

## ORIGINAL ARTICLE

## The evolutionary dynamics of integrons in changing environments

Jan Engelstädter<sup>1</sup>, Klaus Harms<sup>2,3</sup> and Pål J Johnsen<sup>2</sup><sup>1</sup>School of Biological Sciences, The University of Queensland, Brisbane, Queensland, Australia; <sup>2</sup>Faculty of Health Sciences, Department of Pharmacy, UIT - The Arctic University of Norway, Tromsø, Norway and <sup>3</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

**Integrons are genetic elements that are common in bacteria and are hotspots for genome evolution. They facilitate the acquisition and reassembly of gene cassettes encoding a variety of functions, including drug resistance. Despite their importance in clinical settings, the selective forces responsible for the evolution and maintenance of integrons are poorly understood. We present a mathematical model of integron evolution within bacterial populations subject to fluctuating antibiotic exposures. Bacteria carrying a functional integrase that mediates reshuffling of cassette genes and thereby modulates gene expression patterns compete with bacteria without a functional integrase. Our results indicate that for a wide range of parameters, the functional integrase can be stably maintained in the population despite substantial fitness costs. This selective advantage arises because gene-cassette shuffling generates genetic diversity, thus enabling the population to respond rapidly to changing selective pressures. We also show that horizontal gene transfer promotes stable maintenance of the integrase and can also lead to *de novo* assembly of integrons. Our model generates testable predictions for integron evolution, including loss of functional integrases in stable environments and selection for intermediate gene-shuffling rates in changing environments. Our results highlight the need for experimental studies of integron population biology.**

*The ISME Journal* (2016) 10, 1296–1307; doi:10.1038/ismej.2015.222; published online 5 February 2016

## Introduction

Mobile genetic elements (MGEs) allow bacteria to rapidly adapt to changing environments through the acquisition of novel traits that range from single genes to entire catabolic pathways (Top and Springael, 2003). Of particular concern is the ability of MGEs to capture and spread antibiotic resistance determinants by horizontal gene transfer (HGT).

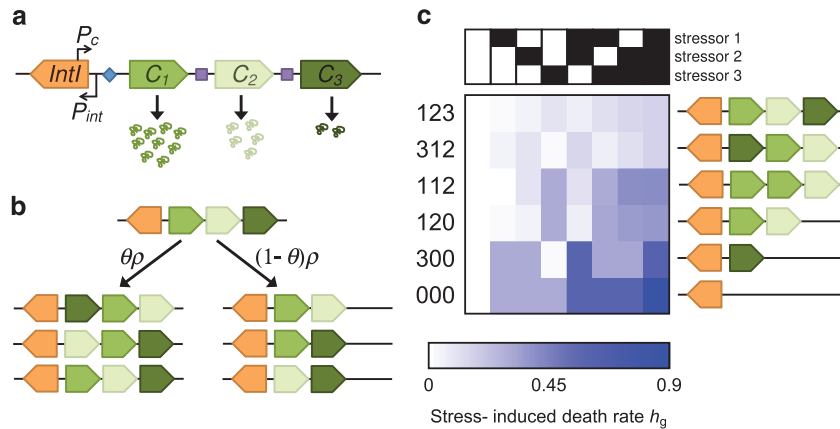
The presence of antibiotic resistance determinants on MGEs makes it increasingly difficult to curb the emergence and spread of drug resistance because these elements have an evolutionary life of their own (Bergstrom *et al.*, 2000; Touchon *et al.*, 2014). Consequently, if frequencies of antibiotic resistance are reduced following, for example, reduced selective pressures, then MGEs harboring resistance determinants can continuously move on to other recipients of the same or different species and promote their persistence (Bergstrom *et al.*, 2000; Baquero *et al.*, 2013). To gain more knowledge on the basic processes controlling the frequencies of antibiotic resistance in bacterial populations, it is

essential that we better understand the population dynamics and evolution of bacteria with acquired MGEs encoding antibiotic resistance in both the presence and absence of antibiotic selective pressures (Johnsen *et al.*, 2009). With the exception of work focusing on plasmids (Bouma and Lenski, 1988; Dahlberg and Chao, 2003; San Millan *et al.*, 2014; Gullberg *et al.*, 2014), surprisingly few studies have addressed the impact of MGE acquisitions on bacterial population dynamics (but see Starikova *et al.*, 2012; Starikova *et al.*, 2013). One class of MGEs with a prominent role in the emergence and spread of antibiotic resistance determinants where studies on the population dynamics following acquisition are particularly limited are the mobile resistance integrons, and especially the class 1 integrons.

Integrons are genetic elements with the ability to capture, express and shuffle one of the smallest MGEs known, the gene cassettes (Stokes and Hall, 1989). Structurally, integrons consist of an integrase (*intI*) gene, an attachment site for acquisition of gene cassettes (*attI*), and an array of gene cassettes that is highly variable in number and composition (Figure 1a). More than 100 integrase genes have been described (Boucher *et al.*, 2007). An integron integrase is capable of site-specific recombination between *attI* and the gene-cassette-borne attachment

Correspondence: J Engelstädter, School of Biological Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia. E-mail: j.engelstaedter@uq.edu.au

Received 29 April 2015; revised 13 October 2015; accepted 16 October 2015; published online 5 February 2016



**Figure 1** Illustration of the main model features. Panel (a) gives a schematic representation of a simple integron consisting of an integrase gene (*IntI*, orange) and three cassette genes ( $C_1$ – $C_3$ , green). The integrase is expressed from the promoter  $P_{int}$ , whereas the cassette array is expressed from the promoter  $P_c$ . Gene expression of the cassettes is assumed to decline with increasing distance from the  $P_c$  promoter. Attachment sites for integrase-mediated gene-shuffling are shown in blue (*attI*) and purple (*attC*). Panel (b) shows how integrase-mediated cassette shuffling can lead to new genotypes. Cassettes are excised from any position at rate  $\rho$  and are then, with probability  $\theta$ , re-inserted in the first position of the array. Starting from the genotype  $C_1C_2C_3$ , six genotypes can be produced depending on which cassette is excised and whether or not it is re-inserted. Panel (c) shows how different stress environments impact the stress-induced death rate  $h_g$ . The top chart shows all eight combinations of presence (black) and absence (white) of three stressors. The chart below shows the stress-induced death rate in these environments for six example array genotypes consisting of different cassettes. These genotypes are shown schematically (right) and as mathematically as vectors (left). The stress-induced death rate ranges from zero (white) to  $3\eta_s = 0.9$  (blue).

site (*attC*), or between two *attC* sites (Collis and Hall, 1992; Collis *et al.*, 2001). One promoter located in this region,  $P_{int}$ , facilitates expression of *intI* and another,  $P_c$ , enables transcription of the gene cassettes. Integrase expression is often controlled by the bacterial SOS response, requiring *lexA* derepression for expression (Guerin *et al.*, 2009; Cambray *et al.*, 2011). Presence of binding sites for repressors, transcription factors or other DNA-binding proteins can further modulate expression of the integrase (Cagle *et al.*, 2011).

