Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Progetto 2: Biologia Cellulare, Molecolare e Industriale

Ciclo XXV

Settore concorsuale di afferenza: 05/E2

Settore scientifico disciplinare: BIO/11

TITOLO TESI

Functional and Genetic characterization of new genomic islands from an *E. coli* strain associated with neonatal meningitis.

Presentata da: Vanja Alberto Mariani Corea

Coordinatore Dottorato
Chiar.mo Prof.
Vincenzo Scarlato

Relatori
Chiar.mo Prof.
Vincenzo Scarlato
Dott.
Roberto Rosini
Dott.ssa
Michèle A. Barocchi

Esame finale anno 2013

ABSTRACT 5
1 INTRODUCTION: 7
1.1 ESCHERICHIA COLI 7
1.1.1 E. COLI CLASSIFICATION METHODS 8
1.2 E. COLI DIVERSITY: WHEN BOUNDARIES ARE NOT SO CLEAR. 10
1.3 PATHOGENESIS OF EXPEC 12
1.4 E. COLI AND GENETIC ISLANDS: EVOLUTION AT A FAST PACE. 15
1.4.1 GENOMIC ISLANDS AND VIRULENCE FACTORS IN EXPEC. 18
1.4.2 DICTYOSTELIUM DISCOIDEUM: A BACTERIAL HUNTER. 20
2 MATHERIALS AND METHODS. 22
2.1 BACTERIAL GROWTH. 22
2.1.1 BACTERIAL STRAINS. 22
2.1.2 ISOLATION OF CHROMOSOMAL DNA. 22
2.1.3 SINGLE AND MULTIPLE GEI-DELETION MUTANTS 23
2.1.3.1 Preparation of electro-competent cells. 23
2.1.3.2 Transformation of bacterial cells by electroporation. 23
2.1.3.3 Single and multiple Genomic island deletion by λ Red recombinase-mediated mutagenesis. 23
2.2 DICTYOSTELIUM DISCOIDEUM GROWTH AND GRAZING ASSAY. 24
2.2.1 WORKING CULTURE. 24
2.2.2 AMOEBA SPORE GENERATION. 25
2.2.3 BACTERIAL GROWTH CURVES. 25
2.2.4 DICTYOSTELIUM D. GRAZING ASSAY. 25
2.3 BIOINFORMATIC ANALYSIS AND PRIMER DESIGN. 26
2.3.1 PRIMER DESIGN. 27
2.3.2 STATISTICAL ANALYSIS. 28
2.4 GEI DISTRIBUTION AND EXCISION STUDIES. 28
2.4.1 PCR AMPLIFICATION AND SEQUENCING. 32
2.4.2 NUCLEASE RESISTANCE OF CIRCULAR INTERMEDIATES. 32
2.4.3 RELATIVE REAL-TIME PCR. 33
2.5 PRIMER LIST. 34
2.6 MEDIA AND BUFFERS. 40
2.6.1 MEDIA. 40
2.6.1.1 LB 40
2.6.1.2 SM 40
2.6.1.3 HL5 40
2.6.1.4 SoC 40
2.6.2 BUFFERS. 41
2.6.2.1 50x Soerensen buffer 41
2.6.2.2 Arabinose 20% 41

2.6.2.3 50x TBE buffer ... 41
2.6.2.4 SM Buffer (Phage extraction) ... 41

3 RESULTS ... 42

3.1.1 ExPEC isolates have a greater number of GEIs than InPEC and non- pathogenic strains. ... 42

3.2 Distribution of IHE3034 islands among a panel of diverse E. coli isolates. ... 43
3.3 Molecular and genetic characterization of IHE3034 genomic islands. ... 48

3.3.1 Deletion of genomic island (GEI) Integrases prevents the formation of circular intermediates. ... 49

3.3.2 GEI 11, 13 and 17 circular intermediates are resistant to DNAse treatment. ... 50

3.3.3 Growth conditions alter the excision rate of IHE3034 genomic islands. ... 51

3.4 IHE3034 genomic islands 13, 17 and 19 are linked to survival in the Dyctiostelium discoideum grazing assay. ... 55

3.4.1 GEI13, 17, 19 deletion affect IHE3034 ability to resist to the Dyctiostelium discoideum grazing assay. ... 55

4 DISCUSSION ... 58
5 BIBLIOGRAPHY ... 65

Abstract

Enterobacteriaceae genomes evolve through mutations, rearrangements and horizontal gene transfer (HGT). The latter evolutionary pathway works through the acquisition DNA (GEI) modules of foreign origin that enhances fitness of the host to a given environment. The genome of *E. coli* IHE3034, a strain isolated from a case of neonatal meningitis, has recently been sequenced and its subsequent sequence analysis has predicted 18 possible GEIs, of which: 8 have not been previously described, 5 fully meet the pathogenic island definition and at least 10 that seem to be of prophagic origin.

In order to study the GEI distribution of our reference strain, we screened for the presence 18 GEIs a panel of 132 strains, representative of *E. coli* diversity. Also, using an inverse nested PCR approach we identified 9 GEI that can form an extrachromosomal circular intermediate (CI) and their respective attachment sites (*att*). Further, we set up a qPCR approach that allowed us to determine the excision rates of 5 genomic islands in different growth conditions. Four islands, specific for strains appertaining to the sequence type complex 95 (STC95), have been deleted in order to assess their function in a *Dictyostelium discoideum* grazing assays.

Overall, the distribution data presented here indicate that 16 IHE3034 GEIs are more associated to the STC95 strains. Also the functional and genetic characterization has uncovered that GEI 13, 17 and 19 are involved in the resistance to phagocitation by *Dictyostelium d* thus suggesting a possible role in the adaptation of the pathogen during certain stages of infection.

1 Introduction:

1.1 *Escherichia coli*

Escherichia coli is a gram-negative bacteria belonging to the *gamma-proteobacteria* class of microorganisms. The vast majority of *Escherichia coli* strains live within the healthy human organism without causing disease; the colonization generally begins a few hours after birth setting up a mutual benefit relationship. *E. coli* are generally non-pathogenic bacteria but in an immune-compromised host they find a way to breach the gastrointestinal barriers it may happen that strains that were harmless in the digestive tracts start to cause diseases. *E. coli* has been identified as a versatile bacteria with a an ability to reorganize its genetic material in order to adapt to the environmental conditions in which it grows[42].

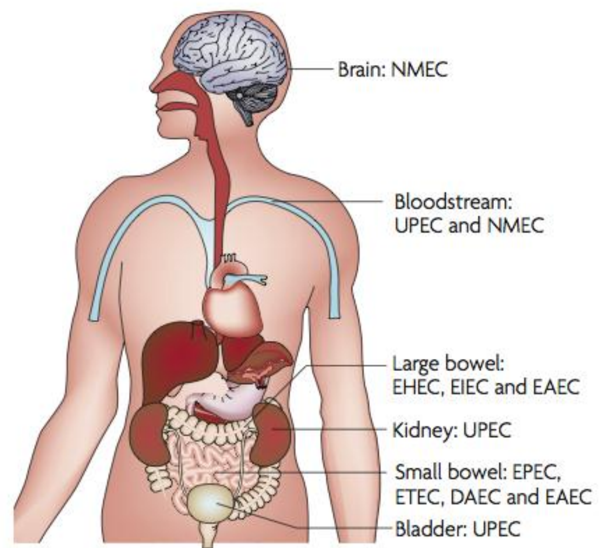


Fig. 1: Sites of pathogenic Escherichia coli colonization. (Croxen *et. al* 2010)
Pathogenic Escherichia coli colonize various sites in the human body. EPEC, ETEC, DAEC colonize the small bowel and cause diarrhoea, whereas EHEC, EIECcause disease in the large bowel; EAEC can colonize both the small and large bowels. UPECenters the urinary tract and travels to the bladder to cause cystitis. Septicaemia can occur with both UPEC and NMEC, whereas the latter can cross the blood–brain barrier into the central nervous system, causing meningitis 3.

Pathogenic *E. coli* can be divided into two major sub-groups depending on the location where they cause disease. The Intestinal Pathogenic *E. coli* (InPEC) cause bowel diseases such as diarrhea, bloody stools and comprises pathotypes such as enterotoxigenic (ETEC), enteropathogenic (EPEC), enterohemorrhagic (EHEC), enteroinvasive (EIEC), diffusely adherent (DAEC) all causing infections to the human intestinal tract. The second group of strains are the Extraintestinal Pathogenic *E. coli* (ExPEC) includes both human and animal pathogen causing urinary tract infections (UPEC) while others cause neonatal meningitides (NMEC) [5, 14].

ExPEC strains represent a major cause of morbidity, as it is responsible for 85-95% of uncomplicated cystitis cases and for over 90% of the episodes of uncomplicated pyelonephritis in premenopausal women. It has been estimated that 40-50% of

women will experience at least one case of UTI due to *E. coli* during the lifetime, with one fourth of these cases becoming a recurrent infection within 6 months of initial infection. Extra-intestinal strains are also responsible for episodes of catheter-associated UTIs (25-35%). NMEC together with *Streptococcus agalactiae* (GBS) are the leading causes of neonatal meningitis, accounting for an estimated 20 to 40% of the cases, with a fatality rate ranging from 25 to 40% and with neurological sequelae affecting 33 to 50% of survivors. These strains account for 17% of the cases of severe sepsis, with a mortality rate of approximately 30%. There are also strains that can be associated with intra-abdominal infections and nosocomial pneumonia and that occasionally participate in other extraintestinal infections, such as osteomyelitis, cellulitis and wound infections[25, 26, 65].

These kinds of diseases have never captured the public attention because they do not cause dramatic epidemics like those that cause food borne illness, thus underestimating the health and economic impact that they have. The high plasticity and rate of mutation of the *E. coli* genome is one variable that leads to the high number of diseases and the increasing antimicrobial resistance of the ExPEC strains. These characteristics translate into a large burden on the healthcare systems increasing the already heavy strain due to the actual economical environment; it is thus clear that a better understanding of *E. coli* is necessary in order to reduce the impact on healthcare[40].

1.1.1 *E. coli* classification methods

In order to classify *E. coli* many different typing methods have been developed; the most used ones are the multi-locus sequence typing (MLST) and the Phylogenetic groups.

MLST is a generic typing method for the molecular characterization of bacterial isolates that is robust and easily accessible. It has been employed principally, but not solely, to type bacterial pathogens and its strength relies on the fact that it is based explicitly on the same population genetics on which was based the multilocus enzyme electrophoresis (MLEE). MLST has the additional aims of providing a unified bacterial isolate characterization approach that generates data that can also be used for evolutionary and population studies of a wide range of bacteria regardless of their diversity, population structure, or evolution[47]. This methodology is based on the sequencing PCR of seven highly conserved housekeeping genes (*purA, adk, icd, fumC,*

recA, mdh, gyrB) as it can be seen in Fig. 2. The sequences are then concatenated and aligned through a program such as eBurst; each unique combination of alleles was assigned a sequence type (ST). Related STs were assigned to so-called ST complexes (STC), using the principles of the eBurst algorithm: each ST complex includes at least three STs that differ from their nearest neighbour by no more than two of the seven loci while ST complexes differ from each other by three or more loci. STs not matching the criteria for inclusion were referred to by their ST designation[81]. *E. coli* demonstrated to be a clonal bacteria and as Wirth and colleagues have pointed out the strains appertaining to a given ST or STC tend to have similar virulence factors and thus phenotypes. This is particularly true for very conserved STCs like: STC10 were almost all the non-pathogenic phylogenetic group A strains group or STC95 strains that are all of the phyl. group B2 and carry the K1 capsule gene cluster.

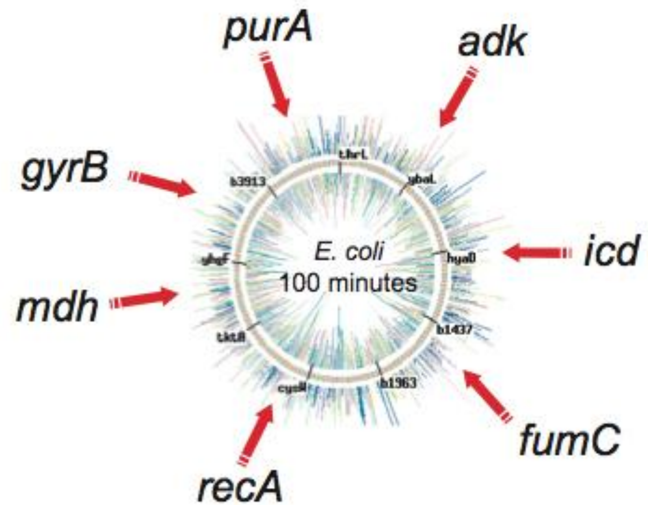


Fig. 2: MLST genes and genomic position (Wirth *et. al* 2010)
Genomic disposition of all the 7 alleles used for the MLST analysis.

This typing method is highly used by clinical microbiologists and epidemiologists as it is very robust and reproducible. The limit of this typing technique is that, in order to create an homogeneous classification of the bacterial populations, it requires a common agreement on the alleles (and their order) to be used [47].

Originally phylogenetical grouping was performed using techniques like MLEE and ribotyping that are complex, time-consuming and also require a collection of typed strains. Thirteen years ago (2000) Clermont *et. al.* proposed a rapid and simple method for typing Escherichia coli that uses a triplex polymerase chain reaction to amplify three targets genes. The markers were: chuA, a gene required for heme transport in enterohemorrhagic O157:H7; yjaA, a gene of unknown function identified in a *E. coli* K-12 strain and an anonymous DNA fragment designated TSPE4.C2.

The results of these three amplifications made it possible to establish a dichotomous decision tree (Fig. 3) that could attribute to any typed strain a phylogenetical group out of the four possible (A, B1, B2, D)[12]. This new typing method used by Clermont allows a faster and easier discrimination of the strains appurtenance to a phylogenetical group with an accuracy ranging from 80-85%[30]. The previous methods, due to not-common pattern of bands, assigned some strains to smaller sister groups (ABD, AxB1) that had a typing profile which was intermediate to between A and B1.

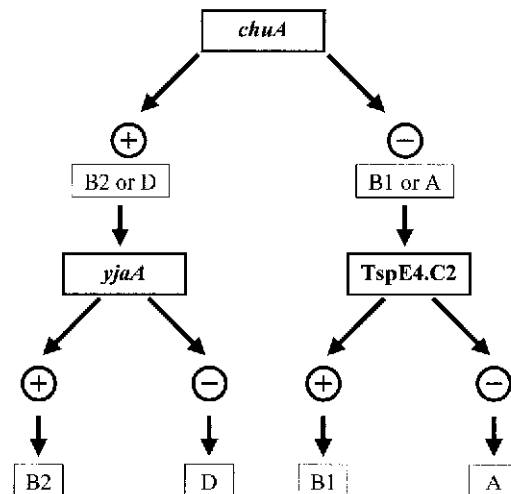


Fig. 3: Phylogenetic group decision tree (Clermont *et al.*)
Dichotomous decision tree to determine the phylogenetic group of an *E. coli* strain by using the results of PCR amplification of the *chuA* and *yjaA* genes and DNA fragment TSPE4.C2.

1.2 *E. coli diversity: when boundaries are not so clear.*

The high rate of mutation and plasticity of *E. coli* genome is the peculiarity that allows this bacteria to survive and thrive in different enviroments ranging from waste water to human/animal body. Bio-informatic analysis of the ever-growing collection of *E. coli* genomes allowed to understand that bacterial genomes comprise stable regions that form the “core” genome and variable regions that form the flexible gene pool. [3] Also genomic comparisons revealed that non-pathogenic *E. coli* genomes size varies from the 4,6Mb of the non-pathogenic strains to the 5.7Mb of the pathogenic and Asymptomatic Bacteriuria (ABU) strains. ExPEC virulence factors exhibit distinct patterns of phylogenetic distribution. This provides evidence of both, vertical and horizontal transmission of the corresponding virulence-associated genes as well as of host-specific associations and strong associations among different virulence-associated genes[18]. The constant typing effort allowed the identification of phylogenetic groups into which the major *E. coli* pathotypes cluster together (Fig. 4).

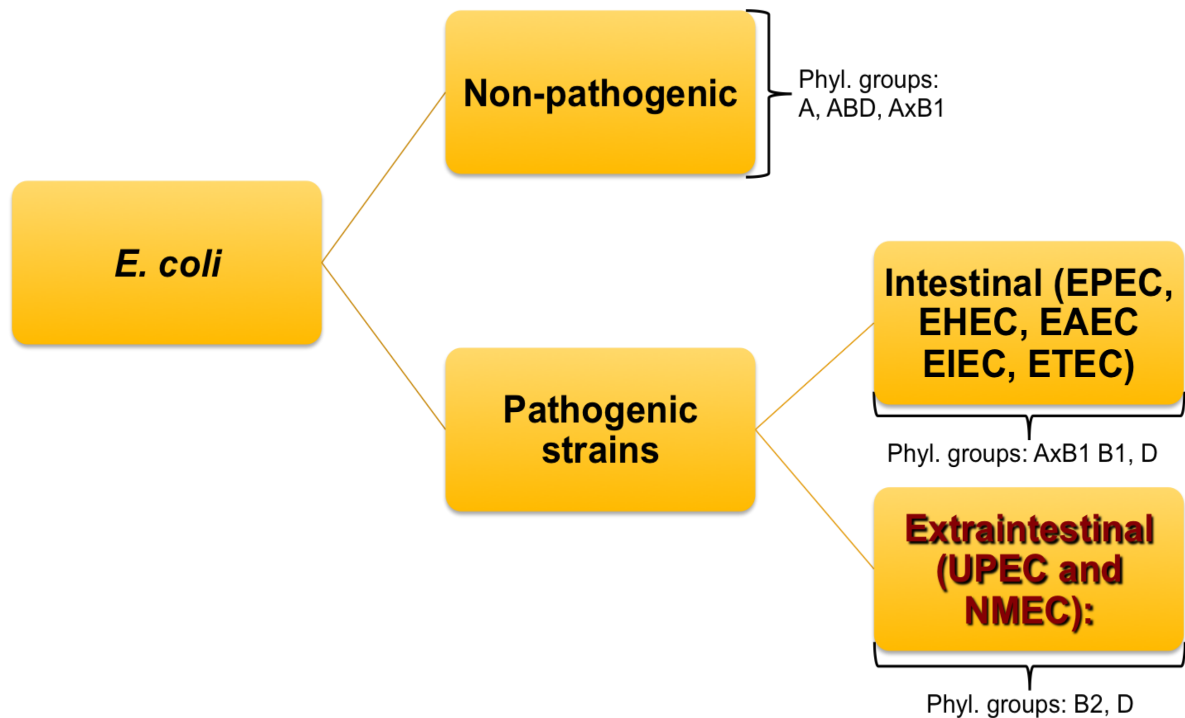


Fig. 4: Schematic representation of the *E. coli* pathogenic organization.

E. coli can be divided in pathogenic and non-pathogenic strains. Pathogenic strains can be further divided in Intestinal pathogenic or in Extra Intestinal pathogenic strains depending on where they cause a disease. In the image are represented to which phylogenetic groups the different pathogroups are associated.

The majority of the non-pathogenic strains cluster together in the group A, ABD, AxB1; while the intestinal pathogenic strains tend to cluster in the AxB1, B1 and D groups. It is important to understand that the groups AxB1 and ABD are sister groups to the B1 group. Intestinal pathogenic *E. coli* strains derive from phylogenetic groups A, B1 or D or from ungrouped lineages and are seldom found in the fecal flora of healthy individuals as the mere acquisition of these bacteria by the naïve host is sufficient for disease to ensue. Each intestinal pathotype possesses a characteristic combination of virulence and fitness factors that allow the colonization of specific niches and results in a unique diarrheal syndrome.

The B2 cluster is where almost all the ExPEC strains group while the remaining strains belong to cluster D. Extraintestinal strains have acquired various virulence genes that allow them to induce infections outside the digestive system in both normal and compromised hosts. ExPEC are incapable of causing gastrointestinal disease, but they can asymptomatically colonize the human intestinal tract and become the predominant strain in approximately 20% of normal individuals[40, 71].

ExPEC strains carry a broad range of virulence factors, distinct from those found in InPECs, that allow them to colonize host mucosal surfaces, avoid or subvert local and systemic host defense mechanisms, scavenge essential nutrients such as iron, injure or invade the host, and stimulate a noxious inflammatory response[40]. Due to extra-intestinal *E. coli* ability to survive either in or out of the gastrointestinal tract the definition of non-pathogenic strains has been hard to define. As colonizing sites outside the gut are unlikely to provide any selective advantage in terms of transmissibility, it is clear that any so-called ''extra-intestinal virulence factors'' are likely to have evolved to enhance survival in the gut and/or transmission between hosts, and therefore will be shared with at least some commensal strains. So this ability to fluctuate between mutualism, commensalism, opportunistic pathogenesis or even specialized pathogenesis make *Escherichia coli* the perfect candidate to study the boundaries between pathogenicity and commensalism[18, 74].

1.3 *Pathogenesis of ExPEC*

Among ExPEC strains, uropathogenic *E. coli* and neonatal meningitis *E. coli* are characterized by different molecular mechanisms of pathogenicity.

Urinary tract infection usually begins with the colonization of the bowel with a uropathogenic strain in addition to the commensal flora these strains, by virtue of its virulence factors, are able to colonize the periurethral area and to ascend the urethra to the bladder. Between 4 and 24 hours after infection, the new environmental conditions in the bladder select for the expression of type 1 fimbriae that allow the adhesion to the uroepithelium[42]. This attachment is mediated by fimbrial adhesin H (FimH), which is located at the tip of type 1 pili. FimH binds to mannose moieties of the receptors uroplakin Ia and IIIa that coat terminally differentiated superficial facet cells in the bladder, stimulating also unknown signaling pathways that induce invasion and apoptosis (Figure 5). Bacteria internalization is also mediated by FimH binding to α3 and β1 integrins that are clustered with actin at the sites of invasion, as well as by microtubule destabilization.

These interactions trigger local actin rearrangement by stimulating kinases and Rho-family GTPases, which results in the envelopment and internalization of the attached bacteria.

