An analysis of the evolutionary relationships of integron integrases, with emphasis on the prevalence of class 1 integrons in *Escherichia coli* isolates from clinical and environmental origins

J. Javier Díaz-Mejía,^{1,2} Carlos F. Amábile-Cuevas,¹ Irma Rosas³ and Valeria Souza⁴

Correspondence

Carlos F. Amábile-Cuevas
carlos.amabile@lusara.org

¹Fundación Lusara, Apartado Postal 102-006, 08930, Mexico City, Mexico

²Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Av. Universidad 2001, Col. Chamilpa, 62210 Cuernavaca, Morelos, Mexico

³Centro de Ciencias de la Atmósfera, Universidad Nacional Autónoma de México, Circuito Exterior, 04510, Mexico City, Mexico

⁴Instituto de Ecología, Universidad Nacional Autónoma de México, Circuito Exterior, 04510, Mexico City, Mexico

Integrons are genetic elements that allow the mobilization and expression of smaller elements called gene cassettes, and are considered to be key elements in the evolution of antibiotic resistance among enteric bacteria. Although in nature integrons appear to be abundant, the presence of class 1 integrons in *Escherichia coli* has been reported to be much less frequent among isolates of non-human origin than among clinical ones. Searching for integrons in a wide variety of *E. coli* isolates we found a steep decline in class 1 integron prevalence when going from clinical strains to environmental ones, from outdoor urban dust to the microbiota of wild animals. Attempting to assess the causes of this decline, we addressed the evolution of integron integrases, comparing the amino acid sequence of various of these enzymes, the key proteins in gene-cassette mobilization. We found that all integrases are homologues, but different classes have been recruited by enteric bacteria, supporting the notion that integrons can frequently be gained and lost. Additionally, we found that phylogenetically distant organisms that bear *intI1*, such as *E. coli* and other enteric bacteria, but also the Gram-positive corynebacteria, have a similar preferential genomic codon usage (CU), suggesting that CU might play an important role in the acquisition and/or maintenance of integrons. In fact, the CU of *intI1* is more similar to the preferential genomic CU of non-enteric bacteria than it is to that of *E. coli*. CU has been proposed to be involved in the retention of horizontally transferred genes; integrons in *E. coli* are often plasmid-borne. This might explain the reduced prevalence of integrons in enteric bacteria when not under the selective pressure of antibiotics. Collectively, our results provide evidence that class 1 integrons are important gene mobilizers within *E. coli*, but are not acquired and/or stably maintained without selective pressure. Thus, although not effective to reduce the prevalence of resistance itself, decreasing the use of antibiotics could be useful to diminish the presence of gene-mobilization machineries.

Received 30 March 2007

Revised 16 September 2007

Accepted 17 September 2007

INTRODUCTION

Integrons are bacterial genetic elements that allow the shuffling of smaller mobile elements called gene cassettes; they have been termed 'a genetic construction kit for

bacteria' (Bennett, 1999). The main components of an integron are: an integrase enzyme (IntI), responsible for gene cassette integration; a recombination site (*attI*) that is the target of the integrase; and a promoter located upstream of the integration site. Integrons are involved in the evolution and spread of antibiotic-resistance genes in enteric bacteria, as many such genes mobilize as gene cassettes, are expressed in foreign cellular backgrounds

Abbreviations: CAI, codon adaptation index; CU, codon usage; HMM, hidden Markov model; RMSD, root mean squared difference.

(due to the promoter included in the integron), and can be co-selected by mobile and non-mobile resistance genes linked to the integron. There are at least eight classes of integrons (Nield *et al.*, 2001), but those found in clinical isolates belong to four main classes according to their integrases and associated cassettes (Hall *et al.*, 1996; Rowe-Magnus *et al.*, 2001). Further classification attempts, as well as some of the evolutionary history of integrons, have recently been reviewed (Boucher *et al.*, 2007).

Class 1 integrons have been described as the main mobilizers of antibiotic-resistance genes amongst enteric bacteria, and particular attention has been given to measuring their incidence in *Escherichia coli* isolates. Although Gram-positive bacteria have been recently highlighted as 'reservoirs' for integrons in some environments (Nandi *et al.*, 2004), the general assumption is that integrons are natives of enteric bacteria, or at least stably maintained without defined selective pressure. In fact, integrons are thought not to play a significant role in the evolution of resistance in Gram-positive bacteria (Firth, 2003).

