



**HAL**  
open science

# Unmasking the ancestral activity of integron integrases reveals a smooth evolutionary transition during functional innovation

Jose Antonio Escudero, Celine Loot, Vincent Parissi, Aleksandra Nivina, Christiane Bouchier, Didier Mazel

## ► To cite this version:

Jose Antonio Escudero, Celine Loot, Vincent Parissi, Aleksandra Nivina, Christiane Bouchier, et al.. Unmasking the ancestral activity of integron integrases reveals a smooth evolutionary transition during functional innovation. *Nature Communications*, 2016, 7, pp.Article number 10937. 10.1038/ncomms10937 . pasteur-01292328

**HAL Id: pasteur-01292328**

**<https://pasteur.hal.science/pasteur-01292328>**

Submitted on 22 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

ARTICLE

Received 5 Oct 2015 | Accepted 3 Feb 2016 | Published 10 Mar 2016

DOI: 10.1038/ncomms10937

OPEN

# Unmasking the ancestral activity of integron integrases reveals a smooth evolutionary transition during functional innovation

Jose Antonio Escudero<sup>1,2</sup>, Celine Loot<sup>1,2</sup>, Vincent Parissi<sup>3</sup>, Aleksandra Nivina<sup>1,2,4</sup>, Christiane Bouchier<sup>5</sup> & Didier Mazel<sup>1,2</sup>

Tyrosine (Y)-recombinases have evolved to deliver mechanistically different reactions on a variety of substrates, but these evolutionary transitions are poorly understood. Among them, integron integrases are hybrid systems recombining single- and double-stranded DNA partners. These reactions are asymmetric and need a replicative resolution pathway, an exception to the canonical second strand exchange model of Y-recombinases. Integron integrases possess a specific domain for this specialized pathway. Here we show that despite this, integrases are still capable of efficiently operating the ancestral second strand exchange in symmetrical reactions between double-stranded substrates. During these reactions, both strands are reactive and Holliday junction resolution can follow either pathway. A novel deep-sequencing approach allows mapping of the crossover point for the second strand exchange. The persistence of the ancestral activity in integrases illustrates their robustness and shows that innovation towards new recombination substrates and resolution pathways was a smooth evolutionary process.

<sup>1</sup>Institut Pasteur, Unité de Plasticité du Génome Bactérien, Département Génomes et Génétique, 28 Rue du Dr Roux, 75015 Paris, France. <sup>2</sup>CNRS, UMR3525, 28 Rue du Dr Roux, 75015 Paris, France. <sup>3</sup>Laboratoire de Microbiologie Fondamentale et Pathogénicité UMR-CNRS 5234. Bat 3A 3ème étage, 146 rue Léo Saignat, 33076 Bordeaux, France. <sup>4</sup>Université Paris Descartes, Sorbonne Paris Cité, 75006 Paris, France. <sup>5</sup>Institut Pasteur, Genomic Platform, 28 Rue du Dr Roux, 75015 Paris, France. Correspondence and requests for materials should be addressed to D.M. (email: mazel@pasteur.fr).

Tyrosine (Y)-recombinases mediate recombination between two specific DNA sites to deliver a variety of biological functions (for a review see (ref. 1)). They are present in the three domains of life but are most widespread among prokaryotes, where they play fundamental roles such as the integration/excision of viral genomes<sup>2</sup>, the alternation of gene expression<sup>3</sup> and the resolution of deleterious chromosome dimers arising during replication<sup>4</sup>. Members of this family show a conserved fold of the catalytic domain including a three-dimensional clustering of highly conserved RKHRH residues. Y-recombinases typically recognize specific DNA sites comprising two inverted binding domains separated by a 6–8 bp spacer. The recombination reaction is a stepwise process that starts with the cleavage and transfer of one strand, leading to the transient formation of a Holliday junction (HJ) that is further resolved through a second exchange of strands. This sequence of steps, which we will refer to as the classical pathway (Fig. 1a), is highly conserved among Y-recombinases. Type IB topoisomerases and telomere resolvases share the fold and catalytic chemistry, and can therefore be considered as distant relatives within the same superfamily, despite sequence divergence<sup>1,5–7</sup>. This implies that recombinases have evolved to deliver mechanistically distinct reactions (for example, the cleavage of only one strand by topoisomerases), and to recognize structurally different substrates (for example, hairpins in telomeres). Such evolvability is likely facilitated by the robustness of the fold<sup>6</sup> (its capacity to accommodate mutations while preserving its tertiary structure), as well as the modularity of protein function through the combination of different functional domains. Nevertheless, information on evolutionary pathways that enable functional innovation among these enzymes is scarce.

Integrations are recombination platforms that allow bacteria to evolve and adapt rapidly through the acquisition, stockpiling and reordering of genes embedded in mobile genetic elements named integron cassettes<sup>8–10</sup> (Fig. 1b). They are ancient structures that have driven bacterial evolution for hundreds of millions of years<sup>11</sup>, and, through their mobilization onto conjugative plasmids, they have played a critical role in the rise of multidrug resistance among clinically relevant bacteria<sup>12</sup>. Recombination in integrons is governed by integron integrases<sup>13</sup> (hereafter integrases), a distinct group of Y-recombinases closely related to XerCD recombinases<sup>9</sup>. Integrases show the conserved fold of the family, as well as the RKHRH pentad of residues in the active site (Fig. 1c). Incorporation of cassettes into the integron platform occurs through site-specific recombination between the *attC* site in the cassette and the *attI* site in the integron<sup>14</sup>. Genes in cassettes are then expressed from the *P<sub>c</sub>* promoter within the integron platform (Fig. 1b). Cassette recombination represents an exception to the archetypical reaction mediated by Y-recombinases, as a consequence of the structural peculiarities of *attC* sites. These sites show no sequence conservation, except for the crossover point on the bottom strand (bs) (5'-AAC-3'), that provides microhomology with *attI* sites, necessary for strand transfer<sup>15</sup>. Instead, *attC* sites share a conserved palindromic structure, and are only recognized as a folded single strand (ss)<sup>16</sup> (Fig. 1d). Recombination between a ss-*attC* and the double strand (ds)-*attI* site forms an asymmetric HJ (aHJ) (Fig. 1a) that cannot be resolved through a second exchange of strands since it would linearize the replicon and be abortive<sup>17</sup>. During *attC* site folding, the imperfect pairing of both arms of the bs allows for the protrusion of a set of extrahelical bases (EHBs) (Fig. 1d) that are key in strand selectivity, favouring the recombination of the bs<sup>18,19</sup>. EHBs also induce conformational changes in one integrase monomer during the reaction, pulling apart the tyrosine residue and avoiding the nucleophilic attack on the L box of the *attC* site<sup>20</sup> (Fig. 1d). This prevents the second strand exchange,

and the aHJ is then resolved through replication<sup>21</sup> (Fig. 1a). To accommodate EHBs, integrases possess a 20-residue-long domain dubbed I2 (ref. 22; Fig. 1c) that has not been identified elsewhere, suggesting that it could have emerged recently<sup>23</sup> among Y-recombinases. Such domain insertions are well-characterized phenomena in protein evolution<sup>24</sup>, allowing for a modular organization of protein function.

ssDNA is central to the integron, also controlling integrase expression through the SOS response<sup>25</sup>. Integrons are, therefore, unique recombination platforms streamlined to recombine ssDNA through a distinct resolution pathway. They represent a molecular example of functional innovation<sup>26</sup>.

While *attC* sites have been extensively studied<sup>17,19,20,27,28</sup>, knowledge is lacking on the recognition and processing of *attI* sites. These are more canonical than *attC* sites, since they are recognized as dsDNA<sup>16</sup>. They have two binding domains, the L and R boxes, although one is degenerate (Fig. 1d), and are generally recognized specifically by their cognate integrases<sup>29</sup>. Some *attI* sites possess accessory binding domains<sup>30</sup>.

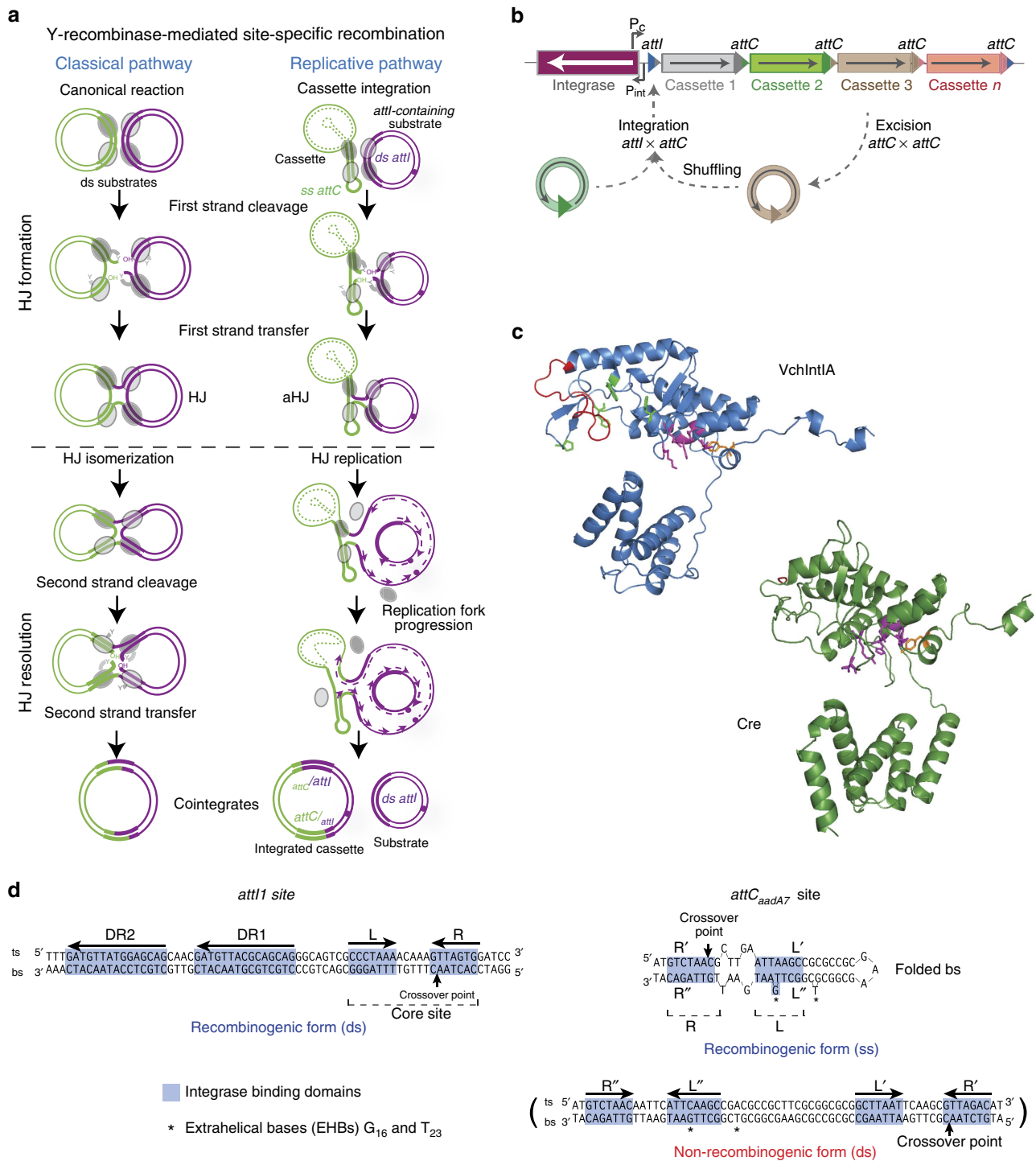
Integrase site recognition underscores a remarkable conformational dynamism<sup>31</sup>. Identical monomers recognize ds-*attI* sites by sequence and ss-*attC* sites through structural landmarks. From an evolutionary standpoint, this duality could reflect a single specialized activity (the recombination of ss/ds DNA partners) or two distinct ones: the canonical Y-recombinase activity of integrase ancestors on *attI*, and the functional innovation, that is, the recombination of ssDNA structures, on *attC* sites. For the latter to be true, both classical and innovative activities must be complete: integrases should be capable of resolving HJs through a second strand exchange.