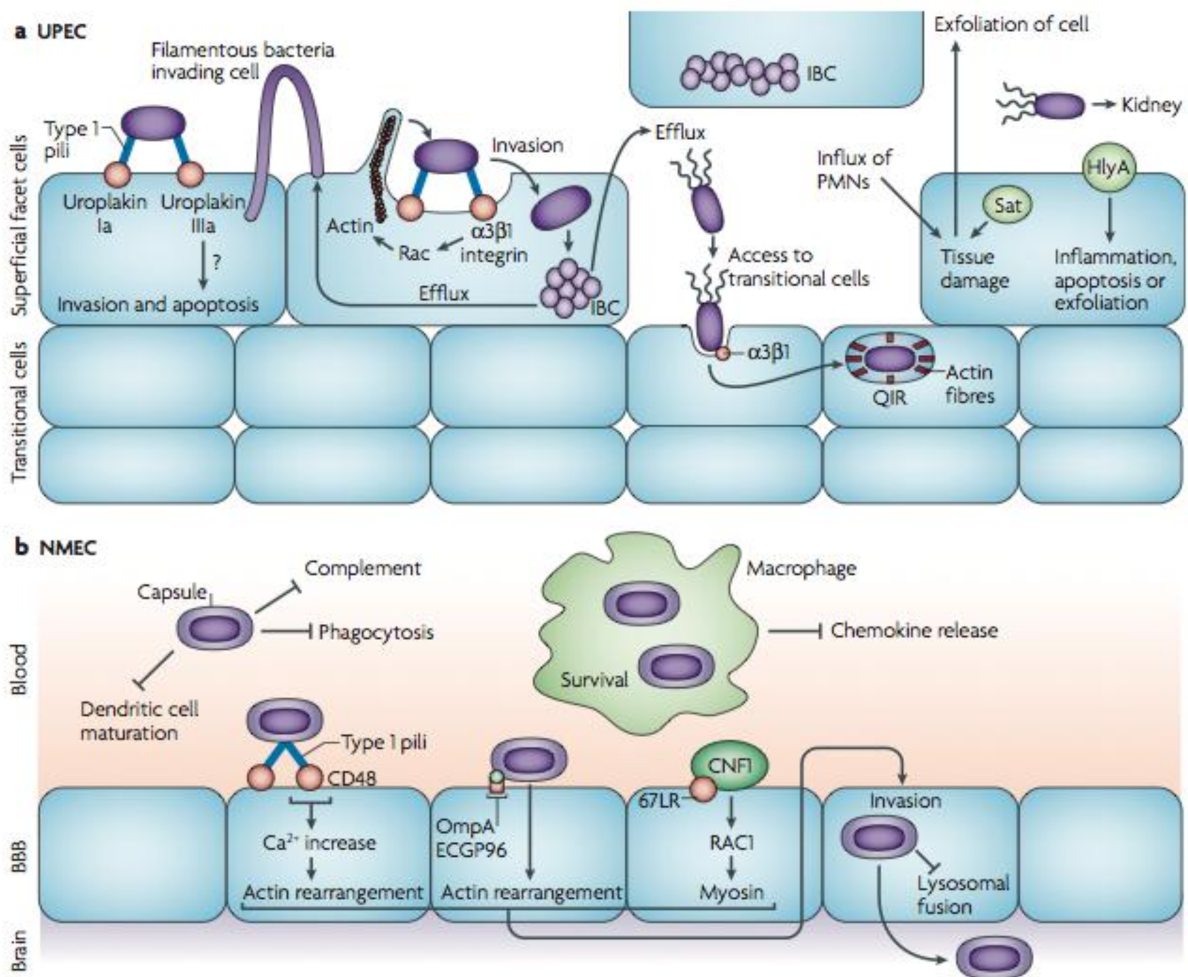


Fig. 5: Pathogenic mechanisms of ExPEC (Croxen and Finlay, 2010). The different stages of extraintestinal *E. coli* infections are shown. (*A*) UPEC attaches to the uroepithelium through type 1 pili, which bind the receptors uroplakin Ia and IIIa. Sublytic concentrations of the pore-forming toxin HlyA can inhibit the activation of Akt proteins and lead to host cell apoptosis and exfoliation. Exfoliation of the uroepithelium exposes the underlying transitional cells to further UPEC invasion. (*B*) NMEC is protected from the host immune response by its K1 capsule and outer-membrane protein A (OmpA). Invasion of macrophages may provide a replicative niche for high bacteremia, allowing the generation of sufficient bacteria to cross the blood-brain barrier (BBB) into the

Once internalized, UPEC can rapidly replicate and form biofilm-like complexes called intracellular bacterial communities (IBCs), which act as transient, protective environments. UPEC can also leave the IBCs through a fluxing mechanism and enter again the lumen of the bladder. Filamentous UPEC has also been observed fluxing out of an infected cell, looping and invading surrounding superficial cells in response to innate immune responses. During infection, the influx of polymorphonuclear leukocytes (PMNs) causes tissue damage, while apoptosis and exfoliation of bladder cells can be induced by UPEC attachment and invasion, as well as by sublytic concentrations of the pore-forming toxin HlyA. This breach of the superficial facet cells temporarily exposes the underlying transitional cells to UPEC invasion and

dissemination. Invading bacteria are trafficked in endocytic vesicles enmeshed with actin fibers, where replication is restricted. Disruption of host actin allows rapid replication, which can lead to IBC formation in the cytosol or fluxing out to the cell. This quiescent state may act as a reservoir that is protected from host immunity and may, therefore, permit long-term persistence in the bladder, as well as recurrent infections[14]. In strains causing cystitis, type 1 fimbriae are continuously expressed and the infection is confined to the bladder. In strains that are able to cause pyelonephritis, the invertible element that controls type 1 fimbriae expression turns to the "off" position and type 1 pili are less well expressed. This releases the UPEC strain from bladder epithelial cell receptors and allows the microorganism to ascend through the ureters to the kidneys, where it can attach by P fimbriae to digalactoside receptors that are expressed on the kidney epithelium. At this stage, hemolysin could damage the renal epithelium inducing an acute inflammatory response with the recruitment of PMNs to the infection site. Hemolysin has also been shown to cause calcium oscillations in renal epithelial cells, resulting in increased production of interleukin-6 (IL-6) and -8 (IL-8). Secretion of the vacuolating cytotoxin Sat damages glomeruli and is cytopathic for the surrounding epithelium. In some cases, bacteria can cross the tubular epithelial cell barrier and penetrate the endothelium to enter the bloodstream, leading to bacteremia[42]. The pathogenesis of NMEC strains is a complex mechanism, as the bacteria must enter the bloodstream through the intestine and ultimately cross the blood-brain barrier (BBB) into the central nervous system, which leads to meningeal inflammation and pleocytosis, that means presence of a higher number of cells than normal, in the cerebrospinal fluid (Fig. 5). Bacteria can be acquired perinatally from the mother and, after the initial colonization of the gut, they can translocate to the bloodstream by transcytosis through enterocytes. The progression of disease is dependent on high bacteremia ($>10^3$ colony forming units per ml of blood), therefore survival in the blood is crucial. NMEC is protected from the host immune responses by its K1 antiphagocytic capsule, made up of a homopolymer of polysialic acid, and by outer membrane protein A (OmpA), which confers serum resistance through manipulation of the classical complement pathway. NMEC has also been shown to interact with immune cells: invasion of macrophages and monocytes prevents apoptosis and chemokine release, providing a niche for replication before dissemination back into the blood. Bacterial attachment to the BBB

is mediated by FimH binding to CD48 and by OmpA binding to its receptor, ECGP96. Invasion of brain microvascular endothelial cells involves CNF-1 binding to the 67 kDa laminin receptor (67LR), which leads to myosin rearrangement, as well as OmpA and FimH binding to their receptors, which results in actin rearrangement. The K1 capsule, which is found in approximately 80% of NMEC isolates, also has a role in invasion by preventing lysosomal fusion and thus allowing delivery of live bacteria across the BBB. Collectively, these mechanisms allow NMEC to penetrate the BBB and gain access to the central nervous system, where they cause edema, inflammation and neuronal damage[14].

1.4 E. coli *and genetic islands: evolution at a fast pace.*

The ability to adapt and thrive across a huge diversity of hosts both human and animal make microbial pathogens a considerable threat all around the world [3]. The versatility that pathogens show is caused, at a molecular level, by the ability of the bacteria to adapt and evolve to evade detection.

Bacterial genome evolution is a continuous process that can be analysed from two points of view: a long-term 'macroevolution', which leads to the development of new species or subspecies over millions of years, and short-term 'microevolution', which spans shorter time frames (days or weeks) and leads to the alteration of genes and traits[84]. Bacterial evolution takes place following three main mechanisms of large-scale genome alteration: DNA deletions, rearrangements a point mutations, gene duplication and gene acquisition through

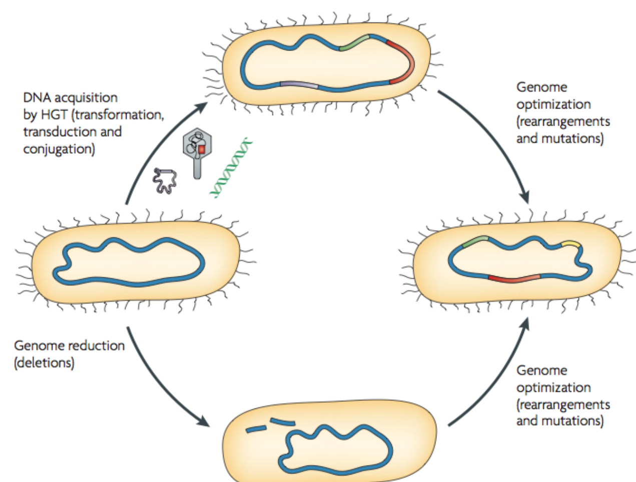


Fig. 6: Mechanisms that contribute to bacterial genome evolution(Ahmed *et. al.* 2004)

Genome plasticity results from DNA acquisition by horizontal gene transfer (HGT; for example, through the uptake of plasmids, phages and naked DNA) and genome reduction by DNA deletions, rearrangements and point mutations. The concerted action of DNA acquisition and gene loss results in a genome-optimization process that frequently occurs in response to certain growth conditions, including host infection or colonization.

horizontal gene transfer (HGT). Upon selection, such modifications to the genome, create subgroups of strains able to resist environmental stress and possibly cause a diseases using a common set of virulence/fitness factors (pathotypes) (Fig. 6).

As previously noted pathogenic genomes are bigger than non-pathogenic ones, bioinformatic analysis has shown and that these areas of difference are generally very variable, thus dividing the bacterial genome in very conserved "core" areas and variable areas that are more susceptible to rearrangements[22]. Such areas can be hotspots for insertions and stabilization pieces of DNA carried by phages, transposons and larger mobile chromosomal elements such as genomic islands (GEIs).

GEIs are very long non replicative mobile elements, ranging from 10Kbp to 120Kbp, that have features taken by other mobile elements (ICEs, prophages, plasmids,…) allowing them to integrate and excise from the genome. Given the great amount of genes carried by such mobile elements, the acquisition of a GEI, is generally considered to be a big evolutionary event that may cause a marked variation in the microorganism phenotype[31]. After such an event the genomic islands become integrant part of the bacteria and are subsequently subject to mutation to prevent further transmission and integration depending on the usefulness of the island itself[32]. Of course, the line that separates these conditions can be very subtle, according to the niche and to the right combination of factors.

Genomic islands take different names based on the kind of fitness advantage they furnish with the genes encoded on them: help the microorganism to live in the environment (ecological islands) or to persist as saprophyte (saprophytic islands), to colonize the host and provide benefit (symbiosis islands) or to cause disease (PAIs)[31, 60]. Most notably the recent German *E. coli* outbreak was caused by a mildly pathogenic InPEC strain integrating the Shiga toxin-encoding genomic island (*stx* island) in its own genome thus creating a new microorganism more fit to survive against the immune system.

Among the most characterized islands such as: the shiga phage or the high pathogenicity island (HPI). The *stx* island is a GEI of prophagic origin that carries the shiga toxin genes. This island has been throughoutly studied as it carries genes that significantly enhance the pathogenicity of the host and that are highly over expressed when DNA interfering or oxidative agents are added to the media[46]. The HPI island

or *Yersinia* island is devided in two portions: a conserved “core” portion and a variable AT-rich. The stable encodes a functional cluster of genes coding for biosynthesis, transport and regulation of the siderophore yersiniabactin, the recombinase gene and siderophore (int_{HPI}); while the AT-rich carries genes carries the excisionase (Xis_{HPI}) and Hex two genes foundamental for the island mobilization. It is of interest that this island is able to successfully colonize *Enterobacteriacea* such as *E. coli*, but in the majority of this strains (*Yersinia* excluded) this AT-rich zone is truncated and missing the *attR* site and thus is immobilized[4].

Genomic islands are identified by bioinformatic means as this genetic elements have very distinct features such as: the presence of an integrase gene, a GC content lower than the surrounding core DNA, the presence of a tRNA (facultative), the presence of direct repeat sites (*att* sites) at each side of the area. The integrase gene and the *att* sites play a fundamental role in the island mobilisation as they are the molecular machinery that allows GEIs to mobilize themselves. There are GEIs though missing some of this features that have been stabilized by the selective pressure; all these islands are not able to mobilize themselves anymore and generally carry virulence/fitness factors. To excide from the genome and release the plasmid-like structures, called circular intermediates (CI), in the cytoplasm the integrase protein brings the *att* sites close to each other, thus allowing for a site-specific recombination event to happen[9, 37, 50]. This ability to excise from the genome and create discrete CIs is thought to be an adaptation of the one used by bacteriophages to integrate and excide from the genome[50]. Continuous non-perfect integration and mobilization events may also have been the cause for the creation of this stretches of DNA that do still carry some prophagic elements, but should not able to create a fully working prophage. This assumption may be considered true to the point that genomic islands can be divided in two groups depending on their gene content and integrase gene (GEI-encoded, Phage-encoded)[55]. Also If we take into account that, for the prophages, the passage from a lysogen to a lytic cycle is considered to be their way to survive to stressful conditions[56, 78] we can also understand that the variations of genomic island excision rates in bacteria may be affected by external stress conditions (temperature, minimal medium, iron depletion, oxidative stress).

E. coli has been selected as the representative the pathogen genomic fluidity due to its high concentration of plasmid-mediated and phage-encoded virulence factors and

GEIs that have been fully described. Plasmids, phages and PAIs all play a crucial part in the evolution of different *E. coli* pathotypes[18, 42].

One main feature of the different intestinal *E. coli* pathotypes is the presence of pathotype-specific plasmids that often encode toxins. The characteristic protein toxins of enterotoxigenic, enteroaggregative, enteroinvasive, enterohaemorrhagic and enteropathogenic *E. coli* (and also extraintestinal pathotypes) are plasmid-encoded.

Also it is important to understand that as whole GEIs have a mosaic-like, modular structure and, although many of them superficially resemble each other (presence of certain virulence determinants), a great variability exists with regard to GEI composition, structural organization and chromosomal localization among strains even if they are of the same patho- or sero- type.

1.4.1 Genomic islands and virulence factors in ExPEC.

As previously stated genomic islands are the main effectors of the HGT due to their ability to transfer themselves from a donor to a host; their importance is also due to the high amount of open reading frames (ORFs), many of which of unknown origin, that encode for fitness or virulence factors.

IHE3034 is a neonatal meningitis strain appertaining to the phylogenetical group B2 and to the clonal complex (STC) 95 that it has been sequenced in 2010. The bioinformatic analysis carried out by Moriel *et. al.* uncovered 19 possible genomic islands present in IHE3034 (Tab. 1) [53].

Table 1

GEI	Virulence / Fitness Factors	Kbp	Related islands
1	Putative type VI secretion system	30	PAI II_{APECO1}
2	Prophage DNA	57	F-CFT073-smptB
3	Prophage DNA	22	Moriel DG et al.
4	Prophage DNA	33	Moriel DG et al.
5	S-perfimbriae, IroN, putative TonB-dependant receptor, Antigen 43	61	PAI III_{536}, PAI-CFT073-serX, PAI $I_{Nissle1917}$
6	sitABCD	47	PAI-CFT073-icdA
7	Prophage DNA	46	F-CFT073-potB
8	Yersiniabactin and cdtABC	78	PAI-CFT073-asnT
9	Colibactin gene cluster	54	PAI VI_{536}, GI-CFT073-asnW
10	Putative TonB-dependent receptors and ibrAB	44	PAI VI_{536}, GI-CFT073-cobU, PAI IV_{APECO1}
11	Prophage DNA	37	Moriel DG et al.
12	Prophage DNA	40	Moriel DG et al.

13	Enterohemolysin 1	39	Moriel DG et al.
14	Prophage DNA	43	Moriel DG et al.
15	Putative type VI secretion system	36	PAI-CFT073-metV, PAI-536-metV
16	T2SS and K1 capsule	28	PAI V_{536}, PAI I_{APEC01}
17	Prophage DNA	16	Moriel DG et al.
18	IbeA and IbeT	20	GimA
19	Prophage DNA	46	Moriel DG et al.

Among virulence factors carried by ExPEC GEIs, a fundamental role is played by adhesins (GEI 5), which allow the strict interaction of the pathogen with the host, facilitating the colonization and invasion processes and avoiding clearance by the host immune defences. Also the presence of group K1 (GEI 16) capsule confers additional selective advantages to ExPEC strains. Indeed, their molecular mimicry to host tissue components helps the bacteria to evade the immune response, providing protection against phagocytic engulfment and complement-mediated bactericidal activity[24, 79]. GEI 16 though is not a genomic island but a hotspot of integration; bioinformatic analysis has shown that it is a highly variable region and that the tRNA present in the middle of it is a typical insertion point for mobile elements. Other proteins are also associated with the virulence of ExPEC strains. For example IbeA and IbeT (GEI 18) that are involved in the invasion of brain microvascular endothelial cells [38, 85]. Antigen 43 (Ag43 – GEI 5) is associated with a strong aggregation phenotype and with biofilm formation, promoting long-term persistence in the bladder, although its relevance and contribution in the pathogenesis are far from clear[76].

Growth of ExPEC strains in iron-limited conditions, such as urine, requires successful mechanisms for the scavenging of iron, which rely on siderophores and iron-complex receptors [80]. Several iron and siderophore receptors, which are highly expressed during infection of the urinary tract, have already been described in *E. coli*, for example the salmochelin siderophore receptor IroN[33] and the ferric and manganese receptor sitABCD[83].

Eight out of nineteen islands of IHE3034 islands are or prophagic origin and have been identified for the first time by Moriel *et. al.*; many of the ORFs on these GEIs are of unknown function. These islands altogether account for the 0,5% of the genome of IHE3034 and the percentage of ORFs of known function ranges from 50% to 75%. Understanding the mechanisms behind GEI mobilization and functions is a key point

to develop preventive and therapeutic approaches that could aim to selectively induce PAI deletion and reduce the incidence of *E. coli* diseases.

1.4.2 *Dictyostelium discoideum*: a bacterial hunter.

Bacteria like *E. coli* are mainly environmental microorganism; they live in the soil were they are constantly threatened by bacteria-eating predators such as protozoa and nematodes. These evolutionary pressures may affect bacterial populations in multiple ways, like creating defense strategies that allow them to survive and to establish new replicative niches. For example, to protect themselves from predators, produce biofilms thus preventing engulfment and phagocytosis, or use molecular machinery to avoid lysosomal killing[34]. As a soil amoeba and a phagocyte *Dictyostelium discoideum* can be a natural host of opportunistic bacteria that may have developed strategies to invade, survive and replicate intracellularly inside the amoeba itself[10].

D. discoideum is a fascinating member of the amoebozoa, its natural habitat is deciduous forest soil and decaying leaves, where the amoebae feed on bacteria, yeast and grow as separate, independent, single cells. The organism offers unique advantages for studying fundamental cellular processes with powerful molecular genetic, biochemical, and cell biological tools[23].

Phagocytosis is a very complex, evolutionarily conserved mechanism that is used by higher eukaryotes to clear dead cells and cell debris and to counter the constant threat posed by pathogens. For this purpose they harbour specialized cells such as macrophages, neutrophils or dendritic cells that have the ability to rapidly and efficiently internalize a variety of organisms and particles and degrade them. For lower eukaryotes like *D. discoideum* phagocytosis is a means to internalize bacteria that are used as food source. The ingested microorganism is trapped in a phagosome and, via the phago-lysosomal pathway, is ultimately delivered to a lysosome where it is degraded by a cocktail of hydrolytic enzymes[11, 13].

Bacterial pathogenicity was certainly largely developed to resist predatory bacteriovorous microorganisms in the environment, and this accounts for the fact that a large number of bacterial virulence traits can be studied using *Dictyostelium* as a host.

The increasing number genome sequences and the genetic tractability of *E. coli* generate many opportunities for the study of host-pathogen interactions. The use of

Dictyostelium cells as a screening system for bacterial virulence combining the use of *E. coli* mutant cells will allow to identify determinants of susceptibility and resistance to infection providing a particularly powerful, simple and animal-free system [13].

2 Matherials and Methods.

2.1 *Bacterial growth.*

2.1.1 Bacterial strains.

IHE3034 (O18:K1:H7), ST95, is a neonatal meningitis-associated strain isolated in Finland in 1976[1]. Its genomic sequence has been revealed in 2012[53].
The 132 E. coli strains extra intestinal, intestinal pathogenic and non-pathogenic that have been used in this analysis are described in table 3. The eleven ST131 strains have been kindly provided by Marina Cerquetti from the Istituto Superiore di Sanità (Rome). 77 strains mixed ExPEC, InPEC and non-pathogenic have been kindly provided by Lothar H. Wieler from the Freie Universität Berlin. 35 strains ExPEC, InPEC and non-pathogenic have been kindly provided by Ulrich Dobrindt from the Universitätklinikum of Münster. The collection is composed of strains belonging to the A, AxB1, ABD, B1, B2 phylogenetic group[39]. Throughout the manuscript, GEI deletion mutants (partial or whole) of E. coli strain IHE3034 are named by the symbol "ΔG" followed by the numbers of the deleted GEI and if needed, the letter of the deleted portion. The numbers indicating the GEI are expressed in the Arabic form instead of the Roman one (classically used to number GEIs in the literature) to ease readability. Bacteria were routinely grown in LB broth at 37°C except when otherwise stated. Ampicillin (Amp 100μg/ml), Kanamycin (Kan 25μg/ml), Trimethoprim (Trim 100μg/ml), Mitomycin C (Mit. C 0,5μg/ml) or Chloramphenicol (Clm 8 μg/ml) were added to the media when necessary.

2.1.2 Isolation of chromosomal DNA.

Genomic DNA was prepared by culturing bacteria overnight at 37 °C, with antibiotics added when needed, and left overnight in an orbital shaker. The extraction took place the following day using either the GenElute Bacterial Genomic DNA Kit (Sigma) according to the manufacturer's instructions or preparing a raw genomic extract. Final DNA concentration, of the genomic kit preparation sample, was assessed by optical density determination at 260 nm.
Raw genomic extract preparations were prepared adding 200μl of ON culture (c.a. $4,6*10^7$ CFU) in a clean Eppendorf. The culture was centrifuged and the supernatant was removed; the pellet was re-suspended in 100μl of PCR-grade water and boiled

for 10 minutes. The raw extraction was then centrifuged for 5 minutes at 1100 g in a table top centrifuge (Eppendorf) and the supernatant was transferred in a new tube.

2.1.3 Single and multiple GEI-deletion mutants .

2.1.3.1 Preparation of electro-competent cells.

For electro-competent IHE3034 cell preparation, 2ml LB were inoculated starting from the glycerol stocks and set to grow overnight at 37°C shaking at 180rpm. If the red recombinase (p434, pKOBEG) or the flipase plasmid (pCP20) were present the culture was grown at 30°C. The next day 25ml of fresh LB were inoculated to a final OD/ml of 0,1 and left to grow, with Arabinose to a final concentration 0,2%, up to 0,6-0,8OD/ml. When the OD/ml of 0,6-0,8 is reached the culture was poured into a Falcon tube and the cells precipitated for 30min at 3650g at 4°C in a Heraeus MULTIFUGE 3 S-R centrifuge with a 75006445 rotor. The pellet was then washed 3 times with 25ml of cold sterile water (4°C) and one time with 25ml of cold a 10% glycerol solution. The pellet was resuspended in 500μl of 10% glycerol solution and divided in 60μl aliquots in Eppendorf tubes and stored at -80°C.

2.1.3.2 Transformation of bacterial cells by electroporation.

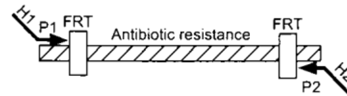
For electroporation 60μl of electro-competent cells were thawed on ice and mixed with 1-12μl of plasmid or purified PCR constructs to the final concentrations of up to 100ng plasmid, 1μg PCR product. Cells were transferred in a 1-mm-wide GenePulser electroporation cuvette and then electroporated using program Ec1 of the Biorad GenePulser Xcel (1,8kV). Transformations with a time constant no lower than 4 were recovered in 250μl of SoC media and set to grow 1-2hrs at 37°C (30°C for temperature-sensitive plasmids) in a shaking thermal block before being plated on selective LB-Agar plates. The following day colonies were PCR screened after being streaked in a fresh plate.

2.1.3.3 Single and multiple Genomic island deletion by λ Red recombinase-mediated mutagenesis.

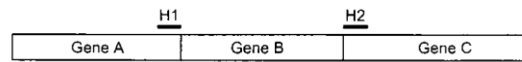
Δ 4, Δ 13, Δ 17 and Δ19: each single GEI deletion mutant was generated using the λ-Red recombinase gene inactivation method [16]. Flipase recognition target (FRT)-flanked kanamycin or chloramphenicol cassettes were generated by PCR using as a template pKD4 or pKD3. Primers carried tails of 60 to 71 bases homologous with both ends of the GEI to be deleted (Fig. 7).

