CHARACTERIZATION OF THE *Enterobacteriaceae*-ASSOCIATED ROD21-LIKE FAMILY OF EXCISABLE GENOMIC ISLANDS AND STUDY OF EXCISION IN THE PATHOGENICITY ISLAND ROD21

Tesis entregada a la Pontificia Universidad Católica de Chile en cumplimiento parcial de los requisitos para optar al Grado de Doctor en Ciencias con mención en Genética Molecular y Microbiología por:

LUIS ALEJANDRO PIÑA ITURBE

Directora de Tesis: Susan M. Bueno, PhD

Codirector de Tesis: Pablo A. González, PhD

Enero de 2021

ACTA DE APROBACIÓN DE COMITÉ DE LA TESIS

DEDICATION

A Dios, porque

pedí y recibí, busqué y hallé, llamé a la Puerta y ésta me fue abierta.

A María Elena, Víctor, André,

Liliana, Carlos, Luis Alberto, Sonia, José Luis, Soledad y Hernan Luis,

porque por su amor soy quien soy.

A Johanna,

porque para ella soy.

ACKNOWLEDGEMENTS

A Dios, por el sustento, guía y ayuda en las decisiones y desarrollo de esta tesis.

A la Dra. Susan Bueno y Dr. Pablo González por la guía directa, análisis crítico y apoyo proporcionados a lo largo del desarrollo de esta Tesis. Y también al Dr. Carlos Santiviago, Dra. Beatriz Díez, Dr. José O'Brien y Dr. Rodrigo De La Iglesia, por su revisión crítica y orientación, proporcionados durante las distintas etapas de seguimiento de este trabajo.

A los miembros actuales y pasados del Laboratorio de Patogénesis Microbiana de la Pontificia Universidad Católica de Chile, especialmente a Irenice Coronado, Catalina Pardo, y Yaneisi Vásquez, y a Paulina Fernández del Laboratorio de Microbiología de la Universidad de Chile, por sus enseñanzas sobre las bases de las distintas técnicas experimentales empleadas en esta Tesis, incluyendo, pero no limitándose a, la amplificación, extracción y purificación de ácidos nucleicos, PCR y qPCR, extracción y purificación de proteínas, técnicas de detección y cuantificación de ácidos nucleicos y proteínas, y ensayos de unión proteína-ADN.

A Diego Ulloa y Guillermo Hoppe por su apoyo técnico en el desarrollo de los distintos análisis bioinformáticos y trabajo experimental.

A la Vicerrectoría de Investigación y Colegio de Programas Doctorales (ahora Escuela de Graduados) de la Pontificia Universidad Católica de Chile, al Instituto Milenio en Inmunología e Inmunoterapia, y al Programa de Formación de Capital Humano avanzado de la Comisión Nacional de Investigación Científica y Tecnológica de Chile (ahora Agencia Nacional de Investigación y Desarrollo) por el financiamiento recibido a través de la Beca VRI-CPD de Ayudante Becario, Beca de Tesista Doctoral IMII y Beca de Doctorado Nacional 21172030.

INDEX

INDEX OF FIGURES

SECTION 2: ROD21 belongs to a family of excisable genomic islands found in *Enterobacteriaceae*

<u>SECTION 3</u>: The EARL genomic islands are highly AT-rich islands that carry homologs of the xenogeneic silencer H-NS

<u>SECTION 4</u>: The excision of ROD21 and expression of its recombination directionality factor are negatively correlated and are modulated by the growth phase

INDEX OF TABLES

ABBREVIATIONS

AMP:	ampicillin
ANOVA:	analysis of variance
AT content:	adenine plus thymine content
CAP:	"*Candidatus* Accumulibacter phosphatis"
cat:	chloramphenicol acetyl transferase encoding gene
cDNA:	complementary DNA
ChAP-Seq:	chromatin affinity purification coupled to sequencing
CHL:	chloramphenicol
Ctd:	C-terminal domain
DNA:	deoxyribonucleic acid
DNase:	deoxyribonuclease
DRS:	direct-repeated sequence
EARL:	*Enterobacteriaceae*-associated ROD21-like
EBPR:	enhanced biological phosphorus removal
EDTA:	ethylenediaminetetraacetic acid
EHEC:	enterohemorrhagic *Escherichia coli*
EMSA:	electrophoretic mobility shift assay
EPEC:	enteropathogenic *Escherichia coli*
FC:	fold change
FRT:	Flippase recognition target
GC content:	guanine and cytosine content
GI:	genomic island
HGT:	horizontal gene transfer

ICE:	integrative and conjugative element
IMAC:	immobilized metal affinity chromatography
IME:	integrative and mobilizable element
IPTG:	isopropyl β-D-1-thiogalactopyranoside
LEE:	Locus of Enterocyte Effacement
mRNA:	messenger RNA
MWCO:	molecular weight cut-off
NAP:	nucleoid-associated protein
ORF:	open reading frame
PCI:	phenol:chloroform:isoamyl alcohol
PICI:	Phage-inducible chromosomal island
RDF:	recombination directionality factor
RNA:	ribonucleic acid
RNase:	ribonuclease
ROD21:	Region of Difference 21
RT-qPCR:	reverse transcription quantitative polymerase chain reaction
SD:	standard deviation
SPI:	*Salmonella* Pathogenicity Island
ST:	sequence type
T3SS:	type-III secretion system
Tcp:	TIR-domain containing protein
TIR domain:	Toll/interleukin-1 receptor/resistance domain
TLR:	Toll-like receptor
tmRNA:	transfer-messenger RNA
tRNA:	transfer RNA
UPEC:	uropathogenic *Escherichia coli*
VPI:	*Vibrio* Pathogenicity Island
VSP:	*Vibrio* Seventh Pandemic Island
XS:	xenogeneic silencer

RESUMEN

Las islas genómicas (IGs) son elementos adquiridos horizontalmente que juegan un papel importante en la evolución del genoma y la adaptación bacteriana a diferentes ambientes, ya que estos elementos transfieren genes que codifican diferentes funciones ventajosas para la bacteria, por ejemplo: la capacidad de metabolizar nuevos sustratos, la resistencia a los antimicrobianos o la virulencia, entre otras. Además de estas funciones, muchas IGs también contienen un módulo de escisión/integración que incluye los genes que codifican la maquinaria molecular (la enzima integrasa y el factor de direccionalidad de recombinación) y las regiones reguladoras necesarias que promueven una reacción de recombinación específica de sitio. Esta reacción da como resultado la escisión de la isla del cromosoma como un elemento circular que puede ser el sustrato para la transferencia por conjugación o puede reintegrarse. De manera interesante, se ha encontrado evidencia que el estado escindido/integrado de algunas IGs tiene un rol en la virulencia de las bacterias en su hospedador, aunque por ahora, este fenómeno se ha observado solamente en dos IGs. La creciente disponibilidad de genomas bacterianos secuenciados ha permitido la identificación y caracterización de varias IGs las cuales han sido clasificadas en diferentes familias que comparten un set conservado de genes implicados en los procesos de transferencia y escisión, que comparten vías reguladoras comunes. La identificación y clasificación de las IGs permite comprender mejor su evolución, sus interacciones con su hospedador y los mecanismos reguladores que rigen su transferencia a otras bacterias. Sin embargo, muchas IGs aún no han sido clasificadas y la comprensión de su biología sigue siendo limitada. La Región de Diferencia 21 (ROD21, por sus siglas en inglés) es una isla de

patogenicidad escindible y movilizable que codifica algunos factores de virulencia involucrados en la infección producida por *Salmonella enterica* serovar Enteritidis, un patógeno transmitido por alimentos y una de las serovariedades de *Salmonella* de mayor prevalencia en el mundo. Los experimentos realizados con cepas mutantes de *Salmonella* ser. Enteritidis que tienen una capacidad reducida o alterada para escindir ROD2, mostraron un defecto en estas cepas para invadir órganos internos en el modelo murino de infección. Un resultado que puede estar relacionado con los cambios observados en la expresión de los genes en ROD21 y en otros loci genómicos. Sin embargo, aunque la escisión de ROD21 tiene un rol en la virulencia de *Salmonella* ser. Enteritidis, se conoce muy poco acerca de la regulación de este proceso. ROD21 también está presente en los serovares Gallinarum y Dublin, y aunque se ha reportado la existencia de "islas relacionadas con ROD21", no se conoce más acerca de este tema. En esta Tesis Doctoral, se utilizaron diferentes herramientas bioinformáticas para identificar islas relacionadas con ROD21 y analizar su filogenia y contenido genético. Estos análisis revelaron que ROD21 pertenece a una familia de islas genómicas presentes en diferentes cepas de bacterias patógenas de plantas y animales del orden Enterobacterales, la cual fue denominada *Enterobacteriaceae*-associated ROD21-like (EARL). El análisis filogenético sugiere que la familia EARL se originó en bacterias patógenas de plantas y que la adquisición de genes específicos permitió su diseminación a diversos hospedadores bacterianos. Se encontró también que, si bien estas islas codifican grupos de genes variables y conservados, el módulo de escisión/integración fue la única región conservada en las 56 islas identificadas, incluida ROD21, hallazgo que destaca la naturaleza escindible de la familia EARL como se demostró para dos de sus miembros albergados por *Klebsiella pneumoniae* ST258 resistente a carbapenémicos cepa KP35 y *Escherichia coli* enteropatógena O127:H6 cepa E2348/69. Cabe

destacar que las islas EARL poseen genes que codifican homólogos cortos y de longitud completa del regulador transcripcional global H-NS, una proteína involucrada en la regulación de funciones bacterianas importantes como la virulencia y la transferencia horizontal de genes. Como parte del módulo de integración /escisión conservado, se caracterizaron el gen *SEN1998* y su producto proteico purificado para evaluar su rol en la escisión de ROD21. Si bien *SEN1998* codifica un factor de direccionalidad de recombinación capaz de unirse a las regiones de ADN involucradas en el proceso de escisión, la deleción de este gen solo produjo una ligera reducción en el nivel de escisión. Sorpresivamente, los niveles de expresión de *SEN1998* resultaron estar correlacionados negativamente con la escisión de ROD21, una característica que también se observó para el gen *SEN1970* que codifica la integrasa. A pesar de estos hallazgos, no está claro si la represión de *SEN1998* y *SEN1970*, la cual ocurrió inmediatamente antes que comenzara a aumentar la escisión, está relacionada a la regulación de este proceso. En resumen, los resultados de esta Tesis amplían el repertorio de islas genómicas escindibles identificadas, presenta información sobre su evolución y proporciona un primer acercamiento a la comprensión de la regulación de la escisión en la isla de patogenicidad ROD21. Estos hallazgos podrían contribuir a una mejor comprensión de los patógenos actuales y emergentes como lo son las diferentes serovariedades de *Salmonella*, *E. coli* patógena y *K. pneumoniae* ST258 resistente a carbapenémicos, entre otros.

ABSTRACT

Genomic islands (GI) are horizontally acquired elements that play a major role in genome evolution and bacterial adaptation to different environments because these elements transfer genes encoding different advantageous functions such as the metabolism of new substrates, antimicrobial resistance, or virulence, among others. In addition to those functions, many GIs also contain an excision/integration module comprising the genes encoding the molecular machinery (the integrase enzyme and the recombination directionality factor) and the regulatory regions required for a site-specific recombination reaction. This reaction results in the excision of the island from the chromosome as a circular element, which can be the substrate for conjugal transfer or can be reintegrated. Interestingly, there is evidence that the excised/integrated state of some GIs plays a role in the virulence of the bacteria in the host, although more research is required as this phenomenon has been observed only for two GIs. The increasing availability of sequenced bacterial genomes has allowed the identification and characterization of several GIs classified in different island families with a conserved core of genes involved in the transfer and excision processes, which share common regulatory pathways. The identification and classification of GIs allows a better understanding of their evolution, their interactions with their host, and the regulatory mechanisms that govern their dissemination. Nevertheless, many GIs remain unclassified and the understanding of their biology remains limited. The Region of Difference 21 (ROD21) is an excisable and mobilizable pathogenicity

island that encodes virulence factors involved in the infection produced by *Salmonella* serovar Enteritidis, a foodborne pathogen and one of the most prevalent serovars of *Salmonella* worldwide. Experiments carried out with mutant strains of *Salmonella* ser. Enteritidis with a reduced or impaired capacity to excise ROD21 showed a defect in these strains to invade internal organs in the mouse model, an outcome that may be related to the observed changes in gene expression within ROD21 and other genomic loci. However, although the excision of ROD21 plays a role in the virulence of *Salmonella* ser. Enteritidis, the regulation of this process remains poorly understood. The pathogenicity island ROD21 is also present in the serovars Gallinarum and Dublin, and although the existence of "ROD21-related islands" was previously reported, there is a lack of knowledge about the relationships between ROD21 and other GIs. In this thesis, different bioinformatic tools were used to search and identify ROD21-related islands, and analyze their phylogenetic relationships and gene content. These analyses revealed that ROD21 belongs to a family of genomic islands carried by different strains of plant- and animal pathogenic bacteria from the order Enterobacterales denominated the *Enterobacteriaceae*-associated ROD21-like (EARL) family of GIs. The phylogenetic analysis suggests that the EARL family originated in plant-pathogenic bacteria and that the acquisition of specific genes allowed their dissemination to diverse bacterial hosts. Interestingly, while these islands carry variable and conserved sets of genes, the excision/integration module was the only region conserved across the 56 identified islands including ROD21, highlighting the excisable nature of the EARL family as demonstrated for two of its members harbored by carbapenem-resistant *Klebsiella pneumoniae* ST258 KP35 and enteropathogenic *Escherichia coli* O127:H6 E2348/69. Noteworthy, the EARL GIs are carriers of genes encoding short and full-length homologs of the global transcriptional regulator H-NS, a protein involved in regulation of important bacterial

functions such as virulence and facilitation of horizontal gene transfer. As part of the conserved integration/excision module, the gene *SEN1998* and its purified protein product were characterized to assess its role in the excision of ROD21. While *SEN1998* encodes a recombination directionality factor with DNA-binding affinity for the DNA regions involved in the excision process, deletion of this gene produced only a slight reduction in the excision level. Unexpectedly, the expression levels of *SEN1998* resulted to be negatively correlated to the excision of ROD21, a feature that was also observed for the integrase-encoding gene *SEN1970*. However, it remains unclear whether the repression of *SEN1998* and *SEN1970*, which occurred immediately before the excision started to increase, plays a role in regulation of this process. In summary, the results of this thesis expand the repertory of identified excisable genomic islands, presenting insights about their evolution, and provide a first approach to the understanding of the regulation of excision in the pathogenicity island ROD21. These findings may contribute to the better understanding of current and emerging pathogens such as *Salmonella* serovars, pathogenic *E. coli* and carbapenem-resistant *K. pneumoniae* ST258, among others.

1. INTRODUCTION

The next generation sequencing, bioinformatics and the consequent availability of a greater number of bacterial genomes have allowed comparative studies at different taxonomic levels, including the level of subspecies and even strains. These analyses showed the great extension of the acquisition and transfer of discrete DNA regions in a process known as horizontal gene transfer, highlighting the major role that this process plays in genome evolution and bacterial adaptation.

1.1. Horizontal gene transfer

Horizontal gene transfer (HGT) refers to the transmission of DNA from one cell to another besides inheritance from parent cells to their offspring (Soucy et al., 2015). Recent studies addressing the extent of HGT in Bacteria and Archaea showed that, among 2472 completely sequenced genomes, between 10 to 35% of all genes in most prokaryotic phyla participated in HGT (Jeong et al., 2016). Together with the finding that the larger the genomes, the higher the number of genes that participated in HGT (Jeong et al., 2016), this finding highlights the major role of HGT in shaping prokaryotic genomes and also in microbial adaptation, since a substantial number of genes encoding advantageous traits can be acquired by HGT. Indeed, it is currently known that several genes participating in niche colonization, symbiosis, catabolism of different substrates, antimicrobial resistance and virulence, were

acquired by HGT (Carraro et al., 2016; Curraize et al., 2020; Fernández-Alarcón et al., 2011; Maeda et al., 2003; Montero et al., 2017; Sullivan and Ronson, 1998).

Different mechanisms can mediate the transfer of foreign DNA, the first step for the acquisition of the possibly advantageous traits encoded on it. These mechanisms can be: transformation, transduction and conjugation. In natural transformation, the foreign naked DNA, likely from a death donor, is taken by the recipient cell directly from the environment (Johnston et al., 2014). In transduction, infective viral particles that have packaged portions of the DNA from their host, are released to infect a new host, introducing their own and/or the other cell's DNA (Chiang et al., 2019). Finally, in conjugation, a direct contact between donor and recipient cells is required for the transfer of the DNA through specialized molecular channels that connect the cytoplasm of both bacteria (Cabezón et al., 2015). Although these three mechanisms are the most studied, there are other means for HGT, for example, through membrane vesicles which can contain different molecules, including DNA (Domingues and Nielsen, 2017); or through gene transfer agents, which are viral-like particles that contain fragments of their host DNA but, in contrast to what happens in transduction, the gene transfer agents cannot encapside their own DNA (Lang et al., 2012).

Different mobile genetic elements can be substrates for HGT. Among the most studied are plasmids, bacteriophages and genomic islands. Plasmids are self-replicating extrachromosomal elements whose copies are produced in parallel with, but uncoupled from, bacterial chromosome replication. Specialized partitioning systems are responsible for the distribution of plasmid copies in daughter cells after cell division (Shintani et al., 2015). Bacteriophages are viruses that infect bacteria and hijack their replication and transcription

machinery to produce new infective particles. This lytic process can be switched to a lysogenic one, in which the phage genome integrates into the bacterial chromosome and is replicated once in each chromosomal replication cycle. Certain stimuli can activate the viral genome, inducing its excision and resuming the lytic cycle (Frost et al., 2005). Finally, genomic islands (GIs) are genetic elements inserted in the bacterial chromosome that share features with plasmids and prophages (Almagro-Moreno et al., 2010; Lesic et al., 2004; Murphy and Boyd, 2008; Quiroz et al., 2011; Salazar-Echegarai et al., 2014). As genomic islands are the central study subject in this thesis, a description of their main features will be presented below.

1.1.1. Genomic islands, genomic island excision and recombination directionality factors.

GIs are horizontally acquired elements usually found integrated at the 3'-end of genes encoding tRNAs and tmRNAs (Marcoleta et al., 2016; Williams, 2002), although the 3'- or 5'-end of other genes can also be used as insertion sites for different island families (Bellanger et al., 2014; Coluzzi et al., 2017; Daccord et al., 2013). Given its foreign origin, GIs can be recognized for having sequence signatures that differ from the average of their host genome, such as different G+C content, codon usage bias and dinucleotide frequencies (Che et al., 2014). With sizes ranging between 10 to 500 kbp (Hacker and Kaper, 2000; Sullivan and Ronson, 1998), GIs can encode proteins related to different advantageous functions such as antibiotic resistance and virulence factors, as well as proteins involved in the integration and excision of the island from the chromosome, and horizontal transfer (Daccord et al., 2012; Montero et al., 2017; Nepal et al., 2018). Similar to bacteriophages, GIs can encode integrases and recombination directionality factors (RDFs), the island-encoded proteins that promote excision (Haskett et al., 2018; Lesic et al., 2004; Ramsay et al., 2006). Integrases are enzymes that

catalyze a site-specific recombination reaction between the direct-repeated sequences located at the ends of the island in the left and right attachment regions (*attL* and *attR*) (Landy, 2015). This reaction produces a circular form of the island, and the regeneration of the GI attachment region (*attP*) and the chromosomal insertion site (*attB*) (Carpenter et al., 2016; Salazar-Echegarai et al., 2014). While integrases have the enzymatic activity, the RDFs of GIs are believed to play structural roles by binding at least one *att* region of the island to promote the assembly of the nucleoprotein complex required for excision, as observed for prophages (Abbani et al., 2007). After excision, the circular extrachromosomal state of genomic islands can be transferred to a new host using the molecular machinery encoded by the GI, or by the host or other mobile elements if the island does not encode a complete type-4 secretion system (Haskett et al., 2016; Ramsay and Firth, 2017; Salazar-Echegarai et al., 2014; Waldor, 2010).

RDFs, small proteins of approximately 50-180 amino acids (Lewis and Hatfull, 2001), not only play a role in island excision but these can also function as transcriptional repressors of integrases. This was first observed for the prophages P4 and KplE which encode the RDFs Vis and TorI. Since in both phages the promoter region of the integrase-encoding gene is located immediately downstream of one of the *att* regions, the RDF binding site overlaps with the core promoter, negatively affecting the expression of the integrase (Panis et al., 2010; Piazolla et al., 2006). For GIs, it was observed that deletion of the three genes encoding the VefA, VefB and VefC RDFs, encoded by the VPI-2 and VSP-II pathogenicity islands of *Vibrio cholerae* serotype O1 biovar El Tor, produces a significant increase in the mRNA levels of the integrases encoded by the VPI-1 and VPI-2 pathogenicity islands (Carpenter et al., 2016). There is an important difference between phage- and GI-encoded RDFs: while those encoded by phages are essential for prophage excision, deletion of the GI-encoded RDF does not inhibit island excision. The

reasons for this difference is not known and, since RDFs are difficult to identify for their small size, the GI-encoded RDFs have received little attention and only a few have been described and studied (Carpenter et al., 2016; Haskett et al., 2018; Lesic et al., 2004; Lewis and Hatfull, 2001; Poulin-Laprade et al., 2015).

1.2. Xenogeneic silencers

While HGT allows the gaining of beneficial traits, the acquisition of new genes which are not part of the regulatory network in the host cell can impose a metabolic cost, with deleterious consequences, as a result of the uncontrolled expression of the newly acquired genes. Therefore, different bacterial taxa encode in their genomes a functionally related family of proteins involved in the silencing of foreign gene expression.

The nucleoid-associated proteins (NAPs) are a group of proteins encoded by the core genome, which play major roles in the supercoiling of the bacterial nucleoid, by bending and bridging the chromosome, and also in DNA replication and gene expression, by reorganizing the nucleoid to allow the entrance of the DNA and RNA polymerases. Included in the NAP superfamily, a group of functionally and structurally, but not phylogenetically, related proteins have been identified to participate in the acquisition of foreign DNA by HGT, and in the integration of the newly acquired genes in the host transcriptional network. These NAPs are known as xenogeneic silencing (XS) proteins and belong to four families identified so far: the H-NS and MvaT families in Gram-negative bacteria (Alpha-, Beta- and Gamma-proteobacteria; only Gamma-proteobacteria for the MvaT family), and the Lsr2 and Rok families in Gram-positive bacteria (Actinobacteria and *Bacillus*, respectively).

The XS proteins facilitate the acquisition of foreign DNA by suppressing its expression following the HGT event, thus decreasing the fitness impact on the cell that results from the unrestrained transcription of the acquired genes. To target the horizontally acquired DNA, the XS proteins have higher binding affinity for DNA sequences with an AT-content above the average of the host genome, a characteristic usually found in horizontally acquired elements. Through their C-terminal DNA-binding domain, the XS proteins enter the deeper and narrower minor groove of AT-rich DNA which, based on analysis of the H-NS protein, is thought to provide favorable electrostatic interactions for binding of the XS, especially in sequences that contain a TpA step. After binding of high affinity AT-rich sequences, additional subunits of the XS can bind the DNA through cooperative binding, forming oligomers as a result of the interaction of their N-terminal oligomerization domains. As a consequence of the dimerization and oligomerization capacity of XSs, bridges can be formed within and between different regions of the bacterial chromosome, producing nucleoprotein structures that can interfere with the function of the RNA polymerase. Since the binding of the XSs proteins to DNA can be relieved by changes in temperature, pH or growth phase, the silencing of horizontally acquired DNA by the XSs allow the expression of the transferred genes in response to changes in the environment that surrounds the bacterial cell, facilitating their integration in the transcriptional network under conditions in which the expression of the foreign DNA could be beneficial.

In addition to the XS encoded by the core genome, homolog proteins encoded by plasmids, phages and genomic islands have also been reported and studied. These horizontally acquired homologs belongs to the four families of XSs and share the main structural and functional features observed in those encoded by the core genome: the binding affinity for AT-rich sequences and the dimerization/oligomerization capacity. Therefore, the horizontally

acquired XSs can form homodimers and repress the expression of genes in AT-rich regions. Moreover, some XS homologs encoded in mobile elements can physically interact with the XSs to form heterodimers and possibly heterooligomers. As a consequence, the horizontally acquired XSs can bind and modulate the expression of a subset of the XS regulon; although, in some instances, the contribution of the XS homolog to the expression of some target genes is only observed after deletion of the XS encoded in the core genome. Nevertheless, since temperature changes and growth phase can modulate the expression of some horizontally acquired homologs, these proteins may play important roles in regulation of virulence factors and horizontally acquired genes under certain environmental conditions found during the bacterial life cycles. Additionally, some horizontally acquired XSs also participate in reducing the fitness burden produced after the acquisition of AT-rich mobile elements by preventing the sequestering of the core-genome encoded XS to the horizontally acquired element.

1.3. *Salmonella enterica* serovar Enteritidis and the pathogenicity island ROD21

Salmonella enterica serovar Enteritidis (*Salmonella* ser. Enteritidis) is a Gram-negative bacterium and one of the most prevalent serovars of *Salmonella* worldwide. When ingested in contaminated food (e.g., contaminated eggs or poultry), this pathogen can produce a self-limited gastroenteritis. However, in susceptible individuals such as children under 5 years old, immunocompromised patients and the elderly, this bacterium can produce a severe invasive disease which can lead to death. The number of cases of infection produced by *Salmonella* ser. Enteritidis increased worldwide since the decade of 1980 (Rodrigue et al., 1990) and different measures and practices were implemented and promoted by countries to prevent the release of contaminated food to decrease the number of cases. However, despite these efforts, new

outbreaks are still occurring globally, such as the ongoing multi-country outbreak that currently involves 18 European and started in 2017 (ECDC and EFSA, 2020). To colonize eggs, poultry, the human host and produce disease, *Salmonella* ser. Enteritidis encodes different virulence factors in its genome, many of which are carried by horizontally acquired elements such as pathogenicity islands.

The Region of Difference 21 (ROD21) is a genomic island harbored by *Salmonella* serovars Enteritidis, Gallinarum (including the biovar Pullorum) and Dublin. This 26.5 kbp island is found integrated at the 3'-end of an Asn-tRNA-encoding gene and carry 30 putative coding sequences (*SEN1970* to *SEN1999*), most of which are predicted to encode proteins of unknown function. Among these genes, ROD21 encode virulence factors involved in colonization of the liver and spleen of mice, such as *SEN1976*, encoding a putative type-IV prepilin (Silva et al., 2012); and also in evasion of the immune system, such as *SEN1975* which encodes TlpA, a protein that mimics eukaryotic TIR-domain-containing proteins and interferes with Toll-like receptor signaling, activation of NF-κB and production of IL-1β (Newman et al., 2006; Xiong et al., 2019). In addition to the virulence factors, ROD21 harbor genes involved in the excision of the island from the chromosome of *Salmonella* ser. Enteritidis: an integrase protein encoded in *SEN1970* and a putative RDF encoded in *SEN1998*. Previous research have shown that deleting *SEN1970* or *SEN1970* and *SEN1998* significantly reduces the excision frequency of ROD21 within the bacterial population (Pardo-Roa et al., 2019; Tobar et al., 2013). Nevertheless, a detailed study of the role played by these proteins in the excision of ROD21 is lacking.

Research from our laboratory has found a link between the excision of ROD21 and the invasion capacity of *Salmonella* ser. Enteritidis. Mutant strains of this bacterium, with a reduced or impaired capacity to excise ROD21, show a reduced capacity to colonize the liver and spleen of infected mice as compared to the wild-type strain. This phenomenon could be a consequence of the altered expression observed in some genes carried by ROD21, such as *SEN1975* and *SEN1976*, and by the SPI-1 pathogenicity island, such as *invA*, which encodes a structural component of the type-3 secretion system 1 required for invasion of the intestinal epithelium (Pardo-Roa et al., 2019; Tobar et al., 2013). These findings suggest that the integrated/excised state of ROD21 modulates the expression of the genes it encodes, likely as a result of the supercoiling of the island excised form (Neale et al., 2018; Nieto et al., 2016).

Although ROD21 and its excision plays a role in the pathogenicity of *Salmonella* ser. Enteritidis, most of the genes encoded in this island remain without assigned function and there is a lack of knowledge regarding the factors involved in the excision process and its regulation. When ROD21 was discovered, it was also reported the existence of other four genomic islands, found in *Enterobacteriaceae*, that share some genes with ROD21, including homologs of the global regulator and xenogeneic silencer H-NS (Thomson et al., 2008). More recently, other research provided preliminary evidence of the involvement of *SEN1998* in the excision of ROD21, since the overexpression of this gene from an inducible plasmid increased the proportion of bacteria with and empty insertion site of ROD21 (Salazar-Echegarai et al., 2014; Tobar et al., 2013). The ever-increasing number of sequenced bacterial genomes provides the opportunity, through comparative analysis, to gain insight about the relationship of ROD21 with other genomic islands and to identify important conserved genes that may play a role in horizontal transfer, virulence and the excision process.

Based on the presented background and preliminary observations, the following hypotheses were proposed:

a. ROD21 belongs to a family of excisable genomic islands present in different bacterial species.
b. The ROD21-related islands are carriers of H-NS homologs within *Enterobacteriaceae*.
c. The protein encoded in *SEN1998* is a recombination directionality factor which modulates the excision of ROD21 by regulating the expression of the integrase SEN1970.

These hypotheses were addressed, and the resulting data, interpretation and conclusions are presented in each of the following Sections in the form of published articles (Piña-Iturbe et al., 2018, 2020) and one manuscript in preparation.

2. ROD21 BELONGS TO A FAMILY OF EXCISABLE GENOMIC ISLANDS FOUND IN *Enterobacteriaceae*

[*Scientific Reports*. 2018;8(1):10292.]

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

Alejandro Piña-Iturbe[1], Diego Ulloa-Allendes[1], Catalina Pardo-Roa[1], Irenice Coronado-Arrázola[1], Francisco Salazar-Echegarai[1], Bianca Sclavi[2], Pablo A. González[1], Susan M. Bueno[1]*

[1]Millennium Institute on Immunology and Immunotherapy, Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile.

[2]Laboratoire de Biologie et Pharmacologie Appliquée, Centre National de la Recherche Scientifique UMR 8113, École Normale Supérieure Paris-Saclay, Cachan, France.

*Correspondence to: sbueno@bio.puc.cl, Tel: +56-2-23542841.

Abstract

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 this GIs are excisable and may play a role in bacterial interactions with their hosts.

Introduction

Genomic Islands (GIs) are horizontally transferred DNA segments integrated into bacterial chromosomes (Boyd et al., 2009). They are characterized by a G+C content, codon usage bias, dinucleotide frequencies, among other sequence signatures, which usually differs from those of the genome (Che et al., 2014). Many of them are found integrated at the 3'-end of tRNA and tmRNA genes (Williams, 2002), although different families of GIs can show preference for other genes as integration sites (Wozniak et al., 2009; Daccord et al., 2013; Ambroset et al., 2016; Coluzzi et al., 2017). GIs range in size from approximately 10 to 500 kbp (Sullivan and Ronson, 1998; Hacker and Kaper, 2000) 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 (Dobrindt et al., 2004). Many GIs contain an excision/integration module that includes an integrase gene, usually belonging to the tyrosine recombinase family (Boyd et al., 2009; Hudson et al., 2015) and a Recombination Directionality Factor (RDF), which is a small protein of approximately 60-180 amino acids (Lewis and Hatfull, 2001). 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 DRS (*attP*), while another DRS (*attB*) remains in the host DNA (Lesic et al., 2004; Murphy and Boyd, 2008; Quiroz et al., 2011; Daccord et al., 2013). 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 genome (Sentchilo et al., 2009). 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 (Penadés and Christie, 2015), or transferred by conjugation (Salazar-Echegarai et al., 2014; Haskett et al., 2016). There is evidence that some GIs are replicative in their circular form (Lovell et al., 2011; Vanga et al., 2012; Carraro et al., 2015) and that others lack this feature (Almagro-Moreno et al., 2010).

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 (Daccord et al., 2013) and the SXT/R391 family of integrative and conjugative GIs (Wozniak et al., 2009) present in different Gram-negative bacterial families (Wozniak et al., 2009; Li et al., 2016), or the conjugative and mobilisable elements recently found in streptococci (Ambroset et al., 2016; Coluzzi et al., 2017) and the Phage-Inducible Chromosomal Islands (PICIs) of *Staphylococcus aureus* and other Gram-positive (Penadés and Christie, 2015).

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) (Newman et al., 2006; Thomson et al., 2008; Quiroz et al., 2011), one of the most prevalent serotypes of *Salmonella* in humans and other hosts, such as poultry (Hendriksen et al., 2011; Ao et al., 2015; European Food Safety Authority and European Centre for Disease Prevention and Control, 2015). This genomic island was identified by Thomson *et al.* (2008), 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 (Thomson et al., 2008). Genes contained within ROD21 encode potential

virulence factors, such as the TlpA protein (Quiroz et al., 2011), which has a toll/interleukin-1 receptor (TIR) domain (Newman et al., 2006). A previous study has shown that this protein is involved in bacterial survival within macrophages by disrupting intracellular signaling events that coordinate NF-κB activation and IL-1β secretion (Newman et al., 2006). 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 (Tobar et al., 2013).

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 (Quiroz et al., 2011).

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_1) O:38(P), O:44(V), O:3,10(E_1) and O:54. Although the size of the GIs ranges from 19 kb (*Photorhabdus 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* RNS08.42.1A and *P. parmentieri* WPP163, which have two (Table 1).

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: names of genomic islands already identified in the Islander database ([a]) or in the literature with their indicated references ([b]), and the predicted length in amino acids for the full-length ([c]) and short versions ([d]) of H-NS homologues, as well as TIR-like domain containing proteins ([e]) when present. Note: *Salmonella* ser. Pullorum is considered as a biovar of serovar Gallinarum and the Pwa2.30N island was found only by searching in the Islander database.

Bacterium	Accession N°	Island Length (bp)	Islander[a]	Literature[b]	H-NS[c] (aa)	H-NS-t[d] (aa)	Tcp[e] (aa)
Photorhabdus luminescens subsp. *laumondii* TTO1	BX571865.1	18,997	-	-	-	-	-
Salmonella serovar Typhi ERL024120	LT906494.1	19,825	-	-	-	-	-
Salmonella serovar Typhi P-stx-12	CP003278.1	20,309	Sen353.21N	-	-	-	-
Klebsiella michiganensis M1	CP008841.1	20,363	-	-	-	78	-
Escherichia coli S1	CP010226.1	20,717	-	-	-	-	-
Escherichia coli M8	CP019953.1	20,970	-	-	134	80	-
Salmonella serovar Inverness ATCC 10720	CP019181.1	21,047	-	-	-	-	-
Serratia marcescens UMH3	CP018925.1	21,268	-	-	-	78	-
Salmonella serovar Agona 24249	CP006876.1	21,485	Sen74.21N	-	-	-	-
Salmonella serovar Borreze SA20041063	CP019407.1	21,485	-	-	-	-	-
Escherichia coli Ecol_517	CP018965.1	21,622	-	-	134	80	-
Escherichia coli MEM	CP012378.1	21,706	-	-	-	-	-
Escherichia coli GE3	CP012376.1	22,458	-	-	134	80	-
Salmonella serovar Typhi Ty2	AE014613.1	22,588	Sen355.23N	(Deng et al., 2003)	-	-	-
Salmonella serovar Sloterdijk ATCC 15791	CP012349.1	22,804	-	-	134	80	169
Escherichia coli CI5	CP011018.1	23,855	-	-	134	81	-
Salmonella serovar Typhi M223	LT904854.1	23,951	-	-	-	-	-
Escherichia coli D7	CP010150.1	24,691	-	-	134	80	-
Salmonella serovar Quebec S-1267	CP022019.1	24,817	-	-	134	-	-
Cedecea neteri FDAARGOS_392	CP023525.1	25,678	-	-	-	90	-
Escherichia coli 0127:H6 E2348/69	FM180568.1	25,819	Eco626.26N	IE3 (Iguchi et al., 2009)	-	80	-
Salmonella serovar Typhi PM016/13	CP012091.1	26,314	-	-	134	-	-
Salmonella serovar Pullorum ATCC 9120	CP012347.1	26,359	Sen215.26N	-	134	-	293

Table 1 (continued). *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: names of genomic islands already identified in the Islander database ([a]) or in the literature with their indicated references ([b]), and the predicted length in amino acids for the full-length ([c]) and short versions ([d]) of H-NS homologues, as well as TIR-like domain containing proteins ([e]) when present. Note: *Salmonella* ser. Pullorum is considered as a biovar of serovar Gallinarum and the Pwa2.30N island was found only by searching in the Islander database.

Bacterium	Accession N°	Island Length (bp)	Islander[a]	Literature[b]	H-NS[c] (aa)	H-NS-t[d] (aa)	Tcp[e] (aa)
Salmonella serovar Gallinarum 287/91	AM933173.1	26,404	Sen212.27N	ROD21 (Thomson et al., 2008)	134	-	293
Salmonella serovar Enteritidis P125109	AM933172.1	26,496	Sen204.27N	ROD21 (Thomson et al., 2008)	134	-	293
Salmonella serovar Dublin CT_02021853	CP001144.1	26,498	Sen82.26N	ROD21 (Porwollik et al., 2005)	134	-	293
Salmonella serovar Nitra S-1687	CP019416.1	26,500	-	-	134	-	293
Citrobacter freundii 18-1	CP022273.1	27,132	-	-	-	-	169
Salmonella serovar Anatum USDA-ARS-USMARC-1175	CP007483.2	27,139	-	-	134	-	-
Escherichia coli ETEC-2265	CP023346.1	27,200	-	-	134	-	-
Escherichia coli YD786	CP013112.1	27,239	-	-	134	80	-
Enterobacter sp. FY-07	CP012487.1	27,410	-	-	134	-	-
Klebsiella pneumoniae 30684/NJST258_2	CP006918.1	27,495	-	ICEKp258.2 (DeLeo et al., 2014)	134	-	139
Pectobacterium wasabiae CFBP 3304	CP015750.1	27,690	-	-	133	90	-
Pectobacterium atrosepticum SCRI1043	BX950851.1	28,081	Pat3.28N	HAI13 (Bell et al., 2004)	133	83	-
Escherichia coli SF-088	CP012635.1	28,452	-	-	134	-	-
Klebsiella pneumoniae strain AR_0148	CP021950.1	28,777	-	-	134	-	-
Serratia marcescens SM39	AP013063.1	28,938	-	-	134	-	-
Serratia marcescens CAV1492	CP011642.1	29,010	-	-	134	-	-
Kluyvera intermedia CAV1151	CP011602.1	29,250	-	-	134	-	-
Escherichia coli ED1a	CU928162.2	29,355	-	-	134	-	-
Klebsiella pneumoniae 1084	CP003785.1	29 ,853	Kpn29.30N	-	134	-	-
Yersinia intermedia Y228	CP009801.1	30,275	-	-	-	83	-
Pectobacterium parmentieri WPP163	CP001790.1	30,288	Pwa2.30N	-	-	-	-
Raoultella ornithinolytica Yangling I2	CP013338.1	30,440	-	-	134	-	-

Table 1 (continued). *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: names of genomic islands already identified in the Islander database ([a]) or in the literature with their indicated references ([b]), and the predicted length in amino acids for the full-length ([c]) and short versions ([d]) of H-NS homologues, as well as TIR-like domain containing proteins ([e]) when present. Note: *Salmonella* ser. Pullorum is considered as a biovar of serovar Gallinarum and the Pwa2.30N island was found only by searching in the Islander database.

Bacterium	**Accession N°**	**Island Length (bp)**	**Islander**[a]	**Literature**[b]	**H-NS**[c] **(aa)**	**H-NS-t**[d] **(aa)**	**Tcp**[e] **(aa)**
Escherichia coli ABWA45	CP022154.1	30,538	-	-	-	-	-
Yersinia rohdei YRA	CP009787.1	30,663	-	-	-	81	-
Pectobacterium parmentieri RNS08.42.1A	CP015749.1	33,037	-	-	133(2)	80	-
Pectobacterium carotovorum subsp. *brasiliense* BC1	CP009769.1	33,571	-	-	133	86	-
Pectobacterium parmentieri RNS08.42.1A	CP015749.1	33,684	-	-	-	-	-
Pectobacterium parmentieri WPP163	CP001790.1	33,698	Pwa2.34N	-	133	81	-
Pectobacterium atrosepticum 21A	CP009125.1	35,819	-	-	132	87	-
Pectobacterium wasabie SCC3193	CP003415.1	37,046	-	-	-	-	-
Pectobacterium carotovorum SCC1	CP021894.1	37,050		-	-	-	-
Klebsiella oxytoca AR_0147	CP020358.1	37,983	-	-	-	-	-
Pectobacterium atrosepticum SCRI1043	BX950851.1	40,824	Pat3.41N	HAI7 (Bell et al., 2004)	132	80	-

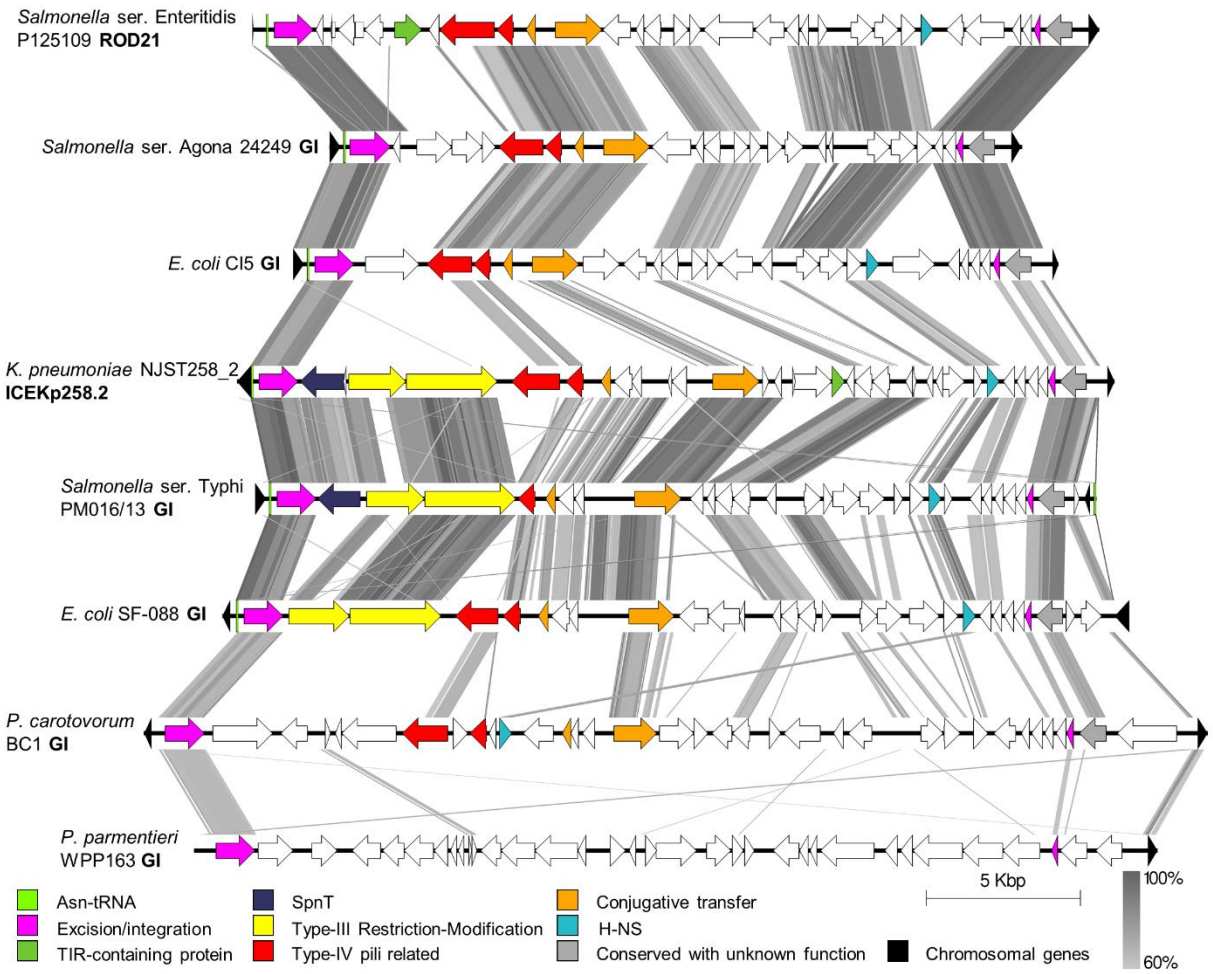


Figure 1. EARL genomic islands have a conserved excision/integration module. Sequence comparison (all three-frames translated for each DNA strand) of genomic islands carried out with tBLASTx. This figure shows eight representative GIs. Gray lines indicate regions with 60-100% identity. Note that the 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).

