-4143, 05-3106, and 2011C-4180) (Supplementary Figure S2), supporting a previous report that CDT type I and IV genes are framed with lambdoid prophage genes [56].

Consistent with previous reports documenting *Stx2f* to be the predominant subtype in *E. albertii* [7,11,33,60], all wild bird *E. albertii* strains examined in this study carried phage-borne *stx<sub>2f</sub>* genes. The complete genome sequences from this study allowed the *Stx2f*-prophage integration sites to be mapped and compared with the ones in *E. coli*. The most common site, the tmRNA gene *ssrA*, is also commonly utilized in *E. coli* and in other members of the family *Enterobacteriaceae* for integration of bacteriophages, including *Stx*-phages and phages carrying other virulence genes, e.g., T3SS effector genes [61–63]. Although the tRNA gene *serT* has not been reported as an *Stx*-phage integration site in *E. coli*, the tRNA gene *serU* is known for *Stx*-phage integration in an *E. coli* O121 strain [43]. The insertion sites in the wild bird strains RM9973 and RM9974 are located within a hypothetical gene that is widespread in *E. albertii* and *E. coli* populations.

The substantial variation in *Stx2f* production observed in this study could not be explained by the difference in regulatory elements of the *stx<sub>2f</sub>* genes or the genetic makeup of *Stx2f*-prophages. The *Stx2f*-prophages in strains RM9973 and RM9974 were nearly identical and utilized the same chromosome integration site. Furthermore, the DNA sequences upstream of the *stx<sub>2f</sub>* coding sequence, as well as the gene encoding antitermination protein, were identical in the two strains. The vast difference in the spontaneous *Stx2f* production between the two strains may be due to the difference in the physiological state of host bacteria, which is a known factor that influences *Stx* production [64,65]. Strain variation in spontaneous induction of *Stx1a* and *Stx2a* was reported in a group of STEC O145:H28 strains carrying the *Stx*-prophages within the same lineage [66]. Two groups of antitermination protein were identified in *E. albertii* strains examined in this study, and the majority of STEA strains carried a 123 aa protein belonging to pfam 06350, while strains RM9976 and 2011C-4180 carried a 276 aa protein belonging to pfam 03589. Further studies are needed to understand whether, like many *Stx*-prophages in *E. coli*, *Stx2f*-prophages in *E. albertii* are agent inducible, the conditions that induce the production of *Stx2f*, as well as the role of the antitermination proteins in the production of *Stx2f*.

Given that *stx*<sub>2f</sub>-positive STEC strains have rarely been associated with HUS, the clinical significance of STEA is likely underestimated due to the misclassification of this pathogen as *E. coli*. In the first report of a clinical *E. albertii* strain expressing Stx2f, a re-examination of previously identified EPEC or EHEC strains using routine classification systems adopted in clinical labs revealed that 15% of them were *E. albertii* [7]. Stx2f is thought to be less toxic to mammalian cells than Stx2a but more stable at low pH and high temperature, according to a study using purified toxins [67]. Our study uncovered that the majority of STEA wild bird strains exhibited higher cytotoxicity than the *stx*<sub>2f</sub>-positive *E. coli* strains, and several of them displayed a comparable cytotoxicity level with the *stx*<sub>2a</sub>-positive EHEC outbreak strains and an EHEC strain carrying both *stx*<sub>1a</sub> and *stx*<sub>2a</sub>, consistent with the finding that the Stx2f-expressing *E. albertii* strain AKT5 had a similar level of cytotoxicity as the EHEC outbreak strain Sakai [68]. The increased cytotoxicity in the *stx*<sub>2f</sub>-positive *E. albertii* strains might be attributed to other toxins, such as CDT. CDT produced from *E. coli* and *Campylobacter* spp. can induce cellular distension, cell cycle arrest, and cell death in the mammalian Vero cells, although the receptors mediating the entry of CDT are unknown [58]. A BLASTn search of the genomes of the EHEC strains examined in this study and the genome of the strain Sakai did not identify any homologs of *cdtABC* genes. Further genetic and biochemical studies are required to reveal if CDTs have contributed to the increased cytotoxicity in the STEA strains detected in our study.

