Accepted Manuscript

Bringing them together: plasmid pMV158 rolling circle replication and conjugation under an evolutionary perspective

Fabián Lorenzo-Díaz, Cris Fernández-López, M. Pilar Garcillán-Barcia, Manuel Espinosa

PII: S0147-619X(14)00036-5
DOI: http://dx.doi.org/10.1016/j.plasmid.2014.05.004
Reference: YPLAS 2208

To appear in: Plasmid

Received Date: 24 March 2014
Accepted Date: 22 May 2014

Please cite this article as: Lorenzo-Díaz, F., Fernández-López, C., Pilar Garcillán-Barcia, M., Espinosa, M., Bringing them together: plasmid pMV158 rolling circle replication and conjugation under an evolutionary perspective, Plasmid (2014), doi: http://dx.doi.org/10.1016/j.plasmid.2014.05.004

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Bringing them together: plasmid pMV158 rolling circle replication and conjugation under an evolutionary perspective

Fabián Lorenzo-Díaz\textsuperscript{a*}, Cris Fernández-López\textsuperscript{b*}, M. Pilar Garcillán-Barcia\textsuperscript{c§} and Manuel Espinosa\textsuperscript{b§}

\textsuperscript{a}Unidad de Investigación, Hospital Universitario Nuestra Señora de Candelaria and Instituto Universitario de Enfermedades Tropicales y Salud Pública de Canarias, Centro de Investigaciones Biomédicas de Canarias, Universidad de La Laguna; Santa Cruz de Tenerife, Spain; \textsuperscript{b}Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu, 9, E-28040 Madrid, Spain; \textsuperscript{c}Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC); Universidad de Cantabria–CSIC-IDICAN, Santander, Cantabria, Spain

*: Equal contribution
§: Corresponding authors

e-mail addresses: florenzo@ull.edu.es (F. Lorenzo-Díaz); cflopez@cib.csic.es (C. Fernández-López); maria.garcillan@unican.es (M.P. Garcillán-Barcia); mespinosa@cib.csic.es (M. Espinosa)

\textbf{Key words:} Conjugative transfer / Relaxases / Rolling-circle replicating plasmids / origins of transfer / Firmicutes

\textbf{Running title:} pMV158 family of plasmids
Abstract

Rolling circle-replicating plasmids constitute a vast family that is particularly abundant in, but not exclusive of, Gram-positive bacteria. These plasmids are constructed as cassettes that harbor genes involved in replication and its control, mobilization, resistance determinants and one or two origins of lagging strand synthesis. Any given plasmid may contain all, some, or just only the replication cassette. We discuss here the family of the promiscuous streptococcal plasmid pMV158, with emphasis on its mobilization functions: the product of the \textit{mobM} gene, prototype of the MOB\textsubscript{V} relaxase family, and its cognate origin of transfer, \textit{oriT}. Amongst the subfamily of MOB\textsubscript{V1} plasmids, three groups of \textit{oriT} sequences, represented by plasmids pMV158, pT181, and p1414 were identified. In the same subfamily, we found four types of single-strand origins, namely \textit{ssoA}, \textit{ssoU}, \textit{ssoW}, and \textit{ssoT}. We found that plasmids of the rolling-circle Rep\textsubscript{2} family (to which pMV158 belongs) are more frequently found in Lactobacillales than in any other bacterial order, whereas Rep\textsubscript{1} initiators seemed to prefer hosts included in the Bacillales order. In parallel, MOB\textsubscript{V1} relaxases associated with Rep\textsubscript{2} initiators tended to cluster separately from those linked to Rep\textsubscript{1} plasmids. The updated inventory of MOB\textsubscript{V1} plasmids still contains exclusively mobilizable elements, since no genes associated with conjugative transfer (other than the relaxase) were detected. These plasmids proved to have a great plasticity at using a wide variety of conjugative apparatuses. The promiscuous recognition of non-cognate \textit{oriT} sequences and the role of replication origins for lagging-strand origin in the host range of these plasmids are also discussed.
Contents

1. Introduction
2. Replication, conjugation, and mobilization: common themes
3. Modular construction of RCR-plasmids
4. The family: pMV158 and relatives
5. The MOB_V family
   5.1. The origins of transfer (oriT)
   5.2. The MOB_V1 relaxases
5. The single strand origins (sso)
6. Conclusions and perspectives

Abbreviations: Antibiotic resistance(s), AbR; Gram-negative, G-; Gram-positive, G+;
Integrative Conjugative Elements, ICEs; Leading-strand initiation and control of replication, LIC; primer RNA, pRNA; rolling circle, RC; rolling circle-replicating plasmids, RCR-plasmids; RNA polymerase, RNAP; single-stranded, ss; double-stranded, ds; single strand origin, sso; transferred DNA strand, T-DNA; Type IV secretion systems, T4SSs; T4SS-coupling protein, T4CP; Vertical/Horizontal Gene Transfer, VGT/HGT

§ Corresponding authors at: Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu, 9, E-28040 Madrid, Spain; Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC); Universidad de Cantabria–CSIC-IDICAN, Albert Einstein 22, PCTCAN, 39011 Santander, Cantabria, Spain
Highlights

- The replication and transfer of rolling circle-replicating plasmids is reviewed.
- Comparisons of replication and transfer cassettes are presented.
- The current understanding of the pMV158 DNA transfer mechanism is reviewed.
1. Introduction

The concept developed by Marshall McLuhan in the early 1960’s on considering planet Earth as becoming a Global Village (http://projects.chass.utoronto.ca/mcluhan-studies/v1_iss2/1_2art2.htm) has proved to be true fifty years afterwards. Humankind has become global, indeed, not only virtually but physically as well: in addition to the World Wide Web, economical, commercial, and touristic activities have led to unquestionable benefits for the exchange of cultures, but it has also imposed a heavy burden on the rest of the biosphere. A huge part of it is formed by the microbial world, so that going global has increased health risks in the way of outbreaks of epidemics, and the appearance of new transmissible diseases: the swine and avian flu, and the Severe Acute Respiratory Syndrome, to name the most known (see, for instance, the reports by the US-Center for Disease Prevention and Control: http://www.bt.cdc.gov/publications, and its European counterpart: http://www.ecdc.europa.eu/en/Pages/home.aspx). Within this global scenario, not only viruses but also bacterial plasmids have increased their relevance as vehicles to disseminate genetic information among different species: the human activities and human contacts have resulted to be brutal selection processes that have accelerated the horizontal transfer of genetic information among microorganisms (Baquero, 2004; Wellington et al., 2013). Selection of novel bacterial traits derives not only from the need of better starters for fermentation/food production but also from the abuse and misuse of broad-spectrum antibiotics. These last activities have led to the selection of bacteria harboring genetic elements with multiple antibiotic-resistances (AbR), which are
rapidly spread (even as epidemic outbursts in hospitals) by horizontal gene transfer, HGT (Anderson and Seifert, 2011; Baquero, 2009). Genes responsible for AbR frequently cluster in the bacterial mobile elements (the mobilome) within the Integrative and Conjugative Elements (ICEs) and transmissible plasmids that, in turn, are platforms to recruit various smaller mobile elements, such as insertion sequences, transposons and integrons that can also encode AbR genes (Frost et al., 2005). Thus, HGT mediated by self-replicating plasmids or by transfer of the islands, plays an essential role in the bacterial, and consequently in the global biodiversity.

In the Plasmid Biology field, the processes pertaining to the dissemination of genetic information stored in plasmids are of special relevance. Those processes involve the two main ways of inheritance, namely vertical (VGT, from mother to daughter cells) and HGT (cell-to-cell) gene transfer (Thomas, 2000). Whereas the former includes DNA replication and partition, this latter being coupled to the cell division, HGT is usually achieved by conjugation between cells of the same or different species (del Solar et al., 1998; Grohmann et al., 2003; Lanka and Wilkins, 1995). Replication and conjugation are, in addition to genetic recombination, the most important sources of genetic variability among plasmids and their hosts. The entire genetic content of the bacterial mobilome amounts up to 25% of the total DNA circulating among bacteria, thus being a shared strong task force for the evolution of the bacterial populations (Ochman et al., 2000; Thomas and Nielsen, 2005). A substantial part of the mobilome is constituted by plasmids, the so-called plasmidome (Walker, 2012).
Plasmids are much more than cloning vectors or tools to over-produce proteins. In addition to their role in the spread of genetic information, bacterial plasmids are excellent models to study a number of biological processes, such as transactions involving macromolecular interactions: protein-DNA, protein-protein, and DNA-RNA (Espinosa, 2013). They also constitute a wealth of information on control of gene expression (del Solar and Espinosa, 2000; del Solar et al., 2002), intracellular distribution of DNA molecules (Reyes-Lamothe et al., 2014), and can be considered as useful models for system biology (Paulson and Ehrenberg, 1998). Excellent books dealing with various aspects of the biology of plasmids have been published (Funnell and Phillips, 2004; Thomas, 2000) and a new one is in press (Alonso and Tomalsky, 2014). In the present review we will concentrate on plasmids replicating by the rolling circle mechanism (RCR-plasmids) because they may encode up to two proteins with endonuclease/topoisomerase-like activity that cleave supercoiled DNA at two different regions on the same molecule, making their study thought-provoking.