To unmask the ancestral activity of integrases, we have studied the reaction between two *attI* sites. This is a symmetrical reaction involving exclusively ds partners for which resolution through a second strand exchange is theoretically possible, since it is neither abortive nor impeded by EHBs. We show that the top strand of *attI* is reactive and that HJ resolution can take place through a second strand exchange. We have determined the crossover point through a novel deep-sequencing approach that is applicable to other recombination systems, and validated this approach *in vitro*. Finally, we prove that *attC* sites govern the directionality of the reaction. To our knowledge, integrases are the first example among recombinases of a smooth evolutionary transition between substrates and processing pathways. We discuss the possible origins and the evolutionary implications of the dual activity of integrases, as well as the importance of *attI* × *attI* reactions in antibiotic resistance.

## Results

For this work we chose to study *attI*, the integration site of the most widespread and best-studied integron platform, the Class 1 integron.

**Dispensability of known HJ-processing host factors.** The *attI* site possesses two direct repeats upstream the L–R boxes that make the site ~65-bp long<sup>30</sup>. It is therefore possible that host factors specialized in homologous recombination, branch migration and resolution of HJ-like structures (such as stalled replication forks), could influence the resolution of the *attI* × *attI* HJ, biasing our results. To assess the impact of such host machinery, we have tested the influence of RecA and two well-characterized HJ-processing complexes, RecG<sup>32</sup> and RuvABC<sup>33,34</sup>, using a classic suicide conjugation assay<sup>35</sup>. Recombination rates of mutants of these complexes were not statistically different from those of the wild-type (WT) strain, suggesting that known host HJ-processing systems do not

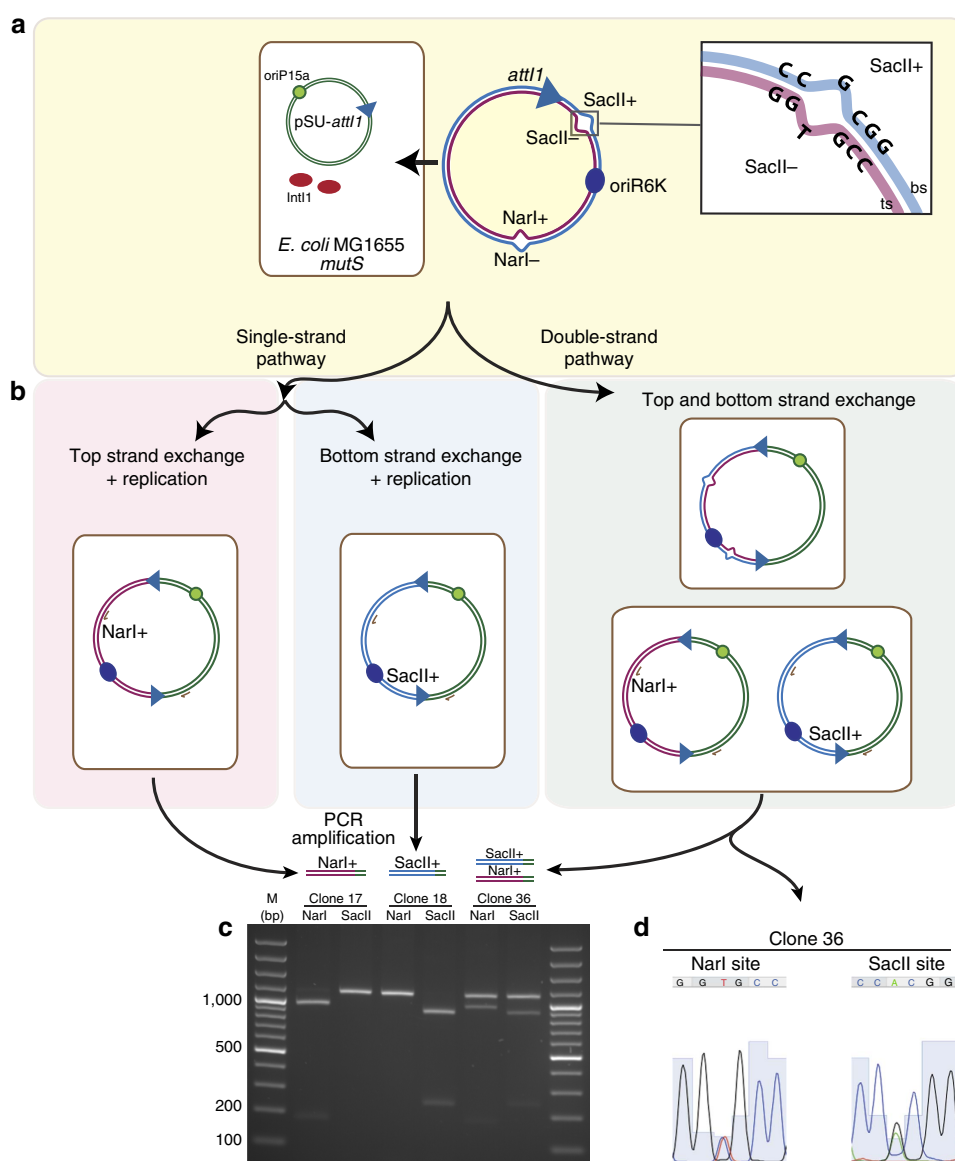


**Figure 1 | Recombination in the integron.** (a) Diagram showing the two recombination models among tyrosine recombinases. The left column shows the recombination pathway shared by a majority of Y-recombinases, except the integron integrase for which the pathway is represented in the right column. (b) Schematic representation of the integron architecture and recombination reactions. Open reading frames are represented as boxes, with arrowheads showing the direction of transcription. Arrows marked as P<sub>c</sub> and P<sub>int</sub> represent the promoters of the cassette array and the integrase, respectively. Acquisition of new cassettes involves the recombination of the integron attI site and the cassette attC site, while cassette excision involves two attC sites. Sequential coupling of excision and integration reactions shuffles the order of cassettes within the array. (c) Crystal structure of *V. cholerae* integrase<sup>20</sup> (blue) and Cre<sup>60</sup> (green). RKHR residues are depicted as purple sticks. The catalytic tyrosine is shown in orange. Red residues represent the I2 domain in VchIntI and the residues in Cre that align with those delimiting I2 in VchIntI. Residues in VchIntI that interact with EHBs are represented in green. (d) Sequence of the Class 1 integron attI site (attI), and the attC<sub>aadA7</sub> site. aHJ, asymmetric HJ; bs, bottom strand; DR, direct repeats; ds, double stranded; HJ, Holliday junction; ss, single stranded; ts, top strand.

significantly influence the *attI1* × *attI1* reaction (Supplementary Fig. 1).

**Both strands of *attI1* are reactive.** In cassette integration reactions, only the bottom strands of *attC* and *attI* sites are exchanged, with the crossover point located within the R boxes of both sites. To test the hypothesis that the *attI1* × *attI1* reaction can take place through both the bottom and the top strand, we used an *attI1* carried on a mismatched covalent circle<sup>21</sup>. These synthetic circles mimic non-replicative plasmids assembled using complementary strands produced separately and hybridized together *in vitro*. This allowed us to specifically tag each strand with a mutation in a restriction site (SacII or NarI) so that the restriction pattern of cointegrates reveals which strand is transferred during recombination (Fig. 2a). In our setting, the transfer and replication of the bottom and the top strands leads,

respectively, to a SacII<sup>+</sup>/NarI<sup>-</sup> and SacII<sup>-</sup>/NarI<sup>+</sup> restriction pattern. The classical pathway of double-strand exchange would yield cointegrates containing both mismatches that, once replicated, would produce a mixed population of plasmids and hence partial restriction profiles with both enzymes (Fig. 2b). We verified the correct assembly of the DNA circles (as in ref. 21 (Supplementary Fig. 2)) and transformed them into a *mutS*<sup>-</sup> MG1655 *Escherichia coli* (*E. coli*) strain containing an *attI*-bearing pSU plasmid (p929) as well as a plasmid with the integrase under the control of a P<sub>BAD</sub> promoter (pBAD-*intI1*). The *mutS*<sup>-</sup> background ensures that the mismatches are not repaired. By selecting for the circle resistance marker, we recovered clones in which these molecules had recombined with the resident pSU plasmid through an integrase-mediated *attI1* × *attI1* reaction. The mismatch-containing region was PCR amplified from 100 recombinant clones and used for the restriction pattern analysis (Fig. 2c), revealing that both strands of



**Figure 2 | The top strand of *attI1* is reactive.** (a) Experimental set-up used for mismatched-covalent-circles recombination. Mismatched circles bearing an *attI1* site are transformed in an *E. coli* MG1655 *mutS*<sup>-</sup> strain bearing a pSU-*attI1* plasmid and expressing the integrase in trans. (b) Representation of the recombination products resulting from the three possible HJ resolution pathways. The mismatch-bearing regions were PCR amplified and subjected to NarI and SacII restriction. (c) Restriction pattern analysis of three clones obtained, representing the three resolution pathways. (d) Sequencing of clone 36 PCR product revealed the expected mismatches.

the *attI* site are reactive and can be transferred independently, without a clear bias on strand specificity. This is contrary to what is observed for the *attI1* × *attC* reaction where only the bs of *attI1* is reactive<sup>17</sup>. Some clones showed partial restriction patterns and mixed plasmid populations (Fig. 2d), suggesting the possibility of HJ resolution through second strand exchange. Nevertheless, it could not be ruled out that this pattern was the consequence of two independent recombination reactions involving two different plasmids.