PCR conditions are listed in paragraph 2.4.1. Since tRNA and certain genes deletions have been shown to result in a decreased fitness[21, 63, 72] the primers were designed, for all but GEI 17, so that this regions flanking the GEIs were left untouched (Primer #184 to #223). Strains with GEIs partially deleted were also created following the lambda-red protocol with GEIs have divided in 2 to 4 regions named A to D. Each section has been selected so that gene continuity and operons were not interrupted.

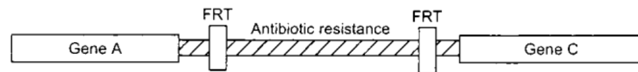
Step 1. PCR amplify FRT-flanked resistance gene



Step 2. Transform strain expressing λ Red recombinase



Step 3. Select antibiotic-resistant transformants



Step 4. Eliminate resistance cassette using a FLP expression plasmid

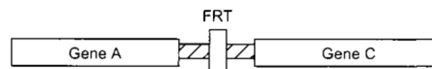


Fig. 7: One-step gene inactivation (Datsenko *et. al.* 2000)

H1 and H2 refer to the homology tails. P1 and P2 refer to priming sites. FRT sites refer to the recombinations sites used by the Flipase (pCP20) to excide the resistance.

Multiple Knock-out strain: t□□ Δ4 mutant was used to successively remove GEIs, 13, 17, and 19 using the λ Red recombinase method as for single GEI deletion mutants. After the deletion of two GEIs, the antibiotic resistance cassette was removed using Flp before proceeding to the next GEI deletion [16].

2.2 *Dictyostelium discoideum growth and grazing assay.*

Amoeba spore aliquots where thawed and allowed to grow for three days in 10ml of HL5 medium in 25cm^2 cell-culture flasks in order to generate pre-cultures[27, 70]. All the amoeba cultures were grown at room temperature (21-25°C) in a thermally controlled laboratory.

2.2.1 Working culture.

After all the spores have germinated 5 ml of pre-culture were inoculated in 15ml of fresh HL5 medium to generate a new Working culture a 75cm^2 cell-culture flask. The new culture was then left to grow for three to four days in a temperature controlled room at 21-25°C. [27, 70].

2.2.2 Amoeba spore generation.

A shaking culture of 25ml HL5 was inoculated in a conical cylinder using 1,5-2ml of matured pre-culture. After 3 to 4 days the cells are counted using the Neubauer improved counting chamber using the following formula:

$$\frac{\#\, cell \ * Dil.factor \ * 10^4}{\#\, counted\ squares} = {cell}/{ml}$$

If needed cell were concentrated in Soerensen Buffer 1x[28] to a final concentration no less than $1x10^7$cell/ml. In each Soerensen-Agar Plate were then plated 400µl of amoeba resuspension and gently distributed by rotating the plate with circular motions.

The plates were put in a closed-lid box with wet towel papers in the bottom and left to incubate for 3-4 days at room temperature. The *Dyctyostelium d.* yellowish spore heads were harvested by rinsing the plates with 2ml of Soerensen buffer 1X until the majority of the spore heads were in solution. Each stock cryotube was filled with 1ml of spore suspension and stored at -80°C[28].

2.2.3 Bacterial growth curves.

The *Dyctiostelium discoideum* grazing assay is heavily influenced by the bacterial fitness. In order to confirm that the assay is not biased by such a problem, both IHE3034 wild type and all the mutant strains have been tested in a growth curve assay. This analysis has been carried out to assess if the deletions that have been made reduce the fitness of the bacteria.

For each strain, a 0,01 OD/ml LB or SM inoculum was prepared in a 96 well plate with a final volume of 200µl. The plate was then sealed tight with parafilm and placed in a pre-warmed Tecan Infinite 200 with a Heating and Shaking module. The instrument temperature has been pre-set to 37°C shaking at 180rpm; the OD measurements (600nm) were taken every 10 min for 24hrs. The data gathered were then plotted in an OD – Time graph with the X axis (OD) expressed in logarithmic scale.

2.2.4 *Dictyostelium d.* grazing assay.

The IHE3034 wild type and the deletion mutant *E. coli* strains were inoculated from the glycerol stock and let to grow for 3-4 hrs in LB media at 37°C shaking at 180rpm. 300µl of bacterial growth (c.a. $4,5x10^8$CFU) were then evenly plated on fresh SM-Agar petri dishes using a plating spatula and left to dry in a laminar flow hood for

1 hr. The amoeba population in the working culture was then quantified and four working dilutions were prepared in HL5 medium: 10^6cell/ml, $2x10^5$cell/ml, $2x10^4$cell/ml and $2x10^3$cell/ml. The bacterial SM plates were divided in four areas in which different amount of amoeba cells (5000, 1000, 100 and 10) were plated. In order to reduce the experimental variability two 5μl drops were placed on each plate; each working dilution was vortexed after each use to reduce the error caused by the fast sedimentation rate of the amoeba cells. The plates were then left at room temperature for a variable period between 3 and 7 days and the plaque formation assessed daily; each plate was photographed with an acquiring time of 230ms using the Protocol 2 (Synbiosis) cell counter. The plating assay was repeated 3 times for each bacterial strain under study and for the two control strains. As a negative control, immune to the amoeba grazing, the highly virulent UPEC strain 536 was used and due to its susceptibility to *Dictyostelium*, the non-pathogenic *E. coli* DH5α has been selected as positive control.

2.3 *Bioinformatic analysis and primer design.*

All of the *E. coli* genomes sequences used for the analysis were gathered from the NCBI website (www.ncbi.gov) and have been imported in the Geneious 5.6 software[49] (for the list of the genomes refer to Tab. 2). Full genome alignments have been carried out using the plug-in MAUVE program implemented on Geneious using the default settings[15]. The primer designs have been carried out using the Primer 3 plugin for Geneious [67]. Real time data analysis and images generation has been don using the REST2009 program[57, 62]

Table 2

	Name	Pathotype	Group	Sequence Length (Bp)
1	ABU 83972	ABU	Non-pathogenic	5131397
2	LF82	AIEC	InPEC	4773108
3	O83:H1 str. NRG 857C	AIEC	InPEC	4747819
4	UM146	AIEC	InPEC	4993013
5	APEC O1	APEC	ExPEC	5082025
6	SMS-3-5	AREC	InPEC	5068389
7	042	EAEC	InPEC	5241977
8	55989	EAEC	InPEC	5154862
9	O26:H11 str. 11368 DNA	EHEC	InPEC	5697240
10	O103:H2 str. 12009	EHEC	InPEC	5449314
11	O157:H7 EDL933	EHEC	InPEC	5528445

12	O157:H7 str. EC4115	EHEC	InPEC	5572075
13	O157:H7 str. Sakai	EHEC	InPEC	5498450
14	O157:H7 str. TW14359	EHEC	InPEC	5528136
15	O111:H- str. 11128	EHEC	InPEC	5371077
16	O55:H7 str. CB9615	EPEC	InPEC	5386352
17	O127:H6 str. E2348/69	EPEC	InPEC	4965553
18	E24377A	ETEC	InPEC	4979619
19	H10407	ETEC	InPEC	5153435
20	UMNK88	ETEC	InPEC	5186416
21	ATCC 8739	Laboratory strain	Non-pathogenic	4746218
22	BL21(DE3)	Laboratory strain	Non-pathogenic	4570938
23	B str. REL606	Laboratory strain	Non-pathogenic	4629812
24	DH1 (ME8569)	Laboratory strain	Non-pathogenic	4621430
25	BW2952	Laboratory strain	Non-pathogenic	4578159
26	DH10B (K-12)	Laboratory strain	Non-pathogenic	4686137
27	MG1655 (K-12)	Laboratory strain	Non-pathogenic	4639675
28	W3110 (K-12)	Laboratory strain	Non-pathogenic	4646332
29	KO11	Laboratory strain	Non-pathogenic	4920168
30	MDS42 (K-12)	Laboratory strain	Non-pathogenic	3976195
31	W	Laboratory strain	Non-pathogenic	4900968
32	IHE3034	NMEC	ExPEC	5108383
33	O7:K1 str. CE10	NMEC	ExPEC	5313531
34	S88	NMEC	ExPEC	5032268
35	HS	Non-pathogenic	Non-pathogenic	4643538
36	SE11	Non-pathogenic	Non-pathogenic	4887515
37	SE15 DNA	Non-pathogenic	Non-pathogenic	4717338
38	IAI1	Non-pathogenic	Non-pathogenic	4700560
39	ED1a	Non-pathogenic	Non-pathogenic	5209548
40	536	UPEC	ExPEC	4938920
41	CFT073	UPEC	ExPEC	5231428
42	IAI39	UPEC	ExPEC	5132068
43	NA114	UPEC	ExPEC	4971461
44	str. 'clone D i2'	UPEC	ExPEC	5038386
45	str. 'clone D i14'	UPEC	ExPEC	5038386
46	UMN026	UPEC	ExPEC	5202090
47	UTI89	UPEC	ExPEC	5065741

2.3.1 Primer design.

The distribution studies were carried out using 3 sets of primers; the first set was a single pair of primers while the second and third sets were combined together in a Multiplex PCR design. Primers (#72 to #183) have been designed on three open reading frames (ORF) spanning along the whole island. Each ORF was selected only if it was present, in the BLAST analysis, in IHE3034 or in pathogenic strains and absent

in non-pathogenic strains.

Primers for the identification of the circular intermediates (CI) and for the exclusion PCR were designed to be functional only if the CI was formed and thus the island was not integrated in the genome (Primer #1 to #71).

Primers for the generation of knock out strains were designed to have a fixed portion that could anneal on pKD3/4[16] and had a tail spanning between 60 and 71bp that was completely homologous to the flanking regions that were to be knocked-out (#184-#223).

Primers designed for the screening of the knock-out strains were a forward annealing at the 5' region of the KO region and a reverse annealing either on the inserted resistance gene or on the 3' of the insertion region. The external oligonucleotides were also used to test if the islands had lost their resistances after pCP20 transformation (#241-#262).

Primers for the Relative Real Time PCR were designed to be no more than 20bp and each pair had to have a Tm of 60°C, equal GC content when possible, a Ta difference within 0,5°C and a final amplicon of no more than 350bp. Before being used in the real-time experiment a test PCR was run and the product of these primers were analysed on an agarose gel for aspecific bands. All the products were sequenced to check for the specificity of the reaction (#224-#240).

2.3.2 Statistical analysis.

Data were analyzed by Fisher's exact tests to evaluate associations. Results with p values lower than 0.5 indicate a low significance, 0,05 indicate good significance while a p value lower than 0,01 indicates a strong significance. The statistical significance of expression ratios of the real time data has been calculated using the integrated randomization and bootstrapping methods in the REST2009 program[77].

2.4 *GEI distribution and excision studies.*

Using the PCR approach a panel of 132 *E. coli* isolates (listed in table 2) were examined for the presence of 3 genes carried on the genomic islands of IHE3034[53]. The GEIs have also been analyzed for their ability to form circular intermediates, their dependence on the *int* gene to excide and for CI resistance to DNase.

Table 3

	Strain	ST	STC	Pathotype	Group	Phyl. Group
1	042	n.a.	n.a.	EAEC	InPEC	n.a.
2	764	14	14	UPEC	ExPEC	B2
3	3970	155	155	ETEC	InPEC	B1
4	40956	n.a.	n.a.	EAEC	InPEC	n.a.
5	A5/10	40	40	STEC	InPEC	B1
6	05-07839-2	38	n.a.	EAEC	InPEC	D
7	06-04456-2	131	n.a.	EAEC	InPEC	B2
8	08-24489	10	n.a.	EAEC	InPEC	A
9	303/89	29	29	EHEC	InPEC	B1
10	312/00	n.a.	n.a.	EPEC	InPEC	n.a.
11	350 C1A	10	n.a.	ETEC	InPEC	n.a.
12	37-4	n.a.	n.a.	EPEC	InPEC	n.a.
13	413/89-1	113	29	STEC	InPEC	B1
14	537/89	298	306	STEC	InPEC	nt
15	540/00	n.a.	n.a.	EPEC	InPEC	n.a.
16	5477/94	n.a.	n.a.	EAEC	InPEC	n.a.
17	7476A	58	n.a.	ETEC	InPEC	n.a.
18	76-5	n.a.	n.a.	EIEC	InPEC	n.a.
19	B10363	95	95	NMEC	ExPEC	B2
20	B13155	390	95	NMEC	ExPEC	B2
21	B616	390	95	NMEC	ExPEC	B2
22	E 34420 A	1312	n.a.	ETEC	InPEC	n.a.
23	E1392-75	2353	n.a.	ETEC	InPEC	n.a.
24	E22	20	20	EPEC	InPEC	B1
25	E2348/69	n.a.	n.a.	EPEC	InPEC	n.a.
26	E457	95	95	Commensal	Non-pathogenic	B2
27	Ecor19	48	10	Commensal	Non-pathogenic	A
28	Ecor20	48	10	Commensal	Non-pathogenic	A
29	Ecor34	58	155	Commensal	Non-pathogenic	AxB1
30	Ecor35	59	59	Commensal	Non-pathogenic	ABD
31	Ecor64	14	14	UPEC	ExPEC	B2
32	EDL1284	n.a.	n.a.	EIEC	InPEC	n.a.
33	F630	23	23	APEC	ExPEC	B1
34	F645	62	n.a.	SEPEC	ExPEC	ABD
35	F911	12	12	SEPEC	ExPEC	B2
36	H10407	n.a.	n.a.	ETEC	InPEC	n.a.
37	IHE3034	95	95	NMEC	ExPEC	B2
38	IHE3036	390	95	NMEC	ExPEC	B2
39	IHE3080	390	95	NMEC	ExPEC	B2
40	IHIT0578	29	29	EHEC	InPEC	B1
41	IHIT0608	28	28	EHEC	InPEC	ABD
42	IHIT2087	21	29	STEC	InPEC	B1
43	IMT10651	10	10	ExPEC	ExPEC	A
44	IMT10666	58	155	Commensal	Non-pathogenic	AxB1

45	IMT10740	1159	n.a.	Commensal	Non-pathogenic	B2
46	IMT14782	69	69	Commensal	Non-pathogenic	D
47	IMT14967	73	73	UPEC	ExPEC	B2
48	IMT14973	12	12	UPEC	ExPEC	B2
49	IMT14993	127	127	UPEC	ExPEC	B2
50	IMT15000	95	95	UPEC	ExPEC	B2
51	IMT15006	88	23	UPEC	ExPEC	B2
52	IMT15007	141	n.a.	UPEC	ExPEC	B2
53	IMT15009	80	568	UPEC	ExPEC	B2
54	IMT15010	12	12	UPEC	ExPEC	B2
55	IMT15014	117	117	UPEC	ExPEC	ABD
56	IMT15019	127	127	UPEC	ExPEC	B2
57	IMT15020	88	23	UPEC	ExPEC	B1
58	IMT15146	95	95	Commensal	Non-pathogenic	B2
59	IMT15150	131	131	Commensal	Non-pathogenic	B2
60	IMT15991	32	32	STEC	InPEC	ABD
61	IMT16101	73	73	Commensal	Non-pathogenic	B2
62	IMT17424	10	10	UPEC	ExPEC	A
63	IMT1930	88	23	APEC	ExPEC	B1
64	IMT1932	23	23	APEC	ExPEC	B1
65	IMT1939	155	155	APEC	ExPEC	B1
66	IMT2111	38	38	APEC	ExPEC	D
67	IMT2113	101	101	APEC	ExPEC	B1
68	IMT2120	356	23	APEC	ExPEC	B1
69	IMT2121	357	n.a.	APEC	ExPEC	n.a.
70	IMT2283	23	23	APEC	ExPEC	B1
71	IMT2312	10	10	APEC	ExPEC	A
72	IMT2358	915	117	APEC	ExPEC	ABD
73	IMT2470	95	95	APEC	ExPEC	B2
74	IMT2487	69	69	APEC	ExPEC	D
75	IMT2490	117	117	APEC	ExPEC	ABD
76	IMT5112	127	127	APEC	ExPEC	B2
77	IMT5124	369	23	APEC	ExPEC	B1
78	IMT5155	140	95	APEC	ExPEC	B2
79	IMT5214	95	95	APEC	ExPEC	B2
80	IMT5215	93	168	APEC	ExPEC	A
81	IMT8103	10	10	UPEC	ExPEC	A
82	IMT8897	141	n.a.	APEC	ExPEC	B2
83	IMT9087	131	131	UPEC	ExPEC	B2
84	IMT9096	73	73	UPEC	ExPEC	B2
85	IMT9213	88	23	SEPEC	ExPEC	B1
86	IMT9258	73	73	UPEC	ExPEC	B2
87	IMT9286	80	568	UPEC	ExPEC	B2
88	IMT9650	372	372	UPEC	ExPEC	B2
89	IMT9713	372	372	APEC	ExPEC	B2
90	IMT9884	372	372	UPEC	ExPEC	B2

91	IN16/R	131	131	SEPEC	ExPEC	B2
92	IN22/R	131	131	SEPEC	ExPEC	B2
93	IN30/R	131	131	SEPEC	ExPEC	B2
94	IN31/R	131	131	SEPEC	ExPEC	B2
95	IN33/R	131	131	SEPEC	ExPEC	B2
96	IN36/R	131	131	SEPEC	ExPEC	B2
97	IN40/R	131	131	SEPEC	ExPEC	B2
98	IN6/R	131	131	SEPEC	ExPEC	B2
99	MG1655	10	10	Lab Strain	Non-pathogenic	A
100	Ref. Str. O164	270	n.a.	EIEC	InPEC	n.a.
101	RL318/96	17	20	EPEC	InPEC	B1
102	RS168	59	59	NMEC	ExPEC	ABD
103	RS179	62	n.a.	NMEC	ExPEC	ABD
104	RS226	95	95	Feacal	Non-pathogenic	B2
105	RW1374	17	20	EHEC	InPEC	B1
106	RW2297	113	29	STEC	InPEC	B1
107	St5119	141	n.a.	SEPEC	ExPEC	B2
108	TB156A	335	n.a.	EPEC	InPEC	n.a.
109	U3454	95	95	UPEC	ExPEC	B2
110	U4252	48	10	UPEC	ExPEC	A
111	U5070	69	69	UPEC	ExPEC	D
112	UEL31	101	101	APEC	ExPEC	B1
113	Uli 2038	59	59	Human Fec.	Non-pathogenic	B2
114	Uli 2039	405	405	Human Fec.	Non-pathogenic	A
115	Uli 2040	93	168	Human Fec.	Non-pathogenic	A
116	Uli 2041	10	10	Human Fec.	Non-pathogenic	A
117	Uli 2042	1497	n.a.	Human Fec.	Non-pathogenic	A
118	Uli 2043	59	59	Human Fec.	Non-pathogenic	B2
119	Uli 2044	95	95	Human Fec.	Non-pathogenic	B2
120	Uli 2045	73	73	Human Fec.	Non-pathogenic	A
121	Uli 2046	1298	469	Human Fec.	Non-pathogenic	A
122	Uli 2047	1298	469	Human Fec.	Non-pathogenic	A
123	Uli 2048	59	59	Human Fec.	Non-pathogenic	D
124	Uli 2049	636	n.a.	Human Fec.	Non-pathogenic	A
125	Uli 2050	59	59	Human Fec.	Non-pathogenic	D
126	Uli 2051	350	350	Human Fec.	Non-pathogenic	B2
127	Uli 2052	567	n.a.	Human Fec.	Non-pathogenic	B2
128	Uli 2053	93	168	Human Fec.	Non-pathogenic	A
129	UR14/R	131	131	UPEC	ExPEC	B2
130	UR3/R	131	131	UPEC	ExPEC	B2
131	UR40/R	131	131	UPEC	ExPEC	B2
132	W9887	48	10	SEPEC	ExPEC	A

2.4.1 PCR Amplification and Sequencing.

For screening and distribution purposes 100ng of chromosomal DNA were used as the template for the amplification of the target genes. The amplification of the knockout inserts was carried out in two PCR steps: a 50µl reaction with 4ng of plasmid (pKD3 or pKD4) and a second PCR in 100µl using 1µl of the previous reaction. The amplification enzymes used were either the Phusion® DNA Polymerase (Finnzymes) for sequencing, KO amplification or the GoTaq® Green Master Mix (PROMEGA) for the screening and distribution studies. The amplification of the GEI portions to create the complementation vectors was carried out using the *PfuUltra* II Fusion HS (Agilent). Primers were designed in conserved DNA region and the sequences are reported in Table #. The sequencing and KO PCRs were run for 30 cycles of denaturation at 98 °C for 10 s, annealing at 57-60°C for 20 s, and elongation was carried out depending on the length of the amplicon at 72 °C considering the speed of the Taq enzyme to be around 17bp/sec. The distribution PCRs were carried out in 20µl for 30 cycles, denaturation at 94 °C for 30 s, annealing at 57-60 °C for 30 s, and elongation at 72 °C for 1 min 10 s. The complementation vector's inserts has been amplified using the following cycles: 92°C for 2 min, 10 cycles with T_m of 92°C for 10 s T_a of 57°C for 30 s and an elongation at 68°C of between 14 min / 34 min and 24 s, 20 cycles T_m=92°C for 10 s - T_a=57°C for 30 s and elongation at 68°C between 14 min / 34 min and 24 s with 10 s added after the end of each cycle.
PCR products used for the knockout generation and the complementation were treated for 2Hrs with DpnI at 37°C to erase any leftover of template plasmid/genomic DNA. All the products were purified with Wizard® SV Gel and PCR Clean-Up System protocol (PROMEGA) and sent for sequencing at the in-house facility. Sequences were assembled with Geneious 5.6 (Biomatters), aligned and analyzed using its clustalW plugin.

2.4.2 Nuclease resistance of circular intermediates.

To examine whether the circular DNA intermediates were nuclease resistant, we extracted DNA from culture supernatants. To this aim *E. coli* IHE3034 strain was grown until late exponential phase in a 250ml conical flask with 50ml of LB. Next, cells were precipitated at 4000g for 5 minutes at 4°C in a Heraeus MULTIFUGE 3 S-R centrifuge with a 75006445 rotor. The supernatant was then transferred in a 50ml

syringe with a 0,45μm PES filter attached to it. After filtration the supernatant was split in two 25ml aliquots and one was treated with DNAse I (Roche) and RNAse A (PROMEGA) to the final concentration of 25μg/ml for one hour at 37°C while the other Falcon was left in ice. Following enzymes inactivation for 10 minutes at 75°C the prophage particle extraction protocol was started. NaCl and Polyetilene Glycol 8000 were added to the final concentration of 1M and 10%vol respectively to both the samples The supernatants were then gently mixed and left overnight at 4°C to precipitate. Samples were centrifuged at 11000g for 30 minutes at 4°C, the supernatant gently poured away and the tubes left to dry upside down on a sheet of paper. The transparent pellet was gently re-suspended in 400μl of SM Buffer and, after the addition an equal volume of Chloroform, lightly vortexed for 30 seconds. The solution is then left to separate in two phases for 5-10 minutes and the superior acqueos portion is recovered and tested by PCR or frozen a -20°C.

2.4.3 Relative Real-Time PCR.

The frequency of detection of GEI excision was assessed by relative quantification using real-time PCR on IHE3034 raw DNA extractions. The PCR mix was composed of 2μl of raw DNA extract (see chapter 2.1.2.2), 2μl of each primer (final conc. 0,3μM), 12.5μl of FastStart Universal SYBR green master mix (ROX)(Roche), and PCR-grade water to a final volume of 25μl in a LightCyclerII 480 real-time PCR system (Roche). The reaction was initiated by enzyme activation and DNA denaturation at 95°C for 10 min and 40-45 cycles at 95°C for 10 s, annealing at 60°C for 8 s, and extension at 72°C for 14 s. The specificity of the reaction was assessed by melting curve analysis using the LightCycler 4.5 software V1.5.5 and by running the qPCR results in 1,2% agarose gels. The melting curve has been carried out by heating the PCR amplicons to 95°C for 1 s, then by cooling them down at 65°C for 15 s and heated again slowly with a ramping temperature of 0.1°C/s to 99°C under continuous fluorescence monitoring. Each real-time experiment included in the plate a negative control without DNA. The C_q values were extrapolated from the software and the fold-change of the frequency of excision was estimated using the $\Delta\Delta C_q$ method. All the conditions have been tested in three independent experiments and each sample had three technical replicates. The primers used for the real time analysis were from #224 to #240.