Rapid identification of foodborne pathogens that have the potential to cause severe human disease is one of the biggest challenges in food safety. For STEC, a combination of Stx subtypes and the adherence capacity has been used traditionally to predict the pathogenicity potential of a strain [69]. Because the subtypes Stx1a, Stx2a, and Stx2d are most frequently implicated in causing severe human illness, including HUS [70], strains that possess one of the above Stx subtypes and also express the intimin adhesin, such as in EHEC strains, are considered high-risk STEC strains. Although the majority of the STEA strains characterized to date have been found to carry *stx*<sub>2f</sub> genes only, an *stx*<sub>2a</sub>-positive *E. albertii* strain was isolated from a patient who had developed bloody diarrhea [22], suggesting that, similarly to STEC, *E. albertii* acquires virulence genes via horizontal gene transfer in a lineage-independent manner. In fact, our study revealed the significant contribution of MGEs in shaping the virulence gene repertoire in *E. albertii*. Besides LEE, pathogenicity islands OI-122 and HPI were detected. In addition to genes encoding Stx and CDTs, genes encoding the T3SS effectors were frequently associated with the prophage genomes in *E. albertii*. Although the enterohemolysin genes (*hlyCABD*), often present on the large virulence plasmid in EHEC strains, were not detected in *E. albertii* strains examined in this study, *hlyCABD*-positive *E. albertii* strains were reported in China [25]. In addition to vast genetic diversity, phenotypic variation, including Stx2f production and cytotoxicity revealed in our study, implies that *E. albertii* strains are likely varying in their pathogenicity, even among the strains belonging to the same genomic lineage. Loss-of-function mutations in genes encoding master transcriptional regulators are known to contribute to the strain variation in stress resistance and biofilm formation in EHEC strains. Knowledge about the regulation of virulence gene expression in *E. albertii* is limited. Further studies are needed to illustrate the impact of bacterial and environmental factors on the expression of virulence genes and pathogenicity traits, as well as to determine how ecological niches may impose a selective pressure for these traits. Such information will be valuable in developing a better understanding of the pathogenesis of *E. albertii* and an effective mitigation strategy for the detection, identification, and control of this emerging avian and human pathogen.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/microorganisms1112803/s1>, Table S1. Detection of *E. coli* T3SS effector genes in *E. albertii*. Table S2. Detection of toxin-related genes in *E. albertii*. Figure S1. Phylogeny of *E. albertii* strains based on sequences of the MLST loci. The concatenated sequences of the seven loci (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*) were aligned using Clustal Omega, and the consensus sequence was used to construct phylogeny tree in Geneious Prime® as detailed in Materials and Methods. The sequence type (ST) was determined using MLST 2.0 service at the Center for

Genomic Epidemiology with the *Escherichia coli* #1 configuration. Figure S2. Phylogeny of *E. albertii* *cdtB*. Boldface indicates reference genes for five subtypes. The GenBank accession numbers are shown in parentheses. The *cdtB* gene located on the Stx2f-prophage in strain 2012EL-1823 was not included in the analyses due to a large insertional mutation in the *cdtB* gene (Loucs tag, AYN53\_RS14910).

**Author Contributions:** Conceptualization, M.Q.C., P.S.G.C. and R.L.L.; Methodology, M.Q.C., B.Q., X.H., A.P., D.C., M.B.C., C.-C.L., P.S.G.C., R.L.L. and J.L.B.; Formal analysis, M.Q.C., B.Q., X.H., A.P., D.C., M.B.C., C.-C.L., P.S.G.C. and J.L.B.; Resources, M.Q.C., M.B.C. and R.L.L.; Data curation, B.Q., X.H., A.P., D.C., C.-C.L., R.L.L. and J.L.B.; Writing—original draft, M.Q.C.; Writing—review & editing, M.Q.C., B.Q., X.H., A.P., D.C., M.B.C., C.-C.L., P.S.G.C., R.L.L. and J.L.B.; Project administration, M.Q.C.; Funding acquisition, M.Q.C., B.Q., X.H., M.B.C. and R.L.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by USDA-ARS CRIS projects 2030-42000-049-00D, 2030-42000-052-000D, and 2030-42000-055-000D, and the Advanced Molecular Detection (AMD) Initiative grant number AMD-21 at the Centers for Disease Control and Prevention.

