The chromosomal accommodation and domestication of mobile genetic elements

Marie Touchon, Louis-Marie Bobay, Eduardo P C Rocha

To cite this version:

The chromosomal accommodation and domestication of mobile genetic elements

Marie Touchon\textsuperscript{a,b}, mtouchon@pasteur.fr
Louis-Marie Bobay\textsuperscript{a,b,c}, louis-marie.bobay@pasteur.fr
Eduardo PC Rocha\textsuperscript{a,b,*}, erocha@pasteur.fr

\textsuperscript{a}Microbial Evolutionary Genomics, Institut Pasteur, 75724 Paris, France
\textsuperscript{b}CNRS, UMR3525, 75724 Paris, France
\textsuperscript{c}Sorbonneuniversités, UPMC Univ Paris06, IFD, 4 place Jussieu, 75252 Paris cedex05, France.
\textsuperscript{*}Corresponding author. Tel: +33 1 45 68 33 53; Fax: +33 1 44 27 97 79; Email: erocha@pasteur.fr; Present Address: Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris

Keywords: Evolution, prokaryotes, viruses, comparative genomics.

Abstract

Prokaryotes are constantly being infected by large mobile genetic elements such as conjugative elements and temperate phages. The fitness of these elements is tightly linked with the evolutionary success of the host. This leads to selection against disruptive effects their integration might have on the organization and structure of the chromosome. Seamless genetic accommodation of the mobile elements also involves silencing infectious mechanisms and expressing functions adaptive to the host. Ironically, these characteristics favor the host ability to domesticate the mobile element. Recent data suggests that the domestication of mobile elements might be frequent. Importantly, it might affect the evolution of chromosome organization and drive the diversification of social traits.
Introduction

The genomes of Prokaryotes are extremely plastic because of the joint action of horizontal gene transfer (HGT) and differential gene loss [1-3]. Most genetic information arrives into the genome in large mobile genetic elements (MGEs), such as plasmids, integrative conjugative elements (ICEs) and temperate phages [4,5]. These large elements are the focus of the present review. MGEs are important motors of bacterial adaptation, rapidly spreading new functions and social traits, even if they have short residence times in genomes [6]. Several striking examples show the profound impact of MGEs. ICEs and insertion sequences make a third of the genomes of Orientia tsutsugamushi [7]. Certain genomes of Escherichia coli encode up to 18 prophages [8]. The symbiotic ICE of Mesorhizobium loti is ~500 kb [9], which is more than certain bacterial chromosomes. The integration of such large genetic elements may impose a significant burden to the cell [10]. The cost of transfer, in particular, leads to a trade-off between vertical and horizontal transmission [11]. Too much horizontal transfer decreases bacterial fitness and therefore vertical transmission. Infrequent transfer increases vertical at the cost of horizontal transmission. Importantly, both the element and the host are favored by the vertical transmission of host-MGE associations. The interaction between integrative MGE and the host is therefore favored when it leads its smooth accommodation in the host genetic background.

Chromosomal accommodation

The genetic information encoded in prokaryotic chromosomes is highly organized [12] (Figure 1). A very large fraction of the genome encodes genes or regulatory sequences needed for the interactions between the chromosome and cellular molecular machines. Since random MGE integration is likely to impose a fitness cost on the host, these elements have evolved strategies to integrate at permissive locations using site-specific recombinases. MGEs integrating at ubiquitous highly conserved sequences are also more likely to find an integration site in another host and thus proliferate both vertically and horizontally. Different mobile elements encoding closely related integrases tend to use the same integration regions in the chromosome. For example, the highly conserved tmRNA site in Escherichia, Salmonella, and Klebsiella is an integrative hotspot for pathogenicity islands, prophages, and integrative conjugative elements. The high sequence conservation of stable RNA genes might contribute to the frequency with
which they are used as integration sites[13,14]. Interestingly, several temperate phages integrate the genome of *E. coli* in intergenic regions that are highly conserved and used by other phages in *Salmonella*[15]. The high conservation of these non-coding sequences suggests the existence of selection for their preservation as integration sites. Since vertical replication is in general less costly than horizontal propagation, site-specific integration might result from host-MGE co-evolution towards a seamless and reproducible accommodation of the mobile element.

