Whole-genome sequence analysis of environmental *Escherichia coli* from the faeces of straw-necked ibis (*Threskiornis spinicollis*) nesting on inland wetlands

Ethan R. Wyrsch[1,2], Piklu Roy Chowdhury[1,2], Louise Wallis[1], Max L. Cummins[1,2], Tiziana Zingali[1,2], Kate J. Brandis[3] and Steven P. Djordjevic[1,2,*]

Abstract

Wildlife, and birds in particular, play an increasingly recognized role in the evolution and transmission of *Escherichia coli* that pose a threat to humans. To characterize these lineages and their potential threat from an evolutionary perspective, we isolated and performed whole-genome sequencing on 11 sequence types (STs) of *E. coli* recovered from the desiccated faeces of straw-necked ibis (*Threskiornis spinicollis*) nesting on inland wetlands located in geographically different regions of New South Wales, Australia. Carriage of virulence-associated genes was limited, and no antimicrobial resistance genes were detected, but novel variants of an insertion element that plays an important role in capturing and mobilizing antibiotic resistance genes, IS*26*, were identified and characterized. The isolates belonged to phylogroups B1 and D, including types known to cause disease in humans and animals. Specifically, we found *E. coli* ST58, ST69, ST162, ST212, ST446, ST906, ST2520, ST6096 and ST6241, and a novel phylogroup D strain, ST10208. Notably, the ST58 strain hosted significant virulence gene carriage. The sequences of two plasmids hosting putative virulence-associated factors with incompatibility groups I1 and Y, an extrachromosomal integrative/conjugative element, and a variant of a large *Escherichia* phage of the family *Myoviridae*, were additionally characterized. We identified multiple epidemiologically relevant gene signatures that link the ibis isolates to sequences from international sources, plus novel variants of IS*26* across different sequence types and in different contexts.

DATA SUMMARY

The eleven short-read whole-genome sequences of *Escherichia coli* generated in this project have been deposited in the National Center for Biotechnology Information (NCBI) databases under BioProject PRJNA591373, including the Sequence Read Archive under accessions SRR10560910–SRR10560920, and GenBank under accessions WOEK00000000–WOEU00000000.

BACKGROUND

Members of the Gram-negative family *Enterobacteriaceae* exist as commensals or pathogens in the gut of warm-blooded animals, soil, water and vegetation. Despite comprising

approximately 0.1% of the human gut microbiome [1], *Escherichia coli* is perhaps the best-characterized member of the *Enterobacteriaceae* because of its ability to cause intestinal and extra-intestinal disease, and to acquire genes encoding resistance to all classes of antibiotics. Laterally transmissible elements including plasmids, genomic islands and prophages play a central role in the evolution of *E. coli* and other members of the *Enterobacteriaceae* as they harbour beneficial, often niche-adapted, traits. Different lineages of the species, represented in part by multi-locus sequence type (ST), exchange genes between environments, especially different mammalian, avian and environmental reservoirs [2].

Whole-genome sequencing (WGS) has emerged as the primary method to trace these evolutionary events and the pathogens that result [3]. Public repositories of bacterial whole-genome sequences are heavily biased toward pathogenic bacteria, particularly dominant STs of *E. coli* causing disease in humans (e.g. ST131). In addition to this pathogen focus, genome sequence data describing clinically important antimicrobial resistance (AMR) genes are disproportionally represented. To understand *E. coli* as both commensal and pathogen, and its influence on the spread and evolution of AMR, there is a need to resolve both the physical and the phylogenetic distribution of *E. coli*, particularly the less clonal STs belonging to the so-called commensal phylogroups A (e.g. ST10) and B1 (e.g. ST58), which predominate in commensal faecal populations in intensive food animal production [4, 5] and have been observed in a diverse range of hosts [6]. ST58 is a good example, as studies have demonstrated its wide geographical distribution, as well as an affinity for human [7, 8] and animal [9] hosts, plus environmental reports [10]. ST58 has also been associated with episodes of multiple-drug-resistant human extraintestinal infections in Australia [8] and elsewhere [11, 12]. It will be important to keep track of these lineages as they develop to prevent the emergence of critical pathogens borne from the acquisition of virulence-associated genes (VAGs) and antimicrobial resistance. It is notable that there has been a rise in the reporting of *E. coli* that have acquired diverse arrays of VAGs, that challenge existing *E. coli* pathotype designations from cases of human intestinal and extra-intestinal disease [13, 14], and in the capture of multiple antibiotic resistance genes by both commensal and pathogen *E. coli* populations [5, 15–17]. The interplay between microbial populations from agriculture and human disease, and bacterial communities in the broader environment, influences how they encounter and acquire ARGs and VAGs and evolve as potential pathogens.

E. coli is often a causative agent of disease in avian agriculture [18, 19], with multiple common extraintestinal infection sites that appear to often stem from initial respiratory infections [20]. A range of virulence-associated genes, including the serum survival gene *iss* and other protectins, iron acquisition systems and iron regulation genes (e.g. *sit* and *iro* operons), are prominently detected in a range of *E. coli* STs causing avian infections [18]. Recent studies have also been exploring some common links between these STs, these virulence factors and their capacity to cause extraintestinal infections in humans [21, 22], but these studies have been hindered by a lack of consistency in VAG profile between infection cases, indicating a heterogenous pool of potential pathogens and perhaps multiple pathways to successful colonization. Epidemiological investigations are now becoming key to identifying reservoirs of these specific genes and gene profiles to assist in unravelling these associations. While the focus of these studies has necessarily been at agricultural sites and in cases of human infection, the role of wildlife has been under-investigated [23].

Here we generated and characterized whole-genome sequences of *E. coli* from the faeces of straw-necked ibis

Impact Statement

Whole-genome sequences of *Escherichia coli* strains from an environmental avian source were investigated for the presence of key virulence and resistance indicator genes and carriage of mobile genetic elements. Among the 11 strains isolated, there were both commonly observed and novel chromosome sequence types. Virulence gene carriage was low among the strains, except for an ST58 isolate hosting the subtilase cytotoxin amongst other virulence content. No mobilized resistance genes were detected, but several single-copy variants of IS*26* were observed in both chromosomal and extra-chromosomal loci. Four extra-chromosomal elements were also characterized: virulence-associated Y family and I1 family plasmids, a putative replicative integrative conjugative element, and a large *Escherichia* phage. Our findings demonstrated linkage between these environmental Australian strains and the known international microbiome and highlighted the potential of ibis to harbour bacteria commonly associated with human and animal disease.

(*Threskiornis spinicollis*) nesting in three separate inland wetlands in New South Wales, Australia. We characterized their STs, phylogeny, *in silico* serotypes, plasmid and phage content, and identified several genes of interest, particularly some individual IS*6* family sequences, and characterized their genetic context. Analyses were also performed to identify carriage of ARGs and VAGs and determine if these environmental avian isolates are related to known Australian human or agricultural pathogens.

METHODS

E. coli strains were isolated from 5 of 44 faecal samples collected from nests of juvenile *Threskiornis spinicollis* (straw-necked ibis) in Australian wetlands. Faecal samples were taken in 2016 from three sites in New South Wales; named here Barmah, Booligal and Macquarie Marshes. Samples were collected from active nests during routine nest surveys and placed into ziplock sealable plastic bags. Samples were stored at −20 °C. Sampling of faeces occurred throughout the nesting period (approximately 3 months). These waterbirds are thought to have little contact with humans; however, they are known to breed in mixed colonies with *Threskiornis moluccus* (Australian white ibis), which has significant human contact. During standardization of the protocol for isolation of *Enterobacteriaceae*, an aliquot of the faecal sample was taken out of the freezer and held at 4 °C.

E. coli strain isolation was performed by first resuspending 1 g of the faecal samples in 5 ml Lysogeny (LB) broth and incubating it for 18 h at 37 °C. Three 1:1 000 000 dilutions were then made from each overnight culture, and three replicates of each dilution were then plated onto McConkey agar plates.

Single colonies that matched *E. coli* colony morphology were purified on McConkey agar plates. Six individual colonies per sample (one per McConkey plate) were grown overnight in LB broth and boiled lysates from 100 µl of the overnight cultures were subjected to a multiplex polymerase chain reactions (PCRs) (MyTaq Red, Bioline) to screen for the gene encoding an *E. coli*-specific universal stress protein *uspA* [24] and the class 1 integrase gene *intI1* [25]. Cultures that returned positive hits for the *uspA* gene were stored in 0.5× M9 salts supplemented with 25% glycerol as −80 °C frozen storage. We also screened for IS*26* by PCR using primers IS26F: GGCATCAGTTACCGTGAGC and IS26R: CGTGTTGAT-GAATCGTGG that spans 400 bp of the element. Both IS*26* and the class 1 integron (*intI1*) are important elements that play key roles in the capture and assembly of antimicrobial resistance genes in *Proteobacteriaceae* [26] and in the creation of novel resistance gene loci [27].

Genomic DNA was extracted (ISOLATE II Genomic DNA kit, Bioline) following the manufacturer's protocol for both WGS and as templates for confirmatory PCR. The final sequencing pool comprised isolates that returned a positive hit for the *uspA* gene, with some positive for the IS*26* transposase gene. The *intI1* gene was not detected in any of the samples. Confirmatory PCR reactions comprised the following reagents: MyTaq Red Mix (10 µl), PCR-grade water (5 µl), *uspA* primers (2 µl), *intI1* primers L2 and L3 (2 µl) and 1 µl of stock DNA. PCR was performed using the following protocol for 30 cycles: preheating for 2 min, 94 °C; denaturation for 20 s, 94 °C; annealing for 20 s, 61 °C; extension for 20 s, 72 °C, followed by a post-cycling extension (5 min, 72 °C).

Paired-end short-read WGS was performed on an Illumina HiSeq Sequencer at the Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research. Library preparation was performed with a Nextera DNA library preparation kit. Reads were checked for quality using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequences were assembled by A5-miseq v. 20150522 with a minimum 45× coverage coming from 2×125 bp paired-end reads. Genome sequence data were uploaded to the National Center for Biotechnology Information (NCBI) Sequence Read Archive and GenBank database under BioProject PRJNA591373.