2. Replication, conjugation, and mobilization: common themes

Replication and transfer generally require that plasmid-encoded protein(s) interact with their cognate DNA sites to initiate the process. The key initiator players in these two processes are: i) the generically called Rep proteins involved in vegetative replication and that interact with their cognate origin of replication, oriV, and ii) the relaxases, usually termed Tra or Mob proteins, involved in transfer, and that recognize their cognate origin of transfer, the oriT. Thus, interplays between Rep-oriV and Tra/Mob-oriT will define the
VGT and the HGT processes, respectively. The early discovery of conjugation (Lederberg and Tatum, 1946; Lederberg and Tatum, 1953) and the existence of plasmid-encoded proteins that relax the donor DNA as the first stage in the transfer process, led to a historically broad interest in plasmid transfer among Gram-negative (G-) bacteria. Conjugation in Gram-positive (G+) bacteria was studied relatively later and it is still a matter of active research (Goessweiner-Mohr et al., 2013; Grohmann et al., 2003). Replication involves a number of different strategies devoted to the melting of the DNA strands at the origin of replication (namely theta, strand-displacement and rolling-circle mechanisms), and is usually mediated by the Rep initiator alone or with the help of plasmid- or host-encoded proteins (del Solar et al., 1998). Although DNA melting to initiate the transfer is also required, conjugation, however, involves a single strategy devoted to the unidirectional transfer of mobile elements (plasmids or ICEs) from a donor bacterial cell to a recipient one through physical cell-to-cell contact (de la Cruz et al., 2010; Zechner et al., 2000). Based on their transfer machinery we can distinguish between: i) self-transmissible (conjugative) plasmids and integrative conjugative elements (ICEs), which codify all the functions required for their HGT, and ii) mobilizable plasmids and integrative and mobilizable elements (IMEs), which ‘travel light’ because they encode only the relaxase and its cognate oriT. These latter plasmids and IMEs make use of functions provided by either the host chromosome or by other auxiliary (also termed ‘helper’) plasmid for their transference.

The assembly of plasmid- and host-encoded proteins in a specific DNA region initiates replication and transfer, thus initiation of both processes requires the generation of multiprotein-DNA-complexes, the replisome or the
relaxosome, respectively (del Solar et al., 1998; Lanka and Wilkins, 1995; Pansegrau and Lanka, 1996a). In the case of RCR-plasmids (del Solar et al., 1987; Khan et al., 1981; Novick, 1998; Puyet et al., 1988; te Riele et al., 1986a; te Riele et al., 1986b), initiation of replication and mobilization are mechanistically similar processes. Both, Rep and Tra/Mob proteins have endonuclease/topoisomerase-like activities on their DNA targets, the double-strand origin (dso) or the oriT. These proteins cleave their cognate DNA at the phosphodiester bond of a specific di-nucleotide (the nic site) generating a stable amino acyl-DNA adduct (Chandler et al., 2013; de la Campa et al., 1990; Guasch et al., 2003; Khan, 2003; Koepsel et al., 1986; Moscoso et al., 1997; Pansegrau and Lanka, 1996a). The nick introduced by Rep or Tra/Mob proteins generates a free 3’-OH end, which acts as a primer for leading-strand synthesis in both cases, VGT and HGT. In the RCR-plasmids, proteins from the host replicative machinery, at least DNA-polymerases I and III, single-strand (ss) DNA-binding protein and PcrA helicase, participate in the elongation from the 3’-OH end generated by the plasmid-encoded Rep initiator (Anand et al., 2005; Anand and Khan, 2004; Díaz et al., 1994; Khan, 2003; Khan, 2005; Machón et al., 2010; Ruiz-Masó et al., 2006; Soultanas et al., 1999; Thomas et al., 2013). To process their DNA substrates, initiators of replication and transfer require the nic site being exposed in a single-stranded configuration, which can be achieved by DNA melting and generation of hairpin structures. Hairpin formation would be mediated either by the binding of auxiliary associated proteins, and/or by binding of the initiator or the relaxase. In both cases, the di-nucleotide to be cleaved will be exposed in an unpaired form (Jin and Novick, 2001; Lorenzo-Díaz et al., 2011; Lucas et al., 2010; Noirot et al., 1990; Ruiz-Masó et al., 2007).
Besides the initiator (relaxase) and its binding site (oriT), conjugation requires an additional machinery that RCR does not, namely the protein complex that completes the conjugation apparatus. This is a highly specialized protein machinery encoded by the donor plasmid DNA (or the host chromosome in some cases) that includes the coupling protein (T4CP) and the Type IV Secretion System (T4SS) which translocate the relaxase-DNA complex to the recipient cell (Draper et al., 2005; Garcillán-Barcia et al., 2007). T4SSs recruit their substrate by mechanisms still not fully understood for conjugation, and they also participate in other processes involving DNA trafficking between prokaryotic donors and recipient (prokaryotic or eukaryotic) cells, like in the case of Ti and related plasmids (Baron et al., 2002; Bhatt et al., 2013; Cascales et al., 2013; Chen et al., 2005; de Paz et al., 2005; Goessweiner-Mohr et al., 2013; Gomis-Ruth et al., 2004; Hamilton et al., 2000; Llosa et al., 2009; Zhang et al., 2012). The T4CP and the T4SS proteins participate in pumping the transferred DNA (T-DNA)-relaxase complex into the recipient cell (Gomis-Rüth and Coll, 2006; Guasch et al., 2003; Llosa et al., 2002; Matilla et al., 2010). In RCR as well as in transfer, the parental DNA strand will be displaced until either the dso or the oriT are reconstituted. This last stage involves DNA strand transfer reactions that will terminate either leading strand replication (in the donor cell) or plasmid transfer (in the recipient cell). Intermediates of both processes will be ssDNA molecules that correspond to the parental plus strand (del Solar et al., 1998; Grohmann et al., 2003; Pansegrau and Lanka, 1996a; Pansegrau and Lanka, 1996b; Wilkins and Lanka, 1993). Synthesis of the lagging strand initiates from the so-called single-strand origins (sso), as described below.
3. Modular construction of RCR-plasmids

A relevant part of the mobilizable plasmids replicate by the RC mechanism (del Solar et al., 1987; Khan et al., 1981; Novick, 1998; Puyet et al., 1988; te Riele et al., 1986a; te Riele et al., 1986b), although not all of them are mobilizable. In the case of small plasmids, isolated primarily from G+ bacteria, two pioneer discoveries were made: i) the identification of specific single-stranded DNA molecules (ssDNA) as intermediates of plasmid replication (Puyet et al., 1988; te Riele et al., 1986a; te Riele et al., 1986b), and ii) the finding that their Rep initiator proteins exhibited a sequence-specific relaxing activity on supercoiled DNA (Koepsel et al., 1985). These results led to the discovery of RCR-plasmids, which constituted a new class of plasmids that replicate by the asymmetric rolling circle mechanism that share similarities with the replication of ssDNA coliphages (Novick, 1998). Furthermore, the presence of RCR-plasmids in G- bacteria (Yasukawa et al., 1991) helped to cast off the idea of a genetic barrier between the two types of bacteria (del Solar et al., 1993). Thus, the conclusion that plasmid RCR is mechanistically similar to conjugative transfer was soon achieved (Waters and Guiney, 1993). It was interesting to learn that some of the RCR-plasmids encoded, in addition to the Rep topoisomerase-like initiator, another protein with the ability to relax DNA (Caryl et al., 2004; Grohmann et al., 2003; Guzmán and Espinosa, 1997; Smith and Thomas, 2004). These proteins were thought to be involved in interplasmidic recombination (plasmid recombination enzymes, Pre; (Projan and Novick, 1988), although it was later shown that Pre proteins were required for
plasmid mobilization (Priebe and Lacks, 1989). Further, the relaxase activity of the pMV158_Pre protein (renamed MobM) on supercoiled DNA, and its nic site were demonstrated in vitro and in vitro (Grohmann et al., 1997; Guzmán and Espinosa, 1997). However, even up today, these proteins are grouped into a family termed Mob-Pre at the database of protein families Pfam (PF01076), and no further investigations on their participation in plasmid recombination have been performed.

