

# SCIENTIFIC REPORTS



OPEN

## Comparative and phylogenetic analysis of a novel family of *Enterobacteriaceae*-associated genomic islands that share a conserved excision/integration module

Alejandro Piña-Iturbe<sup>1</sup>, Diego Ulloa-Allendes<sup>1</sup>, Catalina Pardo-Roa<sup>1</sup>, Irenice Coronado-Arrázola<sup>1</sup>, Francisco J. Salazar-Echegarai<sup>1</sup>, Bianca Sclavi<sup>2</sup>, Pablo A. González<sup>1</sup> & Susan M. Bueno<sup>1</sup>

Genomic Islands (GIs) are DNA regions acquired through horizontal gene transfer that encode advantageous traits for bacteria. Many GIs harbor genes that encode the molecular machinery required for their excision from the bacterial chromosome. Notably, the excision/integration dynamics of GIs may modulate the virulence of some pathogens. Here, we report a novel family of GIs found in plant and animal *Enterobacteriaceae* pathogens that share genes with those found in ROD21, a pathogenicity island whose excision is involved in the virulence of *Salmonella enterica* serovar Enteritidis. In these GIs we identified a conserved set of genes that includes an excision/integration module, suggesting that they are excisable. Indeed, we found that GIs within carbapenem-resistant *Klebsiella pneumoniae* ST258 KP35 and enteropathogenic *Escherichia coli* O127:H6 E2348/69 are excised from the bacterial genome. In addition to putative virulence factors, these GIs encode conjugative transfer-related proteins and short and full-length homologues of the global transcriptional regulator H-NS. Phylogenetic analyses suggest that the identified GIs likely originated in phytopathogenic bacteria. Taken together, our findings indicate that these GIs are excisable and may play a role in bacterial interactions with their hosts.

Genomic Islands (GIs) are horizontally transferred DNA segments integrated into bacterial chromosomes<sup>1</sup>. They are characterized by a G + C content, codon usage bias and dinucleotide frequencies, among other sequence signatures, which usually differs from those of the genome<sup>2</sup>. Many of them are found integrated at the 3'-end of tRNA and tmRNA genes<sup>3</sup>, although different families of GIs can show preference for other genes as integration sites<sup>4-7</sup>. GIs range in size from approximately 10 to 500 kbp<sup>8,9</sup> and encode sets of genes that encompass a wide range of functions for the host bacterium, such as niche colonization, catabolism of diverse substrates, symbiotic relationships, resistance to antimicrobial agents or enhanced virulence<sup>10</sup>. Many GIs contain an excision/integration module that includes an integrase gene, usually belonging to the tyrosine recombinase family<sup>1,11</sup> and a Recombination Directionality Factor (RDF), which is a small protein of approximately 60–180 amino acids<sup>12</sup>. Together, these proteins can promote recombination reactions between Direct Repeated Sequences (DRS), also known as Left and Right attachment sites (*attL* and *attR*) at both ends of GIs. Recombination leads to GI excision from the chromosome and the consequent formation of a circular episomal element that carries one copy of the

<sup>1</sup>Millennium Institute on Immunology and Immunotherapy, Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile. <sup>2</sup>Laboratoire de Biologie et Pharmacologie Appliquée, Centre National de la Recherche Scientifique UMR 8113, École Normale Supérieure Paris-Saclay, Cachan, France. Correspondence and requests for materials should be addressed to S.M.B. (email: [sbueno@bio.puc.cl](mailto:sbueno@bio.puc.cl))

DRS (*attP*), while another DRS (*attB*) remains in the host DNA<sup>5,13–15</sup>. After excision the *attB* and *attP* sites can act as substrates for integrase-mediated recombination, resulting in the re-integration of the GI into the bacterial chromosome<sup>16</sup>. In addition to this, excised islands can also be transferred to other hosts by exploiting co-resident prophages for high-frequency transduction inside their capsids<sup>17</sup>, or transferred by conjugation<sup>18,19</sup>. There is evidence that some GIs are replicative in their circular form<sup>20–22</sup> and that others lack this feature<sup>23</sup>.

Importantly, GIs are susceptible to the loss and gain of genes during their dissemination from one bacterium to another. However, genes encoding the key functions of excision/integration, mobilisation and their regulation remain as a conserved core, as reported for different families of GIs such as the Mobilisable Genomic Islands<sup>5</sup> and the SXT/R391 family of integrative and conjugative GIs<sup>4</sup> present in different Gram-negative bacterial families<sup>4,24</sup>, or the conjugative and mobilisable elements recently found in streptococci<sup>6,7</sup> and the Phage-Inducible Chromosomal Islands (PICIs) of *Staphylococcus aureus* and other Gram-positive strains<sup>17</sup>.

The Region of Difference 21 (ROD21) is an excisable pathogenicity island that has been shown to be important for the virulence of the food-borne pathogen *Salmonella enterica* subsp. *enterica* serovar Enteritidis (*Salmonella* ser. Enteritidis)<sup>15,25,26</sup>, one of the most prevalent serotypes of *Salmonella* in humans and other hosts, such as poultry<sup>27–29</sup>. This genomic island was identified by Thomson *et al.* (2006), in a study that searched for genomic regions that were present in the genome of *Salmonella* ser. Enteritidis, but absent in the genome of *Salmonella* ser. Typhimurium<sup>25</sup>. Genes contained within ROD21 encode potential virulence factors, such as the TlpA protein<sup>15</sup>, which has a toll/interleukin-1 receptor (TIR) domain<sup>26</sup>. A previous study has shown that this protein is involved in bacterial survival within macrophages by disrupting intracellular signaling events that coordinate NF- $\kappa$ B activation and IL-1 $\beta$  secretion<sup>26</sup>. Furthermore, recent studies from our group have shown that changes in the excision rate of ROD21 affect the virulence of *Salmonella* ser. Enteritidis, since mutant strains with reduced rate of excision take more time to cause 100% of mortality in mice, as compared to the wild type strain, likely as a result of changes in the expression of the genes located within the island. In addition, expression of some genes within ROD21, including *tlpA*, showed at least a 3-fold increase when the strain has a reduced or blocked ability to excise the island<sup>30</sup>.

Due to the features described above for ROD21, we decided to perform a computational analysis to identify specific GIs related to this excisable pathogenicity island in order to determine whether these elements are present among other pathogenic members of the *Enterobacteriaceae* family and could thus play a role in their virulence. Since many GI databases use bacterial genomes stored within the RefSeq database for genomic island search, which limits the number of genomes interrogated for island identification, we used a BLASTn-based approach to search a non-redundant database that harbors a larger set of genomes. Using this approach, we found and analyzed different genomic islands within pathogenic and non pathogenic *Enterobacteriaceae* that share a conserved syntenic core with ROD21. These GIs share conserved genes encoding the excision/integration processes of the islands and the majority also share genes encoding putative proteins that are likely involved in conjugal transfer. Importantly, we experimentally corroborated that these islands are able to excise from the chromosome, as previously observed in *Salmonella* ser. Enteritidis<sup>15</sup>. Additionally, some of the GIs identified here were found to carry genes encoding TIR-domain containing proteins and homologues of the global regulator H-NS. Phylogenetic analysis revealed that these islands represent a novel family of GIs present among animal- and plant-pathogenic members of the *Enterobacteriaceae* family.

## Results

**Identification of putative excisable genomic islands with a ROD21-like excision/integration module among pathogenic *Enterobacteriaceae*.** Since integrases and DRS are key factors for the excision of GIs, we decided to perform BLASTn searches using as query a nucleotide sequence from the excisable pathogenicity island ROD21. This 285 nucleotide-long sequence includes the *attL* DRS of ROD21 and the promoter and first 82 nucleotides of the coding sequence of its integrase gene (*SEN1970*). This BLASTn search allowed us therefore to identify putative excisable islands (See Methods). A total of 56 GIs associated with the 3'-end of Asn-tRNA genes were identified among 335 genomes (Table 1). The bacterial genomes found to have ROD21-like GIs belong to some strains of 18 bacterial species from 12 different genera, all members of the *Enterobacteriaceae* family (Table 1). In *Salmonella enterica*, we found islands of interest in 11 different serovars, corresponding to serogroups O:2(A), O:4(B), O:9(D<sub>1</sub>) O:38(P), O:44(V), O:3,10(E<sub>1</sub>) and O:54. Although the size of the GIs ranges from 19 kb (*Photobacterium luminescens* TTO1), to 41 kb (*Pectobacterium atrosepticum* SCRI1043), the majority (73%) of the islands identified have sizes ranging between 21 and 30 kb. These islands are inserted in different Asn-tRNAs but share highly similar DRS with >85% of identity and sizes from 24 to 37 bp (Table S5 in Supplementary File S1). Interestingly, a small number of the ROD21-like GIs seems to be inserted in one of the attachment sites of a different GI located at the 3'-end of an Asn-tRNA gene (e.g. in *P. carotovorum* BC1 and *P. parmentieri* WPP163, Fig. 1).

Notably, the identified ROD21-like GIs are found in many plant and animal pathogenic bacteria, such as phytopathogens from genus *Pectobacterium*, entomopathogenic *Ph. luminescens*, extraintestinal pathogenic *E. coli*, antibiotic-resistant clinical isolates of *E. coli* and *K. pneumoniae* ST258, invasive *K. pneumoniae*, and foodborne disease- and typhoid fever-causing *Salmonella* (Table S2 in Supplementary File S1). Each of the 335 genomes harbor one ROD21-like GI, except for the phytopathogens *P. atrosepticum* SCRI1043, *P. parmentieri* RNS08.42.1 A and *P. parmentieri* WPP163, which have two (Table 1).