In their simplest form, gene cassettes contain a promoterless open reading frame followed by an *attC* site. The integron integrase can excise gene cassettes, generating transiently free circular DNA molecules, and insert free gene cassettes preferentially at *attI* (reviewed in Cambray *et al.*, 2010; Escudero *et al.*, 2015). There is a negative correlation between the distance of the gene cassettes from  $P_c$  and their expression levels, except for rare occasions where internal promoters are present (for example, Weldhagen, 2004). Genes encoded by gene cassettes can be antibiotic resistance determinants or other genes that confer selective advantage, but also addiction modules or genes of unknown functions ('ORFans'). The region downstream of the gene cassettes is variable, and in the clinically relevant class I integrons this region often contains a truncated *qacE1* and a *sul1*; both genes are thought to have been important in the early evolution of these elements (Gillings, 2014; Gillings *et al.*, 2014). Integrons have been traditionally grouped into those associated with MGEs (mobile integrons) and the large chromosomal superintegrons, but there is no structural difference that distinguishes superintegrons from MGE-borne integrons (Boucher *et al.*,

2007; Cambray *et al.*, 2011). However, the MGE-associated integrons are often involved in the acquisition and spread of antibiotic resistance determinants, and a number of different classes have been described (reviewed in Gillings, 2014). Class 1 integrons have been extensively studied mechanistically with respect to gene-cassette acquisition. They are widely dispersed in Proteobacteria (including most clinical Gram-negatives) and occasionally found in Gram-positives (Fluit and Schmitz, 1999; Partridge *et al.*, 2009).

The potential adaptive effects of gene-cassette acquisition as well as the clinical relevance of antibiotic resistance-encoding gene cassettes have received much attention since integrons were discovered in the late 1980s (Stokes and Hall, 1989). Moreover, the main mechanistic aspects of integration, excision and expression of gene cassettes have been elucidated in detail (Dubois *et al.*, 2007; Dubois *et al.*, 2009; Loot *et al.*, 2012). Yet, the selective forces responsible for the evolution and maintenance of integrons are poorly understood. The recent documentation of substantial fitness costs imposed by newly acquired integrase genes (Starikova *et al.*, 2012) highlights the need to identify evolutionary benefits of integrases and integrons that can overcome such fitness costs. Experimental evidence indicates that through reshuffling of cassettes within the integron, the bacteria gain access to genes that were previously not expressed because they were too distant from the promoter (Collis and Hall, 1995; reviewed in Cambray *et al.*, 2010), but this idea has never been formally scrutinized.

Here, we construct and analyze mathematical models to gain a better understanding of integron evolutionary dynamics. Specifically, we consider a

population of bacteria that is subject to a number of stochastically changing stressors (for example, antibiotics). The bacteria may carry integrons consisting of an integrase gene as well as a number of cassette genes providing resistance to these stressors. We focus on short and relatively simple integrons, as best exemplified by clinical class 1–3 integrons (see discussion section). We are interested in the selective forces that maintain functional integrases in the face of fitness costs and mutations rendering them inactive. Three versions of the model are considered: a basic model where the integrase reshuffles cassette genes at a constant rate, a model extension where reshuffling only occurs under stressful conditions and a second extension where new cassettes can be incorporated through HGT.

## Methods

We constructed a mathematical model of a bacterial population in which bacteria with a functional and with an inactivated integrase compete with each other (see Figure 1 for an illustration). For a full description of this model, see the Supplementary Information. In brief, in bacteria carrying the functional integrase, the gene cassettes are reshuffled at a certain rate. This affects their patterns of gene expression and, as a result, the level of resistance to a number of stressors that the bacteria experience and that fluctuate stochastically through time. Our model is described by the following system of ordinary differential equations:

$$\begin{aligned}\frac{dX_g}{dt} &= X_g \left( 1 - \frac{N}{K} \right) - (\eta_0 + \eta_I + h_g) X_g - \mu X_g \\ &\quad + \rho \left( \sum_{j \in G} X_j \left( (1 - \theta) M_{jg}^{\text{exc}} + \theta M_{jg}^{\text{int}} \right) - X_g \right) \\ \frac{dY_g}{dt} &= Y_g \left( 1 - \frac{N}{K} \right) - (\eta_0 + h_g) Y_g + \mu X_g\end{aligned}$$

**Table 1** Parameters of the model with their assumed standard values

Parameter	Description	Standard value
$n$	Number of different gene cassettes	3
$k$	Maximum number of cassettes in integron	3
$K$	Carrying capacity of population	$10^9$
$\eta_0$	Baseline death rate	0.1
$\eta_I$	Increase in death rate caused by functional integrase	0.001
$\eta_S$	Increase in death rate caused by stressor in absence of resistance gene	0.3
$\rho$	Rate of integrase-mediated gene reshuffling	0.001
$\theta$	Probability at which the integrase re-inserts an excised gene cassette	0.5
$\beta$	Parameter determining how fast gene expression declines with increasing distance from promoter	0.5
$\gamma$	Shape parameter determining how expression level of a resistance gene affects death rate	3
$\sigma_{\text{mean}}$	Average fraction of time that a given stressor is present in the population	0.2
$\sigma_{\text{vel}}$	Average rate at which switches between presence and absence of a stressor occur	0.01
$\mu$	Mutation rate from functional to non-functional integrase	$10^{-5}$
$\varphi$	Threshold level of stress-induced death above which the integrase becomes active (model extension 1)	0.3
$\tau$	Rate of HGT (model extension 2)	$10^{-16}$

Abbreviation: HGT, horizontal gene transfer.

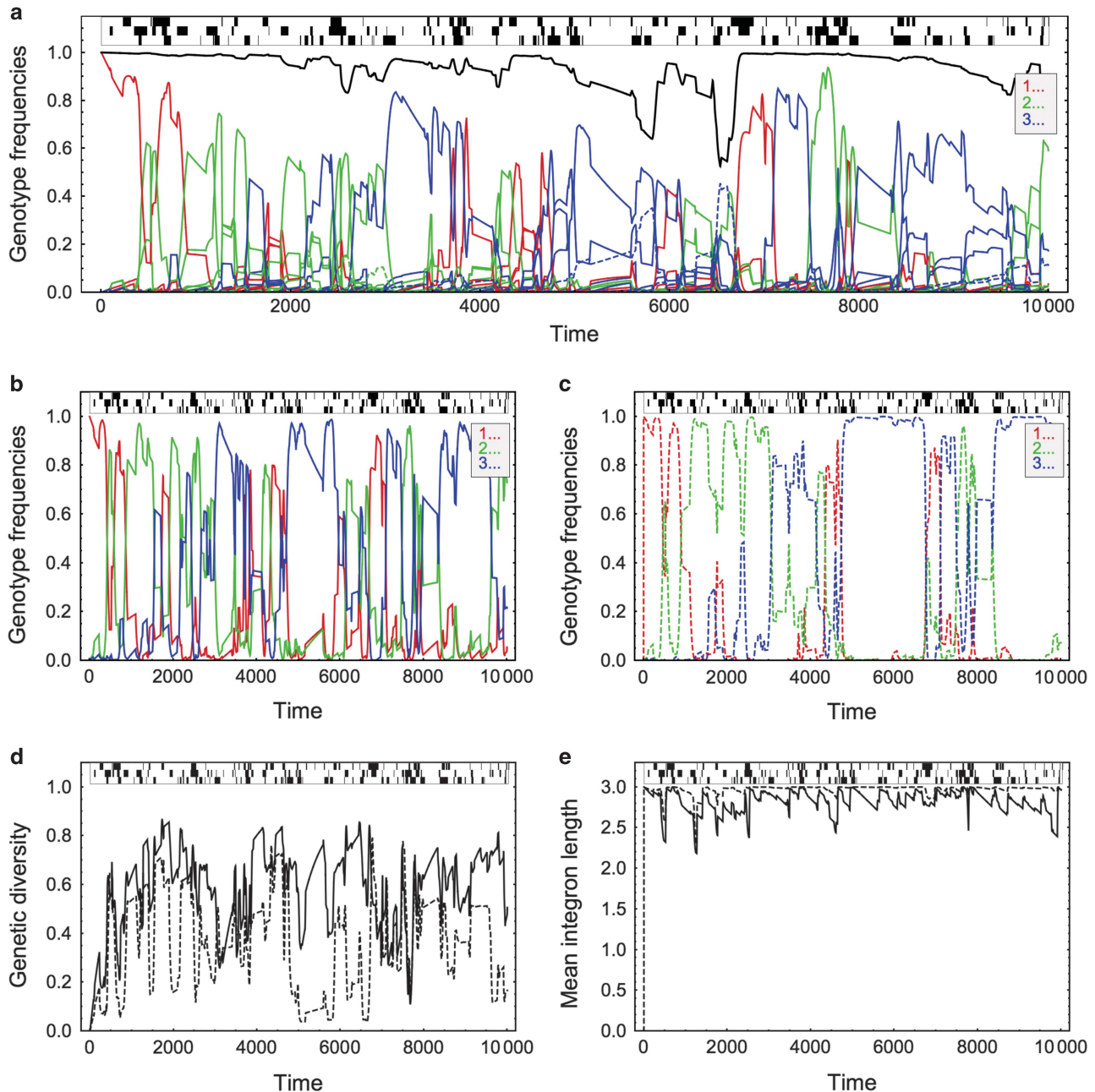
Here,  $X_g$  and  $Y_g$  are the abundances of bacteria with and without a functional integrase, respectively, for different genotypes  $g$ . The four terms in the equations for  $dX_g/dt$  correspond to bacterial growth, death, mutational inactivation of the integrase and integrase-mediated cassette shuffling. The model equations were solved numerically using the software package Mathematica v10.0 (Wolfram Research, Inc., Champaign, IL, USA). Unless stated otherwise, all simulations were run for 10 000 time units and in 100 replicates; simulations were generally started from a homogeneous initial population of  $10^6$  bacteria carrying a functional integrase. All parameters of the model with their standard values are listed in Table 1. Parameter values chosen are based on experimental results where available (Collis and Hall, 1995; Starikova *et al.*, 2012; Discussion section). We added two extensions to the model in which we included stress-dependent integrase activity, and HGT of gene cassettes.