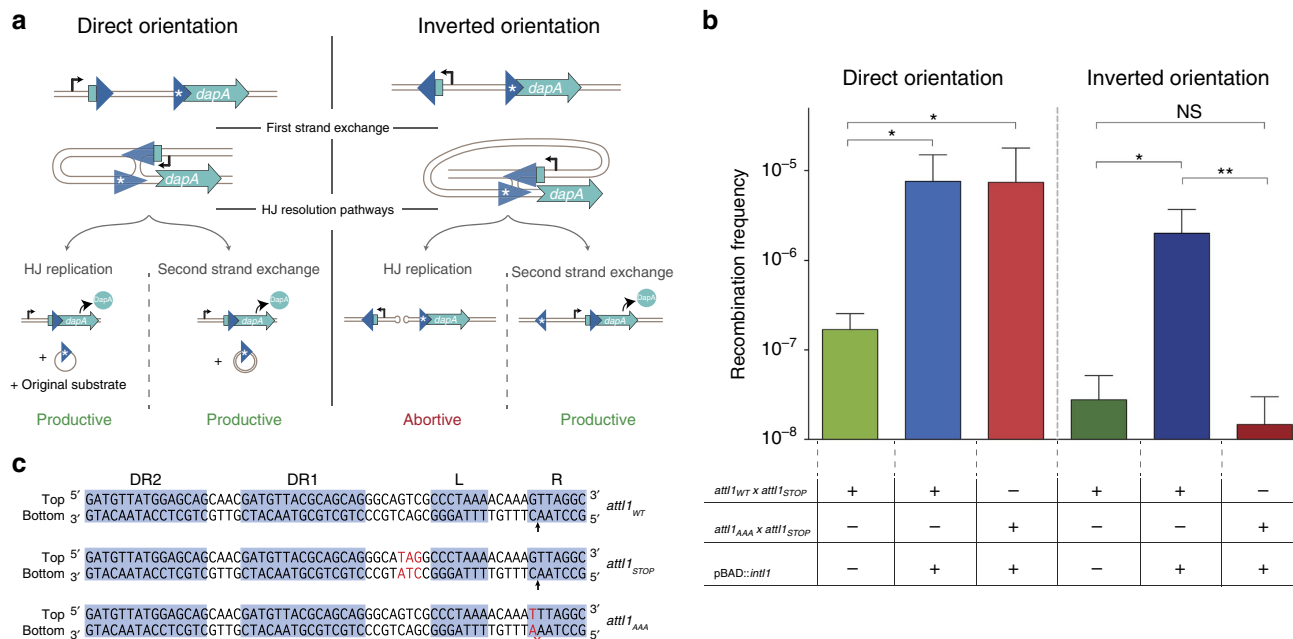
**The integrase resolves HJs through a second strand exchange.**

To further assess if both strands of the *attI1* site could be transferred during the same reaction, we used a modified version of the *dapA* recombination reporter<sup>36</sup>. In this system, recombination between *attI1* sites leads to reconstitution of the essential gene *dapA*, allowing recombinants to grow on media lacking diaminopimelic acid (DAP, the reticulating agent of peptidoglycan in *E. coli*). Comparison of the number of clones growing with and without DAP yields a recombination frequency for a given reaction. We produced two versions of this reporter system, one with two *attI1* sites in direct orientation, and the other with sites in inverted orientation. Both structures were inserted ectopically in single copy in the chromosome of an *E. coli* MG1655  $\Delta$ *dapA* strain. The rationale of this experiment is that, when both sites are in inverted orientation, viable recombinants can only be recovered if HJ resolution takes place through a second strand exchange that inverts the region between both sites. In this orientation the replicative resolution pathway linearizes the chromosome and is hence abortive (Fig. 3a; Supplementary Movies 1 and 2). On the other hand, when sites are in direct

orientation both resolution pathways result in the excision of the region between the sites and are hence productive, serving as a control for the experiment. Our results show that integrase expression increased significantly the yield of recombinant clones for sites in direct orientation (44-fold) but also for sites in inverted orientation (72-fold) (Fig. 3b) (verified by PCR). This demonstrates that the *attI1* × *attI1* HJ can be resolved through a second strand exchange, that is, through the classical Y-recombinase pathway. Furthermore, comparison of the increases in recombination frequencies for both orientations proves that the classical resolution pathway is extremely efficient.

To avoid expression of *dapA* before recombination, one *attI1* site was modified 1 bp upstream of the L box to include a stop codon (*attI<sub>STOP</sub>*) (Fig. 3c). As a consequence, resolution through branch migration is not possible, suggesting again that host factors were not necessary in the reaction. It implies too that resolution of the HJ had taken place not further than 1 bp upstream of the L box, as confirmed by the change of location of the stop codon after recombination.

A similar experiment was performed using two *attC* sites, *attC<sub>aadA7</sub>* and *attC<sub>ereA2</sub>* (Supplementary Fig. 3). When sites were in direct orientation, we observed an increase in recombination frequencies of five orders of magnitude due to the presence of the integrase. Contrary to our observations for the *attI1* × *attI1* reaction, no clear increase in recombination frequency was observed for *attC* sites in inverted orientation. Sequencing of the few events (*n* = 8) in which the central structure was found inverted showed that the site had not been recombined in a double-stranded form. Indeed, the resulting sites were largely conserved, with the presence of moieties of the original sites only visible after the canonical 5'-AAC-3' cleavage point. Although



**Figure 3 | The integrase can resolve HJs through second strand exchange.** (a) Graphic representation of the structural set-up of our experiment in which an *attI1* × *attI1* reaction reconstitutes the *dapA* gene, with *attI1* sites in a direct orientation (left column) and in an inverted orientation (right column). Asterisks indicate *attI* sites modified to include a STOP codon (*attI<sub>STOP</sub>*). The topology of HJs resulting from the first strand exchange is shown for both orientations. Recombination products following the two possible HJ resolution pathways are shown below. (b) Recombination frequency of all set-ups. Columns represent the mean and error bars the s.d. (*n* ≥ 3). Significance testing was performed with Kolmogorov–Smirnov test (\**P* value < 0.05; \*\**P* value < 0.01; NS, not significant). For the direct orientation set-up, strains used are B36, B82 and C162 (first, second and third column, respectively). For the inverted orientation set-up, strains used are B37, B83 and C163 (fourth, fifth and sixth column, respectively); their relevant genotypes are described in the table beneath the graph. (c) Sequence of the *attI* sites used in this experiment: *attI<sub>WT</sub>* (the cleavage point on the bottom strand is marked with an arrow), *attI<sub>STOP</sub>*, a variant not encoding an ORF, that allows to reduce background noise in the experiment while remaining recombinogenic; and *attI<sub>AAA</sub>*, a variant in which the bottom strand is inactive.

these inversion events require a second cleavage that could not be elucidated, and therefore seem to be the consequence of illegitimate recombination reactions (possibly between sites on different replicating chromosomes), they are strong proof that *attC* sites are exclusively recombined in their structured form and confirms previous observations<sup>17</sup>.

To further prove that the second strand exchange occurs and is mediated by the integrase, we explored the possibility of impeding the bottom strand recombination. To do so, we mutated the cleavage point in the R box of one *attII* site in direct and inverted orientation from 5'-AAC-3' to 5'-AAA-3' (see *attII<sub>AAA</sub>* in Fig. 3c). It has been observed that a C to A modification drops recombination rates for *attII* with both *attC* and *attII* partners<sup>15,37</sup>. Our results show a complete abolition of recombination when both sites are in the inverted orientation (that is, when the second strand exchange is needed to resolve the HJ), proving that the bs is no longer reactive. In contrast, recombination rates remain constant when the sites are in direct orientation, proving that integrase-mediated recombination through the ts is still possible for the *attII<sub>AAA</sub>* site, and supporting the previously observed lack of clear bias in strand selection.