Site-specific integration may facilitate the organization of genome plasticity in relation to the structure of the chromosome, which is highly condensed in the nucleoid to be able to fit the prokaryotic cell[16]. Depending on the methods and the species, chromosomes are structured at different scales that include macrodomains (>500 kb) [17], chromosome interaction domains (30-420 kb, CID) [18], and negatively supercoiled loops (2-66 kb) [19]. The association between chromosome structure and transcriptional networks renders gene expression dependent on chromosome organization [20]. In a similar way, chromosome structure is tightly linked with segregation [21]. The integration of MGEs alters the structure of the chromosome and might thus affect many cellular processes. While there are few reports on this subject, it has recently been observed that phage Mu, which integrates almost randomly the chromosome, forms an autonomous stable chromosomal domain in *E. coli* [22]. This might facilitate transposition of the element and favor the expression of adaptive genes harbored by the prophage. Interestingly, phage lambda shows strong preference for integration at one very specific site in the chromosome and does not form a folding domain. It is tempting to speculate that MGEs integrating at specific sites in the chromosome, such as phage lambda, have co-evolved with the chromosome such that their integration does not lead to deleterious consequences in terms of chromosome structure. Elements integrating randomly in the chromosome, such as phage Mu, must evolve ways of doing so without disturbing chromosomal structure, like producing autonomous structural units.

In *E. coli* and *Salmonella*, prophages tend to integrate at the macrodomains closer to the terminus of replication [15], even though this region encodes fewer typical integration targets such as stable RNA genes [23]. Several MGEs even integrate specifically at *difs* sites using the host Xer recombination system [24]. These patterns suggest that integration tropism and/or natural selection drive the large-scale distribution of MGEs in the
chromosome. Integration of elements carrying their own recombinases could be favored by DNA accessibility because Ter-proximal macromdomains are at the periphery of the *E. coli* nucleoid [25], or by the region of the chromosome closer to the infection site [26]. MGE integration sites could also be counter-selected in certain chromosomal regions. For example, MGEs integration close to highly expressed genes might be disfavored because transcription spillover from these genes into the MGE might reduce repression on its genes and favor excision and transfer. Accordingly, the tRNAs most frequently used as integration targets by *E. coli* phages are those corresponding to the rarer codons, which are also the least expressed [15]. Genes near the origin of replication are on average more expressed. They can also be over-expressed by replication-associated gene dosage effects, especially in periods of fast-growth or stress [23, 27]. This may produce an effect similar to transcriptional spillover and explain why these regions usually encode few prophages. Non-lethal mobile elements, whose expression might be highly adaptive in certain circumstances, show different patterns of distribution. For example, pathogenicity islands concentrate in the origin-proximal half of the *E. coli* chromosome (Figure 1). Finally, chromosomal genes flanking frequent MGE integration sites often have higher recombination rates [28, 29], and this might lead to selection for integration close genes encoding specific functions. These few studies on the global and local patterns of distribution of MGEs suggest that site-specific integration has evolved towards lowering its disruptive effects on chromosome organization. Further studies will be required to disentangle the effects of integration tropism and natural selection on this co-evolutionary process.

MGEs endure selection to adapt to their genetic background. This involves the ability of the temperate phage to integrate and manipulate the host genetic networks. A classical example concerns the stress-induced excision of prophages and integrative conjugative elements by sensing the host SOS response [30, 31]. Recent works have shown that the integration or manipulation of the host genetic networks by the MGE may stabilize its integrative state. For example, the stabilization of several prophages in *E. coli* is under the control of the highly conserved transcriptional terminator Rho protein [32]. Several *Staphylococcus aureus* prophages are regulated by the host alternative σH factor, also involved in the regulation of natural competence [33]. The high mutation rates of prophages provide them with ample opportunity to avoid DNA motifs used by hosts to control their gene expression. The conservation of these regulatory dependencies
suggests the existence of co-evolution between host and MGE to stabilize the integrative state. The integration of MGEs in the host genetic network allows them to regulate the trade-off between horizontal and vertical transmission in relation to cell physiology.