## Results

### Basic model

Figure 2 shows an example for the evolutionary dynamics emerging from our model. It can be seen that owing to the fluctuating selection pressure on the population caused by the stochastically changing environment (that is, absence or presence of three stressors), genotype frequencies in the population change rapidly. Specifically, whenever a single stressor is present, genotypes that carry the corresponding resistance cassette gene at the first position within the integron are selectively favored. Similarly, when two stressors are present, genotypes carrying the two corresponding resistance cassette genes in the first two positions are favored.

It can also be seen that throughout the simulation, the functional integrase is maintained at a high frequency in the population (black line in Figure 1a), with one drop to about 0.5 but subsequent recovery.



**Figure 2** Example evolutionary dynamics of an integron. Throughout, solid lines denote bacteria with an intact integrase, whereas dashed lines denote bacteria without a functional integrase. The bar above each plot shows the presence (black bars) or absence (white) of each of three stressors to which the cassette genes 1, 2 and 3 provide resistance, with stressor 1 shown on top and stressor 3 at the bottom. Panel (a) shows the frequencies of the different genotypes in the population. Colors denote integron genotypes with different cassettes in the first position of the integron, as indicated in the legend; genotypes with the same leading cassette are shown in the same color. The bold black line gives the total frequency of genotypes carrying a functional integrase. Panels (b and c) give the total frequencies of genotypes carrying a particular cassette in the first position among all bacteria with or without a functional integrase, respectively. (d) gives the genetic diversity (computed as the probability that two randomly drawn cassette array genotypes are different) and (e) the mean number of cassettes in integrons within the two subpopulations. All parameters take the standard values given in Table 1.

This is despite a number of fitness costs imposed on bacteria carrying a functional integrase, including the assumed constitutive increase in death rate  $\eta_i$ , but also the fact that often the ‘wrong’ cassette gene will be incorporated in the first integron position and that often cassette genes are lost following excision (leading to integrons that are shorter on average;

Figure 2e). The reason that the functional integrase is maintained in spite of these fitness costs is that by shuffling of cassette genes, the integrase increases genetic diversity. This can be seen by comparing genotype frequencies among bacteria with and without the functional integrase (Figures 2b vs c), and also by comparing the levels of genotype diversity in

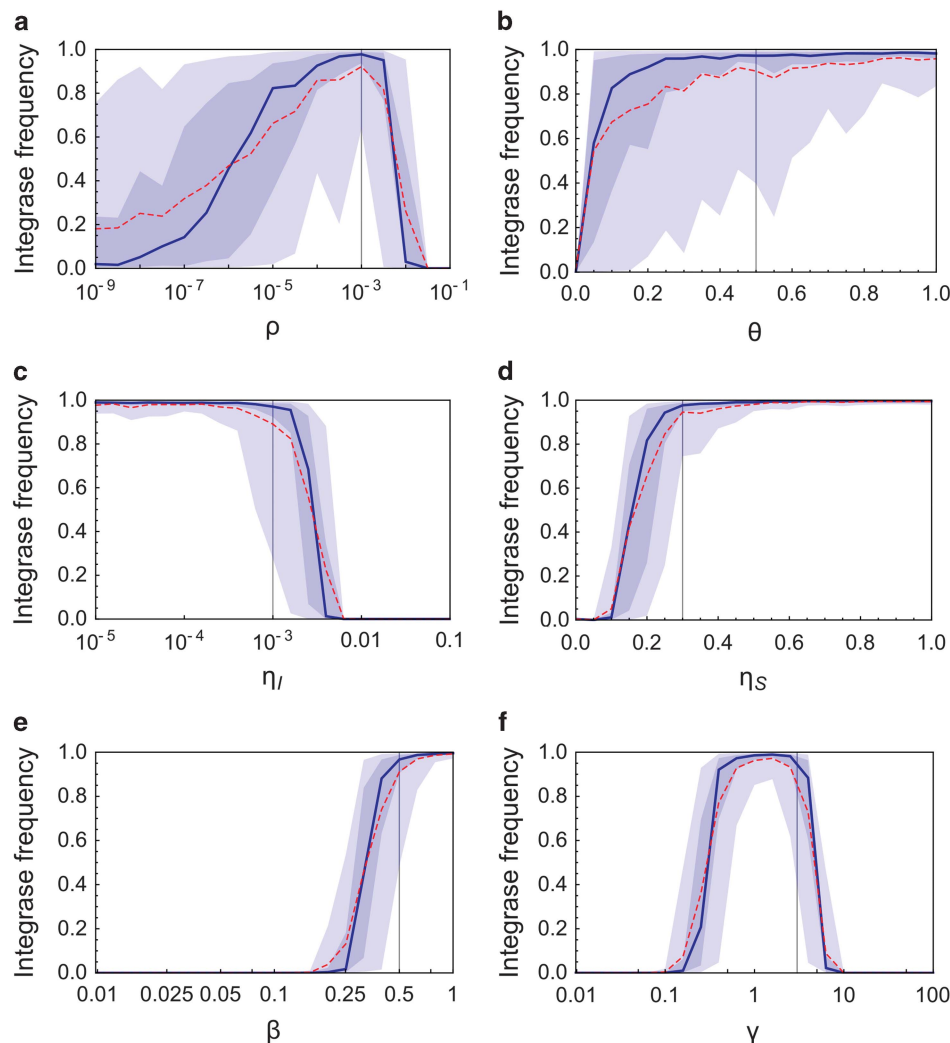


these two subpopulations (Figure 2d). As a result of this increased genetic diversity and the ensuing higher variance in fitness, the bacterial subpopulation with the functional integrase can respond more rapidly to changing selection pressure.