**Data Availability Statement:** The data presented in this study are openly available in GenBank under the BioProject numbers PRJNA561565 and PRJNA218110. The GenBank accession numbers are listed in Table 1.

**Acknowledgments:** We wish to thank Nicole Laniohan and Wenxian Du for providing technical assistance and Bertram Lee for his assistance in statistical analyses.

**Conflicts of Interest:** The findings and conclusions in this report are those of the author(s) and do not reflect the view of the Centers for Disease Control and Prevention, the Department of Health and Human Services, or the United States government. Furthermore, the use of any product names, trade names, images, or commercial sources is for identification purposes only, and does not imply endorsement or government sanction by the U.S. Department of Health and Human Services.

## References

- Albert, M.J.; Alam, K.; Islam, M.; Montanaro, J.; Rahaman, A.S.; Haider, K.; Hossain, M.A.; Kibriya, A.K.; Tzipori, S. *Hafnia alvei*, a probable cause of diarrhea in humans. *Infect. Immun.* **1991**, *59*, 1507–1513. [[CrossRef](#)] [[PubMed](#)]
- Huys, G.; Cnockaert, M.; Janda, J.M.; Swings, J. *Escherichia albertii* sp. nov., a diarrhoeagenic species isolated from stool specimens of Bangladeshi children. *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 807–810. [[CrossRef](#)] [[PubMed](#)]
- Hyma, K.E.; Lacher, D.W.; Nelson, A.M.; Bumbaugh, A.C.; Janda, J.M.; Strockbine, N.A.; Young, V.B.; Whittam, T.S. Evolutionary genetics of a new pathogenic *Escherichia* species: *Escherichia albertii* and related *Shigella boydii* strains. *J. Bacteriol.* **2005**, *187*, 619–628. [[CrossRef](#)] [[PubMed](#)]
- Walk, S.T.; Alm, E.W.; Gordon, D.M.; Ram, J.L.; Toranzos, G.A.; Tiedje, J.M.; Whittam, T.S. Cryptic lineages of the genus *Escherichia*. *Appl. Environ. Microbiol.* **2009**, *75*, 6534–6544. [[CrossRef](#)]
- Foster, G.; Ross, H.M.; Pennycott, T.W.; Hopkins, G.F.; McLaren, I.M. Isolation of *Escherichia coli* O86:K61 producing cyto-lethal distending toxin from wild birds of the finch family. *Lett. Appl. Microbiol.* **1998**, *26*, 395–398. [[CrossRef](#)]
- Pennycott, T.W.; Ross, H.M.; McLaren, I.M.; Park, A.; Hopkins, G.F.; Foster, G. Causes of death of wild birds of the family *Fringillidae* in Britain. *Vet. Rec.* **1998**, *143*, 155–158. [[CrossRef](#)]
- Ooka, T.; Seto, K.; Kawano, K.; Kobayashi, H.; Etoh, Y.; Ichihara, S.; Kaneko, A.; Isobe, J.; Yamaguchi, K.; Horikawa, K.; et al. Clinical significance of *Escherichia albertii*. *Emerg. Infect. Dis.* **2012**, *18*, 488–492. [[CrossRef](#)]
- Gomes, T.A.T.; Ooka, T.; Hernandes, R.T.; Yamamoto, D.; Hayashi, T. *Escherichia albertii* Pathogenesis. *EcoSal Plus* **2020**, *9*, 10–1128. [[CrossRef](#)]
- Masuda, K.; Ooka, T.; Akita, H.; Hiratsuka, T.; Takao, S.; Fukada, M.; Inoue, K.; Honda, M.; Toda, J.; Sugitani, W.; et al. Epidemiological Aspects of *Escherichia albertii* Outbreaks in Japan and Genetic Characteristics of the Causative Pathogen. *Foodborne Pathog. Dis.* **2020**, *17*, 144–150. [[CrossRef](#)]
- Konno, T.; Yatsuyanagi, J.; Takahashi, S.; Kumagai, Y.; Wada, E.; Chiba, M.; Saito, S. Isolation and identification of *Escherichia albertii* from a patient in an outbreak of gastroenteritis. *Jpn. J. Infect. Dis.* **2012**, *65*, 203–207. [[CrossRef](#)]
