Spread and Persistence of Virulence and Antibiotic Resistance Genes: A Ride on the F Plasmid Conjugation Module

GÜNThER KORAIMANN
Institute of Molecular Biosciences, University of Graz, Graz, Austria

ABSTRACT
The F plasmid or F-factor is a large, 100-kbp, circular conjugative plasmid of Escherichia coli and was originally described as a vector for horizontal gene transfer and gene recombination in the late 1940s. Since then, F and related F-like plasmids have served as role models for bacterial conjugation. At present, more than 200 different F-like plasmids with highly related DNA transfer genes, including those for the assembly of a type IV secretion apparatus, are completely sequenced. They belong to the phylogenetically related MOBF12A group. F-like plasmids are present in enterobacterial hosts isolated from clinical as well as environmental samples all over the world. As conjugative plasmids, F-like plasmids carry genetic modules enabling plasmid replication, stable maintenance, and DNA transfer. In this plasmid backbone of approximately 60 kbp, the DNA transfer genes occupy the largest and mostly conserved part. Subgroups of MOBF12A plasmids can be defined based on the similarity of TraJ, a protein required for DNA transfer gene expression. In addition, F-like plasmids harbor accessory cargo genes, frequently embedded within transposons and/or integrons, which harness their host bacteria with antibiotic resistance and virulence genes, causing increasingly severe problems for the treatment of infectious diseases. Here, I focus on key genetic elements and their encoded proteins present on the F-factor and other typical F-like plasmids belonging to the MOBF12A group of conjugative plasmids.

INTRODUCTION
Bacterial conjugation is the most sophisticated form of horizontal gene transfer (HGT) in bacteria and provides a platform for the spread and persistence of antibiotic resistance and virulence genes (1). The vehicles for transfer are conjugative plasmids (CPs) and integrative conjugative elements (ICEs). Based on phylogenetic analyses, distinct types of CPs can be differentiated. In the following review, I will focus on a family of conjugative plasmids that can be termed F-like because, in respect to DNA transfer genes, they share a common ancestry with the well-known F plasmid (or F, or F-factor). Ever since the first description of bacterial conjugation by Lederberg and Tatum (2), F and F-like plasmids were central in unraveling molecular mechanisms and structures facilitating the transfer of DNA between bacterial cells. Whereas conjugation is ubiquitous in the bacterial and archaeabacterial world (as are ICEs and CPs), F-like plasmids seem to be restricted to closely related enterobacterial genera such as Escherichia, Salmonella, Klebsiella, Shigella,
Enterobacter, and Citrobacter. These Gram-negative, facultative anaerobic bacteria can survive a broad range of environmental conditions, including the gastrointestinal tract of humans and animals, where they live as commensals or cause mild to severe diseases. Notably, F-like plasmids are present in commensal and various types of pathogenic Escherichia coli strains. With the advent of next-generation sequencing (NGS) techniques it has become evident that most E. coli isolates originating from the environment or from clinical samples contain at least one F-like plasmid. According to a plasmid classification scheme based on comparison and phylogenetic analyses of relaxase (traL) and coupling protein (traD) genes, they fall into the MOB\textsubscript{F\textsubscript{12}} group (3), which can be further classified into five major subgroups, or shades of F, with F and other “classical” F-like plasmids such as R100 or R1 belonging to the MOB\textsubscript{F\textsubscript{12}A} group (4). According to a recent analysis, the MOB\textsubscript{F\textsubscript{12}} group of plasmids is the most active player among E. coli ST131 ExPEC isolates involved in urinary tract infections. MOB\textsubscript{F\textsubscript{12}} plasmids were found to be pervasive in ST131 isolates, adding by themselves more than 350 protein families to the ST131 pangenome (5). The result of an analysis of the largest worldwide collection of 215 sequenced E. coli ST131 isolates demonstrated that emergence and spread of the resistance against broad spectrum cephalosporins in that pathogenic strain resulted from capture of CTX-M resistance genes on a conjugative F-like plasmid. From the original event that presumably happened in the early 1990s, the plasmid has been kept in and expanded with the population (6). I here present an overview of this important family of self-transmissible plasmids with a special focus on the genetic organization of F and typical F-like plasmids, including a review of the most important advances obtained from genetic, biochemical, and structural studies during the past decade.

**THE BACKBONE GENOME OF F AS A MODEL FOR MOB\textsubscript{F\textsubscript{12}A} PLASMIDS**

Figure 1 illustrates a genetic map generated from the F plasmid sequence (AP001918) with the origin of the map set to the first nucleotide following the stop codon of *slt*, the specialized lytic transglycosylase gene (formerly designated gene 19, gene X, ORF169) which is highly conserved among F-like plasmids. Its gene product can locally degrade peptidoglycan, thereby facilitating the assembly of the large type IV secretion (T4S) apparatus in the cell envelope (7, 8). In the clockwise direction, the origin of DNA transfer (oriT) and the DNA transfer (tra) genes, including T4S genes, are represented. The *finO* gene at the distal end of the *tra* operon is disrupted by an IS3 element making F “de-repressed” and highly transferable (9). Note that F is the exception among F-like plasmids, which in all other cases known so far possess an intact *finO* gene; F-like plasmids belonging to the MOB\textsubscript{F\textsubscript{12}A} (4) group, such as its prototype plasmid R1 (KY749247), are therefore “repressed” or “fertility inhibited,” meaning that only a few cells in a population (1 to 10 of 10,000 cells) express *tra* genes and DNA transfer functions and thereby become transfer competent (10, 11). Tight control and sophisticated regulation of DNA transfer genes and transfer competence development is not only the rule for F-like plasmids but is a common feature of conjugative plasmids (CPs) and integrative conjugative elements (ICEs) in general (10, 12). Whereas many research groups have investigated regulation, biological function, and structure of DNA transfer genes and proteins, little is known about genes situated in the “leading region” that, nevertheless, represent a conserved block important for the establishment of the plasmid in a new host (13). Replication of F is facilitated by the RepFIA (also known as IncFIA) replicon, whereas most F-like plasmids have an intact RepFII replicon that is disrupted and nonfunctional in F itself. Toxin-antitoxin (TA) systems prevent segregational loss of the plasmid. F is stabilized by two type I TA systems with an RNA antitoxin (*flm* and *srn* loci with similarity to the paradigmatic R1 encoded *hok-sok* system) and one well-characterized type II TA system (*ccdAB*, with a protein antitoxin). As described below, a fourth TA system (the second type II TA system, *vapBC*), which so far has been unnoticed, is located in the DNA transfer region of F. Partitioning of plasmid DNA molecules before cell division is mediated by a dedicated type I plasmid-partitioning system (*sapABC*).

**DIFFERENT GROUPS OF PLASMIDS EXIST IN THE MOB\textsubscript{F\textsubscript{12}A} PLASMID FAMILY**

F is a typical MOB\textsubscript{F\textsubscript{12}A} plasmid, with the exception that the fertility inhibition gene, *finO*, is disrupted, making transfer genes derepressed and conjugation rates high. Other classical F-like plasmids are the antibiotic resistance plasmids isolated already back in the 1960s, such as plasmid R1 (KY749247) or plasmid R100 (AP000342) (14). More recently, pSLT (AE006471) plasmids from Salmonella enterica joined this group as well as plasmids isolated from Klebsiella (e.g., pKDO (JX424423), pKPN3...
It was noticed early on that there are important differences in the DNA transfer genes that were specific for a given plasmid. One gene that encodes a key regulatory gene for transfer gene expression, TraJ, varies extensively (11, 15). Representative plasmids of MOB\textsubscript{F\textsubscript{12}}A subgroups that can be defined based on TraJ sequence variation are listed in Table 1. F plasmid encoded TraJ, for instance, does not share more than 23% identical amino acids with any of the other TraJ proteins. A consequence of TraJ variations, the plasmid subgroup-specific activation of DNA transfer genes, is discussed later in this review.