In general, RCR-plasmids appear to be constructed as gene cassettes (del Solar et al., 1993; Khan, 1997; Khan, 2005) that may have up to four independent modules involved in: i) Leading-strand initiation and control of replication (LIC); ii) AbR determinant (DET); iii) Mobilization (MOB), and iv) One or two origins for lagging strand replication, sso (Fig. 1). This latter kind of origins varies among the different RCR described, but they have been categorized in four types: ssoA, ssoU, ssoW, and ssoT depending on their DNA sequence (Khan, 2000; Khan, 2005; Kramer et al., 1998a; Kramer et al., 1997; Meijer et al., 1995b; van der Lelie et al., 1989), or in their role in plasmid promiscuity (Kramer et al., 1995; Lorenzo-Díaz and Espinosa, 2009). Their structure and roles will be described below (see Table 1). RCR-plasmids may harbor all the cassettes, like the streptococcal plasmids pMV158 and pRW35 (Priebe and Lacks, 1989; van der Lelie et al., 1989; Woodbury et al., 2008), or just the LIC cassette, as in the mycoplasma plasmids pADB201 (Bergemann et al., 1989) and pKMK1 (King and Dybvig, 1992), which are considered to be the smallest RCR-plasmids, unless we consider the hybrid phage-plasmid phasyl (Seufert et al., 1988) as one of them.
Many of these plasmids have left their ‘print’ on the chromosome of the hosts they might have colonized and, in fact, a bioinformatics survey detected RCR plasmids–related MOB\textsubscript{V}-relaxase genes in 63 out of 1207 chromosomes analyzed (Guglielmini et al., 2011). The AbR trace could be considered as a thought-provoking approach to follow the RCR-plasmid ‘fate’ along evolutionary history. Although we have not performed any further search, a homolog of the pMV158-*tetL* determinant was found in the chromosome of *Bacillus subtilis* (Lacks et al., 1986), whereas a *cat* gene, homologous to the one harbored by plasmids pC194 and pC221 has been described to be present in the chromosome of G+ bacteria, like *Bacillus pumilus* (Harwood et al., 1983) or *Streptococcus pneumoniae* (Pepper et al., 1988), and even in the chromosome of *Clostridium perfringens* (Bannam and Rood, 1991). These findings argue in favor of the role of these RCR-plasmids in the integration and dispersion of resistance genes in the chromosome of hosts that they have colonized.

4. The family: pMV158 and relatives

In general, plasmids have been grouped according to their replicon, since this is the hallmark of the plasmid (Chang et al., 2000; del Solar et al., 1998; Khan, 1996; Nordström et al., 1984; Novick, 1989). As mentioned above, RCR-plasmids were grouped in several families, although the most studied are the Rep\_1 family (PF01446), whose prototypes are the staphylococcal plasmids pC194, and pUB110, the Rep\_2 family (PF01719) of which the streptococcal plasmid pMV158 and the staphylococcal plasmid pE194 are the representatives, and the Rep\_trans family (PF02486) represented by the
staphylococcal plasmids pT181 and pC221 (del Solar et al., 1998; del Solar et al., 1993; Gruss and Ehrlich, 1989; Khan, 1997; Khan, 2000; Khan, 2005; Novick, 1989). These three families harbor all the cassettes with the exceptions of pC194 that does not contain a MOB module and the lagging-strand origin $ssou$, which is reported to be present only in pUB110 and pMV158 (Figure 1).

Curiously, these two latter plasmids also share identical oriT sequences. *In vitro*, the MobM-protein from pMV158 was able to relax supercoiled DNA from both plasmids (Fernández-López et al., 2013a). Out of the representatives of the different families, pC194 has been reported to be one of the most promiscuous, since it was shown to replicate not only in staphylococci, streptococci and bacilli (Ballester et al., 1990; Horinouchi and Weisblum, 1982), but also in *Escherichia coli* and in the yeast *Saccharomyces* as well (Goursot et al., 1982). In the case of pMV158, it has been shown to exhibit an extraordinary host range. The plasmid was primarily isolated from *Streptococcus agalactiae* (Burdett, 1980) and, along the years, it has been transferred in the laboratory (either by mobilization or by transformation) to more than 20 bacterial species, from Firmicutes to α-Proteobacteria and in all of them it replicates stably (del Solar et al., 1998; Espinosa, 2013; Lacks et al., 1986).

After a BlastP search using prototypes of the three RCR-plasmid families, we found the following distribution: i) in Rep_1 (prototype RepA from pC194), 334 non-redundant homologs in Bacillales, and 157 in Lactobacillales; ii) in Rep_2 (prototype taken, RepE initiator from pE194), 44 non-redundant homologs in Bacillales and 242 in Lactobacillales; iii) in Rep_trans (RepC from pT181 as prototype), 176 non-redundant homologs in Bacillales and 241 in Lactobacillales. These figures suggest that whereas the Rep_1 family of RCR-
plasmids would prefer to colonize Bacillales and the Rep_2 family of pMV158 would be better fitted to Lactobacillales, the Rep_trans family of RCR-plasmids would be distributed more evenly. In general, we found a good correlation between the G+C content of RCR-plasmids and their respective hosts (Espinosa et al., 1995), with the exception of the staphylococcal plasmid pUB110 which has a G+C content close to 45%, making it a plasmid which has been considered more like a bacilli than a staphylococci replicon (Alonso et al., 1988; McKenzie et al., 1986).

Traditionally, RCR-plasmids have been classified according their LIC module (del Solar et al., 1998; del Solar et al., 1993). According to this criterion, we performed a PSI-BLAST search for plasmids of the pMV158 family using the initiator RepB_{pMV158} as the query. The results are compiled in Supplementary Table S1. The 78 plasmids retrieved belong to the Rep_2 family. Out of them, 55 did not encode any relaxase whereas the remaining 23 contained relaxase genes belonging to the MOB family (see below).

Many of the retrieved plasmids lack any distinguishable marker, although resistance to a variety of antibiotics, and even one instance of resistance to arsenate was found. A phylogenetic tree of the 78 plasmids retrieved was constructed (Supplementary Figure S1), using the protein GpA initiator from phage ΦX174 as the out-group representative. There will be a detailed in-depth analysis on the LIC module of plasmids of the pMV158 family to be published elsewhere (G. del Solar and J.A. Ruiz-Masó, personal communication).

Furthermore, most of the retrieved plasmids were primarily isolated from Lactobacillales, but as shown in Supplementary Figure S1 they are also widely distributed in other taxonomic orders of Firmicutes, Tenericutes and
Proteobacteria, with notable examples of plasmids from *Mycoplasma* (eight plasmids). Although not retrieved in this search, there is an RCR-plasmid from *Mycoplasma yeatsii*, pMyBK1, which deserves a mention here because of two features (Kent et al., 2012; Breton et al., 2012). Firstly, it is unique at encoding an initiator of RCR different from the ones described here. Secondly, pMyBK1 is the only example of mobilizable plasmid in genus *Mycoplasma*, and it is precisely a MOB$_V$ plasmid. Inspecting the sequence of the *mob* gene of pMyBK1, we found that it harbors 39 UGA codons (out of a total of 520 codons). Since in *Mycoplasma* UGA specifies “tryptophan” instead of “stop” (Halbedel and Stülke, 2007; Inamine et al., 1990; Yamao et al., 1985), we can conclude that the transfer of pMyBK1 would be limited to *Mycoplasma* species.

5. The MOB$_V$ family

Classification of plasmids according to the relaxases and origins of transfer they carry has supposed a novel definition of plasmid families (Francia et al., 2004; Garcillán-Barcia et al., 2009). This classification provided a more global view of HGT by conjugation, whereas classification by replicons would point more to the vertical transfer. The HGT-based classification has showed that there are many plasmids that do not encode transfer functions (Garcillán-Barcia et al., 2011; Smillie et al., 2010). RCR-plasmids of Firmicutes fall mostly in the so-called MOB$_V$ family, being the MobM relaxase from the streptococcal plasmid pMV158 the representative of the family (Garcillán-Barcia et al., 2009). It was believed that plasmids are DNA molecules that result from shuffling of various gene cassettes, evolving independently one of each other (Osborn et
al., 2000). As we will discuss below, analysis of evolutionary roots of these cassettes motivates to think that there could be crosstalk between them, as some module combinations prevail among others.

To update the inventory of elements composing the MOB$_{V1}$ subfamily we have performed a PSI-BLAST search using MobM$_{pMV158}$ as the query. The retrieved MOB$_{V1}$ plasmids ranged from 2.7 to 30 kb (median = 5.1 kb; Figure 2). Roughly half of them coded for antibiotic resistance genes (mainly to a wide variety of protein synthesis inhibitors, such as macrolides, lincosamides, aminoglycosides, streptogramins, amphenicols and tetracyclines) or other virulence traits like resistance to heavy metals and production of bacteriocins (Table 1). The organization of the mobilization region was found to be similar in all members: the oriT located upstream, close to the mob gene. Furthermore, the nic site was placed on the same strand as the mob gene, as expected.

5.1. The origin of transfer (oriT)

Plasmid conjugation initiates through the assembly of the relaxosome on the oriT, a region that contains inverted (IR) and/or direct (DR) repeats, A+T-rich tracts and, most importantly, the nic site. The relaxase requires that its target is presented as ssDNA to cleave it and generate the relaxase-DNA adduct, which will be pumped through a T4SS. Once in the recipient, the oriT is reconstituted by a transesterification reaction mediated by the relaxase to close the incoming molecule (reviewed in (Chandler et al., 2013)).