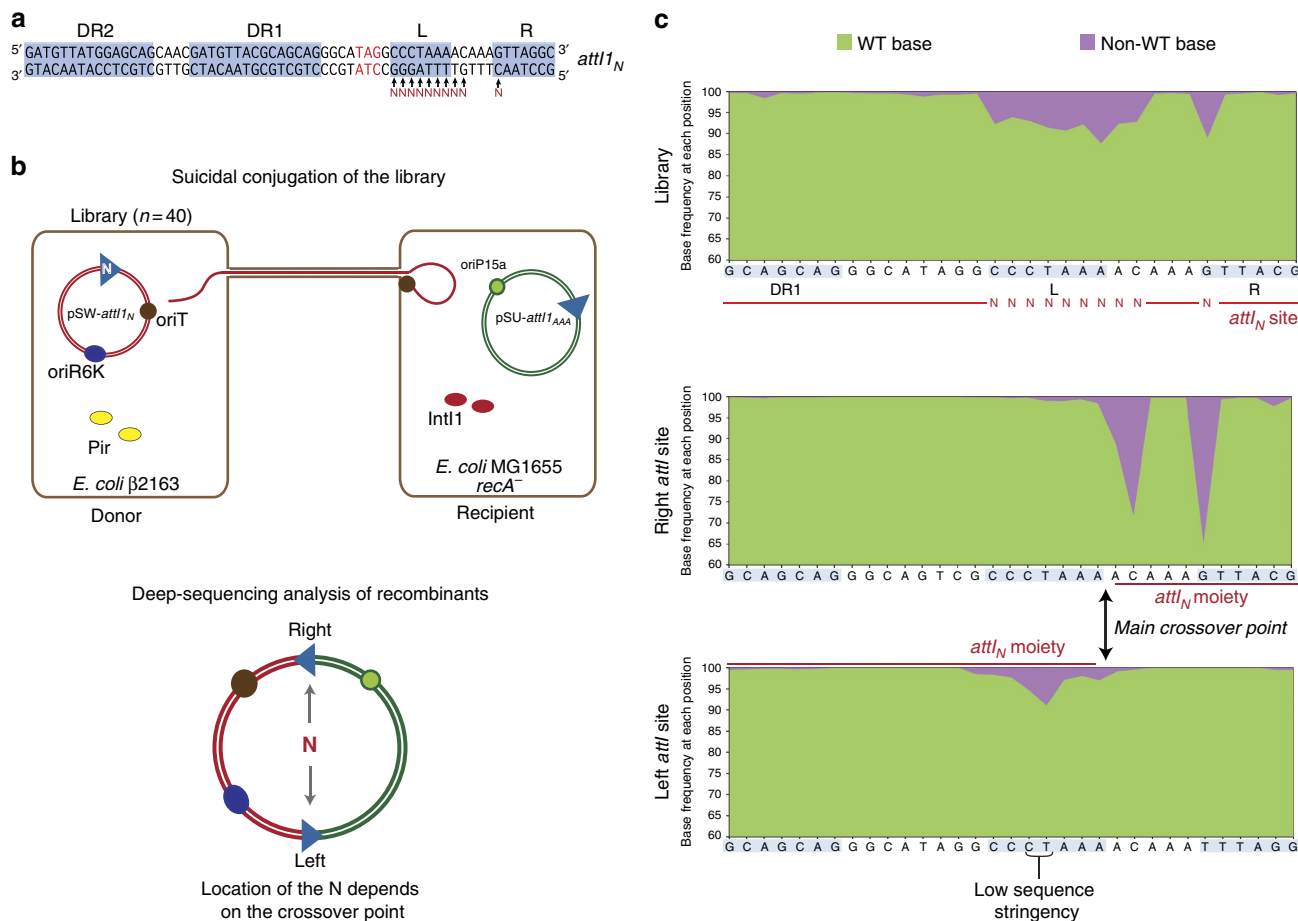
**Mapping the ts crossover point using deep sequencing.** Recombination between sites on different replicons results in co-integrate molecules bearing two sites. Each of them is composed of two moieties of the initial sites, but when the latter are identical, moieties are not easily identified and the crossover point of the reaction remains cryptic. Such is the case of the *attII* × *attII* reaction. Hence, to map the crossover point on the top strand of *attII*, we built a library of *attII* sites in which we randomized all base pairs in the L box, the two adjacent base pairs within the spacer region and the C in the crossover point of the R box (Fig. 4a). Using N-containing primers, 10 *attII<sub>N</sub>* site sets were produced, each bearing random bases in one of the ten positions. These sites were cloned in *pir*-dependent pSW plasmids and mixed to form a small library ( $n = 40$ ) of pSW-*attII<sub>N</sub>* plasmids to be used in our suicide conjugation assay. The rationale of this experiment is that the pattern of segregation of random bases between the two resulting sites in the co-integrate molecule will reveal the precise crossover point in the *attII* top strand (L box). Furthermore, since base composition of the L box can influence IntI1 binding, some *attII* mutants might be poorly recognized. A skew in the proportion of bases in a given position after recombination would underscore the importance of a specific base in the binding process (Fig. 4b). The shortcoming of this experimental set-up is that, as seen before (Fig. 2), the reaction can take place exclusively through the bottom strand (bs-exchange plus replication), without splitting the randomized L box. To overcome this, having confirmed the lack of bs exchange for *attII<sub>AAA</sub>*, we used a pSU-*attII<sub>AAA</sub>* plasmid as recombination partner for the library, forcing recombination to occur at the randomized L box. Deep sequencing of pSW-*attII<sub>N</sub>* before recombination confirmed the expected distribution of 7.5% mutated bases in the randomized positions (explained in Methods) (Fig. 4c). Sequencing of the resulting left and right sites in the co-integrate revealed a complementary distribution of random bases among sites, allowing location of the crossover point in the L box between both adenines of the 5'-AAC-3' triplet (Fig. 4c). It also allowed us to infer the importance of each base for protein–DNA interactions, revealing lower sequence stringency in the central 5'-CT-3' base pairs of the L box (5'-CCCTAAA-3'), in accordance with previous interference data for the central T base<sup>38</sup>. To confirm the crossover position, we performed the *attII* × *attII* reaction *in vitro* using purified

MBP-IntI1 protein, which is known to be fully active *in vivo*<sup>39</sup>, and *attII* fragments bearing a mismatch between the top and the bottom strand in the adenine that is transferred during the reaction (ts sequence 5'-ATC-3', bs sequence 5'-GTT-3') (Fig. 5a). Such a mismatch eliminates the microhomology necessary for docking of the transferred strand, and should abolish the transfer of the top strand. While the reaction between two WT *attII* sites is poorly detected *in vitro*<sup>40</sup> (likely because HJ intermediates are quickly resolved), the reaction between L-box-mismatched substrates led to a significant accumulation of HJs (Fig. 5b), proving the role of the L-box cleavage in the resolution of these structures, and the dispensability of host factors in the reaction. The complementary effect (that is, the accumulation of HJs due to an R-box mismatch) was not observed, probably because mismatches in the R box led to a significant decrease in integrase binding (Supplementary Fig. 4a). The fact that binding is only affected by R-box mismatches is in accordance with the sequential and cooperative binding of monomers, starting on the R box<sup>38</sup>. As a control, all reactions were performed in parallel with the catalytically inactive version of the integrase, IntI1<sub>Y312F</sub>, for which no transfer activity was observed.

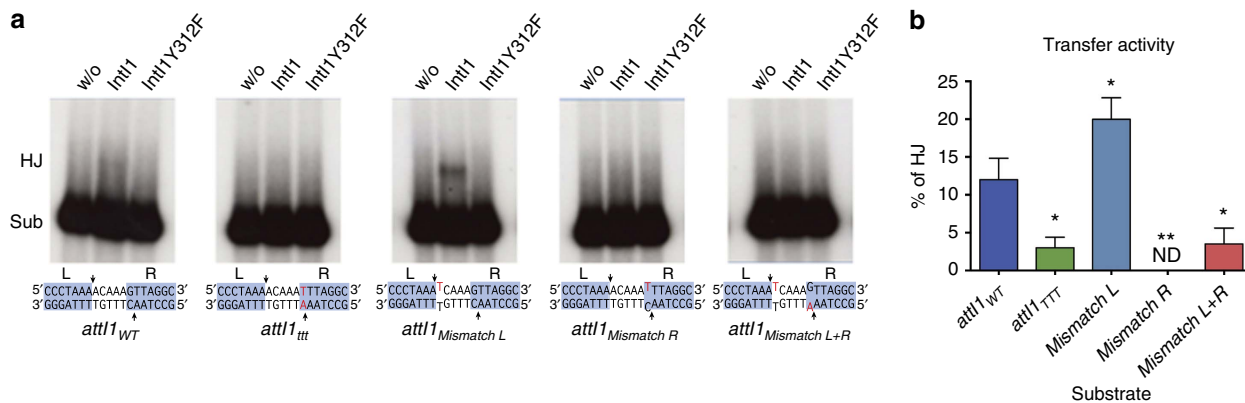
***attC* sites govern the directionality of the reaction.** During *attC* site folding, the inverted repeats of the dsDNA molecule hybridize to form the R and L boxes in the hairpin (Fig. 1d). These boxes are present in hairpins arising from the folding of both the top and bottom strands. Despite a clear bias towards bs recombination, top strands are also recombined, although at much lower frequencies. In all cases, the crossover in the *attII* × *attC* reaction takes place in the R boxes of both sites. This means that top-strand recombination integrates cassettes opposite to the Pc promoter and are likely non-functional. Therefore, the similar frequency at which L and R boxes of *attII* recombine in the *attII* × *attII* reaction is in clear opposition to that found for the *attII* × *attC* reaction. An event that has never been reported is the recombination of *attC* sites through the L box. We sought to detect such recombination events with *attII*, since they would presumably occur through the L box of *attII* ts. We reasoned that, since bottom strands of both sites recombine very efficiently through the R boxes, it would be easier to observe such events using the ts of *attC*. Using our conjugation assay, we delivered the ts of *attC<sub>aadA7</sub>* and checked a total of 607 recombinants. Among them, an estimate of 251 had recombined through the ts (the rest having recombined through the resynthesized complementary bs) and only 1 (0.16%) had recombined through the L box with the top strand of *attII* (Supplementary Fig. 5). Since the EHB T<sub>23</sub> on the bottom strand impedes the nucleophilic attack on the L box, we assessed if the EHB complementary to T<sub>23</sub> on the top strand plays a role in this bias, using a ΔT<sub>23</sub> mutant site. We found three L-box recombinants out of 642 colonies tested (0.47%) suggesting a role (although minor) for the complementary A base in avoiding L-box recombination of ts *attC* sites. The crossover point in these recombinants could not be determined precisely due to an adenine stretch, but it was not coherent with the cleavage site determined previously for the *attII* L box. This is probably related to conformational differences between *attII* and *attC* sites that force the crossover to shift. This need for a different crossover point could be itself at the basis of the low frequency of L-box events in the *attII* × *attC* reaction.

## Discussion

In this work we show that integron integrases can perform the second strand exchange in reactions involving exclusively



**Figure 4 | Mapping of the crossover point on the top strand.** (a) Positions randomized in the attI<sub>STOP</sub> site to produce the attI<sub>N</sub> library (n = 40). The theoretical frequency of bases in every randomized position is of 92.5% WT base and 2.5% of every other base. (b) Diagram of the experimental set-up and recombination reaction. Recombination between attI<sub>N</sub> and attI<sub>AAA</sub> takes place through the top strand, splitting the randomized bases between moieties in the resulting sites of the cointegrate (right and left). (c) Graphical representation of base frequency within the attI<sub>N</sub> library before recombination, and the right and left sites after recombination. Green, WT base; Purple, non-WT base.



**Figure 5 | Mismatches in the top strand abolish HJ resolution in vitro.** In vitro activity of IntI1 on different attI substrates. (a) Gel Strand transfer test allowing the observation of HJ intermediates. Resolution of these structures through a second cleavage converts them back to the original substrate (sub). A diagram of the L and R boxes of the site is shown below its cognate gel (the rest of the site is identical to the wild type in all cases), where red bases represent changes with respect to the wild-type sequence. Crossover points in both strands are indicated with black arrows. (b) Transfer activity quantified as the amount of signal retained in the filter (HJs) over the total amount of signal. Columns represent the mean and error bars the s.d. (n = 3). Transfer activity for IntI1<sub>Y312F</sub> was undetectable in all assays. Statistical analysis was performed using Dunnett's test with the activity on the WT site as the control (\*significant (α = 0.05); \*\*, significant (α = 0.01); ND, not detected).

double-stranded sites. We found that integrases recombine the top strand of attI sites at frequencies matching those of the bottom strand and that there is no order in the use of strands

during the reaction. The crossover point in the top strand of the attI site has been elucidated using a novel technique that involves deep sequencing. This approach has allowed overcoming



technical difficulties found with classical *in vitro* methods due to the low recombination activity observed. With deep sequencing becoming more accessible, this technique could be useful for others working with recombinases that show low activity or for which optimal conditions *in vitro* have not been elucidated.