**ROD21-like genomic islands share a conserved excision/integration module.** Given that the identified ROD21-like GIs share a conserved integration site and DRS, and that the annotations revealed the presence of different conserved genes with similar distribution, we carried out comparisons of the translated open reading frames (ORFs) (six in total, three reading frames on each DNA strand) using tBLASTx. The majority of ROD21-like GIs (52 out of 56) share a conserved set of genes that could be classified in three main groups: a

Bacterium	Accession N <sup>o</sup>	Island Length (bp)	Islander <sup>a</sup>	Literature <sup>b</sup>	H-NS <sup>c</sup> (aa)	H-NS-t <sup>d</sup> (aa)	Tcp <sup>e</sup> (aa)
<i>Photobacterium luminescens</i> subsp. <i>laumondii</i> TTO1	BX571865.1	18,997	—	—	—	—	—
<i>Salmonella</i> serovar Typhi ERL024120	LT906494.1	19,825	—	—	—	—	—
<i>Salmonella</i> serovar Typhi P-stx-12	CP003278.1	20,309	Sen353.21N	—	—	—	—
<i>Klebsiella michiganensis</i> M1	CP008841.1	20,363	—	—	—	78	—
<i>Escherichia coli</i> S1	CP010226.1	20,717	—	—	—	—	—
<i>Escherichia coli</i> M8	CP019953.1	20,970	—	—	134	80	—
<i>Salmonella</i> serovar Inverness ATCC 10720	CP019181.1	21,047	—	—	—	—	—
<i>Serratia marcescens</i> UMH3	CP018925.1	21,268	—	—	—	78	—
<i>Salmonella</i> serovar Agona 24249	CP006876.1	21,485	Sen74.21N	—	—	—	—
<i>Salmonella</i> serovar Borreze SA20041063	CP019407.1	21,485	—	—	—	—	—
<i>Escherichia coli</i> Ecol_517	CP018965.1	21,622	—	—	134	80	—
<i>Escherichia coli</i> MEM	CP012378.1	21,706	—	—	—	—	—
<i>Escherichia coli</i> GE3	CP012376.1	22,458	—	—	134	80	—
<i>Salmonella</i> serovar Typhi Ty2	AE014613.1	22,588	Sen355.23N	<sup>63</sup>	—	—	—
<i>Salmonella</i> serovar Sloterdijk ATCC 15791	CP012349.1	22,804	—	—	134	80	169
<i>Escherichia coli</i> CI5	CP011018.1	23,855	—	—	134	81	—
<i>Salmonella</i> serovar Typhi M223	LT904854.1	23,951	—	—	—	—	—
<i>Escherichia coli</i> D7	CP010150.1	24,691	—	—	134	80	—
<i>Salmonella</i> serovar Quebec S-1267	CP022019.1	24,817	—	—	134	—	—
<i>Cedecea neteri</i> FDAARGOS_392	CP023525.1	25,678	—	—	—	90	—
<i>Escherichia coli</i> 0127:H6 E2348/69	FM180568.1	25,819	Eco626.26N	IE3 <sup>64</sup>	—	80	—
<i>Salmonella</i> serovar Typhi PM016/13	CP012091.1	26,314	—	—	134	—	—
<i>Salmonella</i> serovar Pullorum ATCC 9120	CP012347.1	26,359	Sen215.26N	—	134	—	293
<i>Salmonella</i> serovar Gallinarum 287/91	AM933173.1	26,404	Sen212.27N	ROD21 <sup>25</sup>	134	—	293
<i>Salmonella</i> serovar Enteritidis P125109	AM933172.1	26,496	Sen204.27N	ROD21 <sup>25</sup>	134	—	293
<i>Salmonella</i> serovar Dublin CT_02021853	CP001144.1	26,498	Sen82.26N	ROD21 <sup>65</sup>	134	—	293
<i>Salmonella</i> serovar Nitra S-1687	CP019416.1	26,500	—	—	134	—	293
<i>Citrobacter freundii</i> 18-1	CP022273.1	27,132	—	—	—	—	169
<i>Salmonella</i> serovar Anatum USDA-ARS-USMARC-1175	CP007483.2	27,139	—	—	134	—	—
<i>Escherichia coli</i> ETEC-2265	CP023346.1	27,200	—	—	134	—	—
<i>Escherichia coli</i> YD786	CP013112.1	27,239	—	—	134	80	—
<i>Enterobacter</i> sp. FY-07	CP012487.1	27,410	—	—	134	—	—
<i>Klebsiella pneumoniae</i> 30684/NJST258_2	CP006918.1	27,495	—	ICEKp258.2 <sup>66</sup>	134	—	139
<i>Pectobacterium wasabiae</i> CFBP 3304	CP015750.1	27,690	—	—	133	90	—
<i>Pectobacterium atrosepticum</i> SCRI1043	BX950851.1	28,081	Pat3.28N	HAI13 <sup>67</sup>	133	83	—
<i>Escherichia coli</i> SF-088	CP012635.1	28,452	—	—	134	—	—
<i>Klebsiella pneumoniae</i> strain AR_0148	CP021950.1	28,777	—	—	134	—	—
<i>Serratia marcescens</i> SM39	AP013063.1	28,938	—	—	134	—	—
<i>Serratia marcescens</i> CAV1492	CP011642.1	29,010	—	—	134	—	—
<i>Kluyvera intermedia</i> CAV1151	CP011602.1	29,250	—	—	134	—	—
<i>Escherichia coli</i> ED1a	CU928162.2	29,355	—	—	134	—	—
<i>Klebsiella pneumoniae</i> 1084	CP003785.1	29,853	Kpn29.30N	—	134	—	—
<i>Yersinia intermedia</i> Y228	CP009801.1	30,275	—	—	—	83	—
<i>Pectobacterium parmentieri</i> WPP163	CP001790.1	30,288	Pwa2.30N	—	—	—	—
<i>Raoultella ornithinolytica</i> Yangling I2	CP013338.1	30,440	—	—	134	—	—
<i>Escherichia coli</i> ABWA45	CP022154.1	30,538	—	—	—	—	—
<i>Yersinia rohdei</i> YRA	CP009787.1	30,663	—	—	—	81	—
<i>Pectobacterium parmentieri</i> RNS08.42.1 A	CP015749.1	33,037	—	—	133(2)	80	—
<i>Pectobacterium carotovorum</i> subsp. <i>brasiliense</i> BC1	CP009769.1	33,571	—	—	133	86	—
<i>Pectobacterium parmentieri</i> RNS08.42.1 A	CP015749.1	33,684	—	—	—	—	—
<i>Pectobacterium parmentieri</i> WPP163	CP001790.1	33,698	Pwa2.34N	—	133	81	—
<i>Pectobacterium atrosepticum</i> 21 A	CP009125.1	35,819	—	—	132	87	—
<i>Pectobacterium wasabiae</i> SCC3193	CP003415.1	37,046	—	—	—	—	—
<i>Pectobacterium carotovorum</i> SCC1	CP021894.1	37,050	—	—	—	—	—
<i>Klebsiella oxytoca</i> AR_0147	CP020358.1	37,983	—	—	—	—	—
<i>Pectobacterium atrosepticum</i> SCRI1043	BX950851.1	40,824	Pat3.41N	HAI7 <sup>67</sup>	132	80	—

**Table 1.** *Enterobacteriaceae*-associated ROD21-like genomic islands. ROD21 islands found in different serovars of *Salmonella* or other bacterial species are listed in order of decreasing length. Letters in superscript indicate: <sup>(a)</sup>names of genomic islands already identified in the Islander database or <sup>(b)</sup>in the literature with their indicated references and <sup>(c)</sup>the predicted length in amino acids for the full-length and <sup>(d)</sup>short versions of H-NS homologues, as well as <sup>(e)</sup>TIR-like domain containing proteins when present. Note: *Salmonella* ser. Pullorum is considered as a biovar of serovar Gallinarum and the Pwa2.30 N island was found only by searching in the Islander database.

first group includes ORFs encoding putative type-IV pilin and a type-IV pilus-related protein; a second group includes ORFs encoding putative relaxase belonging to the MobA/MobL family and a TraD homologue, and the third group includes ORFs encoding a P4-like integrase from the tyrosine-recombinase family, together with a putative RDF (Fig. 1). Four ROD21-like GIs showed exceptions, namely those found in *P. wasabie* SCC3193, *P. carotovorum* SCC1, *P. parmentieri* RNS08.42.1 A and *P. parmentieri* WPP163, which only encode the excision/integration module (Fig. 2). In all cases, the coding sequences of the integrases and putative RDFs are located near to each attachment site.

Interestingly, in most ROD21-like GIs the putative RDF-coding sequence is accompanied by an 843 bp gene of unknown function. An aspect that is not shared by all ROD21-like GIs, but is present in many of them, is a group of genes encoding putative H-NS and SpnT homologues, Toll/Interleukin-1 receptor (TIR)-domain containing proteins (Tcps) and type-III restriction-modification (R-M) systems (Figs 1, 2 and S5). Regarding the H-NS homologues identified, we found sequences that are predicted to encode full-length (132–134 amino acids), as well as truncated forms of the protein (78–90 amino acids) (Table 1). Notably, 14 of the 33 ROD21-like GIs encoding full-length H-NS homologues also encode a truncated form of this protein, whose coding sequence is located in the opposite direction. On the other hand, six other ROD21-like GIs only encode truncated forms of the H-NS homologues (Fig. 2).

Since the identified ROD21-like GIs are only found in *Enterobacteriaceae* and that the genomic island most studied to date within this group is ROD21, we named this group of GIs the *Enterobacteriaceae*-associated ROD21-like (EARL) genomic islands.