Analyses of genome sequence data were performed using several pieces of software, accessed through miniconda3 (https://docs.conda.io/en/latest/miniconda.html), plus online databases and utilities. Single-gene identification using the ResFinder, VirulenceFinder and PlasmidFinder databases (sourced from http://www.genomicepidemiology.org/ on 14 May 2018), multi-locus sequence typing (Achtman *E. coli* scheme [28]), *in silico* serotyping [29] and phylogrouping [30] were performed on unassembled reads using ARIBA v2.13.3 [31] with unaltered parameters, with post-processing using ARIBAlord (https://github.com/maxlcummins/ARIBAlord). Sequence comparisons were performed using the NCBI BLASTN/BLASTX (https://blast.ncbi.nlm.nih.gov/Blast.cgi) servers and multiple

progressiveMauve [32] alignments. Phylogenetic analyses were performed with the Harvest software suit [33], including Parsnp v1.2, which generates a conserved core phylogeny of assembled sequences, and Gingr v1.2, which is used to visualize the data. Parsnp was run using unaltered software parameters. Phylogenetic tree outputs were analysed in FigTree 1.4.4 (https://github.com/rambaut/figtree). Bacteriophages were identified using PHASTER [34]. Insertion sequences were identified using ISFinder [35]. PlasmidFinder and the I1 plasmid pMLST scheme [36] were accessed through the CGE server (http://www.genomicepidemiology.org/). Chromosomal resistance genes were detected using the Comprehensive Antibiotic Resistance Database (CARD) RGI [37]. Read alignments for testing scaffold circularity were performed with BWA v0.7.17 [38] and visualized in Tablet [39]. To do this, 1 kb of sequence from either end of the tested assembled scaffold (suspected to represent a full extrachromosomal element based on reference comparisons) was concatenated and aligned to the sequencing reads. If there was read contiguity over the scaffold break, it was inferred that this scaffold represented the full, circular sequence of the element.

Reference sequences for phylogenetic analyses were sourced from Enterobase (WGS) (https://enterobase.warwick.ac.uk/) and GenBank (plasmids) (https://www.ncbi.nlm.nih.gov/nucleotide/), attempting to avoid including clonal database entries to better resolve the overall distribution of certain strain and plasmid subtypes. WGS representatives were included based on date, country and host metadata to reveal an underlying phylogenetic structure of the ST and assist in placing Australian sequences amongst international reference. Lists of reference plasmids were generated using the plsdb PlasmidFinder metadata [40] and then that selection of sequences was refined based on BLASTN similarity so that an accurate phylogeny could be constructed.

Figure components were generated with SnapGene v4.1.9 (https://www.snapgene.com/), BRIG v0.95 [41] and iTOL (https://itol.embl.de/).

RESULTS

Genomic analysis of *E. coli* from wild ibis

Whole-genome sequences of 11 *E. coli* strains from 3 different ibis scat sources in Australian wetlands – Booligal, Barmah, and Macquarie Marshes – were generated (Table 1). The strains sourced belonged solely to phylogroups B1 and D and included STs known to cause human and animal disease, including two ST69 strains with different serotypes. All sequences except IBS62 assembled in under 200 scaffolds. Typical of *E. coli* genomes, assembly sizes ranged from 4.8 Mb to 5.5 Mb. Of the STs we identified, those with the highest representation in the Enterobase database are ST58, ST69, ST162 and ST906, with ST212 and ST446 having fewer entries in comparison. Types ST2520, ST6096 and ST6241 appear much less frequently in the database. We also isolated a novel ST, designated ST10208, part of phylogroup D and quite distinct from any other strain on

Table 1. Summary of details for *E. coli* short-read assemblies

Strain	Site	Group	ST	*In silico* serotype	Size (bp)	Scaffolds	Extrachromosomal features	IS6 family
IBS28	Booligal*	B1	58	ONT:H25	4884028	140	• pIBS28_1	+
IBS1	Booligal	D	69	O15:H18	5054477	184	• F replicon (F67:B38)	+
IBS62	Barmah	D	69	O17:H18	5150726	220		
IBS67	Barmah	B1	162	O8:H19	4892495	111	• F replicon (F19) • Y replicon	
IBS19	Macquarie Marshes	B1	212	O18:H49	5066990	97	• Phage IBS	
IBS44	Booligal*	B1	446	O7:H21	4888326	158		
IBS63	Barmah	B1	906	O10:H8	5056979	166	• F replicon (F89:B54:C1)	+
IBS46	Booligal*	B1	2520	O116:H49	4797535	100	• I1 replicon	
IBS12	Macquarie Marshes	B1	6096	O39:H28	5543924	167	• Phage IBS • F replicon • I1 replicon	
IBS69	Barmah	B1	6241	ONT:H16	5077592	135	• F replicon (F56:B16)	
IBS49	Booligal	D	10208	O3:H48	5191708	124	• I1 replicon • F replicon (F96:B54)	

*, same sample.

Enterobase. Full details for each strain are presented in Table S1 (available in the online version of this article).

Identification of chromosomal antimicrobial resistance genes

Each of the *E. coli* strains isolated here hosted suites of chromosomal antimicrobial resistance genes common to many modern strains. Details and a comparison of these data are presented in Table S2. The efflux operons seen in each strain included *acrABDEFRS*, *baeRS*, *mdtABCEFGHMNOP* and *emrABKRY*. Chromosomal *ampC* and *ampH* beta-lactamase genes, efflux-associated singular genes such as *tolC* and resistance-associated SNPs in various genes were also identified commonly, as detailed in the Supplementary Material. Less frequently identified resistance associations were fosfomycin-conferring SNPs in *cyaA*, efflux gene *emrE* and, found only in novel ST10208, a fosfomycin resistance-associated single-nucleotide polymorphism (SNP) in *uhpT*, as well as an unusual partial hit for *gyrB* from *Morganella morganii*.

Characterization of IS elements

Across the 11 strains, 30 IS element families were identified, although most strains hosted 5 or fewer families of IS elements. Short-read assemblies make it impossible to resolve the exact number of elements for each family identified. Strains IBS44 (ST446) and IBS1 (ST69) hosted 7 and 10 families of IS element, respectively. Three strains – IBS28 (ST58), IBS63 (ST906) and IBS1 (ST69) – each hosted a single copy of an unreported IS6 family element, likely IS26 sequences with SNPs in previously unidentified sites. In IBS28 (ST58), this element is inserted chromosomally near the O-antigen lipopolysaccharide gene cluster. Compared to the full 820 bp IS26 sequence, the chromosomal IS6 element in IBS28 had a single C573T alteration. This specific allele can be seen in the GenBank nucleotide database twice; out of context in a *Citrobacter freundii* strain (CP022151) (Swiss wastewater, 2015) and in a surprisingly homologous context in the *E. coli* ST101:ONT:H21 strain KSC9 (CP018323) chromosome (porcine sample, USA 2014). The sequence data for IBS1 (ST69) and IBS63 (ST906) suggest that these IS6 family elements were carried on unresolved F plasmids. The IS6 family sequences from IBS1 and IBS63 share G48A and G614A polymorphisms. The element in IBS1 has an additional C85T alteration. Interestingly, only the element with the three SNPs has a homologue on GenBank, matching to an *E. coli* F plasmid (CP024151) from a human clinical sample in PR China, 2014. None of these IS6 family elements had bordering direct repeats.

Carriage of virulence–associated genes

At least one VAG was detected in each strain (Table S1), with long polar fimbriae subunit *lpfA* the most common identification, present in all but the ST10208 strain. The ST58 isolate hosted the most virulence-associated content, with immune survival factor *iss*, iron-regulated virulence outer-membrane protein *ireA,* microcins *mch* and *mca*, the F17 adhesin major subunits and *subAB* encoding the subtilase cytotoxin. Of these, only *iss* was identified in the other strains, being seen in 6 of the 11.

Genomic phylogeny of individual STs

Phylogenetic analyses based on conserved single-nucleotide polymorphic sites were performed on each ST sourced from

the ibis samples. This enabled an analysis of general subclade structure and the identification of any links to isolates with publicly available whole-genome sequences. IBS69 and IBS49 were excluded, as these STs lacked a significant number of examples to compare. The core genome selected for each SNP alignment was consistently 70–80% of each genome, with the IBS isolate used as a reference in each case.

A phylogenetic analysis of ST58 was performed to compare strain IBS28 to known Australian human and avian pathogen strains, as well as a diverse subset of national and international strains sourced from Enterobase (Fig. 1). The tree generated using IBS28 as reference splits ST58 into two broad clades. Strain IBS28 placed amongst a subset of ST58 strains of diverse origin, with only a single other Australian sequence distantly related. The other Australian pathogen sequences included were placed in a separate clade. IBS28 had four close relatives, each from North America, sourced between 1988 and 2018.

Similarly, a mid-point phylogeny was constructed from a diverse set of ST69 sequences from Enterobase and the two ST69 strains sequenced here, reference IBS1 (O15:H18) and IBS63 (O17:H18), with the earliest strain isolated in 1982 (Fig. 2). The Parsnp alignments draw three primary clades for this distribution of ST69 strains and show the ibis strains are not closely related despite the shared H antigen allele. IBS62 sat distant from the root, in a small, internationally diverse clade composed of strains of distant time points. It placed nearest strains ESC_RA2043AA_AS from Poland (2014) and ESC_HA9259AA_AS from Luxembourg (2016), each with unknown sources. IBS1 sat in a separate, equally diverse, clade, but most strains in the clade were isolated post-2010, with only one earlier representative from 1997. IBS1 had no very close relatives, but its local subclade was generally human-sourced, and included another Australian strain (ESC_TA7275AA) with matching serotype (ST69:O15:H18) in Enterobase, isolated in 2017. A progressiveMauve alignment between the short-read assemblies of these strains (Fig. S1) revealed almost full chromosomal synteny, although there were alterations in F plasmid content, phage content and various chromosomal islands. While IBS1 only encoded *espP*, a serine protease, ESC_TA7275AA hosted numerous virulence factors (including *iroN* and *ireA* iron acquisition genes, the secreted autotransporter toxin *sat* and an *iha* variant coding for adherence). In total, the progressiveMauve alignment demonstrated ESC_TA7275AA carried approximately 450 kb of DNA that IBS1 did not, including F family and Q1 family plasmid replicon genes, suggesting that it may have accessed a larger repertoire of mobilized genes. This additional sequence included numerous antimicrobial resistance and related genes, with sulphonamide resistance genes *sul1* and *sul2*, cephalosporin resistance gene $bla_{CTX-M-15}$, and class 1 integrase *intI1* of note.