ROD21-like genomic islands share a conserved excision/integration module

Given that the identified ROD21-like GIs share a conserved integration site and a DRS, and that the annotations reveal the presence of different conserved genes with similar distribution, we carried out comparison analyses 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 first group includes ORFs similar to a putative type-IV pilin and a type-IV pilus-related protein; a second group includes ORFs similar to a putative relaxase belonging to the MobA/MobL family and a TraD homologue, and a third group including ORFs similar to 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* RNS08.42.1A 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.

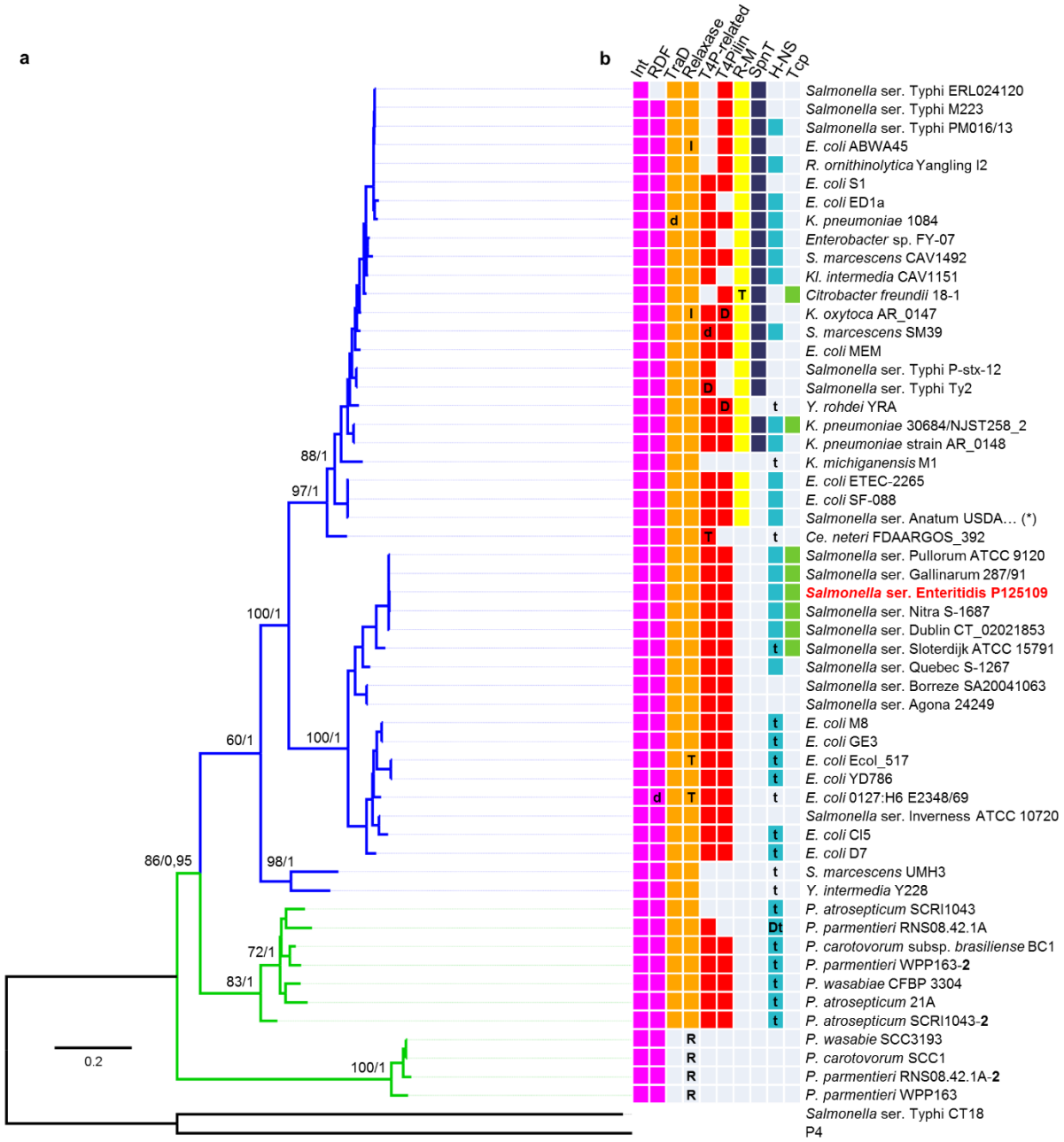

a
b
Int
RDF
TraD
Relaxase
T4P-related
T4Pilin
R-M
SpnT
H-NS
Tcp
Salmonella ser. Typhi ERL024120
Salmonella ser. Typhi M223
Salmonella ser. Typhi PM016/13
E. coli ABWA45
R. ornithinolytica Yangling I2
E. coli S1
E. coli ED1a
K. pneumoniae 1084
Enterobacter sp. FY-07
S. marcescens CAV1492
Kl. intermedia CAV1151
Citrobacter freundii 18-1
K. oxytoca AR_0147
S. marcescens SM39
E. coli MEM
Salmonella ser. Typhi P-stx-12
Salmonella ser. Typhi Ty2
Y. rohdei YRA
K. pneumoniae 30684/NJST258_2
K. pneumoniae strain AR_0148
K. michiganensis M1
E. coli ETEC-2265
E. coli SF-088
Salmonella ser. Anatum USDA… (*)
Ce. neteri FDAARGOS_392
Salmonella ser. Pullorum ATCC 9120
Salmonella ser. Gallinarum 287/91
Salmonella ser. Enteritidis P125109
Salmonella ser. Nitra S-1687
Salmonella ser. Dublin CT_02021853
Salmonella ser. Sloterdijk ATCC 15791
Salmonella ser. Quebec S-1267
Salmonella ser. Borreze SA20041063
Salmonella ser. Agona 24249
E. coli M8
E. coli GE3
E. coli Ecol_517
E. coli YD786
E. coli 0127:H6 E2348/69
Salmonella ser. Inverness ATCC 10720
E. coli Cl5
E. coli D7
S. marcescens UMH3
Y. intermedia Y228
P. atrosepticum SCRI1043
P. parmentieri RNS08.42.1A
P. carotovorum subsp. brasiliense BC1
P. parmentieri WPP163-2
P. wasabiae CFBP 3304
P. atrosepticum 21A
P. atrosepticum SCRI1043-2
P. wasabie SCC3193
P. carotovorum SCC1
P. parmentieri RNS08.42.1A-2
P. parmentieri WPP163
Salmonella ser. Typhi CT18
P4
88/1
97/1
100/1
60/1
100/1
86/0,95
98/1
72/1
83/1
100/1
0.2

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 colored 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 (R). The genome harboring ROD21 is in bold and red font. The number **2** after *Pectobacterium* strains WPP163, SCRI1043 and RNS08.42.1A is intended to identify the largest of the two EARL GIs harbored by these bacteria. **(*)** strain USDA-ARS-USMARC-1175.

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 (Fig. 1, 2 and Supplementary Fig. 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 in the same GI, which 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 the sequence of this protein to perform phylogenetic analysis using it as a marker to study the evolutionary history of EARL GIs (Boyd et al., 2009). 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 clustered EARL GIs present only in phytopathogens, while the other clade clusters EARL GIs from non-pathogenic strains and animal and plant pathogens (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 an 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 *hns* homologues (genes that encode H-NS 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).

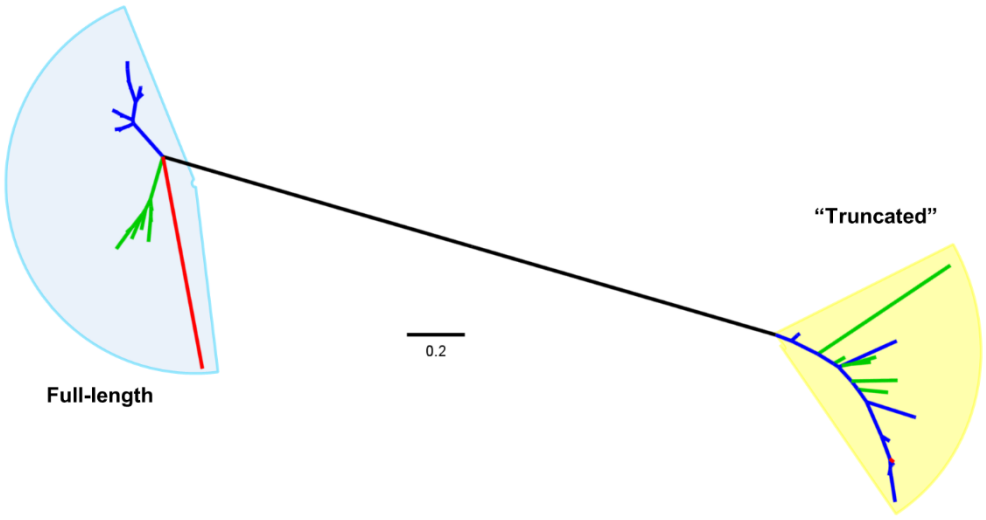


Figure 3. Full-length and truncated *hns* homologues belong to distantly related clades. Unrooted maximum likelihood phylogenetic tree constructed with codon-aligned *hns* homologue sequences found in EARL GIs. Blue and green branches correspond to animal- and plant-associated bacteria, respectively.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

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 the *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 the *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 integrase and RDF was generated. As shown in Fig. 4a, deletion of these genes prevents ROD21 excision. Both wild-type and mutant strain 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 a H-NS full-length homologue, respectively. As show in Fig. 4b to 4d, impairment of ROD21 excision prevented the proper expression of ROD21 genes on 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.

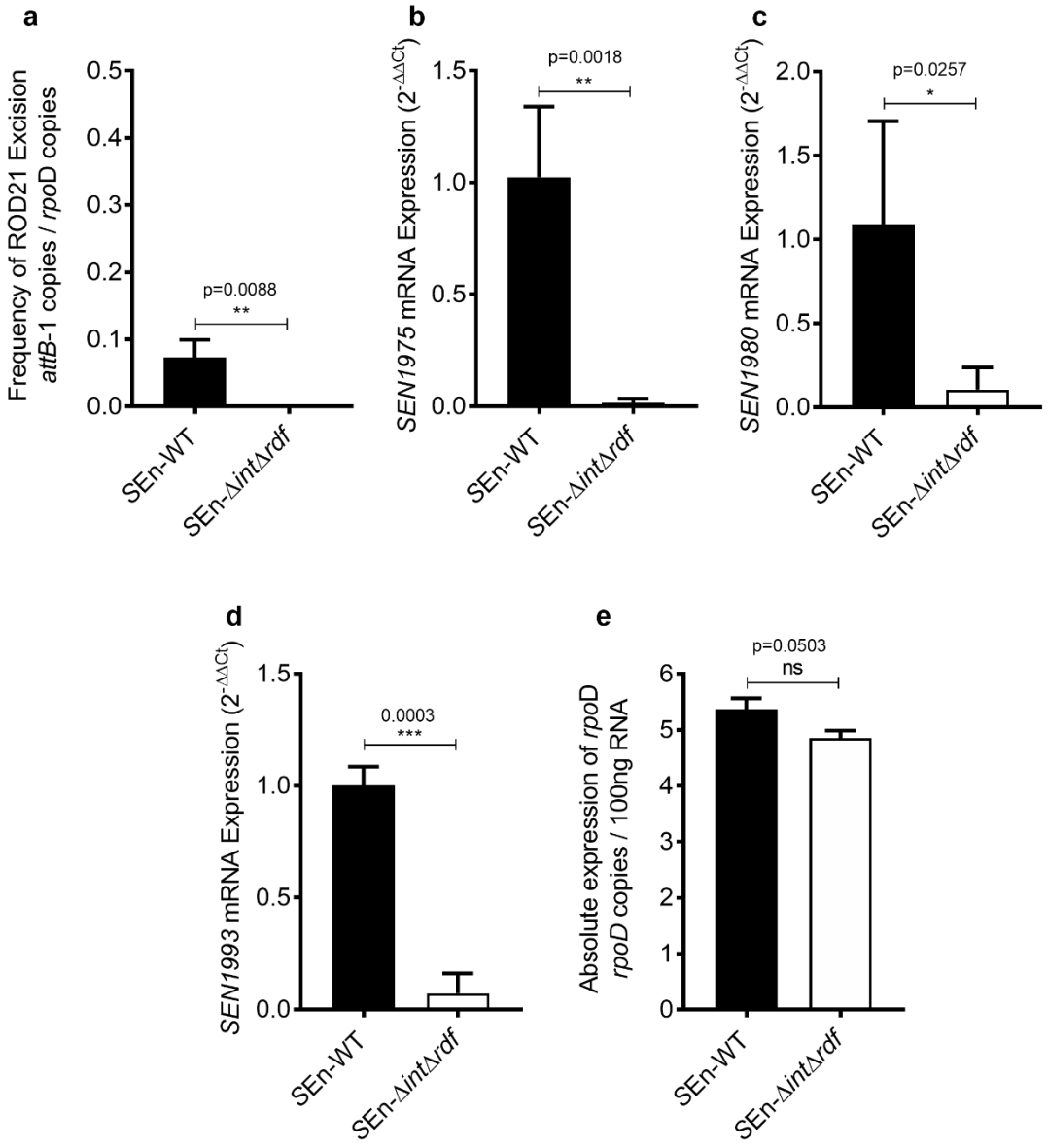


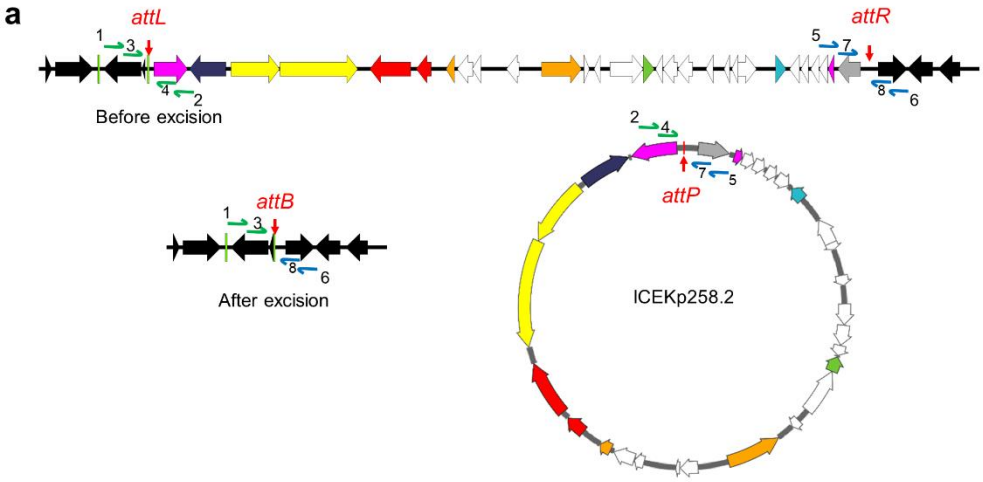
Figure 4. ROD21 excision affects the gene expression inside the island. (a) Deletion of the integrase (SEN1970) and putative RDF (SEN1998) coding sequences from *Salmonella* ser. Enteritis results in impairment ROD21 excision and reduction in expression of (b) SEN1975, (c) SEN1980 and (d) SEN1993 under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

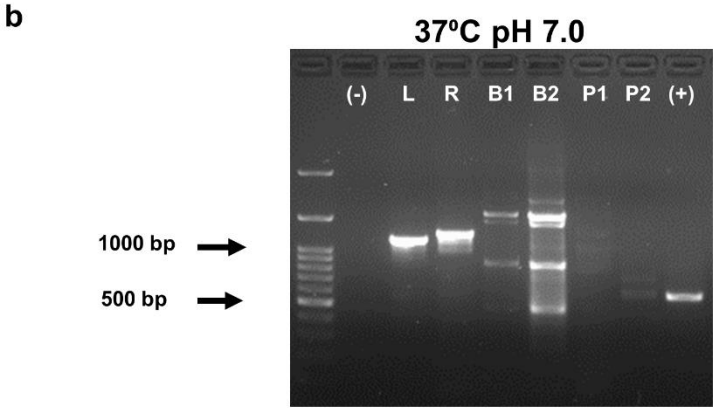
The EARL GIs from *K. pneumoniae* ST258 strain KP35 and enteropathogenic *E. coli* O127:H6 strain E2348/69 are excisable

Since ROD21 can be excised from the bacterial chromosome and this 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 23ºC/pH 7, 37ºC/pH 5.4, and 37ºC/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 37ºC/pH7 (Fig. 5b). For conditions 23ºC/pH 7 and 37ºC/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* (Fig. 6b and Supplementary Fig. 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 unspecific products. Therefore, to corroborate that the PCR products obtained were specifically the *attB* and *attP* sites expected, the amplicons were purified and sequenced. As shown in

Fig. 5c, the sequenced 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 sequence 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.


a
attL
attR
1
3
4
2
5
7
8
6
Before excision
attB
After excision
attP
ICEKp258.2


b
37ºC pH 7.0
(-)
L
R
B1
B2
P1
P2
(+)
1000 bp
500 bp

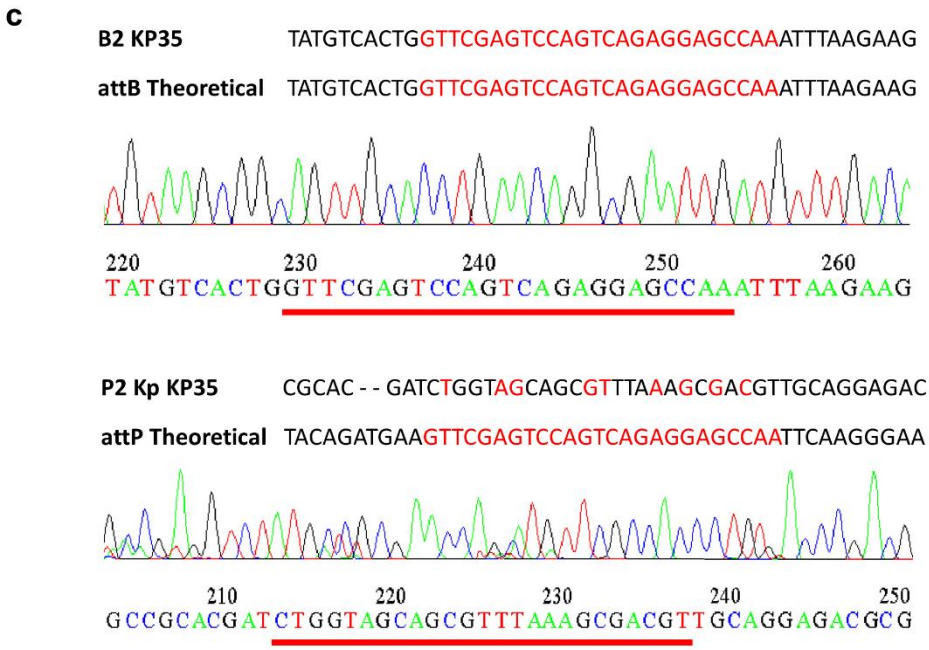
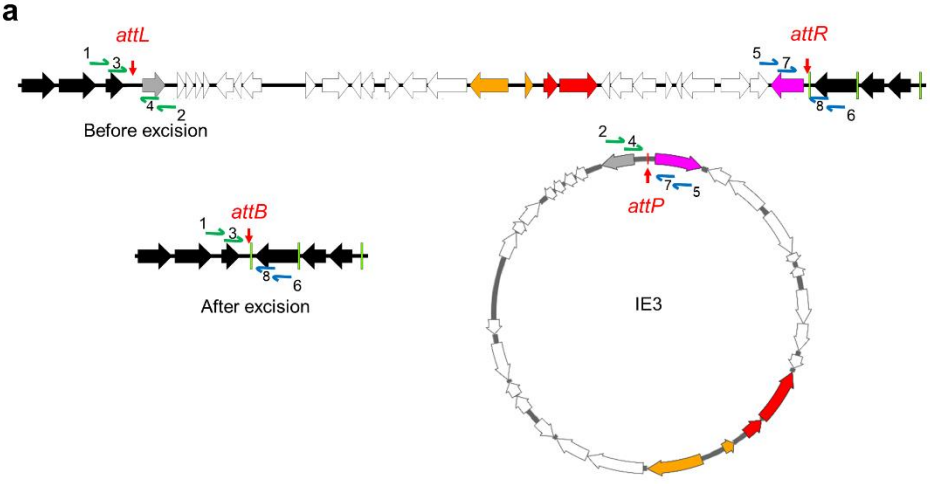
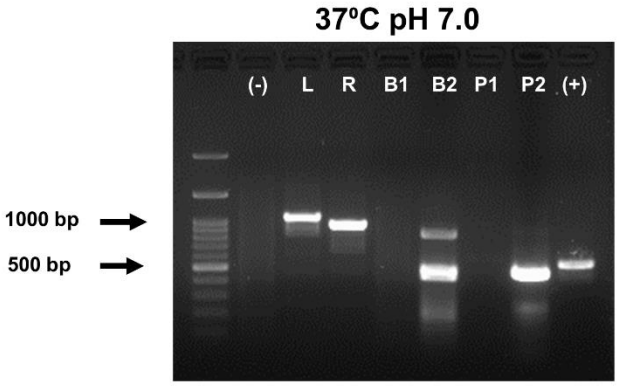

c
B2 KP35 TATGTCACTGGTTCGAGTCCAGTCAGAGGAGCCAAATTTAAGAAG
attB Theoretical TATGTCACTGGTTCGAGTCCAGTCAGAGGAGCCAAATTTAAGAAG
220
230
240
250
260
TATGTCACTGGTTCGAGTCCAGTCAGAGGAGCCAAATTTAAGAAG
P2 Kp KP35 CGCAC - - GATCTGGTAGCAGCGTTTAAAGCGACGTTGCAGGAGAC
attP Theoretical TACAGATGAAGTTCGAGTCCAGTCAGAGGAGCCAATTCAAGGGAA
210
220
230
240
250
GCCGCACGATCTGGTAGCAGCGTTTAAAGCGACGTTGCAGGAGACGCG

Figure 5. Detection of ICEKp258.2 excision in *K. pneumoniae* ST258 KP35. (a) Location 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. Amplicons containing *attB* or *attP* can be obtained only if the 27 kb island is excised from the chromosome. Gene colouring is similar as in Fig.1. (b) Agarose gel showing amplification products (expected sizes of PCR products are: L: 1123 bp, R: 1220 bp, B1: 1029 bp, P1:1340 bp,) and nested PCR reactions (B2: 519 bp, P2: 668 bp) using genomic DNA obtained from *K. pneumoniae* KP35 grown at 37ºC, 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. *att*B and *attP* specific sequences are highlighted in red and chromatograms are shown.

a



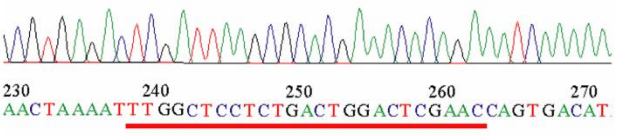
b



c

B2 Ec E2348/69 AACTAAAATTTGGCTCCTCTGACTGGACTCGAACCAGTGACAT

attB Theoretical AACTAAAATTTGGCTCCTCTGACTGGACTCGAACCAGTGACAT



P2 Ec E2348/69 CTTTCCTTAAATTGGCTCCTCTGACTGGACTCGAACTTCATCTGTA

attP Theoretical CTTTCCTTAAATTGGCTCCTCTGACTGGACTCGAACTTCATCTGTA

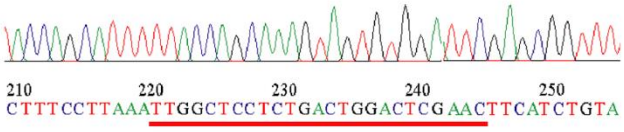


Figure 6. Detection of IE3 excision in EPEC E2348/69. (a) Location of the eight primers used for the nested PCR are indicated with blue and green arrows next to the *attL*, *attR*, *attB* and *attP* sites. Primers 1, 2, 5, 6 and 3, 4, 7, 8 were used for the first and second round in nested PCR. Amplicons containing *attB* or *attP* can be obtained only if the 26 kb island is excised from the chromosome. Gene colouring is similar as in Fig.1. (b) Agarose gel showing amplification products (L: 1057 bp, R: 937 bp, B1: 985 bp, P1: 1009 bp,) and nested PCR reactions (P2: 502 bp, B2: 507 bp), using genomic DNA obtained from EPEC E2348/69 grown at 37ºC, 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.

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 (Pitman et al., 2005; Godfrey et al., 2011; Quiroz et al., 2011; Tobar et al., 2013). Given the role of ROD21 excision in the virulence *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 *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 bacterial genomes available to date (October 2017). Although there are bioinformatic tools currently available to identify GIs with high accuracy, their prediction by current computational methods still is a challenging task, and existing GIs may be overlooked (Che et al., 2014; Lu and Leong, 2016). For instance, the Islander tool (Islander Database of Genomic Islands at Sandia National Laboratories) only identified 12 of the 55 islands identified in this study (Table 1). The reason for these significant differences is that the majority of the bacterial genomes that harbor EARL GIs are not in the RefSeq database used by Islander for GI prediction and thus omission of an important group of bacteria occurs in the analyses. However, Islander found an island which shares the features of EARL islands, namely Pwa2.30N, which could not be identified by our BLASTn search, likely because of a low degree of similarity between its DRS with the query sequence. Nevertheless, we generally found consensus between

the location and length of the 12 GIs predicted by Islander and our approach, except for GIs found in *Salmonella* serovars Enteritidis, Gallinarum and Typhi, for which Islander included a tRNA and remnant of integrase genes adjacent to the tRNA flanking the island, which are not part of the islands.

Dissemination of many GIs relies on bacterial conjugation (Salazar-Echegarai et al., 2014; Johnson and Grossman, 2015) 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 (Burrus et al., 2002), the integrative and mobilisable elements (IMEs; also known as mobilisable genomic islands) utilize the transfer machinery coded by ICEs or conjugative plasmids (Doublet et al., 2005; Daccord et al., 2010; Carraro et al., 2014).

Most EARL islands encode a TraD homologue and a putative relaxase of the MobA/MobL (MOB_Q superfamily) (Garcillán-Barcia et al., 2009). 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 coresident conjugative element(Ramsay and Firth, 2017). Transfer of ROD21 has been observed from *Salmonella* ser. Enteritidis to mutant strains lacking the island, as well as to *Salmonella* ser. Typhimurium (Salazar-Echegarai et al., 2014). 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 (Salazar-Echegarai et al., 2014). 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 as increased amounts of DNA in the bacterial genome require its replication and maintenance. Furthermore, a lack of regulatory mechanisms for controlling the expression of the newly acquired genes poses metabolic challenges for the bacterium (Singh et al., 2016). Here we support the notion that the ability of EARL islands to excise from the chromosome might be a mechanism conserved among these GIs that modulates gene expression. We have demonstrated here that expression of genes within the EARL is affected in a mutant strain of *Salmonella* ser. Enteritidis unable to excise ROD21. 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), previous work assessing the effects of ROD21 excision found the opposite effect (Tobar et al., 2013). 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 (Δ*SEN1970*), and the entire region spanning the integrase and the Asn-tRNA which has the *attL* site of the island [Δ*(asnT2-SEN1970)*], respectively; while our mutant strain lacks the integrase and putative RDF (Δ*SEN1970* Δ*SEN1998*). Therefore, 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 (Carpenter et al., 2016). Further research is required to elucidate the relation between SEN1998 and 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 (Godfrey et al., 2011; Nieto et al., 2016).

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 (Lucchini et al., 2006; Ali et al., 2012; Kotlajich et al., 2015; Higashi et al., 2016; Singh et al., 2016). 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 their preservation, by providing a regulation 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, since the Hfp protein, an H-NS homologue encoded in the *serU* island of some UPEC strains participate in the regulation of virulence factors encoded outside the island (Müller et al., 2010). 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 the H-NST encoded in the IE3 island of enteropathogenic *E. coli* E2348/69 (Williamson and Free, 2005). 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 (Williamson and Free, 2005). The H-NST homologs encoded in 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 alignment (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 (Newman et al., 2006). Bacterial Tcps are known virulence factors able to disrupt the innate immune response by interacting with mammalian TIR domain-containing proteins of the TLR signaling pathway, such as TLRs, IL-1R, MyD88 and TIRAP, thus suppressing transactivation of the NF-κB transcriptional activator, with the consequent alteration of proinflammatory cytokine secretion (Cirl et al., 2008). Interestingly, we identified an ORF encoding a putative Tcp in the ICEKp258.2 island found in carbapenemase-resistant *K. pneumoniae* ST258 strains (Chen et al., 2014), which may have a role in the resistance of this ST against phagocytic killing (Ahn et al., 2016). 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.

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 the diversity of EARL GIs. Our phylogenetic analysis, based on the nucleotide sequence of the integrase open reading frame, 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 while 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 some 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 (Quiroz et al., 2011). 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 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 (Tobar et al., 2013; Salazar-Echegarai et al., 2014).

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) (Penadés and Christie, 2015; Martínez-Rubio et al., 2017) and the Staphylococcal Chromosomal Cassette *mec* (SCC*mec*) (Liu et al., 2016) in Gram-positive bacteria and the SXT/R391 family in Gram-negative γ-Proteobacteria (Wozniak et al., 2009; Li et al., 2016), which encode genes responsible for virulence and antimicrobial resistance. The

EARL islands described in this study adds 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 coded 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(Zhang et al., 2000) was performed using as query the sequence "*attL*+82 nt" from *Salmonella enterica* ser. Enteritidis P125109 (Accession Nº 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 (≤ 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 non-coding 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 (Altschul et al., 1997) with the standalone BLAST v2.6.0+ software run in EasyFig v2.2.2 (Sullivan et al., 2011) and then visualized and analyzed using EasyFig and Artemis Comparison Tool v13.0.0 (Carver et al., 2005).

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 (Kumar et al., 2016) and gapped columns were deleted (the aminoacidic sequence obtained from *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/) (Trifinopoulos et al., 2016) 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) (Ronquist et al., 2012) trees, using as an outgroup 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 coloring. 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 coloring 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 -80ºC. When required, overnight cultures of *Klebsiella pneumoniae* ST258 strain KP35, enteropathogenic *Escherichia coli* O127:H6 strain E2348/69, *Salmonella* ser. Enteritidis phage type 4 strain P125109 or the *Salmonella* ser. Enteritidis Δ*SEN1970*::*FRT* Δ*SEN1998*::*FRT* mutant strain were prepared inoculating 3 mL of LB broth with an aliquot/bead of the stock culture and incubated at 37ºC and 120 rpm overnight.

Assessment of ICEKp258.2 and IE3 excision.

Island excision was assessed under three conditions (37°C/pH 7.0; 23°C pH 7.0 and 37°C 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 μL of an overnight culture each bacterium 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 (Sambrook and Russell, 2001). 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/μL, 0.5 μM of each primer pair (IDT), 0.2 mM dNTP Mix, 1.5 mM $MgCl_2$ and 1U of *Taq* DNA Polymerase (Invitrogen™) in a reaction volume of 25 μL. A second round PCR was performed with Platinum™ *Pfx* DNA Polymerase (Invitrogen™) of *attB* and *attP* sequences, using 2 μL 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 90V 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 11.0.

Assessing the effect of ROD21 excision on gene expression

Overnight cultures of wildtype *Salmonella* ser. Enteritidis PT4 strain P125109 (NCTC13319) and the Δ*SEN1970*::*FRT* Δ*SEN1998*::*FRT* isogenic strain (lacking the integrase and putative RDF coding sequences), were inoculated in LB broth at pH 7.0 until 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°C. Two independent experiments performed in duplicate were carried out.

RNA and DNA extraction were carried out as described by manufacturer with some modifications. Briefly, after DNA precipitation and wash, 300 µL of TE buffer and 300 µL 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°C and the aqueous phase (100 µL) was recovered. DNA was precipitated by adding 60 µL of propan-2-ol and 10 µL of 3M sodium acetate followed by a 30 min incubation at -20°C and centrifugation at 14,800 rpm for 15 min at 4°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 20 µL 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 ($2^{-\Delta\Delta Ct}$).

Data Availability

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

Acknowledgements

We are grateful to Dr. Alice Prince from Columbia University Medical Center, New York, United States, and Dr. Roberto Vidal from Universidad de Chile, Santiago, Chile, for kindly providing bacterial strains *K. pneumoniae* KP35 and EPEC E2348/69, respectively. Authors of this study were supported by Fondo Nacional de Investigación Científica y Tecnológica de Chile CONICYT (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. API, CPR and ICA are also supported by Beca de Doctorado Nacional CONICYT (21172030, 21140169 and 63140215).

Author Contributions

API, DUA and SB conceived the study; API, DUA, CPR, BS and FS designed the methodology; API and DUA collected the data, API, DUA, CPR and FS analysed the sequences, API, DUA, CPR and ICA performed the experiments and made the figures and tables; PG and SB acquired funding and supervised the research. All authors contributed in manuscript writing and revision, and approved the final version.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

References

Ahn D., Peñaloza H., Wang Z., Wickersham M., Parker D., Patel P., Koller A., Chen E. I., Bueno S. M., Uhlemann A.-C., and Prince A. (2016). Acquired resistance to innate immune clearance promotes *Klebsiella pneumoniae* ST258 pulmonary infection. *JCI Insight, 1*(17), e89704. https://doi.org/10.1172/jci.insight.89704

Ali S. S., Xia B., Liu J., and Navarre W. W. (2012). Silencing of foreign DNA in bacteria. *Curr. Opin. Microbiol.*, *15*, 175–181. https://doi.org/10.1016/j.mib.2011.12.014

Almagro-Moreno S., Napolitano M. G., and Boyd E. F. (2010). Excision dynamics of *Vibrio* pathogenicity island-2 from *Vibrio cholerae*: role of a recombination directionality factor VefA. *BMC Microbiol.*, *10*, 306.

Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W., and Lipman D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, *25*(17), 3389–3402.

Ambroset C., Coluzzi C., Guédon G., Devignes M. D., Loux V., Lacroix T., Payot S., and Leblond-Bourget N. (2016). New insights into the Classification and Integration Specificity of *Streptococcus* Integrative Conjugative Elements through Extensive Genome Exploration. *Front. Microbiol.*, *6*, 1483. https://doi.org/10.3389/fmicb.2015.01483

Ao T. T., Feasey N. A., Gordon M. A., Keddy K. H., Angulo F. J., and Crump J. A. (2015). Global Burden of Invasive Nontyphoidal *Salmonella* Disease, 2010. *Emerg. Infect. Dis.*, *21*(6), 941–949. https://doi.org/10.3201/eid2106.140999

Bell K. S., Sebaihia M., Pritchard L., Holden M. T. G., Hyman L. J., Holeva M. C., Thomson N. R., Bentley S. D., Churcher L. J. C., Mungall K., Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J., Fraser A., Hance Z., Hauser H., … Toth I. K. (2004). Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. *Proc. Natl. Acad. Sci. U. S. A.*, *101*(30), 11105–11110.

Boyd E. F., Almagro-Moreno S., and Parent M. A. (2009). Genomic islands are dynamic, ancient integrative elements in bacterial evolution. *Trends Microbiol.*, *17*(2), 47–53. https://doi.org/10.1016/j.tim.2008.11.003

Burrus V., Pavlovic G., Decaris B., and Guédon G. (2002). The ICE*St1* 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. https://doi.org/10.1016/S0147-619X(02)00102-6

Carpenter M. R., Rozovsky S., and Boyd E. F. (2016). Pathogenicity island cross talk mediated by recombination directionality factors facilitates excision from the chromosome. *J. Bacteriol.*, *198*(5), 766–776. https://doi.org/10.1128/JB.00704-15

Carraro N., Matteau D., Luo P., Rodrigue S., and Burrus V. (2014). The Master Activator of IncA/C Conjugative Plasmids Stimulates Genomic Islands and Multidrug Resistance Dissemination. *PLoS Genet.*, *10*(10), e1004714. https://doi.org/10.1371/journal.pgen.1004714

Carraro N., Poulin D., and Burrus V. (2015). 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*(6), e1005298. https://doi.org/10.1371/journal.pgen.1005298

Carver T. J., Rutherford K. M., Berriman M., Rajandream M. A., Barrell B. G., and Parkhill J. (2005). ACT: The Artemis comparison tool. *Bioinformatics*, *21*(16), 3422–3423. https://doi.org/10.1093/bioinformatics/bti553

Che D., Hasan M. S., and Chen B. (2014). Identifying Pathogenicity Islands in Bacterial Pathogenomics Using Computational Approaches. *Pathogens*, *3*, 36–56. https://doi.org/10.3390/pathogens3010036

Chen L., Mathema B., Pitout J. D. D., DeLeo F. R., and Kreiswirth B. N. (2014). Epidemic *Klebsiella pneumoniae* ST258 Is a Hybrid Strain. *MBio*, *5*(3), e01355-14. https://doi.org/10.1128/mBio.01355-14

Cirl C., Wieser A., Yadav M., Duerr S., Schubert S., Fischer H., Stappert D., Wantia N., Rodriguez N., Wagner H., Svanborg C., and Miethke T. (2008). Subversion of Toll-like receptor signaling by a unique family of bacterial Toll/interleukin-1 receptor domain-containing proteins. *Nat. Med.*, *14*(4), 399–406. https://doi.org/10.1038/nm1734

Coluzzi C., Guédon G., Devignes M. D., Ambroset C., Loux V., Lacroix T., Payot S., and Leblond-Bourget N. (2017). 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. https://doi.org/10.3389/fmicb.2017.00443

Daccord A., Ceccarelli D., and Burrus V. (2010). 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*(3), 576–588. https://doi.org/10.1111/j.1365-2958.2010.07364.x

Daccord A., Ceccarelli D., Rodrigue S., and Burrus V. (2013). Comparative Analysis of Mobilizable Genomic Islands. *J. Bacteriol.*, *195*(3), 606–614. https://doi.org/10.1128/JB.01985-12

DeLeo F. R., Chen L., Porcella S. F., Martens C. A., Kobayashi S. D., Porter A. R., Chavda K. D., Jacobs M. R., Mathema B., Olsen R. J., Bonomo R. A., Musser J. M., and Kreiswirth B. N. (2014). Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 *Klebsiella pneumoniae*. *Proc. Natl. Acad. Sci. U. S. A.*, *111*(13), 4988–4993. https://doi.org/10.1073/pnas.1321364111

Deng W., Liou S.-R., Plunkett III G., Mayshew G. F., Rose D. J., Burland V., Kodoyianni V., Schwartz D. C., and Blattner F. R. (2003). Comparative Genomics of *Salmonella enterica* Serovar Typhi Strains Ty2 and and CT18. *J. Bacteriol.*, *185*(7), 2330–2337. https://doi.org/10.1128/JB.185.7.2330

Dobrindt U., Hochhut B., Hentschel U., and Hacker J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.*, *2*(5), 414–424. https://doi.org/10.1038/nrmicro884

Doublet B., Boyd D., Mulvey M. R., and Cloeckaert A. (2005). The *Salmonella* genomic island 1 is an integrative mobilizable element. *Mol. Microbiol.*, *55*(6), 1911–1924. https://doi.org/10.1111/j.1365-2958.2005.04520.x

European Food Safety Authority, and European Centre for Disease Prevention and Control. (2015). The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2013. *EFSA J.*, *13*(1), 3991. https://doi.org/10.2903/j.efsa.2015.3991

Garcillán-Barcia M. P., Francia M. V., and De La Cruz F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.*, *33*(3), 657–687. https://doi.org/10.1111/j.1574-6976.2009.00168.x

Godfrey S. A. C., Lovell H. C., Mansfield J. W., Corry D. S., Jackson R. W., and Arnold D. L. (2011). The Stealth Episome: Suppression of Gene Expression on the Excised Genomic Island PPHGI-1 from *Pseudomonas syringae* pv. *phaseolicola*. *PLoS Pathog.*, *7*(3), e1002010. https://doi.org/10.1371/journal.ppat.1002010

Hacker J., and Kaper J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.*, *54*, 641–679.