**Phylogenetic analyses of EARL GIs.** Because integrases are involved in GI excision/integration and dissemination between bacterial species or genera, we used them as markers to study the evolutionary history of EARL GIs<sup>1</sup>. We performed a reconstruction of EARL GIs phylogeny based on the nucleotide sequences of their cognate integrases. The resulting maximum likelihood tree shows that the integrases form two major clades: one at the bottom of the tree, which clusters EARL GIs present only in phytopathogens, while the other clade clusters EARL GIs from both non-pathogenic and pathogenic strains infecting both animals and plants (Fig. 2a). All the phytopathogenic strains belong to the genus *Pectobacterium*. Interestingly, the subclade that groups the non-pathogenic and pathogenic animal strains branches into two clusters: one includes EARL GIs present in *Escherichia coli* and *Salmonella enterica*, while the other one harbors EARL GIs carried by a greater diversity of species, such as *Klebsiella pneumoniae*, *K. oxytoca*, *K. michiganensis*, *Raoultella ornithinolytica*, *Kluyvera intermedia*, *Serratia marcescens*, *Yersinia rhodei*, *Y. intermedia*, *S. enterica* serovars Anatum and Typhi, *Enterobacter* sp., *Citrobacter freundii*, *Cedecea neteri* and *E. coli* (Fig. 2). This cluster includes the carbapenem-resistant *K. pneumoniae* NJST258-2, which harbors a EARL GI previously denominated ICEKp258.2 (Table 1).

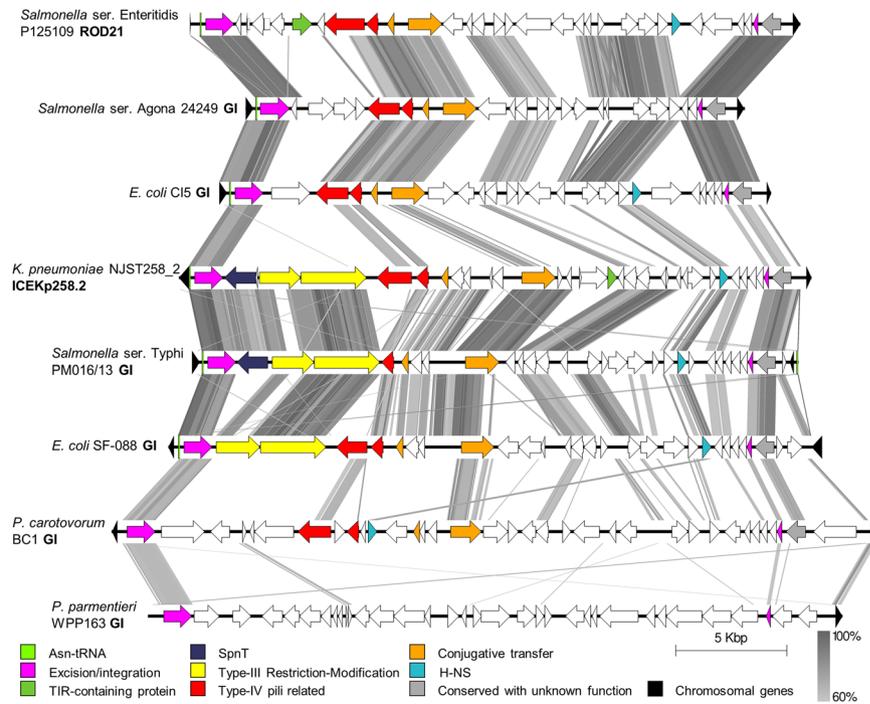
The presence of conserved genes within EARL GIs correlates with the integrase-based EARL GI phylogeny. While all EARL GIs encode an integrase and have RDF genes, as well as DRS (Figs 1 and 2b), the major clades are differentiated by the presence of genes encoding putative proteins related to type-IV pili, conjugative transfer and H-NS homologues (Fig. 2b). In the animal pathogen subclade, phylogeny correlates with the acquisition of genes encoding a type-III R-M system and homologues of the SpnT encoding gene.

Additionally, we constructed a phylogenetic tree based on the *hns* homologues carried by EARL GIs in order to gain a better understanding of the relationships between the full-length forms of these proteins and the truncated ones. The unrooted maximum likelihood tree shows that full-length and truncated forms of *hns* separate into two distantly related clades (Fig. 3). As with the integrase-based phylogeny, the full-length homologues form two sub-clades, comprised by the genes carried by plant and animal pathogenic bacteria. In contrast, the truncated homologues show a different branching pattern, in which *hns* from phytopathogens are more closely related to genes of some animal pathogenic bacteria (Fig. 3). It is noteworthy that full-length homologues are more closely related between themselves than with the chromosomal *hns* encoded by *E. coli* K-12 (Fig. 3).

To corroborate that the genomes included in this study were fully syntenic, we calculated the distance of the GI insertion site relative to the chromosomal replication origin (*oriC*) for each EARL GI (Table S2 in Supplementary File S1). Most of the EARL GIs are located far from *oriC*, at 66.4–98.2% (mean = 77.5%) of the maximum possible distance (considered as one half of the chromosome length) (Supplementary Fig. S1). However, the EARL GI from *Salmonella* ser. Inverness ATCC 10720 is located at 39.7% of the maximum distance, probably as a result of a chromosomal rearrangements that locate the *Asn*-tRNAs near *oriC* (Supplementary Fig. S2).

**Inability to excise ROD21 modulates expression of genes within this pathogenicity island.** In order to corroborate the hypothesis that ROD21 excision modulates the expression of genes within this genomic island, a *Salmonella* ser. Enteritidis strain lacking the genes coding for the integrase and RDF was generated. As shown in Fig. 4a, deletion of these genes prevents ROD21 excision. Both wild-type and mutant strains were grown in LB under standard laboratory conditions (37 °C, pH 7.0) and RNA was purified to evaluate, by RT-qPCR, the expression of the ROD21 genes *SEN1975*, *SEN1980*, *SEN1993*, which encode the TlpA protein, the putative relaxase and the H-NS full-length homologue, respectively. As show in Fig. 4b–d, impairment of ROD21 excision prevented the proper expression of ROD21 genes in the mutant strain. As a control, expression of the housekeeping gene *rpoD*, which is located on the *Salmonella* ser. Enteritidis genomic core, and outside ROD21, was also evaluated for both strains. We observed no changes in expression of *rpoD* (Fig. 4e), suggesting that the impairment of ROD21 excision affects specifically genes within this island. These results support the notion that ROD21 excision influences expression of genes within the GI.

**The EARL GIs from *K. pneumoniae* ST258 strain KP35 and enteropathogenic *E. coli* O127:H6 strain E2348/69 are excisable.** Since ROD21 can be excised from the bacterial chromosome and this



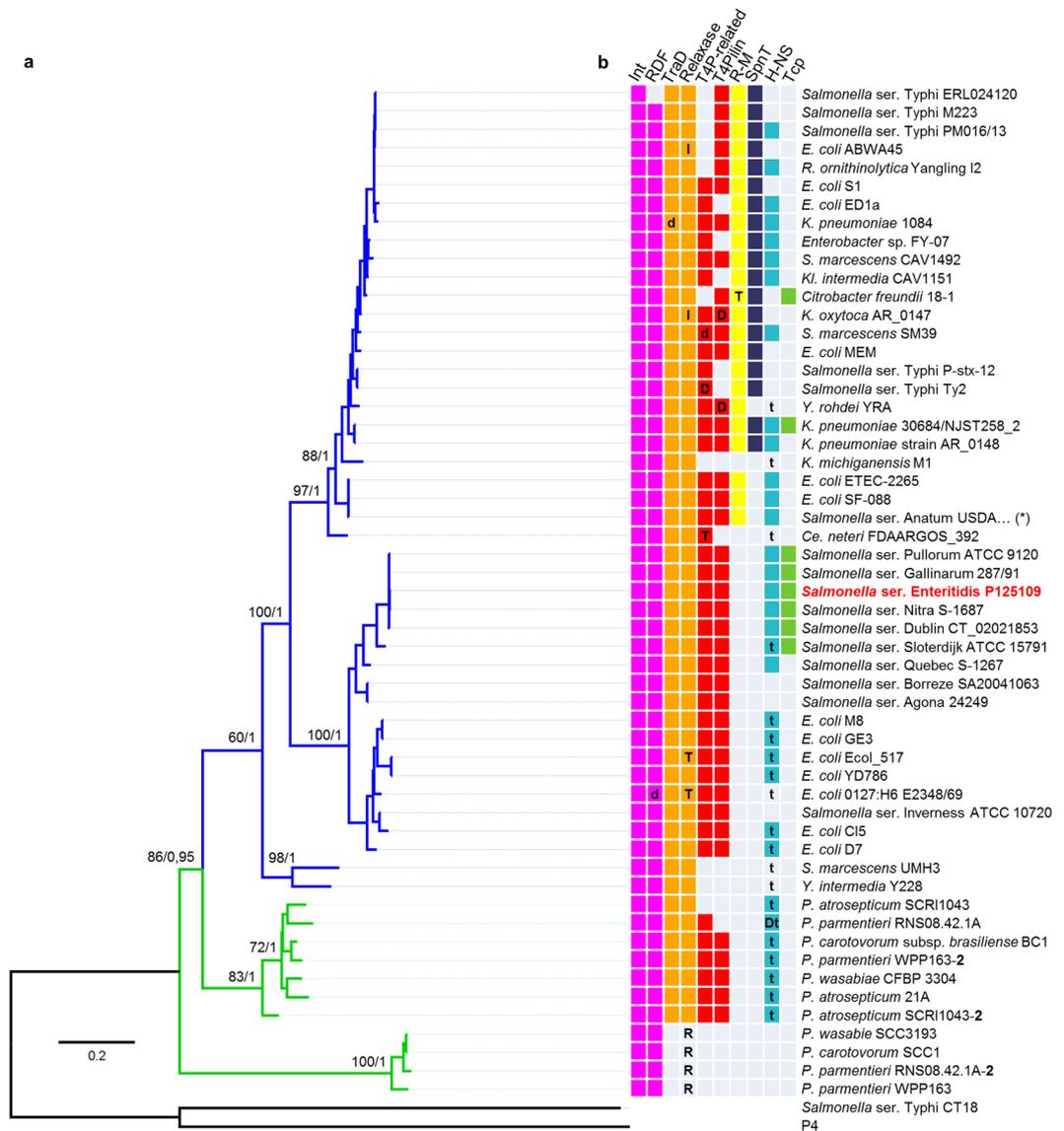
**Figure 1.** EARL genomic islands have a conserved excision/integration module. Sequence comparison (all three-frames translated for each DNA strand) of genomic islands carried out with tBLASTx. This figure shows eight representative GIs. Gray lines indicate regions with 60–100% identity. Note that the islands of *P. carotovorum* BC1 and *P. parmentieri* WPP163 are not inserted at Asn-tRNA genes (the green bars in the left side of the islands).

