MICROBIAL GENOMICS

RESEARCH ARTICLE
Wyrsch *et al.*, *Microbial Genomics* 2020;6
DOI 10.1099/mgen.0.000387

MICROBIOLOGY SOCIETY

OPEN DATA OPEN MICROBIOLOGY

Duplication and diversification of a unique chromosomal virulence island hosting the subtilase cytotoxin in *Escherichia coli* ST58

Ethan R. Wyrsch[1,2], Piklu Roy Chowdhury[1,2], Veronica M. Jarocki[1,2], Kate J. Brandis[3] and Steven P. Djordjevic[1,2,*]

Abstract

The AB_5 cytotoxins are important virulence factors in *Escherichia coli*. The most notable members of the AB_5 toxin families include Shiga toxin families 1 (Stx_1) and 2 (Stx_2), which are associated with enterohaemorrhagic *E. coli* infections causing haemolytic uraemic syndrome and haemorrhagic colitis. The subAB toxins are the newest and least well understood members of the AB_5 toxin gene family. The subtilase toxin genes are divided into a plasmid-based variant, *subAB1*, originally described in enterohaemorrhagic *E. coli* O113:H21, and distinct chromosomal variants, *subAB2*, that reside in pathogenicity islands encoding additional virulence effectors. Previously we identified a chromosomal *subAB2* operon within an *E. coli* ST58 strain IBS28 (ONT:H25) taken from a wild ibis nest at an inland wetland in New South Wales, Australia. Here we show the *subAB2* toxin operon comprised part of a 140 kb tRNA–Phe chromosomal island that co-hosted *tia*, encoding an outer-membrane protein that confers an adherence and invasion phenotype and additional virulence and accessory genetic content that potentially originated from known virulence island SE-PAI. This island shared a common evolutionary history with a secondary 90 kb tRNA–Phe pathogenicity island that was presumably generated via a duplication event. IBS28 is closely related [200 single-nucleotide polymorphisms (SNPs)] to four North American ST58 strains. The close relationship between North American isolates of ST58 and IBS28 was further supported by the identification of the only copy of a unique variant of IS26 within the O-antigen gene cluster. Strain ISB28 may be a historically important *E. coli* ST58 genome sequence hosting a progenitor pathogenicity island encoding *subAB*.

DATA SUMMARY

The completed assembly of the chromosome and raw sequencing reads for strain *Escherichia coli* IBS28 have been uploaded to GenBank under BioProject PRJNA591373, with the assembly under accession CP049979 and raw reads in the Sequence Read Archive under accession PRJNA591373.

INTRODUCTION

Escherichia coli is a globally distributed commensal bacterium, colonizing most warm-blooded mammals and avian species, as well as diverse environmental niches [1]. The capacity to successfully colonize such a broad range of environments stems largely from its ability to acquire and integrate genetic cargo from lateral sources. The acquisition of plasmids [2–4], phagemids [5], phages [4, 6–8] and genomic islands [2, 9] has played an important role in shaping the evolution of *E. coli* and other clinically important proteobacteria. Lateral gene transfer (LGT) events underpin the formation of novel hybrid *E. coli* pathogens [10, 11]. The ability to capture genetic material horizontally gives the prospect of closely related bacterial lineages with very different phenotypes; outer-membrane polysaccharide structure (O-antigen) [12], metabolism [13] and the production of toxins and adhesins [14], to name a

Received 25 March 2020; Accepted 18 May 2020; Published 10 June 2020
Author affiliations: [1]ithree institute, University of Technology Sydney, Ultimo, NSW 2007, Australia; [2]The Australian Centre for Genomic Epidemiological Microbiology, University of Technology Sydney, Ultimo, NSW 2007, Australia; [3]Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, University of New South Wales, Kensington 2052 NSW, Australia.
***Correspondence:** Steven P. Djordjevic, Steven.Djordjevic@uts.edu.au
Keywords: *Escherichia coli*; virulence; genomic island; subtilase; cytotoxin.
Abbreviations: CDS, coding sequence; HUS, haemolytic uremic syndrome; IS, insertion sequence; LGT, lateral gene transfer; LT, heat-labile enterotoxin; ORF, open reading frame; PAI, pathogenicity island; SNP, single nucleotide polymorphism; STEC, Shiga-toxigenic *Escherichia coli*; Stx1, Shiga toxin 1; Stx2, Shiga Toxin 2; VAGs, virulence-associated genes.
Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Three supplementary tables and four supplementary figures are available with the online version of this article.
000387 © 2020 The Authors

This is an open-access article distributed under the terms of the Creative Commons Attribution NonCommercial License.

few. In *E. coli*, LGT has shaped pathotype lineages, including the differentiation of intestinal and extra-intestinal pathogens [15, 16]. The AB_5 toxins are particularly significant in this regard and include the Shiga toxins, Stx_1 and Stx_2, that distinguish enterohaemorrhagic *E. coli* [8] and the heat-labile (LT) enterotoxin family [17] that are responsible for significant enteric disease in both humans and porcine agriculture.

In 2004, a new AB_5 toxin known as the subtilase toxin SubAB was described [18]. The genes were found in a Shiga-toxigenic *E. coli* (STEC) strain 98NK2 with serotype O113:H21 that caused haemolytic uraemic syndrome (HUS). The SubAB AB_5 toxin is a serine protease and the endoplasmic reticulum chaperone protein BiP is one of its known targets [19]. Immunoprecipitation experiments using the SubAB toxin as bait with a Vero cell lysate identified several interaction partners, including α2 and β1 integrin, leading the authors to suggest that α2β1 integrin is a receptor for SubAB [20]. The ability of SubAB to bind to α2β1 integrin is likely important to how SubAB enters host cells, engages with the endosomal pathway and traffics intracellularly. Furthermore, the SubB component of the toxin is atypical for AB_5 toxins in that it has specificity for glycans terminating with α2–3-linked N-glycolylneuraminic acid (Neu5Gc) [21].