Haskett T. L., Terpolilli J. J., Bekuma A., O'Hara G. W., Sullivan J. T., Wang P., Ronson C. W., and Ramsay J. P. (2016). Assembly and transfer of tripartite integrative and conjugative genetic elements. *Proc. Natl. Acad. Sci. U. S. A.*, *113*(43), 12268–12273. https://doi.org/10.1073/pnas.1613358113

Hendriksen R. S., Vieira A. R., Karlsmose S., Lo Fo Wong D. M. A., Jensen A. B., Wegener H. C., and Aarestrup F. M. (2011). 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*(8), 887–900. https://doi.org/10.1089/fpd.2010.0787

Higashi K., Tobe T., Kanai A., Uyar E., Ishikawa S., Suzuki Y., Ogasawara N., Kurokawa K., and Oshima T. (2016). H-NS Facilitates Sequence Diversification of Horizontally Transferred DNAs during Their Integration in Host Chromosomes. *PLoS Genet.*, *12*(1), 1–31. https://doi.org/10.1371/journal.pgen.1005796

Hudson C. M., Lau B. Y., and Williams K. P. (2015). Islander: A database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic Acids Res.*, *43*(D1), D48–D53. https://doi.org/10.1093/nar/gku1072

Iguchi A., Thomson N. R., Ogura Y., Saunders D., Ooka T., Henderson I. R., Harris D., Asadulghani M., Kurokawa K., Dean P., Kenny B., Quail M. A., Thurston S., Dougan G., Hayashi T., Parkhill J., and Frankel G. (2009). Complete Genome Sequence and Comparative Genome Analysis of Enteropathogenic *Escherichia coli* O127:H6 Strain E2348/69. *J. Bacteriol.*, *191*(1), 347–354. https://doi.org/10.1128/JB.01238-08

Johnson C. M., and Grossman A. D. (2015). Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Annu. Rev. Genet.*, *49*, 577–601. https://doi.org/10.1146/annurev-genet-112414-055018

Kotlajich M. V., Hron D. R., Boudreau B. A., Sun Z., Lyubchenko Y. L., and Landick R. (2015). Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria. *Elife*, *4*, e04970. https://doi.org/10.7554/eLife.04970

Kumar S., Stecher G., and Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Lesic B., Bach S., Ghigo J. M., Dobrindt U., Hacker J., and Carniel E. (2004). Excision of the high-pathogenicity of *Yersinia pseudotuberculosis* requires the combined of actions of its cognate integrase and Hef, a new recombination directionality factor. *Mol. Microbiol.*, *52*(5), 1337–1348. https://doi.org/10.1111/j.1365-2958.2004.04073.x

Lewis J. A., and Hatfull G. F. (2001). Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res.*, *29*(11), 2205–2216.

Li X., Du Y., Du P., Dai H., Fang Y., Li Z., Lv N., Zhu B., Kan B., and Wang D. (2016). SXT/R391 integrative and conjugative elements in *Proteus* species reveal abundant genetic diversity and multidrug resistance. *Sci. Rep.*, *6*, 37372. https://doi.org/10.1038/srep37372

Liu J., Chen D., Peters B. M., Li L., Li B., Xu Z., and Shirliff M. E. (2016). Staphylococcal chromosomal cassettes mec (SCCmec): A mobile genetic element in methicillin-resistant *Staphylococcus aureus*. *Microb. Pathog.*, *101*, 56–67. https://doi.org/10.1016/j.micpath.2016.10.028

Lovell H. C., Jackson R. W., Mansfield J. W., Godfrey S. A. C., Hancock J. T., Desikan R., and Arnold D. L. (2011). In planta conditions induce genomic changes in *Pseudomonas syringae* pv. *phaseolicola*. *Mol. Plant Pathol.*, *12*(2), 167–176. https://doi.org/10.1111/J.1364-3703.2010.00658.X

Lu B., and Leong H. W. (2016). Computational methods for predicting genomic islands in microbial genomes. *Comput. Struct. Biotechnol. J.*, *14*, 200–206. https://doi.org/10.1016/j.csbj.2016.05.001

Lucchini S., Rowley G., Goldberg M. D., Hurd D., Harrison M., and Hinton J. C. D. (2006). H-NS Mediates the Silencing of Laterally Acquired Genes in Bacteria. *PLoS Pathog.*, *2*(8), e81. https://doi.org/10.1371/journal.ppat.0020081

Martínez-Rubio R., Quiles-Puchalt N., Martí M., Humphrey S., Ram G., Smyth D., Chen J., Novick R. P., and Penadés J. R. (2017). Phage-inducible islands in the Gram-positive cocci. *ISME J.*, *11*(4), 1029–1042. https://doi.org/10.1038/ismej.2016.163

Müller C. M., Schneider G., Dobrindt U., Emödy L., Hacker J., and Uhlin B. E. (2010). Differential effects and interactions of endogenous and horizontally acquired H-NS-like proteins in pathogenic *Escherichia coli*. *Mol. Microbiol.*, *75*(2), 280–293. https://doi.org/10.1111/j.1365-2958.2009.06995.x

Murphy R. A., and Boyd E. F. (2008). Three Pathogenicity Islands of *Vibrio cholerae* Can Excise from the Chromosome and Form Circular Intermediates. *J. Bacteriol.*, *190*(2), 636–647. https://doi.org/10.1128/JB.00562-07

Newman R. M., Salunkhe P., Godzik A., and Reed J. C. (2006). Identification and Characterization of a Novel Bacterial Virulence Factor That Shares Homology with Mammalian Toll/Interleukin-1 Receptor Family Proteins. *Infect. Immun.*, *74*(1), 594–601. https://doi.org/10.1128/IAI.74.1.594

Nieto P. A., Pardo-Roa C., Salazar-Echegarai F. J., Tobar H. E., Coronado-Arrázola I., Riedel C. A., Kalergis A. M., and Bueno S. M. (2016). New insights about excisable pathogenicity islands in *Salmonella* and their contribution to virulence. *Microbes Infect.*, *18*, 302–309. https://doi.org/10.1016/j.micinf.2016.02.001

Penadés J. R., and Christie G. E. (2015). The Phage-Inducible Chromosomal Islands: A Family of Highly Evolved Molecular Parasites. *Annu. Rev. Virol.*, *2*, 181–201. https://doi.org/10.1146/annurev-virology-031413-085446

Pitman A. R., Jackson R. W., Mansfield J. W., Kaitell V., Thwaites R., and Arnold D. L. (2005). Exposure to Host Resistance Mechanisms Drives Evolution of Bacterial Virulence in Plants. *Curr. Biol.*, *15*, 2230–2235. https://doi.org/10.1016/j.cub.2005.10.074

Porwollik S., Santiviago C. A., Cheng P., Florea L., Jackson S., and McClelland M. (2005). Differences in Gene Content between *Salmonella enterica* Serovar Enteritidis Isolates and Comparison to Closely Related Serovars Gallinarum and Dublin. *J. Bacteriol.*, *187*(18), 6545–6555. https://doi.org/10.1128/JB.187.18.6545

Quiroz T. S., Nieto P. A., Tobar H. E., Salazar-Echegarai F. J., Lizana R. J., Quezada C. P., Santiviago C. A., Araya D. V, Riedel C. A., Kalergis A. M., and Bueno S. M. (2011). Excision of an Unstable Pathogenicity Island in *Salmonella enterica* Serovar Enteritidis Is Induced during Infection of Phagocytic Cells. *PLoS One*, *6*(10), e26031. https://doi.org/10.1371/journal.pone.0026031

Ramsay J. P., and Firth N. (2017). Diverse mobilization strategies facilitate transfer of non-conjugative mobile genetic elements. *Curr. Opin. Microbiol.*, *38*, 1–9. https://doi.org/10.1016/j.mib.2017.03.003

Ronquist F., Teslenko M., van der Mark P., Ayres D. L., Darling A., Höhna S., Larget B., Liu L., Suchard M. A., and Huelsenbeck J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst. Biol.*, *61*(3), 539–542. https://doi.org/10.1093/sysbio/sys029

Salazar-Echegarai F. J., Tobar H. E., Nieto P. A., Riedel C. A., and Bueno S. M. (2014). Conjugal Transfer of the Pathogenicity Island ROD21 in *Salmonella enterica* serovar Enteritidis Depends on Environmental Conditions. *PLoS One*, *9*(4), e90626. https://doi.org/10.1371/journal.pone.0090626

Sambrook J., and Russell D. W. (2001). *Molecular Cloning: A Laboratory Manual* (Third). Cold Spring Harbor Laboratory Press.

Sentchilo V., Czechowska K., Pradervand N., Minoia M., Miyazaki R., and van der Meer J. R. (2009). Intracellular excision and reintegration dynamics of the ICEclc genomic island of *Pseudomonas knackmussii* sp. strain B13. *Mol. Microbiol.*, *72*(5), 1293–1306. https://doi.org/10.1111/j.1365-2958.2009.06726.x

Singh K., Milstein J. N., and Navarre W. W. (2016). Xenogeneic Silencing and Its Impact on Bacterial Genomes. *Annu. Rev. Microbiol.*, *70*, 199–213. https://doi.org/10.1017/CBO9781107415324.004

Sullivan J. T., and Ronson C. W. (1998). Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Genetics*, *95*(April), 5145–5149. https://doi.org/10.1073/pnas.95.9.5145

Sullivan M. J., Petty N. K., and Beatson S. A. (2011). Easyfig: A genome comparison visualizer. *Bioinformatics*, *27*(7), 1009–1010. https://doi.org/10.1093/bioinformatics/btr039

Thomson N. R., Clayton D. J., Windhorst D., Vernikos G., Davidson S., Churcher C., Quail M. A., Stevens M., Jones M. A., Watson M., Barron A., Layton A., Pickard D., Kingsley R. A., Bignell A., Clark L., Harris B., Ormond D., Abdellah Z., Brooks K., Cherevach I., Chillingworth T., Woodward J., Norberczak H., Lord A., Arrowsmith C., Jagels K., Moule S., Mungall K., Sanders M., Whitehead S., Chabalgoity J. A., Maskell D., Humphrey T., Roberts M., Barrow P. A., Dougan G., and Parkhill J. (2008). 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. https://doi.org/10.1101/gr.077404.108.1

Tobar H. E., Salazar-Echegarai F. J., Nieto P. A., Palavecino C. E., Sebastian V. P., Riedel C. A., Kalergis A. M., and Bueno S. M. (2013). Chromosomal Excision of a New Pathogenicity Island Modulates *Salmonella* Virulence *In Vivo*. *Curr. Gene Ther.*, *13*(4), 240–249.

Trifinopoulos J., Nguyen L., von Haeseler A., and Minh B. Q. (2016). W-IQ-TREE : a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.*, *44*, W232–W235. https://doi.org/10.1093/nar/gkw256

Vanga B. R., Butler R. C., Toth I. K., Ronson C. W., and Pitman A. R. (2012). Inactivation of PbTopo IIIβ causes hyper-excision of the Pathogenicity Island HAI2 resulting in reduced virulence of *Pectobacterium atrosepticum*. *Mol. Microbiol.*, *84*(4), 648–663. https://doi.org/10.1111/j.1365-2958.2012.08050.x

Williams K. P. (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res.*, *30*(4), 866–875. https://doi.org/10.1093/nar/30.4.866

Williamson H. S., and Free A. (2005). A truncated H-NS-like protein from enteropathogenic *Escherichia coli* acts as an H-NS antagonist. *Mol. Microbiol.*, *55*(3), 808–827. https://doi.org/10.1111/j.1365-2958.2004.04421.x

Wozniak R. A. F., Fouts D. E., Spagnoletti M., Colombo M. M., Ceccarelli D., Garriss G., Déry C., Burrus V., and Waldor M. K. (2009). Comparative ICE genomics: Insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genet.*, *5*(12), e1000786. https://doi.org/10.1371/journal.pgen.1000786

Zhang Z., Schwartz S., Wagner L., and Miller W. (2000). A Greedy Algorithm for Aligning DNA Sequences. *J. Comput. Biol.*, *7*(1/2), 203–214.

Supplementary Information

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

Alejandro Piña-Iturbe[1], Diego Ulloa-Allendes[1], Catalina Pardo-Roa[1], Irenice Coronado-Arrázola[1], Francisco Salazar-Echegarai[1], Bianca Sclavi[2], Pablo A. González[1], Susan M. Bueno[1]*

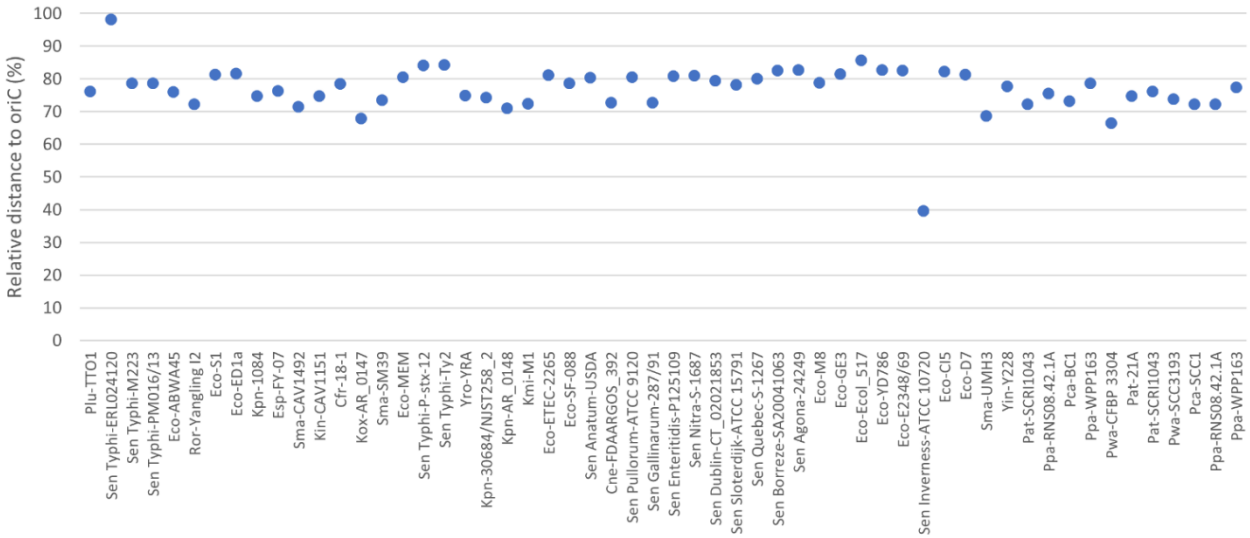
[1]Millennium Institute on Immunology and Immunotherapy, Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile.

[2]Laboratoire de Biologie et Pharmacologie Appliquée, Centre National de la Recherche Scientifique UMR 8113, École Normale Supérieure Paris-Saclay, Cachan, France.

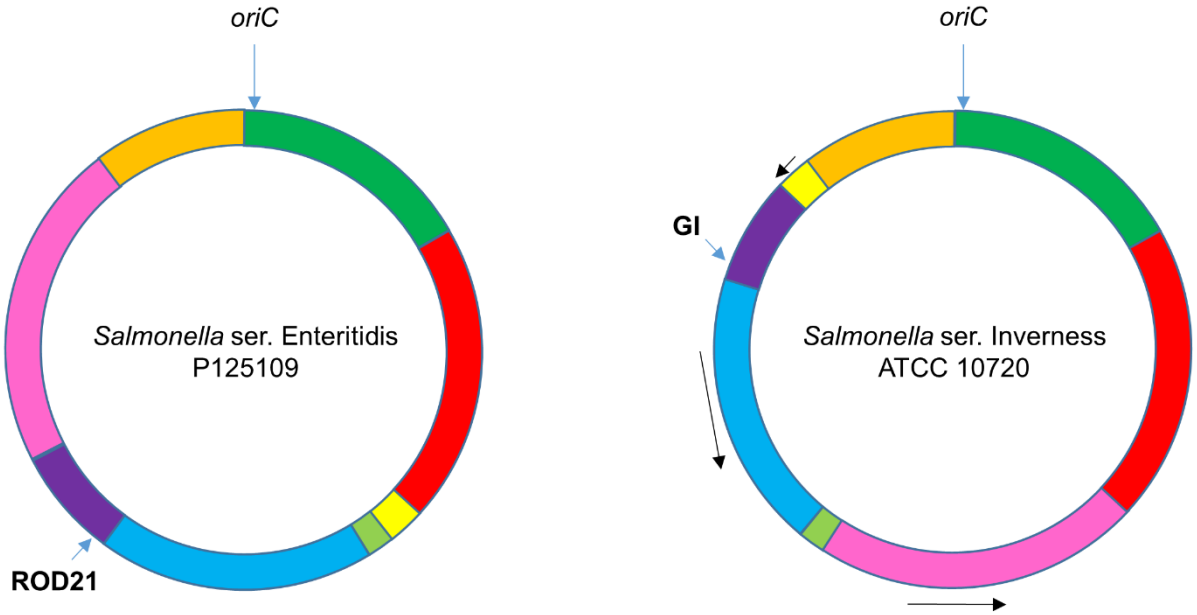
*sbueno@bio.puc.cl

Supplementary File S1. *Enterobacteriaceae*-associated ROD21-like genomic islands, nested PCR and qPCR. Location, length, and host genomes of EARL islands (**Table S1**). Location and length of genes encoding H-NS homologues and Tcps, island G+C content and distances from *oriC* (**Table S2**). Primers and probes used for quantification of ROD21 excision and gene expression (**Table S3**). Primers used for detection of island excision (**Table S4**). Sequence and location of the left and right attatchment sites of EARL islands, considering the orientation of the Asn-tRNA (**Table S5**). This information is provided in a separate .xlsx file.

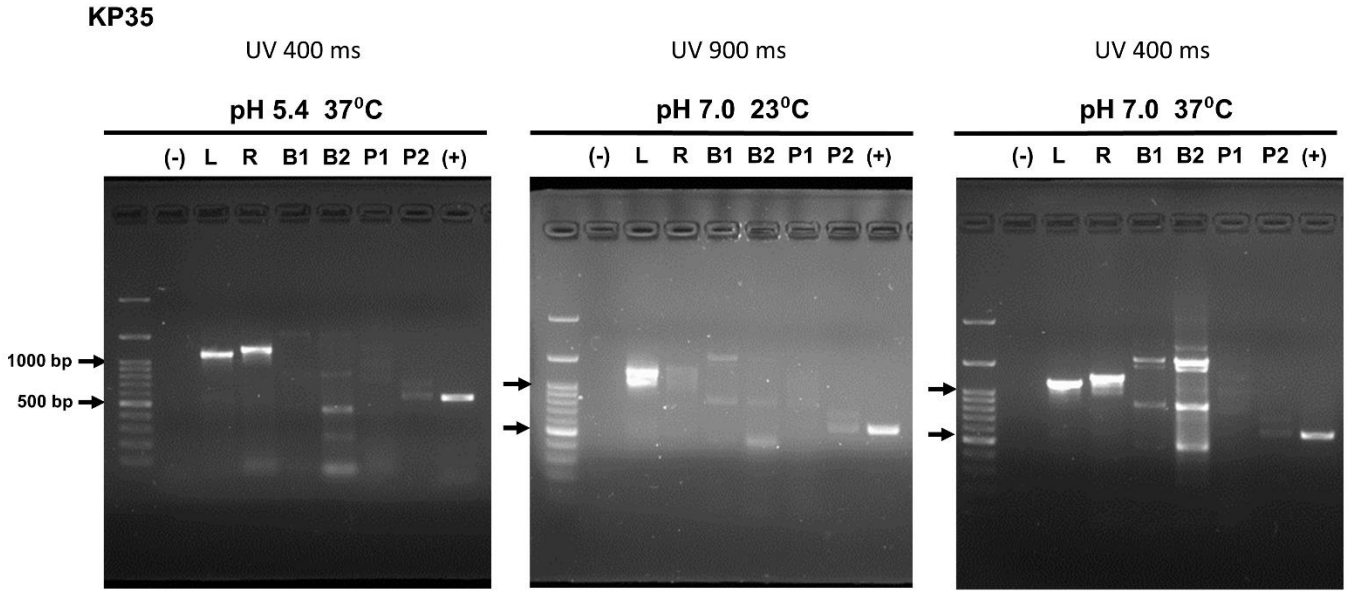
Available at https://www.nature.com/articles/s41598-018-28537-0#additional-information



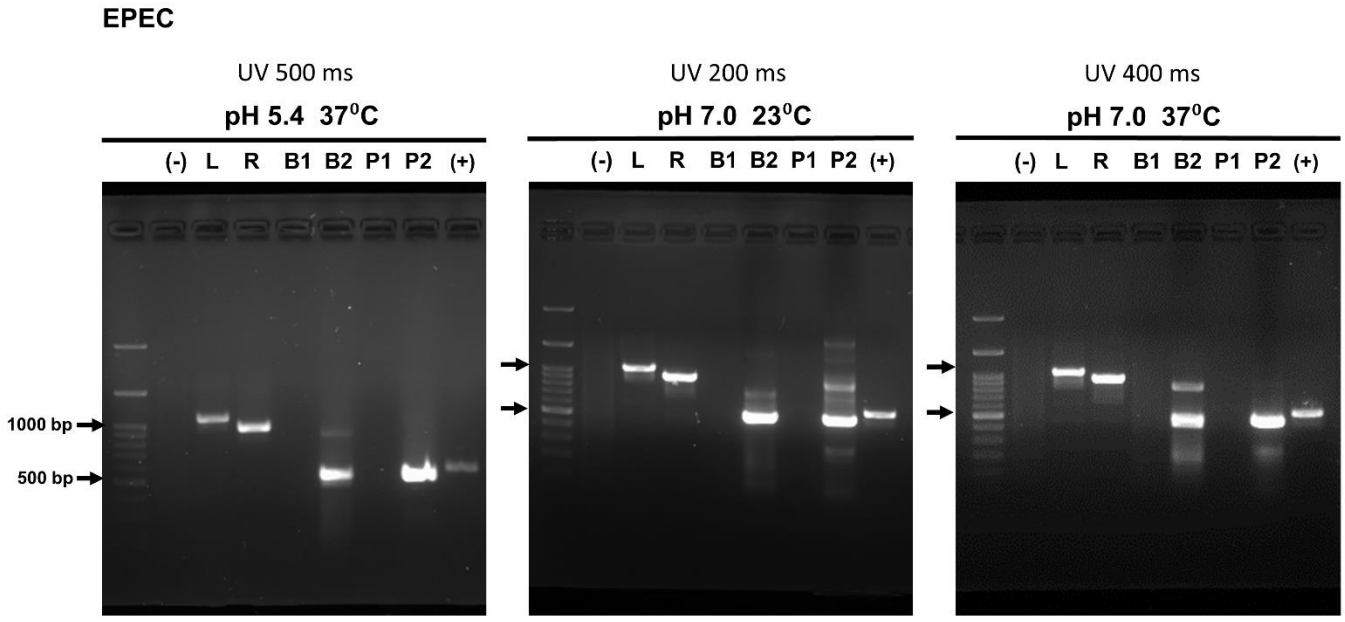
Supplementary Figure S1. EARL GIs are inserted far from the chromosomal replication origin (*oriC*). Relative distance of the insertion site of each EARL island from the chromosomal *oriC* expressed as a percentage of the maximum possible distance (one half of the chromosome length). The location and length of *oriC* were calculated using Ori-Finder (http://tubic.tju.edu.cn/Ori-Finder/)[1]. 0% represents the midpoint of *oriC* and 100% represent the farthest location in the chromosome.



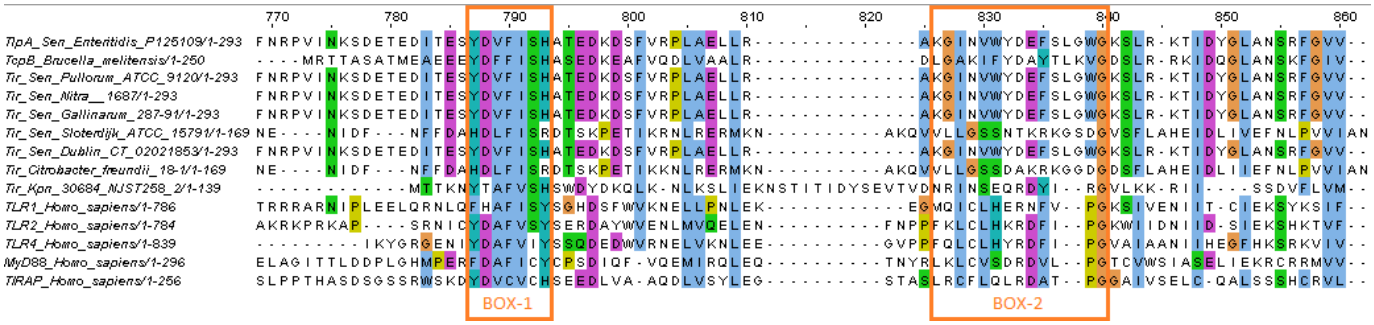
Supplementary Figure S2. The EARL island from *Salmonella* ser. Inverness ATCC 10720 is located near the oriC as the result of a chromosomal rearrangement. Circular representation of Mauve alignment of the chromosomes from *Salmonella* ser. Inverness ATCC 10720 and ser. Enteritidis P125109. The colored regions correspond to the blocks of homology identified using Mauve v2.4.0[2] and black arrows indicate the regions found in different orientation. ROD21 and GI indicate the approximate location of the EARL GIs found in each chromosome. The origin of replication identified with Ori-Finder is also indicated.



Supplementary Figure S3. Detection of ICEKp258.2 excision in different pH and temperature. Agarose gels showing amplification products (expected sizes of PCR products are: L: 1123 bp, R: 1220 bp, B1: 1029 bp, B2: 519 bp, P1:1340 bp, P2: 668 bp) of nested PCR reactions using genomic DNA obtained from *K. pneumoniae* strain KP35 grown at pH 5.4/37ºC, pH 7.0/23ºC and pH 7.0/37ºC. (-), 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 in milliseconds is indicated above each gel.



Supplementary Figure S4. Detection of IE3 excision in different pH and temperature. Agarose gel showing amplification products (L: 1057 bp, R: 937 bp, B1: 985 bp, B2: 507 bp, P1: 1009 bp, P2: 502 bp) of nested PCR reactions using genomic DNA obtained from enteropathogenic *E. coli* O127:H6 strain E2348/69 grown at pH 5.4/37ºC, pH 7.0/23ºC and pH 7.0/37ºC. (-), 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 in milliseconds is indicated above each gel.



Supplementary Figure S5. Multiple sequence alignment (MSA) of the putative TIR-domain containing proteins (Tcps) encoded in EARL islands (Tir), characterized bacterial Tcps (TlpA and TcpB), and human TIR proteins of the Toll-like receptor (TLR) signaling pathway (TLR1, TLR2, TLR4, MyD88 and TIRAP. The conserved Boxes 1 and 2 of the TIR domain are indicated with orange rectangles. MSA were carried out with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

References

1. Gao, F. & Zhang, C.-T. Ori-Finder: A web-based system for finding *oriCs* in unannotated bacterial genomes. *BMC Bioinformatics* 9, 79 (2008).

2. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS One* 5, e11147 (2010).

Sequence *attL*+82nt used as query for BLASTn search of EARL islands.

> AM933172.1:2061160-2061444 Salmonella ser. Enteritidis P125109

GTTCGAGTCCAGTCAGAGGAGCCAAATTAAGGAAAGCAGACGTTCACTGACGTCTGCTTTCTGCATTTAT
ATCAACTGGTTATTCCCTTCTTCAGGTTCACTCTCGTTCACTAAAAACCACTCGAAGCCATACCCTTTTG
CTGGTAAAAATGCTGGTAAAGCTGGTTCGATTTGTGTTTTACCAGCACGCGGAGGGAACCGTCATGTCAC
TTACTGATACCAAAGTAAAAAATACCAGACCATCGGAAAAGGCCGTCAAGCTCACTGACGGGTTTGGCCT
CTATC

Integrases used for construction of island phylogeny. Genome accession number, integrase-coding sequence location and strain.

(The full version, which includes the genes sequences, is available at:

https://www.nature.com/articles/s41598-018-28537-0#additional-information

>BX571865.1:c96668-95405 Photorhabdus luminescens subsp. laumondii TTO1

>LT906494.1:c296327-295053 Salmonella enterica subsp. enterica serovar Typhi strain ERL024120

>CP003278.1:c976446-975172 Salmonella enterica subsp. enterica serovar Typhi str. P-stx-12

>CP008841.1:4827330-4828610 Klebsiella michiganensis strain M1

>CP010226.1:1383693-1384967 Escherichia coli strain S1

>CP019953.1:2346134-2347402 Escherichia coli M8

>CP019181.1:c926034-924766 Salmonella enterica subsp. enterica serovar Inverness str. ATCC 10720

>CP018925.1:2796319-2797593 Serratia marcescens strain UMH3

>CP019407.1:2009888-2011162 Salmonella enterica subsp. enterica serovar Borreze str. SA20041063

>CP006876.1:1996007-1997335 Salmonella enterica subsp. enterica serovar Agona str. 24249

>CP018965.1:c3361590-3360316 Escherichia coli strain Ecol_517

>CP012378.1:c1908906-1907632 Escherichia coli strain MEM

>CP012376.1:2144844-2146112 Escherichia coli strain GE3

>AE014613.1:c978771-977497 Salmonella enterica subsp. enterica serovar Typhi Ty2

>CP012349.1:c1849487-1848213 Salmonella enterica subsp. enterica serovar Sloterdijk str. ATCC 15791

>CP011018.1:c437987-436719 Escherichia coli strain CI5

>LT904854.1:2870642-2871916 Salmonella enterica subsp. enterica serovar Typhi strain M223

>CP010150.1:c4276508-4275234 Escherichia coli strain D7

>CP022019.1:1273301-1274575 Salmonella enterica subsp. enterica serovar Quebec str. S-1267

>CP023525.1:c1720180-1718906 Cedecea neteri strain FDAARGOS_392

>FM180568.1:c2172882-2171614 Escherichia coli 0127:H6 strain E2348/69

>CP012091.1:c4785386-4784112 Salmonella enterica subsp. enterica serovar Typhi strain PM016/13

>CP012347.1:c1856941-1855667 Salmonella enterica subsp. enterica serovar Pullorum str. ATCC 9120

>AM933173.1:2052521-2053861 Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91

>AM933172.1:2061363-2062637 Salmonella enterica subsp. enterica serovar Enteritidis str. P125109

>CP001144.1:2188380-2189654 Salmonella enterica subsp. enterica serovar Dublin str. CT_02021853

>CP019416.1:2061702-2062976 Salmonella enterica subsp. enterica serovar Nitra strain S-1687

>CP022273.1:c5193700-5192426 Citrobacter freundii strain 18-1

>CP007483.2:c1900930-1899659 Salmonella enterica subsp. enterica serovar Anatum str. USDA-ARS-USMARC-1175

>CP023346.1:2221844-2223115 Escherichia coli strain ETEC-2265

>CP013112.1:1592022-1593296 Escherichia coli strain YD786

>CP012487.1:c1917234-1915960 Enterobacter sp. FY-07

>CP006918.1:c1903820-1902546 Klebsiella pneumoniae 30684/NJST258_2

>CP015750.1:59853-61130 Pectobacterium wasabiae CFBP 3304

>BX950851.1:3236381-3237667 Pectobacterium atrosepticum SCRI1043

>CP012635.1:c1941415-1940144 Escherichia coli strain SF-088

>CP021950.1:c5170217-5168943 Klebsiella pneumoniae strain AR_0148

>AP013063.1:2720507-2721781 Serratia marcescens SM39

>CP011642.1:c4518665-4517391 Serratia marcescens strain CAV1492

>CP011602.1:c1111277-1110003 Kluyvera intermedia strain CAV1151

>CU928162.2:2237821-2239095 Escherichia coli ED1a chromosome

>CP003785.1:c1952189-1950915 Klebsiella pneumoniae subsp. pneumoniae 1084

>CP009801.1:c3911165-3909894 Yersinia intermedia strain Y228

>CP001790.1:c1876258-1874981 Pectobacterium parmentieri WPP163

>CP013338.1:c1916947-1915673 Raoultella ornithinolytica strain Yangling I2

>CP022154.1:3112494-3113768 Escherichia coli strain ABWA45

>CP009787.1:c153204-151906 Yersinia rohdei strain YRA

>CP015749.1:2241234-2242511 Pectobacterium wasabiae strain RNS08.42.1A

>CP009769.1:c1733890-1732613 Pectobacterium carotovorum subsp. brasiliense strain BC1

>CP015749.1:2322845-2324122 Pectobacterium parmentieri strain RNS08.42.1A

>CP001790.1:c1909931-1908654 Pectobacterium parmentieri WPP163

>CP009125.1:c1865896-1864619 Pectobacterium atrosepticum strain 21A

>CP003415.1:c1821846-1820572 Pectobacterium sp. SCC3193

>CP021894.1:c1734851-1733577 Pectobacterium carotovorum strain SCC1

>CP020358.1:1490733-1492007 Klebsiella oxytoca strain AR_0147

>BX950851.1:c1926959-1925682 Pectobacterium atrosepticum SCRI1043

>AL513382.1:c4542913-4541654 Salmonella enterica subsp. enterica serovar Typhi str. CT18

>NC_001609.1:c3926-2607 Enterobacteria phage P4

3. THE EARL GENOMIC ISLANDS ARE HIGHLY AT-RICH ISLANDS THAT CARRY HOMOLOGS OF THE XENOGENEIC SILENCER H-NS

[*Genes*. 2020;11(2):142.]

Review

Horizontally acquired homologs of xenogeneic silencers: modulators of gene expression encoded by plasmids, phages and genomic islands

Alejandro Piña-Iturbe[1], Isidora D. Suazo[1], Guillermo Hoppe-Elsholz[1], Diego Ulloa-Allendes[1], Pablo A. González[1], Alexis M. Kalergis[1,2], Susan M. Bueno[1,*]

[1]Millennium Institute on Immunology and Immunotherapy, Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile

[2]Departamento de Endocrinología, Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile

*****Correspondence: sbueno@bio.puc.cl

Abstract

Acquisition of mobile elements by horizontal gene transfer can play a major role in bacterial adaptation and genome evolution by providing traits that contribute to bacterial fitness. However, gaining foreign DNA can also impose significant fitness costs to the host bacteria and can even produce detrimental effects. The efficiency of horizontal acquisition of DNA is thought to be improved by the activity of xenogeneic silencers. These molecules are a functionally related group of proteins that possess affinity for the acquired DNA. Binding of xenogeneic silencers suppresses the otherwise uncontrolled expression of genes from the newly acquired nucleic acid, facilitating their integration to the bacterial regulatory networks. Even when the genes encoding for xenogeneic silencers are part of the core genome, homologs encoded by horizontally acquired elements have also been identified and studied. In this article we discuss the current knowledge about horizontally acquired xenogeneic silencer homologs, focusing on those encoded by genomic islands, highlighting their distribution and the major traits that allow these proteins to become part of the host regulatory networks.

Keywords: Genomic islands; xenogeneic silencers; horizontal gene transfer; transcriptional network; H-NS; MvaT; Lsr2; Rok

1. Introduction

Horizontal gene transfer (HGT) refers to the natural process of transmission of genetic material to an organism in addition to the inheritance of genes from parents to the offspring (Soucy et al., 2015). The availability of an increasing number of sequenced bacterial genomes has revealed the great extension of HGT in these organisms, which highlights the major role played by this process in bacterial adaptation to the environment and to the hosts in which they thrive (Jeong et al., 2016). Several functions, such as those related to niche colonization, symbiotic relationships, catabolism of new nutrients, antimicrobial resistance, and pathogenesis, can be acquired through the DNA gained by HGT (Dobrindt et al., 2004; Darmon and Leach, 2014; Paquola et al., 2018). However, acquisition of some genetic elements can impose fitness costs to the recipient bacteria (Platt et al., 2012; San Millan et al., 2015) likely as a result of the additional DNA that requires replication and repair, in addition to the eventual uncontrolled expression that the newly acquired genes might undergo (Lucchini et al., 2006; Navarre et al., 2006; Lamberte et al., 2017). In the past decades, a group of proteins denominated xenogeneic silencers (XSs) was shown to target the horizontally acquired DNA, repressing the expression and contributing to the integration of the horizontally acquired genes into the host transcriptional network (Navarre et al., 2007; Dillon and Dorman, 2010; Will et al., 2014; Singh et al., 2016).

While XSs are encoded by the core genome (Navarre et al., 2007; Gordon et al., 2010; Yun et al., 2015), homologs encoded by horizontally acquired DNA have also been discovered and characterized (Dorman, 2014b). In recent years, new findings have shed light about the role played by these XS homologs and several more have been identified in numerous mobile elements, which highlights the extension of their presence. In this article, the current knowledge

about the horizontally acquired XS homologs is discussed. After a brief introduction to the main features of the four families of XS proteins, we describe the horizontally acquired homologs and their distribution among bacteria. We also summarize the evidence of the interaction between the horizontally acquired homologs and the core XSs, highlighting the impact produced on the bacterial transcriptional networks. Furthermore, assisted by bioinformatics, we present a more in-depth view about the XS homologs encoded by genomic islands, showing that these proteins are encoded by highly AT-rich islands in *Enterobacteriaceae* and that genes encoding homologs from different XS families are also present in genomic islands. Together, the literature indicates that the horizontally acquired XSs play important roles outside their encoding elements, mostly as a result of the interaction with the core XSs.

2. Xenogeneic silencers

2.1. The H-NS, MvaT, Lsr2 and Rok families of xenogeneic silencers

Xenogeneic silencers are nucleoid-associated proteins (NAPs), a diverse group of proteins involved in condensation of the bacterial chromosome, bending and bridging the DNA to organize it in both micro and macrodomains that can easily fit inside a bacterial cell (Ochman et al., 2000; Helgesen et al., 2016). These proteins are also involved in the regulation of replication and transcription, as well as in the reorganization of the nucleoid to provide accessibility to both the DNA- and RNA-polymerases (Dorman, 2004, 2014a; Scolari et al., 2015; Krogh et al., 2018). Four families of XS proteins have been identified so far, defined by sequence and structure. In Gram-negative bacteria, the H-NS and MvaT families are present among several species of Alpha-, Beta-, and Gamma-proteobacteria, with the second present only in Pseudomonadales (Gamma-proteobacteria). In Gram-positive bacteria, the Lsr2 and Rok

families can be found in Actinobacteria such as *Mycobacterium* spp. and in *Bacillus* spp., respectively (Smits and Grossman, 2010; Navarre, 2016; Duan et al., 2018). Despite their low sequence similarity, these four families of NAPs share some features that are responsible for their ability to target and bind horizontally acquired DNA, as well as to interact with themselves or with homologs, which leads to gene expression silencing (Navarre, 2016).

2.2. Binding to AT-rich DNA and oligomerization are key features for XS function

Recognition of AT-rich DNA (including core promoters and horizontally acquired DNA) and formation of high-order complexes allow H-NS, MvaT, Lsr2, and Rok to act as global transcriptional regulators of hundreds of genes in their respective bacterial hosts, mainly as repressors of gene expression. For this reason, XSs are involved in the regulation of several key functions, such as replication, transcription, translation, chemotaxis, biofilm formation, modulation of diverse biosynthetic pathways, stress responses and virulence, among others (Hoa et al., 2002; Castang and Dove, 2010; Smits and Grossman, 2010; Bartek et al., 2014).

Xenogeneic silencers bind non-specifically to DNA but prefer AT-rich regions, a feature often found in horizontally acquired elements (Marcus et al., 2000; Sherburne et al., 2000; Maeda et al., 2003; Marcoleta et al., 2016; Navarre, 2016; Suzuki-Minakuchi and Navarre, 2019). For years, the basis of this preference remained elusive until nuclear magnetic resonance studies examined the interaction between the *Salmonella* H-NS$_{Ctd}$ and *Burkholderia* Bv3F$_{Ctd}$ (an H-NS homolog) with DNA. These assays showed that a conserved loop in the C-terminal DNA-binding domain contains an AT-hook-like structure defined by the Q/RGR motif that enters the minor groove (Gordon et al., 2011). AT-rich DNA has a narrower and deeper minor groove that probably provides favorable electrostatic interactions with the Q/R and R residues, especially in

sequences that contain contiguous T and A nucleotides (TpA step), which provide the optimal minor groove narrowness and increased DNA flexibility for H-NS binding (Gordon et al., 2011). The TpA step can be found in many -10 elements of core promoters bound by RNA polymerase, which account for the H-NS capacity to silence gene expression (Lang et al., 2007; Landick et al., 2015). While the AT-hook-like motif is also present in Lsr2 (Gordon et al., 2011), the MvaT and Rok proteins use different mechanisms for binding to DNA. In MvaT, the amino acid sidechains that enter the minor groove come from two different loops of the protein, forming an "AT-pincer" (Ding et al., 2015). In Rok, a winged-helix domain binds the DNA along the minor groove (Duan et al., 2018). As for H-NS, the DNA-binding affinity of the other XSs is favored by the TpA step (Lang et al., 2007; Ding et al., 2015; Duan et al., 2018), except for Lsr2, which is insensitive to its presence (Ding et al., 2015). Despite these differences, the DNA-binding domain of the four xenogeneic silencers is always found in their C-terminal domains.

Formation of dimers and oligomers enable cooperative binding of XSs to DNA and the formation of bridges within or between different regions of the chromosome. Both the dimerization and oligomerization domains of H-NS, MvaT, and Lsr2 are in the N-terminal region. In these proteins the dimerization/oligomerization results from the interaction of four α-helices in H-NS, two α-helices in the MvaT-homolog TurB, and one antiparallel β-sheet and one α-helix in Lsr2 (Arold et al., 2010; Summers et al., 2012; Suzuki-Minakuchi et al., 2016). Antiparallel homodimers and heterodimers, as well as higher-order interactions, can be formed as a result of oligomerization, which is important for nucleoid compaction and modulation of gene expression (Johansson et al., 2001; Bhat et al., 2014; Yun et al., 2015). Less is known about the oligomerization of Rok, but recent data show that it can form oligomers through the N-terminal domain (Duan et al., 2018). It is important to mention that truncated versions of the

XS proteins that maintain the N-terminal dimerization domain can exert a dominant-negative effect on the other XS homologs present in the bacterial cell. For instance, in the presence of a null hns mutation, the H-NS paralog StpA can function as a backup that represses genes originally targeted by H-NS (Picker and Wing, 2016); however, if the truncated version of H-NS is present, it forms dimers with StpA, disrupting its function as a result of the absence of the H-NS binding domain (Williams et al., 1996).

Changes in temperature, pH, or growth phase can alleviate the repression exerted by the XSs on AT-rich genes, enabling expression control in response to environmental stimuli (Shahul Hameed et al., 2019), a characteristic which provides a mechanism to integrate horizontally acquired genes into the host transcriptional network. For instance, AT-rich mobile elements can be bound by XSs to repress and avoid deleterious effects resulting from uncontrolled gene expression (Navarre, 2016), then the silencing could be relieved in response to particular environmental stimuli in which the functions encoded by the newly acquired genes result beneficial (Queiroz et al., 2011; Leh et al., 2017). Interestingly, the same features that enable xenogeneic silencers to regulate expression of foreign genes also enable their horizontally acquired homologs to interact with the host transcriptional network (see the next Section).

3. Horizontally acquired xenogeneic silencers

Several genes encoding members of the four families of XSs have been identified in hundreds of different horizontally acquired elements, specifically in plasmids, bacteriophages and genomic islands, which highlights their widespread distribution (Table 1) (Skennerton et al., 2011; Shintani et al., 2015; Piña-Iturbe et al., 2018).