process can be modulated by temperature and pH, we hypothesized that EARL GIs, which share conserved excision/integration modules with ROD21, would also be excisable. For this purpose, the excision of the ICEKp258.2 island from *K. pneumoniae* ST258 KP35 and the IE3 island from the enteropathogenic *E. coli* O127:H6 (EPEC) E2348/69 was assessed by PCR. After overnight growth in LB broth at 23 °C/pH 7, 37 °C/pH 5.4, and 37 °C/pH 7, genomic DNA was purified and assessed by nested PCR with primers targeting the attachment sites that would result in an amplicon after island excision (Figs 5a and 6a).

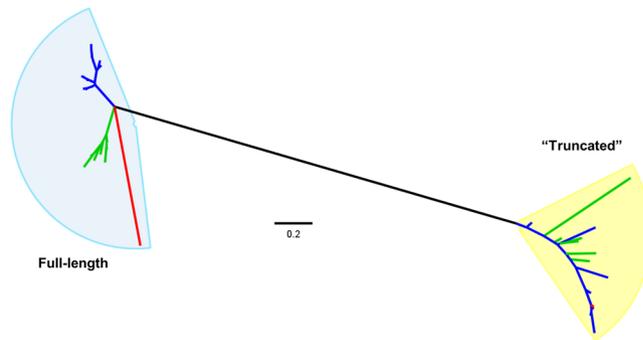
For ICEKp258.2 excision, according to the sequence analyses, the PCR products for *attB* and *attP* should be 519 bp and 668 bp long, respectively. As shown in Fig. 5b, amplicons with the expected size were obtained. These amplicons can only be found if the EARL GIs are absent, indicating that the island was excised from the bacterial chromosome. Although the *attB* and *attP* sites were detected under all tested conditions, the product that indicates GI excision was scarcely detectable at 37 °C/pH 7 (Fig. 5b). For conditions 23 °C/pH 7 and 37 °C/pH 5.4 see Supplementary Fig. S3. Similar results were obtained for EPEC E2348/69 with amplicons matching the predicted size of 507 bp and 502 bp for *attB* and *attP* (Figs 6b and S4). However, for both bacteria, we found that other amplicons were observed after electrophoresis of the *attB* and *attP* PCR products, suggesting that the primers might produce nonspecific products. Therefore, to corroborate that the PCR products obtained were specifically from the *attB* and *attP* sites as expected, the amplicons were purified and sequenced. As shown in Fig. 5c, the sequence obtained for the PCR product correspond to the predicted *attB* site generated in the chromosome of *K. pneumoniae* after ICEKp258.2 excision. However, the sequence of the amplicons obtained for the *attP* site did not match the predicted sequence of ICEKp258.2 *attP* (Fig. 5c). Since several amplicons were obtained after the second round of PCR to detect the *attP* site, it is possible that the PCR conditions used in this assay were not optimal for the amplification of the expected *attP* sequence. On the contrary, for EPEC E2348/69, the sequences obtained for both *attB* and *attP* amplicons were identical to the predicted sites after IE3 island excision (Fig. 6c). These findings support the notion that the EARL elements are excisable islands.

## Discussion

Studies in the PPHGI-1 island of *Pseudomonas syringae* pathovar *phaseolicola* and ours in ROD21 indicate that GI excision from bacterial chromosomes can modulate the expression of genes contained within the islands and influence the virulence of the bacteria that harbor them<sup>15,30–32</sup>. Given the role of ROD21 excision in the virulence of *Salmonella* ser. Enteritidis, we carried out computational analyses to identify similar GIs in the bacterial genomes available at the time of this study. Here, we found that ROD21-like GIs are a novel group of genomic islands encoded by members of *Enterobacteriaceae*, including plant and animal pathogenic strains. Given their presence in this bacterial family, we have designated these GIs as the *Enterobacteriaceae*-associated ROD21-like genomic islands (EARL GIs). Importantly, the bioinformatics approach performed in this study used the non-redundant nucleotide database of GenBank, which has allowed us to identify EARL GIs among all



**Figure 2.** Genomic island phylogeny correlates with the distribution of conserved genes. **(a)** Maximum likelihood tree based on codon-aligned nucleotide sequences of island integrases with support values corresponding to 100 bootstrap replicates and posterior probabilities (BS/p; shown only for basal nodes). The integrase from the SPI7 island and phage P4 were used as outgroups. Blue and green branches correspond to GIs harbored by animal and plant bacterial pathogens, respectively. **(b)** Distribution of conserved genes coloured as in Fig. 1. Bold letters inside the squares indicate whether the corresponding ORF is duplicated (D), disrupted (d), truncated (T) or has an insertion (I). For H-NS, the presence of full-length (colored square) or truncated (t) homologues is indicated. For the relaxases, the presence of a relaxase different from that of the Moba/MobL family is also specified



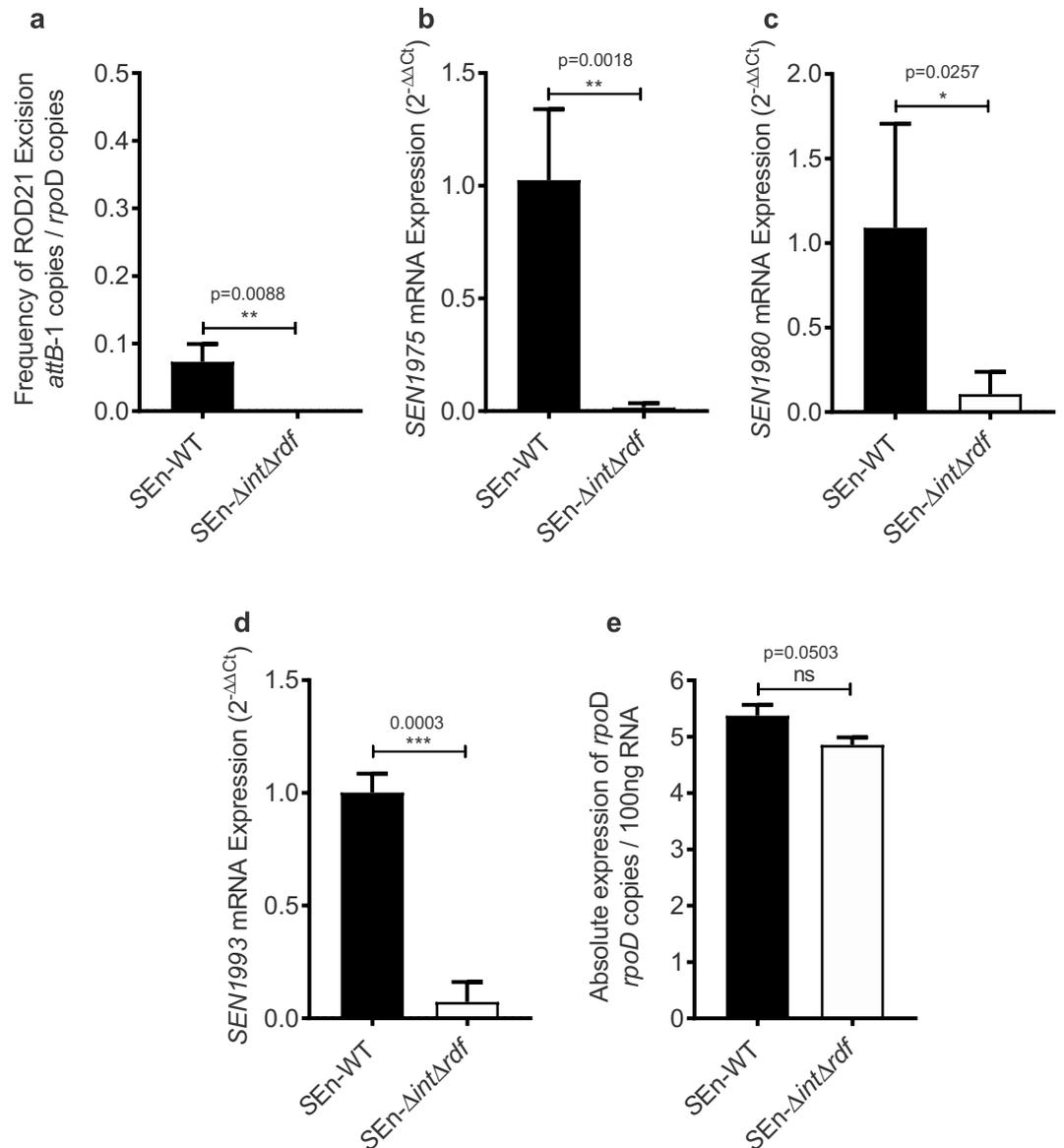
**Figure 3.** Full-length and truncated *hms* homologues belong to distantly related clades. Unrooted maximum likelihood phylogenetic tree constructed with codon-aligned *hms* homologue sequences found in EARL GIs. Blue and green branches correspond to animal- and plant-associated bacteria, respectively. The red branches represent chromosomal *hms* of *E. coli* K-12 strain MG1655 and truncated *hms* carried by the IE3 island of EPEC E2348/69.