SubAB is increasingly seen as a significant addition to the virulence genes of *E. coli*. Although the clinical significance of SubAB toxin in humans is poorly understood, the pathological effects of SubAB toxins on mice [22, 23] and rats [24] are substantial and mimic the pathological features of enterohaemorrhagic uraemic syndrome . The *subAB* operon is often described in association with strains that are either LEE-negative STEC or Shiga toxin-negative *E. coli* [25]. The isolation of an intimin-negative (*eae*$^-$), *subAB*-positive (*subAB*$^+$) O128:H2 STEC strain from a patient with prolonged bloody diarrhoea [26] may be a significant observation. O128:H2 is a serotype of *E. coli* typically sourced from sheep and deer but not cattle [27, 28]. Ovine STEC strains are considered to be less pathogenic because they lack intimin and carry variants of *stx*, such as stx_{1c} and stx_{2b} [29–32], which are infrequently associated with STEC that cause HUS. It was suggested that carriage of *subAB* linked to a chromosomal pathogenicity island (PAI) in the human O128:H2 isolate may have ramifications for its ability to cause disease in humans [26], but more comprehensive genomic epidemiological studies are needed.

Currently, four variants of *subAB* (*subAB1*, *subAB2-1*, *subAB2-2* and *subAB2-3*) have been described [25, 33]. *subAB1* is a plasmid-encoded variant found on plasmid pO113 that was originally described in *E. coli* O113:H21 strain 98NK2. STEC with serotype O113:H21 have a reservoir in cattle [28] that spills over into humans, fresh produce and in the environment but they are rarely isolated from sheep [27, 32, 34, 35]. The remaining three variants (*subAB2-1*, *subAB2-2* and *subAB2-3*) reside chromosomally. In Shiga toxin-negative *E. coli* strains ED32 and ED 591 from unrelated cases of childhood diarrhoea, pathogenicity island SE-PAI, which carries *subAB2-1*, spans about 8 kb and is located between *yjhS* and tRNA–Phe [25, 36]. The SE-PAI also carries several putative virulence-associated genes (VAGs) including *tia* that encodes an epithelial cell adhesin and invasin [37], ShiA, whose function in *Shigella flexneri* is to attenuate inflammation by suppressing the innate T-cell immune response [38, 39], a putative sulfatase and an integrase that presumably is involved in the mobilization/capture of the PAI [36].

Impact Statement

Tracing the source, distribution and evolution of specific genetic components from the bacterial pan-genome is proving critical to understanding the development of infectious disease. Here we undertook a comparative analysis of a bacterial strain sourced from an 'environmental' niche hosting the *Escherichia coli* AB_5 protein, subtilase cytotoxin. This Australian strain was closely related to a set of sequences from North America. While only the Australian strain hosted the subtilase toxin, each related ST58 strain hosted the same IS*26* allele within the O-antigen lipopolysaccharide region. This O-antigen region (including the insertion) was also observed in an unrelated *E. coli* ST58 strain, and a separate *E. coli* ST101 strain, both also from North America. The subtilase cytotoxin was hosted in a large genomic island alongside other virulence content often reported in *E. coli*. This genomic island appeared to be formed from a homologous duplication event leading to two large genomic islands with partial shared content. Finally, context of the subtilase toxin suggests that it is a genetic progenitor of many other instances of the toxin that have been reported based on the co-presence of invasin *tia* and other genes, as through this duplication we see the contextual formation of this *tia-subAB* gene configuration.

E. coli ST58 belongs to a phylogroup B1 commensal lineage that has been isolated from diverse animal, environmental and human hosts [40–42], demonstrating links with human urinary tract infections and urosepsis, and it can carry a colV-like virulence resistance plasmid [40]. Recently we characterized an environmental ST58 isolate recovered from the faeces of a straw-necked ibis nest in inland wetlands, New South Wales, Australia [43].

The short-read assembly analysis of this ST58:ONT:H25 isolate (IBS28) indicated that it was closely related to 4 North American strains from a collection of 69 diverse ST58 genomes [43]. Strain ISB28 was distinct from known Australian human ST58 pathogen *E. coli* 2009–52 [40] and avian pathogen *E. coli* A193 [44], which have been described recently. It hosted various VAGs that were uncommon among the other strains recovered from the ibis nests. The profile included immune survival factor *iss*, microcins *mch*, major subunits of the F17 adhesin and the AB_5 subtilase cytotoxin genes *subAB*. Strain IBS28 hosted an ICE with no known VAGs; a relative of an ICE identified in an Australian *E. coli* ST38 strain taken from

hospital-sourced human infection. Strain IBS28 was host to a single IS*6* family element, an insertion sequence (IS) heavily involved in the evolution and dissemination of antimicrobial resistance genes [45]. This insertion was chromosomal within the acquired O-antigen island (non-typeable *in silico*). Here we report the long-read sequencing, assembly and annotation of *E. coli* ST58 strain IBS28 for the purpose of resolving the genetic context of *subAB*, IS*26* and numerous other virulence genes.

METHODS

The isolation and initial analysis of strain *E. coli* IBS28 has been reported previously [43]. The strain was derived from a desiccated faecal sample initially taken from an ibis nest at an inland wetland during 2012, in New South Wales, Australia.

To isolate genomic DNA for long-read sequencing, the strain was first plated onto Lysogeny broth (LB) agar for single colonies and incubated at 37 °C for 16 h. A single colony was then sub-cultured in 2 ml LB broth at 37 °C, with shaking at 220 r.p.m. for 16 h. Genomic DNA was then extracted from lysed cells using XS buffer [46] and purified gently with a phenol/chloroform procedure designed to preserve the integrity of DNA strands for sequencing.

Long-read sequencing was performed on a PacBio Sequel at the Ramaciotti Centre for Genomics at the University of New South Wales, as part of a multiplex protocol. Assembly was performed by the service provider using the HGAP4 pipeline. Genome sequence data were deposited under accession PRJNA591373.