2.5 *Primer List.*

Table 4

#	Thesis NAME	Sequence (5'->3')	Bp	Tm
1	3034P01ciF	GTTATGGTCTTTTGTTTGATGTTATTG	27	57°C
2	3034P01ciF	GCTTTCATCTTTGTTTTTGTCTTTATT	27	57°C
3	3034P02ciF	TTTCCGTCATACCTTTCTCTTTCAG	25	57°C
4	3034P02ciF	GTATCAACTCAGACAAAGGCAAAGC	25	57°C
5	3034P03ciF	CGGACTGATTTACCTTTTCTCAATATG	27	57°C
6	3034P03ciF	ATCGTTCAGTATGGTTGTAAATGTGTG	27	57°C
7	3034P04ciF	GTAAAACTGAAACGCAAAAAGAAAGA	26	57°C
8	3034P04ciF	GTAAATCGGCATCTGGCAATAATG	24	57°C
9	3034P05ciF	AAACGATGATGTCAGATATCACAATCTC	28	57°C
10	3034P05ciF	TGTAAGTATCACCATTAATAACAGTGCG	28	57°C
11	3034P06ciF	TTTTTATTGTTTTATCTTGTTGACTTTG	28	57°C
12	3034P06ciF	AAACCCAAAACTCCAAAGGATAATC	25	57°C
13	3034P07ciF	TAAAACCCAGTTCCAACACCAATATC	26	57°C
14	3034P07ciF	GGAAAAGGATGGTTACTTTTTACAG	25	57°C
15	3034P08ciF	ATATCATCGTTTTCAGGTTCTTTTTAC	27	57°C
16	3034P08ciF	GAAGAGGAGCAAGAAGATGAAAACAG	26	57°C
17	3034P09ciF	TTTTAGCGGAGAACAACAACAGATAG	26	57°C
18	3034P09ciF	TATCTTGTTTCGTGTTGTATCCCATCT	27	57°C
19	3034P10ciF	AGTAAATCTTAACCACCGATAAGGAG	26	57°C
20	3034P10ciF	GTCAACCACCAAAAGAAAATACAATAC	27	57°C
21	3034P11ciF	ACTAAACCAAAAGGATAACAAAATGAAA	28	57°C
22	3034P11ciF	TCAAAAGATAGCTGAAGGATTGAAAC	26	57°C
23	3034P12ciF	CGCTATAAAGGTGAATATCGACAATG	26	57°C
24	3034P12ciF	CGAGATACTGAGCATGGTTGTAAATAC	27	57°C
25	3034P13ciF	GCATCACCAGCAGATTTAAGAAAATG	26	57°C
26	3034P13ciF	AAACATGGAGATTAAACAATTCCAGC	26	57°C
27	3034P14ciF	TAAGGACTACACCAACAAAAACAGGAA	27	57°C
28	3034P14ciF	CAATAACAACCTTCACTTTTCCTTCC	26	57°C
29	3034P15ciF	CAAACAAACCAAGACTAACAATGAAATC	28	57°C
30	3034P15ciF	GTACAGACATCAGCATTTCCTTTTCAG	27	57°C
31	3034P17ciF	TGAACATACTGCGATAGTTATCAACCTC	28	57°C
32	3034P17ciF	AACCGATATGAGGGAATATATAAAGCTC	28	57°C
33	3034P18ciF	CTTATTGTTCTGTGCTTTACCTTTTTG	27	57°C
34	3034P18ciF	GTGTATTGTTCTGTTGCTCAGGCTTT	26	57°C
35	3034P19ciF	CTATGCTCTGATACCTCCAAAATGTA	26	57°C
36	3034P19ciF	ATGACCTAGCATTATTTCTGCAATATG	27	57°C
37	3034P01esF	CTACGAAATAGATAACAGTAAACGA	25	57°C
38	3034P01esR	CTAGCACAAGATGACGTAGTGAAC	24	57°C
39	3034P02esF	AACGTTCCTCTTGCGGTAAAGACAC	25	57°C
40	3034P03esF	TGGAATAATGAGCGAAAATATCTTC	25	57°C
41	3034P03esR	AAGACAATATTGAAATGCAAGGTA	24	57°C
42	3034P04esF	ATCGAGTTTGTATTCTTCACCCATTG	26	57°C
43	3034P04esF	AAATCTTCAACGGTAACTTCTTTA	24	57°C
44	3034P04esR	GTTGTGTATGGTAAGAAAACTGGTAA	26	57°C
45	3034P04esR	ACTTCTAACGTTGTGTATGGTAAG	24	57°C
46	3034P05esF	GAATAAAGTTAGTGAAAACACAAAAC	26	57°C
47	3034P05esR	GACTAAAGCATAATCAGCAGAGTC	24	57°C
48	3034P06esF	AAAAATCCACACAGGTTTATGGTCAG	26	57°C
49	3034P06esF	ATTGAAGATGTAGAAAATAATAAACC	26	57°C

50	3034P07esF	TATTAACTACACCACCTTCTTTGATA	26	57°C
51	3034P07esR	TTCTGATGAAATAGTCAAAGGCTCTA	26	57°C
52	3034P08esF	TTTTAACCTGATTATTCATGAAGTC	25	57°C
53	3034P08esF	GTTCATATCTTCGGCGAGCAGAGA	24	57°C
54	3034P09esF	CTAAATGCTTTATTTATGCCTATTT	25	57°C
55	3034P09esR	AGAATATGAGTCTGACCGAAAAGT	24	57°C
56	3034P11esF	AGAGTGGCGTATGGAAAGTCAGAATA	26	57°C
57	3034P11esR	GTAAATACTCCCCTAATTGCGCTAAAAG	28	57°C
58	3034P12esF	GTTCTGTTGAAATCCTCTATCTGGTGTT	28	57°C
59	3034P12esF	GATAGACAAACAGAGGGTTATCAG	24	57°C
60	3034P13esF	GAACAGCAAGGTAAAAATGAAGAGCA	26	57°C
61	3034P13esF	GTAAGTATGACTGGGTTGTCTCTCT	25	57°C
62	3034P14esF	ACGATAAACGTTCAGATATCAAAG	24	57°C
63	3034P14esR	AAAACGCAGAAACGATTTGTACTC	24	57°C
64	3034P15esF	TATAAACATTACTCTGGTTGCCATAC	26	57°C
65	3034P15esR	TTTATCAGTAACGTTGAGGAAGAG	24	57°C
66	3034P17esF	ATCGTACTCATAAACTTCCAGTTC	24	57°C
67	3034P17esR	AGGATAATAAAGTCACAGTACAAAAC	26	57°C
68	3034P18esF	ATCATAAAGACGCCGTACAATC	22	57°C
69	3034P18esR	CAAAGACTATGATTTCAGTATCAGC	25	57°C
70	3034P19esF	TGAAGATTTTCAGGACTATCAGG	23	57°C
71	3034P19esR	GTCAAGTTTGTTCATAAAGGTGAG	24	57°C
72	unGI01aF	TCTATGTGCTGATGGAGGCGCTGT	24	57°C
73	unGI01aR	CCTGATTCGGATTGTGATGGCGGG	24	57°C
74	unGI01bF	TAATAACCACAACTGCCTTG	20	57°C
75	unGI01bR	ATGCCCTACTTTACTTCCAG	20	57°C
76	unGI01cF	AAACTCCCACAAATAACCAG	20	57°C
77	unGI01cR	TAATACAACTCAGCACTCCTTC	22	57°C
78	unGI02aF	TGCCCATCACCATTTATTGT	20	57°C
79	unGI02aF	ATAATACCGGAGCCGAAGTC	20	57°C
80	unGI02bF	ACGATTACCGAAAAGAGAAC	20	57°C
81	unGI02bR	GTGATGCGAGTAACCTTCTA	20	57°C
82	unGI02cF	GAGTTGAAATCGGAAATATG	20	57°C
83	unGI02cR	AAAGGTGGGGTAAGTAAAAC	20	57°C
84	unGI03aF	TATTCAGAGCACAGGGCCAC	20	57°C
85	unGI03aR	AGGCTGTTTCTGGTCGTGTA	20	57°C
86	unGI03bF	GTAACTGGTTGATATTTTCG	20	57°C
87	unGI03bR	TTTTTCTGTGTTGTGGCTAT	20	57°C
88	unGI03cF	GTTAACAATGCAATTAGCCA	20	57°C
89	unGI03cR	GCCTTTCTTCTGTAGCAACT	20	57°C
90	unGI04aF	GCGTAAGGTGGCATCAGGTATGGC	24	57°C
91	unGI04aR	GCCTTGAGCACCATTGCGGTTTTC	24	57°C
92	unGI04bF	GGTTTTACTCAGTTAAGCAG	20	57°C
93	unGI04bR	TGTCTGGATATACGATTCAA	20	57°C
94	unGI04cF	CTCTTTGACTGTTTGGTTGA	20	57°C
95	unGI04cR	TTACGCATCCTGTTTTTATC	20	57°C
96	unGI05aF	ACAGGAATTGTTCTTTCTGACACTA	25	57°C
97	unGI05aR	CTTCGCCAGCGTATCCCACTTCAC	24	57°C
98	unGI05bF	GCTTTGGTGTTTATTACGAG	20	57°C
99	unGI05bR	TAGTATATTTCGGGATGACC	20	57°C
100	unGI05cF	ATGAAATGACAATGAAAAGC	20	57°C
101	unGI05cR	CAGTGAAAAGGAATGTATGG	20	57°C
102	unGI06aF	GGACAGATAGTTTTGGTTTTACTT	24	57°C

103	unGI06aR	AACTTACCACTACCTCTGTTGATT	24	57°C
104	unGI06bF	AAGTGGGAAAGAGTTTAGGA	20	57°C
105	unGI06bR	CATCATCCAAGCATTCATAG	20	57°C
106	unGI06cF	TTGGAGAATGGTAGAACTTG	20	57°C
107	unGI06cR	ATACTCGCATCAACTTTGTC	20	57°C
108	unGI07aF	TTGTCCGTATGAAAGTAGAGAAG	23	57°C
109	unGI07aR	GTCGATTACCTTTTTAGTGCTATG	24	57°C
110	unGI07bF	CCTAACATACTTCGCATCAA	20	57°C
111	unGI07bR	TATTTTCATTAGGTCGCTCA	20	57°C
112	unGI07cF	GCCCTTCTATCTCCAGTTTA	20	57°C
113	unGI07cR	GCGTTACATAAGTTCACTGG	20	57°C
114	unGI08aF	CGGAGGTCACGCAACTGGAAGAAG	24	57°C
115	unGI08aR	GGCATATCAATAACACCACTGTAAA	25	57°C
116	unGI08bF	CGTTAGACATCATCCAGTTC	20	57°C
117	unGI08bR	TCCCTGTTATTTGGTATCAC	20	57°C
118	unGI08cF	AATTACCGAAAACTCCAGAA	20	57°C
119	unGI08cR	TTATATGGGCTGTCTTGGAT	20	57°C
120	unGI09aF	TACAAGTAACGCAGCCGGGTCTCA	24	57°C
121	unGI09aR	AAATCTCTCCTTCCACCCCGACCG	24	57°C
122	unGI09bF	TTTTCCACTTGTATCACTCG	20	57°C
123	unGI09bR	TACTATCGATTTACCGCAGA	20	57°C
124	unGI09cF	GTCGTTGAGTGGAGTGATAG	20	57°C
125	unGI09cR	CTGCACAGAATATTGAACGT	20	57°C
126	unGI10aF	GGGGAAGGCAATGTGGATGACAGC	24	57°C
127	unGI10aR	TGCGGCGTAAATGTCACCGTATCG	24	57°C
128	unGI10bF	TACAATGCTCAGAAAGAACG	20	57°C
129	unGI10bR	TTTTATCGCTATCATTGCTTC	21	57°C
130	unGI10cF	CACCATCCACTATCACCATC	20	57°C
131	unGI10cR	ATCTGGCAATGAACTACCTC	20	57°C
132	unGI11aF	GCCTGATGGGGCAGTTTGGTGACT	24	57°C
133	unGI11aR	GCCTGAACGCGGGACATCTCTT	22	57°C
134	unGI11bF	GCTGGTCAAATCAGGCATCA	20	57°C
135	unGI11bF	ATAATGGTATTGGCGATGTG	20	57°C
136	unGI11cF	TCCAACATTTACTCCATCTG	20	57°C
137	unGI11cR	GATGTAGTAATGGATGTGTGC	21	57°C
138	unGI12aF	ATGATTCTGGCCTTCGATTC	20	57°C
139	unGI12aF	GGCAGAAAACACACCAGAAG	20	57°C
140	unGI12bF	TCTCTTGTGCTGATAACCTC	20	57°C
141	unGI12bR	AAGGTTTTGGATGATGTTTA	20	57°C
142	unGI12cF	ATTTTCCTGTTTGTGTCGTT	20	57°C
143	unGI12cR	ACTGACTACACTGACACGCT	20	57°C
144	unGI13aF	AGACTCGTTGGTCGGGCTGGTTTC	24	57°C
145	unGI13aR	ACCTGTGCCATCTTCCGCATTTCA	24	57°C
146	unGI13bF	GTTTTTCTTCTTTCATTTCGA	21	57°C
147	unGI13bR	AGTCGAATCTCTACCAGTCTCT	22	57°C
148	unGI13cF	CAGAAGACGAAGAGAAAAAGT	21	57°C
149	unGI13cR	CAAGAAAGAAAAAACCATGC	20	57°C
150	unGI14aF	CCGCTGGTATCGTTCATCTCGGTC	24	57°C
151	unGI14aR	GTCGAATACGCTGGTTCCGCTGTT	24	57°C
152	unGI14bF	CCTATGTATGGACTCAGCAA	20	57°C
153	unGI14bR	GCTCTTCCACTCATTTTCAT	20	57°C
154	unGI14cF	ACTGGCCTGTTTATTCATCT	20	57°C
155	unGI14cR	TTGTCCTTCTTCACTAAAACC	21	57°C

156	unGI15aF	CAAGAAAAGATCGCGGCTGGGGAG	24	57°C
157	unGI15aR	GCGTTCTTCGGGGGCAATACCTTC	24	57°C
158	unGI15bF	TCCGGGAATGTTTATTATTG	20	57°C
159	unGI15bR	AAACTGACTCGTGAACTGCT	20	57°C
160	unGI15cF	TAAACAACTCTGGCAACACT	20	57°C
161	unGI15cR	CAAATGACGATGAGGAGATA	20	57°C
162	unGI16aF	GTGTACGTGACCGGATTTGTGCGT	24	57°C
163	unGI16aR	AGGCTGGCTTCTTCTTTGGTGGA	23	57°C
164	unGI17aF	TGAGGTAGCCAATTACACCGGAAGA	25	57°C
165	unGI17aR	CCAGGAGAATCACGCAATCACACT	24	57°C
166	unGI17bF	GAGTCCTCCCTTTGATAATG	20	57°C
167	unGI17bR	CTATCCAGCCCTAAGAACAC	20	57°C
168	unGI17cF	TAATGAAATGGTTGTTGCTAA	21	57°C
169	unGI17cR	AACTTAGATGCCAAAACCTC	20	57°C
170	unGI17F	GGTCTGAAGCGTTTTAAACA	20	57°C
171	unGI17F	CACCAAGATAATGATTGCAC	20	57°C
172	unGI18aF	CCGATGATTTCTGCCGTATTTTGCC	25	57°C
173	unGI18aR	GGTTCACTCTCACTATCTGCCCGT	24	57°C
174	unGI18bF	ATCGCTGTATGTGAAGTTGT	20	57°C
175	unGI18bR	ATCTTGCTCTGTGTGCTAAA	20	57°C
176	unGI18cF	GCTACTATGCTGATTGAACG	20	57°C
177	unGI18cR	CGTCACATCTTTTGGAATAA	20	57°C
178	unGI19aF	ATTTATATTTATGACGAGATTGGTT	25	57°C
179	unGI19aR	AGACTTACATTATCTGTGGAGATTTT	26	57°C
180	unGI19bF	ACTAGCCTATCCACCAAGAG	20	57°C
181	unGI19bR	CAGTGGCAAGTAATGGTAGA	20	57°C
182	unGI19cF	TTAAAAGTCTCCGCTCTACC	20	57°C
183	unGI19cR	CGATACGTTGATGACATTCT	20	57°C
184	KO_GI04F	GGTGATAAAGCGAATACCCGGGCCGTCTACGGTTCCACAGGA TTCAAAGGAGTGAATGCGGTGTAGGCTGGAGCTGCTT	79	57°C
185	KO_GI04R	GCCTTCGAGCTGCGCACCAACACGGCCTCAGATGGGCCACAT CTGGAGAAACACCGCAATCATATGAATATCCTCCTTA	79	57°C
186	KO_GI06F	CGAAATATGCCGGACAGGACAAAGTAAACCCAGGCTCTATTA TTCTCTCCGCTGAGATGAGTGTAGGCTGGAGCTGCTT	79	57°C
187	KO_GI06R	GTCTTCGCGTTGATTGCACCTTCCATACCTTTAACAATCAGG TCTGCGGCTTCAGTCCAGCATATGAATATCCTCCTTA	79	57°C
188	KO_GI07F	TCTTGGGGAGCTGCCGGGCAGGTGGGGGTTGATTATCTGATT AACCGTGACTGGTTGGTTGTGTAGGCTGGAGCTGCTT	79	57°C
189	KO_GI07R	CCCGTAATTACGGGGTCATTTTTGTGCGGAATTAAAAACGAT ATCCTGCTGAGAACATAACATATGAATATCCTCCTTA	79	57°C
190	KO_GI13F	TTAGGATAAAAAAACCCTCTGTAGTAACAGAGGGTTTTGTT CATTCATAGTGCAGGGTCAGTGTAGGCTGGAGCTGCTT	79	57°C
191	KO_GI13R	AATGAAGTGAATGGTATTTCCCGCGTGGTGTATGACATCAGC GGCAAGCCACCAGCAACTCATATGAATATCCTCCTTA	79	57°C
192	KO_GI17F	AGTGGCGAAATCGGTAGACGCAGTTGATTCAAAATCAACCGT AGAAATACGTGCCGGTTCGTGTAGGCTGGAGCTGCTT	79	57°C
193	KO_GI17R	ATATGGGTGATTTCAGACACAAAAAAAGCCGCTCTTGAGCG ACTCGATTTGCATACGGTGCATATGAATATCCTCCTTA	79	57°C
194	KO_GI19F	GCGCGGCGCGATGCCGCTTACTCAAGAAGAAAGAATTATGAC GTTGTCTCCTTATTTGCAGTGTAGGCTGGAGCTGCTT	79	57°C
195	KO_GI19F	GCGCGGCGCGATGCCGCTTACTCAAGAAGAAAGAATTATGAC GTTGTCTCCTTATTTGCAGTGTAGGCTGGAGCTGCTT	79	57°C

196	KO_GI19R	GTCTGAATGGCCTGTCCGAACAGCAGCACCTTCTCGGTGATG GTAGTCTTACCGGCGTCCCATATGAATATCCTCCTTA	79	57°C
197	KO_G13A_F	GCTATTGCCTGATATTTATTTCAGATAATAAATATTCACCCA TAAGGTAACAAAAATCAAGTGTAGGCTGGAGCTGCTT	79	57°C
198	KO_G13B_F	ATCAATGACTCCGTACGCAATTAAATTATTACCAATTTAACC ACATATGATTTATTTATCGTGTAGGCTGGAGCTGCTT	79	57°C
199	KO_G13B_R	GATGATTCAATTTAAAGCAATATTACCCAACAGGTAAATGC ACCCCACAGGTAACTATCCCATATGAATATCCTCCTTA	79	57°C
200	KO_G13C_F	AGCCAGTGGGCTGATTTCTGTTGGGGCAGTGATAAAGATGT GCATACCCACAGACTACTGGTGTAGGCTGGAGCTGCTT	79	57°C
201	KO_G13C_R	GAATGAGCCCTTTGGTTACCTGAAAGGTAATAATTAACGCGT TAAATGTCAACCTTCTACCATATGAATATCCTCCTTA	79	57°C
202	KO_G13D_R	CGTCCACAAAAAAGCCCGCGCTGCGGGCTTCTATTAATGCAG TTTATCTTTGCTTATAACCATATGAATATCCTCCTTA	79	57°C
203	KO_G17A_R	GTAAGCCGGTTTTTCCTGCGACCTTTTCCTGGCTTGCCGGTC TGAGGATGAGTCTCCTGTCATATGAATATCCTCCTTA	79	57°C
204	KO_G17B_F	AGAAAATTGTGACGTACACCGGACAACAACACGACGCATTGC AGATGTGCCAGCCCTGACGTGTAGGCTGGAGCTGCTT	79	57°C
205	KO_G19A_R	ACTGCAATATCTTCAGAAGGCCTGATTAAATGCTGTTTTTCA CTTGTCCACCAGCGTGTTCATATGAATATCCTCCTTA	79	57°C
206	KO_G19B_F	GTGATAGCAACCCGCCACTGAGCGGGTTTTTTGTACCTGTAA ACTTGGTGCAGTACAGTAGTGTAGGCTGGAGCTGCTT	79	57°C
207	KO_G19B_R	CATGTTCTCCACCTGCAAAAAAGCCCCGGATAACCGGGGCAA ATGATGAGTATCGTCCTGCATATGAATATCCTCCTTA	79	57°C
208	KO_G19C_F	ATGACTTCTGACGGCGATTTTGTGGCAGTGGCTACGGTGGCT GTCAGCGCCGCAGGTTAAGTGTAGGCTGGAGCTGCTT	79	57°C
209	KO_GI19R	GTCTGAATGGCCTGTCCGAACAGCAGCACCTTCTCGGTGATG GTAGTCTTACCGGCGTCCCATATGAATATCCTCCTTA	79	57°C
210	KO_intGI04_F	CGAAAGAAAATTGCATTAATTTTCAAGTAGTAGAAGTAAAC AGCGTCATCGGAGGGCTTTCATATGAATATCCTCCTTA	79	57°C
211	KO_intGI04_R	GTTGTATTACAATTAGTTAAATTACTCATATCGCTTCAATTG GCTCTAGTTAACTCTGGTGTGTAGGCTGGAGCTGCTT	79	57°C
212	KO_intGI06_F	CGCTTGCAGAACCGCAACTCCCAATAAACGCAAACCCAAAAC TCCAAAGGATAATCGCTGCATATGAATATCCTCCTTA	79	57°C
213	KO_intGI06_R	GATGATGTAAAATCTTCCCCAAAACTTTCCCCAAAACCCTTC CCCAAAACTGGCTATTTTGTGTAGGCTGGAGCTGCTT	79	57°C
214	KO_intGI07_F	TGCTAAACAAGCCGGAGGTGATCGCCACAGATCACCCTGCTT TGAAGAGGATACTGGAAGCATATGAATATCCTCCTTA	79	57°C
215	KO_intGI07_R	GGATATCGTTTTTAATTTTCTCTGCAAAACCTCTGCAAAACC CCTCTGCAAAACTGGTCAGTGTAGGCTGGAGCTGCTT	79	57°C
216	KO_intGI11_F	CTGATTATTGCTGGAAAGAGGCTGGAGAAATTGTAGAGCGG TTCATTGGAGGGCTTCGTACATATGAATATCCTCCTTA	79	57°C
217	KO_intGI11_R	GGCAATACTGAAAAATTGTTATTCAGTATCGCCCTTTAAGCC GTTTCAGGCTATATAAAAGTGTAGGCTGGAGCTGCTT	79	57°C
218	KO_intGI12_F	TTGTTTCATAGCCTATGAGACACACAAGGCTTTGTGCTCTTC GATAGTTGTTAAGGCGGACATATGAATATCCTCCTTA	79	57°C
219	KO_intGI12_R	GCCGCACTACATGACGGGTAAAAAGTGGATAAAATAATTTT ACCCACCGGATTTTTACCCGTGTAGGCTGGAGCTGCTT	79	57°C
220	KO_intGI13_F	GTATTACCTTAAAGGTATACTCTCATACCGTCATGAAAATGG TTTCTATACGGGTGAATTCATATGAATATCCTCCTTA	79	57°C
221	KO_intGI13_R	AAAATTGAATCACTGGCAGCAGAAGCAGAGATGGAAGAAAA TCAGCAGAACTATTAAACGGTGTAGGCTGGAGCTGCTT	79	57°C