E. coli ST162 (Fig. S2) had a broad set of available sequences, with the earliest coming from swine in the USA, isolated 1980. While the overall clade structure indicates several globally distributed sub-lineages, IBS67 was placed early in the tree alongside six strains from the USA and a single isolate from Kenya, from a variety of animal sources isolated in the years 2000–17. Interestingly, an Australian avian–human pair of isolates from 2017 can be observed in a separate ST162 subclade. ST212 (Fig. S3) comprised a more limited set of strains than the previous STs, with the earliest isolated from Japan in 2008. The clade structure indicated a closely related set of isolates, with IBS19 related to isolates from Australian poultry meat sourced in 2014. A progressiveMauve alignment (Fig. S4) indicates a closer genetic relationship between the strains isolated from meat than to strain IBS19, but also demonstrates their similarity. A small subclade also highlighted matches between an Australian avian strain (2017) and three Gambian avian strains (2016). ST446, while comprising a smaller number of strains in Enterobase as compared to ST58 or ST69, its earliest isolation was in the USA, 1979, and had few human-associated representatives (Fig. S5). In this case, IBS44 had no near relatives, with the closest strain of US bovine origin, 2018. ST906 was again an internationally distributed clade, though with heavy representation of US data available online, with the first isolation available from the USA in 1983 (Fig. S6). Here again, IBS63 was not very closely related to any specific strains, but was part of an internationally observed subclade, with its closest neighbours from the USA more recently. Both ST2520 (Fig. S7) and ST6096 (Fig. S8) were smaller clades again, with strains IBS46 and IBS12 most closely related to strains from the USA isolated in 2003 and 2012, respectively.

Characterization of an unusual integrative/conjugative element

The IBS28 (ST58) WGS resolved an extrachromosomal DNA molecule, named here pIBS28_1, identified as a potential integrative/conjugative element (ICE) formed on a rarely observed variant of the type 4 secretion system (T4SS), but also encoded an F family plasmid replication (*rep*) gene. A diagram of the element with comparisons to available homologues is presented in Fig. 3, the closest of which was identified from Australian human pathogen *E. coli* ST38 strain HoS72 (SRR11495744), while the second was identified from *E. coli* strain CIP106223, RSC93_pII (LT985310). The third closest homologue, pJARS36, is distantly related at 70% sequence identity, and was taken from *Yersinia pestis* strain JAVA9 isolated in Indonesia (1957). Despite the low identity, it appears to be the only other element with a similarly structured T4SS in the GenBank nucleotide database. pIBS28_1 resolved at 37020 bp, approximate to its homologues, and encodes a *ccgAII*-like gene for cobalamin (vitamin B12) synthesis. While this gene is shared with the element from Australian strain HoS72, it is not carried by RSC93_pII.

IncI1 plasmid analysis

Short-read assembly data for ST2520 strain IBS46 resolved an I1 plasmid contig comprising 109519 bp, referred to from here as the IBS46 I1 contig. Read alignments of the assembly scaffold are not contiguous across the break, however, indicating that this contig does not represent the full plasmid

Fig. 1. Mid-point root phylogenetic tree of *E. coli* ST58 strains. Country, year and source metadata are presented to the right of each tree tip. *E. coli* IBS28 is highlighted in red. Node confidence is represented by dot size, with high confidence across the primary split points. Tree scale is in SNPs per site.
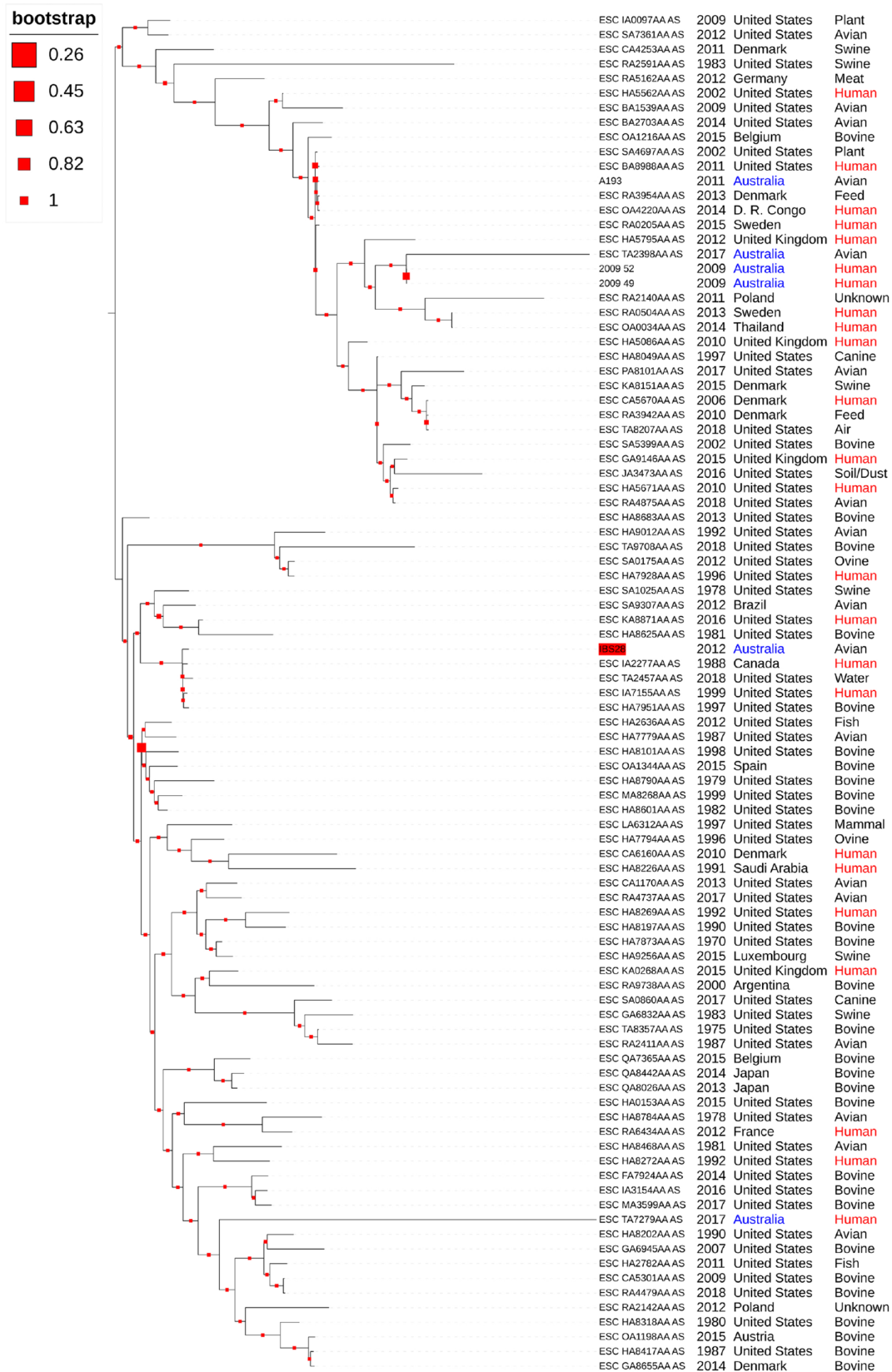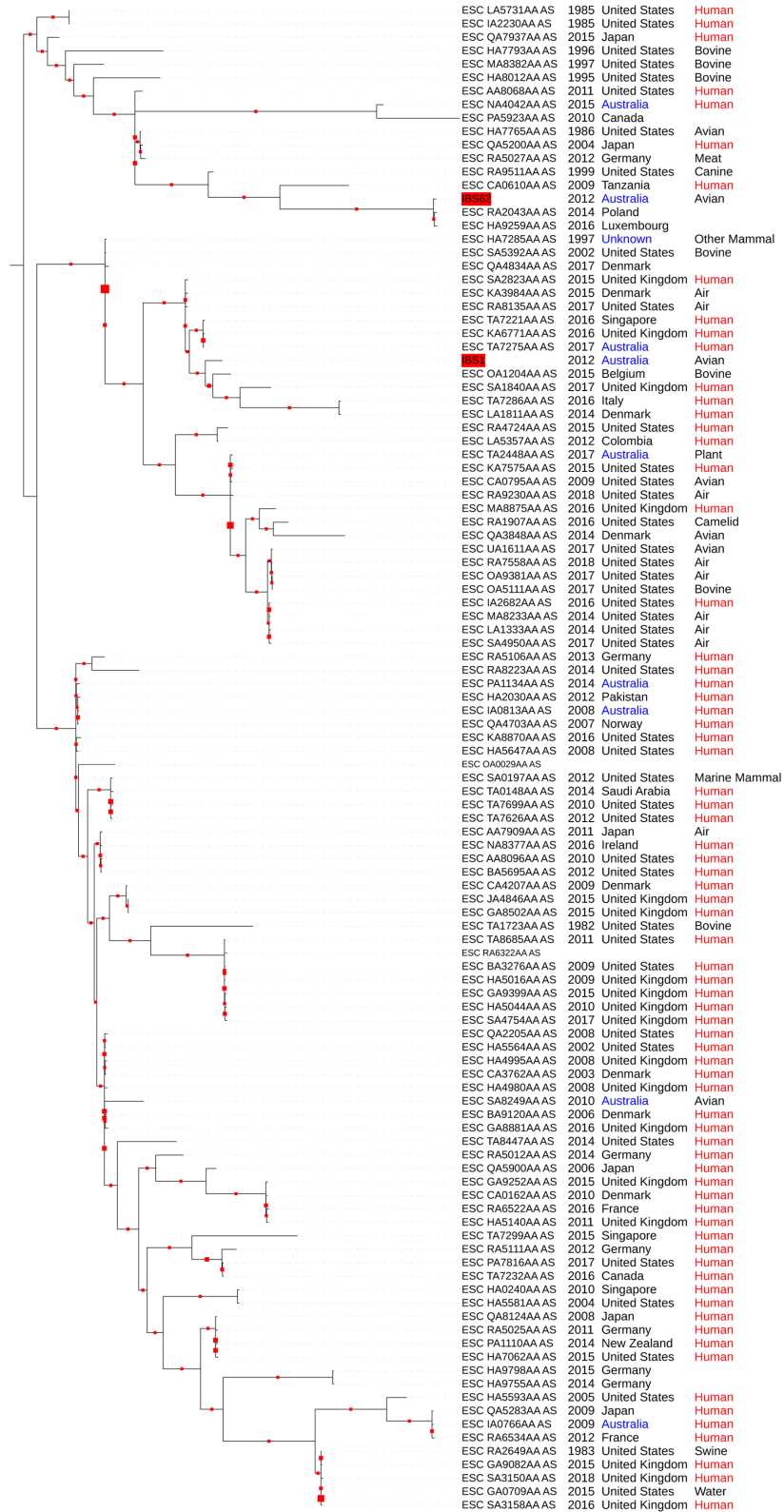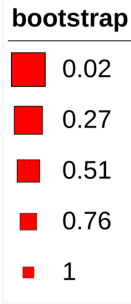
Fig. 2. Mid-point root phylogenetic tree of *E. coli* ST69 strains. Country, year and source metadata are presented to the right of each tree tip. *E. coli* strains IBS1 and IBS62 are highlighted in red. Node confidence is represented by dot size, with high confidence across the primary split points. Tree scale is in SNPs per site.
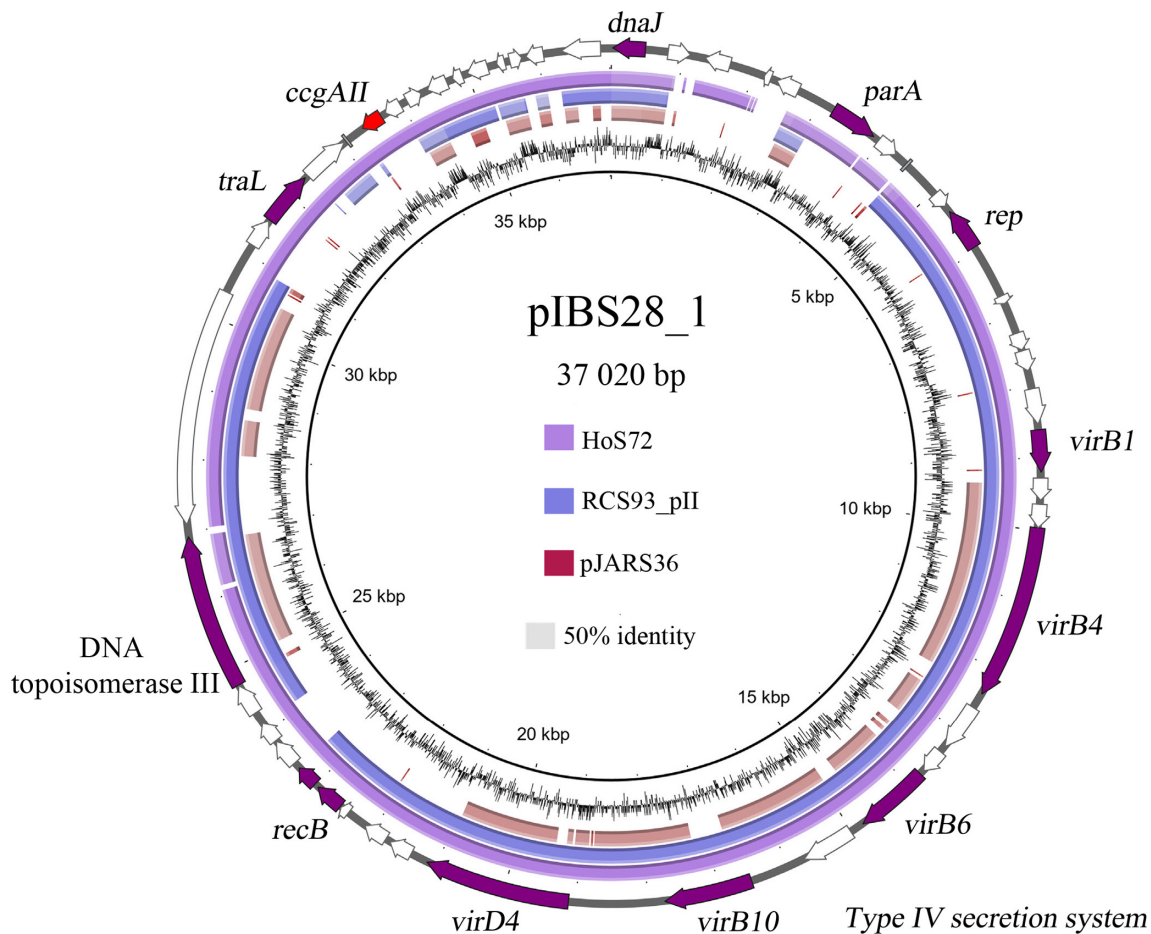
Fig. 3. Plasmid map of pIBS28_1 with BLASTN comparisons to related mobile elements. Plasmid annotations are indicated as arrows in an external ring, with inner rings representing BLASTN values to colour-coded sequences. The centre histogram (black) indicates GC content. Known gene families are annotated as purple arrows, hypothetical CDS are shown in white. The Vitamin B12 synthesis gene *ccgAll* is shown in red.