3.1. Xenogeneic silencers encoded by plasmids

The most extensively studied plasmid-encoded homologs are Sfh, Pmr and Acr2. The H-NS family protein Sfh is encoded on the R27-like plasmids associated with antibiotic resistance in strains of *Salmonella* serovars Typhimurium and Typhi, and on the 99.7% identical plasmid pSf-R27 of *Shigella flexneri* serotype 2a strain 2457T, which lacks the antibiotic resistance genes (Beloin et al., 2003; Forns et al., 2005; Doyle et al., 2007). Pmr is an MvaT homolog encoded on the pCAR1 plasmid, which confers carbazole-degrading capacity to *Pseudomonas putida* KT2440 (Miyakoshi et al., 2007; Yun et al., 2010; Singh et al., 2016). Acr2 is an H-NS family protein which acts as a negative regulator of conjugative transfer of plasmids belonging to the A/C incompatibility group, which are responsible for the spreading of antibiotic resistance in several species of Gammaproteobacteria (Fernández-Alarcón et al., 2011; Carraro et al., 2014; Shintani et al., 2015; Lang and Johnson, 2016). Acquisition of the plasmids that encode these XS homologs produces transcriptional alterations in the recipient bacteria, however, different changes in the expression of chromosomal and plasmid-encoded genes take place upon acquisition of plasmid variants lacking the corresponding XS-encoding gene, indicating a role of the plasmid-encoded homologs in regulation of gene expression both on and off of their mobile element.

Table 1. Horizontally acquired xenogeneic silencer homologs found in genomic islands, plasmids and bacteriophages.

XS homolog	Extent of similarity (with XS)	Host	Location	Protein-protein interaction with[1]	References
Hfp/H-NSB	full-length (H-NS)	Family *Enterobacteriaceae*	*serU* island and EARL GIs	H-NS	(Williamson and Free, 2005; Müller et al., 2010; Piña-Iturbe et al., 2018)
H-NST	N-terminal (H-NS)	Family *Enterobacteriaceae*	*serU* island and EARL GIs	H-NS	(Williamson and Free, 2005; Piña-Iturbe et al., 2018)
Ler	C-terminal (H-NS)	Attaching and effacing pathogens (*Enterobacteriaceae*)	Locus of Enterocyte Effacement GI	Ler	(Mellies et al., 2011)
Pmr	full-length (MvaT)	*Pseudomonas putida* strain KT2240	pCAR1 plasmid	Pmr, TurA, TurB, TurE	(Yun et al., 2010)
Acr2	full-length (H-NS)	Class Gammaproteobacteria	IncA/C plasmids	Unknown	(Shintani et al., 2015; Lang and Johnson, 2016)
Sfh	full-length (H-NS)	*Shigella flexneri* 2a strain 2457T	pSf-R27 and R27-like plasmids	H-NS, StpA, Sfh	(Beloin et al., 2003; Deighan et al., 2003; Forns et al., 2005; Doyle et al., 2007; Picker and Wing, 2016)
$H\text{-}NS_{EPV1}$	full-length (H-NS)	"*Candidatus* Accumulibacter phosphatis"	Phage EPV1	Unknown	(Skennerton et al., 2011; Flowers et al., 2013)
Rok_{LS20}	C-terminal (Rok)	*Bacillus subtilis* strain IFO3335	pLS20 plasmid	Unknown	(Singh et al., 2012)
Lsr2 homologs	full-length (Lsr2)	Phylum Actinobacteria	Plasmids (unclassified) and mycobacteriophages	Unknown	(Shintani et al., 2015; Pfeifer et al., 2019)

[1] Protein-protein interaction with other chromosomally encoded XS homologs or the horizontally acquired XS homolog itself.

Introduction of the pSf-R27 plasmid in *Salmonella* ser. Typhimurium SL1344 results in altered transcription of several genes related to different functional categories. However, when the Δ*sfh* variant was introduced, an increase in the expression level and the number of gene categories affected was observed along with changes in motility and virulence which resembled those caused by a dominant-negative *hns* allele (Doyle et al., 2007). Similarly, the acquisition of the plasmid pCAR1 altered the log-phase transcriptional profile of *P. putida* KT2440, modifying the expression of 112 genes, in which both up- and down-regulation were observed (Yun et al., 2010). Interestingly, disruption of *pmr* on pCAR1 altered the transcription of 140 additional genes present in the KT2440 chromosome, including horizontally acquired genes (Yun et al., 2010). The extent of the effect produced by the absence of Sfh and Pmr, is likely a result of their structural similarities to the xenogeneic silencers H-NS and MvaT which allow these homologs to oligomerize and bind AT-rich DNA. The Sfh protein has the capacity to form heterodimers with H-NS and its paralogue StpA (Deighan et al., 2003; Picker and Wing, 2016), and the occupancy of Sfh expands to include 645 of the 745 H-NS-unique target genes in a dominant-negative *hns* background (Dillon et al., 2010). Likewise, Pmr can interact with other MvaT family proteins present in the KT2440 chromosome, i.e. the TurA and TurB proteins, which possess overlapping binding sites with Pmr (Yun et al., 2010, 2015; Sun et al., 2018). Less is known about the Acr2 protein compared with Sfh and Pmr. However, recent RNA-Seq experiments carried out with an *Escherichia coli* DH10B strain that carries the pAR060302-derivative plasmid pARΔ*acr2* (modified to eliminate the antibiotic resistance determinants and the Acr2-coding gene) have identified several plasmid and chromosomal genes with altered expression resulting from the *acr2* deletion (Lang and Johnson, 2016). Nevertheless, further chromatin affinity purification (ChAP-Seq) assays showed that only three chromosomal genes

with altered expression were bound by a 6xHis-tagged Acr2 (Lang and Johnson, 2016). It remains to be addressed whether Acr2 can interact with H-NS or StpA. It is worth mentioning that, although Dillon *et al.* (Dillon et al., 2010) observed a complete inclusion of the Sfh-targeted genes within the H-NS regulon, Doyle *et al*. (Doyle et al., 2007) found that the Δ*sfh* mutation altered the expression of several genes not recognized as regulated by H-NS (Navarre, 2010). The findings of Doyle *et al*. suggest that Sfh, and perhaps other XS homologs, might be playing regulatory roles independently of H-NS and other XSs.

Lsr2 and Rok homologs have also been identified in plasmids harbored by members of the phylum Actinobacteria and *Bacillus subtilis*, respectively (Singh et al., 2012; Shintani et al., 2015). It was found that a short homolog of Rok, encoded by the conjugative plasmid pLS20 of *Bacillus subtilis* strain IFO3335, also serves as a repressor of the master regulator of competence ComK like the full-length Rok encoded by the chromosome (Hoa et al., 2002; Singh et al., 2012). Overexpression of Rok_{LS20} in *B. subtilis* strain 168 significantly decreases the transformation efficiency of the strain and reduces gene expression from the *comK* promoter as a result of the Rok_{LS20} binding to this regulatory region. Interestingly, the Rok_{LS20} protein , which corresponds to the C-terminal DNA-binding domain (Duan et al., 2018), can complement the absence of Rok at least in the competence pathway of *B. subtilis* 168 (Singh et al., 2012). Although it was found that Rok_{LS20} likely binds several regions of the *B. subtilis* chromosome (Singh et al., 2012), experiments addressing the function of Rok_{LS20} in other loci, or its interaction with chromosomal Rok are still lacking. Regarding the plasmid-encoded Lsr2 homologs, to the best of our knowledge, there are no published data about their function.

3.2. Xenogeneic silencers encoded by bacteriophages

The increasing number of sequenced bacteriophage genomes has unveiled the presence of phage-encoded homologs of H-NS and Lsr2. Metagenomic analyses of phage-enriched samples from an enhanced biological phosphorus removal (EBPR) bioreactor lead to the identification of the EPV1 virus, a parasite of "*Candidatus* Accumulibacter phosphatis" (CAP), which is a member of the community that carries out the EBPR process in that bioreactor (Skennerton et al., 2011; Flowers et al., 2013). The genome of this virus harbors a homolog of the *hns* gene closely related to the *hns* encoded in the CAP chromosome (Skennerton et al., 2011). Although experimental evidence is still lacking, prediction of H-NS binding sites and the identification of low GC regions in the CAP genome led the authors to propose the hypothesis that $H\text{-}NS_{EPV1}$ could participate in the modulation of the CRISPR and/or the Type-III restriction-modification systems encoded by CAP, which are known phage-defense systems (Skennerton et al., 2011).

Homologs of the Lsr2 protein have also been identified in genomes of the phages Cjw1, 244, Porky, Kostya and Omega, which infect *Mycobacterium smegmatis* strain mc^2155 (Pedulla et al., 2003; Hatfull et al., 2010), and in the CGP3 prophage of *Corynebacterium glutamicum* strain ATCC 13032 (Pfeifer et al., 2016). In the latter, ChAP-Seq experiments found that a Strep-tagged version of the CGP3-encoded homolog CgpS bound preferentially to AT-rich regions mainly at the CGP3 prophage, but also in other chromosomal regions likely acquired by HGT. When a truncated version of CgpS, spanning the N-terminal region, was overexpressed in *C. glutamicum*, derepression of several CGP3-encoded genes and the induction of this prophage was observed (Pfeifer et al., 2016). The effect of the truncated version of CGP3

underlines the importance of dimerization/oligomerization in the activity of xenogeneic silencers and their horizontally acquired homologs. A recent review article reported that horizontally acquired Lsr2 homologs are encoded by a great number of other actinobacteriophages sequenced to date, whose hosts belong to genera *Mycobacterium*, *Microbacterium*, *Gordonia* and *Streptomyces* (Pfeifer et al., 2019). As described by the authors of that review, the finding that CgpS is essential to *C. glutamicum* only when CGP3 was present (Pfeifer et al., 2016), together with the relatively high frequency of Lsr2 homologs ($Lsr2_{Actinophage}$) in lysogenic versus lytic phages, suggests that these XS homologs play a major role in the integration of prophages in bacteria (Pfeifer et al., 2019). Nevertheless, CgpS can also bind and modulate the expression of genes outside the CGP3 prophage in *C. glutamicum* (Pfeifer et al., 2016).

3.3. Xenogeneic silencers encoded by genomic islands

3.3.1. Genomic islands

Genomic islands (GIs) are horizontally transferred genetic elements of about 10 to 500 kbp, that can integrate in bacterial chromosomes providing their hosts with advantageous traits (Soucy et al., 2015), such as new metabolic functions, resistance to antibiotics, or virulence factors, among others, which often improve the strain overall fitness (Sullivan and Ronson, 1998; Hacker and Kaper, 2000; Dobrindt et al., 2004; Soucy et al., 2015). Usually characterized by a different GC content, codon usage bias and dinucleotide frequency relative to their host chromosome (Che et al., 2014), GIs are often found at the 3'-end of genes encoding tRNAs and tmRNAs (Williams, 2002); nevertheless, different GI families may prefer other genes as integration sites (Wozniak et al., 2009; Daccord et al., 2013; Coluzzi et al., 2017). Under certain

conditions, GIs can be excised from the chromosome through the site-specific recombination between the direct repeated sequences (DRSs) that flank the element, in a reaction catalyzed by the integrase protein encoded within or outside the island (Manson and Gilmore, 2006; Murphy and Boyd, 2008; Quiroz et al., 2011; Lautner et al., 2013). This reaction, which is usually promoted by a recombination directionality factor also encoded inside or outside the GI, results in the formation of a circular element (Carpenter et al., 2016; Haskett et al., 2018). The insertion site is reconstituted in the chromosome, and a copy remains as part of the excised GI. These sequences can subsequently take part in re-integration of the GI into the host chromosome (Sentchilo et al., 2009). Besides re-integration, GIs in their circular form can be transferred from one cell to another by means of transduction by co-resident prophages (Penadés and Christie, 2015) or by conjugation either using their own transfer machinery or taking advantage of the conjugation system encoded by a self-transmissible element (Doublet et al., 2005; Haskett et al., 2016).

Full-length and short homologs belonging to the H-NS family have been identified in genomic islands. While full-length homologs comprise the dimerization/oligomerization and DNA-binding domains of H-NS, the short ones share similarity to one domain only, a feature that provides the latter with anti-H-NS properties that can relieve the H-NS-mediated silencing. Again, similar to what was observed for XS-homologs encoded by plasmids and bacteriophages, the dimerization and DNA-binding capacity of the GI-encoded XSs allow these proteins to regulate the expression of genes located outside their encoding GI.

3.3.2. H-NST, Ler and Hfp (H-NSB)

First reported in 2005, H-NST is a short homolog of the H-NS protein encoded in the so-called *serU* island, a 22.5 kbp pathogenicity island harbored by different strains of pathogenic *E. coli*, such as the uropathogenic (UPEC) strains CFT073 and 536, and also in IE3, a 25.8 kbp genomic island found in the enteropathogenic *E. coli* (EPEC) strain E2348/69 (Williamson and Free, 2005). H-NST corresponds to the dimerization/oligomerization domain of H-NS and behaves as an antagonist of this protein, as observed in experiments where the activity of the H-NS-repressed promoters of *proU* and *bgl* fused to *lacZ,* increased in the presence of a low-level expression of H-NST$_{EPEC}$ (Williamson and Free, 2005). The antagonistic effect is most likely caused by the heterodimers that H-NS and H-NST can form, which would have an altered DNA-binding capacity due to the absence of an H-NS-like DNA-binding domain in the H-NST component (Williamson and Free, 2005). H-NST is also able to increase the expression of genes encoded in the Locus of Enterocyte Effacement (LEE), a genomic island responsible of the attaching/effacing lesions caused by EPEC, enterohemorrhagic *E. coli* (EHEC), *E. albertii* and *Citrobacter rodentium* (Elliott et al., 1998; Deng et al., 2001; Levine et al., 2014; Ooka et al., 2015). An increase of the amount of the EspA and EspB proteins (exported by the type-III secretion system encoded in LEE) was observed only when H-NST was expressed from a plasmid in the EHEC strain TUV93-0, which lacks the *serU* island, but not when expressed in the EPEC strain E2348/69, which already had high basal expression levels of the LEE-encoded proteins (Levine et al., 2014). Whether the high basal expression of the LEE-encoded proteins in EPEC E2348/69 is linked to the presence of the *serU*-encoded H-NST is unknown. Interestingly, it was found through electrophoretic mobility shift assays that H-NST had an intrinsic DNA-binding capacity on two regulatory regions encoded in the LEE island (Levine

et al., 2014). It is believed that H-NST-binding to these regions might help Ler, another H-NS homolog, to displace the bound H-NS oligomer.

Ler is encoded in the LEE pathogenicity island and, although it has a length comparable to that of H-NS (123 and 137 aa, respectively), the similarities with this XS are restricted to the C-terminal DNA-binding domain (Mellies et al., 1999). Ler is a master activator of gene expression that works by alleviating the H-NS-mediated repression of the type-III secretion system and its effectors encoded within LEE. Upon binding of Ler to DNA in a non-cooperative manner, a displacement of the bound H-NS takes place (Winardhi et al., 2014). However, the role of Ler is not limited to its genomic island and can activate other virulence factors in pathogenic *E. coli*, such as EspC, which regulates the translocation of effector proteins to host cells and pore formation by EPEC (Guignot et al., 2015); and StcE, encoded in the EHEC O157:H7 virulence plasmid, which contributes to adherence to host cells (Elliott et al., 2000; Lathem et al., 2002). A similar interplay between Ler and H-NS was observed to modulate the expression from the promoter of the Long Polar Fimbriae in the EHEC strain EDL933, where H-NS represses expression and Ler acts as an antisilencer (Torres et al., 2007). Although the formation of dimers and high order oligomers of Ler in solution have been observed (Mellies et al., 2011), interaction with H-NS or other xenogeneic silencers has not been reported.

Hfp, also known as H-NSB, is an XS homolog also encoded in the *serU* island but, unlike H-NST, it is a full-length H-NS homolog (Williamson and Free, 2005; Müller et al., 2010). In the UPEC strain 536, Hfp was found to play a role in bacterial growth, autoaggregation, hemolytic activity, and specifically in the downregulation of the S and P fimbriae major subunits, the K15 capsule, and expression in the *bgl* operon, which indicates that Hfp acts in a similar

fashion as H-NS. Nevertheless, these effects were observed only when the *hfp* mutation was accompanied by an *hns* mutation, suggesting a shared role in the regulation of some genes which has been supported by the finding of cross-regulation between Hfp and H-NS and their capacity to form heteromeric complexes (Müller et al., 2010).

3.3.3. The *Enterobacteriaceae*-associated ROD21-like genomic islands

Besides the first report of the occurrence of the H-NS homologs H-NST and H-NSB (Hfp) in the *serU* and IE3 genomic islands made by Williamson and Free in 2005 (Williamson and Free, 2005), the assembly of the complete genome sequence of *Salmonella* ser. Enteritidis strain P125109 by Thomson *et al.* in 2008, revealed the presence of other genomic islands harboring H-NS homologs. The authors identified several regions present in the chromosome of the strain P125109 that were absent from *Salmonella* ser. Typhimurium strain LT2 which were denominated Regions of Difference (Thomson et al., 2008). Among these horizontally transferred regions, the Region of Difference 21 (ROD21), which includes the genes *SEN1970* to *SEN1999*, was found to encode a putative homolog of H-NS (*SEN1993*). Moreover, the authors showed that other GIs related to ROD21, and found in *Photorhabdus luminescens* strain TTO1, *Pectobacterium atrosepticum* strain SCR1043 and UPEC strain CFT073, also encoded H-NS homologs.

ROD21 is a 26.5 kbp pathogenicity island inserted in the 3-end of the Asn-tRNA-encoding gene *asnW* of the global *Salmonella* ser. Enteritidis epidemic strains and in the serovars Gallinarum, Dublin and Nitra (Thomson et al., 2008; Feasey et al., 2016; Piña-Iturbe et al., 2018). This GI encodes TlpA, a TIR-domain-containing protein required for the intracellular survival in THP-1 macrophages and the efficient colonization of the murine spleen

(Newman et al., 2006). Other putative virulence factors likely involved in the colonization of bird and murine internal organs (liver and spleen) are also encoded in ROD21 (Coward et al., 2012; Silva et al., 2012). ROD21 is also an excisable island, a feature that has reached a special relevance due to its role in the virulence of *Salmonella* ser. Enteritidis since different mutant strains, with a reduced excision capacity, show a reduced colonization of the liver and spleen of infected mice (Quiroz et al., 2011; Pardo-Roa et al., 2019). The DRSs that flank ROD21 participate in the site-specific recombination reaction that excises the island and produces a circular form of the element (Quiroz et al., 2011; Salazar-Echegarai et al., 2014). The excision process is likely promoted by the products of the genes *SEN1970* and *SEN1998*, which are predicted to encode a tyrosine recombinase and a putative recombination directionality factor, respectively. Indeed, compared with the wild-type strain, the fraction of bacteria with the excised island is reduced in the Δ*SEN1970*::*FRT* and the Δ*SEN1970*::*FRT* Δ*SEN1998*::*FRT* population (Tobar et al., 2013; Piña-Iturbe et al., 2018; Pardo-Roa et al., 2019).

The location of *SEN1970*, downstream of the insertion site in *asnW*, allowed the sequences spanning the corresponding DRS, the *SEN1970* promoter and the first 82 nucleotides of *SEN1970* to be used to search for genomic islands using BLASTn against the GenBank non-redundant database. This approach identified several genomic islands, phylogenetically related to ROD21, in different species belonging to the family *Enterobacteriaceae* including plant- and animal-pathogenic strains of *Pectobacterium* spp., *Serratia marcescens*, intestinal and extraintestinal *E. coli*, *Enterobacter* sp., carbapenem-resistant *Klebsiella pneumoniae* ST258 and different *Salmonella* serovars, among others (Piña-Iturbe et al., 2018). Since ROD21 is the most studied member of this group, it was denominated the *Enterobacteriaceae*-associated ROD21-like (EARL) family of genomic islands. All these GIs share, among other features, the

location in an Asn-tRNA-encoding gene and the excision/integration module (the DRSs, and the integrase- and putative RDF-encoding genes), characteristics which allow their excision (Bueno et al., 2004; Carpenter et al., 2016; Piña-Iturbe et al., 2018). Other genes, such as those encoding putative type-4 pilus-related proteins, relaxases and type-III restriction-modification systems (type-III R-M), are conserved only in closely related subgroups within the EARL family likely as a result of their acquisition by an EARL GI followed by the spreading of the island by HGT. The comparative analysis of these GIs also revealed that most of them also have genes encoding full-length and short homologs of the H-NS protein. Indeed, the islands previously reported by Thomson *et al.* and Williamson and Free as carriers of *hns* homologs, namely HAI7 and HAI13 (*P. atrosepticum* strain SCRI1043) and IE3 (EPEC E2348/69) belong to the EARL family (Williamson and Free, 2005; Thomson et al., 2008; Piña-Iturbe et al., 2018). Although the *serU* island is also related to ROD21 and encodes H-NSB (Hfp) and H-NST_{UPEC} (Williamson and Free, 2005; Müller et al., 2010), it possesses a different integrase and, therefore, a different integration site. This GI is likely a derivative from IE3, the EARL GI from EPEC E2348/69 which encodes two integrases, the one present in all others EARL islands, and the one present in the *serU* island (Williamson and Free, 2005).

The EARL GI-encoded full-length and short homologs of H-NS are homologs of H-NSB (Hfp) and H-NST, but represent two different and distantly related clades (Piña-Iturbe et al., 2018), and, henceforth, they will be denoted as H-NSB_{EARL} and H-NST_{EARL}. The full-length homologs share many similarities at the amino acid sequence level with the chromosomal H-NS, including the region interacting with the small nucleoid-associated protein Hha (the Hha signature (García et al., 2006; Madrid et al., 2007; Ali et al., 2013)), the residues that enters the DNA minor groove for DNA-binding (the Q/RGR motif (Gordon et al., 2011)) and key amino

acids at the linker domain also important for DNA-binding (Gao et al., 2017) (Fig. 1A). Since the similarity between H-NS and the H-NSB$_{EARL}$ proteins spans the entire sequence, it is most likely that the secondary and tertiary structures will also be conserved, as suggested by the capacity of the H-NSB protein encoded in the *serU* island to form heterodimers with H-NS and participate in the regulation of known H-NS targets, including virulence factors (Müller et al., 2010). Because of the different accessory gene pools present in different species, the regulated or co-regulated genes by the H-NSB$_{EARL}$ proteins are expected to also be different. This represents a subject of further research aimed at better understanding the regulation of virulence in bacterial pathogens. For example, the type-III secretion system 1 (T3SS-1) encoded in the *Salmonella* pathogenicity island 1 (SPI-1), is a key virulence factor that, by translocating effector proteins, allows the invasion the host cells (Desai et al., 2013). Recent findings show that the *Salmonella* ser. Enteritidis Δ*SEN1970*::*FRT* Δ*SEN1998*::*FRT* mutant, which have a reduced expression of several ROD21-encoded genes, also have a significantly increased expression of *invA*, a gene encoding a structural component of the T3SS-1 (Pardo-Roa et al., 2019). *invA*, as well as other SPI-1-encoded genes, are upregulated by proteins also encoded by SPI-1 which are, at the same time, negatively regulated by H-NS (Lou et al., 2019), raising the possibility that the observed link between the two pathogenicity islands could be the result of the H-NSB$_{ROD21}$ protein interacting with H-NS or the H-NS binding sites within SPI-1. As another example, the type-3 pili and the capsular polysaccharide are known H-NS-regulated virulence factors of *K. pneumoniae* (Ares et al., 2016). Since the ICEKp258.2 EARL island of the globally spread carbapenem-resistant *K. pneumoniae* ST258 encodes an H-NSB$_{EARL}$ protein (Fig 1A; (Piña-Iturbe et al., 2018)), this GI-encoded homolog might be playing a role in the virulence of ST258 as well.

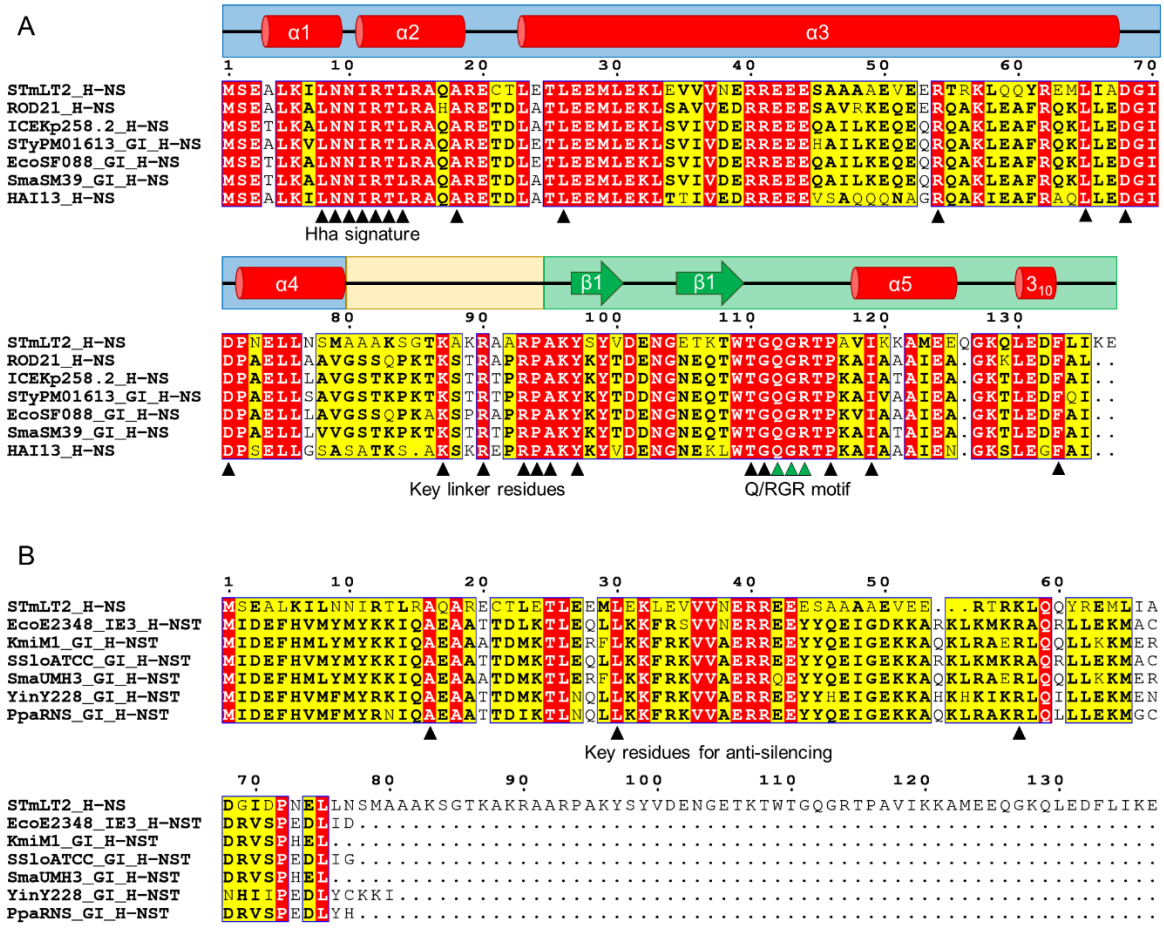

A
α1
α2
α3
STmLT2_H-NS
ROD21_H-NS
ICEKp258.2_H-NS
STyPM01613_GI_H-NS
EcoSF088_GI_H-NS
SmaSM39_GI_H-NS
HAI13_H-NS
Hha signature
α4
β1
β1
α5
3 10
Key linker residues
Q/RGR motif
B
STmLT2_H-NS
EcoE2348_IE3_H-NST
KmiM1_GI_H-NST
SSloATCC_GI_H-NST
SmaUMH3_GI_H-NST
YinY228_GI_H-NST
PpaRNS_GI_H-NST
Key residues for anti-silencing

Figure 1. Multiple sequence alignment of the H-NS protein from *Salmonella* ser. Typhimurium LT2 and selected full-length (A) and short (B) homologs found in EARL genomic islands. The secondary structure elements of H-NS are represented with red cylinders (α-helices) and green arrows (β-strands). The blue, yellow and green background represent the dimerization/oligomerization, linker and DNA-binding domains respectively. Numbers indicate the corresponding residue of the H-NS protein. The triangles under the alignment indicate the amino acid residues required for function of H-NS (A) or H-NST (B) that are also conserved in the homologs encoded by the EARL GIs. Aligned proteins correspond to H-NS from *Salmonella* ser. Typhimurium LT2 (STmLT2; WP_001287383.1), full-length homologs encoded in the EARL islands from *Salmonella* ser. Enteritidis P125109 (ROD21; WP_001287371.1), *Klebsiella pneumoniae* ST258 NJST258-2 (ICEKp258.2; WP_001588074.1), *Salmonella* ser. Typhi PM016/13 (StyPM01613_GI; WP_045354302.1), *Escherichia coli* SF-088 (EcoSF088_GI; WP_000005143.1), *Serratia marcescens* SM39 (SmaSM39_GI; WP_041035854.1), *Pectobacterium atrosepticum* SCRI1043 (HAI13; WP_011094424.1), and the short homologs encoded in the EARL islands from *E. coli* O127:H6 E2348/69 (EcoE2348_IE3; WP_000564595.1), *K. michiganensis* M1(KmiM1_GI; WP_038424693.1), *Samonella* ser. Sloterdijk ATCC 15791 (SsloATCC_GI; WP_023201852.1), *S. marcescens* UMH3 (SmaUMH3_GI; WP_089187391.1), *Yersinia intermedia* Y228 (YinY228_GI; WP_042569548.1) and *P. parmentieri* RNS08.42.1A (PpaRNS_GI; WP_033071994.1). The multiple alignment was made in MEGA X (Kumar et al., 2018) using MUSCLE, and the graphic representation was made using ESPript3 (Robert and Gouet, 2014). The complete alignment of all 34 full-length and 20 short homologs found in EARL islands is provided as supplementary Figures S1 and S2.

Compared with the H-NSB$_{EARL}$ proteins, the short homologs are less related to H-NS, having similarity with the dimerization/oligomerization domain only (Fig. 1B), a feature which allows the H-NST$_{EARL}$ proteins to exert a dominant-negative effect that relieves the H-NS-mediated silencing [Section 3.3.2; (Williamson and Free, 2005; Levine et al., 2014)]. These proteins have sequence identities ranging from 56.6% to 98.8% and share many conserved amino acids some of which are important for the activity of the H-NST$_{EPEC}$ protein encoded in the IE3 island from EPEC E2348/69 (Fig. 1B). For example, A16, L30 and R60 (positions according to the H-NST$_{EPEC}$ sequence), which are important for the antisilencing and DNA-binding capacity of H-NST$_{EPEC}$ (Williamson and Free, 2005; Levine et al., 2014), are also conserved in the other H-NST$_{EARL}$ proteins (Fig. 1B). R63 was shown to be important for the DNA-binding capacity of the IE3-encoded protein (Levine et al., 2014), however, in the other H-NST$_{EARL}$ proteins this position is occupied by different amino acids, the most frequent being K and Q, as previously observed by Levine *et al.* (Levine et al., 2014). It is possible that an H-NS-H-NST interplay, similar to that observed in the LEE island, might be regulating other H-NS-controlled virulence factors in the other enterobacterial pathogens carrying an EARL GI.

Genomic islands usually have a lower GC content than the average of their host chromosome (Zhang and Zhang, 2004b, 2004a; Murphy and Boyd, 2008; Marcoleta et al., 2016), a feature also present in the GIs harbored by different species of the family *Enterobacteriaceae* (Fig. 2A). Remarkably, the GC content of the EARL islands is significantly lower than the median value of the other *Enterobacteriaceae* GIs (Fig. 2A), remaining low even when the host chromosomal GC increases (Fig 2B). This feature agrees with what has been reported for plasmids and actinobacteriophages that encode homologs of H-NS and Lsr2, respectively, which have a lower GC content compared with those which do not encode an XS homolog (Pfeifer et

al., 2019). The low GC of the EARL islands and its narrow range of variability (35.7-39.2%) could be the result of their relatively rapid spread within *Enterobacteriaceae*, an idea supported by the fact that ROD21 can be transferred by conjugation (Salazar-Echegarai et al., 2014) and that the ICEKp258.2 island from *K. pneumoniae* ST258, which may represent an early member of the SpnT/type-3 R-M-encoding clade within the EARL phylogeny (Piña-Iturbe et al., 2018), was acquired by this sequence type approximately 20-25 years ago (Bowers et al., 2015; Marsh et al., 2019; Zhang et al., 2019). Since the XSs show preference for AT-rich DNA, we speculate that these elements have a selective pressure to acquire genes encoding factors that could interact with the silencing effectors from the host cell in order to relieve silencing and provide the opportunity to be incorporated in the host regulatory network.

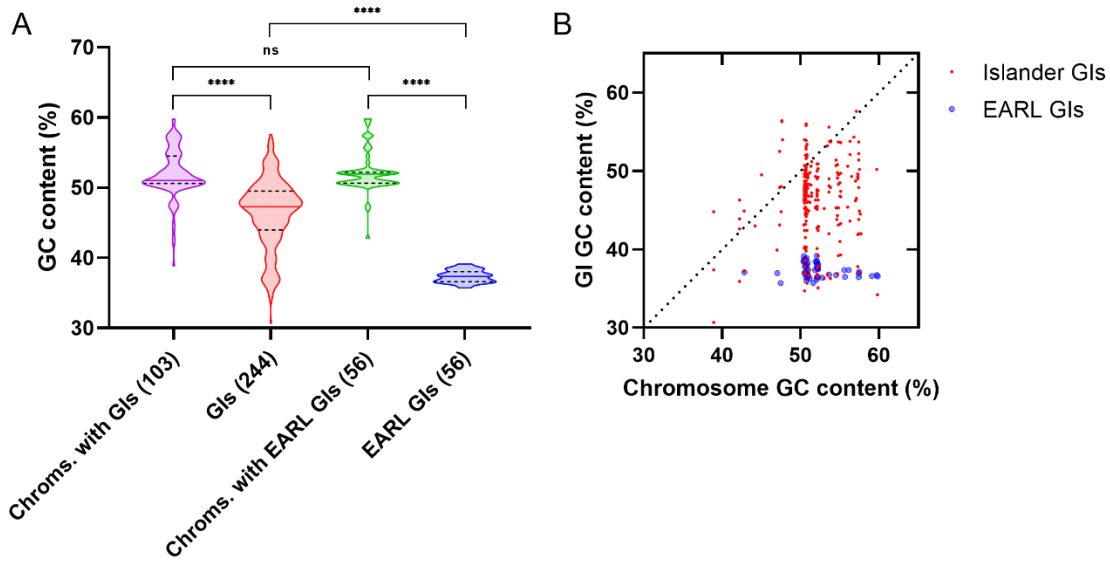


Figure 2. The EARL genomic islands possess a significantly low GC content. (A) Violin plots of the GC content of the genomic islands found in *Enterobacteriaceae* in the Islander database (GIs), their host chromosomes (Chroms. with GIs), the EARL genomic islands (EARL GIs) and the chromosomes that harbor the EARL GIs (Chroms. with EARL GIs). Kruskall-Wallis test followed by Dunn's multiple comparisons were used to assess differences between medians (α=0.05, ****p<0.0001). (B) Comparison of the GC content of genomic islands and the host chromosome. The dotted line represents a 1:1 correspondence. Data was obtained from the Islander database (15/10/19) (Hudson et al., 2015) and the supplementary information from Piña-Iturbe *et al.* (2018). The 597 *Enterobacteriaceae* genomic islands stored in Islander were manually filtered to eliminate possible false positives (indicated in the database), putative prophages (>20% overlap with a PHAST call), duplicated genomic islands (islands with length ≥ 300 pb in the same species, the same integration site and GC difference < 1%), and islands found in plasmids, resulting in 244 genomic islands present in 103 host chromosomes. All the 56 EARL islands identified by Piña-Iturbe et al. (2018) were used.

3.3.4. Genomic islands encode XS homologs from different families

To assess whether GIs encode homologs that belong to the other families of xenogeneic silencers, a tBLASTn search against the entire Islander database of genomic islands (4065 islands) was conducted followed by manual examination of the resulting hits (Fig. 3; Supplementary Table S1). Surprisingly, only 29 genomic islands were found to encode XS homologs, most belonging to the H-NS family, followed by the MvaT and Lsr2 families. No Rok homolog was detected. In agreement with the distribution of the chromosomal XSs among different bacterial taxa (see Section 2.1), the H-NS homologs were found in α-, β-, and γ-proteobacteria; the MvaT homologs in γ-proteobacteria (*Pseudomonadaceae*), and the Lsr2 homologs in Actinobacteria. The different numbers of GIs encoding XS homologs can be in part explained by the overrepresentation of the islands from Proteobacteria versus Actinobacteria (64% and 14% of the GIs from Bacteria), and *Enterobacteriaceae* versus *Pseudomonadaceae* (55% and 9% of γ-proteobacteria) in the Islander database (Hudson et al., 2015). Since *Bacillus* is the only genus in which Rok proteins have been identified, the absence of Rok homologs is likely due to the small number of GIs from *Bacillus* in Islander (46 GIs).

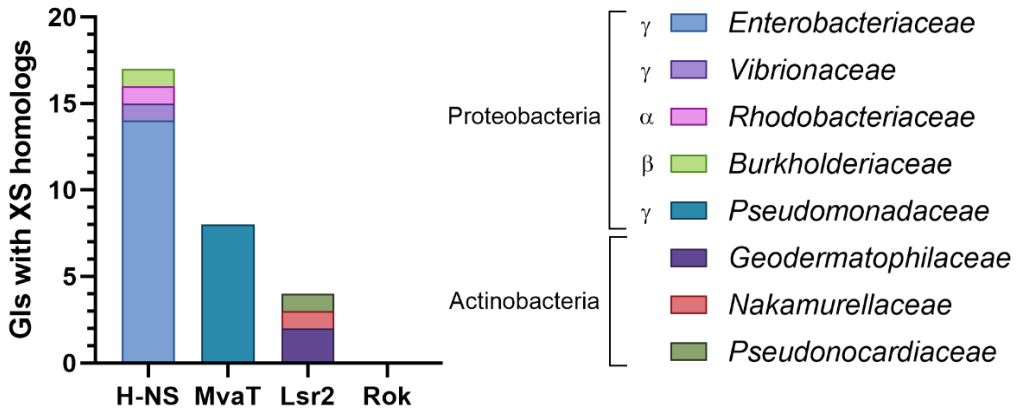


Figure 3. Xenogeneic silencer homologs encoded by genomic islands. The entire Islander database (4065 genomic islands) was interrogated for the presence of xenogeneic silencer homologs using tBLASTn (E-value cutoff=10) using the amino acid sequences of H-NS (WP_001287383.1), MvaT (WP_003093888.1), Lsr2 (WP_003419513.1) and Rok (WP_003232378.1). The genomic islands corresponding to the BLAST hits were manually filtered to exclude possible false positives, putative prophages, duplicated islands, and islands in plasmids using the same criteria as in Figure 2. Then, the individual hits were manually examined to exclude those corresponding to pseudogenes, alignments outside coding sequences, or alignments in a reading frame different of the annotated coding sequence.

3.4. Xenogeneic silencers homologs and the environmental conditions

The observed interaction of the XS homologs with the regulatory networks seems to be more relevant under specific environmental conditions and in different stages of the bacterial life cycle. For instance, in UPEC strain 536, the *hnsB* gene is highly expressed in the stationary phase and at temperatures below 37°C, while the opposite is observed in the logarithmic phase and 45°C (Müller et al., 2010). Moreover, the effect of an *hnsB* mutation on the generation time, capsule expression and *bgl* expression is stronger at 25°C compared with 37°C and 42°C (Müller et al., 2010). Contrary to the *hfp* pattern, the expression of *sfh* and *pmr* is high during the log phase, although it was observed that the amount of the Sfh protein increases during the stationary phase and Pmr remains relatively constant along the growth curve (Doyle and Dorman, 2006;

Yun et al., 2010; Sun et al., 2017). Further research is required to address the different conditions and contexts that might be modulating the expression of the horizontally acquired XS homologs.

4. Concluding remarks

The main features of the XS proteins (i.e. the preference for AT-rich sequences, a relative lack of binding specificity and the capacity to form homo and heterooligomers) allow their horizontally acquired homologs to modulate a subset of their regulon through the interaction with the XS proteins and their bindings sites, as exemplified by Sfh, Pmr and Hfp. Although less is known about the other horizontally acquired XS homologs, Rok_{LS20}, CgpS and H-NST also modulate the expression of several genes outside the mobile elements that encode these proteins. While most research has focused on the homologs encoded by plasmids and, more recently, by bacteriophages, the XS homologs encoded by GIs have received less attention. Nevertheless, the GI-encoded homolog H-NSB may become of special interest due to its presence in several pathogenic members of the family *Enterobacteriaceae* including the globally spread carbapenem-resistant *K. pneumoniae* ST258. Moreover, GIs also encode members of the MvaT and Lsr2 families of XSs and, as next-generation sequencing is continuously providing us with more bacterial genomes, additional mobile elements encoding XS homologs will emerge. The current literature shows that the horizontally acquired homologs of XSs play important roles as modulators of gene expression in bacteria, which facilitate horizontal gene transfer, participate in virulence and provide, in some instances, additional growth-phase/environmental-responsive regulation mechanisms.

Supplementary Materials: The following are available online att http://www.mdpi.com/2073-4425/11/2/142/s1, Figure S1: Multiple sequence alignment of H-NSB$_{EARL}$ proteins, Figure S2: Multiple sequence alignment of H-NST$_{EARL}$ proteins, Table S1: Xenogeneic silencer homologs found in genomic islands of the Islander database.

Author Contributions: Conceptualization, A.P.I. and S.B.; writing—original draft preparation, A.P.I., I.D.S. G.H.E and D.U.A.; writing—review and editing, AP.I., I.D.S., P.G., A.K. and S.B.; funding acquisition, A.P.I., I.D.S, P.G, A.K. and S.B.

Funding: Authors of this article were supported by the Millennium Institute on Immunology and Immunotherapy (P09/016-F) and by Comisión Nacional de Investigación Científica y Tecnológica de Chile, CONICYT, through Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT grants 1190864, 1190830 and 1170964). AP and IS were also supported by CONICYT PFCHA/Beca de Doctorado Nacional 21172030, and the PhD Scholarship VRI-CPD of the Vice-Rectory of Research from Pontificia Universidad Católica de Chile, respectively.

Conflicts of Interest: The authors declare no conflict of interest.