Genomic analysis was performed using several software packages and online servers. Primarily, sequence comparisons and gene identification were performed using the National Center for Biotechnology Information (NCBI) BLASTN/BLASTX (https://blast.ncbi.nlm.nih.gov/Blast.cgi) servers and progressiveMauve [47]. Bacteriophages were identified using PHASTER [48], insertion sequences were identified using ISFinder [49] and genomic islands were identified using IslandViewer 4 [50]. Genomic islands were only annotated if they appeared in outputs from at least two detection algorithms. Single-nucleotide polymorphism (SNP) analysis of *subAB* was performed using AliView [51] utilizing the MUSCLE aligner [52]. GC content analysis was performed using GC Content Calculator (https://www.biologicscorp.com/tools/GCContent) Reference sequences for comparative alignments and phylogenetic analyses were sourced from GenBank (https://www.ncbi.nlm.nih.gov/nucleotide/) and Enterobase (https://enterobase.warwick.ac.uk/).

RESULTS

Sequence summary

The long-read assembly of *E. coli* strain IBS28 resolved two contigs, a 4934027 bp chromosome, plus the ICE pIBS28_1, as described previously [43]. Automated annotation identified 4877 open reading frames (ORFs) and 112 RNA coding sequences in the chromosome. Automated genomic island analyses identified 25 acquired regions, and comparisons to the closely related North American ST58 genomes, Australian ST58 genomes and ST101 strain KSC9 (Fig. 1) suggests the presence of others highlighted by a lack of BLASTN coverage amongst the references. The genomic islands identified in the IBS28 chromosome encode previously identified operons including secretion systems, VgrG-Rhs protein operons, putative adhesins and DNA repair enzymes. Three complete phages were similarly identified, two reported as lambda enterophages with one encoding the serum resistance gene *iss*/*bor*, and one mEp460-like enterophage, each approximately 50 kb in size.

Characterization is IS6 family insertion sequence

A key observation regarding the IBS28 chromosome was the presence of an IS*6* family element, an IS*26* variant with a single C573T SNP (820 bp sequence). The long-read assembly confirms that this element lies within the variable O-antigen region. Alignment analysis revealed that this same O-antigen and IS*6* element were present across the entire local clade of North American ST58 strains most closely related to IBS28, which range from isolations in Canada 1988 to the USA in 2018. Coincidentally, this insertion was identified in another ST58 strain, short-read assembly ESC_LA6312AA_AS (USA, mammal, 1997) from the Enterobase collection utilized in our prior phylogenetic analysis. This additional ST58 strain is phylogenetically distinct from the local clade IBS28 belongs to. This unusual insertion signature also led to the identification of *E. coli* strain KSC9, an ST101 from a 2014 porcine sample from the USA, which also hosted this specific LPS system carrying the IS*26* allele. Our comparative analysis in Fig. 1 demonstrates its presence there. These examples place this chromosomal IS*6* element in human, porcine, bovine and environmental contexts isolated over a 30 year span.

Summary of pathogenicity islands PAI-1 and PAI-2

Critically, amongst the genomic islands we identified in IBS28 were two large related PAIs of approximately 140 and 88 kb. These genomic islands, defined from here as PAI-1 and PAI-2, share much of their genetic content and we hypothesize that they were once a single genomic island that underwent duplication and then diversification through the capture of additional genetic content.

From what could be observed in the sequence data, the progenitor island sat at one of two separate tRNA–Phe–GAA sites in the *E. coli* ST58 chromosome, generated by a DUF4102 domain-containing integrase. At the time of duplication, this island was approximately 21 kb in size. One scenario suggests a recombination event involving a second tRNA–Phe–GAA sequence elsewhere in the chromosome led to the duplication of this region, leaving a second copy in the reverse orientation at the second tRNA–Phe–GAA site. This event also apparently led to the inversion of a 1.3 Mb region of chromosome that sits between the two PAIs (See Fig. 1). Over the course of time, PAI-1 and PAI-2 appear to have subsequently captured diverse sets of unrelated genes. Shared genetic content encoded by

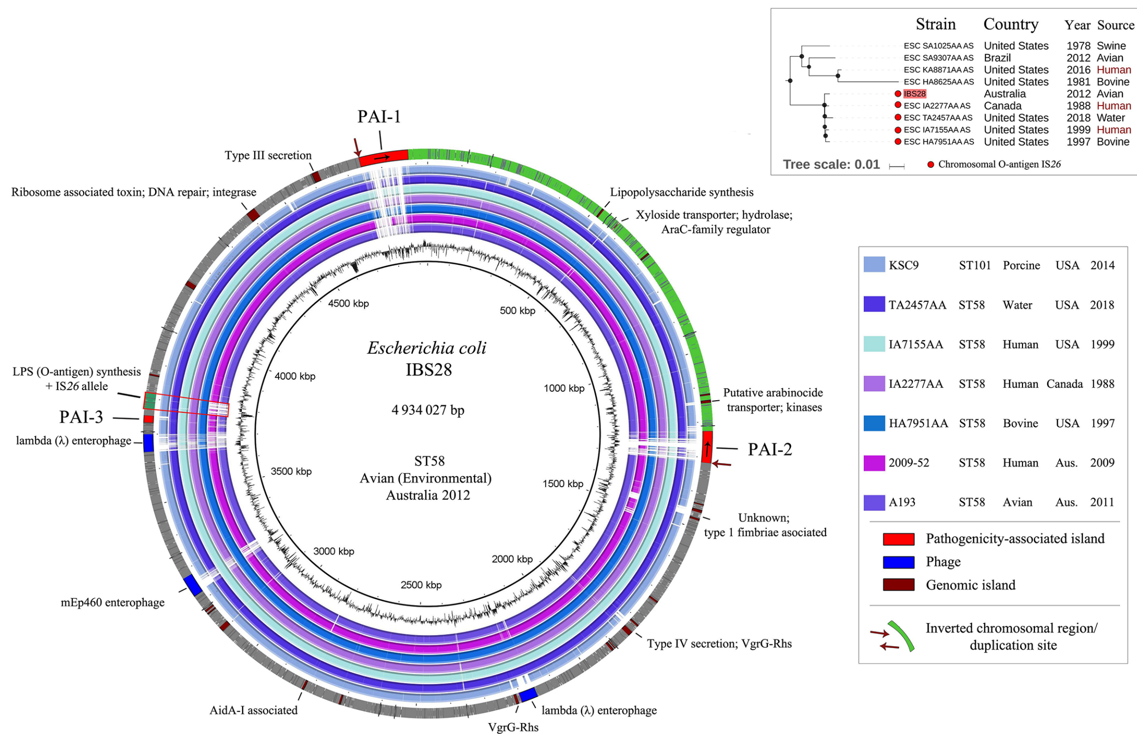

Fig. 1. Map and phylogeny of the *E. coli* strain IBS28 chromosome highlighting genomic islands and other insertions. Coloured rings on the map visualize BLASTN alignment data of other *E. coli* sequences, with the black centre ring indicting GC content for IBS28. A snapshot of a larger ST58 SNP phylogeny including IBS28 is presented in the top right. BLASTN alignments were performed against the most closely related sequences from North America, plus two additional Australian ST58 pathogens.