sequence. The plasmid hosts two potential virulence regions: one encoding a type 1 fimbrial adhesin (approximately 10.2 kb), and the other a putative efflux transporter (approximately 5.5 kb). Alignments are presented in Fig. 4 against two I1 plasmids from *Salmonella enterica* in the USA; pSNE-1928 (CP025239) isolated in 2011 and hosting the type 1 fimbrial adhesin, and pSAN1-1727 (CP014622) from 2005 hosting the putative efflux transporter. A phylogeny of I1 alpha plasmids is presented in Fig. 5, showing the plasmids separate into three primary clades. The IBS46 I1 contig has a novel plasmid ST and was associated with ST2 and ST136 plasmids, amongst others, including pSAN1-1727. Its branching suggests some evolutionary distance from the other plasmids. pSNE-1928 resolved to a clade comprising ST7/ST271 plasmids. Interestingly, ST3 appeared to be the most clonal type amongst the plasmids. Compared to the I1 backbone sequence in pSAN1-1727 and pSNE-1928, the IBS46 I1 contig backbone encoded an additional NTD biosynthesis operon, two hydrolase ORFs and an ATPase involved in DNA repair that oddly shares a homologue with FII plasmid pEC14I (KU932025).

For the region encoding the putative ABC efflux transporter, 17 homologues in the same I1 plasmid context have been uploaded to GenBank, from a range of *E. coli* and *S. enterica*, each at 98% identity or greater. Interestingly, there were some additional high-identity matches, suggesting this same autotransporter can be found in F plasmid sequences (e.g. CP010214). In contrast, the putative type 1 fimbriae region had only two close homologues in the GenBank database, hosted in pSNE2-1928 utilized in the alignment and pCFSAN000752 (CP039490), also isolated in the USA, 2012. In each example the region encoding the fimbriae is bordered by IS91-like fragments, although in the IBS46 I1 contig one of the flanking elements has undergone significantly less sequence loss compared to the other partial elements. This island is additionally adjacent to a *relBE*-like toxin/antitoxin system in the IBS46 I1 contig, not observed as part of the usual I1 backbone. Regarding the potential fimbrial ORFs, homologues to transcriptional regulators and usher proteins on GenBank range from >85% amino acid identity in multiple *E. coli, S. enterica* and *Yersinia* spp. sequences.
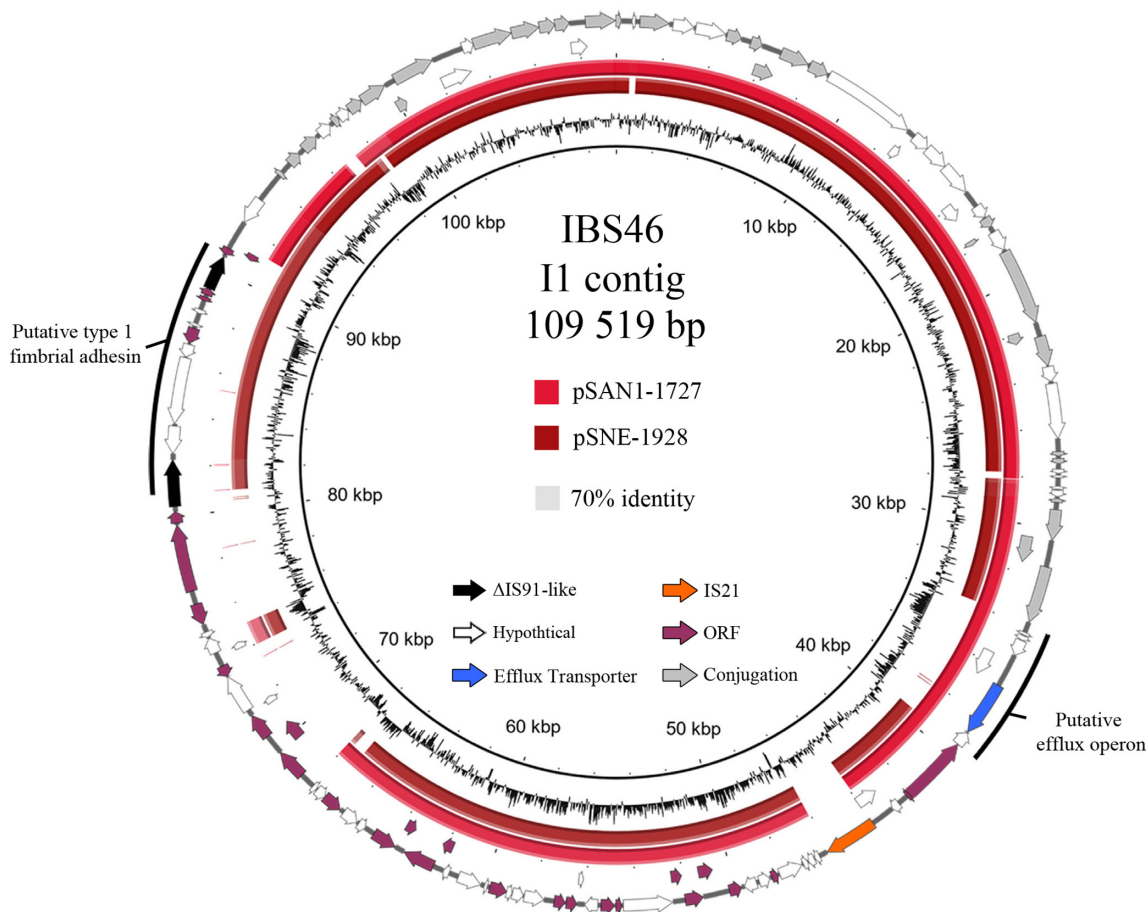
Fig. 4. Plasmid map of the IBS46 I1 contig with BLASTN comparisons to two reference I1 plasmids. Plasmid annotations are indicated as arrows in an external ring, with inner rings representing BLASTN values to colour-coded sequences. The centre histogram (black) indicates GC content. Interesting features are indicated.