**VARIABILITY AND MOSAICISM IN MOB\textsubscript{F\textsubscript{12}}A PLASMIDS**

All the above-introduced genetic modules provide elementary functions in F and F-like plasmids and may therefore be viewed as the plasmid backbone, or core,
encompassing approximately 60 kbp of DNA. Consequently, 60 kbp represents the minimal genome size of a typical F-like plasmid. Interestingly, one recently sequenced plasmid from a *S. enterica* subsp. enterica serovar Heidelberg strain 12-4374, plasmid p12-4374_62 (NZ_CP012928) is approximately 63 kbp in size. Its TraJ coding sequence (CDS) is 98% identical to that of F; the bacterium was isolated in 2012 in Canada from human stool. In addition to the plasmid backbone, it solely carries colicin toxin and immunity genes (3 kbp) as cargo. Cargo gene load on F-like plasmid backbones varies remarkably, and the total size can be more than 300 kbp, as is the case with pKPN-262 (NZ_CP000734), a plasmid isolated in 2012 from *Klebsiella pneumoniae* in the United States. Its TraJ CDS is 97% identical to that of pKPN3. One mid-sized example is multidrug resistance plasmid pECwhn14 (CP012197) with a still impressive size of 185 kbp. Its TraJ CDS again is 99% identical to that of F. This plasmid was present in an *E. coli* strain isolated in 2014 in China from healthy chickens and carries ten different antibiotic resistance genes. Since plasmids are genetically highly dynamic and mosaic, backbone elements come in various flavors (see below) and can be exchanged among F-like plasmids, creating a large genetic pool that builds the base for many combinations. Together with the variability of additional genetic material that is usually considered as the “cargo” that confers phenotypic traits, often flanked by or embedded in mobile elements (transposons, IS sequences, integrons), this genetic mosaicism makes a classification of these plasmids and epidemiological studies at least difficult (16). Typical core elements and some important cargo gene features of F and other F-like plasmids from the MOBF12A group (4) are described in this review, followed by an overview on bacterial conjugation mediated by F and F-like plasmids.

<table>
<thead>
<tr>
<th>Name*</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: R1 (228) KY749247</td>
<td>100</td>
<td>73</td>
<td>20</td>
<td>18</td>
<td>18</td>
<td>20</td>
<td>18</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>2: pSLT (228) AE006471</td>
<td>73</td>
<td>100</td>
<td>19</td>
<td>21</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>3: R100 (223) AP000342</td>
<td>20</td>
<td>19</td>
<td>100</td>
<td>21</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>4: pKPN3 (231) CP000648</td>
<td>18</td>
<td>21</td>
<td>21</td>
<td>100</td>
<td>25</td>
<td>26</td>
<td>21</td>
<td>19</td>
<td>23</td>
</tr>
<tr>
<td>5: pKDO (239) JX424423</td>
<td>18</td>
<td>20</td>
<td>21</td>
<td>26</td>
<td>100</td>
<td>61</td>
<td>22</td>
<td>22</td>
<td>28</td>
</tr>
<tr>
<td>6: pKOXM1B (230) CP008843</td>
<td>20</td>
<td>20</td>
<td>22</td>
<td>25</td>
<td>61</td>
<td>100</td>
<td>21</td>
<td>21</td>
<td>28</td>
</tr>
<tr>
<td>7: F (229) AP001918</td>
<td>18</td>
<td>20</td>
<td>23</td>
<td>21</td>
<td>22</td>
<td>21</td>
<td>100</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>8: pKF3-70 (215) FJ494913</td>
<td>18</td>
<td>20</td>
<td>22</td>
<td>19</td>
<td>22</td>
<td>21</td>
<td>23</td>
<td>100</td>
<td>37</td>
</tr>
<tr>
<td>9: pIP1206 (215) NC_010558</td>
<td>21</td>
<td>19</td>
<td>21</td>
<td>23</td>
<td>28</td>
<td>25</td>
<td>22</td>
<td>37</td>
<td>100</td>
</tr>
</tbody>
</table>

*Plasmid names are given in the first column. Number of amino acids in TraJ are indicated in parentheses. To access a complete plasmid sequence in the NCBI nucleotide database (https://www.ncbi.nlm.nih.gov/nuccore), please use the accession number provided. An identity matrix of pairwise amino acid comparisons of TraJ proteins is shown. Numbers indicate percentage of identical amino acids. The table was derived from a Clustal Omega phylogenetic sequence analysis (https://www.ebi.ac.uk/Tools/msa/clustalo/) of the listed TraJ protein sequences.
FEATURES OF F AND TYPICAL F-LIKE PLASMIDS FROM THE MOB$_{F_{12}A}$ GROUP

Plasmid Replication and RepF Loci: It’s Better to Have More Than One

The primary RepFIA replication module (Fig. 1) enabling the maintenance of F at a very low copy number of one to two per cell was a focus of investigation in the early research on the F plasmid (17). A 3-kbp fragment flanked by PstI and KpnI restriction enzyme sites was found to be sufficient for providing replication and incompatibility features. It encodes the replication initiator protein RepE that is needed for DNA strand separation and recruitment of host replication proteins. The repE sequence is flanked by direct repeated 19-bp binding sites for RepE (18), termed iterons, an AT-rich region and DnaA boxes (oriS) that represent well-characterized features of an origin of replication (19). With the use of purified RepE protein, replication of F could be reconstituted in vitro (20), and, based on the crystal structure of RepE, its function in stringent regulation of F plasmid replication has been proposed (21). In brief, in the proposed model, RepE can bind as a dimer to its operator sites leading to transcriptional repression and to prevention of replication initiation. Replication initiation by RepE requires monomerization that is realized by the DnaK-DnaJ-GrpE chaperone system. When the plasmid copy number is low, this leads to RepE (monomer) binding to the oriS iterons and initiation of plasmid replication (21). Besides this primary replication module, F harbors the RepFIB replicon about which little is known. The repB gene encodes a replication initiator protein with a topoisomerase motif; however, it is unknown how and when the RepFIB replicon is operational but it can sustain replication in the absence of RepFIA (13). The third replicon found on the F plasmid is RepFII (originally termed RepFIC); however, a Tn1000 insertion disrupts the repA gene, rendering this replication initiation region nonfunctional. In the F-like plasmid R1 (KY749247), it is the only replicon that is present enabling replication of the plasmid at a low copy number of one to two per chromosome. Replication of plasmid R1 as well as the functions of replication genes have been investigated in great detail, making R1 prototypic for plasmids carrying the RepFII replicon (22). In contrast to the RepFIA iteron-type control, the RepFII-encoded replication initiation protein RepA is controlled by a negative regulator of transcription, CopB, and a small regulatory antisense RNA, CopA (23, 24). RepA synthesis in plasmid R1 is regulated by CopA RNA through inhibition of leader peptide (tap open reading frame [ORF]) translation (25). RepFII replication presumably enables rapid replication when plasmid R1 is transferred to a new host by conjugation because of a temporary lack of CopB repressor; furthermore, replication is connected to a type II toxin-antitoxin module (26).

The presence of multiple replicons is found in several cases among the MOB$_{F_{12}A}$ group of conjugative F-like plasmids. For instance, plasmid pECC-1470_100 (100 kbp, R1 subgroup, CP010345), a virulence plasmid from an E. coli strain isolated from bovine mastitis (27), harbors all of the above-described replicons in an undisrupted form. Most of the MOB$_{F_{12}A}$ plasmids, however, contain a RepFII replicon that comes in many sequence variants, presumably enabling independent replication and plasmid compatibility. Interestingly, from 493 completely sequenced plasmids classified as MOBF, 359 (approximately 73%) were listed as plasmids containing a RepFII (IncFII) replicon (28).

In this context, it is important that, besides the already-introduced MOB-based classification scheme (3), another widely used classification method uses the replicon of a given plasmid. In the current replicon-typing efforts, F-like plasmids are defined by the presence of one of the above-described replicons or sequence variants thereof (29, 30). Efforts to work out a taxonomy of large conjugative plasmids (31) will be based on a combination of mobility and replicon-typing schemes as well as analyses of complete plasmid sequences available in the nucleotide databases. This will definitely help researchers to better categorize newly sequenced plasmids or track epidemiological paths. Also important are efforts to improve the annotation quality of new plasmid sequences that will be deposited by using naming conventions such as those that have been recently suggested (32).