- Ori, E.L.; Takagi, E.H.; Andrade, T.S.; Miguel, B.T.; Cergole-Novella, M.C.; Guth, B.E.C.; Hernandes, R.T.; Dias, R.C.B.; Pinheiro, S.R.S.; Camargo, C.H.; et al. Diarrhoeagenic *Escherichia coli* and *Escherichia albertii* in Brazil: Pathotypes and serotypes over a 6-year period of surveillance. *Epidemiol. Infect.* **2018**, *147*, e10. [[CrossRef](#)] [[PubMed](#)]
- Bengtsson, R.J.; Baker, K.S.; Cunningham, A.A.; Greig, D.R.; John, S.K.; Macgregor, S.K.; Seilern-Moy, K.; Spiro, S.; Chong, C.C.; De Silva, P.M.; et al. The genomic epidemiology of *Escherichia albertii* infecting humans and birds in Great Britain. *Nat. Commun.* **2023**, *14*, 1707. [[CrossRef](#)] [[PubMed](#)]

13. Iguchi, A.; Takemura, T.; Ogura, Y.; Nguyen, T.T.H.; Kikuchi, T.; Okuno, M.; Tokizawa, A.; Iwashita, H.; Pham, H.Q.A.; Doan, T.H.; et al. Genomic characterization of endemic diarrheagenic *Escherichia coli* and *Escherichia albertii* from infants with diarrhea in Vietnam. *PLoS Negl. Trop. Dis.* **2023**, *17*, e0011259. [CrossRef] [PubMed]
14. Ooka, T.; Tokuoka, E.; Furukawa, M.; Nagamura, T.; Ogura, Y.; Arisawa, K.; Harada, S.; Hayashi, T. Human gastroenteritis outbreak associated with *Escherichia albertii*, Japan. *Emerg. Infect. Dis.* **2013**, *19*, 144–146. [CrossRef]
15. Muchaamba, F.; Barnettler, K.; Treier, A.; Houf, K.; Stephan, R. Microbiology and Epidemiology of *Escherichia albertii*—An Emerging Elusive Foodborne Pathogen. *Microorganisms* **2022**, *10*, 875. [CrossRef]
16. Albert, M.J.; Faruque, S.M.; Ansaruzzaman, M.; Islam, M.M.; Haider, K.; Alam, K.; Kabir, I.; Robins-Browne, R. Sharing of virulence-associated properties at the phenotypic and genetic levels between enteropathogenic *Escherichia coli* and *Hafnia alvei*. *J. Med. Microbiol.* **1992**, *37*, 310–314. [CrossRef]
17. Ooka, T.; Ogura, Y.; Katsura, K.; Seto, K.; Kobayashi, H.; Kawano, K.; Tokuoka, E.; Furukawa, M.; Harada, S.; Yoshino, S.; et al. Defining the Genome Features of *Escherichia albertii*, an Emerging Enteropathogen Closely Related to *Escherichia coli*. *Genome Biol. Evol.* **2015**, *7*, 3170–3179. [CrossRef]
18. Ren, C.P.; Chaudhuri, R.R.; Fivian, A.; Bailey, C.M.; Antonio, M.; Barnes, W.M.; Pallen, M.J. The ETT2 gene cluster, encoding a second type III secretion system from *Escherichia coli*, is present in the majority of strains but has undergone widespread mutational attrition. *J. Bacteriol.* **2004**, *186*, 3547–3560. [CrossRef]
19. Shulman, A.; Yair, Y.; Biran, D.; Sura, T.; Otto, A.; Gophna, U.; Becher, D.; Hecker, M.; Ron, E.Z. The *Escherichia coli* Type III Secretion System 2 Has a Global Effect on Cell Surface. *mBio* **2018**, *9*, e01070-18. [CrossRef]
20. Jinadasa, R.N.; Bloom, S.E.; Weiss, R.S.; Duhamel, G.E. Cytolethal distending toxin: A conserved bacterial genotoxin that blocks cell cycle progression, leading to apoptosis of a broad range of mammalian cell lineages. *Microbiology* **2011**, *157*, 1851–1875. [CrossRef]
21. Fedor, Y.; Vignard, J.; Nicolau-Travers, M.L.; Boutet-Robinet, E.; Watrin, C.; Salles, B.; Mirey, G. From single-strand breaks to double-strand breaks during S-phase: A new mode of action of the *Escherichia coli* Cytolethal Distending Toxin. *Cell Microbiol.* **2013**, *15*, 1–15. [CrossRef]