The oriT of pMV158 spans 41 bp upstream of the mobM gene, exhibits a high A+T content (75.6%) and has three IRs (IR1 to IR3; Lorenzo-Díaz et al.,}
Furthermore, a 6-bp sequence (5’-ACTTTA-3’) is repeated in the IR1/IR3 left-arm and the IR2 right-arm (Figure 3). The nic site was firstly mapped in vitro between coordinates 3595-3596 (dinucleotide 5’-GpT-3’) in the pMV158 sequence (GenBank Acc. No. NC_010096; (Guzmán and Espinosa, 1997)). Then, the nic sites for the pMV158 and pE194 oriTs were determined in their respective hosts in vivo (Grohmann et al., 1997), mapping exactly in the same position as already suggested by the previous in vitro results. In silico analysis revealed that the IRs could generate three alternative stem-loop structures in which the position of the nic site would be placed in different positions: i) located 8-bp upstream to IR1, ii) in the IR2 inter-arm region, or iii) at the 3’-end of the IR3 (Lorenzo-Díaz et al., 2011). IR3 includes the IR1 sequence and, since IR1/3 and IR2 partially overlap (see Figure 3), the generation of cruciform secondary structures by one of them would hinder the formation of the other, indicating that the target DNA accessibility by the relaxase could depend on the plasmid DNA superhelicity (Fernández-López et al., 2014). Our in vitro analysis demonstrated that MobM binds specifically to ssDNA encompassing IR1/3 with high affinity (Lorenzo-Díaz et al., 2011), which allowed the MobM protein to repress its own synthesis (Lorenzo-Díaz et al., 2012). Functional relevance of the IRs in the different steps of the conjugative process is currently under exploration. We hypothesize that IR1/3 may be involved in the Mob-recognition of the oriT at the initiation of the relaxosome formation (in the donor cell) and IR2 at the termination reaction to close the T-strand (in the recipient cell).

Based on the sequence and structure of oriT_{pMV158}, we inspected the oriT regions of the 93 MOBv-plasmids listed in Table 1. A total of 97 sequences were manually identified upstream to their respective relaxase encoded genes (five of
them exhibiting two different MOBv-related oriTs: pSTE1, pKKS285, pSCFS1, unnamed (GenBank Acc. No GG692894.1), and pTB19. Analysis of the region showed a high degree of sequence conservation in the majority of the oriTs, being composed by three IRs and exhibiting the consensus sequence ‘GTGBG↓T’ for the nic site (B denoting a G, C or T following IUPAC code; ↓ being the nic site). Two other minor plasmid groups, represented by pT181 and p1414, grouped in different clusters given their differences respect to the oriT sequence of pMV158 (Figure 3 and Supplementary Figure S2). However, these two clusters maintain the number and distribution of the IRs. Unfortunately, the experimental evidences for mapping the specific nic site in these plasmids are scarce and only a few oriT predictions, such as for plasmids like pSMQ172 and pPB1, have been published (de las Rivas et al., 2004; Turgeon and Moineau, 2001).

Given that the oriT_{pMV158} is the only element required in cis for pMV158 to be transferred (Farías and Espinosa, 2000), it is plausible to assume that any RCR-plasmid containing an oriT-like sequence would be potentially mobilizable. Such is the case of pXY3 (Zhou et al., 2010) and pCl411 (Coffey et al., 1994) plasmids (not included in Table 1), which only seem to contain an orphan and non-canonical oriT. In fact, the prototype RCR-plasmid pC194, which lacks both a mob gene and a canonical oriT, has been mobilized by using the conjugative transposon Tn916 as a helper (Naglich and Andrews, 1988; Showsh and Andrews Jr, 1999). It is assumed that the relaxase of the helper should be able to recognize a suitable oriT in the â"\Delta mob" plasmid. It is more intriguing the finding that pC194 and ΔoriT-Δmob-derivatives of plasmids pUB110 and pTA1060 were efficiently mobilized by the mating apparatus of ICEBs1, but without the intervention of the
ICEBs1 NicK relaxase (Lee et al., 2012). Thus, in this case, the plasmids were not transferred by cross-recognition of an oriT by the relaxase of the helper. They were neither transferred by forming cointegrates with ICEBs1. Unexpectedly, the RCR initiation proteins of the mobilized plasmids were crucial for transfer (Lee et al., 2012). Thus, the strategy we have followed to search for plasmids containing MOBv1 relaxases might underestimate the real population of mobile elements.

5.2. The MOBv1 relaxases

The second element that belongs to the transfer module of the pMV158 RCR-family is the mobM gene, which codifies the MobM relaxase (Guzmán and Espinosa, 1997). Thus, in addition to using the homology of the Rep proteins to define the pMV158 RCR-plasmid family (Supplementary Figure S1 and Supplementary Table S1), we decided to find out whether any relationship between the Rep- and the Mob-proteins of the plasmid family existed. The classification system for mobilizable plasmids based on the amino acid sequence of the relaxases allowed defining MobM from pMV158 as the prototype of the MOBv superfamily (Francia et al., 2004; Garcillán-Barcia et al., 2011; Garcillán-Barcia et al., 2009). This superfamily is composed of more than 200 relaxases, of which about 140 were located in plasmids and the rest in bacterial chromosomes (Guglielmini et al., 2011).

Five subfamilies were described, and MobM from pMV158 was taken as the prototype of the MOBv1 subfamily (Garcillán-Barcia et al., 2009). Members of this subfamily showed three conserved motifs: i) Motif I (HxxR), of yet unknown function; ii) Motif II (NY(D/E)L), which is the proposed catalytic domain, and iii)
Motif III (HxDE...PHxH), which corresponds to the metal-coordination motif, also known as the 3H motif (Chandler et al., 2013). The MOB\textsubscript{V2} family, represented by the Mob protein of the theta-replicating plasmid pBBR1 (Szpirer et al., 2001), exhibited Motifs I and III but lacked Motif II. The three motifs are located in the N-terminal moiety of the MOB\textsubscript{V} relaxases. This moiety harbors the DNA binding and nicking activities, since a truncated version of MobM, which only contains the first N-terminal 199 residues, retained the relaxase activity on supercoiled DNA (Fernández-López et al., 2013a; Fernández-López et al., 2013b; Lorenzo-Díaz et al., 2011). The C-terminal moiety of MobM could be involved in, at least, two functions: i) protein-protein interactions (dimerization and interactions with the auxiliary plasmid-encoded coupling protein) and ii) association with the cell membrane, through a proposed coiled-coil region located between residues 400 and the C-terminal end of MobM. Disruption of the alpha helical-rich region by mutations (changes to Pro residues) resulted in failure of MobM-association with membranes; further, the pMV158-derivative harboring these mutations lost its ability to be transferred (de Antonio et al., 2004). There was no indication of a helicase activity in the C-terminal moiety of MobM (our unpublished observations).

We have updated the inventory of elements that cluster with pMV158 into the MOB\textsubscript{V1} subfamily. 97 non-redundant plasmid MOB\textsubscript{V1} relaxases were retrieved from a PSI-BLAST search using the N-terminal 300 amino acids of MobM from pMV158 as a query (Table 1). All of them shared the three Motifs described above (Figure 4A). The inferred phylogeny of MOB\textsubscript{V1} relaxases, rooted by a MOB\textsubscript{V2} relaxase, showed well-supported external clusters of highly related sequences and poorly-supported internal nodes (Figure 4B). It is a fact that reflects the low overall similarity of the taxa, mainly circumscribed to the three relaxase motifs.
described (Francia et al., 2004; Garcillán-Barcia et al., 2009). The plasmids coding
these relaxases are primarily distributed in several genera of Lactobacillales and
Bacillales, with a few members out of the phylum Firmicutes. Curiously, the
plasmids encoding relaxases comprised in a monophyletic clade in the
phylogenetic tree shown in Figure 4B are generally hosted in a single taxonomic
order (either Lactobacillales or Bacillales), suggesting less inter-order transfer than
expected.

No genes associated with conjugative transfer others than the relaxase
ones were encoded by these plasmids. Thus, they are classified as mobilizable,
requiring the conjugative machinery of other plasmids to be transferred (namely, a
coupling protein and a mating-pair formation apparatus composed by a T4SS and
a conjugative pilus). Eight mating-pair formation (MPF) types have been
phylogenetically described (Guglielmini et al., 2011); three of them, MPF_{FATA},
MPF_{FA} and MPF_{T} were able to mobilize MOB_{V1} plasmids. Specifically, pMV158
was mobilized between G+ bacteria by functions supplied by helper MPF_{FATA}
plasmids of the Inc18 family, like pIP501 and pAMβ1 (Grohmann et al., 1999;
Priebe and Lacks, 1989; van der Lelie et al., 1990), and even to the G− bacterium
*Escherichia coli* by MPF_{T} plasmids RP4 and R388 but not by MPF_{F} plasmid F
(Farías and Espinosa, 2000). Other elements, such as pC194 and pUB110, were
mobilized by auxiliary MPF_{FA} elements such as *Tn*916 (Lee et al., 2012; Naglich
and Andrews, 1988), and MPF_{FATA} plasmids, like pLS20, mobilized pUB110 and
pBC16 (Koehler and Thorne, 1987; Selinger et al., 1990).