Partitioning: Positioning of Plasmid Molecules Before Cell Division

For all large, low-copy-number plasmids, as well as bacterial chromosomes, a dedicated DNA distribution system, termed partitioning system, exists. Partitioning systems are active after plasmid replication and before cell division. They ensure that plasmid copies are positioned in such a manner that after cell division both daughter cells contain at least one plasmid. Partitioning complexes consist of an ATPase (or GTPase), a centromere sequence-binding protein (CBP), and the centromere-like DNA sequences (33). From the three known classes
of partitioning systems two are found in F-like plasmids: Type I ATPase (P-loop) system represented by F plasmid SopA (ATPase) SopB (CBP) sopC. The prototype of type I ATPase partitioning systems is the P1 plasmid ParA/ParB/parC system. The second type is represented by the R1 plasmid type II ATPase (actin-like) system consisting of ParM (ATPase) ParR (CBP) parC. Both systems, although genetically organized in a similar fashion, fundamentally differ in sequence and structure of the ATPase as well as in the molecular mechanism underlying plasmid partitioning (33, 34). Regarding its mechanism, the type I ATPase system, which is the most widespread form of plasmid- and chromosome-partitioning system across the bacterial kingdom, has remained elusive. Nevertheless, from in vivo and in vitro studies of different type I ATPase systems, a few models for plasmid partitioning have been developed. Note that, for simplicity reasons, the F plasmid nomenclature is used. In one model, SopA filaments formed on chromosomal DNA depolymerize upon contact with the SopB-bound plasmid, thereby pulling plasmids toward opposite poles of the nucleoid (35). Another is the diffusion-ratchet mechanism of plasmid partition (33) in which the SopB-plasmid DNA complex would move along a gradient of SopA toward the pole of a nucleoid (34, 36). Finally, results of a very recent super resolution microscopy study led the authors to propose a “Venus flytrap” model that envisions the SopB-plasmid DNA complex entrapped inside a nucleoid-localized 3D meshwork of SopA clouds that oscillate between nucleoid poles. At the edge of a nucleoid, one plasmid drops from the mesh, whereas the sister plasmid remains tethered and is shifted to the opposite end where it eventually is released (37).

The type II ATPase systems have been investigated in great detail using the plasmid R1-encoded ParMRC system, which is also the prototype of these plasmid-partitioning systems (38). In brief, the ParM ATPase in that system found in plasmids of Gram-positive and Gram-negative bacteria is an actin-like protein. In its ATP-bound form, ParM can dynamically form filamentous polymers, resembling actin protofilaments, that are specifically used to move ParR-bound plasmids to opposite cell poles before cell division (39). By a combination of structural studies and fluorescence microscopy, it was later shown that ParR bound to parC could accelerate growth at only one end of a polar helical ParM protofilament (40). Furthermore, two ParR-plasmid-capped filaments were demonstrated to associate in an antiparallel orientation, forming a bipolar spindle (40). Recent cryo-electron microscopy (cryo-EM) studies refined the model and revealed how subunits and filaments associate within cells to form a simple yet very elegant and efficient mitotic machinery (41).

Toxin-Antitoxin Modules: Deadly Partners for Stable Maintenance

Despite highly efficient DNA partition modules F and F-like plasmids can additionally prevent loss from a bacterial cell. This is achieved via the biological function termed postsegregational killing (PSK) indicating that cells that have lost the plasmid after cell division die because of the activation of a toxin. The underlying molecular mechanism involves toxin-antitoxin (TA) pairs in which the toxin is not active in plasmid-bearing cells and controlled by the activity of a neutralizing antitoxin. These modules can be viewed as selfish DNA elements, but, since they are also present in multiple locations on bacterial chromosomes, their preservation during evolution may derive from their general effect as sentinels of genome integrity. Indeed, it has been demonstrated they can diminish large-scale genome reductions (42). It is noteworthy that toxins of chromosomally encoded TA systems can be activated stochastically or by environmental cues to form a small subpopulation of slow-growing persister cells that are tolerant to antibiotics (43, 44). However, in a recent study, no evidence for the involvement of TA systems in the formation of persister cells under unstressed conditions was found (45). Six types of TA systems have been grouped according to the mechanism of the antitoxin and whether the antitoxin is an RNA or a protein. The target of the toxin protein varies, but, in most cases, an essential function of a bacterial cell such as replication, translation, or membrane integrity is affected (46, 47).

The best-known TA systems are the type I and type II TA systems. Indeed, both types are commonly found on F-like plasmids where they act as effective stabilizers of plasmid carriage in a bacterial population. The canonical hok (host killing)–sok (suppressor of killing) type I system was originally found on plasmid R1 (48). Mechanistically, this system and the related srnB type I TA system of plasmid F work as follows. Plasmid-bearing cells produce a primary transcript encompassing the mok and hok open reading frames. This nontranslatable mRNA is very stable but can be slowly processed at its 3′ terminus to form a truncated form which is, in principle, translationally active. Translation of Mok and the
translationally coupled Hok mRNA, however, is blocked by the binding of a small regulatory RNA, Sok, to the translation initiation region (TIR) of Mok. Upon loss of a plasmid carrying this TA system, the newborn cell inherits the stable primary transcript that after processing gets translated, because Sok RNA has already been degraded because of its inherent instability (49). The small toxin protein (CDD protein domain family: cl27487) can insert itself into the cytoplasmic membrane, causing perturbations of membrane integrity and a breakdown of the electrochemical gradient and hence ATP synthesis.

In addition to the type I TA system, F-like plasmids typically harbor at least one representative of the various subfamilies of type II TA systems. On the basis of structural and functional features of the toxin, bacterial type II TA modules were categorized into eight superfamilies: RelBE, MazEF, VapBC, CcdAB, ParDE, HigAB, HipBA, and Phd–Doc (47). F itself is known for the CcdAB TA system. The ccd (coupled cell division) locus is adjacent to the origin of replication of the F plasmid and enhances F plasmid stability by coupling host cell division to plasmid proliferation (50). It is now well established that CcdB encodes a toxin that, similar to ParE toxin (from plasmid RK2) and quinolone antibiotics, inhibits replication as well as transcription by targeting DNA gyrase (51). Under normal conditions, i.e., when F is present, CcdB is neutralized by the CcdA antitoxin protein. The toxin is unleashed when the plasmid is lost because of rapid antitoxin degradation by Lon protease (52). The kis/kid (killing suppression/killing determinant) type II TA system (pemI/pemK in plasmid R100 is identical to kis/kid) is encoded adjacent to the RepIII replicon of plasmid R1 and represents an example where the Kid (PemK) toxin is a sequence-specific RNase, thereby inhibiting translation of cellular mRNAs (53). It has also been shown that in case of a too-low-copy-number state before cell division, the Kid toxin can not only inhibit cell growth and division, but also stimulate plasmid replication by specifically inactivating the mRNA encoding the negative regulator CopB (26, 54, 55). The Kis/Kid type II TA system belongs to the MazEF superfamily (MazF CDD protein domain family: cl00995).

As with some bacterial chromosomes, multiple type II TA systems of different superfamilies can be present on the same plasmid. This is, for instance, the case with the MOB<sub>F</sub><sub>pl</sub> A plasmid pECC-1470_100 (100 kbp, R1 subgroup, CP010345) that not only harbors three replicons but, in addition to the CcdAB TA system, carries a VapBC TA pair where VapC is a toxin that inhibits translation and cell growth by cleavage of initiator tRNA (56). The large F-like virulence plasmid pSLT (NC_003277) which is highly prevalent in pathogenic Salmonella enterica serovar Typhimurium strains also harbors these type II TA systems that were designated CcdAB<sub>st</sub> and VapBC<sub>2st</sub> respectively (57). Whereas the nonfunctional CcdB toxin still could stabilize the plasmid (58), VapC<sub>2st</sub> was additionally capable of promoting bacterial survival inside fibroblasts and epithelial cells (57). Therefore, some distinct plasmid located TA modules do not only provide a means to stabilize the plasmid in a population, but in addition do play a role in other established functions of TA systems, i.e., antibiotic tolerance, persistence, pathogenicity, or biofilm formation (59).

In pSLT the VapBC<sub>2st</sub> TA system is encoded within the trbH gene which is located in the DNA transfer operon (see below) between traD and traI. Interestingly, trbH, which is not essential for conjugation, is also present in F and some other F-like plasmids, but only in about 10% of all MOB<sub>F</sub><sub>pl</sub> plasmids (60). A quick BLAST search with the ORFs present in the opposite direction of the trbH ORF revealed the presence of the hitherto unnoticed second type II TA system, VapBC, on plasmid F. The VapC amino acid sequence between plasmids pSLT and F is highly conserved (94.7% sequence identity).