22. Brandal, L.T.; Tunsgo, H.S.; Ranheim, T.E.; Lobersli, I.; Lange, H.; Wester, A.L. Shiga toxin 2a in *Escherichia albertii*. *J. Clin. Microbiol.* **2015**, *53*, 1454–1455. [CrossRef] [PubMed]
23. Nataro, J.P.; Kaper, J.B. Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* **1998**, *11*, 142–201. [CrossRef]
24. Batisson, I.; Guimond, M.P.; Girard, F.; An, H.; Zhu, C.; Oswald, E.; Fairbrother, J.M.; Jacques, M.; Harel, J. Characterization of the novel factor paa involved in the early steps of the adhesion mechanism of attaching and effacing *Escherichia coli*. *Infect. Immun.* **2003**, *71*, 4516–4525. [CrossRef] [PubMed]
25. Luo, L.; Wang, H.; Payne, M.J.; Liang, C.; Bai, L.; Zheng, H.; Zhang, Z.; Zhang, L.; Zhang, X.; Yan, G.; et al. Comparative genomics of Chinese and international isolates of *Escherichia albertii*: Population structure and evolution of virulence and antimicrobial resistance. *Microb. Genom.* **2021**, *7*, 000710. [CrossRef]
26. Hineno, A.; Li, X.P.; Zeng, X.; Sahin, O.; Moxley, R.A.; Logue, C.M.; Gillespie, B.; Yamasaki, S.; Lin, J. Isolation and characterization of *Escherichia albertii* in poultry at the pre-harvest level. *Zoonoses Public Health* **2021**, *68*, 213–225. [CrossRef]
27. Cooley, M.B.; Quinones, B.; Oryang, D.; Mandrell, R.E.; Gorski, L. Prevalence of shiga toxin producing *Escherichia coli*, *Salmonella enterica*, and *Listeria monocytogenes* at public access watershed sites in a California Central Coast agricultural region. *Front. Cell. Infect. Microbiol.* **2014**, *4*, 30. [CrossRef] [PubMed]
28. Cooley, M.B.; Jay-Russell, M.; Atwill, E.R.; Carychao, D.; Nguyen, K.; Quinones, B.; Patel, R.; Walker, S.; Swimley, M.; Pierre-Jerome, E.; et al. Development of a robust method for isolation of Shiga toxin-positive *Escherichia coli* (STEC) from fecal, plant, soil and water samples from a leafy greens production region in California. *PLoS ONE* **2013**, *8*, e65716. [CrossRef]
29. Perna, N.T.; Plunkett, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **2001**, *409*, 529–533. [CrossRef] [PubMed]
30. Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Blattner, F.R.; Sofia, H.J. The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157: H7. *Nucleic Acids Res.* **1998**, *26*, 4196–4204. [CrossRef]
31. Blattner, F.R.; Plunkett, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F. The complete genome sequence of *Escherichia coli* K-12. *Science* **1997**, *277*, 1453–1462. [CrossRef]
32. Lindsey, R.L.; Fedorka-Cray, P.J.; Abley, M.; Turpin, J.B.; Meinersmann, R.J. Evaluating the occurrence of *Escherichia albertii* in chicken carcass rinses by PCR, Vitek analysis, and sequencing of the *rpoB* gene. *Appl. Environ. Microbiol.* **2015**, *81*, 1727–1734. [CrossRef] [PubMed]
33. Lindsey, R.L.; Rowe, L.A.; Batra, D.; Smith, P.; Strockbine, N.A. PacBio Genome Sequences of Eight *Escherichia albertii* Strains Isolated from Humans in the United States. *Microbiol. Resour. Announc.* **2019**, *8*, e01663-18. [CrossRef] [PubMed]
34. Bachmann, B.J. *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology; Neidhardt, F.C.E.A., Ed.; ASM Press: Washington, DC, USA, 1996.