Dissemination of many GIs relies on bacterial conjugation<sup>18,34</sup> and thus these elements are classified in two groups based on whether they encode the molecular machinery required for their excision and conjugation. While the integrative and conjugative elements (ICEs) encode both sets of genes<sup>35</sup>, the integrative and mobilisable elements (IMEs; also known as mobilisable genomic islands) utilize the transfer machinery encoded by ICEs or conjugative plasmids<sup>36–38</sup>.

Most EARL islands encode a TraD homologue and a putative relaxase of the MobA/MobL (MOB<sub>Q</sub> superfamily)<sup>39</sup>. Since EARL GIs are not self-conjugative elements, because they don't encode the Type-IV secretion system (T4SS) required for conjugal transfer, the encoded relaxases may provide the means for island mobilisation by recognizing their cognate origin of transfer (*oriT*) and recruiting the island DNA to a T4SS encoded by a co-resident conjugative element<sup>40</sup>. Transfer of ROD21 has been observed from *Salmonella* ser. Enteritidis to mutant strains lacking the island, as well as to *Salmonella* ser. Typhimurium<sup>18</sup>. Experiments assessing ROD21 mobilisation from *Salmonella* ser. Typhimurium LT2 with and without its virulence plasmid pSLT, have shown that transfer of the island is only achieved from bacteria harboring pSLT to a pSLT-positive strain<sup>18</sup>. Therefore, similar mechanisms could be responsible for the horizontal transfer of other EARL GIs.

Importantly, GI acquisition may pose a metabolic burden for bacterial hosts since and increased amount of DNA in the bacterial genome require its replication and maintainance. Furthermore, a lack of regulatory mechanisms for controlling the expression of the newly acquired genes poses metabolic challenges for the bacterium<sup>41</sup>. Here we support the notion that the ability of EARL islands to excise from the chromosome might be a mechanism conserved among these GIs to modulate gene expression. We have demonstrated here that expression of genes within ROD21 is affected in a mutant strain of *Salmonella* ser. Enteritidis unable to excise this GI. However, while we found a reduction in expression of the genes encoding TlpA, the putative relaxase and the H-NS full-length homologue (Fig. 4), a previous work assessing the effects of ROD21 excision found the opposite effect<sup>30</sup>. Differences in the mutant strains might account for observed results. In that work, the strain with reduced excision and the one that lacked this capacity were constructed by deletion of the integrase ORF ( $\Delta$ SEN1970), and the entire region spanning the integrase and the Asn-tRNA which has the *attL* site of the island [ $\Delta$ (*asnT2*-SEN1970)], respectively; while our mutant strain lacks only the integrase and putative RDF ( $\Delta$ SEN1970  $\Delta$ SEN1998). Therefore, the absence of SEN1998 may be causing the different outcome. Previous studies on the *Vibrio* Pathogenicity Islands 1 and 2, which also have their integrase gene located downstream the attachment site as in ROD21, provided evidence that their RDFs act as transcriptional repressors of the integrase by binding the *att* region which overlaps the integrase promoter<sup>42</sup>. Further research is required to elucidate the relationship between SEN1998 and gene expression in ROD21. Nevertheless, inability to excise ROD21 has always resulted in altered gene expression inside the island in all mutant strains tested to date. This phenomenon might account for a mechanism used by pathogens to modulate the expression of newly acquired virulence genes, useful in specific stages of their infective cycle<sup>32,43</sup>.

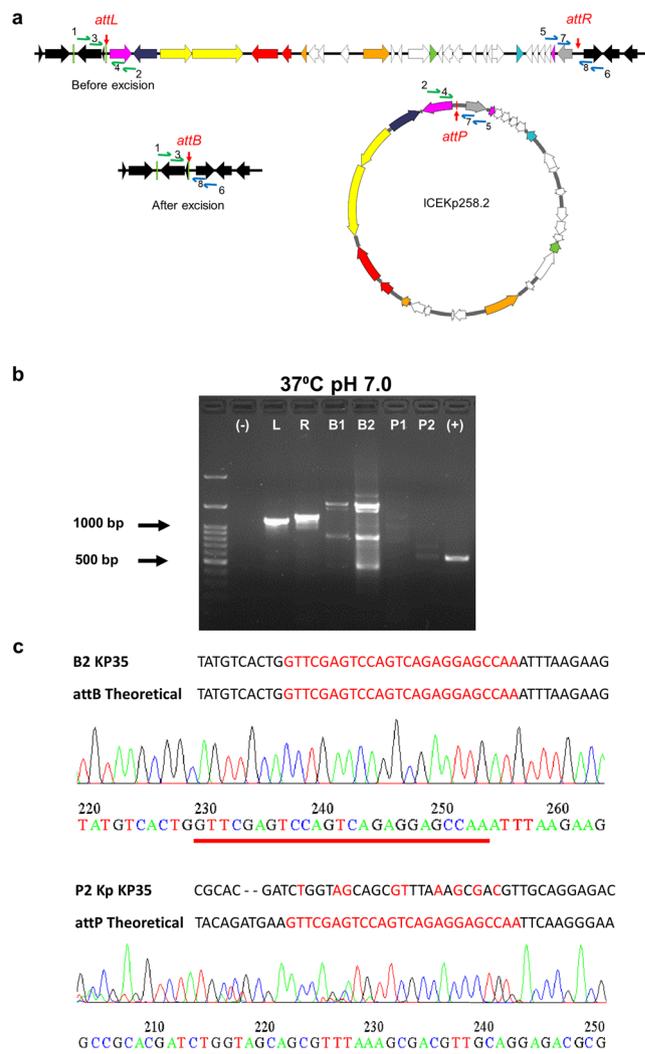
Another mechanism for regulation of gene expression inside GIs is mediated by the heat-stable nucleoid structuring (H-NS) protein, which plays crucial roles due to its ability to recognize and bind to DNA sequences with low G + C content in such a way to form nucleoprotein complexes capable of limiting RNA polymerase activity, in a process designated xenogeneic silencing<sup>41,44–47</sup>. This may explain why most (33/54) of EARL GIs encode predicted full-length H-NS homologues, since they have low G + C content. The presence of such genes could facilitate island dissemination onto new bacterial hosts and its preservation, by providing a regulatory mechanism for EARL GI-gene expression. Further research is required to assess whether these homologues are integrated in the global transcriptional network of their hosts. For example, the Hfp protein, an H-NS homologue encoded in the *serU* island of some UPEC strains participates in the regulation of virulence factors encoded outside the island<sup>48</sup>. It is noteworthy that 20 of the 54 identified islands also carry genes predicted to encode truncated forms of the H-NS protein such as H-NST, encoded in the IE3 island of enteropathogenic *E. coli* E2348/69<sup>49</sup>. This 80 amino acid-long protein (compared with the 137 aa full-length H-NS) corresponds to the N-terminal dimerization domain of H-NS, which was shown to interact with H-NS as an antagonist<sup>49</sup>. The H-NST homologues encoded in



**Figure 4.** ROD21 excision affects the gene expression inside the island. (a) Deletion of the integrase (*SEN1970*) and putative RDF (*SEN1998*) coding sequences from *Salmonella* ser. Enteritidis results in impairment of ROD21 excision and reduction in expression of (b) *SEN1975*, (c) *SEN1980* and (d) *SEN1993* under *in vitro* conditions (LB, pH 7.0, 37°C, OD<sub>600</sub> = 0.6). (e) Expression of *rpoD*, located outside the island, was not affected. Unpaired, two-tailed *t* test was used with  $\alpha = 0.05$ . Error bars represent the standard deviation.

the other EARL GIs could also be playing a role in relieving H-NS-mediated repression. Our phylogenetic analysis of short and full-length *hns* homologues found in EARL GIs, including *hnsT* of EPEC E2348/69, shows both groups clustering as separate, distantly related clades evidencing that the short ones are not simply remnants that resulted after an event of duplication or acquisition. The differing topology within each clade also suggests that acquisition of the truncated and full-length versions of *hns* homologues by EARL GIs represents two separate, independent events.