References

Ali S. S., Whitney J. C., Stevenson J., Robinson H., Howell P. L., and Navarre W. W. (2013). Structural Insights into the Regulation of Foreign Genes in *Salmonella* by the Hha/H-NS Complex. *J. Biol. Chem.*, *288*(19), 13356–13369. https://doi.org/10.1074/jbc.M113.455378

Ares M. A., Fernández-Vázquez J. L., Rosales-Reyes R., Jarillo-Quijada M. D., von Bargen K., Torres J., González-y-Merchand J. A., Alcántar-Curiel M. D., and De la Cruz M. A. (2016). H-NS Nucleoid Protein Controls Virulence Features of *Klebsiella pneumoniae* by Regulating the Expression of Type 3 Pili and the Capsule Polysaccharide. *Front. Cell. Infect. Microbiol.*, *6*, 13. https://doi.org/10.3389/fcimb.2016.00013

Arold S. T., Leonard P. G., Parkinson G. N., and Ladbury J. E. (2010). H-NS forms a superhelical protein scaffold for DNA condensation. *Proc. Natl. Acad. Sci. U. S. A.*, *107*(36), 15728–15732. https://doi.org/10.1073/pnas.1006966107

Bartek I. L., Woolhiser L. K., Baughn A. D., Basaraba R. J., Jacobs W. R., Lenaerts A. J., and Voskuil M. I. (2014). *Mycobacterium tuberculosis* Lsr2 Is a Global Transcriptional Regulator Required for Adaptation to Changing Oxygen Levels and Virulence. *MBio*, *5*(3), e01106-14. https://doi.org/10.1128/mBio.01106-14.Editor

Beloin C., Deighan P., Doyle M., and Dorman C. J. (2003). *Shigella flexneri* 2a strain 2457T expresses three members of the H-NS-like protein family: characterization of the Sfh protein. *Mol. Genet. Genomics*, *270*(1), 66–77. https://doi.org/10.1007/s00438-003-0897-0

Bhat A. P., Shin M., and Choy H. E. (2014). Identification of high-specificity H-NS binding site in LEE5 promoter of enteropathogenic *Escherichia coli* (EPEC). *J. Microbiol.*, *52*(7), 626–629. https://doi.org/10.1007/s12275-014-3562-x

Bowers J. R., Kitchel B., Driebe E. M., MacCannell D. R., Roe C., Lemmer D., de Man T., Rasheed J. K., Engelthaler D. M., Keim P., and Limbago B. M. (2015). Genomic analysis of the emergence and rapid global dissemination of the clonal group 258 *Klebsiella pneumoniae* pandemic. *PLoS One*, *10*(7), e0133727. https://doi.org/10.1371/journal.pone.0133727

Bueno S. M., Santiviago C. A., Murillo A. A., Fuentes J. A., Trombert A. N., Rodas P. I., Youderian P., and Mora G. C. (2004). Precise Excision of the Large Pathogenicity Island, SPI7, in *Salmonella enterica* Serovar Typhi. *J. Bacteriol.*, *186*(10), 3202–3213. https://doi.org/10.1128/JB.186.10.3202

Carpenter M. R., Rozovsky S., and Boyd E. F. (2016). Pathogenicity island cross talk mediated by recombination directionality factors facilitates excision from the chromosome. *J. Bacteriol.*, *198*(5), 766–776. https://doi.org/10.1128/JB.00704-15

Carraro N., Matteau D., Luo P., Rodrigue S., and Burrus V. (2014). The Master Activator of IncA/C Conjugative Plasmids Stimulates Genomic Islands and Multidrug Resistance Dissemination. *PLoS Genet.*, *10*(10), e1004714. https://doi.org/10.1371/journal.pgen.1004714

Castang S., and Dove S. L. (2010). High-order oligomerization is required for the function of the H-NS family member MvaT in *Pseudomonas aeruginosa*. *Mol. Microbiol.*, *78*(4), 916–931. https://doi.org/10.1111/j.1365-2958.2010.07378.x

Che D., Hasan M. S., and Chen B. (2014). Identifying Pathogenicity Islands in Bacterial Pathogenomics Using Computational Approaches. *Pathogens*, *3*, 36–56. https://doi.org/10.3390/pathogens3010036

Coluzzi C., Guédon G., Devignes M. D., Ambroset C., Loux V., Lacroix T., Payot S., and Leblond-Bourget N. (2017). 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. https://doi.org/10.3389/fmicb.2017.00443

Coward C., Sait L., Williams L., Humphrey T. J., Cogan T., and Maskell D. J. (2012). Investigation into the role of five *Salmonella enterica* serovar Enteritidis genomic islands in colonization of the chicken reproductive tract and other organs following oral challenge. *FEMS Microbiol. Lett.*, *336*(1), 73–78. https://doi.org/10.1111/j.1574-6968.2012.02652.x

Daccord A., Ceccarelli D., Rodrigue S., and Burrus V. (2013). Comparative Analysis of Mobilizable Genomic Islands. *J. Bacteriol.*, *195*(3), 606–614. https://doi.org/10.1128/JB.01985-12

Darmon E., and Leach D. R. F. (2014). Bacterial genome instability. *Microbiol. Mol. BBology Rev.*, *78*(1), 1–39. https://doi.org/10.1128/MMBR.00035-13

Deighan P., Beloin C., and Dorman C. J. (2003). Three-way interactions among the Sfh, StpA and H-NS nucleoid-structuring proteins of *Shigella flexneri* 2a strain 2457T. *Mol. Microbiol.*, *48*(5), 1401–1416. https://doi.org/10.1046/j.1365-2958.2003.03515.x

Deng W., Li Y., Vallance B. A., and Finlay B. B. (2001). Locus of Enterocyte Effacement from *Citrobacter rodentium*: Sequence analysis and evidence for horizontal transfer among attaching and effacing pathogens. *Infect. Immun.*, *69*(10), 6323–6335. https://doi.org/10.1128/IAI.69.10.6323-6335.2001

Desai P. T., Porwollik S., Long F., Cheng P., Wollam A., Clifton S. W., Weinstock G. M., and McClelland M. (2013). Evolutionary Genomics of *Salmonella enterica* Subspecies. *MBio*, *4*(2), e00579-12. https://doi.org/10.1128/mBio.00579-12.Editor

Dillon S. C., Cameron A. D. S., Hokamp K., Lucchini S., Hinton J. C. D., and Dorman C. J. (2010). Genome-wide analysis of the H-NS and Sfh regulatory networks in *Salmonella* Typhimurium identifies a plasmid-encoded transcription silencing mechanism. *Mol. Microbiol.*, *76*(5), 1250–1265. https://doi.org/10.1111/j.1365-2958.2010.07173.x

Dillon S. C., and Dorman C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.*, *8*(3), 185–195. https://doi.org/10.1038/nrmicro2261

Ding P., McFarland K. A., Jin S., Tong G., Duan B., Yang A., Hughes T. R., Liu J., Dove S. L., Navarre W. W., and Xia B. (2015). A Novel AT-Rich DNA Recognition Mechanism for Bacterial Xenogeneic Silencer MvaT. *PLoS Pathog.*, *11*(6), e1004967. https://doi.org/10.1371/journal.ppat.1004967

Dobrindt U., Hochhut B., Hentschel U., and Hacker J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.*, *2*(5), 414–424. https://doi.org/10.1038/nrmicro884

Dorman C. J. (2004). H-NS: a universal regulator for a dynamic genome. *Nat. Rev. Microbiol.*, *2*(5), 391–400. https://doi.org/10.1038/nrmicro883

Dorman C. J. (2014a). Function of nucleoid-associated proteins in chromosome structuring and transcriptional regulation. *J. Mol. Microbiol. Biotechnol.*, *24*(5–6), 316–331. https://doi.org/10.1159/000368850

Dorman C. J. (2014b). H-NS-like nucleoid-associated proteins, mobile genetic elements and horizontal gene transfer in bacteria. *Plasmid*, *75*, 1–11. https://doi.org/10.1016/j.plasmid.2014.06.004

Doublet B., Boyd D., Mulvey M. R., and Cloeckaert A. (2005). The *Salmonella* genomic island 1 is an integrative mobilizable element. *Mol. Microbiol.*, *55*(6), 1911–1924. https://doi.org/10.1111/j.1365-2958.2005.04520.x

Doyle M., and Dorman C. J. (2006). Reciprocal transcriptional and posttranscriptional growth-phase-dependent expression of *sfh*, a gene that encodes a paralogue of the nucleoid-associated protein H-NS. *J. Bacteriol.*, *188*(21), 7581–7591. https://doi.org/10.1128/JB.00685-06

Doyle M., Fookes M., Ivens A., Mangan M. W., Wain J., and Dorman C. J. (2007). An H-NS-like Stealth Protein Aids Horizontal Transmission in Bacteria. *Science.*, *315*(5809), 251–252. https://doi.org/10.1126/science.1137550

Duan B., Ding P., Hughes T. R., Navarre W. W., Liu J., and Xia B. (2018). How bacterial xenogeneic silencer rok distinguishes foreign from self DNA in its resident genome. *Nucleic Acids Res.*, *46*(19), 10514–10529. https://doi.org/10.1093/nar/gky836

Elliott S. J., Sperandio V., Girón J. A., Mellies J. L., Wainwright L., Hutcheson S. W., McDaniel T. K., and Kaper J. B. (2000). The Locus of Enterocyte Effacement (LEE)-Encoded Regulator Controls Expression of Both LEE- and Non-LEE-Encoded Virulence Factors in Enteropathogenic and Enterohemorrhagic *Escherichia coli*. *Infect. Immun.*, *68*(11), 6115–6126. https://doi.org/10.1128/IAI.68.11.6115-6126.2000.Updated

Elliott S. J., Wainwright L. A., McDaniel T. K., Jarvis K. G., Deng Y. K., Lai L. C., McNamara B. P., Donnenberg M. S., and Kaper J. B. (1998). The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Mol. Microbiol.*, *28*(1), 1–4. https://doi.org/10.1046/j.1365-2958.1998.00783.x

Feasey N. A., Hadfield J., Keddy K. H., Dallman T. J., Jacobs J., Deng X., Wigley P., Barquist L., Langridge G. C., Feltwell T., Harris S. R., Mather A. E., Fookes M., Aslett M., Msefula C., Kariuki S., Maclennan C. A., Onsare R. S., Weill F.-X., … Thomson N. R. (2016). Distinct *Salmonella* Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. *Nat. Genet.*, *48*(10), 1211–1217. https://doi.org/10.1038/ng.3644

Fernández-Alarcón C., Singer R. S., and Johnson T. J. (2011). Comparative genomics of multidrug resistance-encoding IncA/C plasmids from commensal and pathogenic *Escherichia coli* from multiple animal sources. *PLoS One*, *6*(8), e23415. https://doi.org/10.1371/journal.pone.0023415

Flowers J. J., He S., Malfatti S., Del Rio T. G., Tringe S. G., Hugenholtz P., and McMahon K. D. (2013). Comparative genomics of two "*Candidatus* Accumulibacter" clades performing biological phosphorus removal. *ISME J.*, *7*(12), 2301–2314. https://doi.org/10.1038/ismej.2013.117

Forns N., Baños R. C., Balsalobre C., Juárez A., and Madrid C. (2005). Temperature-dependent conjugative transfer of R27: Role of chromosome- and plasmid-encoded Hha and H-NS proteins. *J. Bacteriol.*, *187*(12), 3950–3959. https://doi.org/10.1128/JB.187.12.3950-3959.2005

Gao Y., Foo Y. H., Winardhi R. S., Tang Q., Yan J., and Kenney L. J. (2017). Charged residues in the H-NS linker drive DNA binding and gene silencing in single cells. *Proc. Natl. Acad. Sci. U. S. A.*, *114*(47), 12560–12565. https://doi.org/10.1073/pnas.1716721114

García J., Madrid C., Juárez A., and Pons M. (2006). New Roles for Key Residues in Helices H1 and H2 of the *Escherichia coli* H-NS N-terminal Domain: H-NS Dimer Stabilization and Hha Binding. *J. Mol. Biol.*, *359*(3), 679–689. https://doi.org/10.1016/j.jmb.2006.03.059

Gordon B. R. G., Li Y., Cote A., Weirauch M. T., Ding P., Hughes T. R., Navarre W. W., Xia B., and Liu J. (2011). Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. *Proc. Natl. Acad. Sci.*, *108*(26), 10690–10695. https://doi.org/10.1073/pnas.1102544108

Gordon B. R. G., Li Y., Wang L., Sintsova A., van Bakel H., Tian S., Navarre W. W., Xia B., and Liu J. (2010). Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.*, *107*(11), 5154–5159. https://doi.org/10.1073/pnas.0913551107

Guignot J., Segura A., and Tran Van Nhieu G. (2015). The Serine Protease EspC from Enteropathogenic *Escherichia coli* Regulates Pore Formation and Cytotoxicity Mediated by the Type III Secretion System. *PLoS Pathog.*, *11*(7), e1005013. https://doi.org/10.1371/journal.ppat.1005013

Hacker J., and Kaper J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.*, *54*, 641–679.

Haskett T. L., Terpolilli J. J., Bekuma A., O'Hara G. W., Sullivan J. T., Wang P., Ronson C. W., and Ramsay J. P. (2016). Assembly and transfer of tripartite integrative and conjugative genetic elements. *Proc. Natl. Acad. Sci. U. S. A.*, *113*(43), 12268–12273. https://doi.org/10.1073/pnas.1613358113

Haskett T. L., Terpolilli J. J., Ramachandran V. K., Verdonk C. J., Poole P. S., O'Hara G. W., and Ramsay J. P. (2018). Sequential induction of three recombination directionality factors directs assembly of tripartite integrative and conjugative elements. *PLOS Genet.*, *14*(3), e1007292. https://doi.org/10.1371/journal.pgen.1007292

Hatfull G. F., Jacobs-Sera D., Lawrence J. G., Pope W. H., Russell D. A., Ko C.-C., Weber R. J., Patel M. C., Germane K. L., Edgar R. H., Hoyte N. N., Bowman C. A., Tantoco A. T., Paladin E. C., Myers M. S., Smith A. L., Grace M. S., Pham T. T., O'Brien M. B., … Hendrix R. W. (2010). Comparative Genomic Analysis of 60 Mycobacteriophage Genomes: Genome Clustering, Gene Acquisition, and Gene Size. *J. Mol. Biol.*, *397*(1), 119–143. https://doi.org/10.1016/j.jmb.2010.01.011

Helgesen E., Fossum-Raunehaug S., and Skarstad K. (2016). Lack of the H-NS protein results in extended and aberrantly positioned DNA during chromosome replication and segregation in *Escherichia coli*. *J. Bacteriol.*, *198*(8), 1305–1316. https://doi.org/10.1128/JB.00919-15

Hoa T. T., Tortosa P., Albano M., and Dubnau D. (2002). Rok (YkuW) regulates genetic competence in *Bacillus subtilis* by directly repressing *comK*. *Mol. Microbiol.*, *43*(1), 15–26. https://doi.org/10.1046/j.1365-2958.2002.02727.x

Hudson C. M., Lau B. Y., and Williams K. P. (2015). Islander: A database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic Acids Res.*, *43*(D1), D48–D53. https://doi.org/10.1093/nar/gku1072

Jeong H., Sung S., Kwon T., Seo M., Caetano-Anollés K., Choi S. H., Cho S., Nasir A., and Kim H. (2016). HGTree: database of horizontally transferred genes determined by tree reconciliation. *Nucleic Acids Res.*, *44*(D1), D610–D619. https://doi.org/10.1093/nar/gkv1245

Johansson J., Eriksson S., Sondén B., Wai S. N., and Uhlin B. E. (2001). Heteromeric interactions among nucleoid-associated bacterial proteins: Localization of StpA-stabilizing regions in H-NS of *Escherichia coli*. *J. Bacteriol.*, *183*(7), 2343–2347. https://doi.org/10.1128/JB.183.7.2343-2347.2001

Krogh T. J., Møller-Jensen J., and Kaleta C. (2018). Impact of chromosomal architecture on the function and evolution of bacterial genomes. *Front. Microbiol.*, *9*, 2019. https://doi.org/10.3389/fmicb.2018.02019

Kumar S., Stecher G., Li M., Knyaz C., and Tamura K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Lamberte L. E., Baniulyte G., Singh S. S., Stringer A. M., Bonocora R. P., Stracy M., Kapanidis A. N., Wade J. T., and Grainger D. C. (2017). Horizontally acquired AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase. *Nat. Microbiol.*, *2*, 16249. https://doi.org/10.1038/nmicrobiol.2016.249

Landick R., Wade J. T., and Grainger D. C. (2015). H-NS and RNA polymerase: A love-hate relationship? *Curr. Opin. Microbiol.*, *24*, 53–59. https://doi.org/10.1016/j.mib.2015.01.009

Lang B., Blot N., Bouffartigues E., Buckle M., Geertz M., Gualerzi C. O., Mavathur R., Muskhelishvili G., Pon C. L., Rimsky S., Stella S., Babu M. M., and Travers A. (2007). High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. *Nucleic Acids Res.*, *35*(18), 6330–6337. https://doi.org/10.1093/nar/gkm712

Lang K. S., and Johnson T. J. (2016). Characterization of Acr2, an H-NS-like protein encoded on A/C2-type plasmids. *Plasmid*, *87–88*, 17–27. https://doi.org/10.1016/j.plasmid.2016.07.004

Lathem W. W., Grys T. E., Witowski S. E., Torres A. G., Kaper J. B., Tarr P. I., and Welch R. A. (2002). StcE, a metalloprotease secreted by *Escherichia coli* O157:H7, specifically cleaves C1 esterase inhibitor. *Mol. Microbiol.*, *45*(2), 277–288. https://doi.org/10.1046/j.1365-2958.2002.02997.x

Lautner M., Schunder E., Herrmann V., and Heuner K. (2013). Regulation, integrase-dependent excision, and horizontal transfer of genomic islands in *Legionella pneumophila*. *J. Bacteriol.*, *195*(7), 1583–1597. https://doi.org/10.1128/JB.01739-12

Leh H., Khodr A., Bouger M.-C., Sclavi B., Rimsky S., and Bury-Moné S. (2017). Bacterial-chromatin structural proteins regulate the bimodal expression of the Locus of Enterocyte Effacement (LEE) pathogenicity island in enteropathogenic *Escherichia coli*. *MBio*, *8*(4), e00773-17. https://doi.org/10.1128/mBio.00773-17

Levine J. A., Hansen A.-M., Michalski J. M., Hazen T. H., Rasko D. A., and Kaper J. B. (2014). H-NST induces LEE expression and the formation of attaching and effacing lesions in enterohemorrhagic *Escherichia coli*. *PLoS One*, *9*(1), e86618. https://doi.org/10.1371/journal.pone.0086618

Lou L., Zhang P., Piao R., and Wang Y. (2019). *Salmonella* Pathogenicity Island 1 (SPI-1) and Its Complex Regulatory Network. *Front. Cell. Infect. Microbiol.*, *9*, 270. https://doi.org/10.3389/fcimb.2019.00270

Lucchini S., Rowley G., Goldberg M. D., Hurd D., Harrison M., and Hinton J. C. D. (2006). H-NS Mediates the Silencing of Laterally Acquired Genes in Bacteria. *PLoS Pathog.*, *2*(8), e81. https://doi.org/10.1371/journal.ppat.0020081

Madrid C., García J., Pons M., and Juárez A. (2007). Molecular evolution of the H-NS protein: Interaction with Hha-like proteins is restricted to *Enterobacteriaceae*. *J. Bacteriol.*, *189*(1), 265–268. https://doi.org/10.1128/JB.01124-06

Maeda K., Nojiri H., Shintani M., Yoshida T., Habe H., and Omori T. (2003). Complete nucleotide sequence of carbazole/dioxin-degrading plasmid pCAR1 in *Pseudomonas resinovorans* strain CA10 indicates its mosaicity and the presence of large catabolic transposon Tn*4676*. *J. Mol. Biol.*, *326*(1), 21–33. https://doi.org/10.1016/S0022-2836(02)01400-6

Manson J. M., and Gilmore M. S. (2006). Pathogenicity island integrase cross-talk: A potential new tool for virulence modulation. *Mol. Microbiol.*, *61*(3), 555–559. https://doi.org/10.1111/j.1365-2958.2006.05262.x

Marcoleta A. E., Berríos-Pastén C., Nuñez G., Monasterio O., and Lagos R. (2016). *Klebsiella pneumoniae* Asparagine tDNAs Are Integration Hotspots for Different Genomic Islands Encoding Microcin E492 Production Determinants and Other Putative Virulence Factors Present in Hypervirulent Strains. *Front. Microbiol.*, *7*, 849. https://doi.org/10.3389/fmicb.2016.00849

Marcus S. L., Brumell J. H., Pfeifer C. G., and Finlay B. B. (2000). *Salmonella* pathogenicity islands: big virulence in small packages. *Microbes Infect.*, *2*(2), 145–156. https://doi.org/10.1016/S1286-4579(00)00273-2

Marsh J. W., Mustapha M. M., Griffith M. P., Evans D. R., Ezeonwuka C., Pasculle A. W., Shutt K. A., Sundermann A., Ayres A. M., Shields R. K., Babiker A., Cooper V. S., Tyne D. Van, and Harrison L. H. (2019). Evolution of Outbreak-Causing Carbapenem-Resistant *Klebsiella pneumoniae* ST258 at a Tertiary Care Hospital over 8 Years. *MBio*, *10*(5), e01945-19.

Mellies J. L., Benison G., McNitt W., Mavor D., Boniface C., and Larabee F. J. (2011). Ler of pathogenic *Escherichia coli* forms toroidal protein-DNA complexes. *Microbiology*, *157*(Pt4), 1123–1133. https://doi.org/10.1099/mic.0.046094-0

Mellies J. L., Elliott S. J., Sperandio V., Donnenberg M. S., and Kaper J. B. (1999). The Per regulon of enteropathogenic *Escherichia coli*: identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler). *Mol. Microbiol.*, *33*(2), 296–306.

Miyakoshi M., Shintani M., Terabayashi T., Kai S., Yamane H., and Nojiri H. (2007). Transcriptome analysis of *Pseudomonas putida* KT2440 harboring the completely sequenced IncP-7 plasmid pCAR1. *J. Bacteriol.*, *189*(19), 6849–6860. https://doi.org/10.1128/JB.00684-07

Müller C. M., Schneider G., Dobrindt U., Emödy L., Hacker J., and Uhlin B. E. (2010). Differential effects and interactions of endogenous and horizontally acquired H-NS-like proteins in pathogenic *Escherichia coli*. *Mol. Microbiol.*, *75*(2), 280–293. https://doi.org/10.1111/j.1365-2958.2009.06995.x

Murphy R. A., and Boyd E. F. (2008). Three Pathogenicity Islands of *Vibrio cholerae* Can Excise from the Chromosome and Form Circular Intermediates. *J. Bacteriol.*, *190*(2), 636–647. https://doi.org/10.1128/JB.00562-07

Navarre W. W. (2010). H-NS as a Defence System. In *Bacterial Chromatin* (pp. 251–322). Springer Netherlands. https://doi.org/10.1007/978-90-481-3473-1

Navarre W. W. (2016). The Impact of Gene Silencing on Horizontal Gene Transfer and Bacterial Evolution. In R. K. Poole (Ed.), *Advances in Microbial Physiology* (Vol. 69, pp. 157–186). Academic Press. https://doi.org/10.1016/bs.ampbs.2016.07.004

Navarre W. W., McClelland M., Libby S. J., and Fang F. C. (2007). Silencing of xenogenic DNA by H-NS - facilitation of lateral gene transfer in bacteria by a defence system that recognizes foreign DNA. *Genes Dev.*, *21*(12), 1456–1471. https://doi.org/10.1101/gad.1543107

Navarre W. W., Porwollik S., Wang Y., McClelland M., Rosen H., Libby S. J., and Fang F. C. (2006). Selective Silencing of Foreign DNA with Low GC Content by the H-NS Protein in *Salmonella*. *Science.*, *313*(5784), 236–238. https://doi.org/10.1126/science.1128794

Newman R. M., Salunkhe P., Godzik A., and Reed J. C. (2006). Identification and Characterization of a Novel Bacterial Virulence Factor That Shares Homology with Mammalian Toll/Interleukin-1 Receptor Family Proteins. *Infect. Immun.*, *74*(1), 594–601. https://doi.org/10.1128/IAI.74.1.594

Ochman H., Lawrence J. G., and Groisman E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature*, *405*(6784), 299–304. https://doi.org/10.1038/35012500

Ooka T., Ogura Y., Katsura K., Seto K., Kobayashi H., Kawano K., Tokuoka E., Furukawa M., Harada S., Yoshino S., Seto J., Ikeda T., Yamaguchi K., Murase K., Gotoh Y., Imuta N., Nishi J., Gomes T. A., Beutin L., and Hayashi T. (2015). Defining the genome features of *Escherichia albertii*, an emerging enteropathogen closely related to *Escherichia coli*. *Genome Biol. Evol.*, *7*(12), 3170–3179. https://doi.org/10.1093/gbe/evv211

Paquola A. C. M., Asif H., Pereira C. A. de B., Feltes B. C., Bonatto D., Lima W. C., and Menck C. F. M. (2018). Horizontal Gene Transfer Building Prokaryote Genomes: Genes Related to Exchange Between Cell and Environment are Frequently Transferred. *J. Mol. Evol.*, *86*(3–4), 190–203. https://doi.org/10.1007/s00239-018-9836-x

Pardo-Roa C., Salazar G. A., Noguera L. P., Salazar-Echegarai F. J., Vallejos O. P., Suazo I. D., Schultz B. M., Coronado-Arrazola I., Kalergis A. M., and Bueno S. M. (2019). Pathogenicity island excision during an infection by Salmonella enterica serovar Enteritidis is required for crossing the intestinal epithelial barrier in mice to cause systemic infection. *PLOS Pathog.*, *15*(12), e1008152. https://doi.org/10.1371/journal.ppat.1008152

Pedulla M. L., Ford M. E., Houtz J. M., Karthikeyan T., Wadsworth C., Lewis J. A., Jacobs-Sera D., Falbo J., Gross J., Pannunzio N. R., Brucker W., Kumar V., Kandasamy J., Keenan L., Bardarov S., Kriakov J., Lawrence J. G., Jacobs W. R., Hendrix R. W., and Hatfull G. F. (2003). Origins of highly mosaic mycobacteriophage genomes. *Cell*, *113*(2), 171–182. https://doi.org/10.1016/S0092-8674(03)00233-2

Penadés J. R., and Christie G. E. (2015). The Phage-Inducible Chromosomal Islands: A Family of Highly Evolved Molecular Parasites. *Annu. Rev. Virol.*, *2*, 181–201. https://doi.org/10.1146/annurev-virology-031413-085446

Pfeifer E., Hünnefeld M., Popa O., and Frunzke J. (2019). Impact of Xenogeneic Silencing on Phage–Host Interactions. *J. Mol. Biol.*, *431*(23), 4670–4683. https://doi.org/10.1016/j.jmb.2019.02.011

Pfeifer E., Hünnefeld M., Popa O., Polen T., Kohlheyer D., Baumgart M., and Frunzke J. (2016). Silencing of cryptic prophages in *Corynebacterium glutamicum*. *Nucleic Acids Res.*, *44*(21), 10117–10131. https://doi.org/10.1093/nar/gkw692

Picker M. A., and Wing H. J. (2016). H-NS, Its Family Members and Their Regulation of Virulence Genes in *Shigella* Species. *Genes (Basel).*, *7*(12), 112. https://doi.org/10.3390/genes7120112

Piña-Iturbe A., Ulloa-Allendes D., Pardo-Roa C., Coronado-Arrázola I., Salazar-Echegarai F. J., Sclavi B., González P. A., and Bueno S. M. (2018). Comparative and phylogenetic analysis of a novel family of *Enterobacteriaceae*-associated genomic islands that share a conserved excision/integration module. *Sci. Rep.*, *8*(1), 10292. https://doi.org/10.1038/s41598-018-28537-0

Platt T. G., Bever J. D., and Fuqua C. (2012). A cooperative virulence plasmid imposes a high fitness cost under conditions that induce pathogenesis. *Proc. R. Soc. B.*, *279*(1734), 1691–1699. https://doi.org/10.1098/rspb.2011.2002

Queiroz M. H., Madrid C., Paytubi S., Balsalobre C., and Juárez A. (2011). Integration host factor alleviates H-NS silencing of the *Salmonella enterica* serovar Typhimurium master regulator of SPI1, *hilA*. *Microbiology*, *157*(9), 2504–2514. https://doi.org/10.1099/mic.0.049197-0

Quiroz T. S., Nieto P. A., Tobar H. E., Salazar-Echegarai F. J., Lizana R. J., Quezada C. P., Santiviago C. A., Araya D. V, Riedel C. A., Kalergis A. M., and Bueno S. M. (2011). Excision of an Unstable Pathogenicity Island in *Salmonella enterica* Serovar Enteritidis Is Induced during Infection of Phagocytic Cells. *PLoS One*, *6*(10), e26031. https://doi.org/10.1371/journal.pone.0026031

Robert X., and Gouet P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.*, *42*(Web Server issue), W320–W324. https://doi.org/10.1093/nar/gku316

Salazar-Echegarai F. J., Tobar H. E., Nieto P. A., Riedel C. A., and Bueno S. M. (2014). Conjugal Transfer of the Pathogenicity Island ROD21 in *Salmonella enterica* serovar Enteritidis Depends on Environmental Conditions. *PLoS One*, *9*(4), e90626. https://doi.org/10.1371/journal.pone.0090626

San Millan A., Toll-Riera M., Qi Q., and MacLean R. C. (2015). Interactions between horizontally acquired genes create a fitness cost in *Pseudomonas aeruginosa*. *Nat. Commun.*, *6*, 6845. https://doi.org/10.1038/ncomms7845

Scolari V. F., Sclavi B., and Lagomarsino M. C. (2015). The nucleoid as a smart polymer. *Front. Microbiol.*, *6*, 424. https://doi.org/10.3389/fmicb.2015.00424

Sentchilo V., Czechowska K., Pradervand N., Minoia M., Miyazaki R., and van der Meer J. R. (2009). Intracellular excision and reintegration dynamics of the ICEclc genomic island of *Pseudomonas knackmussii* sp. strain B13. *Mol. Microbiol.*, *72*(5), 1293–1306. https://doi.org/10.1111/j.1365-2958.2009.06726.x

Shahul Hameed U. F., Liao C., Radhakrishnan A. K., Huser F., Aljedani S. S., Zhao X., Momin A. A., Melo F. A., Guo X., Brooks C., Li Y., Cui X., Gao X., Ladbury J. E., Jaremko Ł., Jaremko M., Li J., and Arold S. T. (2019). H-NS uses an autoinhibitory conformational switch for environment-controlled gene silencing. *Nucleic Acids Res.*, *47*(5), 2666–2680. https://doi.org/10.1093/nar/gky1299

Sherburne C. K., Lawley T. D., Gilmour M. W., Blattner F. R., Burland V., Grotbeck E., Rose D. J., and Taylor D. E. (2000). The complete DNA sequence and analysis of R27, a large IncHI plasmid from *Salmonella typhi* that is temperature sensitive for transfer. *Nucleic Acids Res.*, *28*(10), 2177–2186. https://doi.org/10.1093/nar/28.10.2177

Shintani M., Suzuki-Minakuchi C., and Nojiri H. (2015). Nucleoid-associated proteins encoded on plasmids: Occurrence and mode of function. *Plasmid*, *80*, 32–44. https://doi.org/10.1016/j.plasmid.2015.04.008

Silva C. A., Blondel C. J., Quezada C. P., Porwollik S., Andrews-polymenis H. L., Toro C. S., Mcclelland M., and Santiviago C. A. (2012). Infection of Mice by *Salmonella enterica* Serovar Enteritidis Involves Additional Genes That Are Absent in the Genome of Serovar Typhimurium. *Infect. Immun.*, *80*(2), 839–849. https://doi.org/10.1128/IAI.05497-11

Singh K., Milstein J. N., and Navarre W. W. (2016). Xenogeneic Silencing and Its Impact on Bacterial Genomes. *Annu. Rev. Microbiol.*, *70*, 199–213. https://doi.org/10.1017/CBO9781107415324.004

Singh P. K., Ramachandran G., Durán-Alcalde L., Alonso C., Wu L. J., and Meijer W. J. J. (2012). Inhibition of *Bacillus subtilis* natural competence by a native, conjugative plasmid-encoded *comK* repressor protein. *Environ. Microbiol.*, *14*(10), 2812–2825. https://doi.org/10.1111/j.1462-2920.2012.02819.x

Skennerton C. T., Angly F. E., Breitbart M., Bragg L., He S., McMahon K. D., Hugenholtz P., and Tyson G. W. (2011). Phage encoded H-NS: A potential achilles heel in the bacterial defence system. *PLoS One*, *6*(5), e20095. https://doi.org/10.1371/journal.pone.0020095

Smits W. K., and Grossman A. D. (2010). The Transcriptional Regulator Rok Binds A+T-Rich DNA and Is Involved in Repression of a Mobile Genetic Element in *Bacillus subtilis*. *PLoS Genet.*, *6*(11), e1001207. https://doi.org/10.1371/journal.pgen.1001207

Soucy S. M., Huang J., and Gogarten J. P. (2015). Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.*, *16*(8), 472–482. https://doi.org/10.1038/nrg3962

Sullivan J. T., and Ronson C. W. (1998). Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Genetics*, *95*(April), 5145–5149. https://doi.org/10.1073/pnas.95.9.5145

Summers E. L., Meindl K., Usón I., Mitra A. K., Radjainia M., Colangeli R., Alland D., and Arcus V. L. (2012). The Structure of the Oligomerization Domain of Lsr2 from *Mycobacterium tuberculosis* Reveals a Mechanism for Chromosome Organization and Protection. *PLoS One*, *7*(6), e38542. https://doi.org/10.1371/journal.pone.0038542

Sun Z., Vasileva D., Suzuki-Minakuchi C., Okada K., Luo F., Igarashi Y., and Nojiri H. (2017). Growth phase-dependent expression profiles of three vital H-NS family proteins encoded on the chromosome of *Pseudomonas putida* KT2440 and on the pCAR1 plasmid. *BMC Microbiol.*, *17*(1), 188. https://doi.org/10.1186/s12866-017-1091-6

Sun Z., Vasileva D., Suzuki-Minakuchi C., Okada K., Luo F., Igarashi Y., and Nojiri H. (2018). Differential protein-protein binding affinities of H-NS family proteins encoded on the chromosome of *Pseudomonas putida* KT2440 and IncP-7 plasmid pCAR1. *Biosci. Biotechnol. Biochem.*, *82*(9), 1640–1646. https://doi.org/10.1080/09168451.2018.1484277

Suzuki-Minakuchi C., Kawazuma K., Matsuzawa J., Vasileva D., Fujimoto Z., Terada T., Okada K., and Nojiri H. (2016). Structural similarities and differences in H-NS family proteins revealed by the N-terminal structure of TurB in *Pseudomonas putida* KT2440. *FEBS Lett.*, *590*(20), 3583–3594. https://doi.org/10.1002/1873-3468.12425

Suzuki-Minakuchi C., and Navarre W. W. (2019). Xenogeneic Silencing and Horizontal Gene Transfer. In H. Nishida & T. Oshima (Eds.), *DNA Traffic in the Environment* (pp. 1–27). Springer, Singapore. https://doi.org/10.1007/978-981-13-3411-5

Thomson N. R., Clayton D. J., Windhorst D., Vernikos G., Davidson S., Churcher C., Quail M. A., Stevens M., Jones M. A., Watson M., Barron A., Layton A., Pickard D., Kingsley R. A., Bignell A., Clark L., Harris B., Ormond D., Abdellah Z., Brooks K., Cherevach I., Chillingworth T., Woodward J., Norberczak H., Lord A., Arrowsmith C., Jagels K., Moule S., Mungall K., Sanders M., Whitehead S., Chabalgoity J. A., Maskell D., Humphrey T., Roberts M., Barrow P. A., Dougan G., and Parkhill J. (2008). 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. https://doi.org/10.1101/gr.077404.108.1

Tobar H. E., Salazar-Echegarai F. J., Nieto P. A., Palavecino C. E., Sebastian V. P., Riedel C. A., Kalergis A. M., and Bueno S. M. (2013). Chromosomal Excision of a New Pathogenicity Island Modulates *Salmonella* Virulence *In Vivo*. *Curr. Gene Ther.*, *13*(4), 240–249.

Torres A. G., López-Sánchez G. N., Milflores-Flores L., Patel S. D., Rojas-López M., Martínez De La Peña C. F., Arenas-Hernández M. M. P., and Martínez-Laguna Y. (2007). Ler and H-NS, regulators controlling expression of the long polar fimbriae of *Escherichia coli* O157:H7. *J. Bacteriol.*, *189*(16), 5916–5928. https://doi.org/10.1128/JB.00245-07

Will W. R., Bale D. H., Reid P. J., Libby S. J., and Fang F. C. (2014). Evolutionary expansion of a regulatory network by counter-silencing. *Nat. Commun.*, *5*, 5270. https://doi.org/10.1038/ncomms6270

Williams K. P. (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res.*, *30*(4), 866–875. https://doi.org/10.1093/nar/30.4.866

Williams R. M., Rimsky S., and Buc H. (1996). Probing the structure, function, and interactions of the *Escherichia coli* H-NS and StpA proteins by using dominant negative derivatives. *J. Bacteriol.*, *178*(15), 4335–4343. https://doi.org/10.1128/jb.178.15.4335-4343.1996

Williamson H. S., and Free A. (2005). A truncated H-NS-like protein from enteropathogenic *Escherichia coli* acts as an H-NS antagonist. *Mol. Microbiol.*, *55*(3), 808–827. https://doi.org/10.1111/j.1365-2958.2004.04421.x

Winardhi R. S., Gulvady R., Mellies J. L., and Yan J. (2014). Locus of enterocyte effacement-encoded regulator (Ler) of pathogenic *Escherichia coli* competes off histone-like nucleoid-structuring protein (H-NS) through noncooperative DNA binding. *J. Biol. Chem.*, *289*(20), 13739–13750. https://doi.org/10.1074/jbc.M113.545954

Wozniak R. A. F., Fouts D. E., Spagnoletti M., Colombo M. M., Ceccarelli D., Garriss G., Déry C., Burrus V., and Waldor M. K. (2009). Comparative ICE genomics: Insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genet.*, *5*(12), e1000786. https://doi.org/10.1371/journal.pgen.1000786

Yun C.-S., Suzuki C., Naito K., Takeda T., Takahashi Y., Sai F., Terabayashi T., Miyakoshi M., Shintani M., Nishida H., Yamane H., and Nojiri H. (2010). Pmr, a histone-like protein H1 (H-NS) family protein encoded by the IncP-7 plasmid pCAR1, is a key global regulator that alters host function. *J. Bacteriol.*, *192*(18), 4720–4731. https://doi.org/10.1128/JB.00591-10

Yun C. S., Takahashi Y., Shintani M., Takeda T., Suzuki-Minakuchi C., Okada K., Yamane H., and Nojiri H. (2015). MvaT family proteins encoded on IncP-7 plasmid pCAR1 and the host chromosome regulate the host transcriptome cooperatively but differently. *Appl. Environ. Microbiol.*, *82*(3), 832–842. https://doi.org/10.1128/AEM.03071-15

Zhang C.-T., and Zhang R. (2004). Genomic islands in *Rhodopseudomonas palustris*. *Nat. Biotechnol.*, *22*(9), 1078–1079. https://doi.org/10.1038/nbt0904-1078b

Zhang R., and Zhang C.-T. (2004). A systematic method to identify genomic islands and its applications in analyzing the genomes of *Corynebacterium glutamicum* and *Vibrio vulnificus* CMCP6 chromosome I. *Bioinformatics*, *20*(5), 612–622. https://doi.org/10.1093/bioinformatics/btg453

Zhang X., Wan W., Yu H., Wang M., Zhang H., Lv J., Tang Y.-W., Kreiswirth B. N., Du H., and Chen L. (2019). New Delhi Metallo-β-Lactamase 5–Producing *Klebsiella pneumoniae* Sequence Type 258, Southwest China, 2017. *Emerg. Infect. Dis.*, *25*(6), 1209–1213.

Supplementary Materials

Horizontally acquired homologs of xenogeneic silencers: modulators of gene expression encoded by plasmids, phages and genomic islands

Alejandro Piña-Iturbe[1], Isidora D. Suazo[1], Guillermo Hoppe-Elsholz[1], Diego Ulloa-Allendes[1], Pablo A. González[1], Alexis M. Kalergis[1,2], Susan M. Bueno[1,*]

[1]Millennium Institute on Immunology and Immunotherapy, Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile

[2]Departamento de Endocrinología, Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile

*****Correspondence: sbueno@bio.puc.cl

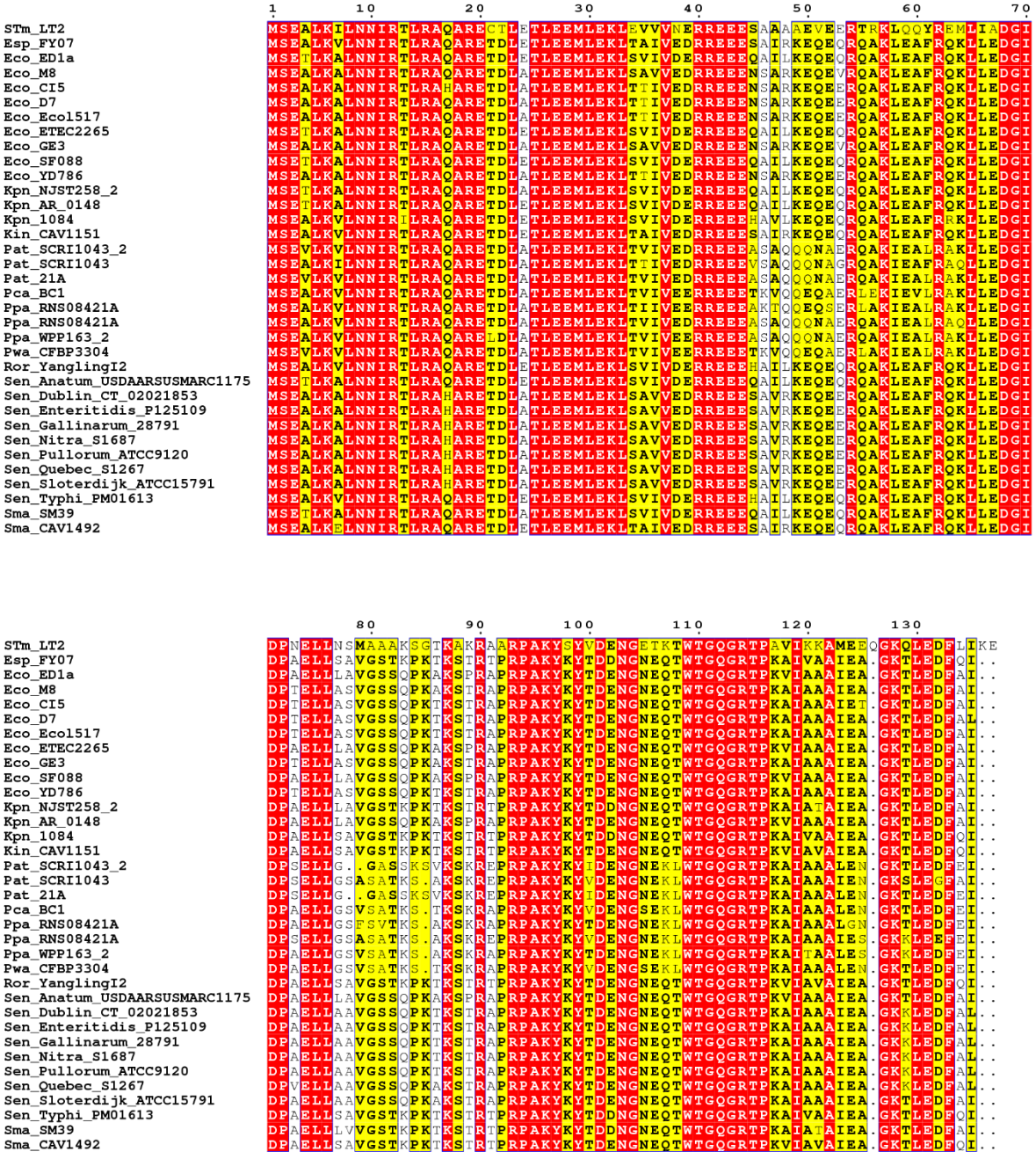


Figure S1: Multiple sequence alignment of H-NSB$_{EARL}$ proteins.