The prevalence of integrons in *E. coli* isolates, however, seems to relate to how close the isolates are to a human-influenced environment. Many, if not most, *E. coli* clinical isolates harbour class 1 integrons, along with the *sulI* gene, which mediates sulfonamide resistance, and which is linked to nearly all class 1 integrons. Among hospital enteric isolates in an earlier study, up to 48 % were found to carry the *intI1* gene, which encodes the integrase of class 1 integrons, as well as 13 % of community isolates from patients just entering hospital (the majority of these isolates were *E. coli*; Leverstein-Van Hall *et al.*, 2002). Of *E. coli* strains isolated from healthy people, 11 % carried class 1 integrons (Skurnik *et al.*, 2005). Among *E. coli* isolates from livestock fed with antibiotics, where nearly all (97 %) were resistant to sulfamethoxazole, 73 % of the strains carried the *intI1* gene (Goldstein *et al.*, 2001). However, amongst *E. coli* strains isolated from irrigation water and sediments at the Rio Grande river, only 1 % (13 % of the 10 % multi-resistant isolates) had class 1 integrons (Roe *et al.*, 2003). Even comparing closely related environments, such as indoor and outdoor dust in urban areas, class 1 integrons are more prevalent among outdoor strains (16 vs 2 %), which are also more likely to be of human origin, judging from other resistance and enteropathogenicity traits (Rosas *et al.*, 2006). On the other hand, when looking for integrons in metagenomic DNA samples, even those from areas without the likely influence of antibiotics, they appear to be widely distributed (Nield *et al.*, 2001), although the presence and prevalence of *E. coli* within these samples has not been assessed. Other selective pressures, such as heavy metals, could be at work in these environments (Nemergut *et al.*, 2004). Also, a wide variety of putative gene cassettes can be PCR-amplified from soil samples, indicating the existence of a gene pool capable of being mobilized by integrons in the environment (Michael *et al.*, 2004). When a set of bovine faecal samples were

PCR-tested for *intI1*, 86 % were positive; but using media that enriched the *Enterobacteriaceae* fraction, only 50 % of strains recovered were *intI1* PCR-positive (Barlow *et al.*, 2004). These results suggest that class 1 integrons are only acquired and/or maintained in *E. coli* under the direct selective pressure of antibiotics, and that class 1 integrons are not native elements of *E. coli*.

To provide further experimental support to the 'fading' of integrons in *E. coli*, we searched for them in a wide variety of isolates, ranging from hospital clinical isolates to environmental ones. Our findings confirm the above observations on the prevalence of integrons. We also found a broad diversity in the size of gene cassettes among clinical isolates, compared with the same-sized cassettes found in 'environmental' strains. An evolutionary perspective, using bioinformatic tools, indicates that enteric bacteria have acquired integrases multiple times during evolution. A different codon usage between *intI1* and the one preferred by *E. coli* might play an important role in the acquisition and/or maintenance of these genetic elements that are mostly plasmid-borne in this species.

METHODS

Strains. *E. coli* clinical strains (114) were obtained from various hospital microbiology laboratories in Mexico City, immediately after isolation. Strains from urban dust (183) were isolated from samples of indoor settled dust from within family houses, collected using a vacuum cleaner (VK121, Vorwerk) with sterile bags, and from soil samples from the outdoors of the same houses (116) collected on a plastic dustpan and a sterile brush. Strains from sewage treatment plants (193) were isolated from water samples of an activated-sludge tank in a waste-treatment plant. Strains from wild animals (134) came from the Institute of Ecology Collection of *E. coli* (IECOL)/UNAM collection, which has *E. coli* strains from different continents (mostly the Americas and Antarctica) and from different hosts (reptiles, birds and mammals), almost all living in their natural habitats (Souza *et al.*, 1999). All strains were identified as *E. coli* using standard biochemical techniques, using either automatic (Vitek, bioMérieux) or manual systems, testing carbohydrate and amino acid utilization, indole and hydrogen sulfide production, urea hydrolysis, etc., and kept at -70°C in liquid media containing 25 % (v/v) glycerol. Strains were tested for susceptibility to sulfonamides by the disk-diffusion method, using sulfadiazine 250 µg disks (BBL) and following the manufacturer's guidelines.

Class 1 integron PCR assays. The *intI1* integrase gene was searched for by PCR, using the method and primers reported elsewhere (Roe *et al.*, 2003) for sulfonamide-resistant strains [although it is possible for a sulfonamide-susceptible strain to carry a class 1 integron, very few integron-carrying strains do not also carry and express the *sulI* gene (Goldstein *et al.*, 2001; Partridge *et al.*, 2001), so we only PCR-tested resistant strains]. The assay of Lévesque & Roy (1993) was used to amplify inserted gene cassettes. Amplification products were separated electrophoretically in a FlashGel device (Cambrex).