Given both the natural variation observed in some of the parameters of our model and the dearth of experimental estimates of others (Discussion section), it is important to ascertain how varying the different parameters over a wide range of plausible values impact the resulting dynamics. Figure 3 shows how varying the main parameters of the model while keeping the other parameters fixed affect the final frequency of the functional integrase in the population. At very low rates  $\rho$  of integrase-mediated gene shuffling, the selective advantage of integrase-carrying bacteria will be very small and the functional integrase will be gradually lost from the population (Figure 3a). On the other hand, at very high rates of gene shuffling,

the integrase becomes strongly deleterious and is also lost. This is because favorable genotypes will rapidly be eliminated through integrase action in this case, including through excision events that are not followed by reintegration of cassettes. The highest integrase frequency, close to one, is therefore seen with intermediate rates of gene shuffling.

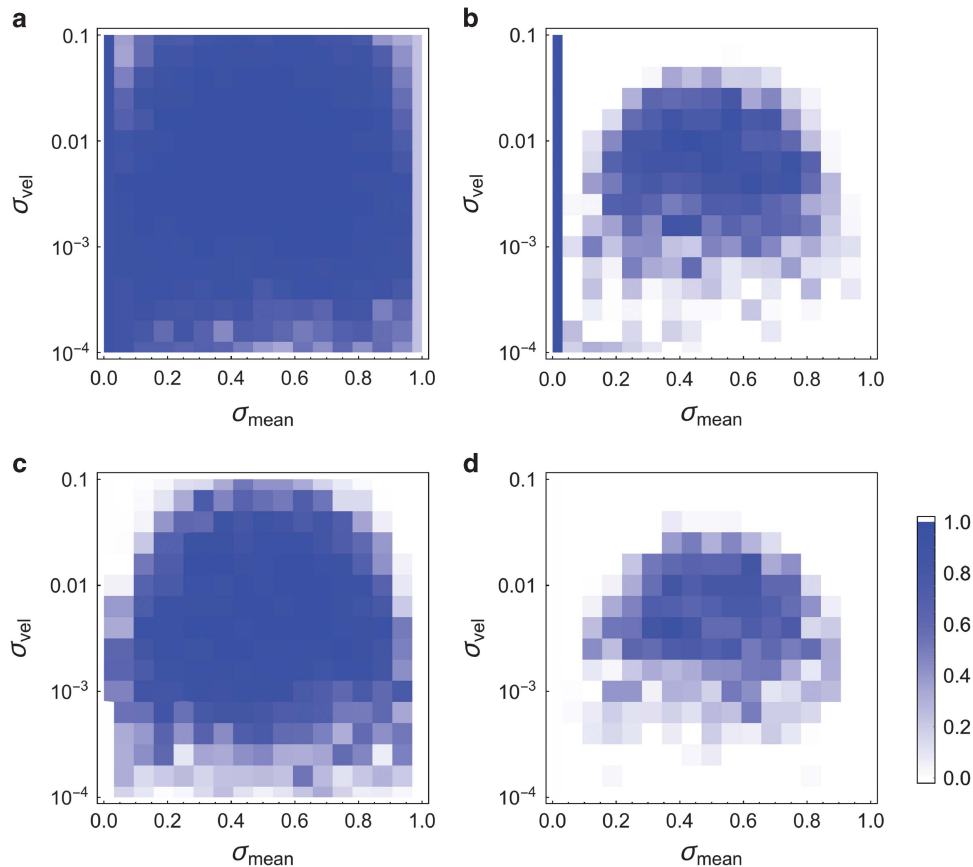
Figure 3b shows the impact of the parameter  $\theta$ , the fraction of cassette genes that are re-inserted following excision. This plot shows that the integrase is selected for even if a relatively high proportion (up to 80%, corresponding to  $\theta = 0.2$ ) of cassettes are not re-inserted. In Figures 3c and d the impact of the fitness cost  $\eta_I$  of the functional integrase as well as the mortality  $\eta_S$  imposed by the stressors is explored. As expected, the integrase frequency decreases from one to zero with increasing  $\eta_I$ , but increases from zero to one with increasing  $\eta_S$ . This means that an integrase that is very costly can only be maintained



**Figure 3** Impact of different parameters on the final integrase frequency. The plots show how (a) the rate  $\rho$  of integrase-mediated cassette shuffling, (b) the cassette reintegration probability  $\theta$ , (c) the fitness cost  $\eta_I$  of the integrase, (d) the stress-induced death rate  $\eta_S$ , (e) the gene expression parameter  $\beta$  and (f) the stress resistance parameter  $\gamma$  influence the final frequency of the functional integrase. For each plot, one parameter was varied, whereas all other parameters take standard values. The blue line gives the median frequency, the dashed red lines the mean frequency and the shaded areas the interquartile and 90% interquartile range of final integrase gene frequencies in the population. Gray vertical lines indicate the standard value of each parameter.

at low frequencies or is not maintained at all in the population. On the other hand, the higher the stress-induced mortality rate  $\eta_S$  is, the stronger does selection for the integrase become. Finally, Figures 3e and f show the impact of the parameters  $\beta$  and  $\gamma$  that determine how gene order affects fitness in the presence of stressors. The integrase frequency increases monotonically with increasing  $\beta$ , the parameter that determines how strongly gene expression declines within the integron. When  $\beta$  is too low, gene expression of a cassette will only weakly be affected by its position in the integron, so that gene shuffling becomes irrelevant and there is no selection for a functional integrase. The parameter  $\gamma$ , determining how gene expression affects resistance levels to a given stressor, needs to take intermediate values to lead to selection for the integrase: when  $\gamma$  is too low, even strongly expressed genes provide little resistance and when  $\gamma$  is too high even weakly expressed genes provide strong resistance; in both cases gene order within the integron becomes irrelevant for fitness (Supplementary Figure S1B). Example dynamics where the parameters  $\rho$ ,  $\theta$ ,  $\eta_I$  and  $\eta_S$  are individually varied relative to their standard values are shown in Supplementary Figure S2.

We next investigated how environmental change affects selection for the functional integrase. To this end, we screened both the average stress level,  $\sigma_{\text{mean}}$  and the velocity  $\sigma_{\text{vel}}$  with which the stressors appear and disappear (Figure 4). We observed that the integrase frequency is maximized for intermediate values of both of these parameters. This result can be understood intuitively by considering extreme values of both parameters. When  $\sigma_{\text{mean}}$  is very low, there will usually be no stressor present, so that on average there will be very weak selection acting on cassette genes and hence for the integrase. When  $\sigma_{\text{mean}}$  is very high, all stressors will be present at once most of the time, so that gene order does not matter much and selection for the integrase will also be weak. (Note, however, that even though gene-cassette order does not matter much in this case, the presence of the cassettes is of course important. This means there will be selection against integrases that excise but do not reinsert cassettes.) When the velocity at which the environment changes is very low, time periods in which reshuffling of gene cassettes is beneficial will be rare, so that again there will be only weak selection for the integrase. Finally, when the velocity of environmental change is very high, the population will essentially



**Figure 4** Mean frequency of the functional integrase gene for different mean stress levels  $\sigma_{\text{mean}}$  and rates  $\sigma_{\text{vel}}$  at which stress conditions change. The integrase is assumed to impose no fitness cost in **a** and **b** ( $\eta_I = 0$ ), but is costly in panels **c** and **d** ( $\eta_I = 0.001$ ). The activity of the integrase is moderate in panels **a** and **c** ( $\rho = 0.001$ ) and relatively high in panels **b** and **d** ( $\rho = 0.01$ ). Other parameters take standard values as indicated in Table 1.

experience only a homogeneous average stress level in which all three stressors are present; gene order also does not matter in this case.