Y plasmid analysis

The assembly of strain IBS67 resolved a Y plasmid contig 94474 bp in length, referred to here as the IBS67 Y contig. As with the IBS46 I1 contig, read alignments do not completely resolve the IBS67 Y contig. An annotation is presented and aligned against five diverse Y plasmid representatives in Fig. 6. The IncY family of plasmids are IncF–phage hybrid elements, and variations across regions of the phage backbone and within the plasmid replicon were clear. A conserved core phylogeny of a set of Y plasmids (Fig. 7) identified a core averaging only 28.1% coverage, composed of multiple regions distributed around the IBS67 Y contig reference sequence. This phylogeny shows two primary clades, the smaller of which included several closely related elements. The IBS67 Y contig, however, belonged to the larger clade, which has branching indicative of a divergent set of plasmids, which is consistent with the selection process and low core sequence coverage. The most closely related element to the IBS67 Y contig was pFORC64.1 (CP023732.1) from the Republic of Korea, isolated from a cucumber in 2014. The IBS67 Y contig hosted a region of accessory phage proteins

of unknown function adjacent to a shared toxin–antitoxin gene pair, plus sequence variations can be observed in a phage tail fibre protein, within the FIB plasmid replicon region (including low homology to *parA* and no homology to *parB* versus the references), and the addition of a putative lipoprotein.

The IBS67 Y contig hosted a putative kappa fimbrial adhesin that is not present in any of the elements with the highest overall nucleotide identity selected for BLASTN comparisons (Fig. 6). Two Y plasmids in the GenBank nuccore database did host close variants of this adhesin, Unnamed1 (CP030784.2) and p2014C-4015–2 (NZ_CP034164), both from *Escherichia albertii* strains isolated by the Centers for Disease Control and Prevention (CDC) in the USA in 2012 and 2014, respectively. The putative adhesin resides in an operon composed of eight open reading frames (ORFs), two of which were identified as hypothetical proteins, sitting within an insertion 8205 bp in length. This operon shares its highest sequence match with the porcine oedema adhesin F18, at 91% nucleotide identity. The ORFs encode regulatory, major fimbrial subunit and usher
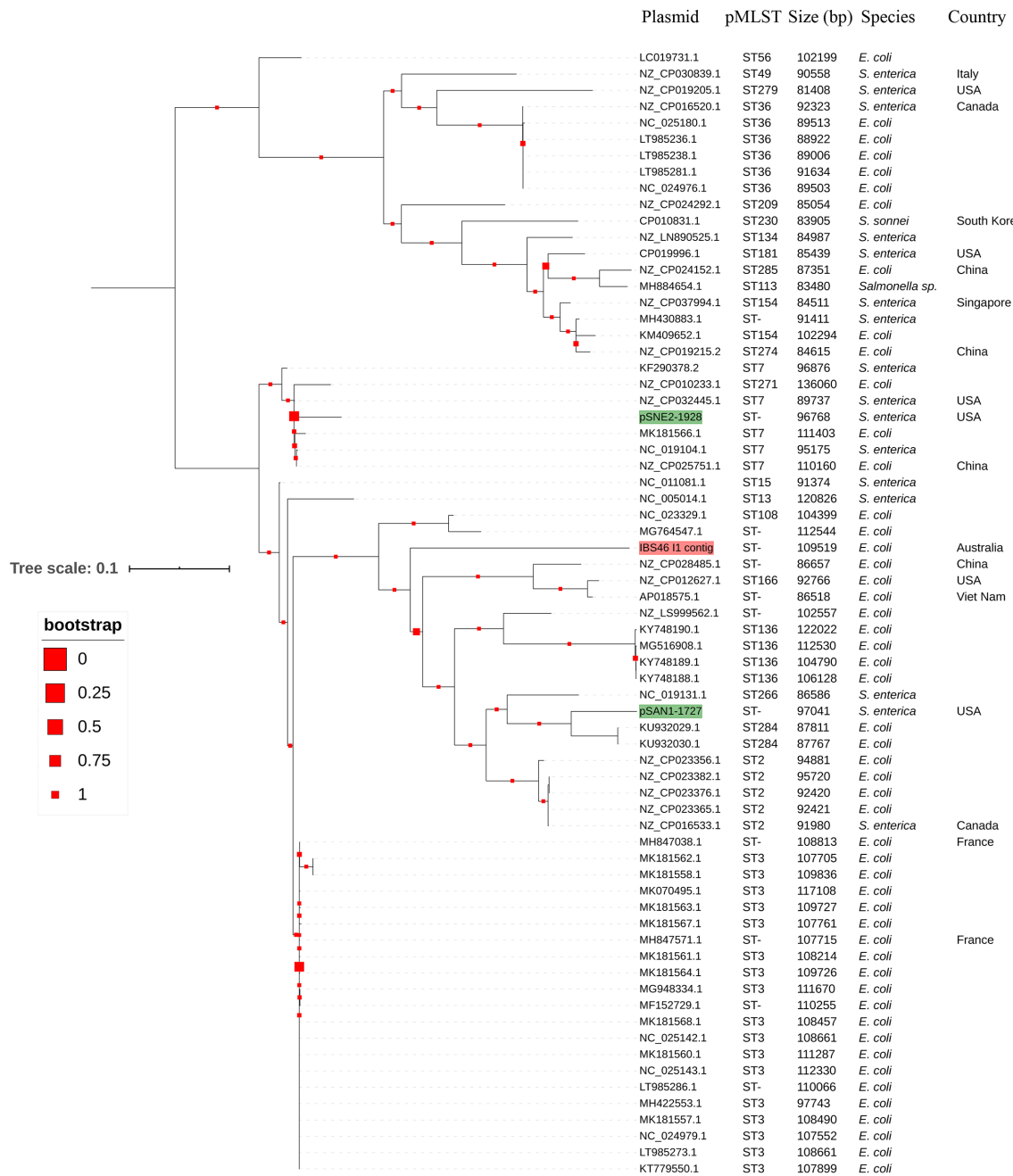
Plasmid	pMLST	Size (bp)	Species	Country
LC019731.1	ST56	102199	*E. coli*	
NZ_CP030839.1	ST49	90558	*S. enterica*	Italy
NZ_CP019205.1	ST279	81408	*S. enterica*	USA
NZ_CP016520.1	ST36	92323	*S. enterica*	Canada
NC_025180.1	ST36	89513	*E. coli*	
LT985236.1	ST36	88922	*E. coli*	
LT985238.1	ST36	89006	*E. coli*	
LT985281.1	ST36	91634	*E. coli*	
NC_024976.1	ST36	89503	*E. coli*	
NZ_CP024292.1	ST209	85054	*E. coli*	
CP010831.1	ST230	83905	*S. sonnei*	South Korea
NZ_LN890525.1	ST134	84987	*S. enterica*	
CP019996.1	ST181	85439	*S. enterica*	USA
NZ_CP024152.1	ST285	87351	*E. coli*	China
MH884654.1	ST113	83480	*Salmonella sp.*	
NZ_CP037994.1	ST154	84511	*S. enterica*	Singapore
MH430883.1	ST-	91411	*S. enterica*	
KM409652.1	ST154	102294	*E. coli*	
NZ_CP019215.2	ST274	84615	*E. coli*	China
KF290378.2	ST7	96876	*S. enterica*	
NZ_CP010233.1	ST271	136060	*E. coli*	
NZ_CP032445.1	ST7	89737	*S. enterica*	USA
pSNE2-1928	ST-	96768	*S. enterica*	USA
MK181566.1	ST7	111403	*E. coli*	
NC_019104.1	ST7	95175	*S. enterica*	
NZ_CP025751.1	ST7	110160	*E. coli*	China
NC_011081.1	ST15	91374	*S. enterica*	
NC_005014.1	ST13	120826	*S. enterica*	
NC_023329.1	ST108	104399	*E. coli*	
MG764547.1	ST-	112544	*E. coli*	
IBS46 I1 contig	ST-	109519	*E. coli*	Australia
NZ_CP028485.1	ST-	86657	*E. coli*	China
NZ_CP012627.1	ST166	92766	*E. coli*	USA
AP018575.1	ST-	86518	*E. coli*	Viet Nam
NZ_LS999562.1	ST-	102557	*E. coli*	
KY748190.1	ST136	122022	*E. coli*	
MG516908.1	ST136	112530	*E. coli*	
KY748189.1	ST136	104790	*E. coli*	
KY748188.1	ST136	106128	*E. coli*	
NC_019131.1	ST266	86586	*S. enterica*	
pSAN1-1727	ST-	97041	*S. enterica*	USA
KU932029.1	ST284	87811	*E. coli*	
KU932030.1	ST284	87767	*E. coli*	
NZ_CP023356.1	ST2	94881	*E. coli*	
NZ_CP023382.1	ST2	95720	*E. coli*	
NZ_CP023376.1	ST2	92420	*E. coli*	
NZ_CP023365.1	ST2	92421	*E. coli*	
NZ_CP016533.1	ST2	91980	*S. enterica*	Canada
MH847038.1	ST-	108813	*E. coli*	France
MK181562.1	ST3	107705	*E. coli*	
MK181558.1	ST3	109836	*E. coli*	
MK070495.1	ST3	117108	*E. coli*	
MK181563.1	ST3	109727	*E. coli*	
MK181567.1	ST3	107761	*E. coli*	
MH847571.1	ST-	107715	*E. coli*	France
MK181561.1	ST3	108214	*E. coli*	
MK181564.1	ST3	109726	*E. coli*	
MG948334.1	ST3	111670	*E. coli*	
MF152729.1	ST-	110255	*E. coli*	
MK181568.1	ST3	108457	*E. coli*	
NC_025142.1	ST3	108661	*E. coli*	
MK181560.1	ST3	111287	*E. coli*	
NC_025143.1	ST3	112330	*E. coli*	
LT985286.1	ST-	110066	*E. coli*	
MH422553.1	ST3	97743	*E. coli*	
MK181557.1	ST3	108490	*E. coli*	
NC_024979.1	ST3	107552	*E. coli*	
LT985273.1	ST3	108661	*E. coli*	
KT779550.1	ST3	107899	*E. coli*	

Tree scale: 0.1

bootstrap

- 0
- 0.25
- 0.5
- 0.75
- 1

Fig. 5. Mid-point phylogenetic distribution of I1 family plasmids. Plasmid metadata are presented in the right-hand side columns, including bacterial source, pMLST result, size (in base pairs) and country, if available. The IBS46 I1 contig is highlighted in red, with plasmids used in the BLASTN comparison (Fig. 4) in green. Node confidence is represented by dot size, with high confidence across the primary split points. Tree scale is in SNPs per site. The IBS46 I1 contig was used as reference, selecting 53.9% average core sequence coverage amongst other input plasmid sequences.

proteins, plus a *fedE*-like fimbrial protein and a periplasmic thiol:disulfide interchange protein.