Most of the MOB_{V1} plasmids replicate by the rolling circle mechanism
(Figure 5 and Table 1). MOB_{V1} relaxases are predominantly linked to RCR
initiators of the three different subgroups: Rep_1 (PF01446), Rep_2 (PF01719),
and Rep_trans (PF02486). A small fraction of MOB\textsubscript{V1} relaxases is linked to a wide variety of theta replication families, of which Rep\textunderscore 3 is the most abundant. Congruently, with the abovementioned taxonomic bias in the abundance of RCR initiators and the MOB\textsubscript{V1} relaxase distribution, most relaxases encoded in Rep\textunderscore 2 RCR plasmids grouped separately from those encoded in Rep\textunderscore 1 RCR plasmids (clade Lactobacillales vs. Bacillales in Figure 6 and Supplementary Table S2). An example of this bias is found in Enterococci (Lactobacillales), where Rep\textunderscore 1 initiators are commonly found in multireplicon plasmids, and may not be functional, as it is the case for plasmid pAM\textalpha\textsubscript{1} (Clewell et al., 2014). Nevertheless, a few functional Rep\textunderscore 1 initiators can be also found in the Lactobacillales clade, such as those grouped in the cartooned pSMA23 cluster of Figure 4B. Curiously, within the pSMA23 group some exceptions also exist: plasmid pCD034-2 encodes a Rep\textunderscore 2 instead of a Rep\textunderscore 1 initiator, an indication of recent recombination events that led to new backbone combinations. Besides, more ancient recombination events could give rise to arrangements present in plasmids of the Bacillales MOB\textsubscript{V1} clade, since the highly divergent relaxases included in that group are linked to initiators of the three RCR subgroups (Figure 6A).

Figure 6 also provides an interesting example of the switch of plasmids from and to integrative elements. ICE\textit{Sgal} is an integrative element highly similar to plasmid pUB110 (both in relaxase and initiator) but located in the chromosome of \textit{Streptococcus galolyticus} UCN34 (Lactobacillales order). It is tempting to speculate that this element integrated in the chromosome once transferred from Bacillales to the Lactobacillales background (maybe helped by a Tn916-like
element located close to ICE-SgaI), where it was not able to replicate and became an integrative and mobilizable element (IME).

A significant proportion (one third) of MOB$_V^1$ plasmids recorded in Table 1 encoded more than one replication initiation protein (Figure 5). Besides, 10 out of the 93 plasmids coded more than one relaxase gene. Both facts suggest the frequent arising of plasmid cointegrates. Precisely, it is known the ability of RCR-plasmid encoded relaxases to promote site-specific DNA recombination at oriT rendering plasmid cointegrates during conjugative mobilization, where the host rec system may function to stimulate the recombination process (Gennaro et al., 1987; Novick et al., 1984; Projan and Novick, 1988). Further, it is worth recalling that the staphylococcal plasmid pE194 was able to integrate into the *Bacillus subtilis* chromosome by a RecA-independent recombination mechanism, and using as little as 6-14 bp homologies (Dempsey and Dubnau, 1989). This, in conjunction with the finding of RCR-plasmids integrated into other bigger plasmids (Oskam et al., 1991), raises the question of whether RCR-plasmids that are integrated on IMEs would participate in their replication and mobilization; in fact, we have found a putative initiator of replication of the Rep$_1$ family within an streptococcal island which also encode a MOB$_V^1$ protein (our unpublished observations). Besides, the cross-recognition of heterologous oriTs by MOB$_V$ relaxases (Fernández-López et al., 2013a) could be favored in the multi-relaxase cointegrates, potentiating their spreading.

Recombination seems to be the most likely cause of the different topologies exhibited in the dendrograms of initiators and relaxases (compare trees in panels A and B of Figure 6). Despite the absence of a strong coevolution between initiators and relaxases, there is a clear tendency to find stable
backbones Rep_2-MOB_{V1} and Rep_1-MOB_{V1} differentially adapted to different taxonomic orders.

6. The single-strand origins (sso)

Rolling circle replication in ssDNA coliphages and in plasmids requires the existence of origins that are involved in lagging strand synthesis (Kornberg and Baker, 1992; Novick, 1998). When RCR-plasmids were discovered (te Riele et al., 1986a; te Riele et al., 1986b), it was apparent that the ssDNA intermediates should harbor signals for the conversion of the ssDNA to plasmid dsDNA. This was first reported for plasmid pT181 (Gruss et al., 1987) and soon afterwards for the pMV158-derivative plasmid pLS1 (del Solar et al., 1987). In these two plasmids, it was shown that their ssoAs were located in non-coding 200-300 base-pair-long regions that have the potential to generate one or several secondary structures on ssDNA. These signals were orientation-dependent (this means that they were functional only when placed in the displaced strand). Deletion of the region encompassing the ssoA led to accumulation of ssDNA intermediates and to plasmid instability, but the plasmids were still able to replicate (del Solar et al., 1987; Dempsey et al., 1995; Gruss et al., 1987; Kramer et al., 1995; Murray; Seegers et al., 1995). This finding led to the hypothesis that alternative, albeit less efficient, ssos could replace the genuine conversion signal (del Solar et al., 1987; Kramer et al., 1998a; Meijer et al., 1995a; Meijer et al., 1995b). Although accumulation of ssDNA intermediates and plasmid unstable inheritance by VGT were thought to be related phenomena (Meijer et al., 1995a; Meijer et al., 1995b), cloning of
signals that allowed an efficient ssDNA $\rightarrow$ dsDNA conversion, at least in plasmid pLS1, did not lead to stable plasmid inheritance (Hernández-Arriaga et al., 2000). Apart from ssoA, other three sso types have been described based on sequence similarity and structure analysis (Khan, 1996; Khan, 2000; Kramer et al., 1998a): ssoU (pUB110, pMV158), ssoT (pTA1060, pBAA1), and ssoW (pWV01). A fifth type was described in plasmid pM4 from Lactobacillus plantarum, which exhibited no significant sequence or structural similarity with any of the four classical sso (Yin et al., 2009). Plasmids pMV158 and pRW35 have the unusual feature of harboring two sso, ssoU and ssoA. Out of the two origins, the former has a more complex structure than the latter (Figure 7).

Two regions were mapped in the ssoA: the recombination site B (RS$_B$) and a 6-nucleotide consensus sequence (CS-6) (del Solar et al., 1987; Gruss et al., 1987). These two regions acted as efficient signals only in ssDNA configuration (replicative intermediates). In these molecules, the ssoA would adopt a long stem-loop structure where the RS$_B$ would be located at the stem and the CS-6 in the loop of the hairpin (Figure 7). Biochemical and genetic analyses demonstrated that RS$_B$ was recognized as the binding site of the host RNA polymerase (RNAP), thus acting as ssDNA promoter (Kramer et al., 1997). These promoters were shown to generate within paired secondary structures on ssDNA molecules and harbor sequences that are recognized by the host RNAP to initiate lagging strand synthesis (Glucksmann-Kuis et al., 1992; Masai and Arai, 1997). From the sso$_{A_{pMV158}}$ promoter, the RNAP synthesized a short 20 nt-long primer RNA (pRNA) that stopped at the CS-6 sequence, which acted as a transcription terminator (Kramer et al., 1997). The pRNA was processed by
DNA polymerase I and was proposed to be elongated by the host DNA polymerase III to finish the lagging DNA strand synthesis (Kramer et al., 1998b).

Efficient ssDNA → dsDNA conversion is also needed in the conjugative process. Within the recipient cell the transferred ssDNA should be converted into dsDNA either prior or after the circularization of the T-DNA by a strand-transfer reaction. During transfer, a fast conversion of ssDNA intermediates would be critical to finish the process, since no proteins essential for vegetative replication and control would be synthesized in the recipient cell until the first dsDNA plasmid copy is generated (Lorenzo-Díaz and Espinosa, 2009). Despite its importance, the synthesis of the lagging strand in the recipient cell remains an unresolved issue. For plasmids F and ColIb-P9 single-stranded promoters were identified in the leading strand region, which is the one that enters first into the recipient cell, and that includes genes that promote the establishment of the incoming plasmid (Bates et al., 1999; Masai and Ara, 1997; Nasim et al., 2004).

We have explored the sso diversity and distribution among MOB\textsubscript{V1} plasmids. Based on the ssos previously identified by homology or characterized by in vivo and/or in vitro approaches, we annotated 35 sso elements in 33 out of 93 plasmids in Table 1. This limited number is due to the little overall homology among ssos, even those of the same type. All the reported sso types were found: sso\textsubscript{T} (13 plasmids), sso\textsubscript{A} (11 plasmids), sso\textsubscript{U} (6 plasmids) and sso\textsubscript{W} (4 plasmids). Two plasmids (pMV158 and pRW35) contained two ssos (sso\textsubscript{A} and sso\textsubscript{U}). No new members arose in our search of the new type of sso reported for plasmid pM4 (Zhai et al., 2009). Curiously enough, we have observed that while sso\textsubscript{A} and sso\textsubscript{W} are generally found downstream and close to the mob gene, the sso\textsubscript{U} is always upstream with respect to the ori\textsubscript{T} sequence. In the case of
the \textit{ssoT} element we have found that it can be located upstream (n=5) or downstream (n=8) of the MOB module.