35. Wirth, T.; Falush, D.; Lan, R.; Colles, F.; Mensa, P.; Wieler, L.H.; Karch, H.; Reeves, P.R.; Maiden, M.C.; Ochman, H.; et al. Sex and virulence in *Escherichia coli*: An evolutionary perspective. *Mol. Microbiol.* **2006**, *60*, 1136–1151. [CrossRef]
36. Carter, M.Q.; Quinones, B.; Laniohan, N.; Carychao, D.; Pham, A.; He, X.; Cooley, M. Pathogenicity assessment of Shiga toxin-producing *Escherichia coli* strains isolated from wild birds in a major agricultural region in California. *Front. Microbiol.* **2023**, *14*, 1214081. [CrossRef] [PubMed]

37. Clawson, M.L.; Keen, J.E.; Smith, T.P.; Durso, L.M.; McDaneld, T.G.; Mandrell, R.E.; Davis, M.A.; Bono, J.L. Phylogenetic classification of *Escherichia coli* O157:H7 strains of human and bovine origin using a novel set of nucleotide polymorphisms. *Genome Biol.* **2009**, *10*, R56. [[CrossRef](#)]
38. Luo, H.; Zhang, C.T.; Gao, F. Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes. *Front. Microbiol.* **2014**, *5*, 482. [[CrossRef](#)] [[PubMed](#)]
39. Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E.P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.D.; Borodovsky, M.; Ostell, J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **2016**, *44*, 6614–6624. [[CrossRef](#)]
40. Carter, M.Q.; Laniohan, N.; Lo, C.C.; Chain, P.S.G. Comparative Genomics Applied to Systematically Assess Pathogenicity Potential in Shiga Toxin-Producing *Escherichia coli* O145:H28. *Microorganisms* **2022**, *10*, 866. [[CrossRef](#)]
41. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [[CrossRef](#)]
42. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [[CrossRef](#)] [[PubMed](#)]
43. Carter, M.Q.; Laniohan, N.; Pham, A.; Quinones, B. Comparative genomic and phenotypic analyses of the virulence potential in Shiga toxin-producing *Escherichia coli* O121:H7 and O121:H10. *Front. Cell. Infect. Microbiol.* **2022**, *12*, 1043726. [[CrossRef](#)] [[PubMed](#)]
44. Aguilar-Bultet, L.; Nicholson, P.; Rychener, L.; Dreyer, M.; Gozel, B.; Origgi, F.C.; Oevermann, A.; Frey, J.; Falquet, L. Genetic Separation of *Listeria monocytogenes* Causing Central Nervous System Infections in Animals. *Front. Cell. Infect. Microbiol.* **2018**, *8*, 20. [[CrossRef](#)]
45. Akarsu, H.; Aguilar-Bultet, L.; Falquet, L. deltaRpm: An R package for a rapid detection of differential gene presence between related bacterial genomes. *BMC Bioinform.* **2019**, *20*, 621. [[CrossRef](#)] [[PubMed](#)]
46. Arndt, D.; Grant, J.R.; Marcu, A.; Sajed, T.; Pon, A.; Liang, Y.; Wishart, D.S. PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **2016**, *44*, W16–W21. [[CrossRef](#)]
47. He, X.; Kong, Q.; Patfield, S.; Skinner, C.; Rasooly, R. A new immunoassay for detecting all subtypes of Shiga toxins produced by Shiga toxin-producing *E. coli* in ground beef. *PLoS ONE* **2016**, *11*, e0148092. [[CrossRef](#)]
48. Skinner, C.; Patfield, S.; Stanker, L.H.; Fratamico, P.; He, X. New high-affinity monoclonal antibodies against Shiga toxin 1 facilitate the detection of hybrid Stx1/Stx2 in vivo. *PLoS ONE* **2014**, *9*, e99854. [[CrossRef](#)]