Escherichia **phage of the family** *Myoviridae*

Two strains of *E. coli* with ST6096 and ST212 were sourced from the Macquarie Marshes site. Quite interestingly, and despite coming from different samples, each hosted

Escherichia phage IBS (344, 801 bp), a variant of an internationally observed large *Escherichia* phage of the family *Myoviridae*. Read alignments indicated that the assembled scaffolds represented the full, closed phage as an extrachromosomal element. Five other variants of this phage have been identified from *E. coli* strains in North America and across Asia. An alignment of these sequences is presented
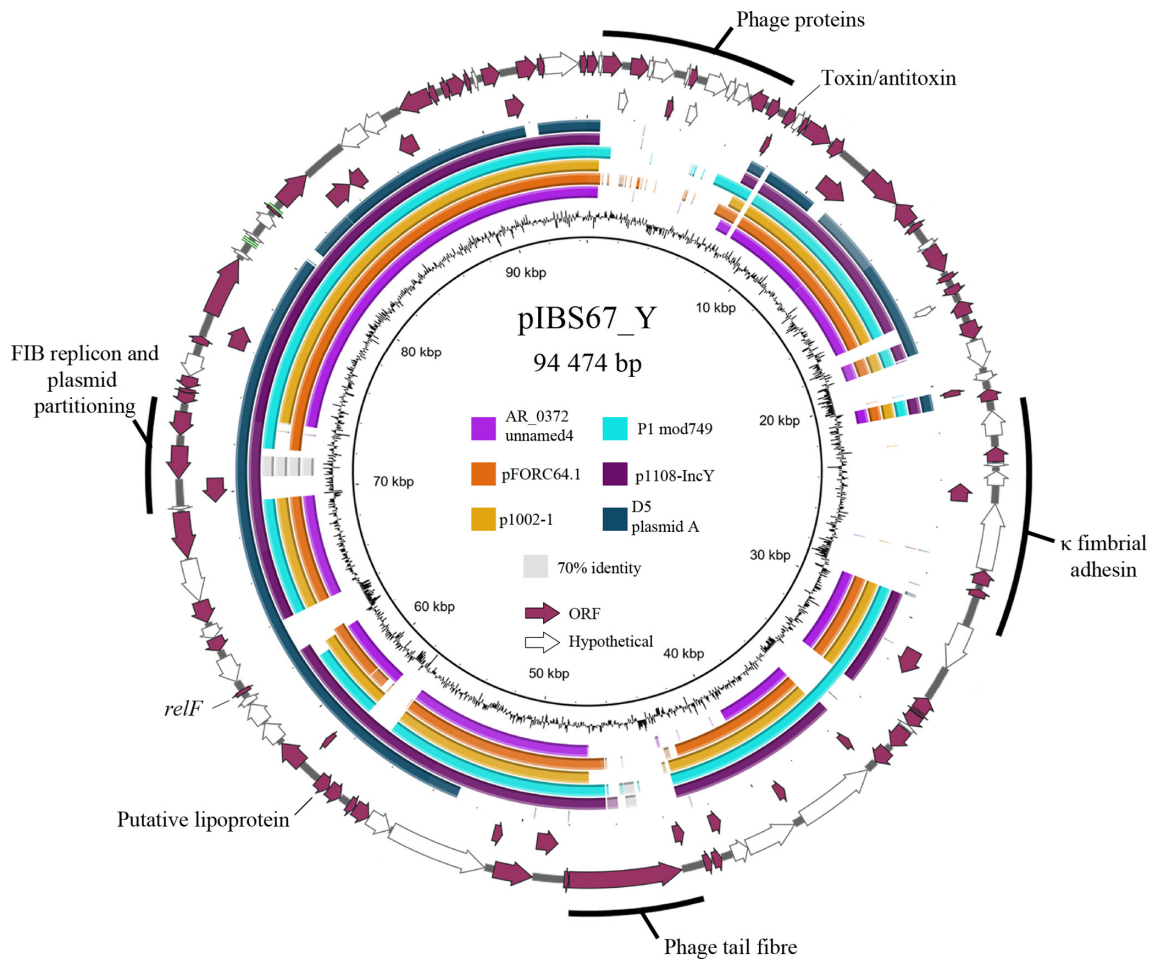
Fig. 6. Plasmid map of pIBS_Y with BLASTN comparisons to close homologues. Plasmid annotations are indicated as arrows in an external ring, with inner rings representing BLASTN values to colour-coded sequences. The centre histogram (black) indicates GC content. Interesting features are indicated.

against well-annotated reference phage 121Q (348, 532 bp) (KM507819), isolated in Canada in 1979 (Fig. 8). The most notable variations observed across the phage sequence are in a tail fibre protein, absent from some phages but present in 121Q, IBS, UB and PBECO_4, plus a set of ORFs encoding for the regulation of chromosome condensation. These ORFs and the region of hypothetical proteins adjacent show significant sequence variation across the family. The remaining smaller sites that show a lack of identity often centre on tRNAs, so are likely hotspots for recombination. These phages host several potentially influential genes, including a dihydrofolate reductase and *telA*, a tellurium resistance gene, as well as many hypothetical ORFs.

Carriage of plasmid replicons

F plasmid replicons were detected in six IBS strains, none of which resolved with obvious contiguity. Individual pMLST results were obtained for five of the six strains (Table 1). Similarly, three I1 plasmid replicons were detected but the plasmids were not resolved through short-read assembly,

two of which gave novel I1 plasmid pMLST profiles, identifying unreported *ardA*, *trbA* and *pilL* alleles. IBS12 (ST6096) was positive for both F and I plasmid replicons, yet it lacked the majority of pMLST genes for both schemes, indicating a potential hybrid plasmid.

DISCUSSION

We describe here an opportunistic sampling and WGS of *E. coli* from the faeces of straw-necked ibis nesting on three geographically distant wetland sites in NSW, Australia. The isolates only hosted chromosomal resistance to antibiotics, generally in the form of efflux pumps [42], and exhibited only limited VAG carriage outside the ST58 isolate. Comparisons to sequence data from Enterobase and the GenBank nucleotide database suggest some of these strains and the mobile genetic elements they carry have close homologues to both Australian and international strains.
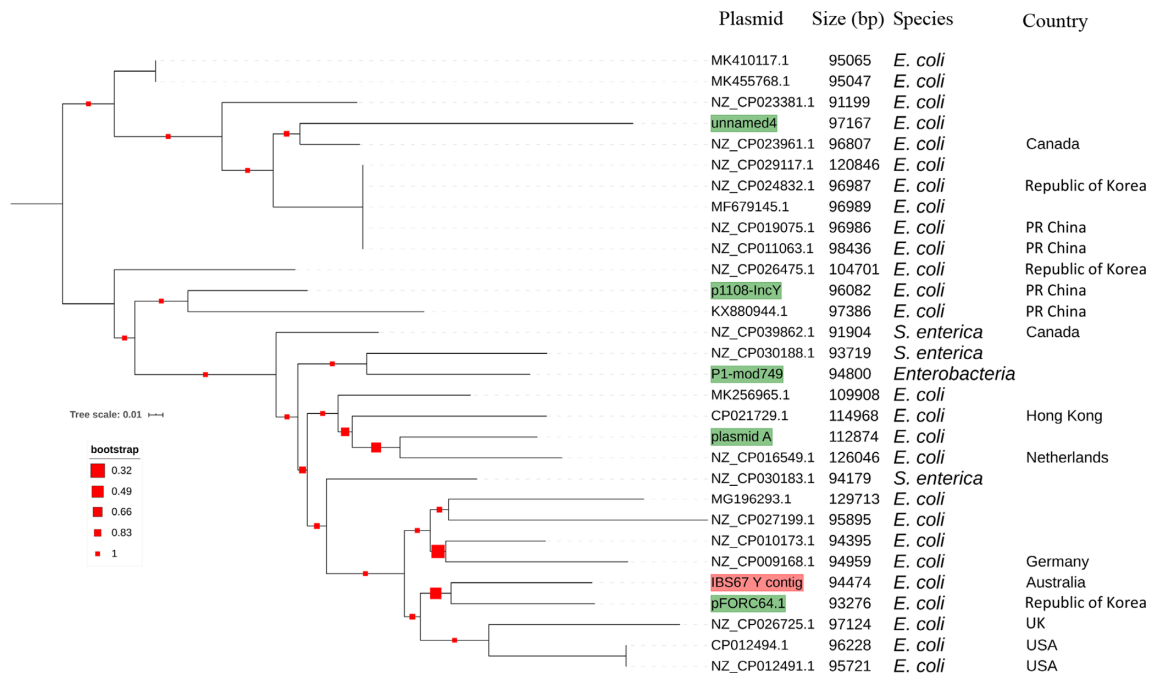
Fig. 7. Mid-point root phylogenetic tree of Y family plasmids. Country, size and species metadata are presented to the right of each tree tip. The IBS67 Y contig is highlighted in red, with plasmids used in the BLASTN comparison (Fig. 6) in green. Node confidence is represented by dot size, with high confidence across the primary split points. Tree scale is in SNPs per site. The conserved core phylogeny (the IBS67 Y contig used as reference) identified a core sequence subset averaging only 28.1 % coverage for each plasmid.