both islands includes 12 coding sequences (CDS) and 16 hypothetical proteins, including invasin *tia*, autotransporter adhesin Ag43 (Antigen 43) involved in biofilm formation, a YeeVU toxin/antitoxin system and a haemolysin expression-modulating protein. Both islands also have an ISS*fl8*-like element inserted into a Z1226 protein (putative RNA-directed DNA polymerase activity), likely a once unique feature now shared by both PAI-1 and PAI-2, supporting the contention these islands were once a single element. Given these observations, we suggest that PAI-1 and PAI-2 were once a single genomic island, the inferred structure of which will be referred to as PAI-0.

A third partial copy of PAI-0, named PAI-3, was found inserted near the acquired O-antigen lipopolysaccharide (see Fig. 1), but this copy did not include the Z1226 protein or the insertion, so the evolutionary relevance of its presence remains unresolved. A comparison of GC content of this potentially homologous region (Fig. S1, available in the online version of this article) demonstrates some similarity, with each sitting at 52% GC over the 14–17 kb included.

The integrase and much of the other shared content initially forming PAI-0 sat at 94% nucleotide identity between PAI-1 and PAI-2, suggesting the duplication was not a recent event, but the tRNA sequences remained identical.

PAI-1 (139839 kb) encoded 91 ORFs plus 93 hypothetical proteins, with some ORFs broadly separated into virulence-associated (*n*=21), membrane-associated (*n*=11) and metabolism-associated (*n*=10) genes (Fig. 2). The virulence genes of primary concern that PAI-1 has acquired include the subtilase cytotoxin, subtyped as a *subAB2-1*, invasin *tia*, Ag43 and a serine protease autotransporter (*espL*), an RTX toxin, haemagglutinins and immunoglobulin-binding protein *eibA*, an F17 fimbrial adhesin operon (*fimACDG*) and microcin H47 (*mch*/MccH47). In addition to these, it has acquired outer membrane-associated proteins of the TonB, TolC, OmpF and ABC transporter families, amongst others. Metabolism genes included *iroBD*, PTS system components and an [NiFe]-hydrogenase operon. The remaining ORFs comprising the island included polymerases, a toxin/antitoxin system, gene co-activators and neurotensin receptor r8. Insertion sequences of the following families were also present: IS*110*, IS*66*, IS*3* and ISS*fl8*, as mentioned.

PAI-2 (87839 bp) likewise encoded 64 ORFs and 48 hypothetical proteins, with relatively reduced virulence-associated content (*n*=7) and membrane-associated proteins (*n*=4), but an increased number of metabolism-associated genes (*n*=20). Primary virulence genes include putative haemolysin activator *hecB*, an AidA-1 like adhesin, a large filamentous haemagglutinin coupled with a haemolysin activator, and

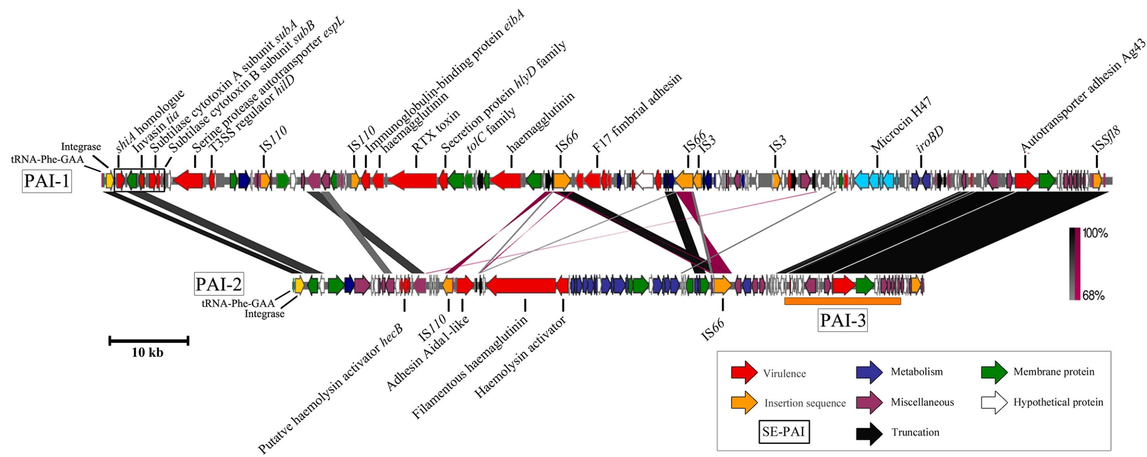

Fig. 2. Maps of large pathogenicity islands PAI-1 and PAI-2. Annotations of the virulence islands with ORFs coloured by function. BLASTN similarity between islands is indicated in black, with reversed sequence indicated in red. Key features are annotated. The region of sequence identified in PAI-3 is indicated by the orange bar under the PAI-2 diagram.

another copy of Ag43 as part of the original PAI-0 sequence. Comparisons between PAI-1 and PAI-2 are shown in Fig. 2.

The full list of genes identified in PAI-1 and PAI-2 can be found in Data File S1, tables S1 and S2 respectively. Additionally, the sporadic presence of these genes can be observed in the ST58 reference sequence BLASTN data presented in Fig. 1.