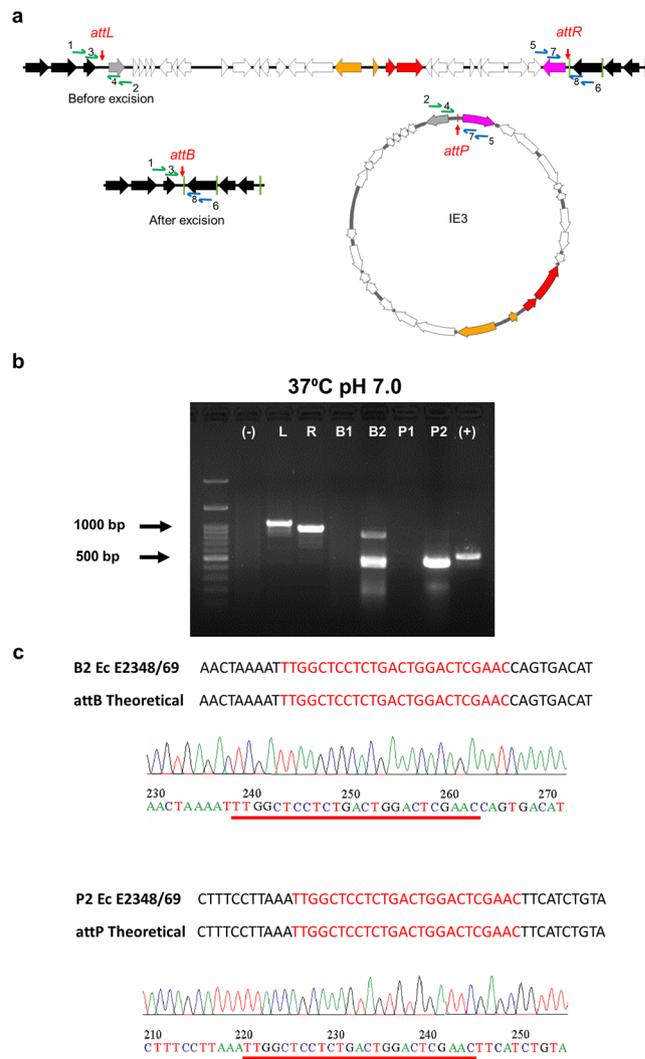
Analysis of EARL islands and multiple sequence alignments (Supplementary Fig. S5) allowed us to identify ORFs likely encoding Toll/Interleukin-1 receptor (TIR)-like domain containing proteins (Tcps) in three different islands (Table 1), in addition to the already characterized TlpA protein encoded in ROD21<sup>26</sup>. Bacterial Tcps are known virulence factors able to disrupt the innate immune response by interacting with mammalian TIR domain-containing proteins of the TLR signalling pathway, such as TLRs, IL-1R, MyD88 and TIRAP, thus suppressing transactivation of the NF- $\kappa$ B transcriptional activator, with the consequent alteration of proinflammatory cytokine secretion<sup>50</sup>. Interestingly, we identified an ORF encoding a putative Tcp in the ICEKp258.2 island found in carbapenem-resistant *K. pneumoniae* ST258 strains<sup>51</sup>, which may have a role in the resistance of this ST against phagocytosis<sup>52</sup>. The other two putative Tcp encoded in the GIs from *Salmonella* ser. Sloterdijk ATCC 15791 and *Citrobacter freundii* 18-1 may also play a role in virulence of these pathogens.



**Figure 5.** Detection of ICEKp258.2 excision in *K. pneumoniae* ST258 KP35. (a) Locations of the eight primers used for the nested PCR are indicated with blue and green arrows next to the *attL*, *attR*, *attB* and *attP* sites. Primers 1, 2, 5, 6 and 3, 4, 7, 8 were used for the first and second round in nested PCR, respectively. Amplicons containing *attB* or *attP* can be obtained only if the 27 kb island is excised from the chromosome. Gene colouring is the same as in Fig. 1. (b) Agarose gel showing amplification products (L: 1123 bp, R: 1220 bp, B1: 1029 bp, P1: 1340 bp, B2: 519 bp, P2: 668 bp) of nested PCR reactions performed using genomic DNA obtained from *K. pneumoniae* KP35 grown at 37 °C, pH 7.0. (–), negative control; L: *attL*, R: *attR*, B1: *attB* from first round of PCR, B2: *attB* from second round of PCR, P1: *attP* from first round of PCR, P2: *attP* from second round of PCR; (+) positive control *rpoD* (577 bp) for the PCR reaction. UV exposure time was 400 ms. (c) The sequence of PCR products from the second nested PCR round for *attB* and *attP* (B2 and P2) were obtained and compared with theoretical sequences. *attB* and *attP* specific sequences are highlighted in red and chromatograms are shown.

As shown in Figs 1 and 2, open reading frames encoding functions related to horizontal transfer (i.e. excision/integration and conjugative transfer genes), as well as other ORFs of unknown function are conserved among EARL GIs, suggesting that an ancestral island gave rise to the diversity of EARL GIs. Our phylogenetic analysis, based on the nucleotide sequence of the integrase open reading frames reveals that island gene content correlates with the evolution of the integrase. Based on the phylogenetic reconstruction, we hypothesize that EARL GIs originated in phytopathogenic bacteria and later, through horizontal gene transfer, disseminated to other bacteria, losing some genes and gaining others that are favorable for the recipient bacteria. Interestingly, a closer relationship between EARL GIs is not necessarily correlated with bacterial species that are closer to each other, which suggests that multiple, independent transfer events have occurred more recently than bacterial speciation.

Because most EARL GIs encode intact excision/integration modules, the genomic region should theoretically undergo excision from the chromosome. Excision of EARL GI has so far only been demonstrated for ROD21 in *Salmonella* ser. Enteritidis P125109, which responds to environmental stimuli<sup>15</sup>. Here, we show that the ICEKp258.2 island from *K. pneumoniae* ST258 KP35 and the IE3 island found in EPEC E2348/69 are excisable elements. These results support the notion that an important feature of EARL GIs is their capacity to excise from



**Figure 6.** Detection of IE3 excision in EPEC E2348/69. **(a)** Locations of the eight primers used for the nested PCR are indicated with blue and green arrows next to the *attL*, *attR*, *attB* and *attP* sites. Primers 1, 2, 5, 6 and 3, 4, 7, 8 were used for the first and second round in nested PCR, respectively. Amplicons containing *attB* or *attP* can be obtained only if the 26 kb island is excised from the chromosome. Gene colouring is the same as in Fig. 1. **(b)** Agarose gel showing amplification products (L: 1057 bp, R: 937 bp, B1: 985 bp, P1: 1009 bp, P2: 502 bp, B2: 507 bp) of nested PCR reactions performed using genomic DNA obtained from EPEC E2348/69 grown at 37°C, pH 7.0. (–), negative control; L: *attL*, R: *attR*, B1: *attB* from first round of PCR, B2: *attB* from second round of PCR, P1: *attP* from first round of PCR, P2: *attP* from second round of PCR; (+) positive control *rpoD* (577 bp) for the PCR reaction. UV exposure time was 400 ms. **(c)** The sequence of PCR products from second nested PCR round for *attB* and *attP* (B2 and P2) were obtained and compared with theoretical sequences. *attB* and *attP* specific sequences are highlighted in red and chromatograms are shown.

the chromosome. This property may be relevant for the capacity of EARL GIs to mobilise and for modulation of gene expression and virulence, as described for ROD21<sup>18,30</sup>.

With the availability of increasing numbers of sequenced genomes, a great diversity of GIs has been described to date in Gram-positive and Gram-negative bacteria. Some examples of early identified and well studied families of genomic islands are the Phage Inducible Chromosomal Islands (PICIs)<sup>17,53</sup> and the Staphylococcal Chromosomal Cassette *mec* (SCC*mec*)<sup>54</sup> in Gram-positive bacteria and the SXT/R391 family in Gram-negative  $\gamma$ -Proteobacteria<sup>4,24</sup>, which encode genes responsible for virulence and antimicrobial resistance. The EARL islands described in this study add to the already substantial and ever-growing number of GIs and GI-families found in bacterial genomes, and share the overall features described in those well characterized GIs: a modular organization and the conservation (among family members) of genes encoding the functions related to excision/integration, transfer and their regulation. Further research is required to assess the role of the genes encoded in this family of islands both when integrated or excised within their respective hosts. The widespread distribution of EARL GIs among pathogenic *Enterobacteriaceae* suggests that they are mobilisable elements and likely active players in bacterial pathogenesis.

## Methods

**Genomic island search and comparative analyses.** A BLASTn search<sup>55</sup> was performed using as query the sequence “attL + 82 nt” from *Salmonella enterica* ser. Enteritidis P125109 (Accession N° AM933172.1; nucleotides 2,061,160–2,061,444; 285 nt), spanning the left attachment site of ROD21 (*attL*), the integrase *SEN1970* promoter and the first 82 nucleotides of *SEN1970*. BLASTn was performed by aligning with the non-redundant sequence database. Putative excisable genomic islands were identified in the graphic view of BLASTn alignments as DNA regions flanked by two direct repeated sequences (DRS), visualized as the main BLAST hit ( $\leq 285$  bp) immediately followed by a putative integrase-encoding gene, and a second hit of approximately 30 bp located in the same orientation in a noncoding region. Bacterial strains, accession numbers and island coordinates were recorded. Sequences harboring the putative islands were downloaded from GenBank. Nucleotide comparisons of each island with ROD21 and ICEKp258.1 were performed using tBLASTx<sup>56</sup> with the standalone BLAST v2.6.0+ software run in EasyFig v2.2.2<sup>57</sup> and then visualized and analyzed using EasyFig and Artemis Comparison Tool v13.0.0<sup>58</sup>.

**Phylogenetic analyses.** Nucleotide sequences of integrases encoded in each island were downloaded from GenBank (Supplementary Information). Codon alignment was performed using MUSCLE in MEGA v7<sup>59</sup> and gapped columns were deleted (the nucleotide sequence obtained from the *Ph. luminescens* TTO1 integrase harbors nonsense mutations and was therefore excluded from the alignment). The IQ-Tree web server (<http://iqtree.cibiv.univie.ac.at/>)<sup>60</sup> was used for the selection of the best-fitting nucleotide substitution model (SYM + G), according to the corrected Akaike Information Criterion. The SYM + G model was then applied for construction of maximum likelihood (IQ-Tree web server) and Bayesian (Mr. Bayes v3.2.6)<sup>61</sup> trees, using as outgroups the integrases of SPI7 harbored by *Salmonella* ser. Typhi CT18 and of phage P4 (Supplementary Information). Two outgroups were used for the ML tree and one (SPI7 integrase) for the Bayesian tree. Node support was obtained from 100 bootstraps (ML) and posterior probabilities (Bayesian). FigTree v1.4.3 was used for tree visualization and colouring. Maximum likelihood phylogeny for H-NS homologues was constructed based on the complete alignment of their nucleotide sequences, using the TN + G model. Tree construction, visualization and colouring was performed as described for the integrase sequences.