Integron length varies widely across different bacteria (Cambray *et al.*, 2010). As such, it is important to ascertain the impact of cassette numbers on selection for the integrase. Supplementary Figure S4 shows that with increasing number  $n$  of cassette genes, the integrase is maintained at a higher frequency in the population. Not surprisingly, the integrase goes extinct rapidly for  $n=0$  (empty integron) and  $n=1$  (no cassette shuffling possible). For  $n=2$  and  $n=3$ , the integrase clearly provides a fitness benefit but observed frequencies vary widely between zero and one. For  $n=4$ , the integrase is strongly selected for and always maintained at a high frequency. We expect that this trend of increasing strength of selection for the functional integrase with increasing integron length will continue with  $n>4$ , but due to computational restrictions we were unable to demonstrate this. (With  $n=5$ , there are already 7812 possible genotypes in the model.) The reason that selection becomes stronger with increasing  $n$  is that for any given value of  $\beta$ , gene expression at the end of the cassette array will become increasingly weak. This means that those cassettes can contribute little to stress resistance and selection for the functional integrase that can move these cassettes to the first position becomes stronger.

#### Model extension 1: stress-dependent integrase activity

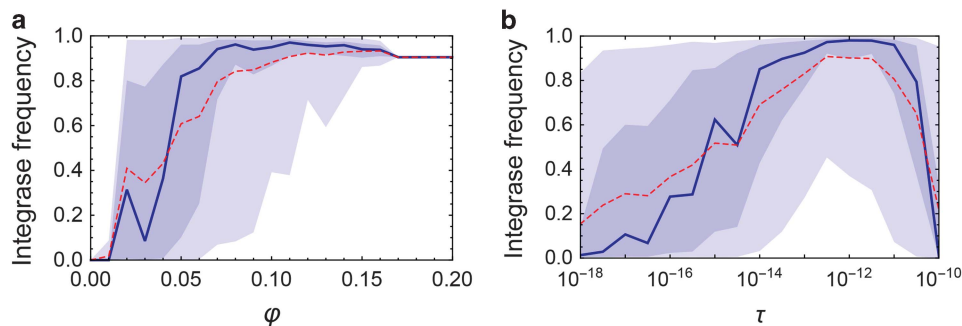
Integrase expression is often not constitutive but controlled by LexA and thus triggered by the SOS response (Guerin *et al.*, 2009; Cambray *et al.*, 2011). Therefore, we next investigated an extension of our model where the integrase is only active (and costly) when the stress-induced mortality exceeds a certain threshold  $\varphi$  (Supplementary Information). Figure 5a shows the impact of this threshold on the evolutionary maintenance of an integrase that is very costly ( $\eta_I=0.01$ ) and as a result not maintained when the other parameters take standard values (Figure 3c). When  $\varphi$  is very low, the integrase is active even at very

low levels of stress so that the model becomes similar (or, with  $\varphi=0$ , identical) to the basic model and the functional integrase is lost. At the other extreme, when  $\varphi$  is very high the integrase is never active and thus not under selection. Rather, the functional integrase slowly declines in frequency due to mutational decay. Therefore, it is at intermediate values of  $\varphi$  that the integrase reaches the highest frequency in the population. Here, the integrase is only activated ‘when needed’ so that the benefit of cassette reshuffling within the integron at times of stress is combined with an absence of fitness costs in times of little stress. This effect is further illustrated in Supplementary Figure S5 showing the evolutionary dynamics with the same stress environment as in Figure 2 and Supplementary Figure S2, but with increasing values of  $\varphi$ .

The advantage of stress-induced compared with constitutive integrase activity is also apparent when the impact of other parameters on the frequency of the integrase is investigated. As Supplementary Figure S6 shows (compared with Figure 3), these parameters have little impact on the frequency of the integrase. In particular, the integrase can be stably maintained at a high frequency unless almost no excised cassettes are re-inserted ( $\theta$  close to zero) or the integrase is very costly (large  $\eta_I$ ). This is because for intermediate stress thresholds  $\varphi$ , the integrase is essentially neutral in the worst case (for example, with low values of  $\rho$ ,  $\eta_S$  or  $\varphi$ ), but can still be highly advantageous in other cases (for example, when  $\eta_S$  is large).

#### Model extension 2: HGT

So far, integrase action was limited to reshuffling genes within the integron. We next considered a model extension where cassettes from other bacteria in the population may also be incorporated into the integron by means of HGT (Supplementary Information). Specifically, we assume here that cassette genes occurring within the population (in living bacteria or as free DNA) may be transferred to recipient bacteria where the integrase incorporates these cassettes into the first position of the integron. In our model, HGT thus represents an



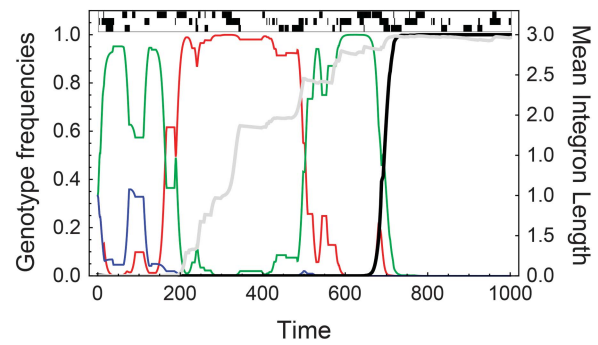
**Figure 5** Final integrase frequency in model extensions 1 and 2. The blue line gives the median frequency, the dashed red lines the mean frequency and the shaded areas the interquartile and 90% interquartile range of final integrase gene frequencies in the population. Panel (a) shows the impact of stress-threshold  $\varphi$  (above which the integrase is expressed) on the final frequency of the functional integrase (model extension 1). Panel (b) shows the impact of the rate  $\tau$  of HGT mediated by the integrase on its final frequency (model extension 2). All parameters take standard values as indicated in Table 1 except (a)  $\eta_I=0.01$  and (b)  $\rho=0$ .

alternative pathway of how cassette arrays can be modified that may operate independently from the array rearrangements triggered by cassette excision considered previously.

To compare these two pathways, simulations were run first in the absence of gene reshuffling ( $\rho = 0$ ; see Supplementary Figure S8 for example dynamics). Figure 5b shows that like reshuffling, HGT can also lead to the stable maintenance of the functional integrase in the population in spite of its fitness costs. Moreover, the final frequency of the functional integrase also increases with increasing rate  $\tau$  of HGT, but decreases at very high rates  $\tau$ , as was the case for very high values of  $\rho$ . To compare the effect of the two rates,  $\tau$  needs to be adjusted by the number of individuals in the population because HGT depends on both the number of donor and recipient individuals in the population. When multiplying  $\tau$  by a factor of  $K = 10^9$  (a rough estimate for population size) and comparing the corresponding values, it can be seen that for low to intermediate rates of  $\tau$ , the effect of HGT is very similar to that of within-genome shuffling (Supplementary Figure S7).