Characterization of the subtilase cytotoxin operon

To characterize the *subAB2-1* allele and context identified here, available *subAB2* sequences (both CDS-only and those present with context) were downloaded from GenBank. Fig. S2 displays a full SNP alignment of available *subA2* gene sequences. The majority of these sequences were initially sourced from two separate studies identifying *subAB2* from *E.coli* in deer samples from Germany (CDS-only, with some reportedly hosted on plasmids) and human samples from Norway. As reported in other studies, this analysis demonstrated both type 1 and type 2 *subA2* alleles, with the IBS28 allele resolving as the type 1 (See Figs 3 and S3). It should be noted that *subB2* shows no such type differentiation (Fig. S4).

Michelacci *et al.* [36] identified the type 1 *subA2* allele co-localized with the genes *tia* and *shiA* and a sulfatase ORF (*orf*) inserted at a phage integrase in the order: tRNA–integrase–*shiA*–*orf*–*tia*–*subAB2*. This region was described as SE-PAI (JQ994271). This same gene co-localization and organization are observed in every available example of contextual type 1 *subA2* genes, including in PAI-1 of IBS28. Initial BLASTN analyses indicate the presence of different integrases at different tRNA sites hosting this virulence island (data not shown). Here, within this PAI duplication example, we appear to have identified a potential source/progenitor of this *shiA*–*orf*–*tia*–*subAB2* virulence region, and thus the now widely distributed type 1 *subA* allele in general. Shared sequence between PAI-1 and PAI-2 (described as PAI-0) encoded the *orf*–*tia* sequence that was observed centrally in SE-PAI. Our analysis suggests that within PAI-1 the *shiA* and *subAB2* were inserted before and after these genes post-duplication. PAI-2 still encodes the sulfatase ORF and *tia* from PAI-0, although the invasin is highly degraded now at 75% nucleotide identity, a comparatively high rate of nucleotide variation to the rest of shared PAI-0 sequence. Added to this is the fact that regions of at least 10 kb of PAI-1 can be found distributed around the chromosomes of strains carrying SE-PAI (Data File S1, table 3)) and also type 1 *subAB* alleles, which implies that this ST58 chromosome may be the source of this and other large gene organisations, which then disseminated as a coherent genomic island through the chromosomal integrase system. The highest coverage of PAI-1 was seen in CP027672.1, which carried 68% of the island at an average 96.75% nucleotide identity, with the highest fragment size being 16501 bp. As a final point, *subA2* genes reported as type 3 (*subA2-3*; LK985413.1, LK999956.1) show the same SE-PAI context as other type 1 *subA* genes, but these alleles share certain SNPs with the type 2 alleles, giving them an observable subclade amongst the type 1 alleles in Fig. 3.

DISCUSSION

Here we characterized an environmentally associated *E. coli* ST58 (ISB28) from the straw-necked ibis. Our initial analysis utilizing short-read whole-genome data indicated that IBS28 was closely related to a series of strains from North America dating back to 1988, and we detected several virulence genes plus an unusual chromosomal IS6 family element [43]. Members of the IS6 family are frequently involved in the dissemination of antimicrobial resistance genes [53], although none were detected within the strain. The only extra-chromosomal element IBS28 carried was an integrative conjugative element, a homologue to an element we identified from an Australian human clinical *E. coli* isolate, giving it a tentative link to the Australian human microbiome.

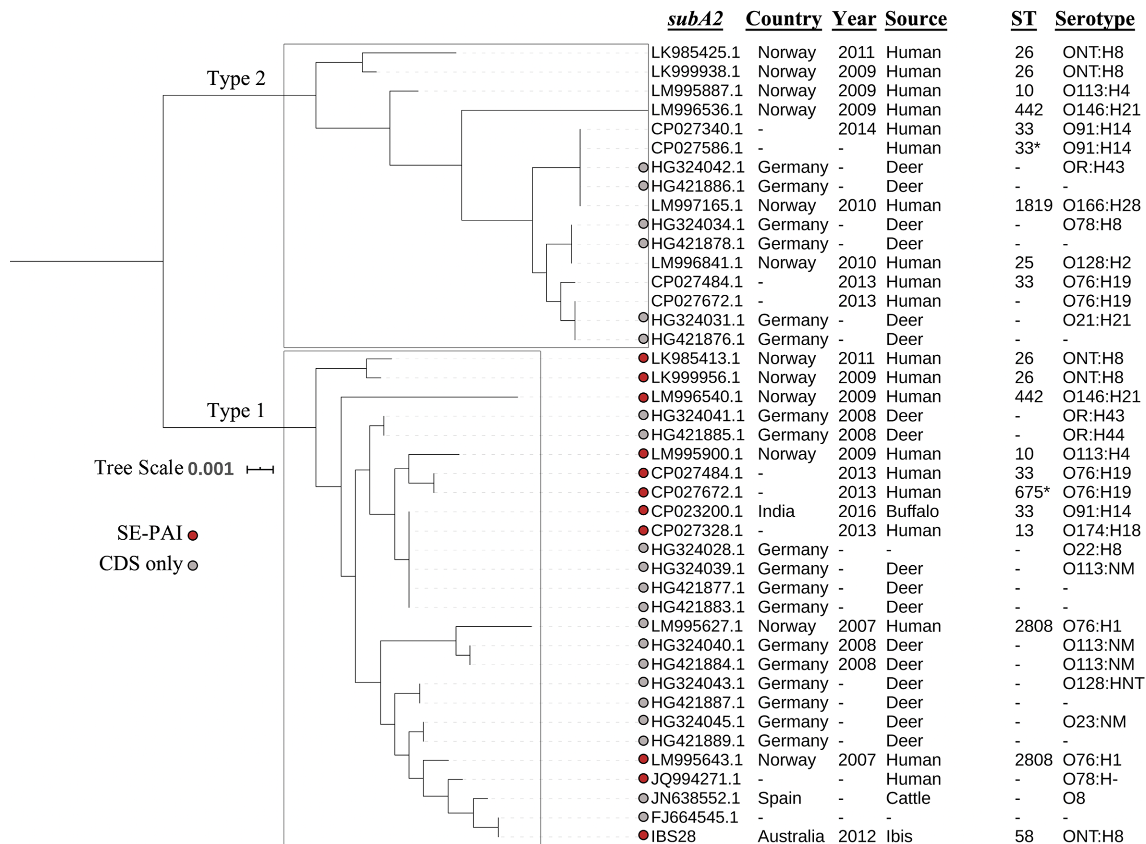

Fig. 3. Phylogeny of *subA2* sequences. Visualization of MUSCLE alignments of *subA2* gene sequences and associated metadata of *E. coli* hosts. Alignment highlights the split between type 1 and type 2 alleles and various subgroups therein. Co-presence of *subA* with *tia* and *shiA* is indicated with a red dot. Grey dots indicate that only the *subAB* CDS were available.