Adaptation to the genetic context of the host can also include selection for MGE-encoded DNA motifs associated with housekeeping cellular functions, such as chromosome replication, recombination, and segregation[34-36]. Accordingly, prophages closer to the terminus of replication are enriched in segregation-related motifs and those in the rest of the chromosome under-represent Ter-macromdomain associated motifs(Figure 1) [15]. The accommodation of MGEto the host chromosome by selecting for DNA motifs that are only relevant in the integrative state may superficially seem in contradiction with their short residence times in genomes. Yet, most well-studied integrative mobile elements have relatively narrow host ranges and integrate at the same genomic position in the different hosts. They will therefore endure the same type of selection pressures in the different hosts. This should favor their accommodation to the traits that are often encountered in its range of host chromosomes.

**Domestication of mobile elements**

MGEs encode many traits useful for the host. These traits increase indirectly the fitness of the element. Famously, many conjugative elements encode antibiotic resistance genes [37], and many phages provide toxins to bacterial pathogens [38]. In some cases it is unclear if a trait is directly advantageous to the MGE or if it is favored because it increases the host fitness. For example, some MGEs favor bacterial growth under certain conditions[39-41]. Interestingly, even costly core MGE functions, such as conjugation and transduction, can provide advantages to the host. Conjugative pili facilitate the formation of biofilms [42], and can be co-opted to secrete virulence factors to eukaryotic cells[43]. Horizontal transfer of the host DNA can occur by co-transfer with the MGE [44,45]. Occasional lysis caused by prophages favors the production of biofilms [46], promotes the bacterial adhesion to eukaryotic cell [47,48], and accelerates niche colonization by removing sensitive strains [49].

Genes encoding the mechanisms of horizontal transmission are strongly repressed some time after transfer, presumably to lower their burden on the host metabolism[50,51]. We have discussed above on how seamless integration of the MGE in the chromosome will lower its cost and therefore increase its fitness. Ironically, long-term repression of the
genes encoding genetic mobilization poses a threat to MGE integrity because inactivating mutations may accumulate silently. Elements rendered incapable of autonomous horizontal transmission are much more dependent on the host fitness and evolve predominantly towards favoring vertical transmission.

Silent genetic information is rapidly lost from bacterial lineages by a joint effect of a bias towards deletions and by selection against non-adaptive genes and pseudogenes [1,3,52]. Most prophages are able to excise and eventually kill the cell but many are incapable of producing autonomous infectious virions [8,48]. One would therefore expect strong selection for the rapid loss of these elements. Surprisingly, degraded prophages of *E. coli* and *Salmonella enterica* evolve predominantly under strong purifying selection - most of the non-synonymous mutations occurring in prophages are lost - suggesting that natural selection removes variants with inactivating mutations [53]. Importantly, this affects most strongly the genes encoding phage housekeeping functions, like lysins and terminases, raising the intriguing possibility that bacteria may systematically select for phage-related functions in degraded genetic elements.

What could be the use of these domesticated phage functions? A sparse but long string of observations shows that inactive prophages or prophage-derived components can be adaptive [54]. They could favor horizontal gene transfer like transducing phages. They could antagonize other mobile elements like many phages, either by competing with them [55], or by inhibiting their entry and expression [56]. They could also make virions unable to produce viable offspring but capable of killing sensitive competitors [54]. Finally, prophages provide many accessory functions, like adhesion, regulation, and defense; several of which are used by bacteria in mutualistic or antagonistic interactions with eukaryotes [48,57-61]. Many of these functions require phage excision, transcription, and packaging, justifying why large fractions of prophages are under purifying selection. The organization of gene expression in phages around few large operons may also lead to counter-selection of gene inactivation to avoid polar mutations.

Elements under purifying selection are expected to remain in genomes for long periods of time. Yet, most phage integration events in *E. coli* are strain-specific, *i.e.*, they occurred very recently [53,62]. How can prophages be under purifying selection when they are also frequently strain-specific? Many prophages may bring no adaptive value upon
degradation and be rapidly lost, leaving only a small fraction of the prophages accumulating mutations. For the latter, the functional redundancy provided by multiple prophages may explain why they may undergo purifying selection and still be frequently lost. The constant afflux of prophages leads to a multiplicity of elements in genomes. This inevitably results in relaxed selection on elements that perform similar MGE-related functions; leading to frequent loss of prophage remnants that were previously under purifying selection. Most processes of mobile elements domestication are thus likely to be short-lived.