**DNA on the Move: A Transfer Region Overview**

The largest part of the backbone of F and F-like plasmids belonging to the MOB<sub>F</sub><sub>pl</sub> A group is occupied by DNA transfer (tra) genes that are required for transferring the plasmid from one bacterium to another. They can be subordinated according to their function in the DNA transfer process (for a list, see Table 2). The tra gene region of plasmid F (details are shown in Fig. 2) encompasses approximately 35 kbp with slt as the first and finO as the last of the DNA transfer genes (in total 38 genes including slt and finP). Only 9 of 38 genes encode proteins with homology to the paradigmatic Agrobacterium tumefaciens Ti plasmid T4S reference system (61), occupying approximately one-third (12 kbp) of the tra region. A VirB11-like ATPase is lacking in F-like plasmids, as well as a clearly defined homologue of VirB8 (62). In contrast to other conjugation systems, the obligatory relaxase gene (traI) and the coupling protein gene (traD) are part of the tra gene array. Most of the other tra or trb genes are specific for F-type T4SS and characteristic for the entire MOB<sub>F</sub><sub>pl</sub> A group of plasmids. The following...
genes present in F but not in all other MOB\textsubscript{F12A} group plasmids (4) were found to be neither essential for pilus biogenesis nor for DNA transfer: \textit{artA}, \textit{traP}, \textit{traR}, \textit{trbA}, \textit{trbD}, \textit{trbF}, \textit{trbG}, \textit{trbH}, \textit{trbJ} (13). Whereas largely conserved (4), \textit{trbB} and \textit{trbE} have also been found to be nonessential for conjugation (13). F-specific genes involved in relaxosome formation and/or regulation of \textit{tra} genes are \textit{traM}, \textit{traJ}, \textit{finP}, \textit{traI}, \textit{traY}, and \textit{finO} (disrupted in F by an IS\textsubscript{3} insertion). Other F-specific \textit{tra} genes enable pilus assembly reactions, F-pilin modification, mating pair stabilization and surface or entry exclusion functions. MOB\textsubscript{F12A} \textit{tra} genes, encoded proteins, and their functions are described in more detail in “Bacterial Conjugation Mediated by F and Other MOB\textsubscript{F12A} Plasmids,” below.

### The Leading Region and Establishment in the New Host

In all known DNA transfer systems involving a dedicated T4SS, single-stranded plasmid DNA (ssDNA) is transferred into a recipient cell. There, the plasmid must be replicated and established. Potentially harmful for a plasmid arriving in a new host are commonly present antiforeign DNA weapons such as restriction enzymes (REs) and the adaptive CRISPR-Cas defense system, or
the induction of the SOS response by the arriving ssDNA. REs may be evaded, in part, by the fact that incoming ssDNA is not recognized and cleaved by RE unless a double-stranded stem containing a recognition site is present. Second, in many cases, one dedicated antirestriction gene is frequently located in the leading region of conjugative plasmids (63). Among those, the ardB (alleviation of restriction of DNA) gene is located on the leading region of F, R1 and many other F-like plasmids (Fig. 1). ArdB is similar to KlcA and exhibits antirestriction activity against type I restriction-modification systems in vivo but not in vitro, suggesting that the inhibitory activity of ArdB is indirect and different from DNA mimicking antirestriction proteins such as ArdA or Ocr (64). Other conserved genetic elements in the leading region are: a gene encoding a putative methyltransferase, single-stranded-initiation (ssi) sites, ssb, psiA, and psiB.

The function of the putative methyltransferase (originally designated ORF227 in F) is unknown; however, it has been noted that the role of Orf227 might be to protect the transferred DNA from restriction endonucleases in the recipient cell (65). The first single-stranded-initiation site (ssi in F) is located proximal to the origin of transfer (oriT). Of this sequence 137 bp are perfectly repeated in ssi2 and represent a part of a large imperfect inverted repeat that forms a duplex DNA in the transferred ssDNA. This dsDNA hairpin forms a promoter sequence that is transcribed by RNA polymerase yielding primer RNA for replication and mRNAs for expression of leading region genes (66). Plasmid R1 is an example where three ssi sites are present in the leading region (60), a situation similar to the IncI1 plasmid ColIB-P9 where promoter activity was demonstrated for single-stranded ssi3 (67). One of the genes that are presumably mainly transcribed by ssi2 and induced immediately after transport of ssDNA into the recipient is psiB (68), it encodes a protein that interferes with induction of the SOS response by specifically binding to and inhibiting activities of the RecA protein, among them filament formation on ssDNA-binding protein (SSB) coated ssDNA, thereby abolishing the SOS response (69). Finally, the first gene to be transcribed from ssi2 is ssb encoding the SSB protein, which is thought to confer important plasmid establishment and replication functions (65). The chromosomal homologue in E. coli, SSB, is involved in DNA replication, repair, and recombination (70). A deletion of the chromosomal ssb gene could be complemented by F plasmid ssb carried on a multicopy plasmid (71), demonstrating functional equivalence. Several other leading region genes and ORFs are conserved, but their function in plasmid stability, establishment, or transfer remains to be established.

An interesting question is whether conjugative elements (ICEs and CPs) including MOBp12,A plasmids are targeted by CRISPR-Cas spacers. Whereas bacterial viruses clearly represent a threat to bacteria, CPs like F can be considered beneficial for the host. A recent bioinformatics-based screen revealed that 80 to 90% of detectable protospacers were mainly found in bacteriophage sequences. Among the re-
mainly protospacers, chromosomal sequences originating from mobile genetic elements (MGEs) were dominant. Top identified spacer targets were VirB4, VirD, and relaxase sequences (72). Although it was found that CRISPR-Cas systems can prevent conjugation in staphylococci (73), horizontal gene transfer in general does not seem to be targeted by CRISPR-Cas on evolutionary timescales (74).

IS Elements, Transposons, and Integrons: Highly Efficient MGEs for Incorporation of Cargo Genes

For a plasmid, genes that confer phenotypic traits to bacterial hosts can be defined as cargo. In many cases, cargo genes enable the host to survive environments that otherwise would be detrimental. Cargo genes of MOB_{\text{P}_{12}}A plasmids usually reside in or are flanked by mobile genetic elements that can jump from one DNA site to another. This movement of IS elements or transposons can be conservative or replicative. Frequently, antibiotic resistance genes are captured into gene cassettes termed integrons (75) that then preferentially reside in often complex assemblies, like genetic Russian dolls, on plasmids (76). One such assembly is the composite transposon Tn21, which for several good reasons has been termed the “flagship” of the floating genome (77). The prototype Tn21 is present on plasmid R100 (also known as NR1, 94 kbp, AP000342), which was originally isolated from Shigella flexneri in the late 1950s (78). Contained within Tn21 are mercury resistance genes as well as a class I integron. Two Tn21 variants of archetypical ABR plasmids R100 and R1 are depicted in Fig. 3. The basic structure shown is from plasmid R100; R1 differs from that only in two aspects: R1 lacks IS1353 but has a Tn3 (79) inserted into merP of the mercury resistance gene cluster, making R1 additionally resistant to ampicillin. Tn21 in both cases is embedded in a Tn9-like transposon that is composed of two flanking IS1 elements carrying a chloramphenicol resistance gene. Contained within Tn21 is a typical class I integron with intI1 encoding the integron integrase capable of capturing resistance gene cassettes into the integron array (75).

Another MGE that is prominent in more recently isolated MOB_{\text{P}_{12}}A plasmids is IS26. This IS element is another very efficient resistance gene-capturing device that can lead to the formation of arrays containing multiple copies of IS26 with various incorporated ABR genes (80, 81). One example of a MOB_{\text{P}_{12}}A plasmid with an IS26 array is plasmid pARS3 (115 kbp, R1 subgroup, AB261016) that was present in an E. coli strain isolated from urine of a hospital patient in Japan in 2003 (82).