In resolving the completed chromosome, it became apparent that IBS28 hosts significant virulence gene content, as aside from hosting smaller, more commonly observed, genomic islands such as secretion systems [54, 55], it had undergone an interesting chromosomal duplication of an integrase-based genomic island, leading to the formation of two extremely large genomic islands. At 148 kb, the pathogenicity islands we describe here are, to the best of our knowledge, the largest integrase-based genomic islands observed in an *E. coli*. We hypothesize that this extreme level of genetic capture occurred due to the duplication, which led to two assumedly functional copies of the integrase. The islands PAI-1 and PAI-2 are composed of many previously reported chromosomal virulence factors, including SubAB2 [36, 56], H47 siderophore microcin [57] and F17 adhesin [58], and many individual genes such as an RTX toxin [59], haemagglutinins and a protease autotransporter [60]. Whilst these genes encoding these factors are not novel, the appearance of large aggregations of these virulence genes suggests high acquisition rates within this chromosome. The potential fitness costs associated with carrying these large islands may have resulted in the lack of plasmid content [61], although this is strictly conjecture.

The IBS28 genome has both a partial and a complete copy of the previously identified virulence island SE-PAI. Encoding *shiA*–orf–*tia*–*subA*–*subB*, the full set of genes are present in the order previously described [36] at the start of PAI-1, with only orf–*tia* found in PAI-2, with *tia* showing a high level of nucleotide variation compared to the rest of the island. We suggest that the sulfatase ORF and *tia* were the initial genes inserted in PAI-0, and we then see *shiA* and the *subAB2* ORFs inserted before and after in PAI-1 Fig. 2), generating this SE-PAI structure observed in multiple strains across continents. We consider this constructive acquisition scenario to be more likely than if the toxin genes were lost within PAI-2, as there is no evidence of mobile genetic element activity, and nor is there any evidence to suggest that the chromosomal integrase gene can extract or rearrange the captured sequence. As a final indicator, sections of PAI-1 that include the SE-PAI region can be found within the reference strains analysed, suggesting that the SE-PAI element described initially is a much-reduced version of the PAI-1 described here.

Here we have identified and characterized an unusual O-antigen in strain IBS28 hosting an IS*26* allele that is present within the local clade of phylogenetically related ST58 strains, and incidentally observed in unrelated ST58 strain LA6312AA_AS from Enterobase plus an ST101 strain KSC9 from the GenBank nuccore database, each from the USA. The role IS*26* plays in the generation and dissemination

of AMR is well described and, combined with its ability to capture and rearrange small genomic regions into new configurations, it has aided in the development of multiple-drug resistance [62]. Recent work has also demonstrated that IS*26* readily self-associates [63], meaning that the likelihood of an IS*26* element inserting at an already present IS*6* family element is much higher than at a random locus. Given this, we view the O-antigen cassette identified here as a potential chromosomal hotspot for the capture of genetic cargo mobilized by IS26, including antimicrobial resistance genes, that can be disseminated alongside the lipopolysaccharide variant.

Neither PAI-0 nor its now duplicated islands are present in any of the closely related North American strains, but the IS*26* allele is, indicating that several closely related lineages of this ST58 strain have been isolated since 1988, and as recently as 2018 in the USA. In identifying this strain from a nest in New South Wales wetlands circa 2012, we demonstrate the breadth of dissemination of bacterial genes that can play a role in human pathogenesis, and suggest that epidemiological work is performed to characterize the spread of this ST58 strain as a source of chromosomal virulence content in *E. coli*.

Funding information
This work was supported by the Australian Research Council, linkage grant LP150100912. This project was partly funded by the Australian Centre for Genomic Epidemiological Microbiology (AusGEM), a collaborative partnership between the NSW Department of Primary Industries and the University of Technology Sydney. This work was additionally supported by the Centre for Ecosystem Science, University of New South Wales. These results are helping to inform the OUTBREAK decision support system funded by the Medical Research Future Fund Frontier Health and Medical Research Program.