It has been demonstrated that the \textit{sso} type is one of the key elements in determining the host range of a plasmid: whereas \textit{ssoU} and \textit{ssoT} support a wide-host range, \textit{ssoA} and \textit{ssoW} seem to evolve in the other direction, working efficiently only in their natural hosts \textit{in vivo} (Khan, 2005). From the analysis abovementioned, it seems apparent that highly related plasmids do not always contain the same type of \textit{sso}, whereas distantly related do. These facts complicate even more the picture of the putative host range a MOBV1 plasmid can reach. Whether the location of the \textit{sso} affects the efficiency of a given plasmid for replication and/or conjugative transfer and, consequently, its host range has not been explored yet.

7. Conclusions and perspectives

RCR-plasmids represent a pool of genetic information that is shared by many bacteria. We have found them in Lactobacillales, Bacillales, Mollicutes, etc. Among the bacteria that host them, we have found several of the most relevant G+ pathogens, namely \textit{Staphylococcus aureus}, \textit{S. pneumoniae}, and \textit{Enterococcus faecalis}. These three bacterial species are important for human health because they: i) exhibit very high rates of human morbidity and mortality; ii) are the cause of many health-care associated infections; iii) have acquired elevated resistance to antibiotics; iv) play an important role as reservoir of AbR and virulence genes, and v) carry conjugative or mobilizable broad host-range plasmids, thus contributing to the spread of resistances. Infections caused by
the above G+ bacteria represent, in addition to a high economic impact, a threat to hospitals, children, elder population and immuno-compromised people (Jones, 2001). Furthermore, plasmid-encoded genes related not only to AbR but also to replication and/or transfer, are also found within the ICEs. However, and in spite of the relevance of these bacteria, very little is known on the genetics and biochemistry of the transfer functions of the RCR-plasmids studied here with the exception of the streptococcal plasmid pMV158 (reviewed in (Espinosa, 2013; Fernández-López et al., 2014) and the staphylococcal plasmid pC221 (Caryl et al., 2004; Caryl and Thomas, 2006; Smith and Thomas, 2004). And even in these two plasmids there is little, if any, information on the interactions of the relaxases encoded by pMV158 or pC221 with the machinery provided by auxiliary plasmids (Arends et al., 2013; Goessweiner-Mohr et al., 2013; Grohmann et al., 2003).

In addition to the above, it would be interesting to explore how the conjugative transfer is regulated in plasmids that contain more than one MOB\textsubscript{V} cassette as well as the level of cross-recognition between the MobM-like relaxases and their non-cognate ori\textsubscript{T}s. We have shown that MobM is able to relax supercoiled DNAs from plasmids with ori\textsubscript{T}s that share total (plasmid pUB110) or partial (plasmid pDL287) homology with the ori\textsubscript{T} of pMV158 (Fernández-López et al., 2013a), a phenomenon that could play an important role in the plasmid spreading between bacteria in natural environments. Furthermore, the fact that RCR initiators relax DNA in a way highly similar to conjugative relaxases and the recent finding of RCR initiators involvement in plasmid mobilization (Lee et al., 2012) open a new research field in the conjugative transmission of RCR plasmids.
Acknowledgements

Thanks are due to members of our labs for information and fruitful discussions. We regret that we have been unable to include all the contributions of people working in our field due to space limitations, and apologize for missing references. Research was funded by the Spanish Ministry of Economy and Competitiveness (grants CSD-2008-00013-INTERMODS to M.E., Sara Borrell CD13/00304 to F.L.-D., and BFU2011-26608 to M.P.G.-B.), and by the 7th Framework Programme (FP7-REGPOT-2012-CT2012-31637-IMBRAIN to F.L.-D. and by 282004/FP7-HEALTH-2011-2.3.1-2 to M.P.G.-B.).
8. References


de las Rivas, B., et al., 2004. Complete nucleotide sequence and structural organization of pPB1, a small *Lactobacillus plantarum* cryptic plasmid that originated by modular exchange. Plasmid. 52, 203-211.


Lucas, M., et al., 2010. Relaxase DNA binding and cleavage are two distinguishable steps in conjugative DNA processing that involve different sequence elements of the nic site. J. Biol. Chem. 285, 8918-8926.


Reyes-Lamothe, R., et al., 2014. High-copy bacterial plasmids diffuse in the nucleoid-free space, replicate stochastically and are randomly partitioned at cell division. Nucleic Acids Res. 42, 1042-1051.


Legends to the figures

Figure 1. Genetic organization of representatives of RCR-plasmids. The modular organization of pMV158 from *S. agalactiae* is depicted in the upper part. Genes and the identified promoters are indicated as arrows (arrowheads pointing to the transcription direction). Negative (-) regulatory elements within the replication and control (green) and the transfer (blue) modules are indicated. Below pMV158, other RCR-plasmids are depicted: pADB201 from *Mycoplasma mycoides*, pWV01 from *Lactococcus lactis* and pE194, pT181, pC221, pC194, and pUB110 from *S. aureus*. The corresponding antibiotic resistance (red module) is indicated with the following abbreviations: MLS$^R$, Macrolide / Lincosamide / Streptogramin B; Tc$^R$, Tetracycline; Cm$^R$, Chloramphenicol; Km$^R$, Kanamycin. Plasmids sharing similar genetic modules are presented in the same color and filling. See Supplementary Tables S1 and S2 for detailed information on the plasmids.

Figure 2. Distribution of MOB$\text{V}_1$ plasmids according to their size. The X axis was built by using the $\log_{10}$ of plasmid size values. Each bar represents the abundance of plasmids for a given size range, which is indicated at each side of the bar. Plasmids encoding genes for metal or antibiotic resistance and/or bacteriocin production are indicated in dark grey. The rest are indicated in light grey. Data was obtained from Table 1.

Figure 3. The origin of transfer (*oriT*) in MOB$\text{V}_1$ representative plasmids. Comparison of the *oriT* sequences located in the pMV158, p1414 and pT181
plasmids, lined up by the position of the nic site (boldface letters). The three overlapping inverted repeats (IR1, IR2 and IR3) are depicted by arrows, and dashed lines indicate the position of those unpaired bases in the predicted secondary structures they could form. Grey background in the pMV158-oriT sequence denotes the position of a conserved repeated region. Consensus of the oriT sequences aligned in Supplementary Figure S2 were prepared using WebLogo (version 2.8.2; (Crooks et al., 2004)).

**Figure 4. Phylogeny of MobM_{pMV158} homologs.** A: Logos of the MOB_{V1} relaxase motifs. MOB_{V1} relaxases were analyzed using WebLogo (version 2.8.2) (Crooks et al., 2004). B: The 300 N-terminal residues of the MobM relaxase of plasmid pMV158 were used as query in a PSI-BLAST search (Altschul et al., 1997) (e-value: 1xE-6 and limited to 100 non-redundant plasmid hits). The search converged in the sixth iteration. The 300 N-terminal residues of the homologs were aligned using MUSCLE (Edgar, 2004). The phylogenetic reconstruction was carried out by maximum likelihood (ML), using RAxML version 7.2.7 (Stamatakis, 2006). 100 ML trees were executed using the JTTGAMMA model. 1000 bootstrap trees were then inferred to obtain the confidence values for each node of the best ML tree. Only bootstrap values > 50% are indicated. The MOB_{V2} relaxase of plasmid pBMYdx (GenBank Acc. No. NP_981974.1) was used as outgroup. Highly related clusters are compressed and a prototype member is indicated. The names and features of all members included in the tree are recorded in Table 1. Clusters grouping plasmids that encode antibiotic resistance traits are underlined. Vertical bars
delimit clades for which most of their members are hosted either in Lactobacillales or in Bacillales.

**Figure 5. Distribution of replication initiation protein families in MOB\textsubscript{V1} plasmids.** The percentage of replication initiator families included in Table 1 is presented. Plasmids with more than one initiator are included in “> 1 replication initiator”. Plasmids with a single initiator are grouped in “Rep\textsubscript{1}”, “Rep\textsubscript{2}” and “Rep\textsubscript{trans}” families (when RCR), or in “Rep\textsubscript{3}” and “Others” (non-RCR).