Our work reveals surprising aspects of the dynamics of *attI* × *attI* reactions, with broad evolutionary implications that are discussed below. Still, other aspects of the *attI*–IntI interaction remain unclear. *attI* sites are recognized for their sequence, as proven by cognate IntI–*attI* recognition<sup>29</sup>. We (and others previously<sup>38</sup>) have observed in our *in vitro* experiments a cooperative binding starting at the R box. It is also known that secondary recombination sites contain the 5′-AAC-3′ signature of the R box of all *attI* sites<sup>37</sup>. Given that the rest of the R box of the site changes its sequence at every cassette integration event, one could presume that this triplet is fundamental for IntI binding. The almost complete loss of binding activity in our experiments when a mismatch is located in this triplet (Supplementary Fig. 4) supports this idea. Nevertheless, this triplet is universal to all *attI* sites and would not allow integrase discrimination of cognate sites. Also, the sequence of the *attII* bears a 5′-AAC-3′ triplet at the border of the L box, but the crossover point occurs between different bases (A/A) compared with that of the bottom strand (A/C), suggesting that topology is more important than sequence in the cleavage of the top strand. Therefore, the relative influence of the universal crossover triplet and of the rest of the R box in the binding remains a subject of interest that would need further studying.

This work shows that the avoidance of the second cleavage on *attC* sites is not due to the mere presence of the integrase-specific I2  $\alpha$ -helix, but rather to its interaction with *attC* sites. The extremely low frequencies at which *attC<sub>aadA7</sub>* recombines through the L box highlights the strong constraints that *attC* sites impose on the reaction. It is noteworthy that in natural conditions L-box reactions between *attC* and *attI* sites would place the *attI* site downstream of the cassette. Hence, such tight control on directionality likely serves to ensure the correct functioning of the integron by keeping the *attI* site next to the Pc promoter after consecutive cassette insertions.

From a functional perspective, the reaction between *attI* sites has an unclear adaptive meaning. It could be argued that in chromosomal integrons these reactions have deleterious effects on bacterial fitness, since they lead to the formation of chromosomal dimers that have to be further resolved before cell division. This is probably the reason for the 3–4 orders of magnitude lower recombination frequency for *attI* × *attI* reactions compared with those involving *attC* sites. Nevertheless, in the multicopy world of mobile integrons, one cell can contain more than one integron on the same or different replicons<sup>41–43</sup>. Given the possibility of having Pc's of different strengths, an *attI* × *attI* reaction can lead to the *en bloc* transfer of cassette arrays between integrons. This would alter the expression levels of multiple antimicrobial resistance genes at once, and could have important effects on chemotherapy outcomes. It could be argued that the importance of such rearrangements is probably minor, due to the low recombination frequencies for *attI* × *attI* in laboratory conditions. Nevertheless, one aspect of this reaction that has been constantly overlooked is that, while *attC* sites are normally in their non-recombinogenic form, *attI* sites are recombinogenic during almost the entire cell cycle, with the brief exception of the passage of the replication fork. Despite this low recombination frequency, it is plausible that *attI* × *attI* reactions are more prevalent among mobile integrons than previously envisaged, especially under transient, low-level (or even stochastic) expression of the integrase. These expression pulses typically occur during horizontal gene transfer events<sup>36,44</sup>.

This work shows that despite millions of years of evolution towards a system specialized in structured ssDNA recombination, and the acquisition of a 20-residue-long domain within the catalytic core, integron integrases have conserved the activity of their canonical Y-recombinase ancestors. Many examples of DNA breakage and rejoining enzymes acting on dsDNA (like Y-recombinases) or ssDNA (like HUH endonucleases<sup>45</sup>) can be found, but this is the first report of a recombinase with full dual activity on both types of sites, as well as a substrate-dependent switch in recombination pathways. Integrases are capable of recognizing distinct substrates (ds versus ss hairpins) in different ways (sequence versus structure specific) and processing the reaction through different pathways (double strand exchange versus replicative resolution). This duality has deep biological implications, since one pathway is semiconservative and dependent on host machinery, while the other is not. Integrons could function, at least theoretically, using only the classical recombination pathway on dsDNA to deliver cassette integration and excision reactions. It is therefore tempting to speculate that the force driving the evolution of integrons towards a mixed ss/ds-DNA system derives from the benefits of semiconservative recombination. By producing recombined and non-recombined offspring, it allows testing the adaptive value of incoming DNA, while minimizing the deleterious effect of capturing maladaptive genes. Furthermore, it also represents a mechanism for gene duplication, as observed in the *Vibrio cholerae* superintegron<sup>10</sup> and possibly in some mobile integrons<sup>46</sup>.

From an evolutionary perspective, the functional innovation of integrases towards the recombination of folded ss-*attC* sites, posed important mechanistic constraints in the recognition of the site and the resolution of aHJ to deliver productive reactions. The solution in integrases comes in the form of the additional I2 domain that recognizes the EHBs in *attC* sites. This impedes the second cleavage, allowing for the replicative resolution of the HJ. Interestingly, EHBs and the integrase I2 domain are a good example of coevolution. These structures were probably the first step in the evolutionary transition from the ancestral recombination system to the ss/ds-DNA recombination platform integrons are today. Nevertheless, the acquisition of the I2 domain raises a chicken-egg paradox: how could the domain be selected for in the absence of *attC* sites to provide a gain of function? And how could *attC* sites be selected for before the recognition domain existed? Such a conundrum is found in many evolutionary processes and can easily be understood with the discovery of functional intermediary forms that allow to infer smooth evolutionary transitions between states. We believe integron integrases are representatives of such a bi-functional state. In the light of our results, it now seems possible that an I2-like domain was acquired at a given time by the ancestor of today's integron integrases and, to some degree, permitted the recognition of new ss-substrates while retaining the original activity (or part of it) on the initial site. From this starting point, the streamlining through natural selection, to reach what integrons are today, seems a more straightforward process.

To date, the only recombination reaction that shares some similarities with cassette recombination is the integration of phage CTX $\phi$  of *Vibrio cholerae*. Nevertheless, CTX $\phi$  cannot be considered a complete system since it does not encode an integrase but instead hijacks the host XerCD recombinases<sup>47</sup>. As a consequence, no specialization of the system towards the recombination of ss-DNA is observed: XerCD lack structural adaptations for the recognition and processing of ss-substrates (I2-like domains), and the phage site stem does not have EHBs but rather mimics a ds-DNA molecule. Nevertheless, the resemblances between cassettes and CTX $\phi$  integration, the close phylogenetic relation between integrases and XerCD

recombinases and the fact that both systems can coexist in the same bacterial species, makes it tempting to speculate about a viral origin for integrons. Interestingly, the link to the SOS response found in all integrons is also present in many phages, including CTX $\phi$  (ref. 48). To date, no chromosome-dimer resolution recombinase of the XerCD family is known to possess a domain similar to I2 and directionality is ensured through their interaction with other host proteins such as FtsK<sup>49</sup>. Nevertheless, one could imagine that acquisition of an I2-like domain by a Xer-like ancestor could have provided the function of avoiding the second strand exchange during phage integration. This would protect the integrity of the chromosome while favouring correct integration events, and would hence be of adaptive value to the cell and the phage. We can now imagine that the chromosome dimer resolution activity of the hybrid recombinase would still be conserved. Phages, or any other integrative mobile element exploiting Xer (IMEX<sup>49</sup>) could have then acted as a vehicle for new adaptive genes, just as they currently do<sup>50,51</sup>, representing an ancestor of integron cassettes.

Previous work on evolutionary aspects of Y-recombinases focused on the properties that allow related phage integrases to target specific sequences in the bacterial genome and the possibility of adapting these proteins to recognize non-cognate sites<sup>52–54</sup>. This work sheds light on the evolutionary potential of the Y-recombinase protein fold towards evolutionary innovation (not adaptation). Our results suggest that Y-recombinase-mediated reactions are a more dynamic process than previously assumed. They underscore the robustness of the protein fold, a characteristic that likely confers a high evolvability to the members of this superfamily<sup>6</sup>. It is arguable that the duality in the recognition and binding of integron integrases to the sites has strongly been conserved because cassettes have to be integrated in the *attI* site. Nevertheless, the classical recombination pathway could have been lost through the specialization of integrases towards the single strand exchange pathway, without affecting any of the known aspects of integron dynamics. Therefore, preservation of this pathway after eons of evolution<sup>11</sup> can be seen as the consequence of a lack of negative selection<sup>55</sup>, or as reflecting an important biological function. The fact that top and bottom strand cleavage of *attI* sites occur at similar frequencies (Fig. 3b) suggests that top strand cleavage is not a promiscuous activity of the integrase and favours the biological function hypothesis. The flexibility of the integrase might not be completely unveiled yet, and functions dependent on some aspects of this plasticity might remain unknown. The second strand exchange resolution pathway opens new avenues for possible mechanisms of cassette genesis, a subject for which no reliable model is yet available<sup>10</sup>.

## Methods

**Bacterial media and conditions.** *E. coli* strains were grown in Luria Bertani broth (LB) at 37 °C. Growth at 42 °C was used to induce the expression and force the loss of the plasmid bearing lambda phage integrase. Antibiotics were used at the following concentrations: carbenicillin (Ap), 100  $\mu\text{g ml}^{-1}$ , chloramphenicol (Cm), 25  $\mu\text{g ml}^{-1}$ , kanamycin (Km), 25  $\mu\text{g ml}^{-1}$ , spectinomycin (Sp), 40  $\mu\text{g ml}^{-1}$ , Thymidine (Thy) and DAP were supplemented when necessary to a final concentration of 0.3 mM. Glucose and L-arabinose were added at 10 and 2 mg ml<sup>-1</sup> final concentration, respectively. Chemicals were obtained from Sigma-Aldrich (France).