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

References

1. Zhi S, Banting G, Stothard P, Ashbolt NJ, Checkley S *et al*. Evidence for the evolution, clonal expansion and global dissemination of water treatment-resistant naturalized strains of *Escherichia coli* in wastewater. *Water Res* 2019;156:208–222.
2. Cummins ML, Hamidian M, Djordjevic SP. *Salmonella* Genomic Island 1 is Broadly Disseminated within Gammaproteobacteriaceae. *Microorganisms* 2020;8:E161:161.
3. Johnson TJ, Jordan D, Kariyawasam S, Stell AL, Bell NP *et al*. Sequence analysis and characterization of a transferable hybrid plasmid encoding multidrug resistance and enabling zoonotic potential for extraintestinal Escherichia coli. *Infect Immun* 2010;78:1931–1942.
4. Brown Kav A, Rozov R, Bogumil D, Sørensen SJ, Hansen LH *et al*. Unravelling plasmidome distribution and interaction with its hosting microbiome. *Environ Microbiol* 2020;22:32–44.
5. Venturini C, Zingali T, Wyrsch ER, Bowring B, Iredell J *et al*. Diversity of P1 phage-like elements in multidrug resistant Escherichia coli. *Sci Rep* 2019;9:18861.
6. Tóth I, Nougayrède J-P, Dobrindt U, Ledger TN, Boury M *et al*. 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.
7. Brüssow H, Canchaya C, Hardt W-D. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* 2004;68:560–602.
8. Nakamura K, Murase K, Sato MP, Toyoda A, Itoh T *et al*. Differential dynamics and impacts of prophages and plasmids on the pangenome and virulence factor repertoires of Shiga toxin-producing *Escherichia coli* O145:H28. *Microb Genom* 2020;6.
9. Yousuf FA, Rafiq S, Siddiqui R, Khan NA. The role of genomic islands in Escherichia coli K1 interactions with intestinal and kidney epithelial cells. *Microb Pathog* 2016;93:145–151.
10. Bai X, Zhang J, Ambikan A, Jernberg C, Ehricht R *et al*. Molecular characterization and comparative genomics of clinical hybrid Shiga toxin-producing and enterotoxigenic Escherichia coli (STEC/ETEC) strains in Sweden. *Sci Rep* 2019;9:5619.
11. Mariani-Kurkdjian P, Lemaître C, Bidet P, Perez D, Boggini L *et al*. Haemolytic-Uraemic syndrome with bacteraemia caused by a new hybrid *Escherichia coli* pathotype. *New Microbes New Infect* 2014;2:127–131.
12. Seif Y, Monk JM, Machado H, Kavvas E, Palsson BO. Systems Biology and Pangenome of *Salmonella* O-Antigens. *mBio* 2019;10:e01247-19.
13. Pang TY, Lercher MJ. Each of 3,323 metabolic innovations in the evolution of *E. coli* arose through the horizontal transfer of a single DNA segment. *Proc Natl Acad Sci U S A* 2019;116:187–192.
14. Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* 1998;11:142–201.
15. Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2004;2:123–140.
16. Croxen MA, Finlay BB. Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol* 2010;8:26–38.
17. Fleckenstein J, Sheikh A, Qadri F. Novel antigens for enterotoxigenic *Escherichia coli* vaccines. *Expert Rev Vaccines* 2014;13:631–639.
18. Paton AW, Srimanote P, Talbot UM, Wang H, Paton JC. A new family of potent AB(5) cytotoxins produced by Shiga toxigenic Escherichia coli. *J Exp Med* 2004;200:35–46.
19. Paton AW, Beddoe T, Thorpe CM, Whisstock JC, Wilce MCJ *et al*. Ab5 subtilase cytotoxin inactivates the endoplasmic reticulum chaperone BiP. *Nature* 2006;443:548–552.
20. Yahiro K, Morinaga N, Satoh M, Matsuura G, Tomonaga T *et al*. Identification and characterization of receptors for vacuolating activity of subtilase cytotoxin. *Mol Microbiol* 2006;62:480–490.
21. Byres E, Paton AW, Paton JC, Löfling JC, Smith DF *et al*. Incorporation of a non-human glycan mediates human susceptibility to a bacterial toxin. *Nature* 2008;456:648–652.
22. Wang H, Paton JC, Paton AW. Pathologic changes in mice induced by subtilase cytotoxin, a potent new Escherichia coli AB5 toxin that targets the endoplasmic reticulum. *J Infect Dis* 2007;196:1093–1101.
23. Furukawa T, Yahiro K, Tsuji AB, Terasaki Y, Morinaga N *et al*. Fatal hemorrhage induced by subtilase cytotoxin from Shiga-toxigenic *Escherichia coli*. *Microb Pathog* 2011;50:159–167.
24. Seyahian EA, Oltra G, Ochoa F, Melendi S, Hermes R *et al*. Systemic effects of Subtilase cytotoxin produced by *Escherichia coli* O113:H21. *Toxicon* 2017;127:49–55.
25. Tozzoli R, Caprioli A, Cappannella S, Michelacci V, Marziano ML *et al*. Production of the subtilase AB5 cytotoxin by Shiga toxin-negative *Escherichia coli*. *J Clin Microbiol* 2010;48:178–183.
26. Sánchez S, Beristain X, Martínez R, García A, Martín C *et al*. Subtilase cytotoxin encoding genes are present in human, sheep and deer intimin-negative, Shiga toxin-producing *Escherichia coli* O128:H2. *Vet Microbiol* 2012;159:531–535.
27. Djordjevic SP, Hornitzky MA, Bailey G, Gill P, Vanselow B *et al*. Virulence properties and serotypes of Shiga toxin-producing *Escherichia coli* from healthy Australian slaughter-age sheep. *J Clin Microbiol* 2001;39:2017–2021.
28. Hornitzky MA, Vanselow BA, Walker K, Bettelheim KA, Corney B *et al*. Virulence properties and serotypes of Shiga toxin-producing *Escherichia coli* from healthy Australian cattle. *Appl Environ Microbiol* 2002;68:6439–6445.