**Figure 6. Phylogeny of the Rep and Mob (MOB\textsubscript{V1}) proteins of relevant RCR-plasmids.** Representative RCR plasmids and MOB\textsubscript{V1} elements included in Supplementary Table S2 were used to trace the evolutionary relationships of their relaxase and replication initiation proteins. The homologs were aligned using MUSCLE (Edgar, 2004). The phylogenetic reconstruction was carried out by maximum likelihood (ML), using RAxML version 7.2.7 (Stamatakis, 2006). 100 ML trees were executed using the JTTGAMMA model. 1000 bootstrap trees were then inferred to obtain the confidence values for each node of the best ML tree. Only bootstrap values > 50% are indicated. A: Phylogenetic tree of the N-terminal 300 residues of MOB\textsubscript{V1} relaxases. Each plasmid is shadowed in grey according to the RCR initiator subgroup as indicated in the legend of panel B. ± indicates an exceptional MOB\textsubscript{V1} plasmid, pE33L5, which does not encode an RCR initiator but a HTH\_36 (PF13730) replication initiation protein. Vertical bars delimit clades for which most of their members are hosted either in Lactobacillales or in Bacillales. ¥ indicates an element not hosted in the taxonomic order indicated by the bars. ¶ indicates that plasmid pMRI\_5.2 also encodes a Rep\textsubscript{1} RCR initiator. B:
Phylogenetic tree of the RCR initiators. Plasmids pT181 and pC221 were used as outgroups. A grey color palette was used to indicate clades containing different RCR initiators families: Rep_1 (PF01446) and Rep_2 (PF01719), as well as the Rep_trans (PF02486) used to root the tree. *: According to their GenBank annotated sequences, plasmids pSBO2 and pLFE1 encode truncated MOB\textsubscript{V} relaxases and thus were not included in the MOB phylogeny, neither were the underlined plasmids (pWV01, pSsal-M18 and pC194) since they do not encode relaxases. §: pC221 is a mobilizable RCR plasmid, but it encodes a MOB\textsubscript{P7} instead of a MOB\textsubscript{V} relaxase.