222	KO_intGI17_F	ATTAAAAACTCCAGATTATGTTACTTAACACTCATTTTAAAA ATCAACCAATAACTATTTCATATGAATATCCTCCTTA	79	57°C
223	KO_intGI17_R	CGGATATCATTGGGGGCATAATTGGGGGCATCTTAACTTCGA TTAGAAATGTGCCCCCAAGTGTAGGCTGGAGCTGCTT	79	57°C
224	RTdinBHF	AGCCTATCTCGATGTCACCG	20	60°C
225	RTdinBHR	TTTAGCCGCTGAGACTTTGC	20	60°C
226	rtGEI04F	CGATGTGGTCAATGTGTGGA	20	60°C
227	rtGEI04R	CGCATCCTGGGTCATTCTAA	20	60°C
228	rtGEI06F	GCAAGGCCACACATGTATTG	20	60°C
229	rtGEI06F	CGCTTCAGTCCAACCCATAT	20	60°C
230	rtGEI07F	GAGCGATGATGATTGGCAGT	20	60°C
231	rtGEI07F	CTTAGCCGTGGTATCGATAT	20	60°C
232	rtGEI12F	TGCCGCCTCAAGTAGATGTC	20	60°C
233	rtGEI12F	ATCTGAAGCGAACCATGACG	20	60°C
234	rtGEI13F	AATACCGTCACCTGTACCGC	20	60°C
235	rtGEI13F	CGGAAGAGTCTGCGAAAAAC	20	60°C
236	rtGEI17F	CCCTTCACCCATTATTCACC	20	60°C
237	rtGEI17F	ACTTTCTCGGATGGTCTTGG	20	60°C
238	rtGEI19F	CAAAATGTACAATCAGCGGC	20	60°C
239	rtGEI19F	TCTGCTATGCTCTGATACCTCC	22	60°C
240	rtGEI19R	ATTTGGGCTTAATTATTGGGG	21	60°C
241	AMPscrF	GAGTAAGTAGTTCGCCAGTT	20	60°C
242	AMPscrR	GATCATGTAACTCGCCTTGA	20	60°C
243	CLM3scrF	GCAAGAATGTGAATAAAGGC	20	60°C
244	CLM5scrR	TATGTTTTTCGTCTCAGCCA	20	60°C
245	KAN3scrF	GCAAGGTGAGATGACAGGAG	20	60°C
246	KAN5scrR	GTCATAGCCGAATAGCCTCTC	21	60°C
247	kanG04scrR	AACCTTGCACCACTCAGACC	20	60°C
248	kanG07scrR	TAAATACGCTAAAGCCGGAA	20	60°C
249	kanG17scrR	ATCACACTCACTTCACGTTG	20	60°C
250	KO_G13Ascr_F	TTTGAGCGCAGCCATTGTCT	20	60°C
251	KO_G13Bscr_F	TAGAACTACGACCAGCAGCA	20	60°C
252	KO_G13Bscr_R	GCGTCTGAACACCATTGAAT	20	60°C
253	KO_G13Cscr_F	TTCATGGTTTGCCTCAGATT	20	60°C
254	KO_G13Cscr_R	AATAAAACCGCTCGACTTGC	20	60°C
255	KO_G13Dscr_R	AATGTGGCGTCAATGAGTGT	20	60°C
256	KO_G17Ascr_R	AACAGTGGTGAACAGACGGT	20	60°C
257	KO_G17Bscr_F	CCTTTTCCAGTTGTGCCAGT	20	60°C
258	KO_G19Ascr_R	CGCGCGCAGAATAATAACGT	20	60°C
259	KO_G19Bscr_F	TGTCGGGTTGATGTAGAGCA	20	60°C
260	KO_G19Bscr_R	AACCGGGGCAAATGATGAGT	20	60°C
261	KO_G19Cscr_F	AGTGTCAGCATTGTGGGCAT	20	60°C
262	recAscrF	AGATAGCCACGATAGAGCAG	20	60°C

2.6 *Media and Buffers.*

2.6.1 Media.

2.6.1.1 LB

Tryptone (Difco)	10g
Yeast Extract (Difco)	5g
NaCl	10g
dH_2O	1L
Agarose	20g

pH to 7,6±0,1using NaOH 1M.

Fill up to 1000 ml with MilliQ water and autoclave.

2.6.1.2 SM

Glucose	10g
Proteose Peptone N2 (Difco)	10g
Yeast Extract (Difco)	1g
MgSO4*7H2O	1g
KH_2PO_4	1,9g
K_2HPO_4	0,6g
dH_2O	1L

Elute the Glucose in 100ml dH2O sterile filter it and add it to the SM media after it has cooled down. If SM-Agar is needed add 20g of Agar before autoclaving. [28]

2.6.1.3 HL5

Glucose	15,4g
Proteose Peptone N2 (Difco)	14,3g
Yeast Extract (Difco)	7,15g
Na_2HPO_4*2H2O	1,28g
KH_2PO_4	0,49g
dH_2O	1L

pH to 7,6±0,1using KOH 1N.

Elute the Glucose in 100ml dH_2O sterile filter it and add it to the media after it has cooled down. [70]

2.6.1.4 SoC

Glucose	0,36g
Tryptone	2g
Yeast Extract	0,5g
NaCl	0,05g
1M KCl	0,25ml
1M $MgCl_2$	1ml
1M $MgSO_4$	1ml
dH_2O	100ml

pH to 7,6±0,1using KOH 1N.

2.6.2 Buffers.

2.6.2.1 50x Soerensen buffer

KH_2PO_4	99,86g
Na_2HPO_4	17,80g

Fill up to 1000 ml with MilliQ water and autoclave. [70] Use 1x for experiments

2.6.2.2 Arabinose 20%

L-Arabinose	20g

Fill up to 100 ml with MilliQ water and sterile filter using a 22µm (PES) sterycup.

2.6.2.3 50x TBE buffer

The 50x TBE buffer has been bought from Quiagen and used 1x to prepare Agarose gels and running buffers.

2.6.2.4 SM Buffer (Phage extraction)

NaCl	5,8g
$MgSO_4$ x $7H_2O$	2g
Tris-Cl (1M, pH7,5)	50ml

Fill up to 1000 ml with MilliQ water.

3 Results

3.1.1 ExPEC isolates have a greater number of GEIs than InPEC and non-pathogenic strains.

The distribution analysis allowed to stratify the data to show GEI prevalence in the major pathotypes of *E. coli* (Fig. 8). This representation showed that IHE3034 associated islands are generally more represented in ExPEC isolates than in non-pathogenic ones and almost not present in InPEC isolates. Of the 18 islands identified 10 are completely absent in InPEC (GEI 2, 3, 4, 8, 11, 12, 13, 14, 18, 19). In the 32 intestinal strains studied the genomic island prevalence ranges from 0,8% of GEI 5, 10, 17 to the 3,3% of GEI 7.

The non-pathogenic strains, taken all together, carry almost all the islands of the study but the GEI prevalence is generally lower ranging from 0,8% for GEI 2, 4, 11 and 17 to the 4,6% for GEI 6 and 9. All of the 29 non-pathogenic strains completely lack GEI 19 that is present only in ExPEC isolates.

The extra-intestinal strains are positive for the GEIs under study; the prevalence ranges from 0,8% for very rare genomic islands such as GEI 17 and19 to 24% of GEI 10. This group of strains is the only one that contains GEI19 and also has a high count (>10%) of GEIs that carry known virulence/fitness factors (GEI 1, 5, 15, 18). It is of note that four of these high prevalence islands (GEI 6, 7, 11 and 14) are still able to excide from the genome.

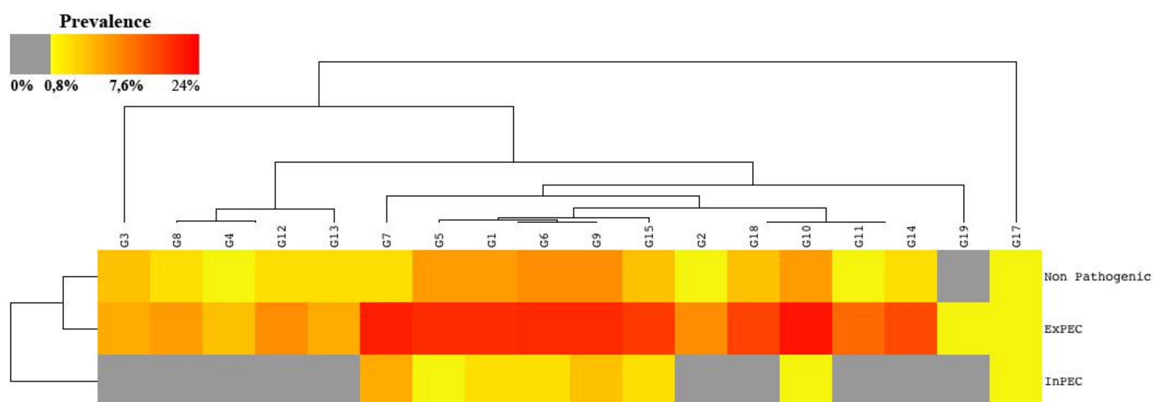


Fig. 8: Genomic islands distribution in *E. coli* major pathotypes.

GEI distribution study, on a collection of 132 strains, grouped by their major pathotype. Strains have been sorted using the average-linkage algorithm of cluster on both the strains and the islands.

3.2 *Distribution of IHE3034 islands among a panel of diverse E. coli isolates.*

The prevalence of the 18 genomic islands (GEIs) present in IHE3034 was assessed in 132 strains representative of the *E. coli* diversity (found in our strain collection) and selected depending on the different pathotypes, phylogenetic groups and MLST types. The panel has been assessed by multiplex PCR with three target regions as probes for each island. These regions have been chosen following two selection criteria: 1) being present and conserved among pathogenic *E. coli* isolates; and 2) being absent in non-pathogenic strains such as K12 MG1655. Primers corresponding to three of these regions were designed, for each bioinformatically predicted GEI, in order to cover the whole island using IHE3034 as reference strain. In the distribution analysis, GEIs have been considered present only if all three PCRs reactions were positive. As expected, testing primers on our reference strain yielded exclusively positive results while no product could be detected in reactions with non-pathogenic K12 MG1655 genomic DNA (Fig. 9). Data were sorted using the average-linkage algorithm of Cluster and the obtained results were used to create a graphical representation with Treeview. The GEI content among the panel was variable; with some groups of island being present in the majority of isolates while others were restricted to strains or groups of strains with a particular phylogenetic group or sequence type. The obtained dendogram demonstrates the presence of two major clusters. In the first cluster, associated with GEI's presence, included the majority of the ExPEC isolates (60/79 75,9%, Fig 9A) while in the second group, due to their lack of genomic islands, InPEC and non-pathogenic isolates were predominant (42/53 79,2%, Fig. 7B). The latter can be further divided into six subgroups (1-6) depending on the GEI content of the studied strains (Fig. 9A).

There are three ExPEC subgroups that may be considered outliers (1_A, 1_B and 2), as they are very divergent from the GEI pattern of the reference strain IHE3034. Group 1_A is composed of six strains with only 1 GEI; five isolates are ExPEC and one non-pathogenic. Of the extra-intestinal bacteria four belong to the phylogenetic groups A, ABD and AxB1. The second cluster 1_B contained the sequence type complex 131 strains that were all positive for the presence GEI 10; strain IN40/R exhibits one of the most rare islands which is particularly related to STC95 strains (GEI04, Fisher $p<0,01$) whereas strain UR3/R contains GEI 3 and 6. It is of note that E457, a non-

pathogenic strain of the same STC, displays the same GEI pattern as UR3/R. The last outlier group 2 is composed of 2 UPEC (B2), 1 APEC, 1 EPEC (both B1) and 2 non-pathogenic strains (B2 and A) with 1 to 3 genomic islands each. All strains except the non-pathogenic Uli2049 (phyl. group A) carry the virulence/fitness factor GEI 5. Strain APEC B1 IMT2120 carries GEI 7 but lacks any other island that is related to this outlier; IMT9096 only harbours island 5 but any other island. In the second sub-cluster of group B all strains carry GEI 1, known to encode for a type VI secretion system related to virulence in *E. coli* and *P. aeruginosa*[51]; this is very interesting as both of the non-pathogenic strains have this island. GEI 8 is a virulence factor island related to the B2 group (Fisher $p<0,01$) and is present in both the UPEC (B2) isolates and, surprisingly, the B2 non-pathogenic strain.

The dendrogram created by cluster analysis shows that the remaining 4 groups (3-6) are linked together and are increasingly related to the reference strain, residing in group 6, due their prevalence of genomic islands. Thus it is not surprising that the majority of APEC (B2), UPEC and SEPEC strains are clustering here.

Cluster 3 exhibits an isolate composition similar to group 2 as it shows 3 APEC, 2 UPEC and 1 ETEC with 1 to 3 GEIs for each isolate. These strains all have in common GEI 7 but two ExPEC B2 strains, IMT2121 and IMT9650, are also positive for the presence of GEI 18 and GEI 13 or GEI 1, respectively; these islands except for GEI13, are highly associated to the B2 phylotype and to the ExPEC pathotype (Fisher $p<0,01$). Moreover, the APEC strain IMT1939 carries the colibactin island GEI 9[36, 44, 61].

The biggest group in this analysis, with 22 strains, is cluster 4. Almost all of the phylogenetic group B2 strains, 16 out of 22 (72,7%), are ExPEC, while the remaining six are either A, B1 (InPEC/non-pathogenic) or non-classified; notably one B1 strains is APEC. All the strains in this group exhibit GEI 9, but this analysis allowed for the identification of three other smaller sub-groups that have different combinations of GEI 1, 5, 6, 10 (sub-cluster 4_A), GEI 7, 9, 10 (4_B) or GEI 5 and 18 (4_C). The latter group is of special note as GEI 5 and 18 are known to carry genes encoding for very important virulence and fitness factors of the bacteria such as S-prefimbriae, IroN, Antigen 43 and the *ibe* cluster [53].

Cluster 5 contains thirteen strains, almost all of them belonging to APEC/UPEC (9/13 – 69,2%), and all harbour GEI 6, 7 and 15; although it is mainly composed of A, B1 and

ABD strains (10/13 - 76,9%) it is the nearest group to the STC95 cluster due to its high GEI content. Within this group of strains there are three identifiable subgroups that have GEI 1, 5 or 11, respectively.

Almost all of the STC95 strains are gathered in a single group (Cluster 6) together with IHE3034. Nevertheless, three sub-clusters (6_A, 6_B, 6_C) could be identified that differ in the number of genomic islands shared between them. The 6_A sub-group is separated from its sister groups 6_B and 6_C that are linked together. These isolates share GEI 1, 6, 14, 15 and 18; strains IMT2470 and IMT2487 (APEC) also have GEI 9, 10 or 12, respectively. The 6_B cluster is composed of four strains, three of which are ST390 (data not shown), that have a mosaic composition of GEIs and completely lack GEI 4, 8, 11, 13, 17 and 19. Strain B13155 and RS226 miss GEI 6 and 7, respectively, while strain IHE3080 contains GEI 2 but does not possess GEI 3. The IHE3034 sub-group 6_C is composed of four strains that show a very similar, but never identical, GEI profile in respect to the reference strain. All of the bacteria forming this sub-cluster have in common GEI 1, 2, 5, 6, 7, 8, 10, 14, 15 and 18. Although IHE3036 harbours GEI 11 but lacks GEI 3, 4, 12, 13, 17, 19. The two strains nearest to IHE3034 IMT5214 (APEC) carry GEI 13 but miss GEI 3, 11, 17 and 19 whereas Uli2044 (Healthy Faecal, B2) lacks GEI 11, 19 in respect to the reference strain.

The distribution analysis has identified four GEIs (4, 13, 17, 19) that have a low prevalence and are almost exclusive for STC95 and ExPEC strains, the islands cluster together in the tree generated with Cluster. These islands were observed for the first time in the paper from Moriel and colleagues [53]. The percentage of unknown ORFs in these islands ranges from 18,2% (8 of 44 ORFs) of GEI 4 to the 56,2% (27 out of 48 ORFs) of GEI 13; in each island there is a variable number of genes of phagic origin ranging from 3 of GEI 17 to 19 of GEI 13. All the islands are present only in ExPEC strains except for GEI 13 that is also present in an EPEC strain; it is of note that this GEI is also the only one carrying a known virulence factor, enterohemolysin 1. GEI 19 with its 48Kbp and is the longest mobile island in IHE3034; the distribution analysis pointed out that this island is the only one that is present only in in our reference strain.

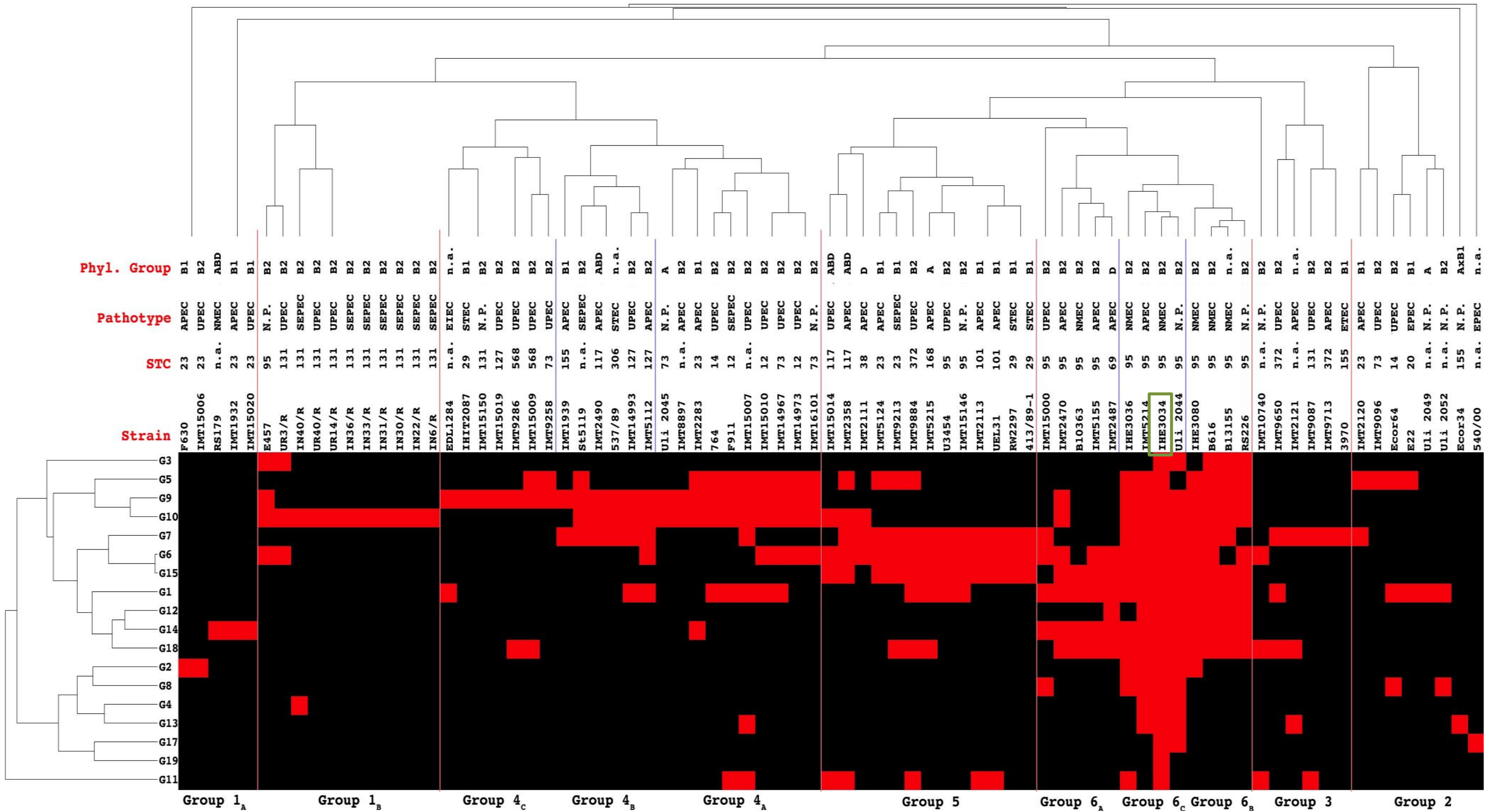


Fig. 9A: Prevalence of genomic islands present in NMEC IHE3034 among *E. coli* isolates belonging to different phylogenetic groups.

GEI distribution study on a collection of 132 strains *E. coli* MG1655 and IHE3034, respectively, were taken as negative (no GEI present) and positive controls. GEIs considered as present are shown in red while GEIs absent are shown in black. The used strains show: name of the strain, STC, pathotype and phylogenetic group; "N.P." strands for Non-Pathogenic while "n.a." for not available. The strains have been sorted using the average-linkage algorithm of cluster on both the strains and the islands. In the green box the reference strain MG1655.

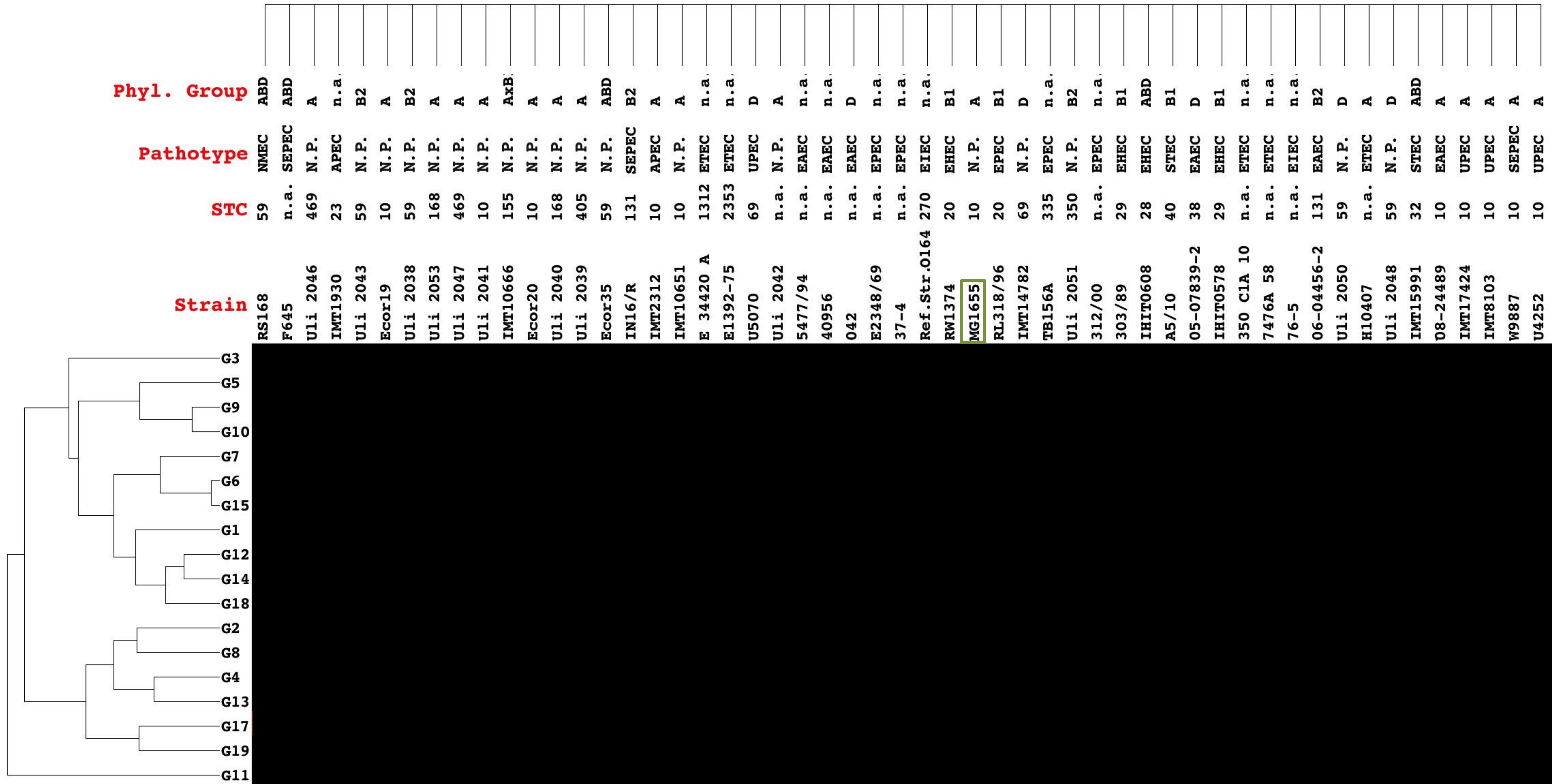


Fig. 9B: Prevalence of genomic islands present in NMEC IHE3034 among *E. coli* isolates belonging to different phylogenetic groups.