Cargo I: Antibiotic, Biocide, and Metal Resistance Genes

From the perspective of the growing threat to human health caused by the rapid spread and persistence of antibiotic resistance, the most dangerous feature on F and F-like plasmids is the presence of antibiotic resistance, the most dangerous feature on F and F-like plasmids is the presence of antibiotic resistance genes. Among those bacteria that are top listed in the threat list published by the WHO in 2017 (http://www.who.int/medicines/publications/global-priority-list

Figure 3 A cargo gene region from the classical MOB_{\text{P}_{12}}A antibiotic resistance plasmids R1 and R100 is shown. In both cases, this resistance gene region is dominated by the composite transposon Tn21 that is flanked by IS1 elements carrying a catA1 gene. Contained within this transposon are mercury resistance genes as well as a class I integron. The basic structure shown is from plasmid R100. R1 does not contain IS1353 in the integron, but additionally has a Tn3 inserted in the merP gene. Thus, this region encodes resistance to chloramphenicol (catA1), sulfonamides (sul1), streptomycin (aadA) in both plasmids, and in R1 additionally to ampicillin (blaTEM-1).
-antibiotic-resistant-bacteria/en/) are extended-spectrum beta-lactamase (ESBL) producing enterobacteria causing sometimes untreatable infections acquired in hospital settings (83). ESBL encoding CTX-M genes as well as K. pneumoniae carbapenemase (KPC) and New Delhi Metallo-beta-lactamase (NDM) are frequently found on F-like plasmids present in enterobacterial hosts. In some cases, more than 10 different ABR genes reside on a single MOB<sub>F</sub>A plasmid (28), eventually giving rise to multi-drug-resistant enterobacteria. Thus, it is obvious that F-like plasmids have widely contributed to the selective-pressure-driven flow of preexisting antibiotic resistance genes from the environment (soil bacteria) to bacterial pathogens (84–86). In the reported cases, the resistance genes reside on MOB<sub>F</sub>A plasmids isolated from E. coli or K. pneumoniae. ESBL genes on F-like plasmids also made their way to Salmonella; a recent whole genome sequencing effort identified the CTX-M-27 carrying MOB<sub>F</sub>A plasmid pESBL931 (68 kbp, R100 subgroup, CP016389) as the cause of cefotriaxone treatment failure for an invasive S. enterica serovar Typhimurium infection (87). Not surprisingly, resistance to an antibiotic of last resort, colistin, was also recently reported on the MOB<sub>F</sub>A plasmid pMR0516mcr (225 kbp, F subgroup, fusion with an IncHI plasmid, KX276657). This plasmid in total harbors eight ABR genes and was present in an E. coli isolate from a patient with a urinary tract infection in the United States (88). It is noteworthy that colistin resistance transmission by CPs has only recently begun. It was first reported in 2015 in the form of the plasmid-borne mcr1 gene in an E. coli isolate from a pig farm in Shanghai (89). Subsequently, mcr1 was discovered on plasmids worldwide (90). Beside ABR genes, biocide and heavy metal resistance genes are frequently present on the same MOB<sub>F</sub>A plasmid. This is largely because class 1 integrons harbor the gene conferring resistance to quaternary ammonium compounds (qacEΔ1), whereas Tn21-like transposons carry genes for mercury (mer) resistance (Fig. 3 and references 75 and 77). Not surprisingly, in a recent study on the co-occurrence of resistance genes to antibiotics, biocides, and metals, it was found that plasmids harboring all these elements were more likely to be conjugative and carry toxin-antitoxin systems (91).

Cargo II: Colicins and Microcins

Colicins are protein toxins that can kill sensitive E. coli bacteria or close relatives by different mechanisms, e.g., by forming pores in membranes or by degrading nucleic acids. Expression of colicin and immunity genes is coupled to SOS induction and can ultimately result in lysis of the colicin-producing cell due to the production and activity of a colicin lysis (or release) protein (92). However, no such release protein-coding gene is associated with group B colicins that are present on large conjugative plasmids. These plasmids can harbor one or two colicin operons side by side (92). A scenario of how and when the colicin toxin is released in case of group B has been recently provided. Prophages that can also be induced by activation of the SOS response have been demonstrated to provide an exit route for a B-type colicin (93). Paradoxically, the combination of two killing systems (bacteriophage and colicin) can thereby promote fitness of S. enterica serovar Typhimurium harboring the group B colicin-encoding plasmid (93, 94). In contrast to group B colicins, colicin V (also known as ColV, MccV), a small pore-forming antimicrobial protein belonging to class Ila of microcins (95), is secreted from producing cells by a dedicated type I secretion machinery (96). Examples of MOB<sub>F</sub>A plasmids carrying group B colicin or MccV (ColV) are pEC14III (80 kbp, R1 subgroup, KU932028) and pS286colv (98 kbp, F subgroup, HF922624), respectively. Interestingly, colicin genes are often associated with virulence genes.

Cargo III: Virulence Genes and Enterotoxins

The classical F plasmid neither carries antibiotic resistance nor colicin genes as cargo. The cargo gene region, however, contains genetic elements that can enhance the virulence of enterobacteria containing this plasmid. The ompP gene located on the F plasmid encodes an outer membrane protease belonging to the ompTs that can degrade cationic antimicrobial peptides, thereby contributing to the ability of the host bacterium to evade the innate immune system (97). OmpP encoded by plasmid F displays a proteolytic activity similar to OmpT to which it is 70% identical (98). Whereas the ompT plasmid (plasminogen activator, approximately 50% identical to OmpP) of Yersinia pestis contributes by several mechanisms to bacterial virulence in plague, a role for E. coli OmpT protease in pathogenicity was originally not clear (99). Recently, it was shown that inactivation of ompT decreased adhesion, invasion, colonization, and proliferation capacities of avian pathogenic E. coli (100). Besides its activity against cationic antimicrobial peptides where subtly different specificities among the closely related ompTs OmpP/OmpT/CroP exist (101), another function may be the proteolytic processing of surface autotransporter adhesins that has been demonstrated to
be required for adhesion/aggregation in case of YapE processing by Pla of Y. pestis (102). The location of two AIDA-I like autotransporter genes (ychA and ychB) next to  ompP on the F plasmid is consistent with such a function. AIDA-I (adhesin involved in diffuse adherence) are classical type Va autotransporters associated with adherence and aggregation functions from enteropathogenic E. coli (103). In addition, TraT is an outer membrane lipoprotein encoded within the tra operon that not only serves as a surface exclusion protein, but has also been implicated in serum complement resistance, thereby providing another F plasmid-encoded virulence trait (104). Hence, together with the presence of the VapBC type II TA system that can function as a virulence determinant in the Salmonella virulence plasmid pSLT (see above), F can be classified as a bona fide virulence plasmid. It is noteworthy that, with respect to its cargo genes, F is highly similar to plasmid pECC-1470_100 (100 kbp, R1 subgroup, CP010345), a virulence plasmid from an E. coli strain isolated from bovine mastitis (27).