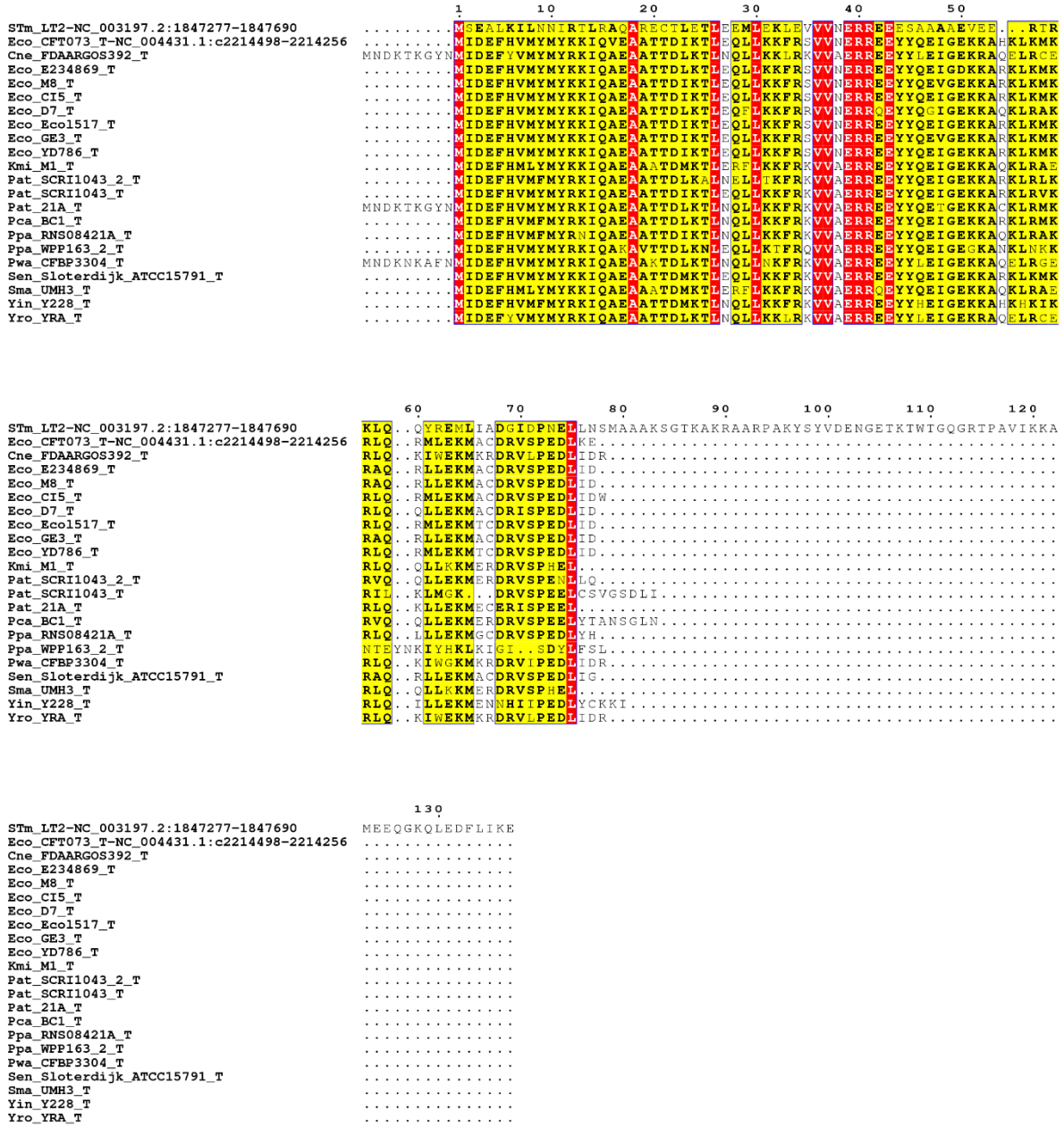


Figure S2: Multiple sequence alignment of H-NST$_{EARL}$ proteins.

Table S1: Xenogeneic silencer homologs found in genomic islands of the Islander database.

N	Accession N°	Strain	GI length	Island	Chrom. GC	Island GC	XS homolog	% Identity	Expect value	Phyllum/Class	Family	tRNA
1	NC_004431	Escherichia coli CFT073	22494	Eco160.22S	50.5	37.8	H-NS	63.7	7.23E-31	Gammaproteobacteria	Enterobacteriaceae	Ser
2	NC_013421	Pectobacterium wasabiae WPP163	33650	Pwa2.34N	50.5	37	H-NS	65.2	6.24E-30	Gammaproteobacteria	Enterobacteriaceae	Asn
3	NC_013716	Citrobacter rodentium ICC168	17184	Cro1.17S	54.7	36.9	H-NS	64.4	7.53E-30	Gammaproteobacteria	Enterobacteriaceae	Ser
4	NC_004547	Pectobacterium atrosepticum SCRI1043	28053	Pat3.28N	51	37.4	H-NS	63.0	9.28E-30	Gammaproteobacteria	Enterobacteriaceae	Asn
5	NC_011274	Salmonella ser. Gallinarum str. 287/91	27346	Sen212.27N	52.2	37.9	H-NS	63.0	1.60E-29	Gammaproteobacteria	Enterobacteriaceae	Asn
6	NC_011745	Escherichia coli ED1a	44629	Eco282.45N	50.7	39.8	H-NS	63.7	4.70E-29	Gammaproteobacteria	Enterobacteriaceae	Asn
7	NC_018522	Klebsiella pneumoniae subsp. pneumoniae 1084	29802	Kpn29.30N	57.4	37	H-NS	63.0	5.47E-29	Gammaproteobacteria	Enterobacteriaceae	Asn
8	NC_004547	Pectobacterium atrosepticum SCRI1043	40774	Pat3.41N	51	36.7	H-NS	60.7	3.40E-27	Gammaproteobacteria	Enterobacteriaceae	Asn
9	NC_014500	Dickeya dadantii 3937	55635	Dda1.56X	56.3	53.8	H-NS	60.0	1.91E-27	Gammaproteobacteria	Enterobacteriaceae	tmRNA
10	NC_016602	Vibrio furnissii NCTC 11218	26509	Vfu2.27L	50.7	42.4	H-NS	47.1	7.04E-18	Gammaproteobacteria	Vibrionaceae	Leu
11	NC_017626	Escherichia coli 042	54499	Eco33.54F	50.6	45.3	H-NS	50.0	2.18E-04	Gammaproteobacteria	Enterobacteriaceae	Phe
12	NC_011748	Escherichia coli 55989	60054	Eco98.60L	50.7	43.6	H-NS	40.0	4.69E-04	Gammaproteobacteria	Enterobacteriaceae	Leu
13	NC_010498	Escherichia coli SMS-3-5	43730	Eco824.44L	50.5	45.9	H-NS	42.6	6.12E-04	Gammaproteobacteria	Enterobacteriaceae	Leu
14	NC_011083	Salmonella ser. Heidelberg str. SL476	57873	Sen233.58F	52.1	45.4	H-NS	41.8	7.27E-04	Gammaproteobacteria	Enterobacteriaceae	Phe
15	NC_017631	Escherichia coli ABU 83972	88587	Eco134.89L	50.6	46.4	H-NS	41.8	7.94E-04	Gammaproteobacteria	Enterobacteriaceae	Leu
16	NC_003911	Ruegeria pomeroyi DSS-3	28536	Rpo1.29C	64.2	53.3	H-NS	30.9	0.036	Alphaproteobacteria	Rhodobacteriaceae	Cys
17	NC_007650	Burkholderia thailandensis E264	18872	Bth30.19S	68.1	58.1	H-NS	32.1	0.34	Betaproteobacteria	Burkholderiaceae	Ser
18	NC_017911	Pseudomonas fluorescens A506	75287	Pfl2.75S	60	43.2	MvaT	52.1	4.85E-25	Gammaproteobacteria	Pseudomonadaceae	Ser
19	NC_010322	Pseudomonas putida GB-1	43959	Ppu7.44S	61.9	49.3	MvaT	54.4	9.31E-23	Gammaproteobacteria	Pseudomonadaceae	Ser
20	NC_007492	Pseudomonas fluorescens Pf0-1	115595	Pfl11.116L	60.5	55.6	MvaT	50.8	1.56E-22	Gammaproteobacteria	Pseudomonadaceae	Leu
21	NC_010501	Pseudomonas putida W619	110820	Ppu18.111G	61.4	58.3	MvaT	50.6	3.25E-21	Gammaproteobacteria	Pseudomonadaceae	Gly
22	NC_017986	Pseudomonas putida ND6	124424	Ppu12.124P	61.8	59.3	MvaT	49.4	5.88E-21	Gammaproteobacteria	Pseudomonadaceae	Pro
23	NC_004129	Pseudomonas protegens Pf-5	16755	Ppr7.17L	63.3	51.5	MvaT	48.7	1.68E-20	Gammaproteobacteria	Pseudomonadaceae	Leu
24	NC_017532	Pseudomonas stutzeri DSM 4166	36084	Pst10.36	64	55	MvaT	50.6	1.28E-19	Gammaproteobacteria	Pseudomonadaceae	Not provided
25	NC_018177	Pseudomonas stutzeri DSM 10701	104468	Pst11.104L	63.2	60	MvaT	49.4	1.33E-18	Gammaproteobacteria	Pseudomonadaceae	Leu
26	NC_013757	Geodermatophilus obscurus DSM 43160	108343	Gob2.108T	74	70.3	Lsr2	47.7	2.27E-16	Actinobacteria	Geodermatophilaceae	Thr
27	NC_013757	Geodermatophilus obscurus DSM 43160	26379	Gob3.26P	74	70.5	Lsr2	44.1	4.37E-16	Actinobacteria	Geodermatophilaceae	Pro
28	NC_013235	Nakamurella multipartita DSM 44233	117538	Nmu1.118L	70.9	68.9	Lsr2	42.9	9.39E-14	Actinobacteria	Nakamurellaceae	Leu
29	NC_015312	Pseudonocardia dioxanivorans CB1190	59607	Pdi4.60I	73.3	71.2	Lsr2	42.6	4.53E-12	Actinobacteria	Pseudonocardiaceae	Ile

4. THE EXCISION OF ROD21 AND EXPRESSION OF ITS RECOMBINATION DIRECTIONALITY FACTOR ARE NEGATIVELY CORRELATED AND ARE MODULATED BY THE GROWTH PHASE

(Manuscript in preparation)

Excision of the pathogenicity island ROD21 is modulated by growth phase and negatively correlates with the expression of its integrase and recombination directionality factor

Alejandro Piña-Iturbe[a], Guillermo Hoppe-Elsholz[a], Paulina A. Fernández[b], Carlos A. Santiviago[b], Pablo A. González[a] and Susan M. Bueno[a]#

[a] Millennium Institute on Immunology and Immunotherapy, Facultad de Ciencias Biológicas, Departamento de Genética Molecular y Microbiología, Pontificia Universidad Católica de Chile, Santiago, Chile

[b] Laboratorio de Microbiología, Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile.

Running head: ROD21 excision negatively correlates to RDF expression

*Address correspondence to Susan M. Bueno, sbueno@bio.puc.cl

Abstract

Genomic islands (GIs) are horizontally transferred elements that shape bacterial genomes and contributes to adaptation to different environments. Some GIs encode the molecular machinery that permits their excision from the chromosome, the first step for their spread by horizontal transfer. Moreover, the excision process of some GIs has been found to play a role in the virulence of bacterial pathogens from plant and animals. While the excision process is thought to be similar to that observed in bacteriophages, this mechanism has been only studied in a few families of GIs. Here, we have characterized a gene-product of the pathogenicity island ROD21 from *Salmonella* ser. Enteritidis which, according to our results, is a recombination directionality factor (RDF). While this protein actually bound to the ROD21 regulatory regions involved in excision, one of which is expected to contain the integrase promoter, deletion of the gene encoding this RDF produced a modest decrease in the excision of the island and integrase expression. Surprisingly, we found that the expression of both the RDF and integrase encoding genes are negatively correlated to the excision of ROD21 which shows a growth-phase-dependent patter. Our results indicate that the excision process of ROD21 and the islands that belong to the same family of excisable GIs may differ to what we currently know about island excision. This study provide insight about the excision regulation of genomic islands in bacterial pathogens.

Importance

The excision of genomic islands from the bacterial chromosome is the first step for their horizontal transfer to other bacteria, spreading the advantageous functions encoded in the islands, such as colonization of new niches, antimicrobial resistance, or virulence factors, among others. Noteworthy, the expression of the genes carried by some genomic islands are modulated in response to the excised/integrated state of the island, as observed for the pathogenicity island ROD21 harbored by the foodborne pathogen *Salmonella* ser. Enteritidis. Since island excision plays important roles in horizontal transfer and pathogenesis, it becomes relevant to understand the mechanism behind its regulation in the different island families. Our research provides insight about the genetic factors that modulate the excision of ROD21 and the members that belongs to the same family of genomic islands which is harbored by a diversity of bacterial genera including current and emerging pathogens.

Introduction

Genomic islands (GIs) are genetic elements acquired by bacteria through horizontal gene transfer that are found integrated at the 5’ or 3’-end of different genes, including tRNA and tmRNA genes, and usually have sequence signatures (GC content, codon usage bias, dinucleotide frequency) that are distinct from those of the host genome (Williams, 2002; Che et al., 2014; Marcoleta et al., 2016; Piña-Iturbe et al., 2020). GIs carry genes that encode advantageous functions for their bacterial hosts such as those related to metabolism of new substrates, antibiotic resistance or pathogenicity (Piña-Iturbe et al., 2018; Regmi and Boyd, 2019; Curraize et al., 2020), which contribute to the survival and adaptation of bacteria in different environments. In addition to advantageous functions, many GIs also encode the molecular machinery responsible for their excision as circular extrachromosomal elements. Then, circular form of GIs can be reintegrated or be the substrate for a new event of horizontal transfer by conjugation or high-frequency transduction in hijacked bacteriophage capsids (Salazar-Echegarai et al., 2014; Fillol-Salom et al., 2019). Interestingly, besides its role in horizontal transfer, the excision of GIs can play a role in the virulence of plant and animal pathogens as found in *Pseudomonas syringae* pathovar *phaseolicola* and *Salmonella enterica* serovar Enteritidis, in which changes in the expression of genes inside and outside the islands have been observed in response to the integrated/excised state of the GI (Neale et al., 2018; Pardo-Roa et al., 2019).

The excision/integration mechanism of GIs is thought to be similar to the mechanism described for lysogenic bacteriophages since the excision/integration module of both genetic elements encode similar genes and regulatory regions. The two attachment regions that flank

the integrated Lambda phage, *attL* and *attR*, are the result of the recombination between the *attB* insertion site in the bacterial chromosome and the *attP* region in the circular phage DNA; *attL* and *attR* are populated by several binding sites for the different proteins involved in the assembly of the excisive intasome, the high-order nucleoprotein complex required for the excision of the prophage (Landy, 2015). Lambda- and host-encoded proteins bind the *attL* and *attR* regions to bend the DNA and promote intasome assembly, allowing the phage-encoded tyrosine recombinase, the integrase, to catalyze a site-specific recombination between the direct repeated sequences (DRSs) located at the end of each attachment region (Seah et al., 2014; Laxmikanthan et al., 2016). This recombination regenerates the insertion site *attB* and the circular form of the Lambda DNA and its *attP* region. While the Lambda integrase, participates in the integration and excision reactions, the phage-encoded recombination directionality factor (RDF) Xis is key for excision since it promotes the assembly of the excisive intasome inhibiting integration (Guarneros and Echols, 1970; Moitoso de Vargas and Landy, 1991). GIs are also flanked by DRSs and encode integrases that catalyzes the excision and integration of the island. Moreover, several GIs also encode one or more RDFs which participate in the excision process. While excision is being studied for some important families of GIs such as the SXT/R391, the MGIs, and more recently the tripartite symbiosis islands of *Mesorhizobium*, there are several other GI families in which the particularities of their excision/integration process and its role in bacterial pathogenesis remain unknown.

The Region of Difference (ROD) 21 is one of the several GIs harbored by *Salmonella enterica* serovar Enteritidis (Thomson et al., 2008). This island carry several genes (*SEN1970* to *SEN1999*) that encode functions related to virulence, conjugal transfer, integration/excision, and others of unknown function (Newman et al., 2006; Quiroz et al., 2011; Salazar-Echegarai

et al., 2014; Xiong et al., 2019). Interestingly, the excision of ROD21 plays an important role in the invasion of the gastrointestinal epithelium and the subsequent colonization of deep organs in mice by *Salmonella* ser. Enteritidis since mutant strains unable to excise ROD21 had a significantly reduced capacity to colonize the spleen, liver and gallbladder as compared to the wild-type strain in which ROD21 excised normally, specially at late stages of infection (Quiroz et al., 2011; Tobar et al., 2013; Pardo-Roa et al., 2019). This phenomenon may be due to a regulatory mechanism in which gene expression within ROD21 is modulated by the supercoiling of the island, which may differ between its integrated and excised states, as proposed for the PPHGI-1 island of *Pseudomonas syringae* pv. *phaseolicola* (Godfrey et al., 2011; Tobar et al., 2013). Furthermore, the excision of ROD21 seems to be upregulated at certain stages of the *Salmonella* ser. Enteritidis infection in the mouse model, hence the ROD21 excision may be part of a fine regulatory mechanism which assures that expression of the ROD21-encoded genes occurs at specific phases of the *Salmonella* ser. Enteritidis infective cycle (Nieto et al., 2016; Pardo-Roa et al., 2019).

The pathogenicity island ROD21 is part of the *Enterobacteriaceae*-associated ROD21-like (EARL) islands, a family of excisable GIs with a conserved excision/integration module, distributed among several strains of different bacterial families of the order Enterobacterales (Adeolu et al., 2016; Piña-Iturbe et al., 2018). The EARL GIs are present in plant and animal pathogens, including several clinically relevant bacteria, such as different *Salmonella enterica* strains including serovars Typhi and Enteritidis, carbapenem-resistant *Klebsiella pneumoniae* ST258 or intestinal/extraintestinal pathogenic *E. coli* (Piña-Iturbe et al., 2018). Therefore, studying the mechanisms important for the horizontal dissemination of genomic islands and their role in virulence, such as the excision process, may become relevant to understand and

counteract current and emerging pathogens. Here, we present an initial characterization of the gene *SEN1998* of ROD21, demonstrating that it encodes a small basic protein with the main features of an RDF, that binds the attachment regions of ROD21. While deletion of *SEN1998* produced a consistent, although modest, reduction of the ROD21 excision level in *Salmonella* ser. Enteritidis, we found that expression of *SEN1998* and the integrase-encoding gene *SEN1970* are negatively correlated with the excision of the island. Our results contribute to the understanding of the excision process in genomic islands and particularly in the EARL family.

Results and Discussion

The gene *SEN1998* of ROD21 encodes a recombination directionality factor

To gain insight about the genes involved in the excision of ROD21, we analyzed the predicted amino acid sequence of the protein encoded by *SEN1998*, a gene conserved across the EARL family of GIs, whose overexpression was found to increase the excision of ROD21 (Tobar et al., 2013; Piña-Iturbe et al., 2018). The multiple sequence alignment of SEN1998 (72 aa) with nine characterized RDFs and transcriptional regulators from phage and island origin showed the presence of conserved amino acid residues along the protein sequence from positions 11 to 60, including the conserved Phe-24, Tyr-30, Pro-39 and Arg-47, which are important for the RDF activity of TorI on the excision of the KplE1 prophage (Panis et al., 2012) (Fig. 1A). The resulting identities relative to SEN1998, ranged from 19.67 to 38.70%, with the higher values belonging to the RDFs of the *Vibrio cholerae* Seventh Pandemic Island II (VefA) and the *Yersinia pestis* High-pathogenicity Island (Hef) (Fig. 1B). Conversely, SEN1998 shared a lower identity with TorI and Xis, the RDFs of the KplE1 prophage and the Lambda phage, respectively. While the isoelectric point of RDFs can be variable, the majority are basic proteins, a feature that may favor their binding capacity to the negatively charged DNA sugar-phosphate backbone (Lewis and Hatfull, 2001). In agreement with this feature, the calculated isoelectric point of SEN1998 resulted to be highly basic (pI=10.20; Fig. 1B).

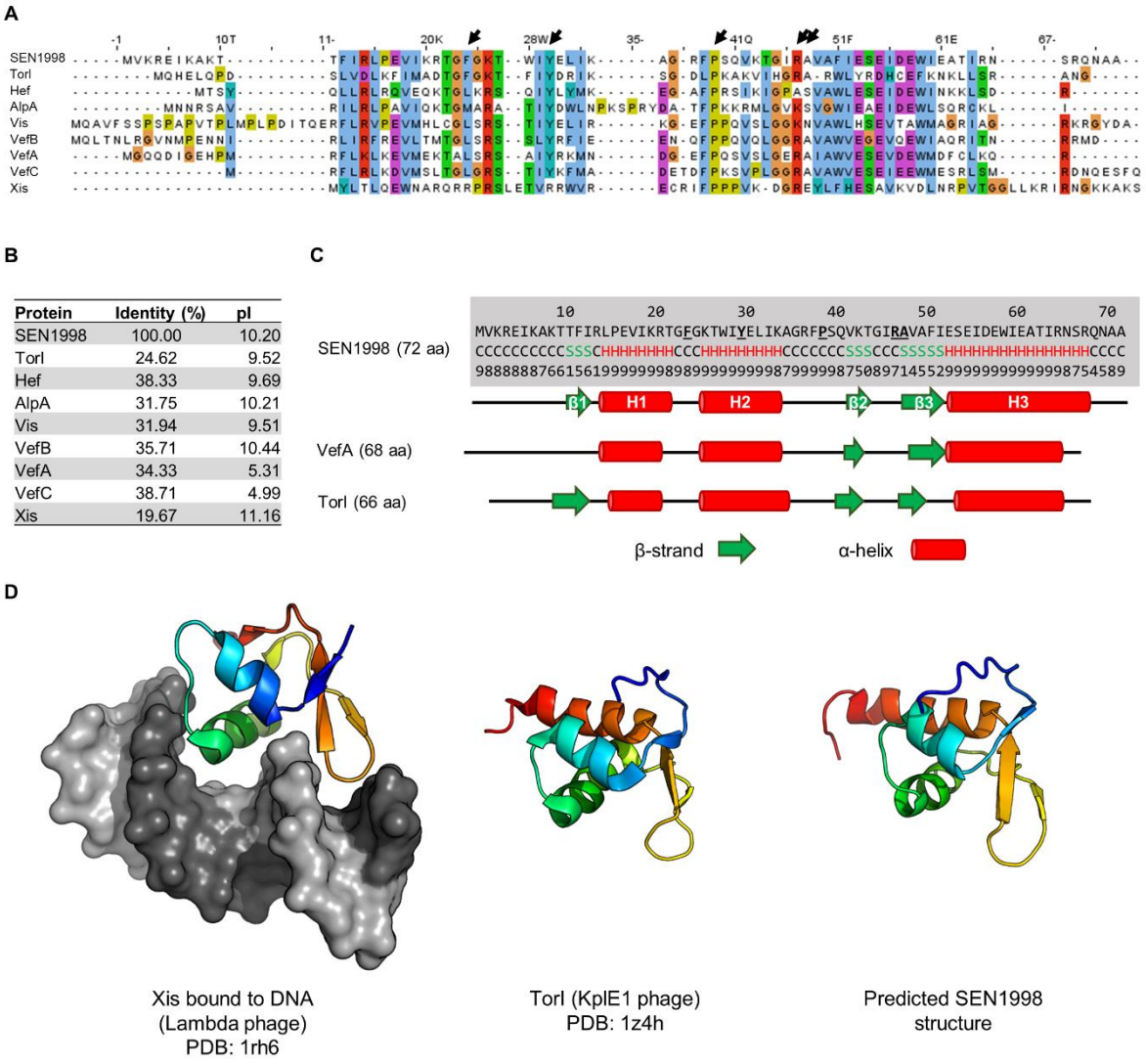


Figure 1. The gene *SEN1998* of ROD21 encodes a putative recombination directionality factor. (**A**) Multiple sequence alignment of the amino acid sequence of SEN1998 and characterized RDFs from phage (TorI, AlpA, Vis and Xis) and island (Hef, VefA, VefB and VefC) origin. Conserved residues are colored according to the ClustalX coloring scheme. The black arrows indicate the conserved residues that has been recognized as important for the RDF function of TorI. (**B**) Percent identity of the aligned proteins regarding SEN1998, and isoelectric point. (**C**) Comparison of the secondary structures of SEN1998 (predicted), VefA (predicted) (Carpenter et al., 2016), and TorI (experimentally determined) (ElAntak et al., 2005). (**D**) Tertiary structures of the RDFs: Xis bound to DNA, (experimentally determined) (Sam et al., 2004); TorI (experimentally determined) (ElAntak et al., 2005) and SEN1998 (predicted).

To further characterize SEN1998, the secondary and tertiary structures were predicted. The secondary structure of SEN1998 consisted of a β-strand (β1: Thr-11 to Ile-13) followed by two α-helices (H1: Leu-15 to Thr-22; H2: Lys-26 to Lys-34), two more β-strands (β2: Val-42 to Thr-44; β3: Ala-48 to Ile-52) and finally an α-helix longer than the previous ones (H3: Glu53 to Arg-68) (Fig. 1C). This arrangement of secondary structures is conserved among RDFs, such as TorI [PDB: 1z4h; (Panis et al., 2012) and the Vef proteins [predicted; (Carpenter et al., 2016)], where the helix H2 and the turn between strands β2 and β3 forms a winged-helix DNA-binding domain that enters the major and minor grooves, respectively, as observed in the structure of Xis [PDB: 1rh6; (Sam et al., 2004)] and TorI (ElAntak et al., 2005) (Fig. 1D). The predicted tertiary structure of SEN1998 is highly similar to that of TorI and Xis, and has the winged-helix motif, suggesting the DNA binding capacity of SEN1998. Together, the sequence analysis suggests that SEN1998 is an RDF.

The SEN1998 protein binds the two attachment regions of ROD21 with different affinities.

To further characterize the product of *SEN1998*, we cloned its coding sequence in a pET-15b vector and purified the 6x-His-tagged recombinant protein (Supplemental Fig. S1). Then, we assessed the DNA-binding capacity of SEN1998 through electrophoretic mobility shift assays (EMSAs). Since most of the studied RDFs bind several binding sites at one of the attachment regions of their prophage or GI to promote excision (Panis et al., 2007; Carpenter et al., 2016), we decided to use the left and right attachment regions of ROD21 as the target DNA. Interestingly, we found that SEN1998 can bind the two attachment regions of ROD21, forming

different complexes with different mobilities (Fig. 2). This suggests the presence of more than one binding site in each *att* region and, therefore, the participation of more than one SEN1998 protein in the assembly of the excisive intasome of ROD21. This finding is in agreement with observations made for prophages, whose RDFs bind at multiple sites to assist formation of the DNA curvature necessary for intasome assembly (Abbani et al., 2007). The SEN1998-*att* complexes became visible at 3.5 µM for *attL* and 0.7 µM for *attR*, suggesting a higher binding affinity of SEN1998 for the *attR* region (Fig. 2B). We cannot rule out that SEN1998 could bind the *att* regions in a cooperative manner (i.e. the binding of one SEN1998 favoring the binding of the second) as observed for Xis in the *attR* region of the integrated Lambda phage (Bushman et al., 1984; Abbani et al., 2007), since our protein harbors an N-terminal 6xHis that could interfere in the establishment of the required interactions. Nevertheless, since cooperativity can be seen in EMSA as the formation of fully saturated protein-DNA complexes in a narrow range of protein concentrations (Bushman et al., 1984), the permanence of the different complexes of SEN1998-*att* across a wide range of protein concentrations (0.7 to 7.0 µM) in our experiment is indicative of a lack of cooperativity. Our results also suggest that the binding sites of SEN1998 are located at the distal part of the *att* regions (i.e. closer to the DRSs), since binding of SEN1998 to shorter target DNAs was observed only for those that span the island boundaries (Fig. 2C).

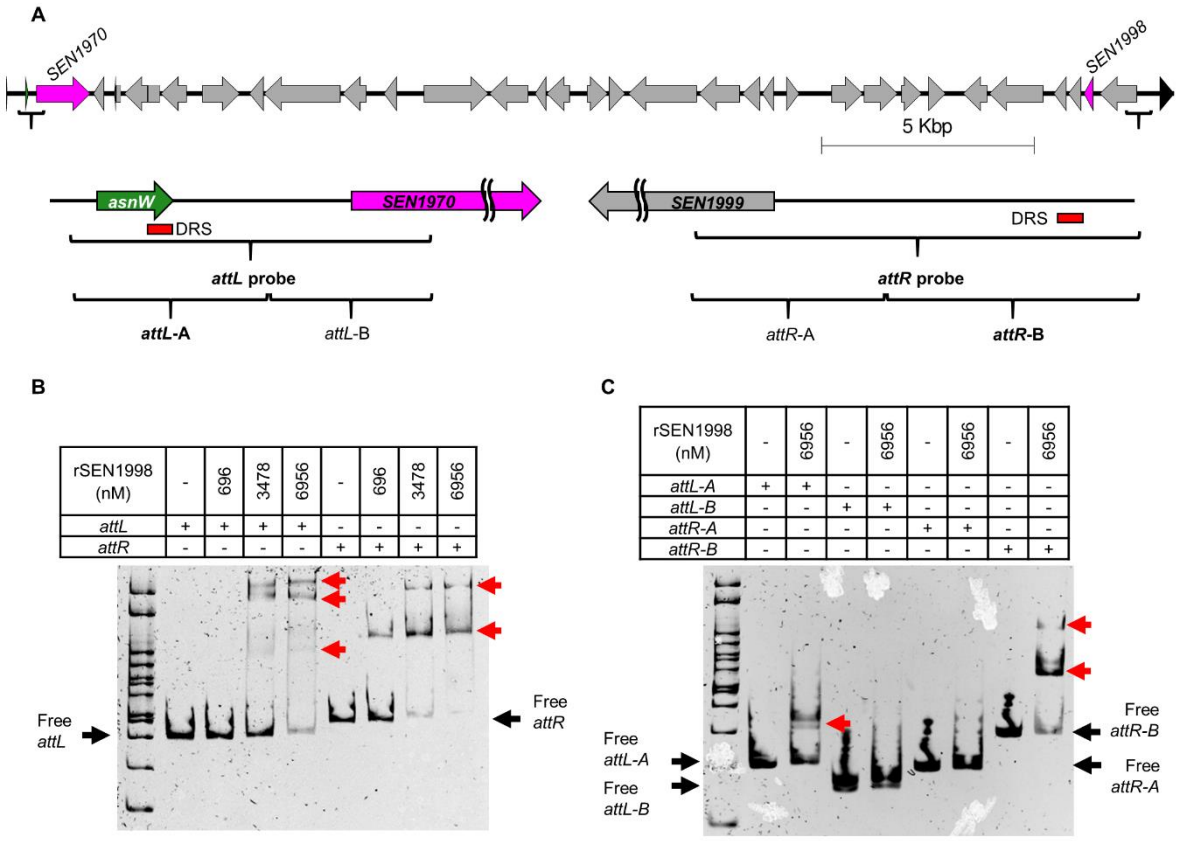


Figure 2. The SEN1998 protein binds the two attachment regions of ROD21. (**A**) Schematic representation of the target DNAs encompassing the left and right attachment regions of ROD21 used for the EMSAs. (**B**) EMSAs showing the binding of SEN1998 to both *att* regions, forming different complexes. The black and red arrows indicate the free and complexed target DNA.

The excision of ROD21 significantly differs in strains lacking and overexpressing *SEN1998*

Previous studies have shown that overexpressing SEN1998 increased the excision level of ROD21 within the bacterial population (Tobar et al., 2013; Salazar-Echegarai et al., 2014). To further assess how SEN1998 contributes to the excision of ROD21, we constructed two strains of *Salmonella* ser. Enteritidis PT4: one mutant strain lacking the coding sequence of SEN1998 (Δ*SEN1998*::*frt*) and the isogenic complemented strain that harbors the plasmid pTrc-*SEN1998* for IPTG-inducible expression (Supplemental Fig. S2).

The excision of GIs is usually expressed as the ratio between the copy number of the empty *attB* site and a single-copy gene (e.g. *attB*/*rpoD*), a value that varies between 0 (all *attB* sites occupied; *attB*=0) and 1 (all *attB* sites empty; *attB*=*rpoD*). Since this ratio is an estimation of the population fraction with an excised GI, we decided to represent it as a percentage of the population (*attB*/100 *rpoD*) to provide a straightforward representation of the excision level within the bacterial population. The excision level (as percent relative frequency) in the wild-type strain *Salmonella* ser. Enteritidis P125109 ranged from 4.8% to 7.7% of the bacterial population (Fig. 3A; blue symbols). While this result is in accordance with previous observations (Quiroz et al., 2011; Tobar et al., 2013; Piña-Iturbe et al., 2018), is to be noted that the excision frequency of ROD21 in LB is considerably high as compared to other GIs whose excision frequencies in uninduced conditions (in different media) are usually very low (<0.1%) (Ramsay et al., 2006; Daccord et al., 2012; Lautner et al., 2013; Marcoleta et al., 2016; Haskett et al., 2018). When we deleted the *SEN1998* coding sequence, the excision decreased slightly

(4.4% to 6.6%) and, although the differences were not statistically significant, this trend was observed in 11 out of 12 different measures (Fig. 3A, red symbols). When deletion of *SEN1998* was complemented from a plasmid, the excision levels increased again, ranging from 5.2% to 8.6%, and equaled or surpassed the excision level of the wildtype strain (Fig. 3A, purple symbols). Expression of *SEN1998* from the pTrc-*SEN1998* plasmid was leaky, and reached very high levels, from 1396 to 37666 times the wildtype level in uninduced and 1.0 mM IPTG conditions, respectively (Supplemental Fig. S3). Is possible that the high expression levels of *SEN1998* could interfere with the synthesis of functional proteins, which may explain why, in other studies, the overexpression of *SEN1998* from the pBAD-*SEN1998* plasmid induced a higher excision level in the wild-type strain (≈20%) (Tobar et al., 2013). Nevertheless, the reduction of the excision frequency in the mutant strain and the recovery in the complemented strain supports the involvement of SEN1998 in the excision of ROD21. Indeed, the two-way ANOVA indicated that statistically significant differences existed between the means regarding the "strain" factor ($p=0.0240$). The Tukey post-hoc test comparing the main effect of the "strain" factor revealed statistically significant differences ($p=0.0462$) between the excision levels of the mutant and complemented strains.

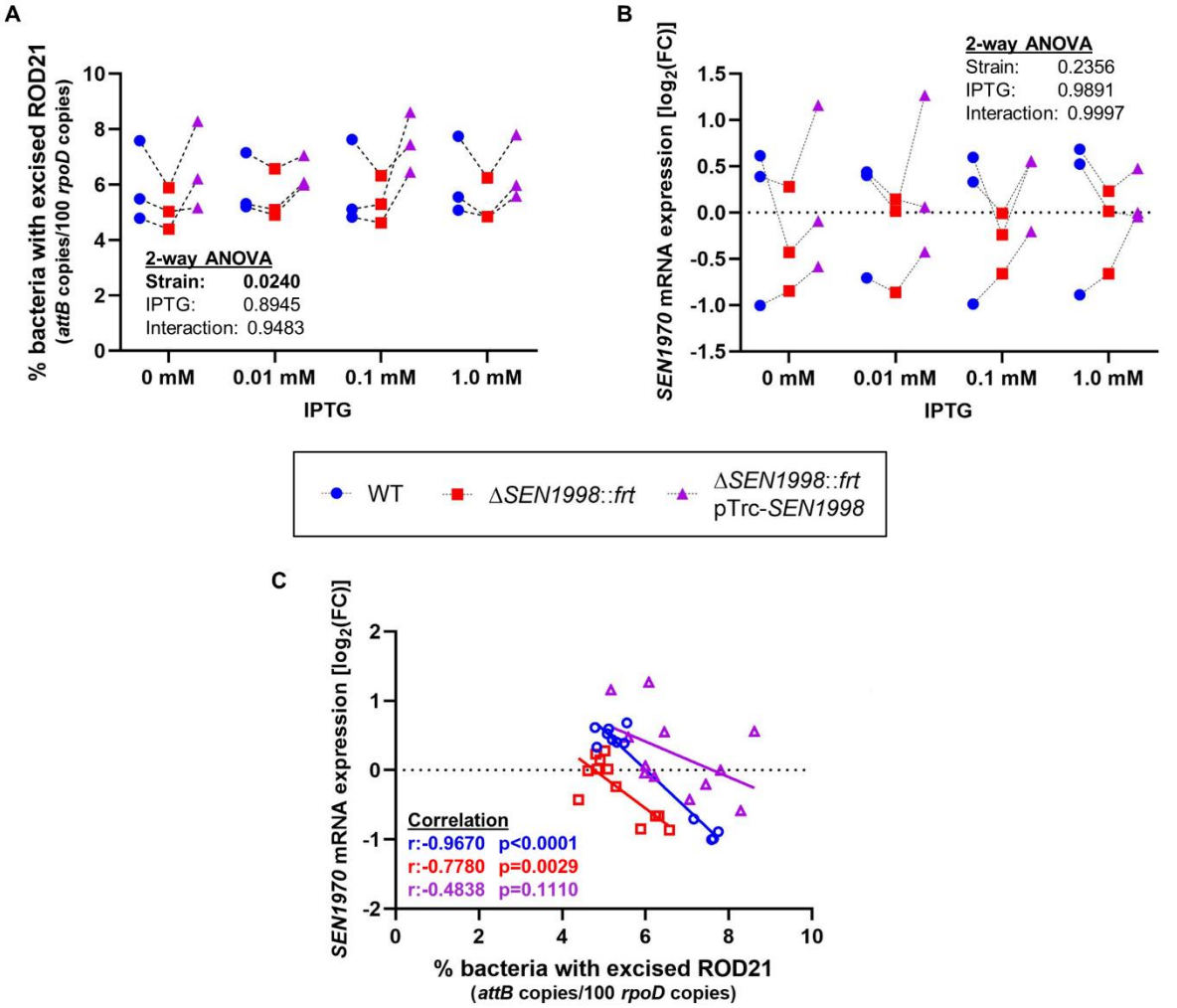


Figure 3. The absence of SEN1998 produces a slight decrease in the excision of ROD21 and the expression of *SEN1970*. (**A**) Excision level of ROD21, represented as the percentage of the bacterial population with an empty *attB* site, in *Salmonella* ser. Enteritidis P125109, Δ*SEN1998*::*frt* and Δ*SEN1998*::*frt* pTrc-SEN1998, under different concentrations of IPTG. (**B**) mRNA expression of the integrase-encoding gene *SEN1970* in the three strains, relative to the expression in P125109. Each point in A and B, at a specific IPTG concentration, represents an independent experiment. The excision or expression values obtained in a same experiment are liked by broken lines. (**C**) Scatter plot of the relative *SEN1970* mRNA expression versus the excision level in the same sample for the three *Salmonella* strains. The Pearson correlation coefficient and p-value are shown for each strain.

It has been reported that deletion of GI-encoded RDFs usually produces a significant to moderate decrease of the excision frequency and, in many cases, the absence of the RDF does not abolish the excision capacity, indicating that these proteins are not essential (Lesic et al., 2004; Antonenka et al., 2006; Ramsay et al., 2006; Daccord et al., 2012; Carpenter et al., 2016; Haskett et al., 2018). For example, the island IE3 carried by *E. coli* E2348/69, is an EARL GI that carries the CDS of a SEN1998 homolog, but a point mutation in the start codon (ATG>ACG) has disrupted the CDS. We have previously demonstrated that although IE3 do not encode its putative RDF, this island is excision-proficient (Piña-Iturbe et al., 2018). It is possible that a different RDF encoded outside ROD21 could participate in its excision (Carpenter et al., 2016); however, a tBLASTn search of SEN1998 homologs in the *Salmonella* ser. Enteritidis P125109 (GenBank accession NC_011294.1) did not retrieve any homolog sequence (data not shown).