An analysis of the globally distributed ST58 and ST69 groups enabled us to construct core phylogenies to observe their clustering against known Australian pathogens. The remaining phylogenetic analyses demonstrated a combination of apparently novel and previously identified sub-lineages, highlighting several Australian and international linkages to the isolates sourced here from wetlands ibis. For ST58, we saw the tree divided into two major clades, with IBS28 sitting within the larger, more varied, clade. Its closest relatives were all from North America, with the earliest isolation having occurred in 1988, giving the strain ample time to disseminate. The human and avian ST58 pathogens included in the tree were placed into the smaller (uppermost) clade, and interestingly ESC_TA2398AA, an isolate from an Australian silver gull, sat on the same branch as the human pathogen strain. This is an interesting link worth exploring between human and local wildlife microbiomes. It is perhaps expected that there would be multiple lineages of ST58 present in Australia, but further investigation will be required to reveal if the lineage sourced here should be of concern at the national level, particularly with its significant virulence gene content. Strain IBS28 (ST58) had a higher chromosomal virulence gene and IS element count compared to the other IBS strains, including cytotoxin *subAB,* first described on plasmid pO113 from an enterohaemorrhagic *E. coli* with serotype O113 that caused a small outbreak of haemolytic uraemic syndrome in Australia in the 1990s [43]. In addition, we also detected iron-associated genes commonly observed in human and avian extraintestinal pathogens [18–20] and the F17 adhesin known to mediate epithelial adhesion in different hosts, particularly cattle [44]. It should be noted that adhesion for toxin delivery is a primary pathogen trait encoded chromosomally in this ST58 strain. These data raise awareness that *E. coli* ST58 may have a complex commensal/pathogen relationship and strains that acquire virulence gene cargo should be tracked [8].

Similarly, the phylogeny of ST69 strains demonstrated that despite sharing an H antigen allele, the two ST69 strains sourced here are phylogenetically distinct, confirmatory to the observation of different gene profiles between the strains. Phylogenetically, each strain clustered on separate, smaller branches of the overall tree, and their placements identify related strains overseas. The general structure of the tree suggests ST69 is a diverse ST with many different lineages of global distribution, and as such pathogens are likely to be explained by virulence gene acquisition over clonal expansion of a specific ST69 clone, similar to ST58 but unlike more dominant ST131 pathogens [45]. It should be noted that the largest clade in the ST69 phylogeny is composed near-exclusively of human-sourced isolates, however, and comparisons may reveal some traits shared by these strains for successful human colonization. Additionally, ST69 strain IBS1 was closely related to ESC_TA7275AA, an Australian human pathogen isolated from a bloodstream infection in 2017. Our analysis demonstrated clear differences in acquired gene content (both chromosomal and extra-chromosomal) that suggest different evolutionary histories have diversified this lineage.
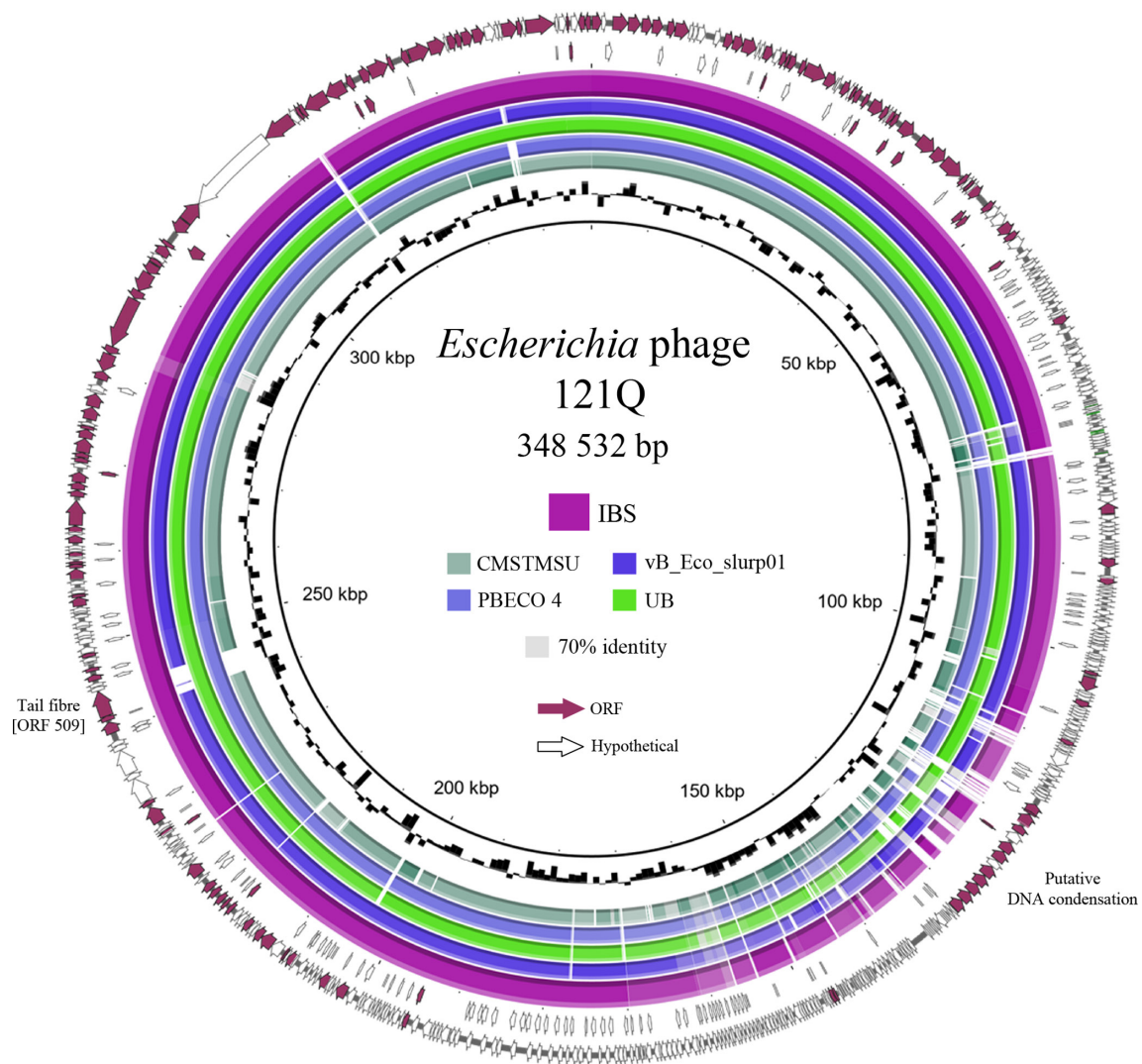
Fig. 8. Map of *Escherichia* phage 121Q with BLASTN comparisons to related phages, including phage IBS. ORF and hypothetical coding sequence (CDS) annotations are indicated as arrows on the external ring, with inner rings representing BLASTN alignment data for colour-coded sequences. The centre histogram (black) indicates GC content.

Regarding the plasmids and other large mobile genetic elements characterized here, it is clear they each have closely related homologues overseas and genetic cargo likely not of environmental avian origin. While the relevance of these elements to cell survival and pathogenicity will require detailed analysis of phenotypes and distribution, their presence in these strains suggests penetration of globally distributed mobile genetic elements into Australian wildlife. A deeper and more expansive understanding of ibis behaviour, diets and flight patterns is needed to enable insight into transmission pathways and potential threats, if any, to humans and food animals. It has been observed previously that the straw-necked ibis does form breeding colonies alongside species of ibis known to co-habitat more with humans, such as the Australian white ibis [46], giving some specific, perhaps traceable, pathways between avian wildlife and humans and the

potential for critical pathogens and antimicrobial resistances to move into the environment.

The ST58 strain IBS58 was host to an extrachromosomal ICE-like element that had a very close homologue in an Australian human pathogen strain found from an *E. coli* collection in our laboratory, and to an ICE-like element sequence deposited in GenBank, but unfortunately with few metadata. Shared gene carriage of the cobalamin synthesis gene on this element suggests that the Australian elements are more closely related compared to RCS93_pII from the database. The isolation of a similarly structured element from a 1957 *Y. pestis* strain (pJARS36) [47] may also assist as a starting point to trace the development of these elements as more examples become available, but for now we are limited to these three examples. These ICE-like elements appear to be benign in captured

content in comparison to the plasmid sequences also observed from the IBS cohort. The presence of the large *Escherichia* phage is similarly interesting, especially as it was identified in two separate strains from different samples at the Macquarie Marshes site. While any functional benefit has yet to be shown for these large phages, their presence is an indicator of the dissemination of these elements, of which few examples have been observed [48, 49].

Hosted on the Y plasmid contig, resolved from the sequence of the ST162 IBS strain, is a kappa fimbrial operon most like the F18 porcine fimbrial adhesin. The kappa fimbriae are a small set of fimbrial operons defined by the presence of kappa-type fimbrial usher protein [50] and include a variety of fimbriae found only in *Enterobacteriaceae*. This island was not observed in any of the Y family elements with the closest backbone nucleotide identity, but was identified in two sequences from the US CDC, in *E. albertii*. Similarly, on the I1 plasmid in IBS46 we see some type 1 fimbriae genes present within one genomic island, and a putative efflux transporter present within another. These islands can be observed in other I family plasmids, but this appears to be the first example of an element carrying both together.

One of the most important observations from this study stems from the identification of individual IS*26*-like elements within both F plasmid-associated and ST58 chromosome-associated scaffolds from these environmental avian strains. IS*6* family elements are the only insertion element of their kind found in Gram-negative bacteria, with a high propensity to form transposable units (both IS*6*-specific and non-site-specific) [51, 52]. This has allowed them, particularly IS*26*, to become one of the most prevalent IS elements involved in the evolution and dissemination of antimicrobial resistance. Given the lack of direct repeats observed here in each of the three cases, we can infer that these elements are remnants of more complex transposition events, in which case they likely hosted mobilized DNA that has now been lost through IS activity and potentially as a result of reduced selective pressure. What is critical is that the remaining elements will act as target sites for the insertion of other IS*6* mobilized genes they encounter, leaving these F family plasmids and the ST58 chromosome as likely targets for the future capture of antimicrobial resistance or integration/fusion of plasmids that also harbour these elements. IS26 is known to mobilise complex resistance gene regions encoding resistance to multiple antibiotic classes into chromosomal locations [53]. Given the ST58 virulence load, a chromosomal capture point for mobilized resistance is a worst-case scenario, particularly given its lack of plasmid carriage. These factors suggest that this strain has an enormous capacity for pathogen development under the right circumstances or may even be a potential pathogen we have captured mid-transition, where it has shed plasmids and other acquired regions no longer under a selective pressure. A second point to note is the high identity match of this chromosomal IS*6* element (and its O-antigen locus) to the *E. coli* KSC9 chromosome, taken from a US porcine sample in 2014 [54]. Both IBS28 (ST58) and KSC9 (ST101) appear to be highly homologous at this locus, with both being O non-typable by *in silico* methods. Given the identical IS*6*-family element context and sequence, we assume this is an IS element being disseminated now with the distribution of the O-antigen cluster and that this is a direct example of identical sequence acquisition observed in strains from Australia in 2012 and the USA in 2014.