Other virulence determinants are iron acquisition and uptake systems that are usually under Fur (ferric uptake regulator) transcriptional control (105). Again, those systems can reside on MOB_F plasmids such as pAPEC-O1-ColBM (174 kbp, F subgroup, DQ381420), a plasmid isolated from an avian pathogenic E. coli (APEC) strain. With the related pAPEC-O2-ColV (184.5 kbp, R1 subgroup, AY545598) plasmid, it shares the possession of colicin and four iron transport and acquisition systems (106). Together with other virulence genes (e.g., iss, increased serum survival, tsh, autotransporter/adhesin) present on the same plasmid, these F-like plasmids are responsible for the pathogenic phenotype of APECs causing colibacillosis in birds (106). It is noteworthy that virulence, colicin, and antibiotic resistance genes can be present on a single MOB_F plasmid such as pCERC3 (144 kbp, F subgroup, KR827684) (107). Plasmid pS88 (also known as pECOS88, 134 kbp, F subgroup, CU928146) is related to the plasmids from APEC mentioned above but plays a role in neonatal meningitis and was present in the NMEC strain S88 isolated from cerebrospinal fluid of a newborn in France. This virulence plasmid was shown to contribute to pathogenicity since a variant of S88 lacking pS88 displayed a marked loss of virulence in a neonatal rat model (108). Cargo genes presumably involved in virulence present on pS88 include three iron uptake systems, the iss gene, the etsABC genes encoding a putative type I secretion system; ompT, encoding an outer membrane protease (omptin, 70% sequence identity to ompP of F), and hlyF. HlyF was demonstrated to encode a hemolysin (109) and, more recently, a protein promoting outer membrane vesicle (OMV) formation (110). In addition, pS88 carries a MccV (ColV) gene cluster for production of immunity to and secretion of the microcin V (colicin V). Another MOB_F plasmid that has been associated with meningitis is pRS218 (114 kbp, R1 subgroup, CP007150). pRS218 was found in E. coli strain RS218 isolated from the cerebrospinal fluid of a neonate with meningitis in 1974, a strain considered a prototype of NMEC. A plasmid-cured strain was significantly less virulent relative to the RS218 wild-type strain in vitro and in vivo (111). The cargo gene region of this plasmid, which is nearly identical to pUTI89 (114 kbp, R1 subgroup, CP000244), a plasmid associated with urinary tract infections (112), contains, among others, iron uptake, type I secretion, enterotoxin production, and copper tolerance genes as virulence factors (111). Apart from the pathovars of extraintestinal pathogenic E. coli strains (ExPEC such as NMEC, APEC, UPEC), a second group of pathogenic E. coli strains can be classified as diarrheagenic (113). Among those, enterotoxigenic E. coli (ETEC) can colonize the small intestine of humans and animals causing diarrheal disease with often fatal consequences. Possession of toxin genes and production of heat-labile (LT) and heat-stable (ST) toxins is considered to be a minimum requirement for pathogenicity (114). Not surprisingly, genes encoding these key virulence factors are frequently found on mobile elements; in the case of ETEC strain H10407 they are located on a MOB_F plasmid, termed pEntH10407 (also known as pETEC666 or p666, 117 kbp, F subgroup, CP022914). Its transfer region is interrupted by IncE8 elements (115, 116), causing a reduction of DNA transfer frequency to still detectable but very low levels (115). An undisrupted tra region is present on LT and ST encoding plasmid pUMNK88_Ent (81.5 kbp, R1 subgroup, CP002732), which was found to be present in porcine ETEC strain UMNK88 (117). Besides pUMNK88_Ent, two other virulence-associated plasmids, an IncA/C multidrug resistance plasmid and an IncI1 plasmid were found in this strain. Remarkably, a second F-like plasmid, pUMNK88_K88 (82 kbp, R1 subgroup, CP002730) with a truncated tra region encoding the K88ac fimbrial operon was one of the two other virulence-associated plasmids. Compatibility of these two plasmids is provided by a set of different backbone genes providing replication, partitioning, and stability functions. In addition, it is very likely that both plasmids can be horizontally transferred by conjugation functions provided by pUMNK88_Ent.
Finally, the MOB<sub>F12</sub>A plasmid pAA (113 kbp, R1 subgroup, CP010239) was found in the prototypical enterogaugregative *E. coli* (EAEC) strain 042. Again, several virulence-associated genes such as the aggregative adherence fimbriae (AAF) and the autotransporter Pet toxin-encoding gene found in this diarrhea-causing strain reside within the cargo region (118).

**Cargo Genes of pSLT, the Virulence Plasmid of *S. enterica* serovar Typhimurium**

The *Salmonella* plasmid virulence (*spv*) locus is strongly associated with strains that cause nontyphoid bacteremia, but is not present in typhoid strains. It is located on the large MOB<sub>F12</sub>A virulence plasmid pSLT (94 kbp, NC_003277) encoding virulence factors that are translocated into the host cell by the *Salmonella* PAI-2 T3SS. Two *spv* encoded effectors, SpvB and SpvC, interfere with the actin cytoskeleton and signal transduction pathways of the host cell (119). Also located on pSLT are the plasmid-encoded fimbrial (*pef*) genes that have been shown to play a role in virulence by mediating adhesion to the small intestine in a mouse model (120). If several compatible plasmids reside in one cell, highly dangerous combinations can arise such as a pSLT plasmid and a second antibiotic resistance plasmid. Because of their high variability in the plasmid backbone, both plasmids can belong to the MOB<sub>F12</sub>A group. Such an ensemble has been recently detected in the already-mentioned case of a *S. enterica* serovar Typhimurium isolated from a hospitalized patient. In a ceftriaxone-resistant isolate, both a pSLT MOB<sub>F12</sub>A and a CTX-M carrying pESBL933 MOB<sub>F12</sub>A plasmid were present (87).

### BACTERIAL CONJUGATION MEDIATED BY F AND OTHER MOB<sub>F12</sub>A PLASMIDS

The most intriguing feature of F-like plasmids is the ability to encode functions sufficient for transfer of the plasmid from a donor to a recipient cell. Our knowledge about plasmid DNA transfer by F-like MOB<sub>F12</sub>A plasmids has been largely accumulated from studies of DNA transfer genes from the “classical” plasmids such as F, R1, and R100 and from studies on the large virulence plasmid from *S. enterica* serovar Typhimurium, pSLT.

#### Regulation of DNA Transfer Genes: ON and OFF Switches

Probably one of the most striking features of F-like plasmids (and also other DNA transfer systems of CPs and ICEs) is that, in a plasmid-carrying population, not all cells express DNA transfer genes, but only a fraction of the population develops transfer competence (10–12). At first sight, repression of DNA transfer genes may seem counterintuitive; however, due to several disadvantages for cells that express DNA transfer genes (cellular stress, high metabolic burden, production of surface appendages, potential attack by bacteriophages), this provides a way to minimize the cost (for the plasmid-carrying bacteria) and maximize success (for the plasmid). As a matter of fact, plasmid loss of F-like plasmids after cell division is very low because of the presence of plasmid-partitioning, toxin-antitoxin, and maintenance systems. For example, for pSLT the loss rate was determined to be less than 1 per 10<sup>6</sup> cells (58), a value that may be generally valid for MOB<sub>F12</sub>A plasmids. It has been mathematically modeled and experimentally shown that, if the plasmid loss rate is very low, persistence of CPs in a bacterial population is provided by any DNA transfer rate that is higher than the plasmid loss rate (121).

A key element for the maintenance of the “OFF” state in F-like plasmids is the FinOP repressor system that controls the plasmid-encoded transcriptional activator of the DNA transfer genes, TraJ. FinP is a small (76 nt) noncoding RNA that is complementary to the 5'-nontranslated part and the first two codons of the TraJ encoding mRNA (122). FinO protein specifically recognizes FinP and the target structure in *traJ* mRNA and promotes RNA-RNA duplex formation and *traJ* mRNA inactivation (123). Like the well-characterized Hfq protein and other FinO-like proteins (124), FinO has RNA chaperone activity and destabilizes internal stem regions of FinP and *traJ* mRNA (125). Escape of the *traJ* mRNA encoding the activator TraJ, from FinOP control, only occurs in a small fraction of cells (1 to 10 out of 1000). The escape frequency of TraJ can be influenced by other factors that resulted in complex regulation models for F-like plasmids published elsewhere (11, 126).

Transcription of DNA transfer genes, most of which are organized in one of the largest bacterial operons known, spanning from *traY* beyond *finO* (approximately 30 kbp of DNA; see Fig. 2), initiates at the *P<sub>γ</sub>* promoter. The *P<sub>γ</sub>* promoter is complex, it is silenced by H-NS, and requires the host-encoded ArcA (SfrA) protein (127–131) as well as the plasmid-encoded TraJ protein for activation (15, 127, 128, 132). It has been proposed that phosphorylated ArcA acts as a desilencing factor that remodels the DNA structure, making the *P<sub>γ</sub>* promoter accessible for TraJ.
and RNA polymerase (RNAP) \cite{15}. The ArcA binding site in the P_\gamma promoter region is conserved and positioned at −68 relative to the transcription start site \cite{15, 127}. ArcA could play an essential role in coupling transcription of DNA transfer genes to the energy state of the cell and to microaerobic conditions present in the gut \cite{133}.

The exact mechanism of how the 25-kDa TraJ protein that possesses a PAS domain with a known 3D structure \cite{134} activates transcription from P_\gamma is not known. It has not been possible to determine the structure of the complete purified protein, neither in a free nor DNA-bound form. Unpublished results from our laboratory indicate that dimeric TraJ forms a tight complex with RNA polymerase (core and holoenzyme with σ70) which then could bind to the P_\gamma promoter and initiate transcription. TraJ contains a second domain that represents the DNA-binding part of the protein as suggested by the presence of a helix-turn-helix DNA-binding motif. It was also shown that this domain, located in the C-terminal half of TraJ, confers DNA-binding specificity \cite{134}. Both the PAS domain and the DNA-binding domain are essential for the function of TraJ \cite{135}, which also served to define subgroups within the group of MOBF12A plasmids (see Table 1). Importantly, the sequence differences between TraJ proteins result in plasmid subgroup-specific activation of the P_\gamma promoter which differs in the proposed sequence (centered at −44 relative to the transcription start site) that is likely contacted by TraJ \cite{15, 128, 134}. Therefore, in situations where two compatible F-like plasmids reside in the same cell, TraJ escape from FinOP in one plasmid would result in transfer gene activation and subsequent DNA transfer of only one of the two plasmids. In addition, the DNA of the second plasmid will not be transferred by the T4SS from the first plasmid because of the plasmid-specific activity of the relaxosomal protein TraM (which acts as a DNA substrate selector). In accordance to its function, TraM expression is coupled to the expression of tra operon genes \cite{136}.