49. Quiñones, B.; Massey, S.; Friedman, M.; Swimley, M.S.; Teter, K. Novel cell-based method to detect Shiga toxin 2 from *Escherichia coli* O157:H7 and inhibitors of toxin activity. *Appl. Environ. Microbiol.* **2009**, *75*, 1410–1416. [[CrossRef](#)]
50. Schubert, S.; Rakin, A.; Karch, H.; Carniel, E.; Heesemann, J. Prevalence of the “high-pathogenicity island” of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. *Infect. Immun.* **1998**, *66*, 480–485. [[CrossRef](#)]
51. Karch, H.; Schubert, S.; Zhang, D.; Zhang, W.; Schmidt, H.; Olschlager, T.; Hacker, J. A genomic island, termed high-pathogenicity island, is present in certain non-O157 Shiga toxin-producing *Escherichia coli* clonal lineages. *Infect. Immun.* **1999**, *67*, 5994–6001. [[CrossRef](#)]
52. Rasko, D.A.; Rosovitz, M.J.; Myers, G.S.; Mongodin, E.F.; Fricke, W.F.; Gajer, P.; Crabtree, J.; Sebaihia, M.; Thomson, N.R.; Chaudhuri, R.; et al. The pangenome structure of *Escherichia coli*: Comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* **2008**, *190*, 6881–6893. [[CrossRef](#)] [[PubMed](#)]
53. Ooka, T.; Seto, K.; Ogura, Y.; Nakamura, K.; Iguchi, A.; Gotoh, Y.; Honda, M.; Etoh, Y.; Ikeda, T.; Sugitani, W.; et al. O-antigen biosynthesis gene clusters of *Escherichia albertii*: Their diversity and similarity to *Escherichia coli* gene clusters and the development of an O-genotyping method. *Microb. Genom.* **2019**, *5*, e000314. [[CrossRef](#)] [[PubMed](#)]
54. Franzin, F.M.; Sircili, M.P. Locus of enterocyte effacement: A pathogenicity island involved in the virulence of enteropathogenic and enterohemorrhagic *Escherichia coli* subjected to a complex network of gene regulation. *Biomed. Res. Int.* **2015**, *2015*, 534738. [[CrossRef](#)] [[PubMed](#)]
55. Phillips, A.D.; Frankel, G. Intimin-mediated tissue specificity in enteropathogenic *Escherichia coli* interaction with human intestinal organ cultures. *J. Infect. Dis.* **2000**, *181*, 1496–1500. [[CrossRef](#)] [[PubMed](#)]
56. Toth, I.; Nougayrede, J.P.; Dobrindt, U.; Ledger, T.N.; Boury, M.; Morabito, S.; Fujiwara, T.; Sugai, M.; Hacker, J.; Oswald, E. Cytolethal distending toxin type I and type IV genes are framed with lambdoid prophage genes in extraintestinal pathogenic *Escherichia coli*. *Infect. Immun.* **2009**, *77*, 492–500. [[CrossRef](#)]
57. Hinenoaya, A.; Yasuda, N.; Hibino, T.; Shima, A.; Nagita, A.; Tsukamoto, T.; Yamasaki, S. Isolation and Characterization of an *Escherichia albertii* Strain Producing Three Different Toxins from a Child with Diarrhea. *Jpn. J. Infect. Dis.* **2017**, *70*, 252–257. [[CrossRef](#)] [[PubMed](#)]
58. Smith, J.L.; Bayles, D.O. The contribution of cytolethal distending toxin to bacterial pathogenesis. *Crit. Rev. Microbiol.* **2006**, *32*, 227–248. [[CrossRef](#)]
59. Hinenoaya, A.; Yasuda, N.; Mukaizawa, N.; Sheikh, S.; Niwa, Y.; Awasthi, S.P.; Asakura, M.; Tsukamoto, T.; Nagita, A.; Albert, M.J.; et al. Association of cytolethal distending toxin-II gene-positive *Escherichia coli* with *Escherichia albertii*, an emerging enteropathogen. *Int. J. Med. Microbiol.* **2017**, *307*, 564–571. [[CrossRef](#)]