29. Friedrich AW, Bielaszewska M, Zhang W-L, Pulz M, Kuczius T *et al*. Escherichia coli harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. *J Infect Dis* 2002;185:74–84.
30. Friedrich AW, Borell J, Bielaszewska M, Fruth A, Tschäpe H *et al*. Shiga toxin 1c-producing Escherichia coli strains: phenotypic and genetic characterization and association with human disease. *J Clin Microbiol* 2003;41:2448–2453.
31. Ramachandran V, Brett K, Hornitzky MA, Dowton M, Bettelheim KA *et al*. Distribution of intimin subtypes among Escherichia coli isolates from ruminant and human sources. *J Clin Microbiol* 2003;41:5022–5032.
32. Brett KN, Ramachandran V, Hornitzky MA, Bettelheim KA, Walker MJ *et al*. stx1c is the most common Shiga toxin 1 subtype among Shiga toxin-producing Escherichia coli isolates from sheep but not among isolates from cattle. *J Clin Microbiol* 2003;41:926–936.
33. Tasara T, Fierz L, Klumpp J, Schmidt H, Stephan R. Draft Genome Sequences of Five Shiga Toxin-Producing *Escherichia coli* Isolates Harboring the New and Recently Described Subtilase Cytotoxin Allelic Variant $subAB_{2-3}$. *Genome Announc* 2017;5:e01582-16.
34. Ramachandran V, Hornitzky MA, Bettelheim KA, Walker MJ, Djordjevic SP. The common ovine Shiga toxin 2-containing Escherichia coli serotypes and human isolates of the same serotypes possess a Stx2d toxin type. *J Clin Microbiol* 2001;39:1932–1937.
35. Djordjevic SP, Ramachandran V, Bettelheim KA, Vanselow BA, Holst P *et al*. Serotypes and virulence gene profiles of Shiga toxin-producing Escherichia coli strains isolated from feces of pasture-fed and lot-fed sheep. *Appl Environ Microbiol* 2004;70:3910–3917.
36. Michelacci V, Tozzoli R, Caprioli A, Martínez R, Scheutz F *et al*. A new pathogenicity island carrying an allelic variant of the subtilase cytotoxin is common among Shiga toxin producing Escherichia coli of human and ovine origin. *Clin Microbiol Infect* 2013;19:E149–E156.
37. Mammarappallil JG, Elsinghorst EA. Epithelial cell adherence mediated by the enterotoxigenic Escherichia coli TIA protein. *Infect Immun* 2000;68:6595–6601.
38. Ingersoll MA, Moss JE, Weinrauch Y, Fisher PE, Groisman EA *et al*. The ShiA protein encoded by the Shigella flexneri SHI-2 pathogenicity island attenuates inflammation. *Cell Microbiol* 2003;5:797–807.
39. Ingersoll MA, Zychlinsky A. ShiA abrogates the innate T-cell response to Shigella flexneri infection. *Infect Immun* 2006;74:2317–2327.
40. McKinnon J, Roy Chowdhury P, Djordjevic SP. Genomic analysis of multidrug-resistant Escherichia coli ST58 causing urosepsis. *Int J Antimicrob Agents* 2018;52:430–435.
41. Reid CJ, Wyrsch ER, Roy Chowdhury P, Zingali T, Liu M *et al*. Porcine commensal Escherichia coli: a reservoir for class 1 integrons associated with IS26. *Microb Genom* 2017;3.
42. Sacramento AG, Fernandes MR, Sellera FP, Muñoz ME, Vivas R *et al*. Genomic analysis of MCR-1 and CTX-M-8 co-producing Escherichia coli ST58 isolated from a polluted mangrove ecosystem in Brazil. *J Glob Antimicrob Resist* 2018;15:288–289.
43. Wyrsch ER, Roy Chowdhury P, Wallis L, Cummins ML, Zingali T *et al*. Whole genome sequence analysis of environmental Escherichia coli from the faeces of straw-necked IBIS (Threskiornis spinicollis) nesting on inland wetlands. manuscript submitted for publication.
44. Cummins ML, Reid CJ, Roy Chowdhury P, Bushell RN, Esbert N *et al*. Whole genome sequence analysis of Australian avian pathogenic Escherichia coli that carry the class 1 integrase gene. *Microb Genom* 2019;5.
45. Harmer CJ, Hall RM. IS26-Mediated formation of transposons carrying antibiotic resistance genes. *mSphere* 2016;1.
46. Leuko S, Goh F, Ibáñez-Peral R, Burns BP, Walter MR *et al*. Lysis efficiency of standard DNA extraction methods for Halococcus spp. in an organic rich environment. *Extremophiles* 2008;12:301–308.
47. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 2010;5:e11147.
48. Arndt D, Grant JR, Marcu A, Sajed T, Pon A *et al*. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;44:W16–W21.
49. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 2006;34:D32–D36.
50. Bertelli C, Laird MR, Williams KP, , Lau BY *et al*, Simon Fraser University Research Computing Group. Simon Fraser University Research Computing Group. Island Viewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res* 2017;45:W30–W35.
51. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 2014;30:3276–3278.
52. Edgar RC. Muscle: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 2004;5:113.
53. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile genetic elements associated with antimicrobial resistance. *Clin Microbiol Rev* 2018;31.
54. Gaytán MO, Martínez-Santos VI, Soto E, González-Pedrajo B. Type three secretion system in attaching and effacing pathogens. *Front Cell Infect Microbiol* 2016;6:129.
55. Ma J, Sun M, Bao Y, Pan Z, Zhang W *et al*. Genetic diversity and features analysis of type VI secretion systems loci in avian pathogenic Escherichia coli by wide genomic scanning. *Infect Genet Evol* 2013;20:454–464.
56. Nüesch-Inderbinen MT, Funk J, Cernela N, Tasara T, Klumpp J *et al*. Prevalence of subtilase cytotoxin-encoding subAB variants among Shiga toxin-producing Escherichia coli strains isolated from wild ruminants and sheep differs from that of cattle and pigs and is predominated by the new allelic variant subAB2-2. *Int J Med Microbiol* 2015;305:124–128.
57. Vassiliadis G, Destoumieux-Garzón D, Lombard C, Rebuffat S, Peduzzi J. Isolation and characterization of two members of the siderophore-microcin family, microcins M and H47. *Antimicrob Agents Chemother* 2010;54:288–297.
58. Buts L, Bouckaert J, De Genst E, Loris R, Oscarson S *et al*. The fimbrial adhesin F17-G of enterotoxigenic Escherichia coli has an immunoglobulin-like lectin domain that binds N-acetylglucosamine. *Mol Microbiol* 2003;49:705–715.
59. Ristow LC, Welch RA. Rtx toxins Ambush immunity's first cellular responders. *Toxins* 2019;11:720
60. Vo JL, Martínez Ortiz GC, Subedi P, Keerthikumar S, Mathivanan S *et al*. Autotransporter adhesins in Escherichia coli pathogenesis. *Proteomics* 2017;17
61. Carroll AC, Wong A. Plasmid persistence: costs, benefits, and the plasmid paradox. *Can J Microbiol* 2018;64:293–304.
62. Roy Chowdhury P, McKinnon J, Liu M, Djordjevic SP. Multidrug Resistant Uropathogenic *Escherichia coli* ST405 With a Novel, Composite IS*26* Transposon in a Unique Chromosomal Location. *Front Microbiol* 2018;9:3212.
63. Harmer CJ, Moran RA, Hall RM. Movement of IS26-associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *mBio* 2014;5:e01801–01814.