Sequence collection. We used the sequences of *IntI1* (from *E. coli* Tn21) and *IntI4* (from *Vibrio cholerae* superintegrons) as seeds for a search of closely related proteins using BLASTP (Altschul *et al.*, 1997) with a cut-off E-value of 10^{-10} . The resulting sequences were aligned using the T-COFFEE program (Notredame *et al.*, 2000) and realigned manually. The 3D structure of *IntI4* (MacDonald *et al.*, 2006) and the

secondary-structure predictions for the rest of the sequences were used as guides for the realignment. The secondary-structure predictions were done using PSI-PRED server (McGuffin *et al.*, 2000). The alignment was used to construct a hidden Markov model (HMM) to represent the relative entropy of each amino acid at each position along the alignment. Importantly, this alignment includes the α_2 -helix and other residues that distinguish integrases from other tyrosine-recombinase superfamily members (Messier & Roy, 2001). With the HMM, a search for integrases and related sequences against the non-redundant protein database was carried out using the HMMER suite of programs (Eddy, 1998) with an E-value of 10^{-5} as threshold. Alternatively, PSI-BLAST was used to search for homologues iteratively, but due to the high divergence in this superfamily, the PSI-BLAST search did not converge after thousands of iterations (cut-off E-value=0.00001). For this reason, we preferred to search for homologues using the structure-based HMM. Additionally, a TBLASTN (Altschul *et al.*, 1997) search using the Tn21-IntI1 sequence as seed was carried out against the NCBI database of finished and unfinished microbial genomes (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi) with a cut-off E-value of 10^{-10} .

Clustering of sequences. To cluster the sequences resulting from the HMM-guided search, an all-against-all comparison was conducted using BLAST (Altschul *et al.*, 1997), as implemented in the CLANS program (Frickey & Lupas, 2004) using default parameters. In the resulting network, each node represents a protein and each link corresponds to the *P* value of paired sequence comparisons. In CLANS, nodes are clustered according to the Fruchterman-Reingold graph layout algorithm to visualize pairwise sequence similarities in either 2D or 3D space. In this way, the sequences having the α_2 -helix were separated from the rest of the superfamily for further phylogenetic analyses.

Phylogenetic tree reconstruction. A test of parsimony was applied to the α_2 -helix-containing sequences. To this end, we employed the CLUSTAL_X program (profile mode; Chenna *et al.*, 2003) to align sequences using the IntI4 structure as template. The alignment was edited manually, eliminating non-informative ambiguously aligned sites (mainly gaps). With the resulting alignment, a phylogenetic tree (parsimony) was constructed using the programs 'SEQBOOT', 'PROTPARS'

and 'CONSENSE' of the PHYLIP package (Felsenstein, 2005). The tree was constructed using XerC/D sequences as outgroups. Additionally the program 'PROML' of the PHYLIP package was used to construct a tree by maximum-likelihood, giving similar results to those obtained by parsimony.

Codon adaptation index (CAI) analyses. We compared the codon usage (CU) of *intI1* sequences against the preferential CU of 302 completely sequenced genomes (25 archaea, 274 bacteria and three fungi) obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000). An all-against-all comparison of preferential genomic CUs was also computed for this set of genomes. To this end, we used the 'CAI' and 'CUSP' programs of the EMBOSS suite (Rice *et al.*, 2000) with default parameters.

RESULTS

Prevalence of sulfonamide resistance and class 1 integrons

In *E. coli* the prevalence of sulfonamide resistance decreased as the source of the isolates moved from the clinical setting to the open environment, going from a maximum of 54 % among clinical isolates to 14 % among strains from indoor dust (Fig. 1). However, the prevalence of class 1 integrons, as revealed by *intI1* PCR, decreased more dramatically, going from 24 % among clinical strains (44 % of sulfonamide-resistant strains) to 0 % among isolates from wild animals (Fig. 1).

Gene cassettes within integrons

All inserted cassettes amplified from dust and sewage isolates were of approximately the same size (900 bp); amplicons of the same size were found in only 53 % of *intI1* PCR-positive clinical isolates, while the rest yielded a

Fig. 1. Prevalence of sulfonamide resistance and integron bearing among *E. coli* isolates of different origins. Abbreviations: *sulR*, sulfonamide resistant strains; *intI1*(+), strains yielding a PCR-amplification product using *intI1* primers.

Fig. 2. PCR-amplified integron cassettes. The first three lanes show typical amplicons obtained from dust and sewage isolates of identical molecular weight (~900 bp); the clinical amplicons in lanes 4–6 are more diverse.

variety of PCR products (Fig. 2). All *intI1* PCR-positive strains yielded one or more amplicons using the method of Lévesque & Roy (1993).