Nevertheless, examples of longstanding domestication of MGE machineries have been described (Figure 2). Many involve only one or a few genes, such as the acquisition of polymerases, lysins, resolvases or recombinases[63], leading to either analogous gene replacement, or to new variants of pre-existing functions. For example, prophage-encoded Redβ recombinases in E. coli compensate the loss of the host recombination machinery while showing higher processivity and tolerance to sequence divergence [64]. Domestication may also produce new sophisticated molecular machines[65]. Conjugative systems have been many times domesticated as type IV protein secretion systems used both by pathogens and mutualists in their interactions with eukaryotes[66,67]. Plasmids have been domesticated as secondary chromosomes in several bacterial clades[68]. Phage-derived tail proteins are part of type VI secretion systems (T6SS) involved in antagonistic interactions with other bacteria and with eukaryotes[69]. Phages have also given rise to full-fledged tailocins (also called pyocins) involved in bacterial competition or pathogenicity [70,71]. Gene transfer agents (GTA) encapsidate randomly the DNA of the bacterial genome (but not their own genome) and are thought to have derived from prophage domestication [72]. Many other observations show important, albeit mechanistically unclear, roles for MGE-derived structures. For example, phage tail-like structures produced by Pseudoalteromonas luteoviolacea trigger the metamorphosis of a tubeworm [60].

**Conclusion**

Most recent very large-scale whole-genome studies have focused on the conserved parts of genomes. These are very helpful to understand epidemiological and mutational patterns but not how adaptation proceeds by horizontal gene transfer. New sequencing technologies producing longer reads should bring back the complete assembled genome
as the standard in bacterial genomics. Larger datasets will facilitate the study of mobile elements accommodation in genomes including their subsequent domestication. They might also guide experimental testing of the many hypotheses that have been put forward regarding these subjects. Which functions often result from the domestication of mobile elements? Available data suggests that functions implicated in social evolution may predominate because mobile elements encode many secreted proteins [73], and complex symbiotic traits [74]. Machineries encoded by MGEs, or derived from them, are also used in bacterial warfare [75] and in symbiosis [47, 76]. Understanding the fate of MGE in genomes is therefore likely to illuminate the constraints imposed by the structure of the chromosome on the dynamics of gene repertoires and also on the evolution of the social lives of bacteria.
**Outstanding questions**

What is the impact of MGE integration on the chromosome structure?

What are the structural and evolutionary traits creating integration hotspots?

What are the main functions provided by domesticated MGE?

What are the rates of MGE domestication?

**Acknowledgements**

We thank members of our lab for discussions. Work in our laboratory is supported by an European Research Council starting grant [EVOMOBIOME no°281605] to EPCR, a grant from the Ministère de l'enseignementsupérieur et de la recherche to LMB, and funding from the CNRS and the Institut Pasteur. We apologize to the authors of studies we could not cite given space restrictions.


This detailed study of the 18 prophages of *Escherichia coli* O157:H7 Sakai shows that most prophages are defective at several levels (see also reference 48).


An excellent review on the costs of horizontal gene transfer.


Temperate phages adapt to the local and global organizational traits of the bacterial chromosome.


MGE increase the rate of recombination in chromosomal genes neighboring their integration sites in Staphylococcus aureus, a species with overall low recombination rates.


Regulation by Rho, an essential E. coli protein, stabilizes prophages.


Shows the various cellular functions of cryptic prophage genes and their impact on bacterial physiology.


**Accurate analysis of the functionality of Enterococcus faecalis prophages with a focus on the interactions between the different prophages (see also reference 8).**


52. Lawrence JG, Hendrix RW, Casjens S: **Where are the pseudogenes in bacterial genomes?** *Trends Microbiol* 2001, **9**:535-540.