**Bacterial strain and growth conditions.** Bacteria were maintained as a stock in LB broth supplemented with glycerol (20% v/v), or in the CRYOBANK Bead System, at  $-80^{\circ}\text{C}$ . When required, overnight cultures of *Klebsiella pneumoniae* ST258 strain KP35, enteropathogenic *Escherichia coli* O127:H6 strain E2348/69, *Salmonella* ser. Enteritidis phagotype 4 strain P125109 (NCTC 13349) or the *Salmonella* ser. Enteritidis  $\Delta$ SEN1970:FRT  $\Delta$ SEN1998:FRT mutant strain were prepared inoculating 3 mL of LB broth with an aliquot/bead of the stock culture and incubated at  $37^{\circ}\text{C}$  and 120 rpm overnight.

**Assessment of ICEKp258.2 and IE3 excision.** Island excision was assessed under three conditions ( $37^{\circ}\text{C}/\text{pH } 7.0$ ;  $23^{\circ}\text{C}/\text{pH } 7.0$  and  $37^{\circ}\text{C}/\text{pH } 5.4$ ). Three tubes containing 3 mL of LB broth (yeast extract 5 g/L, tryptone 10 g/L and NaCl 10 g/L) were inoculated with 20  $\mu\text{L}$  of an overnight culture of each strain and incubated for 16 h in a shaking incubator. One milliliter of each culture was used for extraction of genomic DNA using the phenol-chloroform technique<sup>62</sup>. Detection of GI excision and formation of the circular episomal element were assessed by nested PCR using primers listed in Table S4 (Supplementary File S1). A first round of PCR was carried out with genomic DNA at a final concentration of approximately 0.4 ng/ $\mu\text{L}$ , 0.5  $\mu\text{M}$  of each primer pair (IDT), 0.2 mM dNTP Mix, 1.5 mM  $\text{MgCl}_2$  and 1U of *Taq* DNA Polymerase (Invitrogen™) in a reaction volume of 25  $\mu\text{L}$ . A second round PCR was performed with Platinum™ *Pfx* DNA Polymerase (Invitrogen™) for *attB* and *attP* sequences, using 2  $\mu\text{L}$  of the product of the first round PCR as template, and nested primers, following manufacturer's instructions. PCR products were visualized under UV light after electrophoresis in Tris-acetate-EDTA buffer at 90 V using 1% agarose gels precast with SafeView. To confirm that the amplicons obtained during the nested PCR correspond to the expected *attB* and *attP* sequences, we purified the obtained fragments from agarose gel using the Wizard® SV Gel and PCR Clean-Up System kit (Promega) following manufacturer's instructions. The purified PCR products were sequenced by Macrogen Inc. and raw data obtained was analysed using Vector® NTI v11.0.

**Assessing the effect of ROD21 excision on gene expression.** Overnight cultures of wildtype *Salmonella* ser. Enteritidis PT4 strain P125109 and the  $\Delta$ SEN1970:FRT  $\Delta$ SEN1998:FRT isogenic strain (lacking the integrase and putative RDF coding sequences), were inoculated in LB broth at pH 7.0 until  $\text{OD}_{600} = 0.6$  was reached, and samples of 1 mL were taken and centrifuged at 8,000 rpm for 6 min. Supernatant was discarded and the bacterial pellet was resuspended in TRIzol™ reagent (Invitrogen™) and stored at  $-80^{\circ}\text{C}$ . Two independent experiments performed in duplicate were carried out.

RNA and DNA extraction was carried out as described by the manufacturer with some modifications. Briefly, after DNA precipitation and wash, 300  $\mu\text{L}$  of TE buffer and 300  $\mu\text{L}$  phenol:chloroform:isoamyl alcohol (25:24:1) (Winkler) was added and vigorously shaken, the mix was then centrifuged at 14,800 rpm for 15 min at  $4^{\circ}\text{C}$  and the aqueous phase (100  $\mu\text{L}$ ) was recovered. DNA was precipitated by adding 60  $\mu\text{L}$  of propan-2-ol and 10  $\mu\text{L}$  of 3 M sodium acetate followed by a 30 min incubation at  $-20^{\circ}\text{C}$  and centrifugation at 14,800 rpm for 15 min at  $4^{\circ}\text{C}$ . The DNA pellet was washed with 75% ice-cold ethanol and resuspended in nuclease-free water. Reverse transcription was performed with iScript™ cDNA Synthesis Kit (BioRad) according to the manufacturer's instructions.

Quantitative real-time PCR (qPCR) was performed for the quantification of ROD21 excision, using TaqMan™ probes and TaqMan™ Fast Advanced Master Mix (Applied Biosystems™) following the manufacturer's instructions for a 20  $\mu\text{L}$  reaction mixture. Standard curves for *attB*-1 and *rpoD* were used for quantification of ROD21 excision and *rpoD* expression using serial one-tenth dilutions of genomic DNA from *Salmonella*

ser. Typhimurium strain 14028s, which harbors only one copy of *attB-1* and *rpoD*. Quantitative real-time PCRs (RT-qPCRs) were carried out using specific primers and TaqMan™ MGB probes for genes *SEN1975*, *SEN1980* and *SEN1993* inside ROD21 (Table S4 in Supplementary File S1). A StepOnePlus™ thermocycler was used, employing the cycling conditions established for TaqMan™ Fast reagent. The expression of the target gene was normalized by the housekeeping gene *rpoD* and abundance of each target mRNA was determined by the comparative method ( $2^{-\Delta\Delta C_t}$ ).

**Data availability.** All data generated and analysed during this study are included in this published article and its Supplementary Information files.

## References

- Boyd, E. F., Almagro-Moreno, S. & Parent, M. A. Genomic islands are dynamic, ancient integrative elements in bacterial evolution. *Trends Microbiol.* **17**, 47–53 (2009).
- Che, D., Hasan, M. S. & Chen, B. Identifying Pathogenicity Islands in Bacterial Pathogenomics Using Computational Approaches. *Pathogens* **3**, 36–56 (2014).
- Williams, K. P. Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res.* **30**, 866–875 (2002).
- Wozniak, R. A. F. *et al.* Comparative ICE genomics: Insights into the evolution of the SXT/R391 family of ICEs. *Plos Genet.* **5**, e1000786 (2009).
- Daccord, A., Ceccarelli, D., Rodrigue, S. & Burrus, V. Comparative Analysis of Mobilizable Genomic Islands. *J. Bacteriol.* **195**, 606–614 (2013).
- Ambroset, C. *et al.* New insights into the Classification and Integration Specificity of *Streptococcus* Integrative Conjugative Elements through Extensive Genome Exploration. *Front. Microbiol.* **6**, 1483 (2016).
- Coluzzi, C. *et al.* A Glimpse into the World of Integrative and Mobilizable Elements in Streptococci Reveals an Unexpected Diversity and Novel Families of Mobilization Proteins. *Front. Microbiol.* **8**, 443 (2017).
- Hacker, J. & Kaper, J. B. Pathogenicity Islands and the Evolution of Microbes. *Annu. Rev. Microbiol.* **54**, 641–679 (2000).
- Sullivan, J. T. & Ronson, C. W. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Genetics* **95**, 5145–5149 (1998).
- Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* **2**, 414–424 (2004).
- Hudson, C. M., Lau, B. Y. & Williams, K. P. Islander: A database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic Acids Res.* **43**, D48–D53 (2015).
- Lewis, J. A. & Hatfull, G. F. Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res.* **29**, 2205–2216 (2001).
- Lesic, B. *et al.* Excision of the high-pathogenicity of *Yersinia pseudotuberculosis* requires the combined actions of its cognate integrase and Hef, a new recombination directionality factor. *Mol. Microbiol.* **52**, 1337–1348 (2004).
- Murphy, R. A. & Boyd, E. F. Three Pathogenicity Islands of *Vibrio cholerae* Can Excise from the Chromosome and Form Circular Intermediates. *J. Bacteriol.* **190**, 636–647 (2008).
- Quiroz, T. S. *et al.* Excision of an Unstable Pathogenicity Island in *Salmonella enterica* Serovar Enteritidis Is Induced during Infection of Phagocytic Cells. *Plos One* **6**, e26031 (2011).
- Sentchilo, V. *et al.* Intracellular excision and reintegration dynamics of the ICE<sub>elc</sub> genomic island of *Pseudomonas knackmussii* sp. strain B13. *Mol. Microbiol.* **72**, 1293–1306 (2009).
- Penadés, J. R. & Christie, G. E. The Phage-Inducible Chromosomal Islands: A Family of Highly Evolved Molecular Parasites. *Annu. Rev. Virol.* **2**, 181–201 (2015).
- Salazar-Echegarai, F. J., Tobar, H. E., Nieto, P. A., Riedel, C. A. & Bueno, S. M. Conjugal Transfer of the Pathogenicity Island ROD21 in *Salmonella enterica* serovar Enteritidis Depends on Environmental Conditions. *Plos One* **9**, e90626 (2014).
- Haskett, T. L. *et al.* Assembly and transfer of tripartite integrative and conjugative genetic elements. *Proc. Natl. Acad. Sci. USA* **113**, 12268–12273 (2016).
- Lovell, H. C. *et al.* In planta conditions induce genomic changes in *Pseudomonas syringae* pv. *phaseolicola*. *Mol. Plant Pathol.* **12**, 167–176 (2011).
- Vanga, B. R., Butler, R. C., Toth, I. K., Ronson, C. W. & Pitman, A. R. Inactivation of P<sub>b</sub>Topo III $\beta$  causes hyper-excision of the Pathogenicity Island HAI2 resulting in reduced virulence of *Pectobacterium atrosepticum*. *Mol. Microbiol.* **84**, 648–663 (2012).
- Carraro, N., Poulin, D. & Burrus, V. Replication and Active Partition of Integrative and Conjugative Elements (ICEs) of the SXT/R391 Family: The Line between ICEs and Conjugative Plasmids Is Getting Thinner. *Plos Genet.* **11**, e1005298 (2015).
- Almagro-Moreno, S., Napolitano, M. G. & Boyd, E. F. Excision dynamics of *Vibrio* pathogenicity island-2 from *Vibrio cholerae*: role of a recombination directionality factor VefA. *BMC Microbiol.* **10**, 306 (2010).
- Li, X. *et al.* SXT/R391 integrative and conjugative elements in *Proteus* species reveal abundant genetic diversity and multidrug resistance. *Sci. Rep.* **6**, 37372 (2016).
- Thomson, N. R. *et al.* Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res.* **18**, 1624–1637 (2008).
- Newman, R. M., Salunkhe, P., Godzik, A. & Reed, J. C. Identification and Characterization of a Novel Bacterial Virulence Factor That Shares Homology with Mammalian Toll/Interleukin-1 Receptor Family Proteins. *Infect. Immun.* **74**, 594–601 (2006).
- Hendriksen, R. S. *et al.* Global Monitoring of *Salmonella* Serovar Distribution from the World Health Organization Global Foodborne Infections Network Country Data Bank: Results of Quality Assured Laboratories from 2001 to 2007. *Foodborne Pathog. Dis.* **8**, 887–900 (2011).
- Ao, T. T. *et al.* Global Burden of Invasive Nontyphoidal *Salmonella* Disease, 2010. *Emerg. Infect. Dis.* **21**, 941–949 (2015).
- European Food Safety Authority & European Centre for Disease Prevention and Control. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2013. *EFSA J.* **13**, 3991 (2015).
- Tobar, H. E. *et al.* Chromosomal Excision of a New Pathogenicity Island Modulates *Salmonella* Virulence *In Vivo*. *Curr. Gene Ther.* **13**, 240–249 (2013).
- Pitman, A. R. *et al.* Exposure to Host Resistance Mechanisms Drives Evolution of Bacterial Virulence in Plants. *Curr. Biol.* **15**, 2230–2235 (2005).
- Godfrey, S. A. C. *et al.* The Stealth Episome: Suppression of Gene Expression on the Excised Genomic Island PPHGI-1 from *Pseudomonas syringae* pv. *phaseolicola*. *Plos Pathog.* **7**, e1002010 (2011).
- Lu, B. & Leong, H. W. Computational methods for predicting genomic islands in microbial genomes. *Comput. Struct. Biotechnol. J.* **14**, 200–206 (2016).
- Johnson, C. M. & Grossman, A. D. Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Annu. Rev. Genet.* **49**, 577–601 (2015).