Phylogenetic relationships of integrases

In order to assess whether integrons in enteric bacteria, and particularly in *E. coli*, have been acquired and lost during evolution, we extended our analysis from the contemporary-ecological perspective shown above to an evolutionary-genomic view, addressing the origin and evolution of integrases. We used the sequences of Tn21–IntI1 and IntI4 from *V. cholerae* superintegrons as seeds to search for homologue sequences in the NCBI non-redundant protein sequence database, using BLASTP. The resulting matches were used to construct a structure-based HMM specific for the tyrosine-recombinase superfamily. The HMM contains the α I₂-helix that distinguishes structurally and functionally the integrases from other members of the superfamily (Fig. 3a) (MacDonald *et al.*, 2006). This HMM was used to search for homologues in the NCBI non-redundant database. The collected sequences were compared all-against-all with BLASTP, as implemented in the program CLANS, to find clusters of similar sequences. This analysis showed two major clusters (Fig. 3b). The first cluster is formed by XerC/D recombinases (without an α I₂-helix) and is widely distributed across bacteria and archaea. The second cluster is formed by integrases (with an α I₂-helix) and is restricted to some bacterial species.

Interestingly, there was a subset of XerC/D recombinases highly similar to integrases but lacking the α I₂-helix (Fig. 3b). We suggest that this group could represent an evolutionary transition state between XerC/D recombinases and integrases. There are two types of sequences in this subset. One type are chromosomal sequences from β -proteobacteria, vibionales and other γ -proteobacteria, and the second type are plasmid-encoded proteins (red-framed

nodes in Fig. 3b), mainly from planctomycetes. In fact, the latter are placed between XerC/D recombinases and integrases in a phylogenetic tree (group I in Fig. 3c), supporting our notion that these proteins represent a transition state between these two protein families.

Within the integrase family, some observations can be made. In general, in contrast to the XerC/D proteins, which appear to be typically chromosomal, plasmid and chromosomal integrases appear in this network. In particular, there are two groups of highly related integrases. One group (Fig. 3b) mainly consists of chromosomal integrases from vibionales, including IntI4, β -proteobacteria and xanthomonadales. The other group, including IntI1 (Fig. 3b), is mainly formed by integrases from pseudomonadales and enteric bacteria, and shows a highly inter-mixed content of plasmid and chromosomal integrases. Additionally, we searched for integrons using the sequence of Tn21–IntI1 against the NCBI microbial database of finished and unfinished genomes, using TBLASTN. Almost all the ~500 matched sequences were chromosomally encoded and only five, possessing an α I₂-helix, were plasmid-encoded (three from *Salmonella enterica*, one from *Klebsiella pneumoniae* and one from *Aeromonas salmonicida*).

Finally, we constructed a phylogenetic tree of the sequences that possess an α I₂-helix. The XerC/D recombinases most similar to integrases were used as outgroup (group I in Fig. 3c). The first separating group of integrases (group II) represents highly divergent putative integrases from diverse bacterial species, namely vibionales, actinobacteria and planctomycetes. Group III is formed mainly by integrases from enteric bacteria and vibionales, and shows the closer relationship between IntI2 and IntI4. Additionally, this group possesses putative integrases from spirochaetes and δ -proteobacteria. Group IV is formed mainly by putative integrases from β - and δ -proteobacteria, xanthomonadales, and environmental uncultured bacteria. These environmental integrases, namely IntI6, IntI7 and IntI8, show a high similarity with those from xanthomonadales. Finally, group V is formed by IntI1 and IntI3, and IntI3 represents the most basal branch in this group. IntI1 is widely distributed in γ -proteobacteria (enteric bacteria and pseudomonadales), β -proteobacteria and actinobacteria, while IntI3 appears restricted to enteric bacteria.

CU of the *intI1* gene

It has been suggested that the CAI, a measure of the similarity between the CU of a gene and the preferential CU of a genome, can be used to predict when a gene has been transferred horizontally (Garcia-Vallve *et al.*, 2003; Ochman *et al.*, 2000). A recent study, however, shows that successfully horizontally transferred genes generally exhibit a CU similar to that preferred by the recipient genome; otherwise, they are quickly lost (Medrano-Soto *et al.*, 2004). We assessed whether these proposals agree with the

spread and retention of *intI1* by comparing its CU with those preferred by various genomes. We found high similarity between the CU of *intI1* and the preferential genomic CU of *Porphyromonas gingivalis* (CAI=0.823), a member of the *Cytophaga/Flavobacter/Bacteroides* phylum of anaerobic Gram-negative bacteria; *Xylella fastidiosa* (CAI=0.803), a plant pathogen of the family *Xanthomonadaceae*; and *Corynebacterium diphtheriae* (CAI=0.8), a Gram-positive member of the family *Mycobacteriaceae* and the causative agent of diphtheria. In contrast, *E. coli*

occupies the 43rd place in CU similarity with *intI1* (CAI=0.65), out of the 302 genomes analysed herein. In fact, only 749 out of a total of 4234 ORFs identified in the *E. coli* K12 genome have a lower CAI.