53**. Bobay LM, Touchon M, Rocha EPC: **Pervasive domestication of defective prophages by bacteria.** *Proc Natl Acad Sci U S A* 2014, **accepted pending minor revisions.**

**Identification and study of hundreds of vertically inherited prophages. Most of these are partially degraded and evolve under purifying selection.**


55. Refardt D: **Within-host competition determines reproductive success of temperate bacteriophages.** *The ISME Journal* 2011, **5**:1451-1460.

56. Bondy-Denomy J, Davidson AR: **When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness.** *J Microbiol* 2014, **52**:235-242.
57. Bensing BA, Siboo IR, Sullam PM: **Proteins PblA and PblB of Streptococcus mitis, which promote binding to human platelets, are encoded within a lysogenic bacteriophage.** *Infect Immun* 2001, **69**:6186-6192.


Describes an intriguing phage tail-like structure involved in inducing the metamorphosis of a marine tubeworm by a bacterium. This is an interesting example of a non-antagonistic interaction between bacteria and eukaryotes mediated by phage-like particles.


Reports how a prophage induction does not lead to the production of virions and instead is used as a genetic switch by *Listeria monocytogenes* promote escape from phagosomes.


69. Leiman PG, Basler M, Ramagopal UA, Bonanno JB, Sauder JM, Pukatzki S, Burley SK, Almo SC, Mekalanos JJ: **Type VI secretion apparatus and phage tail-associated


Figure legends.

**Figure 1 - Chromosomal organization (top).** (1) The *E. coli* circular chromosome is organized into 4 structured macrodomains (MD) and 2 unstructured flexible chromosomal regions (NS)[17]. (2) The macrodomainTer (MD-Ter, black) is preferentially located at the periphery of the nucleoid [25]. (3) MatP defines the MD-Ter by binding to *matS* sequence motifs [34]. SlmA binds to DNA motifs frequent in MD-Ori and the flanking NS-regions to prevent chromosome fragmentation during septum formation[35]. (4) Chromosomal replication proceeds bi-directionally (black arrow) from a single origin (*oriC*, yellow) to the opposite termination site (*dif*, grey). The leading (resp. lagging) strands are represented in green (resp. red) and are different in terms of composition (GC skew), number of genes, and Chi and KOPS motifs [12]. (5) Presence of
multiple replication forks in fast-growing bacteria produces a transient replication-associated gene dosage effect that leads to selection of highly expressed genes (violet) near the origin of replication[23]. (6) KOPS (FtsK-orienting polar sequence, brown) are polarized motifs very frequent in the Ter-proximal regions. They orient the last stages of chromosome segregation[36]. MGE accommodation (bottom). (1) Transposable prophage Mu (green) integrates randomly in the genome, whereas most other prophages (red) integrate more frequently in the MD-Ter and the flanking MD-Left and Right. Pathogenicity islands (PAI, blue) are more frequent in the other half of chromosome (i.e. MD-Ori and NS-regions)[15]. (2) Mu forms a stable chromosomal domain. (3) Mobile elements using site-specific recombinases are concentrated in a few integrative hotspots. PAI integrated at tRNA genes, whereas prophages also use other targets. (4) The occupancy rate and the number of integrative hotpots increase with the distance from the origin, in inverse relation with replication-associated gene dosage effects. (5) Prophages show avoidance of MatP and over-representation of KOPS that are relevant only at the prophage state in the context of the biology of the host.
Figure 2. Schematic representation of phage and conjugation-related domesticated elements. **Top row:** elements involved in bacteria-bacteria interactions or gene transfer: transducing phages, gene transfer agents (GTA), killer particles (defective phage particles), tailocins, type VI secretion systems (T6SS), type IV secretion systems (T4SS) involved in protein and DNA secretion and in DNA uptake (competence)[65,67]. **Lower row:** elements involved in interactions with Eukaryotes: *Photorhabdus* virulence cassettes (PVC) and antifeeding prophages (Afp) are toxic to certain eukaryotes[58,59]. Multi-MAC arrays contain phage tail-like contractile structures and induce marine tubeworm metamorphosis[60]. Prophage-encoded platelet-binding factors which promote bacterial binding to human platelets and induce endovascular infection[47,48]. T4SS from conjugative systems and T6SS derived from phages are used by pathogens and mutualists to secrete effectors to eukaryotic cells[43,77].