Portion of the graph displaying all the negative strains that cluster together. The used strains show: name of the strain, STC, pathotype and phylogenetic group; the strains have been sorted using the average-linkage algorithm of cluster on both the strains and the islands. The used strains show: name of the strain, STC, pathotype and phylogenetic group; "N.P." strands for Non-Pathogenic while "n.a." for not available. The strains have been sorted using the average-linkage algorithm of cluster on both the strains and the islands. In the green box the reference strain MG1655.

3.3 *Molecular and genetic characterization of IHE3034 genomic islands.*

The acquisition of genomic islands by horizontal gene transfer (HGT) is an effective mechanism of generating diversity between bacterial species [44]. The study of the plasticity of the genome is important, as it allows a better understanding of how bacteria evolve and acquire new genes contributing to virulence.,

Nine IHE3034 genomic islands were able to excise from the genome and form circular intermediates.

The ability of genomic islands to excide from the genome has been shown to be linked to the mechanism used by bacteriophages [50]; they have been found to mobilize themselves, at least transiently, by forming circular intermediates (CI) which are then transferred from a donor bacteria to a recipient bacteria. In order to understand which of these GEIs are still viable as possible HGT vectors the circular intermediates were identified using an IHE3034-specific PCR assay that entails 2 rounds of screening. The first steps of the analysis were carried out by testing for the presence of circular intermediates using a set of inverted primers that would anneal on the newly formed circular DNA structure. Results indicate that GEI 4, 6, 7, 11, 12, 13, 14, 17 and 19 are still able to excide from the genome and form circular intermediates; exclusion PCR data confirmed the results obtained by the inverted PCR (Fig. 10).

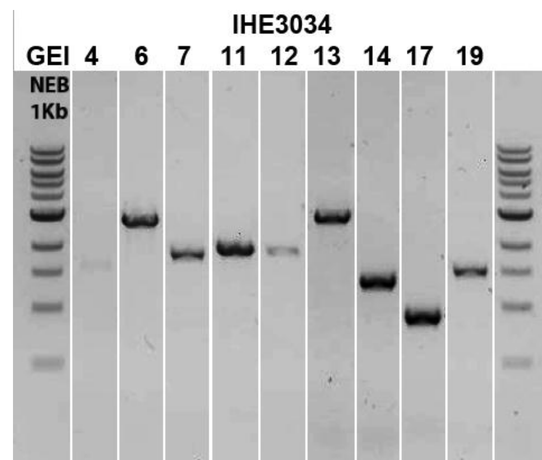


Fig. 10: Detection of Circular intermediates (CI).

Detection of CI PCRs in a SYBRsafe stained 1% agarose gels.

In order to confirm the specificity of the PCR reaction and to characterize the *att* sites of the circularizing GEIs, the inversed and exclusion PCRs were sequenced and analysed by a multiple alignment. This approach allowed the identification of the *att* sites of every genomic island (Tab. 5). The *att* sequences in the reference strain vary between 16 and 51 base pairs in length and their nucleotide identity (left site against right site) ranges from 81% (GEI 19) to 100% (GEI 11 and 13).

Table 5

GEI	*att-L / R*	Identity
4	aatg**cg/tc**accaataactgac	89,5%
6	tgctgcgccatatggg**t/c**tggactgaagc	96,4%
7	aac**a/c**t**gt/ac**cagtg**g/c**tggtacatggatat**c**gataccac	85,7%
11	ttttcatcaacaaggatttt	100%
12	ctgcag**/-**gggacaccatt	94,1%
13	aatcattcccactcaat	100%
14	cgggttcaactcccgccag**ctcca/-----**ccaatcatgattggacggtgtaaggac	90,2%
17	gagtccggccttc**g/-**gcacca	95%
19	agaggt**g/a**gcgaa**gc/aa**g**cc/aa**g**c/a**ac**t/c**tttgc**c/a**attatttctcaccc	81%

ATT sites were identified by aligning (MEGA5, clustalW algorithm) the inverted and the exclusion PCR sequences then by finding the proposed *att* sites on the reference genome sequence of IHE3034. Bases in indicate variations between the two *att* sites.

3.3.1 Deletion of genomic island (GEI) Integrases prevents the formation of circular intermediates.

To study if the excision of the genomic islands is mediated by either the *recA* allele [50] or by the integrase gene that is present in each island a Δ*recA* mutant of IHE3034 has been generated. The circular intermediates analysis carried out on the mutant strain showed that none of the 9 genomic islands was influenced by the absence of *recA*. (data not shown)

To confirm that the excision of the genomic islands is dependant on the integrase, nine *int* deletion mutants of IHE3034 were constructed. The gene was deleted from each GEI that was able to excise from the genome; the ability of the genomic island to create CIs was then assessed by PCR. The results indicate that all the islands except GEI 6 and GEI 7 are completely dependant on their *int* gene to excide from the genome and to create a circular intermediate. This might be explained by a crosstalk mechanism between the two islands as the deletion of either GEI's integrase gene still enables the formation of CIs (Fig. 11).

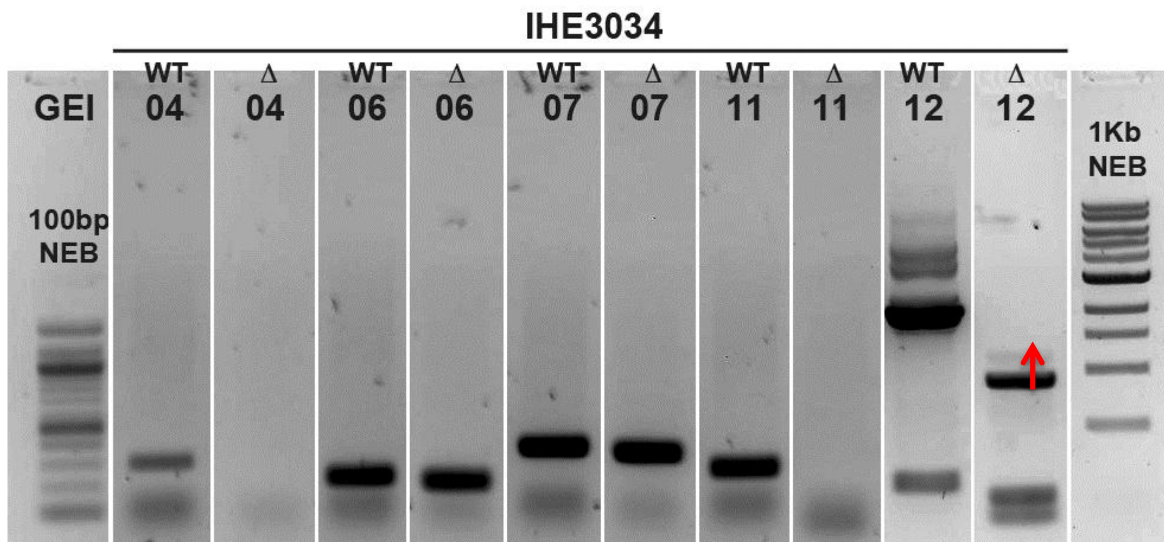


Fig. 11: Deletion of the Integrase gene prevents the formation of CIs.

Detection of CI PCRs in a SYBRsafe stained 1% agarose gels. Circular intermediate production was assessed before and after the deletion of the *int* gene. (A) PCR done using real time primers. (B) PCR of GEI 12 using longer identification primers.

3.3.2 GEI 11, 13 and 17 circular intermediates are resistant to DNAse treatment.

The bioinformatic analysis conducted on the chromosome of IHE3034 allowed Moriel and colleagues to identify many genomic islands of prophagic origin [53]. To further characterize these GEIs the supernatant of an IHE3034 culture was tested for traces of nuclease resistant circular intermediates (nrCI). The presence of nrCIs could indicate the possible production of prophagic molecules into which the circular intermediates are packed, thus shedding light on which genomic islands may be HGT vectors [7, 48, 73].

In the untreated sample all the CIs of genomic islands previously identified as mobile were present (Fig. 12). By contrast, the DNase treated samples yielded three bands of the expected heights for GEI 11, 13 and 17.

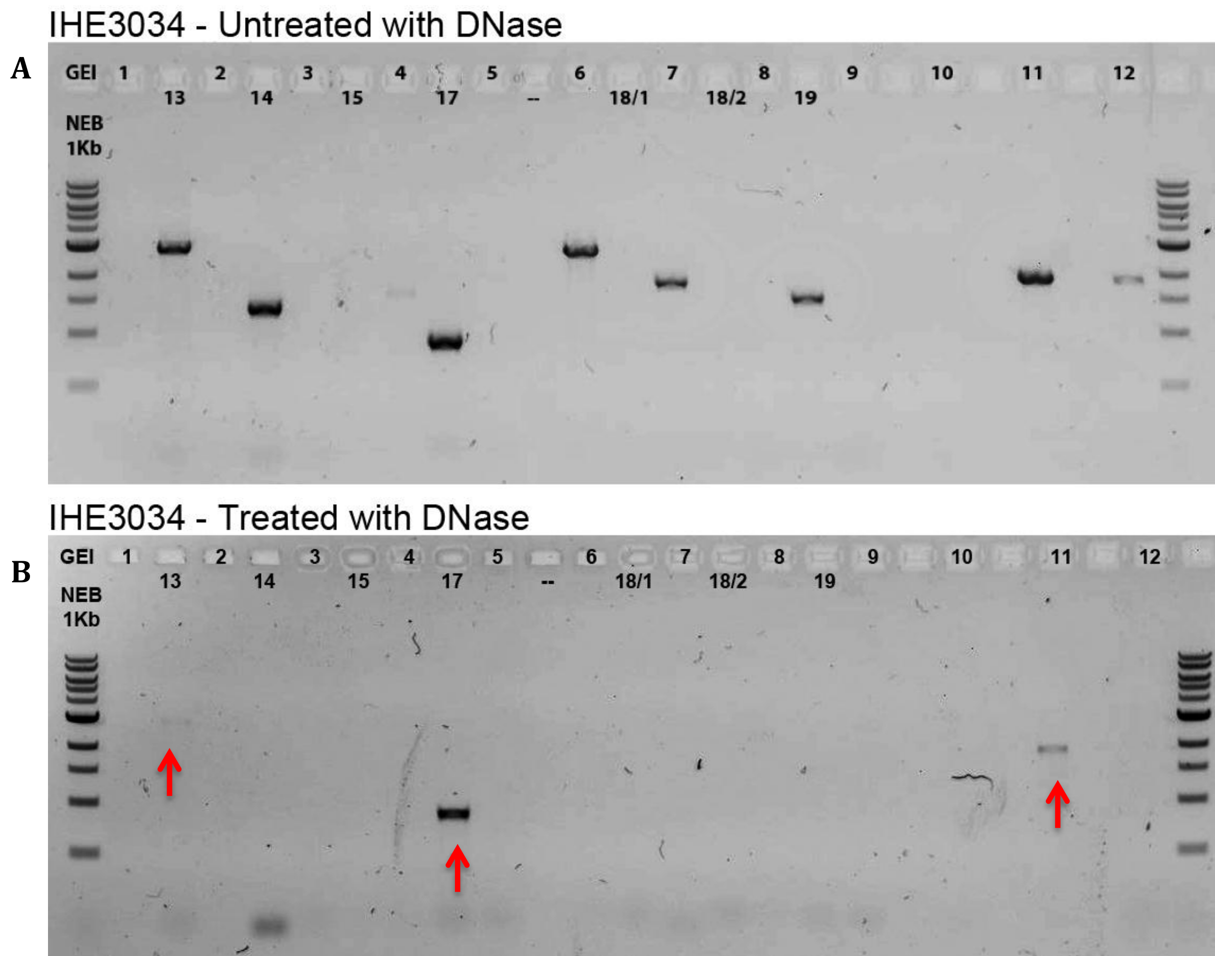


Fig. 12: GEI 11, 13 and 17 circular intermediates are resistant to DNAse treatment.

Detection of CI PCRs in a SYBRsafe stained 1% agarose gels. Circular intermediate production was assessed in both the supernatant samples. (A) supernatant late log growth not treated with DNase. (B) supernatant late log growth treated with DNase.

3.3.3 Growth conditions alter the excision rate of IHE3034 genomic islands.

The HGT mechanisms by which these islands are able to move within and among the genome is still an unresolved question .

The *E. coli* reference strain had been tested under different stress conditions in order to study excision rates variations among the GEIs using a relative qPCR assay. Results provide evidence that all of the stress conditions (temperature, minimal medium, iron depletion and oxidative stress) used for these experiments do not modify GEI 6 and GEI 7 excision rates as the CI production never raised more 1,9 times or reduces itself of more than 1,8 times except when the cells were grown in an iron depleted media such as 2'2'-Dipyridyl (Fig. 13A to F).

Our experimental methods indicate that temperature significantly reduces GEI 4, 13 and 17 (6,2x, 8,5x and 32,3x, $p<0,05$), while the use of sub-lethal quantities of

Mitomycin C, an antibiotic known for its ability to stimulate prophage production, does not cause any changes in the excision rates of the islands, except GEI13, where it increases the CI production by 30,3 times ($p<0,05$) (Fig. 13A/B).
Iron depletion by 2'2'-Dipyridyl greatly affects GEI 4 as it lowers the CI production by more than 30,3 times thus posing the question which kind of ORF may be found to work on this island (Fig. 13C). Additionally, as previously mentioned, this environmental condition increases GEI 06 excision rates (2,3 times) and the relative expression of the *sitABCD* operon carried in this island by 47,3 to 114,6 times (data not shown).
Growth in chemically defined media (RPMI) caused a reduction of circular intermediates by 2,8 times ($p<0,01$) and 4,1 times ($p<0,01$) for GEI 13 and 17, respectively (Fig. 13D).
Reducing compounds that generate free radicals such as H_2O_2 and Menadione had an effect only on GEI 17 by rising its CI production of 3,8 times ($p<0,01$). Genomic island 13 produces 4,4 times more circular intermediates when treated with H_2O_2 ($p<0,01$) (Fig. 13E/F).

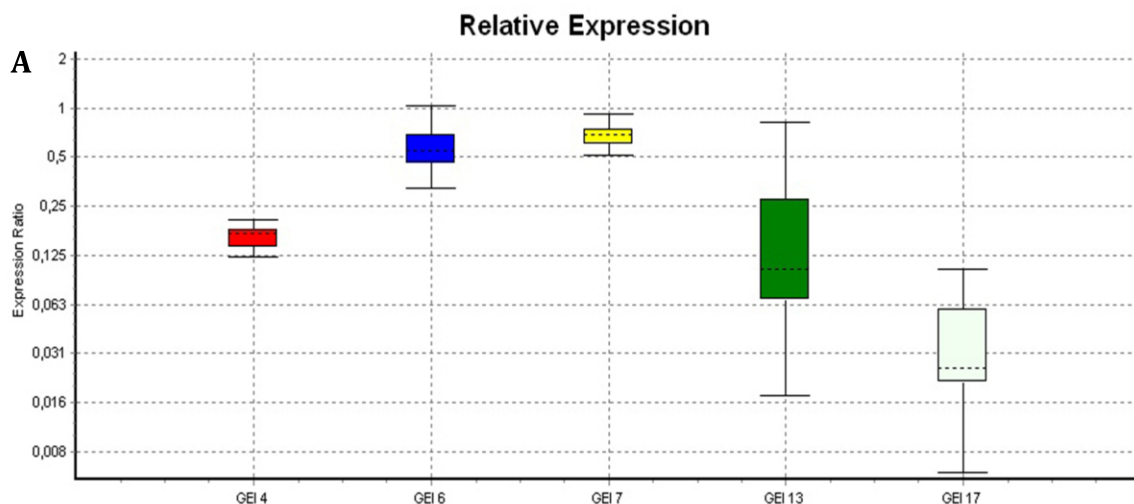
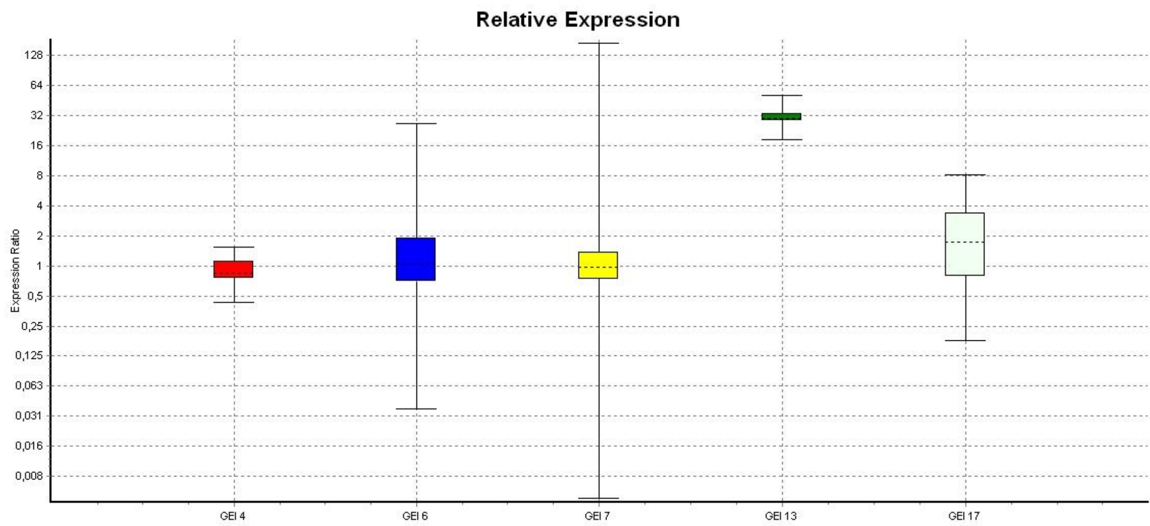
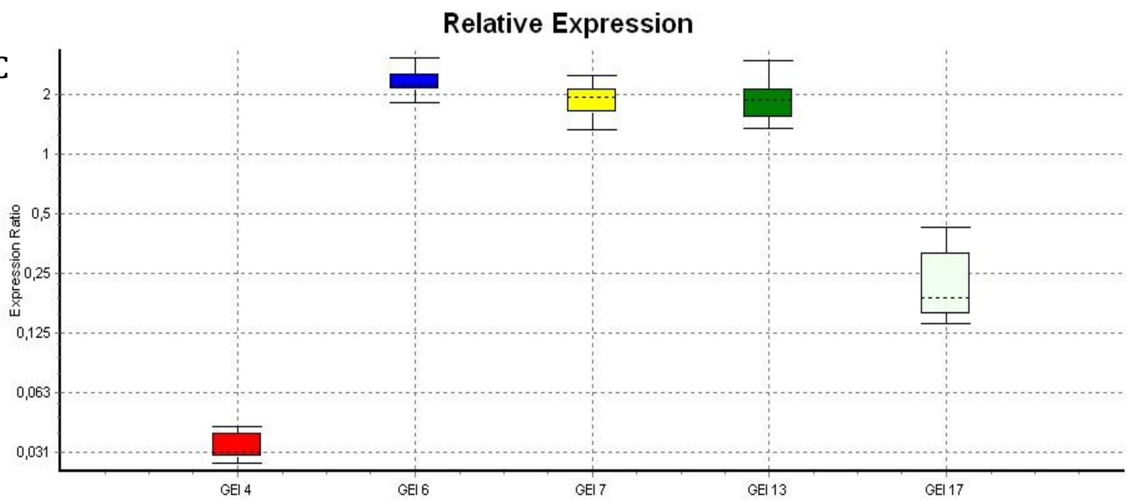


Fig. 13: Growth conditions alter IHE3034 genomic islands excision rates.
Relative circular intermediates production variation in bacterial cells grown in different conditions. The whisker-box plot encompasses 50% of all observations, the dotted line represents the median and the whiskers represent the outer 50% of observations. CI production variations above 2 times are considered as statistically significant. The conditions used were (**A**) growth temperature set to 20°C.

B



C



D

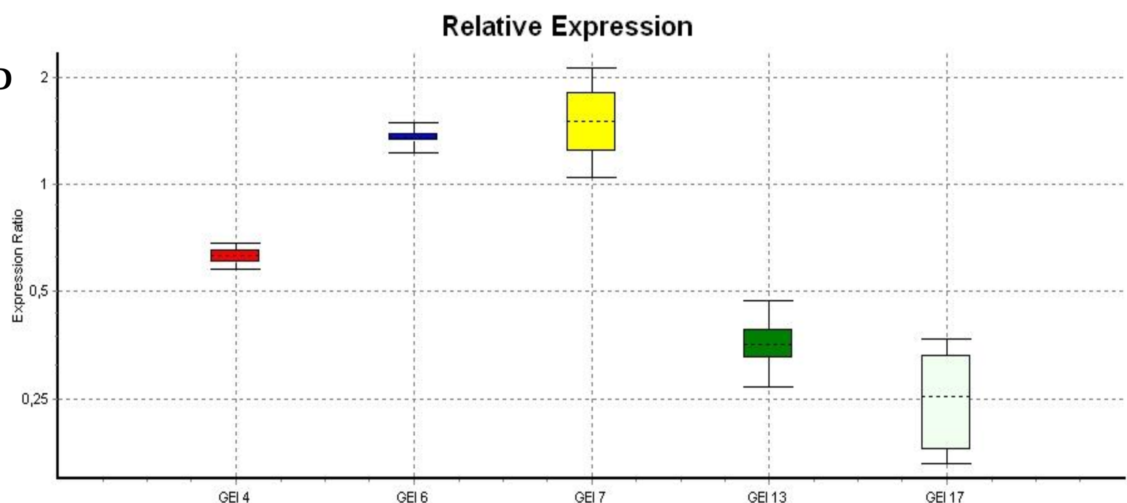


Fig. 13: Growth conditions alter IHE3034 genomic islands excision rates.

Relative circular intermediates production variation in bacterial cells grown in different conditions. The whisker-box plot encompasses 50% of all observations, the dotted line represents the sample median and the whiskers represent the outer 50% of observations. CI production variations above 2 times are considered as statistically significant. The conditions used were (**B**) 0,5 µg/ml mitomycin C, (**C**) 0,25 µM 2'2' Dipyridyl, (**D**) growth in a minimal media, (RPMI).