An all-against-all comparison of preferential genomic CU profiles was carried out for 159 representative prokaryotic genomes. This comparison gives a root mean squared difference (RMSD) for each pair of genomic CU profiles. The RMSD values were exported to a matrix to cluster

Fig. 3. Phylogenetic relationships of integrases. (a) The sequences of IntI1 and IntI4 were used as seeds to search closely related integrases. The resulting sequences were aligned using the 3D structure of IntI4 as template (see Methods). The resulting alignment was used to construct an HMM specific for the tyrosine-recombinase superfamily. The α_2 -helix (cylinder) and other residues (asterisks) that distinguish integrases from other members of the superfamily are shown. (b) With the HMM, a search was carried out for integrases and related sequences against the non-redundant protein database. An all-against-all comparison of the resulting sequences was conducted to cluster similar sequences into a network. In the network, each node represents a protein and each link corresponds to the P-value of paired sequence comparisons. Node background colours denote phylogenetic groups. Red frames indicate plasmid-encoded proteins, otherwise they are chromosomal. Pink arrows indicate the position of IntI1 and IntI4 used as seeds to search for integrases. (c) The sequences possessing the α_2 -helix (integrases) detected in the network were aligned using the HMM as template. A phylogenetic tree (parsimony) was constructed using the α_2 -helix-containing sequences. In the tree, each branch represents a set of sequences depurated at 95 % identity for simplification. The number of sequences representing each phylogenetic group is shown in brackets. Four miscellaneous branches (bold type), representing more than one phylogenetic group, are described: miscellaneous I= β -proteobacteria [1], enteric bacteria [17], pseudomonadales [18], other γ -proteobacteria [1]; miscellaneous II=actinobacteria [2], enteric bacteria [2], pseudomonadales [4]; miscellaneous III=actinobacteria [4], β -proteobacteria [1], enteric bacteria [55], pseudomonadales [33], other γ -proteobacteria [4], environmental bacteria [8]; miscellaneous IV=actinobacteria [4], enteric bacteria [3], pseudomonadales [3]. Five groups of integrases were detected (I–V). The tree was constructed using XerC/D sequences as outgroups (group I).

them hierarchically by similarity (Fig. 4). Interestingly, the cluster including *E. coli* also included distantly related genomes, such as those of *C. diphtheriae*, *X. fastidiosa* and *P. gingivalis*, which also carry *intI1* genes.

DISCUSSION

Bacterial resistance to antibiotics, particularly multi-resistance, is a serious public health problem. To date, the only available strategy to at least try to hinder the evolution of resistance is to diminish the rate of antibiotic usage. However, due to the intricacies of the selection and maintenance pressures at play, resistance has proven to be resistant to these efforts (Heinemann *et al.*, 2000; Salyers & Amábile-Cuevas, 1997). The evolution of resistance, and of multi-resistance in particular, among Gram-negative bacteria has been mainly attributed to mobile genetic

elements and horizontal gene transfer (Amábile-Cuevas & Chicurel, 1992). Integrons seem to have played a central role in the mobilization and assembly of a variety of resistance genes in single genetic elements, which in turn make them easier to spread and to be selected in a wide diversity of hosts.

Our results show that the presence in *E. coli* of class 1 integrons, which seem to be the most frequent of these genetic elements, is closely associated with human-related environments (Fig. 1) that are more likely to be affected by antibiotic selective pressures. Our evolutionary analyses support this idea, showing multiple origins and acquisitions of integrases by enteric bacteria (Fig. 3).

This distribution can be explained by two hypotheses. First, that integrons could be in contact with, and even be acquired eventually by, any *E. coli* strain, but not retained. Two pressures might be at work: the presence of antibiotics

Fig. 4. Clustering of preferential genomic CUs. The CU of 159 representative prokaryotic genomes was calculated. An all-against-all comparison of such CUs was conducted to obtain an RMSD for each pair of genomic CUs. A matrix of RMSD values was used to hierarchically cluster genomes by CU similarity. Two genomes with a similar CU appear as closer branches in the tree (left). Dark regions in the clustering represent groups of genomes with similar CU (centre). The group of genomes with a CU similar to that of *E. coli* is depicted (right). Species in which integrases have been found are highlighted in bold type.