The expression of *SEN1998* and the integrase-encoding gene *SEN1970* are negatively correlated with the excision of ROD21

Is not clear why a gene encoding an RDF-like protein not essential for excision is highly conserved among all the GIs belonging to the EARL family (Piña-Iturbe et al., 2018). Is possible that SEN1998, and their homologs, could play additional roles important for the excision or transfer of their cognate GI. For example, in the KplE1 prophage and the VPI-1 and VPI-2 GIs, it has been observed that their RDFs can negatively modulate the expression of the integrase as a result of the overlap between the *att* region bound by the RDF, and the *int* promoter (Panis et al., 2010; Carpenter et al., 2016). Therefore, we aimed to assess whether *SEN1998* can modulate

the expression of *SEN1970*, since the promoter of the integrase-encoding gene *SEN1970* is expected to be in the *attL* region bound by SEN1998 (Fig. 2B). Quantification of the *SEN1970* expression in the Δ*SEN1998*::*frt* and complemented strains revealed slight variations that were not statistically different from the expression level in the wildtype strain (Fig. 3B). Interestingly, the same trends observed for excision were also present in the expression of the integrase-encoding gene: lower levels of *SEN1970* expression in the strain lacking *SEN1998* (9 out 12 measures; Fig. 3B, red symbols) and a higher expression in the complemented strain (10 out of 12 measures; Fig. 3B, purple symbol). Although these results may suggest a role of SEN1998 in the expression of the integrase SEN1970, it would not be as a negative regulator but an activator. However, is possible for the observed changes to be a consequence of the island excision and not a result of the expression of *SEN1998*. Interestingly, we found a negative correlation between the expression of *SEN1970* and the excision of ROD21 (Fig. 3C), which is: the higher the excision, the lower the expression of the integrase-encoding gene in the same sample. This finding is in agreement with a previous study in which the increase of the ROD21 excision, in presence of oxygen peroxide, correlated with a reduction in the expression of *SEN1970* (Quiroz et al., 2011). Experiments using expression reporters would be useful to evaluate the role of SEN1998 as a modulator of *SEN1970* expression.

The excision of ROD21 and the expression of *SEN1998* and *SEN1970* are modulated by the growth phase

To further explore the role of *SEN1998* in the excision of ROD21, we determined the excision pattern of the island along the growth curve of the wild-type strain and quantified the

expression of *SEN1998* and *SEN1970* (Fig. 4; black line). We found that the excision reaches its lower values during the logarithmic phase (between 2.1-2.6%) and then increases in the transition to stationary phase to stabilize in a maximum value during the stationary phase (around 4.5%). While not all GIs have an excision pattern like this [see for example (Marcoleta et al., 2016)], the excision increase upon entrance to stationary phase has been observed for the ICE*Ml*Sym^{R7A} symbiosis island of *Mesorhizobium loti* R7A (Ramsay et al., 2006) and the ICE*St1* and ICE*St3* islands of *Streptococcus thermophilus* CNRZ368 and CNRZ385 (Carraro et al., 2011). Interestingly, while the excision increase of ICE*St1* and ICE*St3* can be explained by the higher expression levels of the genes encoding their corresponding integrases and RDFs (Carraro et al., 2011), when we quantified the expression of the *SEN1998* and *SEN1970* genes a marked decreases in the expression of both genes was found (Fig. 4; green and orange lines). Relative to time 0.5 h, the expression of *SEN1970* and *SEN1998* showed a reduction of about 5 and 70 times their initial level, respectively. The decrease in the expression of these genes occurred during the log phase while the excision of ROD21 remained nearly constant, and then remained with little variation along the transition and stationary phases (Fig. 4). These results support our previous finding of the negative correlation between the expression of *SEN1970* and the excision of ROD21 and expand it to include the expression of *SEN1998*, which is greatly downregulated as compared to *SEN1970*. In other study, it was found that the supercoiling of the excised form of the PPHGI-1 island of *Pseudomonas syringae* pv. *phaseolicola* was involved in the repression of the genes carried by this GI (Neale et al., 2018). We have previously observed expression changes in the genes contained within ROD21 in response to an impaired excision capacity (Tobar et al., 2013; Piña-Iturbe et al., 2018; Pardo-Roa et al., 2019), which suggest a role of the excised/integrated state of ROD21 in modulation of gene expression.

Nevertheless, the downregulation of *SEN1970* and *SEN1998* observed in our experiment may not be related to the excised/integrated state of ROD21 since the decrease in the expression levels occurred as the excision level was nearly constant (Fig. 4). While these results suggest a still unknown regulatory mechanism that repress the expression of *SEN1970* and *SEN1998* during the phase of active growth, it is not clear how these changes affect the excision of ROD21 which starts to increase as the expression of both genes stabilizes around its minimum. Assessing the protein levels of both SEN1970 and SEN1998, which we speculate do not correlates to gene expression, would shed some light about the conditions modulating the excision of ROD21 along the growth curve.

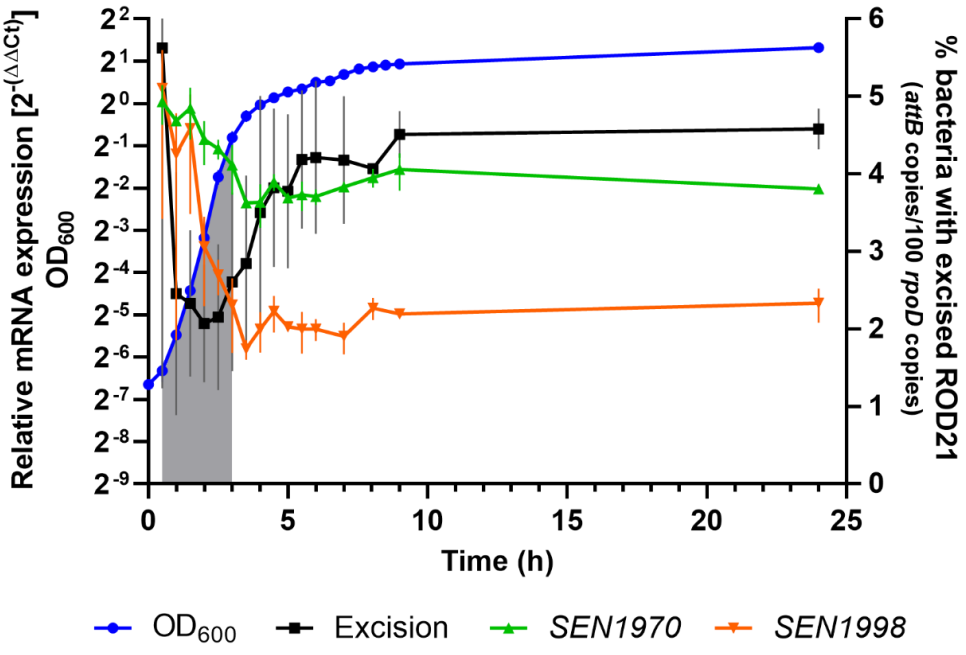


Figure 4. The excision of ROD21 increase during the stationary phase of growth as the expression of *SEN1970* and *SEN1998* decreases. ROD21 excision, and *SEN1970* and *SEN1998* expression patterns during the growth of *Salmonella* ser. Enteritidis P125109 in LB at 37°C with shaking. The grey area represents the logarithmic phase of growth. Values are the mean ± SD of two independent experiments.

Since the genes *SEN1998* and *SEN1970*, in addition to the attachment regions, are conserved across the EARL GIs (Piña-Iturbe et al., 2018), is possible that our findings would not be limited to ROD21. However, as this family of GIs has been recently described, our understanding of their roles within their hosts is still incipient, and more research is needed. A case of interest would be the ICEKp258.2 island harbored by the globally spread carbapenem-resistant *Klebsiella pneumoniae* ST258. The acquisition of this EARL member by the ancestral ST258 strain, in the last decade of the 20th century (Bowers et al., 2015; Zhang et al., 2019), has been suggested to be related to its successful dissemination around the world (Chen et al., 2014). Nevertheless, little is known about this excisable island. We hypothesize that, as observed for ROD21 in *Salmonella* ser. Enteritidis (Tobar et al., 2013; Pardo-Roa et al., 2019), the excision of the EARL GIs, including ICEKp258.2, modulates the expression of the genes carried by them, playing a role in the pathogenesis of their bacterial hosts in response to the conditions found in the environment. The excision of different families of GIs harbored by Gram-negative and Gram-positive bacteria is currently being studied by different groups, and a better understanding of the excision mechanism has proven useful to gain insight about the dissemination of virulence and resistance traits, the interaction of GIs with other horizontally acquired elements, and the evolution of bacterial pathogens (Kiss et al., 2015; Penadés and Christie, 2015; Carraro et al., 2016; Haskett et al., 2018; McKitterick and Seed, 2018). Our findings add to the growing body of knowledge about the excision of GIs, providing insights about a recently discovered family of islands harbored by pathogenic strains of Enterobacterales.

Materials and Methods

Bacterial strains, plasmids and culture conditions

Salmonella enterica serovar Enteritidis phage type 4 P125109, *Escherichia coli* BL21(DE3) and *E. coli* DH5α, were maintained in CRYOBANK tubes (VWR) at -80°C and cultured in LB broth with agitation or LB agar, at 37°C or 30°C, when needed. Plasmid pET15b-*SEN1998*, encoding a N-terminal 6xHis-tagged SEN1998 protein, was synthesized by Epoch Life Science and transformed by electroporation in *E. coli* BL21(DE3) for protein expression. Plasmids pKD3, pKD46 and pCP20, used for deletion of the *SEN1998* coding sequence, were maintained in *E. coli* DH5α and purified using the Wizard SV DNA Minipreps (Promega). When necessary, chloramphenicol (CHL) or ampicillin (AMP) were added to the culture media at 34 µg/mL and 100 µg/mL, respectively.

The mutant strain *Salmonella* ser. Enteritidis PT4 Δ*SEN1998*::*frt* was constructed using the Lambda Red system (Datsenko and Wanner, 2000) using the primers listed in the Supplemental Table S1. The chloramphenicol acetyl transferase (*cat*) cassette was amplified from the template plasmid pKD3 by PCR, using primers H1+P1_SEN1998_Fw and H2+P2_SEN1998_Rev, and purified using the Wizard SV PCR and Gel Clean-up System (Promega). *Salmonella* ser. Enteritidis PT4 P125109, harboring the pKD46 helper plasmid, was grown in LB broth at 30°C until OD_{600}=0.3. Then, 1mM L-arabinose was used to induce expression of the Lambda Red proteins from pKD46 during hours. Electrocompetent cells were prepared and transformed with the purified *cat* cassette, and then incubated in approximately 2 mL LB broth at 30°C during 2 hours. After recovering, cells were pelleted, spread on LB-CHL agar, and incubated at 30°C. After 72 hours, CHL-resistant colonies were selected and grown in

LB-CHL agar at 37 °C overnight. After incubation, one colony per clone was selected and subcultured again in LB-CHL agar at 37°C. Liquid cultures in LB-CHL were prepared and used for DNA extraction and preparation of frozen stocks. Replacement of the *SEN1998* gene by the *cat* cassette was confirmed by PCR using specific primers for *SEN1998* (SEN1998_NcoI_Fw and SEN1998_HindIII_Rev), the *cat* cassette (H+P primers), primers external to the mutation site (SEN1997 3' Fw and SEN1999 5' Rev), and a combination of H1+P1_SEN1998_Fw and SEN1999 5' Rev.

The mutation was transferred to a clean genetic background by generalized transduction using phage P22 H5. Transducing P22 particles were obtained by infecting the *Salmonella* ser. Enteritidis Δ*SEN1998*::*cat* strain in mL of LB-CHL broth with agitation at 37°C, overnight. The culture was vigorously shaken, pelleted, and the clear supernatant was recovered, passed through a 0.22 µm syringe filter and stored at 4°C with. One hundred microliters of an overnight culture of wild-type *Salmonella* ser. Enteritidis P125109 were mixed with 900 µL of the transducing P22 suspension and incubated at 37°C during 60 minutes. Then, 100 µL were plated on LB-CHL agar and incubated at 37°C. CHL-resistant colonies were selected, and the presence of the *cat* cassette was confirmed by PCR as described above. The selected transductants were assessed for the presence of lysogens and pseudolysogens by plating on Evans blue agar and testing susceptibility to phage P22 H5. Finally, the lysogen/pseudolysogen-free transductant was electrotransformed with plasmid pCP20 to eliminate the *cat* cassette by Flp-mediated excision. The resulting *Salmonella* ser. Enteritidis PT4 Δ*SEN1998*::*frt* strain was verified by PCR, and CRYOBANK stocks were prepared and stored at -80°C.

Bioinformatic analysis

Amino acid sequences of proteins SEN1998 (CAR33580.1), VefA (AAF94934.1), VefB (AAF94957.1), VefC (AAF93670.1), Hef (CAB46594.1), AlpA (NP_417113.1), Vis (NP_042041.1), Xis (NP_040610.1) and TorI (WP_001163428.1) were downloaded from GenBank. The isoelectric point of the nine proteins were calculated using the tool Compute pI/Mw from ExPASy (https://web.expasy.org/compute_pi/). The multiple sequence alignment was performed in MEGA X v.10.0.5 (Kumar et al., 2018) using MUSCLE, and the resulting alignment was opened and colored in JalView v.2.10.3b1 (Waterhouse et al., 2009). Finally, prediction of the SEN1998 secondary and tertiary structures was carried out using the I-TASSER server (Roy et al., 2010)

Protein expression and purification

E. coli BL21(DE3) pET15b-SEN1998 was grown in 750 mL of LB-AMP broth at 37°C and ≈220 rpm until OD_{600}=0.8-1.0. The culture was cooled using ice, and the expression of the recombinant N-terminal-6xHis SEN1998 (rSEN1998) was induced with 1 mM IPTG during 20 h at room temperature (19-22 °C) and 220 rpm. Cells were harvested by centrifugation at 8000 rpm at 4 °C during 6 min. and stored at -30 °C until needed.

The cell pellet was resuspended in lysis buffer [20 mM phosphate buffer, 100 mM NaCl, 30 mM imidazole, 0.5 mM Tris(2-carboxyethyl)phosphine, 5% glycerol, ≥ 25U/mL Benzonase, 2 mM $MgCl_2$, 1X cOmplete Protease Inhibitor Cocktail; pH 8.0] and incubated with 1 mg/mL lysozyme during 30 min. at 37°C and 200 rpm. After incubation, the suspension was cooled on ice and then was sonicated with 16 pulses in groups of 4 pulses of 15 seconds each, avoiding

overheating. The lysate was cleared by two steps of centrifugation at 4°C (8000 rpm/6 min, and 11000 rpm/45 min) and the supernatant was recovered and stored at 4 °C overnight.

rSEN1998 was purified by immobilized metal-ion affinity chromatography using HisTrap Fast Flow columns (Merck-Millipore). The column was equilibrated with 10 column volumes (CVs) of buffer A (20 mM phosphate buffer, 500 mM NaCl, 30 mM imidazole, 0.5 mM TCEP, 5% glycerol; pH 8.0), and the recovered supernatant was loaded. The column was washed with 30 CVs of buffer B (20 mM phosphate buffer, 500 mM NaCl, 150 mM imidazole, 0.5 mM TCEP, 5% glycerol; pH 8.0) and rSEN1998 was eluted with 10 CVs of Buffer C (20 mM phosphate buffer, 500 mM NaCl, 300 mM imidazole, 0.5 mM TCEP, 5% glycerol; pH 8.0). Eleven fractions of 1 mL were collected and those with the higher amount of rSEN1998 were pooled and dialyzed against Binding buffer (10 mM Tris-HCl, 150 mM KCl, 0.1 mM EDTA, 0.5 mM TCEP, 5% glycerol; pH 8.0). The protein was concentrated in an Amicon™ Ultra-4 centrifugal filter (MWCO = 3.5 kDa; Merck) and Binding buffer containing 50% glycerol was added to obtain a final concentration of 20% glycerol. Protein concentration was determined using the Pierce™ BCA Protein Assay (Merck). The purity of rSEN1998 was estimated from a 15% polyacrylamide gel stained with Coomassie blue using ImageJ v.1.52a (Schneider et al., 2012). The protein was stored in aliquots at -80 °C.

Electrophoretic mobility shift assays

The *Salmonella* ser. Enteritidis P125109 DNA containing the left and right attachment regions of ROD21 was amplified by PCR using the Phusion High-Fidelity DNA Polymerase (Thermo Scientific) with the HF Buffer and the primers attL_Fw/attL_Rev and attR_Fw/attR_Rev, respectively (Supplemental Table S1). The amplified regions correspond to

the positions 2061048 to 2061443, and 2087253 to 2087698 in the *Salmonella* ser. Enteritidis P125109 chromosome (GenBank Accession Number NC_011294). The *attL* (360 bp) and *attR* (446 bp) PCR products were purified directly from the PCR tube using the Wizard® SV Gel and PCR Clean-Up System (Promega) and stored at -30 °C in nuclease-free water.

One hundred nanograms of the *attL* and/or the *attR* DNA were incubated with 0, 440, 880, 1761, 4402 and 8804 nM rSEN1998 in Binding buffer (10 mM Tris-HCl, 150 mM KCl, 0.1 mM EDTA, 0.5 mM TCEP, 5% glycerol; pH 8.0) at room temperature during 60 minutes in a total volume of 20 µL. After incubation, the total volume was loaded in a native 6% polyacrylamide (37.5:1) gel prepared with 0.5X Tris-Borate-EDTA (TBE) buffer (Merck) that has been pre-run at 100 V during 60 min in 0.5X TBE. Samples were run at 70 V during 150 min and the gel was stained during 20 min in water with 5X GelRed® (Biotium) and visualized with UV light in a myECL™ Imager (Thermo Scientific). EMSAs with target DNAs corresponding to subregions of the *attL* and *attR* targets (*attL*-A, *attL*-B, *attR*-A and *attR*-B) were performed as described above, using purified PCR products obtained with primers attL_Fw/attL-A_Rev, attL-B_Fw/attL_Rev, attR_Fw/attR-A_Rev and attR-B_Fw/attR_Rev (Supplemental Table S1).

Effect of *SEN1998* deletion on the expression of the integrase-coding gene *SEN1970*

Overnight cultures of *Salmonella* ser. Enteritidis PT4 wildtype, Δ*SEN1998*::*frt* and Δ*SEN1998*::*frt* pTrc-SEN1998 strains were grown in LB broth (37°C, with shaking) until OD_{600}≈0.6. Five mL of each culture were transferred to 50 mL conical tubes and IPTG was added to obtain final concentrations of 0.01, 0.1 or 1.0 mM. One tube remained without IPTG as a control. After 2 h incubation at the same conditions, 1.5 mL samples were taken and

centrifuged at 8000 rpm during 6 min at 4°C. Supernatants were discarded, the pellets for DNA extraction were stored at -30°C and the pellets for RNA extraction were resuspended in 1 mL TRIzol (Invitrogen) and stored at -80°C. Three independent experiments, performed in duplicate, were carried out.

Assessment of the *SEN1970* and *SEN1998* expression and excision of ROD21 patterns along the growth curve

An overnight culture of *Salmonella* ser. Enteritidis PT4 was grown in 600 mL of LB broth, starting at OD_{600}=0.01, during 24 h at 37°C. Samples were taken each 30 minutes from time 0.5 h to 6.0 h, each hour from 7.0 h to 9.0 h, and then at time 24 h, and OD_{600} was measured. Samples were centrifuged at 8000 rpm during 6 min at 4°C, the supernatant discarded, and pellets were stored at -30°C for DNA and -80°C in 1 mL TRIzol for RNA. Two independent experiments were carried out.

Extraction and purification of genomic DNA and total RNA

Genomic DNA was purified from bacterial pellets stored at -30 °C. Cells were resuspended in 550 µL of Tris-EDTA buffer (pH 8.0) followed by the addition of µL of RNase A (5 mg/mL), µL of proteinase K (10 mg/mL) and µL of 10% SDS. After incubation at 37°C during 1 h, 600 µL (1 volume) of phenol:chloroform:isoamyl alcohol (PCI; Winkler) were added for DNA extraction. The tubes were vigorously shaken and centrifuged at 14 800 rpm during 15 min at 4°C. The transparent aqueous phase (300 µL) was transferred to a new tube to repeat the extraction with PCI. This step was repeated again with 200 µL of the aqueous phase. After recovering 100 µL of the aqueous phase from the third extraction, 0.1 volumes of 3 M sodium acetate and 1 volume of propan-2-ol were added, followed by incubation at -20 °C during 30

min for precipitation of DNA. Tubes were centrifuged at 14 800 rpm during 5 min. at 4°C and the DNA pellet was washed with 1 mL of 75% ethanol by inverting the tube five times. After centrifugation at 7500 x g during 5 min at 4 °C, the DNA pellet was let to dry for 5 min at room temperature, resuspended in 50 µL of nuclease-free water and stored at -30 °C.

Total RNA was purified from bacterial pellets resuspended in 1 mL of TRIzol reagent stored at -80°C, according to the manufacturer instructions. RNA was resuspended in nuclease-free water, treated with the Turbo^{TM} DNase (Invitrogen) and stored at -80 °C.

Quantification of ROD21 excision and gene expression

The excision of ROD21, expressed as the relative frequency of chromosomes containing an empty *attB* integration site, was measured by quantitative PCR (qPCR) using hydrolysis probes (Supplemental Table S1) and the TaqMan® Fast Advanced Master Mix (Applied Biosystems). The relative frequency resulted from dividing the number of *attB* sites by the number of *rpoD* genes in the sample. As *rpoD*, which encodes the σ^{70} factor of the RNA polymerase, is a single-copy gene, its quantification approximates the total number of chromosomes. Therefore, the frequency of excision can vary from 0 (all chromosomes have an integrated ROD21) to 1 (all chromosomes have an empty *attB* site). The absolute numbers of both *attB* and *rpoD* in each sample were determined using a standard curve made with ten-fold dilutions of the *Salmonella* ser. Typhimurium 14028s genomic DNA, which lacks ROD21 and contain only one copy of *attB* and *rpoD* per chromosome. The attB-1 probe and primers attB1-RT-Fw/Rev were used for detection of the *attB* site and the rpoD probe and primers rpoD-RT-Fw/Rev for *rpoD*.

Gene expression was quantified by qRT-PCR with hydrolysis probes SEN1970-RT and SEN1998-RT, and the primers SEN1970-RT-Fw/Rev and SEN1998-RT-Fw/Rev (Supplemental Table S1). cDNA was synthesized using the iScript^{TM} cDNA Synthesis Kit (Bio-Rad) according to the manufacturer instructions. Relative expression was calculated by the $2^{-(\Delta\Delta Ct)}$ method using the expression of *rpoD* as the endogenous control, and the wild-type strain grown at 37°C, at the time of sampling or at time zero, as the calibrator sample. Then, the relative gene expression was plotted as the $\log_2$(fold change; FC)= $\log_2[2^{-(\Delta\Delta Ct)}]$.

Statistical analyses

The differences in the excision of ROD21 and the expression of *SEN1970* between the wild-type, mutant and complemented strains at different concentrations of IPTG, were analyzed by two-way ANOVA followed by the Tukey post-hoc test. The correlation between the excision and *SEN1970* expression was assessed by the Pearson's correlation test. Both analyses were carried out in GraphPad Prism v.8.4.3 with α=0.05.

Acknowledgements

We are grateful to Diego Ulloa-Allendes from Pontificia Universidad Católica de Chile for his valuable help during DNA and RNA extraction.

The authors of this article were supported by the Millennium Institute on Immunology and Immunotherapy (P09/016-F) and by Agencia Nacional de Investigación y Desarrollo de Chile, ANID, through Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT grants 1190864 and 1170964) and Programa de Formación de Capital Humano Avanzado, Beca de Doctorado Nacional 21172030.

References

Abbani M. A., Papagiannis C. V., Sam M. D., Cascio D., Johnson R. C., and Clubb R. T. (2007). Structure of the cooperative Xis-DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly. *Proc. Natl. Acad. Sci.*, *104*(7), 2109–2114. https://doi.org/10.1073/pnas.0607820104

Adeolu M., Alnajar S., Naushad S., and Gupta R. S. (2016). Genome-based phylogeny and taxonomy of the "Enterobacteriales": proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morgane. *Int. J. Syst. Evol. Microbiol.*, *66*(12), 5575–5599. https://doi.org/10.1099/ijsem.0.001485

Antonenka U., Nölting C., Heesemann J., and Rakin A. (2006). Independent acquisition of site-specific recombination factors by *asn* tRNA gene-targeting genomic islands. *Int. J. Med. Microbiol.*, *296*(6), 341–352. https://doi.org/10.1016/j.ijmm.2006.02.021

Bowers J. R., Kitchel B., Driebe E. M., MacCannell D. R., Roe C., Lemmer D., de Man T., Rasheed J. K., Engelthaler D. M., Keim P., and Limbago B. M. (2015). Genomic analysis of the emergence and rapid global dissemination of the clonal group 258 *Klebsiella pneumoniae* pandemic. *PLoS One*, *10*(7), e0133727. https://doi.org/10.1371/journal.pone.0133727

Bushman W., Yin S., Thio L. L., and Landy A. (1984). Determinants of directionality in lambda site-specific recombination. *Cell*, *39*(3 Pt 2), 699–706. https://doi.org/10.1016/0092-8674(84)90477-X

Carpenter M. R., Rozovsky S., and Boyd E. F. (2016). Pathogenicity island cross talk mediated by recombination directionality factors facilitates excision from the chromosome. *J. Bacteriol.*, *198*(5), 766–776. https://doi.org/10.1128/JB.00704-15

Carraro N., Libante V., Morel C., Decaris B., Charron-Bourgoin F., Leblond P., and Guédon G. (2011). Differential regulation of two closely related integrative and conjugative elements from *Streptococcus thermophilus*. *BMC Microbiol.*, *11*, 238. https://doi.org/10.1186/1471-2180-11-238

Carraro N., Rivard N., Ceccarelli D., Colwell R. R., and Burrus V. (2016). IncA/C Conjugative Plasmids Mobilize a New Family of Multidrug Resistance Islands in Clinical *Vibrio cholerae* Non-O1/Non-O139 Isolates from Haiti. *MBio*, *7*(4), e00509-16. https://doi.org/10.1128/mBio.00509-16.

Che D., Hasan M. S., and Chen B. (2014). Identifying Pathogenicity Islands in Bacterial Pathogenomics Using Computational Approaches. *Pathogens*, *3*, 36–56. https://doi.org/10.3390/pathogens3010036

Chen L., Mathema B., Pitout J. D. D., DeLeo F. R., and Kreiswirth B. N. (2014). Epidemic *Klebsiella pneumoniae* ST258 Is a Hybrid Strain. *MBio*, *5*(3), e01355-14. https://doi.org/10.1128/mBio.01355-14

Curraize C. de, Siebor E., Varin V., Neuwirth C., and Hall R. M. (2020). Two New SGI1-LK Variants Found in *Proteus mirabilis* and Evolution of the SGI1-HKL group of *Salmonella* genomic islands. *MSphere, 5*, e00875-19.

Daccord A., Mursell M., Poulin-laprade D., and Burrus V. (2012). Dynamics of the SetCD-Regulated Integration and Excision of Genomic Islands Mobilized by Integrating Conjugative Elements of the SXT/R391 Family. *J. Bacteriol., 194*(21), 5794–5802. https://doi.org/10.1128/JB.01093-12

Datsenko K. A., and Wanner B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.*, *97*(12), 6640–6645. https://doi.org/10.1073/pnas.120163297

ElAntak L., Ansaldi M., Guerlesquin F., Méjean V., and Morelli X. (2005). Structural and genetic analyses reveal a key role in prophage excision for the TorI response regulator inhibitor. *J. Biol. Chem.*, *280*(44), 36802–36808. https://doi.org/10.1074/jbc.M507409200

Fillol-Salom A., Bacarizo J., Alqasmi M., Ciges-Tomas J. R., Martínez-Rubio R., Roszak A. W., Cogdell R. J., Chen J., Marina A., and Penadés J. R. (2019). Hijacking the Hijackers: *Escherichia coli Pathogenicity Islands* Redirect Helper Phage Packaging for Their Own Benefit. *Mol. Cell*, *75*(5), 1020–1030. https://doi.org/10.1016/j.molcel.2019.06.017

Godfrey S. A. C., Lovell H. C., Mansfield J. W., Corry D. S., Jackson R. W., and Arnold D. L. (2011). The Stealth Episome: Suppression of Gene Expression on the Excised Genomic Island PPHGI-1 from *Pseudomonas syringae* pv. *phaseolicola*. *PLoS Pathog.*, *7*(3), e1002010. https://doi.org/10.1371/journal.ppat.1002010

Guarneros G., and Echols H. (1970). New mutants of bacteriophage λ with a specific defect in excision from the host chromosome. *J. Mol. Biol.*, *47*(3), 565–574. https://doi.org/10.1016/0022-2836(70)90323-2

Haskett T. L., Terpolilli J. J., Ramachandran V. K., Verdonk C. J., Poole P. S., O'Hara G. W., and Ramsay J. P. (2018). Sequential induction of three recombination directionality factors directs assembly of tripartite integrative and conjugative elements. *PLOS Genet.*, *14*(3), e1007292. https://doi.org/10.1371/journal.pgen.1007292

Kiss J., Papp P. P., Szabó M., Farkas T., Murányi G., Szakállas E., and Olasz F. (2015). The master regulator of IncA/C plasmids is recognized by the *Salmonella* Genomic island SGI1 as a signal for excision and conjugal transfer. *Nucleic Acids Res.*, *43*(18), 8735–8745. https://doi.org/10.1093/nar/gkv758

Kumar S., Stecher G., Li M., Knyaz C., and Tamura K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Landy A. (2015). The λ Integrase Site-specific Recombination Pathway. *Microbiol. Spectr.*, *3*(2), MDNA3-0051–2014.

Lautner M., Schunder E., Herrmann V., and Heuner K. (2013). Regulation, integrase-dependent excision, and horizontal transfer of genomic islands in *Legionella pneumophila*. *J. Bacteriol.*, *195*(7), 1583–1597. https://doi.org/10.1128/JB.01739-12

Laxmikanthan G., Xu C., Brilot A. F., Warren D., Steele L., Seah N., Tong W., Grigorieff N., Landy A., and Van Duyne G. D. (2016). Structure of a holliday junction complex reveals mechanisms governing a highly regulated DNA transaction. *Elife*, *5*, e14313. https://doi.org/10.7554/eLife.14313

Lesic B., Bach S., Ghigo J. M., Dobrindt U., Hacker J., and Carniel E. (2004). Excision of the high-pathogenicity of *Yersinia pseudotuberculosis* requires the combined of actions of its cognate integrase and Hef, a new recombination directionality factor. *Mol. Microbiol.*, *52*(5), 1337–1348. https://doi.org/10.1111/j.1365-2958.2004.04073.x

Lewis J. A., and Hatfull G. F. (2001). Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res.*, *29*(11), 2205–2216.

Marcoleta A. E., Berríos-Pastén C., Nuñez G., Monasterio O., and Lagos R. (2016). *Klebsiella pneumoniae* Asparagine tDNAs Are Integration Hotspots for Different Genomic Islands Encoding Microcin E492 Production Determinants and Other Putative Virulence Factors Present in Hypervirulent Strains. *Front. Microbiol.*, *7*, 849. https://doi.org/10.3389/fmicb.2016.00849

McKitterick A. C., and Seed K. D. (2018). Anti-phage islands force their target phage to directly mediate island excision and spread. *Nat. Commun.*, *9*(1), 2348. https://doi.org/10.1038/s41467-018-04786-5

Moitoso de Vargas L., and Landy A. (1991). A switch in the formation of alternative DNA loops modulates λ site-specific recombination. *Proc. Natl. Acad. Sci. U. S. A.*, *88*, 588–592. https://doi.org/10.1073/pnas.88.2.588

Neale H. C., Jackson R. W., Preston G. M., and Arnold D. L. (2018). Supercoiling of an excised genomic island represses effector gene expression to prevent activation of host resistance. *Mol. Microbiol.*, *110*(3), 444–454. https://doi.org/10.1111/mmi.14111

Newman R. M., Salunkhe P., Godzik A., and Reed J. C. (2006). Identification and Characterization of a Novel Bacterial Virulence Factor That Shares Homology with Mammalian Toll/Interleukin-1 Receptor Family Proteins. *Infect. Immun.*, *74*(1), 594–601. https://doi.org/10.1128/IAI.74.1.594

Nieto P. A., Pardo-Roa C., Salazar-Echegarai F. J., Tobar H. E., Coronado-Arrázola I., Riedel C. A., Kalergis A. M., and Bueno S. M. (2016). New insights about excisable pathogenicity islands in *Salmonella* and their contribution to virulence. *Microbes Infect.*, *18*, 302–309. https://doi.org/10.1016/j.micinf.2016.02.001

Panis G., Duverger Y., Courvoisier-Dezord E., Champ S., Talla E., and Ansaldi M. (2010). Tight regulation of the *intS* gene of the KplE1 prophage: A new paradigm for integrase gene regulation. *PLoS Genet.*, *6*(10), e1001149. https://doi.org/10.1371/journal.pgen.1001149

Panis G., Franche N., Méjean V., and Ansaldi M. (2012). Insights into the functions of a prophage recombination directionality factor. *Viruses*, *4*(11), 2417–2431. https://doi.org/10.3390/v4112417

Panis G., Méjean V., and Ansaldi M. (2007). Control and regulation of KplE1 prophage site-specific recombination: A new recombination module analyzed. *J. Biol. Chem.*, *282*(30), 21798–21809. https://doi.org/10.1074/jbc.M701827200

Pardo-Roa C., Salazar G. A., Noguera L. P., Salazar-Echegarai F. J., Vallejos O. P., Suazo I. D., Schultz B. M., Coronado-Arrazola I., Kalergis A. M., and Bueno S. M. (2019). Pathogenicity island excision during an infection by Salmonella enterica serovar Enteritidis is required for crossing the intestinal epithelial barrier in mice to cause systemic infection. *PLOS Pathog.*, *15*(12), e1008152. https://doi.org/10.1371/journal.ppat.1008152

Penadés J. R., and Christie G. E. (2015). The Phage-Inducible Chromosomal Islands: A Family of Highly Evolved Molecular Parasites. *Annu. Rev. Virol.*, *2*, 181–201. https://doi.org/10.1146/annurev-virology-031413-085446

Piña-Iturbe A., Suazo I. D., Hoppe-Elsholz G., Ulloa-Allendes D., González P. A., Kalergis A. M., and Bueno S. M. (2020). Horizontally Acquired Homologs of Xenogeneic Silencers: Modulators of Gene Expression Encoded by Plasmids, Phages and Genomic Islands. *Genes (Basel).*, *11*(2), 142. https://doi.org/10.3390/genes11020142

Piña-Iturbe A., Ulloa-Allendes D., Pardo-Roa C., Coronado-Arrázola I., Salazar-Echegarai F. J., Sclavi B., González P. A., and Bueno S. M. (2018). Comparative and phylogenetic analysis of a novel family of *Enterobacteriaceae*-associated genomic islands that share a conserved excision/integration module. *Sci. Rep.*, *8*(1), 10292. https://doi.org/10.1038/s41598-018-28537-0

Quiroz T. S., Nieto P. A., Tobar H. E., Salazar-Echegarai F. J., Lizana R. J., Quezada C. P., Santiviago C. A., Araya D. V, Riedel C. A., Kalergis A. M., and Bueno S. M. (2011). Excision of an Unstable Pathogenicity Island in *Salmonella enterica* Serovar Enteritidis Is Induced during Infection of Phagocytic Cells. *PLoS One*, *6*(10), e26031. https://doi.org/10.1371/journal.pone.0026031

Ramsay J. P., Sullivan J. T., Stuart G. S., Lamont I. L., and Ronson C. W. (2006). Excision and transfer of the *Mesorhizobium loti* R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. *Mol. Microbiol.*, *62*(3), 723–734. https://doi.org/10.1111/j.1365-2958.2006.05396.x

Regmi A., and Boyd E. F. (2019). Carbohydrate metabolic systems present on genomic islands are lost and gained in *Vibrio parahaemolyticus*. *BMC Microbiol.*, *19*(1), 112. https://doi.org/10.1186/s12866-019-1487-6

Roy A., Kucukural A., and Zhang Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.*, *5*(4), 725–738. https://doi.org/10.1038/nprot.2010.5

Salazar-Echegarai F. J., Tobar H. E., Nieto P. A., Riedel C. A., and Bueno S. M. (2014). Conjugal Transfer of the Pathogenicity Island ROD21 in *Salmonella enterica* serovar Enteritidis Depends on Environmental Conditions. *PLoS One*, *9*(4), e90626. https://doi.org/10.1371/journal.pone.0090626

Sam M. D., Cascio D., Johnson R. C., and Clubb R. T. (2004). Crystal structure of the excisionase-DNA complex from bacteriophage Lambda. *J. Mol. Biol.*, *338*(2), 229–240. https://doi.org/10.1016/j.jmb.2004.02.053

Schneider C. A., Rasband W. S., and Eliceiri K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Seah N. E., Warren D., Tong W., Laxmikanthan G., Van Duyne G. D., and Landy A. (2014). Nucleoprotein architectures regulating the directionality of viral integration and excision. *Proc. Natl. Acad. Sci. U. S. A.*, *111*(34), 12372–12377. https://doi.org/10.1073/pnas.1413019111

Thomson N. R., Clayton D. J., Windhorst D., Vernikos G., Davidson S., Churcher C., Quail M. A., Stevens M., Jones M. A., Watson M., Barron A., Layton A., Pickard D., Kingsley R. A., Bignell A., Clark L., Harris B., Ormond D., Abdellah Z., Brooks K., Cherevach I., Chillingworth T., Woodward J., Norberczak H., Lord A., Arrowsmith C., Jagels K., Moule S., Mungall K., Sanders M., Whitehead S., Chabalgoity J. A., Maskell D., Humphrey T., Roberts M., Barrow P. A., Dougan G., and Parkhill J. (2008). 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. https://doi.org/10.1101/gr.077404.108.1

Tobar H. E., Salazar-Echegarai F. J., Nieto P. A., Palavecino C. E., Sebastian V. P., Riedel C. A., Kalergis A. M., and Bueno S. M. (2013). Chromosomal Excision of a New Pathogenicity Island Modulates *Salmonella* Virulence *In Vivo*. *Curr. Gene Ther.*, *13*(4), 240–249.

Waterhouse A. M., Procter J. B., Martin D. M. A., Clamp M., and Barton G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Williams K. P. (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res.*, *30*(4), 866–875. https://doi.org/10.1093/nar/30.4.866

Xiong D., Song L., Geng S., Jiao Y., Zhou X., Song H., Kang X., Zhou Y., Xu X., Sun J., Pan Z., and Jiao X. (2019). *Salmonella* Coiled-Coil- and TIR-Containing TcpS Evades the Innate Immune System and Subdues Inflammation. *Cell Rep.*, *28*(3), 804–818. https://doi.org/10.1016/j.celrep.2019.06.048

Zhang X., Wan W., Yu H., Wang M., Zhang H., Lv J., Tang Y.-W., Kreiswirth B. N., Du H., and Chen L. (2019). New Delhi Metallo-β-Lactamase 5–Producing *Klebsiella pneumoniae* Sequence Type 258, Southwest China, 2017. *Emerg. Infect. Dis.*, *25*(6), 1209–1213.

Supplemental material

Excision of the pathogenicity island ROD21 is modulated by growth phase and negatively correlates with the expression of its integrase and recombination directionality factor

Alejandro Piña-Iturbe[a], Guillermo Hoppe-Elsholz[a], Paulina A. Fernández[b], Carlos A. Santiviago[b], Pablo A. González[a] and Susan M. Bueno[a]#

[a] Millennium Institute on Immunology and Immunotherapy, Facultad de Ciencias Biológicas, Departamento de Genética Molecular y Microbiología, Pontificia Universidad Católica de Chile, Santiago, Chile

[b] Laboratorio de Microbiología, Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile.

*Address correspondence to Susan M. Bueno, sbueno@bio.puc.cl

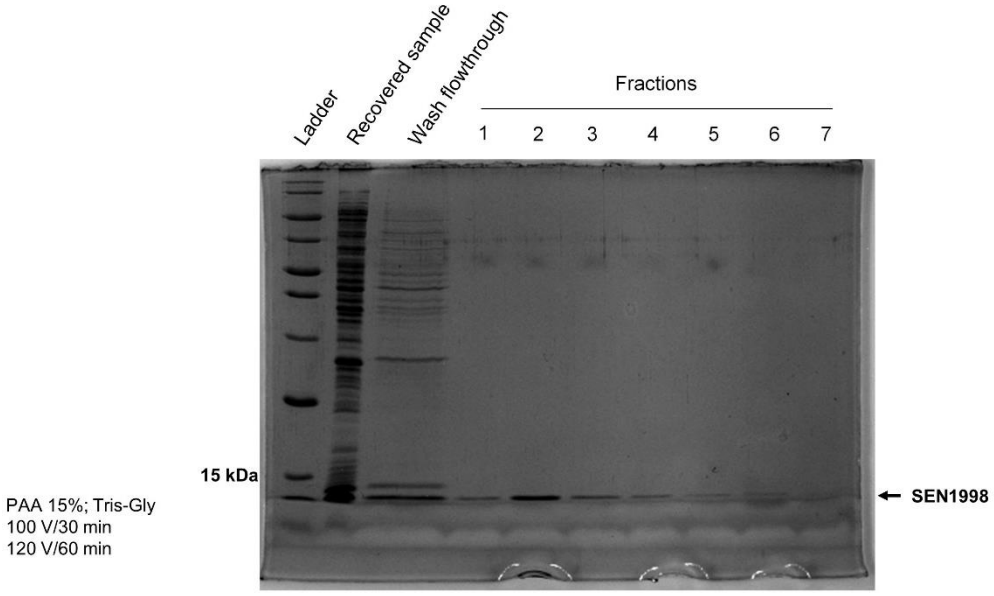
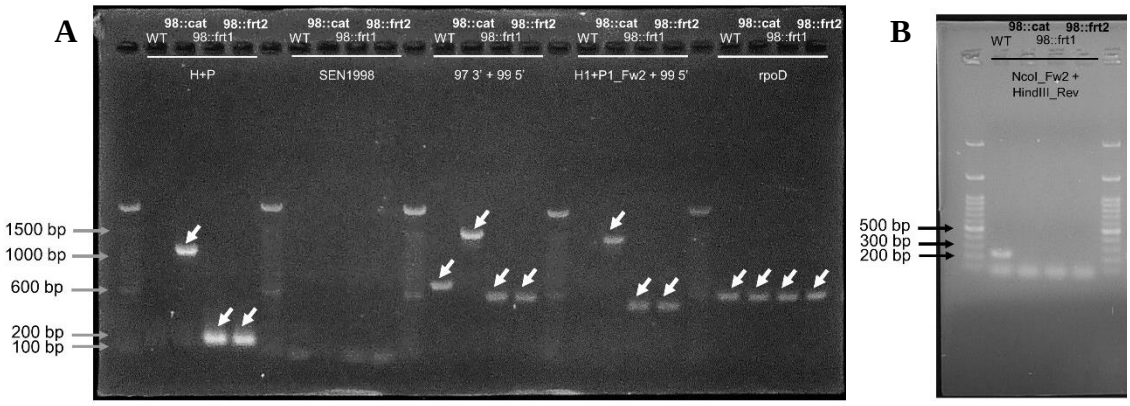


Figure S1. IMAC purification of the SEN1998 protein. Coomassie-stained polyacrylamide gel loaded with samples and fractions recovered after IMAC. Lane 1: molecular-weight marker; lane 2: *Escherichia coli* BL21(DE3) pET15b-*SEN1998* lysate after IMAC; lane 3: recovered wash flowthrough; lane 4-10: fractions eluted from the column.