**Bacterial strains, plasmids and primers.** Bacterial strains in this study are *E. coli* DH5 $\alpha$ , MG1655 (laboratory collection), I1, and  $\beta$ 2163 (ref. 56). Details on strains, plasmids and primers are described, respectively, in Supplementary Tables 1–3.

**DNA procedures.** Standard techniques were used for DNA manipulation and cloning<sup>57</sup>. Restriction and DNA-modifying enzymes were purchased from New England Biolabs and Fermentas (Thermo Scientific). PCRs were performed with Dreamtaq DNA polymerase, and Phusion polymerase (Thermo Scientific)

according to the manufacturer's instructions. Agarose (1%) electrophoresis gels were used to visualize DNA. DNA purification from PCR products and gels, as well as plasmid extractions, were performed using Qiagen kits. When necessary, DNA sequence was verified using an ABI BigDye Terminator v.3.1 sequencing kit and an ABI Prism 3100 Capillary Genetic Analyzer (Applied Biosystem). GATC and EUROFINs sequencing services were also used.

**In silico work.** Protein modelling was performed using Phyre2 (sbj.bio.ic.ac.uk/phyre2) and MacPymol on templates c2a3vA (VchInt1A) and c1ma7A (Cre). Sequence analysis was performed using Geneious 7.1.

**Influence of host machinery in the *attI*  $\times$  *attI* reaction.** To test whether RecA, RecG and RuvABC impact the *attI*  $\times$  *attI* reaction, we performed a suicide conjugation assay based on that of Biskri *et al.*<sup>35</sup> and previously implemented in Bouvier *et al.*<sup>17</sup>. Briefly, the *attI* site provided by conjugation is carried on a suicide vector from the R6K-based pSW family that is known to use the Pir protein to initiate its own replication. This plasmid also contains an RP4 origin of transfer (*oriTRP4*). The donor strain  $\beta$ 2163 carries the transfer functions in its chromosome, requires DAP to grow in rich medium and can sustain pSW replication through the expression of a chromosomally integrated *pir* gene. The MG1655 recipient strains, which contain the pBAD::*intI1* [Ap<sup>R</sup>] (expressing the IntI1 integrase) and the pSU38 $\Delta$ ::*attI1* [Sp<sup>R</sup>] (carrying the *attI1* site), lacks the *pir* gene and therefore cannot sustain replication of the incoming *attI*-containing pSW vector. The only way for this vector to be maintained in the recipient cell is to form a cointegrate by *attI*  $\times$  *attI* recombination. We used the  $\omega$ 10001 strain as the donor and as recipient strains the MG1655 *E. coli*  $\omega$ A197 (control), and its derivatives  $\omega$ 9987 (*recA*-),  $\omega$ 9988 (*recA*-, *recG*-),  $\omega$ 9989 (*recA*-, *ruvABC*-) and  $\omega$ 9990 (*recA*-, *recG*-, *ruvABC*-) (Supplementary Tables 1 and 2). The recombination frequency is calculated as the ratio of transconjugants expressing the pSW marker [Cm<sup>R</sup>] to the total number of recipient clones [Ap<sup>R</sup>, Sp<sup>R</sup>]. Recombination frequencies correspond to the average of at least three independent trials. The *attI*  $\times$  *attI* co-integrate formation was checked by PCR with the appropriate SWbeg and MFD primers (on eight randomly chosen clones per experiment).

**Mismatched covalent circles preparation.** Construction of the p8669 and p8670 phagemid vectors to produce mismatched covalent circles was performed as in ref. 21 with some modifications. Briefly, we exchanged *attC* sites in plasmids p7770 and p7771 for *attI1* sites, giving rise to plasmids p8669 and p8670 (plasmids, carrying the *oriFD* in both orientations ensuring the production of respectively the bottom (*oriFD1*) and top strand (*oriFD2*)). Single strand DNA production was performed using M13K07 Helper Phage as in ref. 21. Both phagemid vectors (p8669 and p8670) are introduced by transformation into *F'* carrying *pir* strain cells ( $\beta$ 2150,  $\omega$ 4446) to respectively obtain the  $\omega$ 8675 and  $\omega$ 8676 strains. M13 infection and purification of ssDNA were performed according to the manufacturer (M13K07 Helper Phage and QIAprep Spin M13 kit). The complementary single strand DNA molecules were annealed, digested by both EcoRI and MfeI restriction enzymes to eliminate *oriFD* (which does not anneal due to the inverted orientation in each molecule), and self-ligated.

**Non-replicative recombination assay.** This assay was developed previously in the laboratory<sup>58</sup>. We transform with 1–2  $\mu\text{g}$  mismatched covalent circles containing the *attI* site into MG1655mutS215 strain containing the pBAD::*intI1* (p3938) and the pSU38 $\Delta$ ::*attI1* (p929) plasmids (strain  $\omega$ 7994). Circles cannot replicate in this genetic background ( $\omega$ 7994). Note that this assay was first envisaged in a *mutS*-, *recA*-strain but transformation efficiency was too low due to a filamenting (sick) phenotype in the double mutant. We hence studied the influence of *recA* and found it to be responsible for <0.5% of all recombinants in our assay (see below recombination frequencies with and without pBAD::*intI1*). This justified the use of a *recA* + strain neglecting the influence of homologous recombination in the recombination process. Competent cells were prepared in the presence of 0.2% arabinose to allow integrase expression. Selection on chloramphenicol (Cm)-containing plates (the mismatched circles marker) yielded, almost exclusively, *attI*  $\times$  *attI* recombination events. The *attI*  $\times$  *attI* cointegrate formation was checked on 101 clones by PCR with SeqattI and SeqNar primers. These PCR products include the mismatch-containing region, and were analysed for their resistance or sensitivity to both SacII and NarI restriction enzymes.

As a control, we performed a similar experiment using an MG1655mutS215 strain containing the pSU38 $\Delta$ ::*attI1* but lacking pBAD::*intI1* (strain  $\omega$ A266). To establish a recombination frequency, we transformed in parallel the same *mutS* strain lacking both the pSU38 $\Delta$ ::*attI1* and pBAD::*intI1* plasmids but containing a Pir-expressing plasmid (pSB118::*pir116*, p1177) ensuring the replication of the incoming circle (strain  $\omega$ 7120). Recombination activity corresponds to the ratio of Cm<sup>R</sup> clones obtained in *pir* – conditions (with and without integrase) to those obtained in *pir* + conditions. Note that the efficiency of transformation of each strain was determined beforehand and used to adjust the final ratio and normalize the results. We obtained a recombination frequency of  $1.17 \times 10^{-3}$  in presence of integrase, and of  $4.24 \times 10^{-6}$  in the absence of integrase.

As a supplementary control, we analysed the Cm<sup>R</sup> clones obtained in *pir* + context (Supplementary Fig. 2). After pooling of 319 clones, plasmids were

extracted and retrotransformed in  $\Pi 1$  *pir*<sup>+</sup> competent cells ( $\omega 1628$ ). Then, 101 transformants were analysed by SacII and NarI restriction and/or by sequencing (using SeqNarI and SeqSacII primers). The segregation pattern was, as indicated in Supplementary Fig. 2, of 49.5% NarI<sup>+</sup>/SacII<sup>-</sup>, 48.5% NarI<sup>-</sup>/SacII<sup>+</sup>, 1% of NarI<sup>+</sup>/SacII<sup>+</sup> and 1% NarI<sup>-</sup>/SacII<sup>-</sup>. This close-to-50% segregation confirms the correct stoichiometry of each strand, and validates the production of the mismatched circles.

**Chromosomal inversion test proves a second strand exchange.** In this experiment we used an assay in which recombination between two sites leads to the reconstitution of the essential *dapA* gene allowing for growth in media not supplemented with DAP<sup>28,36</sup>. DNA molecules containing the sites in direct orientation were designed *in silico* and synthesized by GeneART (Life Technologies). To avoid background noise (expression and translation of *dapA*) before recombination, the right *attI* site (the one immediately upstream *dapA*) was modified to include a STOP codon 1 bp upstream of the L box (Fig. 2). The design also included two XhoI and SmaI restriction sites surrounding the left *attI* site and the  $P_{ndmA}$  promoter (registry part J23100) that drives transcription of *dapA* after recombination. Through XhoI/SmaI double digestion we inverted the left site to obtain the inverted orientation. These constructions were subsequently cloned EcoRI/NruI into the *pir*-dependent plasmid pA669 (Sp<sup>R</sup>) containing a lambda phage *attP* site, and transformed in  $\Pi 1$  *pir*<sup>+</sup> competent cells ( $\omega 1628$ ). A FRT recombination site was added in EcoRI to allow for the resolution of dimers after integration in the chromosome. Sequence was verified and these plasmids were transformed in the  $\omega 8488$  strain, an MG1655 *recA*<sup>-</sup> *dapA*<sup>-</sup> *E. coli* strain containing the replication/expression-thermosensitive plasmid p3153 bearing the lambda phage integrase (Ap<sup>R</sup>). In this strain, *pir*-dependent plasmids do not replicate and are only maintained through their integration in the lambda *attB* site of the chromosome. Integration events where selected for using the Sp resistance marker of the plasmid. Loss of the p3153 plasmid after overnight incubation at 42 °C was verified by replica plating on media containing carbenicillin (selecting for the Ap<sup>R</sup> marker of p3153). Monomeric integration was verified through PCR with primers (1329 and 1698). Sequence was verified in all strains produced. Strains  $\omega B36$  and  $\omega B37$  containing the constructions in direct and inverse orientation were selected for the study and transformed with the pBAD::*intI1* (p3938) plasmid giving rise to strains  $\omega B82$  and  $\omega B83$ . To test the recombination rate of the genetic structures, strains were grown overnight in LB supplemented with DAP, carbenicillin, spectinomycin and glucose, then diluted 1:50 into fresh media containing DAP, carbenicillin (to avoid the loss of p3938) and glucose or arabinose to repress or induce the expression of the *intI1*, and incubated for 3 h at 37 °C with shaking. Cultures were then plated on selective media with and without DAP. The recombination frequency is calculated as the ratio of colonies growing without DAP (recombined) to the total number of colonies growing on DAP-supplemented media. Recombination frequencies correspond to the average of at least three independent assays.