Transcription of the whole tra operon depends on the chromosomally encoded RfaH (SrfB) antitermination factor \cite{130, 131} that was later found to be required for transcription of long operons in \textit{E. coli}. RfaH binds to the operon polarity suppressor (ops) DNA element and stably associates with RNAP \cite{137}. The ops sequence element (GGGCGGTAGCGT) is present in all MOBF12A plasmids, in an intergenic region following traV (virB7). RfaH strongly inhibits Rho-dependent termination via antipaus ing modification of RNAP and exclusion of the Rho cofactor NusG \cite{138}. In addition, RfaH-CTD recruits ribosomal protein S10 and translation of RfaH-controlled operons is strongly enhanced \cite{139}. Antitermination activity of RfaH-modified RNAP is presumably necessary to overcome an efficient transcription terminator following P_\gamma independently transcribed traS and traT surface exclusion/virulence genes \cite{131, 140}. In the case of plasmid R1 and R100, it has been conclusively shown that transcription initiated at the P_\gamma promoter proceeds well beyond the finO gene and includes the rmoA (hha) gene \cite{141}. Hha is a H-NS-like nucleoid-associated protein (NAP) that can modulate gene expression in enterobacteria \cite{142} together with or independently of H-NS \cite{143}. Like Hha, another protein, TraR, which is a RNAP modulator is encoded on the tra operon and is therefore coregulated with DNA transfer genes. The traR gene is located immediately after the ops site in the tra operon of F and other F-like plasmids (but not in all). Like chromosomally encoded DksA, TraR is a secondary channel interactor that modulates transcription but exerts its function independently of the alarmone ppGpp \cite{144, 145}. Based on the finding that σE-dependent genes are activated by TraR, it has been proposed that TraR is activated during conjugation to assist cells in dealing with the upcoming periplasmic stress associated with pilus formation by redirecting the host transcriptional machinery toward σE-dependent transcription \cite{146}. TraR and Hha together could therefore both modulate host gene expression as well as represent elements of the negative feedback loop that switches tra gene expression to the OFF state once DNA transfer competence is reached \cite{10}. Other elements of the negative feedback loop are TraY and FinO proteins as well as the activation of the CpxAR pathway \cite{147}.

### The F-Type T4S Machine and the F-Sex Pilus

Once a cell has reached its transfer-competent state, T4S proteins encoded within the tra operon have been assembled into a large DNA and protein secretion machine spanning from the cytoplasm to the cell surface \cite{61, 148, 149}. The largest part of this dedicated secretion machine has been structurally characterized for the P-type T4SS (from conjugative plasmids pKM101 and R388) in Gabriel Waksman’s laboratory and revealed the outer membrane located core complex \cite{150, 151}, the membrane spanning T4S complex \cite{152}, and the T4S complex together with the inner membrane localized coupling...
protein (153). Based on functional, protein localization and protein-protein interaction studies (154, 155), as well as existing similarities of F-type Tra proteins with the prototypic P-type T4SS encoded on the Ti plasmid of A. tumefaciens, it is assumed that a similar complex is formed in F-type T4SS (62). This machine (schematically depicted in Fig. 4) is capable of assembling the F-pilus (or sex-pilus) and also serves as a conduit for ssDNA. The thin and flexible F-pilus which is extended by polymerization of 7-kDa pilin subunits (up to a length of 20 μm) scans the environment for contacting recipient cells that are then brought to the donor cell by pilus retraction (156, 157). The pilus is also a docking site for bacteriophages that attach to the tip of F-pili (such as bacteriophage M13, a filamentous ssDNA virus) or laterally such as R17 (an icosahedral ssRNA phage) (158). R17 phage infection was, in addition, shown to be dependent on the coupling protein TraD and relaxosomal components, indicating that ssRNA attached to its pilot protein enters transfer competent cells in a reverse route of ssDNA transfer (159). F-pili that constantly undergo cycles of elongation and retraction without any obvious signal (156, 157) have been characterized in atomic detail by cryo-electron microscopy. The structures of F and pED208 (a MOBp127C plasmid) encoded pili revealed that their 3D structures are virtually identical and, remarkably, are assemblies of stoichiometric protein-phospholipid units (160). F-pili are five-start helical filaments that are 8.7 nm in diameter and have an internal lumen of 2.8 nm. For each pilin in the filament, there is a phosphatidylglycerol (PG) moiety with the polar head group oriented to the lumen, making F and pED208 pili unique polymers of protein and PG (160). Inclusion of PG was shown to have a dramatic impact on the electrostatic potential of the pilus lumen, making it moderately electronegative, thereby facilitating transport for negatively charged ssDNA substrate (160). The pilus structure reported by Costa and colleagues from the Waksman laboratory also provides a

Figure 4 A simplified model of a F-type T4S machine with an attached F-pilus is depicted. The overall structure, shape, and dimensions are drawn according to a published P-type T4S structure (152). Protein components as determined for the P-type T4SS are indicated and labeled according to the A. tumefaciens VirB protein nomenclature. Positions of the indicated F-type T4S proteins are inferred from sequence similarity and experimental data (see text). The attached F-pilus is drawn according to a recently published high-resolution cryo-EM structure (160). The F-pilus has a diameter of 8.7 nm and an inner lumen of 2.8 nm. For each pilin, there is a phosphatidylglycerol (PG) molecule in the polymeric pilus filament. Whereas the pilus could be assembled and disassembled through the periplasm as indicated by two black arrows, TraI and the covalently attached ssDNA are transported via the coupling protein, TraD (pink arrow). OMC: outer membrane complex; IMC: inner membrane complex.
structural basis for explaining earlier observations such as the effects of F-pilin mutations (161). Mutations that affected pilus biogenesis locate to protein-protein interfaces, whereas mutations that affected conjugation and phage attachment locate to either the lumen or the periphery of the pilus. Newly identified were mutations that affect pilin-PG interactions (160). Since there are five interwound helical filaments with 12.8 pilins per turn with a height of approximately 20 nm, a simple calculation yields that 3200 pilin and PG subunits are required to form a sex-pilus of 1 μm in length. Pilus extension has been determined to occur at a rate of 40 nm/s, whereas retraction proceeds at 16 nm/s (156). Both processes depend on a functional T4S apparatus which in F is more complex than in P-type T4SS. Many of the proteins additionally expressed from the tra-operon of F (shown in Fig. 2) are involved in pilus formation, extension, retraction, and pilus length control (158). Specific for the F-type T4SS are also TraN and TraG (C-terminal part) which are responsible for stabilization of mating pairs after pilus retraction (13). The N terminus of TraG is similar to VirB6, whereas the large, approximately 600 residue C-terminal domain is thought to functionally substitute for VirB8, which is not present as an orthologue in F-like plasmids (62, 158). The large C-terminal part of TraG is also involved in entry exclusion (eex) via a specific contact to cognate TraS protein (an inner membrane protein) which is expressed in all cells harboring a given F-like plasmid. This contact blocks DNA transfer between cells harboring the same F-like plasmid (162, 163) and therefore unproductive self-transfer, via an unknown mechanism. TraT is an outer membrane lipoprotein that serves as a surface exclusion (sfx) protein preventing mating pair formation between donors harboring the same F-like plasmid but has also been implicated in serum resistance, thereby providing another F plasmid-encoded virulence trait (13, 104). DNA transport from F plasmid-carrying donors to recipient cells is initiated after the formation of a stable cell-to-cell contact (158, 164) although researchers have also reported that ssDNA can be transported between spatially separated cells (165, 166). For an extensive discussion of mating pair stabilization, entry, and surface exclusion functions provided by F plasmid-encoded transfer proteins, the reader is referred to excellent reviews discussing these subjects (158, 167).