Primers	WT	*Δ98:cat*	*Δ98:FRT*
H+P	--	1094 bp	198 bp
SEN1998	235 bp	--	--
97 3'+ 99 5'	702 bp	1497 bp	601 bp
H1P1+99 5'	--	1406 bp	510 bp
rpoD	577 bp	577 bp	577 bp

Figure S2. PCR confirmation of the Δ*SEN1998*::*frt* mutation. (**A**) Deletion of the *SEN1998* coding sequence was assessed by PCR using specific and external primers (see also Table S1). (**B**) The PCR using primers specific for *SEN1998* was repeated with a new set of primers.

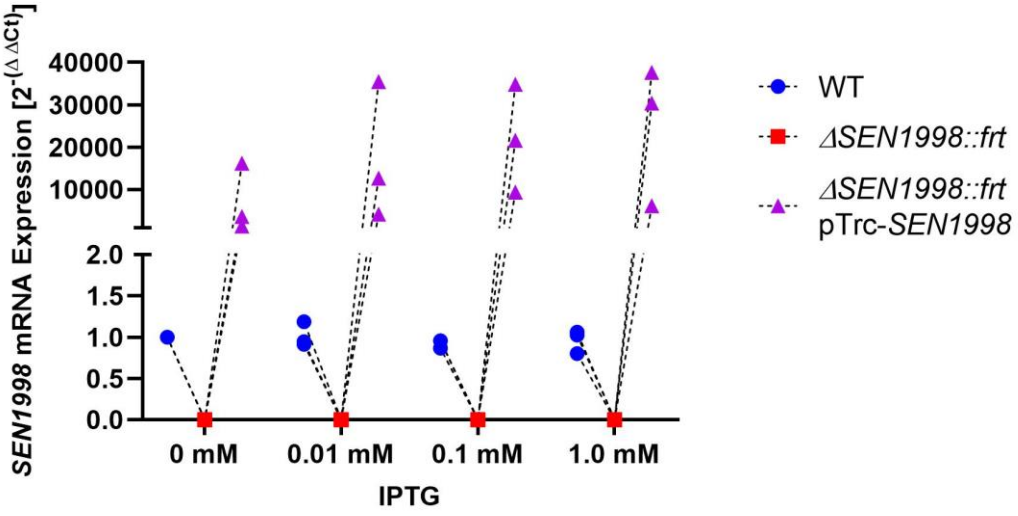


Figure S3. Expression of *SEN1998* in the wildtype, Δ*SEN1998*::*frt,* and Δ*SEN1998*::*frt* pTrc-*SEN1998* strains. The expression of *SEN1998* was measured by RT-qPCR.

Table S1. Sequences of the primers and hydrolysis probes used in this study.

Primers	**Sequence (5'->3')**	**Use**
SEN1998 _H1+P1_Fw2	CGA ACA GAT ATC TGT TCG ACA CGA GAT CCT ATT TCT GTT Tgt gta ggc tgg agc tgc ttc	Construction of mutant strain Δ*SEN1998*::*frt*
SEN1998_H2+P2_Rev2	TCC ATA CAA TTA TCT CCA TCG AAT AAC ACA GGA GAT AAA Tca tat gaa tat cct cct tag	Construction of mutant strain Δ*SEN1998*::*frt*
SEN1998_NcoI_Fw2	GCA GCC ATG GTT AAA AGA GAA ATA AAA GC	Amplification of *SEN1998*
SEN1998_HindIII_Rev	GTT TAA GCT TTT ATG CTG CGT TCT GGC	Amplification of *SEN1999*
3' 1997 Fw	CAT TCG GCT TGT TCA ATA AA	Amplification of the region containing *SEN1998*
5' 1999 Rev	TAT GCG TTA TGC CAG AGG TA	Amplification of the region containing *SEN1998*
attR_Rev	TTA AAC GTC TCC GGA CTC GCC	Amplification of target DNA for EMSA
attR-A_Rev	ACA GAT TAC CGC TAT TGC CCG	Amplification of target DNA for EMSA
attR-B_Fw	GAA ACA TAT CGG GCA ATA GCG G	Amplification of target DNA for EMSA
attR_Fw	CGC GAC AGG CAA TCT TTT TGT C	Amplification of target DNA for EMSA
attL_Fw	TAC TAT GCG CCC CGT TCA CAC	Amplification of target DNA for EMSA
attL_Rev	AT AGA GGC CAA ACC CGT CAG TG	Amplification of target DNA for EMSA
attL-B_Fw	GA AGC CAT ACC CTT TTG CTG G	Amplification of target DNA for EMSA
attL-A_Rev	GA GTG GTT TTT AGT GAA CGA GAG TG	Amplification of target DNA for EMSA
SEN1998-RT_Fw	GAA AAC GCC CGG CCT TAA	Quantification of the *SEN1998* mRNA expression
SEN1998-RT_Rev	CGC ACC GGA TTT GGT AAA A	Quantification of the *SEN1998* mRNA expression
SEN1970-RT_Fw	CGA TAC TGT CTG GAA GCG CCT	Quantification of the *SEN1970* mRNA expression
SEN1970-RT_Rev	TTT TGC TGG ACG GCA TGA C	Quantification of the *SEN1970* mRNA expression
rpoD-RT_Fw	GTT GAC CCG GGA AGG CGA AA	Quantification of the ROD21 excision
rpoD-RT_Rev	CAG AAC CGA CGT GAG TTG CG	Quantification of the ROD21 excision
attB1-RT_Fw	GTT ACT ATG CGC CCC GTT CAC AC	Quantification of the ROD21 excision
attB1-RT_Rev	CCG ATT AAG CCC CAA AAA CTA TG	Quantification of the ROD21 excision
Probes	**Sequence (5'->3')**	**Use**
attB1-RT	TTC GAG TCC AGT CAG AGG A	Quantification of the ROD21 excision
rpoD-RT	CGA CAT CGC TAA ACG	Quantification of the ROD21 excision
SEN1970-RT	TCA CCG CGA TCC TA	Quantification of the *SEN1970* mRNA expression
SEN1998-RT	CAG TTC GTA AAT CCA C	Quantification of the *SEN1998* mRNA expression

5. DISCUSSION

Horizontal gene transfer plays a major role in bacterial evolution and adaptation (Jeong et al., 2016). As horizontally transferred elements, genomic islands (GIs) contribute to acquisition of the genes involved in the metabolism new substrates, the establishment of symbiotic relationships, antibiotic/antimicrobial resistance or virulence factors, among others (Doublet et al., 2005; Montero et al., 2017; Regmi and Boyd, 2019; Sullivan and Ronson, 1998). Despite the relatively difficult identification of GIs in the unceasingly growing 'mountain' of bacterial genomes (Che et al., 2014; Lu and Leong, 2016), different families of GIs have been described and studied since their discovery 26 years ago, especially those involved in antibiotic resistance and virulence, such as the SXT/R391 family of integrative and conjugative elements, the *Salmonella* Genomic Island-1 and related GIs, the Mobilizable Genomic Islands of *Vibrio* spp. or the Phage-Inducible Chromosomal Islands (Bioteau et al., 2018; Blum et al., 1994; Cummins et al., 2020; Daccord et al., 2013; Fillol-Salom et al., 2018). However, there is a great amount of "solitary" GIs that have not been classified in any known family, which are playing important roles for their bacterial hosts, such as the Locus of Enterocyte Effacement of pathogenic *E. coli*, the High-Pathogenicity Island of different Enterobacterales, or the PPHGI-1 of *Pseudomonas syringae* pv. *phaseolicola*, which are some examples (Bach et al., 2000; Elliott et al., 1998; Jackson et al., 2000). Since its discovery in 2008, the genomic island ROD21 has been recognized as a player in the pathogenesis of the foodborne pathogen *Salmonella* serovar

Enteritidis, and different studies have provided evidence that ROD21 encodes genes important for evasion of the immune response and invasion of deep organs (Pardo-Roa et al., 2019; Quiroz et al., 2011; Silva et al., 2012; Thomson et al., 2008). Nevertheless, although the existence of "ROD21-related islands" was reported decades ago (Thomson et al., 2008), the phylogenetic relationship of ROD21 with other GIs remained obscure.

Harnessing the great amount of available sequenced bacterial genomes and current bioinformatic tools, it was possible to search and delimit a set of 54 different GIs related to ROD21 using BLAST (Section 2, Table 1). This was possible since the islands belonging to the same family share the insertion site (determined by the integrase protein encode by them) and the repeated sequences located at both ends of the islands. Therefore, a BLAST search using a query sequence spanning the attachment region of ROD21 and the first nucleotides of the *SEN1970* integrase-coding gene will result in related sequences, corresponding to one attachment region of the GI found in other bacterial genomes, and a second BLAST hit of shorter length corresponding to the repeated sequence located at the other end of the island. While, at the time of publication, it was thought that all the identified GIs were only found in strains of the family *Enterobacteriaceae*, the current taxonomy of these bacterial species show that the *Enterobacteriaceae*-associated ROD21-like (EARL) family of GIs are spread among different bacterial families in the order Enterobacterales, namely *Enterobacteriaceae*, *Yersiniaceae* and *Pectobacteriaceae* (Adeolu et al., 2016). An interesting finding is the likely origin of the EARL GIs in plant-pathogenic bacteria of the genus *Pectobacterium* as suggested by the integrase-based phylogeny, where the islands harbored by different species of *Pectobacterium* clusters at the most basal nodes of the tree (Section 2, Fig. 2A). Another finding supporting the origin of these islands in plant pathogens is that two *Pectobacterium* species (*P. parmentieri* WPP163

and *P. parmentieri* RNS08.42.1A) harbors two different EARL GIs, one of which is an ancestral island that harbors a relaxase-encoding gene that is different from the predominant MobA/MobL relaxase gene of the EARL family (Section 2, Fig. 2B). As relaxases are involved in conjugal transfer of GIs (Vanga et al., 2015; Waldor, 2010), is tempting to speculate that acquisition of the MobA/MobL family relaxase by the ancestral EARL GI allowed its dissemination from plant-pathogenic to animal-pathogenic bacteria. Is possible that acquisition of the genes encoding type-IV-pili-related proteins had also played a role in dissemination of these islands, since some type-IV pili have been associated to the conjugal transfer of the R64 plasmid and the PAPI-1 genomic island (Filloux, 2010). Nevertheless, the analysis of the gene content of the EARL GIs indicate that these islands only encode part of the transfer machinery, consequently relying in a host-encoded transfer system for their dissemination. Therefore, ROD21 and the EARL GIs are mobilizable genomic islands.

The sequence and phylogenetic analysis of the EARL GIs also evidenced that, together with the acquisition of the genes encoding the MobA/MobL relaxase and type-IV-pili-related proteins, the acquisition of genes encoding N-terminal (short) and full-length homologs of the global transcriptional regulator and xenogeneic silencer H-NS was also an important step in the evolution of these GIs (Section 2, Fig. 2) from the ancestral EARL island. The highly AT-rich nature of the EARL GIs (Section 3, Fig. 2) makes these islands a target for gene repression by the XSs encoded by the core genome, i.e. H-NS in Enterobacterales (Navarre, 2016). The short H-NS homologs encoded by some EARL GIs are homologous to the H-NST protein encoded by the *serU* island of UPEC (Section 3, Fig.1), which acts as an H-NS antagonist, when overexpressed, alleviating the silencing imposed by H-NS (Williamson and Free, 2005). Therefore, the presence of such H-NS homologs in the EARL islands may be a mechanism to

ensure expression of the island genes after horizontal transfer (Dorman, 2014; Stoebel et al., 2008). Interestingly, the H-NST protein also participates in derepression of the genes encoded in the Locus of Enterocyte Effacement pathogenicity island (Levine et al., 2014), suggesting that the short H-NS homologs encoded by the EARL GIs may also play virulence regulatory roles outside their own island. In addition to the short H-NS homologs encoded by 20 of the 54 identified EARL islands, 33 GIs encoded full-length H-NS homologs which correspond to the Hfp protein also encoded by the *serU* island (Müller et al., 2010). These proteins were recently recognized as a separate family of homologs, but not paralogs, of H-NS according to their phylogeny, meaning that Hfp and the related proteins encoded by the EARL GIs are not descendants of H-NS but may share a common protein ancestor instead (Fitzgerald et al., 2020). The role played by these full-length H-NS homologs is still not clear. Hfp was found to participate in the expression of different virulence traits in UPEC 536, but the observed phenotypes were only evident in a Δ*hns* background, which is in part explained by the repression of *hfp* by H-NS (Müller et al., 2010). As *hfp* shows an increased expression at temperatures lower than 37°C, and in the exponential growth phase in UPEC 536, the regulatory roles of this protein may be relevant under certain environmental conditions found during the bacterial life/infective cycle (Dorman, 2010; Müller et al., 2010). More research is needed to understand the contribution of the Hfp homolog encoded by the EARLs in the virulence of their hosts and the biology of the islands that encode them.

A feature conserved among all the EARL islands, is their excision/integration module and, therefore, they are excisable as demonstrated for ROD21 from *Salmonella* ser. Enteritidis, ICEKp258.2 from *Klebsiella pneumoniae* ST258 KP35 and IE3 from EPEC E2348/69 (Section 2, Fig.2, 5, 6). This module encompasses the regulatory attachment regions and the genes

encoding an integrase from the tyrosine recombinase family and a small protein predicted to be a DNA-binding AlpA family transcriptional regulator [Section 2, Fig.1; (Quiroz et al., 2011)]. In ROD21, these genes have the locus tag *SEN1970* and *SEN1998*, respectively. Although previous studies have proposed that *SEN1998* encodes a recombination directionality factor (RDF) (Pardo-Roa et al., 2019; Piña-Iturbe et al., 2018; Salazar-Echegarai et al., 2014; Tobar et al., 2013), no analysis of its sequence or its biochemical properties were available. Here, *SEN1998* was characterized as a small basic protein that shares the main characteristics of RDFs and its DNA-binding capacity to the attachment regions of ROD21 was demonstrated (Section 4). Nevertheless, we observed that SEN1998 is not essential for the excision of ROD21, nor acts as a transcriptional regulator of *SEN1970* as suggested by the overlapping of the *SEN1970* promoter with the left attachment region bound by SEN1998 (Section 4, Fig. 3). While the function of RDFs as regulators of their cognate integrase is an accessory role, the slight changes in the excision of ROD21 in the Δ*SEN1998*::*frt* strain resulted unusual when compared to what is currently know about the excision of GIs, which significantly reduce their excision levels in the absence of their RDF, similar to the mechanism observed for Lambda and other bacteriophages (Daccord et al., 2012; Guarneros and Echols, 1970; Haskett et al., 2018; Panis et al., 2012). Also unexpected was the finding of a negative correlation between the mRNA levels of the integrase- and RDF- encoding genes with the excision of ROD21 (Section 4, Fig. 4). Although the mRNA levels may not reflect the protein levels, is unknown whether the transcriptional repression of the integrase and the RDF plays any role in the regulation of excision, as the time frame in which this occurs suggests, i.e. immediately before the excision begins to increase. Further experiments are needed to systematically dissect the excision regulation in ROD21 and the EARL GIs. Assessing the possible interaction between SEN1998

and the integrase SEN1970 and whether it affects DNA binding and excision *in vitro* would provide evidence about intasome assembly and regulation. Additionally, the changes in the excision level along the growth curve provides an opportunity to test whether SEN1998 is required for the excision to increase.

In summary, ROD21 belongs to the the *Enterobacteriaceae*-associated ROD21-like family of genomic islands, a newly identified group of excisable and transferable genetic elements present in plant and animal pathogens of the order Enterobacterales, which carry virulence factors and regulatory proteins that may play specific roles in modulation of virulence traits under specific environmental conditions. The study of this family of GIs may contribute to better understand current and emerging pathogens such as *Salmonella* serovars, pathogenic *E. coli* and carbapenem-resistant *K. pneumoniae* ST258, among others.

6. CONCLUSIONS

Based on the results and data obtained in this Thesis, the following conclusions can be drawn:

From Section 2: ROD21 belongs to a family of excisable genomic islands found in *Enterobacteriaceae*

- The pathogenicity island ROD21 is a member of a family of genomic islands spread among several bacterial families within the order Enterobacterales: the *Enterobacteriaceae*-associated ROD21-like (EARL) genomic islands.
- The EARL GIs are present in clinically relevant bacteria including foodborne, invasive or multidrug-resistant pathogens.
- The EARL GIs have probably originated within plant-pathogenic bacteria, as supported by the integrase-based phylogeny, which places the islands harbored by the *Pectobacterium* species as branching from the basal nodes; and by the finding of *Pectobacterium* species harboring two members of the EARL family that differs in their transfer-related genes.
- The pattern of gene acquisition unveiled by the integrase-based phylogeny suggests a role of the genes encoding the type-IV-related proteins and the conjugation-related proteins in the transfer of the islands from plant-pathogenic bacteria to animal-pathogenic and animal-associated bacteria, and also a role for the genes encoding the type-III restriction-modification system and the SpnT homolog in the broadening of the islands host range.

- The EARL GIs share a conserved excision/integration module encompassing the attachment regions, the integrase-encoding gene, and the RDF-encoding gene, which allows their excision from the chromosome and accounts for the dissemination of this family among different bacterial families. The genes included within this module correspond to *SEN1970* and *SEN1998* in ROD21.

From Section 3: The EARL genomic islands are highly AT-rich islands that carry homologs of the xenogeneic silencer H-NS

- Most EARL GIs, including ROD21, encode a short and/or full-length homolog of the xenogeneic silencer H-NS which shares the key domains and amino acid residues involved in the DNA-binding and oligomerization processes.
- Most GIs harbored by bacteria from the family *Enterobacteriaceae* are significantly more AT-rich than their host genome.
- The EARL GIs have a significantly higher AT-content when compared with the islands harbored by *Enterobacteriaceae,* a feature that may be related to the presence of the H-NS homologs since H-NS has affinity for AT-rich regions. This suggests a role of the H-NS homologs in the biology of the EARL GIs.

From Section 4: The excision of ROD21 and the expression of its recombination directionality factor are negatively correlated, and are modulated by the growth phase

- The gene *SEN1998* of ROD21, highly conserved among the EARL islands, encodes a small basic protein with predicted secondary and tertiary structures that match those of known recombination directionality factors.

- SEN1998 is a DNA-binding protein that binds the two attachment regions of ROD21 with different affinities, forming several complexes which suggest more than one binding site.
- SEN1998 does not modulate the expression of *SEN1970* since the deletion and complementation of *SEN1998* produced only slight differences in the expression of *SEN1970* that were not statistically significant.
- The excision of ROD21 is modulated by the growth phase of *Salmonella* ser. Enteritidis and is negatively correlated to the expression of the integrase-encoding gene *SEN1970* and the RDF-encoding gene *SEN1998*.

7. APPENDIX

7.1. Participation in scientific meetings

- **Piña-Iturbe A**, Hoppe-Elsholz G, González PA, Bueno SM. Characterization of a potential recombination directionality factor encoded in the pathogenicity island ROD21 of *Salmonella enterica* serovar Enteritidis. Póster presentado en: Somich 2019. XLI Reunión Anual de la Sociedad de Microbiología de Chile; 5 al 8 de Noviembre de 2019; Puerto Varas, Chile.

7.2. Participation in scientific papers

- **Piña-Iturbe A**, Ulloa-Allendes D, Pardo-Roa C, Coronado-Arrázola I, SalazarEchegarai FJ, Sclavi B, González PA and Bueno SM. Comparative and phylogenetic analysis of a novel family of *Enterobacteriaceae*-associated genomic islands that share a conserved excision/integration module. Sci Rep. 2018;8(1):10292.
- **Piña-Iturbe A**, Suazo ID, Hoppe-Elsholz G, Ulloa-Allendes D, González PA, Kalergis AM, et al. Horizontally Acquired Homologs of Xenogeneic Silencers: Modulators of Gene Expression Encoded by Plasmids, Phages and Genomic Islands. Genes (Basel). 2020;11(2):142.

- Peñaloza HF, Ahn D, Schultz B, **Piña-Iturbe A**, González LA and Bueno SM. L-arginine enhances intracellular killing of carbapenem-resistant *Klebsiella pneumoniae* ST258 by murine neutrophils. Front Cell Infect Microbiol. 2020;10: 571771.
- **Piña-Iturbe A**, Hoppe-Elsholz G, Fernández PA, Santiviago CA, González PA and Bueno SM. Excision of the pathogenicity island ROD21 is modulated by growth phase and negatively correlates with the expression of its integrase and recombination directionality factor. Manuscript in preparation.

8. REFERENCES

Abbani M. A., Papagiannis C. V., Sam M. D., Cascio D., Johnson R. C., and Clubb R. T. (2007). Structure of the cooperative Xis-DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly. *Proc. Natl. Acad. Sci.*, *104*(7), 2109–2114. https://doi.org/10.1073/pnas.0607820104

Adeolu M., Alnajar S., Naushad S., and Gupta R. S. (2016). Genome-based phylogeny and taxonomy of the "*Enterobacteriales*": proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae, Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.*, *66*(12), 5575–5599. https://doi.org/10.1099/ijsem.0.001485

Almagro-Moreno S., Napolitano M. G., and Boyd E. F. (2010). Excision dynamics of *Vibrio* pathogenicity island-2 from *Vibrio cholerae*: role of a recombination directionality factor VefA. *BMC Microbiol.*, *10*, 306.

Bach S., de Almeida A., and Carniel E. (2000). The *Yersinia* high-pathogenicity island is present in different members of the family *Enterobacteriaceae*. *FEMS Microbiol. Lett.*, *183*(2), 289–294. https://doi.org/10.1016/S0378-1097(00)00005-7

Bellanger X., Payot S., Leblond-Bourget N., and Guédon G. (2014). Conjugative and mobilizable genomic islands in bacteria: Evolution and diversity. *FEMS Microbiol. Rev.*, *38*(4), 720–760. https://doi.org/10.1111/1574-6976.12058

Bioteau A., Durand R., and Burrus V. (2018). Redefinition and Unification of the SXT/R391 Family of Integrative and Conjugative Elements. *Appl. Environ. Microbiol.*, *84*(13), e00485-18.

Blum G., Ott M., Lischewski A., Ritter A., Imrich H., Tschape H., and Hacker J. (1994). Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect. Immun.*, *62*(2), 606–614.

Cabezón E., Ripoll-Rozada J., Peña A., de la Cruz F., and Arechaga I. (2015). Towards an integrated model of bacterial conjugation. *FEMS Microbiol. Rev.*, *39*(1), 81–95. https://doi.org/10.1111/1574-6976.12085

Carpenter M. R., Rozovsky S., and Boyd E. F. (2016). Pathogenicity island cross talk mediated by recombination directionality factors facilitates excision from the chromosome. *J. Bacteriol.*, *198*(5), 766–776. https://doi.org/10.1128/JB.00704-15

Carraro N., Rivard N., Ceccarelli D., Colwell R. R., and Burrus V. (2016). IncA/C Conjugative Plasmids Mobilize a New Family of Multidrug Resistance Islands in Clinical Vibrio cholerae Non-O1/Non-O139 Isolates from Haiti. *MBio*, *7*(4), e00509-16. https://doi.org/10.1128/mBio.00509-16.

Che D., Hasan M. S., and Chen B. (2014). Identifying Pathogenicity Islands in Bacterial Pathogenomics Using Computational Approaches. *Pathogens*, *3*(1), 36–56. https://doi.org/10.3390/pathogens3010036

Chiang Y. N., Penadés J. R., and Chen J. (2019). Genetic transduction by phages and chromosomal islands: The new and noncanonical. *PLoS Pathog.*, *15*(8), 1–7. https://doi.org/10.1371/journal.ppat.1007878

Coluzzi C., Guédon G., Devignes M. D., Ambroset C., Loux V., Lacroix T., Payot S., and Leblond-Bourget N. (2017). 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. https://doi.org/10.3389/fmicb.2017.00443

Cummins M. L., Hamidian M., and Djordjevic S. P. (2020). *Salmonella* genomic island 1 is broadly disseminated within gammaproteobacteriaceae. *Microorganisms*, *8*(2), 161. https://doi.org/10.3390/microorganisms8020161

Curraize C. de, Siebor E., Varin V., Neuwirth C., and Hall R. M. (2020). Two New SGI1-LK Variants Found in Proteus mirabilis and Evolution of the SGI1-HKL group of Salmonella genomic islands. *MSphere*, *5*(2), e00875-19.

Daccord A., Ceccarelli D., Rodrigue S., and Burrus V. (2013). Comparative Analysis of Mobilizable Genomic Islands. *J. Bacteriol.*, *195*(3), 606–614. https://doi.org/10.1128/JB.01985-12

Daccord A., Mursell M., Poulin-laprade D., and Burrus V. (2012). Dynamics of the SetCD-Regulated Integration and Excision of Genomic Islands Mobilized by Integrating Conjugative Elements of the SXT/R391 Family. *J. Bacteriol.*, *194*(21), 5794–5802. https://doi.org/10.1128/JB.01093-12

Domingues S., and Nielsen K. M. (2017). Membrane vesicles and horizontal gene transfer in prokaryotes. *Curr. Opin. Microbiol.*, *38*, 16–21. https://doi.org/10.1016/j.mib.2017.03.012

Dorman C. J. (2010). Horizontally acquired homologues of the nucleoid-associated protein H-NS: Implications for gene regulation. *Mol. Microbiol.*, *75*(2), 264–267. https://doi.org/10.1111/j.1365-2958.2009.06996.x

Dorman C. J. (2014). H-NS-like nucleoid-associated proteins, mobile genetic elements and horizontal gene transfer in bacteria. *Plasmid*, *75*, 1–11. https://doi.org/10.1016/j.plasmid.2014.06.004

Doublet B., Boyd D., Mulvey M. R., and Cloeckaert A. (2005). The *Salmonella* genomic island 1 is an integrative mobilizable element. *Mol. Microbiol.*, *55*(6), 1911–1924. https://doi.org/10.1111/j.1365-2958.2005.04520.x

ECDC, and EFSA. (2020). Multi-country outbreak of *Salmonella* Enteritidis infections linked to eggs, third update - 6 February 2020. https://doi.org/10.2903/sp.efsa.2020.en-1799

Elliott S. J., Wainwright L. A., McDaniel T. K., Jarvis K. G., Deng Y. K., Lai L. C., McNamara B. P., Donnenberg M. S., and Kaper J. B. (1998). The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Mol. Microbiol.*, *28*(1), 1–4. https://doi.org/10.1046/j.1365-2958.1998.00783.x

Fernández-Alarcón C., Singer R. S., and Johnson T. J. (2011). Comparative genomics of multidrug resistance-encoding IncA/C plasmids from commensal and pathogenic *Escherichia coli* from multiple animal sources. *PLoS One*, *6*(8), e23415. https://doi.org/10.1371/journal.pone.0023415

Fillol-Salom A., Martínez-Rubio R., Abdulrahman R. F., Chen J., Davies R., and Penadés J. R. (2018). Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. *ISME J.*, *12*(9), 2114–2128. https://doi.org/10.1038/s41396-018-0156-3

Filloux A. (2010). A variety of bacterial pili involved in horizontal gene transfer. *J. Bacteriol.*, *192*(13), 3243–3245. https://doi.org/10.1128/JB.00424-10

Fitzgerald S., Kary S. C., Alshabib E. Y., Mackenzie K. D., Stoebel D. M., Chao T.-C., and Cameron A. D. S. (2020). Redefining the H-NS protein family: a diversity of specialized core and accessory forms exhibit hierarchical transcriptional network integration. *Nucleic Acids Res.*, *48*(18), 10184–10198. https://doi.org/10.1093/nar/gkaa709

Frost L. S., Leplae R., Summers A. O., and Toussaint A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.*, *3*(9), 722–732. https://doi.org/10.1038/nrmicro1235

Guarneros G., and Echols H. (1970). New mutants of bacteriophage λ with a specific defect in excision from the host chromosome. *J. Mol. Biol.*, *47*(3), 565–574. https://doi.org/10.1016/0022-2836(70)90323-2

Hacker J., and Kaper J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.*, *54*, 641–679.

Haskett T. L., Terpolilli J. J., Bekuma A., O'Hara G. W., Sullivan J. T., Wang P., Ronson C. W., and Ramsay J. P. (2016). Assembly and transfer of tripartite integrative and conjugative genetic elements. *Proc. Natl. Acad. Sci. U. S. A.*, *113*(43), 12268–12273. https://doi.org/10.1073/pnas.1613358113

Haskett T. L., Terpolilli J. J., Ramachandran V. K., Verdonk C. J., Poole P. S., O'Hara G. W., and Ramsay J. P. (2018). Sequential induction of three recombination directionality factors directs assembly of tripartite integrative and conjugative elements. *PLOS Genet.*, *14*(3), e1007292. https://doi.org/10.1371/journal.pgen.1007292

Jackson R. W., Mansfield J. W., Arnold D. L., Sesma A., Paynter C. D., Murillo J., Taylor J. D., and Vivian A. (2000). Excision from tRNA genes of a large chromosomal region , carrying *avrPphB*, associated with race change in the bean pathogen, *Pseudomonas syringae* pv. *phaseolicola*. *Mol. Microbiol.*, *38*(2), 186–197.

Jeong H., Sung S., Kwon T., Seo M., Caetano-Anollés K., Choi S. H., Cho S., Nasir A., and Kim H. (2016). HGTree: database of horizontally transferred genes determined by tree reconciliation. *Nucleic Acids Res.*, *44*(D1), D610–D619. https://doi.org/10.1093/nar/gkv1245

Johnston C., Martin B., Fichant G., Polard P., and Claverys J. P. (2014). Bacterial transformation: Distribution, shared mechanisms and divergent control. *Nat. Rev. Microbiol.*, *12*(3), 181–196. https://doi.org/10.1038/nrmicro3199

Landy A. (2015). The λ Integrase Site-specific Recombination Pathway. *Microbiol. Spectr.*, *3*(2), MDNA3-0051–2014.

Lang A. S., Zhaxybayeva O., and Beatty J. T. (2012). Gene transfer agents: Phage-like elements of genetic exchange. *Nat. Rev. Microbiol.*, *10*(7), 472–482. https://doi.org/10.1038/nrmicro2802

Lesic B., Bach S., Ghigo J. M., Dobrindt U., Hacker J., and Carniel E. (2004). Excision of the high-pathogenicity of *Yersinia pseudotuberculosis* requires the combined of actions of its cognate integrase and Hef, a new recombination directionality factor. *Mol. Microbiol.*, *52*(5), 1337–1348. https://doi.org/10.1111/j.1365-2958.2004.04073.x

Levine J. A., Hansen A.-M., Michalski J. M., Hazen T. H., Rasko D. A., and Kaper J. B. (2014). H-NST induces LEE expression and the formation of attaching and effacing lesions in enterohemorrhagic *Escherichia coli*. *PLoS One*, *9*(1), e86618. https://doi.org/10.1371/journal.pone.0086618

Lewis J. A., and Hatfull G. F. (2001). Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res.*, *29*(11), 2205–2216.

Lu B., and Leong H. W. (2016). Computational methods for predicting genomic islands in microbial genomes. *Comput. Struct. Biotechnol. J.*, *14*, 200–206. https://doi.org/10.1016/j.csbj.2016.05.001

Maeda K., Nojiri H., Shintani M., Yoshida T., Habe H., and Omori T. (2003). Complete nucleotide sequence of carbazole/dioxin-degrading plasmid pCAR1 in *Pseudomonas resinovorans* strain CA10 indicates its mosaicity and the presence of large catabolic transposon Tn*4676*. *J. Mol. Biol.*, *326*(1), 21–33. https://doi.org/10.1016/S0022-2836(02)01400-6

Marcoleta A. E., Berríos-Pastén C., Nuñez G., Monasterio O., and Lagos R. (2016). *Klebsiella pneumoniae* Asparagine tDNAs Are Integration Hotspots for Different Genomic Islands Encoding Microcin E492 Production Determinants and Other Putative Virulence Factors Present in Hypervirulent Strains. *Front. Microbiol.*, *7*, 849. https://doi.org/10.3389/fmicb.2016.00849

Montero D. A., Velasco J., Del Canto F., Puente J. L., Padola N. L., Rasko D. A., Farfán M., Salazar J. C., and Vidal R. (2017). Locus of Adhesion and Autoaggregation (LAA), a pathogenicity island present in emerging Shiga Toxin–producing *Escherichia coli* strains. *Sci. Rep.*, *7*(1), 7011. https://doi.org/10.1038/s41598-017-06999-y

Müller C. M., Schneider G., Dobrindt U., Emödy L., Hacker J., and Uhlin B. E. (2010). Differential effects and interactions of endogenous and horizontally acquired H-NS-like proteins in pathogenic *Escherichia coli*. *Mol. Microbiol.*, *75*(2), 280–293. https://doi.org/10.1111/j.1365-2958.2009.06995.x

Murphy R. A., and Boyd E. F. (2008). Three Pathogenicity Islands of *Vibrio cholerae* Can Excise from the Chromosome and Form Circular Intermediates. *J. Bacteriol.*, *190*(2), 636–647. https://doi.org/10.1128/JB.00562-07

Navarre W. W. (2016). The Impact of Gene Silencing on Horizontal Gene Transfer and Bacterial Evolution. In R. K. Poole (Ed.), *Advances in Microbial Physiology* (Vol. 69, pp. 157–186). Academic Press. https://doi.org/10.1016/bs.ampbs.2016.07.004

Neale H. C., Jackson R. W., Preston G. M., and Arnold D. L. (2018). Supercoiling of an excised genomic island represses effector gene expression to prevent activation of host resistance. *Mol. Microbiol.*, *110*(3), 444–454. https://doi.org/10.1111/mmi.14111

Nepal S., Bonn F., Grasso S., Stobernack T., De Jong A., Zhou K., Wedema R., Rosema S., Becher D., Otto A., Rossen J. W., Van Dijl J. M., and Bathoorn E. (2018). An ancient family of mobile genomic islands introducing cephalosporinase and carbapenemase genes in enterobacteriaceae. *Virulence*, *9*(1), 1377–1389. https://doi.org/10.1080/21505594.2018.1509666

Newman R. M., Salunkhe P., Godzik A., and Reed J. C. (2006). Identification and Characterization of a Novel Bacterial Virulence Factor That Shares Homology with Mammalian Toll/Interleukin-1 Receptor Family Proteins. *Infect. Immun.*, *74*(1), 594–601. https://doi.org/10.1128/IAI.74.1.594

Nieto P. A., Pardo-Roa C., Salazar-Echegarai F. J., Tobar H. E., Coronado-Arrázola I., Riedel C. A., Kalergis A. M., and Bueno S. M. (2016). New insights about excisable pathogenicity islands in *Salmonella* and their contribution to virulence. *Microbes Infect.*, *18*, 302–309. https://doi.org/10.1016/j.micinf.2016.02.001

Panis G., Duverger Y., Courvoisier-Dezord E., Champ S., Talla E., and Ansaldi M. (2010). Tight regulation of the *intS* gene of the KplE1 prophage: A new paradigm for integrase gene regulation. *PLoS Genet.*, *6*(10), e1001149. https://doi.org/10.1371/journal.pgen.1001149

Panis G., Franche N., Méjean V., and Ansaldi M. (2012). Insights into the functions of a prophage recombination directionality factor. *Viruses*, *4*(11), 2417–2431. https://doi.org/10.3390/v4112417

Pardo-Roa C., Salazar G. A., Noguera L. P., Salazar-Echegarai F. J., Vallejos O. P., Suazo I. D., Schultz B. M., Coronado-Arrazola I., Kalergis A. M., and Bueno S. M. (2019). Pathogenicity island excision during an infection by *Salmonella enterica* serovar Enteritidis is required for crossing the intestinal epithelial barrier in mice to cause systemic infection. *PLOS Pathog.*, *15*(12), e1008152. https://doi.org/10.1371/journal.ppat.1008152

Piazolla D., Calì S., Spoldi E., Forti F., Sala C., Magnoni F., Dehò G., and Ghisotti D. (2006). Expression of phage P4 integrase is regulated negatively by both Int and Vis. *J. Gen. Virol.*, *87*(8), 2423–2431. https://doi.org/10.1099/vir.0.81875-0

Piña-Iturbe A., Suazo I. D., Hoppe-Elsholz G., Ulloa-Allendes D., González P. A., Kalergis A. M., and Bueno S. M. (2020). Horizontally Acquired Homologs of Xenogeneic Silencers: Modulators of Gene Expression Encoded by Plasmids, Phages and Genomic Islands. *Genes (Basel).*, *11*(2), 142. https://doi.org/10.3390/genes11020142

Piña-Iturbe A., Ulloa-Allendes D., Pardo-Roa C., Coronado-Arrázola I., Salazar-Echegarai F. J., Sclavi B., González P. A., and Bueno S. M. (2018). Comparative and phylogenetic analysis of a novel family of *Enterobacteriaceae*-associated genomic islands that share a conserved excision/integration module. *Sci. Rep.*, *8*(1), 10292. https://doi.org/10.1038/s41598-018-28537-0

Poulin-Laprade D., Carraro N., and Burrus V. (2015). The extended regulatory networks of SXT/R391 integrative and conjugative elements and IncA/C conjugative plasmids. *Front. Microbiol.*, *6*, 837. https://doi.org/10.3389/fmicb.2015.00837

Quiroz T. S., Nieto P. A., Tobar H. E., Salazar-Echegarai F. J., Lizana R. J., Quezada C. P., Santiviago C. A., Araya D. V, Riedel C. A., Kalergis A. M., and Bueno S. M. (2011). Excision of an Unstable Pathogenicity Island in *Salmonella enterica* Serovar Enteritidis Is Induced during Infection of Phagocytic Cells. *PLoS One*, *6*(10), e26031. https://doi.org/10.1371/journal.pone.0026031

Ramsay J. P., and Firth N. (2017). Diverse mobilization strategies facilitate transfer of non-conjugative mobile genetic elements. *Curr. Opin. Microbiol.*, *38*, 1–9. https://doi.org/10.1016/j.mib.2017.03.003

Ramsay J. P., Sullivan J. T., Stuart G. S., Lamont I. L., and Ronson C. W. (2006). Excision and transfer of the *Mesorhizobium loti* R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. *Mol. Microbiol.*, *62*(3), 723–734. https://doi.org/10.1111/j.1365-2958.2006.05396.x

Regmi A., and Boyd E. F. (2019). Carbohydrate metabolic systems present on genomic islands are lost and gained in *Vibrio parahaemolyticus*. *BMC Microbiol.*, *19*(1), 112. https://doi.org/10.1186/s12866-019-1487-6

Rodrigue D. C., Tauxe R. V, and Rowe B. (1990). International increase in *Salmonella enteritidis*: A new pandemic? *Epidemiol. Infect.*, *105*(1), 21–27.

Salazar-Echegarai F. J., Tobar H. E., Nieto P. A., Riedel C. A., and Bueno S. M. (2014). Conjugal Transfer of the Pathogenicity Island ROD21 in *Salmonella enterica* serovar Enteritidis Depends on Environmental Conditions. *PLoS One*, *9*(4), e90626. https://doi.org/10.1371/journal.pone.0090626

Shintani M., Sanchez Z. K., and Kimbara K. (2015). Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Front. Microbiol.*, *6*, 242. https://doi.org/10.3389/fmicb.2015.00242

Silva C. A., Blondel C. J., Quezada C. P., Porwollik S., Andrews-polymenis H. L., Toro C. S., Mcclelland M., and Santiviago C. A. (2012). Infection of Mice by *Salmonella enterica* Serovar Enteritidis Involves Additional Genes That Are Absent in the Genome of Serovar Typhimurium. *Infect. Immun.*, *80*(2), 839–849. https://doi.org/10.1128/IAI.05497-11

Soucy S. M., Huang J., and Gogarten J. P. (2015). Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.*, *16*(8), 472–482. https://doi.org/10.1038/nrg3962

Stoebel D. M., Free A., and Dorman C. J. (2008). Anti-silencing: Overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. *Microbiology*, *154*(9), 2533–2545. https://doi.org/10.1099/mic.0.2008/020693-0

Sullivan J. T., and Ronson C. W. (1998). Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Genetics*, *95*(9), 5145–5149. https://doi.org/10.1073/pnas.95.9.5145

Thomson N. R., Clayton D. J., Windhorst D., Vernikos G., Davidson S., Churcher C., Quail M. A., Stevens M., Jones M. A., Watson M., Barron A., Layton A., Pickard D., Kingsley R. A., Bignell A., Clark L., Harris B., Ormond D., Abdellah Z., Brooks K., Cherevach I., Chillingworth T., Woodward J., Norberczak H., Lord A., Arrowsmith C., Jagels K., Moule S., Mungall K., Sanders M., Whitehead S., Chabalgoity J. A., Maskell D., Humphrey T., Roberts M., Barrow P. A., Dougan G., and Parkhill J. (2008). 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. https://doi.org/10.1101/gr.077404.108.1

Tobar H. E., Salazar-Echegarai F. J., Nieto P. A., Palavecino C. E., Sebastian V. P., Riedel C. A., Kalergis A. M., and Bueno S. M. (2013). Chromosomal Excision of a New Pathogenicity Island Modulates *Salmonella* Virulence *In Vivo*. *Curr. Gene Ther.*, *13*(4), 240–249.

Vanga B. R., Ramakrishnan P., Butler R. C., Toth I. K., Ronson C. W., Jacobs J. M. E., and Pitman A. R. (2015). Mobilization of horizontally acquired island 2 is induced *in planta* in the phytopathogen *Pectobacterium atrosepticum* SCRI1043 and involves the putative relaxase ECA0613 and quorum sensing. *Environ. Microbiol.*, *17*(11), 4730–4744. https://doi.org/10.1111/1462-2920.13024

Waldor M. K. (2010). Mobilizable genomic islands: Going mobile with *oriT* mimicry. *Mol. Microbiol.*, *78*(3), 537–540. https://doi.org/10.1111/j.1365-2958.2010.07365.x

Williams K. P. (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res.*, *30*(4), 866–875. https://doi.org/10.1093/nar/30.4.866

Williamson H. S., and Free A. (2005). A truncated H-NS-like protein from enteropathogenic *Escherichia coli* acts as an H-NS antagonist. *Mol. Microbiol.*, *55*(3), 808–827. https://doi.org/10.1111/j.1365-2958.2004.04421.x

Xiong D., Song L., Geng S., Jiao Y., Zhou X., Song H., Kang X., Zhou Y., Xu X., Sun J., Pan Z., and Jiao X. (2019). *Salmonella* Coiled-Coil- and TIR-Containing TcpS Evades the Innate Immune System and Subdues Inflammation. *Cell Rep.*, *28*(3), 804–818. https://doi.org/10.1016/j.celrep.2019.06.048