**Killing the bottom strand.** To obtain constructions in which the left recombination *attI* site bore a C to A mutation in the cleavage site, we performed site-directed mutagenesis through overlap extension PCR combining primers attI CxA R and attI GxT F with primers XhoI F and XhoI R and using pA873 as template. The PCR product was cloned in both orientations in plasmid pA873 using XhoI restriction sites, giving rise to plasmids pC131 and pC132. These plasmids were inserted in the chromosome of  $\omega 8488$  as described above and verified clones (strains  $\omega C139$  and  $\omega C140$ ) were further transformed with the pBAD::*intI1* (p3938) plasmid to obtain strains  $\omega C162$  and  $\omega C163$ . Recombination rates were measured as above.

**Determining the crossover point.** To build the library of *attI*<sub>N</sub> sites in the *pir*-dependent pSW plasmid we combined primer 2915 with primers 2916–2925, obtaining 10 *attI*<sub>N</sub> sites, each with a random base at a given position. These sites were mixed, digested XhoI, PstI, cloned into the pSW plasmid pC066 (Sp<sup>R</sup>) and transformed into  $\Pi 1$  *pir*<sup>+</sup> competent cells ( $\omega 1628$ ). >5,000 transformants were mixed to establish the library ( $\omega C351$ ). Sequencing of random clones showed that 18 out of 20 had bases other than WT in the expected positions, validating the library (see also the deep sequencing results in Fig. 4). The library was subcloned into a  $\beta 2163$  donor strain to be deliverable through conjugation ( $\omega C373$ ).

To force L-box recombination, the *attI* site on the resident plasmid was changed to an *attI*<sub>AAA</sub> site using primers 2889 and 2932 and cloned EcoRI/BamHI in p929 (Km<sup>R</sup>), giving rise to pC252. Strain  $\omega 4826$  (MG1655 *recA*<sup>-</sup>) was transformed with pC252 and p3938 (pBAD::*intI1*) ( $\omega C307$ ). As controls of this experiment we included a strain transformed with p929 instead of pC252 ( $\omega C305$ ) and strains without the p3938 ( $\omega C304$  ( $\omega 4826/p929$ ) and  $\omega C306$  ( $\omega 4826/pC252$ )).

The library was delivered through conjugation from  $\omega C373$  to  $\omega C307$ . Recombination frequency was of  $2.63 \times 10^{-4}$  in the presence of arabinose (inducing the expression of the integrase), and of  $1.88 \times 10^{-6}$  in the presence of glucose, confirming the IntI1-mediated recombination of pSW-*attI*<sub>N</sub> into pSU-*attI*<sub>AAA</sub>. More than 7,000 colonies from the conjugation assay were collected ( $\omega C453$  to  $\omega C455$ ) and subjected to plasmid extraction. As a control, conjugation

assays with  $\omega C373$  and  $\omega C305$  (pSU-*attI*) were performed, showing similar recombination frequencies:  $1.64 \times 10^{-4}$  with arabinose and  $2.59 \times 10^{-6}$  with glucose.

The *attI* sites from the library and the recombination products were PCR amplified using primers 1236-*attI*<sub>N</sub>'s right (library); and 1236-738 and MRV-*attI*<sub>N</sub>'s right (recombinant sites). PCR products were used for sequencing on an Illumina MySeq. Resulting sequences were filtered to discard abnormal recombination events and further analyzed using Galaxy (<https://usegalaxy.org>) and IGV 2.3 (<https://www.broadinstitute.org/igv/home>). Base frequency was analysed at every position in the library and the recombinant sites. The expected frequency of bases in each randomized position is of 92.5% of wild-type base and 2.5% of every other base. This is because for a given position only 1 out of 10 sites bears a random base (90% WT-10% N) and 25% of random bases are wild type (92.5% WT and 7.5% not WT). The observed base frequency composition in the library before recombination matched the expected distribution. Sequencing of the resulting sites after recombination showed a split distribution of randomized bases in both sites, allowing to infer the crossover point.

**In vitro experiments.** Translational fusions with a maltose-binding protein were obtained of IntI1 and the catalytically inactive IntI1<sub>Y312F</sub> by cloning in the commercial vector pMAL-CX5 and transforming *E. coli* Top10 strain. Plasmids were further purified and transformed in the BL21 strain for protein production. Colonies from this transformation were collected and grown overnight in 25 ml of LB containing carbenicillin and 0.2% glucose. After incubation, cells were collected by centrifugation at 4 °C, resuspended in 25 ml of fresh media, and inoculated in 3 l of LB containing carbenicillin. When the culture reached an OD = 0.6, IPTG was added at 0.3 mM and the culture was incubated at 14 °C overnight. The cells were then collected by centrifugation and frozen. Cells were thawed to room temperature under cold tap water, and lysed by passage through an Emulsiflex C-5 cell breaker (Avestin Europe GmbH) at 13,000 p.s.i. at 4 °C. The cell lysate was centrifuged at 23,500g for 40 min in an RC-5C using an ss-34 rotor (Sorvall, Thermo Scientific). The IntI1-MBP fusion protein was purified with 5 ml Amylose column (GE Healthcare) on an AKTA-prime FPLC machine (GE Healthcare).

The sequence of the oligonucleotides used for the strand transfer assay is reported in Supplementary Table 3. A total of 100 pmol of oligonucleotides coding for top and bottom strands were labeled using T4 DNA kinase (Promega) and 25 mCi <sup>32</sup>P gamma ATP for 1 h at 37 °C. Once radiolabelled, top and bottom strands were then annealed (85 °C for 3 min and at room temperature for overnight). The strand transfer assay was performed using conditions derived from refs 40,59. Briefly, purified MBP-IntI1 was diluted in 1 M NaCl, 20 mM HEPES, 1 M, pH 7, 10 mM DTT for 30 min on ice. Then incubated with the different recombination sites (25 nM final) for 30 min on ice in a final volume of 5  $\mu$ l. The reaction was then started by adding 5  $\mu$ l of the reaction buffer (final concentrations 100 mM IN, 15% DMSO, 8% PEG, 10 mM MgCl<sub>2</sub>, 100 mM NaCl, 10 mM DTT). Samples were incubated for 2 h at 37 °C. After reaction the products were treated 1 h at 55 °C with proteinase K (Promega) 1 mg ml<sup>-1</sup> final concentration and deproteinized using phenol/chloroform/isoamyl-alcohol (24/25/1 v/v/v). Aqueous phase was run on 12% polyacrylamide gel at 1,000 V for 5–6 h. Gel was then autoradiographed and quantified using Image J software (Supplementary Fig. 4b).

A total of 5 pmol of MBP-IntI1 were incubated for 20 min at room temperature with 1 pmol of radiolabelled double-stranded DNA containing the recombination site in 15% DMSO, 8% PEG, 10 mM MgCl<sub>2</sub>, 100 mM NaCl, 10 mM DTT. The complexes were then filtered on nitrocellulose filters. Filters were washed three times with PBS and counted using a Wallac 1409 liquid scintillation counter.

**L-box recombination in the attC<sub>aadA7</sub> × attI reaction.** This experiment is based on the suicidal conjugation assay explained above. This time the plasmids delivered by conjugation carry the top or bottom strand of the wild type or the  $\Delta T23$  *attC*<sub>aadA7</sub> sites. Maintenance of these plasmids in the recipient cell is only possible through an *attC* × *attI* reaction. We used the  $\omega D060$ ,  $\omega D061$ ,  $\omega D805$  and  $\omega D806$  strains as donors and  $\omega 9669$  as the recipient strain. The recombination frequencies in R' (bs) and R'' (ts) and the corresponding mean deviations were deduced by determining the orientation of cassette integration for at least 20 independent clones for each bs transfer (and multiplying the total recombination frequency by the ratio of R' and R'' recombination) and for 100 clones in the ts transfer assays. The orientation of cassette integration was determined by performing two PCR reactions: the first reaction with SW23begin/MFD primers produced a product only when the recombination took place in the delivered *attC* strand (in R' when bs is delivered, 530 bp; in R'' when ts is delivered, 520 bp); the second reaction with SW23end/MFD primers produced a product only when the recombination took place in the resynthesized *attC* strand (in R' when bs is delivered, 240 bp; in R'' when ts is delivered, 210 bp). In addition, the recombination in the L box was detected by performing two PCR reactions on all clones. The first reaction with Rplp/571 produced a product when the recombination took place in the L' box (340 bp); the second reaction with RppLpp/571 produced a product when the recombination took place in the L'' box (340 bp). Recombination through the L/L'' box of *attC* sites was only detected when the top strand was delivered (data for bs is hence not shown).