At very high rates of  $\rho$  and  $\tau$ , the integrase reaches a slightly higher frequency with HGT than with within-genome shuffling when  $\theta = 0.5$  (cassette loss in 50% of cases), but a lower frequency when  $\theta = 1$  (all excised cassettes are re-inserted). This can be explained by the fact that in contrast to reshuffling, HGT can also produce integron arrays containing cassette duplications. Although this may sometimes confer higher stress resistance, it means that arrays produced by HGT are generally less variable than arrays containing only different cassettes. When HGT occurs in addition to reshuffling (both  $\rho > 0$  and  $\tau > 0$ , see purple line in Supplementary Figure S7), it appears that low to intermediate rates of HGT have little impact on the frequency of the functional integrase, indicating that the two pathways interact non-linearly in producing selection for integrase action.

Finally, we considered a scenario of *de novo* integron evolution. This scenario is highly simplistic but may still capture important aspects of real-life integron evolution that we have not addressed so far, including the emergence of integrons carrying antibiotic resistance genes in clinical settings (Discussion section). We assume that initially, the population consists only of bacteria that carry each of the resistance cassette genes individually, but no (or no functional) integrase. This will result in fluctuating frequencies of these genotypes in the population but no generation of genotypes carrying  $> 1$  cassette. We then assume that bacteria carrying only a functional integrase but no cassette genes enter the population at a low frequency. Specifically, we assume that the initial numbers of bacteria in the population are  $X_{000}(0) = 10^6$  and  $Y_{100}(0) = Y_{200}(0) = Y_{300}(0) = 10^8$ , with all other bacterial numbers being zero initially. By means of HGT, the bacteria carrying the functional integrase can then capture the different cassette genes,



**Figure 6** Example dynamics showing how HGT can favor the invasion of bacteria with a functional integrase. Here, the population consists initially of predominantly integrase-free bacteria carrying individual cassette genes, with only few bacteria that carry a functional integrase but no cassette genes. The three colored lines show the frequencies of bacteria without a functional integrase that carry one of three cassette genes, providing resistance to each of the three stressors (shown on top as in Figure 2). The bold black line gives the total frequency of bacteria carrying a functional integrase. Finally, the bold gray line gives the average number of cassette genes among all bacteria carrying a functional integrase (scale shown on right hand y axis). All parameters take the standard values shown in Table 1, with the exceptions of  $\sigma_{vel} = 0.05$ ,  $\mu = 0$  and  $\eta_S = 0.2$ .

combining them into a single genotype and thus forming an integron of increasing length (gray line in Figure 6). Those bacteria that have formed the integron have achieved a fitness advantage because carrying several cassette genes provides better protection from stressors than a single cassette gene, even if not all of these cassettes are fully expressed. Thus, the bacteria carrying the fully formed integron can now spread to fixation in the population (bold black line in Figure 6).

## Discussion

We have studied a mathematical model exploring the selective maintenance of integron integrases within a population of bacteria exposed to changing stressful conditions. Our results indicate that reshuffling of cassettes within an integron can confer a selective advantage to a functional integrase that comes about through changes in gene expression that are often strong enough to outweigh substantial fitness costs of the integrase. Our model, albeit focusing on a single population, reproduces several empirical observations seen across bacterial strains and species, including a relatively high prevalence of mutationally inactivated integrases in both clinical isolates and experimentally evolved strains (Gillings *et al.*, 2005; Nemergut *et al.*, 2008; Starikova *et al.*, 2012) and cassette arrays that are highly variable in terms of cassette composition and order (reviewed in Gillings, 2014). Taken together, our results support the view that integron integrases are selectively maintained because they enable their hosts to efficiently use cassette arrays as a ‘reservoir of standing genetic variability’ (Cambray *et al.*, 2010).



Integron diversity encompasses attributes that our model does not account for. Our model is based on characterization of class 1 integrons and other mobile resistance Integrons found in clinical settings (reviewed in Gillings, 2014). These integrons are generally short ( $\leq 8$  cassettes), have cassettes that carry single genes that almost always confer antibiotic resistance, and generally carry a single  $P_C$  promoter from which the cassettes are expressed. An important feature that is not included in our model, however, is the mobility of these integrons. Interestingly, integrases in clinical class 2 integrons are almost always inactivated by the same mutation producing an internal stop codon (Hansson *et al.*, 2002). Viewed in the light of our model, this suggests that following their emergence as resistance integrons in pathogens, the ancestors of the bacteria carrying these integrons may have experienced prolonged stable environmental conditions so that selection to maintain a functional integrase was relaxed.

Many integrons are more complex than the clinical class 1–3 integrons. For example, the chromosomal integrons in different *Vibrio* species are much larger (some carry  $> 200$  cassettes) and more diverse. Many carry toxin–antitoxin systems (for example, Rowe-Magnus *et al.*, 2003; Szekeres *et al.*, 2007; reviewed in Cambray *et al.*, 2010), some carry their own promoters or have  $> 1$  open reading frame, and ORFans are prevalent. Interestingly, however, in cases where the function can be inferred it appears that the gene products produced from the cassettes are often involved in interactions with the environment that may be subject to rapid change (for example, functions associated with host interactions and virulence, interaction with phages and biofilm formation). This suggests that our results are relevant beyond shorter mobile integrons of classes 1–3, and that the extraordinary gene-cassette diversity in chromosomal integrons in environmental bacteria (for example, Gillings *et al.*, 2009; Boucher *et al.*, 2011; Koenig *et al.*, 2011) can be explained by positive selection for an active integron integrase.

Our model rests on the assumption that expression levels of cassette genes depend on their position within the integron: with increasing distance from  $P_C$ , expression levels decline. This assumption is well supported. With only few exceptions (Bissonnette *et al.*, 1991; Stokes and Hall, 1991; Naas *et al.*, 2001), the gene cassettes themselves are promoterless, so that their expression relies on  $P_C$  (Cambray *et al.*, 2010). Mechanistically, transcription initiated by  $P_C$  or by other internal promoters can span several cassette genes, and generally expression of cassette genes decreases with distance from the promoter (Collis and Hall, 1995; Coyne *et al.*, 2010). Consequently, integron gene-cassette dynamics modeled here is also relevant for internal gene-cassette promoters other than  $P_C$ . The decreased promoter activity can be attributed to the presence of stem-loop secondary structures at the *attC* sites

that modulate ribosome binding and in turn affect translation (Jacquier *et al.*, 2009), or act as weak transcriptional terminators. For example, Collis and Hall (1995) observed roughly equal amounts of RNA transcripts containing only the first and the first two cassettes. This corresponds to a parameter of  $\beta = \ln 2 \approx 0.69$  in our model, which is close to our standard value of  $\beta = 0.5$ . However, we still lack a general and accurate description of cassette expression levels within integrons that would be essential for more quantitative predictions of integron evolution.

Another knowledge gap becomes evident when modeling gene-cassette shuffling. Reordering of gene cassettes has been observed experimentally (Collis and Hall, 1992) and is believed to occur by excision of a gene cassette and recapture of that cassette (Hall, 2012). Studies performed *in vitro* show that integrase-mediated cassette excision is preferentially catalyzed at two *attC* sites (releasing cassettes at position 2 or downstream), and that free gene cassettes are inserted typically at *attI* (in the first position close to  $P_C$ ) (Collis *et al.*, 1993; Collis *et al.*, 2002). However, to the best of our knowledge, no experimental data are available for the frequency at which excised gene cassettes are actually re-inserted. Although we assume a standard value of  $\theta = 0.5$  for the probability of gene-cassette recapturing, our model demonstrates that for an integrase to be stably maintained, reinsertion efficiencies down to about  $\theta = 0.2$  may be sufficient in changing environments (Figure 2b).