selecting some genes associated to integrons, such as sulfonamide-resistance genes, or any mobile cassette; and the CU, acting as a facilitator of the expression of integrases in the recipient genome and promoting their retention (Medrano-Soto *et al.*, 2004). As shown herein and in earlier work by others, the prevalence of class 1 integrons among *E. coli* isolates drops significantly as their origin is farther from human influence. This is not the case, however, for resistance phenotypes themselves, as can be seen here for sulfonamide resistance: while 44 % of sulfonamide-resistant strains of clinical origin were *intI1* PCR-positive, only 6 % from sewage and none from wild animals appeared to have class 1 integrons. As gene cassettes within *E. coli* integrons appear to be mostly related to antibiotic-resistance traits (even an unusual integron carrying an intron insert also carries a gene for an aminoglycoside-modifying enzyme; Sunde, 2005), it can be presumed that antibiotics exert the main pressure that selects and/or maintains integrons in *E. coli*. Interestingly, environmental *E. coli* strains that carry class 1 integrons are resistant to more antibiotics than integron-less strains, although resistance determinants are not within the integrons (Rosas *et al.*, 2006). Perhaps class 1 integrons, along with sulfonamide-resistance genes, are maintained in *E. coli* while under the selective pressure of antibiotics; but as such pressure decreases, the integron machinery is dropped by the population. This could be the result of the integron genes having CU profiles dissimilar to that preferred by the recipient genome (see below), and of the fact that most integrons in *E. coli* are plasmid-borne, unlike, for instance, the chromosomal superintegrons of vibriales. Furthermore, integrons in enteric bacteria, with the exception of *E. coli*, are apparently generally chromosomally encoded. Together, this indicates that lifting the selection pressure posed by antibiotics might not reduce the prevalence of resistance itself, but instead that of some of the mobile genetic elements by which resistance can be transferred. This can apply to the entire plasmid in which an integron resides; however, the stability of resistance plasmids even in the absence of antibiotic selective pressure has been analysed in a variety of models (reviewed by Heinemann *et al.*, 2000; Salyers & Amabile-Cuevas, 1997), starting with the now classic paper of Bouma & Lenski (1988).

The second possibility is that the sparse distribution of *intI1* PCR-positive isolates among non-human *E. coli* strains might be a consequence of the lesser ability of those strains to acquire integrons, or simply that they do not interact with integron-bearing organisms, rather than a higher loss rate. Commensal *E. coli* that are prevalent in wild animals are less prone to gather antibiotic-resistance genes (Escobar-Paramo *et al.*, 2006), but we also found fewer integrons in isolates from urban sewage and outdoor dust, which are mainly of human origin. Furthermore, integrons have been found in a wide variety of environments with and without human influences, so that these elements are readily available for non-human strains. However, as we cannot prove that our environmental

strains ever did carry integrons, both hypotheses must be tested in the future.

We find particularly interesting the nature of the inserted gene cassettes: while the few integrons found in environmental samples all carry an insert of the same size, a variety of amplicons were obtained from clinical isolates (Fig. 2). The 'standard' class 1 integron might be the most stable form of this mobile genetic element. This resembles the finding that all class 1 integrons found in isolates from healthy people not taking antibiotics carry a very narrow variety of cassettes: *dfr* and/or *aadA* genes conferring trimethoprim and streptomycin resistance, respectively (Skurnik *et al.*, 2005). Three out of four *intI1* PCR-positive strains from irrigation water also had a ~1 kb insert containing the *aadA* gene (Roe *et al.*, 2003). This contrasts, for instance, with the complex array of cassettes of integron In53, found in a hospital *E. coli* isolate, which includes seven different resistance determinants (Naas *et al.*, 2001).

Our evolutionary analyses highlight the fact that integrases in enteric bacteria seem to come from different origins, because they fall into at least three, and possibly up to five classes of integrases. Also, the wide distribution of XerC/D recombinases, almost universal in bacteria and archaea, contrasts with the sparse distribution of integrases in a few bacterial species, and suggests that the former were the ancestors of the latter. Various types of integrases from vibriales appear in the most basal branches of the tree, supporting the idea that superintegron-integrases from vibriales represent the most ancestral stages in the evolution of integrases (Rowe-Magnus *et al.*, 2001).

Given the distribution of *IntI1* among phylogenetically distant genomes, the next step in our analysis was to try to identify biological signals that allow this distribution. Unfortunately, the comparison of *intI1* and genomic CUs did not show a clear trend with respect to the possible origin of this gene, because similar CAI values were obtained when comparing the CU of *intI1* among different phylogenetic groups. However, an interesting observation emerged from this comparison, in that some phylogenetically distantly related species bearing *intI1* possess a similar preferential genomic CU. As suggested by earlier reports (Medrano-Soto *et al.*, 2004), CU can impose a functional constraint in the retention of horizontally transferred genes, independently of their origin. In the case of *intI1*, CU could act together with the selective pressure of antibiotics to retain integrons in recipient genomes. The *intI1* integrase gene appears, however, to be more related to those of bacteria other than *E. coli*, judging from the CU pattern: the CU might be similar enough to allow the acquisition and expression of *intI1* by *E. coli*, but not similar enough to allow its retention in the absence of antibiotics. Interestingly, one of the organisms to which the *intI1* CU is more alike is a species of *Corynebacterium*; 14 out of 23 species of Gram-positive bacteria bearing class 1 integrons were *Corynebacterium* (Nandi *et al.*, 2004). Genes encoding putative class 1 integrases have been found in the

chromosomes of *Corynebacterium* spp. and other species of similar preferential genomic CU. In contrast, *intI1* has been found only in plasmids in *E. coli*. It would be interesting to look for class 1 integrons in other bacterial species with a CU similar to that of *Corynebacterium*.