## References

- Grindley, N. D., Whiteson, K. L. & Rice, P. A. Mechanisms of site-specific recombination. *Annu. Rev. Biochem.* **75**, 567–605 (2006).
- Landy, A. The lambda integrase site-specific recombination pathway. *Microbiol Spectr* **3**, MDNA3-0051-2014 (2015).
- Golden, J. W., Robinson, S. J. & Haselkorn, R. Rearrangement of nitrogen fixation genes during heterocyst differentiation in the cyanobacterium *Anabaena*. *Nature* **314**, 419–423 (1985).
- Val, M. E., Skovgaard, O., Ducos-Galand, M., Bland, M. J. & Mazel, D. Genome engineering in *Vibrio cholerae*: a feasible approach to address biological issues. *PLoS Genet.* **8**, e1002472 (2012).
- Sherratt, D. J. & Wigley, D. B. Conserved themes but novel activities in recombinases and topoisomerases. *Cell* **93**, 149–152 (1998).
- Toth-Petroczy, A. & Tawfik, D. S. The robustness and innovability of protein folds. *Curr. Opin. Struct. Biol.* **26**, 131–138 (2014).
- Kobryn, K. & Chaconas, G. Hairpin Telomere Resolvases. *Microbiol. Spectr.* **2**, MDNA3-0023-2014 (2014).
- Mazel, D. Integrons: agents of bacterial evolution. *Nat. Rev. Microbiol.* **4**, 608–620 (2006).
- Recchia, G. D. & Hall, R. M. Gene cassettes: a new class of mobile element. *Microbiology* **141**, 3015–3027 (1995).
- Escudero, J. A., Loot, C., Nivina, A. & Mazel, D. The Integron: Adaptation On Demand. *Microbiol. Spectr.* **3**, MDNA3-0019-2014 (2015).
- Rowe-Magnus, D. A. *et al.* The evolutionary history of chromosomal super-integrations provides an ancestry for multiresistant integrons. *Proc. Natl Acad. Sci. USA* **98**, 652–657 (2001).
- Rowe-Magnus, D. A. & Mazel, D. The role of integrons in antibiotic resistance gene capture. *Int. J. Med. Microbiol.* **292**, 115–125 (2002).
- Martinez, E. & de la Cruz, F. Genetic elements involved in Tn21 site-specific integration, a novel mechanism for the dissemination of antibiotic resistance genes. *EMBO J.* **9**, 1275–1281 (1990).
- Hall, R. M., Brookes, D. E. & Stokes, H. W. Site-specific insertion of genes into integrons - role of the 59-base element and determination of the recombination cross-over point. *Mol. Microbiol.* **5**, 1941–1959 (1991).
- Frumerie, C., Ducos-Galand, M., Gopaul, D. N. & Mazel, D. The relaxed requirements of the integron cleavage site allow predictable changes in integron target specificity. *Nucleic Acids Res.* **38**, 559–569 (2010).
- Francia, M. V., Zabala, J. C., de la Cruz, F. & Garcia-Lobo, J. M. The IntI1 integron integrase preferentially binds single-stranded DNA of the *attC* site. *J. Bacteriol.* **181**, 6844–6849 (1999).
- Bouvier, M., Demarre, G. & Mazel, D. Integron cassette insertion: a recombination process involving a folded single strand substrate. *EMBO J.* **24**, 4356–4367 (2005).
- Cambray, G., Guerout, A. M. & Mazel, D. Integrons. *Annu. Rev. Genet.* **44**, 141–166 (2010).
- Bouvier, M., Ducos-Galand, M., Loot, C., Bikard, D. & Mazel, D. Structural features of single-stranded integron cassette *attC* sites and their role in strand selection. *PLoS Genet.* **5**, e1000632 (2009).
- MacDonald, D., Demarre, G., Bouvier, M., Mazel, D. & Gopaul, D. N. Structural basis for broad DNA-specificity in integron recombination. *Nature* **440**, 1157–1162 (2006).
- Loot, C., Ducos-Galand, M., Escudero, J. A., Bouvier, M. & Mazel, D. Replicative resolution of integron cassette insertion. *Nucleic Acids Res.* **40**, 8361–8370 (2012).
- Nunes-Duby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T. & Landy, A. Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.* **26**, 391–406 (1998).
- Toll-Riera, M. & Alba, M. M. Emergence of novel domains in proteins. *BMC Evol. Biol.* **13**, 47 (2013).
- Aroul-Selvam, R., Hubbard, T. & Sasidharan, R. Domain insertions in protein structures. *J. Mol. Biol.* **338**, 633–641 (2004).
- Guerin, E. *et al.* The SOS response controls integron recombination. *Science* **324**, 1034 (2009).
- Wagner, A. The molecular origins of evolutionary innovations. *Trends Genet.* **27**, 397–410 (2011).
- Demarre, G., Frumerie, C., Gopaul, D. N. & Mazel, D. Identification of key structural determinants of the IntI1 integron integrase that influence *attC* x *attI1* recombination efficiency. *Nucleic Acids Res.* **35**, 6475–6489 (2007).
- Loot, C. *et al.* The integron integrase efficiently prevents the melting effect of *Escherichia coli* single-stranded DNA-binding protein on folded *attC* sites. *J. Bacteriol.* **196**, 762–771 (2014).
- Collis, C. M., Kim, M. J., Stokes, H. W. & Hall, R. M. Integron-encoded IntI integrases preferentially recognize the adjacent cognate *attI* site in recombination with a 59-base site. *Mol. Microbiol.* **46**, 1415–1427 (2002).
- Partridge, S. R. *et al.* Definition of the *attI1* site of class 1 integrons. *Microbiology* **146**, 2855–2864 (2000).
- Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).
- Lloyd, R. G. & Sharples, G. J. Dissociation of synthetic Holliday junctions by *E. coli* RecG protein. *EMBO J.* **12**, 17–22 (1993).
- Tsaneva, I. R., Muller, B. & West, S. C. ATP-dependent branch migration of Holliday junctions promoted by the RuvA and RuvB proteins of *E. coli*. *Cell* **69**, 1171–1180 (1992).
- Connolly, B. *et al.* Resolution of Holliday junctions in vitro requires the *Escherichia coli* *ruvC* gene product. *Proc. Natl Acad. Sci. USA* **88**, 6063–6067 (1991).
- Biskri, L., Bouvier, M., Guerout, A. M., Boissnard, S. & Mazel, D. Comparative study of class 1 Integron and *Vibrio cholerae* superintegron integrase activities. *J. Bacteriol.* **187**, 1740–1750 (2005).
- Baharoglu, Z., Bikard, D. & Mazel, D. Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. *PLoS Genet.* **6**, e1001165 (2010).
- Hansson, K., Skold, O. & Sundstrom, L. Non-palindromic *attI* sites of integrons are capable of site-specific recombination with one another and with secondary targets. *Mol. Microbiol.* **26**, 441–453 (1997).
- Gravel, A., Fournier, B. & Roy, P. H. DNA complexes obtained with the integron integrase IntI1 at the *attI1* site. *Nucleic Acids Res.* **26**, 4347–4355 (1998).
- Messier, N. & Roy, P. H. Integron integrases possess a unique additional domain necessary for activity. *J. Bacteriol.* **183**, 6699–6706 (2001).
- Dubois, V., Debreyer, C., Litvak, S., Quentin, C. & Parissi, V. A new in vitro strand transfer assay for monitoring bacterial class 1 integron recombinase IntI1 activity. *PLoS ONE* **2**, e1315 (2007).
- Gonzalez-Zorn, B. *et al.* Genetic basis for dissemination of *armA*. *J. Antimicrob. Chemother.* **56**, 583–585 (2005).
- Boyd, D., Cloeckert, A., Chaslus-Dancla, E. & Mulvey, M. R. Characterization of variant *Salmonella* genomic island 1 multidrug resistance regions from serovars Typhimurium DT104 and Agona. *Antimicrob. Agents Chemother.* **46**, 1714–1722 (2002).
- Hamidian, M., Holt, K. E. & Hall, R. M. Genomic resistance island AGI1 carrying a complex class 1 integron in a multiply antibiotic-resistant ST25 *Acinetobacter baumannii* isolate. *J. Antimicrob. Chemother.* **70**, 2519–2523 (2015).
- Baharoglu, Z., Krin, E. & Mazel, D. Connecting environment and genome plasticity in the characterization of transformation-induced SOS regulation and carbon catabolite control of the *Vibrio cholerae* integron integrase. *J. Bacteriol.* **194**, 1659–1667 (2012).
- Chandler, M. *et al.* Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat. Rev. Microbiol.* **11**, 525–538 (2013).
- San Millan, A. *et al.* Sequencing of plasmids pAMBL1 and pAMBL2 from *Pseudomonas aeruginosa* reveals a blaVIM-1 amplification causing high-level carbapenem resistance. *J. Antimicrob. Chemother.* **70**, 3000–3003 (2015).
- Val, M. E. *et al.* The single-stranded genome of phage CTX is the form used for integration into the genome of *Vibrio cholerae*. *Mol. Cell* **19**, 559–566 (2005).
- Quinones, M., Kimsey, H. H. & Waldor, M. K. LexA cleavage is required for CTX prophage induction. *Mol. Cell* **17**, 291–300 (2005).
- Das, B., Martinez, E., Midonet, C. & Barre, F. X. Integrative mobile elements exploiting Xer recombination. *Trends Microbiol.* **21**, 23–30 (2013).
- Midonet, C. & Barre, F. X. Xer site-specific recombination: promoting vertical and horizontal transmission of genetic information. *Microbiol. Spectr.* **2**, MDNA3-0056-2014 (2014).
- Waldor, M. K. & Mekalanos, J. J. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**, 1910–1914 (1996).
- Campbell, A., del-Campillo-Campbell, A. & Ginsberg, M. L. Specificity in DNA recognition by phage integrases. *Gene* **300**, 13–18 (2002).
- Gottfried, P., Yagil, E. & Kolot, M. Core-binding specificity of bacteriophage integrases. *Mol. Gen. Genet.* **263**, 619–624 (2000).
- Yagil, E., Dorgai, L. & Weisberg, R. A. Identifying determinants of recombination specificity: construction and characterization of chimeric bacteriophage integrases. *J. Mol. Biol.* **252**, 163–177 (1995).
- Tawfik, D. S. Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? *Curr. Opin. Chem. Biol.* **21**, 73–80 (2014).
- Demarre, G. *et al.* A new family of mobilizable suicide plasmids based on the broad host range R388 plasmid (IncW) or RP4 plasmid (IncPa) conjugative machineries and their cognate *E. coli* host strains. *Res. Microbiol.* **156**, 245–255 (2005).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbour Laboratory Press, 1989).
- Loot, C., Bikard, D., Rachlin, A. & Mazel, D. Cellular pathways controlling integron cassette site folding. *EMBO J.* **29**, 2623–2634 (2010).
- Dubois, V., Debreyer, C., Quentin, C. & Parissi, V. In vitro recombination catalyzed by bacterial class 1 integron integrase IntI1 involves cooperative binding and specific oligomeric intermediates. *PLoS ONE* **4**, e5228 (2009).
- Martin, S. S., Chu, V. C. & Baldwin, E. Modulation of the active complex assembly and turnover rate by protein-DNA interactions in Cre-LoxP recombination. *Biochemistry* **42**, 6814–6826 (2003).

## Acknowledgements

This work was supported by the Institut Pasteur, the Centre National de la Recherche Scientifique (CNRS-UMR3525), the European Union Seventh Framework Programme (FP7-HEALTH- 2011-single-stage), the 'Evolution and Transfer of Antibiotic Resistance' (EvoTAR) and the French Government's Investissement d'Avenir program Laboratoire d'Excellence 'Integrative Biology of Emerging Infectious Diseases' (ANR-10-LABX-62-IBEID) and the French National Research Agency (ANR-12-BLAN-DynamINT). J.A.E. is supported by the Marie Curie Intra-European Fellowship for Career Development (FP-7-PEOPLE-2011-IEF, ICADIGE). We thank Mick Chandler, Dan Tawfik and Sebastian Aguilar Pierlé for critical reading of the manuscript.

## Author contributions

J.A.E., C.L. and D.M. designed the experiments. J.A.E., C.L., V.P. and A.N. performed the experiments. J.A.E. and C.B. produced and analysed the deep-sequencing data. J.A.E. and D.M. wrote the paper.

## Additional information

**Accession codes:** The crossover point sequencing data for the library, left site and right site has been deposited at the Sequence Read Archive under the accession code SRP071193.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

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

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Escudero, J. A. *et al.* Unmasking the ancestral activity of integron integrases reveals a smooth evolutionary transition during functional innovation. *Nat. Commun.* 7:10937 doi: 10.1038/ncomms10937 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>