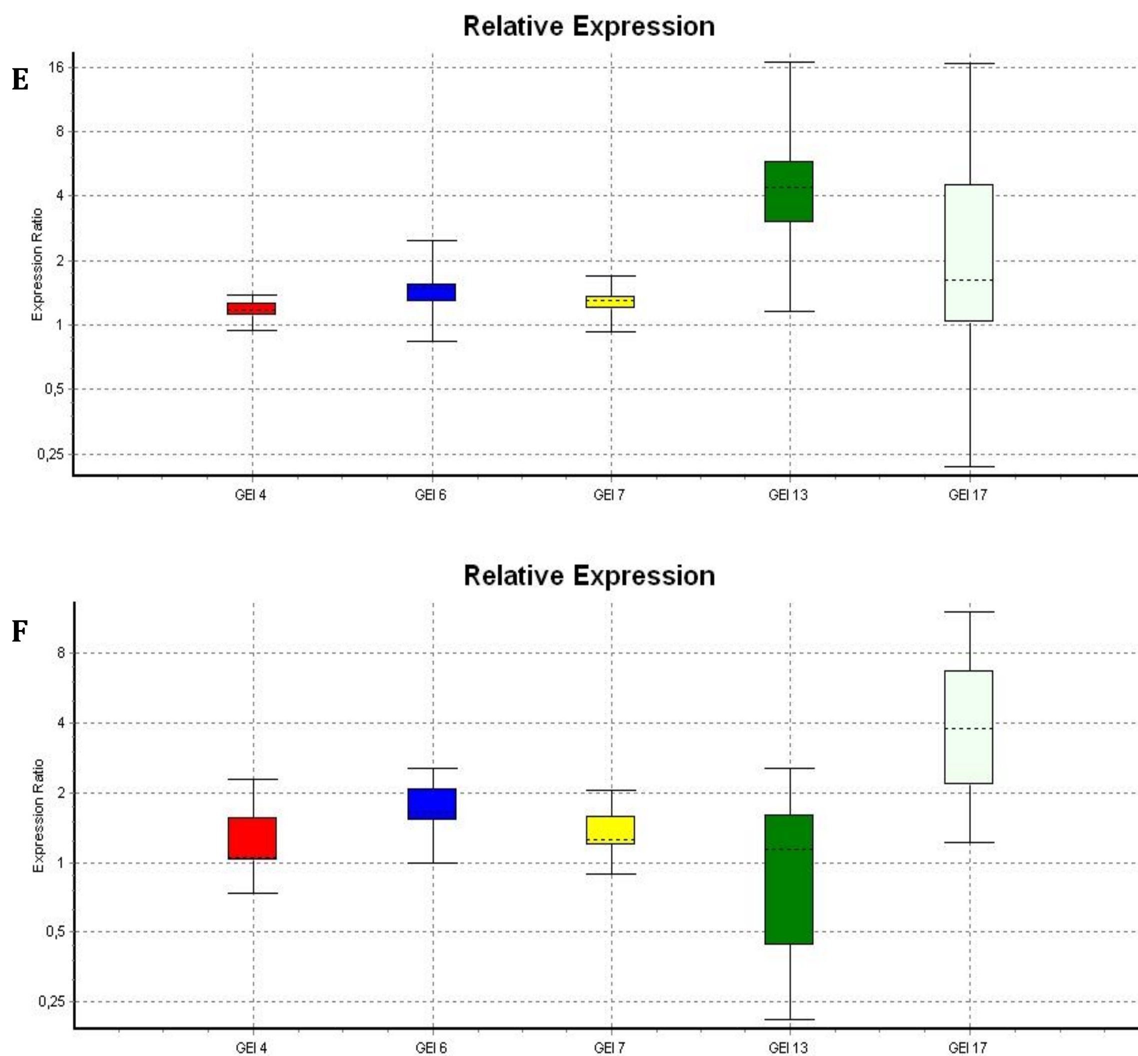


Fig. 13: Growth conditions alter IHE3034 genomic islands excision rates.

Relative circular intermediates production variation in bacterial cells grown in different conditions. The whisker-box plot encompasses 50% of all observations, the dotted line represents the sample median and the whiskers represent the outer 50% of observations. CI production variations above 2 times are considered as statistically significant. The conditions used were (**E**) 60 μM, hydrogen peroxide. (**F**) 75 μM menadione.

3.4 *IHE3034 genomic islands 13, 17 and 19 are linked to survival in the Dyctiostelium discoideum grazing assay.*

Dictyostelium discoideum, is a haploid amoeba that has been extensively used as a model to study host-factors involved in cellular aspects of host-pathogen interactions [2]. *D. discoideum* has been extensively used to define the maintenance and the evolution of genes associated with virulence and fitness in bacteria such as *Legionella pneumophila, Aeromonas salmonicida, Klebsiella pneumonia* and *E. coli*[10, 27, 70]. In this work the functional effects of the genomic islands, almost exclusively associated to the STC 95, are analysed in order to understand their possible contribution to pathogenicity.

3.4.1 GEI13, 17, 19 deletion affect IHE3034 ability to resist to the *Dyctiostelium discoideum* grazing assay.

In this work IHE3034 wild type and its genomic islands mutants have been tested for their resistance to *D. discoideum* grazing in order to assess the possible functions of the deletes GEIs.

In order to assess if the lack of the genomic islands affects the ability to grow of the bacteria, the knock-out strains (IHE3034 ΔG4-13-17-19 / ΔG4/ ΔG13/ ΔG17/ ΔG19) have been tested in a growth curve. The results showed that in both LB and SM the growth rates of all the strains were comparable to that of the wild type strain (Fig. 14).

We tested different bacteria / amoebae cells ratio by measuring grazing capacity at different amoebae population sizes. Over the course of a few days, the bacteria formed lawns on these plates with amoeba embedded in them. The bacteria phagocytized by *D. discoideum* were assessed through the occurrence of bacterial lysis plaques.

The wild type strain showed a resistance to grazing only slightly inferior to the negative control (str. 536). The strain missing island 4 showed a reduction in its ability to resist the amoeba grazing comparable to the wild type thus excluding it as a possible carrier of phagocitation resistance genes (Fig. 15). The ΔG13 strain showed a marked resistance reduction as in the 1000 amoeba area big plaques were readily visible. IHE3034 ΔG17 demonstrated to be the most susceptible strain of all the single GEI deletion mutants as the plaque created by *Dictyostelium* were very big and visible. The lack of GEI 19 caused an impaired resistance to the amoeba phagocytosis.

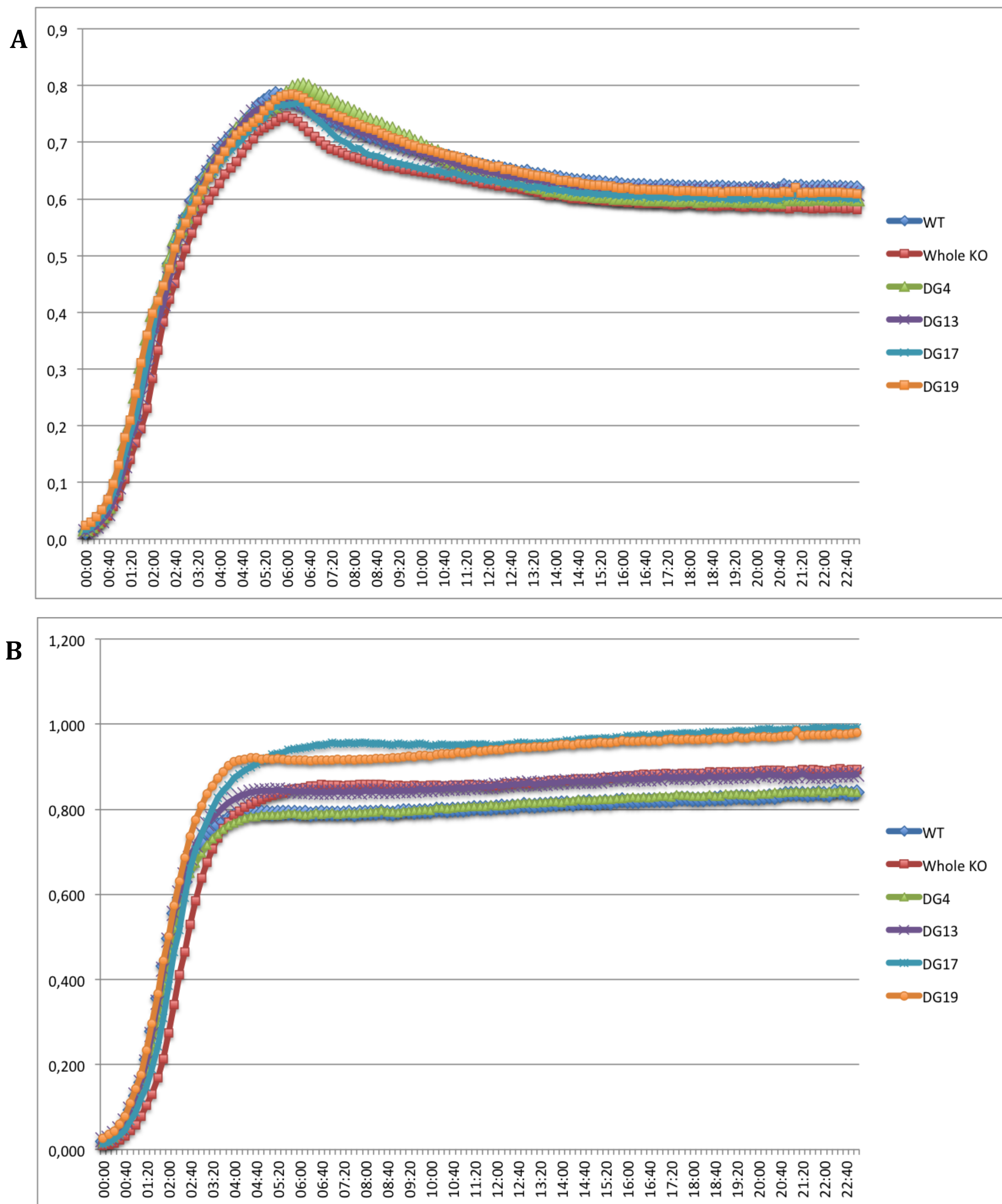


Fig. 14: Single and multiple genomic islands deletion do not alter IHE3034 growth rates. Growth curves of the 6 KO strains generated for the *Dictyostelium discoideum* grazing assay. The strains have been grown for almost 23 hours in a 96 well plate inside a Tecan N200 infinity. The plate was left at 37°C shaking and the OD_{600} has been measured every 10 mintes. (**A**) Growth curve done in LB. (**B**) growth curve done in SM.

Even if reduced, the effect caused by the lack of this island, seemed to be similar to the one of IHE3034 ΔG17. The strain missing all of the genomic islands (IHE3034 ΔG4-13-17-19) was the most susceptible to *D. discoideum* when tested. As you can see in Figure 15, the large plaques resembled the DH5α strain, and in some cases it was

possible to witness spore formation (Fig. 15). It is important to note that no bacterial strain (except 536) was completely resistant in this test (even when 5000 amoebas were used so this dilution was not considered significant).

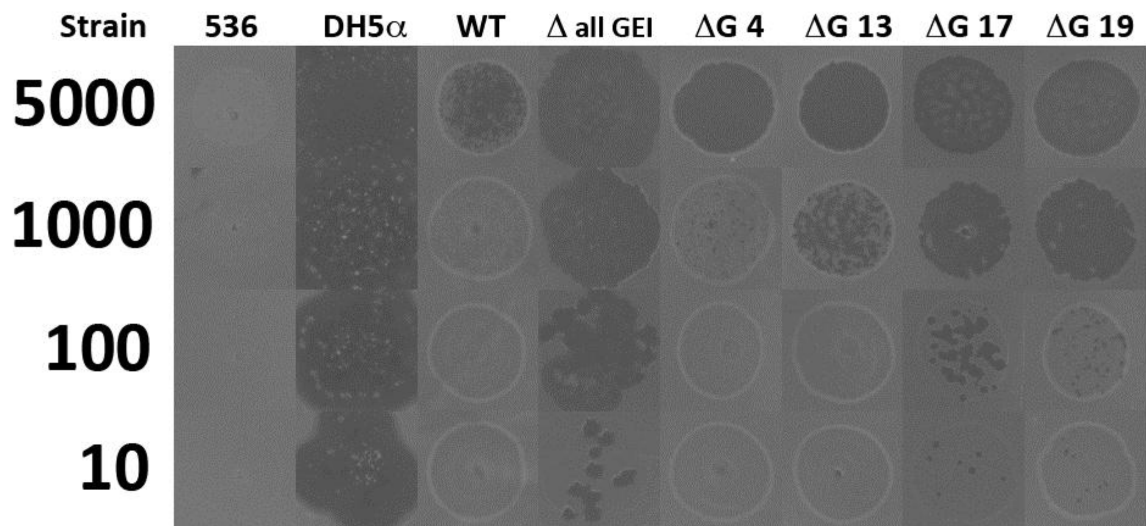


Fig. 15: GEI 13, 17, 19 deletion affect IHE3034 ability to resist to the *Dyctiostelium discoideum* grazing assay.

Dyctiostelium discoideum grazing assay, bacteria were plated on SM plates and challenged with different concentrations of amoeba. 536 is the negative control, DH5a is the positive control; IHE3034 WT to ΔG19 are the samples.

4 Discussion

Genomic islands are often studied from a strain-centric perspective and investigations of the distribution and gene content of these mobile elements in varied strain collections are not frequent. The importance of these studies is paramount as they provide a better understanding of the mechanisms driving the evolution of bacteria with open genomes such as *E. coli* [74].

The distribution analysis proposed in this work has allowed for the study of the prevalence of genomic islands (associated with the NMEC strain IHE3034) in a diverse panel of *E. coli* strains. The data has been stratified by the major pathotypes (ExPEC, InPEC or Non-Pathogenic) in order to detect the differences among the major sub-groups. It comes as no surprise that IHE3034 genomic islands have a higher prevalence (ranging from 0,75% to 24%) in ExPEC strains (Fig. 8). However, it is important to note that the non-pathogenic group also showed the presence of all tested islands (but for GEI19), even if with much lower frequency (0,75% - 11%). The presence of these islands in non-pathogenic *E. coli* strains may be explained by different theories. ExPEC strains are known to live and thrive in the normal intestinal flora of healthy humans without causing damage; such situation is optimal for GEI transmission and creates an ideal environment for horizontal gene transfer to occur [6]. Another theory is that the colonization of the gastrointestinal tract by attenuated strains with reduced pathogenicity that has been developed to enhance host survival [82]. The results obtained in this analysis (Fig. 8) are consistent with other epidemiological findings from Dobrindt *et. al.* and Logue *et. al.* [20, 45]. The almost complete absence of IHE3034 related genomic islands in the InPEC group is to be expected as these bacteria colonize a different niche than the ExPEC ones. Still the low prevalence (0,75% - 7,6%) of islands that are known to carry virulence/fitness factors, like GEI 05, 06, 07, may be of interest to better understand if these strains are able to colonize other niches (Fig. 8).

The distribution analysis studied by a strain specific point of view describes a much more detailed situation (Fig. 9A/B). The InPEC or non-pathogenic strains are generally attributed to the phylogenetic groups A and B1; the lack of NMEC related islands shown in Fig. 9B can be either related to the incapability of the strains to cause harm or to a content of islands different from the one under study. It is of interest though that some intestinal pathogenic isolates are positive for some islands

posing the question if these strains are actually able to survive outside the intestine in either blood or urinary tract.

Of all the strains representing the B2 group the most difficult to understand are the sequence type complex 131 ones. These isolates are emerging as one of the most virulent and antibiotic resistant strains infecting both humans and animals [29, 59, 64]. For this group of isolates the primers used for the prevalence study have failed to consistently amplify almost any island but GEI 10; also considering that the sequences of these strains are not available (at the moment of writing) we can only speculate that the genomic/pathogenic island content is different from IHE3034 one. Still, extensive molecular epidemiology studies have been carried out to understand the possible virulence factors carried by this clonal complex strains; these analysis concluded that no differences can be gleaned from ExPEC strains other than the presence of plasmids encoding for multi-antibiotic resistance genes [58, 64, 75].

It is of interest that the average-linkage algorithm used in this work to analyse the distribution study, was not able to group the APEC strains in a single cluster. This heterogenicity of the avian strains mirrors the previously observed results on APEC's virulence factors by Moulin-Schouleur and colleagues [54]. This data may thus be considered a representation of what really happens in nature where avian strains of E. coli can be isolated from many different hosts.

As previously reported by Middendorf and colleagues genome flexibility also has an impact on the evolution of new bacterial pathogens. The acquisition of new traits by horizontal gene transfer is one of the driving forces in the emergence of new bacterial variants [50]. Comparable to the excision mechanism of bacteriophages, GEIs are thought to exist at least transiently as CIs after excision from the chromosome. The analysis of the circular intermediates formations has shown that nine GEIs (04, 06, 07, 11, 12, 13, 14, 17 and 19) were able to excide from the genome in presence of their respective *int* gene (Fig. 10 and 11) and the *attL-R* sites (Tab. 5).

As considered before GEI 01, 05, 06, 07, 08, 09, 10 and 12, which include both CI forming and non-mobile islands, were associated to the B2 ancestral group ($p<0,5$-$0,01$).

GEI 01 is a non-mobile island of 29-Kb highly associated to the B2 phylogenetic group (Fisher $p<0,01$). This GEI carries an uncharacterized type VI secretion system (T6SS) that has been discovered first in the UPEC strain CFT073; also it has been found in

other strains of different pathotypes of the ExPEC group. Recent studies have proposed this T6SS as a possible NMEC marker due to its action in: inter-bacterial relationships, biofilm formation, cytotoxicity and survival in phagocytic cells[8, 41].

GEI 05 is a genomic island of 61 kilobases inserted at the 5' of a Serine tRNA that is also is not mobile. This GEI has been reported to carry many virulence fitness factors such as S-fimbriae, IroN, putative TonB-dependent receptor, and Antigen 43 [19, 66, 76]. This island is highly associated to the B2 phylotype and to the ExPEC pathotype (Fisher p<0,01). The fact that this GEI is not mobile may be an example of how selective pressure works actively to retain useful pieces of genetic information; this mechanism allows bacteria to have a survival advantage overcoming stress situations not suitable growth conditions [50].

GEI 06 is a mobile island of prophagic origin [55]. In its 47-Kb it carries an active *sitABCD* operon mediating the transport of iron and manganese that confers to the bacteria a boost to resistance to hydrogen peroxide [68, 69]. Real time data showed that operon is highly overexpressed when the growth media gets depleted of iron ions using 2'2'-Dypiridil (data not shown). This islands is mainly associated to B2 phylogenetical group (Fisher p<0,05) and has a higher than usual prevalence in strains not belonging to B2 phylotypes.

As many genomic islands are, GEI 07 is also of prophagic origin and it was first identified by Lloyd and colleagues [44]. This 46-Kb GEI has a high number of unknown ORFs and it's function is unknown. We were able to see that its ability to form circular intermediates is not impaired in absence of the *int* gene and also its excision rates were similar to genomic island 6 ones (Fig 11 and 13). This effect may be explained by an integrase cross-talk mechanism such as the one described by Hochhut *et. al.* in an UPEC strain [35]. It is of note that this effect was present even even if the *att* sites are almost completely different (*att-L* 38,7%, *att-R* 41,9%), sharing though a common pattern of bases that may be a recognized by both the integrases (data not shown). However, more studies will be needed to understand if a similar mechanism is present in an NMEC strain.

GEI cluster 08, 09, 10 is a group of three separated non-CI-forming islands that span 189-Kb. In this work we confirmed the previous observations by Antonenka and colleagues that GEI 08, known as HPI from *Yersinia*, is missing the AT-Rich portion and is not able to excide from the genome [4]. In addition, we were not able to

identify any circular intermediates for GEI 09 and 10 in our conditions. Taken together these data suggests that this cluster could be an example of different genomic islands that have been stabilized into the genome by evolutionary forces in order to maintain their virulence and fitness factors.

All of the GEIs in this cluster are highly associated to the B2 phylotype (Fisher $p<0,01$) even if their prevalence varies among the whole panel (GEI 08 0,5%, GEI 09 2,5%, GEI 10 2,8%). It is of note that 25 out of 46 (54,3%) strains carry both the colibactin island (GEI09) and the TonB island (GEI10) while the remaining 19 strains carry alternatively one of the two indicating a possible association between GEI 09 and GEI 10. In addition, the presence of many tRNA sites of the same type, identical direct repeat sites and of different Insertion sequences (IS) indicates that this zone is a hot spot region for mutation and recombination likely able to be the target of future mutations.

GEI12 is a mobile island of prophagic origins long 40-Kb. Sequencing of the *att* sites showed two sequences of 17bp with an identity of 94,1% (Tab. 5). Also this island is known to carry the *neuO* gene that is involved in the *O*-acetylation of the sialic acid residues of *Escherichia coli* K1, groups W-135, Y, and *C. meningococci*, and group B *Streptococcus* capsular polysaccharides thus modifying their immunogenicity and susceptibility to glycosidases [17]. As previously reported by Mordhorst *et. al*. this island is very poorly represented in non-STC95 strains [52]. The distribution analysis showed that this GEI was present only in ExPEC strains (8/71) and statistical analysis consistently assigns it to the B2 and STC95 groups (7/8 strains - Fisher $p\leq0,01$). However, island excision proved to be not measurable by relative real time means thus not allowing the study of this island's response to cellular stress.

Overall, the study of the prevalence of IHE3034 genomic islands also uncovered four islands (GEI 04, 13, 17, 19) that may be considered as ExPEC specific and very rarely represented or unique outside our reference strain's STC. These islands are all of prophagic origin and still mobile possibly indicating that they have been recently acquired or that the stabilization effort of the cell on these mobile elements is low. Interestingly, as shown in Fig. 12 three islands, GEI 11, 13 and 17, were resistant to DNase digestion. Hence, even if this data has not completely confirmed the production of full-fledged virions, it still indicates that the CIs of these GEIs may be protected by proteins of phagic origin produced by their ORFs. These four islands,

identified by Moriel *et. al.*, have never been described before and their function, as the gene content, is largely unknown.

GEI04 is a medium sized island of 33-Kb that carries no known virulence/fitness factor. Real time analysis indicates that this island's excision rates are stable when cells are treated with Mitomycin C, H_2O_2, or Menadione. When the bacterial culture is grown in cold temperature (20°C) or gets iron depleted using 2'2'-Dipyridyl, the excision rates drop of 6 and 13,5 times (Fig 13A/C). This effect could be due to two mechanisms: the genes on the island give the cell an evolutive advantage so it focuses its efforts to stabilize it into the genome; or the integrase expression levels on this GEI are effectively lower thus leading to a marked reduction of the CI formation.

GEI13, like the other islands of this list, is actively mobile it is 39-Kb long and it has been detected for the first time by Moriel and colleagues [53]. More than 50% of its ORFs are uncharacterized but this is the only GEI, of the less represented ones, that carries a known virulence factor the enterohemolysin 1. This, together with GEI17, is highly susceptible to the stress conditions in which the cell is. The excision rates are variable between the different conditions tested; temperature, a minimal medium (RPMI) and the use of Menadione cause the GEI to be more stable in the genome with rates reduced of 2,3 and 8,1 times (Fig. 13A/D/F). However, Mitomycin C, an antibiotic known for its ability to induce the phage lithic circle [43], and H_2O_2 cause a marked increase in CI production; this is especially true for Mitomycin C as the excision rates rise of 28 times. The fact that the CI is mildly resistant to the DNase treatment and the high presence of prophagic ORFs mark this GEI as of prophagic origin, thus suggesting that it may still be "trying" to switch from a dormant stat to an active one (Fig. 12).