The gene-shuffling rate  $\rho$  is a function of the integrase activity, and we expect that integrase expression is an evolutionarily adjusted process that depends on  $P_{int}$ , the *intI1* sequence, and the host genotype. Indeed, our simulations support the view that natural selection should optimize integrase expression because the integrase is most strongly favored at intermediate shuffling rates. Conversely, very high integrase activity can be deleterious due to occasional loss of excised cassettes as well as through the creation of unfavorable cassette orders within the integron. This shows that overexpressed integrases can be costly not only through pleiotropic effects in the host bacterium (for example, caused by ectopic recombination activities; Recchia *et al.*, 1994; Harms *et al.*, 2013), but also for the integron itself. Both of these costs can be strongly reduced when integrase expression is stress-dependent, as has been suggested by recent studies reporting SOS-induced expression of *intI* (Guerin *et al.*, 2009; Cambray *et al.*, 2011). Our model shows that indeed, for suitable thresholds of stress experienced by the bacteria (model parameter  $\phi$ ), such stress-induced integrase expression can strongly promote the maintenance of functional integrases within bacterial populations.

Our study also highlights a close relationship between integron and mutation rate evolution in bacteria, including mutator strains. Mutators are

characterized by an elevated mutation rate, usually caused by mutations in genes responsible for DNA mismatch correction (Denamur and Matic, 2006; Raynes and Sniegowski, 2014). In a well-adapted population, mutators are disfavored because of their increased genetic load. However, in an adapting population mutations that result in an increased mutation rate may hitchhike to high frequencies along with beneficial mutations that they help produce. Similarly, functional integrases can be expected to decline in frequency when a bacterial population is well adapted to a stable environment. At best, integrases will decay neutrally through mutational pressure, but they may also be selected against due to fitness costs. On the other hand, in populations under selective pressure, active integrases can be indirectly selected for because even though most integron rearrangements their effect may be detrimental, they can still hitchhike along with those that provide a decisive fitness advantage.

HGT allows for gene flow within a population and can lead to acquisition of genes from unrelated taxa. Consequently, HGT is a major factor in the spread of antibiotic resistance determinants (Davies and Davies, 2010; Abel zur Wiesch *et al.*, 2011; Levin *et al.*, 2014). In the extension of our model where we incorporated HGT, the free flow of gene cassettes in the population substantially contributes to the stable maintenance of functional integrases in fluctuating environments. Here, the advantage is achieved both by moving gene cassettes into the first position within the integron, and by capturing new or previously lost gene cassettes (see also Starikova *et al.* (2012) for an earlier and simpler model of integrase-mediated HGT). This model extension can also help explain repeats of identical gene cassettes at different positions in the same integron (for example, Elsaied *et al.*, 2011). Notably, in populations with static integrons that only carry defective integrase genes, for example, as result of long-term stable environmental conditions, HGT is strongly beneficial in our model for genotypes that carry functional integrases but no suitable gene cassettes. Here, HGT provides the gene cassettes required for integron platform assembly and shuffling. As a result, the model infers that carriers of functional integrases can rise to high frequencies over time when the environment has changed from static to fluctuating.

We have presented here the first population biological model for the maintenance of integrase-mediated gene shuffling. Our principal aim was not to present a model capable of making quantitative predictions regarding integron evolution (which seems impossible given the currently available data), but to formally scrutinize hypotheses on integron evolution and to identify parameters that are predicted to be important (see also Servedio *et al.* (2014) for a discussion of ‘proof-of-concept’ models). Clearly, there is much scope for future theoretical work exploring integron evolutionary dynamics.

For example, whereas our model involved competitions between bacteria with and without a functional integrase, it would be interesting to investigate the evolution of integrase activity, taking into account potential constraints of this parameter with respect to  $P_C$  promoter strength (Jove *et al.*, 2010; Guerin *et al.*, 2011). Even more importantly, our study highlights the need for experimental studies estimating crucial parameters influencing integron evolution, including fitness effects, levels of gene expression across cassette arrays and excision/reintegration rates of cassettes.

## Conflict of Interest

The authors declare no conflict of interest.

## Acknowledgements

We would like to thank the editor and three anonymous reviewers for their insightful comments on a previous version of this paper. This work was supported by an Australian Research Council Future Fellowship to JE (project number FT140100907), a postdoctoral fellowship by the Danish National Research Foundation to KH and grants by the Norwegian Research Council (project number 204263) and the Northern Norway Regional Health Authority to PJJ.

## References

- Abel zur Wiesch P, Kouyos R, Engelstädter J, Regoes RR, Bonhoeffer S. (2011). Population biological principles of drug-resistance evolution in infectious diseases. *Lancet Infect Dis* **11**: 236–247.
- Baquero F, Tedim AP, Coque TM. (2013). Antibiotic resistance shaping multi-level population biology of bacteria. *Front Microbiol* **4**: 15.
- Bergstrom CT, Lipsitch M, Levin BR. (2000). Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics* **155**: 1505–1519.
- Bissonnette L, Champetier S, Buisson JP, Roy PH. (1991). Characterization of the nonenzymatic chloramphenicol resistance (cmlA) gene of the In4 integron of Tn1696: similarity of the product to transmembrane transport proteins. *J Bacteriol* **173**: 4493–4502.
- Boucher Y, Cordero OX, Takemura A, Hunt DE, Schliep K, Baptiste E *et al.* (2011). Local mobile gene pools rapidly cross species boundaries to create endemicity within global *Vibrio cholerae* populations. *MBio* **2**: e00335–10.
- Boucher Y, Labbate M, Koenig JE, Stokes HW. (2007). Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends Microbiol* **15**: 301–309.
- Bouma JE, Lenski RE. (1988). Evolution of a bacteria plasmid association. *Nature* **335**: 351–352.
- Cagle CA, Shearer JE, Summers AO. (2011). Regulation of the integrase and cassette promoters of the class 1 integron by nucleoid-associated proteins. *Microbiology* **157**: 2841–2853.
- Cambray G, Guerout AM, Mazel D. (2010). Integrons. *Annu Rev Genet* **44**: 141–166.