This study aimed to source and characterize *E. coli* strains from an environmental avian reservoir, to determine the presence of virulence genes, antimicrobial resistance genes and mobile genetic elements. Despite isolating numerous single colonies from the initial samples, many of the strains we sourced were within two to three SNPs of each other and classified as clonal. The strains presented here represent an example of each lineage observed after WGS. Despite the limited number of lineages in our sampling, the range of STs we observed is generally internationally distributed, with only a single novel ST identified belonging to *E. coli* phylogroup D. Most strains had low virulence carriage, with ST58 being the exception, and a complete lack of known mobilized antimicrobial resistance mechanisms, consistent with a relatively pristine environmental origin. The plasmids resolved from these sequences were host to adhesins and other virulence-associated genes that are yet to be characterized, although these loci have also been observed internationally in both the same and occasionally different contexts, highlighting the interlinked nature of mobilized DNA across time, continents, species and hosts. Finally, the presence of IS*6* family elements highlights a select few of these strains (ST58 and ST69 included) as a potential hotspot for antimicrobial resistance evolution and capture should the necessary selective pressure be applied.

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

Ethical statement
This study was performed under the following ethics approval: Scientific Licence SL101452, UNSW ACEC 14/148A.

Data Bibliography
GenBank accessions and EnteroBase assembly IDs have been included where relevant.

References
1. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L *et al*. Diversity of the human intestinal microbial flora. *Science* 2005;308:1635–1638.
2. Jenkins C, Rentenaar R, Landraud L, Brisse S. Enterobacteriaceae 2017:1565–1578.

3. Wyrsch ER, Roy Chowdhury P, Chapman TA, Charles IG, Hammond JM *et al*. Genomic microbial epidemiology is needed to Comprehend the global problem of antibiotic resistance and to improve pathogen diagnosis. *Front Microbiol* 2016;7:843.

4. Ahmed S, Olsen JE, Herrero-Fresno A. The genetic diversity of commensal *Escherichia coli* strains isolated from non-antimicrobial treated pigs varies according to age group. *PLoS One* 2017;12:e0178623.

5. 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.

6. Reid CJ, DeMaere MZ, Djordjevic SP. Australian porcine clonal complex 10 (CC10) *Escherichia coli* belong to multiple sublineages of a highly diverse global CC10 phylogeny. *Microb Genom* 2019;5.

7. Pietsch M, Irrgang A, Roschanski N, Brenner Michael G, Hamprecht A *et al*. Whole genome analyses of CMY-2-producing *Escherichia coli* isolates from humans, animals and food in Germany. *BMC Genomics* 2018;19:601.

8. 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.

9. Borges CA, Tarlton NJ, Riley LW. *Escherichia coli* from Commercial Broiler and Backyard Chickens Share Sequence Types, Antimicrobial Resistance Profiles, and Resistance Genes with Human Extraintestinal Pathogenic *Escherichia coli*. *Foodborne Pathog Dis* 2019;16:813–822.

10. 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.

11. Ferjani S, Saidani M, Hamzaoui Z, Alonso CA, Torres C *et al*. Community fecal carriage of broad-spectrum cephalosporin-resistant *Escherichia coli* in Tunisian children. *Diagn Microbiol Infect Dis* 2017;87:188–192.

12. Roer L, Hansen F, Thomsen MCF, Knudsen JD, Hansen DS *et al*. WGS-based surveillance of third-generation cephalosporin-resistant *Escherichia coli* from bloodstream infections in Denmark. *J Antimicrob Chemother* 2017;72:1922–1929.

13. 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.

14. Toval F, Schiller R, Meisen I, Putze J, Kouzel IU *et al*. Characterization of urinary tract infection-associated Shiga toxin-producing *Escherichia coli*. *Infect Immun* 2014;82:4631–4642.

15. Coque TM, Baquero F, Canton R. Increasing prevalence of ESBL-producing *Enterobacteriaceae* in Europe. *Euro surveillance: bulletin Europeen sur les maladies transmissibles=European communicable disease bulletin* 2008;13.

16. Wyrsch ER, Hawkey J, Judd LM, Haites R, Holt KE *et al*. Z/I1 Hybrid Virulence Plasmids Carrying Antimicrobial Resistance genes in *S.* Typhimurium from Australian Food Animal Production. *Microorganisms* 2019;7:E299:299

17. Smalla K, Cook K, Djordjevic SP, Klümper U, Gillings M. Environmental dimensions of antibiotic resistance: assessment of basic science gaps. *FEMS Microbiol Ecol* 2018;94

18. 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.

19. Vandekerchove D, Vandemaele F, Adriaensen C, Zaleska M, Hernalsteens J-P *et al*. Virulence-Associated traits in avian Escherichia coli: comparison between isolates from colibacillosis-affected and clinically healthy layer flocks. *Vet Microbiol* 2005;108:75–87.

20. Guabiraba R, Schouler C. Avian colibacillosis: still many black holes. *FEMS Microbiol Lett* 2015;362:fnv118.

21. Johnson JR, Delavari P, Kuskowski M, Stell AL. Phylogenetic distribution of extraintestinal virulence-associated traits in Escherichia coli. *J Infect Dis* 2001;183:78–88.

22. Rodriguez-Siek KE, Giddings CW, Doetkott C, Johnson TJ, Fakhr MK *et al*. Comparison of *Escherichia coli* isolates implicated in human urinary tract infection and avian colibacillosis. *Microbiology* 2005;151:2097–2110.

23. Dolejska M, Literak I. Wildlife is overlooked in the epidemiology of medically important antibiotic-resistant bacteria. *Antimicrob Agents Chemother* 2019;63:e01167–01119.

24. Chen J, Griffiths MW. Pcr differentiation of *Escherichia coli* from other gram-negative bacteria using primers derived from the nucleotide sequences flanking the gene encoding the universal stress protein. *Lett Appl Microbiol* 1998;27:369–371.

25. Márquez C, Labbate M, Raymondo C, Fernández J, Gestal AM *et al*. Urinary tract infections in a South American population: dynamic spread of class 1 integrons and multidrug resistance by homologous and site-specific recombination. *J Clin Microbiol* 2008;46:3417–3425.

26. Harmer CJ, Hall RM. IS26-Mediated formation of transposons carrying antibiotic resistance genes. *mSphere* 2016;1.

27. Wyrsch ER, Reid CJ, DeMaere MZ, Liu MY, Chapman TA *et al*. Complete sequences of multiple-drug resistant IncHI2 ST3 plasmids in Escherichia coli of porcine origin in Australia. *Front Sustain Food Syst* 2019;3.

28. Wirth T, Falush D, Lan R, Colles F, Mensa P *et al*. Sex and virulence in Escherichia coli: an evolutionary perspective. *Mol Microbiol* 2006;60:1136–1151.

29. Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Rapid SF. and Easy *In Silico* Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. *Journal of clinical microbiology* 2015;53:2410–2426.

30. Clermont O, Bonacorsi S, Bingen E. Rapid and simple determination of the Escherichia coli phylogenetic group. *Appl Environ Microbiol* 2000;66:4555–4558.

31. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J *et al*. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom* 2017;3:e000131.

32. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 2010;5:e11147.

33. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* 2014;15:524.

34. 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.

35. 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.

36. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O *et al*. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;58:3895–3903.

37. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M *et al*. Card 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020;48:D517–d525.

38. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–1760.

39. Milne I, Bayer M, Cardle L, Shaw P, Stephen G *et al*. Tablet--next generation sequence assembly visualization. *Bioinformatics* 2010;26:401–402.

40. Galata V, Fehlmann T, Backes C, Keller A. PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Res* 2019;47:D195–D202.

41. **Alikhan N-F**, **Petty NK**, **Ben Zakour NL**, **Beatson SA**. Blast ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 2011;12:402.

42. **Delmar JA**, **Su C-C**, **Yu EW**. Bacterial multidrug efflux transporters. *Annu Rev Biophys* 2014;43:93–117.

43. **Paton AW**, **Woodrow MC**, **Doyle RM**, **Lanser JA**, **Paton JC**. Molecular characterization of a Shiga toxigenic Escherichia coli O113:H21 strain lacking eae responsible for a cluster of cases of hemolytic-uremic syndrome. *J Clin Microbiol* 1999;37:3357–3361.

44. **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.

45. **Pitout J**, **DeVinney R**. Escherichia coli ST131: a multidrug-resistant clone primed for global domination [version 1; peer review: 2 approved]. *F1000. Research* 2017;6.

46. **Brandis KJ**, **Kingsford RT**, **Ren S**, **Ramp D**. Crisis water management and IBIS breeding at Narran lakes in arid Australia. *Environ Manage* 2011;48:489–498.

47. **Eppinger M**, **Radnedge L**, **Andersen G**, **Vietri N**, **Severson G** *et al*. Novel plasmids and resistance phenotypes in Yersinia pestis: unique plasmid inventory of strain Java 9 mediates high levels of arsenic resistance. *PLoS One* 2012;7:e32911.

48. **Kim MS**, **Hong SS**, **Park K**, **Myung H**. Genomic analysis of bacteriophage PBECO4 infecting Escherichia coli O157:H7. *Arch Virol* 2013;158:2399–2403.

49. **Chinnadurai L**, **Eswaramoorthy T**, **Paramachandran A**, **Paul S**, **Rathy R** *et al*. Draft genome sequence of *Escherichia coli* phage CMSTMSU, isolated from shrimp farm effluent water. *Microbiol Resour Announc* 2018;7.

50. **Nuccio S-P**, **Bäumler AJ**. Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. *Microbiol Mol Biol Rev* 2007;71:551–575.

51. **Pong CH**, **Harmer CJ**, **Ataide SF**, **Hall RM**. An IS *26* variant with enhanced activity. *FEMS Microbiol Lett* 2019;366.

52. **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.

53. **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.

54. **Chalmers G**, **Rozas KM**, **Amachawadi RG**, **Scott HM**, **Norman KN** *et al*. Distribution of the *pco* Gene Cluster and Associated Genetic Determinants among Swine *Escherichia coli* from a Controlled Feeding Trial. *Genes* 2018;9:504.