35. Burrus, V., Pavlovic, G., Decaris, B. & Guédon, G. The ICES<sub>1</sub> element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. *Plasmid* **48**, 77–97 (2002).
36. Daccord, A., Ceccarelli, D. & Burrus, V. Integrating conjugative elements of the SXT/R391 family trigger the excision and drive the mobilization of a new class of *Vibrio* genomic islands. *Mol. Microbiol.* **78**, 576–588 (2010).
37. Doublet, B., Boyd, D., Mulvey, M. R. & Cloeckeaert, A. The *Salmonella* genomic island 1 is an integrative mobilizable element. *Mol. Microbiol.* **55**, 1911–1924 (2005).
38. Carraro, N., Matteau, D., Luo, P., Rodrigue, S. & Burrus, V. The Master Activator of IncA/C Conjugative Plasmids Stimulates Genomic Islands and Multidrug Resistance Dissemination. *Plos Genet.* **10**, e1004714 (2014).
39. Garcillán-Barcia, M. P., Francia, M. V. & De La Cruz, F. The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.* **33**, 657–687 (2009).
40. Ramsay, J. P. & Firth, N. Diverse mobilization strategies facilitate transfer of non-conjugative mobile genetic elements. *Curr. Opin. Microbiol.* **38**, 1–9 (2017).
41. Singh, K., Milstein, J. N. & Navarre, W. W. Xenogeneic Silencing and Its Impact on Bacterial Genomes. *Annu. Rev. Microbiol.* **70**, 199–213 (2016).
42. Carpenter, M. R., Rozovsky, S. & Boyd, E. F. Pathogenicity island cross talk mediated by recombination directionality factors facilitates excision from the chromosome. *J. Bacteriol.* **198**, 766–776 (2016).
43. Nieto, P. A. *et al.* New insights about excisable pathogenicity islands in *Salmonella* and their contribution to virulence. *Microbes Infect.* **18**, 302–309 (2016).
44. Ali, S. S., Xia, B., Liu, J. & Navarre, W. W. Silencing of foreign DNA in bacteria. *Curr. Opin. Microbiol.* **15**, 175–181 (2012).
45. Kotlajich, M. V. *et al.* Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria. *Elife* **4**, e04970 (2015).
46. Lucchini, S. *et al.* H-NS mediates the silencing of laterally acquired genes in bacteria. *Plos Pathog.* **2**, e81 (2006).
47. Higashi, K. *et al.* H-NS Facilitates Sequence Diversification of Horizontally Transferred DNAs during Their Integration in Host Chromosomes. *Plos Genet.* **12**, 1–31 (2016).
48. Müller, C. M. *et al.* Differential effects and interactions of endogenous and horizontally acquired H-NS-like proteins in pathogenic *Escherichia coli*. *Mol. Microbiol.* **75**, 280–293 (2010).
49. Williamson, H. S. & Free, A. A truncated H-NS-like protein from enteropathogenic *Escherichia coli* acts as an H-NS antagonist. *Mol. Microbiol.* **55**, 808–827 (2005).
50. Cirl, C. *et al.* Subversion of Toll-like receptor signaling by a unique family of bacterial Toll/interleukin-1 receptor domain-containing proteins. *Nat. Med.* **14**, 399–406 (2008).
51. Chen, L., Mathema, B., Pitout, J. D. D., DeLeo, F. R. & Kreiswirth, B. N. Epidemic *Klebsiella pneumoniae* ST258 Is a Hybrid Strain. *MBio* **5**, e01355–14 (2014).
52. Ahn, D. *et al.* Acquired resistance to innate immune clearance promotes *Klebsiella pneumoniae* ST258 pulmonary infection. *JCI Insight* **1**, e89704 (2016).
53. Martínez-Rubio, R. *et al.* Phage-inducible islands in the Gram-positive cocci. *ISME J.* **11**, 1029–1042 (2017).
54. Liu, J. *et al.* Staphylococcal chromosomal cassettes mec (SCCmec): A mobile genetic element in methicillin-resistant *Staphylococcus aureus*. *Microb. Pathog.* **101**, 56–67 (2016).
55. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A Greedy Algorithm for Aligning DNA Sequences. *J. Comput. Biol.* **7**, 203–214 (2000).
56. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
57. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: A genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).
58. Carver, T. J. *et al.* ACT: The Artemis comparison tool. *Bioinformatics* **21**, 3422–3423 (2005).
59. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
60. Trifinopoulos, J., Nguyen, L., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235 (2016).
61. Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst. Biol.* **61**, 539–542 (2012).
62. Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 2001).
63. Deng, W. *et al.* Comparative Genomics of *Salmonella enterica* Serovar Typhi Strains Ty2 and and CT18. *J. Bacteriol.* **185**, 2330–2337 (2003).
64. Iguchi, A. *et al.* Complete Genome Sequence and Comparative Genome Analysis of Enteropathogenic *Escherichia coli* O127:H6 Strain E2348/69. *J. Bacteriol.* **191**, 347–354 (2009).
65. Porwollik, S. *et al.* Differences in Gene Content between *Salmonella enterica* Serovar Enteritidis Isolates and Comparison to Closely Related Serovars Gallinarum and Dublin. *J. Bacteriol.* **187**, 6545–6555 (2005).
66. DeLeo, F. R. *et al.* Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 *Klebsiella pneumoniae*. *Proc. Natl. Acad. Sci. USA* **111**, 4988–4993 (2014).
67. Bell, K. S. *et al.* Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. *Proc. Natl. Acad. Sci. USA* **101**, 11105–11110 (2004).

## Acknowledgements

We are grateful to Dr. Alice Prince from Columbia University Medical Center, New York, United States, Dr. Carlos Santiviago from Universidad de Chile and Dr. Roberto Vidal from Universidad de Chile, Santiago, Chile, for kindly providing bacterial strains *K. pneumoniae* KP35, *Salmonella* ser. Enteritidis P125109 and EPEC E2348/69, respectively. Authors of this study were supported by Fondo Nacional de Investigación Científica y Tecnológica de Chile (FONDECYT grants 1170964, 1140011 and 1140010), the Collaborative Research Program-ICGEB (grant CRP/CHI14-01) and the Millennium Institute on Immunology and Immunotherapy P09/016-F. A.P.I., C.P.R. and I.C.A. are also supported by Beca de Doctorado Nacional CONICYT (21172030, 21140169 and 63140215).

## Author Contributions

A.P.I., D.U.A. and S.B. conceived the study; A.P.I., D.U.A., C.P.R., B.S. and F.J.S.-E. designed the methodology; A.P.I. and D.U.A. collected the data, A.P.I., D.U.A., C.P.R. and F.J.S.-E. analysed the sequences, A.P.I., D.U.A., C.P.R. and I.C.A. performed the experiments and made the figures and tables; P.G. and S.B. acquired funding and supervised the research. All authors contributed in manuscript writing and revision, and approved the final version.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-28537-0>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018