**Coupling DNA Replication to DNA Transfer**

After formation of a stable cell-to-cell contact, ssDNA and the TraI protein covalently attached to its 5′ end are transported from the transfer-competent donor to the recipient cell via the T4S apparatus. The protein centrally involved in nucleoprotein transport is TraD. Proteins similar to TraD (VirD4 in the *A. tumefaciens* system) are found in all conjugation systems and are ATPases and also known as coupling proteins that direct the protein/DNA substrate to the T4S machine (168). The orthologue of TraD from the IncW plasmid R388, TrwB, as was recently determined, is tightly associated with the cytoplasmic face of the T4S complex (reference 153; Fig. 4). TraD protein from F-like plasmids has early been shown to be required for DNA transfer at a cell-to-cell contact stage but not for pilus synthesis and assembly (169, 170). It was later shown that TraD can specifically interact with a tetrameric sequence-specific DNA-binding protein, TraM, thereby bringing the *oriT* of the plasmid to be transported in physical contact with the T4S machine (126, 171–173). Through its sequence-specific binding to multiple sites in the *oriT* region (173–176), and its interaction with TraD, TraM serves as a substrate selector protein in F-like conjugation systems (172). Since TraM expression is coupled to the expression of the *tra* operon genes (136), this DNA-binding specificity (126) ensures that only the cognate plasmid DNA will be attached to the T4S but not a MOB<sub>fr</sub>-A plasmid from another subgroup which can be present in the same cell.

The relaxase that initiates DNA replication and transfer of ssDNA is the 192-kDa TraI protein. It has a dual role in nicking the DNA strand to be transported at *oriT* and further acts as a helicase separating DNA strands (126). It is a central component of the relaxosome, a nucleoprotein complex consisting of plasmid-encoded Traf, Tram, and TraY proteins as well as host-encoded IHF (126). Traf activity was found to be stimulated by TraD and TraM proteins (177–179), suggesting a close interaction between relaxosomal proteins and the T4S apparatus. TraM encoded by F-like plasmids consists of four domains: (i) a transesterase domain for nicking the DNA substrate strand at *oriT* and covalent attachment to its 5′ end (180); (ii) a vestigial helicase domain that binds to ssDNA (181); (iii) an active 5′- to 3′-helicase domain; and (iv) a C-terminal domain that functions as a platform for interaction with relaxosomal components such as TraM (179, 182). It has also been conclusively shown that Traf is transported via the T4S apparatus into the recipient in a TraD-dependent manner where it serves to recircularize the linear ssDNA (183, 184). In a recent work performed in the Waksman laboratory, Traf from
plasmid R1 bound to oriT DNA could be purified to homogeneity, and a high-resolution, cryo-EM-derived structure revealed the closed conformation of TraI on ssDNA. In that conformation, the two helicase domains entirely enclose the ssDNA, a conformer in which TraI functions as an active helicase (185). A second TraI molecule was also shown to load onto oriT DNA, an observation confirming earlier research findings (181). In that conformation, TraI acts as a transesterase cleaving ssDNA at oriT, with the clamp domains in both the vestigial and the active helicase in the open conformation. In the model presented by Ilangovan et al. (185) this TraI conformer would be transported into the recipient with the 5′ end of the ssDNA covalently attached to a tyrosine residue in the relaxase domain. In the recipient, TraI would switch to the helicase mode helping to transport ssDNA through the T4S apparatus (185). The free 3′-OH end of the ssDNA is generated in the donor cell by a second cleavage by TraI at oriT (184) and therefore could be used in the recipient to reconstitute a circular ssDNA molecule by a nucleophilic attack on the TraI-DNA adduct. Establishment of the plasmid in the recipient, as already outlined above, is achieved with the help of functions encoded on the leading region. Complementary DNA strand synthesis in the donor via DNA Polymerase III can initiate at the 3′ end generated by the first TraI-catalyzed transesterification reaction would continue according to a rolling circle replication (RCR) mode known from ssDNA bacteriophages and plasmids such as pT181 or pMV158 (186). RCR replication involves a dimeric Rep protein that nick the DNA (at a site called nic) and remains covalently bound to its 5′ end until the nic site is encountered again where the newly replicated DNA strand is the substrate for a second transesterification reaction executed by the Rep protein, resulting in the replicated double-stranded plasmid and a single-stranded DNA intermediate that then can be converted to dsDNA by the host replication machinery (186). Thus, in bacterial conjugation mediated by F-like plasmids, monomeric TraI can be regarded as a Rep protein with additional helicase and DNA transport functions. TraI protein helicase activity is remarkably processive unwinding dsDNA at a rate of approximately 1,100 bp/s (187), a value that fits well to the observed time it takes to transfer the entire F plasmid, which is about 2 min at a rate of 45 kb/min (158). Therefore, TraI’s helicase activity, which was found to be essential for DNA transfer (188), could provide the driving force for ssDNA translocation across the T4S channel.

**CONCLUSIONS AND FUTURE PERSPECTIVES**

The F plasmid and the here discussed F-like plasmids belonging to the MOBp112A group represent highly efficient vehicles for horizontal and vertical transmission of antibiotic resistance and virulence genes. The backbone features encoded by approximately 60 kbp of DNA allow stable maintenance in bacterial populations via multiple replication and highly sophisticated partitioning systems that are backed by stability-enhancing toxin-antitoxin systems. In addition, DNA transfer genes encoding a T4SS, and leading region genes for establishment of the plasmid in a new host cell, are key backbone features.

Whole genome sequencing has tremendously impacted our view on MOBp112A plasmids, since it is now clear that they represent major players in the dissemination and persistence of virulence and antibiotic resistance genes in various pathovars of *E. coli*, in *S. enterica*, and in *Klebsiella* strains. Multiple antibiotic resistance genes (among them genes encoding extended spectrum beta-lactamases or resistance to colistin) and various virulence genes can be present on a single F-like plasmid, or on two compatible MOBp112A plasmids that can reside in the same cell. Antibiotic resistance and virulence genes are cargo genes that are flanked by IS elements, transposons, and integrons. According to the functions of cargo genes present on F, it can be classified as a virulence plasmid.

The most intriguing part of the genome of F-like plasmids encompasses genes required for bacterial conjugation, the most sophisticated mode of horizontal gene transfer. Regulatory networks operate to allow transfer gene expression only in a minor fraction of cells harboring a MOBF12A plasmid. Transfer-competent cells contact suitable recipients via the F-sex-pilus that has been recently structurally characterized as a polymer of pilin (protein) and PG (phospholipid) subunits (160). Although the structure of a MOBF12A-encoded T4S apparatus is not known, structures obtained from P-type T4SS (152, 153) will eventually lead the way to elucidate the detailed composition and structure of the F-type T4SS. Due to the presence of additional proteins, the F-type T4SS is expected to be more complex. A large step forward in understanding DNA processing and transfer reactions in F-type bacterial conjugation was made by solving the structure of TraI, the relaxase, helicase protein (185).

Despite such outstanding scientific achievements that are based on knowledge collected over more than 70 years,
many questions remain to be answered by future research endeavors. Cryo-electron tomography may be a method to reveal the structure of an intact T4S machine in the cell envelope of a donor cell or in contact with a recipient. This could lead to a better understanding of the role of the pilus during DNA transfer and reveal the architecture of the contact site between donor and recipient cells. Also, there is little knowledge about how exactly the substrate TraI protein with the attached ssDNA is transported via the T4S apparatus. Another open question is how and when this transport is triggered. Regarding the development of transfer competence in single cells, it will be of interest to specify conditions that modify the probability for TraI escape from FinOP control. Furthermore, a complete structure of the PAS sensory domain containing TraI protein is needed to understand its exact role in the stimulation of transcription of the tra operon.

A better understanding of the molecular biology of F-like plasmids will allow researchers to eventually interfere with DNA transfer and the spread and persistence of virulence and resistance genes via the development and clinical application of conjugation inhibitors (189) or a combination of conjugation inhibitors and substances promoting plasmid loss (121). Such next-generation approaches (90) are urgently needed to develop eco-eco drugs (190) as complementary measures to fight antibiotic resistance that has become a global health problem (again).

REFERENCES

Prophages and growth dynamics confound experimental results with resistance determinants.


López-Villarejo J, Lobato-Márquez D, Diaz-Orejas R. 2015. Coupling between the basic replicon and the Kis-Kid maintenance system of plasmid R1: modulation by Kis antitoxin levels and involvement in control of plasmid replication. Toxins (Basel) 7:487–492.