- Cambray G, Sanchez-Alberola N, Campoy S, Guerin E, Da Re S, Gonzalez-Zorn B *et al.* (2011). Prevalence of SOS-mediated control of integron integrase expression as an adaptive trait of chromosomal and mobile integrons. *Mob DNA* **2**: 6.
- Collis CM, Grammaticopoulos G, Briton J, Stokes HW, Hall RM. (1993). Site-specific insertion of gene cassettes into integrons. *Mol Microbiol* **9**: 41–52.
- Collis CM, Hall RM. (1992). Site-specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. *J Bacteriol* **174**: 1574–1585.
- Collis CM, Hall RM. (1995). Expression of antibiotic-resistance genes in the integrated cassettes of integrons. *Antimicrob Agents Chemother* **39**: 155–162.
- Collis CM, Kim MJ, Stokes HW, Hall RM. (2002). Integron-encoded IntI integrases preferentially recognize the adjacent cognate attI site in recombination with a 59-bp site. *Mol Microbiol* **46**: 1415–1427.
- Collis CM, Recchia GD, Kim MJ, Stokes HW, Hall RM. (2001). Efficiency of recombination reactions catalyzed by class 1 integron integrase IntI1. *J Bacteriol* **183**: 2535–2542.
- Coyne S, Guigon G, Courvalin P, Perichon B. (2010). Screening and quantification of the expression of antibiotic resistance genes in *Acinetobacter baumannii* with a microarray. *Antimicrob Agents Chemother* **54**: 333–340.
- Dahlberg C, Chao L. (2003). Amelioration of the cost of conjugative plasmid carriage in *Escherichia coli* K12. *Genetics* **165**: 1641–1649.
- Davies J, Davies D. (2010). Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* **74**: 417–433.
- Denamur E, Matic I. (2006). Evolution of mutation rates in bacteria. *Mol Microbiol* **60**: 820–827.
- Dubois V, Debreyer C, Litvak S, Quentin C, Parissi V. (2007). A new in vitro strand transfer assay for monitoring bacterial class 1 integron recombinase IntI1 activity. *PLoS One* **2**: e1315.
- Dubois V, Debreyer C, Quentin C, Parissi V. (2009). In vitro recombination catalyzed by bacterial class 1 integron integrase IntI1 involves cooperative binding and specific oligomeric intermediates. *PLoS One* **4**: e5228.
- Elsaied H, Stokes HW, Kitamura K, Kurusu Y, Kamagata Y, Maruyama A. (2011). Marine integrons containing novel integrase genes, attachment sites, attI, and associated gene cassettes in polluted sediments from Suez and Tokyo Bays. *ISME J* **5**: 1162–1177.
- Escudero JA, Loot C, Nivina A, Mazel D. (2015). The integron: adaptation on demand. *Microbiol Spectr* **3**: MDNA3-0019–2014.
- Fluit AC, Schmitz FJ. (1999). Class 1 integrons, gene cassettes, mobility, and epidemiology. *Eur J Clin Microbiol Infect Dis* **18**: 761–770.
- Gillings MR. (2014). Integrons: past, present, and future. *Microbiol Mol Biol Rev* **78**: 257–277.
- Gillings MR, Gaze WH, Pruden A, Smalla K, Tiedje JM, Zhu YG. (2014). Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J* **9**: 1269–1279.
- Gillings MR, Holley MP, Stokes HW, Holmes AJ. (2005). Integrons in *Xanthomonas*: a source of species genome diversity. *Proc Natl Acad Sci USA* **102**: 4419–4424.
- Gillings MR, Xuejun D, Hardwick SA, Holley MP, Stokes HW. (2009). Gene cassettes encoding resistance to quaternary ammonium compounds: a role in the origin of clinical class 1 integrons? *ISME J* **3**: 209–215.
- Guerin E, Cambray G, Sanchez-Alberola N, Campoy S, Erill I, Da Re S *et al.* (2009). The SOS response controls integron recombination. *Science* **324**: 1034–1034.
- Guerin E, Jove T, Tabesse A, Mazel D, Ploy MC. (2011). High-level gene cassette transcription prevents integrase expression in class 1 integrons. *J Bacteriol* **193**: 5675–5682.
- Gullberg E, Albrecht LM, Karlsson C, Sandegren L, Andersson DI. (2014). Selection of a multidrug resistance plasmid by sublethal levels of antibiotics and heavy metals. *MBio* **5**: e01918–14.
- Hall RM. (2012). Integrons and gene cassettes: hotspots of diversity in bacterial genomes. *Ann N Y Acad Sci* **1267**: 71–78.
- Hansson K, Sundstrom L, Pelletier A, Roy PH. (2002). IntI2 integron integrase in Tn7. *J Bacteriol* **184**: 1712–1721.
- Harms K, Starikova I, Johnsen PJ. (2013). Costly Class-1 integrons and the domestication of the functional integrase. *Mob Genet Elements* **3**: e24774.
- Jacquier H, Zaoui C, Sanson-Le Pors MJ, Mazel D, Bercot B. (2009). Translation regulation of integrons gene cassette expression by the attC sites. *Mol Microbiol* **72**: 1475–1486.
- Johnsen PJ, Townsend JP, Bohn T, Simonsen GS, Sundsfjord A, Nielsen KM. (2009). Factors affecting the reversal of antimicrobial-drug resistance. *Lancet Infect Dis* **9**: 357–364.
- Jove T, Da Re S, Denis F, Mazel D, Ploy MC. (2010). Inverse correlation between promoter strength and excision activity in class 1 integrons. *PLoS Genet* **6**: e1000793.
- Koenig JE, Bourne DG, Curtis B, Dlutek M, Stokes HW, Doolittle WF *et al.* (2011). Coral-mucus-associated *Vibrio* integrons in the Great Barrier Reef: genomic hotspots for environmental adaptation. *ISME J* **5**: 962–972.
- Levin BR, Baquero F, Johnsen PJ. (2014). A model-guided analysis and perspective on the evolution and epidemiology of antibiotic resistance and its future. *Curr Opin Microbiol* **19**: 83–89.
- Loot C, Ducos-Galand M, Escudero JA, Bouvier M, Mazel D. (2012). Replicative resolution of integron cassette insertion. *Nucleic Acids Res* **40**: 8361–8370.
- 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.
- Nemergut DR, Robeson MS, Kysela RF, Martin AP, Schmidt SK, Knight R. (2008). Insights and inferences about integron evolution from genomic data. *BMC Genomics* **9**: 261.
- Partridge SR, Tsafnat G, Coiera E, Iredell JR. (2009). Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiol Rev* **33**: 757–784.
- Raynes Y, Sniegowski PD. (2014). Experimental evolution and the dynamics of genomic mutation rate modifiers. *Heredity* **113**: 375–380.
- Recchia GD, Stokes HW, Hall RM. (1994). Characterisation of specific and secondary recombination sites recognised by the integron DNA integrase. *Nucleic Acids Res* **22**: 2071–2078.
- Rowe-Magnus DA, Guerout AM, Biskiri L, Bouige P, Mazel D. (2003). Comparative analysis of superintegrons: engineering extensive genetic diversity in the Vibrionaceae. *Genome Res* **13**: 428–442.

- San Millan AS, Pena-Miller R, Toll-Riera M, Halbert ZV, Mclean AR, Cooper BS *et al.* (2014). Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids. *Nat Commun* **5**: 5208.
- Servedio MR, Brandvain Y, Dhole S, Fitzpatrick CL, Goldberg EE, Stern CA *et al.* (2014). Not just a theory—the utility of mathematical models in evolutionary biology. *PLoS Biol* **12**: e1002017.
- Starikova I, Al-Haroni M, Werner G, Roberts AP, Sorum V, Nielsen KM *et al.* (2013). Fitness costs of various mobile genetic elements in *Enterococcus faecium* and *Enterococcus faecalis*. *J Antimicrob Chemother* **68**: 2755–2765.
- Starikova I, Harms K, Haugen P, Lunde TT, Primicerio R, Samuelsen O *et al.* (2012). A Trade-off between the fitness cost of functional integrases and long-term stability of integrons. *PLoS Pathog* **8**: e1003043.
- Stokes HW, Hall RM. (1989). A novel family of potentially mobile dna elements encoding site-specific gene-integration functions - integrons. *Mol Microbiol* **3**: 1669–1683.
- Stokes HW, Hall RM. (1991). Sequence analysis of the inducible chloramphenicol resistance determinant in the Tn1696 integron suggests regulation by translational attenuation. *Plasmid* **26**: 10–19.
- Szekeres S, Dauti M, Wilde C, Mazel D, Rowe-Magnus DA. (2007). Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection. *Mol Microbiol* **63**: 1588–1605.
- Top EM, Springael D. (2003). The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Curr Opin Biotechnol* **14**: 262–269.
- Touchon M, Bobay LM, Rocha EPC. (2014). The chromosomal accommodation and domestication of mobile genetic elements. *Curr Opin Microbiol* **22**: 22–29.
- Weldhagen GF. (2004). Integrons and beta-lactamases—a novel perspective on resistance. *Int J Antimicrob Agents* **23**: 556–562.

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>).