In conclusion, class 1 integrons in *E. coli* could be seen as transient elements that foster antibiotic resistance in clinical environments, but not in the absence of antibiotic selective pressure. The bioinformatic analyses presented herein support the idea that *E. coli* and other enteric bacteria have acquired and lost integrases multiple times during evolution. It is possible that non-clinical *E. coli* strains are less capable of gaining class 1 integrons, but it is also possible that decreased antibiotic pressure, along with the differences between the CU of *intI1* and that preferred by *E. coli*, and the plasmid-borne nature of class 1 integrons in *E. coli*, make these genetic elements unstable. If that is so, a reduced antibiotic usage may lead to a significant decrease in the carriage of integrons among *E. coli* strains.

ACKNOWLEDGEMENTS

We thank Leticia Martínez, Eva Salinas, Ariadna Cruz, Andrea González, Laura Espinosa-Asuar and Antonio Cruz for technical support, and Peter Gogarten for his helpful comments in the preparation of the manuscript.

REFERENCES

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- Amábile-Cuevas, C. F. & Chicurel, M. E. (1992). Bacterial plasmids and gene flux. *Cell* **70**, 189–199.
- Barlow, R. S., Pemberton, J. M., Desmarchelier, P. M. & Gobijs, K. S. (2004). Isolation and characterization of integron-containing bacteria without antibiotic selection. *Antimicrob Agents Chemother* **48**, 838–842.
- Bennett, P. M. (1999). Integrons and gene cassettes: a genetic construction kit for bacteria. *J Antimicrob Chemother* **43**, 1–4.
- Boucher, Y., Labbate, M., Koenig, J. E. & Stokes, H. W. (2007). Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends Microbiol* **15**, 301–309.
- Bouma, J. E. & Lenski, R. E. (1988). Evolution of a bacteria/plasmid association. *Nature* **335**, 351–352.
- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003). Multiple sequence alignment with the CLUSTAL series of programs. *Nucleic Acids Res* **31**, 3497–3500.
- Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics* **14**, 755–763.
- Escobar-Paramo, P., Le Menac'h, A., Le Gall, T., Amorin, C., Gouriou, S., Picard, B., Skurnik, D. & Denamur, E. (2006). Identification of forces shaping the commensal *Escherichia coli* genetic structure by comparing animal and human isolates. *Environ Microbiol* **8**, 1975–1984.
- Felsenstein, J. (2005). PHYLIP (phylogeny inference package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, USA.
- Firth, N. (2003). Evolution of antimicrobial multi-resistance in Gram-positive bacteria. In *Multiple Drug Resistant Bacteria*, pp. 33–60. Edited by C. F. Amábile-Cuevas. Wymondham, UK: Horizon Scientific Press.
- Frickey, T. & Lupas, A. (2004). CLANS: A Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20**, 3702–3704.
- García-Vallve, S., Guzman, E., Montero, M. A. & Romeu, A. (2003). HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. *Nucleic Acids Res* **31**, 187–189.
- Goldstein, C., Lee, M. D., Sanchez, S., Hudson, C., Phillips, B., Register, B., Grady, M., Liebert, C., Summers, A. O. & other authors (2001). Incidence of class 1 and 2 integrases in clinical and commensal bacteria from livestock, companion animals, and exotics. *Antimicrob Agents Chemother* **45**, 723–726.
- Hall, R. M., Recchia, G. D., Collis, C. M., Brown, H. J. & Stokes, H. W. (1996). Gene cassettes and integrons: moving antibiotic resistance genes in Gram-negative bacteria. In *Antibiotic Resistance: from Molecular Basics to Therapeutic Options*, pp. 19–34. Edited by C. F. Amábile-Cuevas. New York: Chapman & Hall.
- Heinemann, J. A., Ankenbauer, R. G. & Amábile-Cuevas, C. F. (2000). Do antibiotics maintain antibiotic resistance? *Drug Discov Today* **5**, 195–204.
- Kanehisa, M. & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30.
- Leverstein-Van Hall, M. A., Paaauw, A., Box, A. T., Blok, H. E., Verhoef, J. & Fluit, A. C. (2002). Presence of integron-associated resistance in the community is widespread and contributes to multidrug resistance in the hospital. *J Clin Microbiol* **40**, 3038–3040.
- Lévesque, C. & Roy, P. H. (1993). PCR analysis of integrons. In *Diagnostic Molecular Microbiology, Principles and Applications*, pp. 590–602. Edited by D. H. Persing, T. F. Smith, F. C. Tenover & T. J. White. Washington, DC: American Society for Microbiology.
- MacDonald, D., Demarre, G., Bouvier, M., Mazel, D. & Gopaul, D. N. (2006). Structural basis for broad DNA-specificity in integron recombination. *Nature* **440**, 1157–1162.
- McGuffin, L. J., Bryson, K. & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405.
- Medrano-Soto, A., Moreno-Hagelsieb, G., Vinuesa, P., Christen, J. A. & Collado-Vides, J. (2004). Successful lateral transfer requires codon usage compatibility between foreign genes and recipient genomes. *Mol Biol Evol* **21**, 1884–1894.
- Messier, N. & Roy, P. H. (2001). Integron integrases possess a unique additional domain necessary for activity. *J Bacteriol* **183**, 6699–6706.
- Michael, C. A., Gillings, M. R., Holmes, A. J., Hughes, L., Andrew, N. R., Holley, M. P. & Stokes, H. W. (2004). Mobile gene cassettes: a fundamental resource for bacterial evolution. *Am Nat* **164**, 1–12.
- Naas, T., Mikami, Y., Imai, T., Poirel, L. & Nordmann, P. (2001). Characterization of In53, a class 1 plasmid- and composite transposon-located integron of *Escherichia coli* which carries an unusual array of gene cassettes. *J Bacteriol* **183**, 235–249.
- Nandi, S., Maurer, J. J., Hofacre, C. & Summers, A. O. (2004). Gram-positive bacteria are a major reservoir of class 1 antibiotic resistance integrons in poultry litter. *Proc Natl Acad Sci U S A* **101**, 7118–7122.
- Nemergut, D. R., Martin, A. P. & Schmidt, S. K. (2004). Integron diversity in heavy-metal-contaminated mine tailings and inferences about integron evolution. *Appl Environ Microbiol* **70**, 1160–1168.
- Nield, B. S., Holmes, A. J., Gillings, M. R., Recchia, G. D., Mabbutt, B. C., Nevalainen, K. M. & Stokes, H. W. (2001). Recovery of new integron classes from environmental DNA. *FEMS Microbiol Lett* **195**, 59–65.