GEI17 is the smallest island of IHE3034 (16-Kb) and almost all of the ORFs have been identified as of prophagic origin. Of all the predicted genes present in this island, 6 over 13 (46,2%), are labelled as conserved hypothetical and of unknown function. Intrestingly this island's circular intermediate is also protected from the DNase action thus posing the question if this islands is still able to produce full phages. The real time analysis shows that the excision rates of island 17 are increased by 2,6 and 6,8 times when the growing media is treated with agents producing free radicals such as H_2O_2 and Menadione. Low temperature causes a great reduction (31,2 times) of this island's CI; this is possibly due to a preservation mechanism of the possible phage

encoded by this island. Also adding Mit. C to the growth media or growing bacteria in RPMI causes a reduction of the presence of circular intermediates of 2 and 3,5 times.
GEI19 is a very peculiar island that has been found as a whole only in IHE3034. This GEI is still actively mobile but its excision rates were not measurable by real time PCR as the *att* region is highly variable and it is not possible to obtain clean PCR results suitable for real time analysis. This island is 46-Kb long and the bioinformatic analysis identified 54 ORFs; of these 39 are fully annotated, 15 are listed as "unknown origin" or of "prophagic origin". BLAST analysis on the NCBI database identified 7 ORFs to be completely unique to IHE3034.
The capacity of bacteria to modulate their genome structure is an important feature for adapting to changing environmental conditions and thus for the evolution of new bacterial pathogens. GEIs contribute to the formation of new species by providing virulence and fitness genes that allow a better survival in hostile environments [22]. Certain IHE3034 genomic islands have only recently been identified and little is known about their role in pathogenesis. In this study we identified four GEIs (04, 13, 17, 19) with a prevalence almost exclusive in STC95 strains. in order to study their possible involvement in pathogenicity a *D. discoideum* grazing assay has been set up as described by Froquet and colleagues [27].
The results showed that GEI 04 was not involved in the phagocytosis resistance as the phenotype of the mutant missing the island was not different from wild-type strain.
On the contrary, genomic island 13, as GEI 17 and 19, may carry unknown genes involved in survival against internalization and digestion by *Dyctiostelium discoideum* as the knock out strains showed a reduced survivability to grazing (Fig. 15).
Bioinformatic analysis had revealed that all the GEIs under study have large portions that are completely uncharacterized and that carry ORFs of unknown function. In order to better understand which genes are the responsible for the grazing resistance phenotype further localized deletion studies are needed. Nonetheless, the *Dictyostelium discoideum* grazing assay has proven to be a solid tool to study deletion mutants resistance to phagocytosis.
In conclusion, despite the genomic similarities between *E. coli* strains, it is evident that numerous genetic differences exist even among strains of the same pathotype or clonal complex. Our understanding of *E. coli* genome modifications would greatly benefit from distribution and functional studies on genomic islands, as they would

give insights on bacterial pathogenesis, host adaptation and their effects on ExPEC strain's virulence potential.

5 Bibliography

1. **Achtman M.**, **A. Mercer**, **B. Kusecek**, **A. Pohl**, **M. Heuzenroeder**, **W. Aaronson**, **A. Sutton**, **and R. P. Silver**. 1983. Six widespread bacterial clones among Escherichia coli K1 isolates. Infection and Immunity **39**:315–335.
2. **Adiba S.**, **C. Nizak**, **M. van Baalen**, **E. Denamur**, **and F. Depaulis**. 2010. From grazing resistance to pathogenesis: the coincidental evolution of virulence factors. PLoS ONE **5**:e11882.
3. **Ahmed N.**, **U. Dobrindt**, **J. Hacker**, **and S. Hasnain**. 2008. Genomic fluidity and pathogenic bacteria: applications in diagnostics, epidemiology and intervention. Nat Rev Micro **6**:387–394.
4. **Antonenka U.**, **C. Nölting**, **J. Heesemann**, **and A. Rakin**. 2006. Independent acquisition of site-specific recombination factors by asn tRNA gene-targeting genomic islands. Int J Med Microbiol **296**:341–352.
5. **Bergthorsson U.**, **and H. Ochman**. 1998. Distribution of chromosome length variation in natural isolates of Escherichia coli. Mol Biol Evol **15**:6–16.
6. **Bezuidt O.**, **R. Pierneef**, **K. Mncube**, **G. Lima-Mendez**, **and O. N. Reva**. 2011. Mainstreams of horizontal gene exchange in enterobacteria: consideration of the outbreak of enterohemorrhagic E. coli O104:H4 in Germany in 2011. PLoS ONE **6**:e25702.
7. **Bille E.**, **J.-R. Zahar**, **A. Perrin**, **S. Morelle**, **P. Kriz**, **K. A. Jolley**, **M. C. J. Maiden**, **C. Dervin**, **X. Nassif**, **and C. R. Tinsley**. 2005. A chromosomally integrated bacteriophage in invasive meningococci. J. Exp. Med. **201**:1905–1913.
8. **Blondel C. J.**, **H.-J. Yang**, **B. Castro**, **S. Chiang**, **C. S. Toro**, **M. Zaldívar**, **I. Contreras**, **H. L. Andrews-Polymenis**, **and C. A. Santiviago**. 2010. Contribution of the Type VI Secretion System Encoded in SPI-19 to Chicken Colonization by Salmonella enterica Serotypes Gallinarum and Enteritidis. PLoS ONE **5**:e11724.
9. **Blum G.**, **M. Ott**, **A. Lischewski**, **A. Ritter**, **H. Imrich**, **H. Tschäpe**, **and J. Hacker**. 1994. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an Escherichia coli wild-type pathogen. Infection and Immunity **62**:606–614.
10. **Bozzaro S.**, **and L. Eichinger**. 2011. The professional phagocyte Dictyostelium discoideum as a model host for bacterial pathogens. Current Drug Targets **12**:942–954.
11. **Bozzaro S.**, **C. Bucci**, **and M. Steinert**. 2008. Phagocytosis and host-pathogen interactions in Dictyostelium with a look at macrophages. Int Rev Cell Mol Biol **271**:253–300.
12. **Clermont O.**, **S. Bonacorsi**, **and E. Bingen**. 2000. Rapid and simple determination of the Escherichia coli phylogenetic group. Applied and Environmental Microbiology **66**:4555–4558.
13. **Cosson P.**, **and T. Soldati**. 2008. Eat, kill or die: when amoeba meets bacteria. Current opinion in microbiology **11**:271–276.
14. **Croxen M. A.**, **and B. B. Finlay**. 2010. Molecular mechanisms of Escherichia coli pathogenicity. Nat Rev Micro **8**:26–38.
15. **Darling A. E.**, **B. Mau**, **and N. T. Perna**. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE **5**:e11147.
16. **Datsenko K.**, **and B. Wanner**. 2000. One-step inactivation of chromosomal

genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci USA **97**:6640.
17. **Deszo E. L., S. M. Steenbergen, D. I. Freedberg, and E. R. Vimr**. 2005. Escherichia coli K1 polysialic acid O-acetyltransferase gene, neuO, and the mechanism of capsule form variation involving a mobile contingency locus. Proc. Natl. Acad. Sci. U.S.A. **102**:5564–5569.
18. **Dobrindt U.** 2005. (Patho-) Genomics of Escherichia coli. International journal of medical microbiology.
19. **Dobrindt U., G. Blum-Oehler, T. Hartsch, G. Gottschalk, E. Z. Ron, R. Fünfstück, and J. Hacker**. 2001. S-Fimbria-encoding determinant sfa(I) is located on pathogenicity island III(536) of uropathogenic Escherichia coli strain 536. Infection and Immunity **69**:4248–4256.
20. **Dobrindt U., M. G. Chowdary, G. Krumbholz, and J. Hacker**. 2010. Genome dynamics and its impact on evolution of Escherichia coli. Med Microbiol Immunol **199**:145–154.
21. **Dobrindt U., L. Emödy, I. Gentschev, W. Goebel, and J. Hacker**. 2002. Efficient expression of the alpha-haemolysin determinant in the uropathogenic Escherichia coli strain 536 requires the leuX-encoded tRNA(5)(Leu). Mol. Genet. Genomics **267**:370–379.
22. **Dobrindt U., B. Hochhut, U. Hentschel, and J. Hacker**. 2004. Genomic islands in pathogenic and environmental microorganisms. Nat Rev Micro **2**:414–424.
23. **Eichinger L.** 2003. Revamp a model-status and prospects of the Dictyostelium genome project. Curr. Genet. **44**:59–72.
24. **Emödy L., M. Kerenyi, and G. Nagy**. 2003. Virulence factors of uropathogenic Escherichia coli. Int J Antimicrob Agents **22 Suppl 2**:29–33.
25. **Foxman B.** 1990. Recurring urinary tract infection: incidence and risk factors. Am J Public Health **80**:331–333.
26. **Foxman B., and P. Brown**. 2003. Epidemiology of urinary tract infections: transmission and risk factors, incidence, and costs. Infect. Dis. Clin. North Am. **17**:227–241.
27. **Froquet R., E. Lelong, A. Marchetti, and P. Cosson**. 2009. Dictyostelium discoideum: a model host to measure bacterial virulence. Nat Protoc **4**:25–30.
28. **Gaudet P., P. Fey, S. Basu, Y. A. Bushmanova, R. Dodson, K. A. Sheppard, E. M. Just, W. A. Kibbe, and R. L. Chisholm**. 2011. dictyBase update 2011: web 2.0 functionality and the initial steps towards a genome portal for the Amoebozoa. Nucleic Acids Res **39**:D620–4.
29. **George D. B., and A. R. Manges**. 2010. A systematic review of outbreak and non-outbreak studies of extraintestinal pathogenic Escherichia coli causing community-acquired infections. Epidemiol Infect **138**:1679–1690.
30. **Gordon D. M., O. Clermont, H. Tolley, and E. Denamur**. 2008. Assigning Escherichia coli strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex method. Environmental Microbiology **10**:2484–2496.
31. **Hacker J., and E. Carniel**. 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. EMBO Rep **2**:376–381.
32. **Hacker J., and J. B. Kaper**. 2000. Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. **54**:641–679.
33. **Hantke K., G. Nicholson, W. Rabsch, and G. Winkelmann**. 2003. Salmochelins, siderophores of Salmonella enterica and uropathogenic

Escherichia coli strains, are recognized by the outer membrane receptor IroN. Proc. Natl. Acad. Sci. U.S.A. **100**:3677–3682.
34. **Hilbi H.**, **S. S. Weber**, **C. Ragaz**, **Y. Nyfeler**, **and S. Urwyler**. 2007. Environmental predators as models for bacterial pathogenesis. Environmental Microbiology **9**:563–575.
35. **Hochhut B.**, **C. Wilde**, **G. Balling**, **B. Middendorf**, **U. Dobrindt**, **E. Brzuszkiewicz**, **G. Gottschalk**, **E. Carniel**, **and J. Hacker**. 2006. Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic Escherichia coli strain 536. Mol Microbiol **61**:584–595.
36. **Homburg S.**, **E. Oswald**, **J. Hacker**, **and U. Dobrindt**. 2007. Expression analysis of the colibactin gene cluster coding for a novel polyketide in Escherichia coli. FEMS Microbiology Letters **275**:255–262.
37. **Huan P.**, **D. Bastin**, **B. Whittle**, **A. Lindberg**, **and N. Verma**. 1997. Molecular characterization of the genes involved in O-antigen modification, attachment, integration and excision in Shigella flexneri bacteriophage SfV. Gene **195**:217–227.
38. **Huang S. H.**, **Z. S. Wan**, **Y. H. Chen**, **A. Y. Jong**, **and K. S. Kim**. 2001. Further characterization of Escherichia coli brain microvascular endothelial cell invasion gene ibeA by deletion, complementation, and protein expression. J. Infect. Dis. **183**:1071–1078.
39. **Jaureguy F.**, **L. Landraud**, **V. Passet**, **L. Diancourt**, **E. Frapy**, **G. Guigon**, **E. Carbonnelle**, **O. Lortholary**, **O. Clermont**, **E. Denamur**, **B. Picard**, **X. Nassif**, **and S. Brisse**. 2008. Phylogenetic and genomic diversity of human bacteremic Escherichia coli strains. BMC genomics **9**:560.
40. **Johnson J. R.**, **and T. A. Russo**. 2002. Extraintestinal pathogenic Escherichia coli: "the other bad E coli". J. Lab. Clin. Med. **139**:155–162.
41. **Johnson T. J.**, **Y. Wannemuehler**, **S. Kariyawasam**, **J. R. Johnson**, **C. M. Logue**, **and L. K. Nolan**. 2012. Prevalence of avian pathogenic Escherichia coli strain APEC O1 genomic islands among extraintestinal and commensal E. coli isolates. Journal of Bacteriology **194**:2846–2853.
42. **Kaper J. B.**, **J. P. Nataro**, **and H. L. Mobley**. 2004. Pathogenic Escherichia coli. Nat Rev Micro **2**:123–140.
43. **King M. R.**, **R. P. Vimr**, **S. M. Steenbergen**, **L. Spanjaard**, **G. Plunkett**, **F. R. Blattner**, **and E. R. Vimr**. 2007. Escherichia coli K1-specific bacteriophage CUS-3 distribution and function in phase-variable capsular polysialic acid O acetylation. Journal of Bacteriology **189**:6447–6456.
44. **Lloyd A.**, **D. A. Rasko**, **and H. Mobley**. 2007. Defining genomic islands and uropathogen-specific genes in uropathogenic Escherichia coli. Journal of Bacteriology **189**:3532.
45. **Logue C. M.**, **C. Doetkott**, **P. Mangiamele**, **Y. M. Wannemuehler**, **T. J. Johnson**, **K. A. Tivendale**, **G. Li**, **J. S. Sherwood**, **and L. K. Nolan**. 2012. Genotypic and phenotypic traits that distinguish neonatal meningitis-associated Escherichia coli from fecal E. coli isolates of healthy human hosts. Applied and Environmental Microbiology **78**:5824–5830.
46. **Loś J. M.**, **M. Loś**, **A. Wegrzyn**, **and G. Wegrzyn**. 2010. Hydrogen peroxide-mediated induction of the Shiga toxin-converting lambdoid prophage ST2-8624 in Escherichia coli O157:H7. FEMS Immunol. Med. Microbiol. **58**:322–329.
47. **Maiden M. C. J.** 2006. Multilocus sequence typing of bacteria. Annu. Rev. Microbiol. **60**:561–588.

48. **McDonald J.**, **D. Smith**, **P. Fogg**, **A. McCarthy**, **and H. Allison**. 2010. High-Throughput Method for Rapid Induction of Prophages from Lysogens and Its Application in the Study of Shiga Toxin-Encoding Escherichia coli Strains. Applied and Environmental Microbiology **76**:2360.
49. **Meintjes P.**, **C. Duran**, **M. Kearse**, **R. Moir**, **A. Wilson**, **S. Stones-Havas**, **M. Cheung**, **S. Sturrock**, **S. Buxton**, **A. Cooper**, **S. Markowitz**, **T. Thierer**, **B. Ashton**, **and J. Heled**. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics.
50. **Middendorf B.**, **B. Hochhut**, **K. Leipold**, **U. Dobrindt**, **G. Blum-Oehler**, **and J. Hacker**. 2004. Instability of pathogenicity islands in uropathogenic Escherichia coli 536. Journal of Bacteriology **186**:3086.
51. **Morales-Espinosa R.**, **G. Soberón-Chávez**, **G. Delgado-Sapién**, **L. Sandner-Miranda**, **J. L. Méndez**, **G. González-Valencia**, **and A. Cravioto**. 2012. Genetic and phenotypic characterization of a Pseudomonas aeruginosa population with high frequency of genomic islands. PLoS ONE **7**:e37459.
52. **Mordhorst I. L.**, **H. Claus**, **C. Ewers**, **M. Lappann**, **C. Schoen**, **J. Elias**, **J. Batzilla**, **U. Dobrindt**, **L. H. Wieler**, **A. K. Bergfeld**, **M. Mühlenhoff**, **and U. Vogel**. 2009. O-acetyltransferase gene neuO is segregated according to phylogenetic background and contributes to environmental desiccation resistance in Escherichia coli K1. Environmental Microbiology **11**:3154–3165.
53. **Moriel D. G.**, **I. Bertoldi**, **A. Spagnuolo**, **S. Marchi**, **R. Rosini**, **B. Nesta**, **I. Pastorello**, **V. A. M. Corea**, **G. Torricelli**, **E. Cartocci**, **S. Savino**, **M. Scarselli**, **U. Dobrindt**, **J. Hacker**, **H. Tettelin**, **L. J. Tallon**, **S. Sullivan**, **L. H. Wieler**, **C. Ewers**, **D. Pickard**, **G. Dougan**, **M. R. Fontana**, **R. Rappuoli**, **M. Pizza**, **and L. Serino**. 2010. Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic Escherichia coli. Proc Natl Acad Sci USA **107**:9072–9077.
54. **Moulin-Schouleur M.**, **M. Répérant**, **S. Laurent**, **A. Brée**, **S. Mignon-Grasteau**, **P. Germon**, **D. Rasschaert**, **and C. Schouler**. 2007. Extraintestinal pathogenic Escherichia coli strains of avian and human origin: link between phylogenetic relationships and common virulence patterns. Journal of Clinical Microbiology **45**:3366–3376.
55. **Napolitano M. G.**, **S. Almagro-Moreno**, **and E. F. Boyd**. 2011. Dichotomy in the evolution of pathogenicity island and bacteriophage encoded integrases from pathogenic Escherichia coli strains. Infect. Genet. Evol. **11**:423–436.
56. **Panis G.**, **Y. Duverger**, **E. Courvoisier-Dezord**, **S. Champ**, **E. Talla**, **and M. Ansaldi**. 2010. Tight regulation of the intS gene of the KplE1 prophage: a new paradigm for integrase gene regulation. PLoS Genet **6**.
57. **Pfaffl M. W.**, **G. W. Horgan**, **and L. Dempfle**. 2002. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res **30**:e36.
58. **Platell J. L.**, **R. N. Cobbold**, **J. R. Johnson**, **A. Heisig**, **P. Heisig**, **C. R. Clabots**, **M. A. Kuskowski**, **and D. J. Trott**. 2011. Commonality among fluoroquinolone-resistant sequence type ST131 extraintestinal Escherichia coli isolates from humans and companion animals in Australia. Antimicrobial Agents And Chemotherapy.
59. **Platell J. L.**, **J. R. Johnson**, **R. N. Cobbold**, **and D. J. Trott**. 2011. Multidrug-resistant extraintestinal pathogenic Escherichia coli of sequence type ST131 in

animals and foods. Vet Microbiol.
60. **Preston G. M., B. Haubold, and P. B. Rainey**. 1998. Bacterial genomics and adaptation to life on plants: implications for the evolution of pathogenicity and symbiosis. Current opinion in microbiology **1**:589–597.
61. **Putze J., C. Hennequin, J.-P. Nougayrède, W. Zhang, S. Homburg, H. Karch, M.-A. Bringer, C. Fayolle, E. Carniel, W. Rabsch, T. A. Oelschlaeger, E. Oswald, C. Forestier, J. Hacker, and U. Dobrindt**. 2009. Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. Infection and Immunity **77**:4696–4703.
62. **Qiagen, and M. W. Pfaffl**. REST 2009 Software User Guidegene-quantification.de.
63. **Ritter A., G. Blum, L. Emödy, M. Kerenyi, A. Böck, B. Neuhierl, W. Rabsch, F. Scheutz, and J. Hacker**. 1995. tRNA genes and pathogenicity islands: influence on virulence and metabolic properties of uropathogenic Escherichia coli. Mol Microbiol **17**:109–121.
64. **Rogers B. A., H. E. Sidjabat, and D. L. Paterson**. 2011. Escherichia coli O25b-ST131: a pandemic, multiresistant, community-associated strain. J. Antimicrob. Chemother. **66**:1–14.
65. **Russo T. A., and J. R. Johnson**. 2003. Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem. Microbes Infect. **5**:449–456.
66. **Russo T. A., C. D. McFadden, U. B. Carlino-MacDonald, J. M. Beanan, T. J. Barnard, and J. R. Johnson**. 2002. IroN functions as a siderophore receptor and is a urovirulence factor in an extraintestinal pathogenic isolate of Escherichia coli. Infection and Immunity **70**:7156–7160.
67. **S K., and M. S**. *Bioinformatics Methods and Protocols: Methods in Molecular Biology.*, pp. 365–386. *In Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Humana Press.
68. **Sabri M., M. Caza, J. Proulx, M. H. Lymberopoulos, A. Brée, M. Moulin-Schouleur, R. Curtiss, and C. M. Dozois**. 2008. Contribution of the SitABCD, MntH, and FeoB metal transporters to the virulence of avian pathogenic Escherichia coli O78 strain chi7122. Infection and Immunity **76**:601–611.
69. **Sabri M., S. Léveillé, and C. M. Dozois**. 2006. A SitABCD homologue from an avian pathogenic Escherichia coli strain mediates transport of iron and manganese and resistance to hydrogen peroxide. Microbiology (Reading, Engl) **152**:745–758.
70. **Shevchuk O., and M. Steinert**. 2009. Screening of virulence traits in Legionella pneumophila and analysis of the host susceptibility to infection by using the Dictyostelium host model system. Methods Mol Biol **470**:47–56.
71. **Smith J., P. Fratamico, N. Gunther, and A. USDA**. 2007. Extraintestinal pathogenic Escherichia coli. Foodborne Pathog Dis.
72. **Song L., Y. Pan, S. Chen, and X. Zhang**. 2012. Structural characteristics of genomic islands associated with GMP synthases as integration hotspot among sequenced microbial genomes. Comput Biol Chem **36**:62–70.
73. **Sozhamannan S., M. D. Chute, F. D. McAfee, D. E. Fouts, A. Akmal, D. R. Galloway, A. Mateczun, L. W. Baillie, and T. D. Read**. 2006. The Bacillus anthracis chromosome contains four conserved, excision-proficient, putative prophages. BMC Microbiol **6**:34.
74. **Tenaillon O., D. Skurnik, B. Picard, and E. Denamur**. 2010. The population

genetics of commensal Escherichia coli. Nat Rev Micro **8**:207–217.

75. **Totsika M., S. A. Beatson, S. Sarkar, M.-D. Phan, N. K. Petty, N. Bachmann, M. Szubert, H. E. Sidjabat, D. L. Paterson, M. Upton, and M. A. Schembri**. 2011. Insights into a Multidrug Resistant Escherichia coli Pathogen of the Globally Disseminated ST131 Lineage: Genome Analysis and Virulence Mechanisms. PLoS ONE **6**:e26578.
76. **Ulett G. C., J. Valle, C. Beloin, O. Sherlock, J. M. Ghigo, and M. A. Schembri**. 2007. Functional analysis of antigen 43 in uropathogenic Escherichia coli reveals a role in long-term persistence in the urinary tract. Infection and Immunity **75**:3233–3244.
77. **Vandesompele J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman**. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology **3**:RESEARCH0034.
78. **Wagner P. L., and M. K. Waldor**. 2002. Bacteriophage control of bacterial virulence. Infection and Immunity **70**:3985.
79. **Whitfield C.** 2006. Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. Annu. Rev. Biochem. **75**:39–68.
80. **Wiles T. J., R. R. Kulesus, and M. A. Mulvey**. 2008. Origins and virulence mechanisms of uropathogenic Escherichia coli. Exp. Mol. Pathol. **85**:11–19.
81. **Wirth T., D. Falush, R. Lan, F. Colles, P. Mensa, L. H. Wieler, H. Karch, P. R. Reeves, M. C. J. Maiden, H. Ochman, and M. Achtman**. 2006. Sex and virulence in Escherichia coli: an evolutionary perspective. Mol Microbiol **60**:1136–1151.
82. **Zdziarski J., C. Svanborg, B. Wullt, J. Hacker, and U. Dobrindt**. 2008. Molecular basis of commensalism in the urinary tract: low virulence or virulence attenuation? Infection and Immunity **76**:695–703.
83. **Zhou D., W. D. Hardt, and J. E. Galán**. 1999. Salmonella typhimurium encodes a putative iron transport system within the centisome 63 pathogenicity island. Infection and Immunity **67**:1974–1981.
84. **Ziebuhr W., K. Ohlsen, H. Karch, T. Korhonen, and J. Hacker**. 1999. Evolution of bacterial pathogenesis. Cell. Mol. Life Sci. **56**:719–728.
85. **Zou Y., L. He, F. Chi, A. Jong, and S. H. Huang**. 2008. Involvement of Escherichia coli K1 ibeT in bacterial adhesion that is associated with the entry into human brain microvascular endothelial cells. Medical microbiology and ….