60. Murakami, K.; Etoh, Y.; Tanaka, E.; Ichihara, S.; Horikawa, K.; Kawano, K.; Ooka, T.; Kawamura, Y.; Ito, K. Shiga toxin 2f-producing *Escherichia albertii* from a symptomatic human. *Jpn. J. Infect. Dis.* **2014**, *67*, 204–208. [[CrossRef](#)]

61. Creuzburg, K.; Kohler, B.; Hempel, H.; Schreier, P.; Jacobs, E.; Schmidt, H. Genetic structure and chromosomal integration site of the cryptic prophage CP-1639 encoding Shiga toxin 1. *Microbiology* **2005**, *151*, 941–950. [[CrossRef](#)]
62. Bonanno, L.; Loukiadis, E.; Mariani-Kurkdjian, P.; Oswald, E.; Garnier, L.; Michel, V.; Auvray, F. Diversity of Shiga Toxin-Producing *Escherichia coli* (STEC) O26:H11 Strains Examined via *stx* Subtypes and Insertion Sites of Stx and EspK Bacteriophages. *Appl. Environ. Microbiol.* **2015**, *81*, 3712–3721. [[CrossRef](#)] [[PubMed](#)]
63. Tamilmaran, N.; Sankaranarayanan, R.; Selvakumar, A.S.P.; Munavar, M.H. Horizontal transfer of domains in *ssrA* gene among Enterobacteriaceae. *Genes Cells* **2021**, *26*, 541–550. [[CrossRef](#)] [[PubMed](#)]
64. Livny, J.; Friedman, D.I. Characterizing spontaneous induction of Stx encoding phages using a selectable reporter system. *Mol. Microbiol.* **2004**, *51*, 1691–1704. [[CrossRef](#)] [[PubMed](#)]
65. Filipiak, M.; Los, J.M.; Los, M. Efficiency of induction of Shiga-toxin lambdoid prophages in *Escherichia coli* due to oxidative and antibiotic stress depends on the combination of prophage and the bacterial strain. *J. Appl. Genet.* **2020**, *61*, 131–140. [[CrossRef](#)] [[PubMed](#)]
66. Carter, M.Q.; Pham, A.; Du, W.X.; He, X. Differential induction of Shiga toxin in environmental *Escherichia coli* O145:H28 strains carrying the same genotype as the outbreak strains. *Int. J. Food Microbiol.* **2021**, *339*, 109029. [[CrossRef](#)] [[PubMed](#)]
67. Skinner, C.; Patfield, S.; Stanker, L.; He, X. Development of monoclonal antibodies and immunoassays for sensitive and specific detection of Shiga toxin Stx2f. *PLoS ONE* **2013**, *8*, e76563. [[CrossRef](#)] [[PubMed](#)]
68. Hinenoaya, A.; Ichimura, H.; Awasthi, S.P.; Yasuda, N.; Yatsuyanagi, J.; Yamasaki, S. Phenotypic and molecular characterization of *Escherichia albertii*: Further surrogates to avoid potential laboratory misidentification. *Int. J. Med. Microbiol.* **2019**, *309*, 108–115. [[CrossRef](#)]
69. National Advisory Committee on Microbiological Criteria for Foods. Response to Questions Posed by the Food and Drug Administration Regarding Virulence Factors and Attributes that Define Foodborne Shiga Toxin-Producing *Escherichia coli* (STEC) as Severe Human Pathogens (dagger). *J. Food Prot.* **2019**, *82*, 724–767. [[CrossRef](#)]
70. Scheutz, F. Taxonomy Meets Public Health: The Case of Shiga Toxin-Producing *Escherichia coli*. *Microbiol. Spectr.* **2014**, *2*, 15–36. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.