- Notredame, C., Higgins, D. G. & Heringa, J. (2000).** T-COFFEE: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205–217.
- Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000).** Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.
- Partridge, S. R., Brown, H. J., Stokes, H. W. & Hall, R. M. (2001).** Transposons Tn1696 and Tn21 and their integrons In4 and In2 have independent origins. *Antimicrob Agents Chemother* **45**, 1263–1270.
- Rice, P., Longden, I. & Bleasby, A. (2000).** EMBOS: The European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276–277.
- Roe, M. T., Vega, E. & Pillai, S. D. (2003).** Antimicrobial resistance markers of class 1 and class 2 integron-bearing *Escherichia coli* from irrigation water and sediments. *Emerg Infect Dis* **9**, 822–826.
- Rosas, I., Salinas, E., Martinez, L., Calva, E., Cravioto, A., Eslava, C. & Amábile-Cuevas, C. F. (2006).** Urban dust fecal pollution in Mexico City: antibiotic resistance and virulence factors of *Escherichia coli*. *Int J Hyg Environ Health* **209**, 461–470.
- Rowe-Magnus, D. A., Guerout, A. M., Ploncard, P., Dychinco, B., Davies, J. & Mazel, D. (2001).** The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. *Proc Natl Acad Sci U S A* **98**, 652–657.
- Salyers, A. A. & Amábile-Cuevas, C. F. (1997).** Why are antibiotic resistance genes so resistant to elimination? *Antimicrob Agents Chemother* **41**, 2321–2325.
- Skurnik, D., Le Menac'h, A., Zurakowski, D., Mazel, D., Courvalin, P., Denamur, E., Andremont, A. & Ruimy, R. (2005).** Integron-associated antibiotic resistance and phylogenetic grouping of *Escherichia coli* isolates from healthy subjects free of recent antibiotic exposure. *Antimicrob Agents Chemother* **49**, 3062–3065.
- Souza, V., Rocha, M., Valera, A. & Eguarte, L. E. (1999).** Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. *Appl Environ Microbiol* **65**, 3373–3385.
- Sunde, M. (2005).** Class I integron with a group II intron detected in an *Escherichia coli* strain from a free-range reindeer. *Antimicrob Agents Chemother* **49**, 2512–2514.

Edited by: J. Parkhill