Figure 7. Predicted secondary structures of the lagging-strand origins of replication ssoA and ssoU. The ori\textsubscript{T} and the mob\textsubscript{M} gene of plasmid pMV158 are flanked by two lagging-strand origins of replication (ssoA and ssoU). ori\textsubscript{T}\textsubscript{pMV158} sequence (coordinates 3564-3605 from pMV158; GenBank Acc. No. NC_010096) is shown at the center of the image. Its three inverted repeats are represented by arrows and the nic site by a vertical arrowhead. Both sso\textsubscript{s} can generate long hairpin-loop structures that function as ‘ssDNA promoters’ (Kramer et al., 1999; Kramer et al., 1997; Masai and Arai, 1997). The RNAP-binding site (RS\textsubscript{B}), located in the base of the hairpin is recognized by the RNAP to synthesize a short pRNA. A consensus sequence (CS-6), located in the loop of the hairpin, acts as the termination point for the pRNA synthesis. The pRNA is then used by DNA polymerase I for limited extension synthesis, followed by replication of the lagging strand by DNA Pol III. The figure was modified with permission from the American Society for Microbiology from (Fernández-López et al., 2014). No further reproduction or distribution is permitted without the prior written permission of American Society for Microbiology.
<table>
<thead>
<tr>
<th>Plasmid name</th>
<th>Nucleotide GenBank Acc.</th>
<th>Plasmid size (kb)</th>
<th>Relaxase GenBank Acc. No.</th>
<th>Relaxase tree cluster</th>
<th>Host</th>
<th>Replication initiator family</th>
<th>sso type (position related to mob gene)</th>
<th>Virulence or antibiotic / metal resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMV158</td>
<td>NC_010096.1</td>
<td>5.541</td>
<td>YP_001568274</td>
<td></td>
<td>pMV158</td>
<td>Streptococcus agalactiae</td>
<td>Rep_2</td>
<td>ssoA (3') + ssoU (5')</td>
</tr>
<tr>
<td>pEr13</td>
<td>NC_002776.1</td>
<td>4.139</td>
<td>NP_115336.1</td>
<td></td>
<td>pMV158</td>
<td>Streptococcus thermophilus</td>
<td>Rep_2</td>
<td>ssoA (3')</td>
</tr>
<tr>
<td>pSM22.72</td>
<td>NC_004958.1</td>
<td>4.230</td>
<td>NP_852457.1</td>
<td></td>
<td>pMV158</td>
<td>Streptococcus thermophilus</td>
<td>Rep_2</td>
<td>ssoA (3')</td>
</tr>
<tr>
<td>pGB2000</td>
<td>NC_01952.1</td>
<td>4.967</td>
<td>YP_006931085.1</td>
<td></td>
<td>pMV158</td>
<td>Streptococcus pyogenes</td>
<td>Rep_2</td>
<td>ssoA (3')</td>
</tr>
<tr>
<td>pGB2002</td>
<td>NC_015971.1</td>
<td>6.825</td>
<td>YP_004831084.1</td>
<td></td>
<td>pMV158</td>
<td>Streptococcus pyogenes</td>
<td>Rep_2</td>
<td>ssoA (3') + ssoU (5')</td>
</tr>
<tr>
<td>pRW35</td>
<td>NC_010422.2</td>
<td>4.968</td>
<td>YP_001716200.1</td>
<td></td>
<td>pMV158</td>
<td>Streptococcus pyogenes</td>
<td>Rep_2</td>
<td>ssoA (3') + ssoU (5')</td>
</tr>
<tr>
<td>pDRPPS7493</td>
<td>NC_015876.1</td>
<td>4.727</td>
<td>YP_004765941.1</td>
<td></td>
<td>pMV158</td>
<td>Streptococcus pseudoneumoniae</td>
<td>Rep_2</td>
<td>ND</td>
</tr>
<tr>
<td>pS5U1</td>
<td>NC_002140.1</td>
<td>4.975</td>
<td>NP_053061.1</td>
<td></td>
<td>pMV158</td>
<td>Streptococcus suis</td>
<td>Rep_2</td>
<td>ssoA (3')</td>
</tr>
<tr>
<td>pYS8</td>
<td>NC_010936.1</td>
<td>4.973</td>
<td>YP_001967741.1</td>
<td></td>
<td>pYS8</td>
<td>Lactobacillus sakei</td>
<td>Rep_2</td>
<td>ssoT (5')</td>
</tr>
<tr>
<td>pK214</td>
<td>NC_009751.1</td>
<td>29.871</td>
<td>YP_001429536.1</td>
<td></td>
<td>pYS8</td>
<td>Lactococcus lactis</td>
<td>Rep_trans and Rep_3L_lactis_ReplB_C</td>
<td>ND</td>
</tr>
<tr>
<td>pUR2941</td>
<td>HF583290.1</td>
<td>20.876</td>
<td>CCQ43999.1</td>
<td></td>
<td>pYS8</td>
<td>Staphylococcus aureus</td>
<td>RepA_Plasmab_2 and unknown and truncated</td>
<td>ssoA (5')</td>
</tr>
<tr>
<td>pCP549</td>
<td>NC_019142.1</td>
<td>5.292</td>
<td>YP_006958108.1</td>
<td></td>
<td>pYS8</td>
<td>Staphylococcus aureus</td>
<td>Rep_2</td>
<td>ssoA (3')</td>
</tr>
<tr>
<td>pST12</td>
<td>NC_020237.1</td>
<td>11.951</td>
<td>YP_007419104.1</td>
<td></td>
<td>pSTI2</td>
<td>Staphylococcus hyicus</td>
<td>Rep_trans and HTH_Hin_like and truncated</td>
<td>ND</td>
</tr>
<tr>
<td>pKKS825</td>
<td>NC_013034.2</td>
<td>14.363</td>
<td>YP_003084337.1</td>
<td></td>
<td>pKKS825</td>
<td>Staphylococcus aureus</td>
<td>Rep_1 and HTH_Hin_like and Rep_3</td>
<td>ND</td>
</tr>
<tr>
<td>pDB2011</td>
<td>NC_021513.1</td>
<td>7.641</td>
<td>YP_008119849.1</td>
<td></td>
<td>pKKS825</td>
<td>Listeria innocua</td>
<td>Rep_1 and HTH_Hin_like</td>
<td>ND</td>
</tr>
<tr>
<td>pSIL30a</td>
<td>AUPT01000023.1</td>
<td>8.882</td>
<td>EP204218.1</td>
<td></td>
<td>pKKS825</td>
<td>Staphylococcus aureus</td>
<td>HTH_11</td>
<td>ND</td>
</tr>
<tr>
<td>pSC551</td>
<td>NC_005076.1</td>
<td>17.108</td>
<td>NP_889176.1</td>
<td></td>
<td>pKKS825</td>
<td>Staphylococcus sciuri</td>
<td>HTH_Hin_like and Rep_3</td>
<td>ND</td>
</tr>
<tr>
<td>pLB4</td>
<td>M33531.1</td>
<td>incomplete</td>
<td>AAA25252.1</td>
<td></td>
<td>pLB4</td>
<td>Lactobacillus plantarum</td>
<td>Rep_2</td>
<td>ssoT (5')</td>
</tr>
<tr>
<td>pMRI.5_2</td>
<td>NC_019900.1</td>
<td>5.206</td>
<td>YP_007215174.1</td>
<td></td>
<td>pLB4</td>
<td>Lactobacillus plantarum</td>
<td>Rep_1 and Rep_2</td>
<td>ssoT (3')</td>
</tr>
<tr>
<td>pLAC1</td>
<td>NC_014164.1</td>
<td>3.478</td>
<td>YP_003650630.1</td>
<td></td>
<td>pLB4</td>
<td>Lactobacillus acidophilus</td>
<td>Rep_1</td>
<td>ND</td>
</tr>
<tr>
<td>pLAA4</td>
<td>AF304384.2</td>
<td>8.135</td>
<td>ABG23031.1</td>
<td></td>
<td>pLB4</td>
<td>Lactobacillus lactis</td>
<td>Rep_3</td>
<td>ND</td>
</tr>
<tr>
<td>pPB1</td>
<td>NC_006399.1</td>
<td>2.899</td>
<td>YP_138221.1</td>
<td></td>
<td>pPB1</td>
<td>Lactobacillus lactis</td>
<td>Rep_2</td>
<td>ssoT (3')</td>
</tr>
<tr>
<td>pLKS14</td>
<td>NC_014135.1</td>
<td>3.196</td>
<td>YP_003620509.1</td>
<td></td>
<td>pLB4</td>
<td>Leuconostoc kimchii</td>
<td>Rep_2</td>
<td>ND</td>
</tr>
<tr>
<td>pMBL800</td>
<td>NC_019353.1</td>
<td>3.270</td>
<td>YP_006964795.1</td>
<td></td>
<td>pLB4</td>
<td>Leuconostoc mesenteroides</td>
<td>Rep_2</td>
<td>ND</td>
</tr>
<tr>
<td>pLAB1100</td>
<td>MSS222.1</td>
<td>incomplete</td>
<td>P58556.1</td>
<td></td>
<td>pLB4</td>
<td>Leuconostoc hilgardii</td>
<td>Rep_1</td>
<td>ND</td>
</tr>
<tr>
<td>pSMA23</td>
<td>NC_010242.1</td>
<td>3.497</td>
<td>YP_001649176.1</td>
<td></td>
<td>pSMA23</td>
<td>Lactobacillus casei</td>
<td>Rep_1</td>
<td>ND</td>
</tr>
<tr>
<td>pCC88</td>
<td>U31333.1</td>
<td>incomplete</td>
<td>AA74581.1</td>
<td></td>
<td>pSMA23</td>
<td>Lactobacillus casei</td>
<td>Rep_1</td>
<td>ND</td>
</tr>
<tr>
<td>pC141</td>
<td>ABS17606.1</td>
<td>incomplete</td>
<td>BAH73725.1</td>
<td></td>
<td>pSMA23</td>
<td>Lactobacillus plantarum</td>
<td>Rep_1</td>
<td>ND</td>
</tr>
<tr>
<td>pCD304-1</td>
<td>NC_016035.1</td>
<td>3.424</td>
<td>YP_004869658.1</td>
<td></td>
<td>pSMA23</td>
<td>Lactobacillus buchneri</td>
<td>Rep_1</td>
<td>ssoT (3')</td>
</tr>
<tr>
<td>pM4</td>
<td>NC_006662.2</td>
<td>3.320</td>
<td>YP_001621756.1</td>
<td></td>
<td>pSMA23</td>
<td>Lactobacillus plantarum</td>
<td>Rep_1</td>
<td>sso-new</td>
</tr>
<tr>
<td>pF8801</td>
<td>NC_007593</td>
<td>5.558</td>
<td>YP_339641.1</td>
<td></td>
<td>pSMA23</td>
<td>Pediococcus damnosus</td>
<td>Rep_1</td>
<td>ND</td>
</tr>
<tr>
<td>pCD304-2</td>
<td>NC_016034.1</td>
<td>2.707</td>
<td>YP_004869655.1</td>
<td></td>
<td>pSMA23</td>
<td>Lactobacillus buchneri</td>
<td>Rep_2</td>
<td>ssoT (3')</td>
</tr>
<tr>
<td>pG6301</td>
<td>NC_019372.1</td>
<td>3.516</td>
<td>YP_00696557.7</td>
<td></td>
<td>pSMA23</td>
<td>Lactobacillus plantarum</td>
<td>Rep_1</td>
<td>ND</td>
</tr>
<tr>
<td>pLB2095A02</td>
<td>NC_012549.1</td>
<td>3.524</td>
<td>YP_002709052.1</td>
<td></td>
<td>pSMA23</td>
<td>Lactobacillus brevis</td>
<td>Rep_1</td>
<td>ND</td>
</tr>
<tr>
<td>pS511</td>
<td>NC_014919.1</td>
<td>3.225</td>
<td>YP_004134615.1</td>
<td></td>
<td>pSMA23</td>
<td>Lactobacillus brevis</td>
<td>Rep_1</td>
<td>ND</td>
</tr>
<tr>
<td>pG12</td>
<td>NC_016981.1</td>
<td>4.572</td>
<td>YP_005332352.1</td>
<td></td>
<td>pG12</td>
<td>Lactococcus garvieae</td>
<td>Rep_2</td>
<td>ssoW (3')</td>
</tr>
<tr>
<td>pAMalPaha1</td>
<td>NC_050513.1</td>
<td>9.759</td>
<td>NP_863358.1</td>
<td></td>
<td>pAMalPaha1</td>
<td>Enterococcus faecalis</td>
<td>Rep_1+Rep_1 and Rep_3</td>
<td>ssoU (5')</td>
</tr>
<tr>
<td>unnamed</td>
<td>GGE70384.1</td>
<td>incomplete</td>
<td>EEU18290.1</td>
<td></td>
<td>pAMalPaha1</td>
<td>Enterococcus faecalis</td>
<td>Rep_3</td>
<td>ND</td>
</tr>
<tr>
<td>unnamed</td>
<td>GGE692894.1</td>
<td>incomplete</td>
<td>EEU66435.1</td>
<td></td>
<td>pAMalPaha1</td>
<td>Enterococcus faecalis</td>
<td>Rep_1+Rep_1 and Rep_3</td>
<td>ND</td>
</tr>
<tr>
<td>Plasmid</td>
<td>Accession</td>
<td>Size (bp)</td>
<td>Annotation</td>
<td>Replication Protein(s)</td>
<td>Antibiotic Resistance(s)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>-----------</td>
<td>------------</td>
<td>------------------------</td>
<td>--------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pTB53</td>
<td>D14852.1</td>
<td>incomplete</td>
<td>BAA03580.1</td>
<td>pUB110 Bacillus sp.</td>
<td>ND</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pIP1714</td>
<td>AF015628.1</td>
<td>incomplete</td>
<td>AAC61672.1</td>
<td>pUB110 Staphylococcus cohnii</td>
<td>Rep_1+Rep_1 ND PLS(A), MLS(B)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pNM11</td>
<td>NC_019558.1</td>
<td>11.383</td>
<td>pNM11</td>
<td>Planococcus citreus Rep_3</td>
<td>ND</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pBS-03</td>
<td>JQ94981.1</td>
<td>incomplete</td>
<td>pNM11</td>
<td>Bacillus sp. Rep_1</td>
<td>ND Flr/Chlr, Strp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pSS-03</td>
<td>NC_016054.1</td>
<td>7.122</td>
<td>pNM11</td>
<td>Staphylococcus arlettae Rep_1</td>
<td>ND Flr/Chlr, MLS(B)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p612</td>
<td>NC_019186.1</td>
<td>5.048</td>
<td>p612</td>
<td>Haemophilus influenzae Rep_3</td>
<td>ND β-lac</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pA1606</td>
<td>NC_019180.1</td>
<td>5.646</td>
<td>pA1606</td>
<td>Haemophilus influenzae Rep_3</td>
<td>ND β-lac</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Plasmids whose relaxases were retrieved by a PSI-BLAST using MobM_pMV158 (300-N terminal residues) as a query are listed.
- Underlined accession numbers denote those relaxase genes that are probably misannotated in the GenBank database (i.e., extended N-terminal sequence respect to that of MobM_pMV158).
- It locates the corresponding plasmid in one of the cartooned clusters of Figure 4, for which a prototype was selected.
- Replication initiation protein family. When more than one, their names are separated by "and". "+" is used for initiators that contain more than one pfam domain. Further details on replication initiation protein families can be found at http://pfam.sanger.ac.uk/.
- Only the previously identified ssos are annotated. Most of the ss sequences span 200-300 bp and are located in close proximity to the transfer module, with the exception of plasmid pUR2941 (ssosA was mapped 7 kb upstream mob). pM4 plasmid has a new type of ssos as described in (Yin et al., 2009). ND, not determined.  
- Antibiotic or metal resistance to: Tc, tetracycline; MLS(B), macrolide/lincosamide/streptogramin B; Lin, lincosamide; MEP, macrolide efflux protein; Strp, Streptomycin; Chlr, chloramphenicol; Kan, kanamycin; Neo, neomycin; Cu, copper; Cd, cadmium; PLS(A), pleuromutilins/lincosamide/streptogramin A; Trim, trimethoprim; Spec, spectinomycin; Ery: erythromycin; Ble, bleomycin; Flr, florfenicol; β-lac, beta-lactam; Tb, tobramycin.
Highlights

- The replication and transfer of rolling circle-replicating plasmids is reviewed.
- Comparisons of replication and transfer cassettes are presented.
- The current understanding of the pMV158 DNA transfer mechanism is reviewed.